

# Identifying Influential Nodes with Centrality Indices Combinations using Symbolic Regressions

Mohd Fariduddin Mukhtar<sup>1</sup>, Zuraida Abal Abas<sup>2</sup>, Amir Hamzah Abdul Rasib<sup>3</sup>  
Siti Haryanti Hairol Anuar<sup>4</sup>, Nurul Hafizah Mohd Zaki<sup>5</sup>, Ahmad Fadzli Nizam Abdul Rahman<sup>6</sup>  
Zaheera Zainal Abidin<sup>7</sup>, Abdul Samad Shibghatullah<sup>8</sup>

Faculty of Communication and Information Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia<sup>1,2,4,5,6,7</sup>  
Faculty of Mechanical and Manufacturing Engineering Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia<sup>1,3</sup>  
Institute of Computer Science & Digital Innovation, UCSI University, 56000 Cheras, Kuala Lumpur, Malaysia<sup>8</sup>

**Abstract**—Numerous strategies for determining the most influential nodes in a connected network have been developed. The use of centrality indices in a network allows the identification of the most important nodes in the network. Specific indices, on the other hand, cannot search for a network's entire meaning because they are only interested in a single attribute. Researchers frequently overlook an index's characteristics in favour of focusing on its application. The purpose of this research is to integrate selected centrality indices classified by their various properties. A symbolic regression approach was used to find meaningful mathematical expressions for this combination of indices. When the efficacy of the combined indices is compared to other methods, the combined indices react similarly and outperform the previous method. Using this adaptive technique, network researchers can now identify the most influential network nodes.

**Keywords**—Centrality indices; combination; symbolic regressions; influential nodes

## I. INTRODUCTION

For years, the prediction of the most influential nodes has been a source of contention [1]. The node with the most impact is ranked first, and the one with the least effect is ranked last [2]–[4]. Several research had been carried out to enlist the importance of nodes detection which is such as in finding importance suppliers [5], [6], detection of cancer or virus gene [7] or as well as to monitor the terrorist activities [8].

Over the years, over 403 indices have emerged from the four major indices: Degree Centrality (DC), Betweenness Centrality (BC), Eigenvector Centrality (EV), and Closeness Centrality (CC). Individual task and node importance priorities are claimed to impact the development of various indices. Various centrality measures have been employed to predict node outcomes, with the underlying assumption being that the more centrally situated the nodes are in the network, the greater their spreading potential [9], [10].

However, there were limitations to using indices as a single centrality metric because they could focus on one application area. DC for example, is a good indicator of a node's total connections [11]. Still, it does not necessarily imply a node's value in linking nodes or how central it is to the main group. CC on the other hand, determines how close a node is, but an independent network will not profit from its supremacy if two nodes are placed in distinct components [9]. In the case of BC,

the result will be zero if many nodes are not on the shortest possible path to the remainder of the network [12].

A single centrality metric proved to be insufficient for accurately predicting the network's most important nodes [13]. Combining centrality indices to determine the most influential nodes has been floated. According to [7], [14], there is no single centrality measure that can accurately identify key nodes, but the combination of at least two centrality measures is the most accurate. Combining multiple indices is considerably more accurate than using one index while assessing a node's influence capacity [11]. The influence of a node may be evaluated by its location and surroundings.

The researchers have recursively investigated this topic and determined that each indices have distinct features. This attribute, known as network topology, is reflected differently by different methods, and the evaluation results may include flaws or deficiencies. Borgatti [15] observed two forms of network topology: geodesic paths and walking paths. In the following research, Ashtiani [16] determined that centrality measures may be categorized into five classes based on the reasoning and formulas used. This characterization of centrality indices is also used by [17] while accessing the topological structure of student network.

The goal of this research is to investigate the effectiveness of centrality indices combination. Genetic programming-based symbolic regression (SR) is used here to find expressions to combine the selected indices. Two datasets examined to test the performance of a mix of indices based on their individual properties. Vignery's topology principles guided the selection of the centrality indices in this study. Combination outcomes are compared to results from a previous combination strategy to get a better sense of how effective the combinations are. At the end of this research, the possibility on applying symbolic regression to combine centrality indices will be clarified, and whether the categorization of indices according to their characteristics similarities has an impact on the combined indices.

## II. MATERIALS AND METHODS

### A. Data

Zachary and Les Misérables (Les-M) datasets, both weighted and unweighted, were used in this investigation. Thirty-four people from the karate club were included in the

Zachary dataset, which documented 78 connections between members who interacted outside the club. The novel Les-M features 77 nodes and 254 edges, including co-occurring characters. Both networks were depicted in the Fig. 1, with information on the most connected nodes.

**B. Theoretical Topology of Centrality Indices**

The definition of Vignery's eleven centrality indices is simplified and shown in Table I. Several of the indices had the same features and hierarchical clustering analysis is executed to observe whether the indices can be clustered into a single component to justify where the indices are converging. The dendrogram is built up by clustering observations and their similarity levels at each stage and assessing the similarity (or distance) levels of the produced clusters. As a first stage in the modeling technique, the value for each index is computed for each node. Following that, the indices was categorized based on their commonalities. The higher a cluster's similarity level, the more related the variables in that cluster are. These indices are divided into five theoretical groups, explained in Table I.

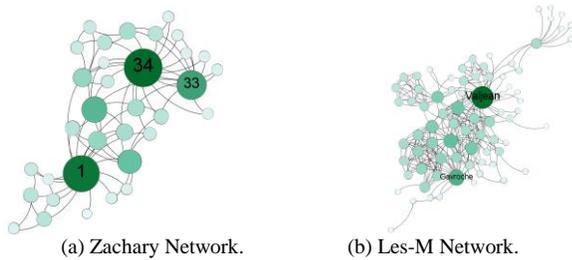


Fig. 1. Representation of both Networks.

TABLE I. VIGNERY'S CENTRALITY INDICES SELECTION

Type of centrality	Definition	Indices involved
Geodesic Distance and Path	A simple measure of the shortest path between two vertices. The length of a geodesic path is called geodesic distance or shortest distance.	Eccentricity (EC), Closeness (CC), Residual closeness (RC), Geodesic K-path (GK), Betweenness (BC), Bottleneck (BN)
Connectivity	Number of direct connections a node has, which means that every pair of vertices has a path linking them.	Eigenvector (EV), Hub & Authority (HUB & AUT), PageRank (PR)
Neighbor's prestige	Measures the number of shortest pathways a specific node has between any other two nodes and is based on the premise that a node with the shortest paths controls communication flows.	EV, HUB & AUT, PR
Nodes' adjacent	The neighbourhood of a node is the collection of all vertices near the vertex.	Cross-clique connectivity (CQ), Maximum Neighborhood Component (MNC)

**C. Combinations of Indices**

Network dynamics are examined as a function of the structure. The best estimate is made by combining a given number of indices based on the features of the component clusters. Genetic programming (GP) with symbolic regression (SR) is employed to generate mathematical expressions that may predict the simulation response values based on the topological indices used. SR is a technique that uses collected data to construct mathematical equations that may be used to test hypotheses [18]. With SR, the parameters and equation form are automatically searched, unlike typical regression methods that require a fixed-form model built from prior knowledge. GP is commonly utilized in SR because of the high computational complexity imposed by a vast search space that generates new solutions using the notion of biological evolution as a meta-heuristic [18], [19]. SR method's results will then be compared with those from other methods that use centrality indices as a comparison metric. Three algorithms are chosen for comparison which is provided by Eq. (1), (2) and (3).

- C(v) algorithm: Wang [20] developed a combination formula with the integration of DC, diffusion degree (DD) and BC as denoted in Eq. 1 with  $a + b + c = 1$ .

$$C(v) = aDC(v) + bDD(v) + cBC(v) \tag{1}$$

- BC and Katz (BKC) algorithm: Zhang [21] merged BC and Katz's centrality. Eq. 2 expresses the relationship between BC and KC.

$$BKC(v) = aBC + bKC \tag{2}$$

- Integrated Value of Influence (IVI): Consider topological dimensions [22], IVI consider six key network centrality measurements (normalizes connectivity (NC), ClusterRank (CR), BC, collective influence (CI), DC and local index (LH)) into account as given in Eq. 3.

$$IVI = ((NC' + CR')(BC' + CI'))(DC' + LH'_{index}) \tag{3}$$

III. RESULT AND DISCUSSION

A. Cluster Analysis

Clustering is used to group comparable data objects using a similarity measure. The similarity is a value that displays the strength of a relationship between two data items; it represents how similar data patterns are. The topological framework classifications from Vignery will be extended. We wanted to see how well the indices matched, so we applied a simple hierarchical clustering algorithm.

The results of clustering are depicted in Fig. 2a and 2b. We discovered that for both networks, all the indices could be clustered into four groups. Take note that the component clustering is quite close to Vinery's recommendations. The final partition specifies how the indices will be clustered. In those eleven indices, both networks were supposed to cluster similarly into four groups, except for Les-M, where PR will be in Cluster 2 rather than Cluster 3 (as in Zachary). EC has the

lowest similarity score and is not assigned to any category. EC was not assigned to any cluster because it has the lowest similarity score (10.56 and 7.00) compared to the other indices.

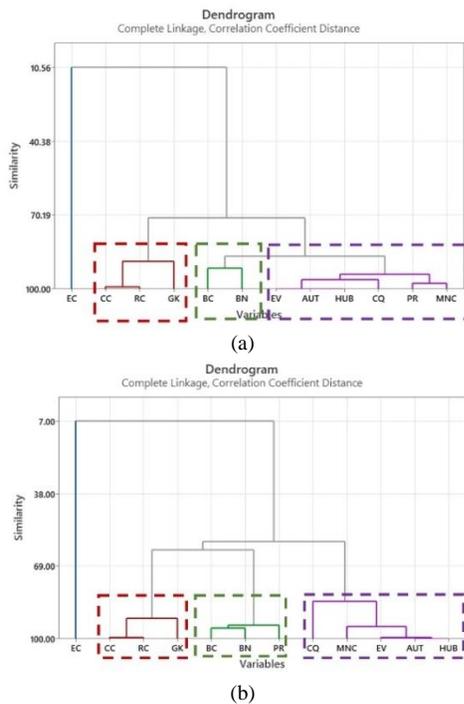


Fig. 2. (a). Hierarchical Clustering of Indices for Zachary. (b). Hierarchical Clustering of Indices for Les-M.

**B. Combinations of Indices**

Turingbot software was applied to execute the symbolic regression codes, which entails combining a set of base functions into simple formulas to produce a regression model. Fig. 3 depicts the steps we take to generate different mathematical formulations for each cluster. Because the output of SR will vary, we choose the phrase with the lowest value in terms of root mean square error (RMS error) and the highest R-squared (R-sq). Finally, we obtained four distinct expressions for each cluster specified, namely C1, C2, C3, and C. C is an expression that includes all the indices involved. As a result of efficient training and shifting, the analytic equations for both networks are shown in Table II are derived.

**C. Pearson Correlation Analysis**

The combined indices and component clusters employed in the earlier approach are compared. The correlation technique can be used to discover the similarity of combined centrality indices. The dataset also includes an average value (AVE) for the average result for each node from the Combined, IVI, BKC, and C(v). This AVE value will serve as the reference result to which the correlation converges.

Correlation for both networks show a significant and favorable relationship as shown in Tables IIIa and IIIb. In Zachary, there is a high correlation between AVE and IVI, BKC, and C(v). There is a high association between IVI and the clustered group and all other combined indices. The concept of combining indices while considering their spreaders and hubs can be extended for future use. With a correlation coefficient of 0.797, the relationship with C is likewise satisfactory. It is fascinating to notice that the C1 and C2 are more closely linked to AVE than is C, while C3 is obviously diverged from correlation toward others.

Results for Les-M also give results like Zachary's network when looking at each cluster component, with C1 and C2 being more correlated than C and C3. However, it is interesting to observe that IVI is quite diverging for this network. According to these findings, the SR modeling combination model has successfully identified influential nodes. In the following section, the node's ranking position is being observed.

**D. Node's Ranking of Position**

In this section, the placements of nodes were analyzed and arranged in descending order. When comparing procedures side by side in Tables IVa and IVb, the top ten ranking position of each approach is considered. Node AVE is a reference column that contains the average positioning value of nodes for each method, as expressed by the average positional value of nodes for each technique. Zachary and Les-M discovered that BKC and C(v) have a very similar node detection to AVE when comparing the two algorithms.

Nodes in IVI react similarly in Zachary, whereas they deviate significantly in Les-M. However, when compared to C for both networks, C1 and C2 show more comparable node detection for both networks when compared to C. To better understand the similarity result, we use the Jaccard similarity (JS) score and Kendall's tau-b to access the top ten nodes' ranking positions. Kendall's tau-b indices are used to measure the strength of a method's ranking position to understand the similarity result from the JS-score better.

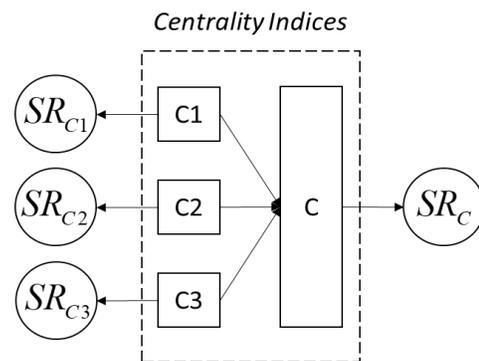


Fig. 3. Illustrations on Indices Combination for each Group.

TABLE II. MATHEMATICAL EXPRESSION OF COMBINED INDICES

Zachary	Les-M
Cluster 1 (C1): CC RC GK $0.00103544 \left( 1.33583 + \left( 0.915295 RC - \left( (-0.122916) (GK - \cos(GK)) \right) \right) \right)$ RMS Error = 0.000248416 R-Sq = 0.987144	Cluster 1 (C1): CC RC GK $(-0.00727966) / \left( 0.0565855 * \left( 0.802185 * \left( 0.121368 * "GK" + "RC" \right) \right) - 2.51961 \right)$ RMS Error = 0.000248416 R-Sq = 0.987144
Cluster 2 (C2): BC BN $\left( 2.12409 - \tanh \left( \text{round} \left( -0.457495 + \tan(BN) \right) + \tan(0.208093BN) \right) \right) \times$ $\left( \text{round} \left( -2.28163 - \cos(BN) \right) + \left( \left( \cos(-0.390946 - \text{floor}(0.542588BN)) \right) / 0.316869 + BN \right) \right)$ RMS Error = 0.841825 R-sq = 0.998876	Cluster 2 (C2): BC BN PR $(-3.95423e-05) \times \left( \tan(0.00897528 + (BN - 1.00369 * BC)) + 169.68 \right) \times$ $\left( \cos \left( (-0.329319) * \left( \cos(BN) + BC \right) \right) - (368.183 + BC) \right)$ RMS error = 0.00283 R-sq = 0.946660
Cluster 3 (C3): EV AUT HUB PR CQ MNC $(-0.00344533 + PR) \times$ $\left( \cos \left( \left( 0.41957 + MNC - (CQ - (MNC / 0.39926)) \right) \right) \times AUT \right) + 0.066212 + 2.55262 * HUB$ RMS Error = 0.0356482 R-Sq = 0.985527	Cluster 3 (C3): EV AUT HUB CQ MNC $\log_2(CQ) + \left( \left( 3.98809 + \tan(1.13497 - (-0.0407873 + (-0.0729471 + CQ))) \right) \right)$ $\left( - \left( AUT / \tan \left( \tan \left( \tan(0.0705538 * (-0.0449964 + CQ)) \right) \right) \right) \right)$ $\left( \times (-0.092534 + AUT) \right)$ RMS Error = 0.2957 R-Sq = 0.99513
Combine (C): EC CC RC GK BC BN EV AUT HUB PR CQ MNC $0.000275397 \times \left( GK - \left( 0.0113009 * \left( \left( \tan(1.23832 - GK) - \cos(MNC) + \left( -7.32811 + 0.979185 + MNC - \left( \tan(RC - 0.814871) + (BN - \tan(BC)) \right) \right) \right) * BN \right) \right) \right)$ $\left( -(\sinh(HUB) - RC - 9.00789) \right)$ RMS Error = 0.008983 R-Sq = 0.97642	Combine(C): EC CC RC GK BC BN EV AUT HUB PR CQ MNC $\left( ("PR" / (0.0140288 + 0.000763124 * "BC" - "EV")) + 4.15684 \right)$ $\times (1.32335 * "MNC" + "RC" - 9.22003) - (-103.165 * "HUB")$ $+ \cos("EC" - 8.01939 * "RC")$ RMS = 5.4994 R-Sq = 0.9941

TABLE III. (A) CORRELATIONS IN ZACHARY

	IVI	BKC	C(v)	C	C1	C2	C3
BKC	0.910						
C(v)	0.823	0.929					
C	0.712	0.736	0.605				
C1	0.752	0.726	0.592	0.975			
C2	0.800	0.856	0.735	0.778	0.762		
C3	0.484	0.525	0.439	0.278	0.268	0.366	
AVE	0.951	0.988	0.943	0.724	0.729	0.839	0.510

TABLE III. (B). CORRELATIONS IN LES-M

	IVI	BKC	C(v)	C	C1	C2	C3
BKC	0.320						
C(v)	0.539	0.943					
C	0.718	0.641	0.852				
C1	0.445	0.691	0.841	0.872			
C2	0.487	0.925	0.954	0.770	0.785		
C3	0.050	-0.004	0.042	0.079	0.087	0.093	
AVE	0.570	0.944	0.997	0.851	0.817	0.953	0.031

TABLE IV. (A) TOP TEN NODES BASED ON RANKING IN ZACHARY

Rank	Node AVE	Node IVI	Node BKC	Node Cv	Node C	Node C1	Node C2	Node C3
1	1	1	1	1	1	1	1	34
2	34	34	34	34	3	34	3	22
3	33	33	33	33	34	3	33	1
4	3	2	3	28	32	33	32	3
5	32	3	32	3	14	32	34	2
6	2	4	9	32	33	2	24	25
7	14	8	2	6	9	9	9	15
8	9	24	14	23	20	14	2	5
9	6	9	20	27	2	20	14	18
10	24	14	6	26	4	4	20	17

TABLE IV. (B) TOP TEN NODES BASED ON RANKING IN LES-M

Rank	Node AVE	Node IVI	Node BKC	Node Cv	Node C	Node C1	Node C2	Node C3
1	Valjean	Gavroche	Valjean	Valjean	Valjean	Valjean	Valjean	Simplice
2	Gavroche	Courfeyra	Myriel	Gavroche	Gavroche	Marius	Myriel	Toussaint
3	Marius	Bahorel	Gavroche	Marius	Marius	Gavroche	Gavroche	Woman2
4	Fantine	Joly	Marius	Fantine	Javert	Javert	Fantine	Tholomye
5	Myriel	Combefer	Fantine	Myriel	Enjolras	Thenardie	Marius	Bamatabo
6	Thenard	Feuilly	Thenardie	Thenardie	Thenardie	Enjolras	Cosette	Mlle.Gille
7	Enjolras	Valjean	Javert	Javert	Bossuet	Bossuet	Enjolras	Gillenorm
8	Javert	Enjolras	Mlle.Gille	Enjolras	Courfeyra	Cosette	Montparn	Fantine
9	Bossuet	Mabeuf	Enjolras	Bossuet	Bahorel	Fantine	Mme.The	Marius
10	Courfeyr	Grantaire	Tholomye	Mme.The	Joly	Babet	Thenardie	Prouvaire

- Jaccard similarity score

Jaccard similarity (JS) score compares two sets of scores by counting the number of elements in each group. JS can be calculated numerically by dividing the intersection of sets by the union of sets [23]. The higher the value, the greater the correlation between the two data sets. The higher the Jaccard similarity indices, the closer two sets of data are to one. Definition of JS is formulated as in Eq. 4.

$$JS(A, B) = \frac{\text{number of observations in both sets}}{\text{number in either set}} = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

For Zachary, C has a higher similarity score than Les-M when looking at the Jaccard scores for the two networks. If we look at Zachary and compare C to the previous combined technique, we see that C has the same top ten ranking entities for IVI (0.6667), BKC (0.8182), and AVE (0.6667). C has a low degree of similarity (less than 0.5) to IVI, BKC, and AVE in Les-M. C1 and C2 are interestingly comparable to C with the IVI, BKC, and C(v), which follow similar patterns. The JS-score heatmap for both networks are shown in the Tables Va and Vb. The greater the degree of resemblance between methods, the darker is the color.

TABLE V. (A) JACCARD SIMILARITY SCORE FOR ZACHARY

	C1	C2	C3	C	IVI	BKC	C(v)	AVE
C1	1							
C2	0.8182	1						
C3	0.25	0.25	1					
C	0.6667	0.8182	0.25	1				
IVI	0.6667	0.6667	0.25	0.6667	1			
BKC	0.8182	0.8182	0.25	0.8182	0.5385	1		
C(v)	0.3333	0.3333	0.1765	0.3333	0.25	0.4286	1	
AVE	0.6667	0.8182	0.25	0.6667	0.667	0.8182	0.4246	1

TABLE V. (B) JACCARD SIMILARITY SCORE FOR LES-M

	C1	C2	C3	C	BKC	AVE	C(v)	AVE
C1	1							
C2	0.5385	1						
C3	0	0.1111	1					
C	0.5385	0.3333	0	1				
IVI	0.1765	0.1765	0	0.4286	1			
BKC	0.5385	0.5385	0.25	0.4286	0.1765	1		
C(v)	0.6667	0.6667	0.4286	0.5385	0.1764	0.6667	1	
AVE	0.5385	0.4286	0.4286	0.4286	0.1765	0.5386	0.6667	1

• Kendall's tau-b

When comparing the rankings of different methods, Kendall's tau-b is applied to determine the ordinal relationship between pairs of observations [24], [25]. Correlation strength and direction are measured using Kendall's tau-b correlation coefficient. According to the theory of rank correlations, the closer two sets of data are linked together, the more closely they are related. Within this range, positive and negative numbers can indicate concordance or discordance, which is characterized by increasing or decreasing values, respectively. The correlation value between two variables increases when the ranks of the observations are similar; the correlation value decreases when the positions of the observations are different. Kendall's tau-b is defined as in Eq. 5.

$$\tau_b(A, B) = \frac{m_c - m_d}{\sqrt{(m - T_0)(m - T_1)}} \quad (5)$$

The top ten ranking of nodes using Kendall's tau-b results for Zachary and Les-M are shown in Tables VIa and VIb. Comparing C with IVI, BKC, C(v), and AVE shows that C has a rather low and moderate positive tau value in the Zachary network, while it has a high-rank similarity for the Les-M network. C also has significance and a strong positive tau correlation with C1 (0.644). C1 also shows the importance and positive correlation with C2. Observe from AVE analysis shows that results were significant in Les-M compared to Zachary except for C3. Since there were differences in C2 and C3 for Zachary and Les-M, it might affect the way on the rank behavior.

TABLE VI. (A) KENDALL'S CORRELATION SCORE FOR ZACHARY

	C1	C2	C3	C	IVI	BKC	C(v)	AVE
C1	1							
C2	0.067	1						
C3	-0.244	-0.467	1					
C	-0.022	0.467	-0.566*	1				
IVI	0.111	-0.111	-0.156	0.067	1			
BKC	0.378	0.244	-0.244	0.156	0.556*	1		
C(v)	0.2	0.244	0.022	0.333	0.467	0.467	1	
AVE	0.244	0.2	-0.378	0.022	0.6*	0.6	0.156	1

\*. Correlation is significant at the 0.05 level (2-tailed).

TABLE VI. (B) KENDALL'S CORRELATION SCORE FOR LES-M

	C1	C2	C3	C	IVI	BKC	C(v)	AVE
C1	1							
C2	0.556*	1						
C3	0.244	-0.2	1					
C	0.644*	0.467	0.156	1				
IVI	0.111	0.111	0.067	0.378	1			
BKC	0.644*	0.644*	0.067	0.644*	0.289	1		
C(v)	0.556	0.467	0.156	0.822*	0.467	0.733*	1	
AVE	0.6*	0.511*	0.111	0.867*	0.511*	0.778*	0.956*	1

\*. Correlation is significant at the 0.05 level (2-tailed).

#### IV. CONCLUSION

The selection of influential nodes is crucial for fostering knowledge and behavior adoption in a network because they can influence other nodes. It is possible to gain a better understanding of network structure and behavior by using prominent nodes. The importance of centrality in identifying influential network spreaders in this scenario cannot be overstated.

Our results show that combining centrality indices can identify the influential nodes in a network. To combine these indices, symbolic regression can identify appropriate mathematical expressions that will fit the network's features. When it comes to recognizing significant nodes, the newly constructed mathematical expression's function performs similarly or better than previous methods (IVI, BKC, and C(v)) that were validated using Pearson correlation, Jaccard similarity score, and Kendall's tau-b correlation of ranking.

It was discovered while clustering indices based on similar attributes that each cluster component may have the same impact as aggregating all indices. Clustering can reduce the total number of indices to be combined while achieving the same overall result. It was also discovered that index selection is critical. A few indices, including Katz centrality and Closeness centrality, could not be computed. Katz's centrality fails to detect connections between high-centrality nodes. To function, closeness centrality requires a well-connected network, and it fails when two nodes belong to different components. Large, complex datasets necessitate more computationally intensive methods. Clustering for the indices involved would be difficult, as it is here.

In future work, the selection of suitable indices to be combine is important. Analyzing their features as well as the computational time required to run each of the indices may be put into factor selection. It is also a good idea to run another run-up for a different network with a different number of weighted nodes if possible.

#### ACKNOWLEDGMENT

The author would like to extend a special thank you to the FRGS-RACER/2019/FTKMP-COSSID/F00412 grant, Centre of Research and Innovation Management (CRIM), Faculty of Mechanical and Manufacturing Technology, Universiti Teknikal Malaysia Melaka, Malaysia for funding this research. Appreciation also goes for Faculty of Communication and Information Technology and Ministry of Education Malaysia.

#### REFERENCES

- [1] Z. A. Abas et al., "Analytics : a Review of Current Trends , Future," *Compusoft*, vol. 9, no. 1, 2020.
- [2] Y. Yang, L. Yu, X. Wang, S. Chen, Y. Chen, and Y. Zhou, "A novel method to identify influential nodes in complex networks," *International Journal of Modern Physics C*, vol. 31, no. 2, Feb. 2020, doi: 10.1142/S0129183120500229.
- [3] A. Zareie, A. Sheikhhahmadi, M. Jalili, M. Sajjad, and K. Fasaie, "Finding influential nodes in social networks based on neighborhood correlation coefficient" vol. 194, p. 105580, 2020, doi: 10.1016/j.knosys.
- [4] M. Alshahrani, Z. Fuxi, A. Sameh, S. Mekouar, and S. Huang, "Efficient algorithms based on centrality measures for identification of top-K influential users in social networks," *Information Sciences*, vol. 527, pp. 88–107, Jul. 2020, doi: 10.1016/j.ins.2020.03.060.

- [5] N. J. Pulles, J. Veldman, and H. Schiele, "Identifying innovative suppliers in business networks: An empirical study," *Industrial Marketing Management*, vol. 43, no. 3, pp. 409–418, 2014, doi: 10.1016/j.indmarman.2013.12.009.
- [6] B. B. M. Shao, Z. (Michael) Shi, T. Y. Choi, and S. Chae, "A data-analytics approach to identifying hidden critical suppliers in supply networks: Development of nexus supplier index," *Decision Support Systems*, vol. 114, pp. 37–48, Oct. 2018, doi: 10.1016/j.dss.2018.08.008.
- [7] M. Jalili et al., "Evolution of centrality measurements for the detection of essential proteins in biological networks," *Frontiers in Physiology*, vol. 7, no. AUG. Frontiers Media S.A., Aug. 26, 2016. doi: 10.3389/fphys.2016.00375.
- [8] K. Basu, C. Zhou, A. Sen, and V. H. Goliber, "A novel graph analytic approach to monitor terrorist networks," *Proceedings - 16th IEEE International Symposium on Parallel and Distributed Processing with Applications, 17th IEEE International Conference on Ubiquitous Computing and Communications, 8th IEEE International Conference on Big Data and Cloud Computing*, 11t, pp. 1159–1166, 2019, doi: 10.1109/BDCloud.2018.00171.
- [9] Q. Ma and J. Ma, "Identifying and ranking influential spreaders in complex networks with consideration of spreading probability," *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 312–330, Jan. 2017, doi: 10.1016/j.physa.2016.08.041.
- [10] H. Zhou, M. Ruan, C. Zhu, V. C. M. Leung, S. Xu, and C. M. Huang, "A Time-Ordered Aggregation Model-Based Centrality Metric for Mobile Social Networks," *IEEE Access*, vol. 6, pp. 25588–25599, Apr. 2018, doi: 10.1109/ACCESS.2018.2831247.
- [11] A. Ibnoulouafi, M. el Haziti, and H. Cherifi, "M-Centrality: Identifying key nodes based on global position and local degree variation," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2018, no. 7, Jul. 2018, doi: 10.1088/1742-5468/aace08.
- [12] J. Wang, C. Li, and C. Xia, "Improved centrality indicators to characterize the nodal spreading capability in complex networks," *Applied Mathematics and Computation*, vol. 334, pp. 388–400, Oct. 2018, doi: 10.1016/j.amc.2018.04.028.
- [13] K. Raman, N. Damaraju, and G. K. Joshi, "The organisational structure of protein networks: Revisiting the centrality-lethality hypothesis," *Systems and Synthetic Biology*, vol. 8, no. 1, 2014, doi: 10.1007/s11693-013-9123-5.
- [14] G. Scardoni, G. Tosadori, M. Faizan, F. Spoto, F. Fabbri, and C. Laudanna, "Biological network analysis with CentiScaPe: centralities and experimental dataset integration," *F1000Res*, vol. 3, no. 0, p. 139, 2014, doi: 10.12688/f1000research.4477.1.
- [15] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005, doi: 10.1016/j.socnet.2004.11.008.
- [16] M. Ashtiani et al., "A systematic survey of centrality measures for protein-protein interaction networks," *BMC Systems Biology*, vol. 12, no. 1, pp. 1–17, 2018, doi: 10.1186/s12918-018-0598-2.
- [17] K. Vignery and W. Laurier, "A methodology and theoretical taxonomy for centrality measures: What are the best centrality indicators for student networks?," vol. 15, no. 12 December. 2020. doi: 10.1371/journal.pone.0244377.
- [18] B. K. Petersen, T. N. Mundhenk, S. K. Kim, C. P. Santiago, and J. T. Kim, "Deep symbolic regression," 2021.
- [19] S. Kim et al., "Integration of Neural Network-Based Symbolic Regression in Deep Learning for Scientific Discovery," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, 2021, doi: 10.1109/TNNLS.2020.3017010.
- [20] W. Jianwei, R. Lili, and G. Tianzhu, "A new measure of node importance in complex networks with tunable parameters," *2008 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2008*, pp. 1–4, 2008, doi: 10.1109/WiCom.2008.1170.
- [21] Y. Zhang, Y. Bao, S. Zhao, J. Chen, and J. Tang, "Identifying Node Importance by Combining Betweenness Centrality and Katz Centrality," in *Proceedings - 2015 International Conference on Cloud Computing and Big Data, CCBBD 2015*, Apr. 2016, pp. 354–357. doi: 10.1109/CCBD.2015.19.

- [22] A. Salavaty, M. Ramialison, and P. D. Currie, "Integrated Value of Influence: An Integrative Method for the Identification of the Most Influential Nodes within Networks," *Patterns*, vol. 1, no. 5, Aug. 2020, doi: 10.1016/j.patter.2020.100052.
- [23] M. R. Hamedani and S.-W. Kim, "JacSim: An accurate and efficient link-based similarity measure in graphs," *Information Sciences*, vol. 414, pp. 203–224, Nov. 2017, doi: 10.1016/j.ins.2017.06.005.
- [24] D. A. Walker, "JMASM9: Converting Kendall's Tau For Correlational Or Meta-Analytic Analyses," 2003.
- [25] Y. Y. Keng, K. H. Kwa, and C. McClain, "Convex combinations of centrality measures," *Journal of Mathematical Sociology*, vol. 45, no. 4, pp. 195–222, 2021, doi: 10.1080/0022250X.2020.1765776.