Journal of Honai Math, 9(2), 349-366, August 2026
https:/doi.org/10.30862/jhm.v9i2.1257

Graph-based similarity measures in patient networks: A qualitative review of structural properties and applications to personalized

Athraa Talib Breesam *

College of Biomedical Engineering, University of Technology, Baghdad, Iraq

*Correspondence: athraatalibbreesam@researchpublication.co

Received: 5 May 2026 | Revised: 17 August 2026 | Accepted: 18 August 2026 | Published: 31 August 2026
© The Authors 2026

Abstract

The increasing volume and heterogeneity of healthcare data have created challenges in identifying meaningful relationships among patients and translating complex clinical information into personalized healthcare decisions. Graph-based approaches provide a relational framework for representing patients and their clinical similarities, enabling patterns that may be difficult to identify through conventional analytical approaches to be examined systematically. This study aims to examine how graph-based similarity measures are used to represent patient relationships, identify clinically meaningful patterns, and support personalized healthcare. A qualitative secondary study was conducted by reviewing peer-reviewed literature published since 2018 and indexed in Scopus, Web of Science, and PubMed. Relevant studies were examined based on their relevance, clarity, and contribution to graph-based patient networks, similarity measures, and personalized healthcare, and the findings were synthesized thematically. The review identified four major contributions of graph-based approaches: patient clustering and pattern discovery, risk prediction and disease progression analysis, treatment selection and clinical decision support, and personalized patient monitoring. The literature also indicates that integrating heterogeneous data, including clinical histories, laboratory results, symptoms, treatment responses, imaging, and genetic information, can provide richer patient representations and strengthen similarity-based analyses. However, practical implementation remains constrained by data incompleteness and inconsistency, privacy and security requirements, limited interoperability, and the technical complexity of integrating multimodal healthcare data. Overall, graph-based similarity measures provide a promising framework for understanding patient variability and supporting data-driven personalized healthcare. Future research should prioritize empirical validation, longitudinal patient networks, explainable graph-based models, multimodal data integration, and privacy-preserving approaches to establish their clinical utility and reliability across diverse healthcare settings.

Keywords: Graph-based Similarity Measures; Graph Neural Networks (GNNs); Patient Networks; Personalized Healthcare; Precision Medicine

Introduction

The volume and heterogeneity of patient data generated within contemporary healthcare systems have increased substantially. Hospitals, clinics, and digital health platforms routinely collect information on symptoms, medical histories, laboratory results, diagnoses, treatments, lifestyle factors, and genetic profiles. Although these data provide valuable information for clinical care, their increasing complexity and distribution across heterogeneous sources make it difficult to identify meaningful relationships among patients, detect clinically relevant patterns, and support reliable decision-making. Conventional analytical approaches often examine individual variables or isolated clinical indicators, which may overlook interactions among multiple patient characteristics. Consequently, potentially informative relationships among patients may remain undetected, limiting the ability of healthcare providers to develop comprehensive and individualized care strategies. Graph theory provides a mathematical framework for representing and analyzing such complex relationships (Schrodt et al., 2020).

Graph-based approaches have increasingly been applied to model complex relationships in healthcare and biomedical research (Paul et al., 2024). In a patient network, patients can be represented as nodes, while edges encode relationships based on shared characteristics, such as symptoms, risk factors, diagnoses, genetic markers, or treatment responses. This representation transforms heterogeneous patient information into a structured relational network in which similarities and patterns can be analyzed systematically. Graph-based methods have consequently been applied to diverse healthcare problems, including drug discovery, epidemiology, disease modeling, and personalized medicine (Sowriraja et al., 2024). By representing relationships explicitly, patient networks can facilitate the identification of patient clusters, disease patterns, treatment-response profiles, and potentially emerging health risks. In addition, graph databases have attracted increasing attention for representing complex disease relationships because of their ability to manage highly connected information with greater flexibility than conventional relational databases (Woodman et al., 2024).

At the center of graph-based patient networks are similarity measures, which determine the degree of resemblance between patients and consequently influence how edges are established within a network. Similarity can be computed from individual features or, more comprehensively, from multiple heterogeneous data sources. This capability is particularly relevant in healthcare, where patient characteristics are multidimensional and may include clinical records, physiological measurements, behavioral information, imaging data, and genomic profiles. For example, in Parkinson's disease research, the integration of wearable sensor data with personal health records has been investigated as a means of monitoring patients and generating clinically relevant alerts (Zafeiropoulos et al., 2024). By quantifying similarities across multiple patient characteristics, graph-based methods can facilitate patient matching, identify clinically meaningful subgroups, and provide evidence for more individualized

treatment decisions. Patients with similar clinical profiles, for instance, may exhibit comparable treatment responses, although such similarity should be interpreted as supportive evidence rather than as a direct basis for transferring treatment decisions from one patient to another.

Electronic health records (EHRs) further expand the potential of graph-based approaches because they provide longitudinal and multidimensional information that can support both clinical assessment and evidence-based decision-making (Boll et al., 2024). However, the effective use of these data remains challenging because healthcare information is often fragmented across systems, varies in quality and completeness, and represents patients at different levels of clinical detail. Such limitations can hinder the identification of patient subgroups and reduce the reliability of prediction and decision-support systems. Recent research has therefore explored the integration of graph theory and machine learning to represent structural relationships within complex medical data and improve diagnostic and predictive performance (Balamurugan et al., 2024). Nevertheless, the growing diversity of graph-based methods also creates a methodological challenge: different similarity measures, graph-construction strategies, and network architectures may capture substantially different aspects of patient relatedness.

This methodological diversity highlights the need for a systematic mathematical examination of graph-based similarity measures. Although existing studies demonstrate the potential of patient similarity networks, graph neural networks (GNNs), knowledge graphs, and related network-based models, the underlying similarity concepts are often discussed within specific clinical applications rather than comparatively from a mathematical perspective. In particular, there remains a need to clarify how different similarity measures represent patient relationships, what structural properties they induce in patient networks, and how these properties influence their applicability to personalized healthcare. A mathematically informed synthesis can therefore help bridge the gap between the formulation of similarity relationships and their practical use in clinical decision support.

Accordingly, this study reviews graph-based similarity measures and their applications to patient networks for personalized healthcare. It addresses the following research question: What value can be realized from graph-based similarity measures in discovering meaningful patient patterns that support accurate and personalized healthcare decisions? The review focuses on the mathematical and structural foundations of patient similarity, including how patient relationships are represented, how similarity measures contribute to network construction and analysis, and how graph-based approaches have been applied to clinical prediction, patient stratification, treatment selection, and disease monitoring. By bringing these perspectives together, this study aims to provide an organized conceptual framework for understanding the role of graph-based similarity measures in analyzing large and heterogeneous patient datasets and supporting data-driven personalized healthcare.

Literature Review

Patient Network Concepts

Patient networks provide a graph-based framework for representing and analyzing relationships among individuals with shared clinical or health-related characteristics. Their development is related to broader traditions of network analysis in sociology, statistics, and health research, although the application of network methods to healthcare systems and patient-level relationships remains an evolving area of research (DuGoff et al., 2018). In a patient network, individuals are represented as nodes, whereas edges represent relationships defined by similarities in characteristics such as symptoms, diagnoses, medical histories, laboratory results, medications, or genetic profiles. This representation enables researchers and healthcare professionals to examine patient relationships collectively rather than analyzing each patient as an isolated observational unit.

The relational structure of a patient network can reveal patterns that may be difficult to identify when clinical variables are examined independently. For example, patients sharing similar combinations of symptoms, diagnoses, or clinical characteristics may form densely connected groups, potentially indicating clinically meaningful subpopulations or disease phenotypes. Graph-structured information is increasingly used as a fundamental representation for machine learning and deep learning tasks because it explicitly preserves relationships among data entities (Khemani et al., 2024). In healthcare, this relational representation can therefore complement conventional statistical approaches by enabling the analysis of dependencies and interactions among patients and their characteristics. Such an approach is particularly relevant to patient-centered healthcare, in which understanding individual differences and patient needs is an important component of healthcare quality (Langberg et al., 2019).

An important advantage of patient networks is their potential to support patient stratification for precision medicine. Patients can be grouped according to similarities in biological characteristics, disease progression, comorbidities, or treatment responses, thereby providing a basis for identifying relatively homogeneous patient subgroups (Pai et al., 2018). Such stratification can help researchers and clinicians investigate whether patients with similar profiles exhibit comparable clinical trajectories or responses to particular interventions. This is especially relevant for chronic diseases and complex conditions involving multiple interacting risk factors, where conventional classifications based on individual variables may not adequately capture patient heterogeneity.

Patient networks can also support the analysis of disease progression and the identification of clinically relevant patterns over time. By examining relationships among patients with similar trajectories or characteristics, researchers can investigate the coexistence of symptoms, patterns of comorbidity, and potential risk groups. In this respect, graph-based representations provide a structured way to integrate heterogeneous information and reduce the conceptual fragmentation of patient data. Recent developments in graph-based and probabilistic representation learning further demonstrate the potential of network-based approaches for

identifying distinct health-management groups within large patient populations (Ding, 2025). Thus, patient networks provide a foundation for transforming heterogeneous patient data into relational structures that can subsequently be analyzed using mathematical, statistical, and computational methods.

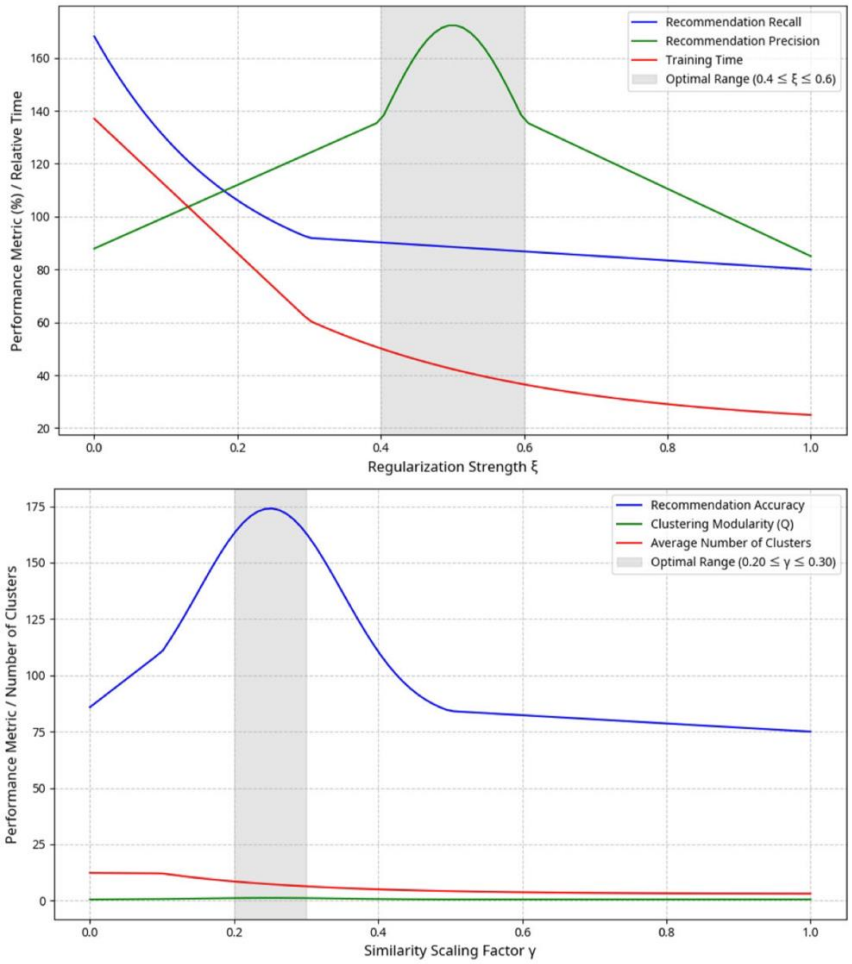


Figure 1. Probabilistic representation learning for identifying distinct health-management groups in a large-scale elderly population (Ding, 2025)

Similarity Measures in Healthcare

Similarity measures constitute a fundamental component of patient-network construction because they determine the degree of resemblance between pairs of patients and, consequently, influence whether and how edges are established between nodes. In mathematical and computational terms, a similarity measure assigns a value representing the degree to which two entities share relevant characteristics. In healthcare applications, these characteristics may include symptoms, laboratory results, medications, diagnoses, genetic information, lifestyle factors, or longitudinal clinical histories. The resulting similarity scores can be used to construct weighted or unweighted patient networks in which stronger similarities correspond to stronger or more significant relationships.

The choice of similarity measure is particularly important because different measures capture different aspects of patient relatedness. Similarity may be calculated from individual variables, combinations of heterogeneous features, or representations learned from complex clinical data. This distinction is important because healthcare data are multidimensional and frequently heterogeneous, meaning that patients may be similar in some dimensions while substantially different in others. The increasing availability of computable electronic medical records and population-level genomic information has consequently expanded opportunities to identify clinically meaningful patient relationships that may not be apparent through conventional assessment alone (Pai et al., 2018). Similarity-based analysis can therefore provide a more structured representation of multidimensional patient profiles and help reveal latent relationships within large datasets.

Similarity measures are also widely used in clinical information processing beyond direct patient-to-patient comparison. For example, semantic similarity measures have been applied to medical terminology for tasks such as terminology disambiguation, prescription analysis, medical question answering, and clinical decision support (Zhang et al., 2021). In patient networks, analogous principles can be applied to identify groups of patients with comparable clinical profiles. Patients with high similarity may subsequently be analyzed as a subgroup to investigate common disease trajectories or treatment responses. However, similarity should not be interpreted as evidence that patients are clinically interchangeable; rather, it provides a quantitative basis for identifying relationships that warrant further clinical or statistical investigation.

The identification and prioritization of patient similarity are therefore important components of personalized healthcare and clinical decision-support systems (Suo et al., 2018). Similarity measures can contribute to clustering algorithms, recommendation systems, predictive models, and patient-stratification methods by providing a formal mechanism for comparing individuals. Their effectiveness, however, depends on the characteristics and quality of the underlying data, the representation of clinical features, and the mathematical properties of the selected similarity metric. Electronic medical records provide an increasingly important source for these applications because they enable the collection and accumulation of large volumes of digitized clinical information (Guo & Chen, 2023). Consequently, understanding the mathematical basis and structural implications of similarity measures is essential for evaluating their suitability for patient-network construction and healthcare applications.

Personalized Care Applications

Personalized healthcare is based on the recognition that patients with the same broad diagnosis may differ substantially in their biological characteristics, disease trajectories, risk profiles, and responses to treatment. Graph-based similarity analysis provides one approach for representing this heterogeneity by identifying patients who share relevant clinical characteristics while distinguishing those whose profiles differ substantially. When combined with network analysis, similarity measures can therefore support patient stratification and provide a data-driven basis for investigating individualized care strategies.

One potential application is treatment-response analysis. For example, a patient network may identify a subgroup of patients who share relevant genetic, clinical, or phenotypic characteristics and exhibit similar responses to a particular intervention. Such information can provide supportive evidence when evaluating treatment options for a new patient with a comparable profile. Importantly, similarity-based evidence should complement rather than replace clinical judgment, since treatment effectiveness may depend on characteristics that are not represented in the network. Nevertheless, the ability to identify clinically relevant patient subgroups can potentially reduce reliance on population-level averages and support more targeted treatment evaluation.

Graph-based similarity analysis can also contribute to early risk identification. When patients are grouped according to shared symptoms, risk factors, or longitudinal trajectories, network analysis can reveal clusters associated with elevated risk or emerging disease patterns. The development of wearable, implantable, and other digital health technologies has further expanded the availability of continuous patient-generated data that can be incorporated into such analyses (Siddiqui et al., 2022). These data can provide additional dimensions for measuring patient similarity and may enable healthcare systems to identify changes in patient states before clinically significant complications emerge.

Individualized care can also benefit from longitudinal patient-network analysis. Digital health and mobile health (mHealth) systems enable the collection, analysis, and use of individual-level health data for monitoring and personalized intervention (Lee et al., 2022). When longitudinal information is represented within a graph-based framework, patients can be compared according to trajectories of disease progression, treatment response, or changes in health status. Such comparisons may help researchers investigate common clinical pathways and support the development of monitoring strategies for patients with similar profiles. These applications are particularly relevant to chronic disease management, oncology, and rare-disease research, where substantial heterogeneity among patients can influence treatment outcomes.

Personalized medicine increasingly relies on integrating diagnostic information from multiple sources to identify interventions that are most appropriate for individual patients (Hassan et al., 2022). In this context, graph-based similarity measures offer a mechanism for integrating heterogeneous patient information and translating it into relational structures that can be analyzed systematically. Their potential applications include patient stratification, treatment-response prediction, disease-risk assessment, and longitudinal monitoring. However, the clinical value of these applications depends not only on the accuracy of the similarity measure but also on the quality, completeness, and representativeness of the underlying patient data. Thus, graph-based similarity should be understood as a computational and mathematical foundation for personalized healthcare rather than as a standalone clinical decision-making mechanism.

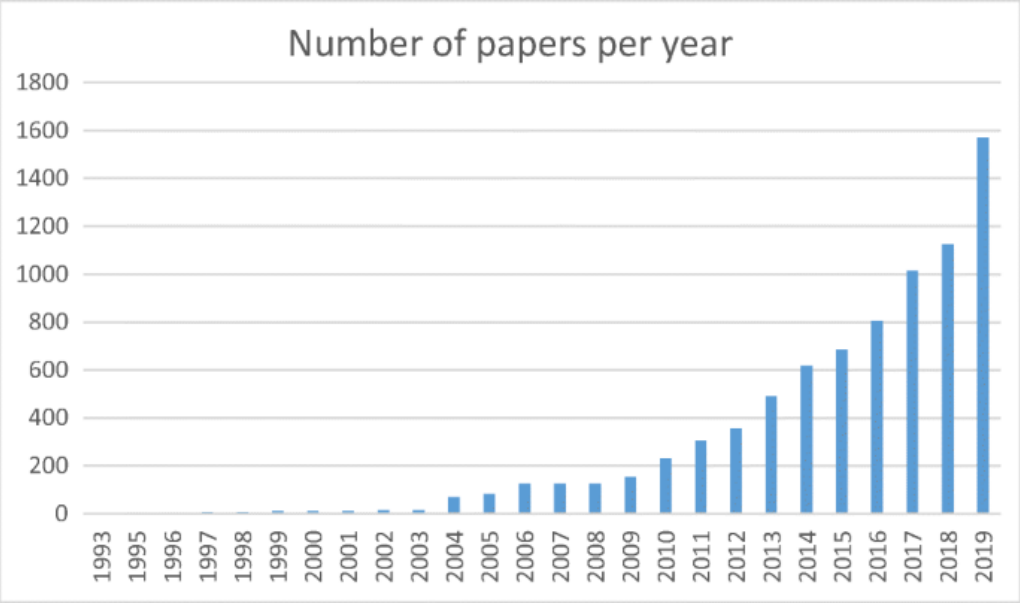


Figure 2. Development of publications using the keyword "Electronic health record" through 2019 (Schrodt et al., 2020)

Methods

This study employed a qualitative literature review to synthesize existing research on graph-based patient networks, similarity measures, and their applications in personalized healthcare. Rather than collecting primary data from patients or healthcare institutions, the study relied on peer-reviewed academic literature to examine how graph-based approaches have been conceptualized, developed, and applied in healthcare contexts. A qualitative review approach was considered appropriate because the primary objective was to identify and synthesize recurring concepts, methodological approaches, and application patterns across the literature rather than to estimate a pooled statistical effect.

Literature Search and Selection

Relevant literature was identified through searches of three major academic databases: Scopus, Web of Science, and PubMed. These databases were selected because they provide broad coverage of research in mathematics, computer science, biomedical sciences, and healthcare. The search focused primarily on peer-reviewed publications addressing graph-based patient networks, patient similarity measures, graph-based healthcare analytics, and personalized or precision medicine. Publications from 2018 onward were prioritized to capture recent methodological and technological developments in graph-based healthcare research.

The selection process considered the relevance of each publication to the research focus, the clarity of its methodological or conceptual contribution, and its potential contribution to understanding graph-based similarity in healthcare. Studies were considered relevant when they addressed at least one of the following aspects: (1) construction or analysis of patient networks,

(2) mathematical or computational approaches for measuring patient similarity, (3) graph-based representations of clinical or biomedical data, or (4) applications of similarity-based networks to patient stratification, prediction, treatment selection, disease monitoring, or personalized healthcare.

Studies that did not substantially address graph-based patient relationships or similarity-based analysis in healthcare were excluded from the synthesis. Duplicate publications and sources that did not provide sufficient information to establish their relevance to the research questions were also excluded.

Data Extraction and Thematic Synthesis

The selected literature was analyzed thematically to identify recurring concepts and relationships across the reviewed studies. The analysis focused on three principal themes: (1) patient-network construction and representation, (2) similarity measures and their role in establishing relationships among patients, and (3) applications of graph-based similarity analysis in personalized healthcare. Within these themes, attention was given to the types of patient data used, the principles underlying similarity computation, the resulting network structures, and the healthcare problems addressed by each approach.

The identified themes were subsequently compared and synthesized to develop a conceptual framework linking patient data, similarity measures, graph construction, network analysis, and personalized healthcare applications. This synthesis was used to examine not only the reported applications of graph-based approaches but also the structural and methodological implications of different approaches to defining patient similarity. The resulting framework provides a basis for understanding how mathematical representations of patient similarity can contribute to the analysis of heterogeneous healthcare data and support data-driven clinical decision-making.

Trustworthiness of the Review

To enhance the trustworthiness of the synthesis, findings were compared across studies and across the three databases rather than being derived from a single source or research group. Particular attention was given to consistency in the reported roles of similarity measures, graph-based representations, and healthcare applications. The synthesis also distinguished between methodological potential and demonstrated clinical application to avoid interpreting computational similarity as direct evidence of clinical equivalence or treatment effectiveness.

Results

Identifying Meaningful Patient Patterns through Graph-Based Similarity

The synthesis of the reviewed literature indicates that graph-based similarity measures provide a useful framework for identifying meaningful patterns within patient populations. In patient similarity networks, patients are represented as nodes, while edges represent similarities based

on shared clinical, demographic, genetic, or behavioral characteristics. This representation enables the identification of patient clusters that may share similar symptoms, disease profiles, risk factors, or treatment responses. Unlike conventional approaches that often examine clinical variables independently, graph-based representations preserve the relationships among patients and therefore provide an alternative means of identifying latent patterns within complex healthcare data.

The reviewed studies further indicate that graph-based representations can support the analysis of complex biomedical relationships, including molecular interactions, signaling pathways, disease comorbidities, and healthcare systems (Li et al., 2022). Such representations are particularly valuable when patient information is distributed across multiple data sources. By integrating laboratory results, medical histories, genetic profiles, imaging data, and lifestyle characteristics, graph-based approaches can represent multiple dimensions of patient similarity within a unified structure. This multidimensional representation may provide a more comprehensive basis for patient stratification than approaches based on individual variables.

Graph-based methods have also been applied to disease-specific problems. For example, graph-based approaches have been used to investigate complex interactions underlying liver diseases across multiple organizational levels (Hu et al., 2024). Similarly, graph and deep learning approaches have been explored for complex biomedical prediction tasks, including cancer diagnosis and treatment (Waqas, 2024). These applications demonstrate the broader potential of graph-based representations for capturing relationships that are difficult to represent using conventional tabular approaches.

Supporting Risk Prediction and Clinical Decision-Making

A second theme emerging from the literature concerns the use of graph-based similarity measures for risk prediction and clinical decision support. Patient similarity networks allow newly observed patients to be compared with previously characterized patients or patient subgroups. Such comparisons can provide additional information for risk stratification, disease progression analysis, and treatment selection. The relational structure of the network enables clinical information to be interpreted not only at the individual level but also in relation to patterns observed among similar patients.

Several studies suggest that graph-based approaches can improve the representation of disease-related relationships and support predictive modeling. Knowledge discovery in health informatics increasingly requires methods capable of processing large and heterogeneous medical datasets (Rostami et al., 2023). Graph-based approaches address this requirement by representing interconnected clinical entities and their relationships within a common analytical structure. Network reconstruction methods similarly provide a means of uncovering complex interrelationships within biomedical and healthcare systems (Singhal et al., 2025).

The reviewed literature also suggests that graph-based patient representations can contribute to the identification of patients at elevated risk. Patients who share combinations of characteristics with previously identified high-risk groups can potentially be prioritized for additional monitoring or assessment. However, the evidence should not be interpreted as

demonstrating that graph-based methods universally outperform conventional approaches. Their effectiveness depends on the quality of the underlying data, the method used to construct the network, and the clinical context in which the model is applied.

Supporting Personalized Healthcare

A third theme concerns the potential of graph-based similarity measures to support personalized and precision healthcare. Patient similarity networks enable clinicians and researchers to identify individuals or subgroups with comparable health profiles and to use information from these groups as contextual evidence for clinical decision-making. This is particularly relevant when patients with the same broad diagnosis may nevertheless differ substantially in disease progression, risk factors, or treatment response.

The literature indicates that similarity-based patient grouping can support treatment recommendation and risk stratification by identifying clinically relevant subgroups. For example, patients with similar disease trajectories or clinical characteristics can be analyzed collectively to identify patterns that may inform treatment decisions. Such approaches may reduce reliance on population-level averages and provide additional evidence for individualized care. However, similarity should not be interpreted as clinical equivalence because patients may differ in characteristics that are not represented in the network.

Graph-based approaches may also support longitudinal monitoring. By incorporating temporal information, patient networks can represent changes in health status and disease progression over time. This is particularly relevant for chronic disease management, where patient conditions and treatment responses evolve continuously. The integration of electronic health records with wearable and other digital health data may further increase the temporal and contextual richness of patient representations (Zafeiropoulos et al., 2024).

Scalability and Integration of Heterogeneous Healthcare Data

Another recurring finding is the potential of graph-based methods to integrate heterogeneous healthcare data within a single analytical framework. Modern healthcare systems generate information from electronic health records, laboratory systems, medical imaging, genomic databases, wearable devices, and patient-generated health data. Graph-based representations can accommodate different types of entities and relationships, thereby providing a flexible structure for integrating these data sources.

Recent studies have demonstrated the application of graph-based learning to electronic health records, medical imaging, disease prediction, and patient risk stratification (Boll et al., 2024; Balamurugan et al., 2024; Sharma et al., 2025; Xu et al., 2024). The reported findings suggest that graph-based models can capture structural and relational information that may not be adequately represented by conventional machine-learning approaches. This capacity is particularly relevant for healthcare environments in which clinical information is heterogeneous, interconnected, and continuously updated.

Nevertheless, scalability should not be interpreted solely in computational terms. Large-scale implementation also requires interoperable data infrastructures, standardized

representations, secure data-sharing mechanisms, and appropriate computational resources. Therefore, although graph-based methods offer a promising framework for large-scale healthcare analytics, their practical implementation depends on the broader technical and institutional environment.

Discussions

The findings synthesized from the reviewed literature indicate that the principal value of graph-based similarity measures lies in their ability to represent healthcare data as a relational structure. Patients, symptoms, diseases, treatments, laboratory findings, and other clinical entities are inherently interconnected. Traditional statistical approaches can capture associations among variables, but graph-based methods provide an explicit representation of relationships among multiple entities. This characteristic makes graph-based models particularly suitable for healthcare applications in which the relationships among patients and clinical entities are themselves informative.

The findings are consistent with the broader development of graph-based artificial intelligence, particularly graph neural networks (GNNs), patient similarity networks, and knowledge graphs. Paul et al. (2024), for example, demonstrated the potential of GNN-based approaches to learn representations from interconnected medical data for tasks such as disease diagnosis, risk stratification, and treatment-response modeling. Similarly, Schrodt et al. (2020) highlighted the value of representing healthcare relationships through network structures to facilitate the identification and interpretation of patterns. Taken together, these studies suggest that graph-based approaches extend conventional patient-level analysis by incorporating relational information into the representation of patient profiles.

A particularly important implication is the role of multimodal data integration. Woodman et al. (2024), Ding (2025), and Li et al. (2022) illustrate how clinical, genetic, laboratory, imaging, and other forms of information can be represented within integrated graph structures. Such integration is important because patient similarity is rarely determined by a single characteristic. Two patients may be similar in terms of diagnosis but differ in genetic profile, disease progression, treatment response, or comorbidities. A graph-based framework can potentially represent these multiple dimensions simultaneously. Moreover, the incorporation of temporal information may further improve patient representation because similarity can change as patients' health conditions evolve over time (Ganzinger et al., 2019).

These findings have direct implications for personalized healthcare. Pai et al. (2018) emphasized the importance of identifying meaningful similarities and differences among patients for advancing precision medicine. Graph-based similarity measures operationalize this principle by positioning an individual patient within a broader population of clinically comparable patients. In principle, clinicians can use information from similar patients as additional evidence when considering treatment options, monitoring strategies, or risk assessments. Importantly, however, this information should complement rather than replace clinical expertise and patient-specific assessment.

The reviewed literature also demonstrates applications across different healthcare domains. Boll et al. (2024) reported the use of patient similarity graphs derived from electronic health records for heart-failure prediction, illustrating how relational patient representations can contribute to risk prediction. Zafeiropoulos et al. (2024) demonstrated the potential of healthcare knowledge graphs for supporting chronic disease management and early-warning applications. Graph-based approaches have also been applied to medical imaging, where structural relationships among features can be incorporated into predictive models (Balamurugan et al., 2024; Sharma et al., 2025). These applications suggest that graph-based methods are not restricted to a particular disease or data modality but can be adapted to diverse healthcare problems.

The literature nevertheless reveals several limitations that constrain the practical deployment of graph-based healthcare systems. First, data quality and standardization remain fundamental challenges. Graph construction depends on the quality, completeness, and consistency of the underlying data. Missing values, inconsistent coding, heterogeneous data formats, and differences in clinical documentation can alter the resulting network structure and consequently influence similarity calculations. Khemani et al. (2024) emphasize the importance of appropriate graph representations for machine-learning applications, while Mesinovic et al. (2025) highlight the broader problem of performance degradation and potential bias when artificial intelligence systems are deployed in real institutional environments.

Second, interoperability and institutional heterogeneity present substantial challenges. Healthcare organizations may use different electronic health-record systems, coding standards, data structures, and governance policies. Reibling et al. (2019) demonstrate the broader heterogeneity of healthcare systems, which can complicate the integration and comparison of patient data across institutions. Consequently, a graph-based model developed from one healthcare environment may not transfer directly to another without appropriate adaptation and validation.

Third, privacy, security, and ethical governance are particularly important because patient similarity networks may encode sensitive relationships among individuals, diseases, treatments, and genetic characteristics. Siddiqui et al. (2022) emphasize the growing importance of secure digital-health infrastructures as wearable, implantable, and other connected technologies increasingly contribute to healthcare data generation. The implementation of graph-based healthcare systems therefore requires not only technical safeguards but also appropriate governance concerning data access, consent, transparency, and responsible use.

Finally, the interpretability of graph-based models remains an important consideration for clinical adoption. Although graph structures can make relationships explicit, increasingly complex GNN architectures may produce predictions that are difficult for clinicians to interpret. For graph-based systems to function effectively as clinical decision-support tools, their predictions should be accompanied by clinically meaningful explanations of why particular patients, features, or relationships contributed to a given prediction.

Overall, the synthesis indicates that graph-based similarity measures offer a promising foundation for understanding patient variability, identifying clinically meaningful subgroups,

and supporting personalized healthcare. Their main contribution is the ability to transform heterogeneous healthcare information into a relational representation that can subsequently be analyzed using graph algorithms, machine learning, or GNNs. However, the clinical value of these approaches depends on the quality and representativeness of the underlying data, the validity of the similarity definition, the robustness of the network-construction procedure, and appropriate external validation. Thus, graph-based similarity measures should be understood not as a replacement for conventional clinical reasoning, but as a complementary analytical framework for relational, data-driven healthcare. Their integration with electronic health records, multimodal clinical data, knowledge graphs, and GNNs provides opportunities for more comprehensive patient representation and more individualized clinical decision support. Future research should therefore focus not only on improving predictive performance but also on developing interpretable, privacy-preserving, transferable, and clinically validated graph-based systems.

Conclusion

This study indicates that graph-based similarity measures provide a useful framework for representing patient relationships, identifying meaningful patterns, and supporting personalized decision-making in healthcare. By representing patients as nodes and their clinical similarities as edges, graph-based approaches can facilitate the identification of patient clusters, potential risk groups, and patterns of disease progression that may be difficult to detect using conventional approaches. The integration of heterogeneous data, including medical histories, laboratory results, symptoms, treatment responses, and genetic information, can further enrich patient representations and improve the contextual interpretation of similarity relationships. Thus, graph-based similarity measures offer a promising basis for data-driven and patient-centered healthcare.

Despite these advantages, several limitations constrain the practical implementation of graph-based approaches. Data quality remains a major concern because healthcare datasets may contain missing, inconsistent, heterogeneous, or outdated information, which can affect the reliability of similarity measures and subsequent predictions. Privacy and security are also critical considerations because patient networks often involve highly sensitive clinical and genomic information. In addition, integrating multimodal data, such as clinical text, medical images, laboratory data, and genomic profiles, requires substantial computational resources, interoperable data infrastructures, and specialized technical expertise. These challenges suggest that the successful implementation of graph-based healthcare systems requires not only methodological advances but also robust data governance, interoperability standards, and ethical safeguards.

Future implementation should therefore focus on improving data quality and interoperability, strengthening electronic health record infrastructures, and establishing standardized procedures for integrating heterogeneous healthcare data. Healthcare professionals and data scientists also require appropriate training to interpret graph-based

analyses and incorporate them responsibly into clinical decision-making. Furthermore, privacy-preserving technologies and secure computational frameworks should be prioritized to ensure that the use of patient networks does not compromise data confidentiality. Integrating graph-based similarity measures with artificial intelligence and machine learning may further enhance patient clustering, risk prediction, and treatment recommendation, although such applications should be evaluated rigorously for accuracy, robustness, interpretability, and potential bias.

Future research should move beyond conceptual and literature-based analyses by empirically evaluating graph-based similarity measures across diverse clinical datasets and healthcare settings. Particular attention should be given to longitudinal patient networks, multimodal data integration, explainable graph neural networks, and privacy-preserving learning frameworks. Dynamic patient networks that incorporate continuously updated clinical information may provide opportunities for more responsive risk assessment and adaptive decision support. However, their clinical value should be established through rigorous validation and prospective evaluation rather than assumed from computational performance alone. Overall, graph-based patient networks represent a promising direction for precision and personalized healthcare, particularly when mathematical modeling, artificial intelligence, and clinical knowledge are integrated within a robust and ethically governed framework.

Conflicts of Interest

The authors declare that no conflict of interest regarding the publication of this manuscript the authors.

References

Balamurugan, M., Muthuraman, M. S., Porkodi, V., Sivaram, M., Sridharan, M., & Gangadevi, E. (2024). Combining graph theory and machine learning for network-based medical image analysis. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1–5). IEEE.

Boll, H. O., Amirahmadi, A., Soliman, A., Byttner, S., & Recamonde-Mendoza, M. (2024). Graph neural networks for heart failure prediction on an EHR-based patient similarity graph. *arXiv*. https://arxiv.org/abs/2411.19742

Chen, A. (2022). A novel graph methodology for analyzing disease risk factor distribution using synthetic patient data. *Healthcare Analytics, 2*, 100084. https://doi.org/10.1016/j.health.2022.100084

Ding, H. (2025). Community clustering and recommendation toward elderly health records: Probabilistic representation learning and large graph search. *IEEE Access*.

DuGoff, E. H., Fernandes-Taylor, S., Weissman, G. E., Huntley, J. H., & Pollack, C. E. (2018). A scoping review of patient-sharing network studies using administrative data. *Translational Behavioral Medicine, 8*(4), 598–625. https://doi.org/10.1093/tbm/ibx038

Ganzinger, M., Schrodt, J., & Knaup, P. (2019). A concept for graph-based temporal similarity of patient data. *Studies in Health Technology and Informatics, 264*, 138–142. https://doi.org/10.3233/SHTI190199

Guo, C., & Chen, J. (2023). Big data analytics in healthcare. In *Knowledge technology and systems: Toward establishing knowledge systems science* (pp. 27–70). Springer Nature Singapore.

Hassan, M., et al. (2022). Innovations in genomics and big data analytics for personalized medicine and healthcare. *International Journal of Molecular Sciences, 23*(9), 4645. https://doi.org/10.3390/ijms23094645

Hu, X., Sun, L., Zheng, R., Xia, X., Liu, M., Chen, W., et al. (2024). Application of graph theory in liver research: A review. *Portal Hypertension & Cirrhosis, 3*(4), 234–248. https://doi.org/10.1097/PHC.

Khemani, B., Patil, S., Kotecha, K., & Tanwar, S. (2024). A review of graph neural networks: Concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data, 11*(1), 18. https://doi.org/10.1186/s40537-024-00818-0

Langberg, E. M., Dyhr, L., & Davidsen, A. S. (2019). Development of the concept of patient-centredness: A systematic review. *Patient Education and Counseling, 102*(7), 1228–1236. https://doi.org/10.1016/j.pec.2019.02.014

Lee, E. Y., Cha, S. A., Yun, J. S., et al. (2022). Efficacy of personalized diabetes self-care using an EMR-integrated mobile app. *Journal of Medical Internet Research, 24*(7), e37430. https://doi.org/10.2196/37430

Li, M. M., Huang, K., & Zitnik, M. (2022). Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering, 6*(12), 1353–1369. https://doi.org/10.1038/s41551-022-00939-x

Mesinovic, M., Buhlan, M., & Zhu, T. (2025). Causal graph neural networks for healthcare. *arXiv*. https://arxiv.org/abs/2511.02531

Pai, S., & Bader, G. D. (2018). Patient similarity networks for precision medicine. *Journal of Molecular Biology, 430*(18), 2924–2938. https://doi.org/10.1016/j.jmb.2018.07.006

Paul, S. G., Saha, A., Hasan, M. Z., et al. (2024). A systematic review of graph neural networks in healthcare applications. *IEEE Access, 12*, 15145–15170. https://doi.org/10.1109/ACCESS.2024.3355678

Reibling, N., Ariaans, M., & Wendt, C. (2019). Worlds of healthcare: A healthcare system typology of OECD countries. *Health Policy, 123*(7), 611–620. https://doi.org/10.1016/j.healthpol.2019.04.001

Rostami, M., Oussalah, M., Berahmand, K., & Farrahi, V. (2023). Community detection algorithms in healthcare applications: A systematic review. *IEEE Access, 11*, 30247–30272. https://doi.org/10.1109/ACCESS.2023.3256478

Schrodt, J., Dudchenko, A., Knaup-Gregori, P., & Ganzinger, M. (2020). Graph-representation of patient data: A systematic literature review. *Journal of Medical Systems, 44*(4), 86. https://doi.org/10.1007/s10916-020-1538-4

Sharma, A., Sharma, A., & Guo, K. (2025). Intelligent medical diagnosis model based on graph neural networks for medical images. *CAAI Transactions on Intelligence Technology*. https://doi.org/10.1049/cit2.12000

Siddiqui, S., Khan, A. A., & Dey, I. (2022). Internet technologies for personalized care. In *ICT frameworks in telehealth* (pp. 173–189). Springer.

Singhal, A. K., Manhas, S., & Singh, A. (2025). Network reconstruction-based model for predicting individual health status. *Computing*, *107*(1), 14. https://doi.org/10.1007/s00607-024-01234

Sowrirajan, R., & Manimekalai, S. (2024). Graph-theoretic approaches in healthcare systems. In *Ubiquitous computing and technological innovation for universal healthcare* (pp. 327–351). IGI Global.

Suo, Q., Ma, F., Yuan, Y., et al. (2018). Deep patient similarity learning for personalized healthcare. *IEEE Transactions on Nanobioscience*, *17*(3), 219–227. https://doi.org/10.1109/TNB.2018.2848733

Waqas, A. (2024). *From graph theory for robust deep networks to graph learning for multimodal cancer analysis* [Doctoral dissertation, University of South Florida]. University of South Florida Digital Commons.

Woodman, R. J., Koczwara, B., & Mangoni, A. A. (2024). Applying precision medicine principles to multimorbidity. *Frontiers in Medicine*, *10*, 1302844. https://doi.org/10.3389/fmed.2023.1302844

Xu, T., Dong, Z., & Luo, Z. (2024). Graph networks for precision medical diagnostics and treatment optimization. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE.

Zafeiropoulos, N., Bitilis, P., Tsekouras, G. E., & Kotis, K. (2024). Ontology-based PD monitoring using knowledge graphs and GNNs. *Information*, *15*(2), 100. https://doi.org/10.3390/info15020100

Zhang, D., Xia, X., Yang, Y., et al. (2021). Word similarity measure for IoT healthcare applications. *Future Generation Computer Systems*, *114*, 209–218. https://doi.org/10.1016/j.future.2020.07.040

Zijlmans, M., Zweiphenning, W., & Van Klink, N. (2019). Changing concepts in presurgical epilepsy assessment. *Nature Reviews Neurology*, *15*(10), 594–606. https://doi.org/10.1038/s41582-019-0236-9