Proceedings of the 17th International Symposium on OPERATIONAL RESEARCH in Slovenia

SOR '23

Bled, Slovenia

September 20-22, 2023

Edited by:

S. Drobne • L. Zadnik Stirn • M. Kljajić Borštnar • J. Povh • J. Žerovnik

Proceedings of the 17th International Symposium on Operational Research in Slovenia

SOR'23

Bled, SLOVENIA, September 20-22, 2023

Edited by:

S. Drobne, L. Zadnik Stirn, M. Kljajić Borštnar,
J. Povh and J. Žerovnik

Slovenian Society INFORMATIKA (SDI)
Section for Operational Research (SOR)

Proceedings of the 17th International Symposium on Operational Research in Slovenia, SOR'23, Bled, Slovenia, September 20-22, 2023.

Proceedings of the scientific symposium SOR'23 (https://sor.fov.um.si/). All contributions were peer-reviewed in a blinded manner.

Editors: Assoc. Prof. Dr. Samo Drobne, UL FGG, Ljubljana, Slovenia
Prof. Dr. Lidija Zadnik Stirn, UL BF, Ljubljana, Slovenia
Prof. Dr. Mirjana Kljajić Borštnar, UM FOV, Kranj, Slovenia
Prof. Dr. Janez Povh, Rudolfovo Institute, Novo mesto, Slovenia
Prof. Ddr. Janez Žerovnik, UL FS, Ljubljana, Slovenia

Revievers are listed below.

Keywords: operations research, mathematical programming & optimization, computational optimization, multicriteria decision-making, graph theory, game theory, econometrics & statistics, data science & artificial intelligence, logistics & transportation, environment & human resources, production & management, agricultural economics & policy, finance & investments, inventory & location research, Slovenian Society Informatika (SSI), Section for operational research (SOR)

Available at: https://www.drustvo-informatika.si/sekcije-drustva?stran=publikacije-sor

Organiser of SOR'23: Slovenian Society INFORMATIKA – Section for Operational Research, SI-1000 Ljubljana, Litostrojska cesta 54, Slovenia (https://www.drustvo-informatika.si/sekcije-drustva)

Co-organiser of SOR'23: University of Maribor, Faculty of Organizational Sciences, SI-4000 Kranj, Kidričeva cesta 55a, Slovenia (http://www.fov.um.si/)

Co-organiser of SOR'23: University of Ljubljana, Faculty of Mechanical Engineering, SI-1000 Ljubljana, Aškerčeva cesta 6, Slovenia (https://www.fs.uni-lj.si/)

Publisher: Slovenian Society INFORMATIKA – Section for Operational Research, SI-1000 Ljubljana, Litostrojska cesta 54, Slovenia (https://www.drustvo-informatika.si/sekcije-drustva)

Ljubljana, 2023

First electronic version.

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 164343555

ISBN 978-961-6165-61-7 (PDF)

Proceedings of the 17th International Symposium on Operational Research in Slovenia (SOR'23) is cited in: ISI (Index to Scientific & Technical Proceedings on CD-ROM and ISI/ISTP&B online database), Current Mathematical Publications, Mathematical Review, MathSci, Zentralblatt für Mathematic / Mathematics Abstracts, MATH on STN International, CompactMath, INSPEC, Journal of Economic Literature

Technical editor: Samo Drobne

Designed by: Samo Drobne

Contents

APPENDICES

GRAPH-BASED PRIORITIZATION OF RELATED CANCER GENES

Peter Juma Ochieng
Institute of Informatics, University of Szeged,
Árpád tér 2, H-6720 Szeged, Hungary
Bánki Donát Faculty of Mechanical and Safety Engineering,
Óbuda University, 1081 Budapest, Népszínház street 8, Hungary
József Dombi
Institute of Informatics, University of Szeged,
Árpád tér 2, H-6720 Szeged, Hungary
Tibor Kalmár
University of Szeged, Albert Szent-Györgyi Health Centre,
Department of Pediatrics and
Pediatric Health Center, H-6725 Szeged, Hungary
András London
Institute of Informatics, University of Szeged,
Árpád tér 2 H-6720 Szeged, Hungary
Miklós Krész
InnoRenew CoE, Livade 6a, 6310 Izola, Slovenia
Andrej Marušič Institute, University of Primorska,
Muzejski trg 2, 6000 Koper , Slovenia
Department of Applied Informatics, University of Szeged,
Boldogasszony sgt 6, H-6725 Szeged, Hungary

Abstract: This paper introduces Graph-Based Prioritization (GBP), a novel computational approach for gene prioritization in cancer genomics. GBP integrates gene interaction networks and mutation data to identify and rank genes based on their relevance to cancer. By considering gene mutations and the influence from neighboring genes in the network, GBP calculates a mutation score and employs GBP-PR for gene prioritization. Experimental results across six cancer types demonstrate the effectiveness of GBP in identifying known and potential novel cancer genes. Overall, GBP offers a valuable tool for understanding tumor mechanisms and advancing cancer research.

Keywords: Cancer, Gene Prioritization, Ranking, Rating

1 INTRODUCTION

Cancer is a devastating disease caused by genetic mutations in cells, and identifying the key mutations that drive its development is a crucial challenge in cancer research. With the advent of computational methods and the wealth of biological data generated through next-generation sequencing technologies, researchers now have powerful tools to tackle this challenge [1]. By analyzing gene interaction networks and employing statistical and graph theory-based algorithms, these computational methods aim to identify the most significant mutations in cancer genes and shed light on their functional roles in cancer biology. Gene interaction network analysis is a key strategy employed by these computational methods. By studying the interactions between mutated genes and their influence on these networks, researchers gain valuable insights into the functional roles of these genes in cancer[2]. Several methods, including Hierarchical HotNet [3], Dendrix [4], and Multi-Dendrix [5], have been developed to identify driver mutations in gene networks. These methods utilize gene graphs, network diffusion algorithms, and weighted functions to identify relevant gene sets with high mutation frequencies in patients. Statistical methods such as CoMEt[6], MEMo [7], and MEMCover[8] have also emerged as important

tools in the identification of mutually exclusive gene sets. These methods employ statistical analysis and network analysis to identify gene modules based on alteration frequency, biological process, and mutual exclusivity. Furthermore, rating algorithms derived from network analysis and graph theory, such as PageRank, Colley, Massey, and Keener[9], have gained prominence in various domains. These algorithms offer valuable tools for analyzing networks and ranking entities, providing insights into the structure and dynamics of complex systems. In the context of cancer research, these algorithms can be applied to prioritize cancer genes based on their mutation data and their influence on gene interaction networks.

This study proposes a graph-based approach that integrates cancer mutation data and gene interaction networks to prioritize related cancer genes. We apply a novel heuristic approaches to build a mutation matrix, calculate mutation scores, create a consensus gene interaction network, generate a gene spreading strength network, extract mutation influence from neighbors, and prioritize genes using a dynamic PageRank algorithm.

2 METHODOLOGY

2.1 Extraction of Gene Spreading Strength (GSS)

Our approach uses multiple gene networks as input, generating an undirected and weighted network (UWN) that preserves original interactions. We then consider gene direct and indirect neighbors to calculate mutations spread from one gene to another in a network by $ss(g_i, g_j) = (1 + r_i \times r_j^{out}) \times p_{ij}$, where $r_j^{out} = \sum_g \in (^{N(g_j)}/_{N(g_i)})p(g_i, g_j)$,$r_i$ is the sum of the edge weights of g_i and r_j^{out} is the sum of the edge weights of g_j that are not edges of g_i.

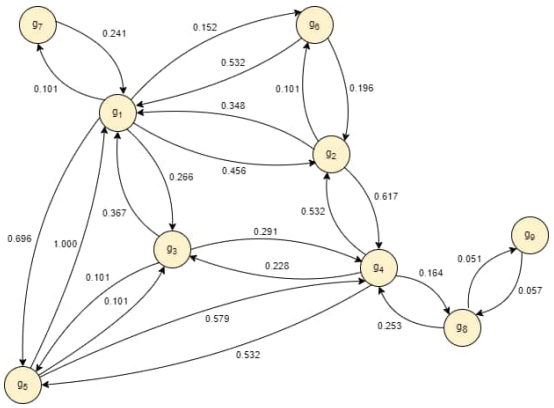

Figure 1: A simple hypothetical graph of extracted GSS

2.2 Extraction of Mutation Neighbors Influence

The spreading strength of genes measures their susceptibility to neighboring mutations and their ability to impact nearby genes. Influence is determined by the function $r(g_i)$ given by $r(g_i) = \sum_{g_k \in N(g_i)} nwm_f(g_k) \times ss(g_k, g_j)$, where $N(g_i)$ are direct neighbors of g_i on GSS. Hence the final gene mutation enrichment score is calculated by $ms(g_i) = wm_f(g_i) + r(g_i)$, where $wm_f(g_i)$ is weighted mutation frequency and $r(g_i)$ is the gene neighbor influence.

2.3 Gene Prioritization

Using the gene mutation enrichment scores obtained in Section 2.2, we generate a rating function to prioritize genes. The mutation enrichment scores matrix of $M \in \mathbb{F}^{g \times g}$ is defined by

$M_{ij} = \#\{ms(g_i) \geq ms(g_j)\}$, where $ms(g_i)$ and $ms(g_j)$ ms score for gene i and j respectively. Next, we prioritize genes using the dynamic PageRank we iteratively calculate gene rating scores by $PR(g_i) = \frac{\lambda}{g} + (1-\lambda)\sum_{g_j \in G^+(g_i)} \frac{PR(g_j)}{ms(g_j)}$, where $G^+(g_i)$ is the set of genes with low mutation score against gene g_i, $ms(g_j)$ is the mutation score of g_j, and $\lambda \in [0,1]$ is a damping factor(usually 0.1 or 0.2) to guarantee convergence. We rewrite the above equation relationship to Markov chains in a vector form as $\mathbf{PR} = \frac{\lambda}{G}[I-(1-\lambda)SD^{-1}]^{-1}\mathbf{1}$, where $\mathbf{PR}$ is the PageRank vector containing values of each gene, D is the diagonal matrix $D = \text{diag}[(D_{ii} = \sum_{\ell=1}^{g} S_{i\ell})_{i=1}^{g})]$, and $\mathbf{I}$ is the $g \times g$ identity matrix.

3 MODEL EVALUATION CRITERIA

3.1 Forward-Looking Approach (FLA)

To evaluate the rating and ranking stability, we applied our previous Forward-Looking Approach with an Expanded Window (FLA-WE)[10]. For rating stability in $S_{EW}^{(K,\Delta k=20)}$ we define Euclidean distance between two rating vectors by $d_{EW}^2(k) = ||\phi_{EW}^{(K\Delta k, k+\Delta k)} - \phi_{EW}^{k})||_2^2$, where EW is window size and $||\cdot||_2$ is the Euclidean norm. Thus, to compute the mean $d_{EW}^2(k)$ for all $top-k$ prioritized genes in EW. For ranking stability we considered the relevance of genes, $\pi_{EW}^{rel_{ct}}$ in consecutive rankings to calculate distance weighted matrix (DWM) Kendall's tau correlation, τ_{EW}^{k} given by

$$\tau_{EW}^{k} = \frac{\sum\limits_{g_i<g_j} w_{g_i,g_j}\left(1 + sgn\left((\pi_{EW}^{1}(g_i) - \pi_{EW}^{2}(g_j))(\pi_{EW}^{rel_{ct}}(g_i) - \pi_{EW}^{rel_{ct}}(g_j))\right)\right)}{2\sum\limits_{g_i<g_j} w_{g_i,g_j}},$$

where w_{g_i,g_j}, is the weight, $\pi_{EW}^{1}(g_i)$ and $\pi_{EW}^{2}(g_j)$ is rank position of genes, g_i and g_j in ranking π_{EW}^{k} and $\pi_{EW}^{(k\Delta k, k+\Delta k)}$, respectively for $top-k$ prioritized genes.

3.2 Ranking Precision and Discounted Cumulative Gain (DCG)

We also evaluated the ranking quality of our proposed model based on precision and DCG. We calculated the precision as the ratio of cancer-related genes that undergo mutation in the top-K predicted set given by $Precision = \frac{TP}{TP+FP}$,where TP is the number of genes prioritized by our method that are in the benchmark. FP is the number of genes prioritized that are not in the benchmark. In addition, we applied DCG to evaluate gene relevance and ranking position, logarithmically decreasing with the position. Hence, DCG score for genes up to position p is calculated by $DCG_p = \sum_{i=1}^{PG_i} \frac{rel_{g_i}^{ct_j}}{\log_2(i+1)}$, where PG_p is the ranking list of p prioritized genes and $rel_{g_i}^{ct_j}$ is the relevance score of gene g_i in cancer type ct_j.

4 RESULTS

Our research compared the GBP-PR method with other peer-rating methods (Colley, Keener, and Massey) and found that GBP-PR demonstrated high rating stability and consistency for gene prioritization across six cancer datasets. The evaluation of gene ranking stability showed that our proposed GBP-PR approach exhibited reliable and robust performance. GBP-PR consistently outperformed other peer methods in precision and Discounted Cumulative Gain metrics for several cancer types. The top 20 prioritized genes by GBP-PR revealed significant genes with crucial roles in cancer development. These findings were validated by benchmark databases and the Cancer Genome Interpreter (CGI) datasets, confirming the biological relevance of our discoveries.

5 CONCLUSION

In conclusion, the GBP-PR method presented in this paper offers a comprehensive and flexible approach to prioritizing significant groups of related genes in cancer. Through the integration of mutation data, gene networks, and asymmetric spreading strength measures, the method effectively identifies potential driver genes and suggests novel genes for further investigation. While future research should focus on incorporating additional biological data and conducting more extensive experimental evaluations, the GBP-PR method has already made a valuable contribution to cancer genomics by providing a robust computational framework for identifying crucial genes involved in cancer.

Acknowledgements

Miklós Krész acknowledges the European Commission for funding the InnoRenew CoE project (Grant Agreement no. 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund). He is also grateful for the support of the Slovenian Research Agency (ARRS) through grants N2-0171 and N1-0223. András London was supported by National Research, Development and Innovation Office—NKFIH Fund No. SNN-135643.

References

[1] L. A. Garraway and E. S. Lander, "Lessons from the cancer genome," *Cell*, vol. 153, no. 1, pp. 17–37, 2013.

[2] B. H. Hristov and M. Singh, "Network-based coverage of mutational profiles reveals cancer genes," *Cell systems*, vol. 5, no. 3, pp. 221–229, 2017.

[3] M. A. Reyna, M. D. Leiserson, and B. J. Raphael, "Hierarchical hotnet: identifying hierarchies of altered subnetworks," *Bioinformatics*, vol. 34, no. 17, pp. i972–i980, 2018.

[4] F. Vandin, E. Upfal, and B. J. Raphael, "De novo discovery of mutated driver pathways in cancer," *Genome research*, vol. 22, no. 2, pp. 375–385, 2012.

[5] M. D. Leiserson, D. Blokh, R. Sharan, and B. J. Raphael, "Simultaneous identification of multiple driver pathways in cancer," *PLoS computational biology*, vol. 9, no. 5, p. e1003054, 2013.

[6] M. D. Leiserson, H.-T. Wu, F. Vandin, and B. J. Raphael, "Comet: a statistical approach to identify combinations of mutually exclusive alterations in cancer," *Genome biology*, vol. 16, no. 1, pp. 1–20, 2015.

[7] G. Ciriello, E. Cerami, C. Sander, and N. Schultz, "Mutual exclusivity analysis identifies oncogenic network modules," *Genome research*, vol. 22, no. 2, pp. 398–406, 2012.

[8] Y.-A. Kim, D.-Y. Cho, P. Dao, and T. M. Przytycka, "Memcover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types," *Bioinformatics*, vol. 31, no. 12, pp. i284–i292, 2015.

[9] A. N. Langville and C. D. Meyer, *Who's# 1?: the science of rating and ranking*. Princeton University Press, 2012.

[10] P. J. Ochieng, A. London, and M. Krész, "A forward-looking approach to compare ranking methods for sports," *Information*, vol. 13, no. 5, p. 232, 2022.