

Visual Analysis Methods for the Exploration of Static and Dynamic Networks

By

SUGEERTH MURUGESAN
B.Tech. (Amrita University) 2013
M.S. (UC Davis) 2015

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:

Professor Bernd Hamann, Chair

Associate Adjunct Professor Gunther Weber

Adjunct Professor Kristofer Bouchard

Committee in Charge

2017

Copyright © 2017 by
Sugeerth Murugesan
All rights reserved.

Dedicated to my parents Murugesan and Bharathi who have sacrificed their happiness to see me succeed.

Contents

List of Figures	vii
List of Tables	xi
Abstract	xii
Acknowledgments	xiv
1 Introduction	1
1.1 Context	1
1.2 Motivation	3
1.3 Research Questions	6
1.4 Contributions	7
1.5 Outline	7
1.5.1 Exploration of Static Networks	9
1.5.2 Exploration of Dynamic Networks	9
1.6 Conclusion	10
2 Background	11
2.1 Brain Imaging Modalities	12
2.1.1 Functional Connectivity	14
2.1.2 Electrocortigography	15
2.2 Visual Analytics	16
2.2.1 Static Networks	16

2.3	Dynamic Networks	20
2.3.1	Communities for Dynamic Graphs	20
2.3.2	Spatio-temporal Data	21
2.3.3	ECoG Data	22
2.3.4	Dynamic Network Simplification	22
2.3.5	Difference Graph Framework	23
2.4	Summary	26
3	Interactive Hierarchical Analysis of Functional Brain Connectivity	27
3.1	Introduction	27
3.2	Brain Modulyzer	31
3.2.1	Anatomical Views	32
3.2.2	Connectivity Views	35
3.2.3	Linking and Interactivity	35
3.2.4	Community Detection and Visualization	35
3.2.5	Dendrogram View	39
3.2.6	Network Measure Table	40
3.2.7	Intra- and Inter-community Analysis	42
3.2.8	Focussing on Data Subsets	45
3.3	Results	45
3.3.1	Case Study 1: Progressive Supranuclear Palsy	45
3.3.2	Case Study 2: Memory Encoding and Retrieval	47
3.4	Conclusions and Future Work	50
4	Multi-scale Visual Analysis of Time-varying ECoG Data Via Clustering	51
4.1	Introduction	51
4.2	Motivation	51
4.3	ECoG ClusterFlow	53
4.4	Cluster detection	56
4.5	System overview	56

4.5.1	Cluster Evolution View	57
4.5.2	Electrode View	60
4.6	Case Studies	65
4.6.1	Synthetic Dataset	65
4.6.2	Epileptic Seizure Dataset	67
4.7	Discussion	70
4.8	Conclusions	73
5	Similarity and Difference Visualization of Dynamic Networks	76
5.1	Introduction	76
5.2	Motivation	78
5.3	Design Goals	79
5.4	Method	80
5.4.1	State-based Similarity	80
5.4.2	States Defined by a Similarity Matrix	81
5.5	Consensus-based State Summary Graphs	82
5.5.1	Our Approach	83
5.6	Visual Representation of the Summary Graph	86
5.7	Difference Graph Framework	90
5.7.1	Community-based Difference Graph Visual Design	90
5.8	McFarland’s Classroom Data Set	91
5.9	TimeSum Used in Neuroscience	94
5.9.1	Case Study 1	95
5.9.2	Case Study 2	100
5.9.3	ESnet: Analyzing Router Topology	104
5.10	Discussion	105
5.11	Comparison with Other Techniques	108
5.12	Conclusions and Future Work	109
6	Context-aware Dynamic Network Analysis for Network Flow Data	110

6.1	Related Work	112
6.1.1	Network Visual Analytics	113
6.1.2	Difference Graphs	114
6.2	Network Flow Data	115
6.3	Goals	115
6.4	Method	117
6.5	Context-aware Difference Graphs	117
6.5.1	Mathematical Formulation	117
6.5.2	Difference Graph Computation	119
6.5.3	Difference Graph Computation	119
6.6	Application Example	124
6.6.1	ESNet Dataset	124
6.7	Conclusion and Outlook	125
7	Conclusion and Future Work	127
7.0.1	Visual Analysis Techniques	128
7.1	Future Directions	128
7.1.1	Simplistic Visualizations	128
7.1.2	Temporal Aspects	128
7.1.3	Network-Specific Parameter Optimizations	129
7.1.4	Multi-modal Visual Analytics	129
7.1.5	Models of Visual Thinking	129
7.2	Conclusion	129

List of Figures

1.1	Prototypes for dissertation	8
2.1	Structure of background section	12
2.2	Different recording methods used to control BMIs	12
2.3	Current state of art in neuro-imaging analysis	13
2.4	Framework for visual analysis tools for graph and network data. This dissertation addresses several components of this framework, split into four major modules, <i>i.e.</i> , Data Space, Analysis Space, Visual Transformation, Visual Exploration	25
3.1	System multiple views showing different data aspects	29
3.2	Typical analysis and visualization methods used for fMRI connectivity data . .	30
3.3	Our system uses the connectivity matrix and its associated parcellation as input. The system operates in two modes—correlation mode and community mode. While the correlation mode allows a user to analyze correlation network data, the Community mode allows a user to perform tasks for modular analysis. . . .	31
3.4	The matrix view	36
3.5	Modules of the system	38
3.6	Modules of the system	40
3.7	Dendrogram View	41

3.8	Functional connectivity of the mesothalamic junction (MTJ) network in 36 healthy controls subjects and 18 PSP patients.	44
3.9	Anatomical and graph views of the memory consolidation network	48
4.1	Overview of the ECoG ClusterFlow pipeline.	55
4.2	Comparison of tracking algorithms in artificial datasets. In image C the maximum overlap algorithm pairs $C_{3,3}$ to $C_{4,1}$, while in Fig. D the globally optimal matching algorithm pairs $C_{3,3}$ to $C_{4,3}$, qualitatively making communities more visually traceable in image D. The scalar values for the links in image A and B are $L_{1,1} = 11$, $L_{1,2} = 11$, $L_{2,1} = 10$, $L_{2,2s} = 2$	58
4.3	Evolution of clusters for four timesteps.	58
4.4	To save display space, our tool crops the electrode view to the region of the brain where electrodes are placed.	61
4.5	Dynamic spatial patterns captured by the electrode views. The colors of the glyphs represent clusters, and the opacity of the glyphs denote electrical activation. The orange cluster in the top, spatially distributes, while the green cluster in the bottom, spatially combines. The yellow cluster in the top activates over time.	61
4.6	Design choices for visualizing cluster membership along with brain electrical activity	62
4.7	Varying the number of timesteps displayed on our glyph design.	62
4.8	Layout for bottom-up analysis of the spatio-temporal data.	66
4.9	The figure shows evolution patterns underlying our generated dataset which we use to test and evaluate our approach	67
4.10	Hierarchical exploration of the seizure dataset at varying levels of granularity	72
4.11	Visualizations of the two pre-processed datasets	74
5.1	Visualizations of a dynamic brain network including summary and difference representations.	77

5.2	The algorithm used to construct the summary graph identifies the most common local and global topological attributes from a set of similar graphs	85
5.3	Our methods applied to a real brain network dataset	87
5.4	We investigate different visual designs to convey uncertainty and variability in the dynamic network data	89
5.5	The two dominant states in the McFarland dataset.	93
5.6	Exploration of <i>posteromedial cortex</i> , node 4, in an integrated state from timesteps 42-54 through the summary graph	97
5.7	We create abstractions of the brain network dataset, considering nine extracted states with similar global and local properties.	98
5.8	Similarity Metric	99
5.9	Similarity and difference topology of network flow	102
5.10	Similarity and difference topology of network flow	103
5.11	Our technique reduces the time-varying data set into summary graphs.	107
6.1	A context-aware difference graph framework showing how the change in graphs affect the dynamic topological structure of the graph.	112
6.2	The figure explains our analysis pipeline that explores the complex structures embedded in dynamics graphs. In the first stage, using the community similarity metrics, we define similar graphs and group them into graph structures. We identify specific local affecting between important time-points. In the third stage, we link the detected contextual topological structure with other detected structures over other timepoints.	116
6.3	The graphs and their community structure are juxtaposed community structure. Our topological-effect algorithm is able to detect the structures that are affected based on the selected node and the change in the difference graph.	118
6.4	Direction-aware modeling of dynamic graphs they diffuse over time.	120
6.5	A small difference in the edges between green nodes induces a topological change, causing a community change between the green and red node.	121

6.6	Color representing the community that a national laboratory/hub belongs to. X-axis the time-scale and the Y-axis the different labs. The detection algorithm found communities within the same spatial locations.	122
6.7	Topological changes happening between two timesteps. The larger the node size is the larger is a change in the laboratory/hubs. Blue dashed lines represent the reduction in network flow and red lines represent the addition of flow. Although significant change happens at each timestep, one can see a core-structure forming between IAR, LSVN, and NSO. They remain stable and then change their structure.	123
6.8	This represents the change happening at WASH. In comparison to all other nodes, this node exhibits substantial a lot of transient activity with sites surrounding the laboratories.	125

List of Tables

1.1	Contributions	7
5.1	Statistical comparison of properties of the detected states.	94

ABSTRACT

Visual Analysis Methods for the Exploration of Static and Dynamic Networks

Representation of networks is fundamental for the understanding of large, complex systems like brains, autonomous systems, and social networks. Major complex systems can be modeled as a network by representing their nodes as real-world objects and edges the relationship between them. In neuroscience, for example, a core component of analysis involves modeling the connectivity of human brain into a network, where the nodes represent brain regions and the edges, the correlation among them. Visualization and statistical analysis of the resulting network data have provided tremendous value such as identifying diagnostic indicators to neuro-degenerative diseases and age-related disruptions in the brain. Interactive visual analytics systems exploring such network data have further provided a powerful means for analysts to investigate complex configurations of the modeled network in real-time. Many of the ways the human connectome is understood are through the intuitive insights that these systems enable. However, existing methods primarily focus on static networks but not satisfactorily visually explore the structural, hierarchical, semantic and time-varying properties of the network. The comprehensive understanding of such properties and their complex dependencies is crucial to holistically understand the structure of networks.

This dissertation contributes novel methods that explore the complex and multi-modal datasets in neuroscience, autonomous systems, and social science domains. First, we focus on static network visual analysis method that allows us to interactively explore the hierarchical modular structure of an fMRI-based brain dataset. Second, we contribute novel designs and a consensus-based clustering framework that analyzes the time-varying modular structure of the ECoG-based brain dataset for an epilepsy dataset. Third, we contribute a novel visual analysis framework that identifies similarities and differences between dynamic graphs for intuitive analysis of temporal states in a time-varying network. Fourth, we focus on developing an algorithm that reveals the topological change and the effect of change happening for a given time interval for an autonomous system dataset.

The dissertation concludes with critical thoughts and guidelines based on the experiences

obtained from performing and implementing the aforementioned methods. High-level concluding ideas are discussed in network analysis, neuroscience and information visualization.

ACKNOWLEDGMENTS

My doctoral studies have been a great experience for me. They have equipped me with major skills that I feel confident to face most uncertainties in life. I was fortunate enough to receive the support of my parents, my advisors, and my friends. Their thoughts and work have deeply inspired me to chart new boundaries with huge confidence. I would like to thank all of them who were a part of this journey. First, I would like to thank my advisor Bernd Hamann. When I was still an undergraduate student, he believed in me and guided me up to this point. All of this started because he responded to my email (which I had sent from India) and offered to work with me a graduate student of his to work some of the most complex scientific visualization problems.

I further would like to thank Gunther Weber, without whom the successful completion of my Ph.D. would never have been possible. His thoughtful guidance has made me cross storms and rivers in academia with much ease. His trust and belief in me have enabled me to focus on contributing the best work that I could ever produce.

I was very fortunate to have Kristofer Bouchard as my external advisor from the neuroscience community. Most of my work was inspired by his view of thinking. The feedback from him and his group were invaluable for improving the quality and standard of my work.

Further, I thank Jesse Brown who guided me early on in my program and exposing me to the art of computational analysis of neural data. Special thanks to my collaborators Edward Chang, Andrew Trujillo, William Seeley and Max Dougherty. At UC Davis and Lawrence Berkeley National Lab (LBNL), I had great the opportunity to take classes and collaborate with other talented researchers, visiting scholars, postdocs, and graduate students. I also would like to thank my friend Harika Sabbella who supported me in many important ways.

Several other people have inspired me and my research. Finally, there are those who deserve the most thanks. I am deeply grateful to my parents, Murugesan Subburam and Bharathi Murugesan, who have always encouraged me to pursue my dreams.

Chapter 1

Introduction

“An approximate answer to the right problem is worth more than an exactly right approach to an approximate problem”

- John Tukey, *Exploratory Data Analysis* [183]

Visual Analysis is Detective Work: A detective investigating a crime needs both understanding and sophisticated tools to carry out complex tasks. The purpose of this dissertation is to provide new visual analytics algorithms and methods that advance the state of the art for carrying out detective work in exploring the organizational properties of complex networks.

1.1 Context

Modern neuroimaging technologies like ECoG, fMRI generate increasingly detailed, high-resolution data of the functional and anatomical connection patterns of the human brain. Other co-existing technological advancements in sensor technologies have similarly generated large connectivity datasets in social, technological and biological fields. Such vast scales of datasets have provided many opportunities for scientists and researchers to perform exploratory and task-based analysis of understanding the communication patterns of the brain and the social structure of the society at a scale unimaginable a few years ago.

Major attempts to characterize such datasets have led to a large multidisciplinary field called complex systems [153]. This unique approach to understanding the data describing the impor-

tant properties of complex systems by analyzing the topographies of the network representations has its origins in the mathematical study of networks such as graph theory.

Correspondingly, the increase in the computation power provides immense opportunities for researchers to develop large-scale data-mining algorithms and tools that explore the complex dataset. While most of these algorithms and statistical techniques are applicable for data and domains that are well-understood or are well-defined, most existing real-world analysis requirements are still exploratory or ill-specified in nature. Such analysis requires methods that explore such complex data beyond traditional formal modeling or hypothesis task. Such methods should ideally summarize the main characteristics, reveal outliers, patterns and be online or interactive in nature.

An approach to solve the aforementioned problems could be a technique known as exploratory data analysis [183]. Such analysis aims to investigate and generate hypotheses on-the-fly, assess limitations and assumptions, and glean and iterate insights based on the evidence of new information. While basic statistical analysis and algorithms quantify major patterns in the datasets, complementary visualization techniques help identify outliers, trends, patterns that other methods may have missed.

To fully comprehend the complex dynamics underlying the connectivity data in complex networks, one can use two major complementary exploratory approaches for analysis: *quantitative analysis* involving statistical measures for empirical investigation and *qualitative analysis* requiring design methodologies for visual depiction. The former takes advantage of our analytical type of thinking for problem-solving while the latter focuses on our visual aspect of information retrieval. While numerous approaches have focused on the quantitative aspects of data analysis, little attention is given to the qualitative aspects of network analysis.

Visual Perception is an important tool of our visual system that has evolved over many decades of successive generations. The purpose of such a system is to interpret and assimilate information from the surrounding environment, making us more aware of our situations. The processing of such information is usually done in parallel and is carried out through a huge information bandwidth reaching the ultimate decision making processes in our brain. Information Visualization is the field in computer science that takes advantage of such a perceptual system

of the brain for effective visual communication of information. An effective visualization truthfully reveals important patterns and outliers in the data without genuine loss of information. To design such methods one needs domain expertise of a particular problem one is solving, algorithmic techniques to organize and depict effective visual elements and most importantly an irrational way of thinking to effectively communicate important aspects of data.

Most of the successful scientific discoveries in network neuroscience is a result of a combination of quantitative and qualitative data analysis. The combination is a very powerful tool for analysts to explore and discover novel insights about the data. This dissertation contributes new methods concerning the visual aspects of quantitative and qualitative methodologies of data exploration revealing new insights in the context of complex networks.

1.2 Motivation

Brain connectivity datasets usually describe the connectivity between neuronal elements through tracts or functional associations. Such data can be categorized in terms of functional or structural relatedness.

More specifically, the two types of connectivity datasets include—the *anatomical connectivity* describing the axonal fibers across cortical regions and *functional connectivity* describing the observed statistical correlations between anatomically segregated brain regions [88]. Although the analysis methods comprehending both the datasets differ, these complex system can be modeled as a simple weighted graph where the nodes represent neuronal substrates of brain regions and the edges depicting the strength of correlation (functional connectivity) or the density of fibers connecting them (structural connectivity). In the context of functional magnetic resonance imaging (fMRI) connectivity, a non-invasive connectivity data acquisition procedure, statistical correlations between brain regions are measured when the subject is performing a task or is at rest. fMRI data has enabled researchers to investigate the intrinsic connectivity network of each subject. While structural or physical connectivity refers to the actual physical connections (like anatomical tracts) between brain regions

In network neuroscience, the communication behavior observed between brain regions is of particular interest when performing specific tasks. Earlier work [39] has investigated net-

work scientific evidence that depicts changes in the functional organization of networks for subjects when performing a certain task versus post-completion of a task during rest. Further, the altered network configuration is also evident for subjects diagnosed with neuro-degenerative diseases. The understanding of such disease mechanisms could be advanced by moving beyond the assessment of atrophy in affected brain areas to a model of disease progression throughout vulnerable brain systems.

Brain connectivity encompasses multiple spatial scales, ranging from individual neurons to entire brain regions and systems [171]. When studying large-scale brain organization, the analysis is often focused on high-level, functional connectivity between brain regions. An example is the study of the network organization of the left and right *motor cortex* [29]. To understand such phenomena for a human connectome, *i.e.*, a map of the neural connections in the brain, data is collected from multiple imaging modalities both invasively and non-invasively. One main goal of the connectomic analysis is to construct an entire representation of the brain that will facilitate an intuitive understanding of how functional brain states emerge from anatomical representations.

When considering functional networks, graph-theoretic methods has shed more light into the organization of the brain enabling succinct understanding of the overall cognitive functioning of the brain. Neuroscience has established the existence of a relatively small number of *intrinsic connectivity networks* (less than 20) that have a dense set of intranetwork connections and a relatively sparse set internetwork connections. These networks underlie specific processes including lower- and higher-order vision, hearing, sensory-motor processing, spatial processing, external attention, saliency processing, and internal mentation [167, 137]. Graph-theoretic techniques, such as community detection [30], can be used to find these intrinsically connected networks or find those hubs that bind multiple communities. Furthermore, brain networks have a hierarchical modular organization. The study of these hierarchical modules has revealed specialized cognitive sub-modules within larger communities. Depending on the cognitive process that the brain engages in, these sub-modules dynamically change their coupling with each other. Furthermore, other properties of the brain that have been proved to exist include—small worldness [55], Rich-club organization [184].

Analyzing fMRI or position emission tomography (PET) data has been a valuable resource in generating numerous scientific hypothesis related to the functioning of the brain, these methods have not been able to delineate the complex spatiotemporal patterns of the brain successfully when examining subjects performing specific tasks. This limitation has given rise to advances in invasive data acquisition procedures like Electrocorticography (ECoG) providing researchers the high-dimensional data with high spatial and temporal resolution. Such data has allowed them to investigate further the activity of the brain about very specific task performed by the subjects. Studies have shown that with such high-dimensional data, we could now deeply understand the functional and dynamic organization of the brain [195, 33]. Studying the hierarchical and modular properties of brain networks is crucial to understand how the brain networks vary across subjects, how the networks change when subjects age [118], and how networks differ under different physiological and neurological disorders [120, 172, 192, 109, 152].

These datasets possessing a rich repertoire of network properties. For a meaningful exploration of such datasets, there is a need for methods that can interactively extract and visually analyze the extracted information. For example, in the analysis of large temporal series of fMRI recordings, visual analysis methods have proven to increase the intuitive pattern-recognition ability in the identification of temporal events based on certain cognitive tasks. To study these relationships as comprehensively as possible, we need more methods for the analysis of complex, functional connectivity data of the human brain that can support an in-depth and modular investigation of fMRI data. Existing analysis methods for functional connectivity are primarily hypothesis-driven or require cumbersome off-line analysis. Moreover, limited tools are available that support these analyses interactively allowing evaluation of analysis algorithms revealing the modules at different levels in the hierarchy.

Further advances in high-resolution data capturing technologies for brain connectivity such as ECoG has further compounded the problem of making sense of this data. The rich repertoire of resolution provided by ECoG needs to be supported with a tool that can analyze complex time-varying patterns. These problems provide new challenges in visualizations that are unique to ECoG [98, 74, 86].

Such exploratory analysis requires novel techniques that can interactively abstract, extract,

condense and expand details on demand. *Interactivity* is a key component of such methods allowing the analyst to perform exploratory querying at an unprecedented scale.

1.3 Research Questions

Motivated by the problems discussed in the above section, the goal of this dissertation is to investigate aspects of visual analysis methods facilitating knowledge discovery and sense-making in the context of complex networks. The high-level questions that we were by motivated were:

- *How do the hierarchical processes interact with each other in complex brain networks? How can we enable multi-level intuitive understanding of hierarchical structure of the brain?*
- *How do abstract functional states in brain networks (clusters) affect epilepsy and evolve over time?*
- *How can we reliably abstract and represent functional states in dynamic brain systems without much genuine information loss?*

The approach that we employ to tackle this higher-level goal is by contributing methods that are a result of a careful consideration of analytical and design thinking for solving a domain-specific network analytic problem.

Analytic methods are based on mathematical formulations mining, abstracting and quantifying hidden relationships in the underlying data while the complementary design methods consist of architecting visual schemes that best communicate the processed data without information loss and not compromising much on aesthetic aspects of data.

As existing methods for network analysis do not satisfactorily explore many organization properties like hierarchical structure exploration, similarity-based abstraction or semantic understanding of real-world elements and their interactions, there is a need to contribute new ones that fully explore the hidden relationships in the data. The impact of such novel methods lies in better characterizations of neuro-degenerative diseases, understanding of cognitive processing in the brain, efficient understanding of functional states in a dynamic network.

Study	Data	Modality	Network	Domain	Cluster
Brain Modulyzer	H	fMRI	St, U	Memory, Neurodege.	St
ClusterFlow	T	ECoG	Dy, U	Epilepsy	Dy
TimeSum	T	fMRI, SS	Dy, U	Salience Network	Dy
NetReduce	T	AS	Dy, D	Interconnect	Dy

Table 1.1. The table mentions the prototypes contributed for a domain specific study. Abbreviations: ‘U’: Undirected, ‘SS’: Social Science, ‘Dy’: Dynamic, ‘AS’: Autonomous Systems, ‘H’: Hierarchical, ‘T’:Temporal, ‘St’:Static.

1.4 Contributions

The contributions of this work (Table 1.3) focuses on different categories: 1) Visual analysis of static networks, 2) Clustering and visualization framework for dynamic networks, 3) Abstraction framework for Dynamic Networks, 4) Context-aware computation framework for dynamic networks.

The main research prototypes introduced in this dissertation (Figure 1.1) fall into one of the aforementioned categories.

Brain Modulyzer presents a linked interactive visual analysis framework to explore the modular hierarchical structures of brain networks in fMRI.

ClusterFlow presents a novel visual analysis method for clustering and visualization of the complex spatio-temporal patterns underlying ECoG data.

Time Sum presents a novel abstraction and difference technique to explore highly scalable temporal data.

NetReduce presents an analysis framework that mines and visualizes important patterns in time-varying autonomous systems.

1.5 Outline

This dissertation consists of two major parts, analysis techniques for static and dynamic networks applied mainly to brain networks. Chapter I and II define the problem, followed by four major studies in this dissertation, each focusing on exploring a specific organizational property of a network. The dissertation concludes with a summary of its main contributions and point

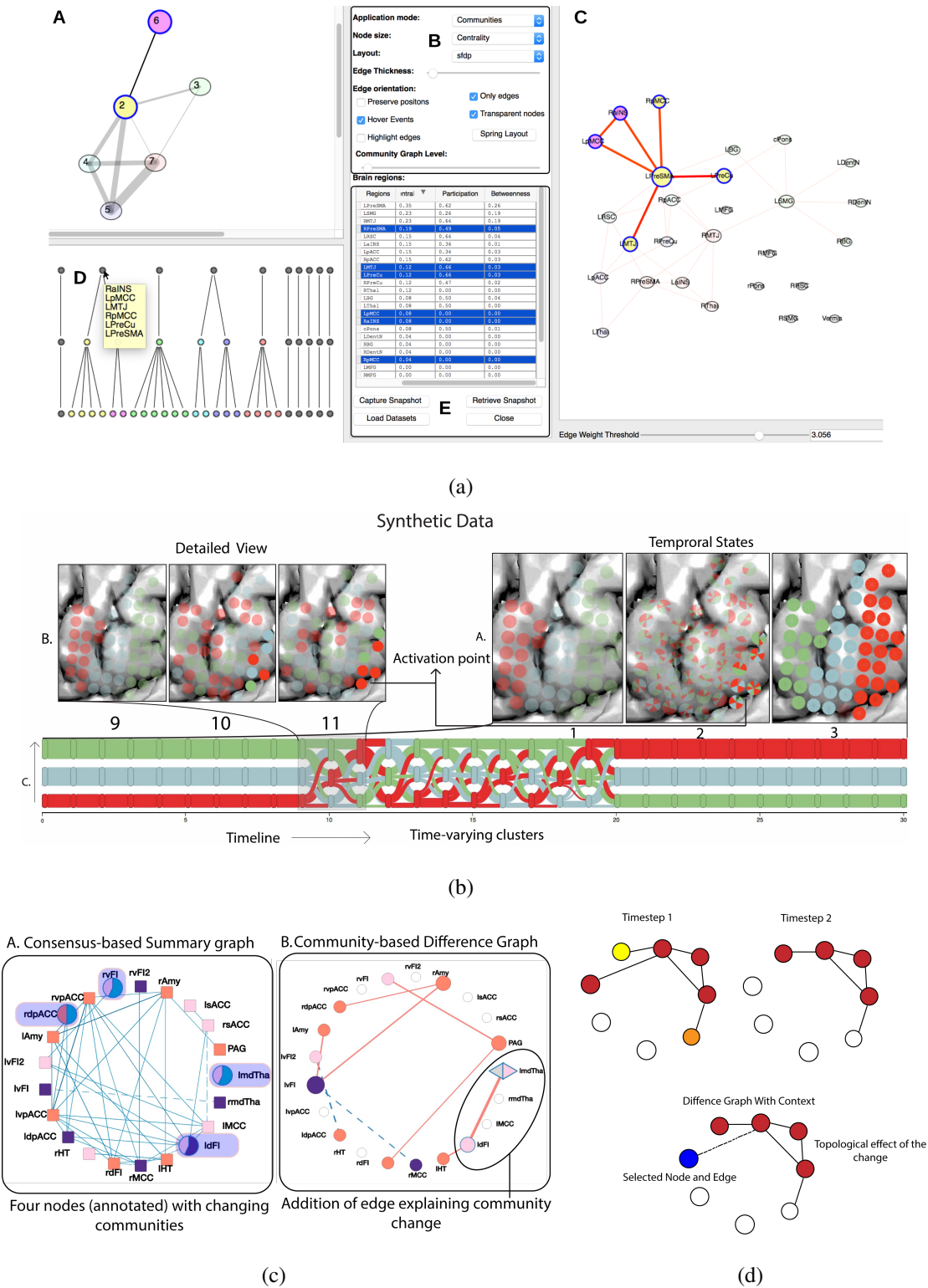


Figure 1.1. Research prototypes that were contributed for the visual analysis of static and dynamic networks.

out possible future research directions.

1.5.1 Exploration of Static Networks

In chapter III, we focus on exploring the hierarchical modular properties of static fMRI brain networks. The contributions include the novel interactive and integrated visual analytic technique that mines and explore a hierarchical structure underlying the network. The amalgamation of such techniques through an intuitive design interface has enabled novel exploratory insights to the neuroscientists closely working on our team. Specifically, the method has enabled use to shed light on the inner workings of the memory system. We explain the design decisions and architecture and the interactive techniques involved in the entire process.

1.5.2 Exploration of Dynamic Networks

Through chapters IV-VII, we explore various aspects of dynamic networks facilitating analysis of time-varying structures. We apply our methods to various modalities within neuroscience, social science, and network science. The major challenge in the visual analysis of dynamic graphs is the characterization of major evolution patterns that cause a major temporal state change in the graph. For example, communities form or die, nodes change their roles over time, becoming more central or secondary. While statistical method mine and quantify higher-level changes in graphs, visualization is used to glean unforeseen low-level patterns in the graph. Furthermore, visualization depicting graphs must show the contextual information to aid the user effectively explore the major evolution patterns in the dataset.

1.5.2.1 Focus on Multi-scale Property

Specifically, in Chapter VI & V, we explore the time-varying modular structure of brain networks applied to epilepsy dataset. Our novel interaction and visual aggregation facilitated understanding and interactive knowledge discovery of Epilepsy dataset. The visual system enabled the neuroscientist in our team to explore major cluster evolution signatures quickly and effectively.

1.5.2.2 Focus on Change Understanding

In chapters VI-VII, we identify the most important ways to understand the change in dynamics happening between subsequent timesteps. We apply our techniques to resting state and social

science datasets. Our unique approach enabled domain experts to interactively explore similarity patterns in dynamic networks.

1.6 Conclusion

In chapter VIII, we close and reflect on the methods contributed. We provide general future directions for research and the important unsolved problems in the visualization of networks, analysis of dynamic networks in neuroscience, social science or distributed computing. We discuss the design lessons learned, visualization models, analysis frameworks to motivate other researchers and practitioners for devising methods that solve related problems.

Chapter 2

Background

“We can only see a short distance ahead, but we can see plenty there that needs to be done”

- Alan Turing, *Quotes*

This dissertation is placed at the intersection of three major fields—*neuroscience*, *visual analytics* and *network science*. We develop novel analysis techniques that in a unique way enhance our understanding of each of these fields. In this chapter, we provide with an overview of the background neuroscience information relevant to understand the importance of the techniques mentioned here. We, illustrate the current imaging technologies in brain connectivity and explain the scientific finding associated with each of these modalities. We explain the existing state of the art in visual analytics methods for network data and specific solutions for neuroscience.

We now describe three major sub-categories (Fig. 2.1) relevant to understand the current state-of-the-art in visual analysis of neuroscience data. The left sub-tree (Imaging Modalities) depicts the rich-repertoire of sensor information recorded from brain networks, while, the right sub-tree (*Network Analytics*) represents the current state-of-the-art in network science visual analysis algorithms that can be applied to the left-subtree.

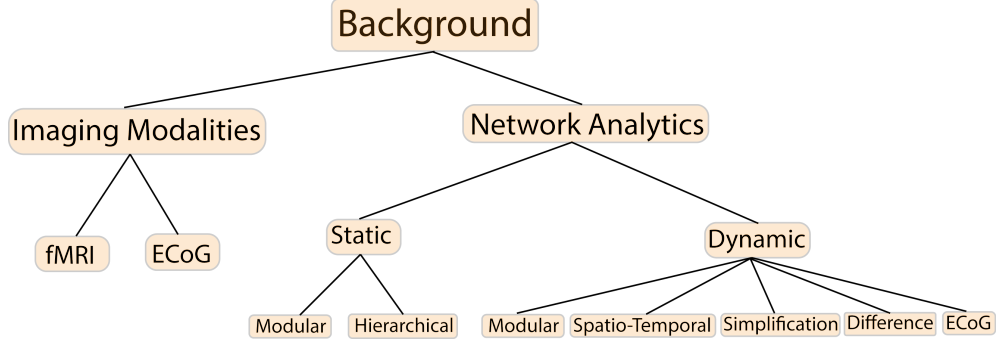


Figure 2.1. The related work section is divided into three major sections—imaging modalities, complex networks and visualization. We describe the state-of-the-art in each of these sub-categories relevant to this dissertation.

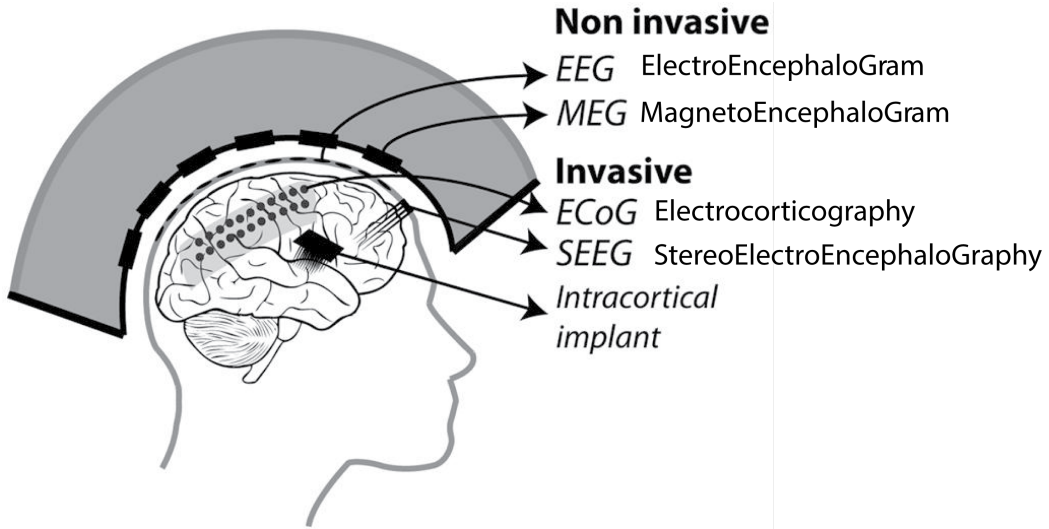


Figure 2.2. Different recording methods used to control BMIs. Invasive methods include: ECoG electrodes placed on the dura, SEEG electrodes placed through the skull into the cortex. Non-invasive methods include: EEG/MEG electrodes placed around the head [17]

2.1 Brain Imaging Modalities

To record complex activation patterns underlying the brain, several recording techniques are introduced. Brain Imaging Modalities recording such high-dimensional information can be described into two major modalities *i.e.*, invasive and non-invasive methods. These categories basically describe the way in which the data is acquired from the brain (refer Fig. 2.2). One of the widely used non-invasive techniques to infer brain activity is through fMRI which leverages from the fact that blood flow and neuronal activation is coupled [71]. Through the indirect measurement of blood flow, one can gain insight into the spatial regions where brain regions are

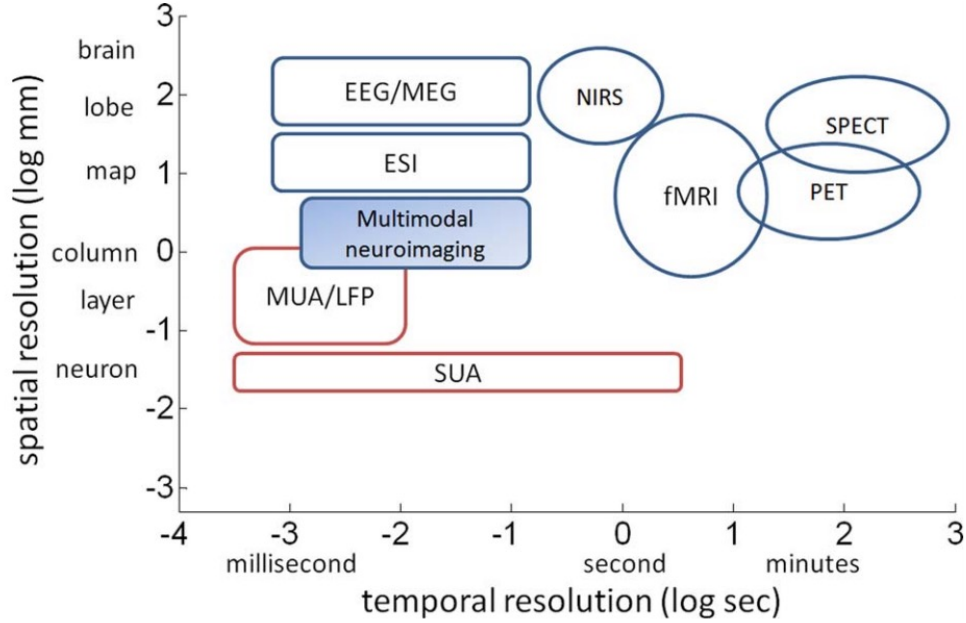


Figure 2.3. Current state of art in the neuro-imaging analysis. Non-invasive (in blue) and invasive (in red) with ranges of spatial and temporal resolutions [80]

activated to perform a particular task.

One of the primary goals of functional neuroimaging is to map the spatially distributed and temporally dynamic neural activity and connectivity with high resolution in space and time domains. Such a mapping will greatly benefit the understanding of how the *cognition* is implemented in the brain, and better understanding the mechanisms of learning, perception, etc [80]

The invasive imaging techniques described in Fig. 2.2 include methodologies that record multi-unit activity (MUA) through electrocorticography (ECoG) and intracranial implants. These technologies have high spatial and temporal resolution but have limited coverage and are invasive in nature. The techniques are applicable to an only select subset of patients (*e.g.*, patients who suffer from epilepsy).

Non-invasive functional brain imaging is an important tool used by researchers in various fields. The important techniques include EEG, MEG, PET, MRI, and fMRI. Each method has its own strengths and weaknesses, and no single method is best suited for all experimental or clinical conditions [78]. We now further provide more details into the invasive and non-invasive modalities of brain research problem statement. As introduced in chapter 1, we categorize

neuroimaging data into two major categories:

- **Functional Imaging** involves imaging the overall functioning of the brain rather than just physical connections. For example, the connections in fMRI are considered as statistical correlations between brain regions.
- **Structural Imaging** deals with imaging the actual structure of the brain, *i.e.*, considering the physical neuronal connections into consideration.

The resulting data from such imaging methods is preprocessed and statistically analyzed to construct a mathematical representation of the inner-workings of the brain. The nodes in such methods represent brain regions and the links anatomical, functional or effective connections [64] depending on the dataset.

2.1.1 Functional Connectivity

The functional network represents cross-correlations between BOLD signals estimated from time series of brain dynamics [153]. Mathematically, brain connectivity based on fMRI analysis is defined in terms of correlation matrices where each matrix entry $c_{i,j}$ encodes the statistical similarity between the time courses of parcellated brain regions i and j .

This data can then be modeled as a 2D node-link diagram where the matrix entry $c_{i,j}$ is encoded by the link between the i -th node and j -th node. In the anatomical view, the brain regions correspond to standard anatomical parcellation data or subject-specific parcellation data. Other imaging modalities used to study connectivity include diffusion tensor imaging (DTI) that map the anatomy of the white matter tracts non-invasively.

Functional connectivity varies with the state of the brain when functional imaging data is acquired. Resting state functional connectivity is particularly useful when studying the connectivity changes due to neurodegenerative diseases [161] or psychiatric disorders [151]. Another application of functional connectivity is to examine the connectivity changes that occur when a test subject performs a task, *e.g.*, mental addition or remembering faces [166].

Furthermore, the human functional brain network is proved to possess many graph theoretic properties like small-worldness, modularity, hierarchical modularity, rich-club organization, and scale-free topology [153, 55, 119]. We focus on the understanding the modular

property of brain networks and its evolution over time.

2.1.1.1 Modularity Analysis in Functional Brain Networks

Modularity in complex networks measures the division of a network into sets of modules where each module possesses dense internal or intra-modular connectivity and sparse external or inter-modular connectivity [130]. Large complex systems like brain networks exhibit modular organization [118] and have a hierarchical network structure at multiple scales [119, 81]. This property of the brain ensures that the network is robust, adaptable and able to evolve [119].

Functional brain networks exhibit hierarchical modular organization[119], *i.e.*, there exist communities within communities. The sub-communities interact with other sub-communities based on the cognitive process of the brain. Several approaches have been proposed [156, 16, 144, 103] to detect communities in complex networks at different scales.

2.1.2 Electrocortigography

In chapters IV and V, we focus on the data recorded from a more invasive monitoring technology, ECoG. Electrocortigography (ECoG) data is derived from individual electrodes implanted on the exposed cortical surface of the brain to record activity (Fig. 2.2). Based on the electrical activity of the individual electrodes, one can . Researchers apply various statistical analysis tools to characterize the structures for the neural activity. The functional connectivity, in this case, is seen as relatedness among specific implanted electrodes that are correlated functionally. As a result, the functional connectivity in ECoG can be represented in terms of a weighted graph.

Such a method is used to capture data at increasing spatial and temporal resolutions and high signal to noise ratio. Currently, in neuroscience, such modalities are regarded as the gold-standard for assessing neuronal activation data for epileptic patients[121]. At such resolutions, existing research and surgical method provide the ability for surgeons to reliably perform presurgical planning and guide the removal of epileptogenic zones in the brain. In our work, we specifically focus on the data collected by such a modality for analysis and visualization (see chapters IV and V).

2.2 Visual Analytics

Visualization of neuroscience data is the cornerstone of communicating results research in cognitive and systems neuroscience [101]. Visual Analytics is a field that was developed as a branch of information visualization, human-computer interaction. The field is regarded as “the science of analysis through interactive visual interfaces.” The origin of the parent field “information visualization” derives from data graphics including statistical plots, maps, diagrams, measurements. The concept of exploratory analysis [183], focuses primarily on the use of visual methods to convey a general-purpose story arising from the underlying data. In addition to static images that clearly depict the underlying information without much information loss, “dynamic” images based on underlying data has enabled multi-part that otherwise would require rigorous offline analysis.

In the context of brain networks, the integration of interfaces and interactive methods into the exploration of complex, high-dimensional data has proven to be crucial in novel scientific discovery [185]. The ability to glean unforeseen patterns in complex network data and perform many queries at real-time, interactive rates has proven to contribute many new discoveries not only in neuroscience but also in social science or computational biology.

In this section, we review the major visual representation techniques closely related to the dissertation. We start with existing state of the art for exploration of functional brain networks for static networks. We focus particularly on visual analysis techniques describing and understanding change in dynamic networks. We eventually move to few more techniques for the exploration of network data.

2.2.1 Static Networks

A deep understanding of network data is crucial to the analysis of scientific neuroscience datasets that better explore and mine the underlying data. Consequently, there has been numerous research in best representing the data such that it best facilitates novel insights which may have been hard to glean from basic techniques.

Graph-based analysis of brain networks has given insight into brain network topology [39]. Results discussed by Crossley *et al.* [47]. indicate that central hub regions in brain networks,

with high participation coefficients, are critical for orchestrating inter-modular communication. Lesions in hub nodes are a common feature across a broad spectrum of neurologic and neuropsychiatric disorders [47] and lead to more global cognitive disturbances than do lesion to peripheral nodes [193]. The modular analysis of functional brain networks provides us with in-depth information about the topological roles of specialized regions that are engaged when performing a particular task. For example, certain brain regions serve as *hubs* within a community (provincial hub) or bind multiple communities (connector hub) to integrate different classes of information [118].

2.2.1.1 Functional Brain Networks

To visualize the whole-brain fMRI correlation data, prior work can be categorized into two common layout techniques, *i.e.*, **spatial** and **non-spatial** layout principles. **Spatial data** layout techniques take into account the actual anatomy of the brain regions while non-spatial techniques do not.

Abstract, non-spatial visualization techniques include spring embedding graphs, matrix bitmaps and scatter plots [158]. Fair *et al.* [60] and Deshpande *et al.* [51] used spring-embedding algorithms to visualize the brain network as 2D node-link diagrams. An alternative approach to visualizing the brain network is to view directly the correlation matrix [59, 25, 159, 2]. However, non-spatial techniques make it difficult to relate the nodes of the graph to the anatomy of the brain.

Three-dimensional (3D) node-link diagrams [196, 28, 41] that place the nodes at the spatial coordinates of the anatomical view address this shortcoming. Tools that utilize anatomical 3D spatial information, along with the brain template data for context, include the CoCoMac Paxinos 3D viewer [28], BrainNet Viewer [198] Brain Voyager QX [70], and Visual Analysis Tool by Li *et al.* [107]. While visualizing functional connectivity data as 3D node-link diagrams allow neuroscientists to inspect visually structural patterns in the brain, they also lead to cluttered and occluded visualizations. One way to overcome clutter and occlusion is to filter edges based on the connection strength threshold [196]. Alternatively, adopting a 2D anatomical layout whose nodes are projections on 2D planes (based on two anatomical axes) reduces clutter to some extent [2].

Spatial representations help neuroscientists to orient themselves with respect to the brain regions while 2D non-spatial graph layouts provide the necessary flexibility for visualizing modifications to the connectivity data [8]. Combining both representations provides more insight and allows neuroscientists to draw more detailed conclusions than spatial or non-spatial views alone. The functional brain connectivity explorer [185] was among the first published tools combining the strengths of non-spatial and spatial visualization techniques. However, there is still a need for additional visualization techniques that can help neuroscientists answer more in-depth questions and perform more specific analysis of modular structures of functional brain connectivity data. Other papers concerned with visual analysis of brain connectivity data include that by Akers *et al.* [5], Bruckner *et al.* [38], Beyer *et al.* [27], Jianu *et al.* [92], Whitfield *et al.* [194], Ribeiro *et al.* [147] and Brown *et al.* [37].

Several research groups over the past decade have been working on the particular problem of visualizing functional maps on the models of the brain; these include work by Cox *et al.* [46], Jin *et al.* [41], Bezgin *et al.* [28], Li *et al.* [107] and many more. These tools address the visualization needs for scientists working on different modalities of brain imaging data such as fMRI, DTI, CT, or scalp-recorded EEG or MEG.

2.2.1.2 Modular Structures

Interactive analysis of brain community structure allows us to formulate answers to questions about evolving brain network architecture as a function of short-term dynamics [7, 202], long-term aging [118, 43], disease progression Alzheimers or schizophrenia [47, 140]. Effective visualizations of the modular and hierarchical nature of the brain are crucial for articulating hypotheses about region-wise and module-wise patterns of integration and segregation.

Early work in visualizing communities [188] by Heer *et al.* [83] paved the path for automatically computing and visualizing communities. Later, research done by Corinna *et al.* [189] sought to solve the problem of visualizing overlapping communities. However, these published approaches do not focus on visualizing modular structures at multiple levels. To address this problem, Herman *et al.* [85] described various techniques to visualize hierarchical modular structures. These techniques include tree-map [93], cone-tree [149] and information cube [146]. ASK-Graph-View [1] deals with the problem of visualizing hierarchical structures by juxtapos-

ing an interactive visualization of a hierarchical tree with a graph and a matrix view. However, this tool only visualizes user-selected clusters from the hierarchical tree and, thus, hides the overview connectivity information between sub-clusters. Such visualizations make it difficult to analyze the interrelationships between sub-clusters. Analyzing these interrelationships is important to interpreting functional brain networks [119]. Summary graph [139], community matrix visualizations [197], and dendrograms [168] provide the foundation for visualizing hierarchical modular organizations of the functional brain networks. We expanded on the aforementioned visual analysis techniques for visualizing hierarchical modular data specifically for the domain of functional brain connectivity data.

Our research builds on these methods and expands their capabilities by emphasizing the visualization of hierarchically organized modular data. A wide range of other open source software packages that also focus on visualization of complex networks include SeeNet [25], BioLayout Express3D [178], NodeTrix [84].

2.2.1.3 Hierarchical Modular Structures

Little progress has been made concerning the interactive visualization of the modular and hierarchical organization of brain networks. The Connectome Visualization Utility (CVU) [104] is a tool used for exploring the modular structure of brain networks. However, CVU is limited to identifying and visualizing modular structures at a single scale. Furthermore, the visualization and interactive features of this tool are limited as they do not provide an overview of the communities or relationships between them.

Current systems typically suffer from information overload, occlusion, or lack of overview visualization. Information overload occurs when views provide few insights because of the overwhelming number of visual elements. Occlusion results when visualizations contain a large number of elements that overlap with each other, inhibiting comprehension of the data. Not providing an overview of the data complicates data comprehension by presenting a view that only captures the relationship of an individual brain region/node but not the global context.

2.3 Dynamic Networks

We now discuss major categories and techniques that explore and analyze the complex spatio-temporal patterns in dynamic network. The major challenges accompanying the analysis is succinct characterization of major properties of the networks, and layouting the mined patterns into a form where it is best perceived for visual understanding. Consequently, related work can be categorized into major sections:

- Communities for Dynamic Graphs
- Spatio-temporal Data
- ECoG data
- Dynamic Network Simplification
- Difference Graph Framework

2.3.1 Communities for Dynamic Graphs

When exploring communities in dynamic graphs, existing techniques primarily use animation (time-to-time mapping) or static timeline-based (time-to-space mapping) visualization methods to depict modular changes over time.

In *animations*, the community structure of the network is shown by color-coding the nodes or partitioning the drawing space into sections [89, 113, 63, 53] or nested blocks [145] (if the data is hierarchical). Due to their reliance on short-term memory, animations increase the cognitive load during analysis [14]. One way to mitigate this problem is to maintain the “mental map” of the layout by minimizing node movement in the animation [138]. An alternate approach to decrease the cognitive load is to place multiple graph representations along a timeline using small multiples [182]. However, this multi-view approach leaves the user with the manual task of assimilating and identifying changes.

To address this problem, several approaches utilize timeline based representations [150, 143, 190], visualizing only the evolution of clusters over time. In a timeline view, each segment along the axis perpendicular to the timeline represents a cluster identified at that particular timestep.

The links between two axes represent the changes in the cluster affiliation of the nodes. The arbitrary ordering of the nodes along the vertical axis may increase link crossings between axes, inhibiting easy comprehension of the evolution patterns. To address this issue, Reda *et al.* [143] and Sallabury *et al.* [157] employ sorting techniques to place active and stable communities at the top of the vertical axis.

To further support the comprehension of transitions between communities, alluvial diagrams [148] model the links between clusters in different vertical axes as split-merge ribbons [150, 187]. This approach enhances the visual traceability of important cluster evolution patterns.

Reda *et al.* [142] visualize the evolution of time-varying clusters while taking into account the spatial context, and by linking a space-time cube with a timeline representation. In contrast, our method provides the spatial context showing a multi-scale dynamic evolution patterns in 2D space, reducing visual clutter and occlusion. Our technique matches clusters through a overall match algorithm, enabling intuitive identification of time-varying community patterns.

2.3.2 Spatio-temporal Data

Previous visual analysis methods for spatio-temporal data utilize either integrated or separated views [12].

Integrated views visualize spatial and temporal data in one view. Superimposing temporal graph data onto a spatial view [11] and visualizing a 3D space-time cube timeline over a 2D spatial view [67] are two examples of integrated views. Another hybrid 2.5D approach proposed by Tominski *et al.* [181] displays temporal information on top of a 2D spatial layout. However, for a large number of timesteps or data points, these views can easily become cluttered and occluded.

Separated views overcome visual clutter by using dedicated views to present different aspects of spatio-temporal data. Plug *et al.* [134] link data in spatial and temporal domains by using small multiples of maps, superimposing a subset of temporal data on each of the spatial maps. Jern *et al.* [91] utilize color to link spatial and temporal data. Other methods [123] for static data use interaction techniques to link data in both domains, requiring substantial and concentrated eye movements for visual analysis.

To overcome such drawbacks, visual glyph designs aggregate spatio-temporal attributes that not only reduce the size of the represented data but also enable intuitive comparison of temporal data. *Glidgets* [97] depict temporal changes by segmenting glyphs into time slices, enabling the comparison of attributes over time. Related efforts by Nan Cao *et al.* [42] and Erbacher *et al.* [58] use glyphs that aggregate temporal data to summarize the entire dataset with the overall goal of detecting anomalous behavior in the network.

ECoG ClusterFlow uses a combination of the aforementioned concepts to provide unique glyph-based designs and visual analysis methods that show the overall modular changes of the network.

2.3.3 ECoG Data

Graimann *et al.* [74] presented methods to visualize event-related desynchronization and synchronization (ERD/ERS) patterns of implanted electrodes. Research done by Korzeniewska *et al.* [98] and Cristhian *et al.* [136] included the visualization of causal relationships among electrode sites. Kubanek *et al.* [101] recently presented a tool for visualizing topographies of ECoG cortical activity on a 3D model of the cortex. Although these approaches satisfactorily portray the spatial layout of the brain, they do not support the visualization of time-varying modular data for functional ECoG brain networks.

There exists a need for tools supporting efficient, high-level data analysis and exploration, including dynamic cluster analysis as a main focus. To aid the process of generating and verifying scientific hypotheses, a thorough visual understanding of the intricate spatio-temporal patterns of ECoG data is necessary. We address this need with a stand-alone application that allows a user to explore cluster community evolution at varying granularity.

2.3.4 Dynamic Network Simplification

One of the important problems in visualizing large-scale dynamic networks is depicting the connectivity information without substantial information loss [112]. Several techniques have been published that deal with the visual complexity problem. These approaches rely on manual filtering of edges and nodes or deriving a minimum spanning tree, when possible, preserving connectivity. The path-oriented simplification [179] method removes edges that do not affect the qual-

ity of best paths between any pair of nodes. Approaches like SEG [163], Apostolico [31] and EgoNetCloud [108] condense large network information without sacrificing connectivity information. Other approaches compress weighted graphs [180] or use motif-based methods [52]. Methods presented by Sun et al. [175] and Eagle et al. [54] use minimum description length and Fourier Transform analysis respectively to compress complex networks respectively.

In several applications, it is important to detect states in dynamic networks, see, for example, Bernaly *et al.* [141], Damaraju *et al.* [48], Calhoun *et al.* [40] and Mutlu *et al.* [127]. In contrast, the abstraction procedure in our method happens at the level of the network, preserving the rich topological structure of the individual state. Further, we compute a summary and difference representation to explore the complex dynamics underlying each state. Such methods are essential for an in-depth comprehension of the overall role of states in the context of system dynamics.

The technique presented by Murugesan *et al.* [124] explores the multi-scale temporal structure of graphs using visual designs along with a flow-based evolution view. In contrast, the approach introduced here uses a simplification algorithm to deal with networks that evolve over a large number of timesteps and utilize summary and difference representations to reveal variabilities and complex dynamics. Table 1 compares our method to other common approaches that analyze dynamic networks.

2.3.5 Difference Graph Framework

A difference graph describes the changes between the graphs of two timesteps. Given two graphs, only the changes concerning edges and nodes are visualized [15]. To deal with the problem of scalability in difference graphs, Archambault *et al.* [13] used hierarchies to show where large areas of a graph change. Subsequent work by Bourqui and Jourdan [34] visualized edges that have similar pathways, focusing on structural similarity. Work by Rufiange [155] is based on a hybrid method involving small-multiples and animations to determine local topological changes between graphs.

Our method focuses specifically on changes in edges (constant nodes), and community membership. We characterize changes by defining the importance of each change in the graph and visualizing the effect of the change on the topology using a community-based difference

graph.

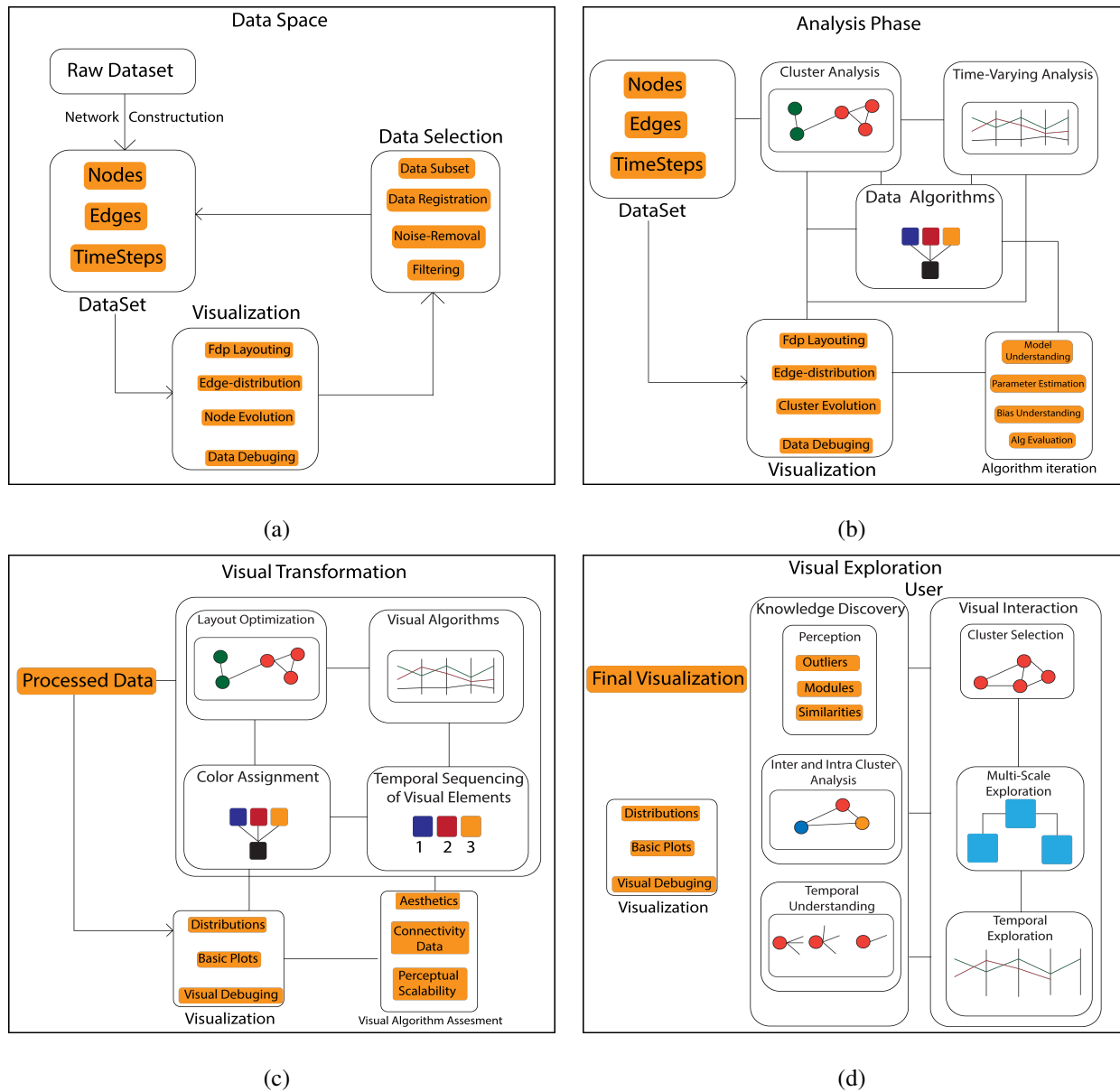


Figure 2.4. Framework for visual analysis tools for graph and network data. This dissertation addresses several components of this framework, split into four major modules, *i.e.*, Data Space, Analysis Space, Visual Transformation, Visual Exploration

2.4 Summary

This chapter reviewed related work and placed it into the context of our work. While there are numerous visual analytics solutions focused on understanding networks, little attention is given to exploring the organizational properties of the networks. In this chapter, more specifically, we start with a focus on the basic imaging methods in neuroscience that helps us construct a functional MRI as well as ECoG datasets. We then explore the major visual analytics techniques in static and dynamic networks. Further, based on existing systems exploring brain networks, we propose a graph analysis “ultimate” framework, see Fig. 2.4.

In the next few chapters, we go in-depth discussing specific contributions in methods for analysis of neural data and other major network datasets. The proposed solutions are primarily inspired by many existing works and surveys discussed in this chapter. The knowledge gleaned in the context of neuroscience, visual analytics, and complex networks have enabled us succinctly glean research gaps and challenges in the field. Chapters III to VII discuss the specific solutions and describe rigorous tests and specific evaluations done by domain-specific collaborators. The experts include researchers, neurosurgeons, network engineers and neuroscientists focused on a specific aspect of network exploration.

Chapter 3

Interactive Hierarchical Analysis of Functional Brain Connectivity

3.1 Introduction

The last chapter focused specifically over the wealth of available computational techniques available for easy network exploration of the underlying brain regions. Existing techniques for analysis of networks focus primarily on providing topological inference of the network or enabling intuitive understanding of the network as a whole. However, the detailed roles of individual brain regions, their group behavior or the interactions between high-level groups is poorly understood.

To better characterize the overall architecture of the network, neuroscientists are especially interested in developing a better understanding of the connectivity among brain regions. Of particular interest is the communication behavior observed when performing specific tasks or the impact of certain diseases on brain connectivity. For example, understanding of disease mechanisms could be advanced by moving beyond the assessment of atrophy in affected brain areas to a model of disease progression throughout vulnerable brain system.

By comparing resting state connectivity with connectivity during or after performing a task, it is possible to gain insights into how individual brain regions communicate when performing a task. Alternatively, tracking resting state connectivity during progression of neurodegenera-

tive disease, such as progressive supranuclear palsy, makes it possible to characterize how the disease affects cognitive abilities by identifying connection hubs that are disrupted.

In order to accommodate and explore the complex dependencies within brain networks, (Fig. 3.1), we implement a system called Brain Modulyzer. The tool supports multiple coordinated views to provide visual analysis techniques for exploring modular structures and hierarchical organization of functional brain connectivity.

The focus of the system is to automatically extract a hierarchical structure from the network and provide a visual way to convey this information as faithfully as possible. The methods and visualization system described are the result of a close collaboration involving neuroscientists and computer scientists, driven by the specifically desirable functionality of such a tool.

Specifically, our system supports (i) abstract views (ii) anatomical views and (iii) community analysis. (i) Abstract views show connectivity information (Figs. 3.1(a), (c)) and analysis results (Figs. 3.1(d), (e), (f)); (ii) physical views relate this information to the anatomy (Fig. 3.1(b)); (iii) community analysis provides integrated results that derive information like modules (Figs. 3.1(d), (e)), their hierarchical structure (Fig. 3.1(f)) and graph measures. Brushing- and-linking relates the abstract views and analysis results to the anatomy. The tool improves the state of the art through:

- An *overview + focus + detail* visualization of the modular structure and hierarchical organization of brain networks by adopting views that explicitly visualize inter- and intra-modular connectivity relationships.
- We present two case studies that demonstrate the usefulness of our visualization tool.

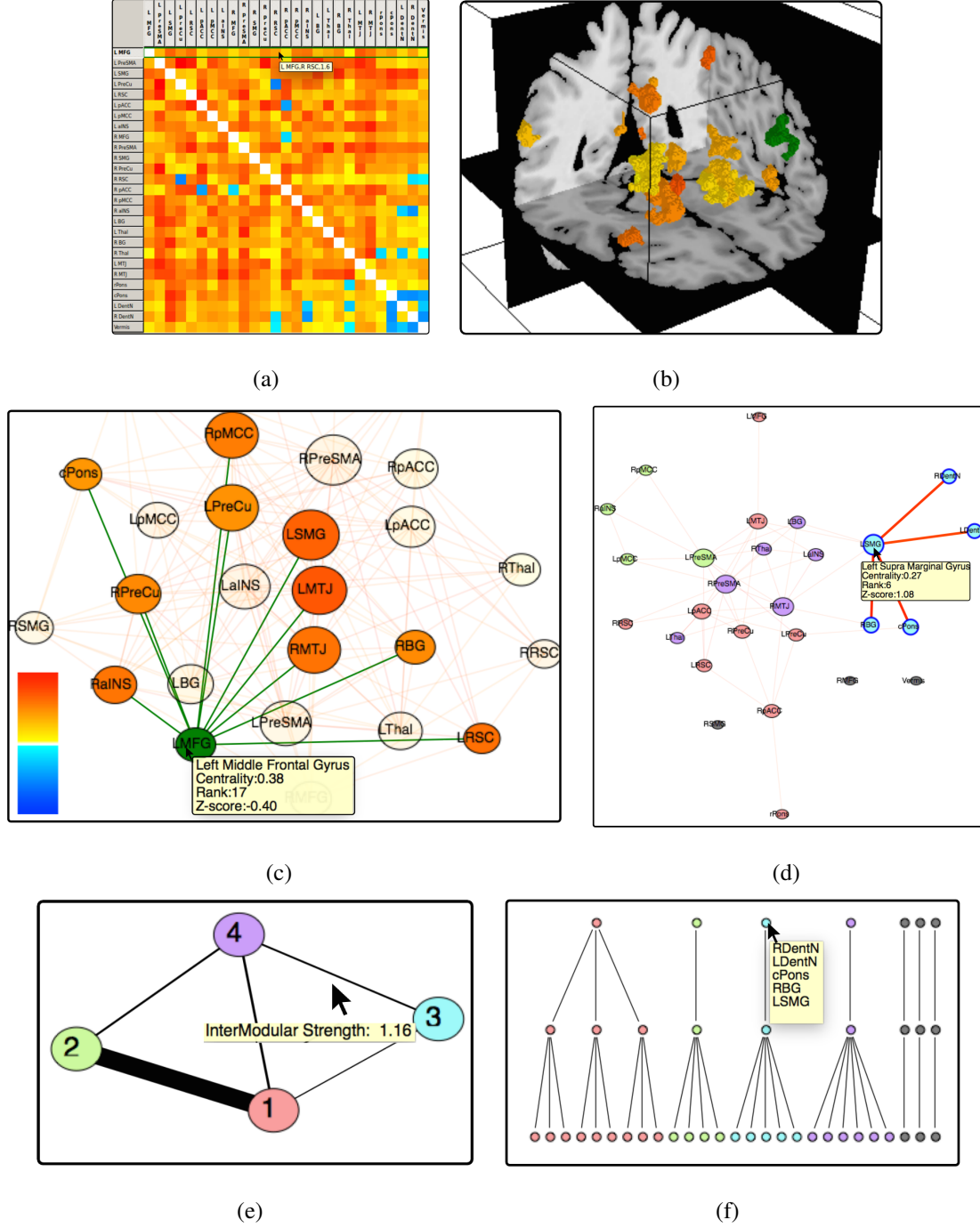


Figure 3.1. Our system links multiple views showing different data aspects to allow neuroscientists to investigate modular and hierarchical organisation of brain networks. (a) Heat maps show the pairwise correlation matrix between parcellated regions. (b) Selecting a region of interest—colored green in the figure—and coloring all other parcellated regions based on their correlation shows connectivity in relation to anatomy. (c) Node-link diagrams considering only connections above a threshold to show connectivity. Integrating community detection highlights groups of brain regions with strong internal connectivity. (d) This graph colors nodes according to their community membership. (e) The community graph provides a high-level overview over communities (b) by representing them as nodes. (f) Dendrograms represent the hierarchical organization of the modules.

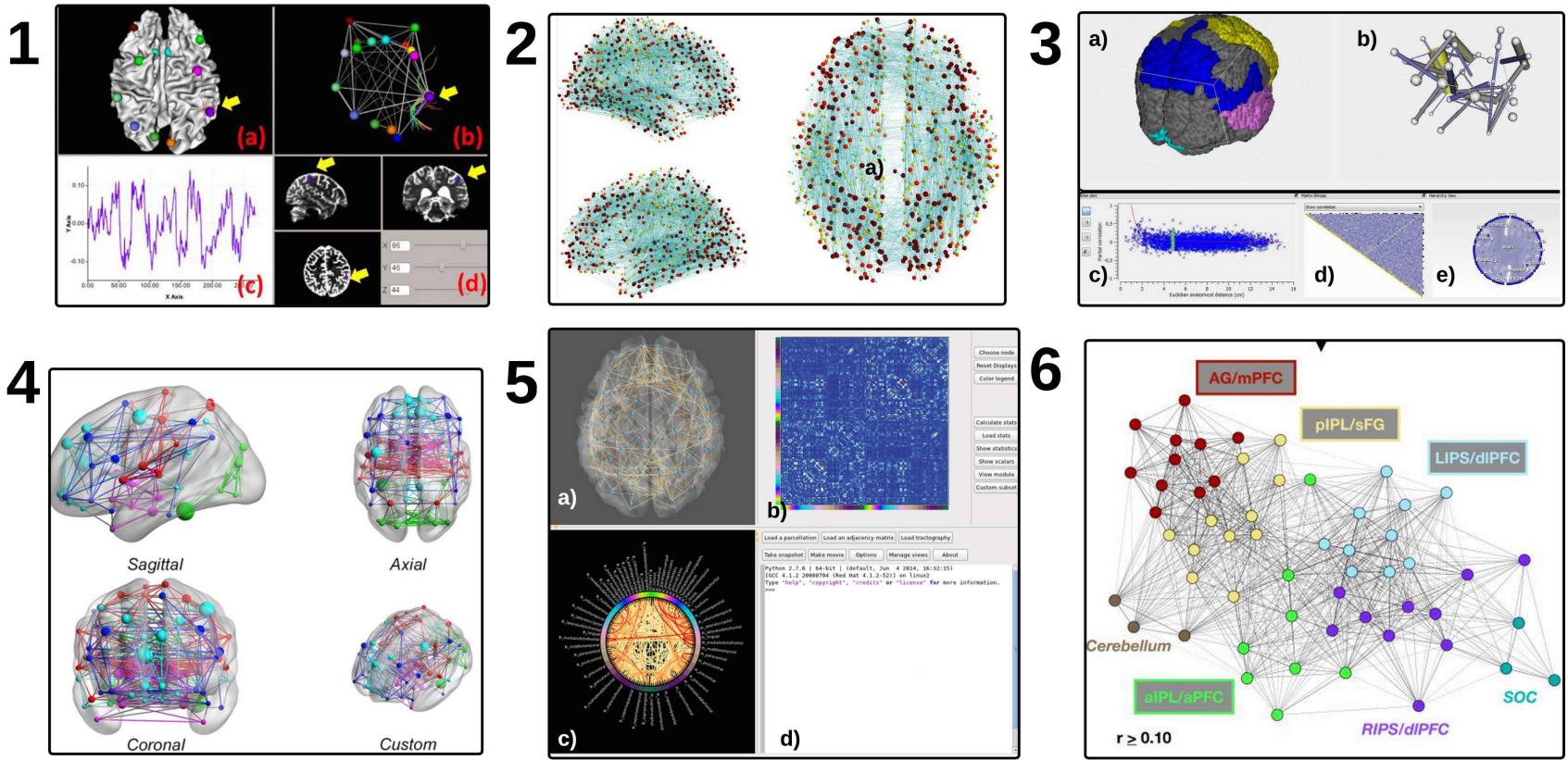


Figure 3.2. Typical analysis and visualization methods used for fMRI connectivity data. 1) The visual analysis tool by Li *et al.* [107], visualizes fMRI/DTI data as a 3D node-link diagram linked with an anatomical view 2) The connectome viewer toolkit [68] embeds a 3D node-link diagram in the anatomical view, the nodes are located at the centroid of their brain region parcellation. 3) The Functional Brain Connectivity tool visualizes functional fMRI data in these views: a) anatomical view, b) anatomical network, c) scatterplot, d) matrix bitmap, e) hierarchical edge bundling view 4) The BrainNet Viewer [198] visualizes 3D node-link diagrams where the nodes are scaled by nodal strengths 5) The Connectome Visualization Utility(CVU) [104] visualizes connectivity information in three different modalities (MEG, fMRI and DTI) through linked views, a) Anatomical View, b) Matrix Views, c) Circle View 6) Nelson *et al.* [128], defined regions based on modularity assignments. The 2D node-link diagram represents modularity assignments based on the color of each node

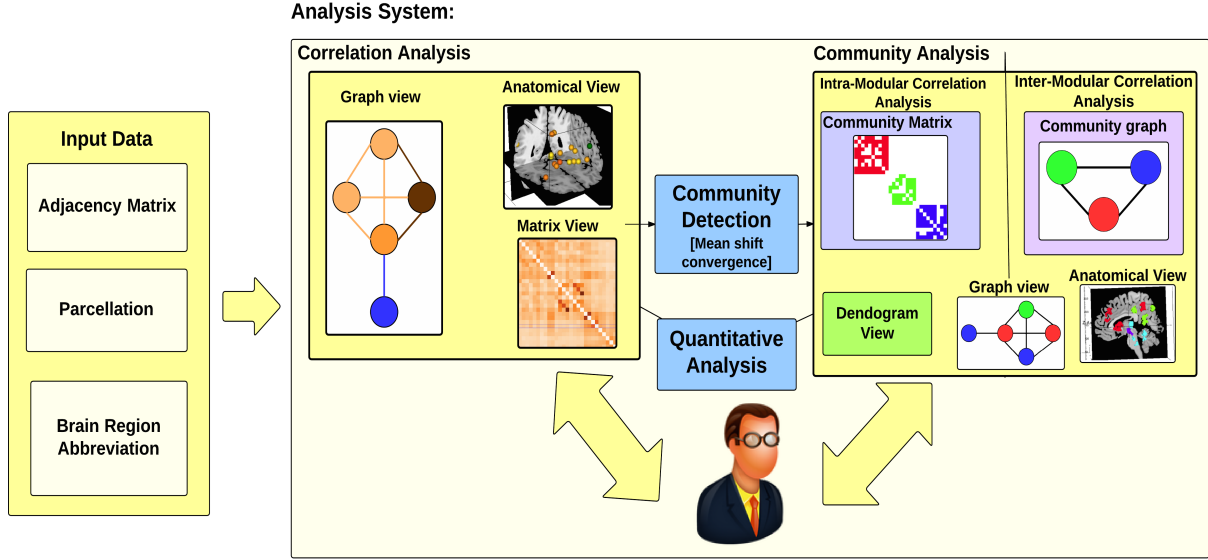


Figure 3.3. Our system uses the connectivity matrix and its associated parcellation as input. The system operates in two modes—correlation mode and community mode. While the correlation mode allows a user to analyze correlation network data, the Community mode allows a user to perform tasks for modular analysis.

3.2 Brain Modulyzer

In the following, we describe the rationale behind the design and implementation of Brain Modulyzer. The design choices were motivated by the following factors:

- **Faithful Representation:** Accomplishing a faithful representation of the underlying graph dataset is challenging as a wrong visual design choices could potentially mislead facts. Based on informal user studies and meetings with the users, we ensured faithful visual representation. Representing visual aspects of datasets is an ongoing research area in graph drawing [131].
- **Simplicity and Usability:** Making a visual interface complex offering a wide range of interactivity tasks is relatively easy, however achieving a powerful interactive interface with a limited complexity is challenging to achieve. This is often accomplished at the cost of simplicity and usability. A tool should allow intuitive exploration of data without much learning and also achieving simplicity providing value simultaneously.
- **Interaction through known Interfaces:** Interactivity is key to understanding large datasets. ■

The ability to query the datasets at the blink of an eye makes the tool more powerful. In the context of graphs, a tool should easily enable sophisticated interaction techniques with ease and simplicity.

- **Aesthetics and Visual Criterion:** When designing a system solving a particular visual analytics problem, care should be taken to focus on the communication of the data itself versus the aesthetic criterion of visual elements. The additional overhead of understanding the aesthetic criterion inhibit easy understanding of the data.

Focused on their design goals, we implemented Brain Modulyzer where the input to our system is the correlation matrix and its associated parcellation data. Fig. 3.3 shows the flow diagram of our tool. Our tool uses the principle of focus+context and detail+overview [44] to visualize connectivity data.

To carry out major network analysis tasks, the tool operates in two major modes, a correlation and community mode, see Fig. 3.3. While the correlation mode provides standard analysis of connectivity data, the community mode provides additional views, *i.e.*, summary graph, community matrix, and dendrogram views for modular and hierarchical analysis of functional brain networks. The overall visualization approach is supported by (i) abstract connectivity—showing non-spatial 2D representations of quantitative relationships between brain regions; (ii) physical anatomy—showing the corresponding brain regions in a spatial context; (iii) community analysis—supporting real-time detection of modules, providing results obtained via interactive analysis and graph metrics results. All views focus on at least one data characteristic—connections or positions—and incorporate the strengths of their associated visualization techniques.

3.2.1 Anatomical Views

Three Anatomical views supported by our tool provide the spatial context for connectivity. These are:

- A **brain parcellation view** shows individual regions in three dimensions (Fig. 3.6, bottom right corner). The brain parcellation is represented as a 3D volume view, where each voxel specifies the identifier of the region containing it (which is zero if the voxel does

not belong to any region). Our system extracts region boundaries using the marching cubes method [110, 132], applied to a binarized version of the volume. Each extracted triangle is associated with an identifier specifying the region from which it was extracted. Our system uses this region identifier to color the regions based on its connection strength or community membership to a selected region, while using the same color map as the correlation matrix heat map.

- A **parcel centroid view** displays a sphere at the centroid of each parcellated region. This representation is useful when the regions fill the entire volume of the brain. The color of the sphere depicts the connection strength of the currently selected region to the other regions or its community membership, see Fig. 3.6(c).
- **Slice views** show a 2D slice in the sagittal, coronal and axial plane, see Fig. 3.6). A grayscale image of the MRI scan—registered with the parcellation volume—provides the spatial context. The parcellated regions are colored based on their connection strength to the currently selected region or its community membership.

To provide additional anatomical context for the 3D views (brain parcellation view and parcel centroid view), our tool displays 2D slice views (sagittal, coronal and axial) both as 2D images and as planes in the 3D views. Furthermore, the user can choose to show a semi-transparent isosurface representing the brain surface in the 3D views.

- A **matrix view** provides users with information about the general correlation or relationships between the brain regions. Comparing the matrices of the functional connectivities for different data sets allows users to quickly assess the trends across the data sets [21, 69, 100] and provides a means of visualizing raw data directly. Each matrix entry $c_{i,j}$ is color-coded according to the correlation strength between the i -th brain region and j -th brain region. As the data is symmetric, the $c_{i,j}$ value is equal to the $c_{j,i}$ value. To aid users in distinguishing the entries in the matrix that are selected, a green rectangular box is placed over the row of the matrix that is selected.
- A **graph view** shows the brain network encoded in the functional connectivity data matrix as a node-link diagram. Nodes depict brain regions, and edges indicate the (pairwise)

correlation strength between the nodes. A force-directed layout of the brain network complements the correlation values, *i.e.*, highly correlated nodes are much closer in the spatial domain.

The following capabilities support performing high-level brain analysis tasks with our tool:

- **Node coloring:** Often neuroscientists are interested in a node and its correlation with the other nodes to investigate its topological role in the network. Our tool allows users to select dynamically brain regions and explicitly visualize their edge strengths with respect to other nodes in the graph. The nodes in the graph view are colored according to their correlation data values, which relate the selected node to the other nodes in the graph. The edges in the graph view are colored according to the correlation data values and combining all of the edges would result in the entirety of the matrix when the minimum threshold value is applied.
- **Thresholding:** Visualizing all the information provided in the matrix produces a complete graph, *e.g.*, for the 27 region graph in the case study in Section 3.3.1, there would be 576 edges. Without thresholding, typical graph rendering techniques would lead to clutter and occlusion. For the purposes of graph analysis, correlations below a certain threshold may not be considered as connections. Hence, to reduce clutter, our system supports dynamic changes to the minimum threshold value for rendering edges. Filtering allows users to specify different threshold values to remove clutter-causing edges around a certain node of interest, which in turn helps the users to focus on the strongest connections between brain regions. Modifying the threshold value changes the overall graph structure, as well as the associated node layout in the graph view.
- **Dynamic layout change:** Neuroscientists use different graph layouts for graph analysis of the brain. Our tool allows the user to dynamically select some of the widely used graph layout algorithms used by neuroscientists, include the neato graph layout [133], force-directed-layout [65], and circular or shell layouts [24].
- **Interactivity:** Interactivity is a key part of our visual analysis tool. All linked views provide interactive tooltips. Hovering over or clicking on graph nodes selects a graph entity.

Interactive annotations in the graph view convey information about network measures, correlation strength and the abbreviation of the region selected.

3.2.2 Connectivity Views

To explore the connectivity between brain regions (without considering spatial information) we use information visualization and graph theory techniques. The non-spatial visualizations are useful because they provide ways to show the explicit changes to connectivity. Namely, the nodes and edges immediately reflect changes made to the connection strength threshold. Two abstract views supported by our tool are matrix views and graph views are discussed in the following.

3.2.3 Linking and Interactivity

Interactive linking between the anatomical views and the abstract views is an important part of our tool. Fig. 3.1 shows an example of selecting the brain region “LSMG” in the abstract view and the resulting data representation in the anatomical views. Users can also choose to select a brain region in the anatomical views, which then gets propagated to the abstract views.

When used in community mode, our system provides three types of linking features to analyze hierarchical relationships: (i) leaf-node-to-brain-region linking, (ii) sub-community-nodes-to-brain-regions linking, (iii) community-nodes-to-brain-regions linking, see Section 3.6. To analyze the modular organization, the interactive features include: (i) inter-modular highlighting between-module connections, (ii) intra-modular highlighting within-module connections. Other interactive features include specifying the connection strength threshold to filter edges in the connection graph, see Fig. 3.1(c), de-cluttering the connection graph, dynamically recalculating the graph structure, graph-theoretical metrics, and hierarchy of communities. Changing the connection strength threshold re-orders the headers of the community matrix and the leaves of the dendrogram such that community relationships of entities are reflected by the placement of their corresponding graphical primitives.

3.2.4 Community Detection and Visualization

We apply the Louvain community detection algorithm [30], to detect communities in the correlation matrix. We pick/choose colors with a maximum perceptive distance [50] for different

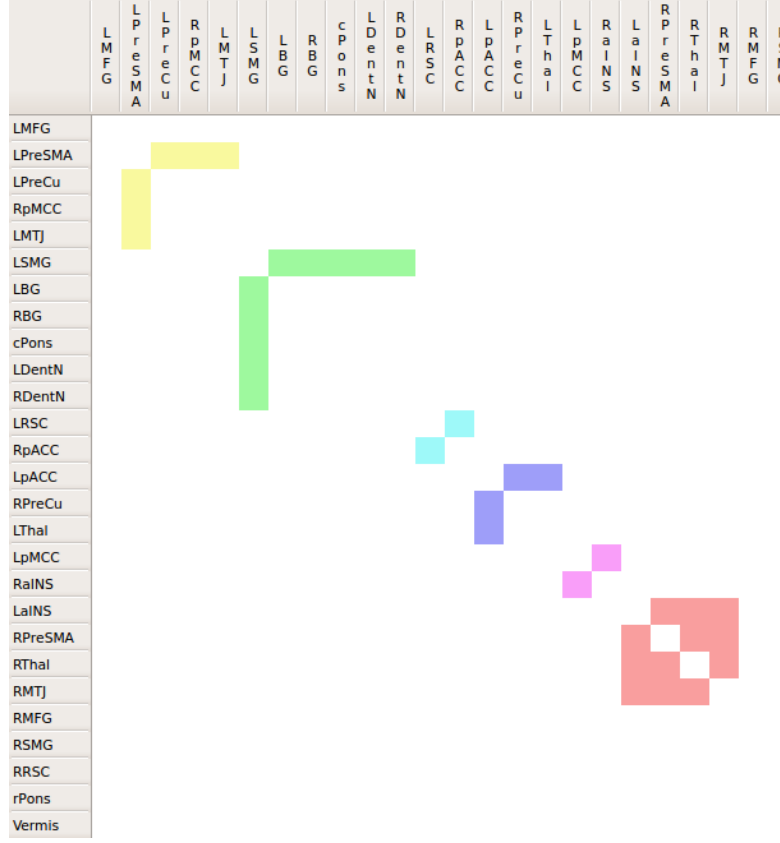


Figure 3.4. The matrix view shows twelve communities identified by the modularity optimization algorithm (Fig. 3.5). Six of the twelve communities are connected internally. Color-coded cells of the matrix view represent intra-modular connections.

communities. When the community mode is selected, the changes that occur in the abstract and anatomical views are discussed next.

- **Abstract views:** As the threshold value is changed, our tool detects and renders communities in real-time. For example, nodes in the graph view are colored according to community membership (Fig. 1(d)).
- **Community matrix:** Our system provides a view of a *community matrix* that visualizes intra-modular connections, see Fig. 3.4. Cell c_{ij} corresponding to a connection within a community above the current threshold is represented by the color of its community. Rows and columns of the matrix are ordered such that sub-matrices associated with particular communities comprise contiguous rows and columns in the matrix. It is important to note that as the system already knows the community

assignments of nodes through the Louvain algorithm, we dynamically sort [87] the rows and columns of the matrix.

- **Community graph:** In this graph view, see Fig. 3.1 (e), nodes represent individual communities that are detected in the primary graph/matrix view after applying a threshold value. The length and the thickness of the edges in the community graph represent the inter-modular correlation edge strength, *i.e.*, an edge in the community graph between a community node A and a community node B in the community graph is the mean of edge lengths of all edges that start from nodes in community A and end at nodes of community B, see Fig. 1(e).

Nodes in the community graph (Fig. 1(e)) are placed at the average node coordinates of the all the nodes in the original graph layout (Fig. 1(d)) comprising their community, ensuring spatially consistent node placement.

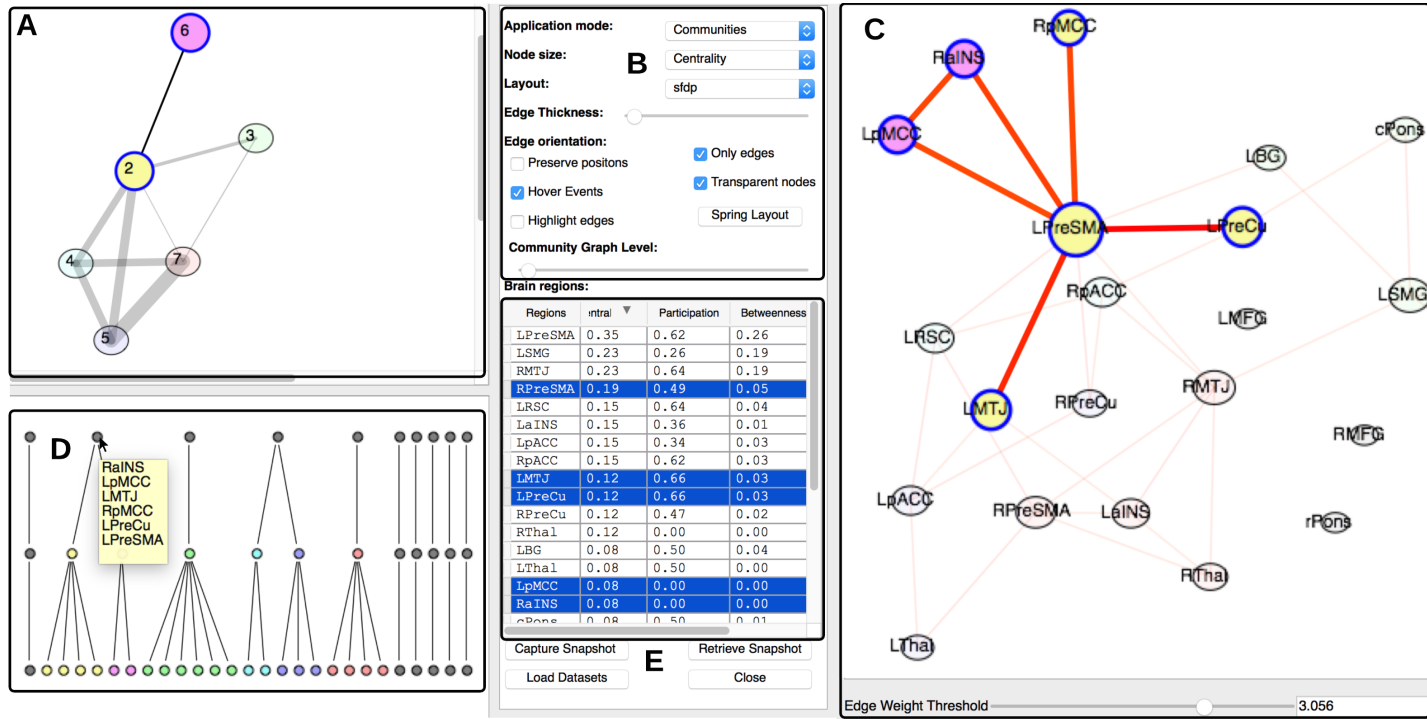


Figure 3.5. (a) The community graph highlights the subcommunities associated with the community selected in the dendrogram view. (b) Configuration options provide various choices for interactivity, such as toggling hovering/clicking, choosing graph layout and varying the opacity of highlighted nodes. (c) The brain region graph displays highlighted nodes associated with selected subcommunities. (d) The dendrogram view displays the hierarchy of communities. The communities that do not have any edges emanating from them are grayed out. (e) A table view lists important graph properties of the graph shown in (c).

- **Anatomical views:** The parcellated brain regions are colored according to the communities identified in the corresponding abstract view. This coloring shows the anatomical context for the detected communities, see Fig. 3.6.

3.2.5 Dendrogram View

The Louvain method [30] outputs a hierarchy of modules that we visualize in the form of a dendrogram. Nodes in the lowest level of the hierarchy denote individual brain regions while the nodes in the highest level denote the resulting communities. The nodes that are neither of these are denoted as sub-communities.

We construct the dendrograms in our system as follows, see Fig. 3.7: given the hierarchical modular information, we maintain a tree-like data structure that maps nodes in the dendrogram view to the nodes in the graph view. We visualize the cluster data using a layout that places the nodes in hierarchical order, *i.e.*, it places nodes in the same hierarchy level at the same height starting with leaf nodes at the bottom and community nodes at the top. Changing the threshold for community analysis interactively results in an updated dendrogram view. The highly interactive features of the dendrogram view are what makes the view a significant part of the modular analysis. For simplicity, we assume that the modularity-optimization algorithm provides us with three levels of hierarchy, *i.e.*, communities, sub-communities and brain regions. The different types of interactive features include:

- **Leaf-node to brain region linking:** Hovering/clicking on leaf-nodes highlights the corresponding brain regions along with the community to which it belongs. The selected entities also include intra-modular edges of the selected community.
- **Sub-community nodes to brain regions linking:** Selecting a sub-community node (*e.g.*, in level 1), highlights not only the corresponding brain regions that belong to the selected sub-community but also the edges within the sub-community.
- **Community nodes to brain regions linking:** Picking a community node (*e.g.*, in level 2) in the dendrogram view highlights all the associated sub-communities. Highlighting a sub-community is similar to the *sub-community nodes to brain regions linking* mechanism.

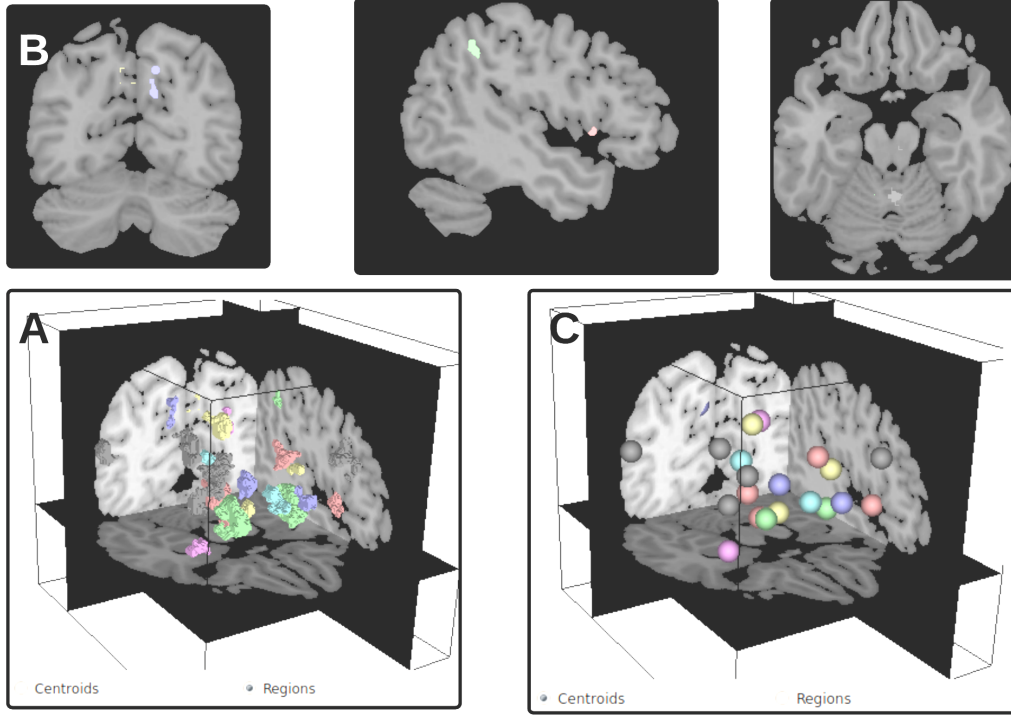


Figure 3.6. Results of community detection (Fig. 3.5) with respect to anatomy. Each community is represented by a distinct color, and each region is colored according to its community membership. Parcellated brain regions can be shown as outlines (a) or centroid depiction via a sphere (c).

Common hierarchical brain analysis tasks often require the correlation information between sub-communities. In order to provide an overview of interactions between sub-communities, the tool allows the user to pick a level in the dendrogram hierarchy to be visualized in the community graph view. For example, in Fig. 3.5(a) the community graph visualizes sub-communities formed by the modularity optimization algorithm at level 1. Highlighted sub-communities colored in yellow and pink later merge into one community in the next stage of the hierarchy.

The leaf nodes and subcommunity nodes of the dendrogram view are color-coded according to their community/sub-community membership. These nodes are also spatially re-ordered based on the spatial positions of the community nodes in the dendrogram view.

3.2.6 Network Measure Table

Measures used in graph theory often convey useful information about connectivity profiles of individual brain regions [154, 39, 169]. The graph measures that we utilize include:

- **Degree centrality:** This measure captures the fraction of nodes a node is connected

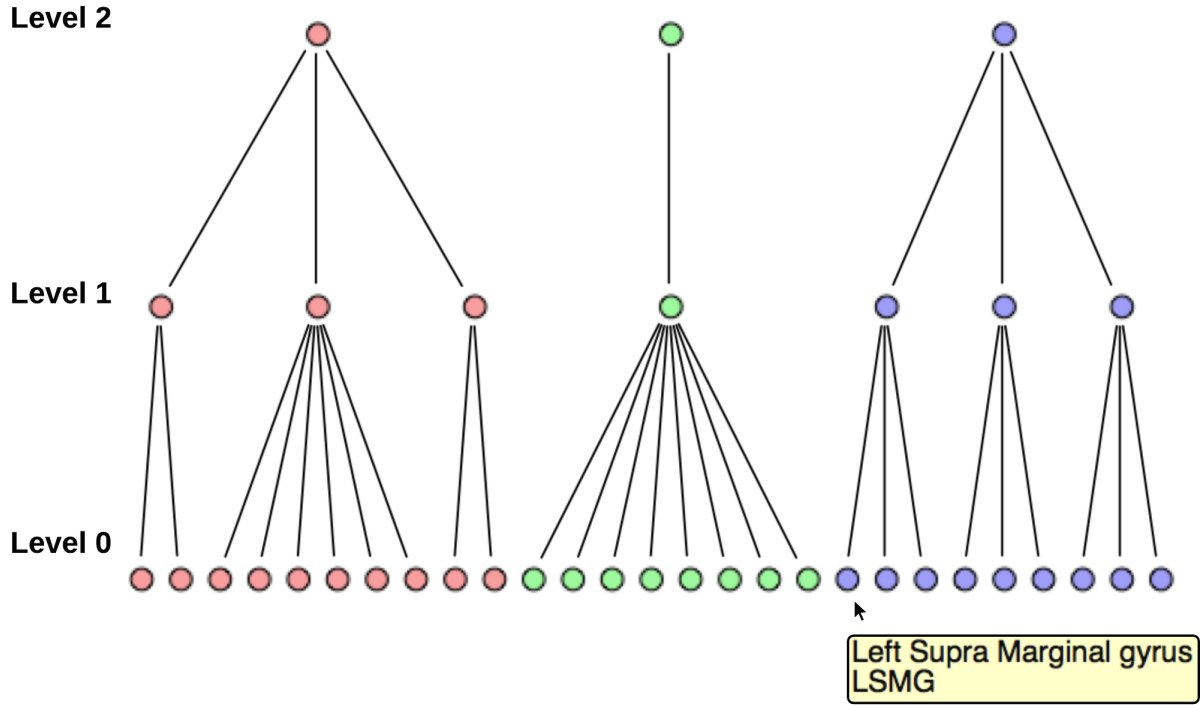


Figure 3.7. The Dendrogram view showing the hierarchical modular information of a community corresponding to threshold value 3.309. Level 1 of the dendrogram represents sub-communities formed when optimizing modularity locally with the nodes in level 0; level 2 nodes represent the final community emerging from the sub-communities identified in level 1.

to [32]. A high value indicates that a node is a local hub that integrates information from its neighbors.

- **Participation coefficient:** This measure reflects the distribution of links of a node with all the modules in the network [77]. A high value indicates that a node acts as a bridge between modules.
- **Betweenness Centrality:** This measure is the sum of the fraction of the shortest paths from all vertices to all others that pass through a node [35, 129]. A high value characterizes nodes that enable rapid distribution of information from one part of a network to another part.
- **Within-Module Degree Z-Score:** This measure represents the connectedness of a node to other nodes in the same module [154]. A high value indicates that the node is a provincial hub within its module.

We present these measures in a table showing one row per brain region, supporting the identification of anomalies or patterns in brain networks. Tooltips provide additional information regarding Z-score values and the relative rank of the measure that is selected. A user can sort the table using any network measure by selecting the table headers.

The table shown in Fig. 3.5(e) provides statistics for communities identified for the threshold value 3.056, sorted by descending centrality. Additionally, one can examine the measures of brain regions that belong to a selected community in the dendrogram view. For example, the brain region PreSMA has a centrality value of 0.19 and is ranked fourth based on the sorting criteria. One can deduce that the node has a relatively average participation index of value 0.49.

Visualizing Graph Measures as Graph Properties Our tool conveys network measures visually by mapping network statistics to graph properties. Fig. 3.5(c), uses *node size* to represent centrality, with larger node sizes indicating higher centrality. This visualization conveys patterns that are not so obvious when presenting network statistics alone.

3.2.7 Intra- and Inter-community Analysis

Analysis of brain networks typically aims at answering high-level questions [154]:

- How diverse are inter-modular connections?
- How does one module specialized in performing a certain task interact with other modules?
- What nodes in a module handle global inter-modular integration?

Exploring and ultimately answering these questions requires both inter-modular, and intra-modular graph analysis and visualization methods. Graph-based analysis techniques provide precise information about the role of a particular node in the network. Our system dynamically computes all the measures pertinent to intra- and inter-modular connectivity.

The community matrix view visualizes connectivity within communities. Color-coded cells of the matrix view represent the intra-modular connections. A larger number of colored cells in the community matrix view indicate strong intra-modular connections.

For example, out of the twelve identified communities shown in Fig. 3.4, only six are connected internally, *i.e.*, have edges within the module. As the other seven communities are single

region communities, they do not possess any colored cells in the community matrix view.

The community graph view focuses on visualizing inter-modular edge connectivity. Edge thickness indicates inter-modular correlation-strength. Selecting an edge in the community graph view highlights the corresponding correlation data in the matrix view and the graph view. A user can distinguish the nodes that are involved in inter- or intra-modular communication visually through dynamic interaction techniques.

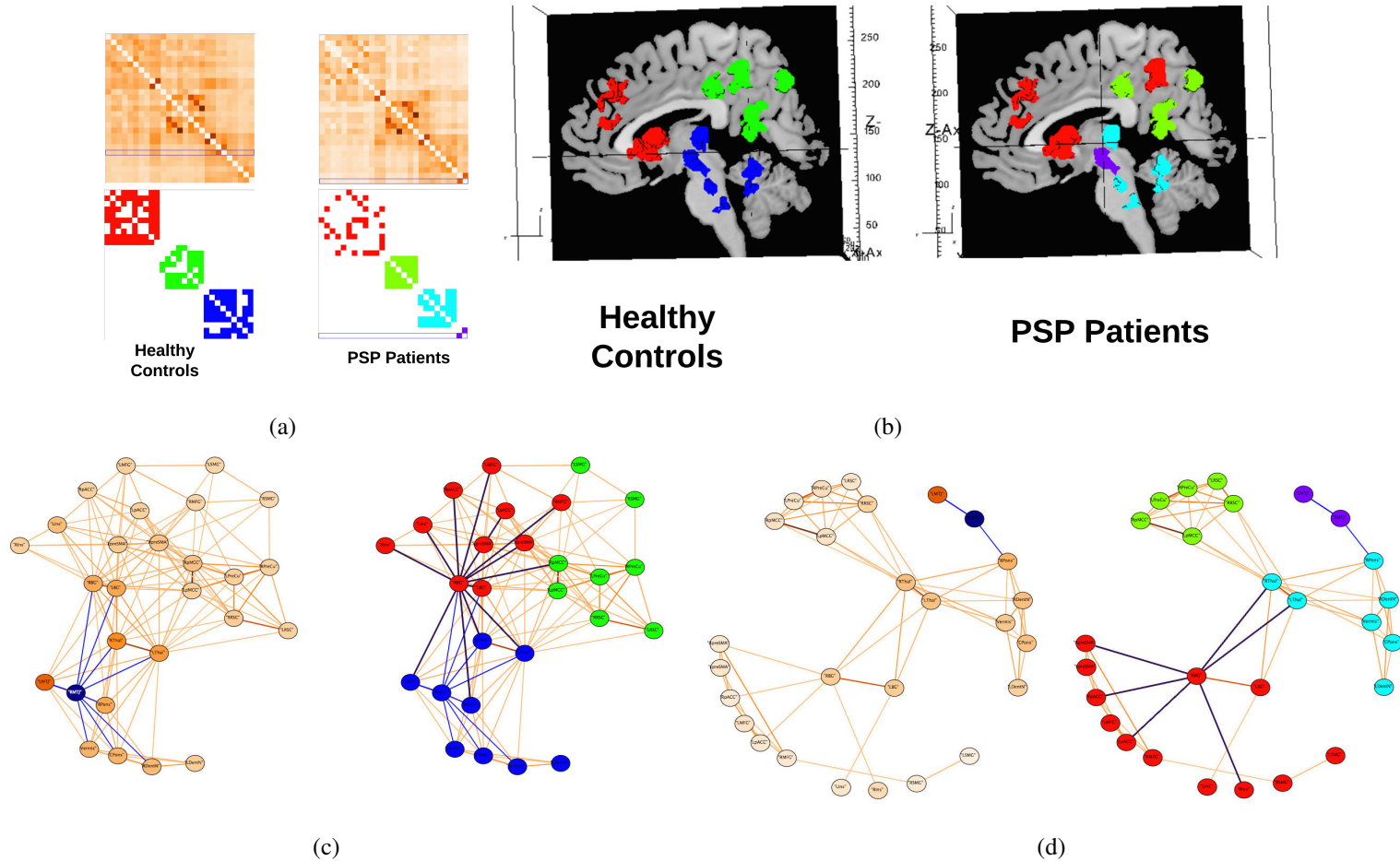


Figure 3.8. Functional connectivity of the mesothalamic junction (MTJ) network in 36 healthy controls subjects and 18 PSP patients. (a) Matrix views of the MTJ network in controls (left column) and PSP patients (right column); the top row shows the connectivity weights as a heat map with the right MTJ node's row highlighted in blue, while the bottom row shows connections above the selected weight threshold as solid colors corresponding to the community they belong to. (b) Anatomical views of the nodes comprising the MTJ network, colored according to their community membership in controls (left) and PSP patients (right). Spring-embedded network views of the control (c) and PSP patient (d) group mean networks, illustrating connectivity strength from the right MTJ (left column) and the community structure of the networks (right column). Right MTJ connections are shown as blue edges in all networks. Right basal ganglia connections are shown as purple edges in the right column.

3.2.8 Focussing on Data Subsets

Our experience with tool showed that when analyzing larger data sets, the matrix cells, and the headers are no longer appropriate for visually capturing all the salient information of interest. In particular, to present a matrix view with a large number of rows/columns., it becomes necessary to omit row and column headers. While tool tips can provide information as to which pair of regions corresponds to a correlation value, this process becomes tedious for larger subregions. To support these use cases, our tool displays a subset of the data, *i.e.*, a block of selected matrix elements, in another view. Restricting the matrix to a subset makes it possible to display row and column headers for orientation. The matrix elements to be displayed in the magnifier window can be based on any number of elements in the matrix views. A tooltip over the entries of the matrix view provides exact information about the associated two brain regions with their corresponding correlation value.

3.3 Results

Our team constitutes experts from visualization as well as neuroscience groups, all of which contributed to the design of the tool. The neuroscientists on our team actively use this tool to formulate new hypotheses and test their findings. Formal as well as informal meetings were held with the neuroscientists on a monthly basis. The questions of the study were: What insights did the neuroscientists gain from using such a tool? How does the interplay of standard visualization techniques help them arrive at the hypotheses?

The tool has helped the domain experts generate new hypotheses and research ideas to analyze neuro-degenerative diseases such as Progressive Supranuclear Palsy and tasks based on memory encoding and retrieval. We describe the data acquisition process as well as the scientific insights gained through the use of our tool by the domain experts.

3.3.1 Case Study 1: Progressive Supranuclear Palsy

Progressive supranuclear palsy (PSP) is a neurodegenerative disease that typically begins around the age of 60 and causes restricted vertical eye movement, balance impairment, axial muscle rigidity, and cognitive-behavioral deficits. Patients demonstrate atrophy of particular brain

structures including the midbrain, cerebellum, thalamus, basal ganglia, and portions of the pre-frontal cortex. The disease course is rapid and there is no available cure [72]. Previous work has identified functional connectivity reductions in a network anchored by the rostral mesothalamic junction (MTJ), connecting to an array of regions which contribute to diverse functions including oculomotor control, skeletomotor control, and social/emotional processing [66]. In this study, 18 patients with PSP and 36 age-matched healthy control (HC) subjects received task-free (resting state) fMRI scans. A total of 27 specific regions of interest comprising this network were used to calculate the pairwise functional connectivity of all regions and define a 27×27 covariance matrix for each subject. To summarize the network connectivity patterns in patients with PSP and healthy controls, we calculated the average connectivity matrices for the 36 controls and 18 patients.

As reported by Gardner *et al.* [66], the mesothalamic junction (MTJ) is an epicenter of PSP disease-related pathology, structural atrophy, and functional connectivity disruption. In that study, the MTJ study exhibited significantly reduced total flow, a graph theory measure similar to the nodal degree. In this study, we visualized the connectivity of this node with other nodes in the MTJ network using the *Brain Modulyzer*. As can be seen in Fig. 3.8, the right MTJ node in control subjects, see Fig. 3.8(c); blue lines emanating from selected node) shows higher connectivity to neighboring nodes than it does in PSP patients, see Fig. 3.8(d). Furthermore, the graph view (color-coded with communities) shows the left and right MTJ fragmented as an isolated community in PSP patients (Fig. 3.8(d), right column, purple nodes). This isolation can be simultaneously recognized in the matrix view (Fig. 3.8(a), right column) and the anatomical location of these nodes can be seen in the anatomical view (Fig. 3.8(b), right). Finally, the graph view (color-coded with communities) illustrates the significant isolation that exists between each community of nodes in PSP patients. Our previous work had shown that connectivity was globally reduced in this network. The visualization tool clarified that the major consequence of reduced connectivity was progressive isolation of the three modules in this network: the sub-cortical (Fig. 3.8(b), left, blue), frontal (Fig. 3.8(b), left, red), and parietal (Fig. 3.8(b), left, green) modules. The cognitive symptoms of PSP include bradyphrenia, an impairment of concentration, and loss of mental flexibility, both of which are components of what has been called

a subcortical dementia [6]. These symptoms are thought to reflect impaired interaction between cortical regions and subcortical regions including the basal ganglia and thalamus [160]. The visualization tool makes simultaneously apparent: (i) the reduced degree centrality of the basal ganglia (“RBG” and “LBG” nodes) in PSP as compared to HC (purple lines in Fig. 3.8(d), right in PSP compared to right in HC, Fig. 3.8(c)) and (ii) the substantial isolation of the subcortical module from the frontal module (blue nodes to red nodes in Fig. 3.8(c), right, vs cyan nodes to red nodes in Fig. 3.8(c), right).

This combination of community detection, nodewise graph theory metrics, anatomical rendering of community membership, and the ability to assess nodewise connectivity edges enables complex, multipart observations that are not otherwise easy to synthesize. Here, this disconnection of the basal ganglia in the face of cortical-subcortical isolation is a candidate mechanism for bradyphrenia in PSP that has not previously been explored in a network context.

3.3.2 Case Study 2: Memory Encoding and Retrieval

Fifteen healthy young subjects participated in an episodic memory encoding and retrieval task (unpublished data). Subjects were asked to learn the associations between images of different famous faces and famous places that involved first studying the pairs of faces and places to be remembered, then immediately afterwards receiving a task-free (resting state) fMRI scan. The following day, subjects returned and performed a memory test to assess how many of the studied pairs they could accurately recall from the previous day. Memory performance was measured in order to relate it to functional connectivity as observed by the pre-encoding and post-encoding fMRI scans. We hypothesized that subjects with superior memory would show elevated connectivity between memory and visual regions in the post-encoding scan, indicative of more active memory consolidation. Subject fMRI scans were parcellated into 199 brain regions spanning the cortical gray matter, subcortical, and brainstem regions. The pairwise functional connectivity between these regions was calculated to derive the 199×199 connectivity matrix for each subject. A *behavioral correlation matrix* was obtained by calculating the correlation of each subject’s performance on the memory task with their functional connectivity between region A and region B, then repeating this procedure for all pairs of regions. This 199×199 groupwise behavioral correlation matrix can be thought of as a network rep-

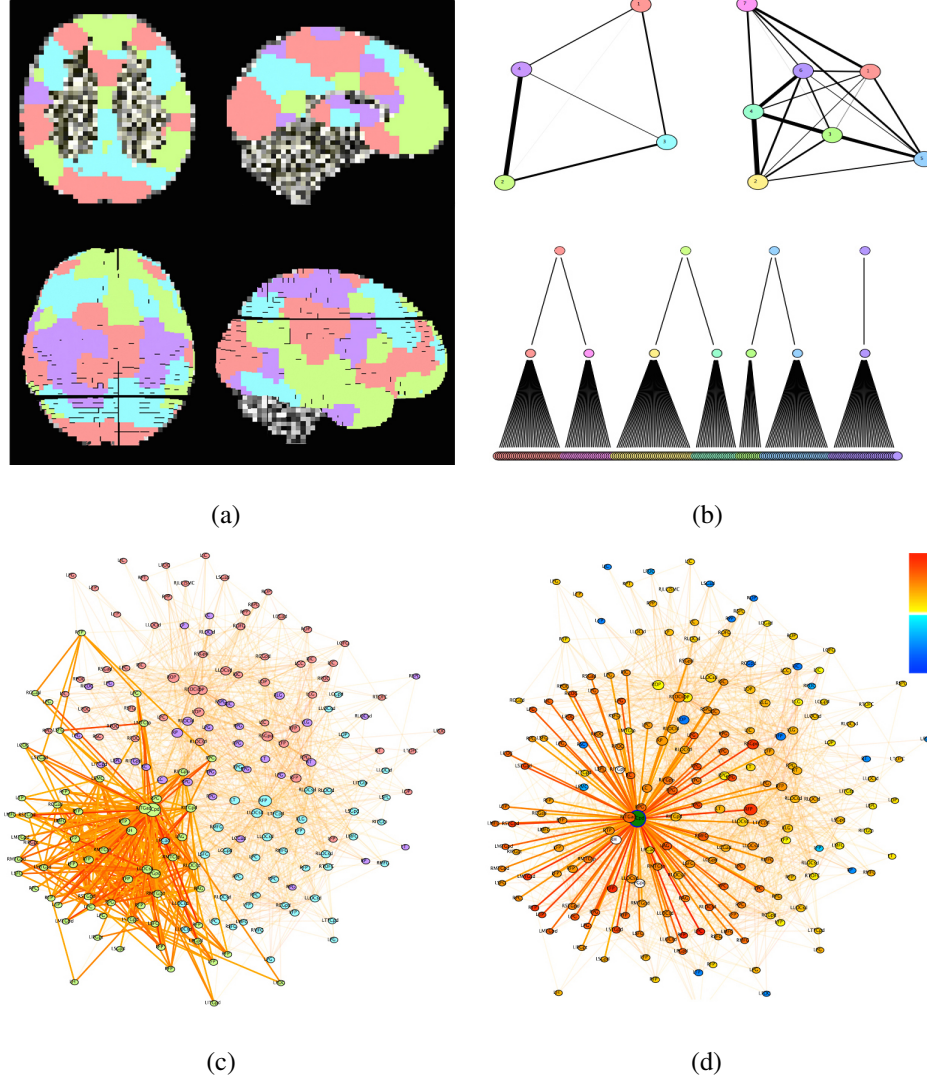


Figure 3.9. Anatomical and graph views of the memory consolidation network (MCN). (a) Planar (top row) and 3D surface (bottom row) views of the 199 regions of interest in the network, overlaid on the anatomical template brain. Node colors correspond to modules as determined by the *Brain Modulyzer*. (b) Network aggregate community graphs (top) and Louvain modular hierarchy dendrogram (bottom). The left network shows the four communities at the top level of the hierarchy while the right network shows the eight subcommunities at the second level of the hierarchy. (c) Whole graph view of the network, showing all nodes with the color corresponding to module membership as in A and B at the top level of the hierarchy. Highlighted are all connections within the default mode network module nodes (green). (d) Whole graph view of the network, showing the connections of the posterior hippocampus/fusiform cortex (green node) to its most strongly connected neighbors ($r > .25$; bottom). Node colors represent the correlation strength to the selected node.

resenting successful memory consolidation. The memory consolidation network (MCN) was analyzed for community structure to determine whether particular regions became more central

hubs in the post-learning network, when memory consolidation was occurring for the subjects with superior memory performance.

Community assignments were visualized in the *Brain Modulyzer* both in abstract and anatomical space. The modules in this network can be recognized as known intrinsic functional connectivity networks—the default mode (Fig.; green), visual/sensory (red), executive control (cyan), and motor networks (purple). The anatomical view is essential for identifying the spatial topography of these networks. For example, in Fig. 3.9(a) and (b), the default mode network (green) can be recognized based on its layout in medial frontal, medial parietal, lateral inferior parietal, and temporal lobes. The executive control network (blue) can be recognized in dorsal lateral prefrontal and superior parietal areas. The axial, sagittal, and slice views, along with the 3D surface view, all provide complementary information. While the spatial pattern of nodes in the same module can be useful for recognizing the system, the anatomical view provides little information about the network interactions within and between these different modules. A graph view is a valuable tool for discerning how the different modules interact. The ability to simultaneously observe this network in linked anatomical, graph, and matrix views enables a rich navigational experience, unlike any currently available tool.

In the network under examination here, the executive control network lies between the default mode and visual/sensory networks, see Fig. 3.9(b), top left. The default mode and visual/sensory networks have extremely sparse direct connectivity to one another, as can be seen by the weak inter-modular connectivity in Fig. 3.9(b), top left, and the nearly perfect separation of the green and red nodes in the spring embedded network in Fig. 3.9(c). From this observation we can generate the hypothesis that during successful memory consolidation, the executive network plays a pivotal role coordinating between these networks, potentially by allocating resources for attention and multimodal sensory binding. One node played a particularly critical role in connecting multiple modules, the posterior hippocampus/fusiform cortex (Fig. 3.9(d), green). This node exhibited the highest combination of degree centrality, participation coefficient, and betweenness centrality of any node in the MCN. This suggests that this node plays an important role binding information from communities supporting memory encoding (default mode; green), visual/sensory processing (red), executive function (blue), and motor function

(purple).

3.4 Conclusions and Future Work

We have presented a visualization tool for functional brain connectivity data. The innovative capabilities of our system make it possible to apply visualization techniques to all relevant data representations and use automated analysis methods for detailed exploration of brain network data. Our tool provides focus+context and detail+overview visualizations supporting an interactive, in-depth analysis of communities along with dynamic calculation of the relevant network measures, community membership of nodes, and layout position of nodes. We have used our system to visualize community hierarchies in real-time for datasets up to 200 brain regions. The tool has already proven to be a valuable resource for the neuroscientists in our collaborative team. Since other groups have voiced a strong interest in our tool, we plan to make it publicly available for download.

While already useful in its current state, numerous extensions of our system are possible. For example, currently the tool supports analyzing one graph at a time. We plan to add the ability to read several connectivity matrices, *e.g.*, a resting state network and a network after performing a mental task and display graphs that highlight the differences.

Chapter 4

Multi-scale Visual Analysis of Time-varying ECoG Data Via Clustering

4.1 Introduction

This chapter focuses on advancing visualization techniques to model and understand changes in Electrocortigraphy (ECoG) data. Understanding how epileptic seizures develop or identifying diagnostic indicators for neurological diseases require the in-depth analysis of neural activity data from ECoG. Such data is multi-scale and is of high spatio-temporal resolution. Comprehensive analysis of this data should be supported by interactive visual analysis methods to enable understanding of functional patterns at varying levels of granularity and comprehend its time-varying behavior.

4.2 Motivation

The human brain is a highly connected, dynamic system comprised of specialized brain regions that coordinate and interact in many complex ways for communication, producing intricate patterns of system behavior [90]. Specifically, such spatio-temporal communication patterns

underlying sensor data can help us gain an understanding of the normal functioning of the brain, how we learn or age, and how neurological disorders develop or can be treated [90, 36]. Brain systems function across a large range of spatial and temporal scales. Investigating how the connectivity patterns vary across these different scales has provided new insights into how low-level signals cause global brain state transformations [117]. To support such analysis and capture these patterns comprehensively, data with high temporal and spatial resolution and the low signal-to-noise ratio is needed.

Recent advances in invasive monitoring technologies such as electrocorticography (ECoG) have risen to this challenge by recording high-resolution electrical signals captured by electrodes placed directly on the cortical surface of the brain. The correlation of electrical activity between these electrodes yields a measure of functional connectivity between them. As the derived functional network changes over time, the topology and the attributes of the network vary as well, making it difficult to analyze and visualize the network.

Developments in graph theoretical methods have made it possible to simplify and characterize the data contained in the connectivity network. For example, through community detection methods, it has been determined that brain networks exhibit modular organization [81], *i.e.*, they consist of clusters—subsets of regions having strong inter-modular connections and sparse inter-modular connections. These clusters represent specialized behavioral systems such as higher-order vision, or sensory-motor processing [137].

One way to explore how these behavioral systems interact when performing a task or are impaired due to neurological disorders is to study how the modules evolve over time [36]. This study involves identifying cluster evolution patterns such as: spatial distribution, or a combination of clusters; electrical activation or deactivation of a cluster; and the birth and death of clusters. In the case of epilepsy, for instance, visual analysis of the cluster data combined with the electrical activity can help differentiate normal and ictal (seizure) states of the brain. These patterns—when validated with statistical analysis—are crucial for a successful treatment of the identified epileptogenic zones.

The spatio-temporal patterns in time-varying clusters appear at different spatial and temporal scales. To capture and analyze these patterns, it is important that the temporal scale of the

analysis matches the temporal scale of the patterns themselves [10]. For example, patterns such as spatial distribution or combinations of clusters are best captured at a finer temporal scale while global transitions of brain states are captured at a coarser temporal scale. Analyzing the patterns at varying granularity is crucial as appropriate scales for evaluation are not obvious *a priori* and a single optimal solution at a particular scale is unlikely to exist [10].

Existing approaches to visualize dynamic spatio-temporal clusters operate mostly at a single temporal scale and do not satisfactorily support the in-depth comparison and evaluation of the evolution patterns underlying the data. They mainly focus on visualizing such data by directly depicting all of the information through visual representations or using computational methods to reduce and summarize the visual data. While direct depiction methods suffer from scalability issues, data reduction methods ignore the low-level details of the dataset that are important in explaining high-level evolution patterns.

4.3 ECoG ClusterFlow

To support a comprehensive and detailed study of ECoG data, we present ECoG ClusterFlow (Fig. 4.1), an interactive system that supports the exploration, comparison and analysis of time-varying community evolution patterns at varying temporal granularity through two major views: 1) an overview (Fig. 4.3) summarizing the overall changes in cluster evolution, where users explore salient dynamic patterns; and 2) a hierarchical glyph-based timeline visualization for exploring the dynamic spatial organizational changes of the clusters that uses data aggregation [62] and small multiples [18] methods.

These techniques allow users to gain insights at many levels of temporal granularity, exploring globally evolving patterns to observing small-scale spatial changes. These are our main contributions:

- A hierarchical multi-scale approach to visualizing temporal modular changes in brain networks
- Unique glyph-based designs that explore spatial organizational changes of the dynamic cluster configuration

Furthermore, the specific design goals and capabilities of our system were articulated in

close collaboration with the neuroscientists and the neurosurgeon on our team, ensuring that our prototype improves the overall data exploration process. Our system was repeatedly evaluated and tested by the users, making possible the development of analysis modules that help gain new insights into the data. We present two case studies using synthetic and epileptic seizure datasets to demonstrate the usefulness of our system.

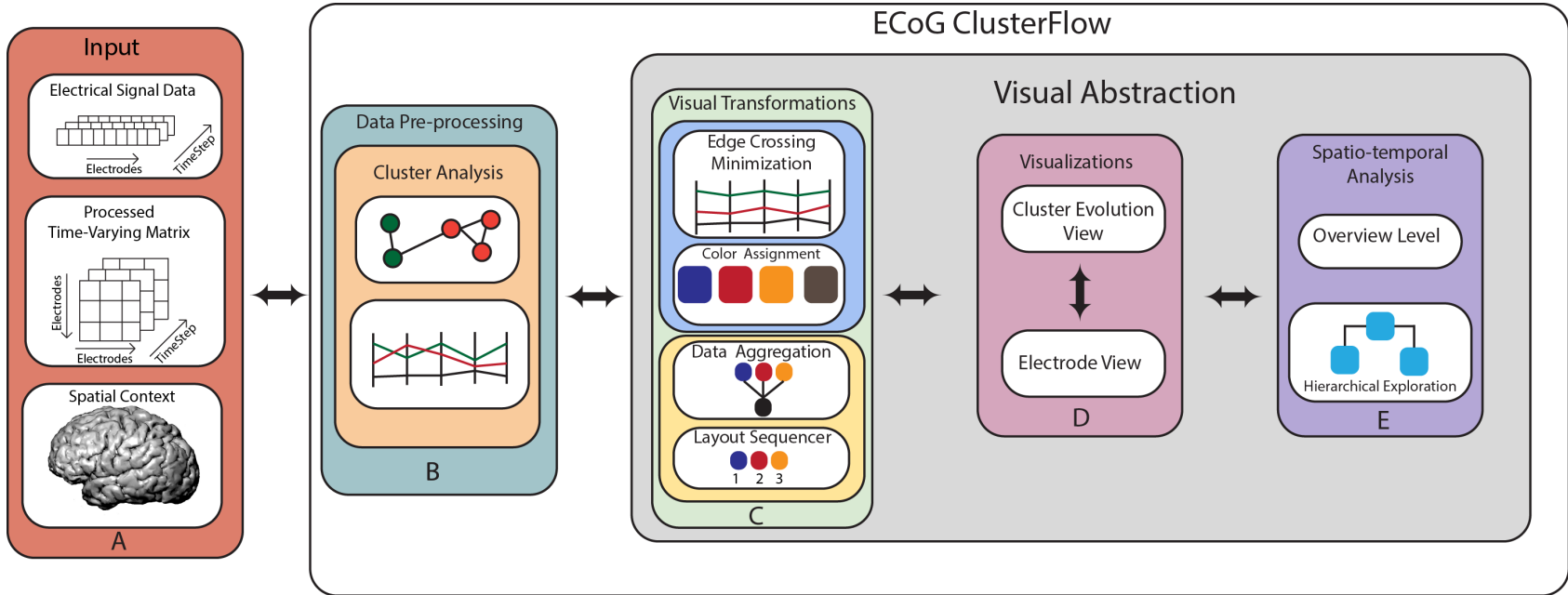


Figure 4.1. Overview of the ECoG ClusterFlow pipeline. A. Raw electrical signals are statistically analyzed to derive the dynamic network data. B. The data pre-processing step identifies and links cluster across timesteps. C. Main modules of the visualization system. D. Users can investigate patterns in two major visualization views. E. Users can perform various types of spatio-temporal analysis based on these views.

4.4 Cluster detection

Our visualization methods are based on sequence of communities detected at each timestep. We call this sequence of communities dynamic communities or dynamic clusters. Given the graph at a particular timestep $G = \{N, E\}$, where N are the nodes that represent electrodes and E are the edges that represents the correlation between the electrodes, the community detection algorithm clusters the data into K non-overlapping and exhaustive communities.

Derivation of time-dependent clusters is an essential task in the analysis of time-varying brain network [170]. Two main approaches [61] are commonly used: 1) A two-stage approach derives communities at each timestep and then tracks them over time using different community tracking methods [75, 157]. 2) An evolutionary clustering approach takes into account the graph topology and the clustering results from previous timesteps. Based on the feedback from the neuroscientists on our team and other existing work [157, 170], we choose the two-stage clustering approach (described in detail in section 4.1.1) with consensus clustering [122] as our primary detection algorithm. This method produces a better quality of clustering results since each timestep is clustered locally (determining the statistically correct number of clusters) [157], and combines the best outputs of multiple runs of the K-means clustering algorithm.

4.5 System overview

We developed ECoG ClusterFlow in close collaboration with neuroscientists (including neurosurgeons) to guide the design of our analysis framework and to ensure that it would be truly valuable as an exploratory tool.

Fig. 4.1 shows the pipeline of our system. The input to our system is: 1) the processed electrical signal data originating from each electrode in the ECoG grid and, 2) its corresponding pairwise dynamic correlation network. The dynamic network data is pre-processed to derive dynamic clusters. Visualization methods, such as data aggregation, are applied to the cluster data in the pre-computation phase and final visualizations are generated.

Based on our conversations with domain experts and the network task taxonomy by Ahn *et al.* [4], we have identified the following domain questions of interest:

Identify temporal brain states, Q1: What activation patterns are consistent over a continuous period of time?

Identify transitions between brain states, Q2: Given the brain states, what patterns characterize their transition to another state?

Compare the evolution patterns associated with different brain states, Q3: What patterns underlie the brain states during normal versus diseased condition?

Assess changes in community membership, Q4: Given a spatial region of interest in the brain, how do the clusters belonging to these regions change over time?

These questions led us to establish the following system design goals:

Timeline-based visualizations, G1: Support views that display the time-varying cluster information on a static display to take advantage of the user’s visual perception instead of cognition (time-to-time mapping)

Multiple levels of detail and abstraction, G2: Support views that enable neuroscientists to explore the data at multiple levels of granularity for analysis

Holistic visualizations, G3: Support visual designs that combine multiple data attributes like cluster membership and its electrical activation

These goals are supported in our system by two major views: the *Cluster Evolution View* and the *Electrode View*.

4.5.1 Cluster Evolution View

The cluster evolution view (Fig. 4.2) highlights the salient patterns of the cluster evolution including the emergence, death, contraction, expansion, merging and splitting of clusters (Q2, Q3, Q4). Through this view, analysts can compare and analyze modular signatures (cluster evolution patterns) over time and identify important time intervals and distinct brain states. The cluster evolution patterns are represented using a flow-based visualization [148, 187] (G1) (alluvial diagram), where the clusters metaphorically flow like a river with split/merge tributaries from left to right.

Formally, at each timestep t on the horizontal axis, rectangular blocks represent clusters $C_{t,i}$ where the height of each block corresponds to the cluster’s size at that timestep. Flow-based transition links $L_{i,j}$, where i is the source community and j is the sink community, connect

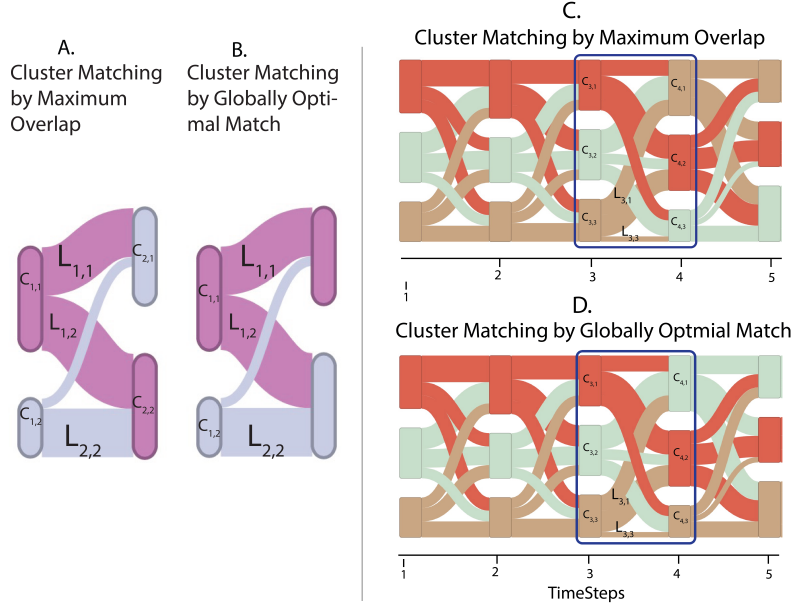


Figure 4.2. Comparison of tracking algorithms in artificial datasets. In image C the maximum overlap algorithm pairs $C_{3,3}$ to $C_{4,1}$, while in Fig. D the globally optimal matching algorithm pairs $C_{3,3}$ to $C_{4,3}$, qualitatively making communities more visually traceable in image D. The scalar values for the links in image A and B are $L_{1,1} = 11$, $L_{1,2} = 11$, $L_{2,1} = 10$, $L_{2,2} = 2$.

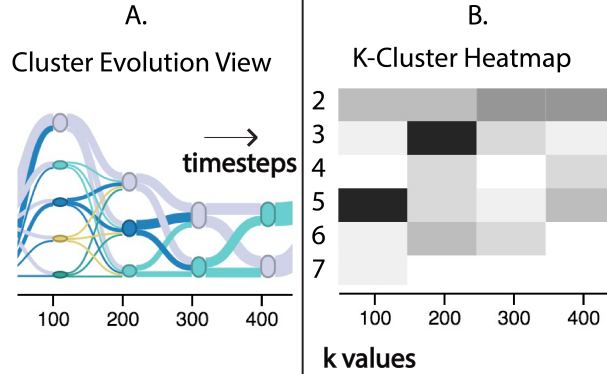


Figure 4.3. Evolution of clusters for four timesteps. A. The cluster evolution view shows clusters and transitions between them. The nodes have colors based on their cluster membership. B. The K-cluster heatmap on the bottom visualizes the likelihood of a range of K values that determine the final number of clusters for a particular timestep, for *e.g.* a clear maxima is evident for 100ms (K as 5) and 200ms (K as 3).

clusters to show changes in the community structure over time. We model these links as Bezier curves, to generate a continuous representation of the transition between successive communities [187]. Fig. 4.2 shows the evolution of dynamic clusters for five timesteps. Furthermore, to easily assess the community membership in dynamic clusters, we color communities using solid coloring, using N perceptually distinct colors from a qualitative colorbrewer [79] colormap.

4.5.1.1 Cluster tracking

To support the two-stage cluster detection approach, it is necessary to determine correspondences between clusters in consecutive timesteps. Based on the input from neuroscientists, we have investigated two approaches to compute this matching: (1) maximum overlap tracking and (2) computing the globally optimal match.

Maximum overlap tracking (Fig. 4.2A, C) is a greedy algorithm that iteratively matches the two clusters in consecutive timesteps that share the maximum number of electrodes. This process is repeated until no overlapping clusters remain. This approach may not always produce an intuitive correspondence between clusters. For example, in Fig. 4.2A and C, clusters $C_{1,2}$ and $C_{2,2}$ have maximum overlap (of 11 electrodes) and are paired in the first iteration. This only leaves $C_{2,1}$ as possible match for $C_{1,2}$ in the second iteration, even though the overlap between $C_{1,2}$ and $C_{2,1}$ is relatively small (only two electrodes).

To find a globally optimal assignment, our second approach determines an optimal overall match between all clusters in consecutive timesteps. We define a similarity measure

$$sim = \frac{|C_{t,i} \cap C_{t+1,j}|}{|C_{t,i} \cup C_{t+1,j}|}$$

between clusters C_i and C_j in consecutive timesteps t and $t+1$, similar to the approaches by Greene et al. [76] and Sallabury et al. [157]. Next, we compute a similarity matrix comprised of the pairwise similarity measures between all possible cluster combinations. To avoid matching of clusters with small overlap, we set to zero those similarity values that are below a threshold θ . To match clusters, we consider all possible cluster matchings between timesteps—by considering all possible permutations of clusters—and compute a global similarity value as the sum of the similarity values for all matched clusters. The overall best match is the permutation that maximizes global similarity. While considering all possible permutations is computationally expensive, we usually consider only a small number of clusters (approximately seven) per time step, keeping this approach tractable. Fig. 4.2B shows the best overall match for our example, matching clusters $C_{1,1}$ and $C_{2,1}$ as well as clusters $C_{1,2}$ and $C_{2,2}$, a more intuitive choice than the result obtained by maximum overlap tracking. Figure 4.2 shows an example of this approach for an artificial dataset with three dynamic communities.. The two approaches differ

in the community results starting at timestep four.

4.5.1.2 Sorting and Ordering of Nodes

To enhance the visual traceability of the clusters, the node layout of the graph should ideally minimize edge crossings with optimal ordering of the nodes (clusters) at each vertical axis. To determine such an ordering, we must take all the timesteps into consideration. Several methods have been proposed to compute such an ordering [187, 157]. Our approach handles more timesteps by not considering the individual elements contained within clusters and dividing the sorting procedure into N individual blocks of T timesteps. To reduce the computational complexity—to achieve the least start-up-time of 40-60 seconds and to scale to up to 60 timesteps—our heuristic solution (barycenter approach [174]) sweeps horizontally across all the clusters over N blocks of T timesteps (where NT is the total number of timesteps in the data) in a front-to-back and back-to-front manner to optimize the order and position of these communities on the vertical axis. This procedure results in a cluster ordering that minimizes the link distance between the timesteps.

4.5.1.3 K-Cluster Heatmap

Cluster analysis results can be sensitive to noise and prone to overfitting [122]. The K-Cluster heat map produces important information for the evaluation of the uniqueness of the number of clusters detected per timestep. Consensus clustering uses a cumulative distribution function (CDF) to determine an appropriate number of clusters K . In most cases, the likelihood for a single value of K will be large compared to the others, and the confidence that the chosen number of clusters is correct is high. The K-Cluster heat map (Fig. 4.3) shows the likelihood for a range of values of K for each time step. Black denotes high likelihood and white low likelihood. Using this heat map, analysts can identify timesteps where multiple values for K are almost equally likely and where confidence in the clustering results is low.

4.5.2 Electrode View

The electrode view shows (Figure 4.4) cluster membership and electrical activity in a spatial context (G3) (Fig. 4.9 A,B), enabling the user to identify important spatial cluster evolution patterns (Q1, Q2 and Q3). These evolution patterns (Fig. 4.5) include 1) *spatial cluster distribution*, i.e., a cluster originally comprised of spatially adjacent electrodes splits into disjoint

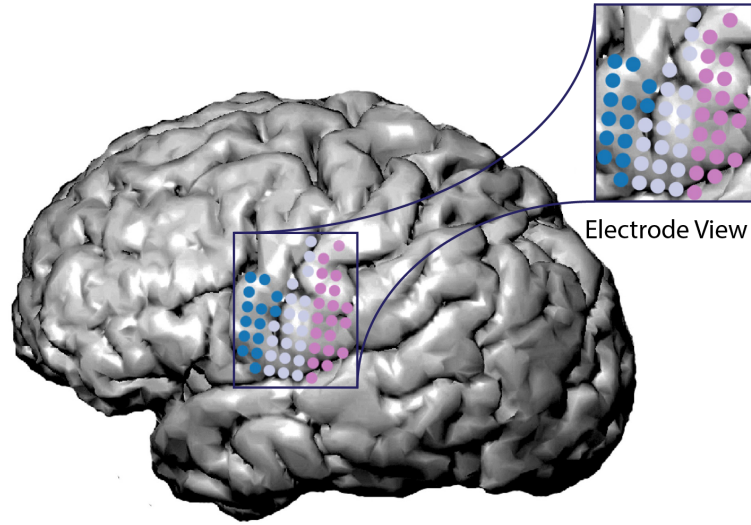


Figure 4.4. To save display space, our tool crops the electrode view to the region of the brain where electrodes are placed.

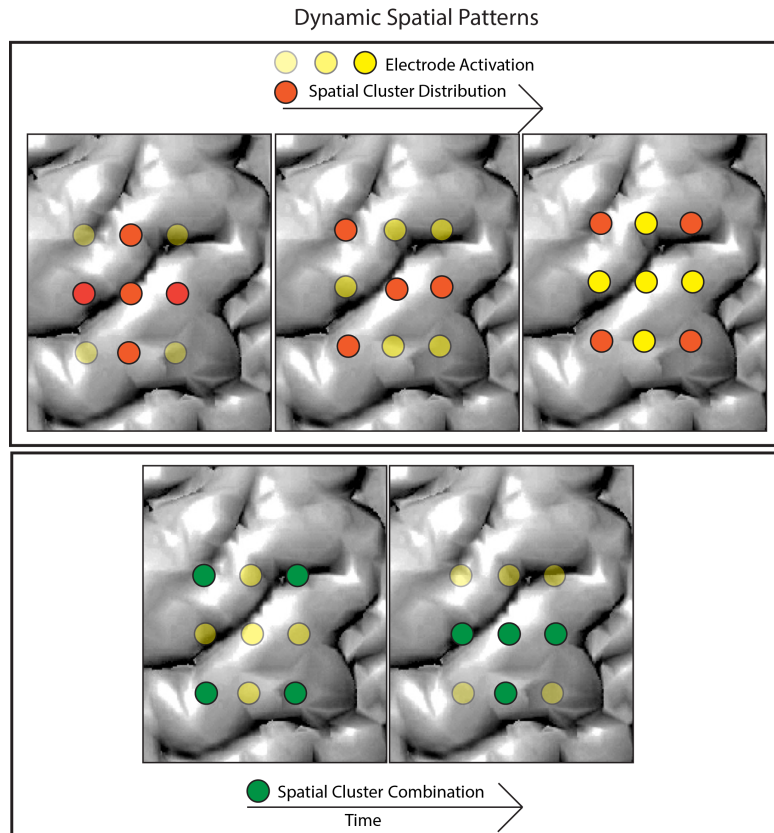


Figure 4.5. Dynamic spatial patterns captured by the electrode views. The colors of the glyphs represent clusters, and the opacity of the glyphs denote electrical activation. The orange cluster in the top, spatially distributes, while the green cluster in the bottom, spatially combines. The yellow cluster in the top activates over time.

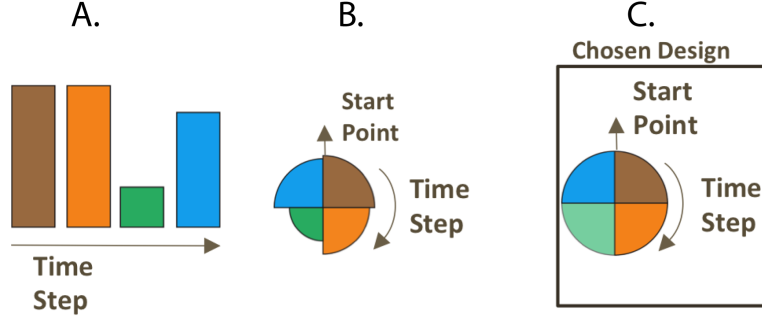


Figure 4.6. Design choices for visualizing cluster membership along with brain electrical activity. A. The height of the individual bar correspond to the electrical activity and the color its cluster affiliation, B. The radius corresponds to electrical activity and the color its cluster membership, C. The opacity of color represent the electrical activity and the color its cluster membership.

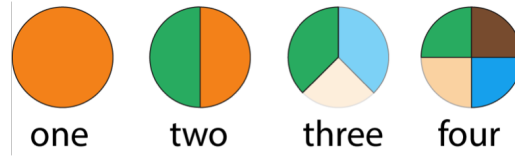


Figure 4.7. Varying the number of timesteps displayed on our glyph design.

parts, 2) *spatial cluster combination*, i.e., a cluster consisting of disjoint regions becomes spatially coherent, and 3) *spatial activation*, i.e., the electrical activity of electrodes increases over time. The electrode view places interactive glyphs—representing electrodes—on a 2D sagittal projection of a subject-specific reconstructed brain MRI model to provide the spatial context.

To support exploration at multiple temporal scales (G2) and effective comparison of cluster characteristics—i.e., cluster membership and electrical activity of the individual electrodes—over the spatial domain, our visualization method aggregates N multiples of continuous timesteps on each electrode view. Furthermore, interactive techniques that can be applied to glyphs, make it possible for the user to define the N multiples to be aggregated on each glyph.

The aggregation techniques provide the user with a detailed overview of the underlying data while not being overwhelmed by the entire data [56]. However, aggregated views have a drawback. They can lead to cognitive information overload. Nonetheless, for the multiples we consider, we found that our visualizations are better suited for the tasks performed by our collaborating neuroscientists.

4.5.2.1 Glyph Design

All our glyph designs display n user-defined timesteps per individual electrode view, to save presentation space and to facilitate comparison between timesteps. We considered three visual designs for our glyphs, following some of the data aggregation guidelines specified by Elmqvist *et al.* [56]. The first design uses vertically stacked bar charts. Each bar represents one time step with its color indicating cluster membership and its height corresponding to electrical activity, see Fig. 4.6A. The second design uses a clock metaphor [97] and subdivides each glyph into n equal slices. Each slice—starting at the top in clockwise order—represents a time step with its color indicating cluster membership and its radius representing electrical activity, see Fig. 4.6B. The third design is a variation of the second design and uses slice opacity instead of radius to represent the electrical activation level, see Fig. 4.6C.

The first design depicts changes in cluster membership intuitively, but requires a large amount of glyph space. In the second design, the varying glyph sizes in the entire spatial layout impede the neuroscientists to assess the relative positions of the electrodes. Overall, our domain science collaborators preferred the third design where the glyph shape remains constant and only color and opacity are used to convey information. We use this design in our system and the remaining discussion in this paper is based on it.

4.5.2.2 Hierarchical Exploration

Patterns in brain activity occur at different temporal scales, where the appropriate scale may not be known *a priori*. To facilitate discovery of these patterns and scales, our tool controls the timesteps that can be displayed in a single electrode view, see Fig. 4.7. At the extremes, the method either displays all timesteps in a single electrode view (low granularity at the top level) or each time step in a separate electrode view (high granularity at the bottom level). Between these extremes, different levels of aggregation are possible. Different approaches exist to search for temporal evolution patterns. *Top-down* analysis starts with all timesteps in a single view and decreases aggregation until a pattern of interest is found. A *bottom-up* approach starts analysis with each time step in a separate view and increases aggregation until a pattern is found. However, exploration may also start at *mid-level*, in situations where the user already has a notion at what temporal scale a pattern may be identified.

4.5.2.3 Layout

To identify low-level changes in a bottom-up based approach, several low-level electrode views need to be examined simultaneously. The system presents the views in a from-left-to-right order, synchronized with the timeline. Visual scalability is problematic as screen resolution is small relative to data resolution. Therefore, ECoG Cluster Flow utilizes a space-saving layout to achieve a tight synchronization between the spatial (electrode) view and temporal attributes (cluster evolution view) of the data. Each electrode view is ordered alternatively above and below the cluster evolution view, see Fig. 4.8, timesteps 3–10, to achieve the desired integration. The cluster evolution view is expanded or contracted based on the combined space allocated to all electrode views. To emphasize the granularity of the electrode view, the space assigned to each electrode view is proportional to the corresponding number of time points.

We use the formula

$$w_i = w_{min} \times S_f, \quad S_f \geq 1,$$

to define the width w_i of each electrode view, and

$$S_f = C \times \frac{W_{max}}{w_{min} \times \max(N_{x_1}, N_{x_2})}$$

to define the expansion factor, S_f , for each electrode view. The variables N_{x_1} and N_{x_2} represent the numbers of the electrode views for the x_1 – and x_2 – axes, representing the “above” and “below” cluster evolution views, respectively. The value of W_{max} is the maximally possible horizontal screen display width, C is a user-specified constant, and w_{min} defines the minimal size of each electrode view.

4.5.2.4 Interactivity

The strength of our system is the fact that it supports fast and intuitive analysis of cluster data at interactive rates. Based on previous work [200] and the suggestions made by our collaborating neuroscientists, we selected interaction techniques satisfying the user-defined design objectives. These are the interaction principles satisfied by our system:

Overview first, then details on demand: The cluster evolution view is an intuitive way for users to first obtain a simple outline of the entire dataset. Overview trends and outliers can be

easily captured, enabling users to quickly determine a time interval of importance, and perform detailed exploration on the visualization.

Focus+context: This technique [45] allows users to focus on small-scale evolution patterns while preserving the overall context. It utilizes smooth, animated changes to track the patterns in focus. Furthermore, the technique facilitates comparison of patterns across time-intervals over the temporal and spatial domains.

Highlighting and linking: The cluster evolution view and electrode views are coordinated using brushing and linking. Users can select timesteps of interest and observe the evolution of patterns via the cluster evolution view or the electrode view. All visual elements are supported by simple tooltips providing information about the meaning of visual encodings.

4.6 Case Studies

We discuss the usefulness of our tool by considering two case studies done in collaboration with the neuroscientists and neurosurgeon on our team of co-authors. We cover two datasets in our case-studies, a synthetic and real-world seizure dataset. The real-world dataset is complicated, containing complex cluster patterns. We have therefore generated a simple, synthetic dataset with known patterns to evaluate how our visual analysis method performs in such a well-understood case. The case studies are meant to serve the purpose of demonstrating the value of our system for gaining relevant scientific insight.

4.6.1 Synthetic Dataset

In our setting, electrodes have (1) three known modes of electrical activity, (2) known intervals of activation patterns, and (3) known likelihood values for the K-cluster heatmap. The controlled data parameters allow us to investigate the features of the visualization outside the context of noisy brain recordings.

We created dynamic network with 54 electrodes with activation values 0.9, 0.6, 0.3 activation values for 30 time steps, and we specified locations of the electrodes with our system. We kept the number of clusters constant for each timestep, and we defined initial clusters for all electrodes in the 30 timesteps.

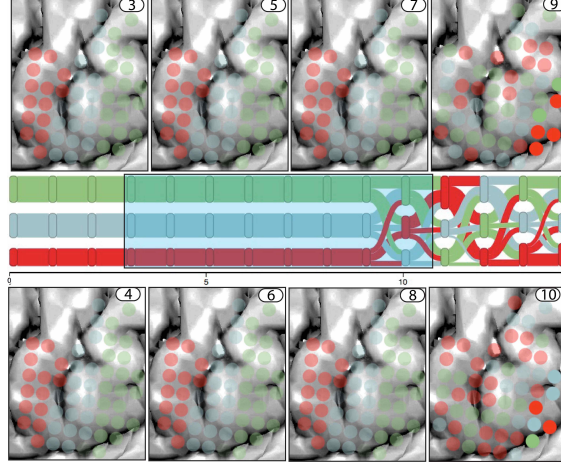


Figure 4.8. Layout for bottom-up analysis of the spatio-temporal data. The cluster evolution view is divided into equal segments and corresponds to the electrode views shown along the two horizontal axes, i.e., above (x1) and below (x2) it.

4.6.1.1 Spatio-temporal Analysis:

We start our data exploration with the evolution view, looking for general evolution trends in the data. In this view, three distinct clusters (colored in red, green and blue in Fig. 4.9) emerge and remain stable throughout timesteps (0-9) (in Fig. 4.9B). These clusters become randomly distributed at (10 to 19) to regain their stable configuration at (20 to 30).

To further examine the spatial configuration of these patterns, we employed a top-down approach exploring various temporal scales progressively to identify a consistent activation pattern (similar activation patterns in one electrode view) across all electrode views. At a granularity of ten (ten-time points in one electrode view, Fig. 4.9A), persistent electrical patterns in each electrode view were found, *e.g.*, glyphs in electrode views *one* and *three* were fully activated and deactivated, respectively. The electrode view *two*, on the other hand, showed a combination of activated and unactivated patterns.

To explore the intricate low-level activational and modular patterns that caused the temporal state change from unactivated to activated state, the granularity of the system was reduced to one (Fig. 4.9B). When examining timesteps (9, 10, 11), an emergent focal activation point (annotated in Fig. 4.9C) in the lower-right corner of the electrode view was evident. Further examination of subsequent timesteps revealed the progressive dominance of the red cluster over the region (Fig. 4.9C, views 9, 10 and 11). In summary, our combined visual analysis approach

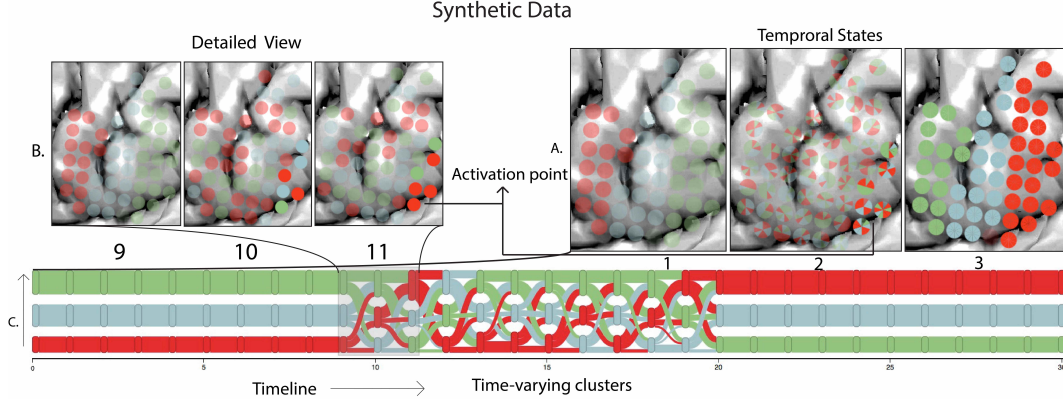


Figure 4.9. Evolution patterns underlying our generated dataset which we use to test and evaluate our approach. Color indicates the cluster configuration at each timestep and the opacities of the glyphs its electrical activity. A. Categorizing different temporal states, i.e., unactivated (View 1), transitional (View 2), activated (View 3). B. Evolution of the cluster assignment changes through the cluster evolution view, stable cluster configuration from timestep intervals (0-10) and (20-30), and, random unstable assignment over timestep interval (10-20). C. Detailed analysis of a time interval selected in the cluster evolution view, activation patterns can be seen in the lower-right corner of the views for the timestep interval (10,11).

helped us categorize temporal states and identify low-level changes and their dependencies with high-level state changes (Additional file 1).

4.6.2 Epileptic Seizure Dataset

Epilepsy is a neurological condition where the normal functioning of the brain is disrupted due to sudden bursts of electrical activity emanating from a certain region of the brain, *i.e.*, seizure-initiating foci. This disruption is characterized by changes in the brain’s modular organization over time [135]. Exploring these differences may provide insight into the genesis and development of the seizures over time [135]. The neuroscientists on our team are primarily interested in: 1) identifying the focal site of seizure genesis and its spatial progression over various stages of the seizure and 2) identifying distinct spatio-temporal evolution patterns that characterize the onset and propagation of the seizure.

Data: The raw signal data from the ECoG electrode array was statistically analyzed to provide two spatio-temporal graph datasets using different pre-processing steps: a high-gamma dataset at a frequency ranging from 70 to 170 Hz, capturing multi-unit neuronal spiking, and a full-range dataset, averaging all frequencies captured by the recording device. We derived communities independently at each timestep using the consensus clustering algorithm [122] that

automatically fits the number of clusters detected. We note that the issue of extracting the *correct* communities from time-varying graphs is more of a question of the statistical data analysis algorithms used in the graph formation and community detection algorithms, than the visualization thereof. With these clusters as input, our system detects and visualizes the dynamic community results in the cluster evolution and spatial view.

4.6.2.1 Domain Expert Analysis:

We now discuss some of the major insights obtained and details of the usage of our system by our collaborating neuroscientist and the neurosurgeon on our team:

Detection of brain states: Using the bottom-up approach, around a temporal scale of ten, we found consistent progressive seizure activation patterns across all electrode views (Fig. 4.10C). Based on these cluster patterns, we categorized the electrode views into four major distinct brain states, i.e., *before-seizure*, *early-seizure*, *mid-seizure*, *late-seizure*.

Detection of transitions between brain states: Based on the gradual changes in activity patterns (2200-2600ms in Fig. 4.10A), the seizure initiation zone (2200ms, Fig. 4.10A) (cause for state change from *before-seizure* to *early-seizure*) was found. The neuroscientists stated, “We could see activated opaque glyph patterns (Seizure Foci) over the Lateral Temporal Cortex both in Fig. 4.10A, and in Fig. 4.11A at 2400ms. The opaque glyphs emerging from this point dynamically spread to other parts of the brain (3000ms-5000ms in Fig. 4.10C).”

Compare and contrast the evolution patterns governing different brain states: The signatures of cluster evolution at three distinct brain states (before-seizure, transition state, late-seizure) were of interest to us.

1) *800-1100ms*: Evolution patterns here appear to be stable and organized (highly confident clusters with a clear maxima in the K-cluster heatmap), with the same number of elements in each community (Fig. 4.11A,B). The spatial organization of clusters at 900ms (Fig. 4.11A) are scattered over the spatial layout and are unactivated with transparent glyphs.

2) *2400-2700ms*: Evolution patterns from this point onward seem to become irregular. There is a noticeable activation of certain electrodes in spatial view and reduction in number of clusters. The neuroscientist commented, “The transitions in the cluster evolution view have a mix of irregular distribution of thin and thick links between timesteps. For example, the

brown cluster in Fig. 4.11A and Fig. 4.11B have comparatively less redistributions of electrode elements over time.”

3) 5200-5500ms: Concerning this period, the scientist’s comment was: *“In the cluster evolution view, a dominant orange ribbon (starting at 4000ms) emerges with few isolated small sized communities.”* The zero likelihood in the heatmap (interval 5500-6000ms) is caused as the algorithm produces only identity (1) values in the consensus matrix, showing no progressive change in the cumulative density plot (section 4.1.4), eventually picking the K value to be one.

Assess changes in community membership: In Fig. 4.10C, when evaluating the last electrode view, *there also seems to be a dominance of a single orange cluster over a significant period of time and space.*

When comparing the evolution patterns of the high-gamma and full-range pre processed datasets (Fig. 4.11A and Fig. 4.11B), the domain experts stated that, *“In general, the full-range dataset exhibits larger numbers of communities in all three sampled intervals (A,B,C). Although a single orange ribbon emerges in both datasets, it only converges clearly in the high-gamma dataset.”*

4.6.2.2 Domain Expert Review

Our collaborating neuroscientists were satisfied with the system and regarded it as intuitive and easy to use. After brief initial training, describing the visual designs and different interactions, they were able to identify regions with coherent cluster configurations for different parameter settings. The ability to switch between different views and drilling down to a level of interest was seen as very helpful by them. One comment was *“I could quickly identify general patterns of interest in the complex dataset and pick the outlier regions of interest through the powerful interactive techniques. The derived patterns are promising and reveal relevant insight into seizures, especially their genesis and overall characterization.”* Further, *“the representation of time-varying community structure and combined with the interactive multi-scale exploration approach has made possible multi-part observations for the ECoG spatiotemporal data. Without this tool, this process would have been very time-consuming and cumbersome.”*

4.7 Discussion

Neuroscientists analyzing and visualizing spatial-temporal graph data commonly use juxtaposed small-multiples or animations of node-link diagrams with a corresponding spatial view (brain maps). We have compared our approach with these baseline visualizations methods and point out how we gain new insights through our interactive system.

Small-multiples of brain maps: To identify low-level community membership changes of a node over time, information in each discrete timestep and its corresponding brain map must be examined. Such static multi-view approaches make it difficult to quickly assess the stability of nodes over a local spatial region of interest. Furthermore, to identify brain states with consistent activation patterns and investigate the global cluster lifetime phenomena, nodes have to be mentally grouped and compared based on their data attributes, further increasing the cognitive load for analysis.

Animations of brain maps: Animations are effective in capturing low-level spatial changes of communities over a small period—a change in community membership or activation is denoted by a sudden change in color. However, the frequently changing data attributes make it challenging to infer meaningful evolution patterns over long time-scales. This complication is due to its reliance on short-term memory that requires users to remember previous static views and then manually identify and compare the relevant changes. The same holds true for identifying brain states or global cluster evolution patterns.

Our approach: The identification of community membership changes of a node can be achieved by choosing an appropriate level of granularity and visually analyzing it over one static view (Fig. 4.10B and C). The analysis is reduced to finding the changes in the color of the consecutive slices in the glyph. Furthermore, the split-merge links in the cluster evolution view ensure that the nodes do not need to be re-identified at each timestep (small multiples) or remembered from previous timesteps (animations). Compared to these methods, interactive techniques in our approach help filter, select, zoom, pan and increase or decrease the level of granularity, help gain a thorough understanding of the significant spatio-temporal patterns in the data. Unlike the small-multiples method, our method significantly reduces the visual scalability by aggregating temporal data over the spatial electrode view.

4.7.0.1 Findings

In the following, we summarize our findings concerning the use of our system for exploring and analyzing ECoG data.

Timeline-based representations enhance perception: The tightly integrated layout mechanism, where a consistent global timeline between the cluster evolution view and the electrode views is maintained, allowed neuroscientists to correlate evolution patterns across the spatial layout and functional organization of the brain network. Furthermore, the side-by-side placement of the views on a timeline helped the scientists with the comparison of relevant salient patterns for discontinuous timesteps.

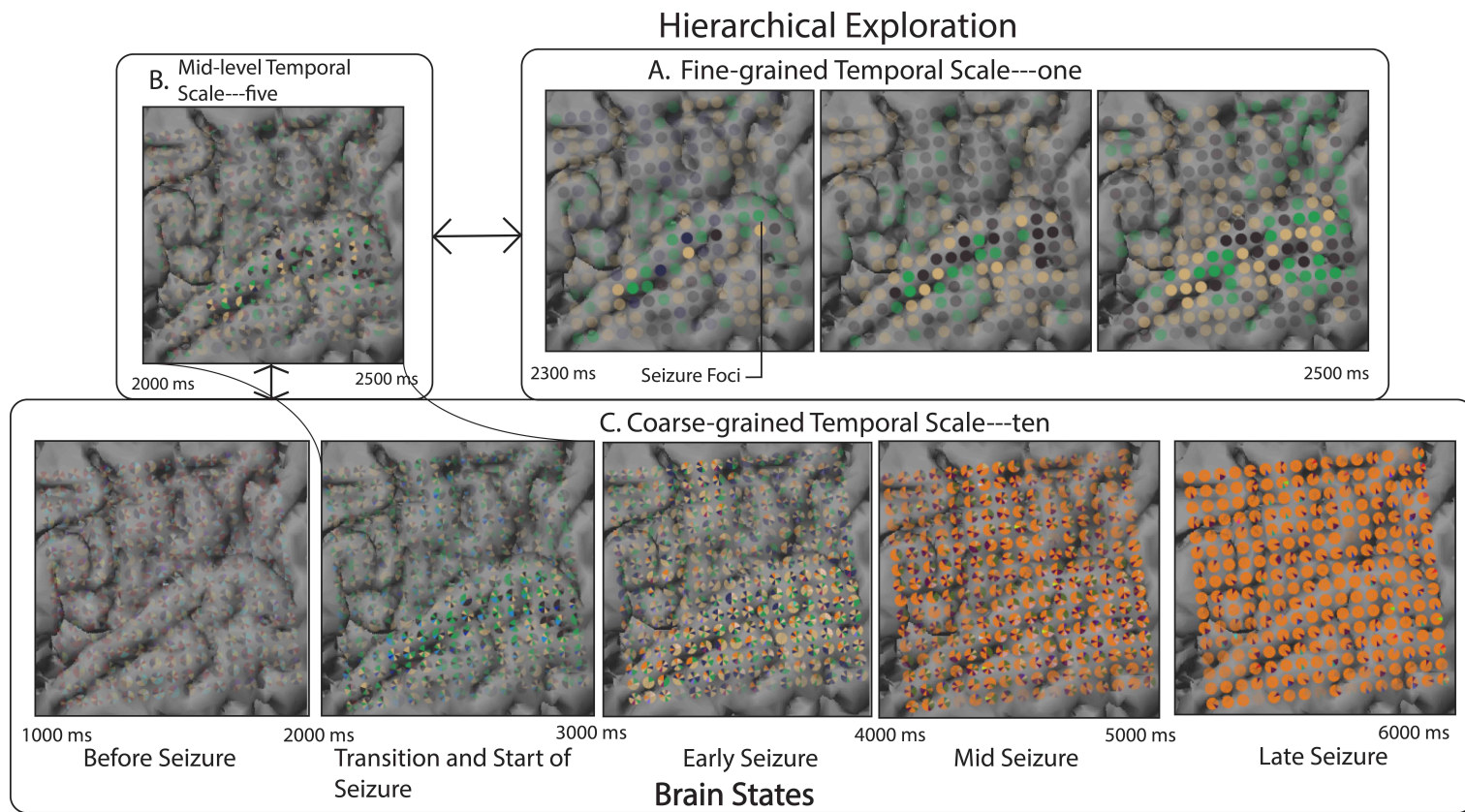


Figure 4.10. Hierarchical exploration of the seizure dataset at varying levels of granularity revealing the various brain states and associated cluster characteristics. Four major brain states are found at C. Upon further examination of their detailed evolution patterns in B and A, provides insight into seizure genesis and the initiation phase. The green cluster (in seizure initiating foci, i.e., ‘Lateral Temporal Cortex’) in A seems to play a prominent role in the seizure initiation phase (2200ms-2500ms).

Bottom-up analysis versus top-down analysis: The neuroscientists preferred using different approaches for different tasks. The bottom-up analysis approach (Fig. 4.10, A-B-C) was preferred over the top-down approach (Fig. 4.10, C-B-A) for tasks relating to Q1 and Q2. As the bottom-up exploration method generated views with increasing complexity, relationships between evolution patterns at various scales could be seen clearly. In contrast, the top-down exploration method, displaying fewer electrode views with high levels of aggregation, was cognitively overwhelming.

Comparison of evolution patterns of multiple datasets: In our case study, we have compared evolution patterns prevalent in two datasets. Sudden changes in cluster membership and activation data were apparent when comparing the cluster evolution view and the electrode view for the datasets. However, as both datasets have different cluster characteristics, it was difficult to determine detailed changes between the datasets. A uniform comparison criterion should ideally assign colors based on an optimization algorithm, where the same colors are used for the dynamically evolving dataset clusters that intersect. We plan to devise a method to achieve this in the future.

Scalability is a main design consideration that we took into account when designing ECoG Cluster Flow. Hierarchical representation/visualization and data aggregation methods are primary techniques that we employ to explore the high-resolution spatio-temporal data. We have visualized up to 64 timesteps and 256 electrodes. A challenge will be to scale the system’s real-time capability to much larger datasets. Our goal is to support interactive, real-time visual analysis for datasets with more than 500 timesteps. Further, to keep the amount of visual representation at a level that can be comprehended by a human, we will devise and employ data reduction and abstraction techniques to simplify data visualization.

4.8 Conclusions

We have presented ECoG ClusterFlow, a hierarchical multi-scale approach for visualizing spatial and functional cluster evolution patterns. Our approach has allowed neuroscientists to investigate the major cluster evolution patterns over space and time. Through our approach, it is possible to examine whether the major evolution events were a result of noise or a sudden

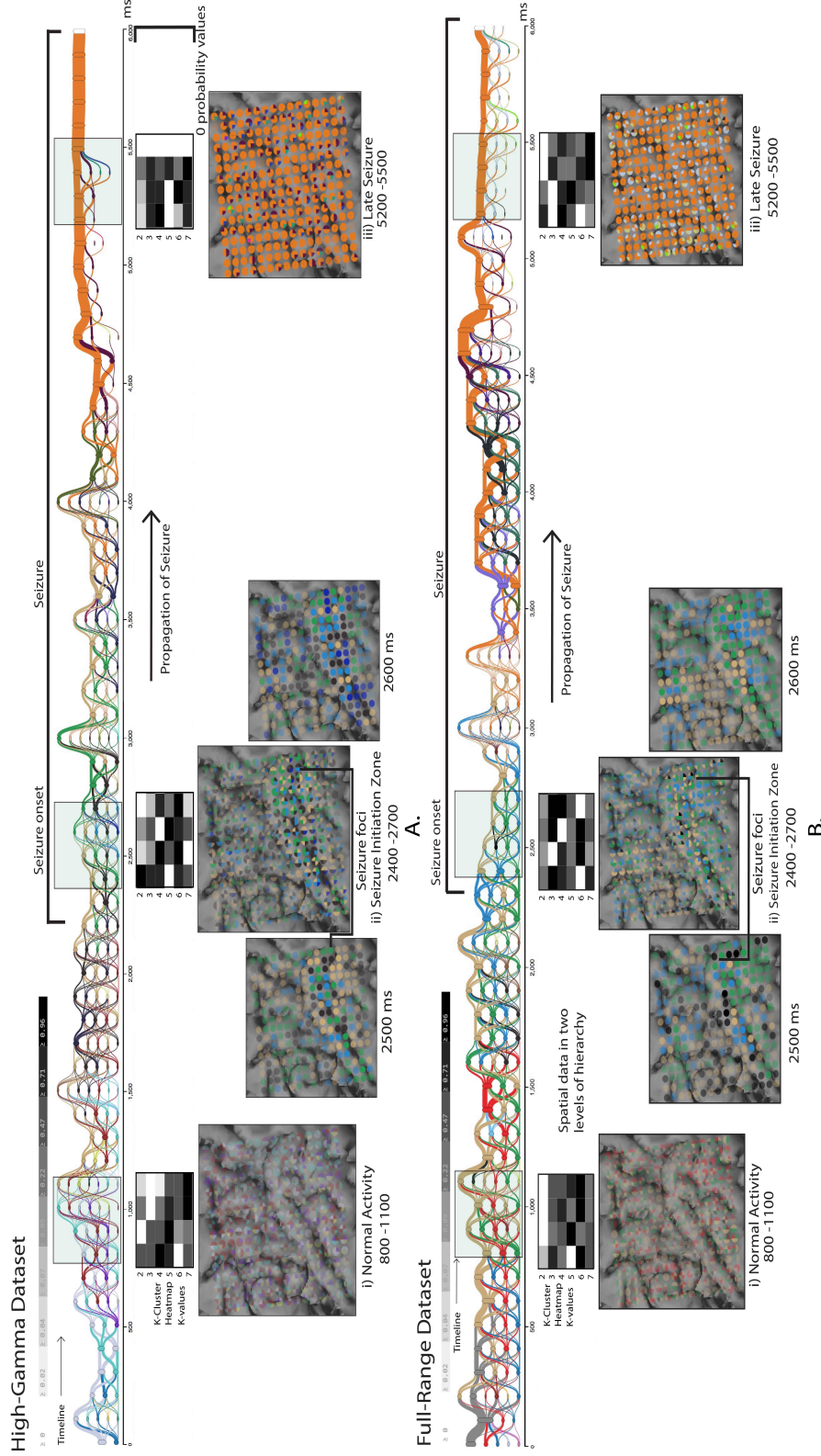


Figure 4.11. Visualizations of the two pre-processed datasets of the same raw input signals. Colors in the cluster evolution view and the electrode view indicate the time-varying communities. Three intervals are chosen for analysis to explore the modular signatures at different brain states. Significant evolution patterns at these stages are 1) normal activity—stable cluster characteristics with equally sized modules, 2) transition state—distinct activation patterns emerging from the ‘Lateral temporal cortex’, 3) late seizure state—single dominance of the orange cluster. When comparing the modular signatures between images A and B, the number of clusters in B seems to be larger

change in functional or spatial properties of the network. Furthermore, we have discussed major neuroscience-driven data analysis tasks and design choices that led to the design of the entire system, done in close collaboration with neuroscientists, insights for the spatial cluster evolution data.

The visual analysis approach greatly supports the comprehension of salient spatio-temporal evolution patterns and provides insight into the life-span of brain states and the clustering stability of the classification algorithm.

We plan to perform an evaluation concerning the usefulness of spatio-temporal clustering, *i.e.*, a clustering algorithm that takes both the spatial as well as the temporal attributes into account. When utilizing both these attributes, interesting research questions arise, such as, “*is the resulting visualization an accurate representation of the data? How can we consistently compare such types of visualizations with a different dataset?*”

Chapter 5

Similarity and Difference Visualization of Dynamic Networks

5.1 Introduction

The previous two chapters focused on research involving visual analytics techniques for cluster exploration in static and dynamic brain networks. However, temporal scalability is still a very important problem in dynamic graph analysis that is not satisfactorily solved.

Networks are often time-dependent, evolving over several hundred timesteps. Network changes are reflected through the addition or deletion of edges, the birth or death of communities or modularity variation. In most cases, such dynamic changes do not affect all the structures in a network, and many core structures (subnetworks that are constant) remain stable over certain periods of time. We call phases with similar local and global topology as *states* of a network. If we can detect and understand these states and core structures, we will be able to answer questions related to their role in the context of an entire system's evolution.

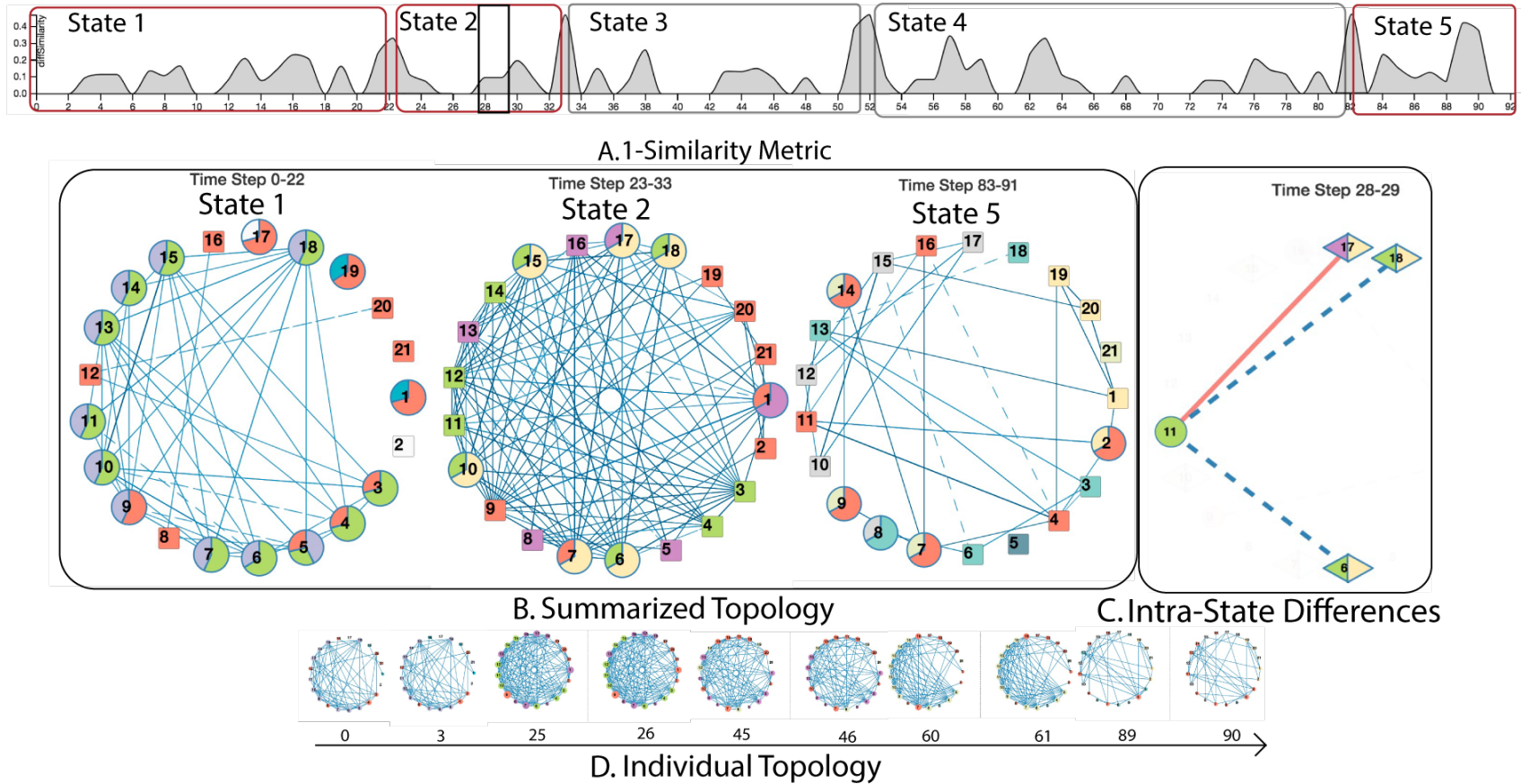


Figure 5.1. Visualizations of a dynamic brain network including summary and difference representations. A. Identification of states. Each of the five states is defined by a time interval corresponding to a relatively stable topological structure. B. Visualizing topology in summary form. Each states possess relatively consistent local and global structures. The brain network evolves from sparse (state 1) to dense (state 2,3,4) to sparse (state 5). C. Topological difference between networks within a state. The disappearance of an edge (from brain region 11 to 18 and to 6) affects global community structure, depicted by the change from green to yellow in brain regions 18 and 6. D. Visualization of original network using small-multiples method.

For example, neuroscientists are keenly interested in such states to comprehend interactions that relate cognitive functions to the functional states; social scientists, studying classroom interactions, are concerned about the correlation of communication patterns and the educational outcomes.

The complex dynamic behavior (within and between states) of a network affects its local and the global topology. Identifying the dependencies between these properties over time can provide us with high-level information about the causes and effects of evolution behavior. In other words, a minor change, such as the addition or deletion of an edge, might explain an entire global state change of a dynamic network. For example, in a classroom social network, the removal of a communication channel between two popular students (who are linked to many other students) could explain a major shift in social dynamics of the classroom.

5.2 Motivation

Analyzing and depicting such data in a comprehensible manner for large, dense networks remains challenging as purely visual approaches become inadequate [94, 9] and require algorithmic techniques to meaningfully present and abstract the data without actual information loss. Many existing approaches that visualize dynamic networks and their community structures depict an entire network and all its changes for every timestep. Such techniques can lead to duplication of visual elements that represent nearly the same information (due to temporal correlation). Further, they require large screen-space for visualization. For example, an ECoG data array capturing neuronal activation of brain regions for just one hour may need several thousand copies of the original network, resulting in a representation with several million nodes and edges. Moreover, most current methods do not effectively show changes in topology between timesteps or their effects on the overall system. Analyzing these changes are essential for understanding the complex time-varying phenomena in dynamic social networks or neurodegenerative diseases, *e.g.*, schizophrenia [36].

We introduce a novel summarization and difference approach that detects important topological changes in the data and abstracts and visualizes similar local and global network structures, *i.e.*, states (Fig. 5.1A). We summarize and represent the underlying dynamic behavior for

each of the detected states by computing a representative summary graph using unique glyph designs (Fig. 5.1B). To describe and explore the dependencies between local and global properties of network states, we introduce a novel network difference technique, see Fig. 5.1C, facilitating holistic comparison of topologies. The major contributions presented are:

Methods that compute similarity and difference topologies in temporal states of a dynamic graph.

Effective visualizations of network graphs with similar and different topologies.

5.3 Design Goals

Based on existing literature regarding dynamic network visualization [23], we identified primary design goals to be met by our approach. We designed our system based our approach on these goals:

Simplification of temporal features Topological patterns in dynamic networks often re-occur, *e.g.*, brain network patterns. Methods are needed that can aggregate data by considering and taking advantage of similarity between graphs.

Support of effective visualization of topological shifts Visualizing series of large, dense graphs causes visual information overload affecting data interpretation. Due to the limited cognitive processing ability of humans, techniques should detect and convey changes in a visually comprehensible manner.

Interactive exploration capability for temporal data Networks function at multiple scales. To efficiently derive insights at these different scales, we methods must allow a user to explore summary and detail information on demand.

Motivated by major network analysis tasks covered in [8, 18], for example, we identified the following important capabilities that our method should support to handle specific tasks:

- Identify and characterize patterns that define temporal states, **T1**
- Analyze summary topologies that represent temporal states, **T2**
- Analyze variations in topologies within states, **T3**
- Analyze local dynamics governing global community and state changes, **T4**

5.4 Method

Our computational framework integrates algorithmic analysis with interactive visualization to extract and summarize reoccurring patterns and enable intuitive visual comparisons of network topologies within states. As shown in Fig. 5.1, in the first stage, we use a similarity computation method to detect time intervals of networks with similar local and global topologies (Fig. 5.1A); in the second stage, we use the state interval information to extract and represent the most common topology and community structure based on all other similar graphs within the state (Fig. 5.1B); in the third stage, to represent such data in a visually concise way, we compute the importance of each topological change and depict them through visual designs (Fig. 5.1B,C). The aggregation and interaction techniques enable users to holistically identify dependencies between local and global changes in topology. The designs of the consensus-based summary graphs and the community-based difference graphs depict the uncertainties in the data, e.g., variance, mean, and stability of a community for a node at a given time interval.

Our technique requires an understanding of change occurring over time. To numerically quantify such change, we define a similarity capturing local and global topological changes. A well-constructed similarity metric allows us to detect states (**T1**) having similar graph-level properties.

5.4.1 State-based Similarity

The first step in detecting topological changes between graphs and grouping them into states considers similarities and changes. To quantify changes between network graphs, appropriate metrics must be used to meet the criteria users would employ when determining groups.

General graph properties, such as degree distribution or community membership, can be used to characterize changes in networks, however, they are too coarse to be helpful. While several graph distances were proposed, we have adapted the metric proposed by Koutra *et al.* [99] to define similarity of network topologies. The reason we use this metric is due to its sound theoretical background using belief propagation in identifying changes in graphs. Further, the method agrees with intuition of taking into four major topological characteristics for comparing graphs, *i.e.*, 1) *edge importance*; 2) *weight awareness*; 3) *edge sub modularity*; and 4) *focus-*

awareness.

We note that any meaningful similarity definition could be adopted and used in our system, as guided by a specific problem. Our method would depict the similarity and difference topologies of the states resulting from such an alternate definition.

Our aim is to compute a similarity value $Sim_i(G_k, G_{k+1}) \in [0, 1]$, where a value of 1 represents two graphs G_i, G_j being identical and a value of 0 represents two graphs being completely dissimilar. To compute this value, we define a column vector $\vec{s}_i = [s_{i,1} \dots s_{i,n}]$, where the influence scores start from the i_{th} node and ends at the n_{th} node. This can then be stacked as an $n \times n$ vector matrix S , where

$$S = [s_{ij}] = [I + \epsilon^2 D - \epsilon A]^{-1}. \quad (5.1)$$

Here, $\epsilon = \frac{1}{1 + \max(d_{ii})}$ is a small constant used to capture the influence between neighboring nodes, and D is an $n \times n$ diagonal matrix, where $d_{ii} = \sum_j a_{i,j}$ is the degree of the node. A is the adjacency matrix, and I is the identity matrix. We compute graph distance as

$$d = RootED(S_1, S_2) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (\sqrt{s_{1,ij}} - \sqrt{s_{2,ij}})^2} \quad (5.2)$$

The final similarity value of two graphs is defined as

$$sim(G_1, G_2) = \frac{1}{1 + d} \quad (5.3)$$

5.4.2 States Defined by a Similarity Matrix

To extract states that have common local and global topological properties, we require a method that captures similar state behavior. Such a method must take into account the similarity information for all pairs of timesteps, also allowing the detection of re-occurring states. We define a distance matrix of size $D = N \times N$, where N is the number of timesteps and matrix entry $c_{i,j}$ is the similarity (Eq. 6.2) of the two graph structures at timesteps i and j . The higher the value is, the more similar the graphs i and j are. As, we are mainly interested in the most common topological properties with the state, we require a thresholding value that is not too low to swamp the matrix with moderately similar entries, and not too high as that will suppress significant

entries. Therefore, we employ a modified change detection mechanism proposed by Mutlu *et al.* [127],

$$I(G_i, G_j) = \begin{cases} Sim_{i,j}, & \text{if } (Sim_{i,j} - (u_{Global} + \delta_2 * \sigma_{Global}) \geq 0), \\ 0, & \text{otherwise} \end{cases}, \quad (5.4)$$

where u_{Global} and σ_{Global} are the mean and standard deviation of the values in the distance matrix and δ_2 is the coefficient for σ_{Global} . Based on the thresholded matrix, we use a connectivity-based clustering algorithm [30] to identify clusters that have similar graph structure. We use this as our definition of states. One of the advantages of this approach is the fact that the states obtained are based on values of the entire time dataset, rather than the values that are based on the current timestep t and previous timestep $t + 1$.

Algorithm 1: Summary Graph Representation

Data: Sequence of similar graphs $G_{t1}, G_{t2}, G_{t3}, \dots$

Result: Representative graph G_{Av} depicting the summary

- 1 Compute adjacency matrix containing pairwise similarity values between timesteps t_k to t_{k+n} using the Eq.6.2;
 - 2 Use standard change detection procedure (Eq. 5.4) to threshold the matrix;
 - 3 Rank all graphs G_{t_i} based on the number and average of non-zero values in i^{th} column, $avgSim_i$;
 - 4 Filter top n candidates for comparison, where $n = \delta_3 * N$;
 - 5 Pick most common community membership for each node in n graphs; in case of a tie, use membership from the graph with highest $avgSim_i$ value;
 - 6 Intersect edgelist of the n graphs to obtain edgelist of G_{Av} ;
 - 7 Weight of edgelist of G_{Av} is set to average over all n graphs;
-

5.5 Consensus-based State Summary Graphs

To understand the dynamics in network topology underlying each state and reduce the amount of information to be visualized for visual inspection, we derive a single representation for each state. The reduced graph sheds light on the overall stability (community changes, edge changes) of the network and the variability and mean of the important changes, for the time interval of

a detected state. Our temporally reduced *summary* graph representations should satisfy the following criteria:

- The graph representation must capture the most common topological properties of the system .
- The graph representation must be indicative of the overall community structure for the interval.
- The representation must highlight nodes that often change communities.

We consider different approaches to generate a state-based temporal graph satisfying the aforementioned criteria. One can accumulate the nodes over time to form edges. However, one of the assumptions here is that the meaning of edges is ignored. The representation of such graphs also depends on the lengths of the time intervals considered. The larger the time-windows are, the larger is the density of the graph.

An alternative approach is to identify changes from a pool of similar graphs and average the weights over the detected state-intervals. This method treats all edges equally and results in a dense network representation, which may not be representative of the *most common* topology.

5.5.1 Our Approach

Our approach for defining the graph summarization (Fig 5.2) produces informative summaries based on salient patterns. We introduce a dynamic graph summarization framework that constructs one representative visualization of the state. This is our problem definition: *Given N similar graph structures, how do we best depict the most representative graph?* These are the main aspects to be considered for the representative graph:

- Identification of the community membership of a node in the graph
- Determination of the presence of edges between all pairs of nodes
- Determination of the weight of an edge

Algorithm 1 and Figure 5.2 describe the procedure that constructs a summary graph from a set of similar networks for a given temporal state. To determine the quality and uncertainties

associated with a graph generated by the summary algorithm, we use a metric, *Average Number of Edges Pruned per Timestep* (AEP), that estimates the average amount of dynamic edges (edges that frequently get added or deleted within the state interval) pruned by the algorithm). This value provides us with insight into the overall variability of the topologies within the state. For example, states with edges exhibiting significant variability (addition or deletion) have relatively high AEP values. Such a metric allows us to determine whether the sparseness of the graph is representative of the underlying data or is the result of the dynamics occurring within the state. The metric is defined as: $(\text{total number of edges pruned by algorithm per state})/(\text{total number of timesteps in state})$.

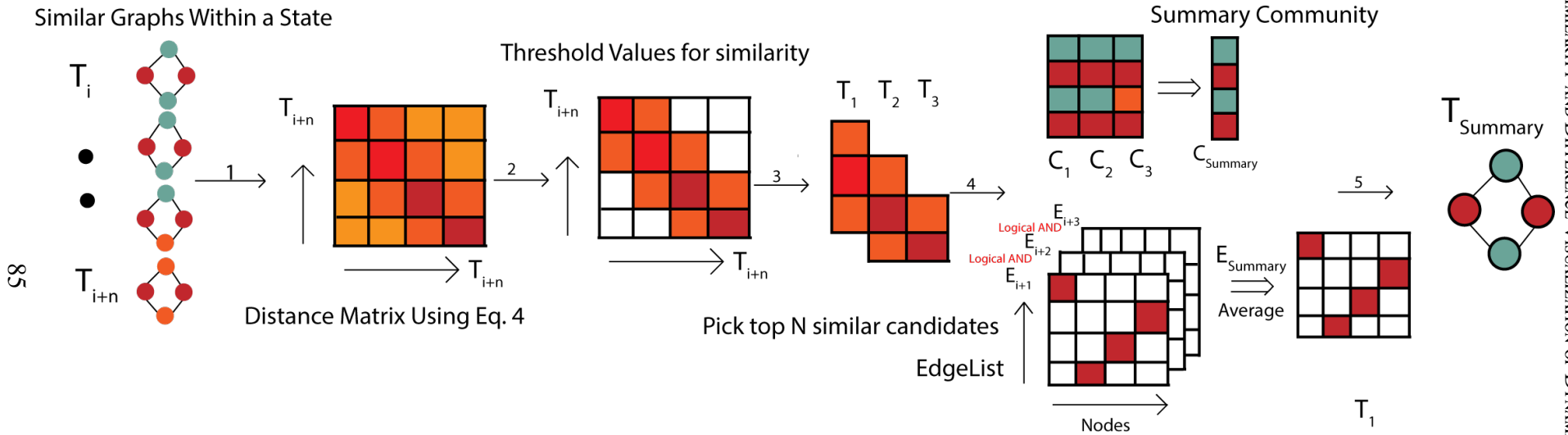


Figure 5.2. The algorithm used to construct the summary graph identifies the most common local and global topological attributes from a set of similar graphs. Networks for a given time interval define the input. Similarities between networks are computed with Eq. 5.2 and define the entries of a similarity matrix. Using a threshold δ_2 , we pick the top N candidates for intersection of edge lists and computation of summary community memberships, characterizing the summary graph.

5.6 Visual Representation of the Summary Graph

The reduced graphs represent a simple representation of a complex dynamic phenomena occurring within the represented time-interval.

To more easily understand such data, we need visual representations that not only depict the summarized topological information, but also depict variances in the properties where there might be one. Such depiction of dynamics is crucial to understanding the cause and effect of the formation of the temporal states. For example, in neuroscience, transient nodes (brain regions) with more flexibility (frequent community memberships changes) may be involved in performing a wide range of cognitive functions and may be particularly involved for changing global brain states [90].

To facilitate such a detailed exploration process in an intuitive manner, we want to represent the following quantities: 1) nodes with frequent community changes, 2) stability of communities, 3) variability in edge weights, 4) mean value of edge weights.

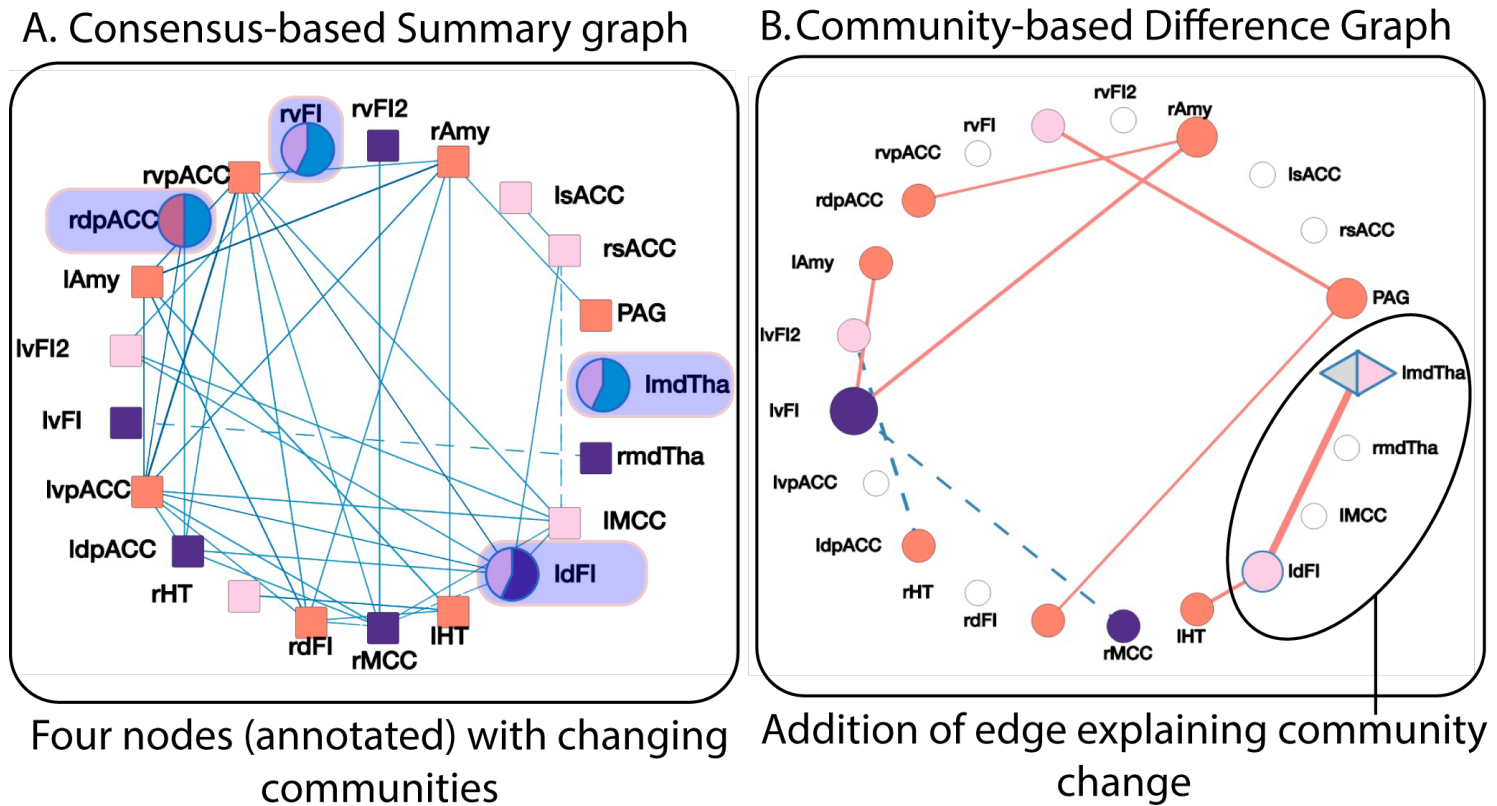


Figure 5.3. Our methods instantiated on a real brain network dataset. The summary graph in conjunction with the difference graph depict similarities and differences in topologies, respectively. A. Summary graph involving four nodes (glyphs) depict the frequent change in communities during a particular time-interval. B. A difference graph explaining the community change of *ldFI*, *left dorsal frontoinsula* and *lmdTha*, *mediodorsal thalamus*, from grey to pink.

We investigated three major visual designs for this purpose, all following the data aggregation principles specified by Elmqvist *et al.* [56] (Figure 5.4). In the first design, to enable comparisons between edges in the same graph, we encode the variance in edge weights using color and the thickness as the mean edge weight. Outlier edges having high variances are denoted using dashed lines. The transient nodes that vary their community membership over time are represented as vertically stacked bars with an enclosed circle. The length of the bars represents the longevity of the community in that node for the given time interval. The second and third design use the same scheme for the edges, however, each node is represented as a pie chart, where each slice represents the longevity of a community in a particular node. In the third design, based on existing research on selective visual attention[105] and to visually distinguish the dynamics of the transient nodes, we encode other stable nodes as static rounded rectangles with color representing their community membership.

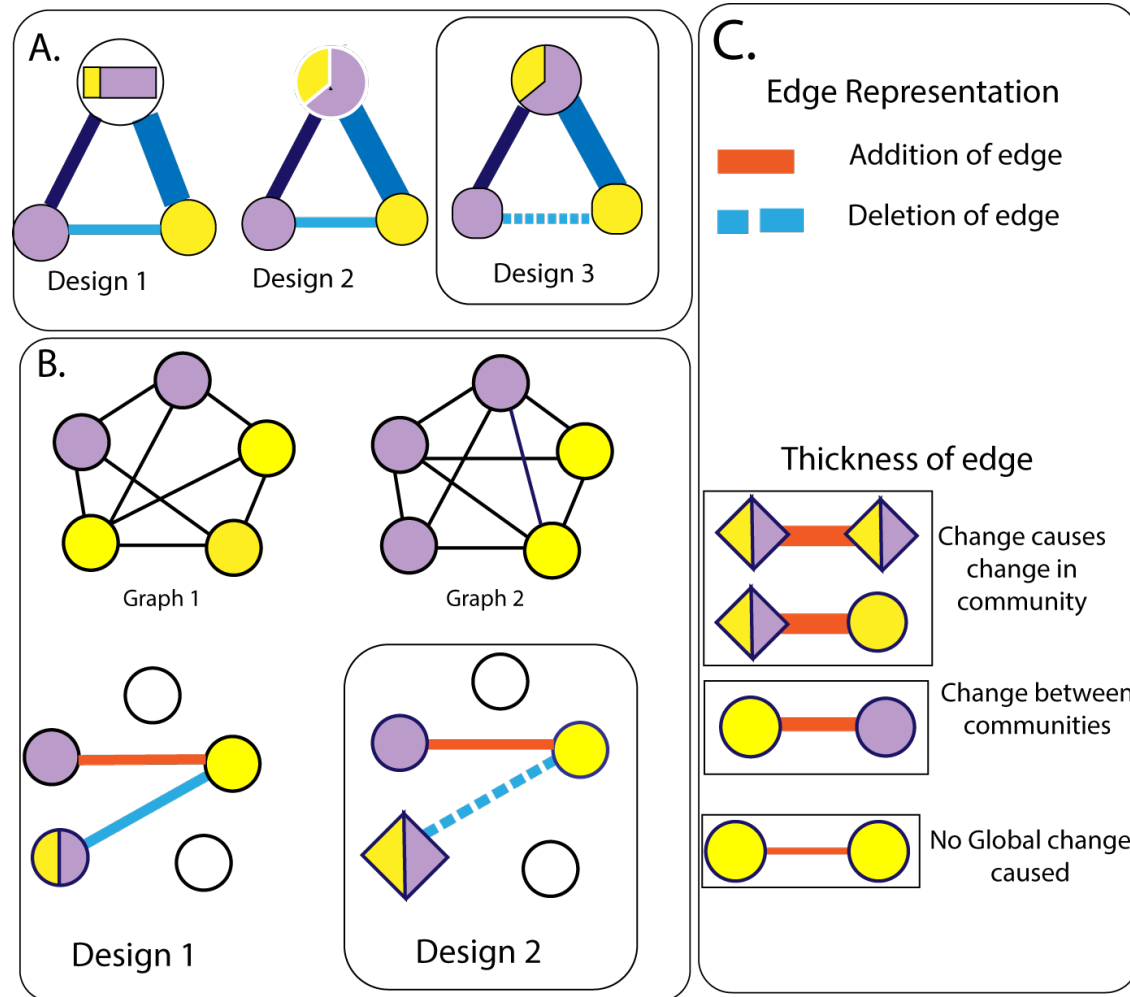


Figure 5.4. We investigate different visual designs to convey uncertainty and variability in the dynamic network data. A. Summary graph representing stable nodes as rectangles and nodes that change their communities as glyphs. B. Difference graphs representing deletion and addition of edges as dashed blue and solid red lines, respectively. C. Visual encodings that depict the importance of each edge change in the difference graph.

Overall, based on different trial runs, we found the third design to express the changes in a visually clear and concise manner. Particularly, the unique visual outlines convey the changes in transient nodes more effectively. We use this as our design for the summary graph (Fig 5.3).

5.7 Difference Graph Framework

Analyzing every graph at each timestep and identifying change is cognitively overwhelming. The problem becomes more pronounced for large graphs with minuscule change occurring between timesteps. The analyst has to manually correlate changes from many views, increasing context-switching costs associated with every view.

Given two graphs, to enable assessment of the functional difference between them, it is crucial that we not only show the change but also show the importance that change has on its overall topology. For example, the disappearance of an edge causing a community-change is far more important than the disappearance of an edge within a community and causing no change in the global structure.

To address this issue, we propose a community-based difference-graph approach to depict the dynamics occurring over a set of similar graphs for a given state.

5.7.1 Community-based Difference Graph Visual Design

Since we do not deal with addition or deletion of nodes, we define the nodes of the difference graph to be the same as the static graph. For edges, we define three major types (in the order of decreasing importance) (Fig. 5.4C). 1) **Modular edge change**—in most cases, they may cause an important change in the overall community structure; 2) **Inter-modular edge change**—such edges connect different communities, and may have a tendency to cause a community change, 3) **Intra-modular edge change**—such edges have little effect on the global topology of the graph.

We use thickness of the edge line to convey the importance of these edges. The higher the thickness, the greater is its importance. Based on previous studies in difference graphs [19, 138, 203], we use the colors red and blue to indicate the addition and the deletion of edges respectively. There are cases when a node in a graph is becoming transient, *i.e.*, changing the community membership. Such events are the pivotal time points of nodes and are of utmost

importance in its evolution. We explored two designs to convey such dynamics occurring in difference graph.

In the first design (Fig. 5.4B), to depict a community change at given node, we subdivide the circle into two equal parts, the left semicircle denoting the community membership of the previous timestep and the right one denoting the membership at current timestep. In our second design to further visually distinguish the most important change, we encode dashed lines to depict deletion of edges and encode a rotated diamond with two sides to depict the change in community membership for a given node. To enhance visual clarity, we only show the nodes that are involved in change. The size of the nodes in our graph represent the amount of local change (change in *Egocentric Topological Overlap*, section 4.3).

Different trials of various visual designs based on real brain network dataset made it clear that the second design showed variations in data more convincingly. However, in cases when there are a lot of change occurring, the user would have to analyze the original raw graph to assess the change in the difference graph. A switching cost is involved in such operations. Nonetheless, we found that this technique is better suited for our use-case.

5.8 McFarland’s Classroom Data Set

We applied our visual analysis technique to a publicly available classroom data set [114], the “McFarland’s classroom data” [115]. This data set contains information concerning conversations between teachers and students in a high school economics class (11th and 12th grades). We used the McFarland data set as a use-case for our system to analyze how students and teachers interacted in this class. We pre-processed the raw data to create a dynamic network dataset, appropriate as input to our system, consisting of 20 nodes (17 students, three teachers) spanning 49 minutes (converted to 82 timesteps). We were able to identify evolving communities, states, and summary topologies for the states.

The state detection algorithm found two dominant conversation states in the dataset, i.e., a state where teachers broadcast information to students and a sociable state where students interact with each other, forming multiple communities (modular) (**T1,T2**). The detected intervals in the states were not continuous. State one has associated intervals (1-2, 10-15, 37-42, 49), and

state two has associated intervals (3-9, 16-36, 42-48, 50-82), see Fig. 5.5A. The states identified with our system are in agreement with those reported in a study concerning this dataset [115].

Specifically, the class starts with the teacher lecturing (state one, time steps 1–2), and then having students perform a group task (state two, time steps 3–9). The students in this state form a modular structure, talking to each other, and forming tightly-knit communities. This state’s structure is dispersed as a consequence of teacher’s intervention, defining the structure for time steps 10–15. The observed fluctuating state patterns with consistent sparse and dense core connectivity is captured by the summary graph, see Fig. 5.5B.

To understand the dynamics of state one, we selected sub-interval (10-15, within state one) and visualized its corresponding summary and difference graphs, see Fig. 5.5 A,B. The broadcasting edges emanating from teachers T1 and T2 represent communication to all students over this entire interval. During this interval, a community (light yellow) involving student group (S13, S15, S18) is detected possibly reflecting conversations among these students during lecture (**T3**). The circles depicting students S5 and S9 indicate that they may have formed their own communities before listening to the lecture of teachers T1 and T2. The summary graph visualizes the consistent dense temporal behavior in this interval.

The difference graph depicting topological change (15–16), see Fig. 5.5C, contains many dashed lines, *i.e.*, edges disappear, indicating that the teachers have stopped lecturing and assigned group tasks to students [115]. The large number of rotated diamonds (seven) in the graph could be indicative of student groups (S4, S7) and (S8, S11, T3 and S16) switching to communities amongst each other to collaborate on a task, which could have caused a system state change (**T3,T4**).

In conclusion, through the summary graphs, we observed consistent core network structure, where either the teachers teach or students perform group tasks (Fig. 5.5B), and through the difference graphs, the topological effect of the teachers pausing lecture is seen, where students form their own communities amongst each other.

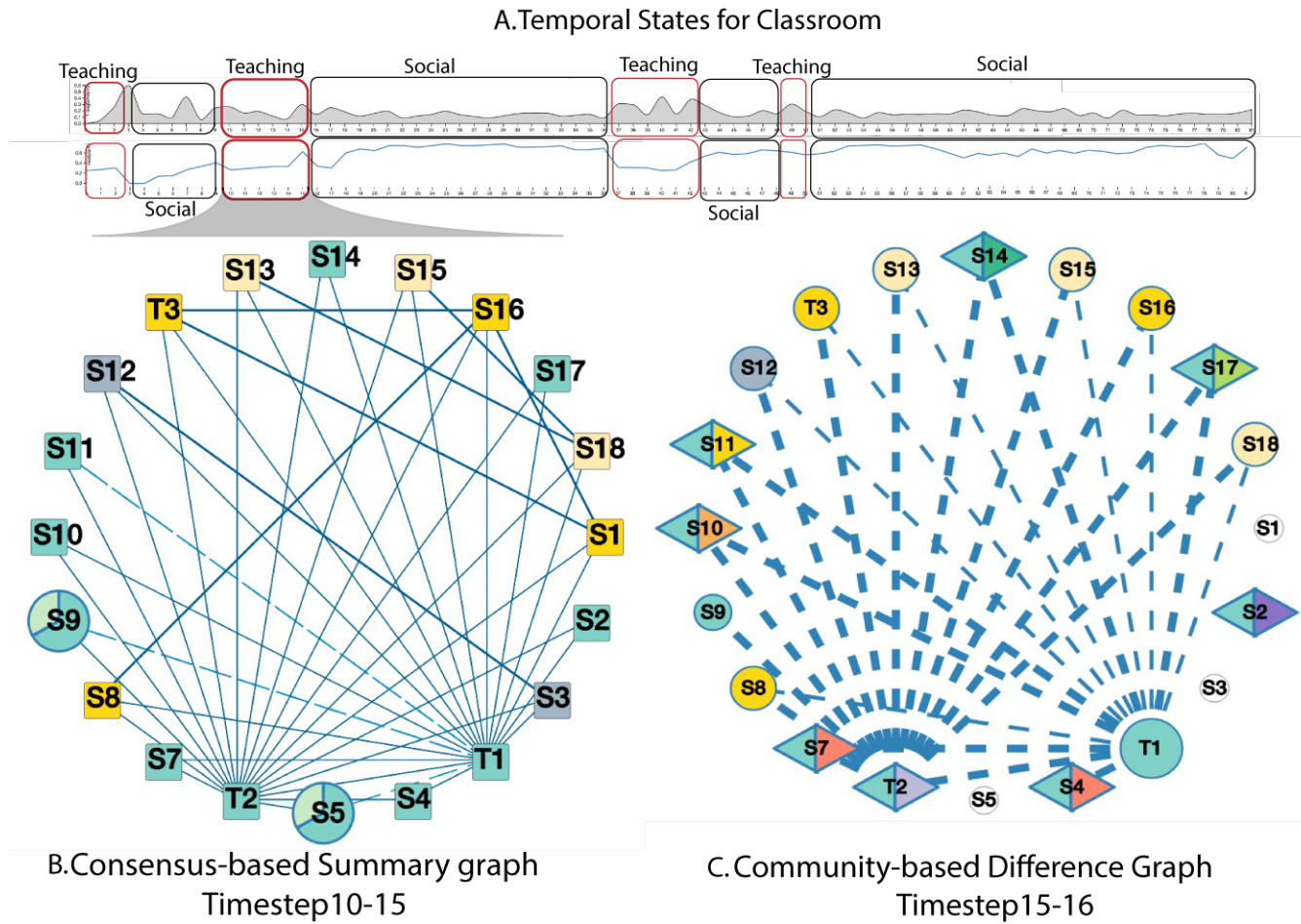


Figure 5.5. The two dominant states in the McFarland dataset. A. Red rectangles represent lecture sessions by the teachers, and black rectangles represent states when the teachers are not lecturing. B. Broadcasting edges from teachers T1 and T2 depict lecturing student groups. C. State change as a consequence of students forming groups and teachers not lecturing.

States	<i>Time</i>	<i>C</i>	<i>Q</i>	<i>ED</i>	<i>AEP</i>
State 1	0-40	0.79	0.28	0.31	3.56
State 2	41-56	0.77	0.22	0.28	6
State 3	57-87	0.83	0.40	0.24	3.8
State 5	100-119	0.78	0.16	0.38	9.3
State 6	120-133	0.74	0.13	0.55	5.42
State 7	134-146	0.72	0.15	0.47	7.14
State 8	147-182	0.84	0.14	0.35	3.04
State 9	183-202	0.74	0.52	0.15	7.33
Entire dataset	0-202	0.78	0.32	0.25	NA

Table 5.1. The table statistically compares the properties of the detected states. We generally find two major types of states, with low modularity and high density or with high modularity and low density. We further find that the difference in modularity is drastic from state 2 to state 3. Abbreviations: ‘*C*’: Average Temporal Correlations; ‘*Q*’: Modularity; ‘*ED*’: Edge Density ; ‘*AEP*’: Average Number of Edges Pruned, see section 5.1

5.9 TimeSum Used in Neuroscience

To demonstrate the value of our approach, we applied it to two brain network datasets. We performed the presented case studies in collaboration with neuroscientists, who are experts in brain network neuroscience. The studies concern exploratory data analysis, where the goal was to visualize global network dynamics and their relationships to brain regions, with the purpose of generating scientific hypotheses that would later be studied with rigorous statistical methods.

Neuroscientists use fMRI to measure whole-brain activity, which can be modeled as a set of brain regions (nodes) connected by edges reflecting the correlations between their activities. Changes in topology over time can be conceptualized as switches between multiple quasi-stable *functional states* [164]. Identifying the temporal intervals with similar topological structure is important to understand the interaction of behavioral systems when performing a task or their impairment due to a disorder [36].

5.9.1 Case Study 1

The objective of this case study was to explore and understand the dynamics concerning the salience network [162] of a healthy older adult. This network represents the correlations between regions in the brain responsible for the monitoring of sensory, visceral, and the reward or threat system. It consists of *input nodes* in anterior insula (e.g., dFI, vFI) that gather sensory information and *output nodes* in anterior cingulate (e.g., dpACC, vpACC) that initiate behavioral actions. A major challenge in analysis is understanding the dynamic fluctuations of connectivity profiles between these two groups or nodes over time.

The processed time-varying data [106] (pairwise time series correlations between 21 regions, 202 timesteps) was used as input to our system to identify dynamic communities, state-intervals, and abstract graph topologies. In identifying the states, we used the following weights: global similarity ($1 - \delta$) set to 0.7 and 0.3 used for local similarity. The system detected nine states with similar graph-level properties. The modularity and (one-similarity) plots, see Fig. 5.9.1B, convey the temporal structure and the extent of the individual states in the dataset (T1).

We can clearly see consistent patterns for each state. Two major structures are revealed, a sparse (highly modular) structure (states 3 and 9) and a densely-integrated structure (states 1, 2, 4, 5, 6, 7, and 8) (T2), which may correspond to existing studies in neuroscience[164]. The state detection method was able to abstract variabilities within those two major categories, see Table 2. One also observes this phenomena as a large number of squares (stable communities) in the summary graph for time 57-87, see Fig. 5.9.1B3, state 3. Most of the other summary graphs have glyphs, representing regions that dynamically change communities.

The more globally integrated and less modular states had earlier been suggested to be more energetically demanding, requiring increased cerebral blood flow, explaining the dense structure [202]. States having high modularity and high stability, like states 3 and 9, require less overall “cost” (fewer/shorter edges) than more globally integrated states. This could explain its maintenance as a stable structure. When analyzing the dynamics of state 3, see Fig. 5.9.1B3, state 3, one can quickly identify two brain regions *right ventral frontoinsula*, i.e., **rvFI2** and *right dorsal frontoinsula*, i.e., **rdFI** with flexible community membership. This could suggest that

these regions may have switched their roles (**T3**).

To explore intrinsic structural change of the network between states two and three, we drill down and visualize the difference graphs from times 55 to 58, see Fig. 5.9.1A3. A drastic topological change is visible. The large number of rotated diamonds at time 56-57 indicates the extent of community changes taking place. In general, many regions lose their correlations (dashed lines), which could explain the network's shift towards a higher degree of modularity (high Q in state 3) (**T3**).

To understand integral node dynamics of brain regions during state transitions, we hover over and interact with the brain region *ventral pregenual anterior cingulate cortex*, *i.e.* **IvpACC**, see Fig. 5.9.1A2. The change of local neighborhood of the node is clearly visible, *i.e.*, changes in the IvpACC's connectivity from the vFI and dFI, its main input nodes, to the dpACC, one of its main output nodes. The node IvpACC eventually changes its membership (**T4**). One explanation of this highly flexible node configuration is that the region oscillates between *receive* (input) and *transmit* (output) modes. A detailed fMRI analysis would be necessary to test this hypothesis.

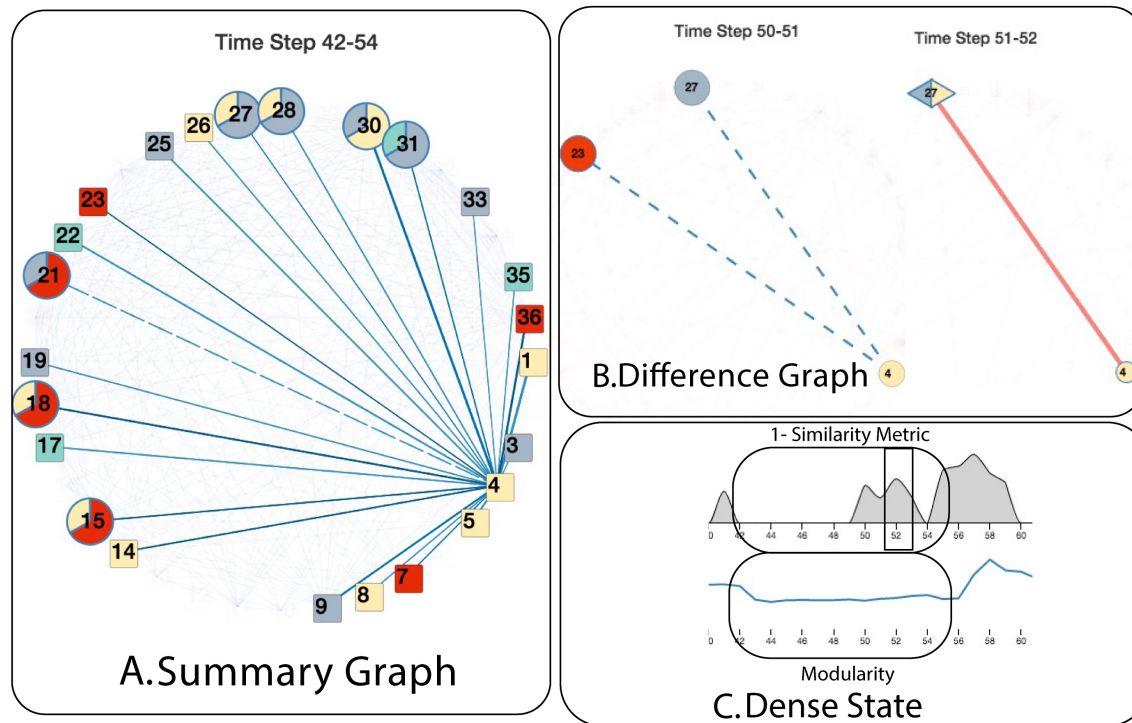


Figure 5.6. A. Exploration of *posteromedial cortex*, node 4, in an integrated state from timesteps 42-54 through the summary graph. B. Detailed exploration of changes in local topology of node 4 through difference graph for timesteps 50-52. C. Plots for 1-similarity and modularity of the identified state. Filtering within-state variability for timesteps 50-52.

Figure 5.7. We create abstractions of the brain network dataset, considering nine extracted states with similar global and local properties. A. Local dynamics of nodes *IsACC* and *IvpACC* (1,2) for time 55-58. During time 55-56 *IvpACC* switches from receive to transmit mode, see section 9.1.

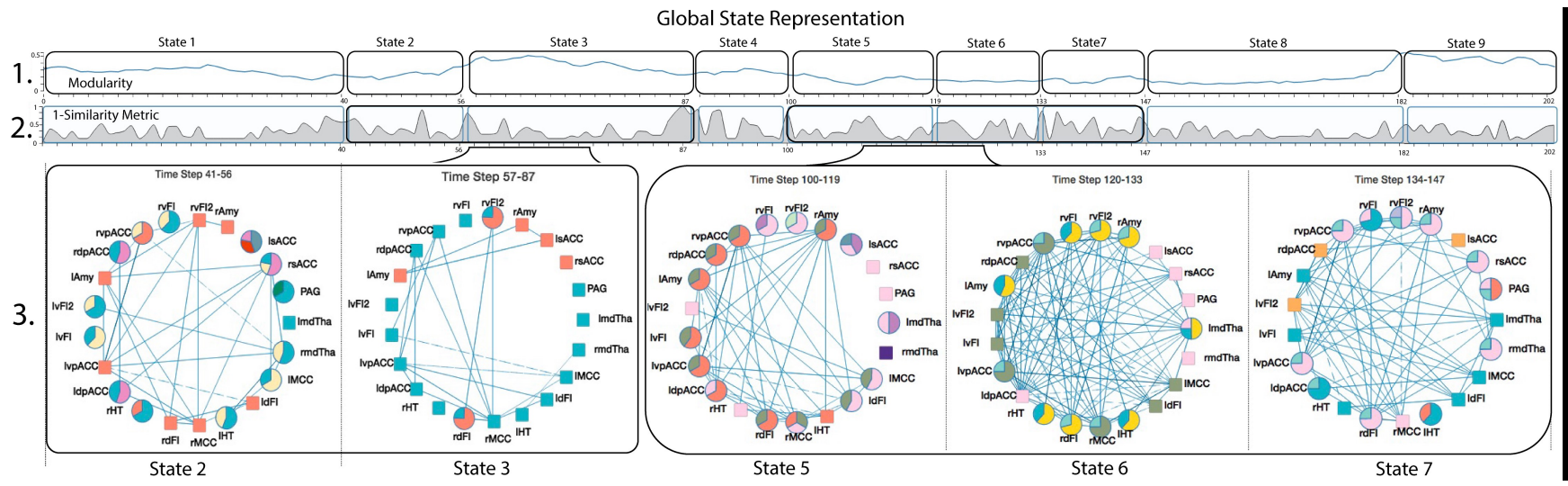


Figure 5.8. Global topological dynamics during time 55-58 defining a pivotal change of brain state from integrated to modular. C. Summary graphs depicting evolution of average topology, *i.e.*, from sparse to dense to sparse.

5.9.2 Case Study 2

Exploratory visual analysis in dynamic networks is key to observe general trends, explore temporal variability, and understand higher-level organization of the network. This case study involved a domain expert who was able to construct and articulate hypotheses on-the-fly, during interactive data analysis process. The goal was to understand the dynamics of the resting state brain network connectivity and its impact on individual nodes and their connections. Specific questions were: What are the major topological changes (spontaneous global dynamics) (**T1**) in the dataset? How do they relate to local changes in the network (**T4**)?

We used a whole-brain, high-temporal resolution (TR=720ms) resting state fMRI data set [186] and extracted time-series from a 36-node functional parcellation [26]. By using *Multiplication of Temporal Derivatives* [164] we were able to calculate 1200 time-varying network matrices. To remove noise, we smoothed the dataset with a 14 times (volume) sliding window.

Based on initial plots (*1-similarity and modularity*) of the full 1200 volume time-series data set, we filtered a 140 times (400-560) window containing consistent similarity patterns, visible as peaks in the similarity plot. Using this dataset as input, TimeSum identified time-varying communities, state-intervals, and summarized topology. Two major classes of network states were identified: one with relatively high density and low modularity (times 42-54, 96-120); one with low density and high modularity (times 0-40, 60-96, 121-140), see Fig. 5.6B (**T1**). This behavior may correspond to the results from a previously performed statistical analysis of the data, which had identified tight, integrated and modular, segregated states [164].

To better understand local and global network changes driving dynamic behavior, we drilled down into a time sub-interval (42-54), investigating a state transition, see Fig. 5.6B. Our domain expert focused on visualizing the dynamics of node 4, encompassing brain region *posteromedial cortex*, one of the most connected areas of the human brain. As shown in the summary graph in Fig. 5.6A, *posteromedial cortex* exhibited diverse connectivity with multiple flexible regions (frequent community changes) within the state, (*i.e.*, nodes 27, 28, 30, 31) (**T2,T4**). Interestingly, node 4 (*posteromedial cortex*) itself does not change its community during this time, and this relative stability matches previous research done regarding the dynamics of this brain region [164].

Within the selected sub-interval, the similarity plot suggested large changes of network structure during time 50-52. Visualization of these changes using a difference graph, see Fig. 5.6B,(T3,T4), only for node 4, may provide an explanation of the relatively large shift in global network structure (visible in the similarity plot), driven by changes in few edges and shifts in community structure. Additional research and rigorous statistical analysis is needed to validate the network dynamics inferred by *TimeSum*.

To summarize, our domain experts could easily identify major topological shifts in the data sets used for the presented case studies. The case studies demonstrate that our approach has great potential for rapidly formulating initial hypotheses for different types of applications requiring complex network analysis. A rigorous domain-specific statistical analysis is still needed to validate hypotheses.

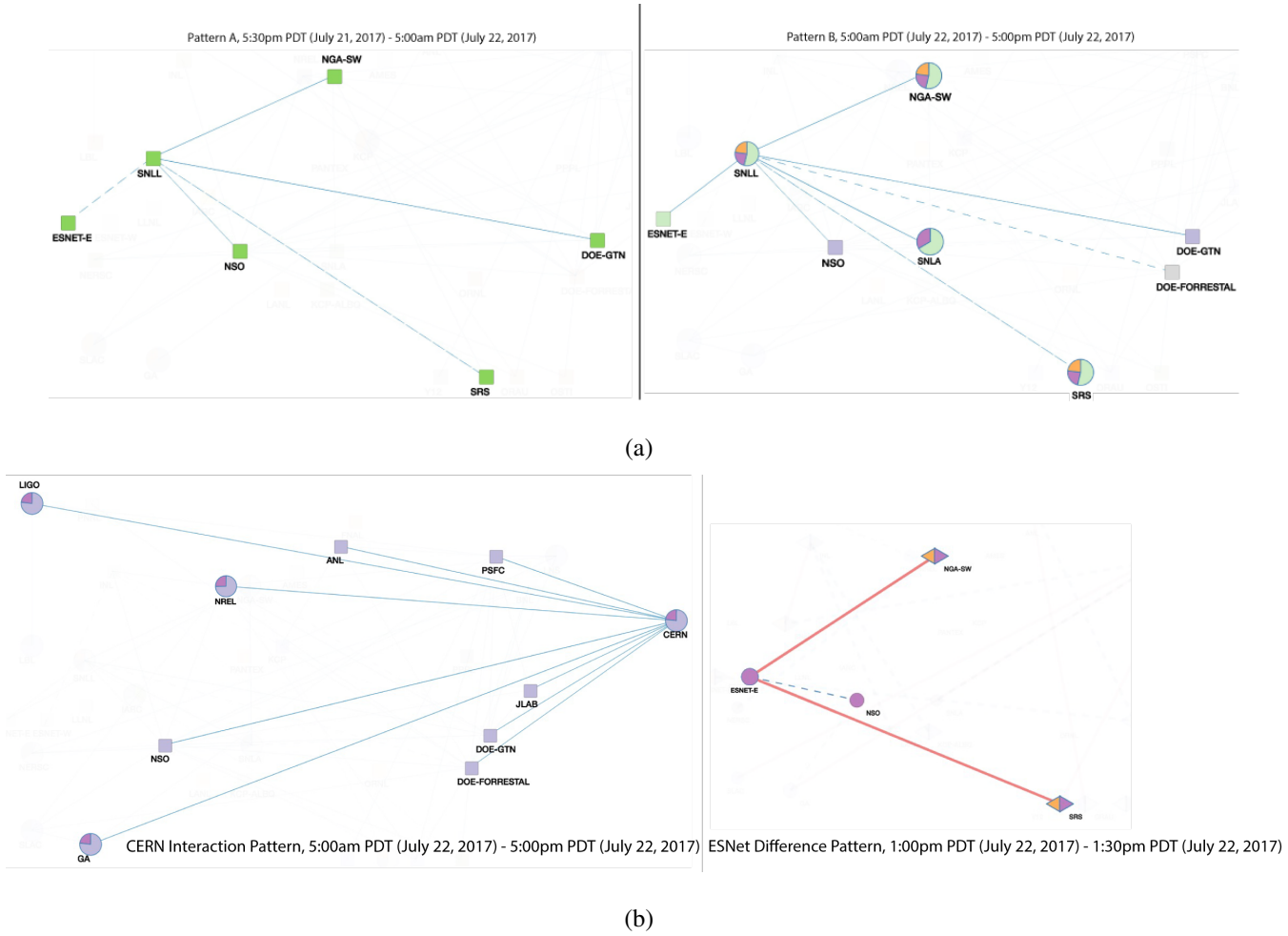


Figure 5.9. Similarity and difference topology of network flow in ESnet. A. Similarity visualization of patterns for period 21 July, 5:30pm PDT to 22 July, 5:00am PDT, and for period 22 July 5:00am PDT to 22 July, 5:00 pm PDT. Nodes SNLL and SNLA switch communities frequently during this period and are stable during the first time interval. B. Interaction patterns between CERN and other sites, and difference topology of ESNET.

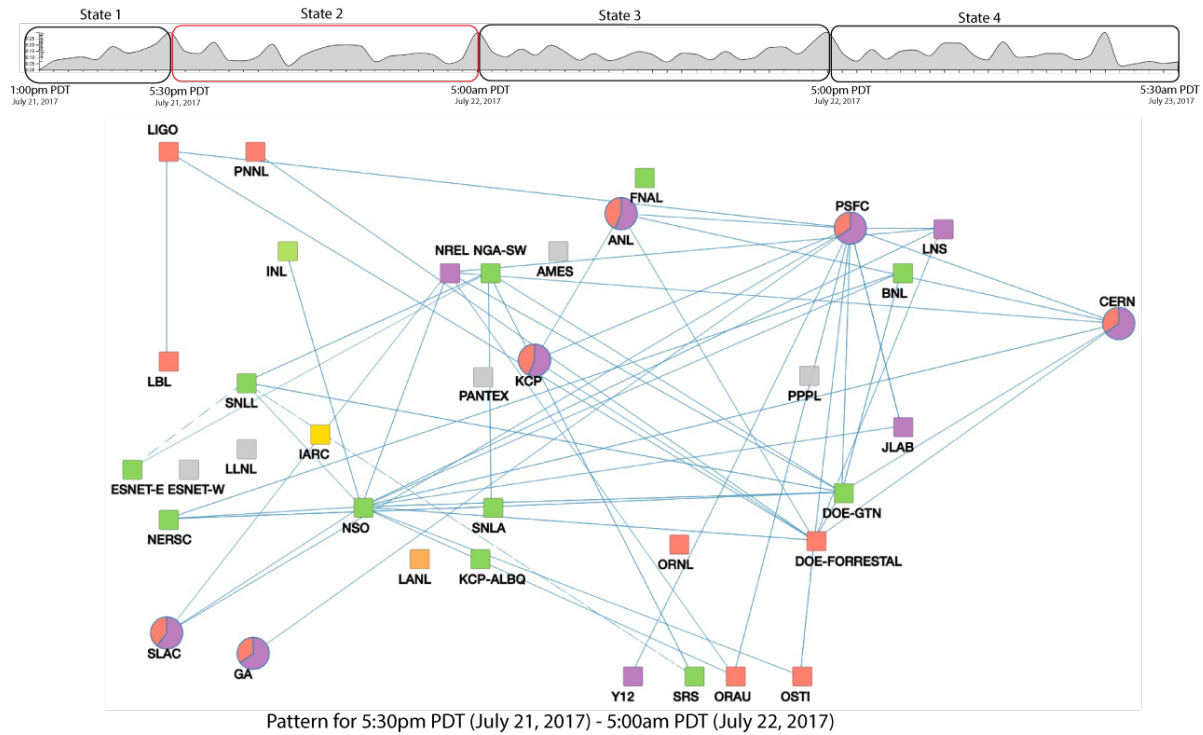


Figure 5.10. Overview of different states detected via corresponding similarity topology. ESNET sites are colored with respect to dynamically changing community affiliation.

5.9.3 ESnet: Analyzing Router Topology

The router topology dataset, shown in Fig. 5.9, examines network connectivity between sites in the United States and Europe. The Energy Sciences Network (ESnet) is a high-speed computer network backbone enabling remote scientific collaboration and experiments. ESnet monitors network connections, collecting statistics on bytes sent/received and link performance logs. One monitoring tool, ESxSNMP, collects router-in and router-out bytes for every interface every 30 seconds. The tool records the packets transferred between sites at different times of a day. While the physical topology is fixed, the virtual topology of data movement changes dynamically, depending on the site's access to data.

In this case study, we examine network connection data for two days collected at a 30-minute intervals, showing packet exchanges between sites. The major evolving ESnet flow patterns are of great interest [49, 95]. Specifically, one seeks answers to these questions: What sites are “friendly”? (**T1**, **T2**) and how do the topologies of these sites vary over time? (**T3**) Do such network topologies reveal common patterns of communication between network sites? (**T4**)

We created a dataset where records contain details, such as: *source*, *destination*, *packets transferred*. The dataset is associated with the data set covers the time period *21 July 2017, 1:00pm PDT to 23 July 2017, 5:00am PDT*. We pre-processed the data to create a derived dataset describing the dynamic network behavior, consisting of 80 timesteps and 33 nodes. We use this derived dataset as input to our method.

With our tool one can identify dynamic communities forming and recognizing temporal states in the recorded period. Using the temporal metric (Eq 6.2), one can find two dominant network states corresponding to communication behavior during day and night, relative to PDT.

The similarity graphs and difference graphs shown in Fig. 5.9 (site abbreviations: <https://my.es.net/sites/list>), depict the consistent topological and community patterns for periods of PDT time: state 1, *1:00pm PDT - 5:30pm PDT, July 21, 2017*; state 2, *5:30am PDT, July 21, 2017 - 5:00am PDT, July 22, 2017*; state 3, *5:00am PDT - 5:00pm PDT, July 22, 2017*; state 4, *5:00pm PDT, July 22, 2017 - 5:30am PDT July 23, 2017*

The similarity graphs depicting each individual state (see Fig. 5.9A.1 and 2) likely represent consistent day-time and night-time operations in PDT time, Radial glyphs in Fig. 5.9A.2

represent the frequently changing nature of sites like SNL, SNLA, and SRS during day-time operations. Further, they depict the exchange of data packets between intermediary sites as experiments develop. Pattern A is more stable between the selected sites (SNLL, SNLA, and SRS). The main “friendly” sites in the identified temporal state (Fig. 5.9B) include CERN, PSFC, JLAB, ANL, and NSO. The similarity graph, shown in Fig. 5.9C, conveys overall topological behavior during night-time, relative to PDT.

In Fig. 5.9B.2, NGA-SW transitions from an orange to a dark-purple community, evident in the difference graph during 1:00 pm - 1:30 pm PDT. The addition of packets from ESNET-E to NGA-SW could have caused the sudden change in community. The same behavior is also observed for SRS. Further, in Fig. 5.9B.1, surprisingly, sites like CERN connect with spatially distant sites like LIGO, SLAC, and GA, see Fig 5.9B, indicating the spreading routing pattern of the network.

Visualizing the time-varying network topology with TimeSum allowed us to determine stable and unstable time-varying connectivity behavior. It is possible to understand temporal data by automatically identifying similarities and differences supporting intuitive pattern identification. The summary graphs show consistent and “friendly” communication patterns, while the difference graph shows the topological cause of a major state change. Detected patterns can be explored further to construct alternative routing paths to better transfer information across sites.

5.10 Discussion

Traditionally, analysts use their expert knowledge, experience and perceptual skill to identify core topological structures and its changes in a complex network. TimeSum, based on its algorithmic and visual methods, is able to characterize important changes happening in a network and compute the core structure characterizing a state. Our approach permits interactive parameter setting and compact visualizations for highly efficient analysis. It significantly speeds up the network data analysis workflow for domain experts.

Domain Expert Feedback Our domain expert (neuroscientist) has used our tool and is convinced of its value. He was quickly able to identify higher-level topological network changes, and identify trend behavior in the connectivity profiles of the nodes. He stated: “*The interactive*

visual techniques introduced, enabled an intuitive understanding of the systems we study. For example, the summary graphs provided high-level overviews of network structure and how it changes. The difference graph allowed me to get a sense of variability within each state and comprehend the dependence between local and global dynamics. It would be great to add more layout techniques to understand modular structure better.” He added that “*there havent really been any good ways to visualize time-varying network data in a way that facilitates detailed and intuitive understanding. TimeSum allows me to perform the kind of deep data exploration that is likely to help substantially for gaining new insights.”*

Challenges Our method was designed primarily to address the temporal scalability problem arising in time-varying network data analysis. In our experiments, we determined that our system can be used as an interactive, near-real-time system for data sets consisting of 36 nodes, 158,400 edges, and 204 timesteps. The processing time required for interactive filtering and switching views is negligible. As all abstraction algorithms are executed online, the system requires approximately five to six seconds to start up. For larger data sets, our technique requires more screen space and computational power. Our method produces intuitively understandable results for data exhibiting temporal correlation, but not all practically relevant networks have such correlated behavior. Further, the threshold used for the computation of summary graphs affects the computation of topology. The resulting graph will be sparse if the algorithm finds similar topology between dissimilar graphs within a state. For example, *AEP*, see section 5.5.1, is an indicator one can use to determine the reliability of the summarization algorithm. Our technique requires a user to specify the values of several parameters, *e.g.*, δ values for weights needed for the similarity measure, threshold values for summary graph generation and state detection, etc. Unfortunately, the specific values that must be used to produce scientifically meaningful and helpful results depend on the particular data set being analyzed and questions to be answered.

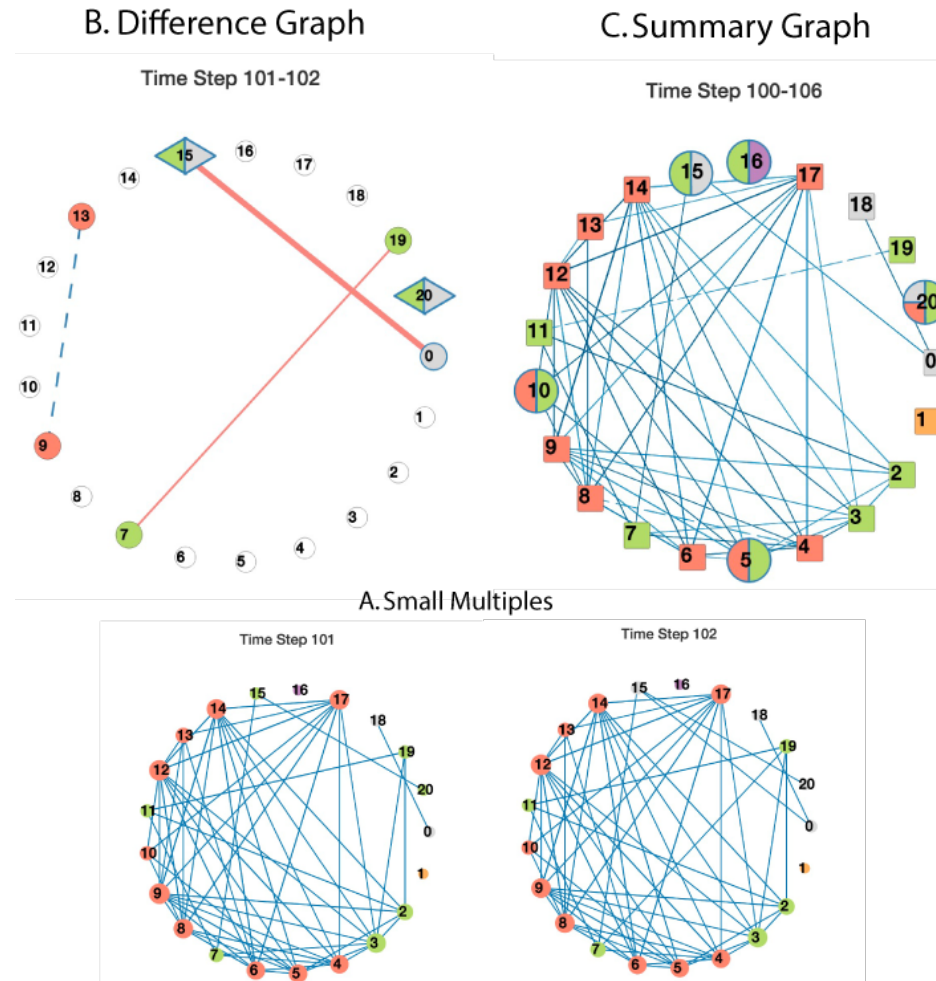


Figure 5.11. Our technique reduces the time-varying data set into summary graphs. Difference graphs convey the causes and effects of local topological changes. Here, the small multiples technique (A) is compared with the summary graph (C) and difference graph (B). C. Considering the summary graph, we can identify the nodes that often change community membership. B. The addition of the edge from node 0 to node 15 causes nodes 15 and 20 to change their community (grey).

5.11 Comparison with Other Techniques

Domain experts commonly use small-multiples or animations to understand the dynamics and evolution of graphs over time, see Figure 5.11. We compare our approach to common methods and explain the conceptual advances our technique offers. Fig. 5.11A shows two graphs (101, 102) in the same state (timestep interval 100-105), with nodes being colored by the dynamically changing community membership. Figs. 5.11B and 5.11C show the difference and summary network computed by our technique.

Considering **small-multiples**, to identify (**T1**) states and analyze summary topologies (**T2**), the method relies on a user’s visual perceptual skills to identify commonalities in the behavior of similar graphs. Generally, a user must manually inspect a large number of views. To understand variations in topology (**T3**) and comprehend complex local dynamics (**T4**) in dense graphs, finding differences is a cognitively overwhelming and extremely time-consuming task.

Animations The use of visual animations, which is limited to a user’s short-term memory, is often ineffective for the analysis of complex time-varying networks. A user must remember changes in community membership, deal with instable layouts necessary to classify states (**T1**), and mentally determine commonalities (**T2**). Further, identifying small-scale changes in the topology of dense graphs is a complex and cognitively demanding task (**T3,T4**).

To support efficient state analysis of complex time-varying networks, our method automatically computes the core network topology and its corresponding time-interval, see Fig. 5.11C. This approach reduces redundancy in visual representations used to depict the same information. It also increases visual comprehension of common connectivity patterns (**T1,T2**) through abstraction and reduction of visual clutter. With our approach, small-scale changes in topology causing community changes can be identified quickly via community-based difference graphs, see Fig. 5.11B, depicting a community change of node 15 (**T3,T4**). The combination of these views supports a user substantially to understand similarities and differences in topology and communities. Without our system, such a detailed analysis of a time-varying network would require a user to spend a significant amount of time for manual inspection of a much larger number of views and develop an instinct to find the common topological structure. We provide a general comparison of our method with other common approaches used for dynamic network

analysis in Table 1.

5.12 Conclusions and Future Work

We have presented *TimeSum*, an approach for exploring dynamic networks and presenting their complex topologies via effective abstractions. We have introduced a unique set of algorithms to identify time-intervals related to similar topological network properties, allowing one to comprehend global trends in a network dataset. TimeSum is a powerful tool for dealing with the temporal visual scalability problem, and it greatly reduces the need for time-consuming manual data analysis steps. With the help of our tool, experts could comprehend rapidly the causes and effects of topological changes occurring in networks, which, using commonly available methods, would require cumbersome off-line data processing.

We plan to focus on the crucially important aspect of scalability for effective network data processing and visualization. As datasets become larger, in terms of nodes and timesteps, methods are needed that can display large data meaningfully. We plan to consider especially hierarchical approaches, *e.g.*, edge bundling techniques. Edge bundling should support effective and loss-less visualization of representative characteristics of a network at a specific time. It is possible to improve our exploration approach by enabling difference graph computation with not only adjacent timepoints but also time-discontinuous time points.

Chapter 6

Context-aware Dynamic Network Analysis for Network Flow Data

Understanding and analyzing network flow traffic information are fundamental to the design, development, and implementation of a network of autonomous systems. Representations and statistical analysis of such data have helped reveal network patterns, trends, structures, and exceptions for even the most complex network datasets. Moreover, the methods inherently discover new patterns and anomalies enlightening the user, providing new knowledge and insight provoking further exploration. A well-designed visual analytic method can potentially summarize a month's worth of flow data, determine changes and provide overall context of the changes in the network.

Network flow information comprising autonomous systems can be modeled as a graph where the sites represent nodes and edges the amount of flow between them. Analyzing patterns detected from such a rich repertoire of information can provide us holistic insights on the overall picture of the system. For example, in an internet network where the nodes represent sites and the edges represent the flow interconnect between sites, the presence of very few hub-nodes generates load balancing issues and further the small-world property of certain networks negatively affects cache performance [173]. However, most existing real world network data change over time producing complex topological dependencies between them.

Devising routing algorithms or traffic management strategies for large physical networks is challenging as the network engineer/analyst does not know the context of the changes happening in a network. In such cases, to efficiently write algorithms that manage routing procedures that reliably utilize most resources of the networks traffic remains challenging. One approach to understanding such general network behavior is to analyze past path traversal and congestion control information to understand how changes in packet transfer have enabled large-scale changes in the routing procedure. Flow networks can help the analyst understand bottlenecks, queuing, and some unique requirements to tailor.

If we can better characterize this information, then will obtain answers to these questions: What are the sites that frequently collaborate with each other? How does the entire network gets affected due to the addition or deletion of collaborations? The analysis of this information is important for network engineers to construct new links or increase the bandwidth capacity between frequently communicating sites. The analysis can also help network engineers perform additional traffic engineering allowing efficient utilization of all the resources for efficient information transfer.

Existing techniques visualizing for dynamic graphs suffer from problems like change blindness [82], non-compliance of mental-map preservation and temporal visual scalability. Further, the methods do not satisfactorily focus on depicting the flow change or the importance of such change happening with respect to the entire global graph structure. This information is crucial to decision makers to assess the dynamics of the network and the its effect on certain sites of the network.

To support a network analyst to perform such comparison tasks, we introduce a novel context-aware difference graph technique to determine major changes between networks and identify its corresponding topological effect of the changes. To best visualize such information we detect the topological context for the most important change and link the information over different time-courses. We further use an umbrella metaphor to encode differences and its frame of reference between individual graphs in certain time periods. To accomplish this, we first compute a difference graph between timesteps and then map a subnetwork to each change. We further visually encode those differences into the dynamic graph visualization leading to a

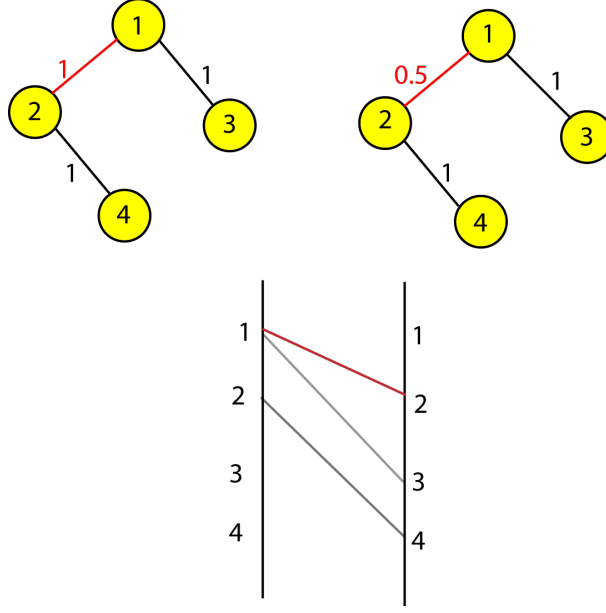


Figure 6.1. A context-aware difference graph framework showing how the change in graphs affect the dynamic topological structure of the graph.

context-preserving approach. The main contribution of this effort are:

- An algorithm that succinctly computes the topological context of the most important change in a traffic network flow graph
- A novel difference graph visualization technique that best visualizes the computed structure by the algorithm

A case study demonstrates the insights that can be obtained about the network system. We have applied the technique to a real-world autonomous system dataset.

Topological context is viewed as a piece of information (combination of nodes and edges) that can be used to characterize the situation of an entity. An entity in our case is an addition or deletion of an edge between time-points in a graph.

6.1 Related Work

Analysis of dynamic networks has gained tremendous attention in the past [191, 22, 3]. Our works span over existing work in the following fields: 1) Network Traffic Visual Analytics, 2) Difference Graphs. We primarily focus on efforts directly related to and affecting our specific use case.

6.1.1 Network Visual Analytics

Several visual analysis techniques are utilized to monitor, understand and model the network patterns underlying the network flow data. Early methods visualizing network data utilized parallel coordinates [201], heatmaps and network maps [165] to understand overall network loads and topology [199]. Methods have become more sophisticated, supporting more effective investigation of networks for scientific and managerial purposes. For example, methods by Erbacher *et al.* [57], Ball *et al.* [20], employ an overview+detail approach to analyze connectivity patterns from the level of intranet down to individual machines. Further, systems like by Goodall *et al.* [73] and Lakkaraju *et al.* [102] utilize aggregation and filtering mechanisms to reduce clutter and help users focus on regions of interest. Few of the other techniques aiding such scalable exploration of the data involve sliders, dynamic queries, brushing and linking.

The aforementioned techniques can be categorized into methods that utilize two or three-dimensional space. Works utilizing 3D techniques for network analysis [176, 96] mostly require sophisticated interactions techniques such as zooming, filtering, rotating, etc [165]. Such methods increase the amount of interaction load, suffer from occlusion and clutter. In contrast, two-dimensional methods such as Port Vis [116] provide an occlusion-free method to identify major events in dynamic networks.

Other methods like seeNet [25], use abstraction techniques to identify and characterize major events in the network flow data and the tool by Teoh *et al.* [177] focus on merging and utilizing multiple visualization views to explore complementary aspects of the data.

Visual methods in other domains such as brain networks employ linked visualization views [123, 40], flow-based techniques [126, 125] to better understand the brain activity.

However, most of the existing systems do not satisfactorily do not focus on temporal aspects of network flow. Further, systems fail to satisfactorily create situational awareness to the analyst using the system. A process must automatically assess the topological effect of the change in flow to better mitigate critical bottlenecks in the network. Further, computational methods should communicate this information as clearly as possible without information loss [165].

6.1.2 Difference Graphs

Traditional dynamic graph visual analysis approaches suffer from change-blindness, overdraw of visual elements, and inability to convey topological change effectively. Difference graph methods solve this specific problem by only depicting change between graphs in two timesteps. Given two graphs, only the changes concerning edges and nodes are visualized [15]. Difference graph has provided new insights into analyzing topological change in dynamic graphs. We focus on networks with constant node and changing edge information.

To deal with the problem of scalability in difference graphs, Archambault *et al.* [13] used hierarchies to depict large areas where the entire graph changes, providing general overview patterns. Work by Bourqui and Jourdan [34] analyzed edges that have similar pathways, focusing on structural similarity. Further work by Rufiange [155] based on a hybrid method involving small-multiples and animations, determined local topological changes between graphs.

However, difference graphs do not satisfactorily provide the context of change happening between graphs, inhibiting the analyst to assess the true topological effect of change and getting a general overview of the portrayed change over time. Missing context information makes interpretation of some alerts difficult or even impossible, and can therefore potentially lead to misjudgment of threats. These are the limitations of many existing systems:

- Hiding or unavailability of contextual information makes interpretation of traffic patterns difficult, potentially leading to misleading the decision-makers
- The effect of a low-level topological change over time and space is often not detected and not satisfactorily visualized.

We have extended traditional visual analytics methods by defining the topological context of the most important change occurring at a given timestep and visualizing them through as forward and backward pointing trees, allowing interactive topological effect querying. We use a multi-class bundling of edges to best represent this information in a visually clear and comprehensible manner.

6.2 Network Flow Data

The data for analysis is the routed UDP, TCP and ICMP network flows, captured by sensors and routers through the network gateway. Each flow describes the time, duration of a network connection at a transport layer. The packet counts and number of bytes transferred are stored, but not the contents. While the network flow data is bi-directional, we use a simplification strategy to simplify the underlying data to make it undirected. Nodes are defined as individual autonomous systems and the connections between the nodes depict a number of packets exchanged between them.

6.3 Goals

We were motivated by the need to understand and analyze the complex time-varying structures found in dynamic temporal graphs. We outline some of the primary design goals to be met by our method were the following:

Rate or growth of change occurring for each timestep. To understand the spreading or contraction dynamics, it is important that we define methods that enable us to see this in a useful way

Linking topological effect structures The linking mechanism of the contextual structures detected between different timesteps help us understand the overall architecture of the system.

Similarity of temporal structures The topologies having similar continuous structures should be grouped together to further enable easy exploration of data.

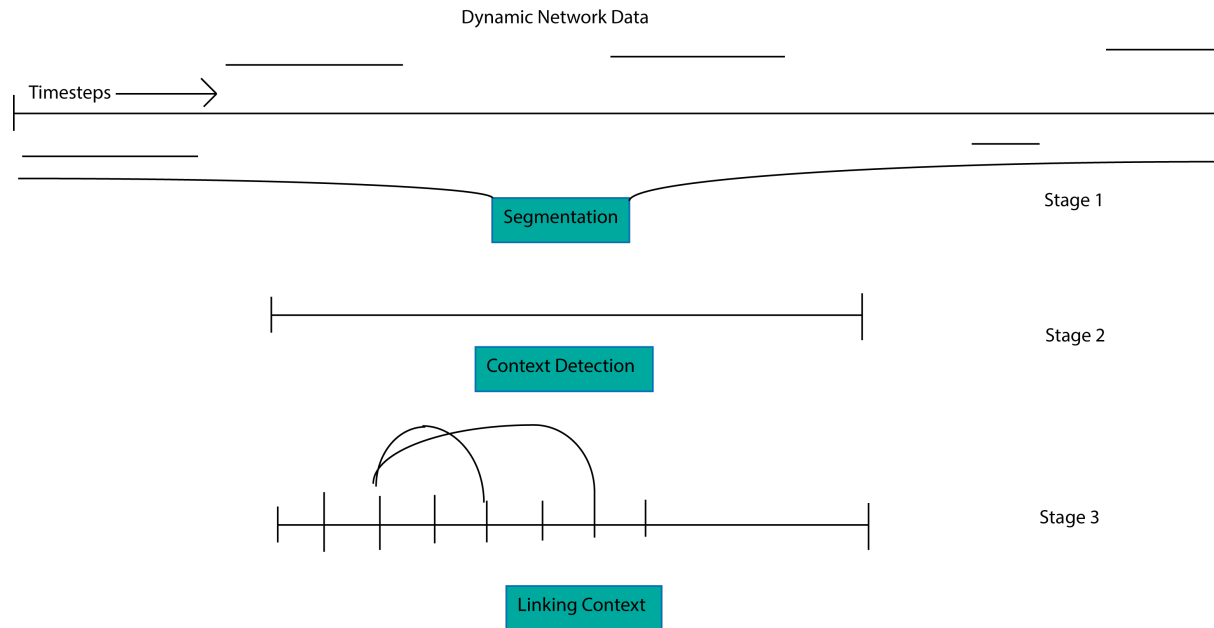


Figure 6.2. The figure explains our analysis pipeline that explores the complex structures embedded in dynamics graphs. In the first stage, using the community similarity metrics, we define similar graphs and group them into graph structures. We identify specific local affecting between important time-points. In the third stage, we link the detected contextual topological structure with other detected structures over other timepoints.

6.4 Method

Our technique involving algorithmic analysis and interactive visualizations easily helps us quantify the change happening at each timestep and explore their dependencies with other parts of the system. This very procedure allows us understand the overall structure of the entire system.

As shown in Fig 6.2, we define three major stages in our analysis methods, first, for each change occurring at every timestep we identify and compute the divergence using symmetric differences. We determine the topological effect happening for each change. We base the topological effect based on criteria that we discuss later Fig. 6.3. Finally, we link the topological effect of the change across all timesteps in the datasets to understand the overall changes caused by the system.

Such methods allow us to easily explore the propagation of events such as viruses, messages or bottlenecks in networks architectures more easily. We can do this on a global scale involving overlapping contextual topological structure.

6.5 Context-aware Difference Graphs

To define the context of change happening between time-sequence in a graph, we define algorithms and visualization techniques characterizing the topological effect of each of the changes.

For each edge change, we consider three major criteria:

- **Existence of graph differences** involves a binary function that mentions whether a change exist or not.
- **Rate of graph differences** involves determines the rate of topological change between timesteps.
- **Change in the community information** considers the amount of change involved in the structural communities over time.

6.5.1 Mathematical Formulation

We model a static graph $G = (V, E)$ as a pair consisting of vertices $V := \{v_1, \dots, v_n\}$ and edges $E \subseteq V \times V := \{e_1, \dots, e_m\}$. The edges may be weighted, *i.e.*, each $e \in E$ is attached by a value

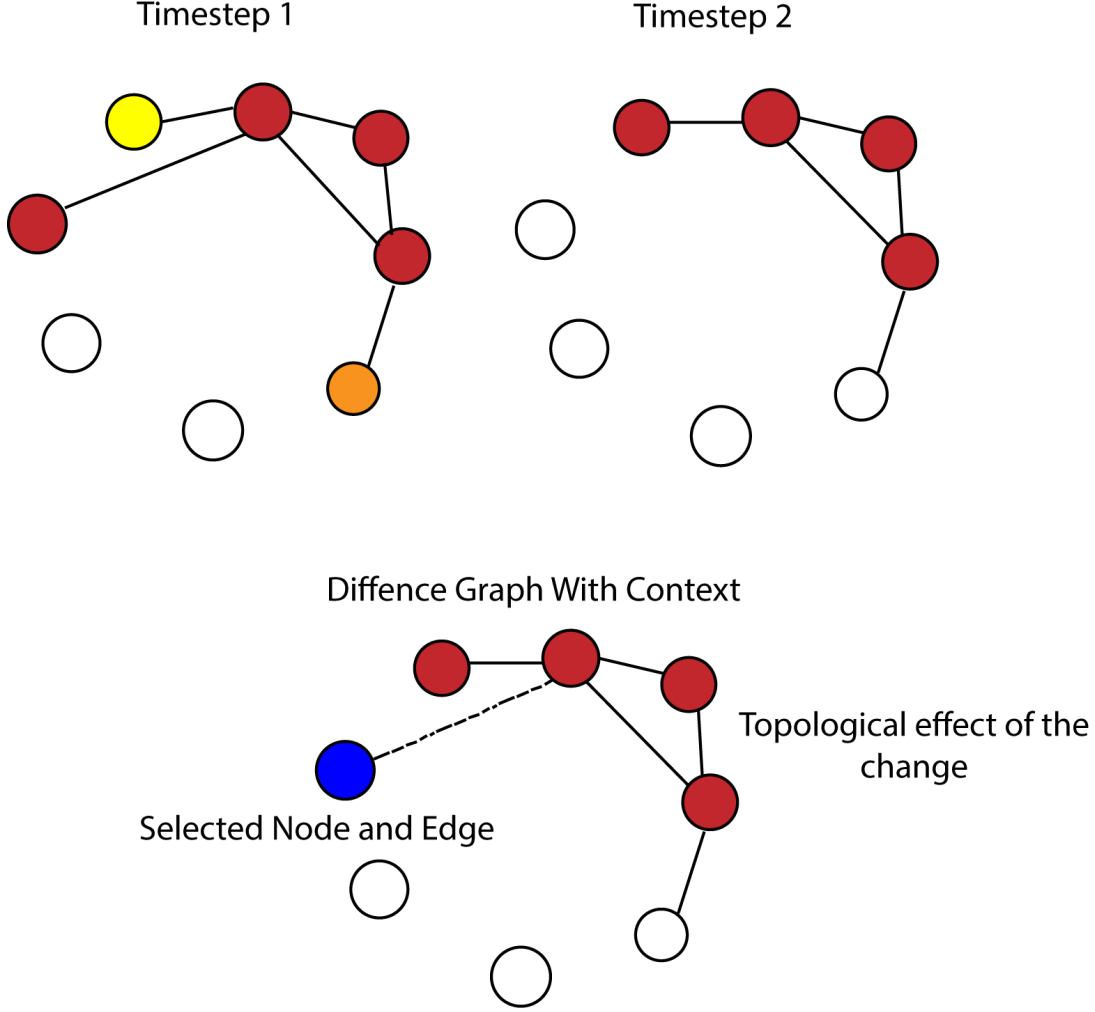


Figure 6.3. The graphs and their community structure are juxtaposed community structure. Our topological-effect algorithm is able to detect the structures that are affected based on the selected node and the change in the difference graph.

$e_w \in \mathbb{R}$ indicating the value of a certain edge e at a fixed time step.

In the dynamic case, we model the graph as a sequence Γ consisting of $k \in \mathbb{N}$ many individual graphs $G_i := (\mathbb{V}, \mathbb{E})$ expressing that each graph G_i consists of the same vertex and edge set while the vertex set $\mathbb{V}$ is obtained by building the union set of all vertices occurring over time. The edge set $\mathbb{E}$ is defined as $\mathbb{E} \subseteq \mathbb{V} \times \mathbb{V}$, but each edge in $\mathbb{E}$ is attached by a time-varying weight function $f : \mathbb{E} \times \mathbb{N} \rightarrow \mathbb{R}$ indicating the weight of each edge at a certain point in time. This data model allows the mathematical notation of a weighted directed dynamic graph (or network) which builds the basis for the difference graph computation and also the context-aware difference graph visualization.

6.5.2 Difference Graph Computation

There are several ways to compute difference graphs based on a graph sequence $\Gamma := \{G_1, \dots, G_k\}$. ■

The most natural way to do this only takes into account neighbored graphs in the sequence, i.e., G_i and G_{i+1} , $1 \leq i \leq k-1$. A more complex way to generate difference graphs observes longer graph subsequences and computes either the difference between first and last graph in the subsequence, or the sum of individual differences between subsequent graphs in the subsequence. To compute the difference value, a certain well-defined difference metric has to be applied. These approaches are technically described in this section.

As we deal with networks having stable node configurations, we keep the number of nodes constant. To analyze the changes in difference graphs, we define the importance of edge-change in the difference graph. Using edge-betweenness centrality [111], for each edge-change, we compute a corresponding sub-network where the context of the change is affected in the graph.

To further enable comparisons of graphs over discontinuous reoccurring time intervals, using cosine -similarity or Euclidean distance, we define a similarity matrix that identifies the dependencies between graphs at subsequent timesteps.

6.5.3 Difference Graph Computation

The underlying graph of stable nodes should not only show the changes in edges, but also the topological effects of those changes in the edges. In order to identify these structures from the underlying network flow data, we follow the criteria explained in section 6.5. We argue that in order to glean useful insights from the difference graphs it is important to not only show the changes in the underlying data, but also changes such as:

- Topological extent of change
- Linkage of the change between timesteps
- Subnetworks that depict contextual knowledge

To compute such a structure, we first identify the directionality of the change happening in the dynamic graph. For *e.g.*, changes of a community that is evolving over a small subnetwork throughout time accumulates more elements in the process, a difference graph depicting this

Difference Importance

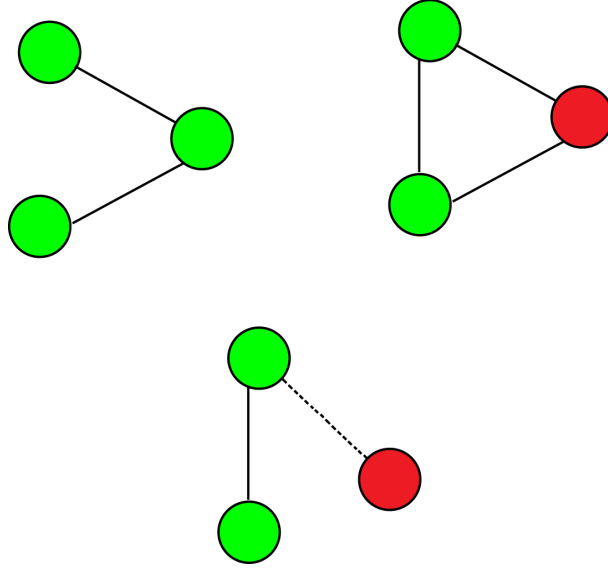


Figure 6.4. Direction-aware modeling of dynamic graphs they diffuse over time.

phenomenon should not only show and compute such a structure but must also compute the contributing elements that may have caused such a structure to form.

A way to quantify such a structure is to first identify the directionality of change happening in the graph. In order to mathematically define these changes, we first model the difference between graphs in terms of directionality. However, as we are interested in the definition of the directionality of impact over time, we further identify changes between detected subnetwork communities and its directionality over time.

We further define the dependencies between the minute small-scale changes with changes that happen in the difference graphs. We define a particular edge-change, $C_i(t_k, t_{k+1})$ between two time steps t_i and t_{i+1} as

$$C_i(t_k, t_{k+1}) \propto \|S_i \in (t_i, E_i, N_i), \| \quad (6.1)$$

where the change causes a specific temporal topological change in the data. The E_i , N_i , t_i are the affected topological elements from the edge-change. Our classification algorithm detects this effect that the change has over the different graph structure between different timesteps.

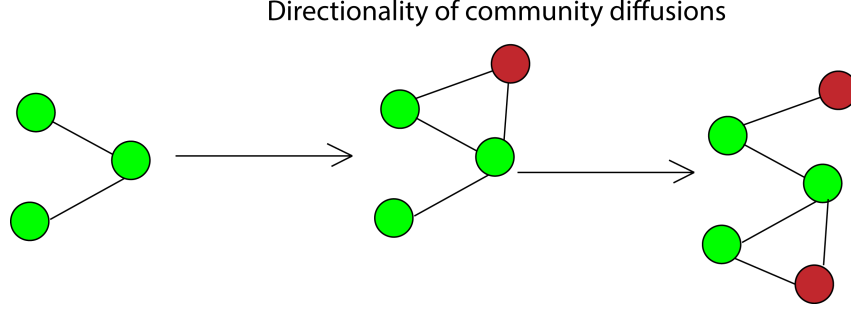


Figure 6.5. A small difference in the edges between green nodes induces a topological change, causing a community change between the green and red node.

$$Directionality(t_k, t_{k+1}) \propto \left(\sum_{i=0}^{N_1} e_i^t * BC_i^t \right) - \left(\sum_{j=0}^{N_2} e_j^{t+1} * BC_j^{t+1} \right) \quad (6.2)$$

where BC_i^t is the betweenness centrality of the i_{th} node at t_{th} timestep.

Algorithm 2: Topological Effect Detection Mechanism

Data: Sequence of graphs $G_{t1}, G_{t2}, G_{t3}, \dots$

Result: $(C_i(t_k, t_{k+1}), S_i)$ pairs defining the change and its effect on topology from timesteps t_k, t_{k+1}

- 1 Compute adjacency matrix containing pairwise similarity values between timesteps t_k to t_{k+n} using Eq.6.2;
 - 2 Identify the edge-betweenness for each change occurring in G_t and G_{t+1}
 - 3 The highest change in edge importance is removed from the analysis
 - 4 Apply local clustering with the identified node and edge that changes
 - 5 Rank all graphs G_{t_i} based on the average of non-zero values in i^{th} column;
 - 6 Aggregate all the edge-pair lists of G_{Av} to define average over all n graphs;
-

We use visualization to explore interesting patterns of repeated behavior, outliers, temporal shifts and counter trends.

Difference Visualizations of ESNET Data

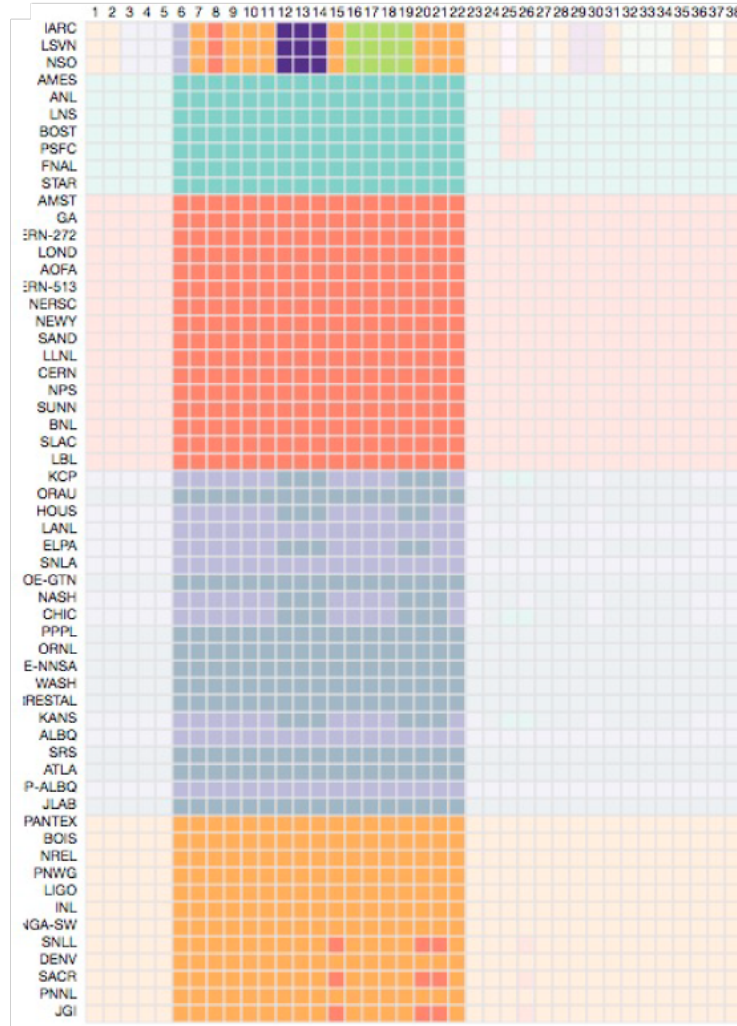


Figure 6.6. Color representing the community that a national laboratory/hub belongs to. X-axis the time-scale and the Y-axis the different labs. The detection algorithm found communities within the same spatial locations.

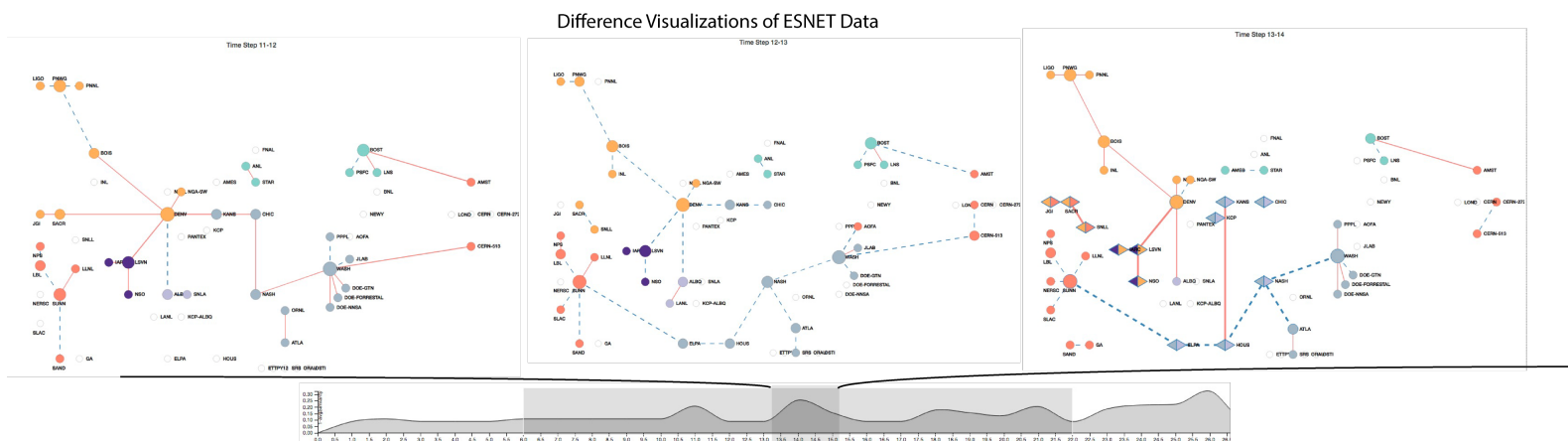


Figure 6.7. Topological changes happening between two timesteps. The larger the node size is the larger is a change in the laboratory/hubs. Blue dashed lines represent the reduction in network flow and red lines represent the addition of flow. Although significant change happens at each timestep, one can see a core-structure forming between IAR, LSVN, and NSO. They remain stable and then change their structure.

6.6 Application Example

We applied our visual analysis technique to the Netflow dataset obtained from ESNet, connecting sites across the United States of America and Europe for transfer of science experiments.

6.6.1 ESNet Dataset

ESnet provides a high-bandwidth, reliable connections that link scientists at national laboratories, universities, and other research institutions. We extracted data from the interconnect of supercomputers over a course of few months. We apply various simplification and thresholding methods to model a dynamic graph explaining the interconnect of the various supercomputers around the world.

The router topology dataset, see Fig. 6.6, examines network connectivity between sites in the United States and Europe. ESnet monitors network connections, collecting statistics on bytes sent/received and link performance logs. One monitoring tool, ESxSNMP, collects router-in and router-out bytes for every interface every 30 seconds. The tool records the packets transferred between sites at different times of a day. While the physical topology is fixed, the virtual topology of data movement changes dynamically, depending on the site's access to data.

Identification of context of the network topology in such cases is crucial to detect how the bottleneck between two nodes in a network affects a subnetwork of the topology. Furthermore, computing and visualizing the effect of this bottleneck over different time-periods of the data is crucial to prevent and diagnose the network configuration from reoccurring. For example, when the network engineers know about a particular bottleneck occurring over a specific two super-computers, they can apply more bandwidth infrastructure for that specific area of interest. One of the problems suffered by the network engineers is the unavailability of tools that help them understand the phenomena occurring over time.

Motivated by the above-mentioned use-cases, we further define the major customizations that we applied in our method for this specific data. The nodes, in this case, mean each supercomputer and the edges depict the communication patterns between the supercomputers. A *context* in such case reveals structures that are most affected by the delay or bottleneck between the communication channel of two supercomputers, Once we have the context of change

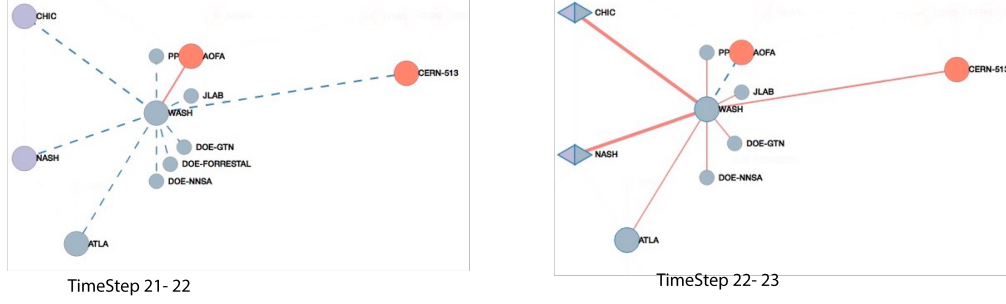


Figure 6.8. This represents the change happening at WASH. In comparison to all other nodes, this node exhibits substantial a lot of transient activity with sites surrounding the laboratories.

occurring at each timestep, the overlapping context structure reveals the most common subnetworks that are affected by the dynamics happening over multiple temporal scales.

In this case study, we examine network connection data for two days collected at a 30-minute intervals, showing packet exchanges between sites. The major evolving ESnet flow patterns are of great interest [49, 95]. Specifically, one seeks answers to these questions: “*What sites are frequently communicating with each other?*” and “*What are the changes in the network topology, how can we reliably depict this information?*”

The color represents the community that the national lab/hub belongs to. X-axis the time-scale and the Y-axis the different laboratories. It is surprising that the detection algorithm found communities within the same spatial locations. The image shows what are the topological changes happening between two timesteps. The larger the node size, the larger there is a change in the lab. Blue dashed lines represent the reduction in network flow and red lines represent the addition of flow. Although significant change happens at each timestep, one can see a core-structure forming between IAR, LSVN, and NSO. They remain stable and then change their structure over time. In Fig. 6.8, the visualization depicts the change happening at WASH. In comparison to all other nodes, WASH seems to have a lot of transient activity.

6.7 Conclusion and Outlook

In this paper, we presented a context-aware difference graph approach that additionally encodes the differences in graph sequences into a visually scalable dynamic graph visualization. To reach this goal, we first compute those changes in subsequent graphs based on certain well-defined graph metrics like edge existence, edge weights, or edge directions. Those can focus

on an individual edge or edge groups. The differences are visually incorporated in form of differently splatted node-link diagrams based on those changes. We discuss scalability issues concerning algorithmic as well as visual aspects while we evaluate our combined visualization by means of two application examples providing dynamic networks consisting of several thousand vertices, edges, and time steps. It would be possible, we plan to further evaluate our visualization technique in a comparative study or to add interaction techniques changing the visualized difference graph metrics on demand.

In the future, we plan to extend our system three major directions. One, focusing on supporting larger, denser and temporally scalable datasets and implementing them in real-time. The challenge is to optimize our algorithms such that the visualizations and analysis are rendered when the data comes in. Second, we would like to extend our algorithms to capture more complex temporal topological points involving more time-varying patterns. Third one should, enable intuitive comparison of patterns among different datasets and efficiently visualize this information in a visually clear manner.

Chapter 7

Conclusion and Future Work

We have introduced and discussed contributions to the visual analysis of data arising in neuroscience, social science, and autonomous systems. The research spanned the exploration of wide range of network properties, ranging from the modular structure of a graph to understanding change of structure of the graph. We analyzed our design problems and provided thoughts and suggestions for practitioners, researchers or scientists analyzing graph or network data arising in a variety of fields.

This dissertation investigated techniques and algorithms for mining, visualizing and understanding complex graph and network dataset. Visualizations are used for data analysis to communicate and explore the most relevant network information and its context. The research carried out had a specific scientific purpose, for *e.g.*, Brain Modulyzer, was driven by the need to explore hierarchical processes in the brain network, TimeSum had to explore the temporal structure of networks. We designed four major analysis systems that explore network data:

Brain Modulyzer is a linked interactive visual analysis framework to explore the modular hierarchical structures of brain networks based on fMRI data.

ClusterFlow is a novel visual analysis method for clustering and visualizing of the complex spatio-temporal patterns underlying ECoG data.

Time Sum is a novel abstraction and difference analysis tool technique to explore scalable temporal data.

NetReduce is an analysis framework that mines and visualizes important network flow pat-

terns in time-varying autonomous systems.

7.0.1 Visual Analysis Techniques

The development of the interactive systems produced major techniques easily applied to other visual analytics tools:

Brain Modulyzer is a hierarchical exploration technique supporting a novel set of interactions to explore modular structure.

ClusterFlow is a novel aggregation technique to explore the multi-scale structure arising in cluster evolution

TimeSum is data algorithms and visual designs to compress and detect topological shifts in a dynamic network data

NetReduce provides a difference analysis framework to explore of the important changes arising in a dynamic network

7.1 Future Directions

There are five main possible future directions that could move this interdisciplinary research forward.

7.1.1 Simplistic Visualizations

It is difficult to devise techniques that can present data through a simple visual narrative that users can easily relate to. Simple designs require minimal training and little computational overhead. Addressing these issue will increase the reachability and expressiveness of visualizations.

7.1.2 Temporal Aspects

The existing analysis algorithms assume that each node in a network may have one specific role, which may not always be the case. Networks are often multi-layered and multi-level, requiring sophisticated and new visualization methods to better represent such data. Visualization research depicting such information in a concise and precise manner is lacking. The challenge is to depict multi-community information in an easy-to-understand narrative.

7.1.3 Network-Specific Parameter Optimizations

The scale of the node data, its edge properties, and the temporal structure affect the layout, colors and the perception of the visualization. Visualizations are often carried out sophisticated by trial-and-error methods. Methods should consider the scale of data and optimize analysis, visualizations and interaction parameters without human intervention. This would decrease the context-switching cost spent for starting up new datasets in the visualization.

7.1.4 Multi-modal Visual Analytics

Data often have a multitude of attributes. For *example*, a functional cluster in an fMRI dataset always has a corresponding anatomical information. Efficient exploration of such multi-modal datasets through superimposition of visual elements is still an under-developed research area in visual analytics research.

7.1.5 Models of Visual Thinking

One can think of devising learning systems that can be trained based on existing space of visual designs and glyphs to produce a design that optimally communicates data without inherent biases for a specific domain. The problem is hard, since visualizations are designed based on human intuition and user studies that best know the domain. Initially, an automated visual design system could benefit from mining important visual design information discussed in the domain literature.

7.2 Conclusion

This dissertation contributes to the multi-disciplinary field of *Visual Network Analytics*. The methods introduced are general to some extent, documented by the fact that we could successfully apply them to social networks and autonomous systems. Methods involving analytical and design thinking are clearly significant to explore and comprehend complex graphs and network data.

References

- [1] James Abello, Frank Van Ham, and Neeraj Krishnsan. Ask-graphview: A large scale graph visualization system. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):669–676, 2006.
- [2] Sophie Achard, Raymond Salvador, Brandon Whitcer, John Suckling, and ED Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience*, 26(1):63–72, 2006.
- [3] Charu Aggarwal and Karthik Subbian. Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1):10, 2014.
- [4] Jae-wook Ahn, Catherine Plaisant, and Ben Shneiderman. A task taxonomy for network evolution analysis. *IEEE transactions on visualization and computer graphics*, 20(3):365–376, 2014.
- [5] David Akers, Anthony Sherbondy, Rachel Mackenzie, Robert Dougherty, and Brian Wandell. Exploration of the brain’s white matter pathways with dynamic queries. In *Visualization, 2004. IEEE*. IEEE, 2004.
- [6] Martin L. Albert, Robert G. Feldman, and Anne L. Willis. The subcortical dementia of progressive supranuclear palsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 37(2):121–130, 1974.
- [7] Elena A. Allen, Eswar Damaraju, Sergey M. Plis, Erik B. Erhardt, Tom Eichele, and Vince D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, page bhs352, 2012.
- [8] Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. Weighted graph comparison techniques for brain connectivity analysis. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pages 483–492. ACM, 2013.

- [9] Gennady Andrienko and Natalia Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*, pages 51–58. IEEE, 2008.
- [10] Gennady Andrienko, Natalia Andrienko, et al. Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10):1577–1600, 2010.
- [11] Gennady Andrienko, Natalia Andrienko, Martin Mladenov, Michael Mock, and Christian Politz. Identifying place histories from activity traces with an eye to parameter impact. *IEEE Transactions on Visualization and Computer Graphics*, 18(5):675–688, 2012.
- [12] Natalia Andrienko, Gennady Andrienko, and Peter Gatalsky. Exploratory spatio-temporal visualization: an analytical review. *J. of Visual Languages & Computing*, 14(6):503–541, 2003.
- [13] Daniel Archambault. Structural differences between two graphs through hierarchies. In *Proceedings of Graphics Interface 2009*, pages 87–94. Canadian Information Processing Society, 2009.
- [14] Daniel Archambault, Helen Purchase, and Bruno Pinaud. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):539–552, 2011.
- [15] Daniel Archambault, Helen C. Purchase, and Bruno Pinaud. Difference map readability for dynamic graphs. In *International Symposium on Graph Drawing*, pages 50–61. Springer, 2010.
- [16] Alex Arenas, Alberto Fernandez, and Sergio Gomez. Analysis of the structure of complex networks at different resolution levels. *New J. of Physics*, 10(5):053039, 2008.
- [17] Elaine Astrand, Claire Wardak, and Suliann Ben Hamed. Selective visual attention to drive cognitive brain-machine interfaces: from concepts to neurofeedback and rehabilitation applications. *Frontiers in systems neuroscience*, 8, 2014.
- [18] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, JD. Fekete, and T. Grabowski. Small multiples: Piling time to explore temporal patterns in dynamic networks. In *Computer Graphics Forum*, volume 34, pages 31–40. Wiley Online Library, 2015.
- [19] Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. Graphdiaries: animated transitions and temporal navigation for dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2014.
- [20] Robert Ball, Glenn A Fink, and Chris North. Home-centric visualization of network traffic for security administration. In *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pages 55–64. ACM, 2004.
- [21] Danielle S Bassett, Jesse A Brown, Vibhas Deshpande, Jean M Carlson, and Scott T Grafton. Conserved and variable architecture of human white matter connectivity. *Neuroimage*, 54(2), 2011.

- [22] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. The state of the art in visualizing dynamic graphs. *EuroVis STAR*, 2, 2014.
- [23] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. A taxonomy and survey of dynamic graph visualization. In *Computer Graphics Forum*. Wiley Online Library, 2016.
- [24] Moritz Y. Becker and Isabel Rojas. A graph layout algorithm for drawing metabolic pathways. *Bioinformatics*, 17(5), 2001.
- [25] Richard A. Becker, Stephen G. Eick, and Allan R. Wilks. Visualizing network data. *IEEE Transactions on visualization and computer graphics*, 1(1):16–28, 1995.
- [26] Pierre Bellec. Mining the hierarchy of resting-state brain networks: selection of representative clusters in a multiscale structure. In *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*, pages 54–57. IEEE, 2013.
- [27] Johanna Beyer, Ali Al-Awami, Narayanan Kasthuri, Jeff W Lichtman, Hanspeter Pfister, and Markus Hadwiger. Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12), 2013.
- [28] Gleb Bezgin, Andrew T. Reid, Dirk Schubert, and Rolf Kötter. Matching spatial with ontological brain regions using java tools for visualization, database access, and integrated data analysis. *Neuroinformatics*, 7(1):7–22, 2009.
- [29] Bharat Biswal, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 1995.
- [30] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Renaud Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10008, 2008.
- [31] Paolo Boldi, Marco Rosa, Massimo Santini, and Sebastiano Vigna. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *Proceedings of the 20th international conference on World wide web*, pages 587–596. ACM, 2011.
- [32] Phillip Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, 2007.
- [33] Kristofer E. Bouchard, Nima Mesgarani, Keith Johnson, and Edward F. Chang. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441):327–332, 2013.
- [34] Romain Bourqui and Fabien Jourdan. Revealing subnetwork roles using contextual visualization: Comparison of metabolic networks. In *Information Visualisation, 2008. IV’08. 12th International Conference*, pages 638–643. IEEE, 2008.

- [35] Ulrik Brandes. A faster algorithm for betweenness centrality*. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.
- [36] Urs Braun, Axel Schäfer, Henrik Walter, Susanne Erk, Nina Romanczuk-Seiferth, Leila Haddad, Janina I Schweiger, Oliver Grimm, Andreas Heinz, Heike Tost, et al. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences*, 112(37):11678–11683, 2015.
- [37] Jesse A. Brown, Jeffrey D. Rudie, Anita Bandrowski, John D. Van Horn, and Susan Y. Bookheimer. The ucla multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. *Frontiers in neuroinformatics*, 6, 2012.
- [38] Stefan Bruckner, Veronika Solteszova, Eduard Groller, Jiri Hladuvka, Katja Buhler, Jai Y Yu, and Barry J Dickson. Braingazer-visual queries for neurobiology research. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6), 2009.
- [39] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.
- [40] Vince D. Calhoun and Tulay Adali. Time-varying brain connectivity in fmri data: Whole-brain data-driven approaches for capturing and characterizing dynamic states. *IEEE Signal Processing Magazine*, 33(3):52–66, 2016.
- [41] Jin Cao, Keith Worsley, et al. The geometry of correlation fields with an application to functional connectivity of the brain. *The Annals of Applied Probability*, 9(4):1021–1057, 1999.
- [42] Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin, and Ching-Yung Lin. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):280–289, 2016.
- [43] Micaela Y. Chan, Denise C. Park, Neil K. Savalia, Steven E. Petersen, and Gagan S. Wig. Decreased segregation of brain systems across the healthy adult lifespan. *Proc. of the National Academy of Sciences*, 111(46):E4997–E5006, 2014.
- [44] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)*, 41(1):2, 2008.
- [45] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)*, 41(1):2, 2009.
- [46] Robert W Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.
- [47] Nicolas A Crossley, Andrea Mechelli, Jessica Scott, Francesco Carletti, Peter T Fox, Philip McGuire, and Edward T Bullmore. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain*, 137(8):2382–2395, 2014.

- [48] Eswar Damaraju, Ellena A. Allen, A Belger, Judith M. Ford, et al. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage: Clinical*, 5:298–308, 2014.
- [49] Eli Dart, Lauren Rotman, Brian Tierney, Mary Hester, and Jason Zurawski. The science dmz: A network design pattern for data-intensive science. *Scientific Programming*, 22(2):173–185, 2014.
- [50] Alberto Del Bimbo, Mauro Mugnaini, Pietro Pala, and Franco Turco. Visual querying by color perceptive regions. *Pattern recognition*, 31, 1998.
- [51] Gopikrishna Deshpande, Priya Santhanam, and Xiaoping Hu. Instantaneous and causal connectivity in resting state brain networks derived from functional mri data. *Neuroimage*, 54(2), 2011.
- [52] Cody Dunne and Ben Shneiderman. Motif simplification: improving network visualization readability with fan, connector, and clique glyphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3247–3256. ACM, 2013.
- [53] Peter Eades and Mao Lin Huang. Navigating clustered graphs using force-directed methods. *J. Graph Algorithms Appl.*, 4(3):157–181, 2000.
- [54] Nathan Eagle and Alex Sandy Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.
- [55] Victor M. Eguiluz, Dante R. Chialvo, Guillermo A. Cecchi, Marwan Baliki, and A. Vania Apkarian. Scale-free brain functional networks. *Physical review letters*, 94(1):018102, 2005.
- [56] Niklas Elmqvist and Jean-Daniel Fekete. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, 16(3):439–454, 2010.
- [57] Robert F Erbacher. Visual traffic monitoring and evaluation. In *ITCom 2001: International Symposium on the Convergence of IT and Communications*, pages 153–160. International Society for Optics and Photonics, 2001.
- [58] Robert F. Erbacher, Kenneth L. Walker, and Deborah A. Frincke. Intrusion and misuse detection in large-scale systems. *Computer Graphics and Applications, IEEE*, 22(1):38–47, 2002.
- [59] Damien A Fair, Alexander L Cohen, Nico UF Dosenbach, Jessica A Church, Francis M Miezin, Deanna M Barch, Marcus E Raichle, Steven E Petersen, and Bradley L Schlaggar. The maturing architecture of the brain’s default network. *Proc. of the National Academy of Sciences*, 105, 2008.

- [60] Damien A Fair, Alexander L Cohen, Jonathan D Power, Nico UF Dosenbach, Jessica A Church, Francis M Miezin, Bradley L Schlaggar, and Steven E Petersen. Functional brain networks develop from a local to distributed organization. *PLoS computational biology*, 5(5):e1000381, 2009.
- [61] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.
- [62] Anna Fredrikson, Chris North, Catherine Plaisant, and Ben Shneiderman. Temporal, geographical and categorical aggregations viewed through coordinated displays: a case study with highway incident data. In *Proc. on new paradigms in information visualization and manipulation*, pages 26–34. ACM, 1999.
- [63] Yaniv Frishman and Ayellet Tal. Dynamic drawing of clustered graphs. In *INFOVIS*, pages 191–198. IEEE, 2004.
- [64] Karl J Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping*, 2(1-2):56–78, 1994.
- [65] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.
- [66] Raquel C. Gardner, Adam L. Boxer, Andrew Trujillo, Jacob B. Mirsky, Christine C. Guo, Efsthios D. Gennatas, Hilary W. Heuer, Eric Fine, Juan Zhou, Joel H. Kramer, Bruce L. Miller, and William W. Seeley. Intrinsic connectivity network disruption in progressive supranuclear palsy. *Annals of Neurology*, 73, 2013.
- [67] Peter Gatalsky, Natalia Andrienko, and Gennady Andrienko. Interactive analysis of event data using space-time cube. In *INFOVIS*, pages 145–152, 2004.
- [68] Stephan Gerhard, Alessandro Daducci, Alia Lemkaddem, Reto Meuli, Jean-Philippe Thiran, and Patric Hagmann. The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Frontiers in neuroinformatics*, 5, 2011.
- [69] Cedric E Ginestet and Andrew Simmons. Statistical parametric network analysis of functional connectivity dynamics during a working memory task. *Neuroimage*, 55(2):688–704, 2011.
- [70] Rainer Goebel, Fabrizio Esposito, and Elia Formisano. Analysis of functional image analysis contest (fiac) data with brainvoyager qx: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, 27(5):392–401, 2006.
- [71] Jozien BM. Goense and Nikos K. Logothetis. Neurophysiology of the bold fmri signal in awake monkeys. *Current Biology*, 18(9):631–640, 2008.
- [72] Lawrence I. Golbe. Progressive supranuclear palsy. *Seminars in Neurology*, 34(02):151–159, 2014.

- [73] John R. Goodall, Wayne G. Lutters, Penny Rheingans, and Anita Komlodi. Preserving the big picture: Visual network traffic analysis with tnv. In *Visualization for Computer Security, 2005.(VizSEC 05). IEEE Workshop on*, pages 47–54. IEEE, 2005.
- [74] Bernhard Graimann, Jane E. Huggins, SP. Levine, and G. Pfurtscheller. Visualization of significant erd/ers patterns in multichannel eeg and ecog data. *Clinical Neurophysiology*, 113(1):43–47, 2002.
- [75] Derek Greene, Daniel Archambault, Václav Belák, and Pádraig Cunningham. Textluas: tracking and visualizing document and term clusters in dynamic text data. *arXiv preprint arXiv:1502.04609*, 2014.
- [76] Derek Greene, Donal Doyle, and Padraig Cunningham. Tracking the evolution of communities in dynamic social networks. In *Advances in social networks analysis and mining (ASONAM), 2010 international conference on*, pages 176–183. IEEE, 2010.
- [77] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028), 2005.
- [78] Yaroslav O. Halchenko, Stephen José Hanson, and Barak A. Pearlmuter. Multimodal integration: fmri, mri, eeg, meg. *Advanced image processing in magnetic resonance imaging*, pages 223–265, 2005.
- [79] Mark Harrower and Cynthia A Brewer. Colorbrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic J.*, 40(1):27–37, 2003.
- [80] Bin He, Lin Yang, Christopher Wilke, and Han Yuan. Electrophysiological imaging of brain activity and connectivity challenges and opportunities. *Biomedical Engineering, IEEE Transactions on*, 58(7):1918–1931, 2011.
- [81] Yong He, Jinhui Wang, Liang Wang, Zhang J Chen, Chaogan Yan, Hong Yang, Hehan Tang, Chaozhe Zhu, Qiyong Gong, Yufeng Zang, et al. Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PloS one*, 4(4):e5226, 2009.
- [82] Christopher G. Healey and James T. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2012.
- [83] Jeffrey Heer and Danah Boyd. Vizster: Visualizing online social networks. In *Infovis*, pages 32–39. IEEE, 2005.
- [84] Nathalie Henry, J. Fekete, and Michael J. McGuffin. Nodetrix: a hybrid visualization of social networks. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1302–1309, 2007.
- [85] Ivan Herman, Guy Melançon, and M Scott Marshall. Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1), 2000.

- [86] Dora Hermes, Kai J Miller, Herke Jan Noordmans, Mariska J Vansteensel, and Nick F Ramsey. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *Journal of neuroscience methods*, 185(2):293–298, 2010.
- [87] Charles AR Hoare. Quicksort. *The Computer Journal*, 5(1):10–16, 1962.
- [88] Barry Horwitz. The elusive concept of brain connectivity. *Neuroimage*, 19(2):466–470, 2003.
- [89] Yifan Hu, Stephen G Kobourov, and Sankar Veeramoni. Embedding, clustering and coloring for dynamic maps. In *Pacific Visualization Symposium (PacificVis), 2012 IEEE*, pages 33–40. IEEE, 2012.
- [90] R. Matthew Hutchison, Thilo Womelsdorf, Elena A. Allen, Peter A. Bandettini, Vince D. Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H. Duyn, Gary H. Glover, Javier Gonzalez-Castillo, et al. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378, 2013.
- [91] Mikael Jern and Johan Franzen. "geoanalytics"-exploring spatio-temporal and multivariate data. In *Tenth International Conference on Information Visualisation (IV'06)*, pages 25–31. IEEE, 2006.
- [92] Radu Jianu, Cagatay Demiralp, and David H Laidlaw. Exploring brain connectivity with two-dimensional neural maps. *Visualization and Computer Graphics, IEEE Transactions on*, 18(6), 2012.
- [93] Brian Johnson and Ben Shneiderman. Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Visualization, 1991. Visualization'91, Proc., IEEE Conference on*, pages 284–291. IEEE, 1991.
- [94] Daniel A Keim, Tamara Munzner, Fabrice Rossi, and Michael Verleysen. Bridging information visualization with machine learning (dagstuhl seminar 15101). In *Dagstuhl Reports*, volume 5. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.
- [95] Mariam Kiran, Eric Pouyoul, Anu Mercian, Brian Tierney, Chin Guok, and Inder Monga. Enabling intent to configure scientific networks for high performance demands. *Future Generation Computer Systems*, 2017.
- [96] Anita Komlodi, Penny Rheingans, Utkarsha Ayachit, John R. Goodall, and Amit Joshi. A user-centered look at glyph-based security visualization. In *Visualization for Computer Security, 2005.(VizSEC 05). IEEE Workshop on*, pages 21–28. IEEE, 2005.
- [97] Brittany Kondo, Hrim Mehta, and Christopher Collins. Glidgets: Interactive glyphs for exploring dynamic graphs, 2014.
- [98] Anna Korzeniewska, Piotr J Franaszczuk, Ciprian M Crainiceanu, Rafał Kuś, and Nathan E Crone. Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (erc) analysis of human electrocorticography (ecog). *Neuroimage*, 56(4):2218–2237, 2011.

- [99] Danai Koutra, Joshua T Vogelstein, and Christos Faloutsos. Deltacon: A principled massive-graph similarity function. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 162–170. SIAM, 2013.
- [100] Mark A Kramer, Uri T Eden, Kyle Q Lepage, Eric D Kolaczyk, Matt T Bianchi, and Sydney S Cash. Emergence of persistent networks in long-term intracranial eeg recordings. *The J. of Neuroscience*, 31(44):15757–15767, 2011.
- [101] Jan Kubanek and Gerwin Schalk. Neuralact: a tool to visualize electrocortical (ecog) activity on a three-dimensional model of the cortex. *Neuroinformatics*, 13(2):167–174, 2014.
- [102] Kiran Lakkaraju, William Yurcik, and Adam J Lee. Nvisionip: netflow visualizations of system state for security situational awareness. In *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pages 65–72. ACM, 2004.
- [103] Renaud Lambiotte, J-C Delvenne, and Mauricio Barahona. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*, 2008.
- [104] Roan A LaPlante, Linda Douw, Wei Tang, and Steven M Stuffelbeam. The connectome visualization utility: Software for visualization of human brain networks. *PloS one*, 9(12):e113838, 2014.
- [105] Nilli Lavie and Sally Cox. On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. *Psychological Science*, 8(5):395–396, 1997.
- [106] Suzee E. Lee, Ana C. Sias, Maria Luisa Mandelli, Jesse A. Brown, Alainna B. Brown, Anna M. Khazenzon, Anna A. Vidovszky, Theodore P. Zanto, Anna M. Karydas, Mochtar Pribadi, Deepika Dokuru, Giovanni Coppola, Dan H. Geschwind, Rosa Rademakers, Maria Luisa Gorno-Tempini, Howard J. Rosen, Bruce L. Miller, and William W. Seeley. Network degeneration and dysfunction in presymptomatic c9orf72 expansion carriers. In *NeuroImage: Clinical*, 2017.
- [107] Kaiming Li, Lei Guo, Carlos Faraco, Dajiang Zhu, Hanbo Chen, Yixuan Yuan, Jinglei Lv, Fan Deng, Xi Jiang, Tuo Zhang, et al. Visual analytics of brain networks. *NeuroImage*, 61(1):82–97, 2012.
- [108] Qingsong Liu, Yifan Hu, Lei Shi, Xinzhu Mu, Yutao Zhang, and Jie Tang. Egonetcloud: Event-based egocentric dynamic network visualization. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*, pages 65–72. IEEE, 2015.
- [109] Yong Liu, Meng Liang, Yuan Zhou, Yong He, Yihui Hao, Ming Song, Chunshui Yu, Haihong Liu, Zhening Liu, and Tianzi Jiang. Disrupted small-world networks in schizophrenia. *Brain*, 131(4):945–961, 2008.

- [110] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics (Proc. ACM SIGGRAPH 87)*, 21(4):163–169, 1987.
- [111] LongJason Lu and Minlu Zhang. Edge betweenness centrality. In *Encyclopedia of systems biology*, pages 647–648. Springer, 2013.
- [112] Daniel S Margulies, Joachim Böttger, Aimi Watanabe, and Krzysztof J Gorgolewski. Visualizing the human connectome. *NeuroImage*, 80:445–461, 2013.
- [113] Daisuke Mashima, Stephen G Kobourov, and Yifan Hu. Visualizing dynamic data with maps. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1424–1437, 2012.
- [114] Dan McFarland. Mcfarland classroom dataset. <https://web.stanford.edu/group/sonia/examples/McFarlaClass.html>, 2001.
- [115] Daniel A McFarland. Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance 1. *American Journal of Sociology*, 107(3):612–678, 2001.
- [116] Jonathan McPherson, Kwan-Liu Ma, Paul Krystosk, Tony Bartoletti, and Marvin Christensen. Portvis: a tool for port-based detection of security events. In *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pages 73–81. ACM, 2004.
- [117] Christian Meisel and Christian Kuehn. Scaling effects and spatio-temporal multilevel dynamics in epileptic seizures. *PLoS One*, 7(2):e30371, 2012.
- [118] David Meunier, Sophie Achard, Alexa Morcom, and Ed Bullmore. Age-related changes in modular organization of human brain functional networks. *Neuroimage*, 44(3), 2009.
- [119] David Meunier, Renaud Lambiotte, Alex Fornito, Karen D Ersche, and Edward T Bullmore. Hierarchical modularity in human brain functional networks. *Frontiers in neuroinformatics*, 3, 2009.
- [120] Andreas Meyer-Lindenberg, Jean-Baptiste Poline, Philip D Kohn, John L Holt, Michael F Egan, Daniel R Weinberger, and Karen Faith Berman. Evidence for abnormal cortical functional connectivity during working memory in schizophrenia. *American J. of Psychiatry*, 2001.
- [121] Kai J Miller, Pradeep Shenoy, John W Miller, Rajesh PN Rao, Jeffrey G Ojemann, et al. Real-time functional brain mapping using electrocorticography. *Neuroimage*, 37(2):504–507, 2007.
- [122] Stefano Monti, Pablo Tamayo, Jill Mesirov, and Todd Golub. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, 52(1-2):91–118, 2003.

- [123] Sugeerth Murugesan, Kristofer Bouchard, Jesse A. Brown, Bernd Hamann, William W. Seeley, Andrew Trujillo, and Gunther H. Weber. Brain modulyzer: Interactive visual analysis of functional brain connectivity. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 14(4):805–818, July 2017.
- [124] Sugeerth Murugesan, Kristofer Bouchard, Edward Chang, Max Dougherty, Bernd Hamann, and Gunther H. Weber. Hierarchical spatio-temporal visual analysis of cluster evolution in electrocorticography data. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB ’16, pages 630–639, New York, NY, USA, 2016. ACM.
- [125] Sugeerth Murugesan, Kristofer Bouchard, Edward Chang, Max Dougherty, Bernd Hamann, and Gunther H. Weber. Hierarchical spatio-temporal visual analysis of cluster evolution in electrocorticography data. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 630–639. ACM, 2016.
- [126] Sugeerth Murugesan, Kristofer Bouchard, Edward Chang, Max Dougherty, Bernd Hamann, and Gunther H. Weber. Multi-scale visual analysis of time-varying electrocorticography data via clustering of brain regions. *BMC Bioinformatics*, 18(Suppl 6):236, 2017.
- [127] Ali Yener Mutlu, Edward Bernat, and Selin Aviyente. A signal-processing-based approach to time-varying graph analysis for dynamic brain network identification. *Computational and mathematical methods in medicine*, 2012, 2012.
- [128] Steven M. Nelson, Alexander L. Cohen, Jonathan D. Power, Gagan S. Wig, Francis M. Miezin, Mark E. Wheeler, Katerina Velanova, David I. Donaldson, Jeffrey S. Phillips, Bradley L. Schlaggar, et al. A parcellation scheme for human left lateral parietal cortex. *Neuron*, 67(1):156–170, 2010.
- [129] Mark EJ Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.
- [130] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69, 2004.
- [131] Quan Nguyen, Peter Eades, and Seok-Hee Hong. On the faithfulness of graph visualizations. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 209–216. IEEE, 2013.
- [132] Gregory M. Nielson. On marching cubes. *IEEE Transactions on Visualization and Computer Graphics*, 9(3):341–351, 2003.
- [133] Stephen C North. Drawing graphs with neato. *NEATO User Manual*, 11, 2004.

- [134] Charlotte Plug, Jianhong Cecilia Xia, and Craig Caulfield. Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis & Prevention*, 43(6):1937–1946, 2011.
- [135] Sophie Ponten, Fabrice Bartolomei, and Cornelis Stam. Small-world networks and epilepsy: graph theoretical analysis of intracerebrally recorded mesial temporal lobe seizures. *Clinical neurophysiology*, 118(4):918–927, 2007.
- [136] Cristhian Potes, Peter Brunner, Aysegul Gunduz, Robert T Knight, and Gerwin Schalk. Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing. *Neuroimage*, 97:188–195, 2014.
- [137] Jonathan D. Power, Alexander L. Cohen, Steven M. Nelson, Gagan S. Wig, Kelly Anne Barnes, Jessica A. Church, Alecia C. Vogel, Timothy O. Laumann, Fran M. Miezin, Bradley L. Schlaggar, and Steven E. Petersen. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.
- [138] Helen C. Purchase, Eve Hoggan, and Carsten Görg. How important is the mental map?—an empirical investigation of a dynamic graph layout algorithm. In *International Symposium on Graph Drawing*, pages 184–195. Springer, 2006.
- [139] Marcus Raitner. Visual navigation of compound graphs. In *Graph drawing*, pages 403–413. Springer, 2005.
- [140] Ashish Raj, Eve LoCastro, Amy Kuceyeski, Duygu Tosun, Norman Relkin, Michael Weiner, Alzheimer’s Disease Neuroimaging Initiative (ADNI, et al. Network diffusion model of progression predicts longitudinal patterns of atrophy and metabolism in alzheimer’s disease. *Cell reports*, 2015.
- [141] Barnaly Rashid, Eswar Damaraju, Godfrey D Pearlson, and Vince D Calhoun. Dynamic connectivity states estimated from resting fmri identify differences among schizophrenia, bipolar disorder, and healthy control subjects. *Frontiers in human neuroscience*, 8:897, 2014.
- [142] Khairi Reda, Chayant Tantipathananandh, Tanya Berger-Wolf, Jason Leigh, and Andrew Johnson. Socioscape—a tool for interactive exploration of spatio-temporal group dynamics in social networks. In *INFOVIS*, 2009.
- [143] Khairi Reda, Chayant Tantipathananandh, Andrew Johnson, Jason Leigh, and Tanya Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. In *Computer Graphics Forum*, volume 30, pages 1061–1070. Wiley Online Library, 2011.
- [144] Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):016110, 2006.

- [145] Florian Reitz, Mathias Pohl, and Stephan Diehl. Focused animation of dynamic compound graphs. In *2009 13th International Conference Information Visualisation*, pages 679–684. IEEE, 2009.
- [146] Jun Rekimoto and Mark Green. The information cube: Using transparency in 3d information visualization. In *Proceedings of the Third Annual Workshop on Information Technologies & Systems (WITS’93)*, pages 125–132, 1993.
- [147] Andre Santos Ribeiro, Luis Miguel Lacerda, and Hugo A Ferreira. Multimodal imaging brain connectivity analysis toolbox (mibca). Technical report, PeerJ PrePrints, 2014.
- [148] Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 233–240. IEEE, 2005.
- [149] George G Robertson, Jock D Mackinlay, and Stuart K Card. Cone trees: animated 3d visualizations of hierarchical information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 189–194. ACM, 1991.
- [150] Martin Rosvall and Carl T. Bergstrom. Mapping change in large networks. *PloS one*, 5(1):e8694, 2010.
- [151] Mikail Rubinov and Ed Bullmore. Fledgling pathoconnectomics of psychiatric disorders. *Trends in Cognitive Sciences*, 17, 2013.
- [152] Mikail Rubinov, Stuart A. Knock, Cornelis J. Stam, Sifis Micheloyannis, Anthony WF. Harris, Leanne M. Williams, and Michael Breakspear. Small-world properties of nonlinear brain activity in schizophrenia. *Human brain mapping*, 30(2):403–416, 2009.
- [153] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059 – 1069, 2010. Computational Models of the Brain.
- [154] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.
- [155] Sébastien Rufange and Michael J McGuffin. Diffani: Visualizing dynamic graphs with a hybrid of difference maps and animation. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2556–2565, 2013.
- [156] Marta Sales-Pardo, Roger Guimera, Andre A Moreira, and Luís A Nunes Amaral. Extracting the hierarchical organization of complex systems. *Proc. of the National Academy of Sciences*, 104(39), 2007.
- [157] Arnaud Sallaberry, Chris Muelder, and Kwan-Liu Ma. Clustering, visualizing, and navigating for large dynamic graphs. In *20th International Symposium on Graph Drawing*, volume 7704, pages 487–498. Springer-Verlag Berlin Heidelberg, 2012.

- [158] Raymond Salvador, John Suckling, Martin R. Coleman, John D. Pickard, David Menon, and ED. Bullmore. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex*, 15(9):1332–1342, 2005.
- [159] Ernesto J. Sanz-Arigita, Menno M. Schoonheim, Jessica S. Damoiseaux, Serge ARB Rombouts, Erik Maris, Frederik Barkhof, Philip Scheltens, and Cornelis J. Stam. Loss of small-world networks in alzheimer’s disease: graph analysis of fmri resting-state functional connectivity. *PloS one*, 5, 2010.
- [160] Jeremy D. Schmahmann and Deepak N. Pandya. Disconnection syndromes of basal ganglia, thalamus, and cerebrocerebellar systems. *Cortex*, 44(8):1037–1066, 2008.
- [161] William W. Seeley, Richard K. Crawford, Juan Zhou, Bruce L. Miller, and Michael D. Greicius. Neurodegenerative diseases target large-scale human brain networks. *Neuron*, 62(1):42–52, April 2009.
- [162] William W. Seeley, Vinod Menon, Alan F. Schatzberg, Jennifer Keller, Gary H. Glover, Heather Kenna, Allan L. Reiss, and Michael D. Greicius. Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, 27(9):2349–2356, 2007.
- [163] Lei Shi, Qi Liao, Xiaohua Sun, Yarui Chen, and Chuang Lin. Scalable network traffic visualization using compressed graphs. In *Big Data, 2013 IEEE International Conference on*, pages 606–612. IEEE, 2013.
- [164] James M Shine, Oluwasanmi Koyejo, and Russell A Poldrack. Temporal metastates are associated with differential patterns of time-resolved connectivity, network topology, and attention. *Proceedings of the National Academy of Sciences*, page 201604898, 2016.
- [165] Hadi Shiravi, Ali Shiravi, and Ali A Ghorbani. A survey of visualization systems for network security. *IEEE Transactions on visualization and computer graphics*, 18(8):1313–1329, 2012.
- [166] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*, 22, 2012.
- [167] Stephen M. Smith, Peter T. Fox, Karla L. Miller, David C. Glahn, P. Mickle Fox, Clare E. Mackay, Nicola Filippini, Kate E. Watkins, Roberto Toro, Angela R. Laird, and Christian F. Beckmann. Correspondence of the brain’s functional architecture during activation and rest. *PNAS*, 106(31), 2009.
- [168] Robert R Sokal and F James Rohlf. The comparison of dendrograms by objective methods. *Taxon*, pages 33–40, 1962.
- [169] Olaf Sporns. From simple graphs to the connectome: networks in neuroimaging. *Neuroimage*, 62(2), 2012.

- [170] Olaf Sporns and Richard F. Betzel. Modular brain networks. *Annual review of psychology*, 67:613–640, 2016.
- [171] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS Comput Biol*, 1(4):e42, 2005.
- [172] CJ Stam, BF Jones, G Nolte, M Breakspear, and PH Scheltens. Small-world networks and functional connectivity in alzheimer’s disease. *Cerebral cortex*, 17(1):92–99, 2007.
- [173] Christian L Staudt and Henning Meyerhenke. Engineering parallel algorithms for community detection in massive networks. *IEEE Transactions on Parallel and Distributed Systems*, 27(1):171–184, 2016.
- [174] Kozo Sugiyama, Shojiro Tagawa, and Mitsuhiko Toda. Methods for visual understanding of hierarchical system structures. *IEEE Trans. on Systems, Man, and Cybernetics*, 2(11):109–125, 1981.
- [175] Jimeng Sun, Christos Faloutsos, Spiros Papadimitriou, and Philip S Yu. Graphscope: parameter-free mining of large time-evolving graphs. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 687–696. ACM, 2007.
- [176] Tetsuji Takada and Hideki Koike. Tudumi: Information visualization system for monitoring and auditing computer logs. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 570–576. IEEE, 2002.
- [177] Soon Tee Teoh, Kwan Liu Ma, S. Felix Wu, and Xiaoliang Zhao. Case study: Interactive visualization for internet security. In *Proceedings of the conference on Visualization’02*, pages 505–508. IEEE Computer Society, 2002.
- [178] Athanasios Theocharidis, Stjin Van Dongen, Anton J Enright, and Tom C Freeman. Network visualization and analysis of gene expression data using biolayout express3d. *Nature protocols*, 4(10):1535–1550, 2009.
- [179] Hannu Toivonen, Sébastien Mahler, and Fang Zhou. A framework for path-oriented network simplification. In *International Symposium on Intelligent Data Analysis*, pages 220–231. Springer, 2010.
- [180] Hannu Toivonen, Fang Zhou, Aleksi Hartikainen, and Atte Hinkka. Compression of weighted graphs. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 965–973. ACM, 2011.
- [181] Christian Tominski, Heidrun Schumann, Gennady Andrienko, and Natalia Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2565–2574, 2012.
- [182] Edward R. Tufte. Envisioning information. *Optometry & Vision Science*, 68(4):322–324, 1991.

- [183] John W Tukey. Exploratory data analysis. 1977.
- [184] Martijn P. van den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *The Journal of neuroscience*, 31(44):15775–15786, 2011.
- [185] Andre F. van Dixhoorn, Bastijn H. Vissers, Luca Ferrarini, Julien Milles, and Charl P. Botha. Visual analysis of integrated resting state functional brain connectivity and anatomy. In *Proceedings of the 2nd Eurographics conference on Visual Computing for Biology and Medicine*, pages 57–64. Eurographics Association, 2010.
- [186] David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy EJ. Behrens, Essa Yacoub, Kamil Ugurbil, WU-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.
- [187] Corinna Vehlow, Fabian Beck, Patrick Auwarter, and Daniel Weiskopf. Visualizing the evolution of communities in dynamic graphs. In *Computer Graphics Forum*, volume 34, pages 277–288. Wiley Online Library, 2015.
- [188] Corinna Vehlow, Fabian Beck, and Daniel Weiskopf. The state of the art in visualizing group structures in graphs. In *Eurographics Conference on Visualization (EuroVis)-STARS*, volume 2. The Eurographics Association, 2015.
- [189] Corinna Vehlow, Thomas Reinhardt, and Daniel Weiskopf. Visualizing fuzzy overlapping communities in networks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2013.
- [190] Christophe Viau and Michael J McGuffin. Connectedcharts: explicit visualization of relationships between data graphics. In *Computer Graphics Forum*, volume 31, pages 1285–1294. Wiley, 2012.
- [191] Tatiana Von Landesberger, Arjan Kuijper, Tobias Schreck, Jörn Kohlhammer, Jarke J. van Wijk, J-D. Fekete, and Dieter W. Fellner. Visual analysis of large graphs: state-of-the-art and future research challenges. In *Computer graphics forum*, volume 30, pages 1719–1749. Wiley Online Library, 2011.
- [192] Liang Wang, Chaozhe Zhu, Yong He, Yufeng Zang, QingJiu Cao, Han Zhang, Qiuhai Zhong, and Yufeng Wang. Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Human brain mapping*, 30(2):638–649, 2009.
- [193] David E Warren, Jonathan D Power, Joel Bruss, Natalie L Denburg, Eric J Waldron, Haoxin Sun, Steven E Petersen, and Daniel Tranel. Network measures predict neuropsychological outcome after brain injury. *Proc. of the National Academy of Sciences*, 111(39):14247–14252, 2014.
- [194] Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, 2(3):125–141, 2012.

- [195] Christopher Wilke, Gregory Worrell, and Bin He. Graph analysis of epileptogenic networks in human partial epilepsy. *Epilepsia*, 52(1):84–93, 2011.
- [196] Keith J. Worsley, Jen-I Chen, Jason Lerch, and Alan C. Evans. Comparing functional connectivity via thresholding correlations and singular value decomposition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):913–920, 2005.
- [197] Han-Ming Wu, Yin-Jing Tien, and Chun-houh Chen. Gap: A graphical environment for matrix visualization and cluster analysis. *Computational Statistics & Data Analysis*, 54(3):767–778, 2010.
- [198] Mingrui Xia, Jinhui Wang, and Yong He. Brainnet viewer: A network visualization tool for human brain connectomics. *PLoS One*, 2013.
- [199] Ling Xiao, John Gerth, and Pat Hanrahan. Enhancing visual analysis of network traffic using a knowledge representation. In *Visual Analytics Science And Technology, 2006 IEEE Symposium On*, pages 107–114. IEEE, 2006.
- [200] Ji Soo Yi, Youn ah Kang, John Stasko, and Julie Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics*, 13(6):1224–1231, 2007.
- [201] Xiaoxin Yin, William Yurcik, Michael Treaster, Yifan Li, and Kiran Lakkaraju. Visflow-connect: netflow visualizations of link relationships for security situational awareness. In *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pages 26–34. ACM, 2004.
- [202] Andrew Zalesky, Alex Fornito, Luca Cocchi, Leonardo L Gollo, and Michael Breakspear. Time-resolved resting-state brain networks. *Proc. of the National Academy of Sciences*, 111(28):10341–10346, 2014.
- [203] Loutfouz Zaman, Ashish Kalra, and Wolfgang Stuerzlinger. The effect of animation, dual view, difference layers, and relative re-layout in hierarchical diagram differencing. In *Proceedings of Graphics Interface 2011*, pages 183–190. Canadian Human-Computer Communications Society, 2011.