

UC Davis

UC Davis Electronic Theses and Dissertations

Title

ML-Assisted Visual Analytics for Reasoning Complex Networks

Permalink

<https://escholarship.org/uc/item/0324v1hk>

ISBN

9798189225444

Author

Lu, Hsiao-Ying

Publication Date

2026-06-25

Peer reviewed|Thesis/dissertation

ML-Assisted Visual Analytics for Reasoning Complex Networks

By

HSIAO-YING LU

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Kwan-Liu Ma, Chair

Dongyu Liu

Jiawei Zhang

Committee in Charge

2026

Copyright © 2026 by

Hsiao-Ying Lu

All rights reserved.

CONTENTS

Abstract	vi
Acknowledgments	ix
1 Introduction	1
1.1 Motivation	1
1.2 Background	3
1.2.1 Networks	3
1.2.2 Network Characteristic Extraction	5
1.2.3 Network Feature Interpretation	7
1.2.4 Visual Analytics of Networks	8
1.3 Content Overview	9
2 Combinatorial Influence Extraction for Multivariate Networks	11
2.1 Introduction	11
2.2 Related Work	14
2.2.1 Learning Network Representation	14
2.2.2 Interpreting Representations	15
2.3 Methodology	17
2.3.1 Design Considerations	17
2.3.2 Workflow Overview	18
2.3.3 Two-class Density Scatterplot	20
2.3.4 Datasets: Students and Adults' Network Resources	22
2.3.5 Learning Representations	23
2.3.6 Understanding Representations	29
2.3.7 Interactive Visual Interface	33
2.4 Case Studies	35
2.4.1 Study 1: Associations with Score Levels	35
2.4.2 Study 2: Differences in Academic Units	38

2.4.3	Study 3: Factors Promoting New Connections	40
2.5	Expert Feedback	42
2.6	Discussion	44
2.7	Conclusion	46
3	Collaboration Effectiveness and Efficiency Assessment through Dynamic Networks	47
3.1	Introduction	48
3.2	Related Works	49
3.2.1	Health Informatics Analysis	49
3.2.2	Network Effectiveness Analysis	50
3.2.3	Network Efficiency Analysis	51
3.3	Data Processing and Analysis	53
3.3.1	Dataset Overview	53
3.3.2	Network Construction	54
3.3.3	Network Analysis	54
3.4	Visual Analytics of EHRFlow	58
3.5	Case Studies	61
3.5.1	Overview of Patient Groups and Analysis Goals	61
3.5.2	Study 1: Identification of Disconnected Communications	62
3.5.3	Study 2: Trade-Off of Notes Underutilization	64
3.5.4	Study 3: Influences from EHR Reading and Writing Duty	67
3.6	Expert Feedback	68
3.7	Discussion and Limitations	70
3.8	Conclusion	71
4	Explanatory Structural Pattern Identification in Network Classification Tasks	73
4.1	Introduction	74
4.2	Background	76

4.2.1	Graphlets and Graphlet Frequencies	76
4.2.2	Graph Neural Networks	77
4.3	Related Works	78
4.3.1	Subgraph Explanations	78
4.3.2	Instance/Model-Level Explanations	79
4.3.3	Visual Analytics for Explaining GNNs	80
4.3.4	GNN Explanation Evaluation Metrics	80
4.4	Methodology	81
4.4.1	Graphlet Frequency: Topological Summary	81
4.4.2	Graph Neural Network: Graph-level Classification	83
4.4.3	Surrogate Model: GNN Performance Approximation	84
4.4.4	Factual and Counterfactual Explanation Evaluation	85
4.5	Visual Analytics of GNNAnatomy	88
4.5.1	Graph Group Selection	89
4.5.2	Substructure Explanatory Ranking	90
4.5.3	Explanatory Substructure Evaluation	90
4.5.4	Substructure Impact on Overall Topology	91
4.6	Case Studies	92
4.6.1	Datasets	92
4.6.2	Study 1: Reddit-Binary	93
4.6.3	Study 2: MUTAG	96
4.6.4	Study 3: BA-House	99
4.7	Discussion and Future Works	102
4.7.1	Scalability in Graphlet Frequency Computation	102
4.7.2	Generalizability	102
4.7.3	Extensibility	103
4.8	Conclusion	103

5 Longitudinal Patient Care Risk Prediction and What-If Analysis for Intervention Planning 105

5.1	Introduction	106
5.2	Related Work	109
5.2.1	Healthcare Team Collaboration Analysis	109
5.2.2	Modeling and Interpreting Dynamic Graphs	110
5.3	Methodology	112
5.3.1	Data Processing	112
5.3.2	Model Pipeline	116
5.3.3	Visual Analytics Design	117
5.4	Experiments	124
5.4.1	Survival Prediction Performance	124
5.4.2	Survival Prediction Interpretation	124
5.4.3	Confounding Factors Correlation Analyses	125
5.5	Case Studies	125
5.5.1	Similar Care-Network Size, Divergent Risk	126
5.5.2	Similar Risk, Distinct Care-Team Evolution	129
5.6	Evaluation	132
5.6.1	Robustness Evaluation of Surrogate Model	133
5.6.2	Expert Assessment of the Tool and Analytic Framing	133
5.6.3	Clinical Grounding of the Surfaced Pattern	135
5.6.4	Summary	136
5.7	Discussion and Limitations	136
5.8	Conclusion	139
6	Discussion	140
7	Conclusion	143
A	Appendix for Chapter 2	145
A.1	Supplementary Comparison of Scatterplot Designs (Sec. 2.3.3)	145
A.2	Additional Statistical Analyses on Composite Variables	147
A.3	Demonstration Video, Source Code, and Sample Data	150

B	Appendix for Chapter 3	151
B.1	Graph Centralization Index for Directed Bipartite Networks	151

ABSTRACT

ML-Assisted Visual Analytics for Reasoning Complex Networks

Network data is pervasive in daily life. Various associations among entities found in real-world applications can be represented as multivariate networks, temporal networks, or heterogeneous networks. However, as data complexity increases, analyzing networks becomes a challenging task. To draw actionable insights, it is important to have the ability to uncover and reason network characteristics associated with behaviors/events of interest. In recent years, machine learning (ML) has proven effective in network feature extraction, while visual analytics (VA) has shown value in dissecting and reasoning complex data. A promising approach to complex network analysis is to combine these two techniques.

This dissertation aims to introduce new ML-assisted visual analytics designs for understanding complex networks. Firstly, to address the challenges of reasoning multivariate networks, I develop a visual analytics system that supports users in selecting network variables of interest and flexibly constructing a composite variable. This composite variable is generated through a series of learning-based optimizations and visually presented to convey the combined associations between a group of variables and another variable of interest.

Secondly, driven by the desire to improve patient care quality, I conduct longitudinal studies on healthcare professionals (HCPs) collaboration networks constructed from patient records. A visual analytics system is designed to investigate the effectiveness and efficiency of each such temporal teamwork network pertinent to the respective cancer patient. This work aims to associate the structural features and dynamics of the HCPs' communication network with patient survival outcomes using network measures. Users can interactively define the analysis timeframe and visually explore how information is transmitted among HCPs and determine whether each piece of patient information has been effectively and efficiently communicated throughout the patient care process.

Third, beyond detecting associations between network measures (e.g., edge density) and patient survival outcomes, identifying the most relevant collaborative patterns can

offer a more intuitive understanding. Graph Neural Networks (GNNs) may be used to extract structural features from complex networks but their complicated data transformation poses challenges in explaining their behavior. To address this shortcoming, I devise a visual analytics system to evaluate and explain the significance of each elementary substructure (i.e., graphlet) to the outputs of the GNN in graph-level classification tasks. In this system, by investigating the relationships between GNN classifications and the composition of elementary substructures, hypothesized explanations can be formulated and evaluated using metrics derived from both factual and counterfactual reasoning processes.

Finally, building on these prior efforts toward interpretable machine learning for patient risk assessment and intervention planning, I develop an interactive visual analytics framework for predictive analysis and what-if simulation. The framework supports the full analytical pipeline, from data curation and mitigation of foreseeable confounding factors to predicting cancer patient morbidity risk using GNNs. It also provides methods for explaining model predictions by identifying critical network structures and key HCP participation patterns that influence decision-making. Based on these explanations, users can simulate targeted interventions and examine projected changes in patient risk. Through this framework, users can explore how collaboration dynamics relate to estimated risk, generate intervention hypotheses, and evaluate them through interactive what-if analyses.

In conclusion, this dissertation advances complex network analysis by integrating machine learning and visual analytics to bridge pattern discovery, interpretation, and decision-making. Through a series of contributions spanning multivariate reasoning, temporal collaboration analysis, interpretable graph learning, and predictive what-if simulation, it demonstrates how complex network data can be transformed into actionable insights. By emphasizing transparency and human-centered reasoning, these efforts establish a foundation for trustworthy, data-driven analysis and decision support across complex networked systems.

ACKNOWLEDGMENTS

Completing this dissertation has been one of the most meaningful journeys of my life, and I would not have reached this point without the people who supported and stood beside me throughout these years.

First and foremost, I would like to thank my advisor, Prof. Kwan-Liu Ma, for his trust, patience, and mentorship throughout my Ph.D. He gave me the freedom to explore different ideas while encouraging me to grow into an independent researcher, and I will always appreciate the opportunity to learn under his guidance. I am also deeply fortunate to have learned from Dr. Takanori Fujiwara and Dr. Yiran Li, both of whom taught me how to approach research with rigor and care and shaped my growth during graduate school.

I would also like to thank my committee members, Prof. Dongyu Liu and Prof. Jiawei Zhang, for their time, insight, and support throughout the dissertation process. I additionally appreciate Prof. Muhao Chen and Prof. Robert Faris for serving on my qualifying exam committee and helping shape my research direction early on. I am equally appreciative of my collaborators, especially Dr. Yang-chih Fu, Dr. Shin-Ping Tu and Dr. Brittany Garcia. Working with researchers from diverse backgrounds broadened both my perspective and the way I think about the impact of research beyond academia.

These years would have been a very different experience without the people around me. I especially want to mention Xiwei, Yiran, Yuhua, Yu-Chia, Yueqiao, and Li-Wen for the conversations, shared experiences, laughter, and support that made these years far more memorable.

To my family, thank you for always being my foundation. To my parents, thank you for believing in me even during moments when I struggled to believe in myself. To my sister, Hsiao-Tung, thank you for always being there through every high and low.

Finally, to my boyfriend, Robert, thank you for walking through every stage of this journey with me. Your patience, steadiness, and care carried me through moments of uncertainty and self-doubt more than words can express.

Chapter 1

Introduction

1.1 Motivation

A network represents entities and their relationships, where entities are presented as nodes and their relations as edges or links. This allows for the modeling of relational data across various fields. Examples include collaborations among healthcare professionals (HCPs), bonds between atoms in molecules, and the interconnection of electricity stations (i.e., power grid). Analyzing networks yields actionable insights. For instance, we can examine how chemical bonds between different pairs of atoms can transform a molecule into a mutagen. Another example involves studying collaborations among HCPs to identify effective teamwork patterns that contribute to improved patient care outcomes. To achieve such insights, two main research topics emerge: network feature extraction and interpretation. Both tasks are challenging due to the simultaneous presence of structural and entity information. Additionally, network data complexity extends further as connections can be dynamic, forming or dissolving over time. Moreover, they can exhibit varying levels of relations, necessitating comprehensive data modeling. Therefore, designing analytical methods to uncover and understand structure-aware characteristics of networks is non-trivial.

To extract network features, scientists have developed measures that quantify various structural functions for either an entity or the entire network. For instance, *betweenness* measures how much an entity (e.g., HCP) serves to bridge two disconnected groups of

entities (e.g., how much the two groups rely on this HCP to facilitate communication and exchange of knowledge between them). In recent years, machine learning (ML) researchers have introduced graph neural networks (GNNs) to tackle this problem. GNNs examine each entity by considering the influences exerted by other entities through their connections. Specifically, GNNs learn a representation for each entity by combining its own information with that of its neighbors. The learned representations for all entities are then aggregated to form a network representation. Utilizing these representations in network-related tasks, such as link prediction or node classification, leads to better performance. However, the large number of network measures (i.e., each node or graph having a value), combined with the complexity of the topological functions involved, makes it difficult to interpret these measures by simply examining the numbers. Furthermore, it is challenging to reason about GNN behavior due to its complicated mechanism of data transformation. These limitations have driven researchers to explore the interpretability of network analytical results.

To comprehend complex analytical results, visual analytics (VA) has proven effective in breaking down intricate data into perceivable and meaningful components. Through interactions, users can efficiently reason about analytical results with the aid of visualizations. For example, if a user chooses to inspect the *betweenness* measure, we can encode the *betweenness* value of each entity as the size of each node in the network visualization. By combining topology overview and network measure values, this visualization facilitates an intuitive understanding of the structural function behind the *betweenness* value. Another approach involves explaining the behavior of ML models. The primary motivation is to leverage the capabilities of ML in transforming discrete networks into compact representations (i.e., vectors) that preserve essential structure-aware characteristics. By unpacking complex yet informative representations, users can understand the features utilized by ML models to make predictions or classifications. For example, this may involve identifying the most relevant subgraph to a GNN’s classification of the network.

To draw tangible findings, it is essential to both analyze complex networks and interpret analytical results effectively. This dissertation focuses on combining ML and VA

techniques with the objective to tailor ML methods for various types of complex networks and design VA systems for embedding and streamlining the analyses. Through interactions with these systems, users are empowered to conduct ML-assisted analyses of interest and interpret the results simultaneously.

1.2 Background

In this section, I provide a brief overview of research topics relevant to this proposal. I start by defining the terminologies for various types of complex networks. Next, I discuss the efforts made to extract network characteristics. Then, I address the techniques developed to reason the extracted characteristics. Finally, I introduce visual analytics works designed to assist in inspecting complex networks and understanding multi-faceted analytical results.

1.2.1 Networks

A network or graph represents entities and their relationships as $G = (V, E)$, where $V = (v_1, v_2, v_3, \dots)$ is the set of entities and $E \subseteq V \times V$ encodes the relations between pairs of entities. An entity is often referred to as a node or a vertex, which can represent, for example, a person in social networks or an atom in a biological network. A relation is referred to as an edge or a link between a pair of entities, which can represent, for example, a collaboration between two colleagues or a chemical bond between two atoms. As the complexity of relational data increases, more network representations arise to accommodate this complexity.

1.2.1.1 Multivariate Networks

Multivariate networks store additional node and edge attributes. Inheriting the representation of a generic network, a multivariate network is given by $G_V = (V, E, X_V, X_E)$, where X_V is the set of node attributes and X_E is the set storing edge attributes. For a network with n node attributes, $X_V = (x_{ij})$ where $i = 1, \dots, n$, $j = 1, \dots, |V|$, and x_{ij} is the i -th attribute of node v_j . Similarly, we can define X_E for edge attributes.

Analyzing these attributes within network topologies enables researchers to derive deeper insights. For instance, in studying the social activities of college students on Face-

book, we can utilize a multivariate network to incorporate each student’s demographics (e.g., age or gender), academic records (e.g., major or GPA), and health data (e.g., depression level) as node attributes. Furthermore, the nature of interactions, such as replies or likes, can be represented as edge attributes. With this wealth of node and edge information, we can answer research questions such as variations in social interaction trends among students from different majors.

1.2.1.2 Dynamic Networks

Dynamic networks represent relational data that evolves over time, meaning that new nodes and edges may join while others may disappear. Therefore, indicating the existence of nodes and edges at each time step is crucial for modeling a time-varying network. This can be simply represented as a sequence of static networks $G_D = (g_1, g_2, \dots, g_T)$, where $g_t = (V_t, E_t)$ represents a network constructed by the existing set of nodes V_t and edges E_t at time step t . A dynamic network can also be represented by encoding timestamps as node and/or edge attributes, denoted as $G_D = (V, E, T_V, T_E)$. Here, $T_V = (t_j)$, $j = 1, \dots, |V|$, and t_j stores the sequence of timestamps or durations during which node v_j appeared in the network. Similarly, T_E can be defined for edge timestamps.

Real-world relational data often evolves over time. For research problems involving temporal aspects, it is crucial to analyze the structure of a network along with its dynamics. For instance, when examining information flows among healthcare professionals (HCPs), timestamps of origin and destination edges, indicating the time where the information exchanges took place, can be used to measure the average time it takes to spread critical patient information. This information can then be leveraged to evaluate collaboration efficiency among HCPs.

1.2.1.3 Heterogeneous Networks

Heterogeneous networks capture relationships among multiple types of entities within a unified graph structure. A common class of heterogeneous networks is the bipartite network, which consists of two node sets, where edges occur only between the two sets. For instance, when modeling healthcare systems, one node set may represent patients and the other healthcare professionals (HCPs), while edges indicate treatment relationships

between patients and providers. Explicitly separating these entity types preserves their distinct roles and relational structure. A bipartite network is denoted as $G_B = (V_1, V_2, E)$, where V_1 and V_2 are two disjoint node sets with $V_1 \cap V_2 = \emptyset$, and E is the set of edges connecting nodes across the two sets.

Analyzing heterogeneous networks provides deeper insights into interactions among different entity types within a system. By preserving the roles of distinct node groups and their cross-type connections, this representation supports the discovery of structural patterns, dependencies, and influential relationships that may be overlooked in homogeneous network models. Such insights can lead to a more precise understanding of complex systems and improve downstream analytical tasks.

1.2.2 Network Characteristic Extraction

Network characteristic extraction involves deriving structure-aware features from networks. These features are typically presented as a representative vector that encapsulates a network’s key topological and entity information in a more manageable format. This process provides a robust foundation for various network analysis tasks, facilitating the examination of complex networks. In this context, I discuss two primary research directions: defining network metrics and generating network representations using machine learning techniques.

1.2.2.1 Network Measures

Network measures quantify various structural functions for either a node or an entire graph. For instance, the *degree* of node k is a node-level measure, indicating the number of nodes connected to k . When examining the social activities of college students on Facebook, the *degree* can gauge how actively a student interacts with others on Facebook. Another example is the graph-level measure, *density*, which represents the edge density of a network. It is calculated as the number of existing edges divided by the maximum possible number of edges. *Density* may offer a rough summary of the social activity level among students in the same network and could be compared to the *density* of students in another network. A full list of network measures and their definitions can be found in [1].

There are also efforts to define temporal measures for dynamic networks. A *time-respecting path* [2] is introduced to incorporate the evolving structure, with an additional condition compared to static paths. This condition requires that any two adjacent edges' timestamps occur in the same temporal order. Additionally, to fully utilize recorded timestamps, *latency* [3] is defined to measure the difference between timestamps of the first and the last edge in a *time-respecting path*. In the context of analyzing collaborations among healthcare professionals (HCPs), *latency* may indicate the time it takes for a piece of patient information to be transmitted from the origin to the target HCP.

1.2.2.2 Machine Learning on Networks

Machine learning (ML) involves computer programs refining their performance on tasks by iteratively adjusting their internal parameters by analyzing patterns in data examples, enabling them to improve decision-making abilities over time. Neural network (NN) [4] is one of the subfield of ML, whose learning mechanism is inspired by the function workflow of human brain. A straightforward example of an NN is the multi-layer perceptron (MLP), which consists of layers of fully interconnected artificial neurons. In an MLP, information flows from the input layer to the output layer through weighted connections and activation functions. As the input data traverses through different combinations of neurons and nonlinear activation functions, it undergoes nonlinear transformations that preserve essential features necessary for the intended tasks, such as classification. During the training process, the MLP adjusts the weights of its connections to minimize prediction errors. This adjustment is achieved using optimization algorithms like gradient descent, which leverages observations from the training data. Once trained, the MLP can generalize its learning to make predictions on similar, unseen data for the same task.

Given the remarkable success of neural networks (NN) in discerning intricate data patterns and relationships, researchers have extended this concept to graph data. Graph Neural Networks (GNNs) are crafted to incorporate graph topology into the training process. GNNs process graph data by iteratively exchanging information among nodes. Each node begins with an initial representation, which is then updated based on the aggregated information from its direct neighbors. The aggregation mechanism resembles that of an

artificial neuron, where information from each neighbor undergoes weighted connections and is combined using a chosen aggregation function (e.g., summation or maximum). This combined information, optionally passed through an activation function, is then integrated with the original representation of the node to update its representation. As nodes repeatedly aggregate information from neighbors and their neighbors, they gradually assimilate both local and global graph details, enabling GNNs to glean informative node representations. The acquired node representations serve as input for diverse downstream tasks, including node classification, edge prediction, and graph classification.

1.2.3 Network Feature Interpretation

Interpretability is often an overlooked aspect when designing network feature extraction methods. Although these extracted features may achieve high performance in network-related tasks, the inability to connect features to actual network topology hinders analysts from making actionable insights. This challenge is particularly significant for machine learning algorithms, often considered black boxes due to their complex learning mechanisms, making it difficult to understand how a network is transformed into a learned representation. In this context, I discuss two ways of deploying interpretation techniques in machine learning algorithms: post-hoc approaches and explainability-by-design approaches.

Post-hoc methods entail interpreting ML models after training, representing the majority of existing interpretation techniques. For instance, when explaining the behavior of a GNN, GNNExplainer [5] trains an additional machine learning model to generate a subgraph deemed most relevant to the classifications made by the trained GNN. As an example, when identifying the network structure associated with a molecule’s mutagenicity, the explanatory subgraph generated by GNNExplainer is expected to include known mutagenic structures, such as a 6-atom carbon ring. Explainability-by-design approaches are integrated directly into the machine learning models’ learning processes, eliminating the need for additional “add-on” models. For instance, gradient-weighted class activation mapping (grad-CAM) [6] utilizes the gradients within the ML model to identify significant regions in the input image that influence the final prediction outcomes in computer vision

tasks. As an example, when classifying an image as a cat, grad-CAM can highlight the specific area of the image that contributes most to the model’s decision, ideally focusing on the cat itself gradually over the learning process.

Beyond distinguishing methods based on how their interpretation techniques are deployed, these techniques can also be categorized into gradient-based, perturbation-based, and surrogate-based methods. Gradient-based methods utilize gradients as approximations of input importance to map outputs back to inputs. For example, grad-CAM [6], which we discussed earlier, leverages gradients within the model to identify significant regions in input images that influence the final prediction outcomes. Perturbation-based techniques examine the impact on outputs when different perturbations are introduced to the inputs. The underlying intuition is that if important input information remains unchanged, the output should remain similar to the original. For instance, GNNExplainer [5] trains an additional ML model to generate masks as perturbations for the original graphs, maximizing the mutual information between the predictions of the original and perturbed graphs. Surrogate-based approaches employ an interpretable surrogate model to approximate the output of a complex ML model for a neighboring data space of the input. Although promising, defining a suitable surrogate model and the neighboring space of the input is often challenging.

1.2.4 Visual Analytics of Networks

Visual analytics (VA) is the science of guiding analytical processes through interactions between users and visualizations. This section provides an overview of why VA is a suitable tool for network analysis and how it is integrated into the analytical process.

Due to the nature of network data, which encapsulates both entities and their relationships, VA is an ideal tool for examining entity proximity and structural information. For example, visualizing a network and encoding the values of a network measure as node colors allows for the clear identification of nodes with different measure values. By combining the observation of node connections and proximity, this visualization facilitates an intuitive understanding of the structural function behind the network measure.

For complex networks, VA has demonstrated its utility in breaking down intricate

data into perceivable and meaningful portions. Considering a dynamic network, where its structure depends on the presence of nodes and edges at any given moment, presenting all information in a single visualization is challenging. One approach, the *1.5D Timeline* [7, 8], adopts an egocentric method that focuses on a single node’s timeline. It connects each neighboring node to this timeline at the timestamp when the edge between the pair of nodes was formed. By incorporating interactive elements, users can explore the evolving relationships over time for each individual of interest.

As hinted in the previous paragraph, VA integrates the intended analysis task into the visualization design process. Depending on the network analysis task, coordinated visualizations in a VA system provide distinct cues (e.g., clusters or correlations) that guide users to perform interactions (e.g., selection) and initiate the next analysis step. For example, GNNLens [9] aims to diagnose error behaviors of GNNs by grouping misclassified nodes based on their properties. Users can select a group of misclassified nodes and examine projections of these nodes based on different structural metrics in separate visualizations. This approach provides users with ample clues to interpret the causes of error patterns within the selected group of nodes.

In sum, VA is designed to be an integral part of analysis processes. Each visualization must guide user interactions and provide an intuitive understanding of the data. By iteratively exploring and investigating network data through interactions with visualizations in the VA system, users can derive analytical insights and gain a deeper understanding of complex networks.

1.3 Content Overview

This proposal comprises of six chapters. Each of the next four chapters introduces a work contributing to analyzing and reasoning a type of complex network under the motivation described in Sec. 1.1.

Chapter 2 presents a method for studying multivariate networks. To uncover the relations of interest in these networks, I design an ML-assisted analysis workflow that captures combinatorial influence from a group of attributes on another. This allows us

to answer questions such as: what combinations of attributes are largely related to the density of a social network? Following this, a visual analytics system is developed to enable analysts to test their hypotheses, thereby facilitating analytical reasoning. The capability of this work is demonstrated through case studies using real-world multivariate social networks, where users are able to unveil and interpret combinatorial effect from attributes of interest. The insights derived from these studies are further evaluated by social scientists through interviews.

Chapter 3 introduces an approach for investigating dynamic networks extracted from collaborations among healthcare professionals (HCPs). I deploy an ML-based method to connect network measures to patient survival outcomes, assessing the effectiveness of HCP collaborations and identifying underperforming teams. Additionally, I design a visual analytics system to explore the evolution of collaboration patterns, facilitating the planning of potential interventions.

Chapter 4 describes a method for systematically generating and evaluating explanations for GNNs. This work uses fundamental substructures as the building blocks for GNN explanations. A VA system is developed to allow analysts to explore the potential of each fundamental substructure as an explanation. The system uses metrics defined in this work to quantify the individual relevance of these substructures to GNN predictions, enabling the generation of trustworthy explanations. By comparing the explanations generated by this method to those produced by the state-of-the-art GNNExplainer [5], the versatility and utility of the approach are demonstrated.

Finally, Chapter 5 presents an interactive visual analytics framework for interpretable predictive analysis and what-if simulation. Building upon the methods developed in Chapters 3 and 4, this chapter extends dynamic healthcare collaboration network analysis into a framework for patient risk prediction, model explanation, and intervention planning. The framework supports exploration of predicted patient morbidity risk using temporal Graph Neural Networks (TGNNs), identification of influential collaboration patterns, and what-if analysis through targeted interventions to observe projected risk changes. This enables users to generate and evaluate hypotheses for improving patient outcomes.

Chapter 2

Combinatorial Influence Extraction for Multivariate Networks

Multivariate networks are commonly found in real-world data-driven applications. Uncovering and understanding the relations of interest in multivariate networks is not a trivial task. This chapter presents a visual analytics workflow for studying multivariate networks to extract associations between different structural and semantic characteristics of the networks (e.g., what are the combinations of attributes largely relating to the density of a social network?). The workflow consists of a neural-network-based learning phase to classify the data based on the chosen input and output attributes, a dimensionality reduction and optimization phase to produce a simplified set of results for examination, and finally an interpreting phase conducted by the user through an interactive visualization interface. A key part of our design is a composite variable construction step that remodels nonlinear features obtained by neural networks into linear features that are intuitive to interpret. I demonstrate the capabilities of this workflow with multiple case studies on networks derived from social media usage and also evaluate the workflow through an expert interview.

2.1 Introduction

Multivariate networks [10] consisting of both topological and semantic information can model complex relations between entities. One common analysis task performed on mul-

tivariate networks is to understand associations among structural and semantic characteristics [10–13]. For example, from social media usage, analysts may want to see how likely each possible combination of individual characteristics (e.g., age, gender, and extraversion) and friendship structures is related to their addiction level to social media. Such analysis can be more complicated when the associations underlie intertwining factors.

To aid in analysis of multivariate networks, researchers have introduced visual analytic support, including network layouts considering semantic information, interactive simplification, and incorporation of coordinated views [10, 14]. Among others, utilizing network representation learning (NRL) [15, 16] is one promising approach as it can capture latent features of networks and highlight essential aspects that should be examined with visualizations. Existing visual analytics methods [17–22] utilize NRL methods to learn networks’ *general* representations—overall summaries of networks (e.g., variance of node degrees). However, to identify associations, a general-purpose NRL is not effective as it is not particularly designed to capture latent features important for associations of interest.

On the other hand, as seen in the machine learning (ML) field, NRL using neural networks (NNs) can learn representations specific to an analysis focus by using appropriate loss functions [23]. While these representations are suitable for fully automated analyses, such as network classification, it is often difficult for analysts to interpret the extracted features. This is not preferable when analysts want to be involved in the analysis process and derive insights into associations.

To address the aforementioned problems, we introduce a visual analytics workflow that provides (1) network representations specific to the structural and semantic associations in multivariate networks as well as (2) interpretation supports for the analysis results.

To learn such representations, we first extract essential structural features by using an NRL method and then train an NN model that is designed to classify values of a user-selected attribute (e.g., high and low addiction levels to social media). After training, the model generates latent features that are highly related to the selected attribute.

For the interpretation of the obtained network representations, we employ dimensionality reduction (DR), interactive visualization, and composite variables [24]. To help

analysts assess the quality of the representations, we simplify the latent features with a linear DR method (specifically, linear discriminant analysis or LDA). We then visualize the information of the simplified features in a 2D plot, where the distribution of networks or their elements (i.e., nodes and links) is shown along a latent direction that is related to the classification. From this plot, analysts can judge which part of networks or elements (e.g., subjects with age 20–30) likely holds clearer associations with the selected attribute. This analysis is almost infeasible if we rely only on the classification quality measures. In addition, we introduce a mechanism of composite variable construction to explain the meanings of the network representations. The mechanism first suggests network structures and attributes highly related to the representations by utilizing a model-agnostic interpretation method, the SHAP [25]. Then, after analysts interactively select structures and attributes from the suggestion, an optimization method automatically generates a composite variable that resembles the network representation. By examining this composite variable, analysts can understand the associations among the selected attributes and the other information.

To support effective analysis of the associations, we develop a new density scatterplot, named *two-class density scatterplot*, that can depict a trend of data distribution as well as inform the mixture and separation of two classes even from numerous instances (e.g., 1000 instances). Incorporating two-class density scatterplots, we develop an interactive interface that links visualizations designed to enhance the visual analytics workflow. We further demonstrate the capabilities of the workflow and interface with three case studies using two real-world datasets on social media usage. Moreover, we validate the usability of the workflow through an expert interview.

In summary, we consider our primary contributions to be:

- a visual analytics workflow considering both generation and interpretation of network representations that are expressive for user-specified analysis targets;
- a workflow of composite variable construction that aids in attribute selection and attribute weight optimization for interpretations with an influence from multiple attributes; and

- a two-class density scatterplot as a versatile visualization to review individual- and group-level data patterns with class information.

We provide source code related to the workflow and a demonstration video of interactive analysis [26].

2.2 Related Work

Our work is closely related to two topics in the visualization field: NRL and the interpretation of learned representations. For the broader discussion on visualizations for analyzing networks and interpreting ML results, refer to existing surveys [10, 27–29].

2.2.1 Learning Network Representation

NRL aims to generate a set of low-dimensional vectors (also called representation) that captures certain important characteristics of networks, nodes, or links [15]. The representation is usually learned for downstream tasks, such as node classification and link prediction. Various NRL methods are developed, including node2vec [30], graph convolutional networks [31], graph neural networks (GNNs) with the self-attention [32], to name but a few [15].

The researchers have been utilizing NRL for visualization to interactively examine complex network datasets. For example, Freire et al. [33] represents one network by a set of network statistics (e.g., degree distribution) to compare many networks in a tabular interface. Gove [34] suggests several network-level features (e.g., density) that are easier to interpret and faster to compute for interactive visualization. Other researchers utilize the occurrences of graphlets [35] (small, connected, non-isomorphic subgraph patterns in a network) to identify visually similar networks [36–38]. When node correspondence exists among networks, another common approach is directly applying DR methods to networks’ adjacency matrices to capture the similarities of networks [39, 40]. Van den Elzen et al. [21] takes a similar DR approach while further incorporating network statistics. Martins et al. [19, 20] uses DR to lay out nodes by their structural and semantic similarities.

Similar to ours, recently, a few works employed NN-based NRL. Fujiwara et al. [18] introduces contrastive NRL (cNRL) by integrating a variant of GNNs and contrastive

learning [41]. cNRL extracts a representation of two networks, where salient characteristics in one network relative to another are highlighted. Utilizing linear DR, they further design an interpretable cNRL method and enhanced it with interactive visualizations [17]. Song et al. [22] uses GNNs to support interactive subgraph pattern search, where GNNs are used to covert each network in a comparable, fixed-length latent vector.

Unlike the above approaches, we use NN-based NRL to obtain representations that are specifically for uncovering the associations of interest in multivariate networks. Also, we address the interpretation of network representations with composite variable construction, which is easier to understand when compared with the approaches referring to complex coefficients in linear DR results [17, 18].

2.2.2 Interpreting Representations

Although the interpretation support for NRL is still sparsely studied (e.g., [17]), a variety of interpretation methods are developed to explain high-dimensional data’s representations that are extracted by DR methods or ML models [16, 42–45]. Existing interpretation methods can be mainly categorized into two approaches: (1) identifying essential information to specific patterns found in representations (i.e., post-hoc explanation approach) and (2) constructing simple, interpretable representations during a learning phase (i.e., explainability-by-design approach [46]).

2.2.2.1 Post-Hoc Approach

Within visual analytics methods, researchers have identified influential attributes on the cluster formation in DR results from statistical charts (e.g., boxplots of attributes for each cluster) [47–49]. As univariate statistics are often insufficient to capture the cluster characteristics, researchers further consider influences from multiple attributes [50–54]. For example, Joia et al. [51] applies PCA to each cluster to examine multivariate influences.

Another common visual analytics strategy is utilizing measures that inform how strongly the change of an input attribute value influences an ML inference result (e.g., how many more bikes are rented if the temperature increases one degree) [55–59]. Such measures include partial dependence (PD) and SHAP values [25]. For example, Krause et al. [55] develops a system to interactively investigate PD. Their system only supports analysis of

PD that is obtained by changing each attribute’s values individually—interpretation only from a single attribute level. Angelini et al. [59] introduces a visual analytics framework that can even deal with PD derived from a simultaneous change of multiple attributes.

Similar to Angelini et al.’s work [59], composite variable construction in our workflow also allows a multi-attribute level interpretation. However, PD or other perturbation-based measures (e.g., SHAP values) merely output a value for each perturbed input. Consequently, it is not trivial to induce numerical relationships between input attributes and ML inference results. In contrast, our workflow’s composite variable provides a linear function that summarizes the associations among multiple attributes. This ability helps us perform the interpretation both efficiently and intuitively.

2.2.2.2 Explainability-By-Design Approach

When compared with the post-hoc approach, the explainability-by-design approach is taken by a smaller number of visual analytics works. Knittle et al. [60] uses NNs consisting of one hidden layer with a small number of NN nodes to extract nonlinear representations that relate input attributes to a target output attribute. This simple NN allows analysts to identify a small number of representations that show clear associations between the input and target attributes. Gleicher [61] produces simple composite variables that are to classify a user-selected attribute. To craft such composite variables, Gleicher performs a support-vector machine-based exhaustive search for the selection of composing variables while considering a balance between their simplicity and expressiveness.

In terms of using NNs to extract the input-output relationships, the work by Knittle et al. [60] is closely related to ours. However, their interpretation of the obtained representations is based only on univariate value distributions, which is insufficient when NNs capture complex input-output relationships. Similar to Gleicher’s work [61], our work crafts simple composite variables, but we do not involve the computationally expensive exhaustive search. Instead, we rank variables based on their contributions to the NNs’ predictions and involve analysts’ knowledge to select attributes of interest.

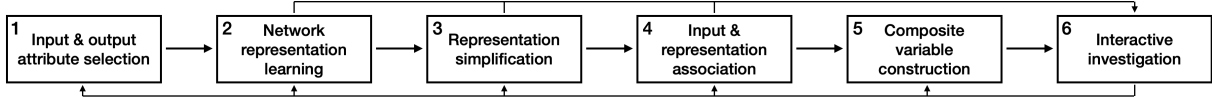


Figure 2.1: The visual analytics workflow for investigating associations in multivariate networks, where Steps 1–4 are executed with a script for machine learning and Steps 5–6 are conducted interactively with our UI.

2.3 Methodology

This section introduces our visual analytics workflow and two-class density scatterplot. We first describe design considerations of the workflow, which we identify based on our literature survey in Sec. 2.2 and targeted analysis on multivariate networks. We then provide an overview of the workflow, followed by the details of each step.

2.3.1 Design Considerations

The following five design considerations (DCs) are identified to support the task of understanding associations of interest and fill the analytical gap that is not covered by existing methods.

DC1: Flexibility. Analysis of multivariate networks requires reviewing intertwined relationships in both structural and semantic information [10]. Also, based on analysis purposes and available datasets, analysts often want to investigate networks from different levels, such as node, link, and network levels (e.g., node: [17, 20]; link: [62]; and network levels: [21, 22, 33, 38]). The workflow should provide flexibility to conduct analyses from multiple aspects at various levels.

DC2: Expressivity. The main objective of our work is to support examining associations among target (e.g., the addiction level to social media) and other related information (e.g., gender, age, and friendships). The workflow should be able to extract features that are expressive for this analysis task (i.e., containing important information to reveal associations) rather than providing a general summary of networks.

DC3: Interpretability. Interpretation is essential to gain insights into analysis results [60, 61]. Because expressive features such as those extracted by NNs are often difficult to interpret only from each individual attribute [17, 63], the workflow should provide more advanced interpretation support that considers the combined influence from multiple attributes.

DC4: Tunability. Conflict often exists between the expressiveness and interpretability of features (i.e., more expressive, more difficult to interpret) [61, 64, 65]. Also, the required preciseness of interpretation can be varied by analysis (i.e., complicated but precise interpretation vs. less precise but simpler interpretation). The workflow should allow control of the balance of expressiveness and interoperability.

DC5: Extensibility. Although we provide a concretized method for each step of the workflow, a different set of methods may be more suitable according to the characteristics of future datasets and analysis goals [15, 66–68]. The workflow should be extensible to incorporate new or different methods.

2.3.2 Workflow Overview

Fig. 2.1 shows our workflow for reviewing associations embedded in multivariate networks. It consists of six steps that are first conducted with a script for ML (Steps 1–4) and then interactively performed with our user interface (UI) designed for visual analysis (Steps 5–6). The UI is shown in Fig. 2.2. While the UI presents the results from Steps 1–4, the UI is not used to perform interactive parameter adjustment for Steps 1–4. These ML steps usually have various adjustable parameters (e.g., the numbers of NN layers and nodes) as well as require relatively long computation (e.g., 5 minutes in total). Consequently, for these steps, adjusting and running an ML script is more convenient to perform flexible analysis.

As we describe in Sec. 2.3.5, we extract structural information as a set of network measures (e.g., degree and betweenness centralities) for each element (i.e., node, link, or network); thus, in the following, we use a term, *attribute*, to indicate both extracted structural (e.g., degree) and semantic information (e.g., age).

Step 1. The workflow begins with the extraction and selection of input attributes and one output/target attribute. For example, to know how students’ grades in a school class are related to their surrounding conditions, an analyst can choose their attending class size, mental health status (e.g., depression level), and centrality in social media connections (e.g., degree) as inputs and their grade as an output.

Step 2. NRL is performed based on the selected inputs and output. A network

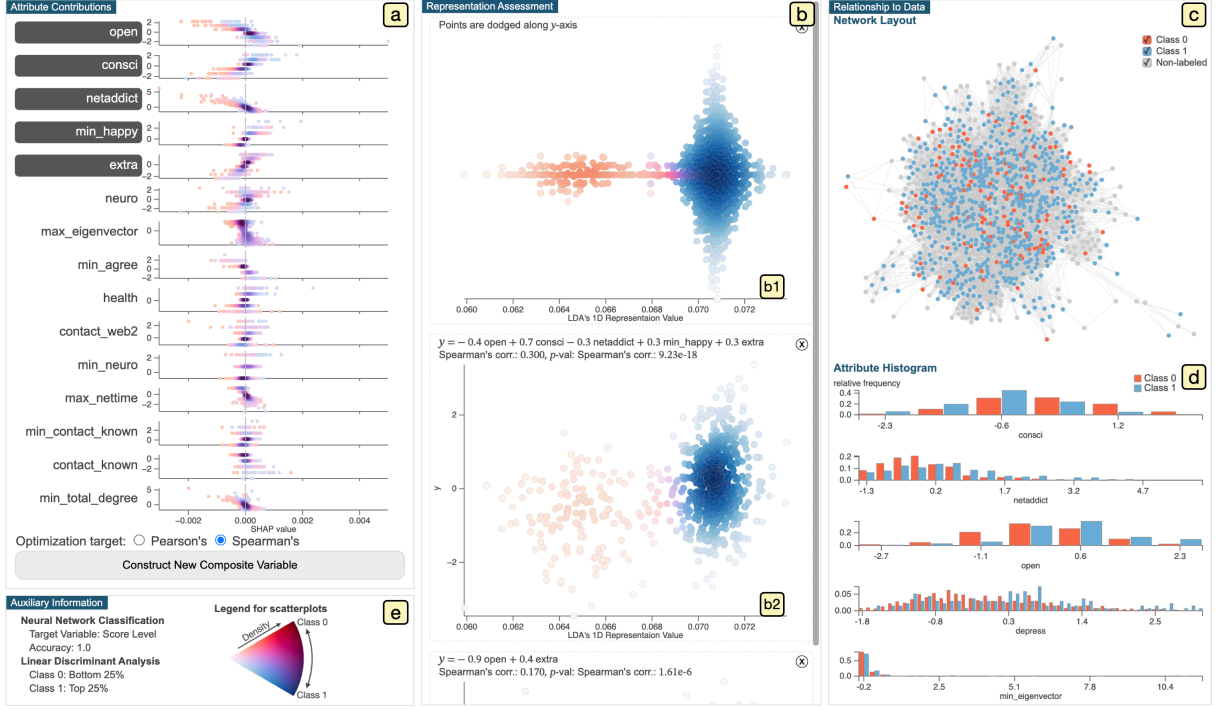


Figure 2.2: The visual interface for facilitating interactive analysis using the workflow in Fig. 2.1. (a) Relating to Step 4, this view visualizes input attributes' contributions to the 1D network representation. (b) Composite variables generated in Step 5 are shown as a list of scatterplots, where the 1D network representation and composite variable correspond to x - and y -axes, respectively (see b2). As shown in b1, when a composite variable is not generated yet, a swarm-plot-like visualization presents the 1D network representation generated through Step 3 to help assess its quality. For a and b, we employ our two-class density scatterplots. (c) A node-link diagram and (d) a set of histograms inform the network structure and attribute distributions. (e) Other auxiliary information is displayed, including the prediction accuracy of NNs trained for Step 2.

representation is generated by an NN trained to perform binary prediction of the output (e.g., good or bad grade) with the inputs. The derived representation directly relates to associations of interest.

Step 3. This step compresses the dimensionality of the network representation into one while preserving the prediction quality as much as possible. This makes the remaining steps simpler to complete and the related results easier to interpret.

Step 4. This step evaluates and ranks each input attribute's contribution to the 1D compressed representation. The obtained ranks can be considered as recommendation levels for the inclusion of the corresponding attributes for composite variable construction in Step 5.

Step 5. Finally, based on the recommendation levels and interests, an analyst manually selects a small set of attributes (e.g., 2–5 attributes) and runs a composite variable construction algorithm. The optimized composite variable (e.g., y -axis in Fig. 2.2-b2)

maximally resembles the 1D compressed representation (e.g., x -axis in Fig. 2.2-b2). This composite variable provides an intuitive explanation of how the selected small set of attributes is related to the output/target attribute.

Step 6. The UI in Fig. 2.2 visualizes the information related to each previous step as well as the detailed structural and semantic information of the multivariate network. By interactively reviewing visualizations, the analyst can gain insights or adjust settings for each step based on their need (the backward arrows in Fig. 2.1).

2.3.3 Two-class Density Scatterplot

We here introduce a *two-class density scatterplot* (Fig. 2.3-d), which we use to visualize results from most of the steps in the workflow. From a scatterplot of two variables (e.g., the 1D compressed representation and the composite variable generated at Steps 4–5), we aim to simultaneously depict (1) patterns related to binary classes; (2) trends including correlations; (3) other supplemental patterns, such as noises, outliers, and clusters. For example, analysts may want to review class separation to assess the quality of network representations, correlation patterns related to the composite variable to gain insights, and patterns only seen in subgroups to consider further analyses.

According to a survey on scatterplot designs [69], we should employ an aggregated-level visual encoding (e.g., binning-based encoding) of instances to reveal trends while we should show individual instances for the other tasks (e.g., finding outliers). However, designing an encoding that satisfies these requirements is not straightforward. For example, in Fig. 2.3-a, class information is encoded with colored dots. This encoding easily suffers from overplotting when analyzing a large dataset (e.g., 1000 instances), hindering the observation of the trend of data. In contrast, density scatterplots are effective in finding trends. For instance, Fig. 2.3-b reveals a correlated pattern. However, from density scatterplots, we cannot grasp distributions related to classes (e.g., whether the dense area mostly consists of a single class). A few existing scatterplot enhancements, splatterplots [70] and winglets [71], can show both aggregated- and instance-level information along with class information. However, based on our experiment [26], the trend of data remains unclear in both enhancements. Also, these enhancements are sensitive to the

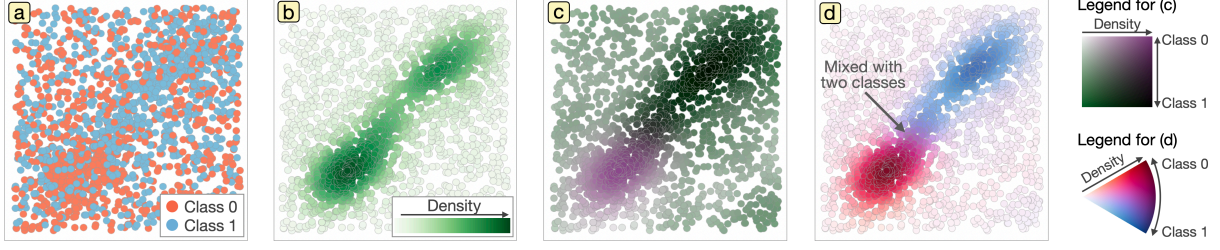


Figure 2.3: The comparison of scatterplot designs: (a) scatterplot with colored classes, (b) density scatterplot, (c,d) scatterplots encoding the total density and the ratio of each class’s density with two different bivariate colormaps. (d) is our final design for a two-class density scatterplot.

choice of parameters of their algorithms (e.g., threshold for contouring) and visualizations (e.g., dot size).

A two-class density scatterplot (Fig. 2.3-d) is designed to solve the above issues. Using a bivariate colormap, a two-class density scatterplot encodes the *total density* from both classes and the *density ratio* of each class with color lightness and hue, respectively. Fig. 2.3-d demonstrates how our design helps identify trends and class-related information from numerous instances. First, it reveals the correlated pattern as in the density scatterplot. In addition, we can confirm that two dense areas at the bottom left (more red colors) and top right (more blue colors) mainly consist of a single class. Moreover, from the purple-color area annotated with the arrow, we can observe that a part of the dense areas mixes a considerable amount of both classes’ instances. We should note that a bivariate colormap [72] has been already utilized to simultaneously show two different measures over scatterplots (e.g., [73]). Our contribution is in the computation of the ratio of each class’s density and the color encoding design that is suitable to depict class and density information with scatterplots.

Computation of density-related information. Similar to an ordinary density scatterplot, we first estimate a 2D probability density function (PDF) from the coordinates of instances, using a Gaussian kernel density estimation. Let f_{all} denote a 2D PDF estimated with all instances (i.e., the total density). Then, the density at a 2D coordinate, $\mathbf{p}$, can be computed with $f_{\text{all}}(\mathbf{p})$. Similarly, let f_0 and f_1 denote 2D PDFs estimated with **Class 0**’s instances and **Class 1**’s instances, respectively. Also, let n_0 and n_1 be the numbers of instances in **Class 0** and **Class 1**. Then, a function that estimates the density ratio of **Class 0** at $\mathbf{p}$ can be written as: $g_0(\mathbf{p}) = n_0 f_0(\mathbf{p}) / (n_0 f_0(\mathbf{p}) + n_1 f_1(\mathbf{p}))$. If

assigning the same total density for each class is more appropriate for an analysis, we can use $g_0(\mathbf{p}) = f_0(\mathbf{p}) / (f_0(\mathbf{p}) + f_1(\mathbf{p}))$, instead.

Visual encoding design. Using a bivariate colormap, we encode $f_{\text{all}}(\mathbf{p})$ and $g_0(\mathbf{p})$. We select color lightness to encode the total density, $f_{\text{all}}(\mathbf{p})$, because of its natural metaphor: the more accumulation of dots, the denser color. We then employ hue to represent the density ratio, $g_0(\mathbf{p})$. To use lightness and hue, we initially try ordinary color spaces such as HSL; however, a subtle difference in the density is difficult to recognize from the generated colors. Thus, we decide to utilize existing carefully designed sequential colormaps. Specifically, we use the red and blue sequential colormaps in Matplotlib [74]. We can obtain a pair of colors corresponding to $f_{\text{all}}(\mathbf{p})$ from these two colormaps and then generate a linearly interpolated color from the pair based on $g_0(\mathbf{p})$. To show this 2D color space as a legend, we use a polar coordinate (see Fig. 2.3). Similar to Correll et al’s colormap [75], this is to clearly indicate the difference between the meanings of two measures (total density and ratio) as well as to inform that the ratio difference is less emphasized in low total-density areas. As shown in Fig. 2.3-c, we also test existing bivariate colormaps such as one introduced by Lespinats and Aupetit [73]. This colormap is not suitable for two-density scatterplots. For example, it is difficult to recognize **Class 1**’s high-density area and the density difference between the high-density areas of **Class 0** and **Class 1**. Sec. A.1 provides more comprehensive comparison of various colormap designs.

2.3.4 Datasets: Students and Adults’ Network Resources

In the rest of this section, our explanations refer to a real-world dataset, Dataset I, as a concrete analysis target. In addition to this dataset, we also analyze Dataset II in our case studies (Sec. 2.4). Both datasets are derived from the Facebook usage and survey data of representative cohorts in Taiwan. While Dataset I is a single multivariate network, Dataset II consists of a set of egocentric multivariate networks. The subjects authorized the use of their contact records on Facebook and survey answers. The use of data is approved by the IRB on Humanities & Social Science Research, Academia Sinica, Taiwan.

Dataset I is about senior college students [76]. We represent each student as a

network node and create a link based on their interactions on Facebook. We construct an undirected link if a “like”, “comment”, or “tagged” was made between students on a post. We consider such links as “friendships” in social media and 1-hop neighbors of each node as “friends”. We further use the answers for the survey as each node’s attributes. The corresponding questions are related to the students’ school life, such as their personalities, moods, and grades. The resultant network consists of 1886 nodes, 64156 links, and 226 survey-based attributes for each node.

Dataset II is collected from sampled adults [77]. This dataset uses Taiwan Social Change Survey, which includes important network-related topics, such as core discussion networks and network resources, as well as personal attributes, such as their backgrounds, personalities, moods, and social lives. We generate multivariate egocentric networks by adding 1- and 2-hop neighbors of each survey respondent based on their contact records (i.e., 1-hop: direct contacts, 2-hop: indirect contacts). We also assign network-level attributes, such as the ego node’s personality, Facebook usage statistics, and network measures (e.g., clustering coefficient). The resultant egocentric networks consist of 345 ego nodes, 3339 1-hop neighbors, 19917 2-hop neighbors, and 11 network-level attributes. In addition, we have information about whether each indirect contact is promoted to direct during a study period, January 1, 2015–June 30, 2017.

Below, from Dataset I, we identify and review college students’ attributes that are highly related to their `scorelevel`—the past three years’ mean score levels reported on a scale of four from the top to bottom. We use 14 survey attributes related to this analysis.

2.3.5 Learning Representations

We explain Steps 1–3, where we learn representations from multivariate networks. To provide a concise explanation, here we describe each process for producing a latent vector of each node (i.e., node feature learning). However, the described learning processes can be easily adjusted to learn a latent vector for each link or network (DC1: Flexibility). Sec. 2.4.3 demonstrates a case for network feature learning.

2.3.5.1 Structural Feature Extraction for Attribute Selection

To select attributes related to the structural information in Step 1, we first precompute a set of structural measures [17, 21]. For node feature learning, such measures can be degree, eigenvector, betweenness centralities, and many others [1].

These centralities inform the importance of each node from different aspects. Degree centrality (i.e., the number of links connected to a node) measures local importance of nodes in the network. Eigenvector centrality (i.e., the eigenvector corresponding to the greatest eigenvalue of the network’s adjacency matrix) more reflects global importance of nodes than degree centrality. This is because eigenvector centrality considers the transitive importance of links (e.g., a link to a node with 100 links is more important than one with 1 link). Betweenness centrality (the number of shortest paths that pass through a node) indicates a node’s importance as a “bridge” connecting other nodes. For details, refer to an introduction to network theory by Newman [1].

To capture more detailed structural information, we precompute statistics of each node’s neighbors, such as the mean and variance of neighbors’ degrees. Such statistics can be computed for semantic attributes as well (e.g., the mean age of neighbors). The neighbor statistics are shown to be useful for various analytical tasks [17, 78]. For this precomputation, we specifically use DeepGL [78] to achieve fast computation as well as to avoid producing redundant measures.

DeepGL summarizes 1-hop neighbors’ centrality/attribute values (called *base features*) with *relational functions*. For relational functions, we can specify multiple neighbor types (e.g., in-, out-, and total-neighbors) and aggregation functions (e.g., mean, variance, and maximum). After obtaining the neighbor statistics for all nodes, DeepGL prunes redundant statistics that show similar value distributions to others. The strength of pruning can be controlled with a hyperparameter of DeepGL. Moreover, DeepGL can repeatedly apply the relational functions and pruning process to summarize farther neighbors’ features. For more details, refer to the work by Rossi et al. [78]. For Dataset I, as base features, we select 3 fundamental network centralities, degree, eigenvector, and betweenness, and the 14 survey attributes; then, we generate 1-hop total-neighbor statistics of

the 17 attributes, resulting in 85 attributes in total after the pruning process.

2.3.5.2 Network Representation Learning

After selecting input and output attributes in Step 1, we first apply the Z-score normalization to each attribute. We then train an NN to predict the output from inputs, and extract representations from an NN layer (DC2: Expressivity). Although we could predict an output value directly by designing the NN for a regression task, we instead categorize output values and perform binary classification of two value ends (e.g., top and bottom 25% of attribute values). Since the classification is performed only on these ends, unrelated instances are not involved in training. If we perform regression instead, the NN tries to fit all instances. Consequently, a resulting representation may be only useful to predict values in the middle range (e.g., when fitting to the two ends is harder than the middle range). While we assume this focus on two ends is reasonable for many analyses, NNs' settings are easy to adjust based on analytical interests. Another benefit of the binary-classification design is enabling a comparison of two categorical groups (e.g., students who major in engineering and business).

By default, we use a multi-layer perceptron (MLP) consisting of five fully-connected layers (one input, three hidden, and one output layers) using the leaky rectified linear unit as an activation function. We then take the last hidden layer's activations as a learned representation. With the consideration of a balance between expressivity and interpretability, we here keep the numbers and sizes of layers sufficiently small based on the size of the input data. On the other hand, we should note that overfitting is acceptable for this step to some extent. We simplify this step's high-dimensional network representation with linear transformation in Step 3 (Sec. 2.3.5.3). This simplification mitigates overfitting when it occurs. The NN-based learning in Step 2 should focus on achieving an accurate alignment of the defined classes in a learned representation (i.e., high classification accuracy) rather than avoiding overfitting. We discuss more details in Sec. 2.3.5.3.

While the simple five-layer MLP is our default NN architecture, this architecture can be easily replaced with more complex ones (DC5: Extensibility). For example, when

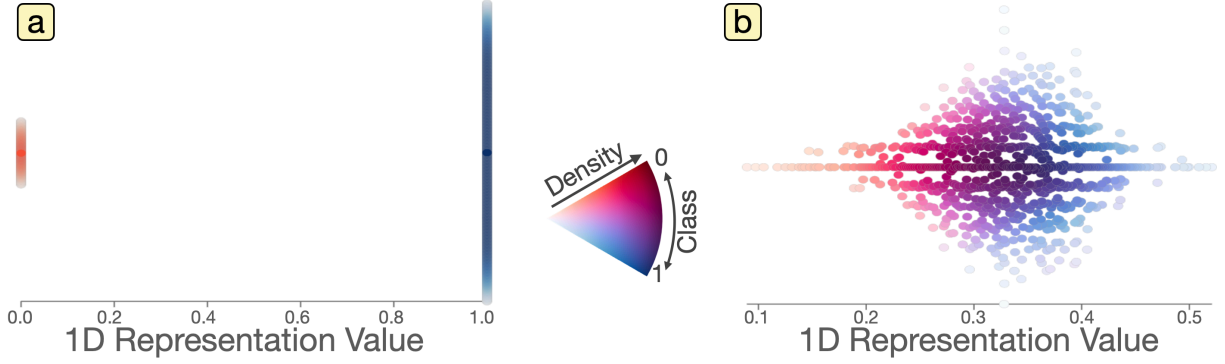


Figure 2.4: 1D representations generated by alternative designs. The same dataset and visualization as Fig. 2.2-b1 are used (red: Class 0, blue: Class 1). (a) is made by applying a softmax activation function at the last layer of an MLP. (b) is generated by directly applying LDA to the dataset.

analysis needs to capture nonlinear neighbor relations in a network, we can incorporate a GNN. If learning latent vectors of links or networks is required, we can replace the precomputation of network measures, accordingly. For instance, we can extract the mean degree and network diameter for network feature learning. Then, we can apply the same procedure employing an NN.

2.3.5.3 Representation Simplification

In Step 3, we simplify the high-dimensional network representation into a 1D representation for the subsequent interpretation steps (DC3: Interpretability). Because the NN is trained for binary classification, this simplification is similar to the process NNs usually perform for their output layer: transforming the last hidden layer’s activations of multiple NN nodes into a single NN node at the output layer. Therefore, we expect that although the simplified representation is 1D, it can preserve sufficient information for the prediction (DC2: Expressivity).

Our simplification employs linear transformation in contrast to nonlinear transformation. This linear transformation is the major difference from conventional NNs. A nonlinear activation function such as a sigmoid or softmax function encourages producing a value of 0 or 1 for classification. As a result, the corresponding 1D representations often consist of many values close to 0 or 1. An example of such cases is shown in Fig. 2.4-a. This close-to-discrete distribution makes it difficult not only to perform optimization for composite variable construction but also to observe insightful patterns from the composite

variable visualization shown in Fig. 2.2-b.

One way to obtain a 1D representation with linear transformation is using a linear activation function between the last hidden and output layers, as in NNs for regression tasks. As stated in Sec. 2.3.5.2, we, however, want to avoid directly performing regression in the NN. Instead, as a post-hoc process, we apply a DR method to the high-dimensional network representation. We specifically use LDA, which places instances of different classes as far as possible in a low-dimensional representation. This placement is achieved by maximizing the distance of each class’s centroid while minimizing data variance within each class. This post-hoc process can efficiently generate a 1D representation with optimal class separation.

On the other hand, even though feature selection methods, such as LDA, are low in computational complexity, LDA can still cause overfitting when the number of instances is relatively small to the number of input features [79]. Thus, we further regularize LDA, which has a regularization mechanism to avoid overfitting [79]. The use of regularized LDA brings an additional benefit for the post-hoc approach: the regularization strength can be controlled outside of the NN training process. Although the $L1$ and $L2$ regularizations can also be used for NNs to avoid overfitting [80], the hyperparameter adjustment for these regularizations often requires many trials and errors, and becomes time-consuming by retraining NNs many times. Applying regularization inside of LDA does not require retraining of NNs and LDA is computationally efficient; thus, we can adjust the regularization strength more conveniently (DC4: Tunability).

The quality of the 1D representation can be investigated with a swarm-plot-like visualization, as shown in Fig. 2.2-b1. This visualization shows an instance as a dot and a 1D representation value as an x -coordinate. Similar to the swarm plot [81], y -direction is used to pile up dots with positional jitters to mitigate overplotting with a limited vertical space. The resultant area height roughly shows the frequency/density around the corresponding x -coordinate (similar to a histogram). We select this instance/point-based visualization to keep it consistent with other visualizations that use x -coordinates to represent the 1D representation values (cf. Fig. 2.2-b2). The color encoding is as with two-class density

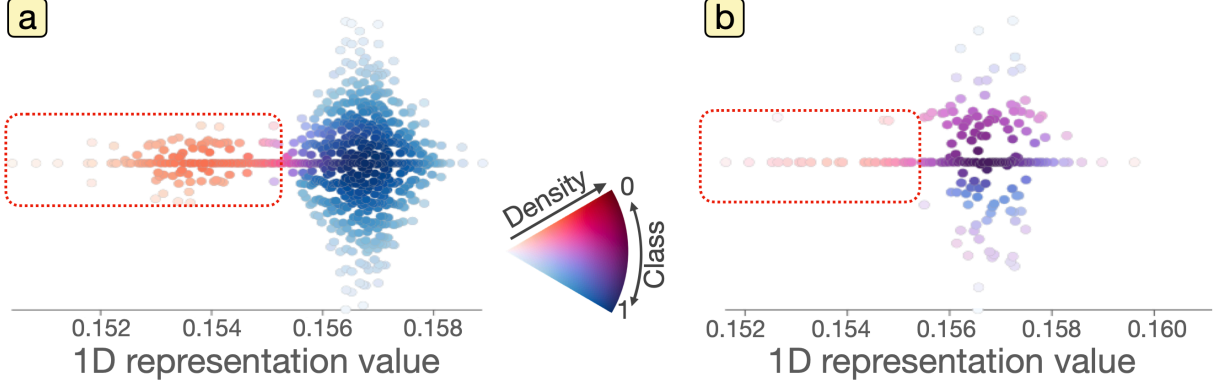


Figure 2.5: 1D representations generated by training the same NRL architecture as the Fig. 2.2-b1. Unlike Fig. 2.2-b1, the NRL architecture was trained with a subset of the data (616 instances). (a) and (b) correspond to the 1D representations of the training set and the remaining test set (170 instances).

scatterplots’. When the 1D representation has good quality, as seen in Fig. 2.2-b1, two classes (**Class 0**: red, **Class 1**: blue) should have small or no overlaps. When the quality is not satisfactory, the analyst can update the NN architecture and/or the selection of input and output attributes. For example, the analyst should consider increasing the number of NN nodes and/or layers when the data has many instances and attributes. Similarly, if there are more attributes that NNs can utilize, the inclusion of them might reduce the overlaps.

The 1D representation in Fig. 2.2-b1 exhibits how the simplification using linear transformation relaxes the overfitted representation. Unlike the representation directly output from the MLP shown in Fig. 2.4-a, Fig. 2.2-b1 shows a continuous distribution of **Classes 0** and **1**. To further check if a simplified representation captures general data trends, we experimentally generated 1D representations while using separate training and test datasets. The resultant 1D representations are shown in Fig. 2.5. In Fig. 2.5-a, we can see the 1D representation generated from only a subset of the data (i.e., training set) still shows a similar pattern to Fig. 2.2-b1. Then, in Fig. 2.5-b, although there is a less clear separation between **Classes 0** and **1**, we can still observe a similar pattern to Fig. 2.5-a. For example, a considerable amount of instances of **Class 0** is placed on the left side. The above observations show that the potential issue of overfitting in MLP becomes less problematic through representation simplification.

In addition to the discussed architectures, we experiment with another alternative

design: directly applying LDA to data to generate a 1D representation. As shown in Fig. 2.4-b, because this design does not exploit the strengths of NNs, the results often mix two classes.

2.3.6 Understanding Representations

We describe Steps 4 and 5, which are mainly for understanding the representation obtained through Steps 1–3 (DC3: Interpretability).

2.3.6.1 Attribute Contribution Measurement

In Step 4, we extract each input attribute’s contribution to the 1D representation, which is useful for two tasks. First, this attribute contribution is useful to understand each attribute’s relationship to the output attribute—the interpretation from a *single attribute* level. Second, the contribution indicates attributes that should be considered for composite variables—the interpretation considering influence from *multiple attributes*.

We employ the SHAP method [25] to measure the attribute contributions. The SHAP method calculates a measure, called the SHAP value for each instance. The SHAP value is computed based on cooperative game theory and shows how much a target value will be changed by adding each attribute value in a model (refer to [25] for details). We apply the SHAP method to a transferred NN model that combines the MLP trained in Step 2 and LDA trained in Step 3. In this case, the SHAP value indicates how much having a corresponding attribute value contributes to moving an instance toward a positive direction of the 1D representation. For example, when an instance’s **age** is 25 and its SHAP value for **age** is 0.1, having this age shifts the instance to a more positive side of the 1D representation by the magnitude of 0.1.

Visualization. As shown in Fig. 2.2-a (also Fig. 2.6-c), we visualize SHAP values with a list of two-class density scatterplots, where each row corresponds to one attribute’s SHAP values. Our design uses *x*- and *y*-axes to depict the SHAP values and attribute values, respectively. As a summary measure of each attribute’s contribution, we use the mean absolute value of all instances’ SHAP values. We then sort the display order of the scatterplots based on this summary measure (i.e., top contributed attributes are listed from the top in order).

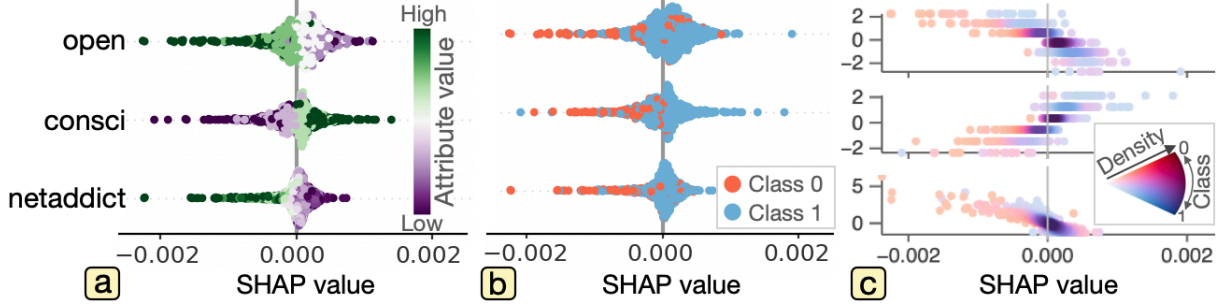


Figure 2.6: Comparison of SHAP value visualizations: (a) the default plot in the SHAP package [82]; (b) the plot colored by class; and (c) our design.

For example, in Fig. 2.2-a, we see the top 3 contributed attributes are `open` (openness), `consci` (conscientiousness), and `netaddict` (internet addiction). Furthermore, from the corresponding scatterplots, we can gain several insights. While larger `open` has a more negative impact on the 1D representation reflecting `scorelevel`, larger `consci` has a more positive impact. Also, `netaddict` shows clearly different patterns by Class 0 (low `scorelevel`) and Class 1 (high `scorelevel`). Larger `netaddict` tends to have a more negative impact for Class 0.

We test different visual designs, as shown in Fig. 2.6. Fig. 2.6-a follows the default plot in the SHAP Python package [82], where the swarm-plot-like visualization is used. x -coordinates correspond to SHAP values, while colors inform attribute values. From this visualization, we can roughly see the relationships between these values (e.g., larger `open` tends to have a more negative SHAP value). But, it is infeasible to grasp patterns that differ by class. Even when juxtaposing a visualization that color-encodes class information, as shown in Fig. 2.6-b, we still cannot clearly see such different patterns. In contrast, our final design (Fig. 2.6-c) can clearly depict the relationships between SHAP and attribute values as well as different patterns by class.

2.3.6.2 Composite Variable Construction

In Step 5, we construct composite variables interactively. The composite variable is commonly used when using a single attribute is not sufficient to understand a target phenomenon (DC2: Expressivity and DC3: Interpretability). For example, in baseball, on-base plus slugging (OPS)—the sum of two different statistics of the player’s performance, on-base percentage and slugging percentage—is used to analyze players’ contributions to

team runs. This is because OPS is a simple composite variable but extremely correlates to team runs [83].

Construction algorithm. We describe the optimization algorithm designed to construct a composite variable from user-selected attributes. To explain the 1D representation with the selected attributes as much as possible, our optimization goal is to maximize a certain *measure of dependence* [84] between the 1D representation and attributes. Such measures include Pearson’s correlation, Spearman’s correlation, the mutual information. We constrain a composite variable to a linear combination of attributes to achieve a simple, intuitive interpretation as well as an efficient optimization for interactive construction.

A linear combination that achieves the highest Pearson’s correlation coefficient can be derived by multivariate linear regression. On the other hand, Spearman’s involves a non-differentiable operation to rank the 1D representation values. Due to non-differentiability, we cannot use simple regression or gradient-based solvers. Thus, we employ a gradient-free solver, specifically COBYLA (or constrained optimization by linear approximations) [85] while using the optimized result for Pearson’s as an initial solution. Although the optimization for Pearson’s is much more efficient, Spearman’s could be more effective in some cases. This is because the 1D representation is constructed from the non-linearly transformed attributes by the NN, and the 1D representation may have a nonlinear correlation with the original input attributes. Although we currently only support Pearson’s and Spearman’s correlations, the algorithm can be easily extended for other measures of dependence (e.g., mutual information) by using COBYLA or a gradient-based solver.

Visualization and interactive use. From the view in Fig. 2.2-a, an analyst can select attributes and a measure of dependence, and then generate a composite variable by clicking “Construct New Composite Variable” located at the bottom. The selected attributes are highlighted with gray backgrounds. Then, a two-class density scatterplot (Fig. 2.2-b2) visualizes the relationships between the 1D representation (x -axis) and the constructed composite variable (y -axis). Also, multiple composite variables can be created to compare their expressivity and interpretability (DC4: Tunability). A newly constructed composite variable’s visualization is appended to a list of the scatterplots in a scrollable

view. Each scatterplot can be discarded by clicking a “×” mark at the top right.

In essence, through this composite variable construction process, analysts can utilize their domain knowledge to review the simplified representation. From the attributes ranked by SHAP values, the analyst can judge attributes they should consider. Then, with the help of the above optimization algorithm, the analyst can interactively examine and fine-tune the composite variables to gain insights. Our contribution to the interpretation of representations lies in this analysis flow, starting from the SHAP ranking of individual attributes, incorporating analysts’ judgment, and to performing the construction and interpretation of composite variables.

Analysis examples. For Dataset I, we compute Spearman’s correlation coefficient between each of the top 5 contributed attributes and the 1D representation. The coefficients are `open`: -0.147, `consci`: 0.237, `netaddict`: 0.134, `min_happy` (the minimum of friends’ happiness levels): 0.093, and `extra` (extraversion): 0.026. As shown in Fig. 2.2-b2, using all these attributes, we can generate a composite variable that has a better correlation coefficient, 0.300, than any single attribute. This coefficient is in a moderate correlation group based on Dancey and Reidy’s categorization [86]. We follow Dancey and Reidy’s categorization in the rest of the chapter when describing correlation strengths.

Since all attributes are normalized, each attribute’s weight in the composite variable shows a contribution level to the composite variable (`open`: -0.4, `consci`: +0.7, `netaddict`: -0.3, `min_happy`: +0.3, `extra`: +0.3). `consci` contributes most, as expected from its higher Spearman’s correlation (0.237) than others. In contrast, `extra` is assigned a relatively large weight (+0.3) in spite of its extremely small correlation (0.026) as a single attribute. The importance of `extra` can be confirmed by excluding `extra` from the composite variable construction. When `extra` is extruded, the composite variable’s correlation coefficient decreases from 0.300 to 0.289. From further interactive investigations on how the inclusion of `extra` improves each other attribute’s correlation, we observe that `open` has a much larger improvement (from 0.147 to 0.170) than others (e.g., `consci` has no improvement). We can infer that, by including `open` and `extra` with different signs (-0.4 and 0.3), the composite variable captures the subtle difference between these two per-

sonalities. Then, `scorelevel` shows higher dependence on this derived difference than either `open` or `extra`. This type of insight cannot be derived if only investigating a single attribute relationship to the 1D representation. Also, this example demonstrates the effectiveness of the SHAP-based recommendation in contrast to relying only on each attribute’s correlation coefficient.

We should emphasize that while we want to make a composite variable correlated to the 1D representation to some extent (e.g., Spearman’s ≥ 0.2 , weakly correlated), we do not have to see extremely correlated results (e.g., Spearman’s ≥ 0.7 , strongly correlated). If a strong correlation can be found with a linear combination of a few attributes, the use of NNs becomes unnecessary. In such a case, linear learning methods such as LDA should be sufficient for the prediction. Our focus is taking a further step from the single-attribute-based interpretation using the SHAP or other methods [25, 47, 48]. Our approach considers the influence of multiple attributes. Such an influence can be analyzed via the signed weights in a composite variable and the interactive adjustment of the composite variable. The correlations to each composite variable should be considered as an indicator of how much of the 1D representation is explained with the composite variable.

2.3.7 Interactive Visual Interface

We have introduced each visual component with its corresponding workflow step in Sec. 2.3.5 and Sec. 2.3.6. In summary, the selected output (Step 1) and performance of NRL (Step 2) are shown as the auxiliary information in Fig. 2.2-f; and the simplified representation (Step 3) is visualized in Fig. 2.2-b1. Then, to aid in understanding the representation, Fig. 2.2-a depicts the attribute contribution information (Step 4) and serves as an interface to construct composite variables (Step 5). The created composite variables are visualized in Fig. 2.2-b2. As we explain in this section, the remaining views in Fig. 2.2-c,d depict structural and semantic information to supplement the interpretation of the results. All visualizations in the UI are fully linked and share the same or similar color encodings (e.g., red represents `Class 0`). Lasso selection can be performed in Fig. 2.2-a, b, c. The selected instances from `Classes 0` and `1` will be highlighted in yellow (e.g., Fig. 2.8-c). With this linking, analysts can investigate specific patterns, such

as outliers and subgroups.

Structural information. Fig. 2.2-c shows network layouts as the structural information. Only in this view, we visualize all instances even including those not belonging to the two ends (i.e., **Classes 0** and **1**). This is because the precomputed network measures (e.g., degree) used for NRL are computed using links among all instances. And, we need to review the entire network to locate patterns related to the structural information. We apply scalable force-directed placement (SFDP) [87] and then use red, blue, and gray colors to represent **Class 0**, **Class 1**, and other instances, respectively. For cases when analysts want to review relationships in a specific class, the UI also supports filtering of instances. The filtering can be applied by using colored checkboxes located at the top right.

Attribute information. We visualize the distribution of each attribute as a double-bar histogram. By default, as shown in Fig. 2.2-d, the UI shows the distribution for each class. When the lasso selection is performed, we show the distributions of selected and non-selected instances (e.g., Fig. 2.7-b1 and b2). Because we have limited screen space, we order histograms based on the two groups' distribution differences measured by the Kolmogorov-Smirnov (KS) statistic.

Implementation. The UI is developed as a web application. For the back end, Python is used to perform Step 1 with DeepGL, Step 2 with MLPs, Step 3 with regularized LDA, Step 4 with the SHAP method, Step 5 with multivariate regression and the COBYLA, and Step 6 with the two-class density scatterplot, SFDP, and KS statistic. The implementation of these algorithms utilizes various libraries, such as deepgl [18], PyTorch [88] (for MLPs), ulca [68] (for regularized LDA), SHAP [82], Scikit-learn [89] (for multivariate regression), NumPy/SciPy [90] (for the COBYLA, KS statistic, and Gaussian kernel density estimation), graph-tool [91] (for SFDP), seaborn [92] (for the swarm-like visualization), and ColorAide [93] (for color generation). The front-end UI is implemented with HTML5, JavaScript, and D3 [94]. WebSocket is used for communication between the front- and back-end modules.

2.4 Case Studies

We demonstrate the effectiveness of our workflow and interactive visualizations with three case studies using the two datasets. For the first two case studies on Dataset I, we show analyses on network nodes. In the third case, we analyze the egocentric networks of Dataset II.

2.4.1 Study 1: Associations with Score Levels

We present a complete version of the analysis we have performed in Sec. 2.3.5 and Sec. 2.3.6. Our analysis target is the identification of attributes that are highly related to college students' `scorelevel1`. We classify the bottom and top `scorelevel` groups with the five-layer MLP using 128, 64, and 64 NN nodes for hidden layers. 138 and 648 students are categorized into the bottom group (**Class 0**, red color) and top group (**Class 1**, blue color), respectively. The prediction results show 1.0 accuracy for Step 2 but 0.98 for Step 3 (refer to Sec. 2.3.5.2 for the reason why we accept such high accuracies).

After completing Steps 1–3, we first confirm that the extracted 1D representation provides a reasonable separation between the two classes, as shown in Fig. 2.2-b1,. Also, the distribution of instances/students shows a sufficient variety in their coordinates. Thus, we expect that the 1D representation is not extremely overfitted for this classification.

We then review attributes' associations with the 1D representation from the view in Fig. 2.2-a. Based on the ranked order of the attributes, contributing attributes to the score level differences are highly related to many of the Big Five personality traits: `open` (openness), `consci` (conscientiousness), `extra` (extraversion), `neuro` (neuroticism), and `agree` (agreeableness). Other highly ranked attributes include `netaddict` (internet addiction level) and `min_happy` (the minimum of the friends' happiness levels). On the other hand, most network centralities are not ranked high. As examined in Sec. 2.3.6.1, the top 5 attributes have clear positive or negative trends with the SHAP values (e.g., higher `open` value tends to have a more negative SHAP value).

Next, we construct a composite variable with the top 5 attributes while using Spearman's as a measure of dependence. Fig. 2.2-b2 visualizes the result. The composite variable is expressed as: $y = -0.4 \text{open} + 0.7 \text{consci} - 0.3 \text{netaddict} + 0.3 \text{min_happy}$

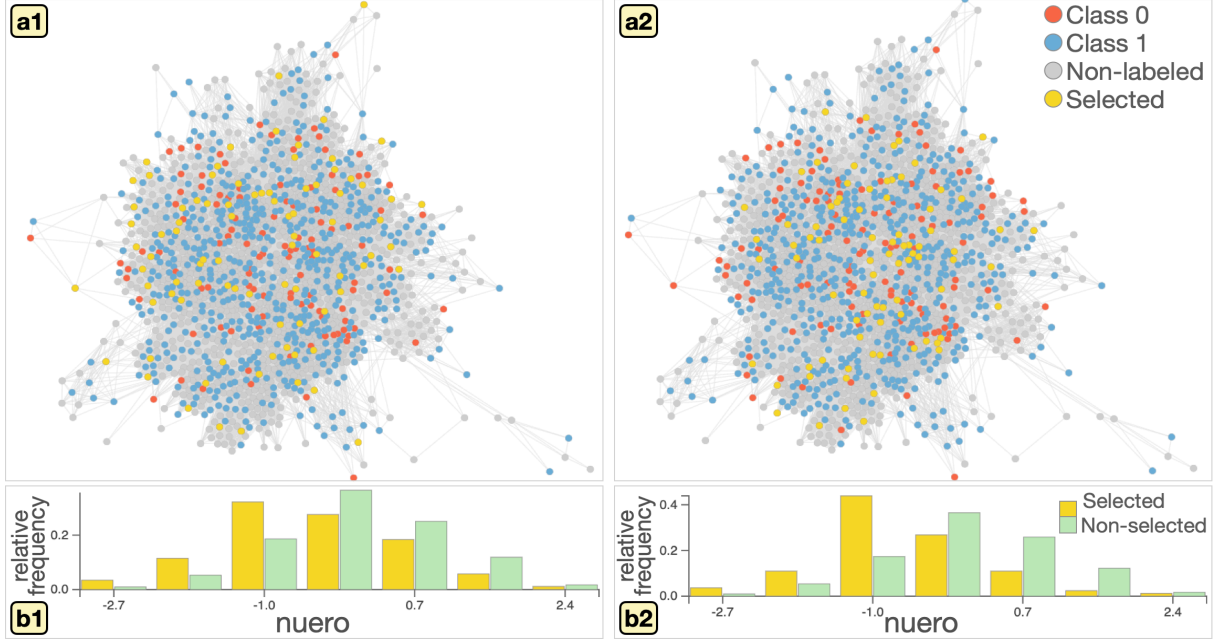


Figure 2.7: Study 1: Networks after selecting students with (a1) low ($-0.9 \text{ open} + 0.4 \text{ extra}$) and (a2) high extra; distributions of *nuero* (b1, b2) corresponding to the selections of (a1, a2).

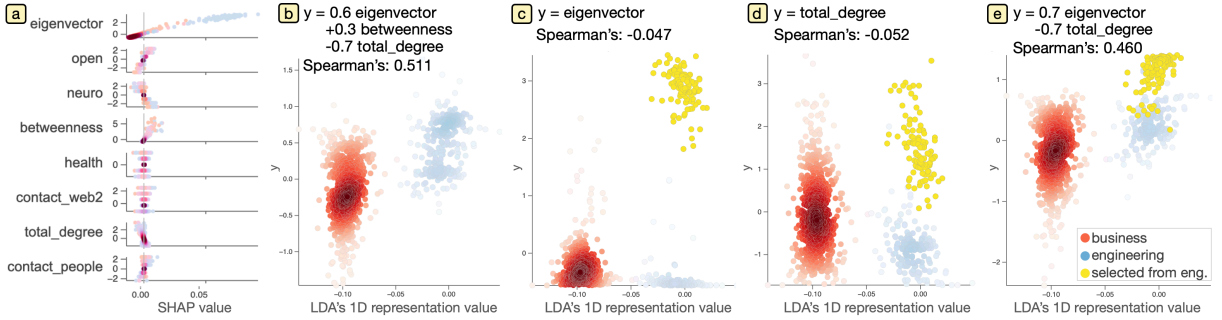


Figure 2.8: Study 2-1: The comparison of business and engineering students: (a) the SHAP values and (b-e) composite variables constructed from network centralities. Combining *eigenvector* and *total_degree* significantly increases the correlation.

+0.3 *extra*. Each weight's sign agrees with the corresponding attribute's positive or negative influence on the 1D representation, which is observed in Fig. 2.2-a. Based on the magnitudes, we can see *consci* contributes most to the 1D representation. Also, as discussed in Sec. 2.3.6.2, we observe that *open* and *extra*—personalities that might have some overlapped aspect—likely derive a new meaningful attribute by having opposite signs with each other. When a composite variable is constructed with these two attributes, the resultant variable, $-0.9 \text{ open} + 0.4 \text{ extra}$, shows a 0.17 correlation coefficient. Thus, having both openness and introversion tends to have clear negative impacts on their score levels.

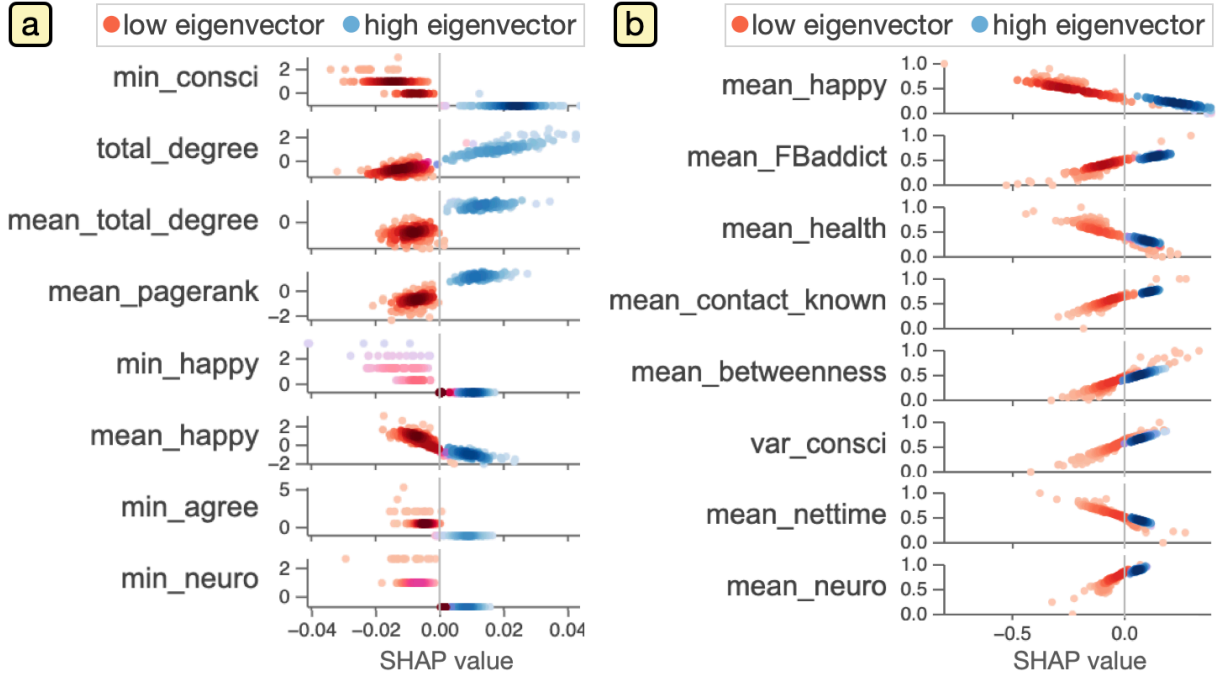


Figure 2.9: Study 2-2: The comparison of the high and low eigenvector groups of engineering students: the SHAP value visualizations (a) before and (b) after improving the feature extraction setting.

We want to investigate another question raised during the above analysis: why network centralities are not listed as highly contributed attributes even though the composite variable, $-0.9\text{open} + 0.4\text{extra}$, shows a clear contribution to the 1D representation. Openness and extraversion could be related to network centralities such as node degree (e.g., extrovert students are likely to have more friends). To examine the relationships to network structure, we highlight students with low ($-0.9\text{open} + 0.4\text{extra}$) in the network layout visualization, as shown in Fig. 2.7-a1. While we see several selected students (colored yellow) are located on the outskirts of the network (i.e., fewer connections to others), we cannot find any clear structural pattern. When selecting students with high **extra** (Fig. 2.7-a2), we cannot see a clear pattern either. Thus, we can expect that openness and extraversion in real life may be difficult to capture only from the connections on Facebook.

We further see the UI suggests **neuro** shows strongly different distributions between the selected and non-selected students. From the distributions of **neuro** shown in Fig. 2.7-b1 and b2, the students with high ($-0.9\text{open} + 0.4\text{extra}$) or high **extra** tend to have lower **neuro** than others. Therefore, **neuro** is also interrelated to **open** and **extra**. How-

ever, based on Fig. 2.2-a, **neuro** has a smaller influence on the score level and does not show consistent positive or negative influences among students. In fact, the inclusion of **neuro** into the composite variable of the top 5 attributes does not improve the correlation coefficient.

The above observations derive several reasonable insights: The student’s conscientiousness is highly related to their score level; internet addiction has a negative association with the score, especially, for those who had bad scores (refer to Sec. 2.3.6.1); if all friends have sufficient happiness, the score tends to be higher, and vice versa; and the openness and extraversion show a clear combinational effect and a high openness with a low extraversion has a more negative relationship to the score.

2.4.2 Study 2: Differences in Academic Units

From Dataset I, we review whether students from different majors have different structural and semantic characteristics. To perform this comparison, we select each pair of all possible different majors (**agricultural**, **business**, **engineering**, **humanities**, **physical sciences**, and **social sciences**). We review all pairs, but here we show one representative analysis: the comparison of **business** (**Class 0**, red, 801 students) and **engineering** (**Class 1**, blue, 287 students). We use the same 85 attributes and MLP as Study 1 and perform classification of **business** and **engineering**. The prediction results show 1.0 accuracy for both Steps 2–3.

Fig. 2.8-a shows the top 8 contributed attributes to the differences between **business** and **engineering**. Unlike Study 1, this list of the top 8 includes network centralities, **eigenvector**, **betweenness**, and **total_degree**. To see the associations between these centralities and the 1D representation, we generate a composite variable with these three centralities, as shown in Fig. 2.8-b. The resultant composite variable shows a moderate correlation (Spearman’s: 0.511). Also, we can see large weights with opposite signs for **eigenvector** and **total_degree** (+0.6 and −0.7). Unlike degree, eigenvector centrality considers the importance of links (discussed in Sec. 2.3.5.1). Taking a subtraction of degree from eigenvector centrality can emphasize this unique characteristic of eigenvector centrality.

As Fig. 2.8-a depicts a clear positive influence of high **eigenvector**, we visualize the relationships between **eigenvector** and the 1D representation. From the result shown in Fig. 2.8-c, we first notice that **eigenvector** itself has a very small correlation coefficient (-0.047). At the same time, there are two clear subgroups in a group of **engineering** as we highlight one group in yellow. On the other hand, as shown in Fig. 2.8-d, a similar but more moderate separation can be seen in **total_degree**. We then construct a new composite variable with only **eigenvector** and **total_degree**. The result shown in Fig. 2.8-e informs a clear correlation of this new composite variable to the 1D representation ($+0.460$). From the visualizations in Fig. 2.8-c, d, and e, we now grasp how this informative composite variable for the difference of **business** and **engineering** is generated from the two centralities. This result emphasizes the importance of interpretation using a combination of multiple attributes. It also highlights the benefits of the design of a two-class density scatterplot, which aids in identifying denser clusters and inferring trends and correlations.

We are further interested in understanding the difference between low and high **eigenvector** groups in **engineering** students. To perform this analysis, we select these two groups as a classification target. Fig. 2.9-a shows the derived top 8 contributing attributes, where red and blue colors are now used for low and high **eigenvector** groups. Most attributes show a clear separation between the groups. Also, many of them are related to the student's personality or mentalities, such as **consci**, **happy**, and **agree**. Meanwhile, the listed attributes in Fig. 2.9-a tend to correspond with the minimum values of their friends', as indicated with the prefix, "**min_**". As the high **eigenvector** group tends to have high **total_degree** (see Fig. 2.8-d), the students in the high **eigenvector** group are likely to have many friends, resulting in a higher chance to have at least one friend with low values for these attributes. Therefore, we consider that this result is caused by inappropriate fittings specific to these two groups. Also, as already observed in Fig. 2.8-d, the two groups have clearly different values in **total_degree**. As the inclusion of these attributes is problematic, we remove them during the feature extraction and rerun the remaining steps. Fig. 2.9-b shows the resultant top 8 contributing attributes. By review-

ing the values of each attribute (i.e., y -coordinates in Fig. 2.9-b), we infer that friends of the high **eigenvector** group tend to have lower happiness levels (**happy**), worse health status (**health**), and more addicted to Facebook (**FBaddict**), while they spend shorter time on the Internet (**nettime**). These results highlight the importance of considering compounded factors in network analysis (e.g., more social media friends but the friends are less happy; more addicted to social media but shorter time spent on the Internet).

From this study, we observe the difference between **business** and **engineering** students in the composite variable constructed with the network centralities. As the composite variable emphasizes the difference between the eigenvector and total degree centralities, **engineering** students' connections tend to be more particular to influential students in Facebook. Also, we demonstrate the case of identifying and resolving potential inappropriate overfitting from the visualized results.

2.4.3 Study 3: Factors Promoting New Connections

From Dataset II containing 345 egocentric networks of adults, we review contributing factors to promoting their indirect contacts (i.e., 2-hop neighbors) to direct contacts (i.e., 1-hop neighbors). For each network, we first compute the ratio of the number of indirect contacts transformed into direct contacts during the study period. We call this ratio the transformation rate and denote it as r . We then group respondents (i.e., ego nodes) with conditions of $r=0$ (i.e., no transformation) as **Class 0** and $r \geq 0.2$ (i.e., high transformation rate) as **Class 1**. **Classes 0** and **1** consist of 56 and 136 respondents, respectively. We then perform the classification of these classes, resulting in accuracies of 1.0 and 0.82 for Steps 2 and 3.

Fig. 2.10-a shows the top 10 contributed attributes to the 1D representation. We see several clear trends between attributes' values and the SHAP values. For example, **age*fbdays** shows an increasing trend (i.e., larger **age*fbdays** has a more positive impact on the transformation). **age*fbdays** is the predefined composite variable by the existing study [77]. This attribute is made by multiplying the ego's **age** and total days of Facebook use, **fbdays**. Due to this definition of **age*fbdays**, all top 3 contributed attributes, **age*fbdays**, **fbdays**, **age**, are likely to interrelate with each other. Then, we construct a

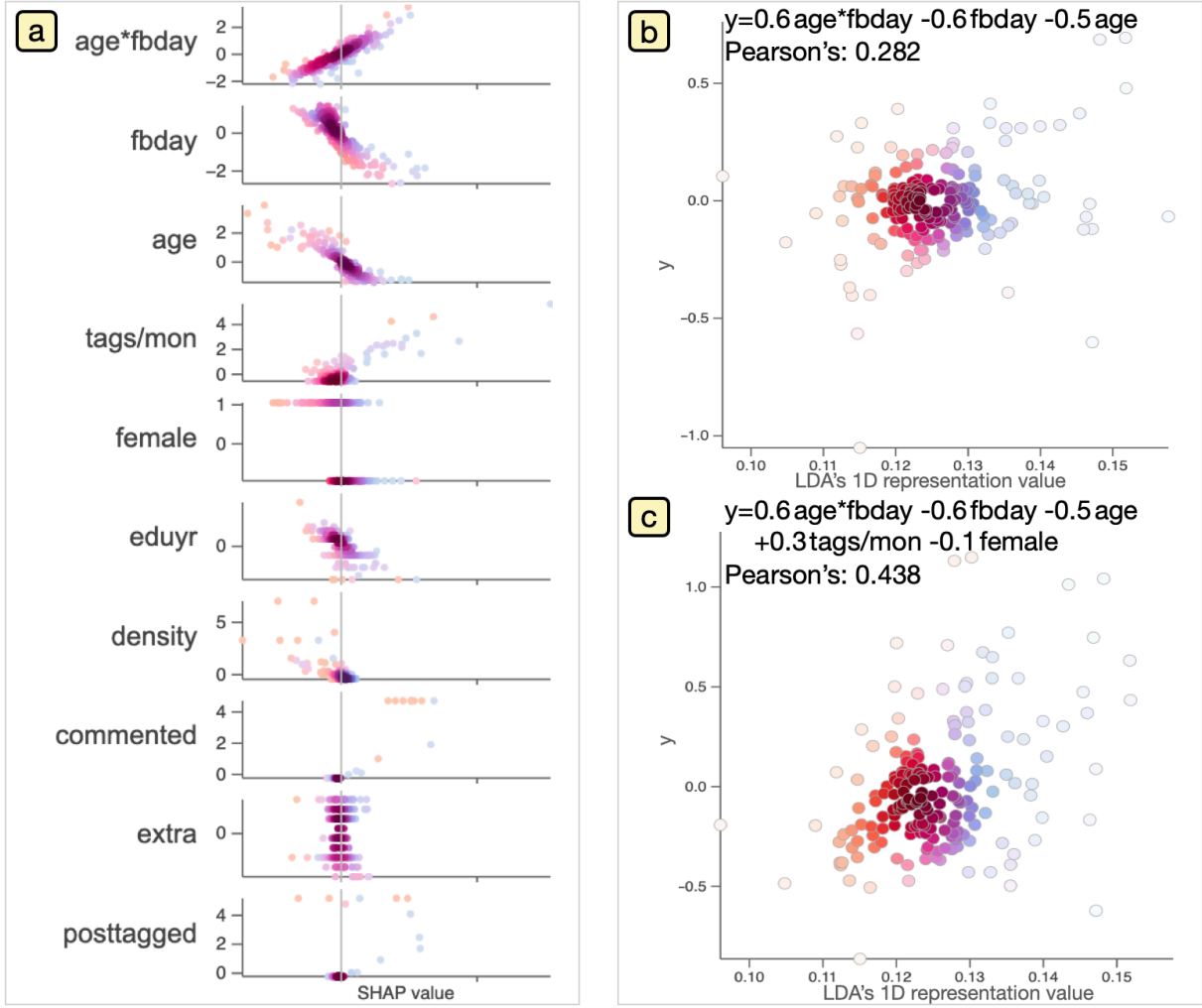


Figure 2.10: Study 3: (a) The top contributing attributes and composite variables constructed with (b) the top 3 and (c) top 5 attributes.

composite variable with these attributes. As shown in Fig. 2.10-b, the constructed composite variable, $0.6 \text{ age*fbday} - 0.6 \text{ fbday} - 0.5 \text{ age}$, presents a weak correlation (Pearson's: 0.282). By investigating three attributes individually, we observe that they show a much weaker correlation to the 1D representation—`age*fbday`: 0.064, `fbday`: 0.052, `age`: 0.123. Thus, the composite variable seems to find more meaningful information by disentangling complex relationships among the three attributes. We further construct a new composite variable by adding other highly ranked attributes, `tags/mon` (the average times of being tagged per month) and `female` (whether ego's gender is female or not). As shown in Fig. 2.10-c, the new composite variable has a significantly stronger correlation than the previous one (0.438 vs. 0.282).

For verification, we compare our findings with the existing study [77]. The existing study employs a mixed-effect model to capture attributes’ contributions to the transformation of indirect contacts. The top 5 contributed attributes suggested by their model are **female**, **extra**, **comments/mon** (the average number of comments made per month), **fbday**, and **age*fbday**. First of all, three attributes, **age*fbday**, **fbday**, and **female**, are seen in both their results and ours. Also, **tags/mon** and **comments/mon** should have semantic similarities with each other based on their meanings. In fact, by replacing **tags/mon** with **comments/mon**, we obtain the composite variable with 0.382 Pearson’s correlation coefficient, which is considerably close to the original (Pearson’s: 0.438). These observations indicate that our NN-based model successfully captured similar information to the mixed-effect model. The unique strength of our approach is in its ability to convey inter-relationships of attributes with their weights in the composite variable. While the existing work [77] crafts **age*fbday**, our study suggests $(0.6 \text{ age*fbday} - 0.6 \text{ fbday} - 0.5 \text{ age})$ or a more simplified version, $(\text{age*fbday} - \text{fbday} - \text{age})$, as a potential composite variable that better capture the influence on the transformation than **age*fbday**.

2.5 Expert Feedback

To further validate our workflow’s usability, we conducted an informal interview with experts in social network studies. The first expert (E1) is a distinguished researcher in an institute of sociology who collected Datasets I and II and also conducted research on these datasets with different focuses from ours. The second expert (E2) is an assistant professor in a department of sociology who formulated the guidelines for collecting the datasets. The other three experts are researchers in institutes of sociology (E3) and statistical science (E4, E5) who also studied the same datasets. The interview was conducted through a video conference setup, where our workflow methods, visualization designs, and the three case studies were presented using an interactive demo combined with static screenshots.

All the experts agreed that our workflow can support a wide range of their analysis targets as well as derive insights with more intuitive interpretations when compared to their current approach. For example, E1 commented, “The results [seen in the composite

variables] are much easier to understand and intuitive than the outputs from the statistical models used in our previous research.” All of them showed strong interest in insights that can be derived from the signed weights in composite variables. On the other hand, E1 noted a potential limitation of the datasets and analyses: “The distribution of the respondents might affect the results related to the internet addiction because the students who granted the use of their data tended to be more addicted to the internet usage than others.” E2 suggested better input attributes for the analyses, and we improved our case studies, accordingly.

There were several discussions on our workflow usage and design. E3 asked, “What if there are attributes with a dominant influence on the result? What if the 1D representation does not have clear separation?” For the former, the visualization shown in Fig. 2.2-a can be used to identify dominant attributes and exclude them from inputs if such attributes exist. Similarly, for the latter case, the visualization shown in Fig. 2.2-b1 is useful to review the quality of the separation. Then, based on the quality, analysts can take actions, such as adding more input attributes, reducing the ranges of the two ends of the output attribute, and improving the NN model. E4 raised a question on our post-hoc simplification of network representations: “Why not only use a neural network by designing the last layer with a linear activation function, instead of using LDA?” Our answer is given in Sec. 2.3.5.3. E5 asked, “Why not use other centralities but these three?” We selected degree, eigenvector, and betweenness centralities as a set of the most fundamental measures of structural characteristics. Our workflow is flexible to employ any other centralities based on analysts’ interests.

Finally, there were comments on our two-class density scatterplot design. E1 and E2 noted its analytical usability: “The design of the two-class density scatterplot facilitates the identification of dense areas, providing better support for observing data distribution and correlations.” E2 further added: “We often wasted a considerable amount of time in validating hypotheses that are formed from misleading visual patterns of correlations and class separations in conventional scatterplots.” E1 also commented: “The design of the two-class density scatterplot can be particularly useful for large-scale data and has

contributions to statistical science research.”

2.6 Discussion

Through the case studies and the expert interview, we discuss the efficacy of our workflow and interactive visualizations. In Appendix A, we further demonstrate the effectiveness of the constructed composite variables for conventional analyses, such as statistical hypothesis testing. Besides these evaluations, to allow readers to test and evaluate our workflow and UI, we provide related source code as well as processed data [26] from a publicly available dataset of faculty networks [95]. Here, we provide additional discussions on our designs.

Applicability to various data types. We have designed our workflow for multivariate networks. As multivariate networks have both high-dimensional and relational characteristics, by their nature, the workflow is even applicable to high-dimensional data and univariate network data. For example, we can analyze high-dimensional data by skipping the extraction of network centrality-related attributes and the visualization of networks. In addition, our workflow design can potentially adapt to advanced models of networks such as those containing meta-nodes and hyperedges. For example, as long as meta-nodes have the same set of node attributes as simple nodes in a network, we can still apply our workflow as is. Our workflow precomputes node- and link-related features before the training utilizing NNs; thus, we can deal with hyperedges during this preprocessing step (e.g., including hyperedges when applying relational functions in DeepGL).

Other potential algorithm designs. While we employ the NNs, DR, and composite variable construction to support our analysis target, there are other potential designs. One common way to understand the associations among a target and other attributes is applying a decision tree (DT) [96, 97]. When compared with a DT-based analysis, our design provides two main strengths in the interpretation step: simplicity and informativity. A DT provides a set of attribute ranges that is useful to classify a target attribute. However, the number of ranges would be easily overwhelming when analyzing networks with many attributes. Also, from the DT results, it is difficult to numerically assess the

influences from multiple attributes, such as those seen in the composite variables we have constructed for the case studies.

Another possible design is enabling the construction of more complicated composite variables, such as those with multiplications and logarithms of attributes. Although this would be more effective to analyze complex relationships among attributes (e.g., `age*fbday` in Sec. 2.4.3), this construction requires much more complicated optimizations than ours. One potential way to perform such advanced constructions while avoiding excessive computation is incorporating analysts’ knowledge more actively (e.g., predetermining a part of a composite variable). We plan to investigate this direction in future research.

Usability of two-class density scatterplots. We develop two-class density scatterplots to depict various patterns (e.g., distributions of class instances, trends, clusters, and outliers) in a single visualization. The use of this scatterplot is not limited to the targeted analyses in this work. As binary classifications or group comparisons are frequently performed for ML and visual analytics, we believe two-class density scatterplots can contribute to these fields. As future research, we plan to evaluate our scatterplot design by conducting a comprehensive user study and designing new quality measures for multiclass scatterplots that convey both aggregated- and instance-level information (e.g., splatterplots [70] and winglets [71]). This evaluation would also identify the two-class density scatterplot’s shortcomings for further improvements. There is one clear limitation: allowing visualization of only two classes. One potential way to support three or more classes is taking similar approaches to the multiclass geographical maps by Jo et al. [98]. For example, we can first partition a 2D space based on the changes in the distribution of class instances and then display the distribution in each partitioned region with a bar-chart glyph. We expect that this design can deal with several classes (e.g., five classes); however, this partition-based approach would not be suitable to reveal outliers and clusters. Thus, we would like to investigate an extension to three or more classes in the future.

2.7 Conclusion

We introduce a new visual analytics workflow designed to help find associations from complex multivariate networks. The workflow integrates neural-network-based representation learning, composite variable construction, and interactive visualizations. Benefiting from these components, the workflow generates expressive as well as interpretable analytical results. The design of workflow is also suited for analyzing other simpler types of data, such as a univariate network and high-dimensional data. Thus, our work potentially contributes to a wide range of applications that involve analyses of large, complex data.

Chapter 3

Collaboration Effectiveness and Efficiency Assessment through Dynamic Networks

In real-world scenarios, relational data frequently undergoes changes over time. When addressing research problems with temporal components, it becomes essential to analyze both the structure and dynamics of a network. This chapter focuses on examining the collaboration among healthcare professionals (HCPs), driven by the goal of enhancing the quality of cancer patient care. Surrounding each patient’s treatment, communication among HCPs can be examined as temporal networks, constructed from Electronic Health Record (EHR) access logs. This work introduces EHRFlow, a visual analytics system designed to study the effectiveness and efficiency of temporal communication networks mediated by the EHR system. I present a method that associates network measures with patient survival outcomes and devises effectiveness metrics based on these associations. To analyze communication efficiency, I extract the latencies and frequencies of EHR accesses. EHRFlow is designed to assist in inspecting and understanding the composed communication effectiveness metrics and to enable the exploration of communication efficiency by encoding latencies and frequencies in an information flow diagram. I demonstrate and evaluate EHRFlow through multiple case studies and an expert review.

3.1 Introduction

Healthcare professionals (HCPs) collaborate as a team in patient care, contributing at different treatment stages and locations. Electronic health records (EHRs) serve as a distributed communication system for asynchronous HCP communications, aiding in storing and retrieving patient information. HCPs collaboration is critical to the treatment quality and hence, can be associated with the patient’s survival outcome [99–102]. Such collaboration around a patient may be examined as a time-evolving bipartite network, in which the nodes depict both the communicating HCPs and the accessed EHR documents, while edges signify information flows. To identify and foster successful teamwork [103–105], assessing the network’s *effectiveness* involves understanding structural patterns that associate with better patient outcomes. Simultaneously, *efficiency* in information flow speed and frequency must be evaluated.

Existing efforts in healthcare-related communication and EHR associations have different motivations from ours. In health informatics literature, some focus on analyzing HCPs’ communication skills with patients [101, 106–109], but little on teamwork evaluation. Some machine learning (ML) methods have been developed for representing relational information in EHRs [110–112]. However, the generated representations are often challenging to interpret, and the information propagation within relations is insufficiently explored. More studies are needed to associate patients’ survival outcomes with HCPs’ teamwork structure and dynamics to enhance patient care quality.

To understand how collaboration structures can influence teamwork outcomes in the realm of network analysis, previous attempts have focused on constructing communication effectiveness metrics from basic network measures [113–115]. These metrics allow researchers to identify and intervene in underperforming teams. However, the metrics constructed in these approaches measure very limited topological patterns, and the criteria are not universally applicable across different domains. Further study into teamwork dynamics can evaluate the efficiency of communication, uncovering potential delays in the transmission of important messages. In efforts to extract these dynamics, network science researchers have developed temporal measures [116–119] to analyze communication

latency and frequency. Visualization designs [28, 120], conversely, aim to illustrate the evolution of network structure over time. However, comparing the latency of individual communications or interpreting the evolution of structure from a single static visualization can challenge human cognitive capabilities. To reduce the perceptual burden, a new design approach for presenting communication efficiency in a perceptible manner is necessary.

In this work, we present a set of network analysis methods as contributions to address the outlined challenges.

- For assessing communication *effectiveness*, we develop a flexible process to extract metrics that evaluate the success based on teamwork structures. This process is adaptable to a wide range of collaboration assessment criteria by incorporating a broader spectrum of network measures. Additionally, to expand the evaluation of a network by including a more comprehensive set of measures, we formulate a method that aggregates node-level measures into network-level expressions, focusing on the distribution imbalance.
- To inspect communication *efficiency*, we employ multiple views that assist users in configuring the collaboration dynamics of interest, thereby simplifying the complexity of information to be perceived at once.
- To support the interpretation of communication *effectiveness* and *efficiency*, we design a visual analytics system that succinctly guides users in understanding effective teamwork structures and exploring the dynamics of communications.

3.2 Related Works

In this section, we will discuss previous efforts in analyzing communications in healthcare and the general assessment of a collaboration’s effectiveness and efficiency.

3.2.1 Health Informatics Analysis

To analyze communications in healthcare, a variety of health informatics analyses have been developed. Some efforts utilize machine learning to encode complex relationships among electronic health records (EHRs) into representations [110–112], which can then be

applied to downstream tasks such as disease diagnosis prediction. While these representations may lead to accurate predictions, the underlying factors driving these predictions are often difficult to interpret, thus offering limited assistance in making meaningful clinical interventions. Another set of approaches has connected treatment outcomes or patient satisfaction to healthcare providers’ (HCPs) communication protocols or skills through the use of natural language processing (NLP)[101, 106–109]. However, the evaluation of HCPs’ teamwork remains a subject of debate, characterized by varying perceptions of interprofessional collaboration among different HCPs [102].

3.2.2 Network Effectiveness Analysis

The investigation of collaboration structures that lead to desired outcomes typically entails comparing the topological characteristics of networks.

Topological characteristics are often derived from basic network measures, which offer structural summaries across various aspects. In the context of teamwork, for example, *betweenness* measures the extent to which an individual facilitates (i.e., bridges) communication within a collaboration. A comprehensive list of network measures and their definitions can be found in Newman’s work [1]. However, previous studies assessing cooperation effectiveness in multiple fields have focused only on a limited set of measures. For instance, in evaluating software developers’ teamwork, Kidane and Gloor [113] consider betweenness centrality and density, while Wolf et al.[114] additionally incorporate degree centrality. In the healthcare literature, Creswick and Westbrook [115] examine medication advice-seeking interactions using density and shortest path distance, while Zhu et al. [121] empirically analyze the distributions of closeness and betweenness centrality in healthcare providers’ collaboration networks. It is vital to resolve the limited exploration of structural features and the evaluation criteria that are not applicable across different domains.

To identify structural features correlated with better outcomes, some approaches involve associating the distribution of selected measures with the network’s outcomes and comparing them among the analyzed networks [114, 115]. Another approach involves comparing the network measures of analyzed networks to a baseline distribution, constructed

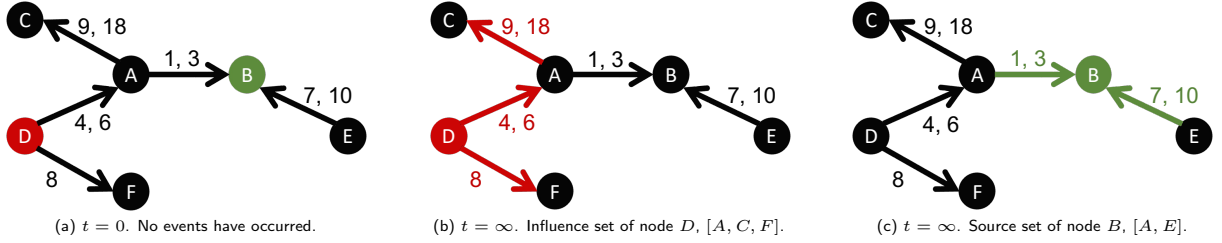


Figure 3.1: Illustrations of time-respecting paths: (a) presents an example of a directed temporal graph. The colored nodes are the analysis targets. (b) follows the red highlighted edges, retrieving nodes that can be influenced by node D . (c) follows the green highlighted edges, retrieving nodes that can influence node B .

using networks simulated by a random graph model [122, 123]. Yet, most simulated random graphs resemble real-world network topology within the locally tree-like structures [124], which may not be suitable for comparing global (network-level) statistics such as density. Furthermore, random graphs are typically simulated to establish specific local nodal properties, such as a scale-free degree distribution [125, 126], which may not adequately represent the complexity of dynamic collaboration networks.

Rather than utilizing network measures to extract structural features, emerging machine learning approaches aim to learn a representation for each network. For example, Fujiwara et al. [18] utilize contrastive learning [41] to obtain network representations that reflect their distinct topological features. However, this approach does not explore the combinatorial effects of diverse structural features. In response to this limitation, Lu et al. [127] present a workflow for extracting associations among nodal topological and semantic features of a multivariate network. They achieve this by constructing a composite variable that is most correlated with the chosen target variable, enabling the investigation of combinatorial influences. However, while this approach has the potential to develop effectiveness metrics applicable across different domains, it is designed for nodal comparison in a static multivariate network.

3.2.3 Network Efficiency Analysis

Network efficiency analysis facilitates the exploration and understanding of information flow dynamics within collaborations. Two primary research directions emerge in this domain, with one emphasizing the discovery of dynamic features and the other focusing on visually conveying these dynamics.

3.2.3.1 Temporal Measures

Unlike network measures that summarize the structural features of static networks, temporal measures are dynamics extracted from time-varying networks. Before computing any temporal measures, it is important to clarify the definition of a path in time-varying networks. In contrast to a static path, which is a series of edges constructed without restrictions, a *time-respecting path* [2] used in temporal measures requires that any two adjacent edges' timestamps occur in the same temporal order. Through these time-respecting paths, the *influence set* of a chosen node k can be traced as the set of nodes that can be reached from k within an observation time window. Conversely, the *source set* is the set of nodes that can reach node k . This concept is illustrated in Fig. 3.1. The subgraph constructed by the nodes in either the *influence set* or the *source set* of k is referred to as the *reachable subnetwork*, and the fraction of the number of nodes in this subgraph to the total number of nodes in the network is defined as the *reachability ratio* [119]. In addition to retrieving the reachable individuals in a network, *latency* and *frequency* further assess how fast and how often connections are made to those individuals. In the context of teamwork, *latency* measures the time it takes for a piece of information to be transmitted [3], computed by the difference in timestamps between the first and the last edge in a time-respecting path. Meanwhile, *frequency* calculates the number of times an interaction involving the same pair of entities occurs (e.g., how many times HCP A has accessed EHR document B within a day). However, there is little discussion on how to efficiently review and compare these abundant temporal measures.

3.2.3.2 Temporal Network Visualization

Temporal networks, or time-varying networks, encompass complex dynamics, including a vast array of temporal measures that require visualization. Designing representations for such data is far from trivial. Previous designs have focused on illustrating the evolution of network structure and can generally be categorized into several main types. We will discuss a simplified version of the taxonomy introduced in [28], specifically concentrating on visualizations presented as node-link diagrams. *Timeline* representations [120, 128] display snapshots of network structure along a timeline, either by stacking node-link

diagrams or placing them side by side. *1.5D timeline* approaches [7, 8] position a node’s timeline at the center, connecting its neighboring nodes to this timeline at the specific times events occur (i.e., the timestamps of the edges). *Super-graph* visualizations [129, 130] incorporate all existing time-respecting paths to offer a comprehensive overview of the network. While these approaches offer an in-depth understanding of network structure evolution, viewers often struggle to absorb such vast amounts of information presented in a single visualization. In light of this issue, *TeCFlow* [131] adopts a step-by-step approach to guide users in narrowing down the dynamics to be displayed at a given time. However, this approach relies on guidance tailored for a specific usage scenario and may not easily generalize to other temporal networks.

3.3 Data Processing and Analysis

As discussed in Sec. 3.2, the existing literature faces challenges in two main areas: **C1.** the extensibility of teamwork effectiveness evaluation criteria, and **C2.** the readability of presentations of temporal measures. Additionally, there is a lack of effort in **C3.** exploring the associations between teamwork effectiveness and efficiency. To tackle these limitations, we develop data processing and analysis methods, along with a visual analytics system, aimed at investigating the EHR-mediated communication effectiveness and efficiency of HCPs.

3.3.1 Dataset Overview

Our raw data consists of EHR digital traces from 518 patients diagnosed with Stage 2 or 3 breast, lung, and colorectal cancers, with approval from the IRB for data use. For each patient, the dataset includes their *basic information* and *access logs* of their EHR data. Basic information encompasses demographics (e.g., age and sex), treatments, and survival outcome (alive/dead). EHR access logs contain timestamped events spanning three months before to one year after the diagnosis date. To maintain a consistent collaboration timeframe, we exclude patients who passed away within a year after diagnosis.

A timestamped EHR access event involves a healthcare provider (HCP) accessing (reviewing or writing) a document, such as a note or a message, on the EHR system. Since

HCPs typically record a patient’s medical conditions in notes, while messages often lack context, we include only the access events involving notes in our analysis. Therefore, throughout this chapter, we use the terms “note” and “HCP/EHR document” interchangeably. To exclude irrelevant information from the evaluation of HCP teamwork, we remove administrative logs, such as billing events. Additionally, focusing on evaluating core teams’ communication, as advised by our medical doctor collaborators, we include only HCPs with titles such as MDs, RNs, ARNPs, PAs, Pharmacists, and Case Managers.

3.3.2 Network Construction

We model these access events by constructing a directed bipartite temporal network for each patient. This network comprises two sets of nodes: the *HCP nodes* and the *note nodes*, representing the HCPs and the EHR documents, respectively. A directed edge from an HCP node to a note node indicates that this core HCP wrote (created or edited) this EHR document at the timestamp(s) recorded in this edge, whereas an edge from a note node to an HCP node means that the HCP reviewed the note at the timestamp(s) recorded. Each edge may contain multiple timestamps since HCPs can review or edit a note multiple times. Notably, we do not aim to model the entire EHR system by constructing a network for each patient. Instead, the networks are specifically designed to focus on fabricating the information flow among HCPs facilitated by the EHR system.

3.3.3 Network Analysis

Our analysis workflow is built upon the collaboration networks of HCPs that we have constructed. It comprises three distinct steps: network effectiveness analysis, network visualization configuration, and network efficiency analysis. In the subsequent sections, we will detail the methods and visual aids developed for each of these steps.

Step 1: Network effectiveness analysis. To initiate the effectiveness analysis, we start by selecting a comprehensive set of basic network measures. These measures are designed to evaluate the quality of teamwork from various perspectives using network-level measures. While there is a plethora of node-level measures offering diverse viewpoints for evaluating HCP teamwork, relying solely on common summary statistics like the

average or maximum often fails to provide a meaningful network-level overview of all nodes. Addressing this challenge, Freeman [132] introduces the graph centralization index, a metric that reflects the imbalance in the distribution of node-level measures across a network. Essentially, a higher graph centralization index indicates that a few nodes dominate the distribution of node centrality, significantly surpassing the centrality values of the remainder (e.g., a 5-node network with a degree distribution of [4, 1, 1, 1]). To enrich our selection of measures, we decide to aggregate node-level centralities by building upon Freeman’s foundational work [132]. We develop equations tailored to calculate the graph centralization index for directed bipartite networks, ensuring their applicability to our collaboration networks. The specifics of these calculations are detailed in Sec. B.1.

Table 3.1: The descriptions of network measures included in the analysis.

Size	The number of HCP/note/total nodes
Density	The fraction of edges present over all possible edges
Hierarchy	The fraction of edges not participating in cycles
Component	The number of disconnected subnetworks
Distance	The average length of all-pair shortest paths
Degree	The number of nodes connected to the target node
Betweenness	The ratio of shortest paths involves the target node
Closeness	The inverse of the length sum of shortest paths to/from the target node to all other nodes

This expanded set of measures, calculated for each patient’s network, is utilized as input for a multi-layer perceptron (MLP) to classify patient outcomes, thereby linking structural patterns within the networks to survival outcomes. In total, there are 12 measures of each graph as shown in Fig. 3.2(b), where Table 3.1 lists their descriptions. However, our goal is not merely to use a machine learning (ML) model to assess the quality of collaboration. Instead, we aim to identify the network measures most closely associated with patient survival, offering insights for potential future interventions. Aligning with this motivation, Lu et al. [127] introduces a workflow to uncover correlations

between nodal topological characteristics in multivariate networks. Adapting their approach, we develop a flexible selection method aimed at identifying the combination of network measures that is most associated with patients’ survival outcomes.

Specifically, we commence by training a four-layer MLP to predict the softmax probabilities of survival and death for each patient’s network. Upon completing the training, we use the MLP’s output probability of survival as a proxy, approximating it with an optimized linear combination of the input network measures. The weights assigned to these measures are determined through linear regression to find a global optimum that yields the highest Spearman’s correlation coefficient between the survival probability and the value of the linear combination. Following this optimization process, we identify a threshold value that effectively distinguishes patients with divergent outcomes based on their linear combination values. This threshold is chosen to maximize the value that results in the greatest multiple of the accuracy for each outcome. Ultimately, these linear combination values serve as scores for our communication effectiveness metric.

In summary, we leverage the nonlinearity of the MLP to uncover intricate associations between network structures and outcomes. We then elucidate these correlations using a semantically interpretable linear combination of network measures. By abstracting the effectiveness analysis into an algorithm that tailors linear combination weights across a comprehensive set of network measures, we facilitate its applicability to a variety of scenarios and diverse evaluation criteria. Hence, we address the first challenge, **C1**. In real practice, we have the flexibility to customize teamwork effectiveness metrics for different patient groups. These groups consist of patients with similar characteristics such as cancer type, cancer stage, age, gender, and insurance (i.e., payer). To emphasize the effects of teamwork patterns, we follow the suggestion of our medical doctor collaborators to minimize the influences from demographic differences among patients.

Step 2: Network visualization configuration. This step serves as a bridge from feature extraction to visual representation, guiding the configuration of structural patterns and dynamics for display in the visualizations. By selecting a network measure to highlight in the node-link diagram of a network, users can interpret the extracted effectiveness



Figure 3.2: The visual analytics interface supporting the interpretation of communication effectiveness and efficiency. (a) is the metric swarmplot that displays the effectiveness metric score of each network (triangle) along the x-axis, guiding users to select a patient’s network to analyze. The brown line is the metric threshold that best separates the data points of two outcomes. (b) presents the composition and summary statistic of the metric. (c) is a series of stacked boxplots, where each shows the measure distribution quantiles per outcome and the vertical line indicates the measure value of the network belonging to the selected patient, allowing comparison and select the measure to inspect. (d) shows a series of network measure evolutions over time, assisting users to select a timestamp for observation. The brushed area indicates the selected observation time window. (e) displays the selected configurations, and legends of (f) and (g). The button click on “reviewed” or “disseminated” triggers the retrieval of source set or influence set reachable subnetwork and the reachability ratios will be updated accordingly. (f) is the global information flow diagram that visualizes the overall network and highlights the reachable subnetwork. (g) is the local information flow diagram that draws the reachable subnetwork, encoding the edge color and the edge thickness with communication latency and frequency.

metric from Step 1. We emphasize the importance of each node (i.e., the importance of an HCP/note) by their values of the selected measure. As discussed in Step 1, assessing the quality of a network requires the use of network-level measures. Therefore, we aggregate node-level measures through graph centralization index. Similarly, to represent the importance of a node, we need to encode each node using a node-level value. Thus, for any network-level measure used in the effectiveness metric, we decompose that network-level value to every node. Our approach follows the intuition that if removing a node k causes a significant change in a network-level measure, node k contributes more to maintaining the current value than a node whose removal does not result in a significant difference. To achieve this, we perturb a network by removing a node, the edges connected to it, and the nodes that only connect to the removed node. Then, we calculate the absolute L1 distance between the original value of a network-level measure and the value after perturbation as the importance of node k regarding this network-level measure. We repeat this

decomposition process for each node in a network and for every network-level measure.

By configuring the dynamics for display, we are able to reduce visual clutter by encoding only a subset of temporal measures, thereby improving the readability of presentations of temporal measures (i.e., addressing challenge **C2**). In our visual analytics system, we employ different visualizations to guide users in selecting a *patient*, a *timestamp*, and an *ego* (i.e., an HCP/note) to retrieve the dynamics bounded by these three parameters.

Step 3: Network efficiency analysis. To retrieve the dynamics up until the selected *timestamp*, we utilize the single-source depth-first search (DFS) algorithm to identify time-respecting paths involving the chosen *ego* node in the network of the selected *patient*. We then provide the *reachability ratio* of either the *source set* or the *influence set* formed by these retrieved time-respecting paths. In consideration of the timeliness of information dissemination and the perceived value of information, we extract the *latency* and *frequency* of each EHR document access. Unlike the *latency* introduced in Sec. 3.2.3.1, we define the first edge in a time-respecting path to have zero latency as the EHR access occurs immediately. Subsequent accesses (i.e., edges) have react latency, defined as the difference between the timestamp of the current edge and the timestamp of the previous edge. This computation is repeated for all retrieved time-respecting paths, and then we average the total latency of an edge based on the number of times it participates in any time-respecting paths. Our latency computation contrasts the reaction time of each EHR access, providing insights into HCPs’ relative speed in responding to updates in the EHR system. Regarding *frequency*, we compute it for each edge as the number of times such access occurred in any time-respecting paths. Our calculation is grounded in the occurrences within distinct time-respecting paths. Each count signifies a distinct propagation, ensuring that an edge with earlier timestamp(s) does not receive preferential treatment over another with a larger count.

3.4 Visual Analytics of EHRFlow

We design a visual analytics system, EHRFlow, to facilitate interactive analysis, incorporating all our network analysis methods. EHRFlow comprises three main components:

effectiveness metric assessment, network visualization configuration, and teamwork effectiveness and efficiency exploration. In this section, we will guide you through these components and illustrate how they support the network analysis steps introduced in Sec. 3.3.3

Component 1: Effectiveness metric assessment. This component is designed to investigate the communication effectiveness metric composed in the Step 1 of network analysis. In the *Metric Swarmplot* illustrated in Fig. 3.2(a), each data point (triangle) represents a patient’s treatment network, with the color encoding the survival outcome. The communication effectiveness metric score of each patient is presented along the x-axis, while the vertical positioning is arbitrary and serves only to prevent overlapping. The brown vertical metric threshold line, which separates the two outcomes, helps identify true positives (i.e., survived/alive), false positives, true negatives (i.e., not survived), and false negatives, aiding users in selecting a *patient* to examine his/her network’s characteristics. To assess the validity of the communication effectiveness metric, we provide the correlation and accuracy per outcome in Fig. 3.2(b), along with the full list of included network measures and their relative associations (i.e., weights in the metric), ranked in the order of their weight magnitudes. Derived from our methods, the interpretations of these statistics are as follows: (1) A higher correlation value suggests that more distinctive topological characteristics exist between networks with contrasting outcomes. (2) A larger weight magnitude of a network measure represents a stronger association with the patients’ survival. (3) A positive/negative weight means a larger value of this network measure potentially improves/deteriorates the chance of survival.

Component 2: Network visualization configuration. In support of setting the three parameters in configuring the perceivable node-link diagram of the selected *patient*, this component guides users to choose a *measure*, a *timestamp*, and an *ego* as introduced in the Step 2 of network analysis.

In the *Network Measure Stacked Boxplots* as presented in Fig. 3.2(c), users can compare each measure’s network-level value within the selected *patient*’s network (represented by the vertical line) with the respective distributions across networks within the same

patient group, characterized by similar demographics. Each stacked boxplot encompasses boxplots displaying the distributions of network-level measures for both surviving (green) and deceased (pink) patients. These distributions include quantiles (25th, 50th, and 75th percentiles), along with the minimum and maximum values indicated by the ends of the horizontal line. The vertical line, color-coded to reflect the survival outcome of the selected *patient*, indicates the network-level measure value within his/her network. Positioned atop the boxplots, this design facilitates inter- and intra-outcome comparisons, guiding users to select a *measure* to highlight in the subsequent node-link diagram of the selected *patient*'s network.

Through the provision of *Network Measure Evolution Line Charts*, as depicted in Fig. 3.2(d), users can track the evolution of each measure's network-level value over the treatment progression of the selected *patient*. In each line chart, we compute and present the network-level measures on a weekly basis, utilizing the communication logs accumulated up to that point as the intermediate teamwork network for the selected *patient*, thus offering insights for users to select a *timestamp*. When a user clicks on a specific week (i.e., a bin in the line chart), all the preceding weeks up to and including the selected bin are highlighted, indicating the analyzed time window from the start to the selected week of the treatment. Weeks associated with critical treatment dates (e.g., the commencement of chemotherapy) are shaded in light green/pink, based on the survival outcome of the selected *patient*, while other weeks are colored in highly saturated green/pink.

The final configuration, *ego*, is chosen in the *Global Information Flow Diagram* as depicted in Fig. 3.2(f) before any dynamics are retrieved and presented. Initially, we draw the nodes in the network of the selected *patient*, positioning them according to the force-directed layout algorithm using every edge (i.e., recorded EHR access) in their treatment history. The opacity of each node is determined by the node-level value of the selected *measure*, which can either be the values of node-level measures or the decomposed values of network-level measures. Additionally, to enhance visual contrast and aid users in selecting an *ego* by assessing the relative importance among nodes, note nodes are encoded using a white-gray color scale with circle shapes, while HCP nodes are encoded

using a white-yellow color scale with cross shapes, based on the node-level value of the selected *measure*. Upon clicking, the color of the selected *ego* node changes to pink or green based on the survival outcome of the chosen *patient*.

Up to this stage in the visual analytics workflow, users have chosen a specific *patient* to examine their configured dynamics, constrained by the selected *timestamp* and *ego*. These selections are listed in Fig. 3.2(e).

Component 3: Teamwork effectiveness and efficiency exploration. Presenting the configured dynamics while highlighting the relative topological importance of nodes, this component aims to address the final challenge **C3**. In the *Global Information Flow Diagram* (Fig. 3.2(f)), node encodings emphasize the relative importance among HCPs and notes regarding the selected *measure*. To illustrate the dynamics, only edges within the retrieved reachable subnetwork, as introduced in Step 3 of network analysis, are displayed, enabling users to observe the information flow to and from the selected *ego*. Additionally, in the *Local Information Flow Diagram* depicted in Fig. 3.2(g), we encode the *frequency* and *latency* of EHR accesses. The local diagram redraws the reachable subnetwork using the force-directed algorithm to illustrate more explicit local structures. Edge thickness represents the information exchange frequency, while edge color represents the reciprocal of latency—the speed at which an HCP reacts to an existing EHR document. This is encoded using a maroon-gray-cyan color scale, with slow to fast speeds mapped accordingly.

3.5 Case Studies

3.5.1 Overview of Patient Groups and Analysis Goals

Patient group A consists of 41 patients ranging from age 65 to 75 with Stage 3 lung cancer, where 27 patients survived and 14 patients did not. The statistics of the teamwork effectiveness metric extracted for this group is listed in Fig. 3.2(b). We conduct Study 1 and Study 2 on this group, analyzing collaboration patterns separately to elucidate both disconnected communications and underutilized EHR documents.

Patient group B contains 45 female patients ranging from age 65 to 75 with Stage

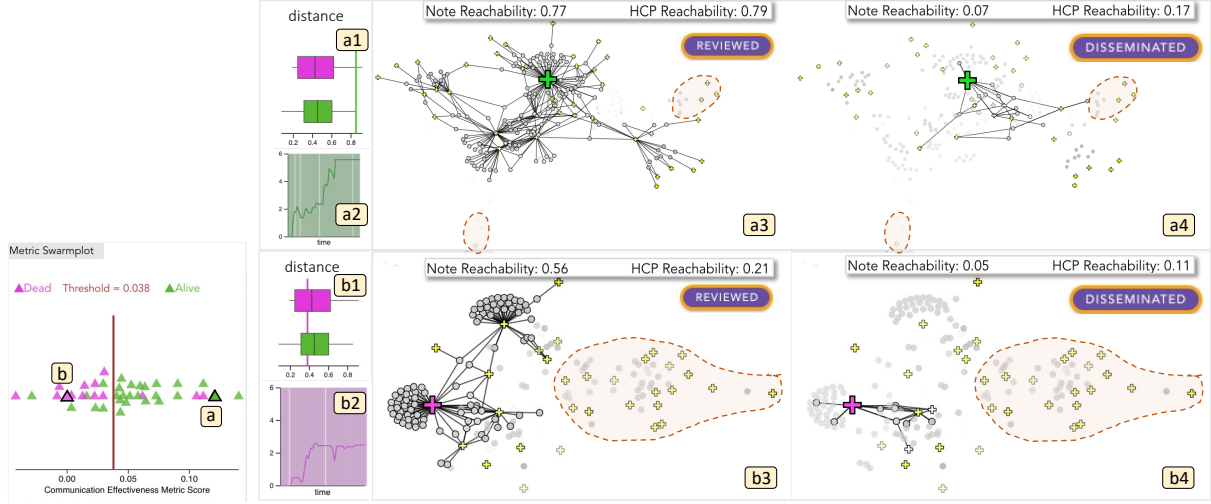


Figure 3.3: Fig. 3: A comparative study focuses on the disconnected communications within two representative patient networks indicated in (a) and (b). (a1) and (b1) present the stacked boxplots of distance measures for these patients. The network-level distance value for each patient is illustrated with a vertical line, positioned above the distance distributions for all patients grouped by different outcomes. (a2) and (b2) depict the measure evolution of the network-level distance for the networks of the two chosen patients. The brushed section marks the specified observation time window. (a3) and (a4) emphasize the reviewed and disseminated reachable subnetworks for the survived patient, whereas (b3) and (b4) are the reachable subnetworks for the deceased patient. The brown circles highlight the individuals disconnected from the ego HCP.

2 breast cancer, where 39 patients survived and 6 patients did not. The metric scores are strongly correlated with the patients' survival, having 0.9618 Spearman's correlation coefficient and the separation accuracy 0.8718 for the survived patients and 1.0 for the deceased patients using the metric threshold. We conduct Study 3 on this group, investigating the influences from the distributions of notes reading and writing duties among HCPs.

Table 3.2: The rank and metric weights assigned to six representative network measures, providing a metric composition overview for patient group B.

Rank#1	#4	#6	#8	#10	#11
In-Closeness	Betweenness	Out-Closeness	Density	HCP size	Hierarchy
-0.4986	-0.3094	0.2899	-0.2637	-0.119	0.1149

3.5.2 Study 1: Identification of Disconnected Communications

In this study on patient group A, we unveil potential information loss among HCPs, characterized by a lack of internal communication throughout the treatment. This revelation stems from the **distance** measure and the global information flow diagram.

As explained in Sec. 3.3.3, each node-level measure, such as **distance**, is aggregated to a network-level representation when composing the effectiveness metric. Specifically, we aggregate node-level **distance** to determine the average number of edges traversed by information before reaching an HCP. In our context, a higher network-level **distance** value indicates that more HCPs are involved in transmitting EHR documents, thereby broadening the accessibility of information. In Fig. 3.2(b), we observe that **distance** is weighted at 0.3868 and ranked as the second most associated network measure with patient survival. This suggests that the more HCPs involved in transmitting information within a patient’s treatment network, the more likely it is for the patient to have a better outcome. To investigate and understand the potential causes further, we select two representative patients (Fig. 3.3(a, b)) with opposite outcomes and high and low network-level **distance** values, as indicated in Fig. 3.3(a1, b1). We choose to examine their entire treatment duration in Fig. 3.3(a2, b2).

To start, we select an ego HCP in both networks. The two ego HCP nodes have node-level **distance** values close to their network averages (i.e., network-level **distance**). This indicates that comparing these two ego HCPs is representative of contrasting the two networks. In Fig. 3.3(a3), the reviewed reachable subnetwork of the survived patient (i.e., information reviewed by the ego HCP) reveals that the HCPs in the brown circles did not transmit any information that reached the ego HCP. Similarly, in Fig. 3.3(a2), the disseminated reachable subnetwork (i.e., individuals who have access to the information generated by the ego HCP) shows that the EHR documents created or edited by the ego HCP were never propagated to the HCPs in the brown circle. Upon observing Fig. 3.3(b3, b4), a considerably larger disconnected portion is evident within the brown circles in the deceased patient’s treatment collaboration. Quantitatively, comparing the reachability ratios in Fig. 3.3(a3, b3), we find that the ego HCP treating the deceased patient had three times lower reachability to other HCPs in the reviewed sets. Additionally, when contrasting the reachable ratios of the two disseminated sets in Fig. 3.3(a4, b4), it is observed that the ego HCP treating the deceased patient had half as much reachability to both HCPs and notes. Had more HCPs participated in transmitting the information, the

network-level **distance** would have been larger, and the information could have reached these disconnected regions. This potential increase in reachability might have led to an improvement in the treatment outcome.

To confirm that the growth in network-level **distance** indeed facilitated HCPs in reaching more others, we observe the survived patient’s reachable subnetworks at three timestamps, where the collaboration networks exhibited progressively increasing values of network-level **distance**. As evident in Fig. 3.4, both the reviewed (top row) and the disseminated (bottom row) reachable subnetworks kept expanding from the diagnosis date through the start date of immunotherapy to the last contact date. This indicates that the ego HCP treating the survived patient actively accessed more information and created more notes after each stage of treatment. Thus, the association between network-level **distance** and information reachability is validated. While more in-depth experiments are necessary, this case study suggests intervening in potential information loss by monitoring the development of the disconnected portion before a flawed treatment decision could be made.

3.5.3 Study 2: Trade-Off of Notes Underutilization

Following Study 1, our continued investigation of the same patients and ego HCPs uncovers the underutilization of EHR documents along with a trade-off between underutilized notes and disconnected communications. This is indicated by the **hierarchy** measure and the local information flow diagram.

To start, **hierarchy** is defined as the count of nodes that are not part of any cyclic paths. Within the scope of our study, a lower **hierarchy** value suggests that more HCPs diligently monitored the information they transmitted, thereby accessing follow-up notes created by other HCPs who had previously reviewed the EHR documents they transmitted. Consequently, this monitoring behavior engenders cyclic paths. We note that **hierarchy** is weighted at -0.1738 in Fig. 3.2(b) and ranks only 7th among all twelve measures. This implies that a greater number of HCPs tracking their previously accessed notes (thus reducing the value of **hierarchy**) correlates with improved patient survival chances. To comprehend the underlying cause, we delve into the local information flow

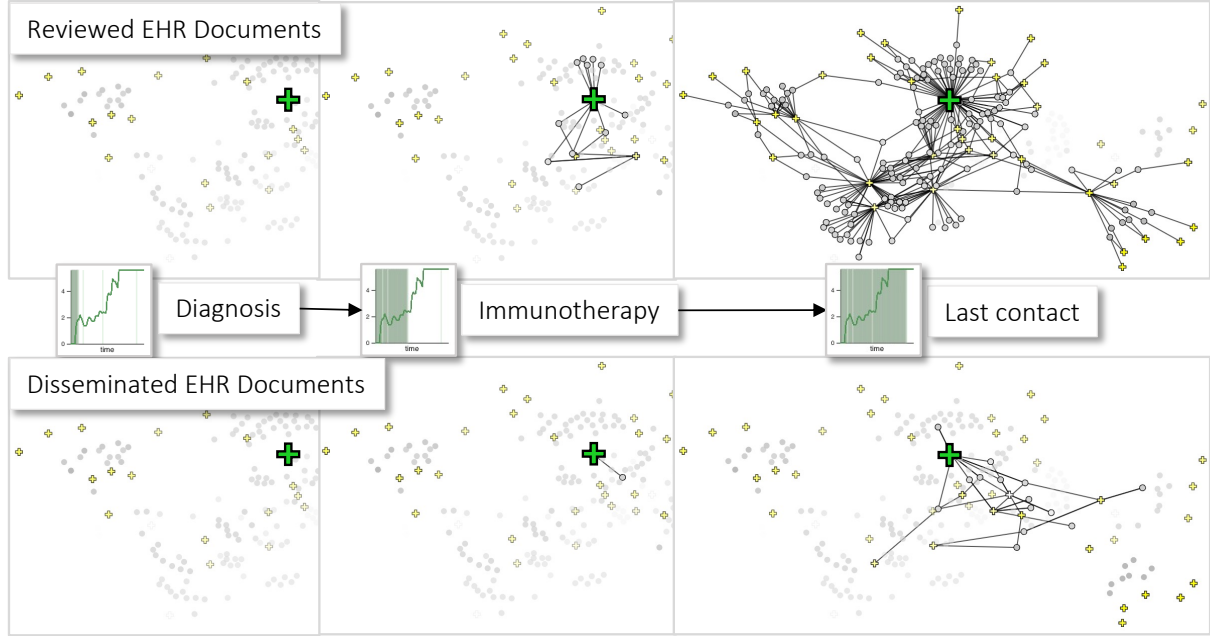


Figure 3.4: The development of reachable subnetworks over time for the survived patient's network chosen in Fig. 3.3(a). The top and bottom rows depict the expanding reviewed and disseminated sets after each critical treatment stage, indicating that the selected HCP actively reviewed and wrote new EHR documents promptly.

diagrams. In Fig. 3.5(b1, c1), the brown circles denote root nodes, serving as the source nodes for information propagation paths but were accessed by only one HCP within each reviewed reachable subnetwork. In Fig. 3.5(b2, c2), the brown circles highlight leaf nodes, representing destinations of paths, which were never accessed by any HCPs in the treatment network. These root and leaf nodes symbolize information transmission that was inadequately tracked by HCPs, resulting from non-cyclic patterns. Their abundance increases with a rise in the **hierarchy** value. Examining the reachable subnetworks of survived patients in Fig. 3.5(b1, b2), we observe that the majority of nodes participate in cyclic paths. In contrast, the reachable subnetworks of deceased patients contain significantly more underutilized root/leaf nodes, as indicated by the brown circles in Fig. 3.5(c1, c2).

The above observation suggests a preference for a lower **hierarchy**; however, a reduction in **hierarchy** is also linked to a decrease in **distance**, which is not ideal for successful teamwork, as it leads to disconnected communications, as explained in Study 1. As depicted in Fig. 3.5(a), the addition of two red-colored edges in the cyclic pattern shortened the **distance** from node X to node Y compared to the non-cyclic pattern.

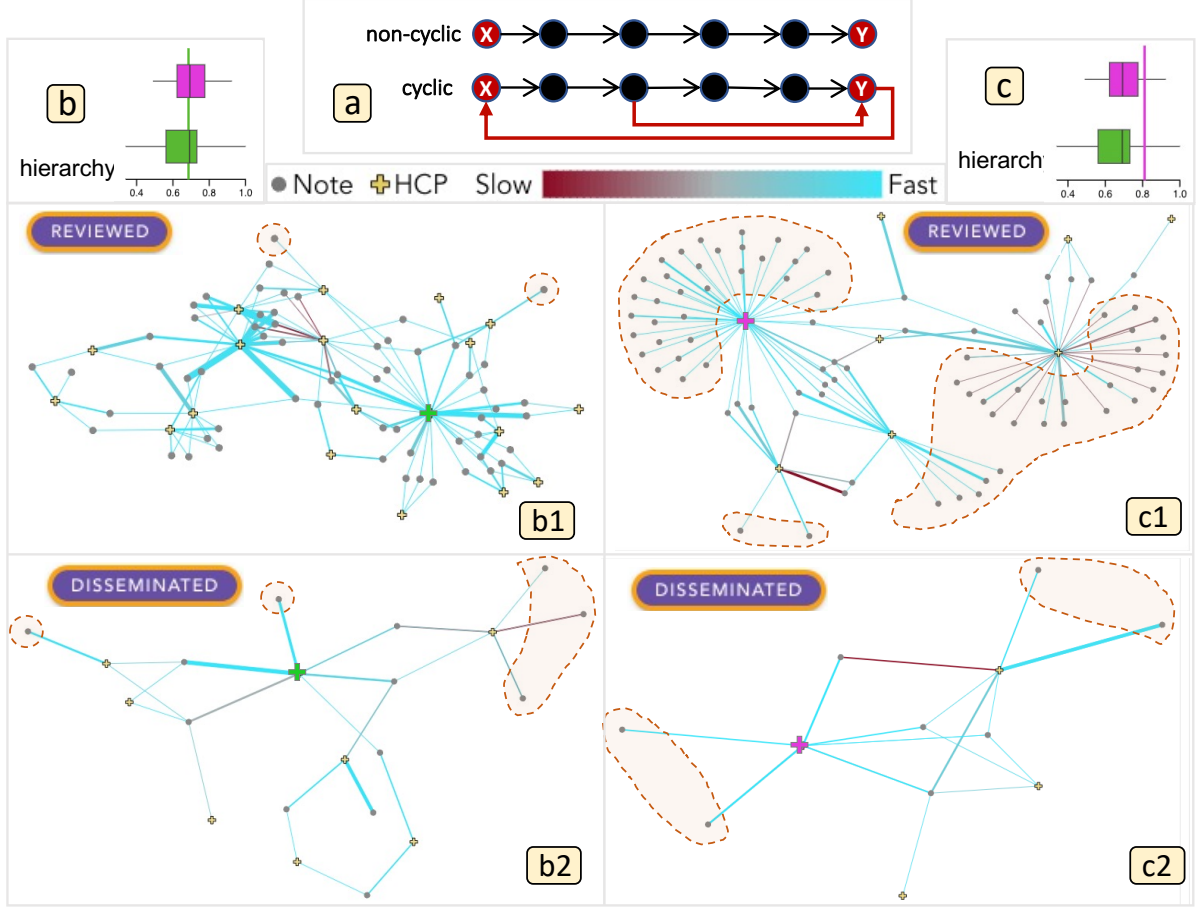


Figure 3.5: (a) compares a non-cyclic pattern with a cyclic pattern. (b) and (c) present the distributions of the *hierarchy* measure for the same two representative networks selected in Fig. 3.3(a) and Fig. 3.3(b). (b1, b2, c1, c2) represent the reviewed and disseminated sets in local information flow diagrams, enabling a comparative study to identify underutilized EHR documents circled in brown.

Consequently, we aim to interpret the potential trade-off between these two measures. As observed in Fig. 3.3(a1) and Fig. 3.5(b), the surviving patient exhibited a high **distance** and low **hierarchy**, whereas the deceased patient displayed the opposite, as illustrated in Fig. 3.3(b1) and Fig. 3.5(c). These contrasting patterns affirm that the two sampled patients serve as fitting representatives for exploring the trade-off. The metric assigning a negative weight to **hierarchy** aims to prioritize minimizing underutilized notes induced by non-cyclic patterns over achieving a larger **distance** by promoting non-cyclic patterns. However, by placing greater emphasis on the magnitude of **distance**, the metric underscores the continued importance of maintaining connected communications. This approach leads to an optimal teamwork pattern by ensuring effective tracking of accessed notes while actively disseminating acquired information to others.

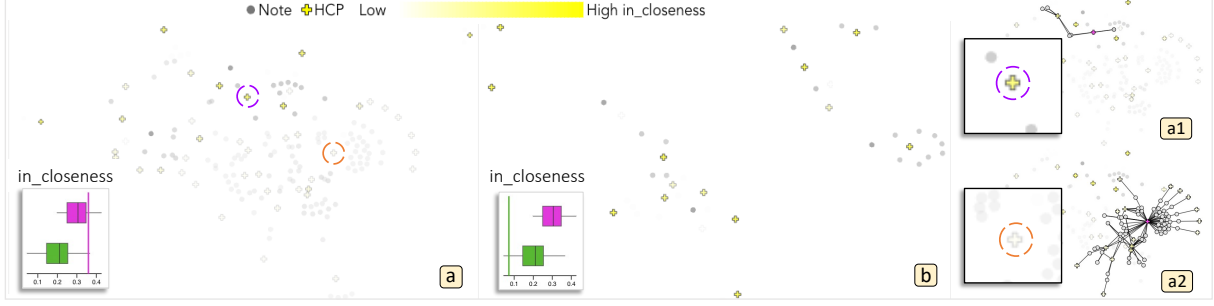


Figure 3.6: Fig. 6: The global information flow diagrams validate the influence from the distribution of HCPs' EHR reviewing accesses. By encoding node opacity and color with node-level *in-closeness*, (a) illustrates unevenly distributed note-reading duties over HCPs in a deceased patient's treatment network, while (b) shows a case of more evenly distributed collaboration in a survived patient's treatment. This encoding helps identify two HCPs with high and low node-level *in-closeness*, circled in purple and orange in (a), where their reviewed reachable subnetworks are shown in (a1) and (a2). (a1) ensures that the HCP with high node-level *in-closeness* had much less EHR reading responsibility compared to the HCP with low node-level *in-closeness* in (a2).

Aside from being able to discern HCPs with slower reaction speeds to EHR documents based on edge colors in the local information flow diagrams, users can pinpoint underutilized EHR documents by leveraging the explicit local structure. This suggests a potential opportunity for intervention to minimize wasted efforts in creating notes.

3.5.4 Study 3: Influences from EHR Reading and Writing Duty

In this investigation of patient group B, we examine the distinct influences from network measures that signify the varying levels of responsibility among HCPs in reading and writing EHR documents. This examination is informed by the global information flow diagram and the measures of *in-closeness* and *out-closeness*.

During the exploration of the effectiveness metric composition of patient group B in Table 3.2, we observe that the network-level *out-closeness* has the opposite weight sign compared to the network-level *in-closeness*. According to their node-level definitions, *in-closeness* and *out-closeness* represent the reciprocal of the sum of all-pairs shortest path distances flowing into and out of a node, respectively. In our context, they consequently reflect the inverse of the information influx and outflux for an HCP, indicating the HCP's EHR reading (influx) and writing (outflux) loads. When aggregated through our graph centralization index derivations, the positive weight of network-level *out-closeness* suggests that concerning the EHR writing duty of every HCP within a treatment network, the more imbalanced the distribution, the more likely the collaboration leads to a better survival outcome. Conversely, with the negative weight of *in-closeness*,

a more unevenly distributed EHR reviewing access over every HCP tends to result in a less favorable chance of survival.

We validate our interpretations by exploring the **in-closeness** measure, comparing two representative treatment networks with opposite outcomes. In Fig. 3.6(a), the network of the sampled deceased patient exhibits a large network-level **in-closeness** value (i.e., the vertical line on top of the stacked boxplot) among all patients (i.e., the stacked boxplot distributions), indicating an uneven distribution of reading load among HCPs. The uneven load is evident in the distinctive node encodings, with two groups of HCP nodes showing clear high (opaque yellow) and low (transparent white) node-level **in-closeness** values. In comparison, the network of the sampled patient who survived shows a low network-level **in-closeness**, demonstrating the expected balanced reading load distribution through the node encodings, where most HCP nodes have high node-level **in-closeness** values. To ensure the node encodings reflect the EHR reviewing load, we further select an ego HCP with higher node-level **in-closeness** (i.e., the inverse of the information influx) to observe their reviewed reachable subnetwork. We expect this ego to have fewer accesses for reading notes. As validated in Fig. 3.6(a1), the reviewed reachable subnetwork of the ego HCP with larger node-level **in-closeness** is much smaller than that of the other ego HCP with lower **in-closeness**, as shown in Fig. 3.6(a2).

In sum, our metric uncovers favorable collaboration patterns, centralizing the EHR writing responsibility on a few HCPs, but requiring every HCP to have similar diligence in reading notes. Through the interactive validation using our visual analytics system, users can not only investigate the suggestions given by the metric, but also identify HCPs being loaded with irregular reading responsibilities, providing clues for potential mediation in reassigning EHR duties.

3.6 Expert Feedback

To evaluate the usability of our analysis system, in addition to extensive case studies, we conducted interviews with experts in the healthcare domain. The first expert (E1) is a medical doctor who also serves on the faculty of internal medicine. Both the second expert

(E2) and the third expert (E3) are faculty members of Public Health. The interview took place via video conference, where we explained the analysis methodology and presented three case studies and corresponding insights in real-time using the EHRFlow interface.

Regarding data usage, E1 raised a concern: “Patient comorbidity is often considered to have a significant association with patient survival, but this information is not included in the analysis.” Although patient comorbidity can be related to teamwork structure and patient survival, we believe that the influences on survival outcomes from comorbidity and collaboration structure should be inspected sequentially or separately. We provide evidence justifying this claim using the following two scenarios:

- In the first scenario, the comorbidities of the analyzed patient group A have a Spearman’s correlation coefficient of 0.866 with patient survival, whereas the effectiveness metric scores have a correlation coefficient of 0.5397. Our effectiveness metric score extracts interpretable collaboration patterns adopted by HCPs to treat patients with worse comorbidities, allowing researchers to investigate potential structural interventions to improve treatment quality.
- In the second scenario, the comorbidities of the analyzed patient group B have a Spearman’s correlation coefficient of 0.6223 with patient survival, but the effectiveness metric scores have a correlation coefficient of 0.9618. In this case, researchers have more opportunity to investigate the excellent teamwork that helped patients with worse comorbidities survive and promote communication patterns in future treatments for this patient group.

These examples validate the separate correlations from comorbidity and communication structure.

Notably, there are many popular comorbidity indices such as the Charlson Comorbidity Index [133] and the Elixhauser Comorbidity Index [134]. To avoid method selection debate, we did not adopt these approaches but used linear regression to find the optimal correlation between comorbidity and survival outcomes. In the future, we plan to consult healthcare experts to incorporate patient comorbidity with domain considerations into our analysis workflow.

E2 considered our analysis design proper because our method always takes into account the period before the patient was diagnosed with cancer. However, E3 questioned, “As your method aims to analyze associations with patient survival, why don’t you utilize survival analysis?” While survival analysis can predict the time to an event (death or cure), our primary goal is to unveil teamwork patterns associated with patients’ survival. This involves contrasting around 40 historical patients’ EHR-mediated communication networks with known outcomes. Extracted trends offer insights into potential patterns for future patients. Domain experts should carefully inspect individual patients before considering interventions based on our suggested insights. Furthermore, we have filtered patient demographics for a specific group and ensured an identical timeframe for deceased and survived patients, eliminating bias in the extracted associations due to sickness levels.

E1 also noted, “As healthcare professionals, we need to fully understand the meaning of each network measure to utilize EHRFlow.” Admittedly, using EHRFlow requires a certain level of learning. To assist healthcare professional users in applying network measures, we plan to provide intuitive interpretations through the user interface. There were several appraisals of our visual analytics interface. E1 commented, “It is valuable that this visual analytics interface enables us to interactively drill down to identify high/low-value EHR documents and observe information dissemination (i.e., over a reachable subnetwork).” E1 added, “The drilling down investigation process exposes the potential interventions that could have been carried out for identified individuals.”

3.7 Discussion and Limitations

Through the case studies and the feedback from experts, we have demonstrated the capability and usability of EHRFlow. In addition to the limitations and justifications of our methods discussed in Sec. 3.6, we also consider additional alternative approaches here.

Alternative for extracting effectiveness. A popular alternative involves using a graph neural network (GNN) to classify network outcomes based on extracted network representations and to generate explanations for the GNN’s behavior. Although GNNs might surpass MLPs in identifying associations within network data, interpreting their

behavior presents a significant challenge. Current works on GNN explanation [5, 135], often highlight a subnetwork as the most critical structure contributing to the classification results. However, these identified substructures can be complex and difficult to rationalize, especially when ground truth is not available. Therefore, a more understandable form of GNN explanation is needed, and we plan to explore this direction in our future research.

Alternative for exploring efficiency. One alternative approach involves utilizing the global and local efficiency measures as derived in [116, 123], which calculate the average speed of information access across the entire network and the edges connecting to a node’s 1-hop neighbors, respectively. However, while the global efficiency measure offers insights into the network’s overall information dissemination speed, it falls short in pinpointing individual HCPs who exhibit lower efficiency. On the other hand, the local efficiency measure captures node-specific efficiency but includes undesirable dynamics by averaging over edges that extend beyond the immediately reachable subnetwork.

Additional node attributes. We also plan to enrich our analysis by incorporating more detailed information about HCPs, beyond their job titles, such as whether they are primary doctors or specialists. Having access to more specific details about the HCP nodes under investigation, or even the content of the examined EHR document nodes, could significantly enhance the depth of insights derived during the exploration process. To ensure privacy and confidentiality, this information will be provided after the data has undergone proper de-identification.

3.8 Conclusion

We introduce a novel approach to leveraging EHR data for potentially enhancing health-care outcomes. Our resulting visual analytics system, EHRFlow, enables the assessment of communication effectiveness and efficiency among healthcare professionals by analyzing and visualizing communication networks constructed from EHR access logs. Our case studies showcase the potential capability and value of EHRFlow, particularly with cancer patient datasets. We verified our analysis results and gathered valuable feedback for further research through reviews by healthcare experts. While our study has primarily

focused on healthcare teamwork using EHR data, the design of EHRFlow may find applications in the analysis of other teamwork scenarios, spanning from software development projects and scientific research collaborations to disaster management and beyond.

Chapter 4

Explanatory Structural Pattern Identification in Network Classification Tasks

In Chapter 3, I examine the collaboration networks among healthcare professionals and linked structural features to patient survival outcomes using network measures. Yet, beyond identifying contributing network measures (e.g., edge density) to favorable patient outcomes, pinpointing the most relevant collaborative patterns can provide a more intuitive understanding.

Graph Neural Networks (GNNs) have demonstrated high effectiveness in various machine learning (ML) tasks related to graphs, including node/graph classification and link prediction. This success underscores GNNs’ ability to extract structural features of complex networks. However, explaining the decisions made by GNNs poses challenges because of the aggregated relational information based on graph structure, leading to complex data transformations. Existing methods for explaining GNNs often face limitations in systematically exploring diverse substructures and evaluating results in the absence of ground truths. To address this gap, this chapter introduces GNNAnatomy, a model- and dataset-agnostic visual analytics system designed to facilitate the generation and evaluation of multi-level explanations for GNNs. In GNNAnatomy, I employ graphlets to elucidate GNN behavior in graph-level classification tasks. By analyzing the associations between

GNN classifications and graphlet frequencies, I formulate hypothesized factual and counterfactual explanations. To validate a hypothesized graphlet explanation, I introduce two metrics: (1) the correlation between its frequency and the classification confidence, and (2) the change in classification confidence after removing this substructure from the original graph. To demonstrate the effectiveness of GNNAnatomy, I conduct case studies on both real-world and synthetic graph datasets from various domains. Additionally, I qualitatively compare GNNAnatomy with a state-of-the-art GNNexplainer [5], demonstrating the utility and versatility of our design.

4.1 Introduction

A Graph Neural Network (GNN) is a machine learning (ML) approach used to transform graphs into vector representations, which are more easily utilized in classification or prediction tasks involving graph data. For instance, consider a teamwork scenario represented as a network, where the participants are nodes and the collaboration between two individuals forms an edge. Each network can be labeled based on historical data indicating whether its outcome was successful or not. By learning the associations between the teamwork structure and outcomes, a GNN can predict the likelihood of success for future teamwork endeavors. To derive actionable insights, the ability to explain the behavior of GNNs is essential; for example, in the teamwork scenario, understanding the underlying reasons for an unsuccessful outcome becomes crucial to implementing interventions to avert the undesirable performance.

To unravel the black box of GNNs, existing explainable GNN approaches, such as [5, 135], typically focus on identifying crucial subgraphs that have the greatest influence on GNN behavior. However, besides investigating subgraphs, many real-world applications necessitate identifying the basic building blocks of a graph (i.e., elementary substructures or recurring patterns) and understanding the function and significance of each substructure. For instance, in biological networks, it is a 6-atom carbon ring that renders a molecule mutagenic [136], rather than a specific subgraph containing a 6-atom carbon ring. Furthermore, previous literature offers limited explorations for alternative expla-

nations. Most methods generate a single subgraph or an ensemble of subgraphs as the explanation without seeking potential alternative explanations. Since such subgraph-level explanations do not always accurately explain the GNN behavior, it is imperative to enable the exploration of diverse substructures and also seek for substructure explanations to derive the most trustworthy explanations.

Another challenge lies in the granularity of explanations. Previous GNN explanation approaches generate either instance- or model-level explanations [137]. While instance-level explanations are tailored for individual graphs [5], they limit the overall understanding of GNN behavior. In response to this limitation, model-level explanations aim to interpret GNN behavior for the entire dataset or for graphs belonging to the same class [138–140]. However, it is often the case that more than one substructure contributes to making a graph distinguishable. For example, the classification of a molecule’s mutagenicity involves both a 3-atom acyclic substructure, NO_2 , and a 6-atom carbon ring [136]. The NO_2 substructure alone can mark some graphs as mutagenic, whereas others require only the carbon ring or a combination of both traits for differentiation. Merely suggesting common substructures in a class fails to encompass these detailed explanations, and thus neither instance-level nor model-level explanations offer sufficient granularities [141]. The ability to identify groups of graphs that require different explanations and determine the most explanatory substructure for each group is thus crucial.

Validating explanations is not a trivial task due to the absence of ground truths regarding which substructures the GNN should be utilizing. Most evaluation metrics are based on the intuition that when explanatory nodes and links are removed from the original graph, there will be a significant change in the GNN prediction [142, 143]. However, removal-based methods pose risks for graph data, as removing a subgraph can often results in the new graphs being out-of-distribution (OOD) compared to the original graphs [144–147]. Moreover, metric scores are typically used for comparative purposes among different GNN explainers instead of assessing potential explanations to determine the best one. A systematic evaluation method is required to examine the relevance of generated explanatory substructures.

Given these limitations and remaining challenges, we present novel methods aimed at addressing each of them within the context of graph-level classification tasks.

- To generate substructure explanations, we employ *graphlets* as explanatory elements. Each graph is represented as a vector composed of graphlet frequencies, facilitating a simpler neighboring data space directly derived from the original graphs. This approach offers a summary of basic substructures for each graph, and is not only to better facilitate explanation but also applicable to any graphs. Subsequently, we approximate the GNN performance using the graphlet frequency vectors.
- To provide granular GNN explanations, we identify groups of graphs that require different classification rules by illustrating the relationships between GNN classifications and graphlet frequencies. By specifying groups of graphs for which explanations are generated, we address the absence of group-level explanations.
- To ensure reliable evaluation and explore potential alternative explanations, we derive two metrics from factual and counterfactual reasoning processes to systematically assess the relevance of each graphlet as a potential explanatory substructure.
- With our methods, we introduce GNNAnatomy, a visual analytics system designed to support interactive generation and evaluation of substructure explanations at the needed granularities.

4.2 Background

This work employs graphlets to explain the behavior of graph neural networks in graph-level classification tasks. In this section, we offer concise background information on the two techniques.

4.2.1 Graphlets and Graphlet Frequencies

Graphlets are small connected non-isomorphic graphs [35, 148]. Given a specific number of nodes in a graphlet, there is a fixed number of graphlet types. For example, we show all 6 graphlet types with 4 nodes in Fig. 4.1. Graphlets have been used in both analysis and visualization of graphs and networks [149, 150].

The relative frequencies of different graphlet types can be used to characterize the

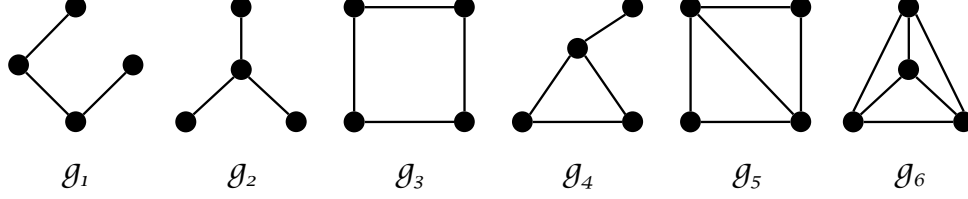


Figure 4.1: Six graphlet types with four nodes.

topology of a graph. For the 6 graphlet types $g_1, g_2, \dots, g_6$ in Fig. 4.1, a 6D vector of graphlet frequencies $\mathbf{f}_G$ is extracted to represent the graphlet distribution in each graph G :

$$\mathbf{f}_G = (f_{g_1}, f_{g_2}, \dots, f_{g_6}), \quad (4.1)$$

where

$$f_{g_k} = \frac{\#(g_k \subseteq G)}{\sum_{i=1}^6 \#(g_i \subseteq G)}. \quad (4.2)$$

In the actual computation of graphlet frequencies, it is too time-consuming to count the exact occurrence of each graphlet $\#(g_i \subseteq G)$ because comparing all possible connected subgraphs is an NP-hard problem. Therefore, we adopt the connected subgraph sampling method in [151], and compute the frequency of each type of graphlet among a set of subgraphs (with 3, 4, and 5 nodes) sampled from G . For any graph G , we concatenate the 2D vector of the 3-node graphlet frequencies, the 6D vector of the 4-node graphlet frequencies, and the 21D vector of the 5-node graphlet frequencies into a single 29D graphlet representation.

4.2.2 Graph Neural Networks

Graph Neural Networks (GNNs) are machine learning models designed to operate on graph-structured data, allowing them to extract relational characteristics inherent in graphs. As depicted in Fig. 4.2, in a GNN, each node undergoes a series of operations across multiple layers. Take node A for instance, in the first layer l_1 , the GNN aggregates messages from the direct neighbors of node A (i.e., nodes B , C , and D). It then combines this aggregated message with the node representation of A from the previous layer l_0 to update the node representation of A in layer l_1 . This process repeats for every node in each subsequent layer, l_N , where $N = 1, 2$ in this example. The learned node representations/embeddings serve as structural features for performing various downstream tasks

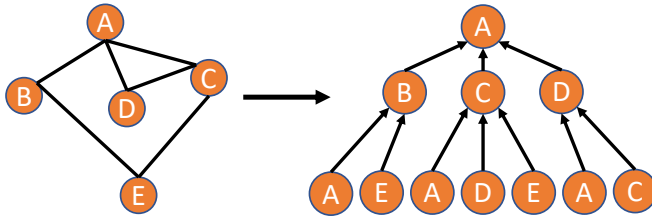


Figure 4.2: GNN aggregate information from layers of neighbors. The right side shows a 2-layer neighbor hierarchy for learning node A embedding.

in GNNs, including node classification, edge prediction, or graph classification. In node classification tasks, the node representations are utilized as input to a classifier, enabling the prediction of the respective node labels. In graph classification tasks, the embeddings of all nodes are aggregated into a single graph embedding, which serves as input to the downstream classifier for predicting the overall graph label.

4.3 Related Works

The field of explainable graph neural networks (GNNs) is still in its early stages of development, with relatively few works exploring this area. Nonetheless, inspired by the success in explaining computer vision models, researchers have approached explainable GNNs from various angles. In this sections, we provide an overview of previous approaches and remaining limitations and challenges. Subsequently, we discuss visual analytics works contributing to explaining GNNs. Finally, we review common metrics used for evaluating GNN explanations.

4.3.1 Subgraph Explanations

Existing GNN explanation works strive to elucidate GNN performance by pinpointing a subgraph that significantly influences GNN classifications or predictions. Perturbation-based methods, such as those introduced in [5, 152–154], utilize generators trained to produce masks (e.g., node or edge masks) applied as perturbations to input graphs. The objective is to perturb graphs in a manner that retains important structural information, thus preserving similar GNN predictions as the original graphs. However, these approaches often face potential biases due to the soft masks, which yield continuous values between 0 and 1, making the generated explanations challenging to interpret. Conversely, surrogate-based methods [135, 155–157] leverage interpretable surrogate models

to approximate GNN predictions using local neighboring data derived from input graphs. Despite their interpretability, defining neighboring data for graphs presents a challenge. The neighboring space should be simpler to comprehend and representative of the original discrete graphs, which contain structural information. Consequently, finding such neighboring data for graphs is non-trivial. Nevertheless, these methods typically yield a specific subgraph as an explanation, whereas in practice, there is often a need to identify the general substructure that influences GNN behavior. For instance, it is more useful to ascertain that it is the ring structure causing a molecule to be mutagenic, rather than merely identifying a subgraph containing a ring as the explanation. Furthermore, constrained by the design of generating subgraph explanations, previous literature has limited exploration into potential alternative explanations. There may exist subgraphs with similar topologies to the explanation that are not adequately justified as to why they are less important than the explanatory subgraph. In essence, it is vital to provide substructure explanations rather than subgraph explanations and allow for the exploration of diverse substructures before determining the most explanatory one.

4.3.2 Instance/Model-Level Explanations

A taxonomic survey conducted by Yuan et al. [137] categorizes GNN explanation methods as either model-level, which elucidates the global behavior of GNNs, or instance-level, which produces explanations specific to each graph instance. While the majority of existing GNN explanation works focus on interpreting why GNNs make predictions or classifications for individual graph instances, as introduced in Sec. 4.3.1, instance explanations alone are insufficient to provide a comprehensive understanding of GNN behavior. Given the multitude of graph instances in a training dataset, deriving a global understanding from individual instance explanations is not feasible. In contrast, several model-level methods have been proposed. To name a few, Xuanyuan et al. [140] treat neurons in GNNs as individual concept detectors capable of interpretation and generalization across multiple instances. PGExplainer [139] learns a parameterized model to rate the importance of each edge in the input space. XGNN [138] learns a graph generator that provides a set of graphs maximizing a target prediction, containing distinct vital graph patterns

relevant to the global behavior of GNNs. However, while most existing model-level methods aim to interpret GNN behavior across entire datasets or graphs from the same class, they often lack granularity. For example, in the classification of a molecule’s mutagenicity, while a 6-atom carbon ring alone may indicate mutagenicity in some graphs, others may require the presence of NO_2 as well. This renders neither instance-level nor model-level explanations sufficient. Therefore, to achieve flexible granularity in explanations, it is crucial to identify groups of graphs that require different explanations and determine the valid explanatory substructure for each.

4.3.3 Visual Analytics for Explaining GNNs

According to a recent survey [43], there has been little visual analytics effort in explainable GNNs. We discuss several methods dedicated to GNN model evaluation, diagnosis, and domain-specific interpretations. Liu et al. [158] proposes an interactive system that connects node embeddings with both the topological neighborhood and node features for model evaluation. Additionally, GNNLens [9] is developed to explain and diagnose prediction results using metrics and correlations between an instance and its neighbors. However, the topological correlations and metrics developed in these works consider only the 1-hop neighbors and focus on explaining the predictions of nodes in a graph. Wang et al. [141] designs a drug repurposing visual analytics system, which uses GraphMask [154] to extract critical meta-paths understandable by domain experts. Similar to our concerns, this work also highlights the need for domain experts to generate explanations for GNNs at needed granularities. Nonetheless, their visualization design choices largely depend on domain preferences, which hinders their applicability to other graph datasets and GNN architectures.

4.3.4 GNN Explanation Evaluation Metrics

Evaluations are necessary to ensure the validity of the generated explanations. Some metrics, such as accuracy (F1 or ROC-AUC scores) [5, 159], require ground truth motifs for computation and are often used to evaluate explanations for synthesized graph datasets. Fidelity [142, 143] refers to the concept that well-suited explanations, if re-

removed, will change GNN predictions. Sparsity [142] pertains to the idea that explanations should be compact and minimally affect the node size and edge density of the original graph when removed. Building upon these two measuring concepts, various metrics can be derived [160, 161]. Yet, when evaluating the fidelity of a generated explanatory substructure, removal-based methods often encounter issues with out-of-distribution (OOD) scenarios [144–147]. Despite efforts to mitigate the OOD problem with emerging generation-based approaches, they still suffer from biases and deviations from the original model behavior [162, 163]. Some other works evaluate their explanations based on the reasoning process, distinguishing between factual and counterfactual explanations. Factual explanations are deemed *sufficient* substructures that lead to the same prediction, while counterfactual explanations are regarded as *necessary* substructures that, once removed, alter the prediction. Commonly derived metrics include probabilities of sufficiency and necessity [164, 165]. Sufficiency measures the percentage of graphs whose factual explanation substructures alone maintain GNN predictions, while necessity measures the percentage of graphs whose counterfactual explanation substructures, if removed, change GNN predictions. Still, there is no standard format for evaluating GNN explanations and addressing additional issues such as OOD and model behavior deviations.

4.4 Methodology

As outlined in Sec. 4.3, existing explainable GNN methods encounter limitations in terms of (1) explorativity, (2) granularity, and (3) evaluability. This section describes how we address these challenges.

4.4.1 Graphlet Frequency: Topological Summary

Given the fundamental role of elucidating graph structures in GNNs, as opposed to explicating node/edge features, which is more aligned with methodologies for non-graph data, this work places emphasis on discerning the importance of substructures.

As outlined in Sec. 4.2.1, our method utilizes connected non-isomorphic small graphs, or graphlets, as the foundational elements for constructing GNN explanations. By calculating the frequency of each graphlet within a graph, we derive a topological summary rep-

resented as a vector. This summarization is achieved through the use of k -node graphlets, specifically for $k = 3, 4, 5$. To illustrate with $k = 5$ as an example, we extract 5-node connected subgraphs from the original graph employing the efficient sampling method detailed in [151]. For each of these sampled subgraphs, we determine their isomorphism to each of the 21 distinct 5-node graphlets. The frequency of each 5-node graphlet within the graph is quantified by the ratio of isomorphic instances to the total sampled subgraphs. This procedure is similarly applied for $k = 3$ and $k = 4$, enabling a comprehensive topological characterization across different graphlet sizes. Ultimately, we compile a 29-dimensional (29D) vector by concatenating the 2D, 6D, and 21D topological summaries derived from the frequencies of the 3-, 4-, and 5-node graphlets, respectively.

Leveraging graphlets as the foundational elements for explanations enhances explainable GNN methodologies in several ways: Firstly, the process of topological summarization is universally applicable to any graph data. Given that current GNN explanations frequently take the form of subgraphs, as reviewed in Sec. 4.3.1, establishing a consistent framework for constructing explanations is beneficial. This uniformity ensures a unique mapping between explanatory substructures and generated explanations, preventing evaluation difficulties arising from slight structural variances. Secondly, the utilization of graphlets as explanatory tools is inherently intuitive due to their compact size. This simplicity makes it easier to understand and differentiate each substructure from other graphlets, facilitating clearer and more straightforward explanations. Lastly, explanations derived from graphlets provide insights into the general substructures that GNN models potentially leverage for making predictions. Existing methods tend to pinpoint a specific subgraph within the original graph as the key contributor influencing GNN predictions. However, in many datasets, it is the substructures that play a pivotal role in shaping the behavior of GNNs, rather than any specific segment of the graph. For instance, it is the 6-atom carbon ring structure causing a molecule graph to be mutagenic, rather than a specific subgraph containing a carbon ring. Our approach extracts the general substructure patterns that GNNs consider important, offering a more generalized understanding of what drives model predictions.

4.4.2 Graph Neural Network: Graph-level Classification

Our approach utilizes graphlets to provide intuitive substructure explanations for GNNs within graph-level classification tasks. Graph-level classification tasks involve GNNs predicting the probability distribution across potential classes for the label of a graph. This means that the GNN model assesses an entire graph as an input and outputs a prediction regarding which category or class the graph belongs to, based on learned structural traits within the graph data. The typical approach involves the GNN first learning to generate a comprehensive embedding for each graph. This embedding encapsulates the essential structural information of the graph. Subsequently, this graph embedding is used as input to a simple linear layer, which serves to reduce the dimensionality of the embedding down to the number of potential classes. The application of a softmax function on this output then yields a probability distribution, where each probability lies between 0 and 1, and the sum of all these probabilities equals 1. The rationale behind employing a simple linear layer for dimensionality reduction of graph embeddings to probabilities is founded on the premise that if the GNN has effectively learned the embedding, it should retain ample structural information from the input graph to facilitate accurate predictions. Therefore, a simple linear transformation suffices for this critical step, translating the rich, learned embeddings into understandable class probabilities.

We adhere to this conventional approach to train our GNN models, ensuring not only accurate classification probabilities but also the preservation of crucial topological characteristics within the graph embeddings. We train a Graph Convolutional Network (GCN) [31] on graph classification tasks using both real-world and synthetic datasets. Prioritizing the aggregation of structural information, we discard any node or edge attributes, opting instead for the one-hot encoding of node degrees as the initial vector for each node. Our GCN architecture consists of four layers, followed by a fully connected layer. The outputs of the four hidden layers, each with a dimensionality of 20, are concatenated to form node embeddings, which are then aggregated into an 80-dimensional graph embedding. The decision to utilize 4-layer architecture is closely tied to our utilization of graphlets. Given that the largest diameter among graphlets is 4 (specifically, the

5-node acyclic graphlet), it becomes necessary to employ up to 4 layers in our GNNs. This enables the models to aggregate information from nodes up to 4 hops away and effectively capture up to 5-node substructures within the graph. Furthermore, by concatenating the output from each hidden layer to form the graph embeddings, we facilitate the expression of substructures captured from lower-hop neighbors in the graph embeddings. This helps mitigate any preferential bias towards the 5-node structures, which would potentially dominate in the last layer of the network. Thus, our approach ensures a more balanced representation of the various substructures within the graph.

4.4.3 Surrogate Model: GNN Performance Approximation

To improve the interpretability of our explanations, we propose the use of surrogate models designed to approximate the behavior of GNNs by leveraging simpler neighboring data derived from input graphs. Specifically, we develop a machine learning (ML) encoder-decoder architecture aimed at reconstructing the graph embeddings produced by GNNs based on the graphlet frequency vectors. Notably, although the use of connected graphlets to approximate GNN behavior may not capture disconnected topological information, if present, this approximation aligns with the aggregation mechanism of GNNs, which gather information from direct neighbors.

For each graph, we utilize its 29-dimensional graphlet frequency vector as input to the *encoder*. Through two 20-dimensional hidden layers, the *encoder* outputs a 10-dimensional bottleneck latent vector. This 10-dimensional latent vector is then fed into a linear layer and softmax function, reducing the dimensionality from 10 to 2D probabilities. The encoder is supervised by the probability outputs produced by a trained GCN using L2 loss, aiming to reconstruct the GCN’s performance. The *decoder* takes the 10-dimensional latent vector as input. Through two 20-dimensional hidden layers with symmetrical residual connections from the *encoder*, it outputs an 80-dimensional reconstructed graph embedding. This reconstructed graph embedding is then fed to the frozen classification model trained alongside our GCN, transferring the 80-dimensional embeddings to 2-dimensional softmax probabilities. The training of the *decoder* is guided by two loss functions: the graph embedding reconstruction loss and the probability reconstruction loss (i.e., mean

squared error/L2 loss).

We train our encoder-decoder model by alternating between training the *encoder* and the *decoder*. Specifically, we train the *encoder* for one step and then freeze its weights to obtain the latent vector, which is fed into the *decoder*. After updating the *decoder* weights for one step, we freeze its weights and commence the next training step for the *encoder*. This training scheme ensures that the *encoder* generates latent vectors that preserve similar structural information as the GCN. These latent vectors can then be reconstructed into similar graph embeddings, resulting in similar classification probabilities. Moreover, our architecture inherently mitigates the possibility of overfitting. By reducing the graphlet information to a 10-dimensional latent vector in the *encoder*, we create a bottleneck that forces the model to learn a compact and general structural representation, which enables successful reconstruction in the *decoder*. This bottleneck also prevents the model from simply memorizing the mapping between graphlet frequency vectors and graph embeddings because most raw graphlet information (i.e., input of encoder) is substantially reduced.

Furthermore, our approach addresses the challenge of defining a suitable neighboring data space for graphs as discussed in Sec. 4.3.1. By extracting graphlet frequencies directly from the original graphs, we ensure that these frequencies accurately represent the neighboring space relative to the original data. Moreover, existing methods mostly focus on identifying local neighboring graphs. In contrast, the graphlet frequency vector acts as a comprehensive structural summary of each graph, applicable across various graph types. Therefore, our method for constructing a neighboring space is not confined to local representations but extends to a global scale, enhancing the overall comprehensiveness.

4.4.4 Factual and Counterfactual Explanation Evaluation

To explore diverse graphlet substructures and formulate hypothesized explanations, we have developed a visual analytics system, elaborated upon in detail in Sec. 4.5. Through examining relationships between GNN classifications and graphlet frequencies, users select a specific graphlet as their hypothesized explanation. Subsequently, we employ two reasoning processes—factual and counterfactual—to assess the validity of this explanation.

For instance, if the hypothesized explanation is *graphlet*₂, it must sufficiently contribute to the GNN’s ability to maintain its current classifications to be deemed successful in the factual reasoning process. To quantify this sufficiency, we employ Spearman’s correlation coefficient. Specifically, we analyze the correlation between the GNN classification confidence scores and the frequency of *graphlet*₂ across all graphs of interest to the user. This correlation coefficient indicates the extent to which the frequency of *graphlet*₂ aligns with the trend observed in the GNN’s classification probabilities. A high correlation coefficient suggests that the presence or absence of *graphlet*₂ significantly influences the GNN’s classifications, supporting its role as a meaningful factual explanatory factor.

A successful counterfactual explanation necessitates demonstrating its indispensability for a GNN to preserve its current classifications. This implies that the removal of the explanatory graphlet should significantly alter the GNN’s classification confidence. To remove a substructure from the original graph, we introduce perturbations in the graphlet frequency vector as input to our surrogate models, as described in Sec. 4.4.3. We then observe how these perturbations influence the classification results. Specifically, we conduct two inference runs for each graph using the trained surrogate model. The first run employs the original, non-perturbed graphlet frequency vector as input, obtaining the classification confidence from the surrogate model. Given that the surrogate model only approximates the GNN’s behavior, this initial run is crucial for establishing baseline comparison standards. The second run involves inputting the perturbed graphlet frequency vector to assess its impact on the surrogate model’s classification confidence. Perturbation involves setting the frequency of the targeted graphlet to zero.

To accurately account for the interdependencies among graphlets of varying sizes, such as between 4-node and 5-node graphlets, our method carefully considers the hierarchical nature of these substructures. For example, a 4-node star graphlet is a subgraph of a 5-node star graphlet. This relationship necessitates adjustments in their frequencies during perturbations. Consider the case where the perturbation targets a 4-node star graphlet. Upon setting its frequency to zero, it is imperative to adjust the frequency of any containing graphlet, such as the 5-node star, to zero as well. Conversely, when targeting

a 5-node star graphlet for perturbation (i.e., set the frequency of the 5-node star graphlet to zero), adjustments are required for the frequency not only of the 4-node star graphlet but also of other 4-node graphlets interdependent on the 5-node star. This adjustment is necessary because a 5-node star can be formed by adding edge(s) and node(s) to any dependent 4-node graphlet. To compute this adjustment, we utilize a softmax function on the frequencies of the dependent 4-node graphlets. This function produces a softmaxed frequency vector, providing an estimate of how much each 4-node dependent graphlet contributes to the formation of the 5-node star graphlet. Subsequently, for each dependent 4-node graphlet, we subtract from it the frequency of the 5-node star graphlet weighted by its softmaxed frequency. This adjustment process is applied across all interdependent graphlets, retrieving and calibrating the frequencies of the dependent graphlets from 5-node to 4-node, and recursively from 4-node to 3-node graphlets.

After introducing perturbations (i.e., target graphlet removal) to the graphlet frequency vector, we compare the classification confidence before and after the removal by calculating the sum of the absolute values of L1 distance for each graph between the two inference runs. A high L1 distance indicates that the selected graphlet serves as a critical substructure, substantially influencing the GNN’s classifications. Additionally, we utilize the confidence scores output by the complete encoder-decoder model to initially project the graphlet frequency vector to the same graph embedding space as the GCN. This approach mitigates the influences stemming from performance differences between the surrogate model and the GCN when computing the L1 distance between the two confidence scores. Furthermore, our strategy of perturbing the original graphs through adjustments to their graphlet frequency vectors, coupled with the identification of graphlet interdependencies, addresses the out-of-distribution (OOD) challenge. The OOD issue arises when the removal of a subgraph leads the residual structure to deviate from the original graph’s distribution, often resulting in inaccurate classifications due to these off-manifold inputs. Employing the original graph’s structural summary (i.e., the graphlet frequency vector) as a baseline for comparison ensures consistent behavior from the surrogate model. By considering the interdependencies between graphlets during perturbation, we enhance the

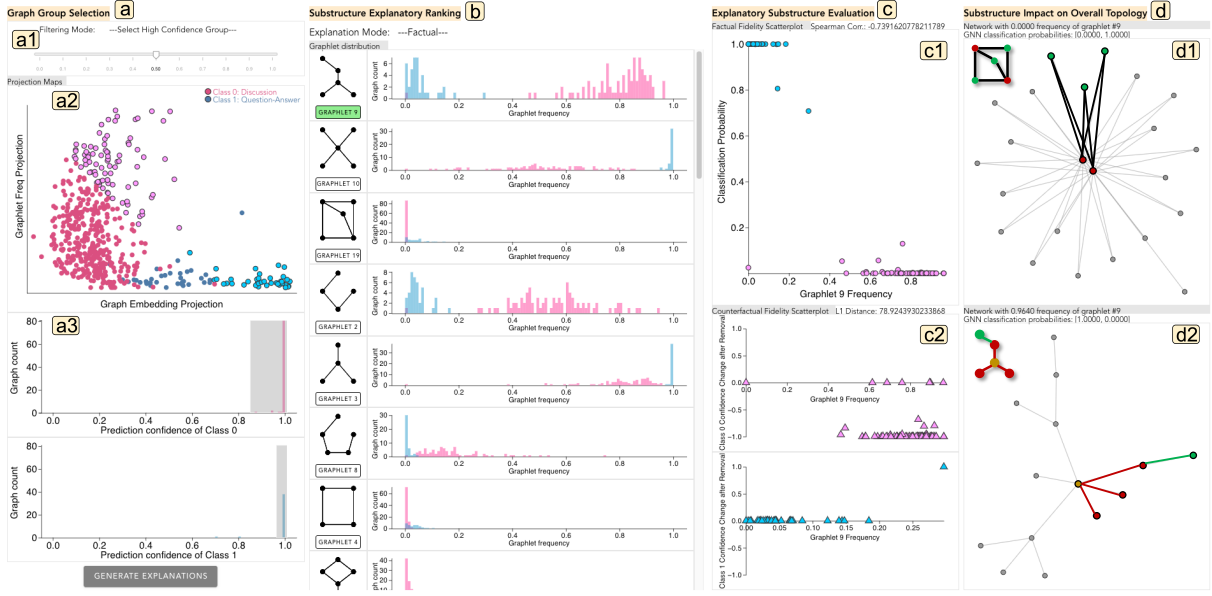


Figure 4.3: The visual interface for facilitating interactive analysis using the workflow in Fig. 2.1. (a) Relating to Step 4, this view visualizes input attributes’ contributions to the 1D network representation. (b) Composite variables generated in Step 5 are shown as a list of scatterplots, where the 1D network representation and composite variable correspond to x - and y -axes, respectively (see b2). As shown in b1, when a composite variable is not generated yet, a swarm-plot-like visualization presents the 1D network representation generated through Step 3 to help assess its quality. For a and b, we employ our two-class density scatterplots. (c) A node-link diagram and (d) a set of histograms inform the network structure and attribute distributions. (e) Other auxiliary information is displayed, including the prediction accuracy of NNs trained for Step 2.

perturbed graphs’ realism.

4.5 Visual Analytics of GNNAnatomy

Our visual analytics system stands as a pivotal tool for the interactive development and assessment of explanations. By employing graphlets as the explanatory elements, the evaluation metrics extend beyond comparing different explanation methods to include an examination of each substructure’s contribution to GNN classifications. Typically, existing explainers utilize a common metric for evaluation, establishing a baseline to identify the most effective method according to the metric used. Yet, this approach lacks indicators to inform users whether the baseline itself is sufficiently robust. Addressing this gap, our system integrates an additional validation layer into the explanation formulation process. This enables users to scrutinize all substructures, assessing their relevance to GNN classifications and determining their potential as explanatory factors through both quantitative and qualitative visual aids provided by our visual analytics system.

The system consists of four columns that facilitate the workflow of GNNAnatomy.

Firstly, in the *Graph Group Selection* column, the *projection map* visualizes the relationship between graphlet frequencies and GNN classifications, aiding users in identifying a group of graphs of interest. Secondly, the *Substructure Explanatory Ranking* column assists in creating an explanatory hypothesis by providing class-wise *graphlet distribution* for each graphlet. Next, in the *Explanatory Substructure Evaluation* column, the validity of the hypothesis can be assessed through the support from *fidelity scatterplots*, which offer both visual and quantitative insights into the sufficiency and necessity of the hypothesis. Finally, to demonstrate the overall impact of substructures on the graph topology, two network visualizations are provided in the *Substructure Impact on Overall Topology* column. Notably, in this interface, the classification probability output for the class a graph belongs to is referred to as “confidence”. Conversely, the classification probability output for $Class_1$ is termed as “classification probability”. For instance, in a Discussion graph (i.e., $Class_0$), if the GCN probability output is $[0.75, 0.25]$, its confidence score is 0.75 and the classification probability is 0.25. In the following paragraphs, we will introduce these visual components in detail, highlighting the specific information they convey.

4.5.1 Graph Group Selection

As depicted in Fig. 4.3(a), this column comprises three components. Firstly, a slider is incorporated to enable users to filter the graphs based on their confidence scores. Users can choose the group of graphs with higher or lower confidence scores than the slider value through the dropdown menu in Fig. 4.3(a1). The remaining graphs are projected onto a scatterplot, as shown in Fig. 4.3(a2). This projection map helps users discern the associations between graphlet frequency vectors and graph embeddings by projecting them from high-dimensional space to a single value. The x-axis value of each graph represents the first principal component (PC1) of its graphlet frequency vector (i.e., graphlet frequency projection), while the y-axis value encodes the PC1 of the graph embedding generated by the GNN (i.e., graph embedding projection). In this scatterplot, users can lasso-select a group of graphs for which explanations are generated, highlighted with outlined pink and blue. To further refine the selection, the chosen group of graphs is redrawn into two histograms, as illustrated in Fig. 4.3(a3). Each histogram displays the graph distribution

over confidence scores, empowering users to select specific sections of confidence scores for different classes through brushing. By providing connections between GCN classifications and graphlet frequencies, this column enables the creation and assessment of GNN explanations at specified levels of granularity.

4.5.2 Substructure Explanatory Ranking

As depicted in Fig. 4.3(b), a sequence of histograms presents the class-wise distributions of graphs within the selected group across the frequencies of each graphlet. The ranking (i.e., the order of display from top to bottom) of graphlets is determined by either our factual or counterfactual metric, based on the chosen explanation mode from the dropdown menu. As elaborated in Sec. 4.4.4, our factual metric sorts graphlets by the descending absolute value of Spearman’s correlation coefficient, while the counterfactual metric ranks graphlets based on the descending sum of absolute L1 distances between the confidence scores from original and perturbed graphs. Hovering over any bin within a histogram automatically highlights all bins of the same class across all histograms by increasing their opacity, providing a clearer visual representation of the distribution. This visual component guides users towards potential explanations for the GNN behavior of the selected group of graphs. Upon clicking a graphlet, users can evaluate the validity of this graphlet in detail.

4.5.3 Explanatory Substructure Evaluation

As depicted in Fig. 4.3(c), we offer two fidelity visualizations: factual and counterfactual fidelity scatterplots.

Factual explanation fidelity. In Fig. 4.3(c1), a scatterplot is presented, where each graph in the selected group is plotted with its frequency of the chosen graphlet encoded as the x-axis value and the classification probability encoded as the y-axis value. This scatterplot enables users to inspect the alignment between graphlet frequency trends and classification probabilities, indicating the efficacy of the graphlet in explaining GNN behavior for classifying the selected group of graphs. Quantitatively, the Spearman’s correlation coefficient, displayed in the top right corner, provides a statistical measure of this relationship. This visualization updates only when users choose “factual” via the

explanation mode dropdown menu in Fig. 4.3(b).

Counterfactual explanation fidelity. In Fig. 4.3(c2), two scatterplots illustrate class-wise L1 distances between confidence scores derived from two surrogate model inferences using original and perturbed graphs, as explained in Sec. 4.4.4. These scatterplots provide detailed insights into changes in classification confidence, distinguishing between positive (improved confidence due to graphlet removal) and negative (decreased confidence due to graphlet removal) effects for each graph, symbolized by triangles along the y-axis. The x-axis denotes the frequencies of the selected graphlet in the original graphs. Additionally, a summation of absolute L1 distances is displayed in the upper right corner, serving as a summary of confidence changes resulting from perturbations (i.e., graphlet removal).

4.5.4 Substructure Impact on Overall Topology

As illustrated in Fig. 4.3(d), two network visualizations serve as examples to depict the overall topological impact resulting from differences in the frequency of the chosen graphlet. In the factual explanation mode, the **top view** (i.e., Fig. 4.3(d1)) showcases the structure of the graph with the lowest frequency of the selected graphlet. This graph is selected from the class with generally lower frequencies of the chosen graphlet. For instance, if there’s a negative correlation between the graphlet frequency and the classification probability, the class with lower frequencies of the selected graphlet is *class*₁. Conversely, the **bottom view** (i.e., Fig. 4.3(d2)) displays the graph with the highest frequency of the selected graphlet from the class with generally higher frequencies. In the case of a negative correlation, this graph would be from *class*₀.

In the counterfactual explanation mode, the **top view** shows the network with the smallest change in classification confidence following a perturbation, while the **bottom view** presents the graph experiencing the largest change in classification confidence after perturbation. By comparing the structures of representative graphs, users can map the impact from substructures to overall topologies, enabling them to reason about the typical appearance of graphs from different classes.

4.6 Case Studies

We demonstrate the capabilities of GNNAnatomy through case studies using both real-world and synthetic graph datasets. Alongside quantitatively measuring the validity of the graphlet explanations using our factual and counterfactual metrics, we compare our explanations to those generated by GNNExplainer [5]. We opt to compare with GNNExplainer because (1) it identifies important motifs for datasets lacking ground truths and (2) its explanatory subgraphs have been widely utilized as baselines for comparison with most other explainers. The quantitative evaluation metric proposed by GNNExplainer focuses on a specific subgraph’s alignment with the ground truth, which isn’t suitable for the substructure explanations generated by GNNAnatomy. Therefore, we provide qualitative comparisons between GNNAnatomy and GNNExplainer’s explanations.

4.6.1 Datasets

4.6.1.1 Real-World Datasets

We utilize two real-world graph-level binary classification datasets. **Reddit-Binary** [166] comprises 2000 graphs, each representing a discussion thread extracted from Reddit. In these graphs, nodes correspond to users engaging in the thread, while edges denote replies between users’ comments. The graphs are labeled based on the nature of user interactions observed in the threads. Specifically, the dataset distinguishes between two types of interactions: Question-Answer, extracted from the subreddits `r/IAmA` and `r/AskReddit`, and Discussion, logged from the subreddits `r/TrollXChromosomes` and `r/atheism`. **Mutagenicity** [136] consists of 4337 molecule graphs, each labeled based on its mutagenic effect on the Gram-negative bacterium *S. typhimurium*.

4.6.1.2 Synthetic Dataset

To construct a graph dataset with ground truth motifs, we draw inspiration from the approach proposed by Ying et al. [5]. They utilize the Barabasi-Albert random graph model (BA) to generate a base graph comprising 300 nodes. This base graph was then enriched by attaching 80 house motifs. The task for GNN is to perform node classifications across four classes: nodes within the top, middle, and bottom sections of a house motif, along with nodes in the base graph itself. Given that our explainer aims to elucidate GNN

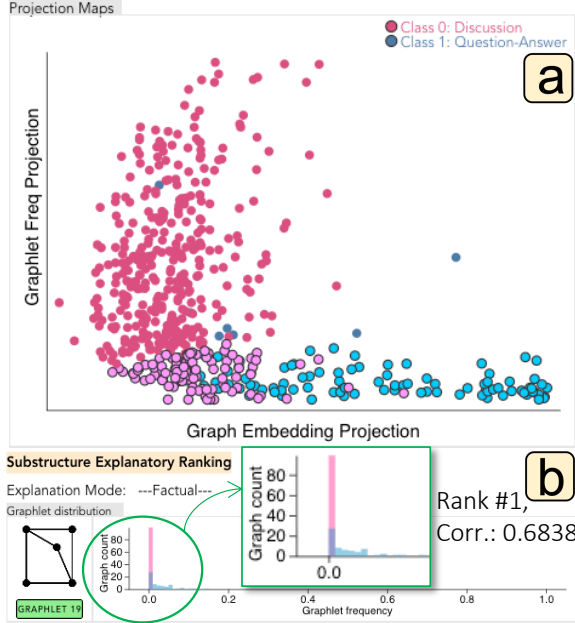


Figure 4.4: (a) illustrates the selection of a group of graphs with similar topologies, aimed at observing subtle substructural differences between the two classes in Reddit-Binary dataset. (b) indicates the top-ranked factual explanatory graphlet and the graph distribution over the its frequency. The correlation coefficient reflects the strong relationship between its frequency and the classification probability.

behavior in graph-level classification tasks, we create **BA-House** with graph labels. We employ the BA model to generate 300 base graphs, each with a random node count ranging from 10 to 40. Subsequently, for half of these base graphs, we attach a random number of house motifs, ranging from 2 to 10. The ratio between the size of a base graph and the number of attached house motifs maintains consistency with the ratio used in [5] (i.e., 300 nodes in the base graph and 80 attached house motifs). These 300 synthesized graphs are then labeled with two classes: those without any attached house motifs and those with attached house motifs. Given the design of the dataset, it is hypothesized that the primary substructure crucial for GNN classifications would be the attached motif, namely, the house substructure.

4.6.2 Study 1: Reddit-Binary

Experiment settings. We train the GCN and surrogate model using graphs with fewer than 100 nodes due to the computational cost associated with graphlet frequency calculation. Among the remaining 554 graphs, consisting of 101 graphs of the Question-Answer (QA) class and 453 graphs of the Discussion class, the GCN model achieves a classification accuracy of 0.9549 (i.e., over 0.5 confidence score), while the classification probabilities output by the surrogate model exhibit a cosine similarity of 0.9322 with the GCN prob-

abilities output.

Group-level explanations. As depicted in Fig. 4.3(a2), the projection map reveals that Discussion graphs display a more diverse range of structural characteristics, as their graphlet frequency projections are scattered along the y-axis. In contrast, the graphlet frequency projections of Question-Answer (QA) graphs remain tightly clustered at the bottom of the y-axis, indicating a more consistent topology among QA graphs. Most graphs from different classes are separable along the x-axis (i.e., graph embedding projection), indicating that the GCN has successfully captured distinguishable structural features even for those graphs at the bottom with similar topologies across classes. We further explore the potential substructures that contribute to the GCN’s ability to make such classifications.

We choose a group of graphs with distinct topologies and graph embeddings (i.e., separated graphlet frequency projections and graph embedding projections) as shown in Fig. 4.3(a2). This selection aims to identify the salient substructure differences between the two classes that are relevant to the behavior of the GCN. In Fig. 4.3(b), the distribution of the top-ranked graphlet, *graphlet*₉, in the factual explanation mode exhibits clear separation across classes. This suggests that classifying graphs based solely on the frequency of *graphlet*₉ could potentially achieve the same high classification accuracy as the GCN, which is 0.9549. The factual fidelity scatterplot, depicted in Fig. 4.3(c1), validates a strong negative correlation (Spearman correlation corr.: -0.74) between the frequency of *graphlet*₉ and the classification probability. Furthermore, in Fig. 4.3(c2), the necessity of *graphlet*₉ in maintaining the current GCN classifications is confirmed. Upon removal of *graphlet*₉, the confidence scores of Discussion graphs decrease while those of QA graphs with unusually high frequencies of *graphlet*₉ increase. These changes indicate that the GCN tends to classify most graphs as QA graphs when *graphlet*₉ is absent from their topologies, thereby validating the explanatory power of *graphlet*₉. In Fig. 4.3(d1, d2), the visualizations of graphs with the lowest and highest frequencies of *graphlet*₉ communicates the topological traits of the two classes: QA graphs (Fig. 4.3(d1)) tend to have multiple centralized individuals that interact with most others, whereas Dis-

cussion graphs (Fig. 4.3(d2)) typically feature branched-out discussions. Annotations in Fig. 4.3(d2) and the context of the Reddit-binary dataset indicate that *graphlet*₉ captures common branched-out discussions, absent in typical QA graphs (Fig. 4.3(d1)).

Further examination of the projection map prompts us to investigate another set of graphs, as shown in Fig. 4.4(a). While the GCN model generally distinguishes between graphs from different classes, they appear scattered along the x-axis. This scattering occurs because some Discussion graphs share similar topologies with QA graphs at the bottom of the y-axis. As a result, capturing the salient structural differences might not be enough for the GCN to make effective classifications, and it is pushed to uncover subtle topological disparities between classes. These subtle structural differences are encoded in the graph embeddings learned by GCN, which further scatter the projections of graphs from each class within their respective clusters. Consequently, we focus on selecting graphs with similar topologies to identify these subtle structural disparities that aid in the GCN’s classification effectiveness. As depicted in Fig. 4.4(b), the frequency of the top-ranked *graphlet*₁₉ correlates with the GCN classification probability with a correlation coefficient of 0.6838. Additionally, the typical topological trait of QA graphs observed in the previous group of graphs is encapsulated by *graphlet*₁₉, as annotated in Fig. 4.3(d1). This suggests a reasoning process for the GCN: while the common structural trait of Discussion graphs (*graphlet*₉) is sufficient to distinguish most graphs, when classifying graphs with similar topologies, the common structural trait of QA graphs becomes the next indicator that the GCN relies on.

In summary, it is evident that GCN employs different substructures to classify separate groups of graphs. GNNAnatomy not only helps users identify these groups but also rank the relevance of graphlets using on our factual and counterfactual metrics. By interactively evaluating the potential explanations, users can formulate a satisfactory explanatory substructure for the chosen group.

Qualitative evaluation. As depicted in Fig. 4.5(a1, b1), our graphlet explanation for the Discussion class incorporates an additional edge and node, highlighted in green, compared to the ground truth motif proposed by GNNExplainer. This difference sug-

gests that our graphlet explanation captures the presence of branch-out substructures, which are uniquely characteristic of Discussion graphs but not addressed by GNNExplainer’s motif. It is worth noting that the Discussion motif proposed by GNNExplainer (Fig. 4.5(a1)) is prevalent not only in typical Discussion graphs but also significantly in typical QA graphs. In Fig. 4.3(d1), although the graph represents a typical QA graph topology, a recurring pattern emerges where one of the two central nodes connects to three peripheral nodes. The fact that this recurring pattern in a representative QA graph is identified as the Discussion motif proposed by GNNExplainer (Fig. 4.5(a1)) underscores the limited explanatory power of their approach. In contrast, our explanations for the two classes (Fig. 4.5(b1, b2)) depict distinct topologies. As illustrated in Fig. 4.3(d1, d2), our Discussion explanation (Fig. 4.5(b1)) is not found in a typical QA graph (Fig. 4.3(d1)), while our QA explanation (Fig. 4.5(b2)) does not occur in a typical Discussion graph (Fig. 4.3(d2)). Moreover, our explanations are mutually exclusive, indicating that a 5-node connected subgraph classified as Fig. 4.5(b1) cannot be categorized as Fig. 4.5(b2) based on the edges among these 5 nodes. This highlights the unique substructures captured by our method. For the QA class, GNNExplainer proposed a recurring substructure as the QA motif (Fig. 4.5(a2)), which can be translated into an equivalent graphlet expression (Fig. 4.5(a3)). In contrast, GNNAnatomy directly captures the same substructure explanation (Fig. 4.5(b2)) in a systematic format, without the need for inference from recurring patterns. It is also crucial to recognize that merely presenting the topological traits of each class does not fully elucidate GNN’s behavior. As demonstrated in this study, different criteria apply to various groups of graphs. GNNAnatomy offers the flexibility to identify the groups and substructures that are most explanatory for the GNN’s behavior at specified granularities.

4.6.3 Study 2: MUTAG

Experiment settings. We train the GCN using the one-hot encoding of node degrees as the initial vector of each node, discarding any node and edge attributes provided in this dataset. Utilizing pure topological characteristics, the GCN model achieves a classification accuracy of 0.7667, while the surrogate model exhibits a cosine similarity of 0.7371 with

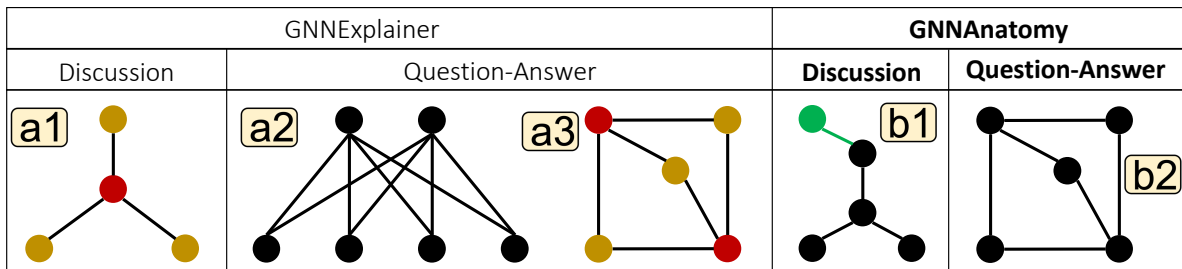


Figure 4.5: (a1, a2) represent the ground truth motifs of each class in the Reddit-Binary dataset proposed by GNNEExplainer. (a3) depicts the translation of (a2) to the graphlet setting. (b1, b2) showcase the extracted explanatory graphlet substructures of each class in Reddit-Binary by GNNAAnatomy.

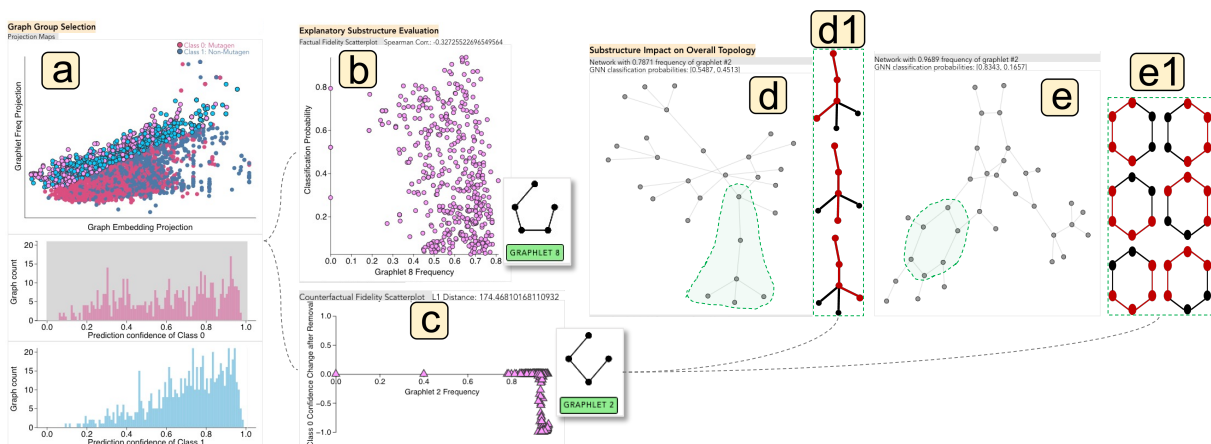


Figure 4.6: (a) presents a selection of mutagenic graphs arranged diagonally, with the aim of identifying the mutagenic substructures that contribute the most to GCN classifications on the MUTAG dataset. In (b) and (c), we evaluate the validity of acyclic substructures *graphlet₈* and *graphlet₂*, which are the top two relevant substructures ranked by our factual metric. (d) illustrates the typical structure of graphs with a lower frequency of *graphlet₂*, while (e) depicts the typical topology of graphs with a higher frequency of *graphlet₂*. We focus on the 6-node subgraphs in (d1) and (e1) to demonstrate that more acyclic patterns can be sampled from a ring than from a non-ring substructure (i.e., 3 vs. 6).

the GCN probability output.

Group-level explanations. In molecular biology literature, certain substructures associated with mutagenicity have been identified, but there’s been limited discussion on non-mutagen-specific substructures. Thus, our study aims to extract traits used by GCN to classify a graph as mutagenic. In Fig. 4.6(a), we selected a group of graphs distributed diagonally and filtered out non-mutagens. This selection aims to identify structural differences (i.e., y-axis deviations) within the mutagenic graphs that contribute to disparities in graph embedding projections.

The top-ranked graphlets in factual explanation mode are both acyclic substructures: *graphlet₈* and *graphlet₂*, as shown in Fig. 4.6(b, c). In Fig. 4.6(b), a moderate negative correlation of -0.33 between the frequency of *graphlet₈* and classification probability

indicates that a graph with a higher frequency of *graphlet*₈ has a higher chance of being classified as a mutagen. To investigate the necessity of the presence of these acyclic graphlets in classifying graphs as mutagens, we observe the impact on confidence scores from the removal of both *graphlet*₂ and *graphlet*₈. Due to the interdependency between *graphlet*₂ and *graphlet*₈, selecting to remove *graphlet*₂ automatically removes *graphlet*₈, as detailed in Sec. 4.4.4. In Fig. 4.6(c), the classification confidence scores decrease for a considerable number of graphs after the removal. This validates the necessity of *graphlet*₂ and *graphlet*₈, as GCN exhibits reduced confidence in classifying a graph as a mutagen when these substructures are absent from its topology.

Comparing Fig. 4.6(d) to Fig. 4.6(e), it is apparent that the presence of a 6-node ring structure is exclusive to the graph with a higher frequency of *graphlet*₂, indicating an association between the two. To validate this observation, we compare the number of *graphlet*₂ that can be sampled from a 6-node non-ring subgraph to that from a 6-node ring subgraph. As depicted in Fig. 4.6(d1, e1), we observe that a single ring structure can produce twice as many *graphlet*₂ as a non-ring structure can do. Furthermore, considering the edges attached to a ring structure, even more *graphlet*₂ can be sampled. Hence, the necessity of the presence of *graphlet*₂ and *graphlet*₈ in classifying graphs as mutagens by GCN may stem from the presence of a 6-node ring substructure.

In conclusion, every graph dataset may contain diverse topological traits that GNNs capture, which can be complex and unexpected. GNNAnatomy provides a systematic approach and ample visual aids to explore the fundamental substructure elements contributing to the significant structural traits associated with GNN classifications.

Qualitative evaluation. In Fig. 4.7(a1, a2, a3), GNNExplainer proposed three substructures commonly associated with mutagenicity: *NO*₂, *NH*₂, and the 6-atom carbon ring. However, as our methods focus on extracting pure structural importance, we do not directly compare our explanations with *NO*₂ and *NH*₂, as they require node attributes to indicate their presence. Instead, we compare with the carbon ring, treated as a simple 6-node ring substructure. As summarized in Fig. 4.7(b1, b2), GNNAnatomy extracts two acyclic patterns as important topologies indicative of mutagenicity, which we infer from

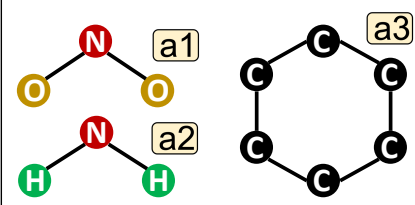
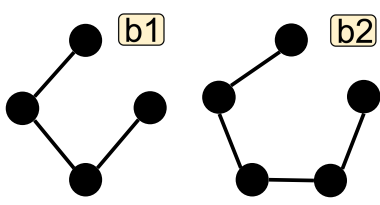
GNNExplainer	GNNAnatomy
Mutagenic	Mutagenic
	

Figure 4.7: (a1, a2, a3) depict the ground truth motifs of the Mutagen class in the MUTAG dataset as proposed by GNNExplainer. (b1, b2) showcase the extracted explanatory graphlet substructures of mutagenicity by GNNAnatomy.

supporting visualizations to be incomplete rings, matching the proposed ground truth motif by GNNExplainer.

In summary, GNNAnatomy explores a diverse spectrum of substructures and uses visualizations to suggest any unexpected substructures that GNNs may utilize for classifications.

4.6.4 Study 3: BA-House

Experiment settings. The GCN is trained to classify whether a base graph generated by the Barabasi-Albert random graph model (BA) is attached with house motifs. The GCN model achieves a classification accuracy of 1.0, while the surrogate model shows a cosine similarity of 0.9421 with the GCN probability output.

Global-level explanations. Since the BA-House dataset is synthetic and the attached house motifs are known topological differences between the two classes, we aim to investigate if the GCN utilizes the house motif to distinguish graphs. We selected all 300 graphs to explore the global-level explanatory graphlets.

In Fig. 4.8(a), $graphlet_8$ is ranked first, indicating its relevance to the classification probability and its higher presence in House graphs. Additionally, in Fig. 4.8(a1), it shows that the frequency of $graphlet_8$ has a negative correlation (-0.4156) with the GCN classification probability. This suggests that a graph with a higher frequency of $graphlet_8$ has a larger chance of being classified as a House graph by the GCN. Moreover, in Fig. 4.8(a2), the confidence scores decrease for the House graphs but increase for Non-House graphs after the removal of $graphlet_8$, highlighting the necessity of its presence in

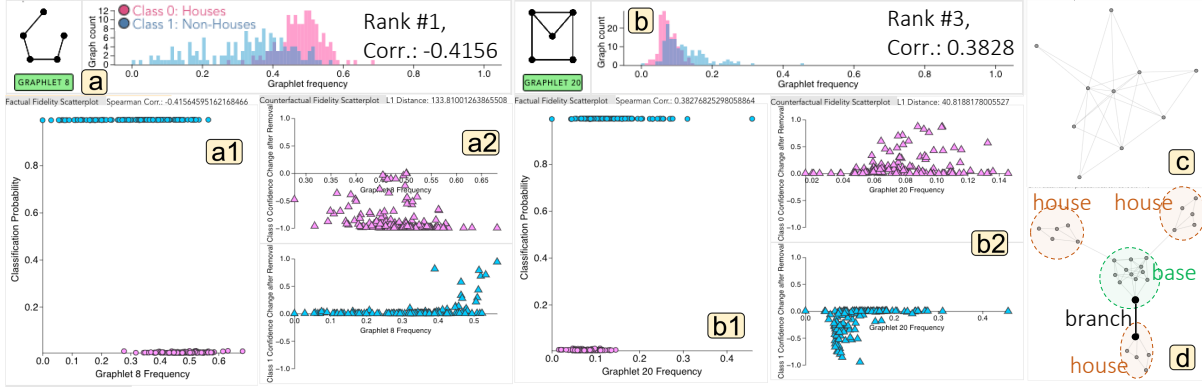


Figure 4.8: In the BA-House dataset experiment, (a) illustrates the distribution of graphs based on the frequency of the house motif (i.e., $graphlet_{20}$), while (b) presents the distribution for $graphlet_8$. We observe opposite correlation signs in (a1) and (b1), along with opposite influences on the classification confidence scores of the two classes after removing the two substructures separately in (a2) and (b2). Further investigation in (c) and (d) reveals the root cause of this phenomenon: attaching house motifs to a BA base graph creates branch edges, as indicated in (d).

classifying graphs as House graphs.

Investigating the influence of the house motif in Fig. 4.8(b), it is evident that Non-House graphs generally have a higher frequency of the house motif ($graphlet_{20}$) compared to graphs with attached house motifs. Additionally, as indicated in Fig. 4.8(b1), the frequency of $graphlet_{20}$ has a positive correlation (0.3828) with the GCN classification probability. This implies that a graph with a higher frequency of the house motif has a higher chance of being classified as, surprisingly, a Non-House graph by the GCN. Furthermore, as presented in Fig. 4.8(b2), the removal of the house motif counterintuitively increases the confidence scores for House graphs but decreases those for Non-House graphs.

Although the house motif indeed proves its relevance as per our hypothesis, its influence appears to be the exact opposite of what we expected. To understand the causes, we examine the overall graph topologies. In Fig. 4.8(c), we observe that a BA-generated base graph (i.e., a Non-House graph) itself contains multiple house motifs due to its high connectivity among nodes. Conversely, a House graph, as shown in Fig. 4.8(d), has one node of each house motif attached to a node in the base graph, resulting in a lower overall edge density compared to a Non-House graph. As indicated in Fig. 4.8(d), the *branch* edges are the root cause of House graphs having a generally lower frequency of the house motif. Only one house motif could be sampled from each house attached. However, the *branch* edges connecting houses to the base graph allow the computation of graphlet frequency

to sample additional 5-node connected subgraphs involving nodes in both the house and the base. These additional subgraph samples are not house motifs. For example, when 3 houses are attached, many more 5-node subgraphs can be sampled compared to a single base graph, but only 3 among those additional samples are house motifs. We validate our finding by computing the frequency of the house graphlet before and after attaching house motifs, which shows the average frequency indeed decreases from 0.1313 to 0.0686.

Upon further reasoning, we find that the *branch* edges may also contribute to *graphlet₈* being the top-ranked explanatory substructure. As discussed, *branch* edges enable subgraph sampling across attached houses and the base graphs, increasing the chance of sampling an acyclic graphlet (i.e., *graphlet₈*). Conversely, due to the high connectivity of a Non-House graph, sampled 5-node subgraphs are rarely acyclic. In summary, despite our assumptions about what the GCN should utilize based on the known structural disparities, GNNAnatomy helps us identify the true decisive factors with ample evidence presented in the supporting visualizations. This study also paves the way for using our methods to validate presumed ground truth motifs for other synthetic graph datasets.

Qualitative evaluation. We posit that the graphlet sampling mechanism naturally leads to a higher frequency of acyclic substructures when attaching houses, but results in a lower frequency of the house graphlet. This raises the question of whether the graphlet sampling mechanism accurately reflects GNN behavior, warranting the connection of the results of the two.

Given that GNNs aggregate node information from direct neighbors in each layer, as explained in Sec. 4.2.2, they can capture a complete house motif within a 3-layer architecture. However, the presence of *branch* edges enables GNNs to aggregate information across both the base graph and the house motif, potentially hindering the GCN’s ability to fully capture a complete house topology in either component. This impact on neighbor sampling mirrors the discussion in this study concerning graphlet sampling. In essence, aggregating information within a Non-House graph encourages GNNs to sample neighboring nodes that collectively form complete house motifs. Therefore, using graphlet distributions to depict the behavior of GNNs is indicative of their functioning.

4.7 Discussion and Future Works

4.7.1 Scalability in Graphlet Frequency Computation

Using 3-, 4-, and 5-node graphlets ensures a diverse range of substructures while keeping the total number of graphlets manageable. For example, in a graph with 94 nodes and 222 edges, efficient sampling methods proposed in [151] yield 1106, 15113, and 172127 sampled 3-, 4-, and 5-node subgraphs, respectively. To compute the frequency of each graphlet for every graph, we iterate over these samples, checking for isomorphism with the target graphlet. Random sampling from this pool of subgraphs was attempted but found to compromise frequency precision. While this process may be time-consuming for larger networks, it is a one-time operation per dataset. The systematic generation and evaluation of explanations using graphlets justify the computational cost involved. To facilitate the more efficient use of GNNAnatomy with other datasets, we are working on developing a graphlet frequency benchmark dataset.

4.7.2 Generalizability

Other GNN models. GNNAnatomy operates as a post-hoc surrogate approach, leveraging graph embeddings and softmax probabilities generated by a trained GCN. This means that it can be applied to any GNN capable of producing a graph embedding for each graph, enabling explanations of its behavior across various levels of granularity.

Multi-class graph classification tasks. Even though we demonstrate GNNAnatomy’s capabilities using binary classification datasets, our approach can be expanded to multi-class classification tasks. Algorithmically, we need to change both factual and counterfactual metrics to computing the correlation and L1 distance for the chosen graphs within a single class. This adjustment will limit part of the versatility of GNNAnatomy, but can still allow systematic generation and evaluation of explanatory substructures.

Node classification tasks. While GNNAnatomy is primarily tailored for graph classification tasks, its methodology could potentially be adapted for explaining GNN behavior in node classification tasks as well. For instance, instead of utilizing graph embeddings and graph classification probabilities, we could employ node embeddings and

node classification probabilities. Regarding graphlet frequency, we could substitute it with the frequency of a node appearing in a specific graphlet type, such as *graphlet*₂₀, relative to all occurrences of *graphlet*₂₀ found in the original graph. This approach would enable the identification of a node’s involvement in the overall graph topology within the context of graphlet representation.

Motifs with more than five nodes. As discussed in Sec. 4.6.3, the limitation of considering up to 5-node graphlets is that it cannot explicitly capture motifs with more than 5 nodes. This limitation may arise when users expect certain large motifs to be utilized by GNNs. Instead of customizing the selection of graphlets to include these expected motifs, GNNAnatomy demonstrates the capability of implicitly capturing these motifs using fundamental substructure elements. With ample visualization support, GNNAnatomy assists users in formulating correct inferences about the explanatory motifs without the restriction of the number of nodes.

4.7.3 Extensibility

As discussed in Sec. 4.6.2, GCNs leverage different substructures to classify various groups of graphs. Specifically, we observe that GCNs utilize the branch-out substructure to distinguish graphs with distinct topologies. However, when there are no significant differences in graphlet distribution between classes, GCNs gradually resort to other suboptimal substructures (as seen in Fig. 4.4(b1)) to achieve satisfactory classifications. This observation suggests a potential extension of our work to develop a decision tree that hierarchically lists the substructure criteria employed by GCNs at each step for each group of graphs to attain the final performance. We intend to explore this application in future research.

4.8 Conclusion

We present a novel GNN explainer designed to facilitate systematic exploration and evaluation of diverse substructures that explain GNN behavior across multiple levels of granularity. Our approach leverages graphlets to generate substructure explanations and integrates factual and counterfactual reasoning logic to assess their validity. Interactive and supporting visualizations are incorporated to streamline the process of formulating

a trustworthy explanation. Through case studies conducted on real-world and synthetic graph datasets spanning various domains, we demonstrate the efficacy of GNNAnatomy and compare the quality of our explanations to a state-of-the-art GNN explainer. While showcasing the versatility and utility of our approach, we outline future research directions for potential extensions of GNNAnatomy.

Chapter 5

Longitudinal Patient Care Risk Prediction and What-If Analysis for Intervention Planning

Extending the methods developed in Chapter 3 and Chapter 4, this chapter presents an interactive visual analytics framework for longitudinal patient care risk prediction and what-if analysis for intervention planning. This chapter unifies dynamic healthcare collaboration network modeling, graph-based risk prediction, and explanation methods into a full analytical pipeline spanning data curation, mitigation of foreseeable confounding factors, patient morbidity risk estimation, and interpretable model reasoning. Building on these capabilities, the framework further supports simulation of targeted interventions and examination of projected changes in patient risk, enabling users to generate and evaluate intervention hypotheses through interactive what-if analysis.

Cancer care is a longitudinal, time-sensitive process that depends on the coordinated activities of diverse healthcare professionals (HCPs), whose patterns of collaboration evolve continuously throughout the course of treatment. Prior studies have shown that care team structure is associated with patient outcomes; however, the temporal dynamics of these collaborations and their relationship to evolving patient risk remain insufficiently understood. To address this gap, I introduce a visual analytics framework for modeling and interpreting time-varying care team interactions in oncology. I construct dynamic col-

laboration networks from electronic health record (EHR)-mediated communication data and employ a temporal graph neural network (TGNN) to estimate patient outcome risk over time. Building on these predictions, I examine how changes in care team composition and specialist participation are associated with shifts in estimated risk across different treatment stages. To support practitioner-driven analysis, I design a multi-view interactive system that integrates TGNN-based risk modeling with coordinated visualizations of care team networks, feature-level explanations derived from a surrogate logistic regression model, and counterfactual what-if analysis. Rather than inferring causality, the system enables analysts to investigate how risk estimates co-evolve with collaboration patterns and to identify the care team characteristics most strongly associated with these temporal changes. I demonstrate the utility of the system through expert-guided analysis scenarios on a cohort of 505 cancer patients spanning breast, colorectal, and lung cancers, and evaluate its effectiveness in uncovering clinically meaningful collaboration patterns. This chapter contributes a practical approach for improving the interpretability of temporal graph-based models in longitudinal, team-based healthcare settings, with broader applicability to domains involving complex, evolving collaborative processes.

5.1 Introduction

Understanding how clinical teams collaborate over time is critical to improving patient care, particularly in complex domains such as cancer treatment where patients interact with multiple healthcare professionals (HCPs) across extended care trajectories [99, 101, 102, 167, 168]. Prior work has examined the relationship between care team structures and patient outcomes, often modeling collaboration as static networks aggregated over an entire episode of care [169, 170]. However, real-world care delivery is inherently dynamic: team composition, roles, and interactions evolve as patient conditions change, motivating the need for representations that capture these temporal dynamics. Electronic health records (EHRs) provide a promising foundation for this purpose. As the primary platform for documenting and coordinating care, EHRs generate fine-grained, timestamped logs that reflect how healthcare professionals interact around patient care.

While most prior research has focused on patient-level physiological data within EHRs [171–173], relatively little attention has been given to the metadata generated through system usage—such as access logs, shared documentation, or co-signatures—as signals of clinical collaboration. These underutilized digital traces offer a scalable and data-driven means to reconstruct time-evolving collaboration networks, providing new opportunities to study how care teams function in real-world settings and how their dynamics relate to patient outcomes.

In this chapter, I leverage timestamped EHR logs to extract digital traces of communication among HCPs, constructing patient-specific, time-evolving collaboration networks that capture how care teams form and evolve throughout treatment [174, 175]. A central challenge is that clinicians need to consider both how collaboration structures change over time and how patient risk changes alongside them, yet these signals are often examined separately and expressed in different ways. In practice, patient risk is commonly estimated from static representations of collaboration networks [176], leading to a misalignment in which dynamic visualizations of collaboration are paired with risk predictions derived from temporally aggregated inputs that do not reflect those evolving dynamics. To address this mismatch, I model patient risk using a temporal graph neural network [177] aligned with the same time-evolving collaboration networks used in the visual analysis. By aligning the model representation with the same time-evolving networks used in the visual analysis, our approach enables a more coherent interpretation of how changes in team composition and interaction patterns are associated with evolving patient risk, supporting a unified analytical framework for human–AI collaborative reasoning.

To support exploration of these dynamics, I introduce a visual analytics system with coordinated visualizations of temporal collaboration networks and model-derived risk predictions, enabling human–AI collaborative analysis of how changes in team structures relate to patient outcomes. Rather than treating model outputs as opaque predictions, our system allows clinicians to interactively examine how changes in team composition and structure correspond to variations in predicted risk over time. This integration supports clinicians in reasoning about both what the model predicts and how those predictions

relate to the underlying care processes reflected in collaboration dynamics. In particular, clinicians often need to compare multiple patient cases, identify patterns across care trajectories, and understand how different trends in collaboration dynamics may contribute to variations in patient outcomes. To support these needs, our system provides coordinated views that enable comparative and longitudinal analysis through side-by-side comparison of patient-specific collaboration graphs, visualization of their temporal evolution, and dynamic clustering of HCPs into domain-meaningful groups. These capabilities allow users to explore how different team configurations emerge and evolve over time, and how such changes align with shifts in predicted patient risk.

Importantly, our goal is not to establish definitive clinical conclusions, but to support insight generation and hypothesis development. Through case studies and expert feedback, I demonstrate how the system enables clinicians to explore complex relationships between collaboration dynamics and patient risk trajectories, uncover patterns of interest, and generate hypotheses for further investigation. In doing so, I position our effort as a step toward more transparent and interpretable integration of predictive modeling into clinical sensemaking workflows.

In summary, this chapter makes the following contributions:

- A transferrable framework for modeling EHR-mediated collaboration networks.
- An approach to analyzing medical collaboration data that mitigate spurious signals, and demonstrate how machine learning models can be applied to the resulting network representations to predict patient survival;
- A visual analytics framework for investigating cancer care as time-evolving collaboration networks, integrating temporal risk modeling with interactive exploration of care-team structure, composition, and feature-level explanations.
- A coordinated multi-view design that supports cohort-level, comparative, and longitudinal analysis of risk trajectories and evolving care-team patterns, including exploratory what-if analysis through linked interactive views.

- Evaluations and expert feedback showing how the system supports practitioner-driven interpretation, insight generation, and hypothesis development around the association between care-team dynamics and model-estimated risk over time.

Finally, I note that part of the system implementation presented in this chapter was carried out in collaboration with a Master’s student, Yuhua Huang, whose contributions focused on development and engineering support. The conceptual design, methodological development, and analytical components remain the primary contributions of the author.

5.2 Related Work

5.2.1 Healthcare Team Collaboration Analysis

Understanding team-based care in healthcare has long been an important topic in healthcare services research. Prior work has shown that team composition, coordination, and information exchange can affect care delivery and patient outcomes, particularly in complex clinical settings where patients interact with multiple providers across extended care trajectories [99, 101, 102, 167, 168]. Related work on patient-sharing and physician networks further demonstrate that structured clinical records can be used to recover meaningful professional relationships among healthcare providers [169].

With the widespread adoption of electronic health record (EHR) systems, researchers have increasingly used digital traces of clinical activity to study care-team collaboration at finer granularity. Prior work has shown that EHR access-log data can be used to construct patient-centered and virtual care-team networks, revealing heterogeneity in care-team structure and communication patterns across patients and clinical contexts [170, 174, 175]. Visual analytics systems have also been introduced to support exploratory analysis of EHR-mediated communication and collaboration networks [176].

Together, these studies established that EHR-derived care-team networks provide a meaningful representation of collaboration and support exploratory analysis of team structure. However, most prior work has emphasized descriptive social-network analysis or aggregate communication measures, leaving limited support for interactive interpretation of patient-level temporal dynamics in predictive modeling settings.

5.2.2 Modeling and Interpreting Dynamic Graphs

5.2.2.1 Modeling dynamic graphs

Dynamic graph learning has been widely studied in the machine learning community through embedding-based, recurrent, and attention-based methods. Early work such as DynGEM [178] models time-evolving node representations, while later methods such as DySAT [179] incorporates temporal modeling directly into graph neural architectures, and EvolveGCN [180] evolves graph convolution parameters over time rather than assuming a fixed node set.

Within healthcare, recent work has begun applying graph-based learning to EHR-derived collaboration networks for patient outcome prediction [181, 182]. Some approaches model collaboration within a fixed observation window, emphasizing predictive signals embedded in aggregated structure and node attributes, whereas more recent temporal formulations represent collaboration as a sequence of cumulative weekly snapshots to capture how care teams expand and change over the course of treatment [182]. These studies show that evolving collaboration structure can be modeled computationally, but they leave open the interpretability challenge of understanding how changing collaboration patterns relate to evolving patient risk and model behavior over time.

5.2.2.2 Interpreting dynamic graphs

As graph neural networks have become more widely adopted, a substantial body of literature has emerged on explainability for graph-based models [5, 183–187]. Among different approaches, surrogate-based methods approximate the behavior of a graph neural network (GNN) using simpler, interpretable models, enabling explanations in terms of human-understandable features. GraphLIME, for example, uses a local surrogate model to explain graph predictions through interpretable features [188]. In healthcare settings, where analysts reason in terms of provider roles, specialties, and documentation behaviors, such feature-centered explanations are particularly useful, as they align with domain-relevant concepts and are more actionable than explanations based solely on subgraph structures.

Yet, feature-centered explanations alone are often insufficient for supporting temporal analytical tasks. While they can reveal which interpretable features are associated with

a prediction at a given time point, they typically do not support longitudinal comparison, temporal drill-down, or reasoning about how feature contributions evolve alongside changes in collaboration structure. This limitation highlights the need for approaches that situate feature-based explanations within a broader temporal and structural analysis context, enabling users to relate model explanations to the underlying dynamics of the data.

5.2.2.3 Visual analytics systems for graphs

Visual analytics provides a powerful approach for exploring complex relationships and generating evidence to support user-driven reasoning. Such interactive systems enable users to interpret graph neural networks, complex predictive models, and time-varying data through coordinated views and interactive exploration. For example, GNNLens and CorGIE support the inspection of GNN predictions, embeddings, and graph structures [189, 190]. In healthcare, RetainVis demonstrates how visual analytics can facilitate the interpretation of temporal predictive models over electronic medical records [191]. Within the care-team collaboration domain, EHRFlow enables the exploration of communication effectiveness and efficiency in EHR-mediated collaboration networks [176].

These systems demonstrate the value of interactive visual reasoning for model interpretation and healthcare collaboration analysis. However, prior work typically examines temporal collaboration networks, model-estimated risk, and feature-level explanations in isolation and represents them inconsistently—for example, pairing dynamic visualizations of collaboration with risk predictions or explanations derived from temporally aggregated inputs. This creates a mismatch that limits practitioners’ ability to reason about how evolving care-team structures relate to changes in predicted patient risk and their underlying feature attributions. As a result, there remains limited support for coordinated, practitioner-driven analysis of patient-specific care-team dynamics in relation to changing risk trajectories and interpretable temporal attributions.

5.3 Methodology

As discussed in Sec. 5.2, to address the challenge of analyzing time-evolving clinical collaboration and its association with model-estimated risk, I design a visual analytics approach that brings together temporal collaboration networks, backend risk estimation, and interpretable feature attributions within a unified analytical framework. Our goal is to support practitioner-driven reasoning by enabling users to examine how care-team structures evolve over time, how these changes align with shifts in model-estimated risk, and how feature-level explanations provide additional context for interpretation. To this end, I construct patient-specific temporal collaboration networks from EHR-derived interaction logs, estimate patient risk using a temporal graph neural network, and develop coordinated visualizations that support comparative and longitudinal analysis.

5.3.1 Data Processing

Data overview. The University of California Institutional Review Board approved our research (IRB #1988738). Our raw data consists of EHR digital traces of patients diagnosed with Stage 2 or 3 breast, lung, and colorectal cancers. For each patient, the dataset includes their *basic information* and *access logs* of their EHR data. Basic information encompasses demographics (e.g., age and gender), treatments, comorbidities, and survival outcome (alive/dead). EHR access logs contain timestamped events spanning three months before to one year after the diagnosis date. A timestamped EHR access event involves a healthcare professional (HCP) accessing (reviewing or writing) a document, such as a note or a message, on the EHR system. Since HCPs typically record a patient’s medical conditions in notes, while messages often lack context, I include only the access events involving notes in our analysis. To focus on assessing the interactions within core teams, following the recommendations of our medical doctor collaborators, I include only HCPs with the titles Medical doctor (MD), Nurse Practitioner (NP), Physician Assistant (PA), Registered Nurse (RN), Pharmacy Technician, Pharmacist, and Case Manager.

This data captures the flow of information among HCPs, allowing us to track which HCP authored a note and who subsequently read it. Those who read the note may then write additional notes, further disseminating the acquired information. I extract

these collaborative interactions that enable information transfer. To ensure a consistent collaboration timeframe, I exclude patients who passed away within a year of diagnosis, resulting in a total of 505 patients in this study, as summarized in Table 5.1. While this may introduce survivor bias by underrepresenting patients with rapidly progressing disease, it enables a more controlled comparison among patients with similar observation windows. In this setting, the predictive signals captured by the model are more likely to reflect meaningful differences in collaboration patterns rather than being driven primarily by acute clinical severity.

Table 5.1: Patient cohort composition and survival status

Cohort	Total	Alive	Deceased
Breast	212	185	27
Lung	150	89	61
Colorectal	143	119	24
Total	505	393	112

Cohort selection. The study cohort consists of 505 patients diagnosed with Stage II or III breast, lung, or colorectal cancers, selected based on the availability of complete EHR access logs and clinical records within the defined study period. Patients were not randomly sampled; rather, the cohort reflects individuals treated at a single academic medical center with sufficient longitudinal data for analysis. As a result, the cohort may be subject to selection bias, including institutional practice patterns and data availability constraints. These factors may limit the generalizability of the findings to other healthcare settings.

Data processing and categorization. To extract the collaboration surrounding a patient, I identify all notes related to the patient and the HCPs who have reviewed, written, or edited these notes using EHR access logs. Each note is further characterized by three variables: the category of the note’s intent, the category of its content, and a label indicating whether it was created during an inpatient period. There are five intent categories, including *Orders* and *Patient Clinical Information*, and 32 content categories,

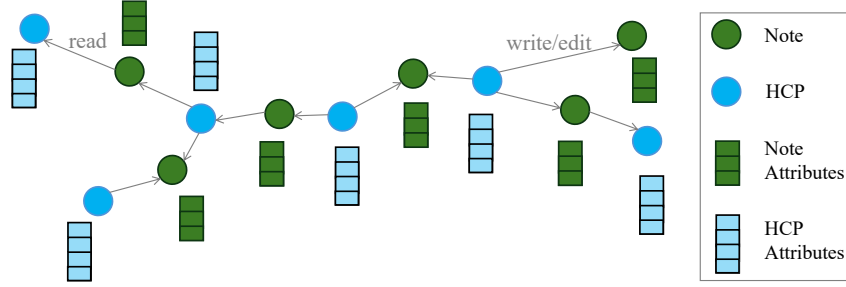


Figure 5.1: The illustration of the attributed bipartite collaboration network.

such as *Order Canceled* and *Note Signed*. For each HCP, I provide context into their role in the collaboration using four variables: title, type, specialty, and a label indicating whether they are a resident. There are seven titles (e.g., *MD* or *RN*), 12 types (e.g., *Physician Faculty* or *Physician Fellow*), and 71 specialties (e.g., *Cardiology* or *Dermatology*).

Bipartite network construction. After identifying all participants in the collaboration surrounding a patient (notes and HCPs), I define the information flow among these entities. This flow is represented as a directed bipartite network, where notes and HCPs serve as network nodes, and edges capture the reviewing and writing events recorded in the EHR system. For example, if HCP_A reviews $Note_B$, an edge is established from $Note_B$ to HCP_A , indicating that the information from $Note_B$ flows toward HCP_A . The bipartite nature of this collaboration network ensures that edges only form between a note and an HCP, but not between two notes or two HCPs. This reflects the fact that interactions among HCPs within the EHR are always mediated through notes rather than direct communication. Finally, as illustrated in Fig. 5.1, the variables extracted to characterize each note and HCP are assigned as node attributes.

Network snapshot sequence construction. To capture how collaboration evolves over time, I transform each patient’s time-stamped interactions into a sequence of graph snapshots. For every note–HCP edge (an edge between a note node and an HCP node), I collect the timestamps at which the interaction occurred in the access logs. Using the earliest interaction time t_0 for a given patient as a reference, I partition the timeline into fixed-length, non-overlapping one-week intervals of duration Δt and assign each interaction at time t to a bin $b = \lfloor (t - t_0) / \Delta t \rfloor$.

From these binned events, I construct a sequence of *cumulative* bipartite graphs. I

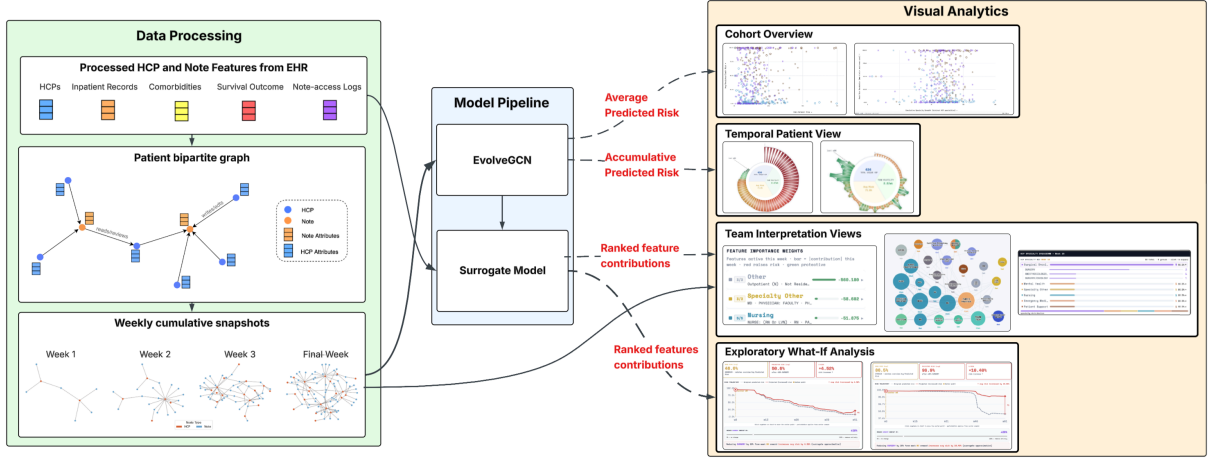


Figure 5.2: Overview of the analysis pipeline from curated EHR data to temporal graph modeling and coordinated visual analysis.

iterate over the bins in chronological order and, for each bin b_t , create a network snapshot G_t that includes all note–HCP edges that have appeared in any bin up to and including b_t (i.e., all edges in $b_0, \dots, b_t$). Consequently, the active care team grows over time: in early snapshots only a few HCPs and notes participate, and in later snapshots additional nodes become active as they first appear in the access logs. The edge set is cumulative, so once a HCP has appeared, their past interactions are retained in all subsequent snapshots.

For each snapshot G_t , I construct a node feature matrix $X_t \in \mathbb{R}^{N_t \times F}$, where N_t is the number of nodes and F is the feature dimension. Each node v has a feature vector

$$x_v^{(t)} = [1, \deg_t(v), a_v],$$

where the first entry is a fixed bias term (set to 1.0 for all nodes), the second entry is the degree of node v in G_t , and $a_v \in \mathbb{R}^{D_{\text{attr}}}$ is the attribute vector of node v . In this case, the feature dimension is $F = 2 + D_{\text{attr}}$, which remains fixed across snapshots, while the number of nodes N_t may grow over time as more HCPs and notes become active. These enriched node features serve as the input to the temporal graph neural network, and thus the per-node attribute matrix a_v is directly used as initial feature by the predictive model.

Fig. 5.2 summarizes the end-to-end pipeline, from curated EHR data and patient-specific bipartite graph construction to temporal risk modeling, surrogate-based attribution, and coordinated visual analysis.

5.3.2 Model Pipeline

5.3.2.1 Temporal Prediction with EvolveGCN.

To model how EHR-mediated collaboration evolves over time, I employ **EvolveGCN** [180], a temporal extension of graph convolutional networks designed for evolving graphs, over the sequence of weekly cumulative bipartite note-HCP snapshots. Among the two versions of EvolveGCN, I choose EvolveGCN-O because our setting involves a graph whose active node set changes over time: as treatment progresses, new notes and new HCPs enter the record, causing both the node set and edge set to grow. Unlike temporal graph models that rely on persistent node embeddings or are most natural for a fixed node set, EvolveGCN evolves the graph convolution parameters themselves across snapshots, making it well suited to this dynamic healthcare collaboration setting.

At a high level, EvolveGCN-O applies graph convolution within each snapshot while using a recurrent update to evolve the convolutional parameters over time. This allows the model to learn from changing collaboration structure without requiring every note or provider to be present across all weeks. Compared with static or aggregated graph approaches, this formulation preserves temporal ordering and accumulated collaboration history. This choice is particularly suitable for our setting because the underlying collaboration structure is not a fixed HCP-only network, but a patient-specific bipartite graph that evolves over time as new notes, providers, and interactions enter the record.

I train the EvolveGCN-O models for each cancer cohort and use their outputs as backend temporal risk estimates over the sequence of cumulative bipartite snapshots.

5.3.2.2 Surrogate-Based Feature Attribution.

Although EvolveGCN captures temporal and structural patterns in the evolving care-network, its predictions are not directly interpretable. To interpret the temporal predictions, I associate each weekly EvolveGCN risk estimate with a week-specific surrogate model defined on attribute summaries from the collaboration observed up to that week. This surrogate provides an interpretable approximation of the model’s weekly behavior and supports attribute-level explanation over time.

Given the surrogate coefficient β_j for feature j and observed cumulative value $x_{j,t}$

at week t , the feature contribution is computed as $c_{j,t} = \beta_j x_{j,t}$. Positive contributions indicate features associated with higher model-estimated risk under the surrogate; negative contributions indicate features associated with lower model-estimated risk. For each patient-week, I retain the top 80 features ranked by $|c_{j,t}|$ for downstream visualization.

Because the surrogate is trained to approximate the GNN’s output rather than the ground-truth binary label directly, its coefficients reflect patterns associated with the risk estimate produced by EvolveGCN and should be interpreted as an approximate linear explanation of GNN model behavior, not a direct decomposition of the original nonlinear model or direct evidence of clinically meaningful effect.

The surrogate additionally supports the system’s exploratory what-if analysis. When a user perturbs a selected feature, I estimate the corresponding directional change in projected risk through the surrogate logit:

$$\hat{p}_t = \sigma \left(\log \frac{p_t}{1 - p_t} - \beta_j \cdot \frac{\delta}{100} \cdot |x_{j,t}| \right) \quad (5.1)$$

where p_t is the original EvolveGCN risk at week t , δ is the user-specified perturbation percentage, and σ is the sigmoid function. The resulting projected trajectory is intended only for exploratory reasoning about potential directional sensitivity in the learned model representation.

5.3.3 Visual Analytics Design

5.3.3.1 Design Rationale

The output of our modeling pipeline is high-dimensional and multimodal. For all patients, the system includes a temporal risk trajectory derived from the EvolveGCN, a sequence of weekly cumulative bipartite note-HCP collaboration networks, per-week feature attributions with ~ 80 features each, and weekly HCP activity snapshots identifying which care team members accessed the patient’s notes during each time interval. Taken together, these outputs provide complementary perspectives on patient-specific care-team dynamics, yet they remain difficult to interpret without a coordinated analytical view.

The model’s raw risk estimates and feature attributions alone cannot address the range of questions that clinicians and health services researchers may bring to this data.

Rather than asking why a risk change occurred in a clinical interpretation, analysts may wish to explore what changes in provider participation, specialty composition, or weekly co-access structure coincide with shifts in the model-estimated risk trajectory. They may also wish to compare how patients with broadly similar cohort context differ in collaboration structure over time, or examine which care-team characteristics are highlighted by the surrogate explanation at a given point in the treatment timeline. These analytical needs are comparative, longitudinal, and interpretive in nature, and they require moving across multiple coordinated representations rather than inspecting any single model output alone.

These goals motivate a human-in-the-loop visual analytics approach in which computation provides temporal modeling and feature-level summaries, while domain experts contribute contextual knowledge, comparison strategies, and interpretive judgment. I design a coordinated multi-view system in which each view supports a specific analytical task while remaining tightly linked to the others. This design allows analysts to move from cohort-level overview to patient-level temporal inspection, structural comparison, and exploratory what-if analysis, enabling them to investigate patterns and relationships that may not be readily apparent from EHR access logs alone. To support this, the interface is organized around six coordinated views, each targeting a distinct analytical task within the workflow described in the following section.

5.3.3.2 Workflow Overview

These phases correspond to the visual analytics workflow summarized in Fig. 5.2.

Phase 1: Cohort Overview and Patient Selection. The analysis begins with the Cohort Scatterplot, which displays all patients by care-network size on the x-axis and average model-estimated risk on the y-axis, with color indicating cancer cohort. By filtering on cohort and survival status, analysts can identify patients or subgroups of interest, inspect outliers or comparison candidates, and select individual patients for more detailed exploration.

Phase 2: Temporal Patient View. After a patient is selected, the system loads the patient’s temporal risk trajectory into the Radial Glyph. Analysts can inspect the

trajectory in either probability mode or delta mode to identify periods of relatively high risk or notable week-to-week change. Selecting a week in the Radial Glyph establishes the temporal focus for downstream analysis and synchronizes the linked interpretation views to the same time point.

Phase 3: Team Interpretation Views. Once a week is selected, the Surrogate Attribution, HCP Distribution, and Ego Network views provide complementary interpretations of that temporal snapshot. Together, these views allow analysts to examine which care-team characteristics are most emphasized by the surrogate explanation, which specialties are active, and how the co-access structure is organized at that point in the treatment timeline. Moving across weeks supports comparative inspection of how feature-level explanations, specialty composition, and collaboration structure evolve over time. The interface also supports synchronized side-by-side comparison of two patients using the same coordinated view layout, enabling direct comparison of risk trajectory, surrogate attributions, specialty composition, and co-access structure at the same relative time point.

Phase 4: Exploratory What-If Analysis. Finally, analysts may select a feature of interest from the Surrogate Attribution Panel and adjust a perturbation slider to explore how the surrogate model responds to a hypothetical change in that feature. The system then displays the original and surrogate-projected risk trajectories together, allowing analysts to inspect possible directional differences in the model-estimated risk curve. This view is intended to support exploratory reasoning and hypothesis generation rather than rigorous causality or clinical recommendation.

5.3.3.3 View Design

Fig. 5.3 presents a composed overview of the coordinated views used for cohort-level selection, temporal risk inspection, feature attribution, specialty composition analysis, and collaboration-structure exploration. I describe each view and its analytical purpose below.

Cohort Scatterplot (Phase 1; panel A of Fig. 5.3). This view serves as the entry point for cohort-level exploration. Each point represents one patient. The x-axis

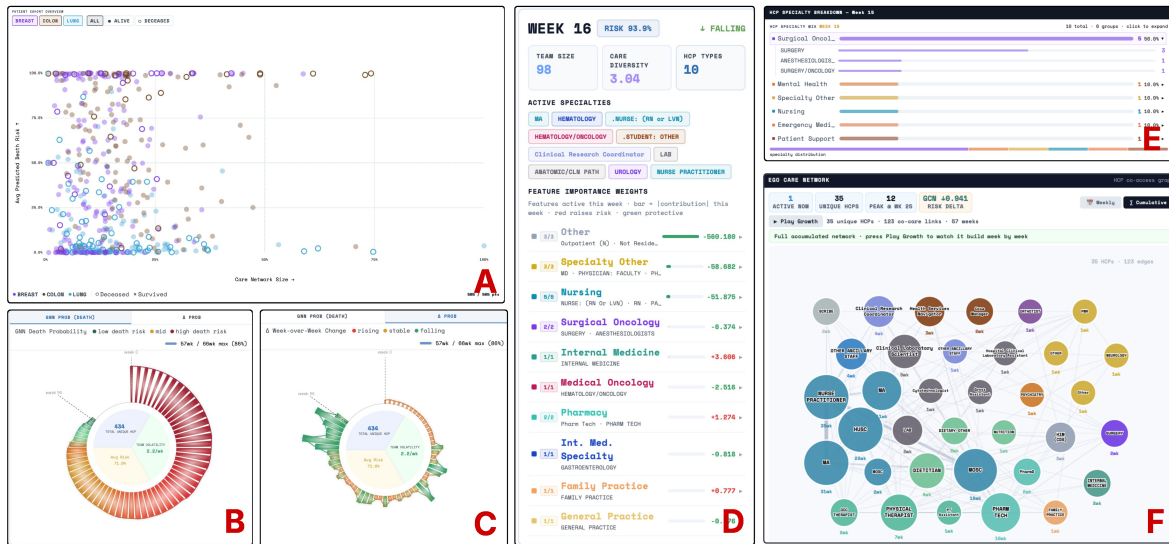


Figure 5.3: Composed overview of the coordinated views in the visual analytics system, including cohort-level selection (A), temporal risk inspection (B and C), feature attribution (D), specialty composition analysis (E), and collaboration-structure exploration (F).

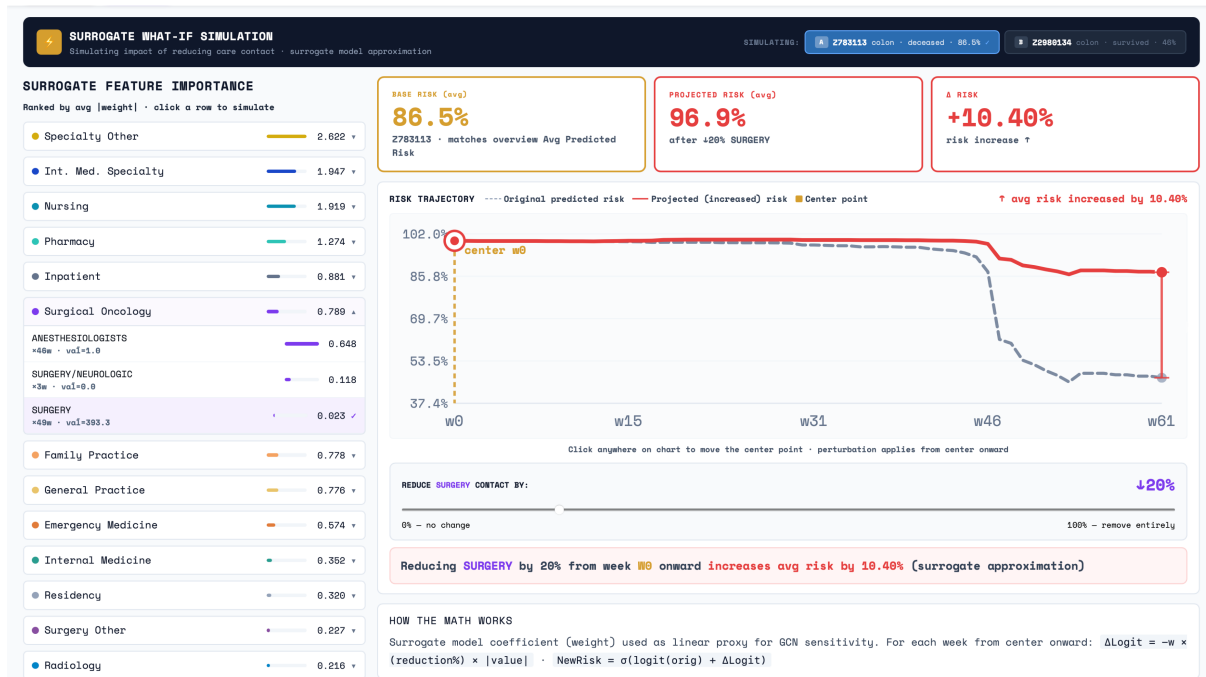


Figure 5.4: Surrogate-based what-if analysis of surgery-related involvement. Reducing surgery-related activity by 20% from week 0 shifts the projected average risk upward under the surrogate approximation.

encodes care-network size, defined as the total number of unique HCPs who accessed the patient's notes and normalized as a percentage of the cohort maximum. The y-axis encodes the patient's average model-estimated risk across all weekly snapshots. Point color

indicates cancer type (purple for breast, brown for colorectal, and blue for lung), while fill style indicates survival outcome (filled circles for surviving patients and hollow circles for deceased patients). This view supports tasks such as identifying outliers, locating patients with similar care-network sizes but different risk profiles, and selecting patients for downstream patient-level analysis and comparison.

In addition to the default cohort projection, the overview scatterplot can be reconfigured using alternative patient-level summaries. One such option maps patients by cumulative specialty breadth on the x-axis, defined as the number of distinct HCP specialties involved across the care trajectory, and by peak risk escalation on the y-axis, defined as the maximum positive week-to-week change in model-estimated risk. This alternative arrangement provides a complementary cohort-level entry point for identifying patients whose risk changes sharply despite narrower or broader care-team composition, supporting flexible case selection for subsequent longitudinal analysis.

Radial Glyph (Phase 2; panels B and C of Fig. 5.3). This view encodes the selected patient’s temporal model-estimated risk trajectory as a radial bar chart centered on a summary disc, where each bar represents one weekly snapshot. A thin overlaid curve traces the outer bar profile to provide a continuous visual guide to the overall trajectory shape across weeks. The glyph supports two toggleable modes to emphasize different aspects of temporal change. In **PROBABILITY** mode, bar height reflects the model-estimated risk for that week, and color follows a low-to-high risk gradient from green to red to provide an immediate sense of absolute risk over time. In **DELTA** mode, bar height reflects the magnitude of week-to-week change in risk, while color indicates direction, distinguishing increasing, relatively stable, and decreasing risk. Rather than always forming a full circle, the glyph spans an angular range proportional to the patient’s observed timeline, allowing follow-up duration to remain visible while preserving week-level detail. The center disc summarizes patient-level statistics across the full trajectory. Its purpose is to help analysts identify salient temporal patterns, including sustained high-risk periods or sharp weekly changes. Selecting any bar sets the temporal focus and synchronizes the linked interpretation views to the same week.

Surrogate Attribution Panel (Phase 3; panel D of Fig. 5.3). This view displays the surrogate linear model’s per-feature contributions at the selected week, where each feature reflects cumulative collaboration attributes observed from the start of treatment through that time point. Each displayed feature merges two underlying components, a `::freq` count capturing how often that activity has occurred and a `::present` indicator capturing whether it has appeared at all. Features are shown as horizontal bars sorted by absolute contribution magnitude, with bar length indicating contribution strength and color indicating direction, where red denotes a risk-increasing contribution and green denotes a risk-decreasing contribution. To support interpretation, the panel also shows the merged cumulative feature value and learned surrogate weight beneath each active feature bar, helping analysts distinguish between features that stand out because they occur frequently and those that reflect stronger model sensitivity.

HCP Distribution Chart (Phase 3; panel E of Fig. 5.3). This view presents a grouped horizontal bar chart showing the distribution of HCPs across a two-level specialty taxonomy, including 22 top-level groups and approximately 150 fine-grained specialties. The top-level categories were defined in collaboration with clinical domain experts to organize the full set of EHR-recorded provider specialties into clinically meaningful groups such as Medical Oncology, Surgical Oncology, Nursing, Pharmacy, and Internal Medicine Specialties. The chart can display specialty composition for the selected week or in aggregate across all weeks, allowing analysts to examine how the functional makeup of the care team changes over time and differs across patients or treatment stages. Hovering over a top-level group reveals the corresponding fine-grained specialty breakdown for more detailed inspection.

Ego Network (Phase 3; panel F of Fig. 5.3). This view visualizes an HCP-HCP co-access network derived from the patient’s bipartite note-HCP graph, where two providers are connected when they both accessed the same patient note, making the note the implicit intermediary of the relationship. Its purpose is to help analysts interpret the structural organization of the care team beyond feature rankings and specialty counts by showing which providers appear connected through shared note activity and how those

relationships evolve over time. The view supports two complementary modes. In **WEEKLY** mode, it shows the co-access structure for a selected week, allowing analysts to inspect whether provider interactions appear sparse, fragmented, or more tightly connected at that point in the care timeline. A scrubber spans the full observation window and enables week-by-week navigation, with blue markers indicating weeks in which HCP-HCP co-access connections were observed and gray markers indicating weeks with provider activity but no co-access ties. For weeks in which providers accessed notes without forming shared-note connections, the view falls back to a simplified layout that places active HCPs around a central patient node, preserving evidence of provider activity even when no co-access network is present. In **CUMULATIVE** mode, the view aggregates all weekly snapshots into a single network to summarize longer-term collaboration patterns; node size encodes how many weeks a provider was active, edge thickness encodes how often a provider pair co-accessed notes across weeks, and node color follows the specialty grouping used in the HCP distribution chart, including split-color arcs for providers associated with multiple groups. Key interactions support both temporal and structural analysis: the scrubber and play control animate changes over time, hovering reveals specialty details, and selecting a node highlights its immediate neighborhood while dimming unrelated parts of the graph.

What-If Panel (Phase 4; Fig. 5.4). This view supports exploratory what-if analysis through surrogate-based feature perturbation. The analyst selects a feature from the Surrogate Attribution Panel and adjusts a perturbation slider to simulate a 0-100% reduction in that feature. The system then computes a surrogate-projected risk trajectory using Eq. 5.1 and overlays it on the original model-estimated trajectory, allowing analysts to compare possible directional changes under the selected perturbation. Its purpose is to support investigation of how highlighted care-team characteristics may relate to temporal variation in the model output. Because the perturbed trajectory is derived from the linear surrogate approximation rather than from re-running the EvolveGCN model itself, it should not be interpreted as a clinical prediction, treatment recommendation, or definitive clinical conclusion. Instead, it serves as a hypothesis-generation tool for exploring model sensitivity in an interpretable feature space.

Table 5.2: Performance of TGNN for each cancer cohort.

Cohort	Accuracy	F1	Precision	Recall
Breast	0.867	0.927	0.884	0.974
Colorectal	0.767	0.863	0.815	0.917
Lung	0.821	0.848	0.875	0.824

5.4 Experiments

In this section, I demonstrate TGNN’s performance in predicting patient survival from EHR-mediated collaboration networks across different cancer types. For interpretability, following the procedure described in Sec. 5.3.2.2, I train cohort-specific surrogate models to obtain attribute-level explanations. I further compute Shapley values [25] to validate and complement the explanations provided by the surrogate model. Together, these components enable us to identify which collaboration factors are most strongly associated with patient survival.

5.4.1 Survival Prediction Performance

First, I measure the predictive performance by F1 scores with complementary metrics. Table 5.2 reports F1-score, accuracy, precision, and recall on the held-out test sets. Across all three cohorts, the trained TGNN achieves non-trivial F1 scores and accuracies, indicating that survival outcomes can be meaningfully predicted from EHR-mediated collaboration patterns. Overall, the classifier effectively distinguishes networks belonging to alive and deceased patients.

5.4.2 Survival Prediction Interpretation

Second, I train attribute-based surrogate models and evaluate how well their predictive behavior aligns with that of the TGNN. I assess surrogate performance on the held-out test set using the same metrics applied to the TGNN. As shown in Table 5.3, the surrogate models achieve predictive performance that is closely aligned with—and only slightly below—that of the TGNN across all three cohorts. This suggests that a substantial portion of the survival signal captured by the TGNN can be represented using HCP and

Table 5.3: Performance of the attribute-based surrogate model for each cancer cohort (test set).

Cohort	Accuracy	Precision	Recall	F1
Breast	0.962	0.773	1.000	0.872
Colorectal	1.000	1.000	1.000	1.000
Lung	0.932	0.919	0.919	0.919

note attributes, supporting the use of surrogate models as interpretable proxies.

5.4.3 Confounding Factors Correlation Analyses

To further ensure the effect of HCP collaboration on patient survival and minimize the influence of confounding factors, I examine four additional variables identified by our medical doctor collaborators as potentially correlated with patient survival: gender, cancer stage (Stage 2 or 3), age, and insurance type (private or public). I assess the alignment between each variable’s distribution and patient survival using both Spearman and Pearson correlation coefficients, along with their corresponding p -values. As shown in Table 5.4, the results from both correlation measures are highly consistent, providing nearly identical indications of the strength of association between these variables and patient survival.

Additionally, the highest correlation is observed between cancer stage and patient survival within the lung cancer dataset. However, even this highest correlation remains below 0.3, indicating little to no meaningful association between the two variables. Furthermore, the relatively high p -values (some exceeding 0.6) suggest that any weak correlation detected is likely due to chance rather than a true relationship. Thus, the likelihood of these four variables being primary drivers of patient survival is minimal. Based on this, I find evidence that the association between HCP collaboration and patient survival is not readily explained by these examined factors.

5.5 Case Studies

I illustrate the analytical workflow through two case studies drawn from the colorectal and breast cancer cohorts. These cases show how an analyst can use coordinated views to investigate specific questions about how model-estimated risk relates to patient-specific

Table 5.4: The correlation analyses examine four additional confounding factors that might be associated with patient survival. The results indicate that these factors exhibit little to no correlation with patient survival, as evidenced by correlation coefficients lower than 0.3. Therefore, the likelihood of these factors being the primary drivers influencing patient survival is minimal.

Correlation between patient survival and	Female Breast Cancer Patients				Lung Cancer Patients				Colorectal Cancer Patients			
	Pearson		Spearman		Pearson		Spearman		Pearson		Spearman	
	corr.	p-value	corr.	p-value	corr.	p-value	corr.	p-value	corr.	p-value	corr.	p-value
gender	-	-	-	-	0.1554	0.0576	0.1554	0.0576	0.0850	0.3128	0.0850	0.3128
stage	0.0858	0.2156	0.0858	0.2156	0.2996	0.0002	0.2996	0.0002	0.1225	0.1450	0.1225	0.1450
age	0.2607	0.0001	0.2574	0.0002	0.1018	0.2149	0.0975	0.2353	0.2401	0.0039	0.2350	0.0047
payer	0.2522	0.0002	0.2529	0.0002	-0.0449	0.5853	0.0426	0.6051	0.2293	0.0059	0.2129	0.0107

collaboration patterns over time. In each case, the interpretation is developed by linking evidence across different views. The resulting observations are intended to support insight generation and hypothesis development, and should therefore be understood as exploratory patterns surfaced through the visual analytics system rather than as prescriptive clinical conclusions.

5.5.1 Similar Care-Network Size, Divergent Risk

I begin with two colorectal cancer patients selected from the cohort scatterplot, where the x-axis encodes care-network size and the y-axis encodes average model-estimated risk across the full timeline. They were selected as an exploratory comparison pair because they have similar care-network size but substantially different average model-estimated risk (86.5% versus 46.0%), as shown in the cohort overview, allowing us to examine how patients with comparable network scale can nevertheless differ in temporal risk patterns and collaboration structure. The recorded outcomes also differ: the labeled deceased patient is predicted to be at a higher risk, whereas the labeled surviving patient is predicted to have a lower risk.

Fig. 5.5 shows the radial glyphs of the chosen patient pair under PROBABILITY mode. The higher-risk patient’s trajectory remains largely red throughout the timeline, with only a modest shift toward orange near the end, suggesting persistently elevated model-estimated risk. In contrast, the lower-risk patient’s trajectory gradually transitions from red toward green, suggesting a steady improvement over the course of treatment. The center summary shows that both patients have the same total number of unique HCPs

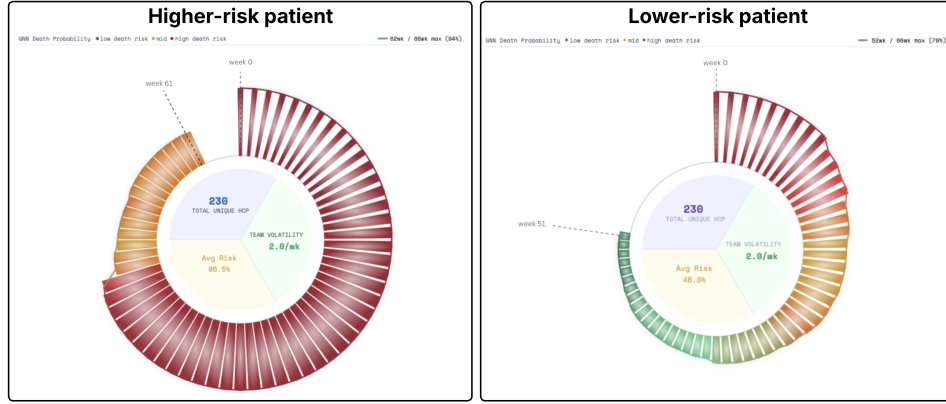


Figure 5.5: GNN Probability-mode radial comparison for the selected patient pair.

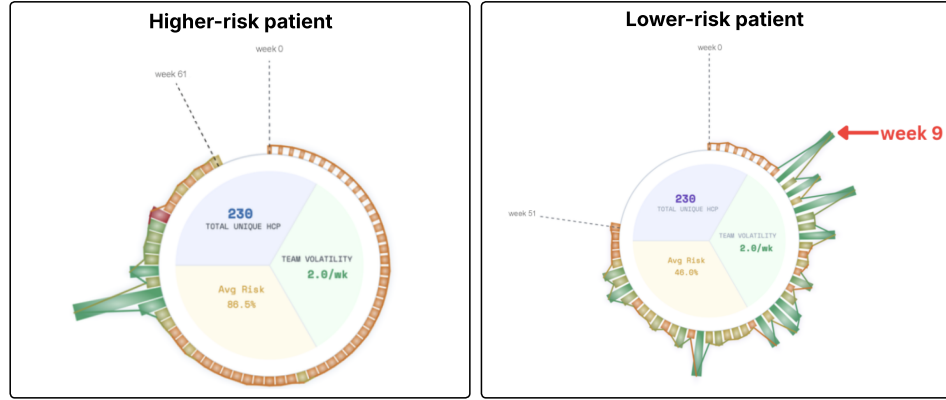


Figure 5.6: Delta-mode radial comparison for the selected patient pair. The higher-risk patient (left) shows persistent orange spikes with little green activity, while the lower-risk patient (right) shows stronger downward risk shifts, particularly around week 9.

(230) and the same average team volatility (2.0/week), suggesting that the difference in model-estimated risk is not explained solely by aggregated statistics of team composition.

Fig. 5.6 shows the pair in DELTA mode. The higher-risk patient is dominated by small orange spikes with relatively little green activity, whereas the lower-risk patient shows multiple green spikes throughout the entire treatment course, indicating gradual week-to-week declines in estimated risk over time. I focus on week 9 to investigate what contributed to this earliest pronounced drop in the lower-risk patient's estimated-risk trajectory.

In the week 9 Surrogate Attribution, the lower-risk patient exhibits a broader set of prominent risk-lowering features, including *Nursing*, *Internal Medicine*, *Internal Medicine Specialty*, and *Radiology*. In contrast, the higher-risk patient shows a narrower feature

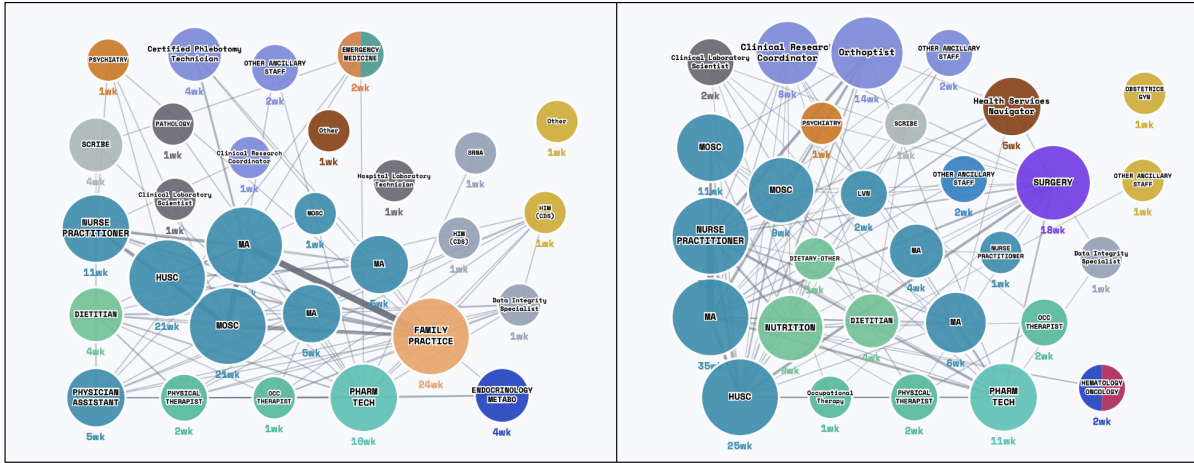


Figure 5.7: Cumulative ego-network comparison. The lower-risk patient (right) shows a denser and more structurally diverse collaboration network than the higher-risk patient (left).

profile and includes a positive contribution from *Family Practice*. Complementing these observations, the summary of *active* HCPs—defined as those with at least one recorded interaction in week 9—indicates that the higher-risk patient has a smaller and less diverse care team (22 providers across 2 HCP types), whereas the lower-risk patient involves 36 providers spanning 11 HCP types. Together, these differences suggest that the lower-risk patient’s improvement in the week 9 risk estimate is associated with a more diverse and actively engaged set of HCPs aligned with lower-risk collaboration patterns. Conversely, the higher-risk patient’s collaboration at week 9 involves fewer and less diverse HCP roles linked to favorable outcomes, which may help explain the absence of a corresponding decrease in predicted risk.

The cumulative ego network further reveals how actively participating HCPs are connected. As shown in Fig. 5.7, the key difference lies in the structure of these connections: the higher-risk patient has 83 co-access edges and a visually sparser network, whereas the lower-risk patient has 99 co-access edges and a more interconnected structure. Nursing remains the dominant specialty group for both patients; however, surgical oncology and medical oncology are more tightly integrated in the lower-risk patient’s network. In contrast, surgery-related specialties appear less integrated in the higher-risk patient’s collaboration structure. Together, these patterns suggest that the contrast between the two patients is not only in the presence of specific roles, but in how oncology- and surgery-

related providers are positioned within the broader cumulative co-access network over time.

With surgery-related provider involvement identified as a contributing factor in lowering predicted risk, I conduct a what-if analysis by applying a 20% reduction in surgery-related activity starting from week 0 as a perturbation. As shown in Fig. 5.4, the average predicted risk for the deceased patient increases by approximately 10.4%, rising from 86.5% to 96.9%. Applying the same perturbation to the surviving patient also leads to an increase in projected risk, from 46.0% to 50.6% (+4.60%). Despite differences in outcome and overall risk trajectories, both patients exhibit a consistent pattern: reducing surgery-related involvement is associated with an increase in predicted risk. Taken together, this case study demonstrates how our system links comparative structural observations with exploratory what-if analysis to support hypothesis generation for further investigation in broader clinical contexts.

5.5.2 Similar Risk, Distinct Care-Team Evolution

Whereas the first case study examined what factors may have contributed to diverging model-estimated risk, this case study examines the temporal phases of care-team evolution of two breast cancer patients who reached similarly elevated final estimated risk.

Fig. 5.8 shows the radial glyph comparison between the two patients in PROBABILITY and DELTA modes. Fig. 5.9 makes the corresponding phase interpretation explicit. Here, I define a *phase* as a consecutive stretch of weeks during which the estimated risk follows a similar pattern. A new phase begins when the trajectory shifts direction or moves into a distinctly different pattern. Under this definition, the left patient shows four phases: an early moderate-risk period, a sharp escalation beginning at week 21 ($\Delta = 0.3302$, reaching 65.7%), a subsequent decline across weeks 29–34, and another rise in estimated risk that reaches a maximum of 98.8% at week 51. The right patient shows three phases: a prolonged low-risk period between 1.6% and 2.0%, a gradual rise in risk, and a late escalation marked by the largest upward shifts at week 33 ($\Delta = 0.3031$, prob = 52.7%) and week 36 ($\Delta = 0.2357$, prob = 76.1%), eventually reaching 98.9% at week 41.

To investigate factors underlying the **left** patient’s early risk escalation, I examine

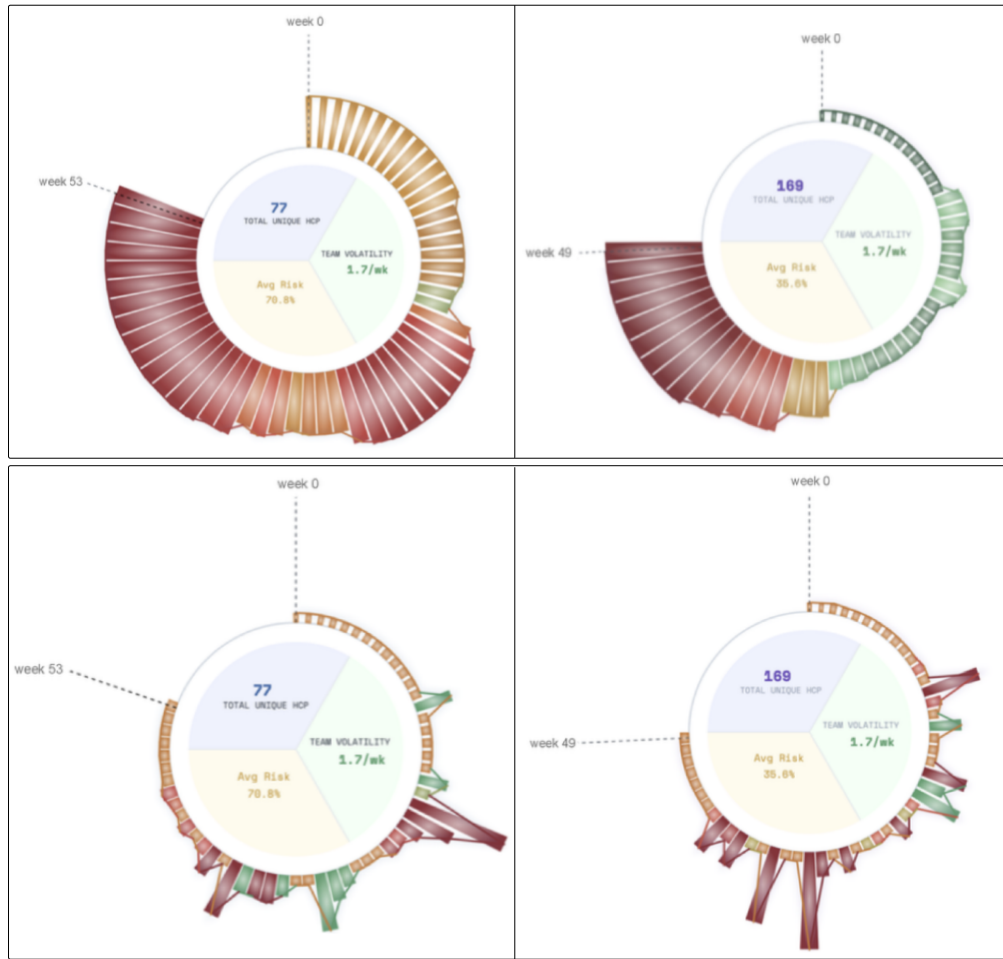


Figure 5.8: Temporal radial comparison of the selected breast cancer patient pair in PROBABILITY mode (top row) and DELTA mode (bottom row). The two modes support comparison of overall model-estimated risk patterns and major week-to-week shifts across the treatment timeline.

the active specialists during the initial weeks of phase 2 (Fig. 5.9). At week 20, only a single Gastroenterology provider is active. By week 21, the care team expands to include Family Practice, Gastroenterology, and Hematology/Oncology. At week 22, an Anatomic Pathology team comprising 13 providers is introduced, followed by additional involvement from Psychiatry and Surgery at week 23.

A similar pattern is observed for the **right** patient during the risk-increasing period (phase 2), where the care team undergoes a rapid shift in composition. At week 33, the team centers on Hematology/Oncology, Occupational Therapy, Physical Therapy, and Surgery; by week 36, it shifts to Anatomic Pathology, Hematology/Oncology, Pharmacy, and Radiation Oncology. Across both patients, these observations indicate expansion and

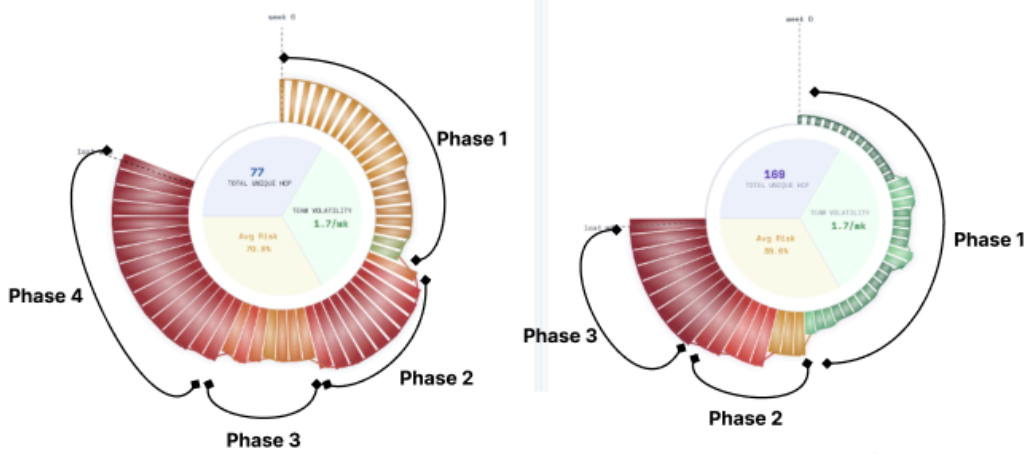


Figure 5.9: Phase-annotated temporal radial comparison of the selected breast cancer patient pair. Under the phase definition used in the text, the left patient shows a four-phase trajectory, whereas the right patient shows a three-phase trajectory with a delayed late-stage escalation. The labeled phase boundaries mark the phase-transition intervals used for the subsequent specialty, ego-network, and surrogate analyses.

reconfiguration of the care team during short periods that elevate model-estimated risk.

To better understand why both patients reach similarly elevated risk despite exhibiting different temporal patterns of care-team evolution, I compare the corresponding weekly ego-network and surrogate views (Fig. 5.10). For the left patient (top row), the early upward transition is accompanied by a rapid buildup of local co-access structure, evolving from no observable weekly co-access pattern at week 20 to more developed sub-networks by weeks 22 and 23. In contrast, as shown in the bottom row, the right patient’s later upward transition appears more gradual at the weekly level, with interactions incrementally and sparsely introduced into the collaboration—showing limited structure at week 33, no observable co-access network during weeks 34–35, and only a small co-access structure at week 36. Despite these differing temporal patterns, both patients exhibit a substantial increase in care-team involvement within a short time window.

The surrogate view provides additional feature-level context for these transitions. Although both patients are characterized by broadly similar risk-raising cumulative care-activity signals, the right patient retains *Surgery::freq* as a visible risk-lowering contributor, whereas no comparable surgery-related buffering signal is evident during the left patient’s early shift. Taken together, these views suggest that, despite differences in how collaboration evolves—either through rapid restructuring or more incrementally intro-

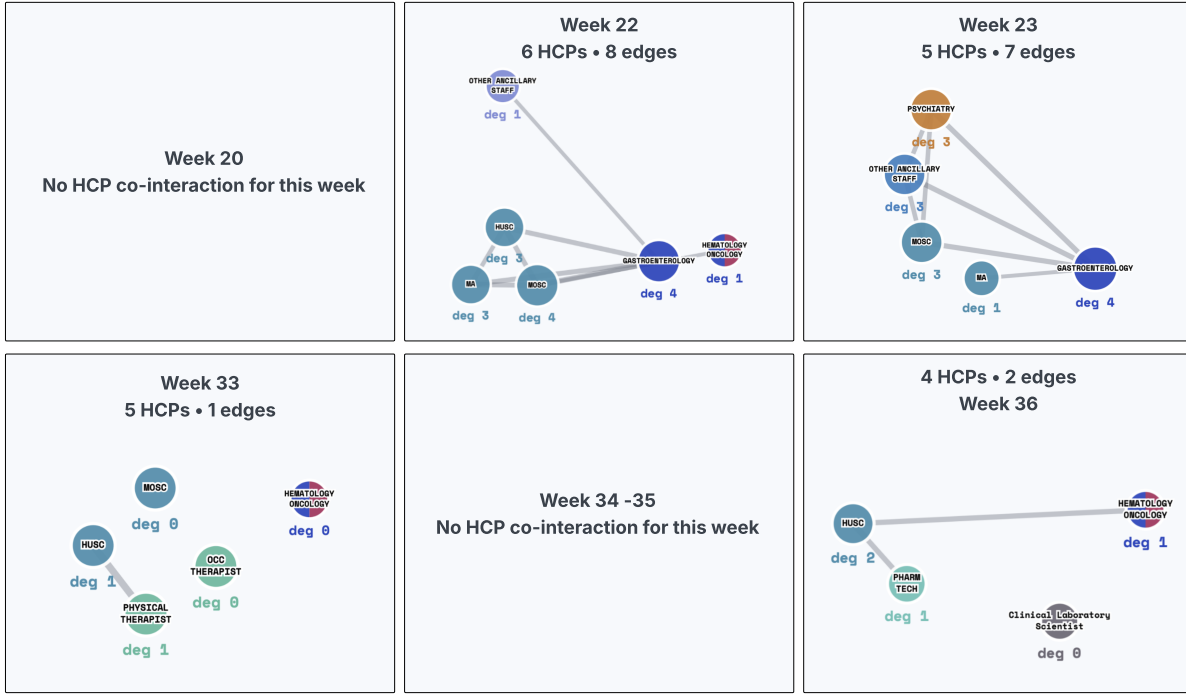


Figure 5.10: Representative weekly ego-network snapshots across transition intervals for two breast cancer patients. The top-row snapshot sequence shows the left patient at weeks 20, 22, and 23, where the co-access structure evolves from no observable HCP co-interactions at week 20 to more developed sub-networks of multiple care teams by weeks 22 and 23. The bottom-row snapshot sequence shows the right patient at weeks 33, 34–35, and 36, where interactions are incrementally and sparsely introduced into the collaboration over time, with limited co-access at weeks 33 and 36 and no observable HCP co-interactions during weeks 34–35.

duced interactions—both patients experience a sharp short-term expansion in care-team size that coincides with elevated predicted risk. This pattern suggests that rapid increases in care-team involvement over a short period, regardless of whether they appear abrupt or gradual at the weekly level, may reflect escalating clinical complexity associated with higher risk. These observations are not intended as clinical conclusions, but rather as hypothesis-generating insights that warrant further investigation within clinical contexts.

In sum, this case study highlights how similar risk outcomes can arise from different temporal patterns of care-team evolution, illustrating the value of comparative analysis for supporting hypothesis generation.

5.6 Evaluation

In this section, I evaluate (1) the robustness of our surrogate model, (2) the expert assessment of our visual analytics system, and (3) the validity if the identified important dynamic collaboration traits.

Table 5.5: Statistics of the attribute “Surgery”: logistic regression (LR) coefficient, LR coefficient rank among all attributes, mean absolute SHAP value, and SHAP value rank among all attributes.

Cohort	LR weight	LR rank	Mean SHAP	SHAP rank
Breast	0.0043	1	0.68	1
Lung	0.0022	2	0.46	1
Colorectal	0.0007	3	0.11	5

5.6.1 Robustness Evaluation of Surrogate Model

To evaluate the robustness of our surrogate models, I re-train each surrogate model using ten random seed. Across the ten runs, the rank of attributes per cohort derived from the learned weights are the same, indicating their relative importance for a cohort is robust to training dynamics and the consistent rank is stable and not affected by the stochastic nature of machine learning approaches. Table 5.5 reports the importance scores (i.e., learned weights and Shapley values) and the respective rank of the attribute “Surgery” among all attributes averaged across ten runs, indicating its consistent and high influence on the patient survival.

5.6.2 Expert Assessment of the Tool and Analytic Framing

I evaluated the system through a qualitative expert review designed to assess whether the coordinated views support plausible interpretations of temporal care-team dynamics and to identify constraints on how such analyses should be presented. The review was conducted as a live demonstration with an interdisciplinary group of collaborators from a multi-site EHR research consortium. The discussion was grounded in the preliminary first case study and the associated system walkthrough used in the demo. The second case study (Sec. 5.5.2) serves as an additional analytic example and was not part of the expert review session. The participants included clinicians, medical informatics researchers, and biostatisticians. The session included a walkthrough of the comparative and longitudinal analysis workflow, followed by discussion of interpretability, clinical plausibility, and limitations of the current framing.

For clarity and anonymity, I summarize the discussion using generalized labels (e.g.,

Expert A) rather than verbatim quotations or named attributions. The expert discussion suggested that the system is useful for moving from high-level patient comparison to temporally focused inspection of care-team organization. In particular, the coordinated workflow linking cohort selection, temporal risk trajectories, week-specific explanation, specialty composition, and collaboration structure was viewed as a promising way to examine how model-estimated risk relates to evolving patterns of EHR-mediated care activity. The experts described the visualization as informative and promising for examining care-coordination patterns through routinely collected EHR activity.

At the same time, the review emphasized several important interpretive constraints. First, Expert A noted that comparative patient analysis would be more convincing when the selected pairs are more tightly matched on major clinical characteristics such as disease stage, age, treatment pathway, and time to first treatment. Strengthening this matching would reduce alternative explanations and make side-by-side interpretation more clinically meaningful.

Second, Expert B highlighted the importance of broadening the analytic framing beyond survival alone. Because patient outcomes reflect many interacting clinical factors, several experts suggested that additional endpoints could make the system more useful for exploring how care-team dynamics relate to different aspects of the care trajectory. This feedback does not weaken the current study, but clarifies an important direction for future extension.

The review also surfaced particular aspects of the system that the experts found especially promising. Among these, the team-volatility metric drew strong interest as a potentially informative summary of changing care-team organization over time. The discussion suggested that this measure may help reveal instability or repeated shifts in coordination that would be difficult to recognize directly from raw access-log records. In this sense, the visual analytics system appeared valuable not only for supporting the presented comparisons, but also for revealing secondary signals that domain experts found intriguing and wanted to investigate more deeply in future analyses.

In addition, the discussion raised an important broader consideration regarding in-

terpretation of EHR interaction traces. Expert C noted that evolving documentation practices, including AI-assisted documentation and changes in provider search behavior, may affect how access-log activity should be interpreted over time. Rather than undermining the current approach, this observation highlights the importance of treating access patterns as situated traces of care activity whose meaning may shift with clinical infrastructure. Because the system is built directly on longitudinal EHR interaction data, it may also provide a useful framework for examining how such infrastructure changes become reflected in care-team activity patterns over longer periods and across settings.

5.6.3 Clinical Grounding of the Surfaced Pattern

The surgery-related pattern surfaced in the case studies should not be interpreted as clinical proof. However, prior colorectal cancer literature provides background showing that this pattern is clinically plausible and therefore reasonable to examine within an exploratory visual analytics workflow.

First, surgery occupies a central role in colon cancer treatment. The National Cancer Institute’s PDQ states that colon cancer is often curable when localized and that surgery is the primary form of treatment [192]. An ASCO guideline likewise identifies surgery as the primary treatment for stage I–IIA colon cancer, and a clinical review of early-stage colon cancer describes surgery as the backbone of curative treatment, with surgery alone curative for many patients [193, 194].

Second, prior work suggests that outcomes are shaped not only by whether surgery occurs, but also by how surgery is coordinated within the broader treatment process. A meta-analysis of colorectal cancer multidisciplinary team care found an association with improved overall survival, supporting the broader importance of coordinated multi-specialty management [195].

Third, the timing of surgery-related care has also been linked to outcomes in the colorectal literature. A systematic review and meta-analysis reported that longer delay to colorectal surgery was associated with poorer survival, while studies of postoperative treatment timing found that delayed adjuvant chemotherapy after surgery was associated with worse overall survival in colorectal and stage III colon cancer [196–198].

Taken together, this literature does not validate the specific mechanism suggested by our case studies, nor does it confirm that the surgery-related signal surfaced by the system reflects a clinically meaningful effect. Rather, it indicates that surgery and surgery-centered coordination are established and outcome-relevant components of colorectal cancer care. It is worth noting that the studies cited above measure surgical care and team coordination directly, whereas our system captures surgery-related involvement through EHR access patterns, which serve as a proxy for participation rather than a direct record of clinical activity. Within the context of this chapter, these findings provide clinical background for why the surgery-related pattern identified by the system is reasonable to inspect, compare, and discuss with domain experts.

5.6.4 Summary

Overall, the expert review suggests that the system is useful for surfacing temporally varying patterns of care-team organization and for helping analysts connect model behavior to underlying collaboration structure through coordinated visual exploration. The discussion also identified clear next steps, including clearer operational definitions in the interface, more rigorous matching in comparative case studies, further study of team volatility, and continued attention to how evolving EHR practices may shape the interpretation of access-log-based collaboration patterns.

5.7 Discussion and Limitations

This chapter presents a visual analytics framework for exploring cancer care as a time-evolving collaboration process, enabling coordinated analysis of temporal risk predictions, feature-level explanations, and evolving care-team structures. By linking these components within a unified workflow, the system supports comparative and longitudinal reasoning about how differences in care-team organization and EHR-mediated communication patterns may be associated with patient risk trajectories over time, helping analysts identify temporally specific changes in care-team structure and communication that are difficult to observe from static summaries alone.

The patterns surfaced through this analysis should be interpreted with appropriate

caution. For example, the surgery-related differences highlighted in the case studies emerge consistently across temporal, compositional, and structural views, but they do not establish that specific care-team configurations directly account for differences in patient outcomes. More broadly, the system is designed to support hypothesis generation and the exploration of plausible clinical associations, rather than to provide prescriptive guidance. This framing supports closer examination of temporally evolving care-team patterns that may merit further clinical interpretation and investigation.

In the following paragraphs, I discuss several considerations that define the interpretive scope of the current study, including aspects related to data representation, study design, and model interpretation.

Data Representation and Observability. The current framework derives collaboration structures from EHR access logs, capturing digital traces of care activity and system-mediated interactions. While this provides a scalable and time-resolved view of care-team dynamics, it does not directly encode richer aspects of clinical communication, such as intent, decision-making context, or care quality. Future work could incorporate additional data sources or clinically grounded signals, including note content, to better contextualize observed collaboration patterns and enrich the interpretation of care-team interactions.

Study Design and Comparative Context. The case studies are designed to illustrate the analytical workflow rather than to establish controlled clinical comparisons. Incorporating stronger patient matching criteria—such as cancer stage, age, treatment pathway, and time to first treatment—would enable more rigorous comparison across patients and reduce confounding factors unrelated to care-team structure. Closer collaboration with domain experts will be essential in identifying and integrating clinically meaningful variables that can strengthen comparative analysis.

Model Behavior and Explanation Fidelity. The surrogate attribution and what-if views provide interpretable approximations of model behavior in a feature-based space, enabling analysts to explore potential associations between care-team activity and predicted risk. I separately evaluated the surrogate model for robustness and observed stable

rankings of influential attributes across retraining with different random seeds. I re-train each surrogate model using ten random seed. Across the ten runs, the rank of attributes per cohort derived from the learned weights are the same, indicating their relative importance for a cohort is robust to training dynamics and the consistent rank is stable and not affected by the stochastic nature of machine learning approaches. This suggests that the feature-level patterns surfaced in the interface are reasonably consistent, while future work may explore alternative explanation strategies or additional domain-informed features to further improve explanation fidelity.

Generalizability and Modularity. An important strength of the framework is its modularity. The current system combines a specific temporal graph model, surrogate explanation strategy, and set of coordinated views, but the analytical workflow is not tied to these choices. This creates opportunities to examine whether similar patterns remain consistent under alternative temporal graph backends, explanation methods, or feature constructions, supporting reuse of the workflow while varying computational components.

Future Work. Several directions follow naturally from this design. Applying the framework to additional institutional datasets with more diverse event information could provide richer temporal context for interpreting changes in care-team communication and risk trajectories, and enable cross-site comparison of collaboration-risk associations. Incorporating stronger patient matching variables, including cancer stage and treatment timing, would further improve the validity of comparative analysis.

On the modeling side, future work may examine whether alternative temporal graph architectures or explanation methods yield consistent or divergent care-team signals, helping distinguish robust findings from model-specific effects. For the explanatory views, further refinement could investigate how dominant feature groups influence surrogate interpretations, including whether re-specifying or constraining communication- or specialty-related features changes the patterns presented to analysts.

5.8 Conclusion

I present a visual analytics framework for investigating cancer care as a time-evolving collaboration process, leveraging EHR-derived communication traces to model how care-team structures evolve throughout treatment. By aligning temporal graph-based risk modeling with the same time-evolving collaboration networks used in visual analysis, the framework addresses the common mismatch between dynamic care processes and static predictive representations, enabling more coherent interpretation of how changes in team composition are associated with patient risk trajectories. Through coordinated visualizations of collaboration structure, specialty composition, and feature-level explanations, the system supports human–AI collaborative analysis across cohort-level patterns, patient-level comparisons, and longitudinal risk inspection. This integrated workflow enables clinicians to move between temporal risk dynamics and underlying care-team evolution, supporting comparative reasoning about how different collaboration patterns are associated with changes in predicted risk over time. The case studies and expert feedback demonstrate how this approach helps surface patterns that are difficult to observe from static or aggregated summaries alone, while enabling interpretable, practitioner-driven analysis of temporal model behavior. Rather than supporting definitive clinical conclusions, the framework is intended to support insight generation and hypothesis development, offering a practical foundation for human–AI analysis of complex, evolving care processes in oncology and a promising analytic workflow for future investigation of care-team dynamics in related healthcare settings.

Chapter 6

Discussion

This dissertation explores how machine learning (ML) and visual analytics (VA) can be integrated to support the reasoning of complex networks. Across multivariate, temporal, and heterogeneous network settings, the presented works collectively address a fundamental challenge in modern network analysis: although ML models can extract informative patterns from complex relational data, these patterns are often difficult to interpret, validate, and operationalize in practice. Through a series of ML-assisted visual analytics approaches, this dissertation demonstrates how interactive visual reasoning can bridge computational pattern extraction with human understanding and decision-making.

A central contribution of this dissertation is advancing *network characterization* for complex relational data. Chapter 2 contributes methods for reasoning multivariate networks by enabling analysts to uncover combinatorial influence among structural and semantic attributes. Rather than relying solely on predefined measures or latent embeddings, this work emphasizes flexible and interpretable characterization of relationships through composite variables and interactive exploration. Chapter 3 extends this direction into temporal networks by demonstrating how dynamic collaboration patterns among healthcare professionals can be examined in relation to patient outcomes. Together, these works highlight the importance of combining ML-driven feature extraction with human-guided visual reasoning to expose meaningful structural patterns and relational dynamics that may otherwise remain hidden in high-dimensional network data.

A second contribution lies in advancing *interpretable prediction* for graph learn-

ing models. Chapter 4 addresses the challenge of understanding Graph Neural Networks (GNNs), whose learned structural representations are often difficult to explain. By representing GNN explanations through elementary substructures and evaluating their relevance using factual and counterfactual reasoning, this work enables analysts to systematically examine how specific structural patterns contribute to graph-level predictions. Building upon this direction, Chapter 5 further integrates interpretable prediction into temporal graph learning for patient risk estimation, connecting predictive outcomes to collaboration and healthcare participation dynamics. These works collectively demonstrate that interpretability is not merely a supplementary feature of ML systems, but a necessary component for trustworthy reasoning on complex networks.

The dissertation further contributes to ***decision support*** through interactive analytical workflows that connect prediction, explanation, and intervention planning. In particular, Chapter 5 moves beyond retrospective analysis by enabling users to explore projected changes in patient risk through what-if simulation and targeted modifications of collaboration dynamics. This contribution illustrates how ML-assisted VA systems can support the full analytical progression from reasoning about historical patterns, to interpreting predictive outcomes, and ultimately to exploring future what-if scenarios and their potential implications. More broadly, these efforts demonstrate the potential of combining ML and VA to transform complex network analysis from passive pattern inspection into an interactive and human-centered decision-support process.

While these contributions advance ML-assisted VA for complex networks, several important challenges remain open. In terms of *network characterization*, scalability and representation complexity continue to pose difficulties as networks grow increasingly large, heterogeneous, and dynamic. Future research may explore adaptive abstraction, hierarchical representations, and interactive summarization techniques that preserve important structural semantics while maintaining analytical tractability.

For *interpretable prediction*, one major challenge is assessing the reliability and generalizability of explanation methods across different datasets, graph structures, and model architectures. Existing approaches, including surrogate-based, attribution-based, and

perturbation-based explanations, may exhibit inconsistent explanatory behaviors depending on the underlying data distribution and learning dynamics, making it difficult to determine whether extracted explanations consistently reflect meaningful structural reasoning beyond the settings evaluated in prior studies. Future directions may involve integrating causal inference, uncertainty quantification, and domain knowledge into explanation frameworks to improve explanation robustness and trustworthiness.

For *decision support*, future ML-assisted VA systems may further strengthen human-AI collaboration to better incorporate domain expertise, contextual reasoning, and evolving analytical goals into the decision-making process. Current workflows still largely separate automated computation from human reasoning, making it difficult for analysts to contextualize predictions, assess uncertainty, and identify potential failure cases during exploration. Future systems may therefore support more tightly coupled analytical workflows in which predictive modeling, explanation, uncertainty assessment, and what-if simulation are continuously integrated within interactive reasoning processes. Such analytical environments may enable analysts to more effectively probe model behavior, examine alternative interpretations, and make more informed decisions through comprehensive inspection of complex networked systems.

Beyond these technical challenges, the increasing deployment of ML-driven network analysis in real-world domains also raises broader concerns regarding transparency, fairness, and responsible use. As predictive models become more influential in high-impact settings such as healthcare and social systems, future research must consider how uncertainty, bias, and ethical implications can be effectively communicated within interactive analytical environments.

Taken together, these challenges and opportunities suggest that the future of ML-assisted visual analytics for complex networks lies in developing scalable and human-centered analytical frameworks with robust explanation mechanisms that tightly integrate predictive modeling, uncertainty assessment, and what-if reasoning.

Chapter 7

Conclusion

As network data becomes increasingly central to modern society, representing interactions in domains such as healthcare, social systems, and communication, a deeper challenge emerges: meaningful progress depends not only on powerful computational models but also on the ability to understand, reason about, and act upon complex relational patterns. While machine learning (ML) has demonstrated strong capability in extracting features and predictive signals from network data, the development of human-centered methods for interpreting these outputs and guiding decision-making remains comparatively limited. This dissertation addresses this gap by developing ML-assisted visual analytics (VA) methods for reasoning complex networks across three essential directions: network characterization, interpretable prediction, and decision support.

First, I develop VA and learning-assisted methods to improve understanding of complex network structures and relationships. For multivariate networks, I introduce an interactive framework that enables users to construct composite variables through learning-based optimization and visually examine how groups of attributes jointly relate to network characteristics. For temporal healthcare collaboration networks, I design a system for exploring evolving teamwork structures among healthcare professionals (HCPs), linking communication dynamics and network measures to patient survival outcomes. Together, these approaches expand network analysis beyond static or isolated measures by supporting richer reasoning about multivariate associations and temporal behaviors.

Second, I develop methods that combine graph learning with visual reasoning to im-

prove transparency and trust in predictive network models. To address the opacity of Graph Neural Networks (GNNs), I introduce a framework that uses elementary substructures, or graphlets, as interpretable building blocks for explanation. Through interactive visual analysis supported by factual and counterfactual metrics, users can examine how structural patterns influence model predictions and assess the significance of candidate explanations. This contribution helps bridge the gap between high-performing graph models and the interpretability required for reliable real-world deployment.

Finally, I integrate these foundations into an interactive framework for predictive analysis and what-if simulation to support intervention planning in healthcare systems. By combining temporal graph neural networks with visual analytics, the framework spans the full analytical pipeline, including data curation, mitigation of foreseeable confounding factors, patient morbidity risk prediction, explanation of model decisions, and simulation of targeted interventions. Users can explore how collaboration dynamics relate to estimated risk, identify influential participation patterns, and evaluate projected changes in outcomes under alternative scenarios. This integration transforms predictive modeling into an actionable decision-support process.

In conclusion, this dissertation advances the study of complex networks by developing and applying ML-assisted visual analytics methods for network understanding, interpretable prediction, and informed decision-making. By emphasizing transparency, interactivity, and domain-centered reasoning, this work demonstrates how machine intelligence and human expertise can be combined to transform complex relational data into actionable knowledge. As network data continues to grow in scale and complexity, the frameworks established here highlight the essential role of interpretable and user-centered analytics in ensuring that future network intelligence remains trustworthy, practical, and impactful.

Appendix A

Appendix for Chapter 2

A.1 Supplementary Comparison of Scatterplot Designs (Sec. 2.3.3)

Scatterplot enhancements. Before I decided to design two-class density scatterplots, I tested various designs, including state-of-the-art enhancements. In Sec. 2.3.3, I show differences between conventional density scatterplots and my two-class density scatterplots. Here, I compare conventional scatterplots, winglets, splatterplots, and two-class density scatterplots. In Fig. A.1, I show the results obtained by applying these designs to several preprocessed datasets used in this chapter.

With winglets and splatterplots, it is still difficult to quantify the density at each location. This is critical for visually grasping trends and correlations. For example, from the splatterplot of Fig. 2.2-b2, analysts might misjudge `Class0` has a higher density in general. Moreover, to effectively use winglets and splatterplots, there are various parameters I need to adjust. For example, winglets' appearance will be radically different based on the length of wings and the size of dots. On the other hand, splatterplots have several sensitive parameters, such as the threshold for contouring and the sampling rate for outliers.

Colormaps. Before deciding the final colormap design (Fig. A.2-d), I also tested various colormaps for two-class density scatterplots. For example, in Fig. A.2-a, I use Lespinats and Aupetit's bivariate colormap. This colormap uses hue to show the ratio of two

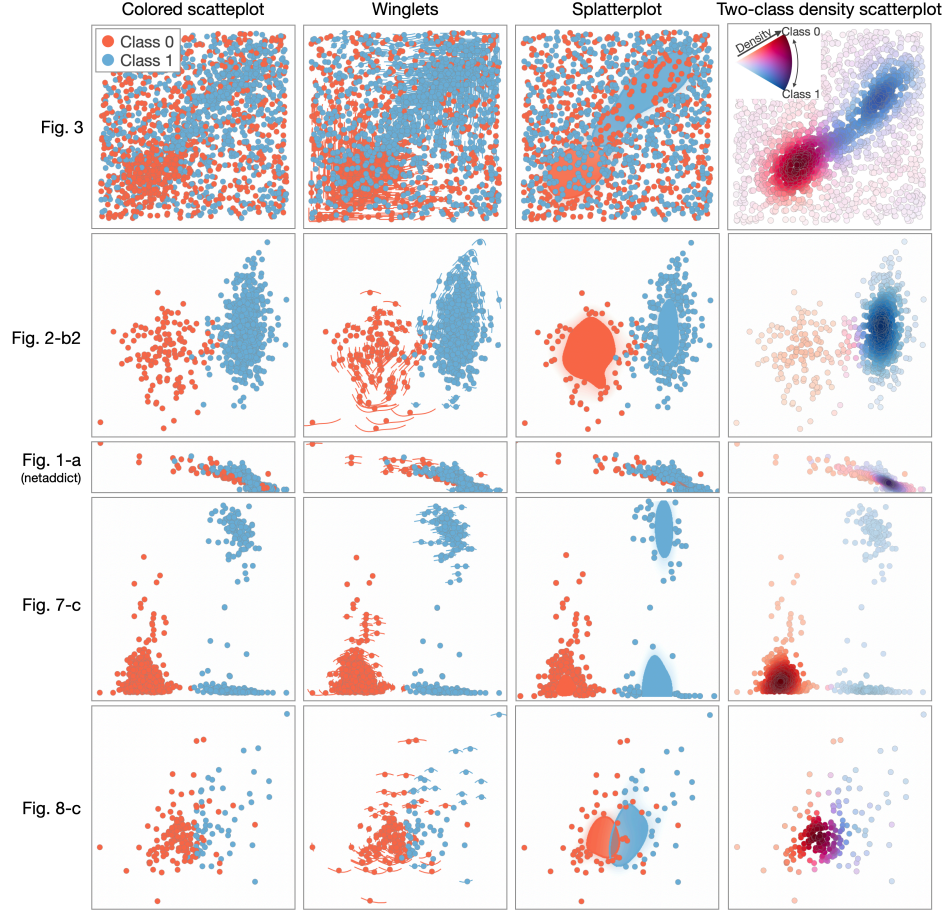


Figure A.1: Comparison of scatterplot designs.

attributes (i.e., the total density and the density ratio of two classes in this case) and lightness to represent the total value of the two attributes. From the visualization result, it is difficult to naturally grasp the two dense regions of **Class 0** and **Class 1** (located at the bottom left and the top right). As this issue is mainly caused by the inappropriate assignment of hue and lightness to the two attributes I use, the same problematic situation can be seen in Fig. A.2-b, where I use a different pair of hues (red and blue).

I can adjust Lespinats and Aupetit's bivariate colormap to assign hue and lightness to the total density and the density ratio, respectively. I can then show this colormap with a polar coordinate. As shown in Fig. A.2-c, this colormap can clearly convey each class's dense regions (dark purple and dark green regions). However, the colormap's hue is sensitive to which class has more density at each location. For example, in Fig. A.2-c, I observe a green region annotated by the arrow. This sensitivity is caused because the

purple is located on the exact opposite side of the green in the CIELAB color space (note: even in other color spaces such as HSV). In this case, I am less interested in binarizing regions based on which class has a higher density. Rather, I want to show the quantity of the density ratio. Thus, I decide to use red and blue as two distinct hues (Fig. A.2-d), with which purple color is produced as each class’s density becomes close to even. Based on this observation, I can create similar colormaps to Fig. A.2-d by avoiding selecting hues located on the opposite side in the color space. Fig. A.2-e shows such a colormap example, where orange and green are employed as two distinct hues (i.e., the mixed color becomes yellow). Both colormaps in Fig. A.2-d,e and similar ones are suitable for the two-class density scatterplot. My implementation uses the one in Fig. A.2-d by default.

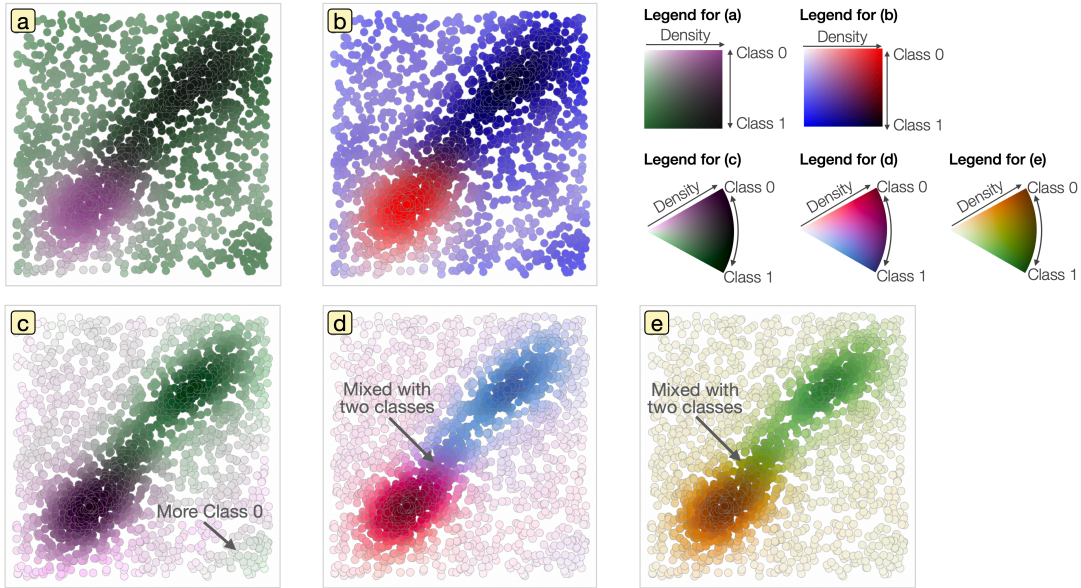


Figure A.2: Comparison of two-class density scatterplots with different colormaps.

A.2 Additional Statistical Analyses on Composite Variables

The workflow uses composite variables to help analysts interpret the learned network representations. Through the case studies, the effectiveness of the constructed composite variables is evaluated both qualitatively (e.g., visual patterns, expert feedback) and quantitatively (e.g., the improvement in correlation coefficients). Here, I further examine

the effectiveness of the composite variables in the context of conventional analyses.

Case 1. I first visually examine data distributions corresponding to the original attributes and composite variable, as shown in Fig. A.3. I can observe that the composite variable, $-0.4\text{open} + 0.7\text{consci} - 0.3\text{netaddict} + 0.3\text{min_happy} + 0.3\text{extra}$, creates more clearly different data distributions between **Class 0** and **Class 1** than any single attribute.

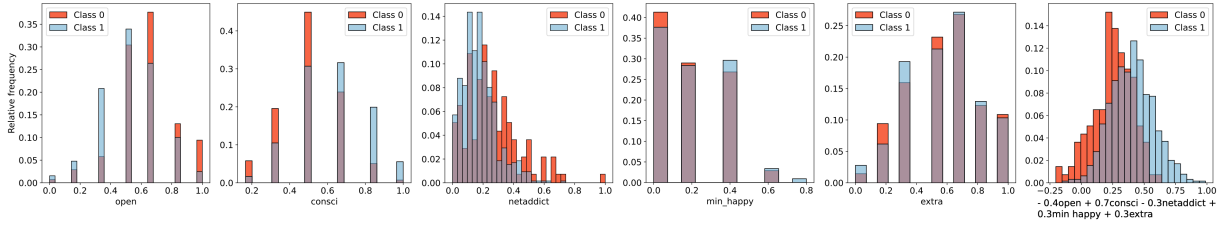


Figure A.3: Case 1: Relative frequency histograms for **Class 0** and **Class 1**. From left to right, sets of histograms correspond to `open`, `consci`, `netaddict`, `min_happy`, `extra`, and $-0.4\text{open} + 0.7\text{consci} - 0.3\text{netaddict} + 0.3\text{min_happy} + 0.3\text{extra}$.

I next perform a set of computational statistical analyses: hypothesis testing of the difference between two groups, analyses of correlation between attribute values and classes, and classification using only one attribute/composite variable. For the hypothesis testing, I apply the Brunner-Munzel test because Levene’s test imply I should not assume equal variances for several attributes (i.e., the Mann–Whitney U test is not appropriate). For correlation analysis, I compute the point-biserial correlation coefficient (r_{pb}) between the class IDs (i.e., 0 and 1) and attribute values. Lastly, for classification, I employ a linear discriminant analysis (LDA) with an oversampling method (specifically, the synthetic minority oversampling technique, or SMOTE) as **Class 1** has many more samples than **Class 0**. The results are shown in Table A.1. I can see the constructed composite variable provides notable improvements in correlation coefficients and classification accuracy.

Case 2. I repeat the same procedure as Case 1. The results are shown in Fig. A.4 and Table A.2. Fig. A.4 shows clear distribution shifts, where **Class 0** tends to have smaller values than **Class 1**. Also, all statistical measures show large improvements achieved by the constructed composite variable. For example, only the composite variable shows the statistical significance difference between **Class 0** and **Class 1**.

Case 3. I followed the same procedure as Cases 1 and 2. The results are shown in Fig. A.5 and Table A.3. Unlike the other two cases, I can only see subtle improvements

Table A.1: Case 1: Statistical analysis results. The bold fonts indicate considerable improvement by the composite variable $(-0.4open + 0.7consci - 0.3netaddict + 0.3min_happy + 0.3extra)$.

	p -value (Brunner-Munzel test)	correlation coefficient (r_{pb})	accuracy (LDA)
open	*** 3.3×10^{-8}	-0.19	0.61
consci	*** 5.5×10^{-13}	0.24	0.64
netaddict	*** 2.7×10^{-7}	-0.24	0.62
min_happy	0.29	0.042	0.51
extra	0.99	-0.0020	0.49
composite variable	*** 0.000	0.32	0.67

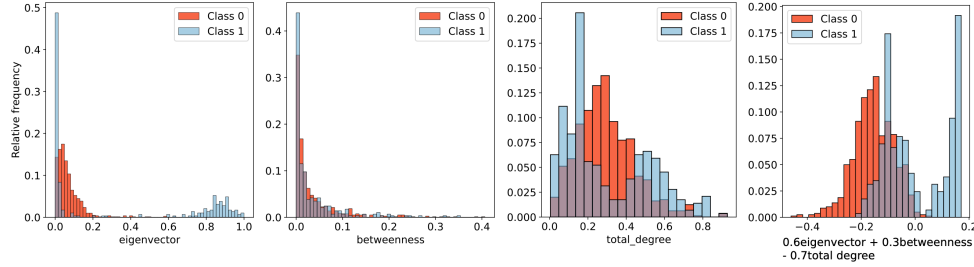


Figure A.4: Case 2: Relative frequency histograms for Class 0 and Class 1. From left to right, sets of histograms correspond to eigenvector, betweenness, total_degree, and $0.6eigenvector + 0.3betweenness - 0.7total_degree$.

Table A.2: Case 2: Statistical analysis results. The bold fonts indicate considerable improvement by the composite variable $(0.6eigenvector + 0.3betweenness - 0.7total_degree)$.

	p -value (Brunner-Munzel test)	correlation coefficient (r_{pb})	accuracy (LDA)
eigenvector	0.074	0.47	0.67
betweenness	0.065	-0.020	0.51
total_degree	0.11	0.012	0.48
composite variable	*** 0.000	0.61	0.74

(e.g., small improvements in correlation coefficients and classification accuracy). This situation of fewer improvements implies the difficulty of identifying the difference between the groups with simple statistical analyses. While the simplified representation in Case 3

utilizes neural networks and their nonlinear transformations (i.e., having high expressiveness), it still has a relatively low accuracy rate (0.82). A potential future improvement of the composite variable to aid in conventional statistical analysis is enabling the construction of more complicated composite variables, as described in Discussion section.

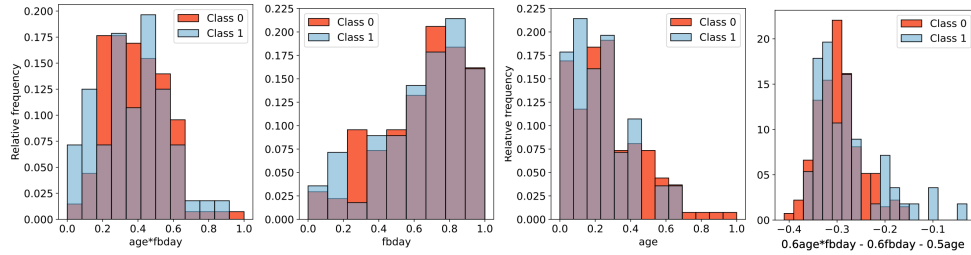


Figure A.5: Case 3: Relative frequency histograms for Class 0 and Class 1. From left to right, sets of histograms correspond to $\text{age} \times \text{fbdays}$, fbdays , age , and $0.6\text{age} \times \text{fbdays} - 0.6\text{fbdays} - 0.5\text{age}$.

Table A.3: Case 3: Statistical analysis results. The bold fonts indicate considerable improvement by the composite variable ($0.6\text{age} \times \text{fbdays} - 0.6\text{fbdays} - 0.5\text{age}$).

	p -value (Brunner-Munzel test)	correlation coefficient (r_{pb})	accuracy (LDA)
$\text{age} \times \text{fbdays}$	0.74	-0.028	0.49
fbdays	0.81	0.0090	0.51
age	0.11	-0.13	0.51
composite variable	0.50	0.14	0.53

A.3 Demonstration Video, Source Code, and Sample Data

I provide a demonstration video of interactive analysis with the designed workflow and UI. Also, to allow readers to test the workflow and UI, I provide related source code and sample data preprocessed from a publicly available dataset of faculty networks [89].

- Demonstration video: <https://www.youtube.com/watch?v=Vro6uFGYBho>
- Source code and sample data: <https://github.com/hylu1994/Network-CV>

Appendix B

Appendix for Chapter 3

B.1 Graph Centralization Index for Directed Bipartite Networks

To aggregate node-level measures, where summary statistics do not provide meaningful representations of all nodes, I extend the graph centralization index [132] definition to directed bipartite networks, which is applicable to our EHR communication networks. Freeman defined that for any node measure X and a graph $\mathcal{G}$ with n nodes, its graph centralization index is given by

$$C_X(\mathcal{G}) = \frac{\sum_{k=1}^n (M_X(p^*, \mathcal{G}) - M_X(p_k, \mathcal{G}))}{\underset{\mathcal{G}^*}{\operatorname{argmax}} \sum_{k=1}^n (M_X(p^*, \mathcal{G}^*) - M_X(p_k, \mathcal{G}^*))}, \quad (\text{B.1})$$

where $M_X(p_k, \mathcal{G})$ is the node measure value of node p_k , and $M_X(p^*, \mathcal{G})$ is the largest node measure value in the graph $\mathcal{G}$. The denominator is the maximum possible sum of node measure differences for any graph $\mathcal{G}^*$ with n nodes. The maximum possible sum of differences is proven by Freeman to occur in an undirected unipartite one-star topology as illustrated in Fig. B.1(a). Inheriting the core concepts from Freeman's equation Eq. B.1, I derive that a two-star topology as illustrated in Fig. B.1(b) leads to the maximum possible sum of differences for degree, betweenness, and closeness centrality in the following paragraphs.

The two-star topology includes a star node i in the node set V_1 containing n_1 nodes, bi-directionally connected to all nodes in the other node set V_2 , along with another star

node j in V_2 containing n_2 nodes, bi-directionally connected to all nodes in V_1 . Using these notations, I will derive the equations for directed bipartite graph centralization indices in the following paragraphs.

Degree centrality for a node k is the number of edges connected to k , which can either be *in-degree* or *out-degree*, counting the incoming or outgoing edges, respectively. Taking *out-degree* as example, for the star node i , the centrality difference from a non-star node in V_2 is $(n_2 - 1)$ and there are $(n_2 - 1)$ non-star nodes in V_2 , which leads to $(n_2 - 1)^2$; the centrality difference from a non-star node in V_1 is $(n_2 - 1)$ and there are $(n_1 - 1)$ non-star nodes in V_1 , which results in $((n_2 - 1) * (n_1 - 1))$; the centrality difference from the star node j is $(n_2 - n_1)$. Finally, I have the maximum sum of degree centrality differences from the star node i in V_1 as the sum of the above three terms $Degree_{diff-V_1} = (n_2 - 1)^2 + (n_2 - 1) * (n_1 - 1) + (n_2 - n_1)$. Deriving the same for the star node j , $Degree_{diff-V_2}$ can be computed and the directed bipartite degree graph centralization index C_D can be obtained by:

$$C_D(\mathcal{G}) = \frac{\sum_{k=1}^{n_1} [M_D(p_{V_1}^*, \mathcal{G}) - M_D(p_{V_1k}, \mathcal{G})]}{Degree_{diff-V_1}} + \frac{\sum_{k=1}^{n_2} [M_D(p_{V_2}^*, \mathcal{G}) - M_D(p_{V_2k}, \mathcal{G})]}{Degree_{diff-V_2}}, \quad (\text{B.2})$$

where $M_D(p_{V_1k}, \mathcal{G})$ can be either the in-degree or out-degree centrality value of node p_{V_1k} in the node set V_1 , and so can $M_D(p_{V_2k}, \mathcal{G})$.

The star structures induce the maximum sum of centrality differences because they maximally centralize the minimum number of node(s). I further validate the usability of our two-star topology by analyzing the potential edge rewiring. Within the two-star directed bipartite topology, the only two rewiring options are (1) removing the edge(s) between star nodes i and j , and (2) rewiring edge(s) to connect two non-star nodes in the different node sets. First, I recompute the three out-degree difference terms in $Degree_{diff-V_1}$ after removing the edge from the star node i to the star node j , which will be $(n_2 - 2) * (n_2 - 1)$, $(n_2 - 2) * (n_1 - 1)$, and $(n_2 - n_1 - 1)$. The total centrality difference after the edge removal is $(n_2 + n_1 - 1)$ less than before. Second, after rewiring the source node of an edge from the star node i to a non-star node in V_1 , the same three terms of out-degree centrality difference will be $(n_2 - 2) * (n_2 - 1)$, $(n_2 - 2) * (n_1 - 2) + (n_2 - 3)$, and $(n_2 - n_1 - 1)$. The total centrality difference after edge rewiring is $(n_2 + n_1)$ less than

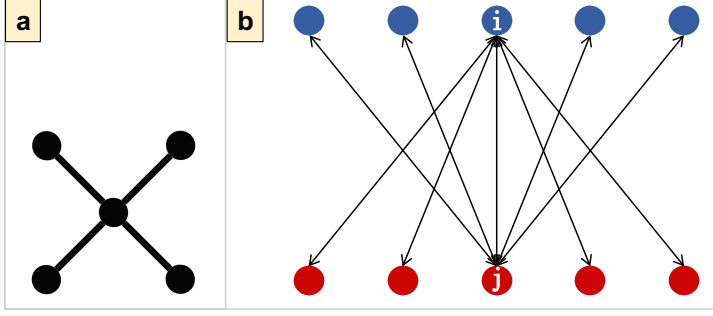


Figure B.1: (a) The star pattern in undirected and unipartite networks. (b) The two-star pattern in directed and bipartite networks. The blue nodes represent one type of nodes and the red nodes represent the other type of nodes. Nodes i and j constitute the two stars in the respective types of nodes.

before. Therefore, the two-star topology remains the structure resulting in the maximum sum of centrality differences.

Betweenness centrality for a node k is the fraction of all-pairs shortest time-respecting paths that pass through node k . I adopt the modified normalization term defined in [199] to compute directed bipartite betweenness centrality $M_{B.V_1}$ and $M_{B.V_2}$. In addition, the denominator of betweenness graph centralization index was defined in [200] as the number of nodes, which will be $n_1 - 1$ and $n_2 - 1$ in bipartite networks. Hence, I derive the directed bipartite betweenness graph centralization index C_B as:

$$C_B(\mathcal{G}) = \frac{\sum_{k=1}^{n_1} [M_{B.V_1}(p_{V_1}^*, \mathcal{G}) - M_{B.V_1}(p_{V_1 k}, \mathcal{G})]}{n_1 - 1} + \frac{\sum_{k=1}^{n_2} [M_{B.V_2}(p_{V_2}^*, \mathcal{G}) - M_{B.V_2}(p_{V_2 k}, \mathcal{G})]}{n_2 - 1}. \quad (\text{B.3})$$

Closeness centrality of a node k is the reciprocal of the sum of all-pairs shortest path distance (i.e., the number of hops) of k to(out-closeness) or from(in-closeness) all the other nodes, normalized by the minimum possible sum of distances [132], [201]. For bipartite networks, the minimum distances in the two node sets are $(2 * (n_1 - 1) + n_2)$ and $(2 * (n_2 - 1) + n_1)$ as discussed in [199], which can be resulted from the two star nodes i and j , respectively. This observation further validates that our two-star topology optimally centralizes the minimum number of nodes, and from which I can derive the closeness node measure for separate node sets as $M_{C.V_1} = \frac{2 * (n_1 - 1) + n_2}{d_{V_1}}$ and $M_{C.V_2} = \frac{2 * (n_2 - 1) + n_1}{d_{V_2}}$, where d_{V_1} and d_{V_2} are the sum of all-pairs shortest path distance.

Since the normalized terms in $M_{C.V_1}$ and $M_{C.V_2}$ are constants w.r.t any edge rewiring, I analyze the potential minimum shortest path distance to find the largest possible closeness measure difference in a directed bipartite network. In our two-star topology, the closeness node measure is 1 of the star nodes i and j . For a non-star node k in V_1 , the shortest path

distance from another node in V_1 is 2 and there are $(n_1 - 1)$ such distance; the shortest path distance from a non-star node in V_2 is 3 and there are $(n_2 - 1)$ such distance; the shortest distance from the star node j is 1. The sum of the shortest path distances of a non-star node k in V_1 is $V_1_nonStar_distance = (2 * (n_1 - 1)) + (3 * (n_2 - 1)) + 1$, and the closeness centrality of the non-star node k will be $M_{C_V_1_nonStar} = \frac{2*(n_1-1)+n_2}{V_1_nonStar_distance}$. Each non-star node in V_1 exhibits the same closeness centrality value as the node k , and hence, there are $(n_1 - 1)$ such closeness centrality difference between the star node i in and a non-star node in V_1 , which is $(1 - M_{C_V_1_nonStar})$, leading to the maximum sum of closeness measure difference in the node set V_1 as $Closeness_{diff_V_1} = (n_1 - 1) * (1 - M_{C_V_1_nonStar})$. Deriving the same for the node set V_2 the same process, I get $V_2_nonStar_distance = (2 * (n_2 - 1)) + (3 * (n_1 - 1)) + 1$, $M_{C_V_2_nonStar} = \frac{2*(n_2-1)+n_1}{V_2_nonStar_distance}$, and $Closeness_{diff_V_2} = (n_2 - 1) * (1 - M_{C_V_2_nonStar})$. Finally, following the computation of Eq. 1, I have the closeness graph centralization index C_C as:

$$C_C(\mathcal{G}) = \frac{\sum_{k=1}^{n_1} [M_{C_V_1}(p_{V_1}^*, \mathcal{G}) - M_{C_V_1}(p_{V_1k}, \mathcal{G})]}{Closeness_{diff_V_1}} + \frac{\sum_{k=1}^{n_2} [M_{C_V_2}(p_{V_2}^*, \mathcal{G}) - M_{C_V_2}(p_{V_2k}, \mathcal{G})]}{Closeness_{diff_V_2}}, \quad (\text{B.4})$$

where $M_{C_V_1}(p_{V_1k}, \mathcal{G})$ can be either the in-closeness (i.e., using the incoming paths to the node k) or out-closeness (i.e., using the outgoing paths from the node k) of the node p_{V_1k} in V_1 , and so can $M_{C_V_2}(p_{V_2k}, \mathcal{G})$.

REFERENCES

- [1] Mark Newman. *Networks*. Oxford university press, 2018.
- [2] David Kempe, Jon Kleinberg, and Amit Kumar. Connectivity and inference problems for temporal networks. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 504–513, 2000.
- [3] Kenneth L Cooke and Eric Halsey. The shortest route through a network with time-dependent internodal transit times. *Journal of mathematical analysis and applications*, 14(3):493–498, 1966.
- [4] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.
- [5] Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. GN-Explainer: Generating explanations for graph neural networks. *Adv. Neural Inf. Process. Syst.*, 32:9240–9251, dec 2019.
- [6] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [7] Lei Shi, Chen Wang, Zhen Wen, Huamin Qu, Chuang Lin, and Qi Liao. 1.5 d egocentric dynamic network visualization. *IEEE transactions on visualization and computer graphics*, 21(5):624–637, 2014.
- [8] Qingsong Liu, Yifan Hu, Lei Shi, Xinzhu Mu, Yutao Zhang, and Jie Tang. Egonet-cloud: Event-based egocentric dynamic network visualization. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 65–72. IEEE, 2015.

- [9] Zhihua Jin, Yong Wang, Qianwen Wang, Yao Ming, Tengfei Ma, and Huamin Qu. Gnnlens: A visual analytics approach for prediction error diagnosis of graph neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 2022.
- [10] Andreas Kerren, Helen C. Purchase, and Matthew O. Ward, editors. *Multivariate Network Visualization*, volume 8380 of *Lect Notes Comput Sci*. Springer, 2014. doi: <https://doi.org/10.1007/978-3-319-06793-3>.
- [11] Martin Atzmueller, Stephan Günnemann, and Albrecht Zimmermann. Mining communities and their descriptions on attributed graphs: A survey. *Data Min Knowl Disc*, 35(3):661–687, 2021. doi: <https://doi.org/10.1007/s10618-021-00741-z>.
- [12] Raj Chetty, Matthew O Jackson, Theresa Kuchler, Johannes Stroebe, Nathaniel Hendren, et al. Social capital I: Measurement and associations with economic mobility. *Nature*, 608:108–121, 2022. doi: <https://doi.org/10.3386/w30313>.
- [13] Qing Chen, Nan Chen, Wei Shuai, Guande Wu, Zhe Xu, Hanghang Tong, and Nan Cao. Calliope-Net: Automatic generation of graph data facts via annotated node-link diagrams. *IEEE Trans Vis Comput Graph*, 30(1):562–572, 2024.
- [14] Carolina Nobre, Miriah Meyer, Marc Streit, and Alexander Lex. The state of the art in visualizing multivariate networks. *Comput Graph Forum*, 38(3):807–832, 2019. doi: <https://doi.org/10.1111/cgf.13728>.
- [15] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. Network representation learning: A survey. *IEEE Trans Big Data*, 6(1):3–28, 2018. doi: 10.1109/TBDATA.2018.2850013.
- [16] Z. Huang, D. Witschard, K. Kucher, and A. Kerren. VA + embeddings STAR: A state-of-the-art report on the use of embeddings in visual analytics. *Comput Graph Forum*, 42(3):539–571, 2023. doi: <https://doi.org/10.1111/cgf.14859>.
- [17] Takanori Fujiwara, Jian Zhao, Francine Chen, and Kwan-Liu Ma. A visual analytics

- framework for contrastive network analysis. In *Proc. VAST*, pages 48–59, 2020. doi: <https://doi.org/10.1109/vast50239.2020.00010>.
- [18] Takanori Fujiwara, Jian Zhao, Francine Chen, Yaoliang Yu, and Kwan-Liu Ma. Network comparison with interpretable contrastive network representation learning. *J Data Sci Stat Vis*, 2(5), 2022. doi: <https://doi.org/10.52933/jdssv.v2i5.56>.
- [19] Rafael Messias Martins, Gabriel Faria Andery, Henry Heberle, Fernando Vieira Paulovich, Alneu de Andrade Lopes, Helio Pedrini, and Rosane Minghim. Multidimensional projections for visual analysis of social networks. *J Comput Sci Technol*, 27(4):791–810, 2012. doi: <https://doi.org/10.1007/s11390-012-1265-5>.
- [20] Rafael Messias Martins, Johannes F Kruiger, Rosane Minghim, Alexandru C Telea, and Andreas Kerren. MVN-Reduce: Dimensionality reduction for the visual analysis of multivariate networks. In *Proc. EuroVis*, pages 13–17, 2017. doi: [10.2312/eurovisshort.20171126](https://doi.org/10.2312/eurovisshort.20171126).
- [21] Stef van den Elzen, Danny Holten, Jorik Blaas, and Jarke J van Wijk. Reducing snapshots to points: A visual analytics approach to dynamic network exploration. *IEEE Trans Vis Comput Graph*, 22(1):1–10, 2016. doi: [10.1109/TVCG.2015.2468078](https://doi.org/10.1109/TVCG.2015.2468078).
- [22] Huan Song, Zeng Dai, Panpan Xu, and Liu Ren. Interactive visual pattern search on graph data via graph representation learning. *IEEE Trans Vis Comput Graph*, 28(1):335–345, 2022. doi: <https://doi.org/10.1109/tvcg.2021.3114857>.
- [23] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proc. NIPS*, pages 1024–1034, 2017. doi: <https://doi.org/10.48550/arXiv.1706.02216>.
- [24] Mi-Kyung Song, Feng-Chang Lin, Sandra E Ward, and Jason P Fine. Composite variables: When and how. *Nurs Res*, 62(1):45, 2013. doi: <https://doi.org/10.1097/nnr.0b013e3182741948>.

- [25] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS’17, page 4768–4777. Curran Associates Inc., 2017. ISBN 9781510860964.
- [26] The supplementary materials: Source code, data, and a demonstration video. <https://github.com/hylu1994/Network-CV>.
- [27] Fintan Mcgee, Mohammad Ghoniem, Guy Melançon, Benoit Otjacques, and Bruno Pinaud. The state of the art in multilayer network visualization. *Comput Graph Forum*, 38(6):125–149, 2019. doi: <https://doi.org/10.1111/cgf.13610>.
- [28] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. A taxonomy and survey of dynamic graph visualization. In *Computer graphics forum*, volume 36, pages 133–159. Wiley Online Library, 2017.
- [29] Angelos Chatzimparmpas, Rafael Messias Martins, Ilir Jusufi, Kostiantyn Kucher, Fabrice Rossi, and Andreas Kerren. The state of the art in enhancing trust in machine learning models with the use of visualizations. *Comput Graph Forum*, 39(3):713–756, 2020. doi: <https://doi.org/10.1111/cgf.14034>.
- [30] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proc. KDD*, pages 855–864, 2016. doi: <https://doi.org/10.1145/2939672.2939754>.
- [31] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proc. ICLR*, pages 1–10, 2017. doi: <https://doi.org/10.48550/arXiv.1609.02907>.
- [32] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? In *Proc. NeurIPS*, volume 34, pages 28877–28888, 2021. doi: <https://doi.org/10.48550/arXiv.2106.05234>.

- [33] Manuel Freire, Catherine Plaisant, Ben Shneiderman, and Jen Golbeck. ManyNets: An interface for multiple network analysis and visualization. In *Proc. CHI*, pages 213–222, 2010. doi: <https://doi.org/10.1145/1753326.1753358>.
- [34] Robert Gove. Gragnostics: Fast, interpretable features for comparing graphs. In *Proc. IV*, pages 201–209, 2019. doi: <https://doi.org/10.1109/iv.2019.00042>.
- [35] Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, 2007. doi: <https://doi.org/10.1093/bioinformatics/btq091>.
- [36] Tatiana von Landesberger, Melanie Gorner, and Tobias Schreck. Visual analysis of graphs with multiple connected components. In *Proc. VAST*, pages 155–162, 2009. doi: <https://doi.org/10.1109/vast.2009.5333893>.
- [37] Martin Harrigan, Daniel Archambault, Pádraig Cunningham, and Neil Hurley. EgoNav: Exploring networks through egocentric spatializations. In *Proc. AVI*, pages 563–570, 2012. doi: <https://doi.org/10.1145/2254556.2254661>.
- [38] Oh-Hyun Kwon, Tarik Crnovrsanin, and Kwan-Liu Ma. What would a graph look like in this layout? A machine learning approach to large graph visualization. *IEEE Trans Vis Comput Graph*, 24(1):478–488, 2018. doi: <https://doi.org/10.1109/tvcg.2017.2743858>.
- [39] Benjamin Bach, Conglei Shi, Nicolas Heulot, Tara Madhyastha, Tom Grabowski, and Pierre Dragicevic. Time Curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Trans Vis Comput Graph*, 22(1):559–568, 2016. doi: [10.1109/TVCG.2015.2467851](https://doi.org/10.1109/TVCG.2015.2467851).
- [40] Takanori Fujiwara, Jia-Kai Chou, Andrew M McCullough, Charan Ranganath, and Kwan-Liu Ma. A visual analytics system for brain functional connectivity comparison across individuals, groups, and time points. In *Proc. PacificVis*, pages 250–259, 2017. doi: <https://doi.org/10.1109/pacificvis.2017.8031601>.

- [41] James Y Zou, Daniel J Hsu, David C Parkes, and Ryan P Adams. Contrastive learning using spectral methods. *Proc. NIPS*, 26, 2013.
- [42] Christoph Molnar. *Interpretable machine learning*. Lulu.com, 2020.
- [43] Biagio La Rosa, Graziano Blasilli, R Bourqui, D Auber, Giuseppe Santucci, Roberto Capobianco, Enrico Bertini, Romain Giot, and Marco Angelini. State of the art of visual analytics for explainable deep learning. In *Comput Graph Forum*, volume 42, pages 319–355. Wiley Online Library, 2023.
- [44] Jaegul Choo and Shixia Liu. Visual analytics for explainable deep learning. *IEEE Comput Graph Appl*, 38(4):84–92, 2018.
- [45] Fred Hohman, Minsuk Kahng, Robert Pienta, and Duen Horng Chau. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Trans Vis Comput Graph*, 25(8):2674–2693, 2018.
- [46] Ronan Hamon, Henrik Junklewitz, and Ignacio Sanchez. Robustness and explainability of artificial intelligence. Technical Report KJ-NA-30040-EN-N (online), Publications Office of the European Union, 2020.
- [47] Bum Chul Kwon, Ben Eysenbach, Janu Verma, Kenney Ng, Christopher De Filippi, Walter F Stewart, and Adam Perer. Clustervision: Visual supervision of unsupervised clustering. *IEEE Trans Vis Comput Graph*, 24(1):142–151, 2018. doi: <https://doi.org/10.1109/tvcg.2017.2745085>.
- [48] Mário Popolin Neto and Fernando V. Paulovich. Multivariate data explanation by jumping emerging patterns visualization. *IEEE Trans Vis Comput Graph*, 30(2): 1549–1563, 2024.
- [49] Christian van Onzenoodt, Pere-Pau Vázquez, and Timo Ropinski. Out of the plane: Flower vs. star glyphs to support high-dimensional exploration in two-dimensional embeddings. *IEEE Trans Vis Comput Graph*, pages 1–15, 2022 (early access). doi: [10.1109/TVCG.2022.3216919](https://doi.org/10.1109/TVCG.2022.3216919).

- [50] Takanori Fujiwara, Oh-Hyun Kwon, and Kwan-Liu Ma. Supporting analysis of dimensionality reduction results with contrastive learning. *IEEE Trans Vis Comput Graph*, 26(1):45–55, 2020. doi: <https://doi.org/10.1109/tvcg.2019.2934251>.
- [51] Paulo Joia, Fabiano Petronetto, and Luis Gustavo Nonato. Uncovering representative groups in multidimensional projections. *Comput Graph Forum*, 34(3):281–290, 2015. doi: <https://doi.org/10.1111/cgf.12640>.
- [52] Cagatay Turkey, Arvid Lundervold, Astri Lundervold, and Helwig Hauser. Representative factor generation for the interactive visual analysis of high-dimensional data. *IEEE Trans Vis Comput Graph*, 18(12):2621–2630, 2012. doi: <https://doi.org/10.1109/tvcg.2012.256>.
- [53] Fangfang Zhou, Juncai Li, Wei Huang, Ying Zhao, Xiaoru Yuan, Xing Liang, and Yang Shi. Dimension reconstruction for visual exploration of subspace clusters in high-dimensional data. In *Proc. PacificVis*, pages 128–135, 2016. doi: <https://doi.org/10.1109/pacificvis.2016.7465260>.
- [54] Zelin Zang, Shenghui Cheng, Linyan Lu, Hanchen Xia, Liangyu Li, Yaoting Sun, Yongjie Xu, Lei Shang, Baigui Sun, and Stan Z. Li. EVNet: An explainable deep network for dimension reduction. *IEEE Trans Vis Comput Graph*, pages 1–18, 2022 (early access). doi: [10.1109/TVCG.2022.3223399](https://doi.org/10.1109/TVCG.2022.3223399).
- [55] Josua Krause, Adam Perer, and Kenney Ng. Interacting with predictions: Visual inspection of black-box machine learning models. In *Proc. CHI*, pages 5686–5697, 2016.
- [56] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. The what-if tool: Interactive probing of machine learning models. *IEEE Trans Vis Comput Graph*, 26(1):56–65, 2019.
- [57] Daniel W Apley and Jingyu Zhu. Visualizing the effects of predictor variables in black box supervised learning models. *J R Stat Soc Series B Stat Methodol*, 82(4):1059–1086, 2020.

- [58] Yiran Li, Takanori Fujiwara, Yong K. Choi, Katherine K. Kim, and Kwan-Liu Ma. A visual analytics system for multi-model comparison on clinical data predictions. *Visual Informatics*, 4(2):122–131, 2020.
- [59] Marco Angelini, Graziano Blasilli, Simone Lenti, and Giuseppe Santucci. A visual analytics conceptual framework for explorable and steerable partial dependence analysis. *IEEE Trans Vis Comput Graph*, Early Access.
- [60] Johannes Knittel, Andres Lalama, Steffen Koch, and Thomas Ertl. Visual neural decomposition to explain multivariate data sets. *IEEE Trans Vis Comput Graph*, 27(2):1374–1384, 2020. doi: <https://doi.org/10.1109/tvcg.2020.3030420>.
- [61] Michael Gleicher. Explainers: Expert explorations with crafted projections. *IEEE Trans Vis Comput Graph*, 19(12):2042–2051, 2013. doi: <https://doi.org/10.1109/tvcg.2013.157>.
- [62] Tarik Crnovrsanin, Chris W. Muelder, Robert Faris, Diane Felmlee, and Kwan-Liu Ma. Visualization techniques for categorical analysis of social networks with multiple edge sets. *Soc Networks*, 37:56–64, 2014. doi: <https://doi.org/10.1016/j.socnet.2013.12.002>.
- [63] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nat Neurosci*, 17(11):1500–1509, 2014. doi: <https://doi.org/10.1038/nn.3776>.
- [64] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *J Comput Graph Stat*, 15(2):265–286, 2006. doi: <https://doi.org/10.1198/106186006X113430>.
- [65] Donato Malerba, Floriana Esposito, and Giovanni Semeraro. A further comparison of simplification methods for decision-tree induction. In Doug Fisher and Hans-J. Lenz, editors, *Learning from Data: Artificial Intelligence and Statistics V*, pages 365–374. Springer New York, 1996. doi: https://doi.org/10.1007/978-1-4612-2404-4_35.

- [66] John P Cunningham and Zoubin Ghahramani. Linear dimensionality reduction: Survey, insights, and generalizations. *J Mach Learn Res*, 16(1):2859–2900, 2015. doi: <https://doi.org/10.48550/arXiv.1406.0873>.
- [67] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst*, 32(1):4–24, 2021. doi: <https://doi.org/10.1109/tnnls.2020.2978386>.
- [68] Takanori Fujiwara, Xinhai Wei, Jian Zhao, and Kwan-Liu Ma. Interactive dimensionality reduction for comparative analysis. *IEEE Trans Vis Comput Graph*, 28(1):758–768, 2022. doi: <https://doi.org/10.1109/tvcg.2021.3114807>.
- [69] Alper Sarikaya and Michael Gleicher. Scatterplots: Tasks, data, and designs. *IEEE Trans Vis Comput Graph*, 24(1):402–412, 2018. doi: <https://doi.org/10.1109/tvcg.2017.2744184>.
- [70] Adrian Mayorga and Michael Gleicher. Splatterplots: Overcoming overdraw in scatter plots. *IEEE Trans Vis Comput Graph*, 19(9):1526–1538, 2013. doi: <https://doi.org/10.1109/tvcg.2013.65>.
- [71] Min Lu, Shuaiqi Wang, Joel Lanir, Noa Fish, Yang Yue, Daniel Cohen-Or, and Hui Huang. Winglets: Visualizing association with uncertainty in multi-class scatterplots. *IEEE Trans Vis Comput Graph*, 26(1):770–779, 2020. doi: <https://doi.org/10.1109/tvcg.2019.2934811>.
- [72] Jürgen Bernard, Martin Steiger, Sebastian Mittelstädt, Simon Thum, Daniel Keim, and Jörn Kohlhammer. A survey and task-based quality assessment of static 2D colormaps. In *Proc. VDA*, volume 9397, page 93970M, 2015.
- [73] Sylvain Lespinsats and Michaël Aupetit. CheckViz: Sanity check and topological clues for linear and non-linear mappings. *Comput Graph Forum*, 30(1):113–125, 2011. doi: <https://doi.org/10.1111/j.1467-8659.2010.01835.x>.

- [74] The Matplotlib development team. Choosing colormaps in matplotlib. <https://matplotlib.org/stable/users/explain/colors/colormaps.html>, 2023. Accessed: 2023-09-25.
- [75] Michael Correll, Dominik Moritz, and Jeffrey Heer. Value-suppressing uncertainty palettes. In *Proc. CHI*, page 1–11. ACM, 2018. ISBN 9781450356206. doi: 10.1145/3173574.3174216.
- [76] Ming-yi Chang and Yang-chih Fu. Social media and network boundaries among college students: Reconstructing companions, conversations, and contact circles. *Taiwanese Sociology*, 37:1–46, 2019. doi: 10.6676/TS.201906_(37).02. Written in Chinese.
- [77] Hsuan-Wei Lee, Ming-Yi Chang, Wan-Yu Chou, Jing-Shiang Hwang, and Yang-Chih Fu. From indirect to direct contacts on Facebook: A big-data approach to the making of triadic network closure. *Can Rev Sociol*, 59(2):207–227, 2022. doi: <https://doi.org/10.1111/cars.12375>.
- [78] Ryan A Rossi, Rong Zhou, and Nesreen Ahmed. Deep inductive graph representation learning. *IEEE Trans Knowl Data Eng*, 32(3):438–452, 2018. doi: 10.1109/TKDE.2018.2878247.
- [79] Yaqian Guo, Trevor Hastie, and Robert Tibshirani. Regularized linear discriminant analysis and its application in microarrays. *Biostat*, 8(1):86–100, 2007. doi: <https://doi.org/10.1093/biostatistics/kxj035>.
- [80] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*, 67(2):301–320, 2005. doi: <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.
- [81] Michael Waskom. seaborn.swarmplot. <https://seaborn.pydata.org/generated/seaborn.swarmplot.html>, 2016. Accessed: 2022-12-27.

- [82] Scott M. Lundberg. SHAP (SHapley Additive exPlanations). <https://github.com/slundberg/shap>, 2018. Accessed: 2022-12-28.
- [83] Jim Albert. Sabermetrics: The past, the present, and the future. In Joseph A. Editor Gallian, editor, *Mathematics and Sports*, page 3–14. Mathematical Association of America, 2010. doi: <https://doi.org/10.5948/UPO9781614442004.002>.
- [84] Yakir A Reshef, David N Reshef, Hilary K Finucane, Pardis C Sabeti, and Michael Mitzenmacher. Measuring dependence powerfully and equitably. *J Mach Learn Res*, 17(1):7406–7468, 2016. doi: <https://doi.org/10.48550/arXiv.1505.02213>.
- [85] Michael JD Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287–336, 1998. doi: <https://doi.org/10.1017/s0962492900002841>.
- [86] Christine P Dancey and John Reidy. *Statistics without Maths for Psychology*. Pearson London, 2017.
- [87] Yifan Hu. Efficient, high-quality force-directed graph drawing. *Math J*, 10(1):37–71, 2005.
- [88] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, et al. PyTorch: An imperative style, high-performance deep learning library. In *Proc. NeurIPS*, pages 8024–8035, 2019. doi: <https://doi.org/10.48550/arXiv.1912.01703>.
- [89] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*, 12: 2825–2830, 2011. doi: <https://doi.org/10.48550/arXiv.1201.0490>.
- [90] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat Methods*, 17: 261–272, 2020. doi: <https://doi.org/10.1038/s41592-020-0772-5>.
- [91] Tiago P. Peixoto. The graph-tool python library. *figshare*, 2014. http://figshare.com/articles/graph_tool/1164194.

- [92] Michael L. Waskom. seaborn: statistical data visualization. *J Open Source Softw*, 6(60):3021, 2021. doi: <https://doi.org/10.21105/joss.03021>.
- [93] Isaac Muse. ColorAide. <https://facelessuser.github.io/coloraide/>, 2020. Accessed: 2022-12-29.
- [94] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D³ data-driven documents. *IEEE Trans Vis Comput Graph*, 17(12):2301–2309, 2011. doi: 10.1109/TVCG.2011.185.
- [95] K Hunter Wapman, Sam Zhang, Aaron Clauset, and Daniel B Larremore. Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature*, 610(7930):120–127, 2022. doi: <https://doi.org/10.1038/s41586-022-05222-x>.
- [96] Jennie E Brand, Jiahui Xu, Bernard Koch, and Pablo Geraldo. Uncovering sociological effect heterogeneity using tree-based machine learning. *Sociol Methodol*, 51(2):189–223, 2021. doi: <https://doi.org/10.1177/0081175021993503>.
- [97] Yiran Li, Erin Musabandesu, Takanori Fujiwara, Frank J Loge, and Kwan-Liu Ma. A visual analytics system for water distribution system optimization. In *Proc. VIS*, pages 126–130, 2021. doi: <https://doi.org/10.1109/vis49827.2021.9623272>.
- [98] Jaemin Jo, Frédéric Vernier, Pierre Dragicevic, and Jean-Daniel Fekete. A declarative rendering model for multiclass density maps. *IEEE Trans Vis Comput Graph*, 25(1):470–480, 2018. doi: <https://doi.org/10.1109/tvcg.2018.2865141>.
- [99] Ayse P Gurses and Yan Xiao. A systematic review of the literature on multidisciplinary rounds to design information technology. *Journal of the American Medical Informatics Association*, 13(3):267–276, 2006.
- [100] M Smits, M Zegers, PP Groenewegen, DRM Timmermans, L Zwaan, G Van der Wal, and C Wagner. Exploring the causes of adverse events in hospitals and potential prevention strategies. *Quality and Safety in Health Care*, 19(5):e5–e5, 2010.

- [101] Annamaria Bagnasco, Barbara Tubino, Emanuela Piccotti, Francesca Rosa, Giuseppe Aleo, Pasquale Di Pietro, Loredana Sasso, Donatella Passalacqua, Laura Gambino, et al. Identifying and correcting communication failures among health professionals working in the emergency department. *International emergency nursing*, 21(3):168–172, 2013.
- [102] Kim J Verhaegh, Annamarieke Seller-Boersma, Robert Simons, Jeanet Steenbruggen, Suzanne E Geerlings, Sophia E de Rooij, and Bianca M Buurman. An exploratory study of healthcare professionals’ perceptions of interprofessional communication and collaboration. *Journal of interprofessional care*, 31(3):397–400, 2017.
- [103] M. Kilduff and D. J. Brass. Organizational social network research: Core ideas and key debates. *Academy of Management Annals*, 4(1):317–357, 2010. ISSN 1941-6520. doi: 10.1080/19416520.2010.494827. URL <GotoISI>://WOS:000286416100005<http://www.tandfonline.com/doi/abs/10.1080/19416520.2010.494827>.
- [104] Nancy Katz, David Lazer, Holly Arrow, and Noshir Contractor. Network theory and small groups. *Small Group Research*, 35(3):307–332, 2004. ISSN 1046-4964. doi: 10.1177/1046496404264941. URL <http://sgr.sagepub.com/content/35/3/307.abstract><http://sgr.sagepub.com/content/35/3/307.full.pdf>.
- [105] Peter R Monge and Noshir S Contractor. *Theories of communication networks*. Oxford University Press, New York, NY, 2003. ISBN 0195160371.
- [106] Phyllis Butow and Ehsan Hoque. Using artificial intelligence to analyse and teach communication in healthcare. *The breast*, 50:49–55, 2020.
- [107] Carla M Pires and Afonso M Cavaco. Communication between health professionals and patients: review of studies using the rias (roter interaction analysis system) method. *Revista da Associação Médica Brasileira*, 60:156–172, 2014.
- [108] Sarah Atkins. Assessing health professionals’ communication through role-play:

- An interactional analysis of simulated versus actual general practice consultations. *Discourse Studies*, 21(2):109–134, 2019.
- [109] Cristina Vaz de Almeida and Célia Belim. Health professionals’ communication competences decide patients’ well-being: Proposal for a communication model. In *Joy*, volume 5, pages 201–222. Emerald Publishing Limited, 2020.
 - [110] Tong Wu, Yunlong Wang, Yue Wang, Emily Zhao, Yilian Yuan, and Zhi Yang. Representation learning of ehr data via graph-based medical entity embedding. *arXiv preprint arXiv:1910.02574*, 2019.
 - [111] Zheng Liu, Xiaohan Li, Hao Peng, Lifang He, and S Yu Philip. Heterogeneous similarity graph neural network on electronic health records. In *2020 IEEE international conference on big data (big data)*, pages 1196–1205. IEEE, 2020.
 - [112] Derun Cai, Chenxi Sun, Moxian Song, Baofeng Zhang, Shenda Hong, and Hongyan Li. Hypergraph contrastive learning for electronic health records. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pages 127–135. SIAM, 2022.
 - [113] Yared H Kidane and Peter A Gloor. Correlating temporal communication patterns of the eclipse open source community with performance and creativity. *Computational and mathematical organization theory*, 13:17–27, 2007.
 - [114] Timo Wolf, Adrian Schroter, Daniela Damian, and Thanh Nguyen. Predicting build failures using social network analysis on developer communication. In *2009 IEEE 31st International Conference on Software Engineering*, pages 1–11. IEEE, 2009.
 - [115] Nerida Creswick and Johanna I Westbrook. Social network analysis of medication advice-seeking interactions among staff in an australian hospital. *Intl. journal of medical informatics*, 79(6):e116–e125, 2010.
 - [116] John Tang, Mirco Musolesi, Cecilia Mascolo, and Vito Latora. Temporal distance

- metrics for social network analysis. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 31–36, 2009.
- [117] Nicola Santoro, Walter Quattrociocchi, Paola Flocchini, Arnaud Casteigts, and Frederic Amblard. Time-varying graphs and social network analysis: Temporal indicators and metrics. *arXiv preprint arXiv:1102.0629*, 2011.
 - [118] Gueorgi Kossinets, Jon Kleinberg, and Duncan Watts. The structure of information pathways in a social communication network. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 435–443, 2008.
 - [119] Petter Holme. Network reachability of real-world contact sequences. *Physical Review E*, 71(4):046119, 2005.
 - [120] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.
 - [121] Xi Zhu, Shin-Ping Tu, Daniel Sewell, Nengliang Aaron Yao, Vimal Mishra, Alan Dow, and Colin Banas. Measuring electronic communication networks in virtual care teams using electronic health records access-log data. *International Journal of Medical Informatics*, 128:46–52, 2019.
 - [122] Adam G Dunn and Johanna I Westbrook. Interpreting social network metrics in healthcare organisations: a review and guide to validating small networks. *Social science & medicine*, 72(7):1064–1068, 2011.
 - [123] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.
 - [124] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Network robustness and fragility: Percolation on random graphs. *Physical review letters*, 85(25):5468, 2000.

- [125] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. Network analysis in the social sciences. *science*, 323(5916):892–895, 2009.
- [126] Stephen P Borgatti. 2-mode concepts in social network analysis. *Encyclopedia of complexity and system science*, 6:8279–8291, 2009.
- [127] Hsiao-Ying Lu, Takanori Fujiwara, Ming-Yi Chang, Yang-chih Fu, Anders Ynnerman, and Kwan-Liu Ma. Visual analytics of multivariate networks with representation learning and composite variable construction. *arXiv:2303.09590*, 2023.
- [128] Florian Reitz. A framework for an ego-centered and time-aware visualization of relations in arbitrary data repositories. *arXiv:1009.5183*, 2010.
- [129] Eloïse Loubier and Bernard Dousset. Temporal and relational data representation by graph morphing. *Safety and Reliability for managing Risk (ESREL 2008), Hammamet*, 14(02):2008–16, 2008.
- [130] Stephan Diehl and Carsten Görg. Graphs, they are changing: dynamic graph drawing for a sequence of graphs. In *International Symposium on Graph Drawing*, pages 23–31. Springer, 2002.
- [131] Peter A Gloor and Yan Zhao. Tecflow-a temporal communication flow visualizer for social networks analysis. In *ACM CSCW Workshop on Social Networks*, volume 6, page 5, 2004.
- [132] Linton C Freeman et al. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge*, 1:238–263, 2002.
- [133] Mary E Charlson, Peter Pompei, Kathy L Ales, and C Ronald MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*, 40(5):373–383, 1987.
- [134] Anne Elixhauser, Claudia Steiner, D Robert Harris, and Rosanna M Coffey. Comorbidity measures for use with administrative data. *Medical care*, pages 8–27, 1998.

- [135] Minh Vu and My T Thai. Pgm-explainer: Probabilistic graphical model explanations for graph neural networks. *Advances in neural information processing systems*, 33:12225–12235, 2020.
- [136] Asim Kumar Debnath, Rosa L Lopez de Compadre, Gargi Debnath, Alan J Shusterman, and Corwin Hansch. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *Journal of medicinal chemistry*, 34(2):786–797, 1991.
- [137] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in graph neural networks: A taxonomic survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [138] Hao Yuan, Jiliang Tang, Xia Hu, and Shuiwang Ji. Xggn: Towards model-level explanations of graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 430–438, 2020.
- [139] Dongsheng Luo, Wei Cheng, Dongkuan Xu, Wenchao Yu, Bo Zong, Haifeng Chen, and Xiang Zhang. Parameterized explainer for graph neural network. *Advances in neural information processing systems*, 33:19620–19631, 2020.
- [140] Han Xuanyuan, Pietro Barbiero, Dobrik Georgiev, Lucie Charlotte Magister, and Pietro Liò. Global concept-based interpretability for graph neural networks via neuron analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10675–10683, 2023.
- [141] Qianwen Wang, Kexin Huang, Payal Chandak, Marinka Zitnik, and Nils Gehlerborg. Extending the nested model for user-centric xai: A design study on gnn-based drug repurposing. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1266–1276, 2022.
- [142] Phillip E Pope, Soheil Kolouri, Mohammad Rostami, Charles E Martin, and Heiko Hoffmann. Explainability methods for graph convolutional neural networks. In *Pro-*

- ceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10772–10781, 2019.
- [143] Sara Hooker, Dumitru Erhan, Pieter-Jan Kindermans, and Been Kim. A benchmark for interpretability methods in deep neural networks. *Advances in neural information processing systems*, 32, 2019.
 - [144] Giles Hooker and Lucas Mentch. Please stop permuting features: An explanation and alternatives. *arXiv preprint arXiv:1905.03151*, 2, 2019.
 - [145] Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. Out-of-distribution generalization on graphs: A survey. *arXiv preprint arXiv:2202.07987*, 2022.
 - [146] Siqi Miao, Mia Liu, and Pan Li. Interpretable and generalizable graph learning via stochastic attention mechanism. In *International Conference on Machine Learning*, pages 15524–15543. PMLR, 2022.
 - [147] Ying-Xin Wu, Xiang Wang, An Zhang, Xiangnan He, and Tat-Seng Chua. Discovering invariant rationales for graph neural networks. *arXiv preprint arXiv:2201.12872*, 2022.
 - [148] Katherine Faust. A puzzle concerning triads in social networks: Graph constraints and the triad census. *Social Networks*, 32(3):221–233, 2010.
 - [149] J. Ugander, L. Backstrom, and J. Kleinberg. Subgraph frequencies: Mapping the empirical and extremal geography of large graph collections. In *Proceedings of the International Conference on World Wide Web*, page 1307–1318, 2013.
 - [150] Oh-Hyun Kwon, Tarik Crnovrsanin, and Kwan-Liu Ma. What would a graph look like in this layout? A machine learning approach to large graph visualization. *IEEE Trans. Vis. Comput. Graph.*, 24(1):478–488, 2018. doi: 10.1109/TVCG.2017.2743858. URL <https://doi.org/10.1109/TVCG.2017.2743858>.

- [151] Nino Shervashidze, SVN Vishwanathan, Tobias Petri, Kurt Mehlhorn, and Karsten Borgwardt. Efficient graphlet kernels for large graph comparison. In *Artificial intelligence and statistics*, pages 488–495. PMLR, 2009.
- [152] Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, and Shuiwang Ji. On explainability of graph neural networks via subgraph explorations. In *International conference on machine learning*, pages 12241–12252. PMLR, 2021.
- [153] Thorben Funke, Megha Khosla, and Avishek Anand. Hard masking for explaining graph neural networks. 2020.
- [154] Michael Sejr Schlichtkrull, Nicola De Cao, and Ivan Titov. Interpreting graph neural networks for {nlp} with differentiable edge masking. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=WznmQa42ZAx>.
- [155] Qiang Huang, Makoto Yamada, Yuan Tian, Dinesh Singh, and Yi Chang. Graphlime: Local interpretable model explanations for graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2022.
- [156] Yue Zhang, David Defazio, and Arti Ramesh. Relex: A model-agnostic relational model explainer. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 1042–1049, 2021.
- [157] Zhaoning Yu and Hongyang Gao. Motifexplainer: a motif-based graph neural network explainer. *arXiv preprint arXiv:2202.00519*, 2022.
- [158] Zipeng Liu, Yang Wang, Jürgen Bernard, and Tamara Munzner. Visualizing graph neural networks with corgie: Corresponding a graph to its embedding. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2500–2516, 2022.
- [159] Benjamin Sanchez-Lengeling, Jennifer Wei, Brian Lee, Emily Reif, Peter Wang, Wesley Qian, Kevin McCloskey, Lucy Colwell, and Alexander Wiltchko. Evaluating

- attribution for graph neural networks. *Advances in neural information processing systems*, 33:5898–5910, 2020.
- [160] Mohit Bajaj, Lingyang Chu, Zi Yu Xue, Jian Pei, Lanjun Wang, Peter Cho-Ho Lam, and Yong Zhang. Robust counterfactual explanations on graph neural networks. *Advances in Neural Information Processing Systems*, 34:5644–5655, 2021.
- [161] Ana Lucic, Maartje A Ter Hoeve, Gabriele Tolomei, Maarten De Rijke, and Fabrizio Silvestri. CF-GNNexplainer: Counterfactual explanations for graph neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 4499–4511. PMLR, 2022.
- [162] Christopher Frye, Damien de Mijolla, Tom Begley, Laurence Cowton, Megan Stanley, and Ilya Feige. Shapley explainability on the data manifold. In *International Conference on Learning Representations*, 2020.
- [163] Ying-Xin Wu, Xiang Wang, An Zhang, Xia Hu, Fuli Feng, Xiangnan He, and Tat-Seng Chua. Deconfounding to explanation evaluation in graph neural networks. *arXiv preprint arXiv:2201.08802*, 2022.
- [164] Juntao Tan, Shijie Geng, Zuohui Fu, Yingqiang Ge, Shuyuan Xu, Yunqi Li, and Yongfeng Zhang. Learning and evaluating graph neural network explanations based on counterfactual and factual reasoning. In *Proceedings of the ACM Web Conference 2022*, pages 1018–1027, 2022.
- [165] Ziheng Chen, Fabrizio Silvestri, Jia Wang, Yongfeng Zhang, Zhenhua Huang, Hongshik Ahn, and Gabriele Tolomei. Grease: Generate factual and counterfactual explanations for gnn-based recommendations. *arXiv preprint arXiv:2208.04222*, 2022.
- [166] Pinar Yanardag and SVN Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1365–1374, 2015.

- [167] Anne Fleissig, Valerie Jenkins, Susan L Catt, and Lesley Fallowfield. Multidisciplinary teams in cancer care: are they effective in the uk? *The Lancet. Oncology*, 7 11:935–43, 2006. URL <https://api.semanticscholar.org/CorpusID:45715603>.
- [168] M Smits, M Zegers, P P Groenewegen, D R M Timmermans, L Zwaan, G van der Wal, and C Wagner. Exploring the causes of adverse events in hospitals and potential prevention strategies. *Qual. Saf. Health Care*, 19(5):e5, oct 2010.
- [169] Michael L Barnett, Bruce E Landon, A James O’Malley, Nancy L Keating, and Nicholas A Christakis. Mapping physician networks with self-reported and administrative data. 46(5):1592–1609, oct 2011.
- [170] Xi Zhu, Shin-Ping Tu, Daniel Sewell, Nengliang Aaron Yao, Vimal Mishra, Alan Dow, and Colin Banas. Measuring electronic communication networks in virtual care teams using electronic health records access-log data. *Int. J. Med. Inform.*, 128:46–52, aug 2019.
- [171] Zhengxing Huang, Wei Dong, Huilong Duan, and Jiquan Liu. A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records. *IEEE Transactions on Biomedical Engineering*, 65(5):956–968, 2017.
- [172] Ali Amirahmadi, Mattias Ohlsson, and Kobra Etminani. Deep learning prediction models based on ehr trajectories: A systematic review. *Journal of biomedical informatics*, 144:104430, 2023.
- [173] Charlotte A Nelson, Atul J Butte, and Sergio E Baranzini. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nature communications*, 10(1):3045, 2019.
- [174] Xi Zhu. Connected cancer care: Ehr communication networks in virtual cancer care teams - final report. Final report, Agency for Healthcare Research and Quality, Rockville, MD, 2022. (Prepared by the University of Iowa under Grant No. R21 HS026075).

- [175] Nengliang Yao, Xi Zhu, Alan Dow, Vimal K Mishra, Allison Phillips, and Shin-Ping Tu. An exploratory study of networks constructed using access data from an electronic health record. *J. Interprof. Care*, 32(6):666–673, nov 2018.
- [176] Hsiao-Ying Lu, Yiran Li, and Kwan-Liu Ma. Ehrflow: A visual analytics approach to studying healthcare professionals’ communication effectiveness and efficiency. In *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 120–131, 2024. doi: 10.1109/CHASE60773.2024.00021.
- [177] Antonio Longa, Veronica Lachi, Gabriele Santin, Monica Bianchini, Bruno Lepri, Pietro Lio, Franco Scarselli, and Andrea Passerini. Graph neural networks for temporal graphs: State of the art, open challenges, and opportunities. *arXiv preprint arXiv:2302.01018*, 2023.
- [178] Palash Goyal, Nitin Kamra, Xinran He, and Yan Liu. Dyngem: Deep embedding method for dynamic graphs, 2018. URL <https://arxiv.org/abs/1805.11273>.
- [179] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. *Proceedings of the 13th International Conference on Web Search and Data Mining*, 2020. URL <https://api.semanticscholar.org/CorpusID:210884001>.
- [180] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, and Charles E Leiserson. Evolvegc: Evolving graph convolutional networks for dynamic graphs. *AAAI*, 2020.
- [181] Hsiao-Ying Lu and Kwan-Liu Ma. Associating healthcare teamwork with patient outcomes for predictive analysis, 2025. URL <https://arxiv.org/abs/2512.03296>.
- [182] Yuhua Huang, Hsiao-Ying Lu, and Kwan-Liu Ma. Modeling and interpreting teamwork dynamics in cancer care outcome prediction. In *Proceedings of the 10th International Conference on Medical and Health Informatics (ICMHI)*, 2026. to appear.

- [183] Dongsheng Luo, Wei Cheng, Dongkuan Xu, Wenchao Yu, Bo Zong, Haifeng Chen, and Xiang Zhang. Parameterized explainer for graph neural network, 2020. URL <https://arxiv.org/abs/2011.04573>.
- [184] Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, and Shuiwang Ji. On explainability of graph neural networks via subgraph explorations, 2021. URL <https://arxiv.org/abs/2102.05152>.
- [185] Wenwen Xia, Mincai Lai, Caihua Shan, Yao Zhang, Xinnan Dai, Xiang Li, and Dongsheng Li. Explaining temporal graph models through an explorer-navigator framework. In *International Conference on Learning Representations (ICLR)*, feb 2023. ICLR 2023 poster, published on OpenReview.
- [186] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in graph neural networks: A taxonomic survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(5):5782–5799, may 2023. ISSN 0162-8828. doi: 10.1109/TPAMI.2022.3204236. URL <https://doi.org/10.1109/TPAMI.2022.3204236>.
- [187] Hsiao-Ying Lu, Yiran Li, Ujwal Pratap Krishna Kaluvakolanu Thyagarajan, and Kwan-Liu Ma. Gnnanatomy: Rethinking model-level explanations for graph neural networks. *arXiv preprint arXiv:2406.04548*, 2024.
- [188] Qiang Huang, Makoto Yamada, Yuan Tian, Dinesh Singh, Dawei Yin, and Yi Chang. Graphlime: Local interpretable model explanations for graph neural networks, 2020. URL <https://arxiv.org/abs/2001.06216>.
- [189] Zhihua Jin, Yong Wang, Qianwen Wang, Yao Ming, Tengfei Ma, and Huamin Qu. Gnnlens: A visual analytics approach for prediction error diagnosis of graph neural networks. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 2020. URL <https://api.semanticscholar.org/CorpusID:238215521>.
- [190] Zipeng Liu, Yang Wang, Jürgen Bernard, and Tamara Munzner. Visualizing graph neural networks with corgie: Corresponding a graph to its embedding. *IEEE*

Transactions on Visualization and Computer Graphics, 28(6):2500–2516, 2022. doi: 10.1109/TVCG.2022.3148197.

- [191] Bum Chul Kwon, Min-Je Choi, Joanne Taery Kim, Edward Choi, Young Bin Kim, Soonwook Kwon, Jimeng Sun, and Jaegul Choo. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, 25(1): 299–309, January 2019. ISSN 2160-9306. doi: 10.1109/tvcg.2018.2865027. URL <http://dx.doi.org/10.1109/TVCG.2018.2865027>.
- [192] Colon Cancer Treatment (PDQ®) — cancer.gov. <https://www.cancer.gov/types/colorectal/hp/colon-treatment-pdq>. [Accessed 22-03-2026].
- [193] Ainhoa Costas-Chavarri, Govind Nandakumar, Sarah Temin, Gilberto Lopes, Andres Cervantes, Marcia Cruz Correa, Rena Engineer, Chisato Hamashima, Gwo Fuang Ho, Fidel David Huitzil, Mona Malekzadeh Moghani, Ala I Sharara, Mariana C Stern, Catherine Teh, Sara E Vázquez Manjarrez, Azmina Verjee, Rhonda Yantiss, and Manish A Shah. Treatment of patients with early-stage colorectal cancer: ASCO resource-stratified guideline. *J. Glob. Oncol.*, 5(5):1–19, feb 2019.
- [194] Sakti Chakrabarti, Carrie Y Peterson, Deepika Sriram, and Amit Mahipal. Early stage colon cancer: Current treatment standards, evolving paradigms, and future directions. *World J. Gastrointest. Oncol.*, 12(8):808–832, aug 2020.
- [195] Dong Peng, Yu-Xi Cheng, and Yong Cheng. Improved overall survival of colorectal cancer under multidisciplinary team: A meta-analysis. *Biomed Res. Int.*, 2021(1): 5541613, may 2021.
- [196] Thomas M Whittaker, Mohamed E G Abdelrazek, Aran J Fitzpatrick, Joseph L J Froud, Jack R Kelly, Jeremy S Williamson, and Gethin L Williams. Delay to elective colorectal cancer surgery and implications for survival: a systematic review and meta-analysis. *Colorectal Dis.*, 23(7):1699–1711, jul 2021.

- [197] Yuchong Yang, Yao Lu, Hui Tan, Ming Bai, Xia Wang, Shaohua Ge, Tao Ning, Le Zhang, Jingjing Duan, Yansha Sun, Rui Liu, Hongli Li, Yi Ba, and Ting Deng. The optimal time of starting adjuvant chemotherapy after curative surgery in patients with colorectal cancer. *BMC Cancer*, 23(1):422, may 2023.
- [198] Peng Gao, Xuan-Zhang Huang, Yong-Xi Song, Jing-Xu Sun, Xiao-Wan Chen, Yu Sun, Yu-Meng Jiang, and Zhen-Ning Wang. Impact of timing of adjuvant chemotherapy on survival in stage III colon cancer: a population-based study. *BMC Cancer*, 18(1):234, mar 2018.
- [199] Stephen P Borgatti and Daniel S Halgin. Analyzing affiliation networks. *The Sage handbook of social network analysis*, 1:417–433, 2011.
- [200] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.
- [201] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Cambridge university press, 1994.