Springer Proceedings in Mathematics & Statistics

Valery A. Kalyagin
Alexey I. Nikolaev
Panos M. Pardalos
Oleg A. Prokopyev *Editors*

Models, Algorithms, and Technologies for Network Analysis

NET 2016, Nizhny Novgorod, Russia, May 2016

Springer

Springer Proceedings in Mathematics & Statistics

Volume 197

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Valery A. Kalyagin · Alexey I. Nikolaev
Panos M. Pardalos · Oleg A. Prokopyev
Editors

Models, Algorithms, and Technologies for Network Analysis

NET 2016, Nizhny Novgorod, Russia, May 2016

Springer

Editors
Valery A. Kalyagin
National Research University Higher School of Economics
Nizhny Novgorod
Russia

Alexey I. Nikolaev
National Research University Higher School of Economics
Nizhny Novgorod
Russia

Panos M. Pardalos
Department of Industrial and Systems Engineering
University of Florida
Gainesville, FL
USA

Oleg A. Prokopyev
Department of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA
USA

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-56828-7 ISBN 978-3-319-56829-4 (eBook)
DOI 10.1007/978-3-319-56829-4

Library of Congress Control Number: 2017937743

Mathematics Subject Classification (2010): 90-02, 90C31, 90C27, 90C09, 90C10, 90C11, 90C35, 90B06, 90B10, 90B18, 90B40, 68R01

© Springer International Publishing AG 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume is based on the papers presented at the 6th International Conference on Network Analysis held in Nizhny Novgorod, Russia, May 26–28, 2016. The main focus of the conference and this volume is centered around the development of new computationally efficient algorithms as well as underlying analysis and optimization of graph (network) structures induced either by natural or by artificial complex networks. Various applications to social networks, power transmission grids, stock market networks, and human brain networks are also considered. The previous books based on the papers presented at the 1st–5th Conferences International Conference on Network Analysis can be found in [1–5]. The current volume consists of three major parts, namely optimization approaches, network models, and related applications, which we briefly overview next.

Part I of this book is focused on optimization problems in networks. In Chapter "Linear Max-Min Fairness in Multi-commodity Flow Networks," a linear max-min fairness (LMMF) approach using goal programming is proposed. This model can be applied to max-min fairness (MMF) problems in networks with applications to multicommodity flows in networks. The proposed model provides a high flexibility for the decision maker to determine the level of throughput and the fairness of flow in the network.

In Chapter "Heuristic for Maximizing Grouping Efficiency in the Cell Formation Problem," Cell Formation Problem in Group Technology with grouping efficiency as an objective function is considered. A heuristic approach for obtaining high-quality solutions is presented. The computational results show the effectiveness of the approach.

In Chapter "Efficient Methods of Multicriterial Optimization Based on the Intensive Use of Search Information," an efficient approach for solving complex multicriterial optimization problems is developed. In particular, it is based on reducing multicriterial problems to nonlinear programming problems via the minimax convolution of the partial criteria, reducing dimensionality by using Peano evolvents, and applying efficient information-statistical global optimization methods. The results of the computational experiments show that the proposed approach reduces the computational costs of solving multicriterial optimization problems.

In Chapter "Comparison of Two Heuristic Algorithms for a Location and Design Problem," the special case of the location and design problem is considered. A Variable Neighborhood Search algorithm and a Greedy Weight Heuristic are proposed. In particular, new best known solutions have been found by applying the proposed approaches.

In Chapter "A Class of Smooth Modification of Space-Filling Curves for Global Optimization Problems," a class of smooth modification of space-filling curves applied to global optimization problems is presented. These modifications make the approximations of the Peano curves (evolvents) differentiable in all points and save the differentiability of the optimized function. Some results of numerical experiments with the original and modified evolvents for solving global optimization problems are discussed.

In Chapter "Iterative Local Search Heuristic for Truck and Trailer Routing Problem," Site-Dependent Truck and Trailer Routing Problem with Hard and Soft Time Windows and Split Deliveries is considered. A new iterative local search heuristic for solving this problem was developed.

Part II of this book presents several network models. Chapter "Power in Network Structures" considers an application of power indices, which take into account the preferences of agents for coalition formation proposed for an analysis of power distribution in elected bodies to reveal most powerful (central) nodes in networks. These indices take into account the parameters of the nodes in networks, a possibility of group influence from the subset of nodes to single nodes, and intensity of short and long interactions among the nodes.

In Chapter "Do Logarithmic Proximity Measures Outperform Plain Ones in Graph Clustering?," a number of graph kernels and proximity measures as well as the corresponding distances were applied for clustering nodes in random graphs and several well-known datasets. In the experiments, the best results are obtained for the logarithmic Communicability measure. However, some more complicated cases are indicated in which other measures, typically Communicability and plain Walk, can be the winners.

In Chapter "Analysis of Russian Power Transmission Grid Structure: Small World Phenomena Detection," the spatial and topological structure of the Unified National Electricity Grid (UNEG)—Russia's power transmission grid—is analyzed. The research is focused on the applicability of the small-world model to the UNEG network. For this purpose, geo-latticization algorithm has been developed. As a result of applying the new method, a reliable conclusion has been made that the small-world model is applicable to the UNEG.

In Chapter "A New Approach to Network Decomposition Problems," a new approach to network decomposition problems is suggested. The suggested approach is focused on construction of a family of classifications. Based on this family, two numerical indices are introduced and calculated. This approach was applicable to political voting body and stock market.

In Chapter "Homogeneity Hypothesis Testing for Degree Distribution in the Market Graph," the problem of homogeneity hypothesis testing for degree distribution in the market graph is studied. Multiple hypotheses testing procedure is proposed and applied for China and India stock markets. The procedure is constructed using bootstrap method for individual hypotheses and Bonferroni correction for multiple testing.

Chapter "Stability Testing of Stock Returns Connections" considers the testing problem of connection stability which is formulated as homogeneity hypothesis testing of several covariance matrices for multivariate normal distributions of stock returns. New procedure is proposed and applied for stability testing of connections for French and German stock markets.

Part III of this book is focused on applications of network analysis. In Chapter "Network Analysis of International Migration," the network approach to the problem of international migration is employed. The international migration can be represented as a network where the nodes correspond to countries and the edges correspond to migration flows. The main focus of the study is to reveal a set of critical or central elements in the network.

In Chapter "Overlapping Community Detection in Social Networks with Node Attributes by Neighborhood Influence," a fast method for overlapping community detection in social networks with node attributes is presented. The proposed algorithm is based on attribute transfer from neighbor vertices and does not require any knowledge of attributes meaning. Computational results show that the proposed method outperforms other algorithms such as Infomap, modularity maximization, CESNA, BigCLAM, and AGM-fit.

In Chapter "Testing Hypothesis on Degree Distribution in the Market Graph," the problem of testing hypotheses on degree distribution in the market graph is discussed. Research methodology of power law hypothesis testing is presented. This methodology is applied to testing hypotheses on degree distribution in the market graphs for different stock markets.

In Chapter "Application of Network Analysis for FMCG Distribution Channels," the approach for multidimensional analysis of marketing tactics of the companies employing network tools is presented. The research suggests omni-channel distribution tactic of a company as a node in eight-dimensional space. Empirical implication is approved on the sample from 5694 distributors from sixteen fast-moving consumer goods-distributing companies from direct selling industry.

In Chapter "Machine Learning Application to Human Brain Network Studies: A Kernel Approach," a task of predicting normal and pathological phenotypes from macroscale human brain networks is considered. The research focuses on kernel classification methods. It presents the results of performance comparison of the different kernels in tasks of classifying autism spectrum disorder versus typical development and carriers versus non-carriers of an allele associated with an increased risk of Alzheimer's disease.

In Chapter "Co-author Recommender System," a new recommender system for finding possible collaborator with respect to research interests is proposed. The recommendation problem is formulated as a link prediction within the co-authorship network. The network is derived from the bibliographic database and enriched by the information on research papers obtained from Scopus and other publication ranking systems.

Chapter "Network Studies in Russia: From Articles to the Structure of a Research Community" focuses on the structure of a research community of Russian scientists involved in network studies by analysis of articles published in Russian-language journals. It covers the description of method of citation (reference) analysis that is used and the process of data collection from eLibrary.ru resource, as well as presents some brief overview of collected data (based on analysis of 8000 papers).

We would like to take this opportunity to thank all the authors and referees for their efforts. This work is supported by the Laboratory of Algorithms and Technologies for Network Analysis (LATNA) of the National Research University Higher School of Economics.

Nizhny Novgorod, Russia — Valery A. Kalyagin
Nizhny Novgorod, Russia — Alexey I. Nikolaev
Gainesville, FL, USA — Panos M. Pardalos
Pittsburgh, PA, USA — Oleg A. Prokopyev

References

1. Goldengorin, B.I., Kalyagin, V.A., Pardalos, P.M. (eds.): Models, algorithms and technologies for network analysis. In: Proceedings of the First International Conference on Network Analysis. Springer Proceedings in Mathematics and Statistics, vol. 32. Springer, Cham (2013a)
2. Goldengorin, B.I., Kalyagin, V.A., Pardalos, P.M. (eds.): Models, algorithms and technologies for network analysis. In: Proceedings of the Second International Conference on Network Analysis. Springer Proceedings in Mathematics and Statistics, vol. 59. Springer, Cham (2013b)
3. Batsyn, M.V., Kalyagin, V.A., Pardalos, P.M. (eds.): Models, algorithms and technologies for network analysis. In: Proceedings of Third International Conference on Network Analysis. Springer Proceedings in Mathematics and Statistics, vol. 104. Springer, Cham (2014)
4. Kalyagin, V.A., Pardalos, P.M., Rassias, T.M. (eds.): Network Models in Economics and Finance. Springer Optimization and Its Applications, vol. 100. Springer, Cham (2014)
5. Kalyagin, V.A., Koldanov, P.A., Pardalos, P.M. (eds.): Models, algorithms and technologies for network analysis. In: NET 2014, Nizhny Novgorod, Russia, May 2014. Springer Proceedings in Mathematics and Statistics, vol. 156. Springer, Cham (2016)

Contents

Contributors

Fuad Aleskerov National Research University Higher School of Economics (HSE), Moscow, Russia; V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences (ICS RAS), Moscow, Russia

Mikhail Batsyn Laboratory of Algorithms and Technologies for Network Analysis, National Research University Higher School of Economics, Nizhny Novgorod, Russia

Hamoud Bin Obaid University of Oklahoma, Norman, OK, USA; King Saud University, Riyadh, Saudi Arabia

Oleg Bulanov National Research University Higher School of Economics, Moscow, Russia

Ilya Bychkov Laboratory of Algorithms and Technologies for Network Analysis, National Research University Higher School of Economics, Nizhny Novgorod, Russia

Pavel Chebotarev Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia; The Kotel'nikov Institute of Radio-engineering and Electronics (IRE) of Russian Academy of Sciences, Moscow, Russia

Vladislav Chesnokov Bauman Moscow State Technical University, Moscow, Russia

Yulia Dodonova National Research University Higher School of Economics, Moscow, Russia

Victor Gergel Nizhni Novgorod State University, Nizhny Novgorod, Russia

Alexander Gnusarev Sobolev Institute of Mathematics SB RAS, Omsk Branch, Omsk, Russia

Alexey Goryachih Lobachevsky State University of Nizhni Novgorod, Nizhni Novgorod, Russia

Ivan S. Grechikhin National Research University Higher School of Economics, Laboratory of Algorithms and Technologies for Network Analysis, Nizhny Novgorod, Russia

Vladimir Ivashkin Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia

Ilia Karpov International Laboratory for Applied Network Research, National Research University Higher School of Economics, Moscow, Russia

P.A. Koldanov Laboratory of Algorithms and Technologies for Network Analysis, National Research University Higher School of Economics, Nizhny Novgorod, Russia

Nadezda Kolesnik National Research University Higher School of Economics, Moscow, Russia

Evgeny Kozinov Nizhni Novgorod State University, Nizhny Novgorod, Russia

Anvar Kurmukov National Research University Higher School of Economics, Moscow, Russia

Valentina Kuskova National Research University Higher School of Economics, Moscow, Russia

J.D. Larushina National Research University Higher School of Economics, Laboratory of Algorithms and Technologies for Network Analysis, Nizhny Novgorod, Russia

Ilya Makarov National Research University Higher School of Economics, Moscow, Russia

Sergey Makrushin Financial University Under the Government of the Russian Federation, Moscow, Russia

Daria Maltseva International Laboratory for Applied Network Research, National Research University Higher School of Economics, Moscow, Russia

Natalia Meshcheryakova National Research University Higher School of Economics (HSE), Moscow, Russia; V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences (ICS RAS), Moscow, Russia

Panos M. Pardalos Laboratory of Algorithms and Technologies for Network Analysis, National Research University Higher School of Economics, Nizhny Novgorod, Russia; Center for Applied Optimization, University of Florida, Gainesville, USA

Anna Rezyapova National Research University Higher School of Economics (HSE), Moscow, Russia

Alexander Rubchinsky National Research University "Higher School of Economics", Moscow, Russia; National Research Technological University "MISIS", Moscow, Russia

D.P. Semenov Laboratory of Algorithms and Technologies for Network Analysis, National Research University Higher School of Economics, Nizhny Novgorod, Russia

Sergey Shvydun National Research University Higher School of Economics (HSE), Moscow, Russia; V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences (ICS RAS), Moscow, Russia

Theodore B. Trafalis University of Oklahoma, Norman, OK, USA

Olga Tretyak National Research University Higher School of Economics, Moscow, Russia

M.A. Voronina Laboratory of Algorithms and Technologies for Network Analysis, Nizhny Novgorod, National Research University Higher School of Economics, Moscow, Russia

Leonid E. Zhukov National Research University Higher School of Economics, Moscow, Russia

Part I
Optimization

Linear Max-Min Fairness in Multi-commodity Flow Networks

Hamoud Bin Obaid and Theodore B. Trafalis

Abstract In this paper, a linear max-min fairness (LMMF) approach using goal programming is proposed. The linear model in this approach is a bi-objective model where the flow is maximized in one objective, and the fairness in flow is maximized for the other objective. This model can be applied to max- min fairness (MMF) problems in networks with applications to multi-commodity flows in networks. The proposed model provides high flexibility for the decision maker to determine the level of throughput and the fairness of flow in the network. The goal of this paper is to find a less-complex approach to find MMF flow in multi-commodity networks. An example is presented to illustrate the methodology.

1 Introduction

The basic approach to apply MMF [1] on a problem is to start with the lowest capacity object of a system to fill with the available resource to its maximum capacity, then fill the second lowest object to its capacity continuing with the next lowest object until the available resource ends. MMF insures that the resource is fairly distributed among the available objects starting with the lower capacity ones. MMF is widely applied in traffic engineering and load-balancing problems. Internet protocol (IP) is an extensively studied application using MMF, where the goal is to maximize the throughput and insure fair distribution of resources. A new approach has not been proposed, according to our knowledge, to maximize the throughput of a flow network and maximize its fairness. A bi-objective model is implemented to optimize the two objectives in a flow network. The proposed approach to solve the bi-objective model is to use the ε-constraint method. In the next section, a brief MMF overview from

H. Bin Obaid (✉) · T.B. Trafalis
University of Oklahoma, 202 W. Boyd St., Lab 28, Norman, OK 73019, USA
e-mail: hsbinobaid@ou.edu; hsbinobaid@ksu.edu.sa

T.B. Trafalis
e-mail: ttrafalis@ou.edu

H. Bin Obaid
King Saud University, P.O. BOX 800, Riyadh 11421, Saudi Arabia

© Springer International Publishing AG 2017
V.A. Kalyagin et al. (eds.), *Models, Algorithms, and Technologies for Network Analysis*, Springer Proceedings in Mathematics & Statistics 197,
DOI 10.1007/978-3-319-56829-4_1

related work is presented. In Sect. 2 the solution approach is discussed. In Sect. 3 a toy example is presented and the results of the model are discussed with a conclusion in Sect. 3.

1.1 MMF in Networks

The early development of the MMF approach in networks appeared in 1974 and is based on lexicographic maximization [2]. MMF is commonly applied to multi-commodity flow networks and is used for internet protocol (IP) throughput optimization. There has been a number of research papers in this field, such as [3–6]. Although the flow in IP network is unsplittable, Multi-Protocol Label Switching (MPLS) technology allows the flow in IP to be splitted among different paths. This technology enables us to design a more relaxed model by not forcing the flow to be routed through one path, which reduces the complexity of solving this problem. The most common approach to apply MMF on multi-commodity flow networks is to solve an LP model a number of times depending on the size of the network [1, 7, 8]. The constraints are classified into two sets: non-blocking constraints and blocking constraints. The blocking constraints set is initialized to be empty. Then an LP is solved until a blocking constraint is found, then the constraint is moved from the non-blocking constraints set to the blocking set until an empty set of non-blocking constraints is reached. This process is computationally demanding for very large networks due to the approach of identifying the blocking constraints in each iteration. The approach to identify the blocking constraints is by using the strict complementary slackness. The constraint is blocking if it is binding which leads to a zero slack value. A positive value of the corresponding dual variable is used as an indicator of a blocking constraint. However, this condition is unnecessary but insures convergence. It indicates that not all blocking constraints in one iteration can be identified, and we call that degeneracy. This leads to a higher number of iterations to find the MMF flow in the network. In [7], the authors proposed another approach called binary search that reduces the number of iterations to identify the blocking constraints. Another method to find MMF in [9] is by the polyhedral approach. Geometry is used to find MMF in networks, where the number of commodities represents the number of dimensions of the polyhedron. Changing the flow of one commodity would result in change of flow for other commodities sharing the same capacity of the network forming a polyhedron. The MMF flow point in the polyhedron can be located by maximizing the distance between the point and the zero axes for each commodity. The drawbacks of this approach is that the routing path for each commodity has to be predefined to identify the right-hand side values of the constraints, and the model is solved iteratively which can lead to high computational time. For the polyhedral approach, there is no efficient model that has been developed yet to find the MMF flow on the network.

In this paper, the overall number of iterations is considerably reduced leading to an efficient approach to find MMF flow for larger networks.

2 Solution Methodology

Given the network N, described through the graph G = (V, E) with the set of nodes V and set of directed arcs E. Each arc e ∈ E, where e = (i, j) and i, j ∈ V. s, t ∈ V where s is a supply node and t is a terminal node. S is the supply node set, and T is the terminal node set where s ∈ S and t ∈ to T. The model is composed of hard and soft constraints since a goal programming approach is used. The set of constraints (1) are known to be the flow conservation constraint, where x_{ij}^k is the commodity flow k from node i to node j. If the incoming flow is greater than the outgoing flow at node i, then node i is a source node. If the outgoing flow is greater than the incoming flow at node i, then node i is a source node. If the incoming and outgoing flows are equal at node i, the node i is a transshipment node.

$$\sum_{j \in N} x_{ij}^k - x_{j,i}^k = \begin{cases} f_i^k & \text{if } i = s, \\ -f_i^k & \text{if } i = t, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

With the additional index k, each commodity flows is in a distinct network. However, the set of capacity constraints (2) link all the commodity flows to one network to share the same capacity resource. It insures that all the commodity flows pass through the arc e = (i, j) but do not exceed the capacity of the arc.

$$\sum_{k \in K} x_{ij}^k \leq C_{ij} \qquad \forall (i, j) \in E, k \in K \tag{2}$$

The set of soft constraints (3) provide a way to reduce the difference among the commodity flows resulting in fair allocation of commodity flows. d^{kl} is the deviational variable of commodity flow k compared with commodity flow l. These deviational variables make up the difference between two commodity flows.

$$\sum_{i \in S} f_i^k - \sum_{j \in S} f_j^l + d^{kl} - d^{lk} = 0 \qquad \forall k \in K \tag{3}$$

If the deviational variables are minimized to 0, all the commodity flows are equal and said to be fairly distributed. The sum of commodity flows $\sum_{i \in S} f_i^k$ appears in the case of having multiple source nodes. The set of constraints (4) are the nonnegativity constraints.

$$x_{ij}^k, f_i^k, d^{kl} \geq 0 \qquad \forall (i, j) \in E, i \in V, k \in K \tag{4}$$

The first objective (5) is to maximize the overall flow to utilize the available capacity resources. Maximizing the first objective does not lead to a fair distribution of flow. However, maximizing the flow as a first step is useful to adjust the value of ε when the first objective is set as a constraint. When the first objective is set as a constraint using the ε-constraint method, the sum of deviations is minimized in the second objective (6).

$$Max\ Z_1 = \sum_{i \in N, k \in K} f_i^k \tag{5}$$

$$Min\ Z_2 = \sum_{k,l \in K} d^{kl} \tag{6}$$

The decision maker then decides what level of fairness is desired considering the tradeoff between fairness and total flow.

2.1 MMF in Networks

Let us consider the simple example of a network (N1) in [1] where we have three routers A, B, and C and three links AB, BC, and AC with capacities 2 MB/s, 3 MB/s, and 1 MB/s respectively. If the flow is maximized, the resulting flow is illustrated in Fig. 1a. The total throughput is 6.

The value of the maximum flow is determined. The next step is to set the first objective (5) as a constraint using the ε-constraint method.

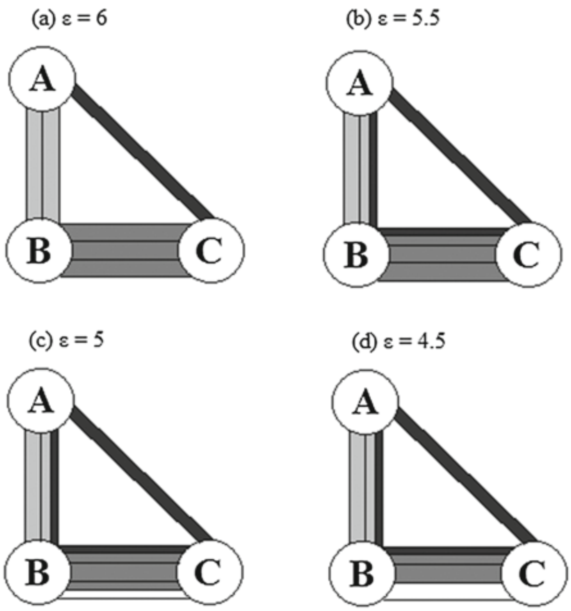

Fig. 1 The result when maximizing the overall throughput. When maximizing the overall flow, the deviation is maximum resulting in unfair distribututiom of commodity flows as seen in (**a**). If the overall flow is reduced to 5.5, the deviation is reduced resulting in MMF solution in (**b**). If the flow is reduced below 5.5, the network is not utilized as seen in (**c**) and (**d**)

Table 1 Summary of the results using different values of ε

ε (Total flow)	6		5.5		5		4.5	
Connection	Path	Flow	Path	Flow	Path	Flow	Path	Flow
A $\longrightarrow$ B	A-B	2	A-B	1.5	A-B	1.5	A-B	1.5
A $\longrightarrow$ C	A-C	1	A-B-C+A-C	1.5	A-B-C+A-C	1.5	A-B-C+A-C	1.5
B $\longrightarrow$ C	B-C	3	B-C	2.5	B-C	2	B-C	1.5
Min deviation	4		2		1		0	

$$Min\ Z = \sum_{k,l \in K} d^{kl} \tag{7}$$

s.t.

$$\sum_{i \in N, k \in K} f_i^k \geq \varepsilon \tag{8}$$

$$\sum_{j \in N} x_{ij}^k - x_{j,i}^k = \begin{cases} f_i^k & \text{if } i = s, \\ -f_i^k & \text{if } i = t, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

$$\sum_{k \in K} x_{ij}^k \leq C_{ij} \qquad \forall (i, j) \in E, k \in K \tag{10}$$

$$\sum_{i \in S} f_i^k - \sum_{j \in S} f_j^l + d^{kl} - d^{lk} = 0 \qquad \forall k \in K \tag{11}$$

$$x_{ij}^k, f_i^k, d^{kl} \geq 0 \qquad \forall (i, j) \in E, i \in V, k \in K \tag{12}$$

When the ε value is set to 6, the maximum flow in this case, the minimum sum of deviations obtained is 4. If ε is reduced to 5.5, the minimum sum of deviations resulted is 2 as shown in Fig. 1b. However, the minimum sum of deviations becomes 1 if we set ε to 5, but the set of capacity constraints (10) are no longer binding, which indicates that the capacity resource is not fully utilized. Table 1 summarizes the results of the tested example. The deviation can be minimized to zero if the ε value is set to 4.5, resulting in equal flows for all commodities with some nonbinding capacity constraints as seen in Fig. 1c, d.

We can observe the results in Table 1, where the sum of deviational variables decreases as the ε value decreases. Reducing the flow reduces the congestion resulted from the competing commodities trying to reach to their destinations resulting in giving space for sharing as seen in the example above. In Fig. 2, there are infinite nondominated solutions creating a Pareto front creating a convex objective space. Additionally, this proposed approach requires less computational time and provides high flexibility in terms of decision-making. This problem is solved in a polynomial time [10].

If the value of ε is continuous, the resulting Pareto front can be observed in Fig. 3. It can be noticed that the slope is different in the intervals [4.5, 5.5) and (5.5, 6]. The most attractive value of ε is 5.5, which is the value we would obtain if the MMF algorithm was used [1]. The reason that 5.5 is the most attractive value is because it gains the most of the two competing objectives. Rationally, if the ε value is decreased below the value 5.5, the gain in fairness is not substantial compared to the gain acquired by creating space for sharing.

The proposed algorithm is applied on a random network with 26 nodes, 83 links, and 109 commodities. The results can be observed in Fig. 4 showing that the level of fairness decreases to 0.1 of the maximum deviation when we reduce the overall flow to 0.6. In Fig. 4, the MMF solution exists with solving the model only several times

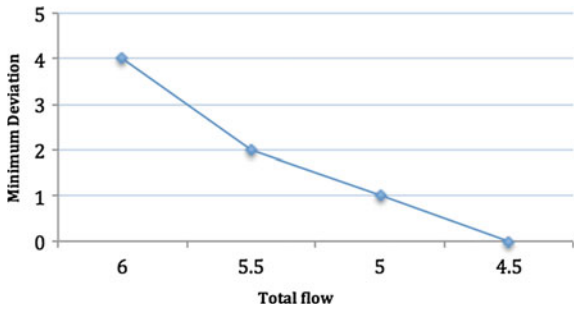

Fig. 2 Pareto front for the 4 values of ε

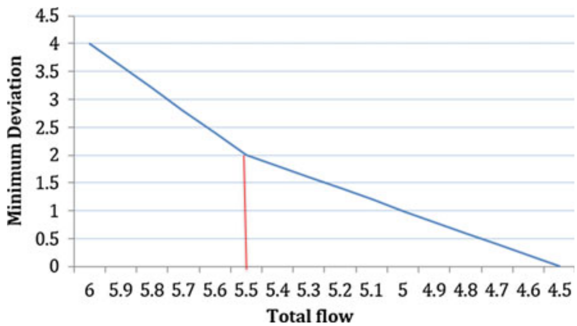

Fig. 3 Pareto front for continuous value of ε

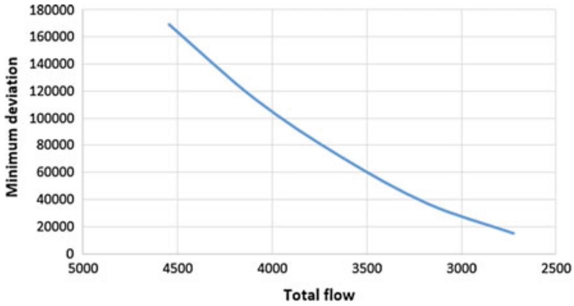

Fig. 4 Pareto front for continuous value of ε

regardless of the size the network. However, in the future version of this model we will propose a robust approach to decide which value of ε gives the MMF solution for any network size.

3 Conclusion

In this paper, a linear max-min fairness (LMMF) approach using goal programming is proposed. The resulting linear model was described as a bi-objective model where we maximize the flow as the first objective, and the fairness in flow as the second objective. A small example from communication networks was used to illustrate the idea. We can conclude that every network structure affects the behavior of the Pareto front of the two objectives. In addition, other performance measures will be included to improve the accuracy when selecting the value of ε. Applying this model to large-scale networks will be the objective of future research. Robust techniques to select the ε value will be also explored. Selection of ε is critical in utilizing the resource capacity of the network. Moreover, another objective is to apply the model on a variety of application such as is congestion control, traffic engineering, and IP networks.

References

1. Nace, D., Pi, M.: Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial. IEEE Commun. Surv. Tutor. **10**(4), 5–17 (2008)
2. Megiddo, N.: Optimal flows in networks with multiple sources and sinks. Math. Program. **7**, 97–107 (1974)
3. Nace, D., Doan, N., Gourdin, E., Liau, B.: Computing optimal max-min fair resource allocation for elastic flows, EEE Trans. Netw. **14**, 1272–1281 (2006)
4. Bertsekas, D., Hall, P.: Data Networks. Prentice-Hall, Engelwood Cliffs, N.J. (1992)

5. Chen, S., Nahrstedt, K.: Maxmin fair routing in connection-oriented networks 1 introduction. In: Euro-Parallel and Distributed Systems Conference, pp. 163–168, Vienna Austria, (July 1998)
6. Nilsson, P., Pioro, M., Dziong, Z.: Link protection within an existing backbone network. In: Proceedings of International Network Optimization Conference (2003)
7. Danna, E., Mandal, S., Singh, A.: A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering. In: Proceedings IEEE INFOCOM, pp. 846–854 (2012)
8. Mollah, Md.A., YuanFast, X., Pakin, S., Lang, M.: Calculation of max-min fair rates for multi-commodity flows in fat-tree networks. In: IEEE International Conference on Cluster Computing, pp. 351–360 (2015)
9. Retvari, G., Biro, J.J., Cinkler, T.: Fairness in capacitated networks: a polyhedral approach. In: IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications, pp. 1604–1612 (2007)
10. Cohen, E., Megiddo, N.: Algorithms and complexity analysis for some flow problems. Algorithmica **11**(3), 320–340 (1994). (March 1994)

Heuristic for Maximizing Grouping Efficiency in the Cell Formation Problem

Ilya Bychkov, Mikhail Batsyn and Panos M. Pardalos

Abstract In our paper, we consider the Cell Formation Problem in Group Technology with grouping efficiency as an objective function. We present a heuristic approach for obtaining high-quality solutions of the CFP. The suggested heuristic applies an improvement procedure to obtain solutions with high grouping efficiency. This procedure is repeated many times for randomly generated cell configurations. Our computational experiments are performed for popular benchmark instances taken from the literature with sizes from 10×20 to 50×150. Better solutions unknown before are found for 23 instances of the 24 considered. The preliminary results for this paper are available in Bychkov et al. (Models, algorithms, and technologies for network analysis, Springer, NY, vol. 59, pp. 43–69, 2013, [7]).

1 Introduction

Flanders [15] was the first who formulated the main ideas of the group technology. The notion of the Group Technology was introduced in Russia by [30], though his work was translated to English only in 1966 [31]. One of the main problems stated by the Group Technology is the optimal formation of manufacturing cells, i.e., grouping of machines and parts into cells such that for every machine in a cell the number of the parts from this cell processed by this machine is maximized and the number of the parts from other cells processed by this machine is minimized. In other words,

I. Bychkov (✉) · M. Batsyn · P.M. Pardalos
Laboratory of Algorithms and Technologies for Network Analysis,
National Research University Higher School of Economics,
136 Rodionova Street, Nizhny Novgorod 603093, Russia
e-mail: ibychkov@hse.ru

M. Batsyn
e-mail: mbatsyn@hse.ru

P.M. Pardalos
Center for Applied Optimization, University of Florida, 401 Weil Hall,
P.O. Box 116595, Gainesville 32611-6595, USA
e-mail: pardalos@ufl.edu

© Springer International Publishing AG 2017
V.A. Kalyagin et al. (eds.), *Models, Algorithms, and Technologies for Network Analysis*, Springer Proceedings in Mathematics & Statistics 197,
DOI 10.1007/978-3-319-56829-4_2

the intra-cell loading of machines is maximized and simultaneously the inter-cell movement of parts is minimized. This problem is called the Cell Formation Problem (CFP). Burbidge [5] suggested his Product Flow Analysis (PFA) approach for the CFP, and later popularized the Group Technology and the CFP in his book [6].

The CFP is NP-hard since it can be reduced to the clustering problem [16]. That is why there is a great number of heuristic approaches for solving CFP and almost no exact ones. The first algorithms for solving the CFP were different clustering techniques. Array-based clustering methods find rows and columns permutations of the machine-part matrix in order to form a block-diagonal structure. These methods include: Bond Energy Algorithm (BEA) of [29], Rank Order Clustering (ROC) algorithm by [20], its improved version ROC2 by [21], Direct Clustering Algorithm (DCA) of [12], Modified Rank Order Clustering (MODROC) algorithm by [9], the Close Neighbor Algorithm (CAN) by [4]. Hierarchical clustering methods at first form several big cells, then divide each cell into smaller ones and so on gradually improving the value of the objective function. The most well-known methods are Single Linkage [28], Average Linkage [39], and Complete Linkage [32] algorithms. Nonhierarchical clustering methods are iterative approaches which start from some initial partition and improve it iteratively. The two most successful are GRAFICS algorithm by [41] and ZODIAC algorithm by [10]. A number of works considered the CFP as a graph partitioning problem, where machines are vertices of a graph. [37] used clique partitioning of the machines graph. Askin and Chiu [2] implemented a heuristic partitioning algorithm to solve CFP. Ng [35, 36] suggested an algorithm based on the minimum spanning tree problem. Mathematical programming approaches are also very popular for the CFP. Since the objective function of the CFP is rather complicated from the mathematical programming point of view most of the researchers use some approximation model which is then solved exactly for small instances and heuristically for large. [25] formulated CFP via p-median model and solved several small size CFP instances, [40] used Generalized Assignment Problem as an approximation model, [44] proposed a simplified p-median model for solving large CFP instances, [22] applied minimum k-cut problem to the CFP, [17] used p-median approximation model and solved it exactly by means of their pseudo-boolean approach including large CFP instances up to 50×150 instance. A number of meta-heuristics have been applied recently to the CFP. Most of these approaches can be related to genetic, simulated annealing, Tabu search, and neural networks algorithms. Among them are works such as: [18, 26, 27, 45–47].

Our heuristic algorithm is based on sequential improvements of the solution. We modify the cell configuration by enlarging one cell and reducing another. The basic procedure of the algorithm has the following steps:

1. Generate a random cell configuration.
2. Improve the initial solution moving one row or column from one cell to another until the grouping efficiency is increasing.
3. Repeat steps 1–2 a predefined number of times (we use 2000 times for computational experiments in this paper).

The paper is organized as follows. In the next section, we provide the Cell Formation Problem formulation. In Sect. 3 we present our improvement heuristic that allows us to get good solutions by iterative modifications of cells which lead to increasing of the objective function. In Sect. 4 we report our computational results and Sect. 5 concludes the paper with a short summary.

2 The Cell Formation Problem

The CFP consists in an optimal grouping of the given machines and parts into cells. The input for this problem is given by m machines, p parts, and a rectangular machine-part incidence matrix $A = [a_{ij}]$, where $a_{ij} = 1$ if part j is processed on machine i. The objective is to find an optimal number and configuration of rectangular cells (diagonal blocks in the machine-part matrix) and optimal grouping of rows (machines) and columns (parts) into these cells such that the number of zeros inside the chosen cells (voids) and the number of ones outside these cells (exceptions) are minimized. A concrete combination of rectangular cells in a solution (diagonal blocks in the machine-part matrix) we will call a cells configuration. Since it is usually not possible to minimize these two values simultaneously there have appeared a number of compound criteria trying to join it into one objective function. Some of them are presented below.

For example, we are given the machine-part matrix [43] shown in Table 1. Two different solutions for this CFP are shown in Tables 2 and 3. The left solution is better because it has less voids (3 against 4) and exceptions (4 against 5) than the right one. But one of its cells is a singleton—a cell which has less than two machines or parts.

Table 1 Machine-part 5 × 7 matrix from [43]

	p_1	p_2	p_3	p_4	p_5	p_6	p_7
m_1	1	0	0	0	1	1	1
m_2	0	1	1	1	1	0	0
m_3	0	0	1	1	1	1	0
m_4	1	1	1	1	0	0	0
m_5	0	1	0	1	1	1	0

Table 2 Solution with singletons

	p_7	p_6	p_1	p_5	p_3	p_2	p_4
m_1	1	1	1	1	0	0	0
m_4	0	0	1	0	1	1	1
m_3	0	1	0	1	1	0	1
m_2	0	0	0	1	1	1	1
m_5	0	1	0	1	0	1	1

Table 3 Solution without singletons

	p_7	p_1	p_6	p_5	p_4	p_3	p_2
m_1	1	1	1	1	0	0	0
m_4	0	1	0	0	1	1	1
m_2	0	0	0	1	1	1	1
m_3	0	0	1	1	1	1	0
m_5	0	0	1	1	1	0	1

In some CFP formulations singletons are not allowed, so in this case this solution is not feasible. In this paper, we consider both the cases (with allowed singletons and with not allowed) and when there is a solution with singletons found by the suggested heuristic better than without singletons we present both the solutions.

There are a number of different objective functions used for the CFP. The following four functions are the most widely used:

1. Grouping efficiency suggested by [11]:

$$\eta = q\eta_1 + (1-q)\eta_2, \tag{1}$$

where

$$\eta_1 = \frac{n_1 - n_1^{out}}{n_1 - n_1^{out} + n_0^{in}} = \frac{n_1^{in}}{n^{in}},$$

$$\eta_2 = \frac{mp - n_1 - n_0^{in}}{mp - n_1 - n_0^{in} + n_1^{out}} = \frac{n_0^{out}}{n^{out}},$$

η_1—a ratio showing the intra-cell loading of machines (or the ratio of the number of ones in cells to the total number of elements in cells).
η_2—a ratio inverse to the inter-cell movement of parts (or the ratio of the number of zeroes out of cells to the total number of elements out of cells).
q—a coefficient ($0 \leq q \leq 1$) reflecting the weights of the machine loading and the inter-cell movement in the objective function. It is usually taken equal to $\frac{1}{2}$, which means that it is equally important to maximize the machine loading and minimize the inter-cell movement.
n_1—a number of ones in the machine-part matrix,
n_0—a number of zeroes in the machine-part matrix,
n^{in}—a number of elements inside the cells,
n^{out}—a number of elements outside the cells,

n_1^{in}—a number of ones inside the cells,
n_1^{out}—a number of ones outside the cells,
n_0^{in}—a number of zeroes inside the cells,
n_0^{out}—a number of zeroes outside the cells.

2. Grouping efficacy suggested by [23]:

$$\tau = \frac{n_1 - n_1^{out}}{n_1 + n_0^{in}} = \frac{n_1^{in}}{n_1 + n_0^{in}} \tag{2}$$

3. Group Capability Index (GCI) suggested by [19]:

$$GCI = 1 - \frac{n_1^{out}}{n_1} = \frac{n_1 - n_1^{out}}{n_1} \tag{3}$$

4. Number of exceptions (ones outside cells) and voids (zeroes inside cells):

$$E + V = n_1^{out} + n_0^{in} \tag{4}$$

The values of these objective functions for the solutions in Tables 2 and 3 are shown below.

$$\eta = \frac{1}{2} \cdot \frac{16}{19} + \frac{1}{2} \cdot \frac{12}{16} \approx 79.60\% \qquad \eta = \frac{1}{2} \cdot \frac{15}{19} + \frac{1}{2} \cdot \frac{11}{16} \approx 73.85\%$$

$$\tau = \frac{20-4}{20+3} \approx 69.57\% \qquad \tau = \frac{20-5}{20+4} \approx 62.50\%$$

$$GCI = \frac{20-4}{20} \approx 80.00\% \qquad GCI = \frac{20-5}{20} \approx 75.00\%$$

$$E + V = 4 + 3 = 7 \qquad E + V = 5 + 4 = 9$$

In this paper, we use the grouping efficiency measure and compare our computational results with the results of [17, 47].

The mathematical programming model of the CFP with the grouping efficiency objective function can be described using boolean variables x_{ik} and y_{jk}. Variable x_{ik} takes value 1 if machine i belongs to cell k and takes value 0 otherwise. Similarly variable y_{jk} takes value 1 if part j belongs to cell k and takes value 0 otherwise. Machines index i takes values from 1 to m and parts index j - from 1 to p. Cells index k takes values from 1 to $c = \min(m, p)$ because every cell should contain at least one machine and one part, and so the number of cells cannot be greater than m and p. Note, that if a CFP solution has n cells then for k from $n + 1$ to c all