HYBRID GLOBAL STRUCTURE MODEL FOR IDENTIFYING IMPACTFUL INFLUENTIAL NODES IN NETWORK ANALYSIS

MOHD FARIDUDDIN BIN MUKHTAR

DOCTOR OF PHILOSOPHY

2024

Faculty of Information and Communication Technology

HYBRID GLOBAL STRUCTURE MODEL FOR IDENTIFYING IMPACTFUL INFLUENTIAL NODES IN NETWORK ANALYSIS

Mohd Fariduddin bin Mukhtar

Doctor of Philosophy

2024

HYBRID GLOBAL STRUCTURE MODEL FOR IDENTIFYING IMPACTFUL INFLUENTIAL NODES IN NETWORK ANALYSIS

MOHD FARIDUDDIN BIN MUKHTAR

A thesis submitted
in fulfillment of the requirements for the degree of
Doctor of Philosophy

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

DECLARATION

I declare that this thesis entitled "Hybrid Global Structure Model For Identifying Impactful Influential Nodes In Network Analysis" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :...

Name : Mohd Fariduddin Mukhtar

Date :. 14/8/2024

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature : ....

Supervisor Name : Prof. Madya. Ts. Dr. Zuraida Abal Abas

Date : 14/8/2024

DEDICATION

In the Name of Allah, the Most Gracious, the Most Merciful.

Bismillah al-Rahman al-Rahim.

I begin this dedication with praise and gratitude to Allah SWT for granting me the strength, wisdom, and guidance throughout this academic journey.

To my beloved wife, your encouragement and sacrifices have kept me focused and determined during the challenges of this thesis. My heartfelt gratitude and love to my precious daughters. Your presence fills my life with joy and purpose. Your innocent smiles have been a source of inspiration, reminding me of the blessings bestowed upon me by the Almighty.

I am forever grateful to my parents, family, and friends for their prayers and encouragement throughout this endeavour. May Allah bless each and every one of you abundantly.

Lastly, I dedicate this work to the Prophet Muhammad (peace be upon him), whose teachings and example continue to illuminate our paths. May this thesis be a reflection of seeking knowledge and understanding, as encouraged in Islam. May Allah accept this effort and make it beneficial for all those who seek knowledge and truth.

Alhamdulillah, all praise is due to Allah alone.

ABSTRACT

Network analysis or graph analytics is crucial in identifying impactful nodes in complex networks, which are prevalent across diverse domains and display intricate structures and interactions. Understanding the significance of nodes within these networks is essential for uncovering their dynamics and functionalities. However, conventional centrality measures often struggle to capture the complexities of real-world networks, necessitating innovative solutions. While combining multiple centrality measures shows promise, optimizing these combinations remains challenging. Existing methods, such as the Global Structure Model (GSM), may require revision to fully assess individual nodes' unique influence. To address these gaps, this research introduces a novel hybrid centrality method called Global Structure Model-Degree-Kshell (GDK), integrating both local and global centrality measures. The aim of this research is to provide a more accurate and detailed evaluation of node influence within complex networks. GDK combines various centrality measures to offer comprehensive insights into node importance. Two variants of GDK are presented: GDK-A (addition) and GDK-M (multiplication). The methodology involves a standardized evaluation analysis to compare the performance of GDK-A and GDK-M against conventional centrality methods. Results indicate that GDK-M outperforms both traditional methods and GDK-A, demonstrating superior accuracy and effectiveness. Specifically, GDK-M shows improved performance percentages, highlighting its capability to better identify impactful nodes. This research significantly contributes to both academia and industry by enhancing network analysis techniques, enabling more informed decision-making across various domains. The introduction of the hybrid centrality method opens new possibilities for advancing the understanding of complex network analysis and its real-world applications. By exploring the hidden intricacies of complex networks, this study sheds light on their potential to shape the interconnected world.

MODEL STRUKTUR GLOBAL HIBRID UNTUK MENGENAL PASTI NOD BERPENGARUH BERIMPAK DALAM ANALISIS RANGKAIAN

ABSTRAK

Analisis rangkaian atau analitik graf adalah penting dalam mengenal pasti nod yang berpengaruh dalam rangkaian kompleks, yang terdapat di pelbagai domain dan mempamerkan struktur serta interaksi yang rumit. Memahami kepentingan nod dalam rangkaian ini adalah penting untuk mendedahkan dinamik dan fungsinya. Walau bagaimanapun, ukuran sentraliti konvensional sering menghadapi kesukaran untuk menangkap kerumitan rangkaian dunia sebenar, yang memerlukan penyelesaian inovatif. Walaupun gabungan pelbagai ukuran sentraliti menunjukkan potensi, mengoptimumkan gabungan ini tetap mencabar. Kaedah sedia ada, seperti Model Struktur Global (GSM), mungkin memerlukan semakan untuk menilai sepenuhnya pengaruh unik nod individu. Untuk mengatasi kekurangan ini, kajian ini memperkenalkan kaedah sentraliti hibrid baru yang dipanggil Model Struktur Global-Degree-Kshell (GDK), yang menggabungkan ukuran sentraliti tempatan dan global. Matlamat kajian ini adalah untuk menyediakan penilaian yang lebih tepat dan terperinci mengenai pengaruh nod dalam rangkaian kompleks. GDK menggabungkan pelbagai indeks sentraliti untuk menawarkan pandangan menyeluruh tentang kepentingan nod. Dua varian GDK diperkenalkan: GDK-A (penambahan) dan GDK-M (pendaraban). Metodologi melibatkan analisis penilaian piawai untuk membandingkan prestasi GDK-A dan GDK-M dengan kaedah sentraliti konvensional. Keputusan menunjukkan bahawa GDK-M mengatasi kedua-dua kaedah tradisional dan GDK-A, menunjukkan ketepatan dan keberkesanan yang lebih baik. Secara khusus, GDK-M menunjukkan peningkatan peratusan prestasi, menonjolkan keupayaannya untuk mengenal pasti nod yang berpengaruh dengan lebih baik. Kajian ini menyumbang secara signifikan kepada kedua-dua akademia dan industri dengan meningkatkan teknik analisis rangkaian, membolehkan pembuatan keputusan yang lebih berinformasi merentasi pelbagai domain. Pengenalan kaedah sentraliti hibrid membuka kemungkinan baru untuk memajukan pemahaman analisis rangkaian kompleks dan aplikasinya dalam dunia sebenar. Dengan meneroka kerumitan tersembunyi rangkaian kompleks, kajian ini mendedahkan potensinya untuk membentuk dunia yang saling berhubung.

ACKNOWLEDGEMENT

Words cannot express my gratitude to my supervisor, Prof. Madya Ts. Dr. Zuraida Abal Abas and co-supervisor, Ts. Dr Amir Hamzah Abdul Rasib for your guidance, invaluable patience and most importantly, providing positive encouragement and a warm spirit to finish this thesis.

I extend my deepest gratitude to my beloved wife, Nor Asyikin Abd Ghani for her unwavering love, understanding, and support. Your encouragement and sacrifices have been the driving force behind my pursuit of knowledge. To my precious daughters, Hidayat Nur Syifa’ and Hayyan Nur Sofea; you are my greatest inspiration. Your innocent smiles and boundless love have been a constant reminder of the purpose behind this endeavour. I dedicate this thesis to you, with the hope that it may serve as a testament to the power of determination and the pursuit of knowledge.

I also extend my heartfelt appreciation to all my family members. I would like to thank my dearest father, Mukhtar Mat Isa, my mother Kaushar Abd Hadi, and all my siblings. Their belief in me has kept my spirits and motivation high during this process.

I am also grateful to my classmates and cohort members, especially my office mates, for their editing help, late-night feedback sessions, and moral support. Thanks should also go to faculty management especially Puan Azean, FTMK and Bahagian Cuti Belajar UTeM for all the support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

UTeM	-	Universiti Teknikal Malaysia Melaka
GDK-A	-	Global Structure Model with the Degree Centrality and the K-shell measures using Addition
AUC	-	Area Under Curve
AUT	-	Authority Centrality
AVE	-	Avereage value of measures
BC	-	Betweenness Centrality
BKC	-	The Betweenness and Katz Centrality
BN	-	Bottleneck Centrality
C(v)	-	Importance of node
CC	-	Closeness Centrality
CCDF	-	Complementary Cumulative Centrality Distribution Function
CQ	-	Cross-clique connectivity Centrality
DC	-	Degree Centrality
EC	-	Eccentricity
EV	-	Eigenvector Centrality
F(t)	-	Total number of infected and recoved
FPR	-	False Positive Rate
GDK-M	-	Global Structure Model (GSM) with the Degree Centrality (DC) and the K-shell (Ks) measures using Multiplication
GI	-	Global influence
GK	-	Geodesic K-path Centrality
GSM	-	Global Structure Model

HITS	-	Hypertext-Induced Topic Search
HUB	-	Hubness Centrality
iGI	-	Improved Global influence
IGSM	-	Improved Global Structure Model
iSI	-	Improved Self-influence
IVI	-	The Integrated Value if Influence
Ks	-	K-shell decomposition
MNC	-	Maximum Neighbourhood Component Centrality
PR	-	PageRank
RC	-	Residual Closeness Centrality
RMS error	-	root mean square error
ROC	-	Receiver operating characteristics
R-sq	-	R-squared
SD	-	Standard deviation
SI	-	Self-influence
SIR	-	Susceptible-Infected-Recovered Model
SR	-	Symbolic Regression
TPR	-	True Positive Rate

LIST OF APPENDICES

LIST OF PUBLICATIONS

The followings are the list of publications related to the work on this thesis:

Mohd Fariduddin Mukhtar, Zuraida Abal Abas, Amir Hamzah Abdul Rasib, Siti Haryanti Hairol Anuar, Nurul Hafizah Mohd Zaki, Ahmad Fadzli Nizam Abdul Rahman, Zaheera Zainal Abidin and Abdul Samad Shibghatullah, 2022. Identifying Impactful Nodes with Centrality Measures Combinations using Symbolic Regressions. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(5), pp. 592 – 599.

Mohd Fariduddin Mukhtar, Zuraida Abal Abas, Amir Hamzah Abdul Rasib, Siti Haryanti Hairol Anuar, Nurul Hafizah Mohd Zaki, Ahmad Fadzli Nizam Abdul Rahman, Zaheera Zainal Abidin and Abdul Samad Shibghatullah, 2023. Hybrid Global Structure Model for Unraveling Impactful Nodes in Complex Networks. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 14(6), pp. 724 – 730.

Mohd Fariduddin Mukhtar, Zuraida Abal Abas, Amir Hamzah Abdul Rasib, Siti Haryanti Hairol Anuar, Nurul Hafizah Mohd Zaki, Zaheera Zainal Abidin, Siti Azirah Asmai, Ahmad Fadzli Nizam Abdul Rahman, 2023. Global Structure Model Modification To Improve Impactful Node Detection. *Journal of Engineering and Applied Sciences (ARPN-JEAS)*, 18(3), pp. 220 – 225.

Mukhtar, M. F., Abal Abas, Z., Baharuddin, A. S., Norizan, M. N., Fakhruddin, W. F., Minato, W., Rasib, A. H., Abidin, Z. Z., Rahman, A. F., and Anuar, S. H., 2023. Integrating local and global information to identify impactful nodes in complex networks. *Scientific Reports*, 13(1), pp.1-12.

Mukhtar, M, Abas, Z, Rasib, A, Anuar, S, Zaki, N et al. Graph analytics' centrality measurement in supply chain. Proceedings Computer Science, 2023(2): 0038, https://doi.org/10.55092/pcs2023020038.

Mohd Fariduddin Mukhtar, Zuraida Abal Abas, Amir Hamzah Abdul Rasib, Siti Haryanti Hairol Anuar, Nurul Hafizah Mohd Zaki, Zaheera Zainal Abidin and Siti Azirah Asmai., 2021. Graph Analytics' Centrality Measurement in Supply Chain. In: *International Conference in Applied Computing, VIRTUAL 2021 - ICAC2021*. 4 – 5 February 2021, AIP Publications.

Mohd Fariduddin Mukhtar, Zuraida Abal Abas, Amir Hamzah Abdul Rasib, Siti Haryanti Hairol Anuar, Nurul Hafizah Mohd Zaki, Zaheera Zainal Abidin and Siti Azirah Asmai, 2022. Strategizing Community Services Network with Graph Analytics. *The 2nd Conference on Community Engagement Project with Theme Community Service to Support Sustainable Development Goals in the Post Covid-19 Pandemic*. 19th January 2022.

CHAPTER 1

INTRODUCTION

1.1 Research Background

Complex networks are a crucial concept across various scientific disciplines, such as computer science, mathematics, and physics. These networks have a unique architecture that differs from a simple network. Complex networks have been extensively studied because of their importance in comprehending the structure and dynamics of complex systems (Dorogovtsev et al., 2005). Examining these networks is a crucial undertaking that can reveal valuable insights and patterns within the interconnected data structures. The investigation of complex networks has resulted in the creation of conceptual and mathematical toolsets, such as graph analytics, which facilitate the examination of the connectivity, structure, and dynamics of these systems (Bian and Deng, 2017; Wang et al., 2022).

Graph analytics plays an important role in this process as it allows for the extraction of valuable insights from complex networks (Wu et al., 2022). Graph analytics can provide a deeper understanding and management of complicated networks, tackling the difficulties linked to complexity of the computational process and memory demands when examining large-scale networks (Xu, 2021).

In order to extract useful insights and information from the interconnected elements, graph analytics comprises a wide range of tools and strategies for analyzing graph data structures. Community detection (Fernandes et al., 2019; Agrawal et al, 2020), path analytics (Pan and Saramäki, 2011; Zhang et al., 2017), connectivity analysis (Welton et al., 2020), and centrality (Kosch et al., 2005) are some of the techniques used in graph analytics. These techniques are useful in many fields, including computational biology, transportation systems, and social network research. These techniques make it possible to draw important conclusions and patterns from complex structures and give useful knowledge on the system.

Centralities are fundamental methods in graph analytics for finding the most important nodes in a network. Centrality measures offer valuable insights into the importance and impact of nodes, assisting in a range of analytical tasks like identifying key players in social networks, understanding critical infrastructure in transportation networks, and pinpointing essential proteins in biological networks (Mishra et al., 2021). Some examples are degree centrality, betweenness centrality, and eigenvector centrality. These metrics provide valuable insights into the significance of nodes in complex networks, revealing impactful nodes and hubs that play crucial roles in the dynamics of the network.

It is important to note that various centrality measures have specific focuses, and their effectiveness can vary based on the network's characteristics and analysis objectives (Borgatti, 2005; Zhuge and Zhang, 2009). Therefore, it is crucial to examine and contrast various centrality measures to gain a thorough grasp of the significance of nodes within complex networks.

Recent research has indicated the importance of incorporating various centrality measures in graph analytics. Through the integration or hybridization of multiple centrality measures, analysts can obtain a comprehensive perspective on the significance and impact of nodes within a network (Salavaty et al., 2020). Understanding the importance of a node's influence in a network is essential for grasping its overall impact. This understanding enables individuals to understand the significance of the node not only in its immediate surroundings, but also in terms of its broader impact on the entire network.

This study aims to build upon established foundations by proposing a novel method that combines multiple centrality measures to assess node influence in complex networks. The proposed hybrid method is designed to enhance the accuracy, robustness, and comprehensiveness of evaluating node importance within such networks. The primary objective of this research is to deepen the comprehension of node significance, identify key influencers, and ultimately enhance the performance and efficiency of complex networks across diverse domains.

1.2 Problem Statement

Identifying the most impactful nodes in complex networks is a challenging task due to the inherent limitations of existing methods. Current centrality measures, while providing valuable insights, often offer only a partial perspective as they fail to account for the unique attributes and influences of individual nodes (Du et al., 2014; Qiao et al., 2017; Rajeh et al., 2023). This partial perspective results from the separate consideration of local and global network properties, each with distinct characteristics and limitations. Neglecting unique node attributes and influences can have several detrimental effects on network analysis. First, it can lead to an incomplete understanding of the network's structure and dynamics. For instance, nodes with similar topological positions might play vastly different roles if their attributes are considered. Second, by not accounting for unique attributes, centrality measures may fail to identify truly impactful or critical nodes. Nodes with high centrality but low relevance in terms of their attributes might be overemphasized, while those with lower centrality but high attribute significance could be overlooked. Third, models that do not incorporate node-specific information may produce less accurate predictions of network behavior, such as the spread of information, disease, or influence. This can lead to ineffective strategies in areas like marketing, epidemiology, or network security. Finally, policies or interventions based solely on traditional centrality measures may be misguided if they do not consider node attributes. For example, targeting nodes based only on their degree might not be effective if those nodes lack the capacity or influence suggested by their connections alone.

Local measures capture node-specific details and immediate connections but often overlook the broader network context. Conversely, global measures provide an overarching view of the network structure but can miss nuanced local influences. Combining these perspectives could leverage the strengths of both approaches, addressing their respective weaknesses. However, the integration of local and global characteristics in a standardized manner remains an unresolved challenge (Ibnoulouafi et al., 2018; Salavaty et al., 2020; Shetty et al., 2022). The absence of standardized evaluation methods for hybrid centrality measures complicates the thorough assessment and comparison of nodes across different studies and domains (Dai et al., 2019; Wu et al., 2019). This lack of standardization can lead to biased conclusions and hinder the development of reliable strategies for network analysis.

1.3 Research Question

The research questions embark from the problem statement are as follows:

i) How different centrality measures and their combinations contribute to the effectiveness of identifying the most important nodes in complex networks?

ii) How to develop a method that effectively incorporates local and global centrality measures by hybridizing them with a global structure model, considering network information, to enhance the identification of key nodes in complex networks?

iii) How can the proposed method be validated through comprehensive evaluation with real-world datasets in identifying important nodes in complex networks?

1.4 Research Objective

The research aims to achieve the following objectives:

i) To analyze different centrality measures and their combinations in terms of the identifications of the most impactful nodes in complex networks.