Elements of graph theory and patient similarity networks (PSN) – A short introduction for ELMED219+BMED365

Arvid Lundervold [1,2] and the Medical AI Assistant (UiBmed - ELMED219 & BMED365)

[1]Department of Biomedicine, University of Bergen, 5020 Bergen, Norway;
[2]Mohn Medical Imaging and Visualization (MMIV) Centre, Department of Radiology, Haukeland University Hospital, Bergen, Norway

Abstract

A quick tutorial on the relevance of graph theory and network science in (bio)medical education.

```
YOU:
Why should medical students learn about graph theory? (in Latex, please)

Medical AI Assistant (UiBmed - ELMED219 & BMED365):
\title{Importance of Graph Theory in Medical Education}
```

1 Introduction

Graph theory, a field of mathematics and computer science, is essential in various aspects of medical studies and research. It offers a robust framework for modeling and analyzing complex systems, which is particularly relevant in understanding biological systems, disease pathways, and healthcare networks.

```
YOU:
Give details about "Graph theory". I.e., concepts, definitions, mathematical formalisms,
and links to internet tutorials suited for a student in medicine or biomedicine
(all in Latex, please).
```

1.1 Introduction to Graph Theory

Graph theory is a field of mathematics and computer science focused on studying graphs, which are mathematical structures used to model pairwise relations between objects. In the context of medicine and biomedicine, graph theory plays a crucial role in modeling complex systems such as genetic interactions, disease pathways, and healthcare delivery systems.

1.1.1 Basic Concepts and Definitions

- **Graph (G):** A graph G is defined as a set of nodes (or vertices) V and a set of edges E that connect pairs of nodes. Formally, $G = (V, E)$.
- **Node (Vertex):** A fundamental unit in a graph representing an entity such as a protein, gene, or patient.
- **Edge (Link):** A connection between two nodes representing a relationship or interaction, such as a biochemical reaction or patient referral.
- **Path:** A sequence of edges that connects a series of distinct nodes, representing a possible route from one node to another.
- **Degree:** The number of edges connected to a node. In directed graphs, this is further divided into in-degree and out-degree.

1.1.2 Mathematical Formalisms

- **Adjacency Matrix (A):** A square matrix used to represent a finite graph. The element A_{ij} is one if there is an edge from node i to node j, and zero otherwise.
- **Laplacian Matrix (L):** Defined as $L = D - A$, where D is the degree matrix and A is the adjacency matrix. It is used in various analyses, including network stability and dynamics.
- **Paths and Connectivity:** The study of paths is crucial in understanding graph connectivity. The shortest path between nodes is often of particular interest.
- **Centrality Measures:** These are metrics designed to identify the most important vertices within a graph. Examples include degree centrality, betweenness centrality, and eigenvector centrality.

1.1.3 Online Tutorials and Resources

For medical and biomedical students new to graph theory, the following online tutorials and resources provide a gentle introduction (`YOU: Confirm the links with Bing`):

1. **Khan Academy Introduction to Graph Theory:** A beginner-friendly series covering basic concepts and applications. Khan Academy Graph Theory
2. **Graph Theory Tutorials by Sarada Herke:** A YouTube playlist offering visual and intuitive explanations of graph theory concepts. Sarada Herke's Graph Theory Tutorials
3. **Introduction to Graph Theory from University of Waterloo:** Detailed course notes suitable for those seeking a more academic approach. University of Waterloo Graph Theory

1.1.4 Conclusion

Understanding the basics of graph theory is essential for analyzing complex networks in medicine and biomedicine. The concepts, mathematical formalisms, and resources provided here offer a starting point for students looking to apply these techniques in their field.

2 Reasons for Learning Graph Theory

2.1 Modeling Biological Networks

Biological systems can be represented as networks of interactions, such as metabolic pathways, protein-protein interaction networks, and genetic regulatory networks. Understanding these complex interactions is crucial for medical students, and graph theory provides the tools to analyze and visualize such structures.

```
YOU:
Give more details about "Modeling Biological Networks", i.e.,  examples, formalisms,
and links to recent references (all in Latex, please)
```

Biological networks provide a comprehensive framework to understand the complex interactions within biological systems. These networks are typically modeled using graph theory, where nodes represent biological entities (such as genes, proteins, or metabolites) and edges represent interactions or relationships between these entities.

2.1.1 Examples of Biological Networks

- **Metabolic Networks:** Nodes represent metabolites, and edges represent biochemical reactions converting one metabolite to another. These networks help in understanding metabolic pathways and fluxes.
- **Protein-Protein Interaction Networks:** Nodes are proteins, and edges represent physical interactions between them. Analyzing these networks aids in understanding cellular processes and identifying crucial proteins.
- **Genetic Regulatory Networks:** These networks consist of genes (nodes) and their regulatory interactions (edges), crucial for understanding gene expression and regulation mechanisms.

2.1.2 Formalisms in Graph Theory for Biological Networks

- **Adjacency Matrix:** A square matrix used to represent a finite graph. The elements of the matrix indicate whether pairs of vertices are adjacent or not in the graph.
- **Incidence Matrix:** A matrix that shows the relationship between nodes and edges, indicating which nodes are connected by which edge.
- **Pathways and Flow:** Algorithms to identify pathways and flow in networks, such as shortest path algorithms, are used to understand metabolic pathways and signal transduction pathways.

2.1.3 Recent References

For the latest in this field, consider the following references, which discuss the current state and advancements in modeling biological networks (The URLs in the references are placeholders and should be replaced with the actual links to the papers or resources mentioned):

1. Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press, New York. `https://math.bme.hu/~gabor/oktatas/SztoM/Newman_Networks.pdf`

2. Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), 101–113. [3]

3. Ideker, T., & Krogan, N. J. (2012). Differential network biology. *Molecular Systems Biology*, 8(1). [8]

2.1.4 Conclusion

Understanding biological networks through graph theory allows for a deeper insight into the functional organization of various biological systems. It aids in identifying key components, understanding disease mechanisms, and discovering new therapeutic targets. As the field evolves, so too does the potential for groundbreaking discoveries in biology and medicine.

2.2 Disease Pathway Analysis

Diseases often involve multiple genes and pathways. Graph theory helps in understanding the interconnectivity and dynamics of disease mechanisms, facilitating the identification of potential therapeutic targets and understanding disease progression.

```
YOU:
Give more details about "Disease Pathway Analysis", i.e.,  examples, formalisms,
and links to recent references (all in Latex, please)
```

Disease pathway analysis is a crucial aspect of understanding the complex mechanisms behind diseases. By utilizing graph theory, researchers can map out the interactions between various biological entities involved in disease processes. This approach helps in identifying potential targets for therapy and understanding the multifactorial nature of diseases.

2.2.1 Examples of Disease Pathways

- **Cancer Signaling Pathways:** Graph theory models the interactions between oncogenes, tumor suppressors, and other cellular factors to understand cancer progression and identify potential points for intervention.
- **Neurodegenerative Diseases:** Networks of protein-protein interactions and metabolic pathways are analyzed to understand diseases like Alzheimer's and Parkinson's, focusing on crucial pathways involved in disease progression.
- **Infectious Disease Pathways:** Interaction networks between pathogens and host cellular machinery are crucial for understanding infection mechanisms and developing strategies for treatment and prevention.

2.2.2 Formalisms in Graph Theory for Disease Pathways

- **Directed Graphs:** These represent pathways with a direction, indicating the flow of biological processes, essential for understanding signaling pathways and metabolic reactions.
- **Weighted Networks:** In these graphs, edges are assigned weights representing the strength or capacity of interactions, which is particularly useful in quantifying the effect of different pathway components.

- **Dynamic Networks:** These models account for changes over time, essential for understanding the progression of diseases and the impact of various treatments.

2.2.3 Recent References

For the most recent and influential work in this area, consider the following references (The URLs in the references are placeholders and should be replaced with the actual links to the papers or resources mentioned):

1. Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature*, 461(7261), 218–223. [17]
2. Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1), 56–68. [2]
3. Ideker, T., & Sharan, R. (2008). Protein networks in disease. *Genome Research*, 18(4), 644–652. [9]

2.2.4 Conclusion

Disease pathway analysis through graph theory offers a powerful approach to understanding and combating various diseases. By modeling biological processes as networks, researchers can uncover the complex interplay of factors involved in disease mechanisms and identify key points for therapeutic intervention.

2.3 Patient Similarity Networks

Graphs can represent similarities between patients, considering various factors like genetics, symptoms, and demographics. This understanding aids in personalized medicine, where treatments are tailored to individual patient profiles.

```
YOU:
Give more details about "Patient Similarity Networks", i.e.,  examples, formalisms,
and links to recent references (all in Latex, please)
```

Patient similarity networks are a novel approach in medical informatics to understand and predict patient outcomes better. By using graph theory, these networks model patients as nodes and similarities between them as edges, aiding in various tasks such as disease prediction, patient cohorting, and personalized treatment strategies.

2.3.1 Examples of Patient Similarity Networks

- **Genomic Similarity Networks:** Patients are linked based on the similarity of their genetic profiles, helping to identify genetic markers for diseases and potential responses to treatment.
- **Clinical Symptom Networks:** These networks connect patients with similar clinical symptoms or disease manifestations, aiding in diagnosis and understanding disease progression.
- **Phenotypic Similarity Networks:** Patients are connected based on shared phenotypes, such as laboratory test results and physical characteristics, to identify subtypes of diseases and predict treatment outcomes.

2.3.2 Formalisms in Graph Theory for Patient Similarity Networks

- **Edge Weighting:** Weights on the edges can represent the degree of similarity between patients, with higher weights indicating greater similarity. Various metrics can be used to calculate these weights, such as Euclidean distance, correlation, or mutual information.
- **Clustering Algorithms:** Techniques like community detection or hierarchical clustering can be used to identify groups of similar patients within the network, which can be crucial for cohort analysis and personalized medicine.
- **Graph Embedding:** Methods to represent patients and their relationships in a low-dimensional space, facilitating the application of machine learning models for predictive tasks.

2.3.3 Recent References

For the latest research and discussions in this area, consider the following references (The URLs in the references are placeholders and should be replaced with the actual links to the papers or resources mentioned):

1. Brown, S-A. (2016). Patient Similarity: Emerging Concepts in Systems and Precision Medicine. *Frontiers in Physiology*, 7, 561. [5]
2. Pai, S. & Bader, G.D. (2018). Patient similarity networks for precision medicine. *Journal of Molecular Biology*, 430(18), 2924-2938. [14]
3. Shaik, T., et al. (2024). A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom. *Information Fusion*, 102, 102040 [18]
4. Lambert, J., et al. (2023). Tracking clusters of patients over time enables extracting information from medico-administrative databases. *Journal of Biomedical Informatics*, 139, 104309 [12]

2.3.4 Conclusion

Patient similarity networks are a powerful tool in the move towards personalized medicine. By applying graph theory to medical data, healthcare providers can gain deeper insights into patient characteristics, improve diagnostic accuracy, and tailor treatments to individual patient profiles, ultimately improving outcomes and care quality.

2.4 Epidemiology and Public Health

Graph theory is fundamental in modeling the spread of infectious diseases and in devising strategies for vaccination and containment. It helps in understanding how diseases spread through populations and in identifying critical control points.

```
YOU:
Give more details about "Epidemiology and Public Health", i.e.,  examples, formalisms,
and links to recent references (all in Latex, please)
```

Epidemiology and public health greatly benefit from graph theory, which provides a framework to model the spread of diseases, understand health-related behaviors, and plan effective interventions. Networks can represent individuals or populations as nodes, with edges indicating interactions or relationships that can facilitate disease transmission.

2.4.1 Examples in Epidemiology and Public Health

- **Infectious Disease Modeling:** Networks model how diseases spread through populations, helping to identify super-spreaders and critical points for intervention.
- **Vaccination Strategies:** By identifying central individuals in a network, public health officials can optimize vaccination campaigns to halt disease spread efficiently.
- **Social Determinants of Health:** Networks analyze the effects of social factors on health, showing how social structures contribute to health disparities.

2.4.2 Formalisms in Graph Theory for Epidemiology and Public Health

- **Compartmental Models in Networks:** Individuals are represented as nodes, and their disease states (susceptible, infected, recovered) as node attributes. The edges determine interaction pathways for disease spread.
- **Percolation Theory:** Used to understand the conditions under which diseases will spread through a population and how changes in network structure can contain outbreaks.
- **Network Metrics:** Various metrics such as degree distribution, centrality, and clustering coefficients are used to analyze the properties of health-related networks and predict the spread and control of diseases.

2.4.3 Recent References

Recent research and discussions in this area can be found in the following references (The URLs in the references are placeholders and should be replaced with the actual links to the papers or resources mentioned):

1. Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, 86(14), 3200–3203. `https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.86.3200` [15]
2. Christakis, N. A., & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, 357, 370-379. [6]
3. Eubank, S., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429, 180-184. [7]

2.4.4 Conclusion

The application of graph theory in epidemiology and public health is vast and vital. It enhances our understanding of disease dynamics, informs public health policies, and aids in the development of strategies to prevent and control diseases, thereby improving population health outcomes.

2.5 Healthcare Systems Analysis

Healthcare systems are complex, with intricate relationships between patients, healthcare providers, and resources. Graph theory provides a framework to analyze and optimize these systems, improving efficiency and patient care.

```
YOU:
Give more details about "Healthcare Systems Analysis", i.e.,  examples, formalisms,
and links to recent references (all in Latex, please)
```

Graph theory plays a critical role in analyzing and optimizing healthcare systems. It provides a structured way to model complex relationships between various healthcare entities, such as patients, providers, and institutions, and to analyze the flow of information, resources, and patients through these systems.

2.5.1 Examples in Healthcare Systems Analysis

- **Patient Referral Networks:** Networks represent how patients are referred between different healthcare providers, helping to identify key nodes and links in the healthcare delivery system.
- **Hospital Collaboration Networks:** These networks model the collaboration between hospitals and other healthcare entities, essential for understanding resource sharing and improving patient outcomes.
- **Resource Allocation Networks:** Graphs model the flow of medical resources, such as equipment and staff, between various departments or facilities, aiding in efficient resource management.

2.5.2 Formalisms in Graph Theory for Healthcare Systems Analysis

- **Network Flow Algorithms:** These algorithms help understand and optimize the flow of patients, information, and resources through healthcare systems.
- **Centrality Measures:** Identifying central nodes (e.g., key hospitals, providers, or resources) in healthcare networks can inform strategies for improving system efficiency and resilience.
- **Community Detection:** Techniques for identifying communities or clusters within healthcare networks can help in understanding the modular structure of healthcare systems and in planning localized interventions.

2.5.3 Recent References

To delve deeper into the current state of research in healthcare systems analysis, consider the following references (The URLs in the references are placeholders and should be replaced with the actual links to the papers or resources mentioned):

1. Uddin, S., & Hossain, L. (2011). Social Networks Enabled Coordination Model for Cost Management of Patient Hospital Admissions. *Journal for Healthcare Quality*, 33(5), 37-48. [20]
2. Kohler, K. & Ercole, A. (2020). Can network science reveal structure in a complex healthcare system? A network analysis using data from emergency surgical services. *BMJ Open*, 10(2), e034265. [11]

3. Ben-Tovim, D. I, et al. (2022). Modular structures and the delivery of inpatient care in hospitals: a Network Science perspective on healthcare function and dysfunction. *BMC Health Services Research*, 22(1), 1-12. [4]

2.5.4 Conclusion

Graph theory offers valuable tools for understanding and improving healthcare systems. By analyzing the structure and dynamics of healthcare networks, stakeholders can make informed decisions to enhance efficiency, collaboration, and patient care.

3 Conclusion

For medical students, graph theory is not just a theoretical concept; it is a practical tool with wide-ranging applications in biology, medicine, and healthcare. Understanding graph theory enhances the ability to tackle complex problems, making it a valuable addition to a medical education.

A Code prompting

YOU:

`Make python code to illustrate a PSN (use 5 patients, a dense graph, and a spring layout)`

Medical AI Assistant (UiBmed - ELMED219 & BMED365):
Python code (using `NetworkX`) to illustrate a dense Patient Similarity Network (PSN) with 5 patients using a spring layout:

```
import networkx as nx
import matplotlib.pyplot as plt
import numpy as np

# Create a dense graph object
G = nx.complete_graph(5)

# Label nodes as patients
mapping = {i: f"Patient{i+1}" for i in range(5)}
G = nx.relabel_nodes(G, mapping)

# Assign random weights to edges to represent similarity
for (u, v) in G.edges():
    G.edges[u, v]['weight'] = np.round(np.random.random(), 2)

# Draw the graph using a spring layout
pos = nx.spring_layout(G)
nx.draw(G, pos, with_labels=True, node_color='skyblue', node_size=500)
edge_labels = nx.get_edge_attributes(G, 'weight')
nx.draw_networkx_edge_labels(G, pos, edge_labels=edge_labels)

plt.show()
```

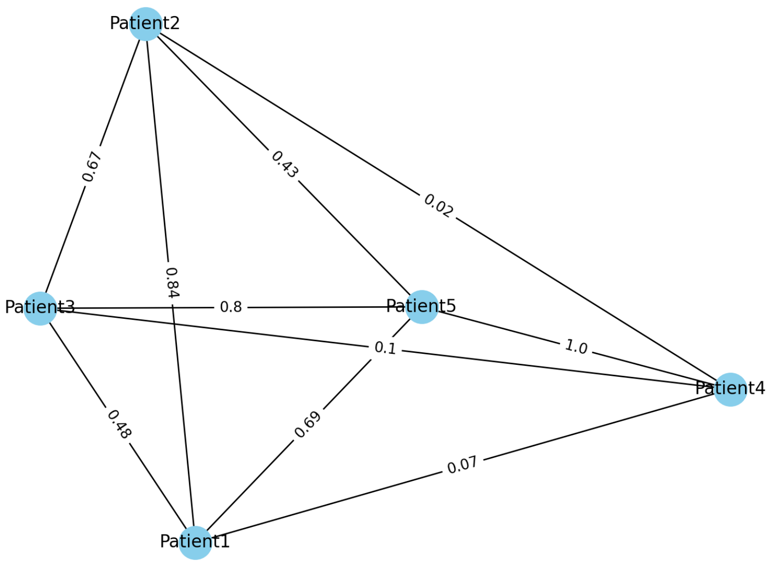

Figure 1: The Python code illustrates a dense Patient Similarity Network (PSN) with 5 patients using a spring layout. Each node represents a patient, and every pair of nodes is connected with an edge, indicating a dense network where all patients are compared with each other. The weights on the edges represent the similarity scores between patients.

Some other relevant references ... [1, 10, 13, 16, 19].

References

[1] M. Askar, R. N. Cañadas, and K. Svendsen. An introduction to network analysis for studies of medication use. *Research in Social and Administrative Pharmacy*, 17(12):2054–2061, 2021. https://www.sciencedirect.com/science/article/pii/S1551741121002382 [Accessed: 27-Dec-2023].

[2] A.-L. Barabási, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature reviews genetics*, 12(1):56–68, 2011. https://www.nature.com/articles/nrg2918 [Accessed: 26-Dec-2023].

[3] A.-L. Barabasi and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113, 2004. https://www.nature.com/articles/nrg1272 [Accessed: 26-Dec-2023].

[4] D. I. Ben-Tovim, M. Bajger, V. D. Bui, S. Qin, and C. H. Thompson. Modular structures and the delivery of inpatient care in hospitals: a network science perspective on healthcare function and dysfunction. *BMC Health Services Research*, 22(1):1–12, 2022. https://pubmed.ncbi.nlm.nih.gov/36494814 [Accessed: 27-Dec-2023].

[5] S.-A. Brown. Patient similarity: emerging concepts in systems and precision medicine. *Frontiers in physiology*, 7:561, 2016. https://www.frontiersin.org/articles/10.3389/fphys.2016.00561/full [Accessed: 27-Dec-2023].

[6] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007. https://www.nejm.org/doi/full/10.1056/nejmsa066082 [Accessed: 27-Dec-2023].

[7] S. Eubank, H. Guclu, V. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004. https://www.nature.com/articles/nature02541 [Accessed: 27-Dec-2023].

[8] T. Ideker and N. J. Krogan. Differential network biology. *Molecular systems biology*, 8(1):565, 2012. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3296360 [Accessed: 26-Dec-2023].

[9] T. Ideker and R. Sharan. Protein networks in disease. *Genome research*, 18(4):644–652, 2008. https://genome.cshlp.org/content/18/4/644.long [Accessed: 27-Dec-2023].

[10] Z. Jia, X. Zeng, H. Duan, X. Lu, and H. Li. A patient-similarity-based model for diagnostic prediction. *International journal of medical informatics*, 135:104073, 2020. https://doi.org/10.1016/j.ijmedinf.2019.104073 [Accessed: 26-Dec-2023].

[11] K. Kohler and A. Ercole. Can network science reveal structure in a complex healthcare system? a network analysis using data from emergency surgical services. *BMJ open*, 10(2), 2020. https://pubmed.ncbi.nlm.nih.gov/32041860 [Accessed: 27-Dec-2023].

[12] J. Lambert, A.-L. Leutenegger, A.-S. Jannot, and A. Baudot. Tracking clusters of patients over time enables extracting information from medico-administrative databases. *Journal of Biomedical Informatics*, 139:104309, 2023. https://www.sciencedirect.com/science/article/pii/S1532046423000308 [Accessed: 27-Dec-2023].

[13] H. Memarzadeh, N. Ghadiri, M. Samwald, and M. Lotfi Shahreza. A study into patient similarity through representation learning from medical records. *Knowledge and Information Systems*, 64(12):3293–3324, 2022. https://doi.org/10.1007/s10115-022-01740-2 see also https://github.com/HodaMemar/Patient-Similarity-through-Representation[Accessed: 26-Dec-2023].

[14] S. Pai and G. D. Bader. Patient similarity networks for precision medicine. *Journal of molecular biology*, 430(18):2924–2938, 2018. https://www.sciencedirect.com/science/article/pii/S0022283618305321 [Accessed: 27-Dec-2023].

[15] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001. https://pubmed.ncbi.nlm.nih.gov/11290142/ [Accessed: 27-Dec-2023].

[16] S. Pokharel, G. Zuccon, X. Li, C. P. Utomo, and Y. Li. Temporal tree representation for similarity computation between medical patients. *Artificial Intelligence in Medicine*, 108:101900, 2020. https://doi.org/10.1016/j.artmed.2020.101900 [Accessed: 26-Dec-2023].

[17] E. E. Schadt. Molecular networks as sensors and drivers of common human diseases. *Nature*, 461(7261):218–223, 2009. https://www.nature.com/articles/nature08454 [Accessed: 26-Dec-2023].

[18] T. Shaik, X. Tao, L. Li, H. Xie, and J. D. Velásquez. A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom. *Information Fusion*, page 102040, 2024. https://www.sciencedirect.com/science/article/pii/S1566253523003561 [Accessed: 27-Dec-2023].

[19] A. R. Sonawane, S. T. Weiss, K. Glass, and A. Sharma. Network medicine in the age of biomedical big data. *Frontiers in Genetics*, 10:294, 2019. https://www.frontiersin.org/articles/10.3389/fgene.2019.00294/full [Accessed: 27-Dec-2023].

[20] M. S. Uddin and L. Hossain. Social networks enabled coordination model for cost management of patient hospital admissions. *Journal for Healthcare Quality*, 33(5):37–48, 2011. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1945-1474.2011.00118.x [Accessed: 27-Dec-2023].