Journal of Physical Chemistry and Functional Materials

Home Page of Journal: https://dergipark.org.tr/jphcfum

An Effective and Robust Approach Based on Malatya Centrality Algorithm for Interpreting Cheminformatics Graphs Using Maximum Clique

Selman Yakut[a]* and Furkan Öztemiz[a]

[a]*Inonu University, Department of Software Engineering, Malatya, Turkey*
** Corresponding author: E-mail: selman.yakut@inonu.edu.tr*

ABSTRACT

Cheminformatics graphs are derived by transforming the atomic nodes and bonds of chemical compounds into graph structures and are used to analyze the chemical and structural properties of molecules. In this study, an effective and robust approach based on the Malatya Centrality Algorithm is proposed for identifying the maximum clique in cheminformatics graphs. The proposed method transforms cheminformatics graphs by taking their complement and calculates the Malatya centrality values for these graphs. Using these values, the minimum independent set is identified in the complemented graph, which corresponds to the set of nodes forming the maximum clique in the original graph. The study demonstrates, through tests on various cheminformatics graphs, including enzyme and molecular graphs, that maximum clique and chromatic number values provide significant insights into the structural properties of these graphs. Notably, the maximum clique value was often calculated as 2 for bipartite graphs. Additionally, it was observed that enzyme graphs exhibit maximum clique and chromatic number values that are optimal or near-optimal, with some graphs possessing perfect graph properties. The proposed approach offers an effective and robust solution for structural analysis in cheminformatics graphs.

ARTICLE INFO

Keywords:
Cheminformatics Graphs
Enzyme Graphs
Malatya Centrality
Maximum Clique Problem
Molecular Graphs

Received: 2024-11-23
Accepted: 2024-12-05
ISSN: 2651-3080
DOI: 10.54565/jphcfum.1590385

1. INTRODUCTION

Cheminformatics graphs are derived by transforming the atomic nodes and bonds of chemical compounds into graph structures, which are utilized to investigate the chemical and structural properties of molecules [1-3]. Furthermore, these graphs are widely employed in chemical graph theory and mathematical chemistry [4]. Molecular networks represent chemical compounds as nodes and visualize relationships by connecting them with edges based on similarities or reaction pathways [5, 6]. Chemical property distribution graphs present the physicochemical properties of molecules as histograms or scatter plots, contributing to the identification of suitable candidate molecules and facilitating database analysis processes [7].

Given the diversity of molecular types with different connections and properties, various graph structures can be developed and utilized for their representation, analysis, and generation [8]. Structure-based networks group molecules based on common substructures or binding motifs, while 3D molecular visualizations represent molecules as three-dimensional structures, aiding in analyses of protein-ligand interactions and molecular dynamics simulations [9]. Additionally, statistical methods are used for classifying molecules based on their chemical properties in areas such as molecular clustering and diversity analysis [10].

In graph theory, a clique refers to a subset of nodes in a graph where each node is directly connected to every other node in the subset. The maximum clique is the largest subset of nodes in which all nodes are directly

connected. The maximum clique problem, which involves identifying all such subsets in a graph, is generally classified as an NP-hard problem [11]. Solving this problem requires exploring numerous possibilities, making it challenging to find solutions in large and complex graphs. Consequently, specialized algorithms and optimization methods have been developed to address the maximum clique problem [12].

Maximum cliques define the largest connected subsets in cheminformatics graphs. The analysis of maximum cliques plays a critical role in identifying molecular similarities. They are also utilized in various cheminformatics applications, such as understanding molecular structures, modeling protein-ligand interactions, and measuring chemical similarities. However, finding maximum cliques in large and complex graphs is computationally demanding. To address this, several techniques, such as depth-first search algorithms like the Bron-Kerbosch algorithm and evolutionary algorithms, have been developed. Cheminformatics software tools such as RDKit and ChemAxon provide functions to facilitate these analyses, particularly for computing maximum cliques and identifying molecular similarities [12], [9]. These analyses are instrumental in the design of new compounds in drug discovery and bioinformatics.

This study proposes an effective and robust approach based on the Malatya Centrality Algorithm (MCA) [13], [14] to determine maximum cliques in cheminformatics graphs [15]. The proposed method involves the steps of graph transformation for molecules, calculating Malatya Centrality (MC) values, and identifying maximum cliques and minimum independent sets using these values. Tests conducted on various cheminformatics graphs, such as molecular and enzyme graphs, demonstrate the efficiency and accuracy of this approach. Analyses of graphs with different sizes, complexities, and classes revealed that the maximum clique value is typically 2 in bipartite graphs. In tested enzyme graphs, the values of maximum clique and chromatic number were found to be optimal or near-optimal, with some graphs exhibiting perfect graph properties. This approach provides significant insights into the relationship between maximum cliques and chromatic numbers in cheminformatics graphs, offering an effective solution for structural analysis. Moreover, the proposed method serves as a practical and efficient tool for simplifying and analyzing complex problems.

This article is organized as follows: Section 2 reviews the relevant literature. Section 3 provides the details of the proposed approach. Section 4 evaluates the experimental results of the proposed study, and Section 5 concludes the study.

2. LITERATURE REWIEV

The graph structure is the fundamental data model of graph theory, which is widely used in modeling various problems across different domains. Among its primary applications are cheminformatics graphs, such as molecular and enzyme graphs [4]. Several approaches have been proposed and developed based on graph theory to construct these graphs [16]. Modeling cheminformatics structures in various ways is extensively utilized in molecule generation, analysis, and chemical processes. By modeling these molecules in 2D and 3D forms, cheminformatics structures, connections, and chemical processes can be studied and analyzed [17].

The literature contains numerous studies that use, create, and examine cheminformatics graphs, primarily focusing on enzyme and molecular graphs. This study reviews works that utilize cheminformatics graphs for different purposes and tools. For instance, Okamoto et al. determined the maximum clique of a relationship graph constructed from the chemical structural formulas of four neuraminidase inhibitors used in treating influenza A and B by examining the Maximum Common Substructure (MCS) [18]. Combining node labels with element types and chemical bonds proved effective in reducing the number of nodes in the graph and computational costs [18]. Robert et al. proposed a new algorithm for identifying disconnected MCS by imposing constraints on the number and size of components. This approach significantly improved structure-based similarity measures while maintaining reasonable computation times [19]. Steve and Douglas investigated the utility of MCS in evaluating structural similarities between marketed drugs and natural human metabolites, demonstrating that MCS offers a chemically meaningful and independent similarity measure compared to fingerprint methods. This provides a valuable strategy for understanding relationships between drugs and metabolites through shared enzymes [20].

Cao et al. proposed a novel backtracking algorithm for predicting biologically active compounds using MCS methods. Tests integrating MCS and traditional similarity measures with support vector machines demonstrated higher specificity and sensitivity in predicting biological activity [21]. Luna et al. introduced a branching and bounding algorithm for efficiently calculating the Merrifield–Simmons index value for grid-like structures and independent sets related to chemical component properties and analysis, applicable to broader network graphs [22]. Ita et al. developed an algorithm for counting independent sets in grid-like structures based on branching and bounding techniques, which have applications in combinatorial mathematics, physics, chemistry, and computer science [23]. Singh et al. examined topological

indices based on degrees and distances of inorganic compounds like lead sulfide (PbS), calculating their metric dimensions, fault-tolerant metric dimensions, and Zagreb indices to analyze the three-dimensional structures of atoms [4]. Raza et al. explored the expected values of Sombor, reduced Sombor, and mean Sombor indices for specific organic compounds, finding these indices effective in modeling chemical thermodynamic structures [24]. Ali et al. investigated various topological properties of invariant graphs of finite groups, calculating Hosoya indices derived from molecular structures representing fundamental structural properties of SL(2, C) finite subgroups [25].

Mercado et al. introduced a graph-based molecular design platform that uses graph neural networks (GNNs) to probabilistically generate new molecules, demonstrating superior performance with a gated GNN [3]. Shui and Karypis proposed heterogeneous molecular graphs and GNNs to model multi-interactions between atoms, achieving superior performance in chemical property prediction by considering interactions involving three or more atoms [26].Wu et al. presented a method to explain molecular property predictions using GNNs by converting structures into molecular graphs and employing chemically meaningful molecular segmentations, enabling the study of structure-activity relationships [27]. Kengkanna and Ohue investigated the impact of different molecular graph representations on model learning and interpretation in molecular property prediction using GNNs, showing that representations such as atom, pharmacophore, and functional groups enhance model performance and provide comprehensive interpretations for applications like drug discovery [28].

David et al. reviewed molecular and macromolecular representations, primarily graph-based, in drug discovery, highlighting their roles in AI-driven applications and offering a guide on chemical representations [29]. Bougueroua et al. examined graph-theory-based methods supported by game theory and reinforcement learning for predicting 3D structures, accelerating conformer identification and transition tracking using 2D topological graphs [30].

Zang et al. proposed a framework for molecular graph representation learning using motif structures and hierarchical self-supervised tasks, capturing chemical information and properties to enhance molecular property predictions [31]. Costa et al. explained the generation and analysis of molecular fragments using graph theory for QSAR and QSPR modeling of chemical compound properties, achieving successful outcomes with this approach [32].

Sharma et al. determined mixed metric dimensions for complex molecular graphs like pentagonal carbon nanocones, demonstrating the sufficiency of three distinct non-adjacent nodes to uniquely identify all edges and nodes in these molecules [33]. Zhang et al. presented a framework that combines 2D topology and 3D geometry for unified modeling in molecular design, excelling in generating drug-like molecules and structure-based drug design [17].

The literature demonstrates the widespread use of cheminformatics graphs in fields like mathematical chemistry. Various cheminformatics graphs exist, and algorithms from graph theory can be applied to them. These foundational algorithms in graph theory are employed in executing processes and analyzing entities related to chemical components. Numerous approaches have been proposed for studying these entities and processes. The computation of the maximum clique set in such graphs is critical for effectively conducting these processes. Thus, efficiently determining the maximum clique in cheminformatics graphs can provide successful solutions for numerous structural and chemical processes associated with these graphs.

3. MATERIAL AND METHOD

The determination of maximum clique members in cheminformatics graphs involves several transformation processes. These processes are outlined in Figure 1 in six stages. In Stage 1, the example molecule C9H12N2O4 is introduced for testing. This molecule consists of 27 atoms and 27 bonds. Additionally, to enhance the comprehensibility of the molecule's structure, its 3D visualization is presented in Stage 2. In Stage 3, the C9H12N2O4 molecule is converted into a molecular graph for testing. The resulting graph comprises 27 vertices and 27 edges. To identify the maximum clique members, the relationship defined in graph theory as MaxClique = Maximum Independent Set ($\overline{G}$) [34] is utilized. This relationship indicates that the independent set of the complement graph corresponds to the maximum clique members in the original graph. In Stage 4, the complement graph of the example molecule is constructed. The complement graph is derived by taking the complement of the original graph's edge set. In other words, it is created by removing the existing edges and adding the missing edges. In Stage 5, the MISA algorithm is applied to the complement graph to identify the independent set members. The connections formed by these identified vertices in the original graph represent the maximum clique. The maximum clique members for the C9H12N2O4 molecule are illustrated in Stage 6. Due to the bipartite nature of the example graph, the sizes of the

maximum cliques will not exceed two. This characteristic also implies that there will be multiple maximum clique results for the given example graph.

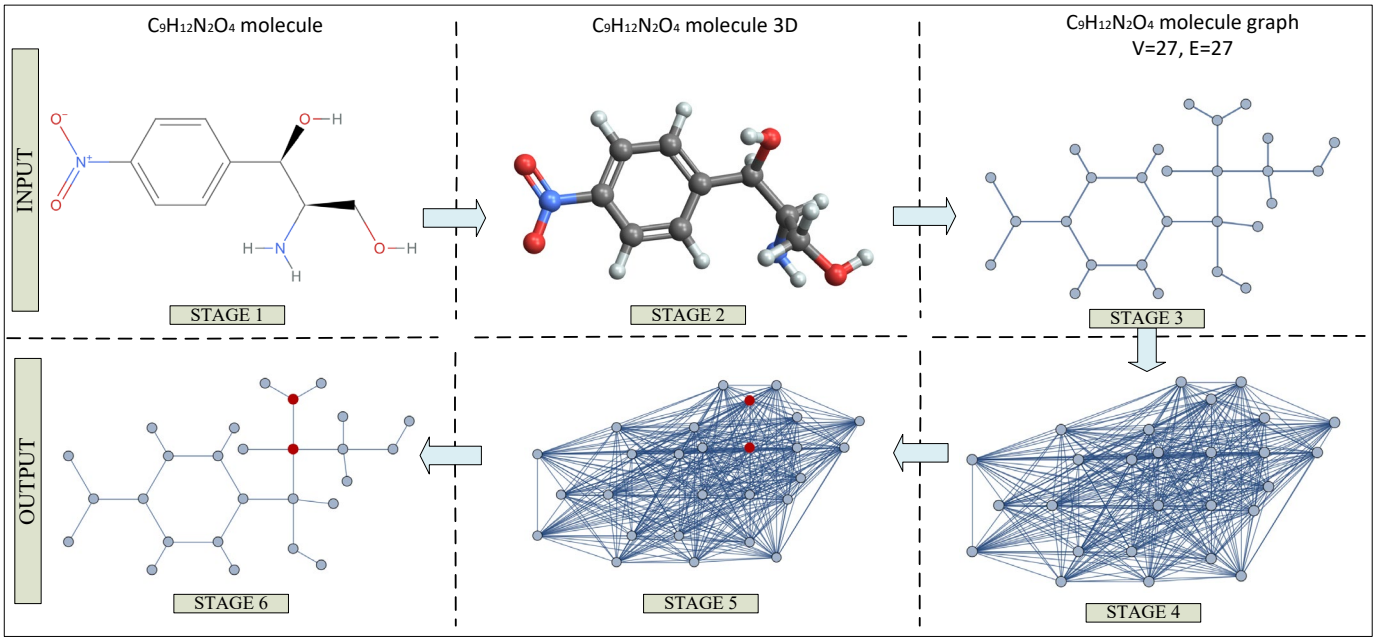

Figure 1 Graphical Abstract of the Proposed Approach

An independent set is a subset of vertices in a graph where no two vertices are adjacent. The maximum independent set is the largest such subset in a graph, and its size is denoted by α(G). Conversely, a clique is a subset of vertices in a graph where each vertex is connected to every other vertex in the subset. The maximum clique is the largest such subset, and its size is denoted by ω(G). Both problems are considered NP-hard, which is why heuristic or approximation methods are often preferred over exact solutions. The relationship between these two problems is significant: the maximum independent set determined in the complement graph corresponds to the maximum clique members in the original graph. The Malatya Independent Set Algorithm (MISA) is an efficient method that produces optimal or near-optimal results regardless of the graph type. The core step of MISA involves the MCA, which provides a priority ranking for vertex selection. MCA calculates the dominance values of vertices as the sum of the ratios of their connections to neighboring vertices, offering a novel contribution to the literature.

The centrality values of vertices in a graph are calculated using MCA through the operations defined in Equation 1 [35]. This algorithm is widely applied to solve fundamental problems in graph theory and beyond [14][35, 36]. The computation of the MC values, denoted as ψ(v), is given in Equation 1, where n represents the number of vertices in the graph, d(vi) denotes the degree of vertex vi, and N(vi) represents the neighboring vertices.

$$\psi(v_i) = \sum_{\forall v_j \in N(v_i)} \frac{d(v_i)}{d(v_j)} \quad (1)$$

The pseudocode for the proposed approach is presented in Algorithm 1. Initially, the complement of the cheminformatics graph is computed. Subsequently, the maximum independent set in the complement graph is identified. This set corresponds to the vertex set that provides the maximum clique solution for the original graph. The MISA selects the vertex with the lowest centrality value using the MCA. This vertex, along with its neighbors, is removed from the graph, and the process is repeated on the updated graph. This procedure continues until no vertices remain in the graph.

Algorithm 1 Pseudocode of the Proposed Approach

Maksimum independent set and maksimum clique

$\bar{G}$ <- Complement (G)
Input: Adjacency matrix of $\bar{G}$ is A and $\bar{G}$ =(V,E)
Output: $V_{ind} \subseteq V$, V_{ind}: It is the independent set solution in the graph $\bar{G}$
$V_{ind} \leftarrow \varnothing$
While $E \neq \varnothing$ do
  i←1, … |V|
    $\boldsymbol{\psi(v_i) = \sum_{\forall v_j \in N(v_i)} \frac{d(v_i)}{d(v_j)}}$
  $\boldsymbol{V_{ind} = V_{ind} \cup \{\min(\psi(v_i))\}}$
  V=V-{v_i}, and E=E-$\forall(v_i,v_j) \in E$
Output=V_{ind}: $V_{ind}(\bar{G}) = V_{clique}(G)$

4. RESULTS AND DISCUSSIONS

Cheminformatics graphs include various types, such as molecular graphs and enzyme graphs. In this study, tests were conducted on numerous enzyme and molecular graphs under the category of cheminformatics graphs. Table 1 presents information on the top 25 enzyme graphs with the highest number of vertices, published in the Network Repository [37]. The table includes the maximum clique values identified using MISA and the upper bound of the chromatic number. Chromatic number values were included to provide more detailed insights into the tested graphs. The chromatic number refers to the minimum number of colors required to color a graph. From an optimal results perspective, the maximum clique value and the chromatic number are not necessarily expected to be equal. Differences between these values may exist; however, they are generally expected to be close to each other. The chromatic number often provides approximate insights into graph coloring. The optimal chromatic number values in the table were obtained from the Network Repository. While the optimal maximum clique value does not guarantee the optimal chromatic number, it can offer approximate results as a baseline. If the chromatic number and maximum clique values in a graph are equal, the graph is considered a perfect graph [38]. Perfect graphs are advantageous for understanding structural properties and solving significant optimization problems efficiently.

Table 1 reveals several insights. For instance, the ENZYMES_g296 graph has 125 vertices and 282 edges. The maximum clique value for this graph, determined using MISA, is 3, with a chromatic number of 3. Another graph, ENZYMES_g103, contains 59 vertices and 230 edges, with a maximum clique value of 5 and a chromatic number of 4. The results indicate that the outcomes are either optimal or near-optimal. The table also includes several perfect graphs, such as ENZYMES-g297, ENZYMES-g293, and others. These graphs are highly valuable due to their regular structures and the advantages they offer in simplifying complex problems.

Table 1 Upper Bound of Chromatic Number and Maximum Clique Test Results in Enzyme Graphs

Cheminformatics Networks	V	E	Upper bound of Chromatic number	MISA Max Clique
ENZYMES_g296	125	282	3	3
ENZYMES_g295	123	278	4	3
ENZYMES-g297	121	298	3	3
ENZYMES-g293	96	218	3	3
ENZYMES-g118	95	242	3	3
ENZYMES-g123	90	254	4	4
ENZYMES-g8	88	266	3	3
ENZYMES-g532	74	240	4	4
ENZYMES-g355	66	224	3	4
ENZYMES-g501	66	160	3	3
ENZYMES-g504	66	240	4	4
ENZYMES-g349	64	236	4	4
ENZYMES-g199	62	216	4	4
ENZYMES-g291	62	208	4	4
ENZYMES-g279	60	214	4	4
ENZYMES-g292	60	200	3	4
ENZYMES-g484	60	160	4	4
ENZYMES-g578	60	206	4	4
ENZYMES-g103	59	230	4	5
ENZYMES-g526	58	220	4	3
ENZYMES-g204	57	210	5	5
ENZYMES-g209	57	202	4	4
ENZYMES-g527	57	214	4	4
ENZYMES-g203	56	200	4	4
ENZYMES-g541	56	152	3	4

The analysis of molecular graphs is illustrated in Figure 2, accompanied by visual representations. In Figure 2, the original form of the molecule, its graph representation, and the visualization highlighting the maximum clique members are provided. For example, the C12H22O11 molecule consists of 45 atoms and 46 bonds. When transformed into a graph, it comprises 45 vertices and 46 edges. After applying MISA, the maximum clique value was determined to be 2. This value represents the optimal maximum clique for the bipartite graph. Upon examining the figure, it is evident that the molecular graphs are bipartite. Consequently, all graphs analyzed yielded an optimal maximum clique value of 2. The results demonstrate that MISA produces optimal or near-optimal outcomes for both enzyme graphs and molecular graphs

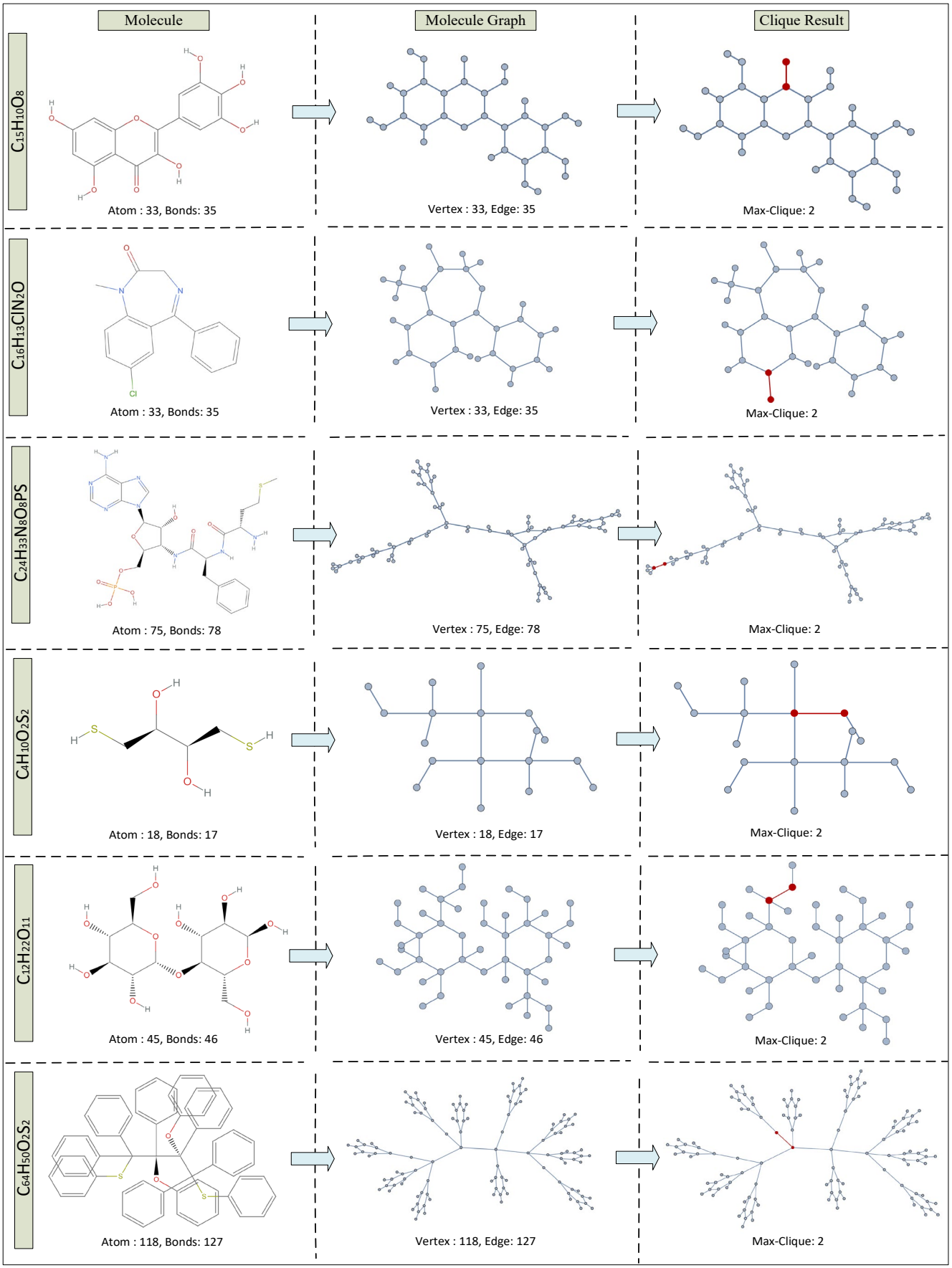

Figure 2 Maximum Clique Test Results and Visualizations for Molecular Graph

5. DISCUSSION

This study presents an effective and robust approach for identifying the maximum clique in cheminformatics graphs, based on the MCA. The modeling of molecular graphs using graph theory principles and the determination of the maximum clique using MC values constitute the core steps of the proposed method. The conducted tests demonstrated the effectiveness and accuracy of the approach on various cheminformatics graphs, including molecular and enzyme graphs. Analysis of graphs with different sizes, complexities, and classes revealed that the maximum clique value is typically 2 in bipartite molecular graphs. The results indicate that the proposed method provides significant insights into the relationship between the maximum clique and chromatic number values in cheminformatics graphs. Furthermore, it was observed that the maximum clique and chromatic number values for the tested enzyme graphs were optimal or near-optimal, with some graphs exhibiting properties of perfect graphs. These findings confirm that the proposed approach offers an effective solution for the structural analysis of cheminformatics graphs and provides simplified solutions for complex problems.

This study highlights the broad application potential of the proposed approach for analyzing cheminformatics graphs in fields such as chemistry, bioinformatics, and data science. The maximum clique determination method can be applied to disciplines such as the design of chemical compounds, biological network analysis, data classification, and machine learning. Additionally, it serves as a foundation for developing more efficient algorithms for complex graph theory problems, contributing to optimization processes.

Competing interests

The authors declare that they have no competing interests.

REFERENCES

[1] J. Bajorath, Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery. Springer Nature, 2013.

[2] P. Bongini, M. Bianchini, and F. Scarselli, “Molecular generative Graph Neural Networks for Drug Discovery,” Neurocomputing, vol. 450, pp. 242–252, 2021, doi: 10.1016/j.neucom.2021.04.039.

[3] R. Mercado et al., “Graph networks for molecular design,” Mach. Learn. Sci. Technol., vol. 2, no. 2, 2021, doi: 10.1088/2632-2153/abcf91.

[4] Y. Singh, S. K. Sharma, and P. Hazra, “Mathematical analysis of one-dimensional lead sulphide crystal structure using molecular graph theory,” Mol. Phys., vol. 120, no. 12, 2022, doi: 10.1080/00268976.2022.2086933.

[5] J. Stumpfe, D., & Bajorath, “Similarity searching and scaffold hopping in chemical space,” Nat. Chem. Biol., vol. 8, no. 2, pp. 118–126, 2012.

[6] N. M. Kriege, “Comparing Graphs,” 2015.

[7] C. A. Lipinski, “Lead- and drug-like compounds: The rule-of-five revolution,” Drug Discov. Today Technol., vol. 1, no. 4, pp. 337–341, 2004, doi: 10.1016/j.ddtec.2004.11.007.

[8] Z. Guo et al., “Graph-based Molecular Representation Learning,” IJCAI Int. Jt. Conf. Artif. Intell., vol. 2023-Augus, pp. 6638–6646, 2023, doi: 10.24963/ijcai.2023/744.

[9] A. J. Goodsell, D. S., Morris, G. M., & Olson, “Automated docking of flexible ligands: applications of AutoDock,” J. Mol. Recognit., vol. 9, no. 1, pp. 1–5, 1996, [Online]. Available: https://doi.org/10.1002/(SICI)1099-1352(199601)9:1%3C1::AID-JMR241%3E3.0.CO;2-6

[10] I. T. Jollife and J. Cadima, “Principal component analysis: A review and recent developments,” Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.

[11] R. M. Karp, “Complexity of Computer Computations,” in Reducibility Among Combinatorial Problems, Springer Berlin Heidelberg, 1972, pp. 85–103.

[12] C. Bron and J. Kerbosch, “Algorithm 457: Finding All Cliques of an Undirected Graph [H],” Commun. ACM, vol. 16, no. 9, pp. 575–577, 1973, doi: 10.1145/362342.362367.

[13] A. Karcı, S. Yakut, and F. Öztemiz, “A New Approach Based on Centrality Value in Solving the Minimum Vertex Cover Problem: Malatya Centrality Algorithm,” Comput. Sci., vol. 55, no. 35, pp. 1–100, Nov. 2022, doi: 10.53070/bbd.1195501.

[14] S. Yakut, F. Öztemiz, and A. Karcı, “A New Approach Based on Centrality Value in Solving the Maximum Independent Set Problem: Malatya Centrality Algorithm,” Comput. Sci., vol. 55, no. 35, pp. 1–100, Jan. 2023, doi: 10.53070/bbd.1224520.

[15] S. Yakut and F. Öztemiz, “A New Maximum Clique Method,” in International Informatics Congress, 2024, pp. 200–208.

[16] M. A. Yirik, E. K. Çolpan, S. Schmidt, M. Sorokina, and C. Steinbeck, “Review on Chemical Graph Theory and Its Application in Computer-Assisted Structure Elucidation,” Pre-print, no. November, 2021, doi: 10.20944/preprints202111.0546.v1.

[17] Z. Zhang, Q. Liu, C. K. Lee, C. Y. Hsieh, and E. Chen, “An equivariant generative framework for molecular graph-structure Co-design,” Chem. Sci., vol. 14, no. 31, pp. 8380–8392, 2023, doi: 10.1039/d3sc02538a.

[18] Y. Okamoto, “Finding a Maximum Common Subgraph from Molecular Structural Formulas through the Maximum Clique Approach Combined with the Ising Model,” ACS Omega, vol. 5, no. 22, pp. 13064–13068, 2020, doi: 10.1021/acsomega.0c00987.

[19] R. Schmidt, F. Krull, A. L. Heinzke, and M. Rarey, “Disconnected Maximum Common Substructures under Constraints,” J. Chem. Inf. Model., vol. 61, no. 1, pp. 167–178, 2021, doi: 10.1021/acs.jcim.0c00741.

[20] S. O’Hagan and D. B. Kell, “Analysis of drug-endogenous human metabolite similarities in terms of their maximum common substructures,” J. Cheminform., vol. 9, no. 1, pp. 1–17, 2017, doi: 10.1186/s13321-017-0198-y.

[21] Y. Cao, T. Jiang, and T. Girke, “A maximum common substructure-based algorithm for searching and predicting drug-like compounds,” Bioinformatics, vol. 24, no. 13, pp. 366–374, 2008, doi: 10.1093/bioinformatics/btn186.

[22] G. De Ita Luna, P. Bello López, and R. Marcial-Romero, “Counting Rules for Computing the Number of Independent Sets of a Grid Graph,” Mathematics, vol. 12, no. 6, pp. 1–14, 2024, doi: 10.3390/math12060922.

[23] G. De Ita, P. Bello, and M. Tovar, “A Branch and Bound Algorithm for Counting Independent Sets on Grid Graphs,” p. 28, 2023, doi: 10.3390/iocma2023-14434.

[24] Z. Raza, K. Naz, and S. Ahmad, “Expected Values of Molecular Descriptors in Random Polyphenyl Chains,” Emerg. Sci. J., vol. 6, no. 1, pp. 151–165, 2022, doi: 10.28991/ESJ-2022-06-01-012.

[25] F. Ali, B. A. Rather, M. Sarfraz, A. Ullah, N. Fatima, and W. K. Mashwani, "Certain Topological Indices of Non-Commuting Graphs for Finite Non-Abelian Groups," Molecules, vol. 27, no. 18, pp. 1–15, 2022, doi: 10.3390/molecules27186053.

[26] Z. Shui and G. Karypis, "Heterogeneous molecular graph neural networks for predicting molecule properties," Proc. - IEEE Int. Conf. Data Mining, ICDM, vol. 2020-Novem, no. Icdm, pp. 492–500, 2020, doi: 10.1109/ICDM50108.2020.00058.

[27] Z. Wu et al., "Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking," Nat. Commun., vol. 14, no. 1, pp. 1–15, 2023, doi: 10.1038/s41467-023-38192-3.

[28] A. Kengkanna and M. Ohue, "Enhancing property and activity prediction and interpretation using multiple molecular graph representations with MMGX," Commun. Chem., vol. 7, no. 1, 2024, doi: 10.1038/s42004-024-01155-w.

[29] L. David, A. Thakkar, R. Mercado, and O. Engkvist, "Molecular representations in AI-driven drug discovery: a review and practical guide," J. Cheminform., vol. 12, no. 1, pp. 1–22, 2020, doi: 10.1186/s13321-020-00460-5.

[30] S. Bougueroua, M. Bricage, Y. Aboulfath, D. Barth, and M. P. Gaigeot, "Algorithmic Graph Theory, Reinforcement Learning and Game Theory in MD Simulations: From 3D Structures to Topological 2D-Molecular Graphs (2D-MolGraphs) and Vice Versa," Molecules, vol. 28, no. 7, 2023, doi: 10.3390/molecules28072892.

[31] X. Zang, X. Zhao, and B. Tang, "Hierarchical Molecular Graph Self-Supervised Learning for property prediction," Commun. Chem., vol. 6, no. 1, pp. 1–10, 2023, doi: 10.1038/s42004-023-00825-5.

[32] P. C. S. Costa, J. S. Evangelista, I. Leal, and P. C. M. L. Miranda, "Chemical graph theory for property modeling in qsar and qspr—charming QSAR & QSPR," Mathematics, vol. 9, no. 1, pp. 1–19, 2021, doi: 10.3390/math9010060.

[33] S. K. Sharma, V. K. Bhat, H. A. E. W. Khalifa, and A. M. Alanzi, "Mixed Metric Dimension of Certain Carbon Nanocone Networks," Polycycl. Aromat. Compd., vol. 44, no. 3, pp. 2046–2061, 2024, doi: 10.1080/10406638.2023.2211734.

[34] Z. Ertem, E. Lykhovyd, Y. Wang, and S. Butenko, "The maximum independent union of cliques problem: complexity and exact approaches," J. Glob. Optim., vol. 76, no. 3, pp. 545–562, 2020, doi: 10.1007/s10898-018-0694-2.

[35] F. Öztemiz and S. Yakut, "An Effective Method for Determining Node Dominance Values: Malatya Centrality Algorithm," 32nd IEEE Conf. Signal Process. Commun. Appl. SIU 2024 - Proc., pp. 1–4, 2024, doi: 10.1109/SIU61531.2024.10601155.

[36] C. T. Bakan and S. Yakut, "Graf Teorisi ve Malatya Merkezilik Algoritmasına Dayalı Haber Metinlerinin Özetlemesi," BİLİŞİM Teknol. DERGİSİ, vol. 17, no. 3, pp. 189–198, 2024, doi: 10.17671/gazibtd.1463107.

[37] "Cheminformatics Graphs, Accesed Time: 02.11.2024, https://networkrepository.com/chem.php."

[38] M. C. Golumbic, "Chapter 3 - Perfect graphs," in Annals of Discrete Mathematics, Elsevier, 2004, pp. 51–80. [Online]. Available: https://doi.org/10.1016/S0167-5060(04)80051-7.