

The role of visualization and analysis of biological data in STEM education using graph structures

Dariia V. Yatseniak¹

¹*Ternopil Volodymyr Hnatiuk National Pedagogical University, 2 M. Kryvonosa Str., 46027, Ternopil, Ukraine*

Abstract

The article substantiates the use of graph structures in STEM education, particularly in the context of bioinformatics, molecular biology, and related fields. The author analyzes the advantages of applying graph models for the visualization and analysis of complex biological systems, such as genetic networks, protein interactions, and metabolic pathways. A comparative analysis of graph-based methods with other traditional approaches to modeling biological systems is presented, demonstrating their effectiveness in representing complex relationships. Special attention is given to the practical aspects of using graphs in the educational process, integrating interactive learning platforms and real-world application cases. The challenges of implementing graph models in educational programs are examined, and solutions are proposed through the integration of artificial intelligence, machine learning, and big data analysis tools. The significant potential of graph structures in developing students' analytical thinking and large-scale data processing skills is highlighted.

Keywords

STEM education, graph structures, bioinformatics, computational biology, data visualization, graph theory, network analysis, educational technology, STEM pedagogy, data-driven learning, biological pathways

1. Introduction

The increasing volume of biological data, driven by advances in science and technology, necessitates the development of effective methods for processing and analysis. The vast data sets generated by modern biological research encompass genomic sequences, protein interactions, metabolic pathways, and cellular signaling networks. There is a need to advocate for new teaching methods that would enable students not only to memorize these data, but also to analyze them in the context of the interrelationships between biological entities. In this regard, the integration of mathematical approaches into research and data analysis has garnered particular attention. It is important to note that traditional methods of teaching biology, which include text descriptions, static diagrams, and linear tables, are informative but insufficiently effective in representing more complex biological systems [1]. These tools lack functionality that could help participants in the learning process identify hidden patterns in biological networks or analyze multilevel interactions between molecules and organisms. Bioinformatics, as an interdisciplinary science, combines biology, mathematics, computer science, and statistics, offering tools for analyzing complex biological systems [2]. One of the most promising methods for analyzing biological data is graph structures, which allow modeling interactions between various elements of biological systems, promoting the development of analytical thinking in students.[3] The use of various branches of discrete mathematics, particularly graph theory, creates new opportunities for modeling and studying interactions between living organisms. Graph structures are effectively used in biological research to model genetic regulatory networks [4, 5], protein interactions [6], and analyze metabolic pathways [7]. In this regard, the question arises: Why do graph models still not occupy a central place in STEM education? The expansion of biological data results from the development of fields such as DNA sequencing, proteomics, systems biology, and other related disciplines. However, despite this, bioinformatics curricula are still focused on traditional methods of data representation, which do not allow for the effective processing of large data sets [8]. Genetic relationships, protein-protein

STEM@ICon-MaSTEd 2025: 4th Yuriy Ramskyi STE(A)M Workshop co-located with XVII International Conference on Mathematics, Science and Technology Education, May 14, 2025, Ternopil, Ukraine

✉ yatsenyak_dv@fizmat.tnpu.edu.ua (D. V. Yatseniak)

ORCID 0000-0002-8427-5532 (D. V. Yatseniak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

interactions, metabolic pathways, and evolutionary connections can be represented using graph models, which can then be analyzed using algorithmic methods. However, this approach is still not sufficiently integrated into STEM education, which limits the potential for applying computational methods to study complex biological phenomena.

Within this context, it is essential to equip future researchers and STEM students with the skills to work with graph models, enabling them to effectively analyze and explore data from various domains, with a particular focus on biological data, as emphasized in this study. We argue that graph structures should be a key tool for visualizing and analyzing biological data in STEM education. Such tools not only facilitate the understanding of complex biological systems, but also contribute to the development of analytical and computational thinking in all participants in the learning process. Moreover, the integration of graph visualization methods into biology and bioinformatics courses will promote the formation of interdisciplinary skills, connecting biology as a foundation with mathematics and programming. Despite a sufficiently extensive research base on the use of graph structures in bioinformatics, their educational potential remains underexplored. Studies [9] and [10] demonstrate the effectiveness of using graphs in predicting protein interactions, but there is still a lack of research on the impact of these methods with a focus on the educational process in STEM education.

The **aim** of this paper is to substantiate the application of graph structure methods in bioinformatics and biological education, as well as to stimulate the development of analytical and computational thinking in STEM students. The use of visualization techniques in biological research not only simplifies the analysis of large datasets but also facilitates a deeper understanding of complex biological processes. The integration of such methods into the educational process enhances the quality of education by combining theory with practice.

To achieve this goal, the paper covers several key aspects. First, it examines the fundamental principles of graph-based biological data analysis, including pathfinding, identifying intermediary elements in biological networks, and clustering. Second, it presents a novel approach to understanding complex biological processes by visualizing relationships between organisms through intuitive graphical models. Third, it highlights the significance of STEM education in fostering students' ability to compare and evaluate various biological systems.

2. Biological data and their representation using graphs

Graph structures are a fundamental tool in discrete mathematics and are actively used for modeling complex systems, including biological processes [11, 12]. A graph consists of vertices representing objects and edges depicting relationships between them, forming a network. The foundations of graph theory also include an essential element—edge weights, which indicate the strength or probability of interactions. Subgraphs, which represent parts of a graph, allow the identification of specific subsystems within complex networks. Depending on their structure and purpose, graphs can be classified into several key types. Directed graphs have edges with a specified direction, establishing a clear sequence of connections, whereas undirected graphs, also known as ordinary graphs, depict symmetric interactions. Weighted graphs include edge weights that quantify the strength or likelihood of interactions between elements. Tree-like structures, or simply trees, that do not contain cycles, are used in phylogenetics and species classification. Network graphs, characterized by intricate relational connections, are employed to model various biological processes, such as gene regulation networks and metabolic pathways [13].

One of the primary advantages of using graph-based structures in bioinformatics is the ability to graphically represent biological data. Protein interaction networks, where vertices represent proteins and edges depict their interactions, can be effectively utilized to study functional protein groups and predict their roles in cellular processes [14]. Phylogenetic trees illustrate evolutionary relationships among organisms and analyze genetic relatedness, serving as a foundation for studying speciation and adaptation. Metabolic pathways and gene regulation networks facilitate the investigation of biochemical processes within cells by modeling molecular interactions and gene expression regulation. Several powerful tools are available for effective graph analysis and visualization, particularly in working

with biological data. Cytoscape is a popular platform for analyzing biological networks, enabling the integration of various types of biological data and performing complex computations. Gephi is used for processing large-scale data networks, visualizing their structure, and uncovering hidden patterns. GraphViz automates the visualization of graph structures and is widely applied in scientific research. In addition to specialized open-source software platforms, programming language libraries are also actively utilized. For instance, the Python NetworkX library allows for the creation, exploration, and visualization of graph structures while supporting a wide range of algorithms for handling large-scale networks. The BioPython toolkit includes modules for processing biological data, such as genome sequence analysis and phylogenetic tree construction. The integration of these or similar tools into STEM education enables the incorporation of computational methods into the learning process, familiarizing students with modern approaches to analyzing biological systems [13].

3. Opportunities for analyzing biological data using graphs

Graph structures serve as a powerful tool for analyzing complex biological data, as they enable the identification of relationships between different elements within large-scale systems, outline patterns, and model more intricate biological processes. The application of graph algorithms in bioinformatics allows for the analysis of protein interactions, the study of evolutionary relationships, the investigation of disease spread, and the optimization of computational resources for processing large datasets [15, 16].

One of the most common tasks in working with biological networks is finding the shortest paths between vertices, which significantly impacts the analysis of metabolic pathways, where it is necessary to determine optimal biochemical reactions, or the modeling of interactions between genes and proteins. Dijkstra's algorithm is widely used to find the shortest path in weighted graphs and is already applied in research on metabolic transport networks. The A* algorithm, an advanced version of Dijkstra's method, demonstrates efficiency in analyzing large biological graphs as it employs heuristic functions to accelerate the search for optimal paths.

Centrality analysis is another crucial method for studying biological networks, as it helps identify key components within complex systems. The PageRank algorithm, initially designed for ranking web pages, is used in bioinformatics to determine central proteins in interaction networks. Centrality measures help identify vertices that connect different parts of a graph and play a critical role in signal transmission or the regulation of biological processes. Betweenness centrality evaluates vertices based on their ability to act as intermediaries between other vertices, aiding in the analysis of biological pathways where molecules or proteins serve as crucial links between different stages of metabolic or signaling pathways [4].

Graph clustering is a fundamental method for identifying groups of functionally related elements in biological networks. Community detection tools, such as the Louvain algorithm, enable the automatic discovery of modules in biological systems, for example, groups of proteins involved in common metabolic processes [6]. This strategy is also used to detect homologous genes, which are important for understanding evolutionary relationships between species.

The subgraph isomorphism algorithm is applied for comparing substructures within two graphs, making it useful for identifying similar biological processes or structures across different organisms. For instance, comparing genetic or metabolic networks of different species helps uncover evolutionary links or shared biological mechanisms. A related but distinct approach is graph matching algorithms, which compare two graph structures to find the best correspondence between their elements. In research, this algorithm can interpret evolutionary connections between different species, particularly in genome comparisons, where identifying structural similarities between genes or proteins helps reveal analogous evolutionary pathways.

Illustrative graph-based models are actively used for simulating the spread of viruses and epidemics. In such simulations, the graph's vertices may represent infected and healthy individuals, while the edges denote possible contacts between them [17, 18]. Graph algorithms allow researchers to predict disease spread and develop strategies for infection control, a critical task in epidemiology. Identifying key nodes

in biological networks makes it possible to determine essential proteins that regulate fundamental, and sometimes critical, biological processes. For example, in cancer research, graph models can help identify regulatory proteins in cancer cells, contributing to the development of new therapeutic approaches.

Repeatingly, the increasing need to process large volumes of biological data requires the development of more efficient methods for their processing and evaluation. The use of graph algorithms significantly optimizes bioinformatics computations, reducing memory and computational resource requirements. Graph partitioning methods allow large biological networks to be divided into smaller components for simplified analysis. One promising direction is the application of graph databases for storing and processing biological data [19]. Platforms such as Neo4j and GraphDB enable efficient work with biological networks, storing genetic interactions as graph structures and quickly identifying relationships between different biological entities. Large-scale research in bioinformatics, previously inaccessible with traditional relational databases, is becoming increasingly feasible.

Reviewing the use of programming languages for biological data analysis, the modern synergy of bioinformatics and STEM education is actively reflected in Python as one of the most popular tools due to its versatility, a rich set of frameworks, and ease of use. For working with graphs tailored to biological data, the NetworkX and BioPython libraries are particularly useful, providing powerful tools for manipulating graphs and biological information. Mastering these libraries helps both students and researchers apply theoretical knowledge in practice, model complex biosystems, and perform large-scale data analysis.

The NetworkX library is a powerful tool for creating, processing, and visualizing graph structures. It also supports working with various types of graphs, including basic ones: directed, undirected, weighted, and tree-like structures. In the field of biology, NetworkX is extremely useful for analyzing protein interaction networks, genetic regulatory pathways, and metabolic networks. An example of this is the following simple code snippet for creating a protein interaction graph:

```
import networkx as nx
import matplotlib.pyplot as plt
# Creating an empty graph
G = nx.Graph()
# Adding nodes (proteins)
G.add_node("Protein_A")
G.add_node("Protein_B")
G.add_node("Protein_C")
# Adding edges (interactions between proteins)
G.add_edge("Protein_A", "Protein_B", weight=1.2)
G.add_edge("Protein_B", "Protein_C", weight=0.8)
# Visualizing the graph
nx.draw(G, with_labels=True, font_weight="bold")
plt.show()
# Centrality analysis (importance of nodes)
centrality = nx.degree_centrality(G)
print("Centrality of each protein:", centrality)
```

The provided code demonstrates the basics of creating a weighted undirected graph of biological networks, where nodes represent proteins, and edges represent interactions. The code also includes centrality analysis, which helps identify key proteins that play a crucial role in signal transmission or other functions within an organism. The NetworkX library in Python also provides algorithms for finding paths between nodes, analyzing common neighbors, and detecting clusters in the graph. As previously mentioned, these features are essential for bioinformatics tasks such as protein function classification or identifying important metabolic pathways.

BioPython is a powerful library focused on processing and interpreting biological data. This component provides tools for working with biological sequences, such as DNA, RNA, and the aforementioned proteins, as well as for handling genomic and protein databases. The BioPython library is particularly

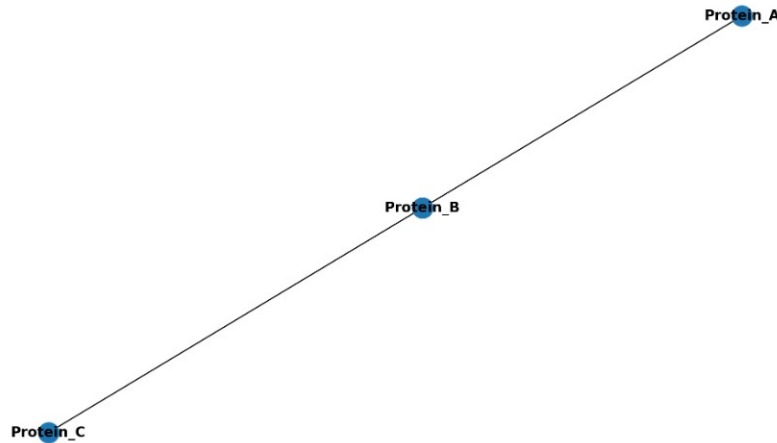


Figure 1: Illustrating the operation of the above code.

useful for extracting specific sequences or searching for particular elements in large genomic databases. Once again, for clarity, here is an example of using BioPython to work with DNA sequences:

```
from Bio import SeqIO
# Loading a DNA sequence from a pre-prepared file
seq_record = SeqIO.read("example_dna.fasta", "fasta")
sequence = str(seq_record.seq) # Extracting the sequence as a string
# Checking for the presence of specific motifs in the sequence
motif = "ATG"
positions = [i for i in range(len(sequence)) if sequence[i:i+len(motif)] == motif]
print(f"Motif '{motif}' found at positions: {positions}")
```

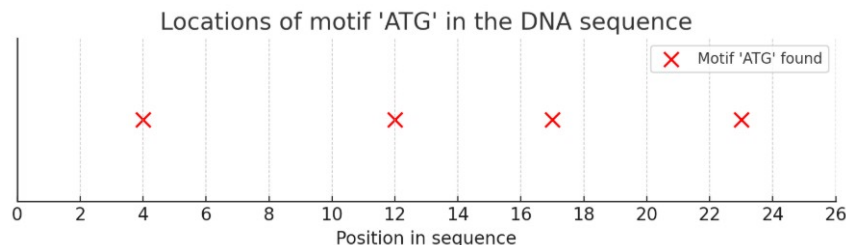


Figure 2: Schematic illustrating the operation of the above code. Red dots indicate positions where the “ATG” motif is found in the DNA sequence.

The lines of code focused on DNA sequence analysis help identify specific motifs that may be crucial for genetic studies. By using this tool, students specializing in the STEM field of biotechnology can learn how to store, analyze, and manipulate biological data for further research or the modeling of biological processes.

One of the key advantages of using Python for bioinformatics is its ability to integrate various libraries for comprehensive data analysis. Thus, it is possible to combine the capabilities of NetworkX for working with graphs and BioPython for handling biological sequences or genomic databases, creating more powerful and versatile tools for analyzing complete biological networks. Summarizing the code examples previously presented, they can be combined into an approach where protein sequences are analyzed to detect specific motifs while protein interactions are represented as a graph. This method enables a comprehensive approach to data analysis and fosters interdisciplinary thinking.

```
import networkx as nx
```



```

import matplotlib.pyplot as plt
from Bio import SeqIO
# Load the protein sequence from a prepared file
seq_record = SeqIO.read("protein_sequence.fasta", "fasta")
sequence = str(seq_record.seq)
# Create a graph for the protein interaction network
G = nx.Graph()
G.add_node("Protein_1")
G.add_node("Protein_2")
G.add_edge("Protein_1", "Protein_2")
# Check for the presence of a peptide motif in the protein sequence
motif = "GTP"
positions = [i for i in range(len(sequence) - len(motif) + 1) if
sequence[i:i+len(motif)] == motif]
print(f"Motif '{motif}' found at positions: {positions}")
# Visualize the protein interaction network
plt.figure(figsize=(5, 5))
nx.draw(G, with_labels=True, node_color='lightblue', edge_color='gray',
node_size=3000, font_size=12)
plt.title("Protein Interaction Network")
plt.show()

```

4. Use of graph theory in STEM education

The integration of graph structures into STEM education opens up new horizons for the effective representation of complex biological processes and the analysis of large volumes of biological data. Graphs are a powerful tool for visualizing and modeling interactions in biological systems, enabling all participants in the learning process to gain a deeper understanding of biological phenomena. The application of graph models in the learning process contributes to the development of computational thinking and an interdisciplinary approach, combining biology, mathematics, and information technologies [20].

Interactive platforms for working with graph objects provide students with the opportunity to analyze biological networks in real time. Modern software tools, such as Cytoscape, Gephi, and GraphViz, allow for the creation and modification of graphs that represent protein interactions, metabolic networks, or phylogenetic trees. Interacting with these graphically represented models helps in better understanding biological patterns and the structural features of complex systems. Graph-based simulations of biological processes allow for the reproduction of the dynamics of biological phenomena, such as the spread of infections, signal transmission between cells, or changes in genetic interactions. Students who have chosen to master the STEM direction of biology can model various scenarios and assess the impact of individual factors on a system, ensuring a deeper and more thorough understanding of the mechanisms of biological processes [9].

The implementation of graphs in bioinformatics, mathematics, computer science, and other related courses significantly expands the boundaries of interdisciplinary research. For example, in bioinformatics courses, graph processing methods can be used for the analysis of genetic sequences and the modeling of metabolic networks, while in discrete mathematics courses, they can be used to study combinatorial and network algorithms. This approach contributes to the comprehensive mastery of natural sciences within STEM education.

Spatial interactions between proteins or the formation of complex branched biological systems can be studied by students through the use of three-dimensional graph visualization. The application of virtual and augmented reality technologies opens up new possibilities for creating learning environments where participants interact with three-dimensional models of biological systems and explore their structure and dynamics [10]. Additionally, the development of analytical thinking and data processing

skills will occur as a result of using new tools. One of the main tasks is the analysis of protein interaction graphs, which allows students to logically identify key proteins involved in cellular processes and study the connections between them, thus gaining a better understanding of cellular functioning mechanisms and identifying potential targets for pharmaceutical developments. The construction of phylogenetic trees using clustering algorithms is an essential component of molecular biology. Students can use graphs to study genetic similarity between species, reconstruct phylogenetic relationships, and explore evolutionary processes. The visualization of metabolic pathways helps learners understand biochemical processes occurring in cells. Implemented graph models allow for the analysis of molecule synthesis and breakdown pathways, identifying key metabolic reactions, and exploring the impact of various factors on metabolism [21]. In summary, the use of graph structures in STEM education significantly improves the learning process, making it more interactive, visual, and research-oriented, while fostering the development of critical thinking, large data analysis skills, and understanding of complex biological systems—competencies necessary for professionals in various fields, including bioinformatics, medicine, and natural sciences.

One of the key areas of integrating graph structures into STEM education is the use of graphs in teaching bioinformatics. Programs from such world-renowned universities as Harvard, the Massachusetts Institute of Technology, and Stanford offer specialized courses that teach students to apply graph algorithms for genomic data analysis, modeling protein interactions, and constructing phylogenetic trees. In these disciplines, students learn pathfinding algorithms in graphs, clustering methods, and centrality analysis to study complex sprawling systems [11]. Interactive laboratories and courses on biological data analysis allow participants in the learning process to work with real datasets used in modern bioinformatics. At Stanford University, students have the opportunity to analyze biological networks using Cytoscape, while the Massachusetts Institute of Technology actively utilizes Python libraries, such as NetworkX and BioPython, to build graph models of metabolic pathways, combining theoretical knowledge with practical skills in programming and large data analysis. In the study of educational modules, work with the real set of information obtained from biological databases is implemented, and clustering algorithms are widely used to identify functionally related protein complexes. The analysis of such graph models allows students to better understand the mechanisms of cellular process regulation and even potential targets for the development of new pharmaceutical drugs.

An example of the use of graph modeling tools in scientific epidemiological experiments is the study of the spread of COVID-19. Research centers around the world used algorithms based on graph structures to model the spread of the virus, analyze contact networks, and predict pandemic waves. The application of methods such as social graph modeling and network node analysis allowed for the evaluation of the effectiveness of quarantine measures and the development of epidemic control strategies [17]. Furthermore, many rare and complex diseases have a genetic nature, and graph-based modeling is actively used for their study. In cancer research, algorithms help identify relationships between gene mutations, determine potential therapeutic targets, and analyze metabolic pathways in tumor cells. Visualizing such data in the form of graphs enables scientists to quickly detect patterns and make informed decisions for the development of new drugs.

5. Prospects and challenges of implementing graphs in STEM education

The use of graph structures in STEM education for the analysis and visualization of biological data is a powerful tool for educational and research purposes. However, as with any new technology, these methods face several challenges that must be addressed for their successful integration into educational programs and research practices.

One of the main obstacles is the complexity of processing large biological graphs. Modern biological systems, such as genetic regulatory networks, protein interaction networks, or metabolic pathways, contain a vast number of elements, complicating their analysis and visualization. Bioinformatic researchers often face the need to solve problems that require significant computational resources, which

can be a problem for educational institutions with limited access to powerful servers and specialized software tools. Additionally, the large volume of data and the complexity of their interactions may lead to student overload if the analysis methods are not presented in an accessible and understandable form. Since the use of graphs in bioinformatics is a relatively new field, educational resources for students, particularly at the early stages of education, are still underdeveloped [16]. In particular, higher educational institutions need to develop a large number of integrated learning materials that combine the theoretical aspects of discrete mathematics with practical tasks in biotechnology and big data analysis.

However, the prospects for the development of graph theory in STEM education appear promising, especially in the context of the growing capabilities of technologies such as artificial intelligence and interactive platforms. Machine learning can significantly ease the processing of large volumes of data, identification of important patterns, and prediction of outcomes. Automated systems that integrate graph models can help students better understand complex biological processes, as well as save time on routine calculations, allowing them to focus on deeper analysis of the results [22]. Modern technologies, such as web platforms and cloud services, already allow students and teachers to work with powerful tools for visualizing and analyzing graph structures. Thanks to the development of powerful platforms such as Cytoscape, Gephi, and GraphDB, students will find it easier to work with real biological networks, and the interactivity of these tools will greatly improve student engagement in the learning process. Moreover, scientific research and bioinformatics laboratories can use these platforms for quick access to data and real-time interaction with it. An inherently important area of development is the creation of educational programs using virtual and augmented reality for modeling biological processes. Virtual and augmented reality technologies have great potential for visualizing complex biological systems and processes that are difficult to grasp using traditional teaching methods. Using these technological tools will allow the creation of three-dimensional models of biological networks and interactions, enabling students to "be inside" these processes and better understand their dynamics. Such programs can be integrated into various educational courses in bioinformatics, genetics, and other STEM disciplines, significantly improving learning effectiveness and increasing student motivation.

Comparing the capabilities of graph-based methods with traditional approaches to biological data analysis, one can highlight their key advantage in effectively representing relationships in complex systems. Traditional statistical methods often focus on identifying patterns in numerical data, while graph-based models allow for a visual assessment of the system's structure and the identification of critical components in biological networks. In particular, pharmacological advancements in the use of graph algorithms serve as an example, as they not only predict potential future interactions between molecules but also analyze the impact of potential drugs on cellular processes.

6. Conclusions

The use of graph structures in STEM education has enormous potential for developing analytical and visualization skills, especially in the context of bioinformatics. A tool that effectively models complex biological systems opens new horizons for learning and research. Due to their ability to reflect interactions between system elements, graphs serve as a foundation for analyzing biological data such as genetic networks, protein interactions, and metabolic pathways. The use of graph models generates a number of advantages over traditional approaches to teaching biological sciences. In particular, the work not only with static descriptions of biological processes but primarily the creation of dynamic models, which thoroughly demonstrate the interactions between elements of the system, stands out. Compared to text descriptions or linear tables, graphs provide visualization, which is crucial for better mastering the learning topics, processing large volumes of information, and discovering hidden patterns.

We do not exclude that, in STEM education, other approaches to visualizing biological data are currently used, mostly analytical models in spreadsheets, simulations in the MATLAB software environment, or integrated platforms for visualizing molecular structures, such as PyMOL [19]. However, there are certain limitations affecting these methods, namely: tabular models are used for simplified

calculations and, therefore, do not always include functionality for easy analysis of connections between elements; with regard to simulation platforms, which are focused on specific types of data, their use is limited for heterogeneous biological networks; tools for molecular modeling work mostly with 3D visualization of individual molecules, rather than with complex interactions between biological systems [18].

However, amidst the implementation of graph-based methods in curricula, several significant challenges arise. The complexity of processing large amounts of data requires powerful computational resources. Even considering modern algorithms for analyzing biological networks, working with large graphs containing hundreds or even thousands of elements can be computationally expensive, necessitating the training in computational optimization methods. A problem remains with limited access to resources for educational institutions. Many educational establishments lack powerful server complexes or access to licensed software applications that facilitate work with large biological data sets. It is necessary to properly adapt methods and tools for students, especially in the early stages of education, since working with graphs requires basic knowledge of discrete mathematics concepts, algorithmic thinking, and programming skills, which may become a barrier for educational programs with a biological focus. Currently, there is a lack of specialized learning materials and standardized courses on topics like graph analysis, which integrate elements of discrete mathematics into bioinformatics education within STEM education frameworks.

On the other hand, integrating graph structures into educational programs opens significant prospects associated with technological advancements. Specifically, the use of artificial intelligence and machine learning for the automatic analysis of large biological graphs allows significantly reducing the burden on students, freeing up their time for more in-depth study of the material. Interactive platforms such as Cytoscape or Gephi also provide access to real biological data and allow working with them in a convenient and visual format. Virtual and augmented reality technologies offer new opportunities for modeling complex biological processes, which were previously difficult to achieve using traditional teaching methods. Participants in the learning process will have the opportunity to work with interactive learning laboratories, where they will analyze biological processes in dynamics. To work with large biological networks in real-time, graph databases, such as Neo4j, can be used to create educational platforms.

In conclusion, the integration of graph structures into STEM education is an important step towards enhancing the effectiveness of learning in bioinformatics, genetics, and other sciences. However, to fully realize this potential, several technical and methodological issues must be addressed. Successfully overcoming these challenges will contribute to the creation of more effective and innovative educational programs, providing students with a deep understanding of complex biological and scientific processes. For the successful implementation of graph structures in STEM education, it is essential to advocate for the development of specialized courses that combine biology, graph theory, and programming. Adapting existing platforms, such as Cytoscape, NetworkX, and BioPython, to educational needs will support the creation of teaching materials and open research resources. The use of graph databases will allow the integration of real biological data, available in the public domain, into the educational project process. A highly promising direction is the development of educational programs using virtual and augmented reality to visualize biological processes in real-time. Further research should focus on assessing the effective impact of integrating graph structures into the educational process through experimental methods and developing new solutions for the integration of graph analytics into educational technologies.

Funding

This research received no external funding.

Data Availability Statement

No new data were created or analysed during this study. Data sharing is not applicable.

Conflicts of Interest

The author declares no conflict of interest.

Declaration on Generative AI

During the preparing this work, the author used GPT-4 and LanguageTool to: Translate text into English and check grammar and spelling. After using these tools/services, the author reviewed and edited the content and takes full responsibility for the content of the publication.

References

- [1] Y. Song, Q. Yuan, S. Chen, Y. Zeng, H. Zhao, Y. Yang, Accurately predicting enzyme functions through geometric graph learning on ESMFold-predicted structures, *Nature Communications* (2024). doi:10.1038/s41467-024-52533-w.
- [2] Y. Yang, Y. Cui, X. Zeng, Y. Zhang, M. Loza, S.-J. Park, K. Nakai, STAIG: Spatial transcriptomics analysis via image-aided graph contrastive learning for domain exploration and alignment-free integration, *Nature Communications* 16 (2025). doi:10.1038/s41467-025-56276-0.
- [3] J. Perna, Possibilities and Challenges of Using Educational Cheminformatics for STEM Education: A SWOT Analysis of a Molecular Visualization Engineering Project, *Journal of Chemical Education* 99 (2022) 1190–1200. doi:10.1021/acs.jchemed.1c00683.
- [4] I. Singh, K. Agarwal, S. Ganapathy, Advancing allergy source mapping: A comprehensive multi-disciplinary framework integrating machine learning, graph theory and game theory, *Applied Soft Computing* 166 (2024). doi:10.1016/j.asoc.2024.112147.
- [5] M. Koutrouli, E. Karatzas, D. Pérez-Espino, G. A. Pavlopoulos, A Guide to Conquer the Biological Network Era Using Graph Theory, *Frontiers in Bioengineering and Biotechnology* 8 (2020). doi:10.3389/fbioe.2020.00034.
- [6] Z. Wu, H. Ma, B. Deng, Z. Li, L. Chang, Co-guided Dual-channel Graph Neural Networks for the prediction of compound–protein interaction, *Applied Soft Computing* 163 (2024). doi:10.1016/j.asoc.2024.111875.
- [7] H. Li, X. Wei, A Concise Review of Biomolecule Visualization, *Curr Issues Mol Biol* (2024) 1318–1334. doi:10.3390/cimb46020084.
- [8] T. Freeman, S. Horwell, A. Patir, J. Harling-Lee, T. Regan, B. B. Shih, J. Prendergast, D. Hume, T. Angus, Graphia: A platform for the graph-based visualisation and analysis of high dimensional data, *PLoS Computational Biology* 18 (2022). doi:10.1371/journal.pcbi.1010310.
- [9] B. Ristevski, S. Savoska, Z. Savoski, Using Graph Databases for Portraying and Analysing Biological and Biomedical Networks, *International Conference on Control, Decision and Information Technologies* 1 (2022) 565–568. doi:10.1109/CoDIT55151.2022.9804139.
- [10] Y. Wu, L. Hu, Y. Wang, Signed attention based graph neural network for graphs with heterophily, *Neurocomputing* 557 (2023). doi:10.1016/j.neucom.2023.126731.
- [11] V. Ruf, Graphical representations of data in STEM education (2024). doi:10.5282/edoc.34588.
- [12] A. T. Khanam, The Role of Graph-Based Data Science Tools in Uncovering Complex Network Relationships 1 (2024) 29–36. doi:10.70849/ijsci27935.
- [13] S. Shen, L. Deng, Y. Du, J. Gao, C. Zhang, Y. Wang, Z. Shen, Y. Li, X. Chen, H. Chen, Analyzing and mapping the research status, hotspots, and frontiers of biological wound dressings: An in-depth data-driven assessment, *International Journal of Pharmaceutics* 629 (2022). doi:https://doi.org/10.1016/j.ijpharm.2022.122385.

- [14] M. Kolomeets, V. Desnitsky, I. Kotenko, A. Chechulin, Graph Visualization: Alternative Models Inspired by Bioinformatics, *Sensors* (Basel, Switzerland) 23 (2023). doi:10.3390/s23073747.
- [15] V. Kolesnikov, Visualizing graph data structures: enhancing understanding in data structures education, *Transactions of Kremenichuk Mykhailo Ostrohradskyi National University* (2024). doi:10.32782/1995-0519.2024.2.7.
- [16] C. O. Yu., A. O. S., Comparison of graph and relational databases, *Twenty-first all-Ukrainian conference of students and young scientists in computer science, information systems and technologies* 21 (2024) 31–32.
- [17] E. C. Snell-Rood, D. Smirnoff, H. Cantrell, K. Chapman, E. Kirscht, E. Stretch, Bioinspiration as a method of problem-based STEM education: A case study with a class structured around the COVID-19 crisis, *Ecology and Evolution* 11 (2021) 16374–16386. doi:<https://doi.org/10.1002/ece3.8044>.
- [18] S. J. Kersting, A. L. Kühn, M. Fischer, Measuring 3D tree imbalance of plant models using graph-theoretical approaches, *Ecological Informatics* 80 (2024). doi:10.1016/j.ecoinf.2023.102438.
- [19] Y. Qiu, J. Lu, A visualization algorithm for medical big data based on deep learning, *journal=Measurement* 183 (2021). doi:10.1016/j.measurement.2021.109808.
- [20] Y. Dariia, STEM approach to learning discrete mathematics through practical tasks based on graph structures, *Modern digital technologies and innovative teaching methods: experience, trends, prospects: materials of the XIV International Scientific and Practical Internet Conference* 14 (2024) 231–233.
- [21] J. Yang, Z. Yu, S. Li, Y. Cao, J. Hu, J. Xu, GGT-SNN: Graph learning and Gaussian prior integrated spiking graph neural network for event-driven tactile object recognition, *Information Sciences* 677 (2024). doi:10.1016/j.ins.2024.120998.
- [22] J. Chem, Possibilities and Challenges of Using Educational Cheminformatics for STEM Education: A SWOT Analysis of a Molecular Visualization Engineering Project, *Journal of Chemical Education* (2022) 1190–1200. doi:10.1021/acs.jchemed.1c00683.