Visual Analysis Techniques for Dynamic Biological Networks

by

Chihua Ma
B.S., Nankai University, China, 2009
M.S., University of Illinois at Chicago, Chicago, 2011

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:
G. Elisabeta Marai, Chair and Co-Advisor
Robert Kenyon, Advisor
Andrew Johnson
Angus Forbes, University of California, Santa Cruz
Tamara Munzner, University of British Columbia

ACKNOWLEDGMENTS

I would first like to thank my advisors Dr. Robert Kenyon and Dr. G. Elisabeta Marai for their guidance and support. Dr. Robert Kenyon always offers support and wisdom through the trials and tribulations of a Ph.D. career. Dr. G. Elisabeta Marai is a source of great motivation and experience, guiding projects to incomparable success. I am very lucky to have such great mentors in my life. I would also like to thank my other committee members Dr. Andy Johnson, Dr. Angus Forbes and Dr. Tamara Munzner for their valuable feedback on my work.

I gratefully acknowledge my collaborators Dr. Tanya Berger-Wolf, Dr. Daniel Llano, Dr. Jie Liang and Anna Terebus. I wouldn't be able to accomplish my work without being able to work on these interesting problems with you.

I would like to thank my labmates at the Electronic Visualization Laboratory at UIC. I appreciate the time that Andrew Burks and Timothy Luciani spent to read parts of my thesis and providing useful feedback. I thank Andrew Burks, Timothy Luciani and Filippo Pellolio for their assistance in developing the visualization applications. I also thank all the EVL faculty, staff and students.

Thanks to my family for their selfless love and support all the time. My parents Jinyuan Ma and Qian Mu give me a home full of love and will always be my harbor no matter where I go. I thank my husband Andrew Burks for understanding and supporting me. Thank you for coming into my life.

CM

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC   Auditory cortex

ACME   Accurate Chemical Master Equation

AD   Alzheimer's disease

CME   Chemical Master Equation

dCME   Discrete Chemical Master Equation

fMRI   Functional Magnetic Resonance Imaging

ICA   Independent Component Analysis

mRNA   messenger RNA

PCA   Principal Component Analysis

PCC   Pearson correlation coefficient

ROIs   Regions of interest

SNA   Social Network Analysis

SUMMARY

Recent neuroscience research indicates that cognitive operations are performed not by individual brain regions working in isolation, but by networks that consist of several discrete brain regions. These discrete brain regions act in synchrony and share functional connectivity. The study of brain networks contains both non-spatial data dynamic functional connections within the brain and spatial data 2D/3D medical scans of the brain. To further understand the relationships between dynamic functional connectivities and physical structures of the brain, researchers analyze dynamic brain networks over multiple experiments or with different models. Similarly, bioinformatics research also contains problems about dynamic biological networks that require the integration of spatial and nonspatial data and multi-run comparison. Due to the complexity of biological data that can be collected, modeled and analyzed thanks to technological and algorithmic advancements, the difficulty in studying biological networks has increased over the last decades. Visualization provides an efficient way to help biologists understand, communicate, and gain insight into their biological data through visual exploration and analysis.

In this dissertation, I propose a set of visual approaches for the analysis of ensemble dynamic biological networks. This set of approaches took shape through the development of multiple prototype visual analytics tools which were aimed to solve complex problems in multiple biological subdomains. The steering of the approach was supported through an in-depth understanding of both the encompassing biological domains and problem

SUMMARY (Continued)

space, as well through a survey of existing visual approaches helping to generalize the

paradigms used when confronting these problems and the complexity of their data.

CHAPTER 1

INTRODUCTION

1.1 Motivation

In the United States, an estimated 5.5 million adults of all ages have Alzheimer's disease

(AD) (1). Unfortunately, there is no single test to diagnose Alzheimer's disease easily,

and little is known about why AD happens. To better understand the causes and

effects of AD, researchers work from what is currently understood about the functionality

of the brain. Recent neuroscience research (2) indicates that cognitive operations are

performed not by individual brain regions working in isolation, but by networks that

consist of several discrete brain regions. These discrete brain regions act in synchrony and

share functional connectivity. Furthermore, research shows evidence that Alzheimer's

disease does not spread from one region to an adjacent one, but rather to a functionally

connected brain region (3). Better understanding functional connectivities within the

brain can, in turn, allow researchers to better diagnose and understand the progression

of Alzheimer's disease and other similar neurological diseases.

The analysis of brain networks is difficult due to the nature of the data and tasks

involved. For example, this analysis requires the integration of functional connections

and brain structures and the comparison of multiple subjects. The functional connec-

tivities between brain regions can change rapidly over time (4; 5), which gives brain

networks a highly dynamic characteristic. The study of brain networks contains both

spatial data — 2D/3D medical scans of the brain — and non-spatial data — dynamic functional connections within the brain. In broader terms, "spatial data" denotes the datasets which specify the position or shape of items — the field and geometry dataset types (6), while "non-spatial data" denotes abstract data in which no spatial position is provided — the table, tree and network dataset types (6). To further understand the relationships between dynamic functional connectivities and physical structures of the brain, researchers examine multiple patients to find similarities and differences in behavior. A recent study (7) shows differential functional connectivity and structural changes in the brains of AD patients with and without cerebrovascular disease. The comparison of AD patients with different conditions reveals the importance of comparing multiple subjects and motivates the modeling of dynamic biological networks over multiple experiments/simulations or with different models. As a result, effective methods that provide both quantitative and qualitative analysis to compare those datasets are necessary, to benefit the exploration of these similarities and differences.

Similarly, bioinformatics research also contains problems about dynamic biological networks that require the integration of spatial and nonspatial data and multi-run comparison. Using the knowledge of molecular structures, we can better understand the temporal and spatial aspects of protein interaction networks. Multiple simulations of stochastic gene networks produce ensemble time-evolving probability landscapes of such networks, which can further biologists' understanding of phenotypic behavior associated with specific genes.

Due to the complexity of biological data that can be collected, modeled and analyzed thanks to technological and algorithmic advancements, the difficulty in studying biological networks has increased over the last decades. These biological data may also contain answers to questions that we do not yet know how to ask. This wealth of data, features, models, and potential hypotheses and experiments exceeds the analytical capabilities of machines. Visualization provides an efficient way to help biologists understand, communicate, and gain insight into their biological data through visual exploration and analysis.

1.2   Visual Computing Challenges

Although a number of visualization tools for biological networks have been developed during the last decades, the theory and design principles behind their development are not clear — in particular with respect to the challenges of visualizing multi-run dynamic biological networks. For example, the challenge of spatial and non-spatial visual integration is common across domains, from geospatial applications to bioinformatics to neuroscience applications (8). In the visualization field, spatial data intrinsically specify the location of each item in 2D or 3D space — for example a medical scan of a human body, while nonspatial data specify abstract information without spatial positions provided — for example tables and networks. However, in contrast to geospatial analysis, where the spatial features (maps) are familiar to users and overlays make sense, bioinformatics and neuroscience users may lack the domain knowledge to understand spatial data — a computational geneticist may not be familiar with molecular structures. Unfortunately, no taxonomy of visual integration of spatial and nonspatial data in biology

exists, to guide researchers with respect to the visualization top-design for particular problems.

Another challenge of visualizing complex biological networks is creating effective combinations of visual representations for multi-scale exploration. As biological data become larger and more complex, the data may be decomposed into a hierarchical data structure which requires multi-scale visualization approaches to support domain experts in the exploration of such large biological data at several levels of abstraction. For example, neuroscientists may explore brain networks at three levels: a high level to get the overall network structure, a middle level to invest particular subregions or communities of the network, and a low level to inspect individual nodes or elements. The features that are interesting to domain experts may vary at different levels. Thus, the visual representations need to display the level of detail appropriate to a particular scale (9). In order to explore detailed levels while still keeping their contexts in overviews, many multi-scale approaches need the efficient use of the screen space, especially when the biological problem features both spatial and non-spatial characteristics.

Supporting comparison across multiple models, datasets, and experiments/simulations is another open problem. Gleicher et al. (10) propose a survey from information visualization domains that presents three visualization solutions for visual comparisons: juxtaposition (side-by-side views or small multiples), superposition (overlays), and explicit encodings for visualizing the relationships between objects directly. Based on this survey, Gleicher (11) also proposes an abstract framework to aid designers by providing the understandings of the comparison problems and strategies to support visual

comparison. According to this framework, comparing multiple objects (multi-run comparison) is more difficult than comparing two objects (pairwise comparison), because neither juxtaposition nor superposition solutions naturally scale to a large number of items (11). A juxtaposition layout separates those items and occupies more space, and thus may be a poor option for biological problems that also contain spatial features in which the spatialization is a crucial component. Overlaying and explicit encodings also increase visual complexity with many objects. Thus, the strategies need to be chosen carefully according to the specific domain problems. The Gleicher framework (11) proposes a Summarize Somehow strategy that builds abstractions of the massive dataset by measuring and quantifying differences through computational and statistical analysis. Taking multi-scale visual exploration into account, the comparison can also be applied at multiple levels of abstraction. Providing detail for multi-scale comparison while preserving the context of these complex datasets is even more challenging. Thus, the selection of visual strategies for comparison may vary with different scales.

Overall, an even more challenging visual computing problem is raised when combining all of three visual challenges discussed above (Figure 1), because spatial-nonspatial visual integration, multi-scale and comparative visual representations all involve arranging visual elements in space and thus require more effective visual designs. Novel visual representations and a good visual strategy to combine solutions for each challenge need to be developed for dynamic biological networks that require spatial-nonspatial visual integration, multi-scale visual exploration, and multi-run visual comparison.
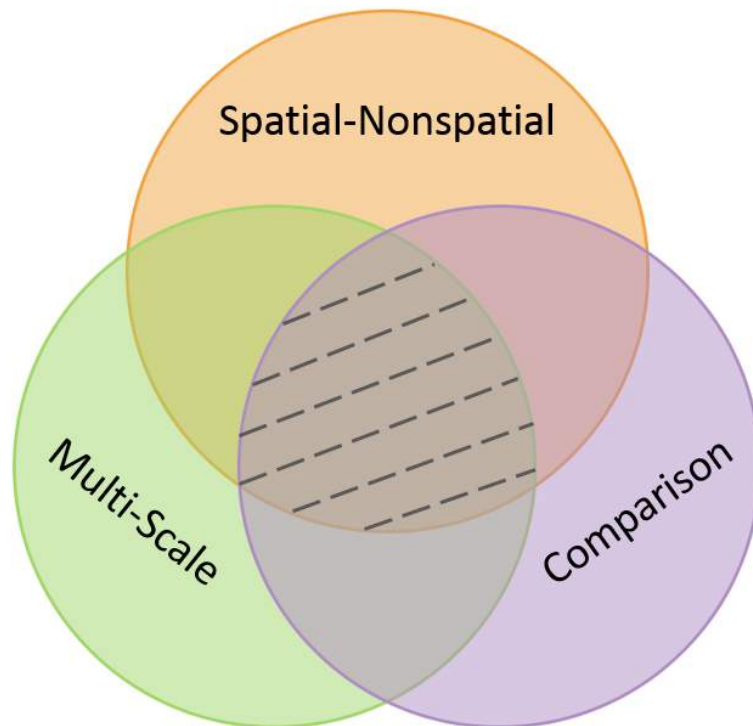
Figure 1. The combination of three challenges belonging to biological networks analysis forms a challenging visual computing problem.

1.3   Contributions Overview

This dissertation makes the following contributions: 1) a survey of spatial-nonspatial integration in biology, 2) a description of the domain data and tasks in dynamic biological networks analyses, 3) several novel visual representations, 4) a visual approach to compare multiple dynamic biological networks at multiple scales, and 5) five visualization applications developed for two biological networks. The principles extracted from this

visual approach can also be applied to other domains, i.e., dynamic multi-run networks with spatial features, especially when those networks can be grouped into a hierarchical structure.

The first contribution of this work, a survey of visual integration of spatial-nonspatial data in biological domain presented in the next chapter, reflects the state of the art and provides a basis to better design visual analysis tools for dynamic biological networks.

The second contribution of this work, a description of domain problems, provides a detailed explanation of the challenges in visualizing dynamic biological networks including spatial-nonspatial visual integration, multi-scale and comparative visual representations. This analysis may benefit researchers who are interested in or currently working on similar problems.

The third contribution of this work, a set of innovative visual designs, enables the visual integration of spatial-nonspatial of dynamic biological networks. Since spatial-nonspatial visual integration is an open research topic and the state of the art does not provide sufficient solutions to the visual analysis of biological networks, new visual designs are necessary. Additionally, those visual designs of spatial-nonspatial integration combined with multi-scale and comparative visual approaches support visual exploration and comparison at multiple scales.

The fourth contribution of this work, a visual approach, supports the visual exploration and comparison at multiple scales. The multi-scale comparative visual approach offers the ability to compare multiple objects at an abstract level through their summarized statistical characteristics and provide detailed pairwise comparisons based on the

user's demand. This visual approach applies to domain problems that contain multiple datasets, including but not limited to dynamic biological networks.

The last contribution of this work, five visualization applications performed in collaboration with domain experts, support two dynamic biological network problems that contain the combination of the challenges of spatial-nonspatial visual integration, multiscale visual exploration, and multi-run visual comparison. These applications were developed to solve complex domain-specific problems in biology.

1.4   Thesis Organization

This dissertation starts with the motivation of visualizing dynamic biological networks along with the three visual computing challenges, followed by a summary of contributions. In order to better design visual analysis tools for dynamic biological networks, we conduct a survey in Chapter 2 to better understand the state of the art in visual integration of spatial-nonspatial biological data. Through the survey, we have not found dynamic biological networks that contain all three visual computing challenges. However, we find that such problems do exist during our collaboration with neuroscientists and biologists.

The following three chapters present five visualization applications consisting of three main tools and two extensions to support two biological network analyses that require the spatial-nonspatial visual integration, multi-scale visual exploration and multi-run visual comparison. Chapter 3 introduces PRODIGEN (12) and its extension — EnsembleProb (13) to visualize the ensemble time-evolving probability landscapes of stochastic gene regulatory networks. This work was developed with Timothy Luciani, Andrew

Burks and G. Elisabeta Marai from Electronic Visualization Laboratory (EVL) at University of Illinois at Chicago (UIC), in collaboration with Anna Terebus and Jie Liang from Bioengineering at UIC.

In addition to stochastic networks in bioinformatics, problems in neuroscience also have the three visualization challenges discussed above. Chapter 4 presents Sword-Plots (14) to explore neuron behavior within dynamic communities of brain networks. This work was designed with Angus Forbes from University of California, Santa Cruz and Robert Kenyon from EVL at UIC, in collaboration with Daniel Llano from Neuroscience at University of Illinois at Urbana-Champaign (UIUC) and Tanya Berger-Wolf from Computer Science at UIC. This work integrates the dynamic behaviors of neurons within their spatial locations in the brain slice using the linked views paradigm. However, the SwordPlots is developed for the visual exploration of a single dynamic brain activity networks, which is not sufficient for comparing multiple networks as our domain experts run multiple experiments with different subjects.

As a follow-up of SwordPlots, chapter 5 discusses RemBrain (15) that enables the visual exploration and pairwise comparison of dynamic brain networks at multiple scales. This work was implemented with Filippo Pellolio from Here Inc, Robert Kenyon, and G. Elisabeta Marai from EVL at UIC. RemBrain 2.0, as the extension of RemBrain, is introduced in Chapter 6 with a focus on visual comparison of multiple dynamic brain networks. This work was implemented with Andrew Burks, Robert Kenyon, and G. Elisabeta Marai from EVL at UIC. Both RemBrain and its extension were developed in collaboration with Daniel Llano from UIUC and Tanya Berger-Wolf from UIC.

Chapter 7 concludes the dissertation and discusses the limitations and assumptions behind this work and future work.

CHAPTER 2

A SURVEY OF VISUAL INTEGRATION OF SPATIAL AND NONSPATIAL DATA IN BIOLOGY

This chapter introduces a brief survey of a set of spatial-nonspatial integration papers in biology visualization literature. We selected 70 relevant papers, mainly published in the proceedings of the visualization conferences and their associated journals. The main conferences and journals are The EG/VGTC Conference on Visualization (Euro-Vis) and Computer Graphics Forum; The IEEE VIS Conference, IEEE Transactions of Visualization and Computer Graphics; and The International Symposium on Biological Data Visualization (BioVis), BMC Bioinformatics, BMC Proceedings, and the Journal of Bioinformatics. These papers span the subdomains of bioinformatics/biochemistry, biomechanics, biomedical, medical image analysis, neuroscience, and epidemiology. In order to maintain focus and avoid an overly ambitious survey, we restrict this chapter to a survey of neuroscience and bioinformatics works which integrate spatial and nonspatial features. This subset contains 16 papers and was selected based on the domain problems in our collaborations with domain experts.

2.1 Neuroscience

In neuroscience, the study of relationships between the function of the nervous system and its structures has become a promising research area. The datasets in this research

11

area often take the form of volumes combined with tables and/or networks. Linked views or a hybrid paradigm are commonly used to integrate these multiple types of datasets.

Early neuroscience studies dealt with data of macroscopic scale which had a time dimension, such as EEG and fMRI. Ten Caat et al. (16) develop an overlay solution to visualizing high-density EEG coherence. A functional unit map displays a color-mapped spatially connected set of electrodes, which is overlaid with the functional connections between functional units drawn as lines. The main task is to present known electrophysiological findings. Domain experts were not involved in the visual design or evaluations. Janoos et al. (17) also propose a hybrid visualization of fMRI data, aimed at exploring temporal relationships in brain function. The visualization includes the main view showing 3D and 2D orthographic views of the structural volume, overlaid with the clusters, and a linked cluster time-series display. In the reported case studies the users seek to identify the effect of certain factors on brain function and to discover the correct parameters for the analysis through comparison. Nowke et al. (18) present a system for the exploration of neural activity data, which supports inspecting connectivity of brain regions and populations. The application uses a hybrid (overlay and linked-view) approach to integrate a control-view which overlays area-activity data on the three-dimensional brain geometry, a population view, a connectivity view, and a flux view in order to visualize dynamic activity exchange between brain areas. The visual encodings include node-link diagrams, bar charts, small multiples, matrices, and animation, and the tasks include discover, annotate, explore, and query. The system was positively evaluated by domain experts, via feedback as well as through case studies.

Spatial-nonspatial integration using more sophisticated visual encodings was further prompted by the advent of connectomics – the study of the connections within an organism's nervous system, typically its brain. Chen et al. (19) introduce a linked views approach for Diffusion Tensor Imaging (DTI) tract clustering. The views link 3D geometry representations of the fibers with scatter plot embeddings to a lower dimensional space and histograms. The users are two neuropsychologists and one cardiologist, who participate in the system evaluation; mainly task success is reported. The tasks are of the Explore, Compare and Summarize type. Jianu et al. (20) pursue a similar linked views approach for DTI tract clustering. The views link 3D geometry representations of the fibers with scatter plots in a lower dimensional space and dendrograms (tree representations of hierarchical clusterings). The authors are the users, and no expert feedback is reported. The tasks are of the Locate, Explore, Compare and Summarize type.

Advances in measuring and computing devices enable neuroscientists and epidemiologists nowadays to collect imaging data at the microscopic scale with higher spatial resolution. Beyer et al. (21) present a linked-views application called ConnectomeExplorer for the query-guided visual analysis and interactive exploration of large volumetric electron microscopy (EM) data. The visualization comprises a 3D volume view for EM, a topological 2D graph view for neuronal connectivity, a slice view for manual segmentation, and other statistics views for attributes and information in detail using scatter plots, histograms, and tables. The case studies discuss several tasks consisting of locating, identifying and comparing synapses, and in general discovering new patterns. Domain experts provide positive feedback. Al-Awami et al. expanded the Connec-

tomeExplorer work into a visual analysis of neurons and their connections with two visualization systems: NeuroLines (22) and NeuroBlocks (23). NeuroLines (22) links a subway map-inspired 2D representation for connections with 2D and 3D volume views of the original and segmented EM data. The main tasks are to explore and identify connectivity, patterns, and trends. Other minor tasks in the two case studies are to discover, browse, compare, and summarize. NeuroBlocks (23) focuses on tracking and managing large volumetric segmentation data, in the context of multiple users with similar expertise. The system comprises multiple linked views consisting of a scalable "pixel-based" visualization of the current segmentation state, an area chart for the timeline, 2D and 3D views of volume rendering, and other views for details-on-demand, such as a node-link diagram. The main task is the exploration of neuronal connectivity, the evolution of segmentation, and overall status. The entire query tasks are also included. Two case studies were used for demonstrating the utility of NeuroBlocks. We notice an increase in the complexity of tasks/visual encodings with the scale of neural data.

2.2 Bioinformatics

Similar to neuroscience research, bioinformatics is a fertile ground for spatial-nonspatial visual integration. In the study of gene expressions, researchers explore gene expression over time in conjunction with the spatial location of the cells where the genes are expressed; they also compare gene expression data across multiple related species. To compare spatial gene expression patterns, Rubel et al. (24) introduce PointCloudXplore, which analyzes 3D gene expression patterns in early Drosophila embryos. The visualization links parallel coordinates (which encode features of gene expression) and a spatial,

physical view showing the relationships among expression patterns. Colormaps overlaid on the physical views represent the gene expression values. No domain experts were involved in the visual design or provided feedback. Meyer et al. (25) introduce MulteeSum, a hybrid, nesting and linked views paradigm example. Their solution includes an Embryo Map to show the cells' spatial information (summary values are encoded by color) and a Curvemap of a small-multiple of filled line charts. Three case studies were successfully conducted with domain experts to demonstrate the efficacy of MulteeSum.

In structural biology, researchers correlate genetic mutations with the impaired function of a protein with known 3D structure. In 2013, BioVis organized a data contest focusing on identifying protein mutations in order to discover their effect on protein function. Among many contest submissions, Mercer et al. (26), Doncheva et al. (27), and Luciani et al. (28) received awards for proposing viable solutions which were endorsed by the domain experts who had posed the problem. Mercer et al. (26) introduce Mu-8 which uses multiple linked views, including score histograms, proximity chords, context sequence, conservation heat map, and 3D structure view. Doncheva et al. (27) also present a visual analysis tool which integrates a 3D protein structure view with network representations and a sequence visualization. Finally, Luciani et al. (28) introduce FixingTIM, which consists of a 3D view of protein structures, a sequence view expression, and a trend image panel for aggregating protein families. All these three solutions support the tasks of identification and comparison, and all use a case study with data from the 2013 BioVis contest to demonstrate their value. In addition to the

case study, Luciani et al. also collected feedback from domain experts; their solution has been adopted as a teaching tool.

Motivated by a similar problem, Lenz et al. (29) present a linked views visual system which includes a global overview using heatmaps and a view of the locations of found patterns in both mutation graphs (node-link diagrams) and 3D protein structures. User tasks include: explore and identify frequent patterns in mutation graphs, and locate and compare the found patterns. Biologists used the tool to analyze real-world data. To analyze and explore protein structures, sequences, and features, Stolte et al. present Aquaria (30) which also uses multiple linked views. Their approach is a combination of a 3D structure panel and a sequence-based visualization of matching structures and features. The domain experts (biologists) were involved in the design and conducted two case studies.

In summary, we notice from the survey that linked views is still the most commonly used paradigm for integrating spatial and non-spatial data in in the field of biology. The hybrid approach, which combines two or more paradigms, has been widely used as well. The use of overlays alone for spatial-nonspatial integration increases the visual complexity, and thus may not be an optimal solution. Additionally, some works in the survey use a newer approach—nesting paradigms, in which spatial representations are embedded within non-spatial representations (non-spatial nesting), and vice-versa (spatial nesting).

In addition to the challenge of integrating spatial and non-spatial features in biological problems, some problems in this survey also face the challenge of multi-scale visual exploration, e.g., NeuroBlocks (23). Some problems require comparative visualization as well, e.g., MulteeSum (25). Although MizBee (31) is not included in the survey since the problem does not contain spatial features, MizBee is a multi-scale comparative visualization. Through this survey, we have found visual approaches that combine any two challenges that discussed in the Introduction Chapter, but not all three of them. The biological problems introduced in the following chapters demonstrate that the problems facing all three challenges do exist, but there are few visual approaches to solve those problems that combine all of them in the field of visualizing biological networks. In the next three chapters, we will introduce two types of biological networks — one in bioinformatics and one in neuroscience, present five visualization applications for these two networks along with several novel visual designs, and propose an visual approach to explore and compare those networks at multiple scales.

CHAPTER 3

PRODIGEN: VISUALIZING THE PROBABILITY LANDSCAPE OF STOCHASTIC GENE REGULATORY NETWORKS IN STATE AND TIME SPACE
(PREVIOUSLY PUBLISHED AS CHIHUA MA, TIMOTHY LUCIANI, ANNA TEREBUS, JIE LIANG, AND G. ELISABETA MARAI. PRODIGEN: VISUALIZING THE PROBABILITY LANDSCAPE OF STOCHASTIC GENE REGULATORY NETWORKS IN STATE AND TIME SPACE. (2017) BMC BIOINFORMATICS, 18(2): 24 AND CHIHUA MA, ANDREW BURKS, TIMOTHY LUCIANI, ANNA TEREBUS, JIE LIANG, AND G. ELISABETA MARAI. VISUALIZING ENSEMBLE TIME-EVOLVING PROBABILITY LANDSCAPES OF STOCHASTIC NETWORKS. ISMB/ECCB 2017 (POSTER))

The first type of biological networks we will discuss are stochastic gene regulatory networks in bioinformatics. During our collaboration with biologists, the importance and difficulty in understanding stochastic gene regulatory networks became apparent. Enabling the analysis of these ensemble probability distributions in both time and space can help scientists better understand the dynamics and stochasticity of gene regulatory processes in nature.

Unlike protein-protein interaction networks, which feature spatial and non-spatial information based in the domain datasets, the research of stochastic gene regulatory networks features spatial and non-spatial information based on domain specific tasks. Biologists study the stochasticity of gene regulatory networks described by probability distributions in space which is defined as a landscape with "hills and valleys" by biologists. They describe the probability distributions in terms of this spatial information much in the same way that a geospatial analyst describes their data in terms of latitude, longitude, height or depth. To further understand the dynamics and stochastics of gene networks, biologists run multiple simulations with stochastic parameters; this approach needs effective visual designs to support ensemble comparison and analysis.

In this chapter, we introduce the domain problem of gene regulatory networks and stochastic network modeling, followed by task analysis. There is a lot of research in stochastic networks in biology, however, we did not find any related problems in our survey of the integration of spatial and non-spatial biological data. Studying the dynamics and stochasticity of such networks is an open research topic, and little work in the state of the art provides suggestions for visually analyzing such a stochastic network. Thus, we propose a novel visual analysis approach and evaluate it through two case studies and domain expert feedback.

3.1 Introduction

Gene regulatory networks encode those interactions among genes and proteins that regulate cellular processes, such as the expression of messenger RNA (mRNA). These interactions dictate the expression levels of genes as well as the production of particular

proteins, and thus play a critical role in regulating biological functions, from metabolism to cell differentiation. Such networks typically involve small copy numbers of the molecular species and large differences in the species reaction rates. Because of these factors, the network interactions have a stochastic nature (32; 33; 34), i.e. they are unpredictable through the influence of a random variable. Modeling the stochasticity of genetic circuits is an important field of research in systems biology, and can help elucidate the mechanisms of cell behavior, which in turn can be the basis of diseases. These models can further enable predictions of important phenotypic cellular states.

The computational study of stochastic properties in gene networks is, however, a challenging task. Ordinary and stochastic differential equations methods may be inadequate in computing accurately the dynamic and steady state probabilistic behavior of gene networks (35), while the Gillespie stochastic simulation algorithm (36) is exact, but can fail in capturing rare events (37; 38; 39). However, the formulation of the discrete Chemical Master Equation (CME) can be used to analyze the dynamics and stochastic nature of gene regulatory networks with low copy number of species. The CME framework can allow, for example, for the dynamics and stochasticity of a network with small copy numbers of molecular species to be fully described by probability distributions in both the state space and time space.

However, the analysis of these probability distributions is difficult due to their spatiotemporal and multidimensional nature, and due to the typically large number of simulations run under varying settings. Moreover, stochastic network researchers often emphasize that what is of biological significance is often not of statistical significance —

numerical analyses often miss small or rare events of particular biological relevance. A visual approach can help, in contrast, in mining the network dynamics through the landscape defined by these probability distributions. For example, visualizing the number, location, and behavior of probability peaks could indicate the number of stable states for a given network and set of model parameters. We note that the state of the art reports in stochastic network modeling only employ visualization post facto. Our collaborators indicate that once the researchers know what they are looking for, they typically use a plotting software like R or Matlab to generate explanatory projection images and animations of the peaks. However, no visual tools exist to support the exploratory analysis of simulated data.

In this chapter, we introduce a web-based visual analysis tool for the exploration of time-varying probability landscapes over the state space in stochastic networks. PRODIGEN (PRObability DIstribution of GEne Networks) supports the exploration of probability distributions in both state and time space. PRODIGEN captures probability distributions with projections at multiple levels, such as spaghetti plots for one dimensional projection and enhanced heatmaps for two dimensional projection. The main contributions of this work are as follows:

- We provide a description of the domain data and tasks in stochastic biological network modeling and analysis.

- We propose several visual encodings to represent the probability landscape in multiple dimensions, including one dimension, two dimensions, and three dimensions.

- We implement an interactive web-based visual explorer, PRODIGEN, that combines these visual encodings to enable the exploration of probability distributions across both state and time.

- We evaluate the visualization system through case studies and report a summary of the feedback provided by domain experts.

3.2 Related Work

3.2.1 Stochastic network modeling

The discrete chemical master equation (dCME) provides a fundamental framework for studying stochasticity in molecular-level networks. Because of the multi-scale nature of many networks, directly solving dCMEs is intractable due to the exploding size of the state space. To address this limitation, the ACME (Accurate Chemical Master Equation) algorithm (40; 41) was introduced as an optimal algorithm for the exhaustive enumeration of discrete microstates. The algorithm is based on the decomposition of stochastic reaction network into multiple independent components called buffer queues, each governed by its own birth-death process. The approach has the advantage of more effective usage of the overall finite state space, rapid estimation of errors, and estimation of required buffer size in order to maintain pre-defined error tolerance.

Several systems have been tested with the ACME method, which we also employ in this work. The steady state and time-evolving behavior has been previously computed for several biologically important networks, including the genetic toggle switch model, the phage-lambda epigenetic circuit, and the 16-node MAPK cascade (42; 40; 41). The

same method has been used for modeling important stochastic network motifs such as Single Input and Coupled Toggle Switch Modules (43).

3.2.2   Probability distribution visualization

Currently, there are only a few systematic ways of visualizing probability distributions at every location and time. The distribution data in 2D space is typically encoded by a 2D color map. Kao et al. (44; 45) present a number of methods for visualizing 2D probability distributions. They color a scalar field to provide an overall impression of distribution data sets over a 2D spatial domain. Luo et al. (46) extend existing visualization methods, such as pseudocolor, to support such distribution data. Potter et al. (47) present a visualization system called ProbVis for exploring differences between distributions across a spatial domain. They also use a color map to encode the distance measure. 3D projections suffer however from overlap problems. To circumvent this problem, in this work we color-encode probability values on a 2D map, as height.

3.2.3   Heatmaps in biostatistics

Heatmaps have been traditionally used to display statistical data, from gene expression and metabolomics to urban and network evolution data (48; 49; 50; 51). In general, heatmaps are used to describe the variables which can be considered as a function of two inputs represented by the rows and columns. We use similar heatmaps representations, in which both rows and columns represent the location in each dimension, and the cell shows the probability value at a particular location in two dimensions.

3.2.4   Spaghetti plots in ensemble visualization

The computation process used in this biology problem generates spatial distribution datasets across multiple time steps. By extension, these datasets can be regarded as ensemble data — a collection of multiple related but different datasets, such as simulations in general. Several techniques have been proposed for ensemble visualization. Spaghetti plots, overlaying plots of individual ensemble members, are a well-known technique for visualizing system flows, including flows in biology, medicine, and meteorology. Obermaier and Joy (52) state that spaghetti plots, as a feature-based visualization, provide an overall impression of the whole ensembles and allow comparisons between ensemble members. Luo et al. (46) present the use of spaghetti plots in meteorology. Potter et al. present Ensemble-Vis (53), a framework to support the visual analysis of ensemble space data. Sanyal et al. propose Noodles (54), a coordinated-view visualization tool, for visualizing weather ensemble uncertainty. In addition to novel uncertainty visualization techniques, they implement spaghetti plots for observing the model uncertainty. Wu and Zhang (55) introduce spaghetti plots for visualizing ensemble uncertainty. Ferstl et al. (56) present a new approach that extends spaghetti plots to extracted ensemble flows. Similar to these works, we adopt spaghetti plots, but in a new context.

3.3   Data and Task Analysis

3.3.1   Stochastic Network Modeling

The study of stochastic gene regulatory networks is a challenging topic, as the system modeled may be large in both the state space and time. Recent developments of the ACME method (40; 41) enable reduction of the state space from $O(b^n)$ to $(\prod_j \binom{b+n_j}{n_j})$,

$\sum_{j=1}^{m} n_j = n$, where n is the number of species, m is the number of Molecular Equivalence Groups (MEGs)— the number of molecular species subgroups in the network, such that member species of a subgroup can be transformed into each other through one or more mass-balanced reactions, n_j is the number of species belonging to group MEG_j, and b is the maximum number of molecules in MEG_j. A buffer of size b is assigned to each of the MEGs involved in the model. These developments allow optimal enumeration of the state space for the system.

We briefly discuss the state space over which we visualize the probability landscape. Assume a system contains n molecular species x_i, and m buffers of size $C(x_j)$. The ACME method assigns a buffer of size $C(x_j)$ to MEG_j. The state space size is then equal to the product of m combinations of $\binom{C(x_1)+n_1}{n_1} \times \binom{C(x_2)+n_2}{n_2} \times \cdots \times \binom{C(x_m)+n_m}{n_m}$, where $\times$ denotes a simple multiplication.

A *microstate* s_k of the system, where $k \in (1, N)$, is defined by the combination of *numbers* of molecules (copy numbers) of every *species* $(n_k(x_1),\ n_k(x_2),\ n_k(x_n))$. The *state space* is formed by all the possible microstates that the system can visit from a given initial condition. The *state space table* (Figure 2) displays a microstate as a single row, for a total of N microstates.

Each row of the *probability matrix* shown in Figure 3 contains the probability values of the corresponding state (row) in the state space table (Figure 2) over time. P_{ij} represents the probability value of state s_i at time t_j, and P_{ij} is a float value between 0 and 1. The sum of the probabilities of all the microstates at any particular time is equal to 1.

		Molecule Species			
	Copy Number	x_1	x_2	...	x_n
States	s_1	$n_1(x_1)$	$n_1(x_2)$	...	$n_1(x_n)$
	s_2	$n_2(x_1)$	$n_2(x_2)$	...	$n_2(x_n)$
	...	...	...	...	...
	s_N	$n_N(x_1)$	$n_N(x_2)$	...	$n_N(x_n)$

Figure 2. The state space of a system with n molecular species x_j and N microstates s_i. $n_i(x_j)$ denotes the copy number of molecular species x_j at state s_i.

The combination of data stored in tables shown in Figure 2 and Figure 3 represents one *simulation* of a system that consists of m molecular species with the state space size of N across T time steps.

In this multidimensional probability space, the biology researchers are interested in identifying both *global and local probability peaks*. Global peaks are in general easily detected. However, local peaks are difficult to notice, when they exist, because they have low probability values. The number of peaks and their probability values can be computed analytically; however, analyzing the locations of peaks requires visualization.

The domain experts have been working with systems that may contain at most 16 molecular species, but over 1 million microstates across thousands of time steps. A system may feature tens of parameter settings, which in turn produce tens of simulations. In this work we use two stochastic network models: a toggle switch system and a transcription regulation network. The toggle switch system features two genes, each

	t_1	t_2	...	t_T
s_1	p_{11}	P_{12}	...	p_{1T}
s_2	P_{21}	P_{22}	...	p_{2T}
...	...	...	...	...
s_N	p_{N1}	p_{N2}	...	p_{NT}
sum	1	1	1	1

Figure 3. The probability matrix displays the probability distributions over N microstates across T time steps.

expressing one protein, with 120 respectively 240 maximum copy number. The transcription network contains three genes which express each one protein. Both proteins and genes (DNA molecules) are reactants in the reaction systems of biomolecules.

3.3.2 Task Analysis

Through multiple interviews and observation sessions with the domain experts, we identified a set of four task groups related to the exploration of the time-varying probability landscape in a stochastic system. Given a simulation run of a stochastic gene regulatory network, the identified tasks are as follows:

- T1: Display the overall plots of probability distributions at multiple dimensional levels: for example, for each molecular species in one dimensional projection and for the combination of any two molecular species in two dimensional projection.

- T2: Discover peaks in multiple dimensional projections. Identify the number of peaks and their probability values. Observe and inspect *small, local* peaks.

- T3: Identify the locations of peaks at multiple dimensional levels. The "location" denotes a collection of corresponding states in the state space.

- T4: Track temporal changes of the system. Observe the probability landscape changes over time, including the number of peaks and their locations.

Based on the visual data analysis taxonomy (6), we match these task groups with four taxonomy categories: Present: T1, T4, Discover: T2, T3, Explore: T2, T4, and Identify: T2, T3. In addition to the functional requirements outline above, we also identified nonfunctional requirements, such as scalability of the system, learnability, and availability of the system on the web.

3.3.3   Data Acquisition and Preprocessing

The simulation results from domain experts are stored in groups of text files: one state space text file in the format shown in Figure 2, and T (time) probability files. Each probability text file, shown by each column in Figure 3, encodes the probability distribution over the state space at a particular time step. The size of a single probability file in a system with a state space size of 680,430 is roughly 18MB.

Because the probability distributions can be high-dimensional while still having a spatial distribution, we preprocess the data through dimensionality reduction to a lower space which can be used for visualization. Furthermore, due to the importance of peaks to the user tasks, we explicitly compute and detect these structures. Due to the large-

scale of the data, we pre-process the data offline when reducing the dimension of the state space for visualization. In contrast, peak detection is computed online.

Dimensionality Reduction. Because the probability distributions of stochastic networks can be potentially defined over a space higher than two dimensions, we explore dimensionality reduction via aggregation and projection. To map the data visually, we compute projections to one dimension and to two dimensions, which is commonly done and accepted in the target domain; an additional projection to three dimensions and visualization using volume rendering was attempted and discarded during the prototyping stage.

To project the landscape defined by the probability distributions to one dimension, we aggregate the microstates with the same copy number of a particular protein to form a new state space in one dimension. We repeat the aggregation process for each species of proteins. For example, in the 2-gene toggle switch system, we obtain two probability distributions over two 1D state spaces: one for protein A and the other for protein B.

To project the probability landscape to two dimensions, we aggregate the microstates for which the combinations of the copy numbers of any two proteins are the same. For example, consider a 3-protein example with only two states possible in the case of 0 copies of protein 1 and 0 proteins for protein 2: state (0,0,0) of probability 0.05 and state (0,0,1) of probability 0.05. The new aggregated state (0,0) will have probability 0.1. We repeat the process for all possible combinations of any two proteins in order to obtain multiple 2D state spaces. For example, the 2-gene toggle switch network only has one 2D state space, while a regulatory gene network with three genes consists of three

state spaces projected to 2D, and a network with four genes would include six 2D state spaces.

Peak Detection. To detect the peaks, we use a gradient estimate approach. We compare all the states to their neighbors, as projected to one dimension. If a state has higher probability value than both its neighbors and its value is above a threshold of 1e-12 (empirically determined based on the simulation threshold for error), the state is defined as a peak. Since the probability distribution along each species dimension is independent, the number of peaks in the entire system is multiplied by the numbers of peaks in each dimension. The peak locations in the system are all the combinations of species locations. The process of peak detection is completed online.

3.4   Visual Design

We have designed the visual analysis tools through a parallel prototyping approach, during which multiple low to mid-fidelity prototypes were sketched in parallel, and presented for feedback to the domain experts. The domain experts are particularly attuned to surface plots of probability distributions, as indicated by their use of "landscape" terminology. However, such plots do not capture directly temporal features: time series of 3D surface plots are usually generated as a movie to display the temporal changes. Thus, visual representations with temporal features embedded were of particular interest. As part of this process, we explored multiple potential visual designs, including star plots, small multiples, animation, space-time-cubes, volume renderings and variations of topographic maps, and converged towards those encodings which best preserved features of interest (such as peaks and temporal behavior) while avoiding occlusion.
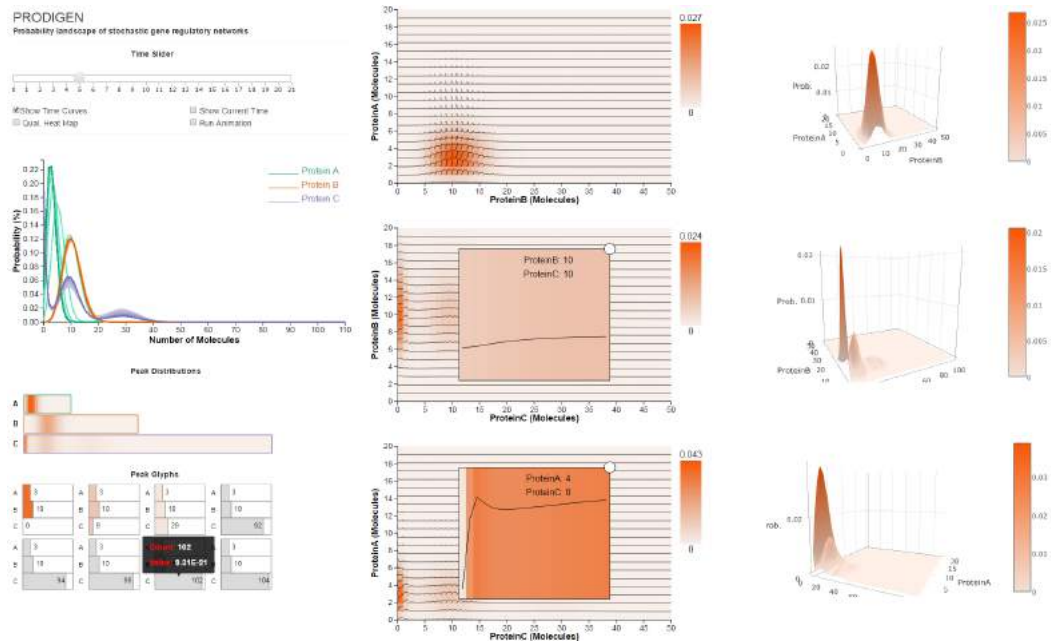
Figure 4. The PRODIGEN interface consists of several visual components: (left top) the ensemble Spaghetti Plots view shows the probability distribution over time, for all the proteins in a system; (left middle) the 1D heatmap view shows a per-protein view of the probability peaks; (left bottom) the Peak Glyph view (displayed as a small multiple) represents all the probability peak states in the system; (center) the 2D heatmap view is enhanced with time-curves, and shows the probability peak correlation between protein pairs over space and time; (right) the animated 3D Surface view describes the shapes of peaks over space and time.

The final prototype of PRODIGEN (Figure 4) consists of a multi-view design with several visual components: 1) a spaghetti plot view that shows the temporal changes of probability values for every gene; 2) a small multiple 1D heatmap view; 3) a small multiple enhanced 2D heatmap view which displays the probability distributions over the state space projected to two dimensions; 4) a peak glyph view which represents the corresponding states of all the probability peaks detected in the system; and 5) a small

multiple 3D surface view. The information shown in the heatmaps and peak glyphs changes according to the user-selected time step, as the user drags a time slider. We describe in detail each visual component below.

3.4.1 Ensemble Spaghetti Plot View

Spaghetti plots have been traditionally used to visualize ensemble data; each single plot represents an individual ensemble member. Color-coding may also be used to differentiate members. In this work, we extend the spaghetti concept to specific simulations over time: by extension, each "ensemble" member represents the probability behavior at a particular time point. The horizontal axis represents the copy numbers of molecules, while the vertical axis represents the probability value. Thus, an individual plot describes the probability distribution over the states with the copy number from zero to the maximum copy number for that protein.

We use color to encode different species of proteins, and the intensity of the color to encode the probability distribution of the corresponding protein at different time steps. The plot intensity from lighter to darker represents the time from the beginning of the simulation to the end.

In our early prototyping stage a third dimension was used to encode time, in the style of space-time cube representations (57). In practice, however, the encoding suffered from occlusions which made difficult the tracking of temporal peak changes, and was later discarded. The domain experts specifically stated that the 2D ensemble spaghetti plots yielded better performance than the cube representations.
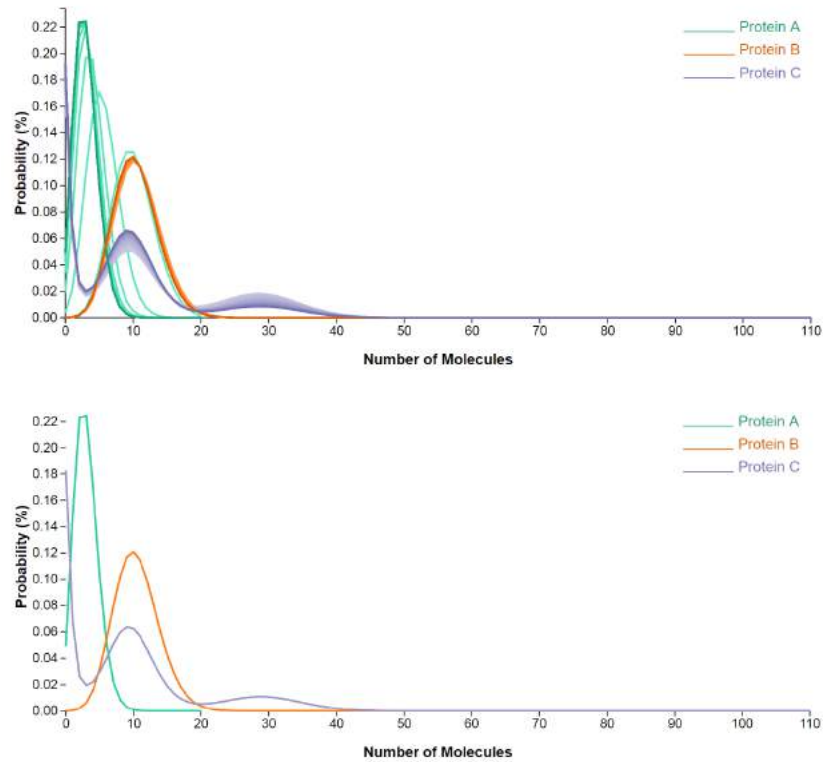
Figure 5. Spaghetti Plots (top) show the probability distribution of each gene and the changes in distributions over time. The Spaghetti Plots view on the bottom displays the probability distribution of each gene at a user selected time step. Protein C has at most 110 molecule copies in this 3-gene regulatory network.

In Figure 5 (top), the Protein A of spaghetti plots represents the temporal probability distribution of protein A. In this representation, the peak changes can be easily tracked in terms of both peak location and value. We notice that protein A has only one peak, whose location shifts to the left in time towards the state with a lower copy number, and whose probability value increases over time. Protein B also has one peak, which stays at the same location without too much change in the probability value. Protein C

has three peaks. The locations of these three peaks do not change over time. However, the probability value of one peak in the middle increases as the other one on the right decreases over time.

In most cases, the peaks either increase or decrease over time, and thus do not cause overlaying or crossing issues. Rare plot overlays and crossings bear in fact meaning, by encoding frequent peak location changes (see Protein A in Figure 5) or peak changes in both directions (increasing and decreasing).

However, it is not easy to detect the probability distribution at a particular time step from the spaghetti plots. To this end, a checkbox filter allows the users to draw the plots at an interactively-selected time step. Figure Figure 5 (bottom) displays the probability distributions of these three proteins at the 10th time step.

3.4.2   Enhanced heatmaps

To show the lower-dimensionality projection of probability landscapes we chose as a basis a heatmap encoding, due to their efficiency in visualizing distribution data. Our domain experts, like most bioinformaticians, were familiar with heatmap representations, compared to other visual representations. We employ both one-dimensional heatmaps and an enhanced version of two-dimensional heatmaps, displayed each time in a small multiples view. The spaghetti view uses a categorical color scheme for the different species, while the aggregated views described below employ a non-overlapping qualitative color scheme to encode probability. The 1-dimensional heatmaps described below (Fig. Figure 6) help bridge the two color schemes. The domain experts did not report or experience during testing any issues related to this dual use of color intensity.

Figure 6. Three 1D heatmaps represent probability distributions projected to one dimension of three proteins A, B and C. A and B have one peak each, while C has three peaks.

1D heatmaps. 1D heatmaps display the probability distribution over the state space projected to one dimensional space. The 1D state space is represented by the horizontal direction in the heatmap. The color intensity of each state represents its aggregated probability value. The heatmaps are stacked up vertically, and each heatmap has a border color-linked to the spaghetti plot for that protein.

Figure 6 shows the probability distribution over one dimensional state space for three proteins at time 0. When projecting to Pa which represents protein A, the only probability peak is located mid-axis, which corresponds to the state in which the copy number of molecule Pa is 10, half of the maximum copy number. Similarly, the peak in the Pb (protein B) projection is also at the state with the copy number of 10, and these two peaks have similar probability values, as indicated by their similar intensities. However, in the Pc (protein C) projection, there are three peaks: a large peak leftmost along the axis (where the state has a copy number of 0), one small peak nine states to the right, and one much smaller peak twenty nine states further right. The smallest

peak in the Pc projection may not be easily observed from the heatmaps, but can be quickly detected through the peak glyphs that will be explained below.
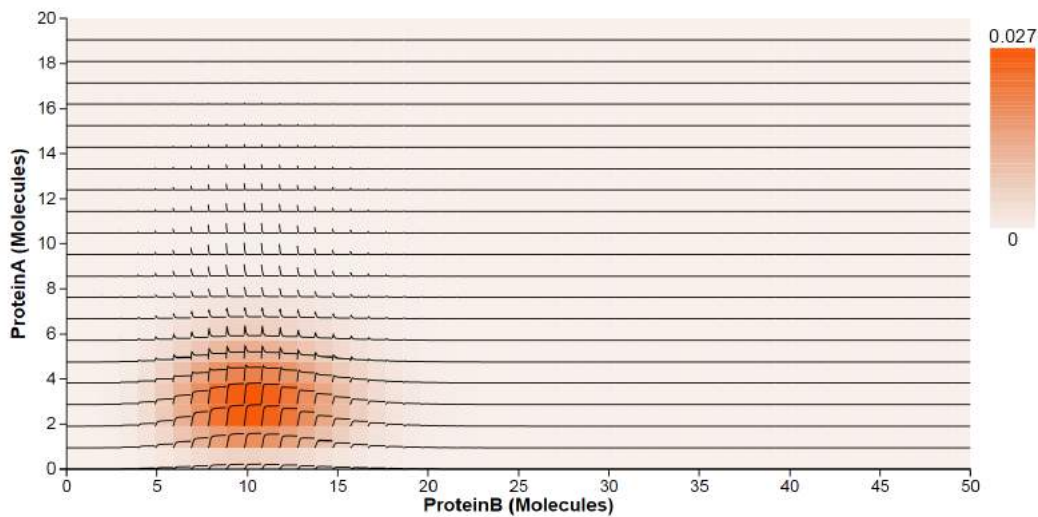
Figure 7. The 2D heatmap displays the probability distribution over the 2D state space projected for the pair of Pa and Pb. Time Curves overlaid on the 2D heatmap indicate how the probability values in the 2D state space change over time.

2D heatmaps. Similar to 1D heatmaps, 2D heatmaps display the probability distribution over the state space, but projected to two dimensional space. The horizontal and vertical axes represent the copy numbers of molecules for any two protein species in the system. The cell at the intersection of two molecule copy numbers represents the state that aggregates the microstates with those copy numbers. The intensity of the cell encodes the probability value. Figure 7 indicates there is only one peak located at

(3, 10) in the 2D state space projected to Pa and Pb. The 2D heatmaps are arranged in a small multiple display, using identical axes mappings to support comparison of the peaks. Per our collaborators request, the heatmap can also be colored using a qualitative colormap (Figure 8), similar to the rainbow maps they had used before in numerical software packages; the qualitative map is derived from ColorBrewer (colorbrewer2.org) and allows the users to highlight different peak classes and thus presumably better identify small peaks.
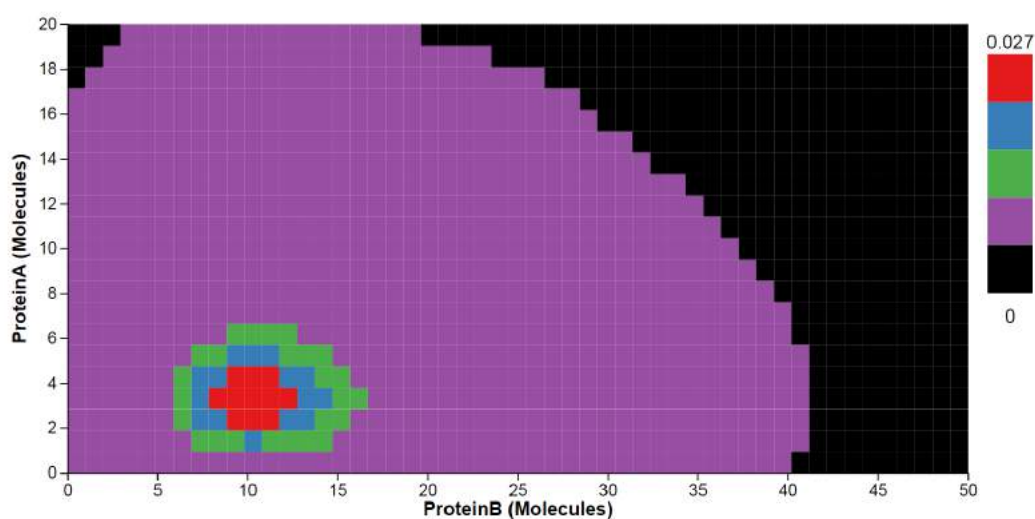
Figure 8. 2D heatmaps using a qualitative colormap.

Overlaid Time Curves. Heatmaps are able to provide an overview of the probability distributions over the state space in one and two dimensions. However, it is

hard to track temporal changes from this aggregated visual representation. Since the changes in probability values over time can be displayed as a line plot, we create an extended heatmap representation by overlaying such line plots on each cell (state) in the 2D heatmaps. The time curves show the dynamics of the system, which include the changes in probability values and peak locations over time. A flat curve means no obvious change, while a steep curve means a big change in the probability value. The user can check the time-curves checkbox to display these curves.

In Figure 7, two groups of steep curves indicate that the peak in the state space projected to Pa and Pb moves in the direction in which Pa has lower copies at the early time, and stays there during the rest of the simulation time. The probability value of the peak increases as the location of the peak changes.
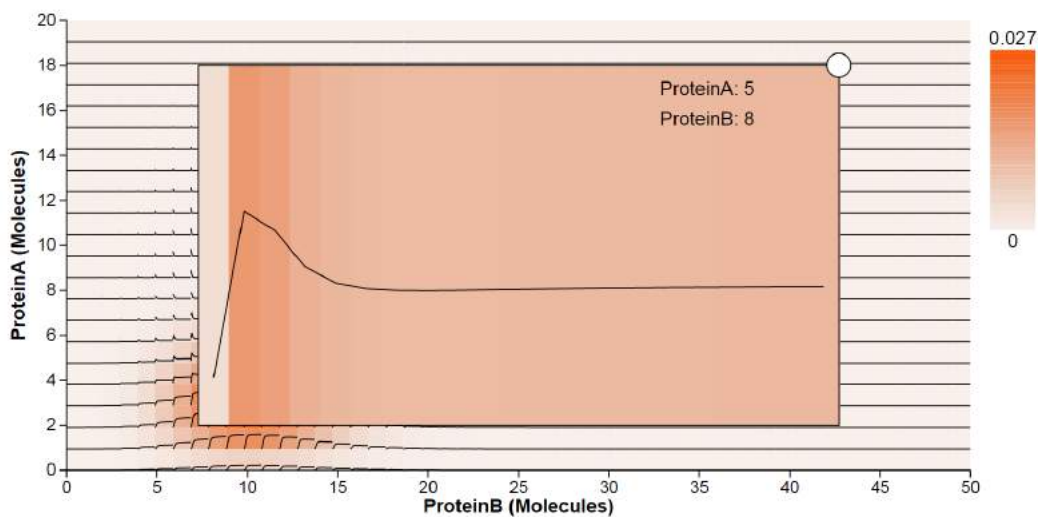
Figure 9. Probability temporal distribution on a user selected state in which Pa has five copies of molecules and Pb has eight.

Details-on-demand. Since the 2D heatmap can display state spaces that contain thousands of states, it may be difficult to observe how the probability of a specific state changes over time. To support detailed inspection of a particular state, we enlarge on demand the size of the cell corresponding to a user-selected state. When the user clicks on an interesting cell, a detail window is overlaid, showing the time curve for that state. Figure 9 is an example of the detail view of the state (5, 8), in which Pa has 5 molecule copies and Pb has 8 copies. We notice that there is a steep and large increase in the probability value during the first two time steps. Later, the probability decreases in the following four time steps, and keeps roughly constant for the rest of time.

3.4.3   Peak Glyphs

Although the heatmaps view can provide an overview which indicates the location of peaks, users often wish to identify the location accurately. Furthermore, the heatmap representations can make small peak detection difficult. To indicate the peak location, we have experimented with a number of existing multidimensional encodings, including star plots. However, since a stochastic gene system consists of a large number of microstates and each microstate is represented by the combination of copy numbers of multiple molecular species at different scales, existing encodings failed to provide the location information of states in the system.

Instead, the end-result of the prototyping phase was a glyph to display the detailed information of a peak state, such as the combination of copy numbers. The peak glyph is composed of a stack of horizontal bars filled with darker or lighter intensities according to the probability value (Figure 10). The number of bars represents the number of
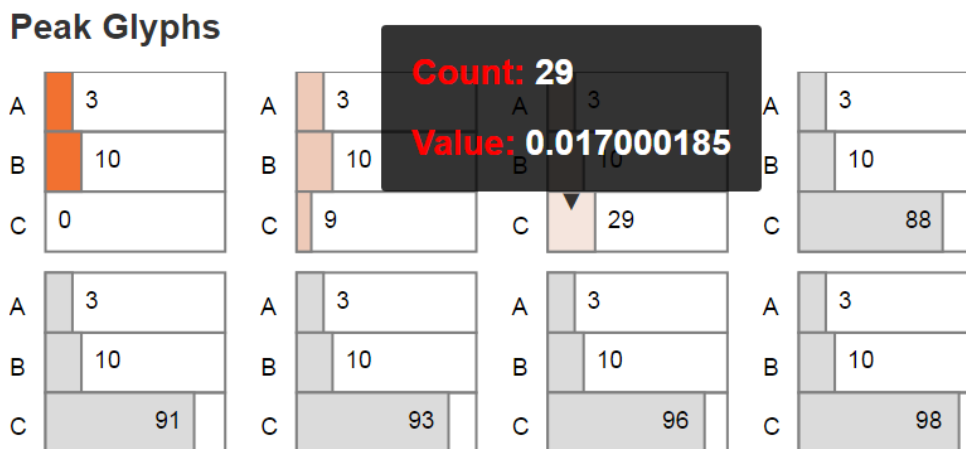
Peak Glyphs

Figure 10. A small multiple of Peak Glyphs with eight peak states, out of which only three are authentic peaks. The remaining five states, shown in gray, were identified as computational errors.

proteins, while the length of each bar represents the number of molecule copies of the corresponding protein. When displayed as a small multiple, the peak glyphs show all the peaks in the system. Because very small probability peaks are typically due to numerical error in the simulation, suspect peaks are displayed in gray. The view is controlled by the user selected timestep. Hovering over a glyph displays the detailed information about the peak.

Figure 10 displays eight potential peak states in a 3-protein system. However, five of these states appear as peaks likely due to numerical error and are shown in gray. Only three of the states are authentic peak states, at the location of (3, 10, 0), (3, 10, 9) and (3, 10, 29). The details on demand indicate, for example, that Pc has a peak at the

state with 29 copies and that its peak value at that location is about 0.017. In general, the peak glyphs can be used as a detailed guide through the peak set generated by a system.

3.4.4   3D Surface View

Similar to the 2D heatmaps, the 3D surface plots also display the probability distributions over the 2D state space, but represent the probability value as height (z-axis) instead of encoding it by the color intensity (Figure 11). As indicated above, the domain experts are particularly receptive to 3D surface plots, which reflect their understanding of probabilistic landscapes. In contrast, the complementary visual encodings were novel to our domain experts, and thus benefited from scaffolding through the familiar surface encodings. Compared to the 2D heatmaps, the advantage is that the surface plots can show the shapes of peaks in a manner similar to existing representations of probability landscapes. Both representations, extended heatmaps and surfaces, capture well the extent of peaks and the relative location of peaks with respect to one another.

We use the library Plotly.js for plotting the 3D surface of probability distribution; the library allows users to rotate and zoom in/out the surfaces. Hovering over the surface displays the location of the state and its probability value. We implement an animation option for playing the probability distributions across the entire time period as a movie, in which the dynamics of peaks including numbers, locations, relationships, values, and shapes can be easily detected. Similar to the heatmaps, the 3D surface plots are also displayed as a small multiple across the potentially multiple projections.
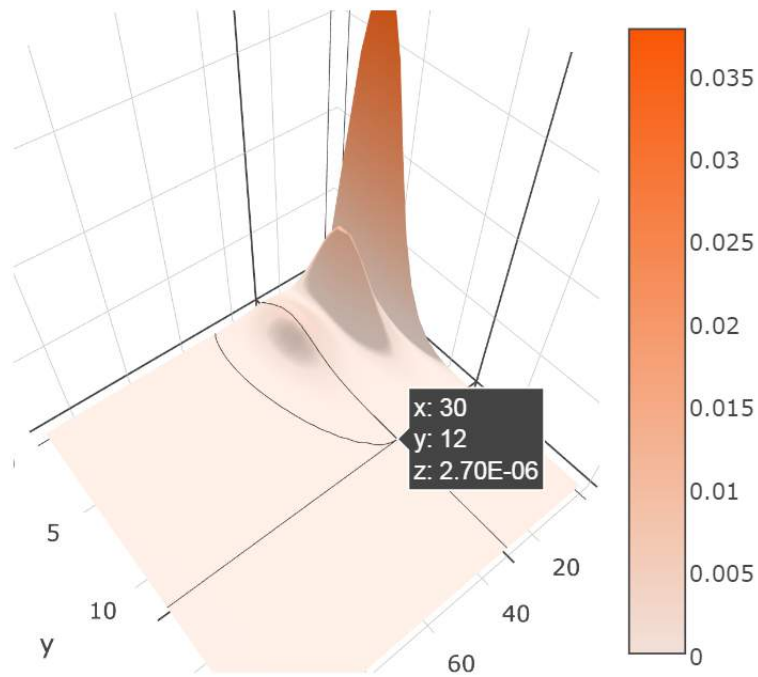
Figure 11. A 3D Surface Plot displays the probability distribution over the state space projected to A and C. Three peaks are distinguishable.

PRODIGEN is implemented as a web-based tool in JavaScript and uses the data visualization libraries D3 and Plotly.js.

3.5   Evaluation

We evaluate the effectiveness of PRODIGEN through two case studies completed with two experts, who are co-authors on this manuscript. One expert is a senior bioinformatics researcher with over thirty years of experience in the field, and the other is a junior researcher in bioinformatics who specializes in stochastic gene network modeling.

These two case studies demonstrate how our visual approach can assist domain experts in the exploration of probability landscapes in stochastic gene networks, and allows them to track the temporal changes as well. The first case study is an exploration of a transcription regulation network with three genes, and the second one studies a toggle switch system with two genes.

3.5.1 Case Study I: Transcription Regulation Network

We illustrate the performance of PRODIGEN using the example of a transcription regulation network. A detailed description of the model architecture will be published by our collaborators. The system consists of three genes GeneA, GeneB, and GeneC, which express ProteinA, ProteinB, and ProteinC, correspondingly. C is the output of the system, and it is regulated by both B and A. Transcription regulations motifs such as this are common in studies of biological systems as, for example, *E. coli* (58), yeast (59), and mammals (60). Their detailed modeling in principle can answer important biological questions such as the existence of multiple phenotypes corresponding to different numbers of stable states, adaptability of the systems to the change in one of the components (61). These phenomena may be relevant for the study of important biological events, such as differentiation of the cell to a new cell type, tumor suppression, HIV latency versus replication, and cell adaptation to a stress. In this simulation, the domain expert decreased the expression level of A by a factor of approximately 3. The goal of the visual analysis was to observe how C responds to this change in the expression level of A.

The spaghetti plot in Figure 4 (left middle) indicates that even a significantly decreased level of A does not affect significantly the behavior of the system. The spaghetti view clearly shows that A has one peak state, B has also one peak state, while C, which is regulated by both A and B, has three peak states. As simulation time advances, the peak-state corresponding to A shifts towards the left along the horizontal axis as the expression level of A decreases. In the same spaghetti plot, the peak corresponding to B is very stable; its clear signature in the plot shows very little change in either location or probability value. The three peaks of C also stay constantly in the same locations, and show little change in probability values over time. The spaghetti plot can be used to detect the number of peaks (T2), and track their temporal changes in one dimension (T4).

The 1-dimensional set of projections (Figure 4 left) captures, as expected, the same number of peaks. The smaller and fainter peaks in the C projection caught the interest of the investigators, who remarked again on the biological significance of small peaks. Small peaks are important since they can correspond to a diseased state of the cell, which could be non-prevalent or rare in the organism and as such have a small number of molecules. It was easier to detect small peaks in the one dimensional view than in 2D, because in 2D the information gets distributed over a larger number of states. The value and location of the peaks are represented in both the spaghetti and 1-dimensional view in a qualitative way.

A step further, the peak glyph view (Figure 4 left bottom) captures the accurate locations and probability values of the peak states (T2 and T3). The peaks with different

values at different locations can correspond to different physiological conditions of the cell, e.g., healthy vs. diseased state or pluripotent vs. differentiated in stem cells. The gray glyphs show the pseudo peaks that are due to numerical errors. As simulation time advanced, the experts noticed an increase in numerical errors. With this observation in mind, the users moved back to exploring relatively more complex visual representations, such as the 2D heatmaps overlaid with the time curves and the 3D surface plots.

Next, we investigate the probability distributions over the entire state space of the transcription regulatory network (Figure 4 center), and again focus on the same small peak. The distributions are displayed through three 2D heatmaps which are the projections of the three possible combinations of the original three components. The time curves overlaid on the heatmaps further capture the trace of peak movements (T4). For example, based on the shapes of time curves, we can conclude that the peak moves down in the heatmap projected to A and B. Last but not least, the surface plot animation (Figure 4 right) confirms the smooth trajectory of the peaks. The 2D enhanced heatmap outperform other visual designs when the user focuses on tracking the temporal changes of probability distributions in 2D state spaces over time.

In conclusion, while the steady state probability landscapes of ProteinA and ProteinB are monostable, ProteinC has three peaks at steady-state. The decrease of the expression level of ProteinA drives the evolution of the system to a new steady state, which has decreased level of ProteinA, with the expression patterns of ProteinB and ProteinC both altered.
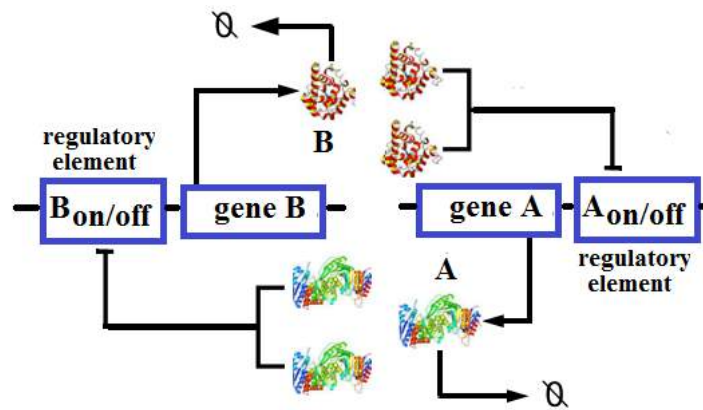
Figure 12. The genetic toggle switch network (62)

3.5.2   Case Study II: Toggle Switch System

The stochastic network module we study here is a genetic toggle switch system consisting of two genes, A and B. Figure 12 shows the network topology. In this second model, single copies of gene A and gene B express protein products A and B. Gene A and gene B can repress the transcription of each other through binding the dimers of their protein products A and B on the promoter sites of the other gene B and A to form protein-DNA complexes (62; 40; 41). Thus, this genetic network includes six molecular species, which are described in detail in Table I. The genetic toggle switch network with two genes has four distinct stable states: on-on representing a state at which both gene A and gene B are unbound, on-off representing a state with unbound gene A and bound gene B, off-on for a state with bound gene A and unbound gene B, and off-off for a state with both bound gene A and gene B. In this work, the domain experts computed the state spaces

Molecule Name	Molecule Structure	Max Copy Number
Pa	Protein A	120
Pb	Protein B	240
Da	GeneA in the unbound state	1
Db	GeneB in the unbound state	1
BDa	GeneA bound by protein B	1
BDb	GeneB bound by protein A	1

TABLE I

A DESCRIPTOR OF THE 2-GENE TOGGLE SWITCH NETWORK.

under the initial condition of 0 copies of protein A and protein B, 1 copy of unbound gene A, 1 copy of unbound gene B, and 0 copies of bound gene A and bound gene B. By using the finite buffer method to solve the dCME of this system, we obtain a state space of size 115,200. After that, we directly compute the probability value of each state at each time point, which forms the probability landscape, and currently output for visualization the first 200 time steps.

The toggle switch system is known to have four peaks in the stable state, as captured by the peak glyphs (Figure 13). However, at the early time, the system has only one peak, at the state (31, 63). The enhanced heatmaps (Figure 14) and the animation offer an opportunity to observe how the peaks move in the state space in both the 2D heatmap and the 3D surface plot (T4). In addition to tracking the location changes, the surface plot shows efficiently the changes in probability values (Figure 15). As simulation time advanced, we noticed three more peaks appear, as the early peak was getting smaller and transitioned to states with higher copy numbers of both protein species. Figure 15 displays the four peaks in the 3D surface plot at different time steps. Later on, the

system becomes more stable. These peaks stay in a roughly fixed location with only small changes in the probability value. Thus, we can confirm that the system reaches a steady state.
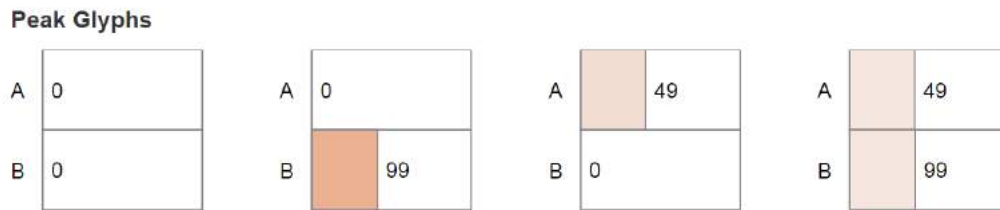
Figure 13. Peak glyphs for the toggle switch network showing four detected peaks and their exact coordinates.

The peak glyphs successfully identified the largest peak (T2). We see one of the three late peaks become the largest peak, at last. Although the toggle switch has more states in the 2D state space, compared to the transcription regulatory network discussed in the case study I, the time curves overlaid on the heatmap, which includes nearly thirty thousand map cells, can still provide an overview of the temporal probability distributions (Figure 14). However, we can only see three peaks according to the curves shapes. The fourth peak is actually a spike at the location (0, 0), which can be seen in the surface plot, and as such was dismissed from the analysis.
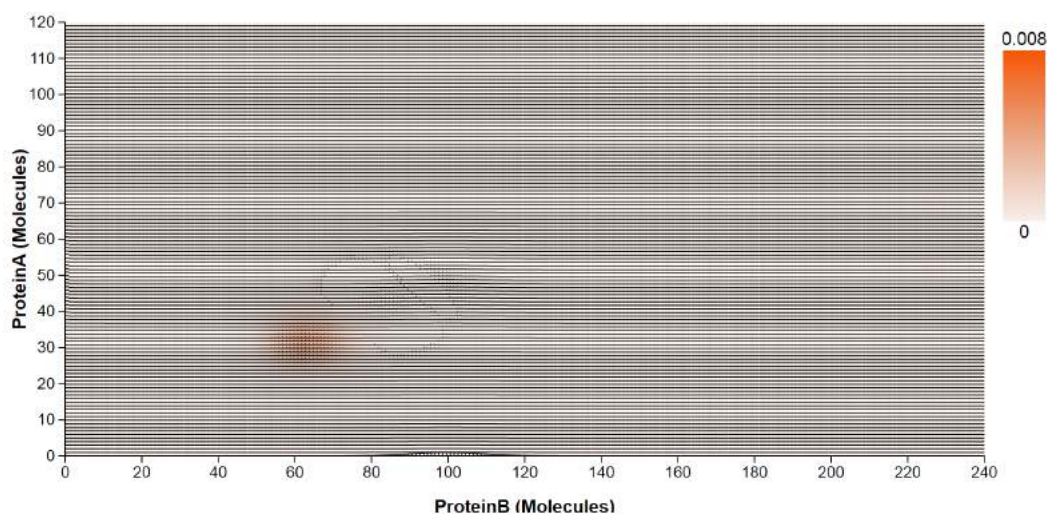
Figure 14. Toggle Switch 2D enhanced heatmap. The heatmap displays the probability distribution over the 2D state space projected for the pair of Pa and Pb. Time Curves overlaid on the 2D heatmap indicate how the probability values in the 2D state space change over time.

3.5.3   Informal Feedback

Over multiple discussion sessions, the domain experts provided positive feedback. The junior researcher pointed out that the visual approach would be beneficial to experts working in the field when exploring the data. In particular, the lower-dimensional plots and encodings go beyond current approaches to showing the data, especially for exploring the dynamics of the system. The feature that generated most excitement for both domain experts was the ability to track the temporal changes of the probability distributions. The senior expert specifically commented on the potential of the tool to assist in deriving new hypotheses, beyond the post-hoc abilities of explanatory visualization. In addition to tracking the peak trajectories in the transcription regulatory network and the toggle
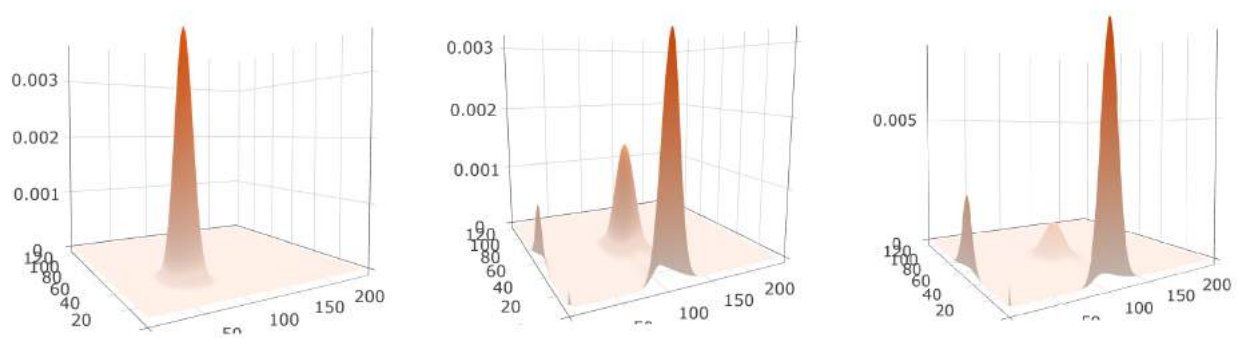
Figure 15. Probability landscape of the toggle switch system at three different time steps, showing four peaks; the least noticeable peak is located at (0,0).

switch system, the senior expert also indicated the potential use of the visual approach to systems with potentially oscillatory peak states, or to track the peak trajectory for other types of stochastic networks.

3.6   Discussion

The two case studies and expert feedback demonstrate that PRODIGEN is an effective visualization tool for the exploration of the probability landscape of stochastic gene regulatory networks. The system successfully allowed stochastic network researchers to perform the T1 through T4 tasks we had identified in the domain characterization stage, while handling gracefully the large scale of the data. In particular, the domain experts were able to: 1) visualize probabilities of stable states, 2) explore the temporal aspect of probability distributions, and 3) discover and analyze small peaks in the probability landscape that had potential relation to specific diseases. In both case studies they were able to identify new findings (including numerical errors) as well as replicate previous

findings. The web-based implementation of the system makes it potentially available to a wide audience. Furthermore, the case studies performed with domain experts indicate the interface is user-friendly and easy to learn.

With respect to the visual design of the system, the overall design successfully supports the exploration of probability distributions in both state and time space. While some of the individual visual encodings employed are not new, the enhanced 2D heatmaps and the peak glyphs are novel contributions. The combination of visual encodings to explore data in multiple dimensions is also novel. As shown by the case studies, the visual encodings proposed were able to capture probability distributions in multiple dimensions and at multiple levels. The spaghetti plots and heatmaps overlaid by time curves are able to display the temporal probability distributions over the state space. In addition to these overviews, detail views embedded in the heatmaps provide the ability to track the probability dynamics of a user selected state. When combined together with the peak glyphs and the animated surfaces, these visual encodings satisfy the requirements from experts for detecting the number of peaks, the locations of peaks, the values of peaks, and their dynamic changes. Because the peak values vary significantly, the experts required we do not normalize axis or color; instead, we clearly indicate each scale in the interface. For example, the fourth small peak in the center plot in Figure 15 is not perceivable when using normalization.

In terms of scalability, the tool can handle with no issues models whose size in terms of species is on par with the ones developed by our collaborators, and was built with visual scalability in mind. Furthermore, we have tested the system on a variety of state

space size and timestep configurations. As discussed in the Task Analysis section, the systems modeled typically contain few molecular species. However, the system modeled may be large in both the state space and time. Our visual approach scales well, in this respect, in either state space or time dimension. The entire system behaves well for stochastic networks with a relatively large state space size of 680,430 states, which is roughly 18MB, as long as the simulation of these systems stays below several hundred timesteps. This last limitation is related to the D3 library limitations regarding loading large datasets. For practical purposes, to circumvent this limitation, simulations can be split into chunks of time and loaded and analyzed sequentially. The time-curve heat maps can further suffer from scalability issues for very large state space sizes, which can lead to small tiles and aliasing. On lower resolution displays, this issue becomes visible for large state spaces with long runs of close to a thousand timesteps, as indicated in Figure 14. Context+focus techniques (63) may help address aliasing issues.

In terms of limitations, the system currently requires a copy of the data to be created locally and loaded for processing and visualization. PRODIGEN can only be used as a post simulation analysis tool, due to this data pre-processing load. A direction of great interest to our collaborators, although beyond the scope of this work, is the ability to run the system directly in situ, on the cluster where the simulations are computed.

PRODIGEN is limited in its capability to enable the visual analysis of these ensemble stochastic simulations, as it is only designed to analyze a single simulation at a time. This is an especially important limitation, as the gene regulatory networks are a complex biological process which requires a comparative analysis of multiple simulations in order

to gain a complete understanding of their mechanisms. Advancing research in stochastic gene networks requires an effective visual approach that allows the exploration and comparison of multiple simulations to find how changes in the probability landscape are associated with the time and system settings. In the next Section, I will introduce a web-based visual analysis tool which uses this approach to enable the analysis of stochastic gene regulatory networks. This tool allows for the exploration of peak distributions over state space and simulation time in such stochastic networks, and the comparison of peak distributions between multiple simulations.

3.7   Extension of PRODIGEN

While allowing for the comparison and analysis of multiple gene regulatory network simulations is necessary, certain visualization challenges arise for high-dimensional ensemble simulations. One such challenge is providing a comparative analysis of a large number of stochastic simulations, which feature many states across many timesteps. This is a common challenge of ensemble visualization of high dimensional data, which is exacerbated by the need to preserve spatial information in order to enable the domain specific tasks.

To support these tasks, we propose a web-based visual analysis tool (Figure 16) that combines multiple linked views to capture ensemble time-evolving probability landscapes at three levels. A peak trajectory cube provides users an overview of peak spatiotemporal distributions across six simulations. A peak projection map shows the exact peak locations of multiple simulations at the user selected time. At a more detailed level, the user can inspect a particular state in the peak projection map to view for each simula-
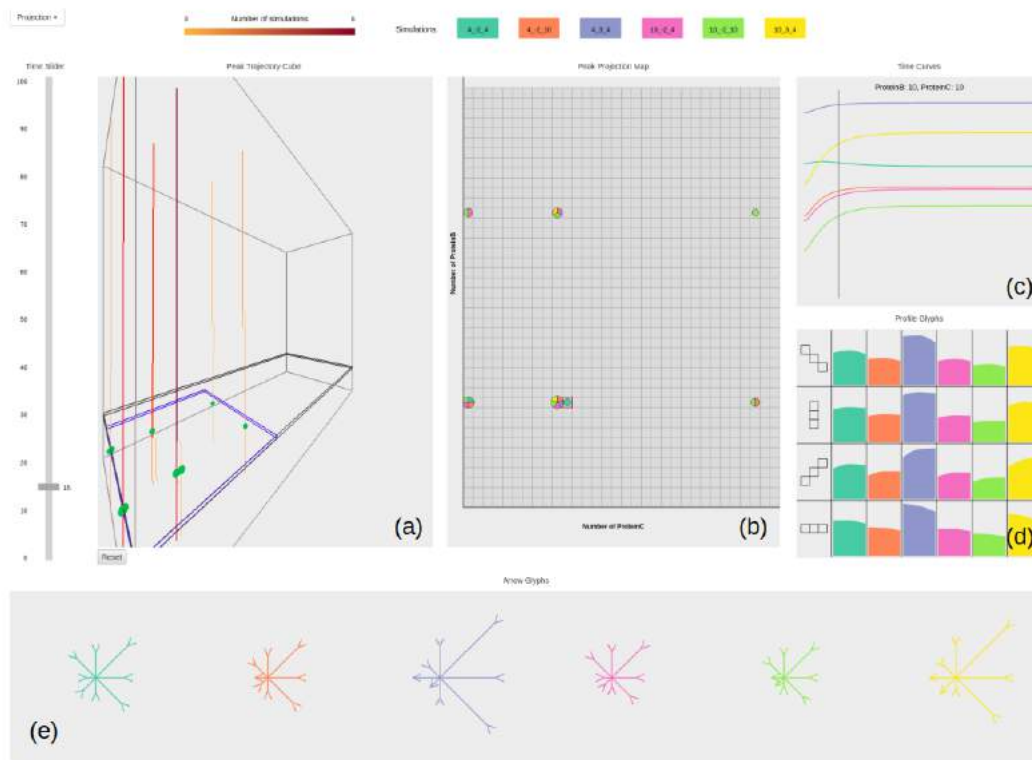
Figure 16. EnsembleProb, a web-based visual analysis tool for the exploration of peak distributions over state space and simulation time in such stochastic networks, and the comparison of peak distributions between multiple simulations

tion both the probability values over time and the local probability landscape shapes. This information is displayed in a small multiple using two glyphs: profile glyphs and arrow glyphs. The profile glyphs show the profiles of the peak from four angles — 0, 45, 90, and 135 degrees, rotating about the center of the peak. The arrow glyph indicates that a state is a peak when all the glyph eight arrows point towards the glyph center. All the views are linked together through interactions. The visual design was finalized according to the feedback from our domain experts.

We have completed a usage scenario within the visualization development team to evaluate this visualization tool. The peak trajectory cube (Figure 16 (a)) maps the 2-protein state space to the X/Z plane, and time to the Y axis of a cube. Yellow-red polylines map the locations of peaks (highest probability points) in this spatiotemporal space, across six simulations. The redder the polyline, the more peaks are sharing that trajectory across the simulations. The trajectory cube also provides interactions like rotation and zooming in/out to overcome the occlusion issue in 3D view. For the selected time (dark gray horizontal rectangle), green dots mark peak locations. Peak locations converge over time. The blue state subarea is mapped to Figure 16 (b). The x/y plane of the peak trajectory view is a detailed view of the blue subarea in Figure 16 (a). A glyph encodes with color which simulation runs contain a peak at that glyph location; colors are mapped to simulations. The green simulation has the most peaks among all six simulations. The Time Curves (Figure 16 (c)) show the orange and pink simulations have very similar probability temporal distributions at the state highlighted in the red box in Figure 16 (b). The Profile Glyphs (Figure 16 (d)) show that the peaks in the blue and yellow simulations are steeper than the other four. The Arrow Glyphs (Figure 16 (e)) indicates that a state is a peak state when all the eight arrows in the glyph point towards the glyph center. The states shown in darker green and pink are peak states. However, the other four states shown are very near a peak state, but are not peaks themselves. In the Figure 16 (b) and (e), we notice a disagreement between the arrow glyphs and the peak projection map. This mismatch demonstrates

that probability distributions over the genes in this system are not independent of each other.

In terms of limitations, using the continuous colormap in the peak trajectory cube may not be a great choice to encode the number of simulations that share the same trajectory, as it becomes difficult to distinguish the multiple discrete values. We could drop a mark in the trajectory when and where peaks converge.

In summary, our visual analysis tool allows bioinformatics researchers to explore and compare the time evolving characteristics of probability landscapes from multiple simulations efficiently, without running many small scripts and computing each characteristic separately.

Besides ensemble stochastic gene regulatory networks, dynamic biological networks also exist in other biological domains, e.g., neuroscience. These networks also feature both spatial and non-spatial information and require comparative visual approach at multiple scales. In neuroscience, the study of the relationship between functional connections of the brain and its spatial structures is an open research topic. In the next two chapters, I will introduce the domain problem of dynamic brain networks, followed by data and task analysis as well. To support those tasks, I present several innovative visual designs and a visual approach for the ensemble exploration and comparison at multiple scales.

CHAPTER 4

SWORDPLOTS: EXPLORING NEURON BEHAVIOR WITHIN
DYNAMIC COMMUNITIES OF BRAIN NETWORKS
(PREVIOUSLY PUBLISHED AS CHIHUA MA, ANGUS G.
FORBES, DANIEL A. LLANO, TANYA BERGER-WOLF, AND
ROBERT V. KENYON. (2016) SWORDPLOTS; EXPLORING
NEURON BEHAVIOR WITHIN DYNAMIC COMMUNITIES OF
BRAIN NETWORKS, JOURNAL OF IMAGING SCIENCE AND
TECHNOLOGY 60(1): 010405-1010405-13

4.1 Introduction

Recent neuroscience research indicates that cognitive operations are performed not by in-
dividual brain regions working in isolation, but by networks consisting of several discrete
brain regions which act in synchrony (2). These networks share "functional connectiv-
ity", meaning that activity in these regions is tightly coupled—in the sense of a statistical
association or dependency among two or more anatomically distinct time-series events.
Functional connectivity between brain regions can change rapidly over time (4; 5), giving
these networks a highly dynamic characteristic. Abnormalities in functional connectiv-
ity have been linked with various degenerative and developmental affections. Evidence
suggests, for example, that Alzheimer's disease spreads from one brain region to a non-
adjacent region within a specific network, which is "activated when a person is recalling

recent autobiographical events" (2). However, even for simple networks, the subtle dynamics of these networks are not fully understood. Therefore, techniques that are able to extract the dynamics of functional connectivity from brain imaging data have high potential value to the neuroscience community.

We worked closely with neuroscientists who use the social network analysis (SNA) model to explore changes in functional behaviors of the brain networks and their resulting effects on community identities of neurons over time. The application of SNA to the study of brain networks has nodes mapped to a single neuron or a group of neurons, while edges between nodes represent a functional connection between neuronal sites. Neurons communicate through electrochemical signals. They process incoming chemical and electrical signals and transmit these signals to other neurons and other types of cells. A neural network is a group of connected neurons that performs a certain function. A community is defined as a group of nodes that tend to interact with other nodes in the same group more often than the nodes outside the group. The research goal is to utilize social network modeling algorithms to analyze mouse brain time-series imaging data to uncover interactions among neurons across time and space. The brain networks can be identified through the application of a linear correlation method and then passing these correlation metrics to a community analysis algorithm. Given the large numbers of potential nodes (>20,000) in a typical brain imaging experiment, effective visual representations are necessarily needed to help domain scientists to study the mouse brain using the model of SNA. Although a number of visualization tools have been developed during the last few years, the development of applications for the visual analysis of

dynamic brain networks still requires further effort. A major challenge in visualizing dynamic brain networks is to reveal temporal features while at the same time accounting for additional constraints imposed by brain architecture, i.e., neuron location.

In this chapter, we present SwordPlots, an interactive multi-view visualization application which assists the domain scientist in exploring neuron behaviors within dynamic communities in the mouse brain. Our visualization design also provides the ability to filter data in a variety of ways in order to make it easier for the user to explore the data. The user can use these filters to focus on a single node of interest. A single node encapsulates many neurons, but neuroscientists assume that the neurons within a single node have similar functionality since they are all in close proximity to the same brain region. Once a node of interest has been identified, the user has the ability to highlight the node to gain more detailed information. We also create an additional visualization view which displays some characteristics of the community model structure in space. We use two case studies to illustrate the effectiveness of our visualization tool. The visualization has the overall effect of reducing the data load on the scientists by allowing them to pare down a large number of neural connections within an experiment to those areas of interest and examine their behavior in time.

The visualization has the contributions of allowing neuroscientists to choose any one of the nodes to track its behavior within dynamic brain communities during a selected time range, providing sufficient interactions at both the individual community level and the node level which can offer the user an overview of the network and the ability to focus

on nodes of interest, and being designed for a very new research topic which combines the SNA with neuroscience.

4.2 Background and Related Work

4.2.1 Domain Background

Network analysis of brain imaging data has proven to be a particularly challenging computational problem. For example, functional magnetic resonance imaging (fMRI) data using blood oxygen level dependent signals involves the analysis of tens of thousands of voxels, each with their own time series. Analysis of pairwise correlations across all of these voxels produces millions of potential interactions between voxels, leading to difficulties in producing a coherent understanding of the patterns of activity. Early efforts to analyze these data sets were focussed on seed-based correlation approaches, whereby a time series in a region of interest was correlated to all other voxels in the brain (64; 65; 66). Such methods have uncovered multiple different large-scale brain networks, but are susceptible to biases based on the initial seeds chosen. Multivariate approaches, such as independent component analysis (ICA), have also been used and are free from the biases inherent in visualizing networks based on an initial seed (67; 68). Most recently, graph-theory-based methods have been applied to brain imaging data and have revealed not only the presence of underlying networks but also the logic underlying the organization of networks by characterizing certain nodes as hubs, and by revealing small-world behavior in certain brain networks (69).

A potential problem with these approaches to brain network analysis is their limited ability to extract the natural dynamics of a network's organization. That is, generally

these techniques are used to generate a single static network map using a window of many minutes of the obtained data. However, brain activity is intrinsically highly dynamic, with functional associations between neurons and brain regions that ebb and flow as the organism's level of arousal, focus of attention or topic of thought changes. Recently, neuroscientists have applied SNA tools initially developed for the study of social dynamics (70; 71) to the analysis of mouse brain imaging data. Among the analysis tools commonly used to understand the relationship between structure and function in social networks is community inference. The definition of "community" in the analysis of brain networks is analogous to that of a neural assembly (72), which we define as a group of neurons that act collectively and are functionally connected.

In network analysis, communities are generally defined as groups of nodes that tend to have either more or stronger connections with each other. Nodes belonging to different communities have few and weaker connections. Community analysis can be applied to a variety of fields from social networks to biological networks (73). Communities appear in networks where nodes join together in tight groups that have few connections between them. Several methods have been proposed to find the structure, of which the best known is called modularity optimization. Modularity is a standard metric for finding and evaluating communities (74). Community detection by modularity optimization on large networks is a computationally challenging problem. Therefore, it needs to use algorithms that find high modularity partitions within large networks in a short time. The Louvain algorithm (75) is ideally suited to finding communities in such cases. Besides the advantage of being extremely fast, its steps are easy to implement.

Because dynamic networks change their topological structure over time, we are interested in analyzing the evolution of communities across time that can be defined as dynamic communities. We use the method of Dynamic Community Inference (CommDy) (70; 76; 77; 78) to study how the interactions and the structure of clusters in dynamic networks change over time. In this method, dynamic communities are essentially viewed as dynamic clusters, where the membership of the individual inside the cluster is determined by a total value of "social cost". The definition of social cost is based on two explicit assumptions about individual behavior, motivated by research in the social sciences (70). First, it assumes that individuals tend not to change their home community affiliation too often (79). Second, it assumes that individuals tend to interact with their respective home communities most of the time (80). These assumptions are translated into three cost parameters potentially incurred by an individual. First, it posits a cost for a switch from one community to another. Second, there is a cost of visiting a community of which one is not a member. Third, in data sets for which not all individuals are observed all of the time, there is a cost of absence for an individual who is not observed at a gathering of the community of which it is a member. A dynamic community is then defined as a time series of sets of individuals among whom the overall social cost of interacting is minimized (70; 76; 78).

In summary, CommDy produces two identification codes: a Home community identifying the community the node belongs to by default, and a Temporary community identifying the community the node currently visits. Figure 17 shows an example network of five individuals during six time steps. In this illustration, the color of the
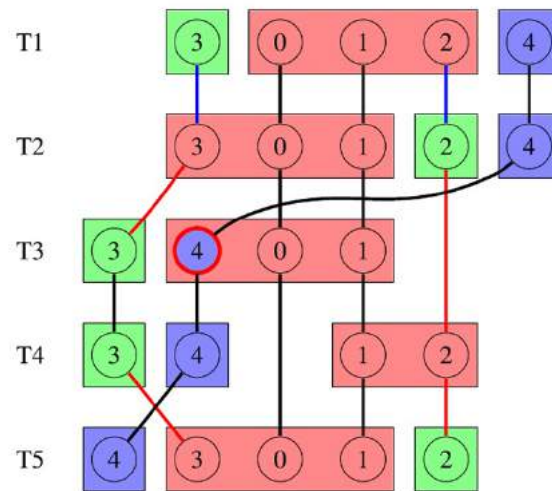
Figure 17. Visualization of CommDy (42) on an example data set including five members across five time steps. Rectangles represent temporary communities, while circles represent home communities. The same colors indicate the same community identifications. At T1, there are three communities: green, pink and blue. At T2, member 3 switches to the pink community and member 2 switches to the green community. At T3, member 2 is absent and member 4 visits the pink community without changing its home community identification. At T4, member 0 is absent, member 2 switches to the pink community, and member 4 comes back to its home community. At T5, member 3 switches to the pink community and member 2 switches to the green community.

circle represents the home community identification code of the node, and its temporary community identification code is encoded by the color of the rectangle surrounding the circles.

4.2.2   Brain Connectivity Visualization

A wide range of visualization tools for biological network analysis has been discussed in a comprehensive survey by Pavlopoulos et al. (81). Cytoscape (82; 83) is a popular bioinformatics package for biological network visualization and data integration. In gen-

eral, the analysis and visualization of brain networks focus on either the cell level or the region of interest (ROI) level. Caat et al. (16) present high-density EEG coherence using a functional unit map. Janoos et al. (17) propose a visual analysis of fMRI data for the exploration of temporal relationships in brain function. Li et al. (84) present a visual analytics toolkit for ROI identification and brain network exploration from fMRI and DTI data. Jianu et al. (85) develop a visualization of two-dimensional neural maps for connectivity exploration and analysis in the human brain. Irimia et al. (86) implement a circle-based visual representation of human connectomics for classifying neuron connectivity relationships in the brain. Margulies et al. (87) provide an overview of various frameworks for visualizing anatomical and functional connectivity in the human brain. Christodoulou et al. present BrainNetVis (88), which effectively quantifies and visualizes brain networks. Xia et al. develop a toolbox called BrainNet Viewer (89) that visualizes the topological properties of brain networks using a ball-and-stick model. LaPlante et al. propose a package (90) for visualizing multimodal human brain networks with the 3D Brain view, the Matrix view, and the Circle view. In addition, the comparison of different connectivity data is another important problem in the neurosciences. Alper et al. (91) discussed matrix-based visual representations to compare brain connectivity patterns. So far, the visualizations discussed above focus on data at the ROI level, such as EEG and fMRI.

Other visualizations focus on data at a microscopic scale with higher spatial resolution, such as the structure or activity of an individual neuron, or integrate brain data at those two scales. Lin et al. developed the Neuron Navigator (NNG) (92) for ana-

lyzing, observing and discovering the connectivity within the neural maze. Sorger et al. implement an interactive two-dimensional graph, neuroMap (93), to render the brain of the fruit fly and its interconnections using a circuit-style wiring diagram. Several other visualizations in neuroscience, such as VisNEST (18), ConnectomeExplorer (21), NeuroLines (22), have been discussed in detail in Chapter 2.1.

4.2.3   Dynamic Network Visualization

Beck et al. (94) state that dynamic data are usually visualized using animation or a timeline-based representation. Animation of a series of dynamic graphs has the advantage that the entire screen can be devoted to the drawing of the graph (95). GraphAEL (96) supports the notion of time slice to visualize evolving graphs with a temporal component. Frishman and Tal developed an algorithm for drawing a sequence of clustered graphs (97) and also created a method for drawing a sequence of graphs online based on a force-directed node-link layout (98). Animation of transitions between time slices may also help users to understand how the graph structure evolves (97). Bach et al. (99) designed a visual interface called GraphDiaries to improve support for identifying, tracking and understanding changes in dynamic networks. Several other studies (100; 101; 102) also support the idea that animation can be an effective way of presenting transitions that are beneficial for the purpose of visualization.

However, Archambault et al. (95) also point out that the high cognitive load caused by animation limits its use. Compared with animation, the timeline approach makes it easier for users to identify changes in the context of the evolution, as they are still visualized on the screen. Classic examples of time-series data visualization techniques such

as line graphs and bar charts focus on presenting univariate data sets. Visualization of multivariate time-series data is complex and requires special effort. A well-known technique dealing with this problem is the ThemeRiver (103; 104) developed for document visualization. It is an intuitive representation of the change between both variables and time steps. Placing node-link diagrams on a timeline is a simple way to visualize dynamic networks. Greilich et al. designed TimeArcTrees (105) for visualizing weighted, dynamic compound digraphs by drawing a sequence of node-link diagrams from left to right in a single view. When visualizing dynamic communities, line graphs are generally used to represent individual vertices, with color-coding presenting community identity. Works (106; 107; 108) use a timeline-based representation only. Reda et al. (108) focus on revealing the community structure implied by the evolving interaction patterns between individuals in dynamic social networks. Vehlow et al. (109) combine the visualization of community evolution and dynamic graphs into a single view.

Besides mapping the time to the 2D space, visualizing temporal data based on the space-time cube has become another popular method for dynamic graph visualization. Gatalsky et al. (110) presented some exploratory analysis of spatio-temporal data by using the space-time cube. Groh et al. (111) used an interactive three-dimensional model to visualize activity and social proximity in streaming event data during a given time period. SocioScape (112) is an interactive tool for the visual exploration of spatially referenced, dynamic social networks. The Wakame visualization system (113) can support discovering anomalies and comparing performance across multiple time series for multi-dimensional, spatio-temporal data. Bach et al. introduced Matrix Cubes (114)

to represent dynamic networks based on the space-time cube metaphor. However, the technique is only scalable to a small size of networks across short periods.

Although a large number of both neuronal connectivity visualizations and dynamic graph visualizations exist, attempting to integrate them is a challenge since the former focus on spatial data and the latter focus on non-spatial data. Maries et al. developed GRACE (8), a framework for the visual integration, comparison and exploration of correlations between spatial and non-spatial data sets. However, our goal is to visualize the dynamic change of multiple abstract attributes (non-spatial) of mouse brain networks (spatial) during a period of time. We found that none of the existing visualizations enabled all of the tasks defined in the section above. Thus, we introduce a novel visualization application to assist domain experts in exploring the dynamics of mouse brain networks at the neuron level without losing their spatial reference. Figure 18 shows an overview of the pipeline to process the mouse brain data.

4.3   Data and Task Analysis

4.3.1   Data Analysis

The data in the target domain include flavoprotein autofluorescence imaging data from mice brain specimens. This imaging modality has been shown to capture the activity of a living slice across a broader field of view than calcium imaging, and with high sensitivity and substantially higher spatial resolution than fMRI (100-200 microns) (115). A pixel contains roughly 100 neurons, and the image acquired at a specific timestep has dimensions of 172×130.
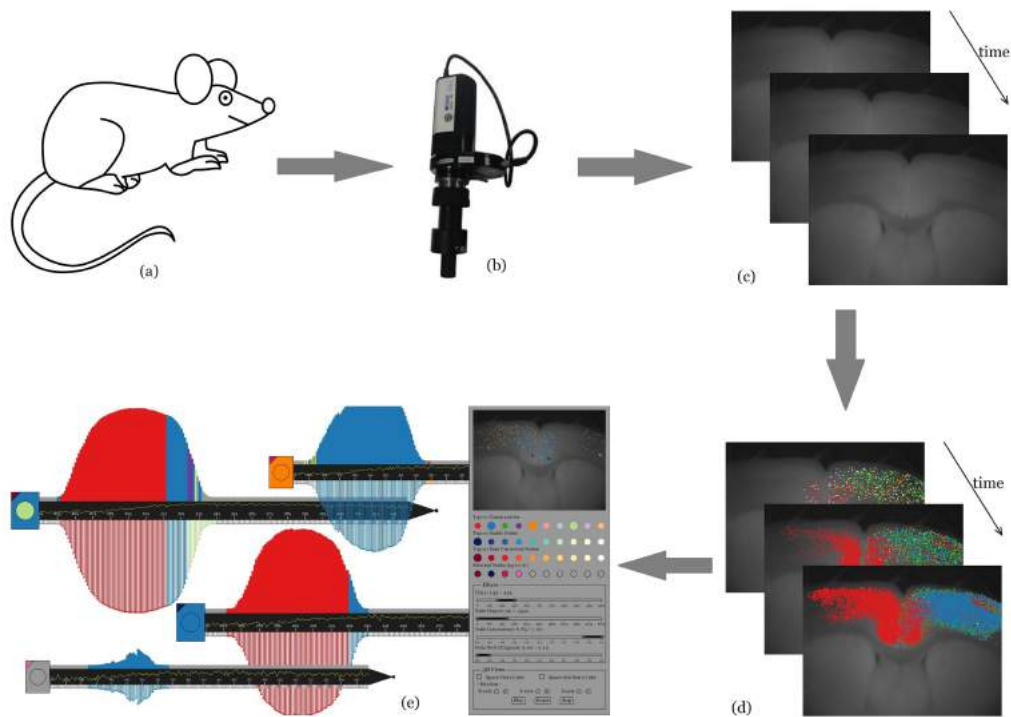
Figure 18. Flow diagram: (a) mouse brain, (b) camera, (c) time-series imaging data of mouse brain slices, (d) generate dynamic communities using SNA and (e) present the analysis results using the SwordPlots visualization.

We begin by computing all pairwise correlations for those 22,360 nodes (pixels) during a certain window using the Pearson product-moment correlation coefficient (PCC). This process leads to a weighted correlation network. A correlation coefficient threshold, empirically determined, is used to filter out noise in order to get a cleaner network. By repeating the computation of correlation while shifting the window one time step for each iteration over the entire timeline T, we obtain a time series of weighted correlation networks.

In the next stage, we find snapshots of functional clusters using the Louvain clustering method, followed by the application of CommDy for identifying dynamic communities. To better understand the neuron behaviors, we need to identify some characteristics of nodes in dynamic brain networks based on the use of CommDy. The SwordPlots application encodes the characteristics listed below.

Raw Data Attributes. In our brain network data, the coordinates of pixels show the real locations (regions in the brain) of the signals, and the pixel intensities represent how strong/weak the signals at the corresponding regions are. Therefore, the pixel values as well as their coordinates within the image are critical features for analyzing the network.

Node Degree. The node degree for many real networks can yield insight into mechanisms underlying system growth (116). Thus, there are significant benefits to visualizing the overall spatial shape of the degree distribution and its variation with time.

Home Community. Each individual has its own home community identification code to identify its membership. The individuals belonging to the same community have the same home community identification code. Neurons belonging to the same community have similar functionality.

Temporary Community. Each individual also has its temporary community identification code to identify the group where it stays at a certain time step. The individuals staying in the same community have the same temporary community identification code. Neurons have similar functionality to their current community, but tend to return to their original home community after a short time.

Consistency. Sometimes, the individual may have a stronger connection with members in other groups. Thus, it visits other communities for a short period of time, without changing its home community identification code. The consistency is a measurement of times when the temporary community of a node is the same as its home community during the entire activity period. For example, node 4 in Figure 17 belonging to the blue community visits the pink community at T3. Its consistency is 0.8. Neurons behave anomalously if they have low consistency.

Switchingness. Switchingness measures the percentage of times when a node changes its membership during the entire activity period. For example, node 3 in Figure 17 changes its home community identification twice, from T1 to T2 and from T4 to T5. Its switchingness is 0.5. Neurons with low consistency and high switchingness have unstable functionality.

4.3.2   Task Analysis

Based on in-depth discussions with domain experts from neuroscience and computational biology, we identified five main tasks that could be enabled using a visualization application.

- Task 1: observe multiple attributes of neurons over time to understand how neurons interact with each other (neuron behaviors).

- Task 2: interact with an individual node (i.e, a group of neurons), such as selecting interesting nodes at any region in the brain.

- Task 3: find relationships between the functional and spatial structures in the brain, to see whether two regions are close to each other or cross two sides of the brain when they are highly correlated.

- Task 4: identify patterns in how dynamic communities change over time in order to show functional connectivity in the brain.

- Task 5: generate hypotheses and make predictions about brain data sets.

4.4   Visual Design

The main goals of our SwordPlots application are to explore neuron behavior over time and discover the relationships among the resulting dynamic communities in both space and time. Our visualization technique uses a combination of different visual representations, and customized interactions to support exploration of the domain data. According to the requirements of the domain experts, we implemented our integrated multi-view visualization system with the following techniques. We choose to use timeline-based representations to observe neuron behaviors over time (Tasks 1 and 4). Animation may be helpful for Task 1 as well. Sufficient interactions are provided for our domain scientists to interact with individual nodes (Task 2). The space-time-cube-based representation is used to find relationships between spatial structures and network properties (Task 3). The SwordPlots application includes four coordinated views: the image control panel; a panel that shows recommendations and enables the filtering of data; the Space-Attribute cube; and a viewport that presents a number of SwordPlot visualizations to represent individual nodes. Each of these views is discussed below.
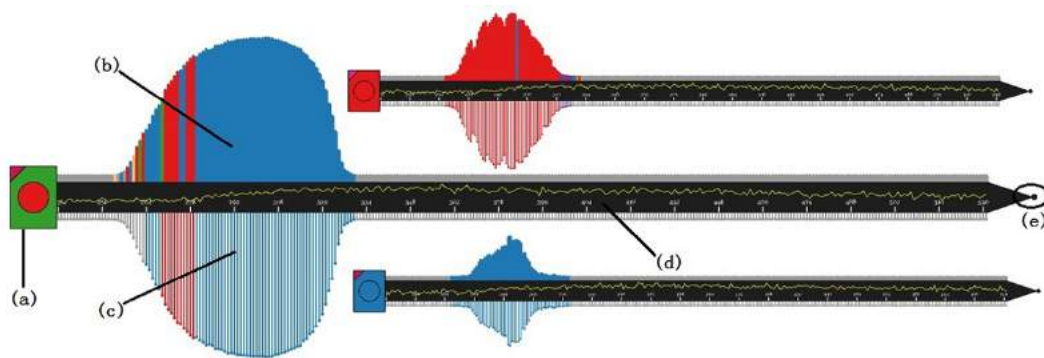
Figure 19. A SwordPlot includes multiple integrated components for representing time-series data and the community membership of a node in a dynamic network. This figure shows the main elements of a SwordPlot: (a) the sword pommel for representing the community identifications (the circle for home community and the rectangle for temporary community) of the node at the current time step; (b) the upper sword cross-guard for representing the node's temporary community identifications over time; (c) the lower sword cross-guard for representing the node's home community identifications over time; (d) the sword blade for representing the raw data (pixel value); and (e) the sword point used for changing the size of a SwordPlot.

4.4.1   SwordPlot

In our design, we evaluated different visual encodings to better represent important attributes of dynamic brain networks. Although the representation of circles and rectangles shown in Figure 17 is effective to visualize the community identifications, it has limitations: it does not scale well to more than a few nodes or a few time steps; it does not clearly track how the home and temporary community identifications of individual nodes change over a long time period; it lacks any spatial references for the nodes. Thus, we created a SwordPlot model (Figure 19), a timeline-based plot for each node in the entire network. A detailed explanation of the sword parts is shown in Figure 20. A SwordPlot provides an overview of a node's behaviors across an entire timeline (Task

1) and makes it easy to identify interesting behaviors at certain time steps, such as when it changes its memberships or temporarily visits other communities. In addition, displaying multiple SwordPlots makes it possible to compare different nodes at once to find patterns within the data (Task 4).We also use a brain image slice view to display the real locations (Task 3) of selected nodes represented by SwordPlots.
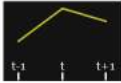
Sword Parts	Sword Pommel	Sword Body	Upper Cross-guard	Lower Cross-guard	Sword Point
Representation	Current Time Status	Raw pixel's value	Temoprary Community	Home Community	——
Visual Encoding					
Interaction	change location	——	——	display detail panel	change size

Figure 20. A detailed explanation of the various visual encodings within the different components of a SwordPlot, showing the change in community over time for a single node.

Sword Pommel. We still keep the visual representation of the community identification shown in Figure 17 using the color of the circle to represent the node's home community and the color of the rectangle to show its temporary community. This works well for a single node at a particular time step. The triangle at the top left corner is used

for mapping the node to its location in the image of the mouse brain slice. Dragging the sword pommel to move the SwordPlot to a new location avoids overlapping by others.

Sword Body. The sword body is used to display the raw pixel value represented by the yellow line during the selected time range. With the pixel values embedded, the relationships between neuron signals and their community identities/structures can be easily explored. This has the benefit of providing users with a view to see the correlations between these two attributes by visualizing the raw data pixel values. The time labels are drawn on the sword body as well. The time slider is used for zooming in/out on the SwordPlot, so that the neuroscientists can focus on the time range of interest.

Sword Cross-guard. The sword cross-guard is the bar-chart based representation. The heights of the bars in both the upper and lower cross-guards are mapped to a statistic, e.g., node degree in this article. Since the neuroscientists are only interested in the node degree distributions in time qualitatively, but not the exact values at certain particular time points, we decided not draw the axes for the cross-guards. As we discussed in the Brain Network Data and Characteristics section, each individual should have two identification codes: the home community represented by the color of the upper cross-guard and the temporary community represented by the color of the lower cross-guard. We display the same color above and below if the member stays in its home community. If the individual is visiting its neighboring community at time t then we display a different color above (home) and below (temporary). The reason why we use two types of bar charts to represent the upper and lower cross-guards is to distinguish the two community identifications in a more obvious way. Missing cross-guards mean

that the node degree is zero. In other words, the node is not active at the corresponding time. Clicking on the end point of the thin bar in the lower cross-guard pops up a detail panel showing some statistical information on the node at the corresponding time step. The statistical information includes some quantitative data, e.g., the pixel value, node degree, time point, etc.

Sword Point. To have more SwordPlots displayed and to avoid overlaps of multiple SwordPlots in the main view, the user is allowed to adjust the size of swords interactively by dragging the sword point. Dragging the sword point toward to the sword pommel decreases the size of the SwordPlot, while dragging it away from the sword pommel increases the size.
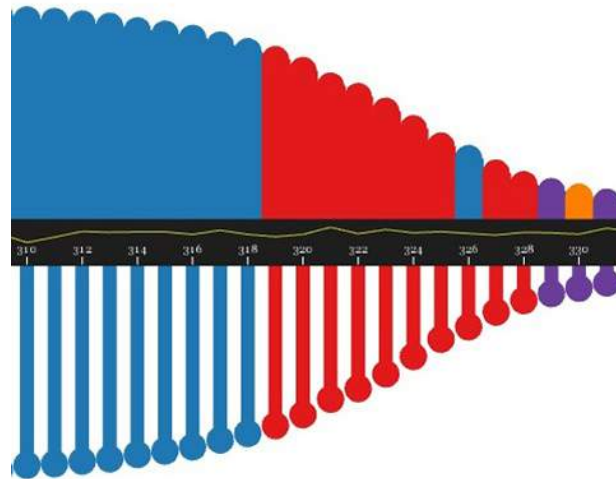
Figure 21. Transitions of the individual memberships of a node's home community (the lower part) and its temporary community (the upper part). The node changes its home community identity from blue to red, and then to purple. The node visits the blue community once when it belongs to the red community, and visits the orange community once when it belongs to the purple one.

Figure 21 is a demonstration of the sequence of changes of a node's individual community identifications. The node with its membership in blue stays in the blue community until the time 319, and then joins the red community with change of its home community identification as well. However, it comes back to the blue community at time 326 without changing its membership. At time 329, it leaves the red community and becomes a member of the purple community. In the next time steps, it visits the orange community and then comes back to its home community.

4.4.2   Image Control Panel

The image of a mouse brain slice shown in Figure 22 is used as a control panel to select an interesting node. The user clicks on a node (pixel) in the image to draw a SwordPlot (Task 2) in the main view and add it to the list of selected nodes as well. They click on the node in the selected list to remove its SwordPlot from the main view. Meanwhile, the image displays the corresponding locations of the selected nodes, which builds the connections between time-dependent attributes and the spatial information. In addition, nodes in the selected communities are also drawn on the image (Figure 22). Playing animation helps the domain scientists to observe the community structures at different time steps and to track the evolution of communities. Theoretically, the SwordPlots for all nodes can be displayed in the main view simultaneously. However, due to the limitations of the distinguishable colors and human's working memory capacity, the number of selected nodes will not exceed a certain value. We discussed this with our domain experts and decided to use ten as the maximum number.
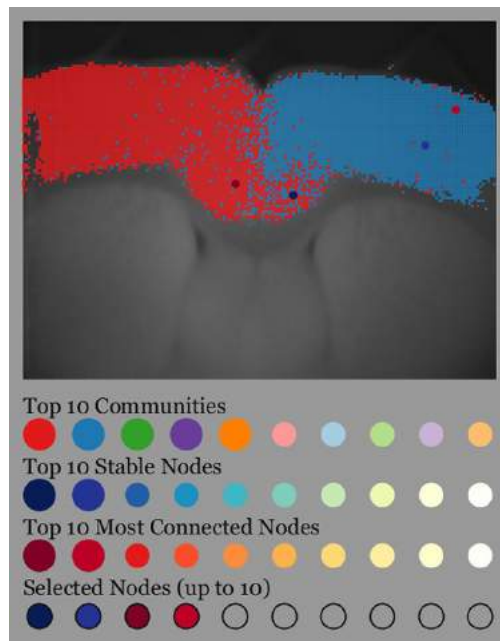
Figure 22. The image control panel is used for selecting interesting nodes and displaying community structures and identities. The recommendation panel includes the top ten communities, top ten stable nodes and top ten most connected nodes. The circles with larger sizes indicate that the corresponding communities or nodes are selected. The selected nodes are also added into the list in the last row.

4.4.3   Recommendation and Filtering

Besides giving the ability for the user to explore some interesting regions randomly, our visualization provides recommendations as well to make the exploration more efficient. The three recommendations are calculated based on three characteristics: the community size, the node stability and the node degree. The community size is based on the total number of members the community has during the entire time period. The stability is calculated by two factors: consistency and switchingness. The higher consistency and lower switchingness the node i has, the more stable it is. The most connected nodes have
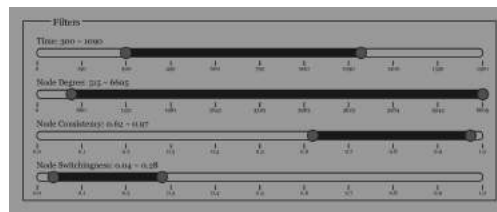
Figure 23. Four filters used to change the ranges of time, node degrees, node consistency and node switchingness.

high accumulated node degrees during the entire time period T. Again, as we indicated in the subsection above, we chose to use the top ten recommendations due to the scalability of the color encoding. We used ten categorical colors for the top ten communities from the D3 color ordinal scales, and used ColorBrewer (colorbrewer2.org) for the top ten stable and top ten most connected nodes. The user clicks on the circle buttons of the top ten communities to draw all of the nodes in the corresponding community. In Figure 22, the top five communities are selected to display on the image of a mouse brain slice. However, only the top two are being observed at the current time point. Figure 22 also shows the two most stable nodes with larger blue dots and the two most connected nodes with larger red dots. Filter sliders (Figure 23) are used to detect interesting features including time, node degree, node consistency and node switchingness. The main contribution of the filtering operation is to offer the user a way to explore the relationships between different attributes. For example, nodes that are located near the boundary of the activation area may have higher switchingness than nodes in the

center. Nodes with higher degree have slightly lower consistency compared with nodes with lower degree.

4.4.4   Space-Attribute Cube

We create a 3D model to represent a network statistic with the spatial reference in the brain networks. This is similar to but has important differences from the idea of the Space-Time Cube. According to the technique of the Space-Time Cube, the vertical dimension represents the positions of an object at different moments in time. In our 3D model, nodes can be represented in a cube as dots placed vertically according to one property value of the dynamic networks, such as node degree. We call this 3D model the Space-Attribute Cube. Color in the Space-Attribute Cube, like the SwordPlot, is used to represent the community identifications. The Space-Attribute Cube reveals the distribution of node degrees of all active individuals in space, and how the distribution is related to the community structure. The left cube in Figure 24 demonstrates such a case, with the nodes belonging to the same community (red) located in different regions of the brain (a)-(c) with very different distributions of node degrees. One explanation is that the red nodes in (a) have strong connections with the nodes at the top of (b) and (c), but few connections with nodes that are also located in (a).

To overcome the problem of occlusion in 3D representations (perspective projection) of data, the visualization provides three options for the user to rotate the Space-Attribute Cube along the X, Y, and Z axes in either the clockwise or the anti-clockwise direction. To discover how the property measured in the third dimension changes over time, animation is applied to the Space-Attribute Cube. The animation starts playing on pressing
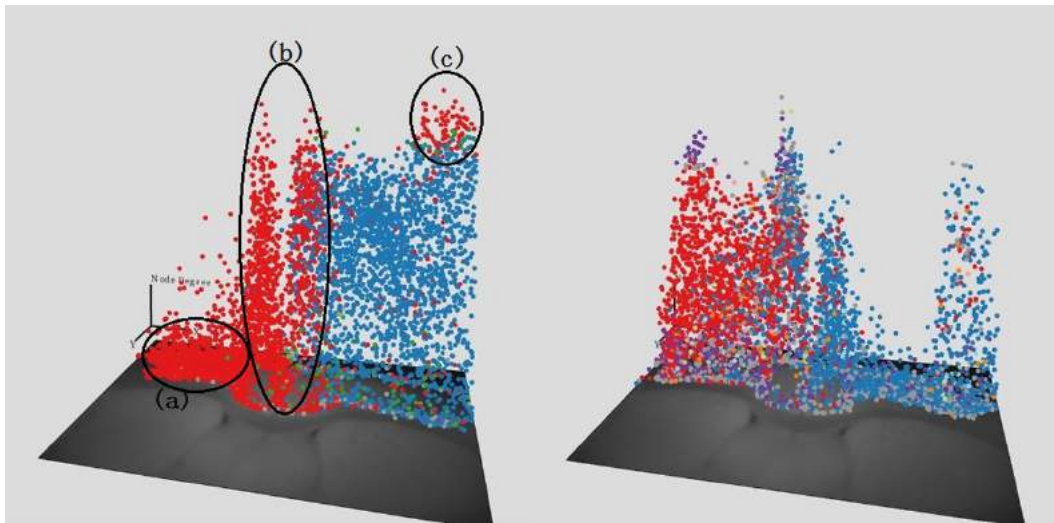
Figure 24. Space-Attribute Cubes at two time steps. Each individual node is represented by a colorful dot. The color indicates the community that the node belongs to. The X and Y coordinates show the real location of the node in the brain. The height of the node represents its corresponding node degree.

the play button and pauses on pressing the pause button. Once the animation is paused, the user can press the left or right arrow key on the keyboard to advance the animation one frame forward or in reverse. Animation used in the Space-Attribute Cube provides a spatial form representation for multivariate time-dependent data and a temporal representation for multivariate spatial data. Figure 24 presents different distributions of node degrees at two time points. The left cube at earlier time indicates that the right side of the brain has stronger connections than the left side, while the right cube at later time shows the opposite situation.

4.5   Evaluation

We used the imaging data of $172 \times 130 \times$T, as described in the Brain Network Data and Characteristics section, to evaluate the SwordPlots application. Here, T is the number of frames, which is different (between 1000 and 2000) for different data sets. We build correlation networks based on these 22,360 pixels per frame captured during T frames, and then apply the CommDy method, as discussed in the Community Analysis section, to the time series of these networks.

4.5.1   Case Study I: Coronal Slice Preparation

For the first case study we looked at a coronal slice preparation of a mouse brain. The data contain mirror image connectivity where the two cerebral hemispheres are connected by a single easily manipulable conduit of information flow: the corpus callosum. One hemisphere of the brain is stimulated, and the activation of correlated areas in the contralateral hemisphere is measured. To assess whether the dynamic communities identified by CommDy correspond to known neuronal networks, after trans-hemispheric networks are characterized, the corpus callosum is cut using microscissors. We then optimize the CommDy analytical parameters based on known changes to the network induced by callosotomy. All data are collected at 70 frames per second using a 2.5X objective and a Retiga Exi camera with 8×8 hardware binning and StreamPix software for image collection.

Figure 25 shows, in an in-vitro preparation, an example of the behaviors of two nodes at analogous locations across the two hemispheres in (a) the pre-cut and (b) the post-cut conditions. From the SwordPlots in Figure 25(a), we find that the two nodes

have similar trends of raw pixel value, similar distributions of node degrees in time, and the same community identifications (red) during most of the time. This indicates the similar behaviors of the two nodes across the two hemispheres. From the observation of the image slice view using animation, we see the partially symmetrical structure of the red community in both hemispheres where the two nodes are located. We then find more nodes with such similar patterns in that region using SwordPlots, illustrating the connection between the left and right sides of the mouse brain. The Space-Attribute Cube is used to explore the community structure, which reveals some information that cannot be retrieved from the image slice view. For example, we may not be able to see any difference between two time steps through the image slice views, but we can obviously see the variation in node degree using the Space-Attribute Cube.

After the corpus callosum is cut (Figure 25(b)), we can easily see that the activation in the two hemispheres happens at different times via the SwordPlots, and moreover that there is no community across the two hemispheres. Since there is a long period of silence between two activations, animation is not an efficient way to identify when the activation occurs. However,we can target the interesting time periods directly by using the SwordPlot view. The views of the Space-Attribute Cube also show an interesting phenomenon, where a community (purple on the right side of the brain or green on the left side) has a distinctly different distribution in node degree, which causes the ring structures shown in the image slice views. One explanation is that nodes in the outer ring have strong connections with the nodes in the upper right region, but few connections with each other.
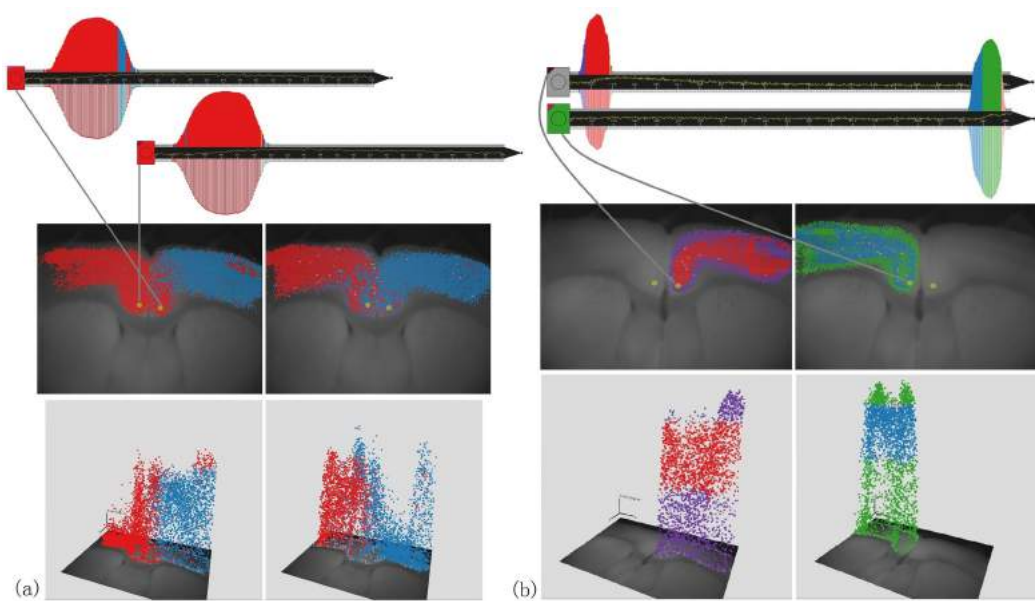
Figure 25. The behaviors of two nodes in the coronal slices of the mouse brain (a) before and (b) after the corpus callosum cut are displayed using the SwordPlots, with the snapshots of 2D and 3D views at two time steps.

4.5.2   Case Study II: Impact of Aging and/or Peripheral Hearing Loss on AC Activity

For our second case study, we investigated the impact of aging and/or peripheral hearing loss on auditory cortex (AC) activity. We examined network activity in the AC in slice preparations (115; 117) taken from young and aged mice. We used the auditory thalamocortical slice, which contains large areas of the AC. For this study we had two main hypotheses: 1. Aging is associated with diminished network associations within the AC.; 2. Changes in network activity with aging are caused, at least in part, by

peripheral hearing loss. Using the SwordPlots visualization tool, we were able to assess the likelihood of the validation of these hypotheses.
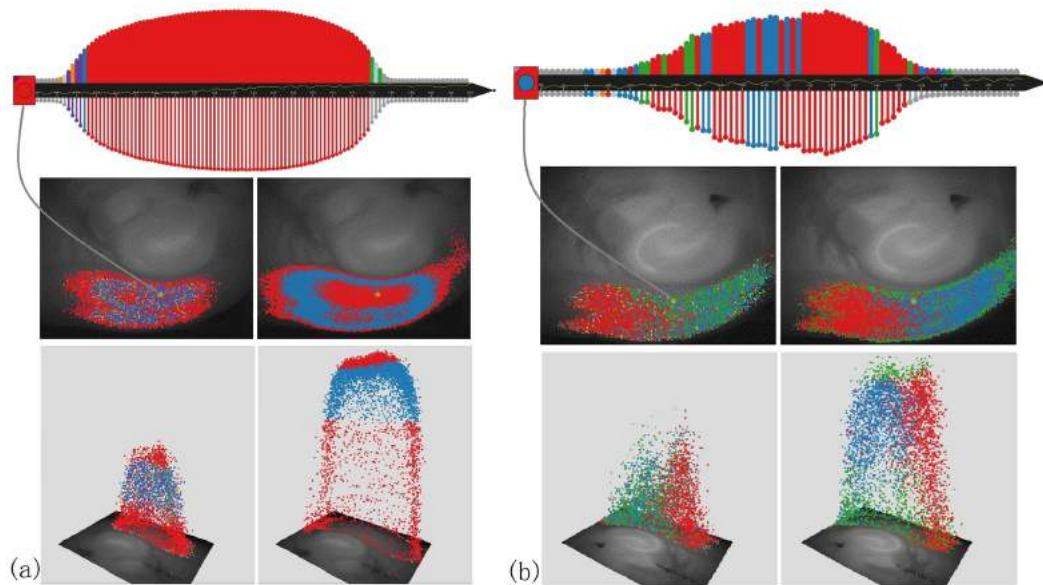
Figure 26. The behavior of one node in the auditory thalamocortical slices of (a) a young mouse and (b) an aged mouse is displayed using the SwordPlots, with the snapshots of 2D and 3D views at two time steps.

In a different animal, using an in-vivo preparation, Figure 26(a) shows the behavior of a node in the AC in a young mouse. Its SwordPlot indicates its consistency within the red community during the active period. We can also see an outer ring structure in this data set. However, the SwordPlot of a node at a similar location in Figure 26(b) of an aged mouse shows frequent changes in colors within the sword cross-guard area

as well as inconsistencies between the upper and lower cross-guard colors, indicating a high switchingness and a low consistency. This demonstrates the instability of the aged mouse brain.

We also found contrasting community structures in the young and aged mice. In the young mouse, the red community is mainly in the center of the AC surrounded by the blue one, while in the aged mouse the red community stays in the left part of the AC and the green/blue one is located on the right side. The views of the Space-Attribute Cubes also show that the young mouse has only one peak of the node degrees, located in the center red community, but the aged mouse has two peaks that are apparent on the left (the red community) and right sides (the green/blue community). These differences between the young and aged mice, shown clearly in the SwordPlots application, provide an initial validation for our hypotheses.

4.5.3   Informal Feedback

To further evaluate our approach, we also interviewed two graduate students who each have extensive experience in visualizing functional neuroimaging data. The first is a graduate student with over five years' experience working with MRI machines in a Speech and Hearing laboratory at a research university. The second is a graduate student in computer science who is working with a laboratory in the Department of Psychiatry at another university to develop techniques to analyze properties of the human connectome. After a brief introduction to the system and an explanation of its features and primary visual encodings, we conducted a cognitive walk-through, asking each of them to carry out tasks using an example data set. During the walk-through we encouraged them to

ask questions and to give us feedback. We were interested primarily in the qualitative response of experts who were not familiar with the specific design goals of the technique, and to see what its utility might be to experts in a different area, but who make use of similar data.

The feedback was generally positive; they both stated that they liked using the system and that they found the visualization approach interesting and potentially useful. One of them said that the SwordPlot representation provides an "easy way" to understand the "changing of community identifications," and that it "packages a lot of information well." However, they were split on the novelty of SwordPlot's visual encodings. One said that he had not seen anything like it, while the other said that "the combination of timelines is not novel," but admitted that it did seem like it would "give an intuitive and effective way" for him to explore his data. One participant mentioned that he liked the way in which raw data were linked to the community identifications and node degrees, and noted that it could be useful for revealing potential correlations between those attributes.

In addition to these general comments, each of the participants provided more detailed feedback about some of our specific design choices. One of the participants wanted to know why we limited users to the selection of a maximum of ten simultaneous Sword-Plots. We acknowledged that, due to limits of screen space, there is a trade off between the ability to show detailed information and displaying more models, which could enable a user to find similar patterns across multiple nodes. The participant then agreed than more than five or six would probably provide an overwhelming amount of information

and would be difficult to perceive. The other participant said that he would rather have more detailed information provided as he could not imagine the need to ever focus on comparing more than two nodes at the same time. One of the participants liked the recommendation operation, as it provides a useful starting point for investigating the data. He told us that "it would be nice to have the top ten least connected nodes available as well," and that he would like to have the ability to choose a node with a specific rank. The other participant told us that he liked the ability to apply a combination of filters, which seemed like a helpful way to discover patterns, for example, to see how certain nodes are sensitive to a certain feature given a fixed range of other features.

Both of the participants indicated that they found the 3D view informative. One of them told us that it would be "totally useful" to see the coordinated image slice view and SwordPlots together. Both of the participants opened the 3D view of the Space-Attribute Cube after they noticed interesting patterns in the image slice view. One told us that they would like to have the ability to interact further with the 3D view of the Space-Attribute Cube. He also wanted to know how the Space-Attribute Cube, and our tool in general, would handle volumetric data rather than image slices. Overall, both of the experts told us that they saw this as a compelling tool overall, and indicated that they would be interested in trying it out on their own data sets.

4.6   Discussion

The two case studies indicated that our interactive multiview visualization application with animation can be of significant help in the exploration of neuron behavior within dynamic communities of mouse brain networks. While the individual visual encoding is

not novel, the combination of visual encoding in a tool to integrate spatial and nonspatial features along with the application of dynamic network analysis to neuroscience research has not been performed before. Our use of the multiple coordinated views through interactive filtering and color-coding provided us with new insight into this domain problem. In addition, it also steered our investigation, and allowed us to identify changes in analysis parameters that aided our visualization of these data sets.

In terms of limitations, although the stable nodes are disjoint from the most connected nodes, the colors for these two types of nodes toward to the right in the control panel are not easy to distinguish. Additionally, the select nodes might not be visible if their colors are very similar to the communities that surround them. Overall, the use of color in the image control panel needs to be reconsidered.

In conclusion, we presented SwordPlots, an interactive application for visualizing neuron behavior within dynamic communities of brain networks. The design of the SwordPlots application coordinates multiple visualization approaches, including the use of animation to show change over time, the introduction of a novel 3D representation called the Space-Attribute Cube, and the SwordPlot itself showing details about individual neuron behavior. The SwordPlots provides the visual exploration at detailed neuronal levels while showing the community structures in the brain slice as an overview of a single brain activation. However, a selection of a subregion in the brain that could benefit the multi-scale visual exploration is not enabled. Additionally, SwordPlots is developed for the visual exploration of dynamic brain networks from a single activation, which is not sufficient for comparing multiple networks as our domain experts run multi-

ple experiments with different subjects. In the next chapter, I will present a novel visual

representation to build the gap between the exploration at an overview and a detailed

level, and propose a side-by-side view paradigm to compare brain activity networks.

CHAPTER 5

REMBRAIN: EXPLORING DYNAMIC BIOSPATIAL NETWORKS WITH MOSAIC-MATRICES AND MIRROR GLYPHS (PREVIOUSLY PUBLISHED AS CHIHUA MA, FILIPPO PELLOLIO, DANIEL A. LLANO, KEVIN AMBROSE STEBBINGS, ROBERT V. KENYON, AND G. ELISABETA MARAI. (2017) REMBRAIN: EXPLORING DYNAMIC BIOSPATIAL NETWORKS WITH MOSAIC MATRICES AND MIRROR GLYPHS, JOURNAL OF IMAGING SCIENCE AND TECHNOLOGY, 61(6): 060404-1060404-13)

5.1   Introduction

This chapter presents the continuing work of visualizing dynamic brain networks that places additional focus on providing a multi-scale comparative visualization, while still integrating spatial and non-spatial features. Advances in imaging technology allow, at increasing pace, the comparative investigation of functional connectivity dynamics at multiple scales, both at the temporal level (time-series, trials) and at the spatial level (neurons, neurons grouped in pixels, regions of interest). In computational neuroscience and computational systems biology, each imaging snapshot captures one activation pattern in the temporal behavior of a biological system. The connectivity is then extracted from these images, in the form of networks with large numbers of nodes (over 20,000).

Next, computational models are employed to calculate the functional network dynamics. In order to study the mechanisms of disease or aging, the process of imaging and modeling is performed repeatedly over multiple subjects, specimens or conditions, leading to a rich tapestry of spatio-temporal imaging and computing data that need to be analyzed. Visual analysis of such complex neuroimaging data can help domain experts understand temporal features along with their spatial references.

In this Chapter, we present the design and implementation of RemBrain, a novel visualization tool for the comparative analysis and exploration of dynamic brain activation networks at multiple scales. RemBrain (named after the intrepid Pixar rodent Remy) is a multi-scale web-based application that supports the tracking of temporal network behaviors. Responding to the novel data characteristics above, RemBrain integrates interactive, multi-scale 2D visualizations of imaging data; displays network connectivity data for each activation snapshot; captures the temporal behavior of a subregion of nodes using novel encodings; and supports pairwise visual comparison of multiple activations. This work, a follow-up to the award-winning Swordplots (14) (published in JIST), enables the exploration and comparison of different activations at multiple levels in dynamic biological networks.

5.2   Related Work

5.2.1   Brain Connectivity Visualization

Many techniques exist for visualizing brain connectivity at either macroscopic (region level) (16; 17; 84; 89; 90) or microscopic (neuron) scale (93; 18; 21; 22; 23). In this work, we design a neural encoding inspired by the sword-plots of Ma et al. (14). However,

to the best of our knowledge, no other visualizations exist for multi-scale biological connectivity data.

Our data further blends spatial and non-spatial features. Marai (**?**) identified two prevalent paradigms for integrating spatial and non-spatial features: overlays and multiple linked views. In neuroscience studies, an overlay approach (16; 118; 119) is commonly used when the non-spatial feature represents only functional connections. However, as the non-spatial data becomes more complex (activation levels, connectivity, clusters, dynamic characteristics, other statistics), the linked-view paradigm (20; 21; 22; 23) becomes the default choice. Nowke et al (18) uses a hybrid approach that consists of both overlays and linked views. We follow a similar hybrid approach to support the exploration of dynamic biological networks.

5.2.2 Dynamic Network Visualization

In static network visualization, the most common visual representations are node-link diagrams (120) and matrix-based visualizations (121). Several projects (122; 123) use a hybrid approach that combines both representations.

Dynamic networks and graphs are usually visualized using either animation or a timeline-based representation (94). Several projects (97; 98; 96; 124) use animation to represent networks with temporal components. To display dynamical changes of networks into a single static view, Greilich et al (105) placed a sequence of graphs onto a timeline. Several other projects (106; 107; 108) use timeline-based representations to visualize the evolution of communities in dynamic networks. Rufiange and McGuffin (125) presented a hybrid approach for visualizing dynamic networks. In addition to mapping

the time to the 2D space, Bach et al. (114) developed a Matrix Cube representation based on the space-time cube metaphor. The Matrix Cube shows the network structure using the 2D matrix and maps time to a third dimension. However, their technique is only scalable to networks that consist of a few nodes across short periods.

While, as shown above, a large number of visualization techniques exist for static brain connectivity, as well as for dynamic non-spatial networks, to the best of our knowledge this novel domain is the first to require visualizing and integrating both types of data.

5.2.3 Multi-scale and Comparative Visualization

In the visual analysis of neuroscience data, VisNEST (18) and NeuroLines (22) integrate data at macroscopic level with microscopic level. We similarly adopt a multiple views approach for different levels using both focus+context and details-on-demand. Visual comparison of brain spatial-nonspatial data is a relatively new research problem in neuroscience. Only a few tools (8; 126) can be found. Maries et al (8) introduced a comparative framework for mining brain geriatric data. Lindemann et al. (126) presented a comparative visualization system that explicitly encodes changes of brain tumor segmentation volumes in shape and size before and after treatment. Outside the application domain, Gleicher et al. (10) proposed a general taxonomy that groups visual designs for comparison into three categories: juxtaposition (side-by-side), superposition (overlay) and explicit. Because of the complexity of our data, in our approach we use side-by-side linked views.

5.3   Data and Task Analysis

5.3.1   Data Analysis and Processing

The input data consists of flavoprotein autofluorescence imaging data collected, in this case, from mice brain specimens captured in the TIFF format. A pixel contains roughly 100 neurons, and the image acquired at a specific timestep has dimensions of 172×130, leading to a file size of 24KB. One activation cycle lasts about 100 timesteps. Figure 27 (a) shows an example of a time series data collected from mouse brain slices. The imaging data is processed in three steps: 1) infer a network model, 2) perform dynamic community analysis, and 3) compute community metrics. The first two steps have been introduced in the last Chapter, so only the computation of community metrics will be discussed below.

Metrics Computation The dynamic community characteristics are used to generate metrics that summarize the behavior of active nodes. CommDy quantitatively describes the characteristics of the inferred networks, at both node and structural level, based on network analysis theory (127). We use 10 relevant metrics to describe the interactions between nodes. These metrics include the average time spent by a node in a community, the number of jumps across communities executed by a node, the fraction of node peers who were its peers in the previous time step and so on. All these characteristics are normalized to a value within the range of 0 to 1. Table II summarizes the full list and definition of the node metrics. Based on the results produced from the current datasets, the number of active nodes varies with different activations of different subjects from approximately 5,000 to nearly 10,000.
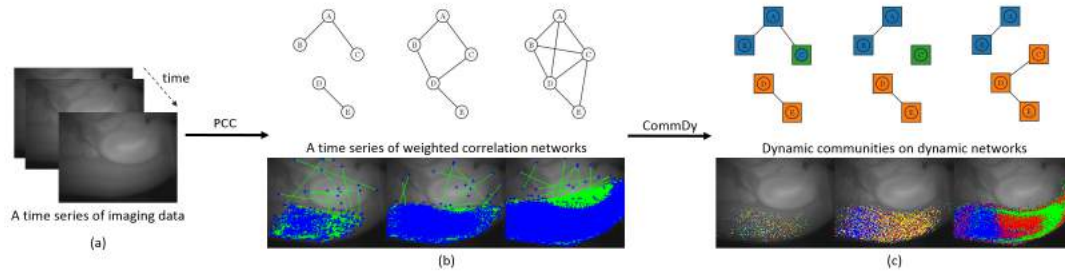
Figure 27. Data processing. (a) Neuroscientists collect time series of biological imaging data (in this instance, mouse brain slices). Bright spots in each image indicate activated (firing) neurons. (b) We use the Pearson correlation method to construct, from these images, an equivalent time series of correlation networks (top, abstraction). The correlation networks (bottom, image overlay) correspond to three time steps; blue dots encode the active nodes and green edges encode the links between correlated pairs. (c) We apply CommDy to these networks to infer dynamic communities (top, abstraction). Four communities (blue, green, orange, red) are overlaid on mouse brain images, for three time steps, in the bottom image.

5.3.2 Task Analysis

Based on several interviews with a domain expert, we identified the following tasks for the comparative analysis of brain activations, and in particular for understanding how aging impacts the auditory cortex (AC) of mice:

- **Task 1**: Explore the community spatial distributions at multi-scale. Brain imaging data contain thousands of nodes. Neuroscientists need to get an overview of the entire dataset, but also to observe a subregion or even an individual node in detail.

- **Task 2**: Track temporal changes at multiple levels. Be able to observe the evolution of communities over a user-defined time window, compare the temporal

behaviors of nodes in the same subregion, or track the behavior of a particular node across the entire time period.

- **_Task 3_**: Explore relationships between functional connections and spatial structures. An interesting and expected finding would be that specific nodes located in different regions of the brain have similar temporal behaviors.

Static/Dynamic	Node Attributes	Attribute Descriptors
Static	Spatial Location	Coordinates in the grayscale image of a brain slice
Dynamic	Signal Value	Pixel (node) intensity value: 0 255 (8 bit)
Dynamic	Node Degree	Number of connections a node has
Dynamic	Home community	The community a node belongs to
Dynamic	Temporary community	The community a node currently visits
Static	Observed	Number of time steps a node is active or observed (normalized by the entire time steps)
Static	Time span	Average span of the communities (the last time step minus the first time step of the community's existence) with which an individual is affiliated (as a member or absent)
Static	Switching	Number of community switches made by an individual (normalized by the number of time steps an individual is observed)
Static	Absence	Number of absences of an individual from a community (normalized by the number of time steps an individual is observed)
Static	Visiting	Number of visits made by an individual to another community (normalized by the number of time steps an individual is observed)
Static	Homing	Fraction of individual's current peers, at each time step, who were peers in the previous time step
Static	Avg group size	Average size of group of which an individual is a member
Static	Avg community size	Average size of community of which an individual is affiliated (as a member or absent)
Static	Avg community stay	Average number of consecutive time steps an individual stays as a member of the same community (normalized by the number of time steps an individual is observed)
Static	Max community stay	Maximum number of consecutive time steps an individual stays as a member of the same community (normalized by the number of time steps an individual is observed)

TABLE II

DATA DESCRIPTORS FOR DYNAMIC BIONETWORK ANALYSIS.

- ***Task 4***: Compare the differences in temporal and spatial behaviors between young and aged animals at multiple levels.

These tasks map to three groups in the visual data analysis taxonomy (6): Explore: ***Task 1, Task 2, Task 3*** and Compare: ***Task 4***.
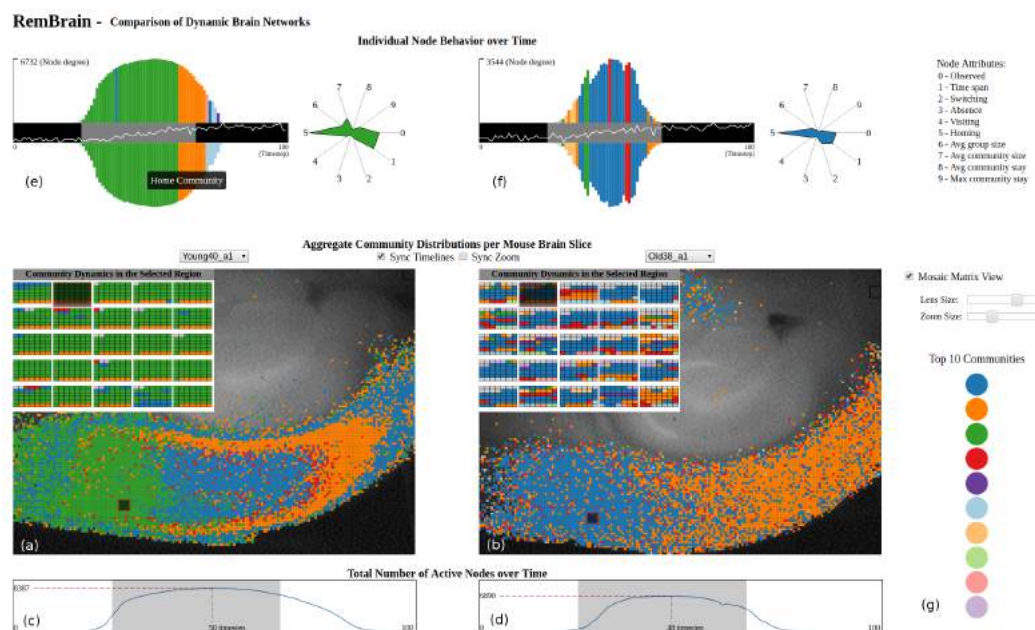
Figure 28. RemBrain implements a visual approach for the analysis of spatiotemporal brain network data. Two aggregate panels (a) and (b) encode the spatial distribution of neuron communities in mouse brains, overlaid with the medical imaging data. Mosaic-Matrix views (top left of panels) encode temporal changes in a selected subregion. Two timeline views (c) and (d) show the number of active nodes over time; (c) shows that the activation has 8387 active nodes at timestep 50, and (d) shows the activation has 6890 active nodes at timestep 48. Two mirror glyphs and Kiviat diagrams (e) and (f) allow tracking dynamic changes over time at a single node level. A control panel (g) enables filtering of node communities; colors are mapped to community IDs.

5.4 Visual Design

The spatio-temporal datasets and comparison tasks captured above are particularly complex. Furthermore, they feature a mix of spatial and nonspatial data, and the experts lack familiarity with complex visual encodings. Because of these combined reasons, we follow a coordinated multi-view top-level design, which has been shown to assist in visual scaffolding (**?**). In this design, a set of multiple views at different scales provides guidance to the domain expert when exploring the data. In addition to the exploration tasks, pairwise comparison is supported by side-by-side views (*Task 4*).

Figure 28 shows the interface of RemBrain, which consists of four main visual components: 1) an overview spatial panel (Figure 28 (a) and (b)) that nests subregion temporal information through a Mosaic-Matrix encoding; 2) an individual behavior panel (Figure 28 (e) and (f)) that includes a novel Mirror glyph to display in detail the dynamic attributes for a particular node, and a Kiviat diagram for the summarized characteristics of the corresponding node; and 3) a timeline representation (Figure 28 (c) and (d)) that controls the spatial panel and the Mosaic-Matrix view. These multi-scale views are linked through interaction. Two sets of each panel are placed side-by-side for visual comparison.

The side-by-side comparison design, in conjunction with the multi-scale views, supports *Task 4*. The slice-based panel displays the community distribution in space and thus explicitly supports *Task 3*. The individual behavior view enables exploration at the level of individual nodes (*Task 2*). The Mosaic-Matrix enables exploration at the level of sub-regions (*Task 2, Task 1*). The three views work together to support *Task*

1 through **Task 3**. Below we describe in detail each visual component. The web-based visualization tool is implemented in JavaScript using the D3 data visualization library.

5.4.1   Aggregate Slice Panel

The slice-based view shows the community distribution map overlaid upon the brain slice image. In this distribution map, nodes are color-coded by their Home community ID. To enable multi-scale temporal analysis (**Task 2**), instead of displaying the community information at a single time point, we aggregate over a user-selected time period and color the active nodes by their most common community during that time period. We assign to each of the 10 largest communities a unique color (Figure 28 (g)) from a qualitative colormap from ColorBrewer2.org. Nodes colored in gray are either inactive or belong to a community not in the top ten. A control panel filters which communities are shown. The view is automatically updated according to the selection in the timeline widget (Figure 28 (c) and (d)).

5.4.2   Individual Panel: Mirror glyphs and Kiviats

The aggregate slice-based view shows the spatial distribution of communities. However, it is also important to display the temporal distribution of communities, along with other temporal attributes. To this end, the individual behavior panel combines in a novel design time-dependent numerical and categorical data. This detail view allows users to explore in detail the dynamic behavior of individual nodes through a timeline-based representation. The panel (Figure 28 (e) and (f)) integrates a mirror glyph for analyzing the temporal node data and a Kiviat diagram for visualizing multiple summarized characteristics.

5.4.2.1   <u>Mirror Glyph</u>

The mirror glyph which is a simplified version of the SwordPlot supports tracking the characteristics of a particular node over time (***Task 2***). These dynamic characteristics include raw signal values, node degrees, and two community identification codes over time, the Home community and the Temporary community (Table II).
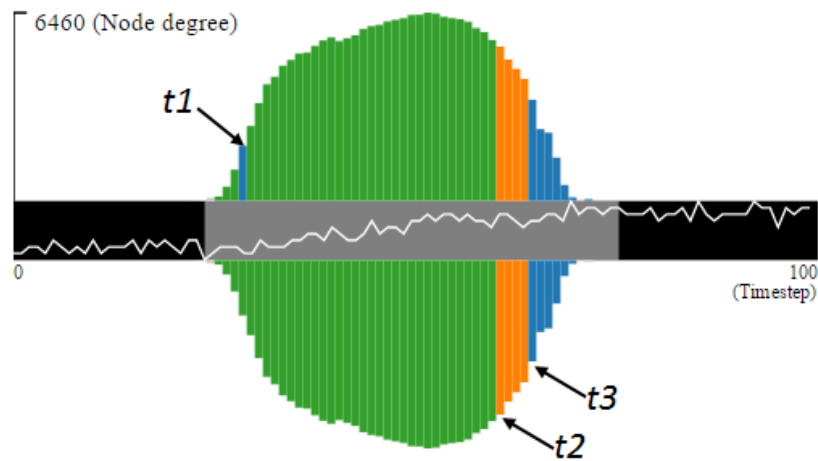
Figure 29. Mirror glyph showing a node that belongs predominantly and consistently to the green community, as its signal increases over time, without much visiting or switching. The only temporary visiting event happens at t1, when the node briefly visits the blue community. Two switching events happen at times t2 and t3, when the node joins the orange community, followed by the blue, just as the node signal is about to peak (middle trunk). The node degree (chart height and content) is almost symmetric: the Temporary community (upper chart) almost mirrors the Home community (lower chart).

Each mirror glyph (Figure 29) has three components: middle black trunk, upper bar chart, and a mirror-like lower bar chart. The height of the upper and lower bar chart represents the node degree over time, because domain experts indicated the degree evolution over time was the important quality in this context, next to the Home and Temporary IDs. The upper chart color encodes the Temporary community while the lower chart color encodes the Home community. The color of bars in the charts indicates the community ID of the node over time. Mouse interaction further shows the type of community represented by a bar. While the upper and lower charts are often almost-symmetric (hence the "mirror" aspect), they can also be asymmetric. Frequent horizontal color changes in this composite glyph indicate node instability. Similarly, a vertical asymmetry between the upper and lower charts indicates high instability.

The line plot in the middle black trunk of the mirror glyph encodes the variation in raw signal intensity over the activation period, from 0 to 100, which is the maximum number of timesteps in our datasets. The trunk's gray segment highlights the user-selected time period. The vertical axis indicates the maximum node degree during the entire timeline. Figure 29 illustrates how this composite glyph can capture dynamic node behaviors. The end mirror glyph result captures a high temporal resolution of the node behavior. The spatial location of the corresponding node is highlighted in the aggregate slice-base view.

We converged towards this streamlined mirror glyph design through parallel prototyping with multiple iterations; some prototypes were completed on paper, and some in software. Driven by the experts' preference for clarity over compactness, the glyph

design converged towards this dual layout, as opposed to stacked graphs or a non-flipped layout. To this end, we note that the data itself was extremely complex and that detecting brief community visits was relevant to the tasks. Similarly, the raw signal intensity and its evolution over time in relation to the community distribution and degree were task-relevant. In this case, the trunk design favored the signal charts that the experts were familiar with.
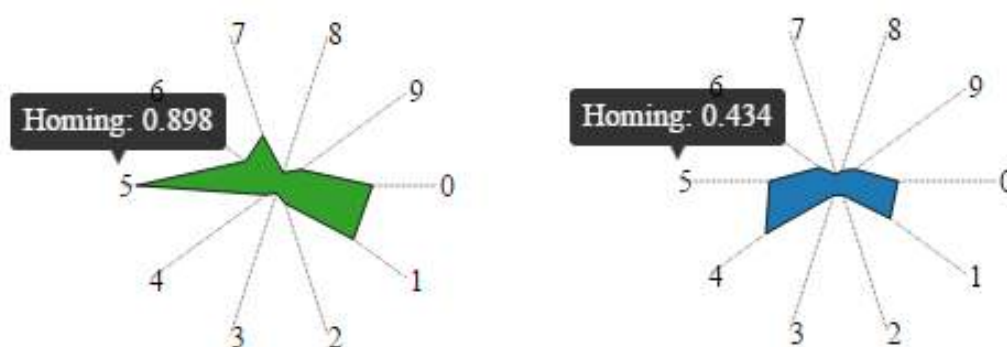
Figure 30. Kiviat diagrams for two nodes that reside most time in the green, respectively blue community. The Kiviat shapes indicate that the green node has longer activation duration, stays in particular communities for longer periods, and is more consistent with its Home community. Conversely, the blue node switches its Home communities and visits other communities more often. Note the details-on-demand and index indirection dictated by real-estate constraints.

5.4.2.2   Kiviat Diagram

To encapsulate the 10 summarized attributes (e.g., observed, time span, switching, etc.) shown in Table II, either parallel coordinates or star plots are natural choices. However,

in the parallel prototyping stage, PCPs were overruled due to low visual literacy among the domain experts and to their less compact appearance. The experts' specific goals, in this case, were detecting significant differences and similarities in the data, to be later analyzed quantitatively (as opposed to precise visual comparison). In fact, one of our two experts remarked on the imprecise nature of measurements computed at the single node scale. Because the experts further favored legibility over precise comparison, we converged towards a star-based Kiviat diagram encoding (Figure 30), as opposed to a line-based star plot, or simple dots on axes. Nevertheless, the Kiviat representations can be optionally superimposed (with transparency) to better support comparison. Each axis of the Kiviat diagram represents one of the ten node metrics. The most common community of a node is encoded in the color of the Kiviat diagram.

5.4.3   Mosaic-Matrix View

The aggregate slice-based view shows the community spatial distribution at an overview level with **high spatial** resolution, but **low temporal** resolution. On the other hand, the individual panel shows the community temporal distribution per node with **high temporal** resolution but **low spatial** resolution. While each of these representations has strengths, our task analysis indicates that visual exploration at a level with both reasonable temporal and spatial resolution was important. To support this type of analysis, we design a Mosaic-Matrix encoding. The encoding captures temporal and regional changes and integrates them into the community spatial distribution map.

Because nodes are densely located in brain slices, using 1D timeline-based representations was not feasible. Instead, we adopt a compact two dimensional layout to

encode time-dependent behavior. The layout is composed of a set of cells that define a Mosaic-Matrix. The set corresponds to a node region, and each cell encodes the temporal behavior of an individual node in that region (Figure 31). Each cell, in its turn, is a 2D dense pixel layout that wraps time into 2D. The sub-cells encode with color the set of communities that node belongs to during the selected time period. Figure 32 shows two Mosaic-Matrices for a subregion of 9 nodes across 33 timesteps (left), respectively 64 nodes across 12 timesteps (right). Figure 31 further demonstrates the nesting of community temporal information across 64 time steps into the brain slice of an aged mouse. The selected area highlighted in the black circle is a region of 9 nodes. Each of the nine nodes within the selected region is represented in the Mosaic-Matrix as a cell. The 64 (8×8) sub-cells encode temporal behavior, with time increasing from left to right and top to bottom. Figure 31 shows an experiment that would only be useful to deploy in cases where there is access to a high-resolution display wall.

However, the integrated temporal features may not be easily observed when displaying the entire brain slice. Additionally, zooming into a small region loses the relevant spatial references. Therefore, we enabled a detailed region view without losing the context of the slice. The resulting Mosaic-Matrix view is composed of several sets of cells.

Figure 32 (a) displays the temporal behavior across 33 time steps, while Figure 32 (b) displays 12 time steps. In Figure 32 (a), the node in the top-left corner of the Mosaic-Matrix is initially part of the green community, then moves to the orange community, and finally joins the pink and blue communities in the last two timesteps. Unlike the
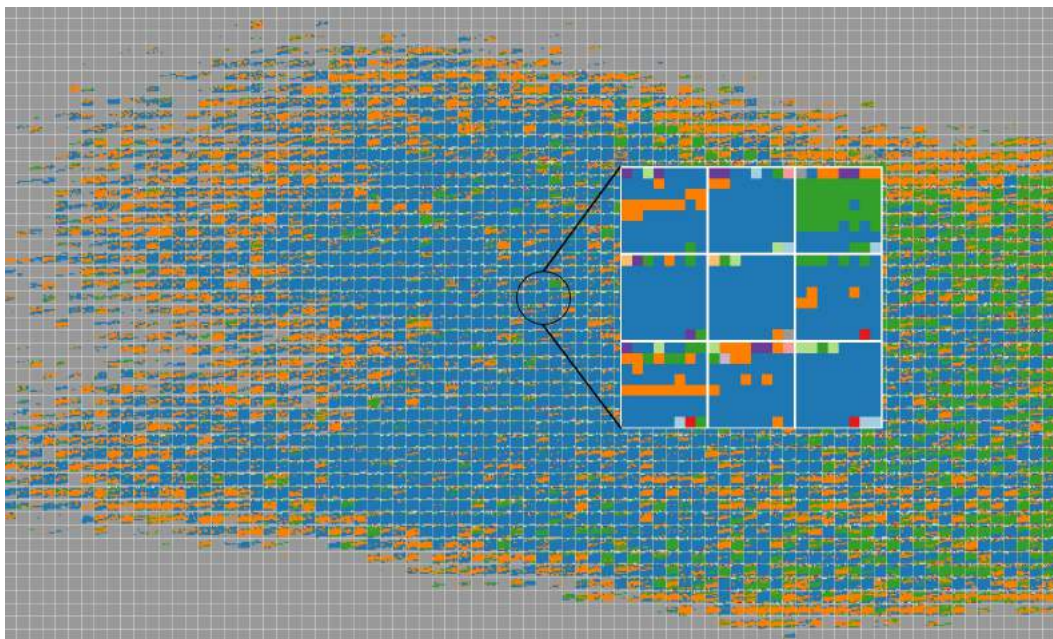
Figure 31. Integration of temporal community characteristics into the brain slice of an aged mouse, across 64 time steps. It is an experiment that would only be useful to deploy in cases where there is access to a high-resolution display wall.

traditional timeline-based (one dimension) visualization for time series data, the Mosaic-Matrix view allows us to effectively nest the temporal features into the spatial structures.

Users can both interactively translate a selection lens in the slice-based view and drag the zoom size slider in the control panel (Figure 28 (g)) to adjust the region size (number of nodes) selected. The Mosaic-Matrix can flexibly capture from one node to 100 nodes, as well as from one to 100 time steps, the length of the entire timeline in this study. Because in this study more than half of the brain slice was inactive at all times, we overlay the Mosaic-Matrix view on the top-half of the slice, to efficiently use space.
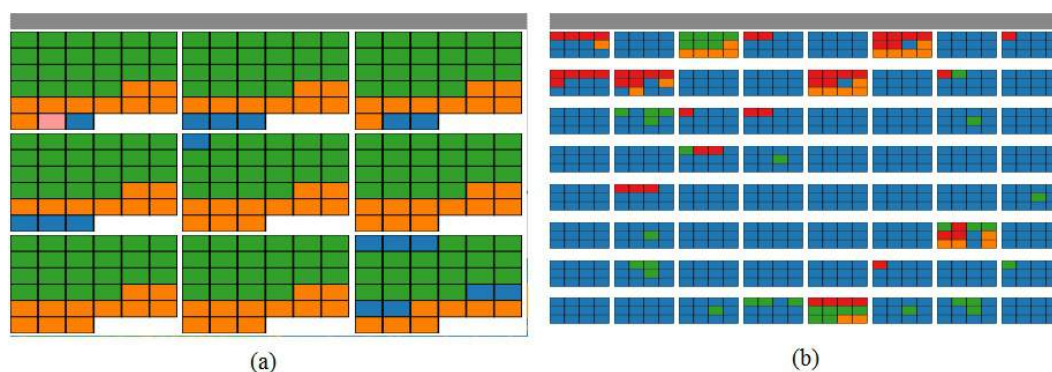
Figure 32. Two Mosaic-Matrix views representing two regions (at different zoom levels) across different time periods: (a) a region of 9 nodes across 33 timesteps; (b) a region of 64 nodes across 12 timesteps. In (a), the node in the top-left corner of the Mosaic-Matrix is initially part of the green community, then moves to the orange community, and finally joins the pink and blue communities in the last two timesteps. In (b), the mosaic captures instability (frequent changes) in the region selected.

In cases where the image activation data may become obscured by the Mosaic-Matrix view, the Mosaic-Matrix window can be moved by dragging the upper gray bar.

5.4.4   Timeline Widget

The timeline widget enables navigation over time in the slice-based view and mosaic-matrix view. Using the widget, a user can click and drag to select the time window. To further help identify the time window during which each slice is most active, the timeline widget also encodes as a plot the total number of active nodes over time (Figure 28 (c) and (d)). Two dashed lines mark peak activity — the timestep with the largest number of active nodes.

5.4.5   <u>Synchronization and Comparison Support</u>

To support the pairwise comparison of activations (***Task 4***), we adopt a coordinated side-by-side dual layout (Figure 28 (a)-(f)). This layout integrates multiple views at different levels: two slice-based views, two timeline views, two Mosaic-Matrix views, and two individual behavior panels. In our experience, because of the data complexity and a large number of differences in the datasets that are typically compared, the use of juxtaposed views effectively reduced visual clutter when compared to superposition. Moreover, the domain experts valued raw data and strongly objected to the superposition of brain slices (via registration) from different specimens. Although the Individual Behavior panel does lend itself to superposition, and we do support Kiviat overlays, the designers and the domain experts converged to a side-by-side layout for all views, in order to maintain consistency. As in other studies that involve domain scientists and require visual scaffolding (128; 8), we found that design clarity and consistency principles take precedence over expressiveness.

Because domain experts perform the comparison in both the spatial and temporal domain, we implement two default options for synchronization: a timeline synchronization (timeline widget) and a region synchronization (aggregate slice view). However, when comparing different specimens, the activations and spatial structures may not be completely aligned. Because of this constraint, we provide an asynchronization option as well, which allows the domain experts to manually align temporal or spatial features.

5.5   Evaluation

We evaluated RemBrain through a combination of multiple demonstrations and case studies (real data, real tasks, real users) with our collaborators: an established neuroscientist researcher who specializes in computational biology, neuroimaging and neurobiology, and a senior researcher in sensory-motor performance, who has a broad background in studying the adaptation of motor systems and imaging data from physiological systems. Both experts have been working together on dynamic brain network analysis for several years. Throughout the evaluation process, we used a "think-aloud" technique (129), which asks users to verbalize their thoughts as they interact with the system, and we collected feedback at the end.

Here we report a case study performed separately by the two scientists, in separate sessions. In this study, the domain experts seek to understand the impact of aging on the auditory cortex (AC) in mouse brains. To this end, they had collected imaging data from a young mouse and an old mouse. Brain slices from each specimen were artificially stimulated, and the resulting activation levels were imaged as a time series. The case study and verbiage reported below have been simplified for a lay audience.

5.5.1   Case Study: Aging Analysis in Mouse Brains

The domain experts wished to investigate how aging relates to auditory processing changes, through the comparison of network activity in the AC from young and aged mice, at multiple scales (*Task 1 through 4*). Each expert started by loading the dataset of the first activation of young mouse No.40 (5.5 months) and the dataset of the first activation of aged mouse No.38 (22 months) in the two side-by-side views.
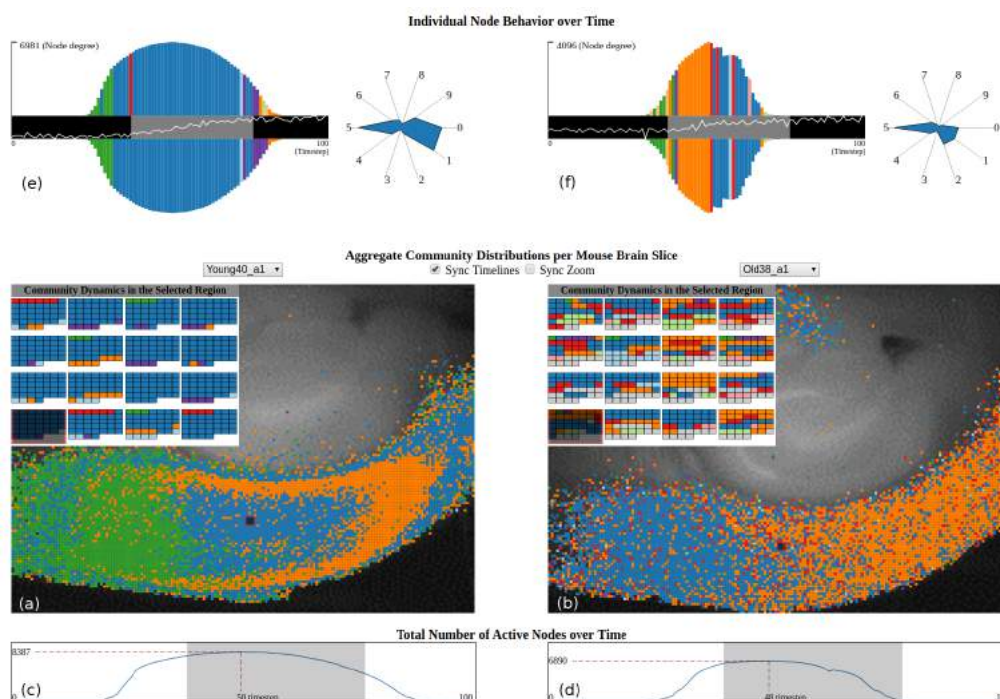
Figure 33. Case Study: Aging Analysis. The slice-based views (a) and (b) capture a difference in the community spatial distributions between young and aged mice. The Mosaic-Matrix views in (a) and (b) present both the spatial and temporal features of communities in two similar regions of young and aged mice. The timeline views (c) and (d) show a higher total number of active nodes for the younger specimen. While the (a) – (d) views compare the two specimens at a high and regional level, the individual behavior views (e) and (f) allow for comparison at the individual node level, both spatially and temporally.

Overview spatial, multi-scale in time exploration. The analysis started at the high overview level of the entire AC. The community distribution differences were immediately noticed in the slice-based panels (Figure 33 (a) and (b)): over the same time window, the young mouse AC features an additional community, shown in green. The young mouse AC (Figure 33 (a)), in particular, featured a ring-type structure of community distributions. That structure was stable even as the experts translated and

scaled their time window selection in the widget (**Task T2, T4**). In contrast, the community distributions in the aged mouse AC (Figure 33 (b)) were less structured. In fact, the neuroscientist expert noted that no contiguous region in this brain image was associated with one single community. The second expert noted that in the timeline views the activations from the two specimens decayed at different rates after reaching their peak. The timeline also captured a higher total number of active nodes for the younger specimen, which was expected. The domain experts concluded that the connectivity between neurons diminishes with age, which "probably correlates with a particular receptor [decay]".

Regional spatial, multi-scale time exploration. The multi-scale analysis moved next smoothly to the regional-scale captured by the Mosaic-Matrix views (**Task T1, T3, T4**). For this analysis, the experts disabled synchronization, and manually selected two regions (marked by red boxes) in roughly the same area of each AC. The difference in the dynamic community behavior between the two regions was striking. The cleaner and predominantly blue Mosaic-Matrix view in Figure 33 (a) captured a homogeneous dynamic behavior in the young mouse AC region. Most nodes in this brain region spend their time in only one community, blue. In contrast, the Mosaic Matrix for the aged brain in Figure 33 (b) indicates significant instability, which revealed the heterogeneity of the aged mouse AC. Only a few nodes stayed in a single community during the user-selected time. Furthermore, the aged AC network was significantly more fragmented over time, especially as the experts enlarged the time window size. The experts moved their attention repeatedly between the regional scale and the overview scale.

Individual spatial, multi-scale in time exploration. In the final analysis stage, the dynamic behavior at the scale of a single node was taken into consideration, in addition to the previously-examined spatial scales (**Task T1, T4**). In several iterations, the experts selected specific cells in the Mosaic Matrix, one at a time, and examined them through the individual panel (Figure 33 (e) and (f)). The nodes examined in this figure are located in the bottom left corner of the Mosaic-Matrix views. Surprisingly, both young and old nodes exhibited symmetric behavior with respect to their Home and Temporary distribution. However, the node degree over time was almost double for the young neuron, when compared to the old. Furthermore, the mirror glyph encoding quickly showed, for example, that sample nodes in the young mouse brain featured few major changes in dynamic communities over the entire time window. In Figure 33 (e), the selected node switches only twice closed to the start of the window, from green to blue and from blue to purple. In contrast, the aged mouse node shown in Figure 33 (f) switches much more frequently at the start of the activation.

The difference noted above was reinforced by the Kiviat diagrams, where the two Kiviat shapes were notably similar in many respects. In the example shown, both nodes score along the normalized average time span of communities (axis 1), but only the aged neuron has a non-zero normalized switching cost (axis 2). Furthermore, the aged mouse also had a shorter activation time (axis-0), and fewer connections than the young mouse. At this point, the neuroscientist expressed interest in seeing the raw signal data. To this end, the experts examined the raw signal plots in the behavior mirror glyphs,

and discovered that the old mouse rising time is generally slower and that the curve during the rising time is less smooth for the old mouse.

Case findings. This multi-scale analysis indicated that aging is associated with a series of changes in node metrics, such as community size and switching cost, and also with temporal changes in individual behavior, such as dynamic community distribution. These changes are consistent at multiple spatial and temporal scales. Together, the domain experts hypothesized that those aging-related changes at multiple scales might be related to changes in intracortical connectivity (***Task T4***). Using the insight from the multi-scale visual exploration, they are designing methodology to capture and quantitatively report these correlations.

5.5.2   Domain Expert Feedback

The domain expert feedback included comments such as "very cool, interesting tool", "fantastic", and "useful to generate hypotheses". Since everything in biology is "so tied to spatial location", the experts found that the integration of the spatial layout and non-spatial network attributes was far more useful than analyses based only on the nonspatial data. In addition, the Mosaic-Matrix provided the ability to explore temporal relationships between nodes with close proximity in the same region, and thus preserved a useful spatial context. While originally unfamiliar to the experts, the mirror glyphs and Kiviats were later praised for their potential to drive hypotheses at the neuron level, once crisper node data become available through the next imaging project. Overall, the experts found that RemBrain augmented their ability to analyze the heterogeneous and multifaceted datasets common in dynamic bionetwork analysis.

Compared to prior analyses of the data, which were done directly with data files and relied on the experts' mental model of the node location within an image, our visual approach succeeded in communicating the spatial findings to others. Also, the experts noted that interactively exploring the imaging data to identify an interesting time step was far more efficient than manually searching an image from a repository.

5.6   Discussion

Meeting the original goals. The case study and expert feedback demonstrate the effectiveness of this multi-scale visualization approach to the comparative exploration of dynamic activation networks across multiple brain imaging datasets at multiple levels. Experts were able to find new, interesting patterns in datasets they had explored using different tools before. They both are eager to adopt the tool for research purposes, both in an exploratory setting and in an explanatory setting (for publication purposes).

The overall design was successful at supporting comparative analysis in a variety of dataset combinations. We note that few guidelines exist in visual comparison design. In most instances in our design we favored juxtaposed (side-by-side) layouts, to attain better clarity and consistency, and to circumvent alignment issues. One exception is the star panel, where the lack of a physical structure supports superposition. Overall, we found that a hybrid approach best supported the tasks revealed by the domain analysis.

The experts considered the inclusion of the spatial context a most valuable feature, and reported the approach was far more useful than analyses based only on the nonspatial data. The chosen visual encodings showed complementary strengths in supporting multi-scale spatio-temporal analyses. When coupled with a coordinated multi-view approach,

these encodings enabled visual analysis across the entire pipeline for dynamic bionetwork data analysis: raw data, network results (node degree), dynamic community analysis based on the results of dynamic networks, and summarized node metrics based on the dynamic community analysis. The experts were able to navigate smoothly between multiple scales in both space and time.

Novelty. The mirror glyphs and the embedding of temporal features in a spatial context through the Mosaic-Matrix views are novel contributions. The composite mirror glyphs, in particular, are not restricted to the presentation of dynamic node behavior in neuroscience. These glyphs can also be applied to general temporal data with multiple variables that include both numerical (height) and categorical data (color). Such datasets exist in other domains where symmetric/asymmetric time-dependent behavior is of interest, for instance in the analysis of spectrograph data in astronomy (130), in the analysis of financial data, or in the analysis of Electronic Health Records. The Mosaic-Matrix nesting approach may find application in other spatially dense temporal datasets.

The combination of visual encodings in a tool to handle multivariate data in dynamic bionetwork analysis and the side-by-side multi-scale design that supports pairwise comparison for spatiotemporal data are also novel. The approach has direct application to the analysis of other spatially-dependent dynamic biological networks, for instance in computational systems biology.

Design lessons and issues. One of the most important lessons from this work relates to limitations arising from increasing model scales and complexity. As scientific models

move from static to dynamic, and single model analysis shifts to the comparison of multiple models with spatial and nonspatial features, even known integration paradigms break down with scale: one cannot keep track effectively of tens of coordinated views. Overlays may similarly fail, and in some instances may not be applicable (in our case, due to domain restrictions related to alignment and the importance of raw data). In the approach illustrated here, we have successfully nested the time-driven behavior into spatial structures and used overlaying and details on demand where possible, to overcome space limitations. Still, the resulting interface is information-dense; on a large tiled display, there was still too little space to attach legible Kiviat labels directly to axes. As the range of data acquisition instruments keeps expanding, these issues will only become more stringent in the visualization field.

The second important lessons arising from this experience relates to the necessity of visual scaffolding (**?**) when dealing with domain experts who are not familiar with sophisticated visual encodings. In our design experience, the application of HCI principles such as clarity and consistency, and the careful consideration of the overall application gestalt were particularly important. For instance, the final design includes dual views for all representations, even though we do support superpositions where possible. We furthermore found success by building upon domain-specific encodings, such as the slice-based views and the timeline widgets. Using those familiar encodings within a linked view framework served as a visual scaffold, allowing the domain experts to harness and expand their previous analysis experience.

Limitations. In terms of limitations, one of our two domain experts noted that interpreting individual node behavior could be too daring given the current imaging done on the datasets. Nevertheless, he did see the future utility of the behavior glyphs in the context of their next imaging project, which will capture node identity more crisply.

Another limitation is that matching precisely communities between different experiments (beyond the size proxy for the ten largest communities) is an open research issue, not just in this work, but in dynamic community analysis in general. This limitation is mainly because dynamic analysis methods currently do not consider the spatial relationship of nodes in different experiments. In this case, the domain experts are well-aware of and willing to accept this current limitation.

Last but not least, the pairwise comparison in a side-by-side layout becomes increasingly inefficient as the number of objects grows, especially when the objects have spatial features. Analysis will devolve to an exhaustive pairwise comparison of all subjects and memorization of activation patterns. Extending the juxtaposition paradigm of our side-by-side views into a small multiple layout is feasible only in an environment with large displays and yet, with the density and complexity of the data, will still quickly become overwhelming. Thus, effective comparative visual designs must address the challenge of scalability in terms of the number of items and the complexity of individual items to be compared at the same time. According to the considerations of comparison (11), extracting features through dynamic behaviors of brain activity networks and comparing those networks based on their summarized characteristics at an abstract level would be a good solution to this limitation. Therefore, we extend the RemBrain application to

support the visual exploration and comparison of dynamic brain networks at multiple scales.

CHAPTER 6

REMBRAIN 2.0: MULTI-SCALE VISUAL COMPARISON OF ENSEMBLE BRAIN ACTIVITY NETWORKS

6.1 Motivation

Network analysis repeatedly performed on different subjects and/or at different conditions forms large datasets for neuroscientists and computational biologists to better understand the functional connectivity of the brain and the dynamics of such connectivities. Although many techniques and tools for visualizing brain activity networks exist, little work focuses on the visual comparative investigation of such networks modeled from multiple subjects or conditions. Most visual comparative approaches only support pairwise comparison due to the difficulty in comparing multiple objects, discussed in the section above. Therefore, we propose a summary+detail visual approach for comparative analysis of brain networks at four scales, using visual abstractions to allow the user to systematically explore similarities and differences in spatial-temporal-functional behaviors across multiple brain activity networks. We demonstrate this approach in our prototype web-based visualization tool, RemBrain 2.0 (Figure 34).

In RemBrain 2.0, we implement the summary+detail visual approach by integrating two complementary high-level summary views with three low-level detailed views. The summary overviews consist of an interactive principal component analysis (PCA) view (Figure 34 (a)) and small-multiple kiviat diagrams (Figure 34 (b)). Together, these two
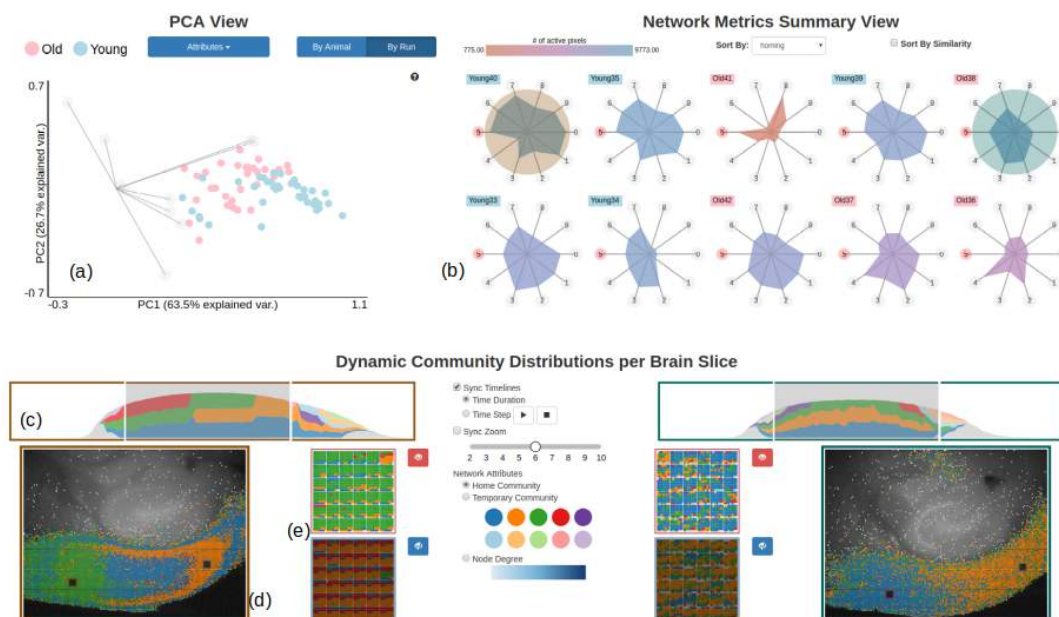
Figure 34. RemBrain 2.0 implements a summary+detail visual approach for the analysis for comparative analysis of multiple brain networks at four scales: groups of subjects, individual subjects, activations of a subject, and ROIs of an activation.

views summarize temporal-functional behaviors of brain activations from multiple subjects and support visual comparisons across those activations/subjects at two scales — groups of subjects and individual subjects. At detailed levels, for specific activations of the selected subject, the dynamic changes of communities along with their distributions in the brain slice are presented separately in a dynamic behavior (temporal) view (Figure 34 (c)) and a slice-based (spatial) view (Figure 34 (d)). A more detailed exploration on the regions of interest (ROIs) nests the temporal behaviors into user-selected regions (Figure 34 (e)). These three views support the exploration of spatial-temporal-functional behaviors at two scales — activations of a subject and smaller ROIs. Overall, these five

visual components work together to enable the exploration and comparison of multiple brain activity networks at four levels of analysis.

To demonstrate the effectiveness of this visual approach, we conduct a case study that seeks to analyze ensemble mouse brain activity networks. Feedback collected from domain experts during this case study indicates that our summary+detail approach can help neuroscientists better understand the different spatial-temporal-functional behaviors across multiple subjects and explore those differences at multiple scales.

6.2   Related Work

In addition to the related work introduced in Section 5.2, several new works about comparative visualizations have been published this year. Yang et al. (131) present a visual analytics method using the inherent block structure of the human brain to compare human brain networks from multiple population groups. However, their visual comparison of multiple subjects is only performed through the pairwise side-by-side comparison. Similarly, Song et al. (132) use a small multiples design (juxtaposition) to support the comparison of gaze data from 14 radiologists examining volumetric images. However, small multiples approach may not work well for a larger dataset. Fujiwara et al. (133) design a visual analytics system to compare the differences in brain networks between individual subjects as well as group averages, which is similar to our research. Fijiwara et al. use multidimensional scaling to show the (dis)similarities between objects in low dimensional space; similarly, we use principal component analysis for dimensionality reduction in RemBrain 2.0. We also follow the Summarize Somehow strategy (11) to build

abstractions of a large amount of data by measuring and quantifying differences through computational and statistical analysis.

6.3   Data and Task Analysis

In this extension of RemBrain, we are dealing with a larger, ensemble dataset which contains 70 activations across five young mice and five aged mice. The initial data analysis process is exactly the same as the process introduced in Section 5.3.1. Additionally, we compute the metrics that summarize the behavior of active nodes for each brain activity network by calculating the mean of all active nodes in that activity network and the metrics for each subject (mouse) by calculating the mean of all activations for that subject. The metrics at these two scales will be presented in the summary overviews. In addition to the tasks discussed in the Task Analysis Section 5.3.2, a major focus of this extension is to summarize and compare dynamic behaviors of multiple subjects (mice) or multiple activity networks from the same subjects to find trends and outliers.

6.4   Visual Design

To support the tasks discussed in Section 5.3.2, RemBrain 2.0 reuses effective visual components from the RemBrain tool, including the slice-based views, timeline widgets, and Mosaic-Matrix views, while extending their functionality. In addition to showing the aggregated information in the slice-based view only, we also provide options to display the community structures at a user-selected time step and play the community evolutions as a movie. The dynamic behavior view (stacked plots) shown in Figure 34 (c) extends the time widget control by embedding the nodal community temporal distribution into the node activity curve. RemBrain 2.0 also uses the Mosaic-Matrix visual encoding to

allow a user to compare two ROIs shown in Figure 34 (e) within one activity network or across two networks, whereas RemBrain only supported a single selection per network.
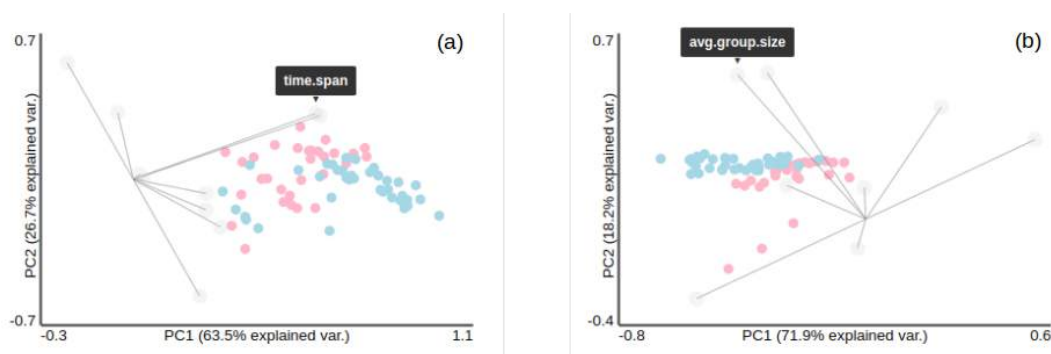
Figure 35. The interactive PCA view supports the selections of attributes to calculate the PCA results. (a) shows the results from all ten attributes and (b) shows the results from eight attributes without time.span and observed.

6.4.1 PCA View

To support the task of comparing multiple brain activity networks or subjects, we use a similar strategy to the Gleicher framework in Summarize Somehow (11) by comparing them through their summarized characteristics at an abstract level. Those summarized characteristics that describe the dynamic behaviors of the networks are the network metrics which contain ten features described in Table II. Our domain experts have previously used PCA to understand the relationships between those activity networks/subjects with those ten attributes. The PCA plot offers the ability to identify characteristics that show as much variation across activations/subjects as possible at an abstract level. By ex-

amining the PCA projections on attribute axes, domain experts speculated that not all attributes contribute a clear separation of aged and young mice. However, common PCA applications do not allow for the interactive selections of attributes. To afford this functionality to our users, we implement an interactive PCA view which dynamically calculates and updates the PCA results according to the new set of attributes. Figure 35 (a) displays the PCA results calculated from all ten attributes and Figure 35 (b) shows the results based on the attributes excluding time.span and observed.
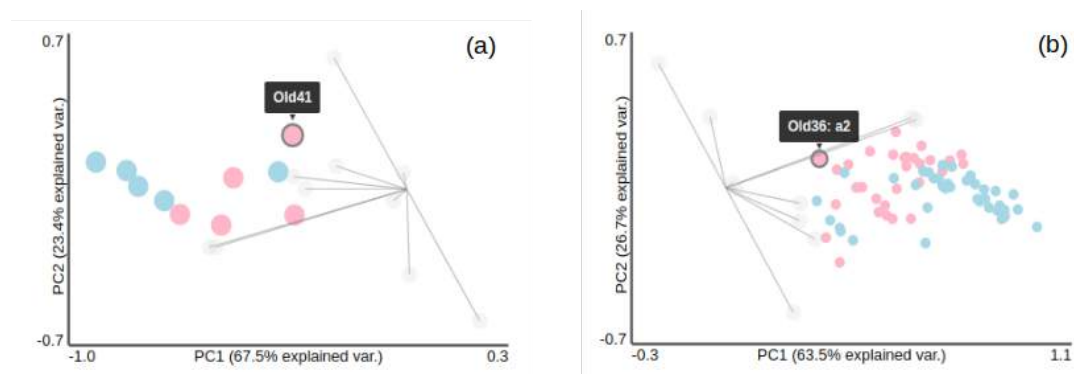
Figure 36. The PCA view includes two modes (a) By Animal that plots the mean value of activations grouped by animals and (b) By Run that plots all 70 activations.

Additionally, the interactive PCA view also supports plotting either all 70 activations (By Run) in Figure 36 (b) or average value of activations grouped by animals (By Animal) in Figure 36 (a). In the mode of By Animal, the user can further inspect all

the activations of a particular subject by clicking on the corresponding dot. The PCA
view then will plot all the activations of the selected subject.
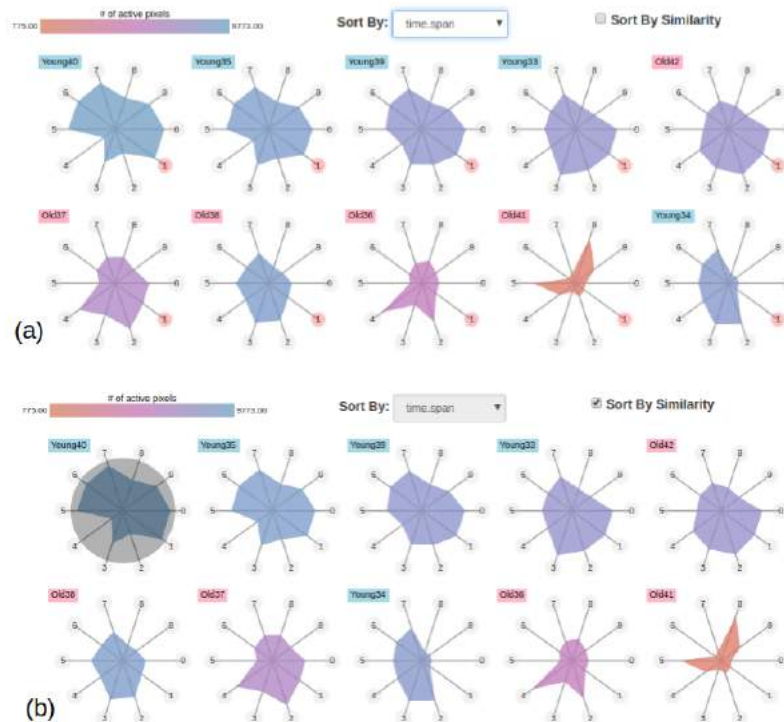
Figure 37. Network Metrics Summary views sort animals by (a) the attribute of
time.span and (b) similarities to the mouse Young 40.

6.4.2   Network Metrics Summary View

PCA is an effective method to visually explore the relationships between activations/subjects
with high-dimensional attributes. However, it cannot provide direct comparisons of par-

ticular attributes across activations/animals. To enable these comparisons, we implement a Network Metrics Summary view in which each kiviat diagram corresponds to an activation/subject and each axis of the kiviat diagram represents one of ten attributes that are used for calculating the PCA. We chose Kiviat diagrams over PCPs or linear bar charts because the shapes of the kiviat diagrams easily illustrate similarities and differences between activations/subjects. The kiviat diagrams are colored by the average node degrees of activations/subjects. The user can also use the RemBrain 2.0 sorting assistant to sort the kiviat diagrams by a user-selected attribute (Figure 37 (a)), or to select a kiviat diagram and sort the rest by their similarity to the selected target (Figure 37 (b)). Overall, the PCA view supports the visual comparisons of multiple subjects/activations across the selected attributes while the Network Metrics Summary view supports the visual comparisons of multiple attributes across multiple subjects/activations. Those two views are linked through several interactions, e.g. mouse over and clicking.

Through the exploration in two summary views, the user may find a particularly interesting activation and want to explore it in detail. He/she can then load that activation into the detailed view of Dynamic Community Distributions per Brain Slice by right-clicking on the activation, either on the PCA plotted dot or the kiviat diagram. Once the activation is loaded, the user can perform similar operations as what he/she performed in RemBrain to interactively explore the community distributions in the brain slice or ROIs and track the evolutions of communities over time.

6.5   <u>Evaluation</u>

We evaluated RemBrain 2.0 through a combination of multiple demonstrations and a case study with our collaborators: a senior computer scientist who specializes in developing discrete modelling and analysis techniques to various areas of computational biology, a Post-doc researcher specializes in neuroimaging and neurobiology, and our two initial collaborators whom we have worked with to develop the RemBrain together. The computer scientist and our two initial collaborators have been working together on dynamic brain network analysis for several years. After several demonstrations for the two initial collaborators, we conducted a formal case study with all four collaborators together during a remote meeting session. All participants have access to the web-based visualization tool and is able to operate the interface directly. However, to make sure all participants are looking at the same thing, we had one user share his/her screen with the group and perform the operations. Throughout the evaluation process, we used the same "think-aloud" technique that we used for RemBrain and we collected feedback at the end.

Domain experts provided positive feedback that the proposed visual approach is an excellent demonstration of the data and tasks by showing the statistics through the PCA view and summary view. The Post-doc researcher pointed out that, for a large set of data, our approach offers an easy way to visualize all of the data at once, instead of going one by one and making pairwise comparisons of activations. Once a scientist notices a pattern through the summary views, then he/she can investigate that hypothesis through detailed views. The computer scientist indicated that activations are very consistent for

an animal by looking at the PCA view, and in which the attribute axes state that the young mice activations last longer than aged mice. She appreciated the freedom to manipulate the PCA view interactively and expresses her interest in the application of this multi-scale comparative visual approach to the analysis of other types of data, e.g. healthy vs. disease datasets in the study of Alzheimer's disease. The two initial collaborators both agreed that RemBrain 2.0 is the most advanced visualization tool they have used to explore and compare dynamic brain networks.

6.6   Discussion

The summary+detail visual approach proposed in the RemBrain 2.0 supports the exploration and comparison of multiple brain activity networks at multiple scales. The development of this approach was guided through past experience, working either with this biological network problem or through learning about challenges in another biological subdomain. Through the process of gradually expanding the applicability of visual analytics applications to the challenges presented by the spatial-nonspatial, multi-scale, and multi-run data, I have grown to better understand techniques which allow for a rich visual analysis of complex biological networks that feature the three main challenges of spatial-nonspatial integration, multi-scale exploration, and multi-run comparisons.

CHAPTER 7

DISCUSSION AND CONCLUSION

This dissertation characterizes three visual challenges behind analyzing dynamic biological networks — spatial-nonspatial visual integration, multi-run visual comparison, and multi-scale visual exploration — and surveys the state of the art in visual integration of spatial-nonspatial biological data. The dissertation introduces two particular domain problems with data and task analysis, proposes a set of visualization techniques and encodings, presents five applications to those two biological problems, and evaluates those visualizations through case studies and usage scenarios.

Specific contributions of this work include:

- a survey of works that integrate spatial and non-spatial data in biology. The survey presented in this dissertation mainly covers the biological subdomains of neuroscience, and bioinformatics. The survey found that the linked views used alone or the hybrid approach, in which one paradigm is linked views, are more popular in the biological domain.

- an analysis of the domain data and tasks in dynamic biological networks. The analysis explains the domain problems in detail that includes how to process and analyze the data and what tasks are required. The analysis may also benefit researchers who are interested in or currently working on similar problems.

- several innovative visual designs. A set of novel visual representations using a hybrid paradigm consisting of the linked views, overlays, and nesting approaches, enables the visual integration of spatial and non-spatial data in dynamic biological networks.

- a visual approach to compare multiple dynamic biological networks at multiple scales. The visual approach supports the comparison of multiple objects at an abstract level through their summarized statistical characteristics and provides detailed pairwise comparisons based on the user's demand.

- five visualization applications developed for two domains that feature dynamic biological networks. These applications developed in collaboration with domain experts address the challenges in visualizing such problems which contain spatial and non-spatial visual integration, multi-scale visual exploration, and multi-run visual comparison.

Regarding limitations to our work, since some visual analytics tools which target biological problems might not be covered in this survey due to publication outside the visualization journals, and as more relative visualization applications are developed, an expanded survey of these applications will affect the current conclusions. Thus, due to the complexity of the biological data and tasks, linked views alone may not be the most popular integration paradigm and might be over-performed by the hybrid approach. Additionally, the taxonomy extracted from the survey that we have been working on is still not fully ready at this point.

The visual designs developed for these two specific biological networks might not be easily generalized to other biological networks. Both PRODIGEN and RemBrain rely on custom visual representations to help integrate the spatial and non-spatial information. The enhanced heatmap in PRODIGEN that nests the temporal changes of probabilities into the landscape is used to study the temporal and spatial distributions of probability. The mosaic-matrix in RemBrain was designed to nest the temporal behavior of neurons into a sub-region of the image slice of the brain. Although the visual encodings are designed for supporting the specific analyses of these biological networks, the principle behind them — nesting non-spatial features (temporal) into spatial layouts — can be further generalized to other biological networks, or even other domains. Last, some of these works, e.g. EnsemblePron — the extension of PRODIGEN, have been mainly evaluated by myself via usage scenarios.

In terms of future work, the survey in Chapter 2 provides a basis for conducting a taxonomy to formalize the understanding of how each visual technique is chosen in specific conditions. A nesting approach, which is adopted in PRODIGEN, RemBrain and RemBrain 2.0, to integrate dynamic behaviors into spatial data may have a wider potential use in other biological problems or even a broader domain. Additionally, RemBrain 2.0, the visualization application in Chapter 6, is not only limited to the study of comparing the auditory cortex of young mice and aged mice but may be applicable to the multi-run visual comparison of other types of imaging data with the use of network analysis model as well. Using the statistical analysis model to extract critical features as the biological networks get larger and more complex, visualization researchers can

further expand the multi-scale comparative visual approach, which is proposed from RemBrain 2.0, to a more generalized approach that supports multiple comparisons of objects at abstract levels and uses overviews to guide the user to explore and compare particular object at more detailed levels.

In conclusion, the dissertation proposes a set of visual approaches for the analysis of ensemble dynamic biological networks. This set of approaches took shape through the development of multiple prototype visual analytics tools which were aimed to solve complex problems in multiple biological subdomains. The steering of the approach was supported through an in-depth understanding of both the encompassing biological domains and problem space, as well through a survey of existing visual approaches helping to generalize the paradigms used when confronting these problems and the complexity of their data.

APPENDICES

Appendix A

SUPPLEMENTAL MATERIAL LINKS

The following are the links to videos of four visualization applications:

A video of PRODIGEN in Chapter 3:
https://youtu.be/xUhuoPcqEds

A video of SwordPlots in Chapter 4:
https://drive.google.com/open?id=1lppVJwiIidW5Urp29mlSVlf3jQYwhefH

A video of RemBrain in Chapter 5:
https://drive.google.com/open?id=1ANUD29OUsOcfxYzFC1uCQJRjGfw94Zhs

Two videos of RemBrain 2.0 in Chapter 6:
https://drive.google.com/open?id=1x7HYtg0ikDFqzOIoWMflVuiHmTIdfWzE
https://drive.google.com/open?id=1dKujauhkLbbzf0j6sqYH4Yqnx981cfj3

Appendix B

PERMISSIONS FOR REUSE

The following are written permission from the publisher or a statement from the journal's/publisher's website outlining their policies.

BMC Bioinformatics

Copyright

- Copyright on any open access article in a journal published by BioMed Central is retained by the author(s).
- Authors grant BioMed Central a license to publish the article and identify itself as the original publisher.
- Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified.
- The Creative Commons Attribution License 4.0 formalizes these and other terms and conditions of publishing articles.

C. Ma, T. Luciani, A. Terebus, J. Liang, and G. E. Marai. PRODIGEN: Visualizing the Probability Landscape of Stochastic Gene Regulatory Networks in State and Time Space. BMC Bioinformatics, 18(2): 24, 2017.

Appendix B (Continued)

Journal of Imaging Science and Technology

March 9, 2018

Ms. Chihua Ma
University of Illinois at Chicago
Electronic Visualization Lab
Dept. of Computer Science M/C 152
851 S Morgan, Room 1120
Chicago, IL 60607-7053

Dear Ms. Ma:

Thank you for your email requesting permission to use the following articles as part of your dissertation entitled "Visual Analysis Techniques for Dynamic Biological Networks."

Chihua Ma, Angus G. Forbes, Daniel A. Llano, Tanya Berger-Wolf, and Robert V. Kenyon, "SwordPlots; Exploring Neuron Behavior within Dynamic Communities of Brain Networks, *Journal of Imaging Science and Technology* **60**, No. 1, PP. 010405-1—010405-13 (2016)

Chihua Ma, Filippo Pellolio, Daniel A. Llano, Kevin Ambrose Stebbings, Robert V. Kenyon, and G. Elisabeta Marai, "RemBrain: Exploring Dynamic Biospatial Networks with Mosaic Matrices and Mirror Glyphs, *Journal of Imaging Science and Technology*, **61**, No. 6, pp. 060404-1—060404-13 (2017.

IS&T hereby grants you permission to use the article as mentioned above. Please note that the Society requires a suitable acknowledgment such as the following: "Reprinted with permission of IS&T: The Society for Imaging Science and Technology sole copyright owners of the *Journal of Imaging Science and Technology*."

Sincerely,

Suzanne E. Grinnan
Executive Director

CITED LITERATURE

1. 2017 ALZHEIMER'S DISEASE FACTS AND FIGURES, 2018 (accessed January 10, 2018).

2. Greicius, M.: Scientists find genetic underpinnings of functional brain networks seen in imaging studies, 2016 (accessed March 30, 2016).

3. Li, Y., Qin, Y., Chen, X., and Li, W.: Exploring the functional brain network of alzheimer's disease: based on the computational experiment. PloS one, 8(9):e73186, 2013.

4. Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., et al.: Dynamic functional connectivity: promise, issues, and interpretations. NeuroImage, 80:360–378, 2013.

5. Robinson, L. F., Atlas, L. Y., and Wager, T. D.: Dynamic functional connectivity using state-based dynamic community structure: Method and application to opioid analgesia. NeuroImage, 108:274–291, 2015.

6. Munzner, T.: Visualization Analysis and Design. CRC Press, 2014.

7. Chong, J. S. X., Liu, S., Loke, Y. M., Hilal, S., Ikram, M. K., Xu, X., Tan, B. Y., Venketasubramanian, N., Chen, C. L.-H., and Zhou, J.: Influence of cerebrovascular disease on brain networks in prodromal and clinical alzheimer's disease. Brain, 140(11):3012–3022, 2017.

8. Maries, A., Mays, N., Hunt, M. O., Wong, K. F., Layton, W., Boudreau, R., Rosano, C., and Marai, G. E.: Grace: A visual comparison framework for integrated spatial and non-spatial geriatric data. IEEE Trans. Visualization and Computer Graphics, 19(12):2916–2925, 2013.

9. O'Donoghue, S. I., Gavin, A.-C., Gehlenborg, N., Goodsell, D. S., Hériché, J.-K., Nielsen, C. B., North, C., Olson, A. J., Procter, J. B., Shattuck, D. W., et al.: Visualizing biological datanow and in the future. Nature methods, 7:S2–S4, 2010.

10. Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., and Roberts, J. C.:
    Visual comparison for information visualization. Information Visualization,
    10(4):289–309, 2011.

11. Gleicher, M.: Considerations for visualizing comparison. IEEE Trans.
    Visualization and Computer Graphics, 24(1):413–423, 2018.

12. Ma, C., Luciani, T., Terebus, A., Liang, J., and Marai, G. E.: Prodigen: visualizing
    the probability landscape of stochastic gene regulatory networks in state and
    time space. BMC bioinformatics, 18(2):24, 2017.

13. Ma, C., Burks, A., Luciani, T., Terebus, A., Liang, J., and Marai, G. E.: Visualizing
    ensemble time-evolving probability landscapes of stochastic networks. In
    ISMB/ECCB (BioVis), pages 1–2, 2017.

14. Ma, C., Forbes, A. G., Llano, D. A., Berger-Wolf, T., and Kenyon, R. V.: Sword-
    plots: Exploring neuron behavior within dynamic communities of brain net-
    works. Journal of Imaging Science and Technology, 60(1):10405–1–10405–
    13, 2016.

15. Ma, C., Pellolio, F., Llano, D. A., Stebbings, K. A., Kenyon, R. V., and Marai,
    G. E.: Rembrain: Exploring dynamic biospatial networks with mosaic-
    matrices and mirror glyphs. Journal of Imaging Science and Technology,
    61(6):60404–1–60404–13, 2017.

16. ten Caat, M., Maurits, N. M., and Roerdink, J. B.: Functional unit maps for
    data-driven visualization of high-density eeg coherence. In EuroVis, pages
    259–266. Citeseer, 2007.

17. Janoos, F., Nouanesengsy, B., Machiraju, R., Shen, H. W., Sammet, S., Knopp,
    M., and Mórocz, I. Á.: Visual analysis of brain activity from fmri data.
    In Computer Graphics Forum, volume 28, pages 903–910. Wiley Online Li-
    brary, 2009.

18. Nowke, C., Schmidt, M., van Albada, S. J., Eppler, J. M., Bakker, R., Diesrnann,
    M., Hentschel, B., and Kuhlen, T.: Visnest-interactive analysis of neu-
    ral activity data. In IEEE Symp. Biological Data Visualization (BioVis),
    pages 65–72. IEEE, 2013.

19. Chen, W., Ding, Z. A., Zhang, S., MacKay-Brandt, A., Correia, S., Qu, H., Crow,
    J. A., Tate, D. F., Yan, Z., and Peng, Q.: A novel interface for inter-

active exploration of dti fibers. IEEE Trans. Visualization and Computer Graphics, 15(6):1433–1440, 2009.

20. Jianu, R., Demiralp, Ç., and Laidlaw, D. H.: Exploring 3d dti fiber tracts with linked 2d representations. IEEE Trans. Visualization and Computer Graphics, 15(6):1449–1456, 2009.

21. Beyer, J., Al-Awami, A., Kasthuri, N., Lichtman, J. W., Pfister, H., and Hadwiger, M.: Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. IEEE Trans. Visualization and Computer Graphics, 19(12):2868–2877, 2013.

22. Al-Awami, A. K., Beyer, J., Strobelt, H., Kasthuri, N., Lichtman, J. W., Pfister, H., and Hadwiger, M.: Neurolines: a subway map metaphor for visualizing nanoscale neuronal connectivity. IEEE Trans. Visualization and Computer Graphics, 20(12):2369–2378, 2014.

23. Al-Awami, A. K., Beyer, J., Haehn, D., Kasthuri, N., Lichtman, J. W., Pfister, H., and Hadwiger, M.: Neuroblocks–visual tracking of segmentation and proofreading for large connectomics projects. IEEE Trans. Visualization and Computer Graphics, 22(1):738–746, 2016.

24. Rübel, O., Weber, G. H., Keränen, S. V., Fowlkes, C. C., Hendriks, C. L. L., Simirenko, L., Shah, N., Eisen, M. B., Biggin, M. D., Hagen, H., et al.: Pointcloudxplore: Visual analysis of 3d gene expression data using physical views and parallel coordinates. In EuroVis, pages 203–210, 2006.

25. Meyer, M., Munzner, T., DePace, A., and Pfister, H.: Multeesum: a tool for comparative spatial and temporal gene expression data. IEEE Trans. Visualization and Computer Graphics, 16(6):908–917, 2010.

26. Mercer, J. D., Pandian, B., Lex, A., Bonneel, N., and Pfister, H.: Mu-8: visualizing differences between proteins and their families. In BMC proceedings, volume 8, page 1. BioMed Central, 2014.

27. Doncheva, N. T., Klein, K., Morris, J. H., Wybrow, M., Domingues, F. S., and Albrecht, M.: Integrative visual analysis of protein sequence mutations. In BMC proceedings, volume 8, page 1. BioMed Central, 2014.

28. Luciani, T., Wenskovitch, J., Chen, K., Koes, D., Travers, T., and Marai, G. E.: Fixingtim: interactive exploration of sequence and structural data to iden-

tify functional mutations in protein families. In BMC proceedings, volume 8, page S3. BioMed Central, 2014.

29. Lenz, O., Keul, F., Bremm, S., Hamacher, K., and von Landesberger, T.: Visual analysis of patterns in multiple amino acid mutation graphs. In IEEE Conf. Visual Analytics Science and Technology (VAST), pages 93–102. IEEE, 2014.

30. Stolte, C., Sabir, K. S., Heinrich, J., Hammang, C. J., Schafferhans, A., and O'Donoghue, S. I.: Integrated visual analysis of protein structures, sequences, and feature data. BMC bioinformatics, 16(Suppl 11):S7, 2015.

31. Meyer, M., Munzner, T., and Pfister, H.: Mizbee: a multiscale synteny browser. IEEE Trans. Visualization and Computer Graphics, 15(6):897–904, 2009.

32. Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S.: Stochastic gene expression in a single cell. Science, 297(5584):1183–1186, 2002.

33. Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J.: Stochasticity in gene expression: from theories to phenotypes. Nature Reviews Genetics, 6(6):451–464, 2005.

34. Liang, J., Cao, Y., Gürsoy, G., Naveed, H., Terebus, A., and Zhao, J.: Multiscale modeling of cellular epigenetic states and tissue patterning: stochasticity in molecular network, chromatin folding in cell nucleus, and cell-cell interactions in tissue patterning. 2016.

35. Duncan, A., Liao, S., Vejchodský, T., Erban, R., and Grima, R.: Noise-induced multistability in chemical systems: Discrete versus continuum modeling. Physical Review E, 91(4):042111, 2015.

36. Gillespie, D. T.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, 81(25):2340–2361, 1977.

37. Cao, Y. and Liang, J.: Adaptively biased sequential importance sampling for rare events in reaction networks with comparison to exact solutions from finite buffer dcme method. The Journal of Chemical Physics, 139(2):025101, 2013.

38. Roh, M. K., Daigle Jr, B. J., Gillespie, D. T., and Petzold, L. R.: State-dependent doubly weighted stochastic simulation algorithm for automatic character-

ization of stochastic biochemical rare events. The Journal of Chemical Physics, 135(23):234108, 2011.

39. Kuwahara, H. and Mura, I.: An efficient and exact stochastic simulation method to analyze rare events in biochemical systems. The Journal of Chemical Physics, 129(16):165101, 2008.

40. Cao, Y., Terebus, A., and Liang, J.: Accurate chemical master equation solution using multi-finite buffers. Multiscale Modeling & Simulation, 14(2):923–963, 2016.

41. Cao, Y., Terebus, A., and Liang, J.: State space truncation with quantified errors for accurate solutions to discrete chemical master equation. Bulletin of Mathematical Biology, pages 1–45, 2016.

42. Cao, Y., Lu, H.-M., and Liang, J.: Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda. Proceedings of the National Academy of Sciences, 107(43):18445–18450, 2010.

43. Terebus, A., Cao, Y., and Liang, J.: Exact computation of probability landscape of stochastic networks of single input and coupled toggle switch modules. In IEEE Conf. Engineering in Medicine and Biology Society (EMBC), pages 5228–5231. IEEE, 2014.

44. Kao, D., Dungan, J. L., and Pang, A.: Visualizing 2d probability distributions from eos satellite image-derived data sets: A case study. In Proc. Visualization (VIS), pages 457–589. IEEE, 2001.

45. Kao, D., Luo, A., Dungan, J. L., and Pang, A.: Visualizing spatially varying distribution data. In Proc. 6th International Conf. Information Visualisation (IV), pages 219–225. IEEE, 2002.

46. Luo, A., Kao, D., and Pang, A.: Visualizing spatial distribution data sets. In Joint EUROGRAPHICS –IEEE TCVG Symp. Visualization, 2003.

47. Potter, K., Kirby, M., Xiu, D., and Johnson, C. R.: Interactive visualization of probability and cumulative density functions. International Journal for Uncertainty Quantification, 2(4), 2012.

48. Saraiya, P., North, C., and Duca, K.: An insight-based methodology for evaluating bioinformatics visualizations. IEEE Trans. Visualization and Computer Graphics, 11(4):443–456, 2005.

49. Wilkinson, L. and Friendly, M.: The history of the cluster heat map. The American Statistician, 2012.

50. Gove, R., Gramsky, N., Kirby, R., Sefer, E., Sopan, A., Dunne, C., Shneiderman, B., and Taieb-Maimon, M.: Netvisia: Heat map & matrix visualization of dynamic social network statistics & content. In Privacy, Security, Risk and Trust (PASSAT) and IEEE Third International Conf. Social Computing (SocialCom), pages 19–26. IEEE, 2011.

51. Ivanisevic, J., Benton, H. P., Rinehart, D., Epstein, A., Kurczy, M. E., Boska, M. D., Gendelman, H. E., and Siuzdak, G.: An interactive cluster heat map to visualize and explore multidimensional metabolomic data. Metabolomics, 11(4):1029–1034, 2015.

52. Obermaier, H. and Joy, K. I.: Future challenges for ensemble visualization. IEEE Computer Graphics and Applications, 34(3):8–11, 2014.

53. Potter, K., Wilson, A., Bremer, P.-T., Williams, D., Doutriaux, C., Pascucci, V., and Johnson, C. R.: Ensemble-vis: A framework for the statistical visualization of ensemble data. In International Conf. Data Mining Workshops (ICDMW), pages 233–240. IEEE, 2009.

54. Sanyal, J., Zhang, S., Dyer, J., Mercer, A., Amburn, P., and Moorhead, R. J.: Noodles: A tool for visualization of numerical weather model ensemble uncertainty. IEEE Trans. Visualization and Computer Graphics, 16(6):1421–1430, 2010.

55. Wu, K. and Zhang, S.: Visualizing 2d scalar fields with hierarchical topology. In IEEE Symp. Pacific Visualization (PacificVis), pages 141–145. IEEE, 2015.

56. Ferstl, F., Burger, K., and Westermann, R.: Streamline variability plots for characterizing the uncertainty in vector field ensembles. IEEE Trans. Visualization and Computer Graphics, 22(1):767–776, 2016.

57. Bach, B., Dragicevic, P., Archambault, D., Hurter, C., and Carpendale, S.: A review of temporal data visualizations based on space-time cube operations. In Eurographics Conf. Visualization, 2014.

58. Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U.: Network motifs in the transcriptional regulation network of escherichia coli. Nature Genetics, 31(1):64–68, 2002.

59. Wuchty, S., Oltvai, Z. N., and Barabási, A.-L.: Evolutionary conservation of motif constituents in the yeast protein interaction network. Nature Genetics, 35(2):176–179, 2003.

60. Cui, Q., Yu, Z., Purisima, E. O., and Wang, E.: Principles of microrna regulation of a human cellular signaling network. Molecular Systems Biology, 2(1), 2006.

61. Ma, W., Trusina, A., El-Samad, H., Lim, W. A., and Tang, C.: Defining network topologies that can achieve biochemical adaptation. Cell, 138(4):760–773, 2009.

62. Cao, Y. and Liang, J.: Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. BMC Systems Biology, 2(1):30, 2008.

63. Card, S. K., Mackinlay, J. D., and Shneiderman, B.: Readings in information visualization: using vision to think. Morgan Kaufmann, 1999.

64. Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S.: Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic Resonance in Medicine, 34(4):537–541, 1995.

65. Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., and Raichle, M. E.: Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proceedings of the National Academy of Sciences, 103(26):10046–10051, 2006.

66. Vincent, J., Patel, G., Fox, M., Snyder, A., Baker, J., Van Essen, D., Zempel, J., Snyder, L., Corbetta, M., and Raichle, M.: Intrinsic functional architecture in the anaesthetized monkey brain. Nature, 447(7140):83–86, 2007.

67. Calhoun, V., Adali, T., McGinty, V., Pekar, J., Watson, T., and Pearlson, G.: fmri activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. NeuroImage, 14(5):1080–1088, 2001.

68. Beckmann, C. F., DeLuca, M., Devlin, J. T., and Smith, S. M.: Investigations into resting-state connectivity using independent component analysis. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360(1457):1001–1013, 2005.

69. Bullmore, E. and Sporns, O.: The economy of brain network organization. Nature Reviews Neuroscience, 13(5):336–349, 2012.

70. Berger-Wolf, T., Tantipathananandh, C., and Kempe, D.: Dynamic community identification. In Link Mining: Models, Algorithms, and Applications, pages 307–336. Springer, 2010.

71. Holme, P. and Saramäki, J.: Temporal networks. Physics Reports, 519(3):97–125, 2012.

72. Abeles, M. and Gerstein, G. L.: Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. Journal of Neurophysiology, 60(3):909–924, 1988.

73. Girvan, M. and Newman, M. E.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12):7821–7826, 2002.

74. Newman, M. E. and Girvan, M.: Finding and evaluating community structure in networks. Physical Review E, 69(2):026113, 2004.

75. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008, 2008.

76. Tantipathananandh, C., Berger-Wolf, T., and Kempe, D.: A framework for community identification in dynamic social networks. In Proc. 13th ACM SIGKDD International Conf. Knowledge Discovery and Data Mining, pages 717–726. ACM, 2007.

77. Tantipathananandh, C. and Berger-Wolf, T.: Constant-factor approximation algorithms for identifying dynamic communities. In 15th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining, pages 827–836. ACM, 2009.

78. Tantipathananandh, C. and Berger-Wolf, T. Y.: Finding communities in dynamic social networks. In IEEE 11th International Conf. Data Mining (ICDM), pages 1236–1241. IEEE, 2011.

79. Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X.: Group formation in large social networks: membership, growth, and evolu-

tion. In 12th ACM SIGKDD International Conf. Knowledge Discovery and Data Mining, pages 44–54. ACM, 2006.

80. Wasserman, S. and Faust, K.: Social network analysis: Methods and applications, volume 8. Cambridge university press, 1994.

81. Pavlopoulos, G. A., Wegener, A.-L., and Schneider, R.: A survey of visualization tools for biological network analysis. Biodata Mining, 1(1):1, 2008.

82. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11):2498–2504, 2003.

83. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T.: Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics, 27(3):431–432, 2011.

84. Li, K., Guo, L., Faraco, C., Zhu, D., Chen, H., Yuan, Y., Lv, J., Deng, F., Jiang, X., Zhang, T., et al.: Visual analytics of brain networks. NeuroImage, 61(1):82–97, 2012.

85. Jianu, R., Demiralp, Ç., and Laidlaw, D. H.: Exploring brain connectivity with two-dimensional neural maps. IEEE Trans. Visualization and Computer Graphics, 18(6):978–987, 2012.

86. Irimia, A., Chambers, M. C., Torgerson, C. M., and Van Horn, J. D.: Circular representation of human cortical networks for subject and population-level connectomic visualization. NeuroImage, 60(2):1340–1351, 2012.

87. Margulies, D. S., Böttger, J., Watanabe, A., and Gorgolewski, K. J.: Visualizing the human connectome. NeuroImage, 80:445–461, 2013.

88. Christodoulou, E. G., Sakkalis, V., Tsiaras, V., and Tollis, I. G.: Brainnetvis: an open-access tool to effectively quantify and visualize brain networks. Computational Intelligence and Neuroscience, 2011, 2011.

89. Xia, M., Wang, J., and He, Y.: Brainnet viewer: a network visualization tool for human brain connectomics. PloS one, 8(7):e68910, 2013.

90. LaPlante, R. A., Douw, L., Tang, W., and Stufflebeam, S. M.: The connectome visualization utility: Software for visualization of human brain networks. PloS one, 9(12):e113838, 2014.

91. Alper, B., Bach, B., Henry Riche, N., Isenberg, T., and Fekete, J.-D.: Weighted graph comparison techniques for brain connectivity analysis. In SIGCHI Conf. Human Factors in Computing Systems, pages 483–492. ACM, 2013.

92. Lin, C.-Y., Tsai, K.-L., Wang, S.-C., Hsieh, C.-H., Chang, H.-M., and Chiang, A.-S.: The neuron navigator: Exploring the information pathway through the neural maze. In IEEE Symp. Pacific Visualization (PacificVis), pages 35–42. IEEE, 2011.

93. Sorger, J., Buhler, K., Schulze, F., Liu, T., and Dickson, B.: neuromapinteractive graph-visualization of the fruit fly's neural circuit. In IEEE Symp. Biological Data Visualization (BioVis), pages 73–80. IEEE, 2013.

94. Beck, F., Burch, M., Diehl, S., and Weiskopf, D.: The state of the art in visualizing dynamic graphs. EuroVis STAR, 2014.

95. Archambault, D., Purchase, H. C., and Pinaud, B.: Animation, small multiples, and the effect of mental map preservation in dynamic graphs. IEEE Trans. Visualization and Computer Graphics, 17(4):539–552, 2011.

96. Erten, C., Harding, P. J., Kobourov, S. G., Wampler, K., and Yee, G.: Graphael: Graph animations with evolving layouts. In Graph Drawing, pages 98–110. Springer, 2003.

97. Frishman, Y. and Tal, A.: Dynamic drawing of clustered graphs. In IEEE Symp. Information Visualization (InfoVis), pages 191–198. IEEE, 2004.

98. Frishman, Y. and Tal, A.: Online dynamic graph drawing. IEEE Trans. Visualization and Computer Graphics, 14(4):727–740, 2008.

99. Bach, B., Pietriga, E., and Fekete, J.-D.: Graphdiaries: animated transitions andtemporal navigation for dynamic networks. IEEE Trans. Visualization and Computer Graphics, 20(5):740–754, 2014.

100. Beiró, M. G., Busch, J. R., and Alvarez-Hamelin, J. I.: Visualizing communities in dynamic networks. In LAWDN-Latin-American Workshop on Dynamic Networks, pages 4–p, 2010.

101. Shi, L., Wang, C., and Wen, Z.: Dynamic network visualization in 1.5 d. In IEEE Symp. Pacific Visualization (PacificVis), pages 179–186. IEEE, 2011.

102. Trier, M. and Bobrik, A.: Analyzing the dynamics of community formation using brokering activities. In Communities and Technologies, pages 463–477. Springer, 2007.

103. Havre, S., Hetzler, B., and Nowell, L.: Themeriver: Visualizing theme changes over time. In IEEE Symp. Information Visualization (InfoVis), pages 115–123. IEEE, 2000.

104. Havre, S., Hetzler, E., Whitney, P., and Nowell, L.: Themeriver: Visualizing thematic changes in large document collections. IEEE Trans. Visualization and Computer Graphics, 8(1):9–20, 2002.

105. Greilich, M., Burch, M., and Diehl, S.: Visualizing the evolution of compound digraphs with timearctrees. In Computer Graphics Forum, volume 28, pages 975–982. Wiley Online Library, 2009.

106. Falkowski, T., Bartelheimer, J., and Spiliopoulou, M.: Mining and visualizing the evolution of subgroups in social networks. In IEEE/WIC/ACM International Conf. Web Intelligence, pages 52–58. IEEE Computer Society, 2006.

107. Rosvall, M. and Bergstrom, C. T.: Mapping change in large networks. PloS one, 5(1):e8694, 2010.

108. Reda, K., Tantipathananandh, C., Johnson, A., Leigh, J., and Berger-Wolf, T.: Visualizing the evolution of community structures in dynamic social networks. In Computer Graphics Forum, volume 30, pages 1061–1070. Wiley Online Library, 2011.

109. Vehlow, C., Beck, F., Auwärter, P., and Weiskopf, D.: Visualizing the evolution of communities in dynamic graphs. In Computer Graphics Forum, volume 34, pages 277–288. Wiley Online Library, 2015.

110. Gatalsky, P., Andrienko, N., and Andrienko, G.: Interactive analysis of event data using space-time cube. In 8th International Conf. Information Visualisation (IV), pages 145–152. IEEE, 2004.

111. Groh, G., Hanstein, H., and Wörndl, W.: Interactively visualizing dynamic social networks with dyson. In IUI Workshop on Visual Interfaces to the Social and the Semantic Web, volume 2, 2009.

112. Reda, K., Tantipathananandh, C., Berger-Wolf, T., Leigh, J., and Johnson, A.: Socioscape–a tool for interactive exploration of spatio-temporal group dynamics in social networks. In Proc. IEEE Conf. Information Visualization (InfoVis), 2009.

113. Forlines, C. and Wittenburg, K.: Wakame: sense making of multi-dimensional spatial-temporal data. In Proc. International Conf. Advanced Visual Interfaces, pages 33–40. ACM, 2010.

114. Bach, B., Pietriga, E., and Fekete, J.-D.: Visualizing dynamic networks with matrix cubes. In Proc. SIGCHI Conf. Human Factors in Computing Systems, pages 877–886. ACM, 2014.

115. Llano, D. A., Theyel, B. B., Mallik, A. K., Sherman, S. M., and Issa, N. P.: Rapid and sensitive mapping of long-range connections in vitro using flavoprotein autofluorescence imaging combined with laser photostimulation. Journal of Neurophysiology, 101(6):3325–3340, 2009.

116. Kairam, S., MacLean, D., Savva, M., and Heer, J.: Graphprism: compact visualization of network structure. In Proc. International Conf. Advanced Visual Interfaces, pages 498–505. ACM, 2012.

117. Llano, D. A., Turner, J., and Caspary, D. M.: Diminished cortical inhibition in an aging mouse model of chronic tinnitus. The Journal of Neuroscience, 32(46):16141–16148, 2012.

118. Ten Caat, M., Maurits, N. M., and Roerdink, J. B.: Data-driven visualization and group analysis of multichannel eeg coherence with functional units. IEEE Trans. Visualization and Computer Graphics, 14(4):756–771, 2008.

119. Böttger, J., Schäfer, A., Lohmann, G., Villringer, A., and Margulies, D. S.: Three-dimensional mean-shift edge bundling for the visualization of functional connectivity in the brain. IEEE Trans. Visualization and Computer Graphics, 20(3):471–480, 2014.

120. Heer, J. and Boyd, D.: Vizster: Visualizing online social networks. In IEEE Symp. Information Visualization (InfoVis), pages 32–39. IEEE, 2005.

121. Henry, N. and Fekete, J.-D.: Matlink: Enhanced matrix visualization for analyzing social networks. In IFIP Conf. Human-Computer Interaction, pages 288–302. Springer, 2007.

122. Henry, N. and Fekete, J.-D.: Matrixexplorer: a dual-representation system to explore social networks. IEEE Trans. Visualization and Computer Graphics, 12(5), 2006.

123. Henry, N., Fekete, J.-D., and McGuffin, M. J.: Nodetrix: a hybrid visualization of social networks. IEEE Trans. Visualization and Computer Graphics, 13(6):1302–1309, 2007.

124. Bastian, M., Heymann, S., Jacomy, M., et al.: Gephi: an open source software for exploring and manipulating networks. ICWSM, 8:361–362, 2009.

125. Rufiange, S. and McGuffin, M. J.: Diffani: Visualizing dynamic graphs with a hybrid of difference maps and animation. IEEE Trans. Visualization and Computer Graphics, 19(12):2556–2565, 2013.

126. Lindemann, F., Laukamp, K., Jacobs, A. H., and Hinrichs, K. H.: Interactive comparative visualization of multimodal brain tumor segmentation data. In Vision, Modeling, and Visualization, pages 105–112. Citeseer, 2013.

127. Rubenstein, D. I., Sundaresan, S. R., Fischhoff, I. R., Tantipathananandh, C., and Berger-Wolf, T. Y.: Similar but different: Dynamic social network analysis highlights fundamental differences between the fission-fusion societies of two equid species, the onager and grevys zebra. PloS one, 10(10):e0138645, 2015.

128. Marai, G. E.: Visual Scaffolding in Integrated Spatial and Nonspatial Analysis. In EuroVis Workshop on Visual Analytics (EuroVA), eds. E. Bertini and J. C. Roberts. The Eurographics Association, 2015.

129. Someren, M. v., Barnard, Y. F., Sandberg, J. A., et al.: The think aloud method: a practical approach to modelling cognitive processes. Academic Press, 1994.

130. Luciani, T. B., Cherinka, B., Oliphant, D., Myers, S., Wood-Vasey, W. M., Labrinidis, A., and Marai, G. E.: Large-scale overlays and trends: Visually mining, panning and zoomingthe observable universe. IEEE Trans. Visualization and Computer Graphics, 20(7):1048–1061, 2014.

131. Yang, X., Shi, L., Daianu, M., Tong, H., Liu, Q., and Thompson, P.: Blockwise human brain network visual comparison using nodetrix representation. <u>IEEE Trans. Visualization and Computer Graphics</u>, 23(1):181–190, 2017.

132. Song, H., Lee, J., Kim, T. J., Lee, K. H., Kim, B., and Seo, J.: Gazedx: Interactive visual analytics framework for comparative gaze analysis with volumetric medical images. <u>IEEE Trans. Visualization and Computer Graphics</u>, 23(1):311–320, 2017.

133. Fujiwara, T., Chou, J.-K., McCullough, A. M., Ranganath, C., and Ma, K.-L.: A visual analytics system for brain functional connectivity comparison across individuals, groups, and time points. In <u>IEEE Symp. Pacific Visualization (PacificVis)</u>, pages 250–259. IEEE, 2017.

134. Grottel, S., Reina, G., Vrabec, J., and Ertl, T.: Visual verification and analysis of cluster detection for molecular dynamics. <u>IEEE Trans. Visualization and Computer Graphics</u>, 13(6):1624–1631, 2007.

135. Lindow, N., Baum, D., Bondar, A., and Hege, H.-C.: Dynamic channels in biomolecular systems: Path analysis and visualization. In <u>IEEE Symp. Biological Data Visualization (BioVis)</u>, pages 99–106. IEEE, 2012.

136. Parulek, J., Turkay, C., Reuter, N., and Viola, I.: Implicit surfaces for interactive graph based cavity analysis of molecular simulations. In <u>IEEE Symp. Biological Data Visualization (BioVis)</u>, pages 115–122. IEEE, 2012.

137. Patro, R., Ip, C. Y., Bista, S., Cho, S. S., Thirumalai, D., and Varshney, A.: Mdmap: A system for data-driven layout and exploration of molecular dynamics simulations. In <u>IEEE Symp. Biological Data Visualization (BioVis)</u>, pages 111–118. IEEE, 2011.

138. de Heras Ciechomski, P., Klann, M., Mange, R., and Koeppl, H.: From biochemical reaction networks to 3d dynamics in the cell: The zigcell3d modeling, simulation and visualisation framework. In <u>IEEE Symp. Biological Data Visualization (BioVis)</u>, pages 41–48. IEEE, 2013.

139. Strobelt, H., Bertini, E., Braun, J., Deussen, O., Groth, U., Mayer, T. U., and Merhof, D.: Hitsee knime: a visualization tool for hit selection and analysis in high-throughput screening experiments for the knime platform. <u>BMC bioinformatics</u>, 13(Suppl 8):S4, 2012.

140. de Leeuw, W., Verschure, P. J., and van Liere, R.: Visualization and analysis of large data collections: a case study applied to confocal microscopy data. IEEE Trans. Visualization and Computer Graphics, 12(5):1251–1258, 2006.

141. Oeltze, S., Freiler, W., Hillert, R., Doleisch, H., Preim, B., and Schubert, W.: Interactive, graph-based visual analysis of high-dimensional, multi-parameter fluorescence microscopy data in toponomics. IEEE Trans. Visualization and Computer Graphics, 17(12):1882–1891, 2011.

142. Pretorius, A., Zhou, Y., and Ruddle, R.: Visual parameter optimisation for biomedical image processing. BMC bioinformatics, 16(Suppl 11):S9, 2015.

143. Ding, H., Wang, C., Huang, K., and Machiraju, R.: Graphie: graph based histology image explorer. BMC bioinformatics, 16(Suppl 11):S10, 2015.

144. Carley, K. M.: Dynamic network analysis. Citeseer, 2003.

145. Stolte, C., Tang, D., and Hanrahan, P.: Multiscale visualization using data cubes. IEEE Trans. Visualization and Computer Graphics, 9(2):176–187, 2003.

146. Lex, A., Streit, M., Kruijff, E., and Schmalstieg, D.: Caleydo: Design and evaluation of a visual analysis framework for gene expression data in its biological context. In IEEE Symp. Pacific Visualization (PacificVis), pages 57–64. IEEE, 2010.

147. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A.: Circos: an information aesthetic for comparative genomics. Genome Research, 19(9):1639–1645, 2009.

148. Meyer, M., Wong, B., Styczynski, M., Munzner, T., and Pfister, H.: Pathline: A tool for comparative functional genomics. In Computer Graphics Forum, volume 29, pages 1043–1052. Wiley Online Library, 2010.

149. Sedlmair, M., Meyer, M., and Munzner, T.: Design study methodology: Reflections from the trenches and the stacks. IEEE Trans. Visualization and Computer Graphics, 18(12):2431–2440, 2012.

150. McKenna, S., Mazur, D. C., Agutter, J., and Meyer, M.: Design activity framework for visualization design. IEEE Trans. Visualization and Computer Graphics, 20(12):2191–2200, 2014.

151. Munzner, T.: A nested model for visualization design and validation. IEEE Trans. Visualization and Computer Graphics, 15(6):921–928, 2009.

152. Ma, C., Kenyon, R. V., Forbes, A. G., Berger-Wolf, T., Slater, B. J., and Llano, D. A.: Visualizing dynamic brain networks using an animated dual-representation. In Proc. Eurographics Conf. Visualization (EuroVis), pages 73–77, 2015.

153. Benesty, J., Chen, J., Huang, Y., and Cohen, I.: Pearson correlation coefficient. In Noise Reduction in Speech Processing, pages 1–4. Springer, 2009.

154. Oeltze, S., Doleisch, H., Hauser, H., Muigg, P., and Preim, B.: Interactive visual analysis of perfusion data. IEEE Trans. Visualization and Computer Graphics, 13(6):1392–1399, 2007.

155. Hamilton, N. A., Wang, J. T., Kerr, M. C., and Teasdale, R. D.: Statistical and visual differentiation of subcellular imaging. BMC bioinformatics, 10(1):94, 2009.

156. Herghelegiu, P.-C., Manta, V., Perin, R., Bruckner, S., and Gröller, E.: Biopsy planner–visual analysis for needle pathway planning in deep seated brain tumor biopsy. In Computer Graphics Forum, volume 31, pages 1085–1094. Wiley Online Library, 2012.

157. Mlejnek, M., Ermest, P., Vilanova, A., Van der Rijt, R., Van den Bosch, H., Gerritsen, F., and Groller, M. E.: Profile flags: a novel metaphor for probing of t/sub 2/maps. In IEEE Visualization (VIS). IEEE, 2005.

158. Mlejnek, M., Ermes, P., Vilanova, A., van der Rijt, R., van den Bosch, H., Gerritsen, F. A., and Gröller, M. E.: Application-oriented extensions of profile flags. In EuroVis, pages 339–346, 2006.

159. Muigg, P., Kehrer, J., Oeltze, S., Piringer, H., Doleisch, H., Preim, B., and Hauser, H.: A four-level focus+ context approach to interactive visual analysis of temporal features in large scientific data. In Computer Graphics Forum, volume 27, pages 775–782. Wiley Online Library, 2008.

160. Streit, M., Schulz, H.-J., Lex, A., Schmalstieg, D., and Schumann, H.: Model-driven design for the visual analysis of heterogeneous data. IEEE Trans. Visualization and Computer Graphics, 18(6):998–1010, 2012.

161. Cedilnik, A., Baumes, J., Ibanez, L., Megason, S., and Wylie, B. N.: Integration of information and volume visualization for analysis of cell lineage and gene expression during embryogenesis. In Visualization and Data Analysis, page 68090J, 2008.

162. Blaas, J., Botha, C. P., and Post, F. H.: Interactive visualization of multi-field medical data using linked physical and feature-space views. In EuroVis, pages 123–130, 2007.

163. Wang, C. and Shen, H.-W.: Lod map-a visual interface for navigating multiresolution volume visualization. IEEE Trans. Visualization and Computer Graphics, 12(5), 2006.

164. Oeltze, S., Kuß, A., Grothues, F., Hennemuth, A., and Preim, B.: Integrated visualization of morphologic and perfusion data for the analysis of coronary artery disease. In EuroVis, pages 131–138, 2006.

165. Raidou, R. G., van der Heide, U. A., Dinh, C. V., Ghobadi, G., Kallehauge, J. F., Breeuwer, M., and Vilanova, A.: Visual analytics for the exploration of tumor tissue characterization. In Computer Graphics Forum, volume 34, pages 11–20. Wiley Online Library, 2015.

166. Zachow, S., Muigg, P., Hildebrandt, T., Doleisch, H., and Hege, H.-C.: Visual exploration of nasal airflow. IEEE Trans. Visualization and Computer Graphics, 15(6):1407–1414, 2009.

167. Weber, G. H., Rubel, O., Huang, M.-Y., DePace, A. H., Fowlkes, C. C., Keranen, S. V., Luengo Hendriks, C. L., Hagen, H., Knowles, D. W., Malik, J., et al.: Visual exploration of three-dimensional gene expression using physical views and linked abstract views. IEEE/ACM Trans. Computational Biology and Bioinformatics (TCBB), 6(2):296–309, 2009.

168. Krekel, P. R., Valstar, E. R., De Groot, J., Post, F. H., Nelissen, R. G., and Botha, C. P.: Visual analysis of multi-joint kinematic data. In Computer Graphics Forum, volume 29, pages 1123–1132. Wiley Online Library, 2010.

169. Pretorius, A. J., Khan, I. A., and Errington, R. J.: Cell lineage visualisation. In Computer Graphics Forum, volume 34, pages 21–30. Wiley Online Library, 2015.

170. Keefe, D., Ewert, M., Ribarsky, W., and Chang, R.: Interactive coordinated multiple-view visualization of biomechanical motion data. IEEE Trans. Visualization and Computer Graphics, 15(6):1383–1390, 2009.

171. Bak, P., Mansmann, F., Janetzko, H., and Keim, D.: Spatiotemporal analysis of sensor logs using growth ring maps. IEEE Trans. Visualization and Computer Graphics, 15(6), 2009.

172. Tominski, C., Schulze-Wollgast, P., and Schumann, H.: 3d information visualization for time dependent data on maps. In Proc. Ninth International Conf. Information Visualisation (IV), pages 175–181. IEEE, 2005.

173. Zhang, Z., Wang, B., Ahmed, F., Ramakrishnan, I., Zhao, R., Viccellio, A., and Mueller, K.: The five ws for information visualization with application to healthcare informatics. IEEE Trans. Visualization and Computer Graphics, 19(11):1895–1910, 2013.

174. Schroeder, D., Korsakov, F., Knipe, C. M.-P., Thorson, L., Ellingson, A. M., Nuckley, D., Carlis, J., and Keefe, D. F.: Trend-centric motion visualization: Designing and applying a new strategy for analyzing scientific motion collections. IEEE Trans. Visualization and Computer Graphics, 20(12):2644–2653, 2014.

175. Klemm, P., Oeltze-Jafra, S., Lawonn, K., Hegenscheid, K., Völzke, H., and Preim, B.: Interactive visual analysis of image-centric cohort study data. IEEE Trans. Visualization and Computer Graphics, 20(12):1673–1682, 2014.

176. Masood, T. B., Sandhya, S., Chandra, N., and Natarajan, V.: Chexvis: a tool for molecular channel extraction and visualization. BMC bioinformatics, 16(1):119, 2015.

177. Lindow, N., Baum, D., Bondar, A.-N., and Hege, H.-C.: Exploring cavity dynamics in biomolecular systems. BMC bioinformatics, 14(S19):S5, 2013.

178. Krone, M., Reina, G., Schulz, C., Kulschewski, T., Pleiss, J., and Ertl, T.: Interactive extraction and tracking of biomolecular surface features. In Computer Graphics Forum, volume 32, pages 331–340. Wiley Online Library, 2013.

179. Krone, M., Kauker, D., Reina, G., and Ertl, T.: Visual analysis of dynamic protein cavities and binding sites. In IEEE Symp. Pacific Visualization (PacificVis), pages 301–305. IEEE, 2014.

180. Bryden, A., Phillips, G., and Gleicher, M.: Automated illustration of molecular flexibility. IEEE Trans. Visualization and Computer Graphics, 18(1):132–145, 2012.

181. Sarikaya, A., Albers, D., Mitchell, J., and Gleicher, M.: Visualizing validation of protein surface classifiers. In Computer Graphics Forum, volume 33, pages 171–180. Wiley Online Library, 2014.

182. Byška, J., Jurčík, A., Gröller, M. E., Viola, I., and Kozlíková, B.: Molecollar and tunnel heat map visualizations for conveying spatio-temporo-chemical properties across and along protein voids. In Computer Graphics Forum, volume 34, pages 1–10. Wiley Online Library, 2015.

183. Palmas, G., Bachynskyi, M., Oulasvirta, A., Seidel, H.-P., and Weinkauf, T.: Movexp: A versatile visualization tool for human-computer interaction studies with 3d performance and biomechanical data. IEEE Trans. Visualization and Computer Graphics, 20(12):2359–2368, 2014.

NAME:	Chihua Ma
EDUCATION:	B.S., Computer Science and Technology, Nankai University, China, 2009 M.S., Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, 2011
EXPERIENCE:	Data Visualization Scientist, Decision Science, Conversant LLC, 2018-present Research Assistant, Electric Visualization Laboratory, University of Illinois at Chicago, 2015-2017 Teaching Assistant, Computer Science Department, University of Illinois at Chicago, 2012-2015 Software Engineering Intern, CME Group, 2013 Graduate Assistant, Office of Social Science Research, University of Illinois at Chicago, 2011-2012
AWARDS:	IS&T 2017 Charles E. Ives Journal Award for Best Paper in Journal of Imaging Science and Technology. 2017 IEEE VGTC Visualization Pioneers Group (VPG) Data Visualization Contest Honorable Mention. 2016 IEEE VIS 2016 Doctoral Colloquium Participant - selected by a committee and had travel & registration paid. 2016 UIC Student Presenter Award to assist w/ travel expense to a conference. 2016 UIC Graduate Student Council Travel Award. 2016 2-week Scholarship to Participate in Mining and Modeling of Neuroscience Data Course at UC Berkeley. 2015
PUBLICATIONS:	**Journal Publications** C. Ma, F. Pellolio, D.A. Llano, R.V. Kenyon, and G. E. Marai. RemBrain: Exploring Dynamic Biospatial Networks with Mosaic-Matrices and Mirror Glyphs. Journal of Imaging Science and Technology, 61(6), January 2016, pp. 060404-1-060404-13 (2017). C. Ma, T. Luciani, A. Terebus, J. Liang, and G. E. Marai. PRODIGEN: Visualizing the Probability Landscape of Stochastic Gene Regulatory Networks in State and Time Space. BMC Bioinformatics, 18(2): 24, 2017. C. Ma, A. Forbes, D.A. Llano, T. Berger-Wolf, and R.V. Kenyon.

SwordPlots: Exploring Neuron Behavior within Dynamic Communities of Brain Networks. Journal of Imaging Science and Technology, 60(1): 10405-1-10405-13, 2016. **Charles E. Ives Journal Award**

Peer-reviewed Conference Papers
D. McNamara, J. Tapia, C. Ma, T. Luciani, A. Burks, J. Trelles, and G. E. Marai. Spatial Analysis of Employee Safety Using Organizable Event Quiltmaps. IEEE VIS 2016 The Event Event: Temporal and Sequential Event Analysis Workshop, pp. 1-4, Oct. 2016.
C. Ma, R.V. Kenyon, A. Forbes, T. Berger-Wolf, B.J. Slater, and D.A. Llano. Visualizing Dynamic Brain Networks Using an Animated Dual-Representation. In Proceedings of the Eurographics Conference on Visualization (EuroVis'15 Short papers), pp. 73-77, May 2015.
C. Ma, S. Liberman, and H. Zheng. GCLViz: Garbage Collection vs. Latency Visualization. In Proceedings of the 5th International Conference on Information Visualization Theory and Applications (IVAPP), pp. 292-299, Jan 2014.

Poster/Abstract Presentations
D. Kirilov, I. Lindmae, A. Burks, C. Ma, and G. E. Marai. MC1: A Bespoke Analysis Tool for Spatio-temporal Park Traffic Data. IEEE Visual Analytics Science and Technology (VAST) Challenge 2017 Proceedings, pp 1-2, 2017.
C. Ma, A. Burks, T. Luciani, A. Terebus, J. Liang, and G. E. Marai. Visualizing ensemble time-evolving probability landscapes of stochastic networks. ISMB/ECCB 2017 (BioVis'17 Poster), pp. 1-2, 2017.
T. Luciani, J. Trelles, C. Ma, A. Burks, M. M. Thomas, K. Bharadwaj, S. Singh, P. Hanula, L. Di and G. E. Marai. Multi-scale Voronoi-based ACT Assessment. IEEE VIS VPG Data Contest, pp. 1-2, Oct 2016. **Honorable Mention**
T. Luciani, C. Ma, J. Trelles, and G. E. Marai. Developing a Data-Driven Wiki of Spatial-Nonspatial Integration Tools. IEEE VIS C4PGV Workshop, pp.1-1, Oct 2016.
C. Ma, R.V. Kenyon, T. Berger-Wolf, and D.A. Llano. Visualizing Communities in Dynamic Mouse Brain Networks. In Proceedings of the IEEE Information Visualization Conference (InfoVis'14 Poster), Nov 2014.

INVITED TALK: "Visual Analysis Techniques for Dynamic Biological Systems", Brookhaven National Lab, NY, USA, April 3, 2017
"Introduction to Data Visualization and Visual Analytics", Nankai University, Tianjin, China, June 13, 2014