

**Proceedings of the 17<sup>th</sup> 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)

© 2023 Samo Drobne – Lidija Zadnik Stirn – Mirjana Kljajić Borštnar – Janez Povh – Janez Žerovnik

Proceedings of the 17<sup>th</sup> 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

Reviewers 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)

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted by any other means without the prior written permission of the copyright holder.

---

Proceedings of the 17<sup>th</sup> 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

Program Committee	I
Organizing Committee	II
Reviewers	III
Chairs	IV
Preface	V
Insights into the growth and impact of operations research in Slovenia over the past 30 (60) years: Building intellectual and social capital in the environment of OR from the perspective of the activities of SSI-SOR	VII
Professor Dr. Janez Grad - 90th Anniversary	XXIII
Professor Ddr. Viljem Rupnik - 90th Anniversary	XXVII
Contents	XXXI
Author index	XXXIX

---

## ***Plenary Lectures*** ***1***

<i>Andrej Kastrin</i> Knowledge Discovery by Literature Mining: From Serendipity to Computational Creativity	3
<i>Victor Magron</i> Sparse Polynomial Optimization: Theory and Practice	4
<i>Mirjana Pejić Bach</i> Operations Research meets Artificial Intelligence: Intersection or Union	5
<i>Marc Sevaux and Alexandru-Liviu Olteanu</i> Julia, a Programming Language for Operations Research	6
<i>Lingling Shi, Suresh P. Sethi and Metin Cakanyildirim</i> Promoting Electric Vehicles: Reducing Charging Inconvenience and Price via Station and Consumer Subsidies	7

---

## ***Special Session 1: Applications of OR in Agricultural Economics*** ***9***

<i>Živa Alif, Tanja Šumrada and Jaka Žgajnar</i> Does Economic Situation Cause Land Abandonment? Estimating Economic Viability of Farming in a Sub-Mediterranean Region in Slovenia	11
<i>Maja Borlinič Gačnik, Boris Prevolšek, Antonio Peláez Verdet, Alfonso Cerezo Medina and Črtomir Rozman</i> Measuring the Efficiency of Spain's Wineries Through Data Envelopment Analysis	15
<i>Jure Brečko and Jaka Žgajnar</i> Price Volatility and Its Impact on Farm Operations, an Example of Analysis with a Farm Model	19
<i>Gregor Kramberger, Matjaž Glavan and Karmen Pažek</i> Building Resilient Agricultural Systems: A Multi-Stakeholder Approach for Sustainable Transformation	23

<i>Nikola Obrenović, Maksim Lalić, Dimitrije Stefanović, Marko Panić, Sanja Brdar, Vladimir Crnojević and Oskar Marko</i>	
Optimised Routing of the Blueberry Cultivating Unmanned Ground Vehicle	27
<i>Maja Petrač, Krunoslav Zmaić and Jaka Žgajnar</i>	
Typical Family Dairy Farms in the Republic of Croatia	31
<i>Boris Prevolšek, Larisa Lorbek, Maja Borlinič Gačnik and Črtomir Rozman</i>	
Multi-Criteria Model for Assessment of SPA Service Quality	35
<i>Jaka Žgajnar and Lidija Zadnik Stirn</i>	
The Use of Operation Research Methods to Support Agricultural Policy	39
<b><i>Special Session 2: Applications of OR in Industry and Mechanical Engineering</i></b>	<b>43</b>
<i>Simon Brezovnik, Darja Rupnik Poklukar and Janez Žerovnik</i>	
Roman and Italian Rainbow Domination Number of Graphs	45
<i>Michaela Chocholatá</i>	
Heating with Solid Fuel in Slovak Dwellings: A GWR Approach	49
<i>Igor Reznichenko, Primož Podržaj and Aljoša Peperko</i>	
Control Theory and Numerical Analysis of Magnetic Field Involving Mechanical Systems	53
<i>Anita Talaja and Krešimir Samac</i>	
The Role of Resource Complementarity and Opportunism in Strategic Alliance Performance	57
<b><i>Special Session 3: Artificial Intelligence in Business: Obstacles and Perspectives</i></b>	<b>63</b>
<i>Mile Bošnjak and Mirjana Pejić Bach</i>	
Factors Affecting COVID-19 Vaccine Uptake of Young Adults: Machine Learning Approaches	65
<i>Aljaž Ferencek and Mirjana Kljajić Borštnar</i>	
Topic Modelling of Open Government Data Impact Areas Using GPT 3.5 Model	71
<i>Tea Šestanović and Tea Kalinić Milićević</i>	
A MCDM Approach to Machine Learning Model Selection: Bitcoin Return Forecasting	77
<i>Lukáš Veverka</i>	
Maximizing Ad Campaign Effectiveness through TV Viewership Analysis: A Machine Learning Investigation	83
<b><i>Special Session 4: Discrete Optimization Methods and Models for Real-world Problem Domain</i></b>	<b>87</b>
<i>Daniil Baldouski, Balázs Dávid, György Dósa, Tibor Dulai, Ágnes Werner-Stark and Miklós Krész</i>	
Managing and Optimizing Container Flow in Port Logistics	89
<i>József Békési, Gábor Galambos and Imre Papp</i>	
Automatic Planning of Vehicle and Driver Schedules for Public Transportation: A Case Study	93
<i>Zuzana Borcinova and Peter Zimmermann</i>	
Edges which are Critical for Emergency Service Systems	97
<i>Peter Zimmermann</i>	
Detection of Critical Vertices for the Designed Service System	101

<i>Balázs Dávid</i>	
Optimization of Tree Bucking with Quality and Volume Requirements	105
<i>Murat Elhüseyni, Balázs Dávid, László Hajdu, and Miklós Krész</i>	
Distributed System Based Sensor Networks and the Connected P-Median Problem	109
<i>Zsolt Ercsey, Zoltán Kovács and Tamás Storcz</i>	
Optimal Schedule of a Sport Shooting Competition	113
<i>Emrecan Erdem, Ayşe Dilek Maden and Masood Ur Rehman</i>	
The Laplacian Energy of some Special Tree Families	117
<i>Jaroslav Horáček</i>	
A General Framework for Modelling Opinion Formation	121
<i>Jaroslav Janáček and Marek Kvet</i>	
Scatter Search For Bi-Criteria Public Service System Design	125
<i>Marek Kvet and Jaroslav Janáček</i>	
Intensification and Diversification for Pareto Front Approximation	129
<i>Giuseppe Lancia and Paolo Vidoni</i>	
Finding the Best 2-OPT Move on Nearly Random Euclidean TSP Tours in Average Linear Time	133
<i>Maciej Machowiak</i>	
The Moldable Tasks in Container Port Terminal	138
<i>Grzegorz Pawlak</i>	
Simulation Model for Cyclic Single Track Railway Problem	142
<i>Małgorzata Sterna and Bartłomiej Popielarz</i>	
Metaheuristic Methods for TV Advertisement Scheduling Problem	146
<b><i>Special Session 5: Game Theory</i></b>	<b>151</b>
<i>Andrew Clausen and Christopher Staphenhurst</i>	
Detering Bribery with Scotch Hold'em Poker	153
<i>Péter Csóka and P. Jean-Jacques Herings</i>	
Non-Cooperative Bargaining on Debt Restructuring	154
<i>Martin Černý</i>	
Bounding Solution Concepts of Incomplete Cooperative Games	155
<i>Ziv Hellman and Miklós Pintér</i>	
Three Variations on Money Pump, Common Prior, and Trade	158
<i>Tamás Solymosi, Ata Atay and Marina Núñez</i>	
On the Core of Many-to-One Assignment Games	159
<b><i>Special Session 6: Industry &amp; Society 5.0: Optimization and Learning in Human and Industrial Environments</i></b>	<b>161</b>
<i>Kolos Cs. Ágoston, Marianna E.-Nagy and Janez Povh</i>	
Comparing Optimum KMEDIAN and MSS Clustering with Ground Truth Clustering	163
<i>Drago Bokal, Špela Tertinek, Anja Šketa, Janja Jerebic, Robert Repnik, Urška Martinc, Edita Rozina, Metka Zaletel, Branka Mirt and Vlasta Krmelj</i>	
Climate Risk Indicators for Small Communities – The Effect of Heat Stress on Mortality	169
<i>Alen Granda and Drago Bokal</i>	
On Code Quality and Code Relevance Metrics	175

<i>Elza Jurun, Daniela Garbin Praničević and Valentina Bašić Androja</i> Croatia Vs EU From the Perspective of Digital Skills	181
<i>Nataša Ošep Ferš and Aleš Zamuda</i> Improvement in Continuous Black-box Setting Search Performance by Tuning L-SHADE Differential Evolution Historical Memory Size Parameter	187
<i>Melani Potrč, Klemen Tršinar, Špela Kajzer, Špela Tertinek, Urška Martinc and Drago Bokal</i> The Maturity Model for Climate Neutrality and Business Process Optimization in Slovenian Companies of the Future	191
<i>Janez Povh and Aljaž Krpan</i> Partitioning Graphs for Advancing Stable Set Problem Solutions through Quantum Annealers	195
<i>Janez Povh and Dunja Pucher</i> Advancing Stable Set Problem Solutions through Quantum Annealers	200
<i>Aleš Zamuda</i> Solving 100-Digit Challenge with Score 100 by Extended Running Time and Parallel Benchmarking	206
<b><i>Special Session 7: Social Innovations in Ageing Studies Supported by OR Models</i></b>	<b>213</b>
<i>Marija Bogataj, David Bogataj, Carmen Rajer, Suresh Sethi and Samo Drobne</i> The Influence of Residential Dispersion on the Optimal Long-Term Care of Senior Citizens	215
<i>Samo Drobne and Marija Bogataj</i> Generational Distinctions in Migrations between Slovenian Municipalities	219
<i>Samo Drobne and Marija Bogataj</i> The Impact of Hierarchical Spatial Levels on Internal Migration by Age Cohorts in Slovenia	223
<i>Terezie Krestová, Aleš Kresta and Lucie Bestová</i> The Application of Age Management Practices in Organisations: Does the Economic Sector, Size, and Family Business Status Matter?	227
<i>Josipa Višić</i> COVID-19 Pandemic and Profitability Determinants in Elderly Care Homes: Evidence from Croatia	231
<i>Josipa Višić and Ivana Tomas Žiković</i> Panel Data Analysis in Predicting Demand for Institutional Long-Term Care for Older Adults	237
<i>Berislav Žmuk</i> Elderly Population Activity in Croatia: Gender Comparison by Text Analysis Approach	243
<b><i>Special Session 8: Unravelling the Business Models of Sharing Economy by Applying Methods of OR and Statistics</i></b>	<b>249</b>
<i>Milica Maricic, Katarina Cvetic, Marina Ignjatovic and Veljko Jeremic</i> Time Series Analysis of Airbnb House Rentals Price in the Balkan Region	251
<i>Milica Maricic, Andrea Popovic, Katarina Cvetic and Marina Ignjatovic</i> Shared Accommodation in Europe: Consumer Behaviour Analysis	255
<i>Veljko Uskokovic, Stefan Zdravkovic, Stefan Komazec and Veljko Jeremic</i> Detecting Trending Topics Captivating Circular Economy: A Bibliometric-Based Approach	259

---

**Session 1: Econometric Models and Statistics** **263**

---

<i>Peter Knížat and Andrea Furková</i> From Linear to Spatial Regression: Parametric Versus Nonparametric Functional Form	265
<i>Tomislav Korotaj, James Ming Chen and Nataša Kurnoga</i> Hierarchical Clustering of CEE Countries According to Educational and Labour Market Indicators	270
<i>Kosovka Ognjenović</i> Using the Stochastic Frontier Approach to Determine the Gender Wage Gap and Labour Market Efficiency	277
<i>Mario Pepur</i> Confirmatory Factor Analysis of the Sports Team Reputation Scale: A Case Study in Croatia	281

---

**Session 2: Human Resources** **287**

---

<i>Özlem Akarçay Pervin, Nimet Yapici Pehlivan and Gerhard Wilhelm Weber</i> Prediction of Heart Disease Using Classification and Feature Selection Methods	289
<i>Luka Goropečnik, Jože Kropivšek, Matej Jošt and Lidija Zadnik Stirn</i> Competency Model Developed by Using a Multi-Criteria Approach and K-means Clustering	293
<i>Marko Hell</i> The General Model for Curricula Structure Presentation - A Matrix Approach to Managing Curricula	298
<i>Nikola Kadoić, Nikolina Žajdela Hrustek and Maja Gligora Marković</i> The Analysis of Eurovision Song Contest Results: The Differences between Public and Jury Votes	303
<i>Lorena Mihelač and Janez Povh</i> Computational Segmentation of Children's Melody	308

---

**Session 3: Finance and Investments** **313**

---

<i>Ksenija Dumičić, Mihovil Anđelinović and Blagica Novkovska</i> Resource Productivity Influenced by Selected Circular Economy and Development Level Indicators: Profiles of Clustered EU Countries	315
<i>Adela Has, Kristina Hodak and Marinela Mokriš</i> Deep Learning for Predicting Corporate Financial Distress of Croatian IT Companies	321
<i>Ivana Jerković, Ana Rimac Smiljanić and Blanka Škrabić Perić</i> Interdependence between Cryptocurrency Adoption and Financial Literacy: A Cross-Country Evidence	329
<i>Aleš Kresta, Bahate Maidiya and Jialei Xiong</i> On AAI Sentiment Index Influence on S&P 500 Stocks	335
<i>Karol Szomolányi, Martin Lukáčik and Adriana Lukáčiková</i> Macroeconomic Facts of the EU27 Countries During the COVID-19 Pandemic	339
<i>Vladimir Šimić</i> The Impact of Globalization on Government Consumption – Worldwide Evidence	343
<i>Ante Toni Vrdoljak and Zdravka Aljinović</i> The Impact of Financial Literacy and Sociodemographic Factors on Investor Preferences	349



<i>Marija Vuković</i> The Role of the Financing Source in the Behavioral Intention to Buy Real Estate: Multigroup Analysis	355
---	-----

---

<b><i>Session 4: Location and Transport, Graphs and their Applications</i></b>	<b>361</b>
--	------------

---

<i>David Bogataj and Francisco Campuzano-Bolarin</i> How Would the Construction of a New Container Port in Cartagena Influence the Reduction of Greenhouse Gas Pollution	363
<i>Elif Garajová and Miroslav Rada</i> Interval Transportation Problem: The Worst Finite Optimal Value is Hard for Inequalities	367
<i>Peter Juma Ochieng, József Dombi, Tibor Kalmár, András London and Miklós Krész</i> Graph-Based Prioritization of Related Cancer Genes	371
<i>Juraj Pekár, Marian Reiff and Ivan Brezina</i> Location of Service Devices at any Point in Two-Dimensional Space	375
<i>Tina Šfligoj and Aljoša Peperko</i> Estimating Node Importance in Public Transport Networks	379
<i>Rabia Taspinar and Burak Kocuk</i> Discretization-Based Solution Approaches for the Circle Packing Problem	383
<i>Janez Žerovnik</i> On $t$ -Rainbow Domination Number of Generalized Petersen Graphs $P(ck, k)$	387

---

<b><i>Session 5: Mathematical Programming and Optimization</i></b>	<b>391</b>
--	------------

---

<i>Helena Gaspars-Wieloch and Katarzyna Wyrębska</i> Impact of the Unitarization Technique on Final Rankings Based on the Goal Programming	393
<i>Milan Hladík</i> A General Approach to Handle Complex Sensitivity Analysis in Linear Programming	397
<i>Alf Kimms, Hédi Király and Christin Münch</i> Computational Geometry and Mathematical Programming: Combined Techniques and Selected Applications in Logistics	401
<i>Vedran Kojić, Mira Krpan and Zrinka Lukač</i> A Simple Approach to Solving the Monopolist's Long-Run Profit Maximization Problem: The Case of the Hyperbolic Inverse Demand and Cobb-Douglas Production Functions with Two Factors of Production	405
<i>Anita Varga and Marianna E.-Nagy</i> Numerical Comparison of Long-Step Interior Point Algorithms for Solving Linear Complementarity Problems	411

---

<b><i>Session 6: Multi-Criteria Decision-Making</i></b>	<b>415</b>
---	------------

---

<i>Andrej Bregar, Anas Husseis and Jose Luis Flores</i> A MCDM Methodology for Cyberattack Mitigation	417
<i>Vesna Čančer</i> Measuring Accuracy of Approximation Methods for Priorities Derivation Based on Pairwise Comparisons	421
<i>Rok Drnovšek, Marija Milavec Kapun and Uroš Rajkovič</i> Risk Management in Healthcare: DEX (Decision Expert) Evaluation Model	425

<i>Petra Grošelj, Gregor Dolinar and Helena Erika Rojc</i>	
Aggregation of Individual Linguistic Evaluations in Spherical Analytic Hierarchy Process	429
<i>Jerzy Michnik</i>	
The Wings Model for Choice of Innovation Strategy	433
<i>Tjaša Šmidovnik and Petra Grošelj</i>	
Aggregating Individual Weights into Group Weights in Best-Worst Method	437
<i>Tadeusz Trzaskalik</i>	
AHP Application to Multistage Bipolar Method	441

## ***APPENDICES***

---

*Authors' Addresses*

*Program of SOR'23*

*Sponsors' Notices*

# 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  $g_i$ .

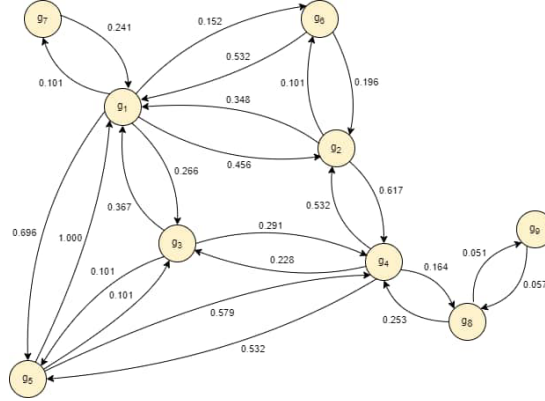


Figure 1: A simple hypothetical graph of extracted GSS

### 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  $r(g_i) = \sum_{g_k \in N(g_i)} num_f(g_k) \times ss(g_k, g_j)$ , 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,  $\tau_{EW}^k$  given by

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

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  $\pi_{EW}^k$  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  $DCG_p = \sum_{i=1}^{PG_p} \frac{rel_{g_i}^{ct_j}}{\log_2(i+1)}$ , 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.