Iterative Exploration of Big Brain Network Data

Florian Ganglberger[1], Nicolas Swoboda[1], Lisa Frauenstein[1], Joanna Kaczanowska[2], Wulf Haubensak[2] and Katja Bühler[1]

[1]VRVIS Research Center
[2]Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC)

Abstract

A current quest in neuroscience is the understanding of how genes, structure and behavior relate to one another. In recent years, big brain-initiatives and consortia have created vast resources of publicly available brain data that can be used by neuroscientists for their own research experiments. This includes microscale connectivity data - brain-network graphs with billions of edges - whose analysis for higher order relations in structural or functional neuroanatomy together with genetic data may reveal novel insights into brain functionality. This creates a need for joint exploration of spatial data, such as gene expression patterns, whole brain gene co-expression correlation, structural and functional connectivities together with neuroanatomical parcellations. Current experimental workflows involve time-consuming manual aggregation and extensive graph theoretical analysis of data from different sources, which rarely provide spatial context to operate continuously on different scales.
We propose a web-based framework to explore heterogeneous neurobiological data in an integrated visual analytics workflow. On-demand queries on spatial gene expression and connectivity data enable an interactive dissection of dense network graphs - with of billion-edges on voxel-resolution - in real-time based on their spatial context. In order to take higher order connections between brain regions into account, queries can be executed in a cascading way. Relating data to the hierarchical structure of common anatomical atlases allows experts to quantitatively compare multimodal networks on different scales. Additionally, 3D visualizations have been optimized to accommodate for the domain experts' need for publishable network figures.
We demonstrate the relevance of our approach for neuroscience by exploring social-behavior and memory/learning functional neuroanatomy in mice.

---

1. Introduction

The quest for understanding the principle organization of the brain and its functional parcellation is constantly changing due to the increasing wealth of multimodal neurobiological data generated by brain initiatives, such as the Allen Institute [all], the Human Brain Project [hum], the WU-Minn Human Connectome Project [VESB*13], and the China Brain Project [mPlDI*16]. They offer rich information about genes, structure and behavior, while understanding their relationship is a key factor in neurocircuit research. Mining these resources can provide researchers with additional context for their experiments. The major challenges for visual analytics in neuroscience involves accessing, fusing and visualizing spatial brain data such as brain-wide gene expression, structural and functional connectivity, and non spatial data like gene lists related to behavior or the functional association of genes.

Our work takes recent advances in circuit neuroscience into account (e.g. neuro- and behavioral genetics, optogenetics, imaging) that identified gene sets underlying a specific behavioral function [KZM*17]. However, there is a lack of tools to explore the mesoscale as well as the global structural and functional brain networks related to these gene sets in silico. We meet this demand by

proposing a framework allowing visualization, iterative exploration and the integration of spatial data like imaging data showing brain-wide gene expressions, fMRI, or structural data, with structural, functional and genetic anatomical relations at different scales and different hierarchical anatomical labels. This data can be retrieved from public resources and integrated in our framework. The same applies to private data generated during experiments in the lab.

The data we handle includes hierarchical parcellations and structural annotations (e.g. Allen Mouse Brain and Allen Human Brain region annotations), classical 3D image data (e.g. brain template, spatial gene expression), 3D aggregated data (e.g. Allen Mouse Brain gene expression data or viral injection traces), region-wise connectomes/relations (e.g. resting state functional connectivity) and voxel-resolution connectomes/relations (structural connectivity, gene co-expression), all aligned to a common reference space. Voxel-wise connectivity/relational data generally has a very dense connectivity with up to billions of connections, so their matrices can take up hundreds of gigabytes.

Our framework visualizes volumetric, geometry and graph data simultaneously in 3D rendering and 2D slice views, linked to views showing quantitative profiles at a hierarchical parcellation level.

The user can interactively navigate the hierarchical levels, which provide spatial context by rendering their brain regions in 2D/3D views. In a previous paper [GKHB18], we described a specialized data structure that organizes and aggregates the voxel-wise connectivity data hierarchically. Via so called *Aggregation Queries* [GKHB18], which represent the aggregated incoming/outgoing connectivity of volumes of interest (*VOI*), the data is made accessible in real time and can be explored interactively on different anatomical scales. The hierarchical organization is anatomy-driven. It can be flexibly generated for different ontologies and their respective spatial region annotations.

In this paper, we used interactive local 3D selection on visualized data to define *VOIs* with high intensity. Based on these, *Aggregation Queries* can be executed on a user selected connectivity data set. The result is the cumulative voxel-wise connectivity of the selected *VOI* that, again is visualized as intensity volume in the 3D rendering as well as quantitative brain-region-wise profile. These are augmented with hierarchical information (i.e. intensity of sub-regions) to provide an overview of different hierarchical levels. This kind of interaction allows the researcher to relate integrated resources, for example incoming/outgoing connectivity at voxel-level and region level, directly to imaging data showing brain-wide gene expressions (spatial gene expression data). Higher order network connections can be targeted by repeatedly cascading over the query results, which can be selected directly in the brain-region-wise profile. This represents an iterative exploration of the networks. The query results can further be visualized as graph representation rendered as 3D and 2D network graphs. This reduces the complexity of voxel-level results to region-level graphs and provides neuroanatomical context, so they can be presented in a way that is particularly suited for neuroscience publications, as advised by domain experts. We integrated the framework components into workflows for interactively fusing volumetric with connectivity data, brain network exploration at different anatomical levels, and intuitive 3D graph visualization. We demonstrate the practical significance of this tool by presenting several use cases based on heterogeneous neurobiological data from large scale brain initiatives that allowed to reproduce several recent biological findings.

Our previous work enabled us to realize, for the first time, methods for interactive exploration of dense voxel-wise brain connectivity data of several gigabytes of size. Based on this, we propose additional visual analytic methods for interactive:

- joint exploration and fusion of brain network data over different scales ranging from multi-resolution voxel-wise connectivity to connectivity defined in respect to different anatomical region hierarchy levels
- joint exploration of data from different brain network types (e.g. structural, functional, gene co-expression)
- identification of higher order connectivity in dense network graphs with billions of edges

2. Background

We have been collaborating with experts interested in neuronal circuits in the mouse brain, how they control emotional states and behavior, and how they are modulated by genes and psychoactive drugs [HKC*10]. The data they use and the related use cases provide a good starting point for understanding how nowadays data-driven research is done in neurocircuit science. Based on previous projects, informal interviews and discussions with our collaborating domain experts we jointly identified the following major analytical workflows and user stories:

Relating spatial gene expression data to different kinds of connectivity: The entry point for many data analytics workflows are 'candidate regions' - brain regions that may be part of a specific brain circuit. Those are affected by genes that are either related to a certain behavior, or targeted by a psychoactive drug. Thus, the knowledge of where a gene affects the brain is a first step in relating it to a particular function. The effect is rather broadly defined. Well documented cases are primary gene expression sites [LHA*07] (sites where the gene creates products, such as proteins) and structural connectivity (regions to which primary gene expression sites project) [OHN*14]. Spatial gene expression and structural connectivity data for the mouse brain is provided e.g. by the Allen Mouse Brain Atlas (*AMBA*) at a voxel-level, that was published by the Allen Institute in 2007 and 2012. This data has been studied for this purpose before [LHA*07]. Here, the primary expression sites of a gene are made available as 8 bit intensity volumes, representing the density of cells expressing the gene (i.e. spatial gene expression data), related to a specific standard brain. To use those resources for hypotheses building, it is necessary to aggregate data manually by querying online databases and literature research. Therefore, an integrated workflow for data fusion is therefore missing. Identifying brain regions that are functionally or structurally connected (either directly or transitively) to this expression site may also contribute to this function or be a second order effect thereof. Performing this tasks at a voxel-level is a particular challenge, since connectivity matrices can easily grow up to hundreds of gigabytes.

Comparing different types of connectivity is of essential importance for identifying neural circuits. For example, two brain regions can have a high structural connectivity (a connection via neurons) but do not necessarily express the same genes (e.g. a so called ligand-receptor binding [YW04]). Therefore, their genetic connectivity for a specific gene set, represented as gene co-expression correlation (the correlation of the gene-expression between two regions) is low. Finding a correlation between regions in their fMRI activity may show their functional relation, but a difference in gene co-expression correlation could reveal that this activity relies on completely different molecular mechanisms. Directional structural connectivity can reveal the flow of information for (undirectional) fMRI connectivity [GSB*15]. The possibilities of brain network comparisons are versatile, but the given examples are in particular relevant for domain experts. Combining specific circuits for larger networks can be done again by manual data aggregation, such as whole brain fMRI studies, but require the expertise of a bioinformatitian.

Exploring the data on different scales: Operating region-wise on the data (for example a region-wise network graph) depends on spatial hierarchy. Brain regions are organized by ontologies as trees with larger regions at the top, and ultimately resolves into voxel-resolution at the lowest level of the related standard brain. In order to keep an overview while comparing global networks
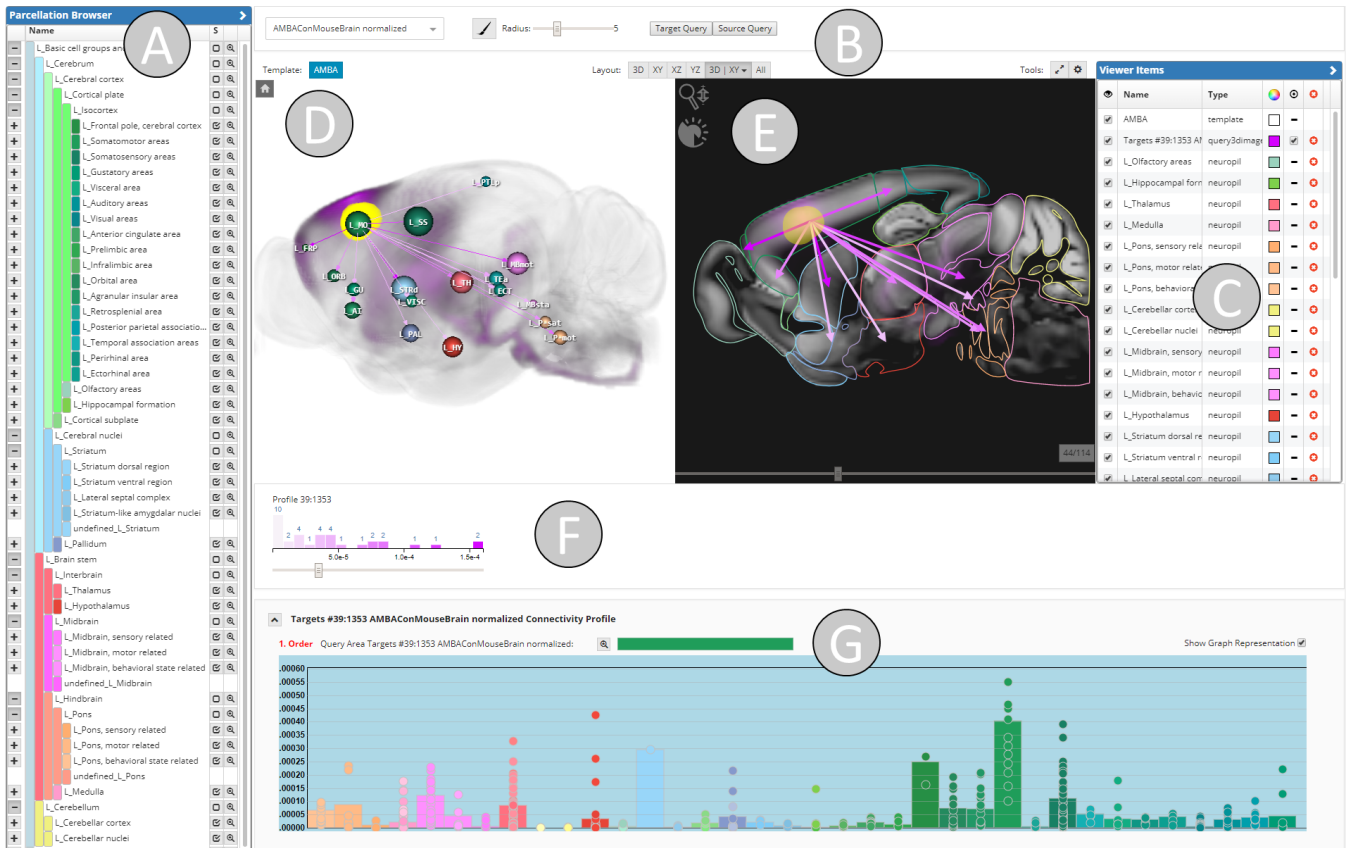
Figure 1: *Web-component for iterative exploration of multiscale brain data. **A)** Parcellation Browser showing a hierarchical (anatomical) brain parcellation. This controls the parcellation level of visualized graphs and profiles. **B)** Query Toolbar to start visual queries for target/source connectivity on different connectivity matrices. **C)** Viewer Items List shows all volumes/geometries visualized in 2D and 3D. **D)** 3D View for rendering volumetric (spatial gene expression and connectivity) data, brain region geometry and network graphs **E)** 2D slice view of 3D representation. **F)** Filtering Toolbar showing histograms of the visualized graphs and threshold sliders for edge filtering. **G)** Expression/Connectivity Profile providing a region-wise quantitative representation of volumetric data.*

with different modalities, larger regions are preferred. On the other hand, visualizing small subnetworks for circuit dissection requires smaller regions of the hierarchy or even voxel-level. A hierarchical organization of the network graph can therefore replace time-consuming pre-selection of relevant regions in current experimental workflows.

3. Related Work

Joint exploration of neurobiological data on different scales:

As various studies have already shown for the mammalian brain, combining information of the macroscale connectome with microscale neuronal architecture does provide a deeper understanding of the brain's organization [SSDV14]. Recently, this has also been confirmed for the living human brain, associating resting state functional connectivity with both gene expression [PLK*15] as well as cortical microstructure such as T1-based myelin content, as obtained from ultrahigh-resolution MRI [HBG*17]. This emphasizes the growing need for visualization tools which incorporate

data of multiple scales. To visualize the relationship between multiscale data, mentioned studies primarily utilize enhanced heatmaps, as it can be seen in Scholtens et al. [SSDV14], where the matrix entries depict anatomical projections. Structural connectivity strength is encoded by color, projection distance by dot size. To provide spatial context, the matrix is presented side-by-side to a brain surface, color-coding anatomical areas. This approach provides a good overview on region-level. However its complexity increases on higher resolution, such as voxel-level.

To integrate macroscopic data at a brain region level with microscale data from simulation of neural activity, Nowke et al. [NSvA*13] have introduced the interactive analytics tool VisNEST, for Macaque monkeys visual cortex. They provide views for visualizing connectivity between brain regions, within-region connectivity representation and time-varying activities across regions. Regions of interest can be selected in a 3D anatomical view and are represented as meshes, where color and opacity depict activity. Although VisNEST allows comparison of region-level connectivity,

it is not tailored for high-resolution connectivity data nor does it relate to volumetric data.

For Drosophila, Lin et al. [LTW*11] introduced Neuron Navigator, a tool that allows queries for connectivity of regions of interest in the Drosophila brain space. These spatial queries are achieved by accessing a 3D neuron image database and matching the region of interest with annotated locations of neuron terminals. Although these queries represent connectivity between neurons, they are not executed on actual graph/network data.

BrainExplorer [LNT*08] is an interactive visualization tool, provided by the Allen Brain Institute to explore the spatial gene expression and structural connectivity data provided by the Allen Mouse Brain Atlas. Brain Explorer provides volume rendering and blob-like visualization of gene expressions mapped to a standard brain. Voxels are colored according to their expression level. The tool enables the execution of pre-computed source/target connectivity queries of brain regions, which is also available via a web-interface on the Allen Institute's website [all]. This work comes closest to our solution. However, they provide (pre-computed) queries for incoming/outgoing connections only on pre-defined sites or anatomical regions. Other tools allowing exploration of connectomic data are CATMAID [SCHT09] and ConnectomeExplorer [BAAK*13], which allow tracing neurons on single EM Stacks, so that they work at a local level of a single network with fixed scale. Similar accounts for Sherbondy et al. [SAM*05], who query pre-computed pathways on diffusion tensor imaging data from volumes of interest. NeuroLines [AABS*14] offers a visualization as simplified skeleton graph, similar to a 2D subway map. Different levels of abstraction allow for multi-scale exploration, but is ultimately tailored to branched, tree like connections in electron microscopy data.

A more generic framework, named neuroVIISAS [SE12] provides multiple ways of organizing, visualizing and analyzing multi-scale brain network simulation data. NeuroVIISAS offers a collection of visualization techniques in 2D such as different representations of hierarchical connectivity matrices, circular connectograms, various layouts for planar graphs, as well as in 3D, where regions are rendered as spheres or true anatomical surface meshes, colored based on the color defined in the reference ontology. Connections are depicted as tubes with or without arrows. Moreover, individual source and target queries can be performed textually by defining filters on a table of all the available connections. The resulting selections can be viewed side-by-side in 2D atlas views. However, as this tool is designed for analysis of simulation data it has a different scope.

Visualization of structural, functional and gene co-expression brain networks: Brain networks are commonly represented as node-link diagrams, connectivity matrices or connectograms, where only the first can be visualized in anatomical space both in 2D, with slice views of the anatomical planes and 3D. In volumetric space, edges have been rendered as straight lines in between nodes depicted as spheres across the brain [XWH13, LDTS14] or curved lines along the brain surface [LFG*15]. However, with a dense number of connections, this produces clutter and obscures the underlying anatomy. Mostly, brain networks are too complex to be able to show all

details up to the highest level of resolution whilst seeing the entire global structure. Thus, hierarchies can define the level of partition. Applied to rendering in 3D space, the hierarchical level can be determined by the distance to the viewpoint in a level-of-detail-visualization [BD07], edge bundling [BSL*14]. Hence, brain network visualization tools [MBB*16] often present network structure and anatomical context side-by-side and achieve reference through linked views and coloring network nodes according to their anatomical region. This provides the flexibility to layout the connectivity network according to network properties such as using a spring-embedded layout where well-connected groups of nodes are pulled together as seen in [PLK*15]. Different graph layouts, such as an anatomical layout, are also supported by NeuroMap [Sor13], which renders potential neuronal connections in the fruit fly's brain as interactive circuit-style wiring diagrams. Spatial context is provided by introducing brain regions as spatial constraints to a 2D layout by providing a 2D abstraction of anatomical organizations and by linking NeuroMap with 3D visualizations, showing the neurons in their 3D anatomical context. Although these tools provide spatial/anatomical context, they are limited to a regional level, and do not scale to voxel-level resolution with billions of edges.

4. Core Tasks

Based on the user stories described in Section 2 we identified the following core tasks and requirements to be supported by our system. It is necessary that all tasks can be performed interactively i.e. without time-consuming computation:

R1: 3D target/source assessment: *What is connected to a certain volume of interest?* The user has 3D volume data that shows activity in certain areas, for example where a certain gene is expressed. She wants to know which areas are connected via structure or by function at a voxel-level. She can directly select the areas that are interesting for her based on the visualization and receives the results instantly. The user can select additional connectivity modalities, which she can visualize similarly or by their overlap.

R2: Higher order target/source assessment: *What is transitively connected to a certain volume of interest?* Originating from the target/sources of R1, the user wants to know what further connections exist i.e. which brain regions (also on a voxel-level) are transitively connected to a certain area. Using R1 iteratively is simply too time consuming for the user, since it involves the accurate selection of potentially large brain areas.

R3: Anatomical context: *To which brain regions is my data related?* 3D volume data, such as spatial gene expression data or target/source connectivity query results need to be related to their anatomical context, so that the user has quantified information of what she is seeing. Therefore, the volume data needs to be summarized to show how much gene expression/connectivity individual brain regions exhibit. The hierarchical nature of the regions makes it necessary to navigate/visualize different hierarchical levels.

R4: Explore 3D network graphs with spatial reference: *How do my networks look like on different scales?* The user wants to explore connectivity of experimental data, such as a fMRI network of a study. She executes target/source queries on the fMRI network and

other modalities and renders a graph representation of the results. She wants to navigate hierarchical brain parcellation until the graph fits her needs. Multiple connectivities need to be represented as a graph for purposes of comparison, i.e. to see the difference from and the overlap with her experiment.

R5: Publishable figures: *Give me a publishable figure!* The region-wise network graph, generated by R4 needs to be visualized in a way with which neuroscientists feel intuitively familiar. The figure needs to be easy to understand, interpretable and suitable for neuroscientific publications/presentations. The domain experts consider the best solution to be a 3D graph visualization overlayed by the outlines of the brain. Furthermore, the workflow for generating the figure has to be traced in order to be reproducible.

5. A Data Structure for real-time Aggregation Queries of Big Brain Networks

Querying large graphs, with billions of edges can be time consuming. This data, represented as connectivity matrices (weighted adjacency matrices) can easily grow up to hundreds of gigabytes, which is why keeping them in memory becomes infeasible with increasing size. This requires special solutions for fast disk access (R1, R2). When operating on different anatomical scales, it is necessary to perform cumulative operations on the connectivity matrices (e.g calculate region-level connectivity from voxel-wise connectivity). In this case large parts of the connectivity matrix need to be loaded and processed. We created a data structure in our previous work [GKHB18] to allow these *Aggregation Queries* to happen in real-time, directly on the hard disk. It uses a specialized *Connectivity Storage* to efficiently manage and access large connectivity data (Figure 2). By exploiting the sparseness of the data and its spatial organization, it optimizes disk-reading speed via read-ahead paging and therefore achieves almost sequential reading-speed for local *Aggregation Queries* (i.e. cumulated connectivity of a brain area). To scale for larger brain areas, we implemented a cache mechanisms that uses pre-computed queries in a hierarchical way, similar to an image pyramid. Moreover, the *Connectivity Storage* provides a mapping to a standard reference space, which allows the retrieval of connectivity data regardless of its original resolution. The *Connectivity Storage* has been combined with a multi-model graph/document-database, which is tailored to hierarchical brain parcellations (i.e. hierarchical anatomical brain atlases) to store region-level connectivity for different hierarchy levels. It is also used to organize spatial data, such as volume data (e.g. spatial gene expression), binary masks and meshes of region definitions.

A REST API is acting as the central access point for the data. It provides calls for querying the connectivity matrices, as well as importing them. Data import can be performed by providing the matrices in a row/column table format as csv. The data structure will then convert the data to a compressed file format for the *Connectivity Storage* and automatically aggregats region-level connectivity in the graph-database (see details in [GKHB18]). It can then be used immediately afterwards for real-time connectivity queries. These can be executed in less than one second for brain areas involving about 1% of the brain, and less than four seconds for larger areas up to the whole brain on SSD. On HDD the queries are approximately three times slower.

6. Visualization Components

6.1. Parcellation Browser

The *Parcellation Browser* (Figure 1A) shows a hierarchically organized brain parcellation in a tree view, where every brain region has a name and a color code. Its main purpose is to define the brain regions that are visualized in the *Expression and Connectivity Profiles* as well as the query result's graph representation (the *selection state*). Navigating the tree adds/removes regions to/from the *selection state* to let the user control the current level of detail (first column), or mask parts of the brain (column *S*), depending on her research question. The uppermost regions are the two hemispheres that can be expanded further down the hierarchy. Each major brain region has a color code that is varied in shade depending on its hierarchical level, providing the user with an indirect reference to the larger anatomical context of a subnode and to the spatial position in the brain. This effect is further enhanced by using established color codes for brain regions together with the brain ontologies, which, for example, available in the *AMBA* context.

6.2. Query Toolbar

The *Query Toolbar* (Figure 1B) can be used to execute target/source queries on connectivity matrices (R1), selected by a dropdown menu. Furthermore, it can start brushes, which is a spherical selection tool in the 2D slice view.

6.3. Viewer Item List

The *Viewer Item List* (Figure 1C) controls the visibility and appearance of spatial volumetric data and geometry that is visible in the application. Visibility of volumetric data can be controlled by setting brightness, contrast and transparency. Moreover, it contains entries for all regions in the selection state, since they are visible in 2D (color-coded in the same way as in the *Parcellation Browser*. For all items, the visibility and color, and for query results (i.e. connectivity data), the visibility of the graph representation can be set.

6.4. 3D Visualization in Anatomical Space

To visualize data in its original spatial, anatomical environment we are using classical volume rendering (Figure 1D) and multi-planar reformation (Figure 1E) for the visualization of volumetric data and graph rendering in a ball-and-stick model, since they are common in the neuroscience community. We further render anatomical context in the planar view, visualized by region contours, and mesh geometry in 3D (if enabled via *Viewer Items List*) in order to provide anatomical context. 3D mesh rendering is turned off by default, since it would obstruct 3D volume rendering.

6.4.1. Visualization of Volumetric Data

Volumetric data, like gene expression data or voxel-level structural connectivity is rendered in 3D (Figure 1D), as well as three planar slice views (Figure 1E). A maximum of four volumes can be blended simultaneously in 3D using single color-based transfer-functions (limited by the four RGBA channels, since the volume is loaded as texture). In the slice views, blending is also used for multiple images. Therefore, the rendering of intensity-overlap converges to white.

1. Spatial Reordering   2. Row Compression   3. Aggregation Query
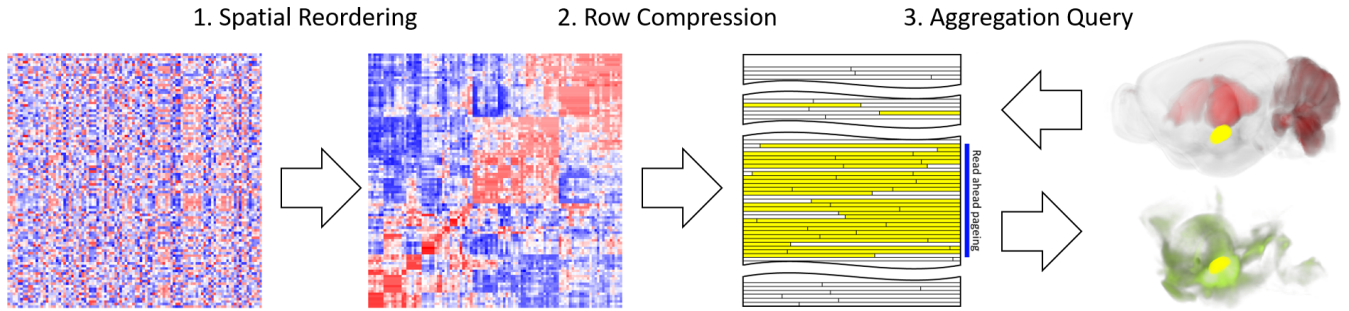
Figure 2: *Principle of the Connectivity Storage and real time Aggregation Queries:* **1.** *Spatial reordering of a (voxel-level) connectivity matrix. Rows/columns that represent the outgoing/incoming connectivity of voxels are reordered, so that rows/columns that belong to spatially close voxel are close to each other.* **2.** *Row Compression: The reordered connectivity matrix is stored with a row-wise compression as Connectivity Storage File on the hard disk. Row-wise compression saves disk space (and therefore reading time) by exploiting sparseness.* **3.** *Aggregation Query: A volume of interest (yellow) is selected on 3D spatial data (e.g. spatial gene expression data, red). Outgoing connectivity is aggregated by reading the corresponding rows (yellow) from the Connectivity Storage File. Since spatially close rows are also close in the file, they can profit from read ahead paging (blue). The result is the aggregated connectivity for every voxel (green).*

6.4.2. Region-level Graph Representation of Voxel-Wise Connectivity

A result of a target/source query can be also visualized as a 3D ball-and-stick-model within its anatomical reference space (R4 and R5). Network nodes are rendered as spheres, where size is determined by the size of the corresponding anatomical region and color reflects their anatomical representation in the originating parcellation. Rendering them as meshes instead, would cause blocking the view by larger brain regions. For directed networks, edges are rendered as tubes with arrows, differentiating incoming from outgoing connections. Undirected network edges are represented by tubes. For weighted networks, connectivity strength is color-coded, ranging from white to the color selected in the *Viewer Item List*. Rendering arrows that are indicating the direction of connections, were specifically requested by our expert users as they are an intuitive way of differentiating incoming from outgoing edges in publications. Edges and nodes are user-adjustable via color-transfer-functions. Moreover, thresholding edge weights enables dynamical network refinement and an exploration of connections of interest. Node labels further support orientation. In addition to the network model in the 3D view, network regions are shown as colored outlines of their anatomical equivalent in the corresponding 2D slice views; arrows (for directed) or lines (undirected networks) are rendered between the region centers in every slice. Although this shows only a part of the network (i.e. in the specific slice), it has the advantage that one can corectly render the regions by anatomy (instead of using a sphere), without facing the problem caused by possible obstruction in 3D.

6.4.3. Filtering Connectivity Strength

Since the mean number of edges increases quadratically with the number of nodes, the user can apply thresholding on edges to highlight stronger connections. This is done with a slider to set a user defined threshold, in combination with a histogram that shows the edge weight distribution in the *Filtering Toolbar* (Figure 1F). The colors of the histogram bars map directly to the visualized edges. Hence, color directly relates to connectivity strength.

6.4.4. Multiple Network Graphs

The comparison and joint exploration of multiple graphs, which relate to the same anatomical parcellation, has been realized by overlaying several graphs, i.e. simply rendering multiple edges between nodes (Figure 4E). Our domain experts found this the most intuitive method, since one can directly see the similarity/contrast of two edges between two nodes. Showing only two graphs even renders up to four arrows between nodes. Therefore, we also implemented an overlap visualization. Between two nodes, a maximum of two edges, whose weight is defined by the multiplication of all edges (at region-level) between those nodes, is rendered. If an overlap is computed, the histograms depict a formula of the multiplication of individual graphs (Figure 4F), so that the calculation can be comprehended by the user. Cascading queries are treated as a single network, depicted by brackets in the formula (see Section 7.2), since they represent connectivity of different orders.

6.5. Connectivity and Expression Profiles

Connectivity Profiles (*C-Profiles*) as well as *Expression Profiles* (*E-Profiles*) (Figure 1G) summarize volumetric voxel-level data in order to make them quantitatively comprehensible for the user (R3). They visualize the mean target/source connectivity or gene expression for brain-regions in bar graphs. These brain regions are defined by the selection made in the *Parcellation Browser*. The mean connectivity/expression of all subregions are rendered as dots. This allows the user both to identify regions that have highly connected/expressed subregions, and to further refine the anatomical hierarchy to focus on relevant subregions. For anatomical context, the colors of the bars and dots are corresponding to the *Parcellation Browser*. Names are made visible via a tooltip. One or multiple regions can be picked for a high-intensity *VOI* selection, which
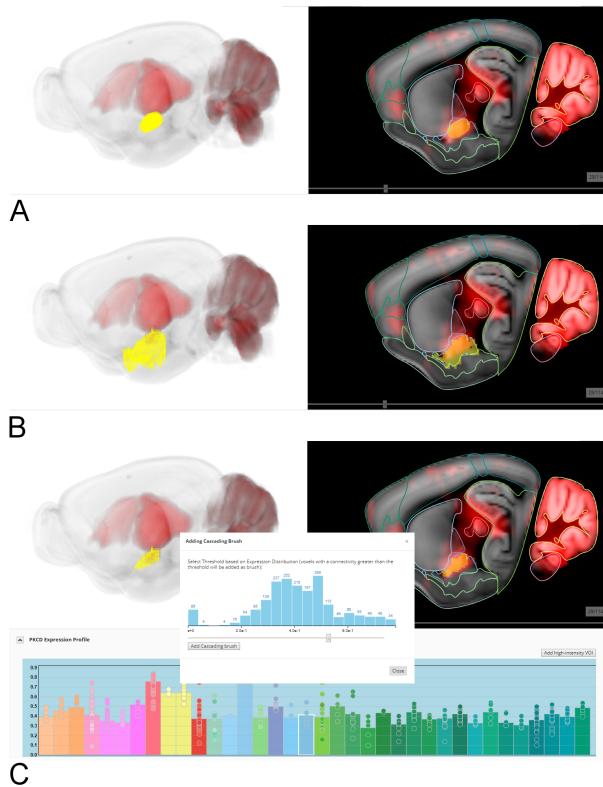
Figure 3: *Selection of parts of a gene expression pattern (red). A) Brush-selection of spherical region (yellow) in 2D slice views and simultaneous rendering in 3D. B) Region-selection initiated by the Parcellation Browser, voxel-level visualization of a region (yellow). C) High-intensity-selection of a brain-region started from an Expression Profile. A dialog which shows a histogram of voxel-level connectivity within the selected region lets the user decide which intensity, and therefore which voxels, will be selected. The result is rendered in 2D and 3D (yellow).*

enables the user to threshold for high expression/connectivity voxels in a separate dialog (see Figure 3). This allows higher order target/source queries in an iterative workflow (R2), as further described in Section 7.1.3. For *C-Profile*s, the percentual composition of the query region is shown in a bar above the profile to indicate the queries anatomical context. It uses the same parcellation/color code as the profile, and reveals its region names via mouse-over. Furthermore the query region can be added again to the 2D/3D viewer, so that a user can reproduce the query with the same or other connectivity matrices.

7. Basic Workflows for Joint Data Exploration

7.1. Visual Queries

A visual query allows API requests to the data structure that are based on selections of a volume of interest directly in volume rendering. The response can then be directly shown in the 2D/3D ren-

derer as well as *C-Profile*. This subsection describes the interaction with volume data, and the query types we have implemented.

7.1.1. Selecting a Volume of Interest

In the volume rendering, the selection of areas can be performed in three different ways: Brush-selection, region-selection, and high-intensity-selection. Brush-selection is performed from the *Query Toolbar*. It lets the user draw pherical areas encoded in transparent yellow in the 2D slice views, which are also directly rendered in 3D. Figure 3A for example shows a gene-expression volume, where the spherical area is drawn on voxels with high gene-expression. Executing queries on this area is therefore acting as link between volume and connectivity data. The region selection (Figure 3B), added by the magnifying-glass-plus-button in the *Parcellation Browser* next to a region, selects only voxels within this region rendered as transparent yellow cubes. This provides the user with the possibility to explore the network without volume data. High-intensity queries are started from the *E-Profiles*. It allows the user to select voxels with high-intensity within user-selected brain regions. The user selects brain regions of interest in the *E-Profiles* (Figure 3 C) with high gene expression. By clicking the "add high-intensity VOI " button, a dialog appears which shows a voxel-level histogram of the intensity values within this region. Here the user defines high-intensity voxels by setting a threshold. In the 3D/2D views, the voxels are instantly selected and visualized similarly to the region selection.

7.1.2. Target/source queries

Target/source queries can be used to link connectivity data with volume data (R1). In the *Query Toolbar*, connectivity matrices for querying can be selected and brush-drawings with a certain radius can be started (although one can also use the other brush-types instead). The selected area (Figure 3A) then acts as an input for a target or source query on the API. The API retrieves the connectivity to all voxels that are either targets or sources from the selection, and then returns a *Connectivity Volume* (connectivity from/to VOI on voxel-level) as compressed JPEG to the web-component which will be rendered instantly in 3D, 2D slice view and as *C-Profile* (R3). This represents the cumulated connectivity to (target) or from (source) the selected area. The *Connectivity Volume* automatically gets assigned a random colorscale, from white to a color that differs the most in its RGB value from other viewer items, to make it visually most distinctive to other visualized volumetric data. *C-Profiles* are shown in a list under each another, with the most recent one on top, so that the user does not need to scroll down.

7.1.3. Higher order target/sources

On the *C-Profile* of the *Connectivity Volume*, the user can select regions of interest with high connectivity to start a high-intensity-selection (Figure 3C). Similarly to the *E-Profile* (Section 7.1.1), the user can choose a threshold in a voxel-level connectivity histogram, which adds all voxel within the region, and a connectivity above the threshold, as *VOI*. After performing a target/source query on the selection, the resulting 2nd-order connectivity is visualized in 2D and 3D as other volume data, while the associated *C-Profile* will be rendered below its originating *C-Profile* (R2). The name is indented

and prefixed with "2nd-order" in red font, so that it is clear to the user that this query belongs to an iterative cascade, while the profile itself is not indented to ensure vertical comparability between multiple profiles (Figures 5 C). The cascading nature of these queries is further emphasized in the region-level graph representation, described in the next section.

7.2. Transformation of Voxel-level Connectivity to Region-level Graph Representation

Changing from voxel-level *Connectivity Volumes* to region-level graph representation can be controlled either directly from a *C-Profile* via "Show Graph Representation" button, or via the *Viewer Item List* (R4). The *Graph Representation* is a mapping of already aggregated connectivity to region-level, and therefore does not reflect the total and individual connectivity between brain regions visible in the graph. It rather represents an abstract view of the query, showing the aggregated connectivity of a *VOI* mapped to a region-level in its 3D/2D spatial context. Moreover, it allows iterative queries for higher order target/sources to be visualized in a cascading way, since the individual graphs are connected by arrows (Figure 5D) (R5). Edges of the graph are drawn from all regions in the query area, to all regions in the *C-Profile* (i.e. the selection state), with their mean connectivity as weights. These edges do not need to be calculated, since they have already been aggregated for the *C-Profiles*.

### 7.3. Region-level Selection and Manipulation

The central control element for the *C-Profile* and *Graph Representation* is the *Parcellation Browser*. Navigating in the tree directly influences the selection state, thereby determining which bars (profile) or nodes (graph) are shown (R3). Opening a subtree of a region will remove the corresponding bars/nodes from the graph, and adding all its children. Closing a region will add it, while removing its children. Since the user is not limited by a rigid parcellation, she can focus on regions of the brain that are relevant to her (e.g. only sub-cortical areas),

8. Case Studies

We conducted two case studies, which were designed in colaboration with our domain experts. The first case study compares different kinds of networks (structural-connectivity vs gene co-expression correlation) related to social-bonding behavior. The second examines neurocircuits related to memory and learning, which we explore iteratively.

8.1. Comparison of multiple networks

In the first case study the domain experts would like to examine brain networks related to social-bonding behavior. In particular, they were interested in oxytocin and vasopressin (neuropeptides known to be related to social behavior in mammals) release effect at the network level [BA15]. For this, they wanted both to examine primary expression sites of both genes, and explore their target sites (i.e. outgoing connectivity of expression sites) on different networks. For this case study we incorporated the following: spatial
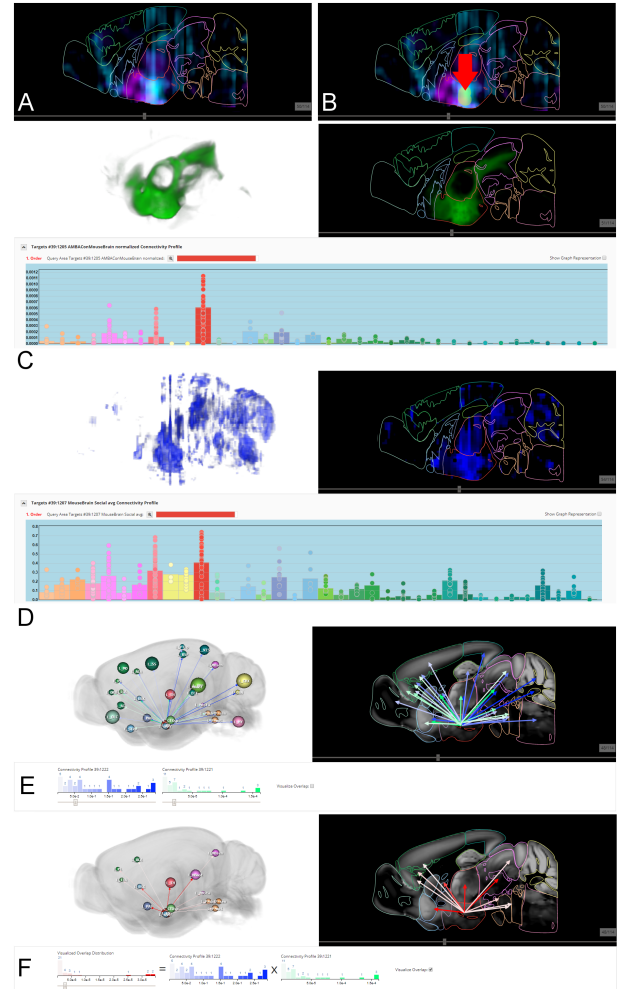
Figure 4: *Case Study 1: A) Gene expression of OXT (cyan) and AVP (purple). B) Selected VOI with brush-selection (yellow, indicated by red-arrow). C) Structural connectivity of the VOI in 3D and 2D (green) and its Connectivity Profile. D) Gene co-expression correlation of the VOI in 3D and 2D (blue) and its Connectivity Profile. E) Graph representation of structural connectivity (white to green) and gene co-expression correlation (white to blue). F) Multiplication (overlap) of the connectivities (red).*

gene expression data (67x41x58 volume on a 200-micron resolution) of oxytocin (OXT) and vasopressin (AVP), a spatial gene co-expression correlation network (also 200-micron resolution, matrix file size is ~ 12 GB) of social-bonding related genes consisting of (gene ENTREZ ID in brackets) Oxt (18429), Oxtr (18430), Avp (11998), V1b receptor (26361), D1R (13488), D2R (13489), Slc6a3 (13162) and Crh (12918); and a 100-micron voxel-level (132x80x114) structural-connectivity from the *AMBA* (file size is ~ 90 GB) [GKHB18]. Although these networks have billions of edges, our data structure [GKHB18] allows real-time retrieval of aggregated connectivity (i.e. cumulated incoming/outgoing connectivity) of a *VOI* on a 132x80x114 standard brain space.

Identify a volume of interest of spatial data: The entry point for experts is the spatial gene expression data of OXT (Figure 4A cyan) and AVP (Figure 4A purple). The user notices an overlap in the hypothalamus (indicated by dark red contours, in the 2D slice view) (R3). She selects the overlapping area with a spherical brush initiated by the *Query Toolbar* (Figure 4B yellow spheres, highlighted by red arrow).

Find connected areas: The user selects the structural connectivity matrix in the *Query Toolbar*, a click on the *Target Query* button executes the selected *VOI* on the data structure (R1). The accumulated connectivity is instantly displayed as *Connectivity Volume* in 3D and 2D (Figure 4C green) and as *C-Profile* (Figure 4C bottom). The profile shows that the strongest connectivity is to the Hypothalamus (dark red) itself (strongest connectivity to itself is common for structural and gene co-expression correlation). Other top connections are the Striatum (light blue), Pallidum (dark blue), Midbrain (pink), Thalamus (light red) and some cortical areas (green) (R3). This process is repeated for the gene co-expression correlation (Figure 4D blue), where its profile depicts hypothalamus as strongest again, but among similar connections also strong connections to Cerebellum (yellow) and Hindbrain (orange).

Compare networks: To compare the connectivities, the user visualizes them as *Graph Representation* (R4). Figure 4E shows the two graphs in 2D and 3D, easily recognizable by their colors. In the *Filtering Toolbar* the user selects a threshold to filter weak connections. Among others, the user sees that both networks show projections to the Septal Complex (LSX), the Anterior Cingulate (ACA) and the Prelimbic/Infralimbic Area (PL and ILA), which are well known nodes involved in social behavior [BA15]. To further highlight this, the user also visualizes the overlap of both graphs (i.e. multiplication of both connectivities). The resulting reduced amount of edges makes it easier for the user to identify regions which are connected via both connectivities. These types of network visualization were familiar to the domain experts, who confirmed their suitability for neuroscientific visualization or publications (R5).

This case study illustrates the joint exploration of gene expression data and different kinds of connectivities on voxel- and region-level. This approach allows real-time visual analytic workflows which are fast and efficient, compared to time-consuming manual data aggregation by querying different online databases, literature research and scripting. A video of the case study is available as *Supplementary Video 1*.

8.2. Higher-order connectivity

We designed this case study with domain experts to showcase iterative higher-order connectivity queries with a well-known relationship between the brain-derived neurotropic factor (BDNF) and hippocampal synaptic plasticity, respectively circuits related to learning and memory [LBD17]. One of these circuits is Dentate Gyrus (DG) → CA2/CA3 → CA1 → Entorhinal cortex/Subiculum [LBD17]. For this purpose, we explored a primary expression site of BDNF in DG, and traversed their first-, second- and third-order targets iteratively. This case-study required the spatial gene expression data at a 200-micron resolution for BDNF (gene ENTREZ ID:
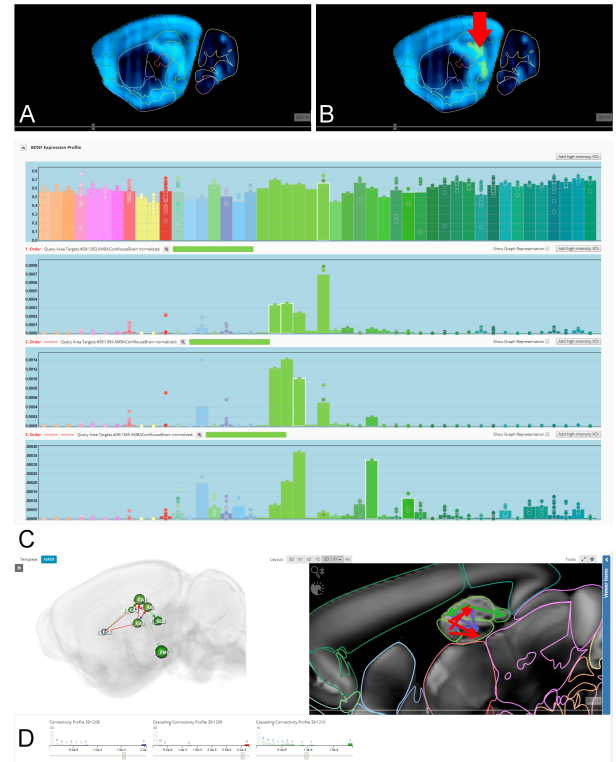
Figure 5: *Case Study 2: A) Gene expression of BDNF (light blue). B) Selected VOI with high-intensity-selection (yellow, indicated by red-arrow). C) Expression Profile of BDNF and its first-, second- and third.order Connectivity Profiles D) Graph representation of the connectivity in 2D and 3D (first-order blue, second-order red, third-order green).*

12064) and structural connectivity at a 100-micron resolution from *AMBA* in the data structure [GKHB18].

Identify a volume of interest of spatial data: The experts started their investigation by visualizing the spatial gene expression of BDNF in 2D/3D and as *E-Profile* (Figure 5A). Since the circuit has its origin in DG, the user navigates in the *Parcellation Browser*. In the 2D slice view the user sees that there is a high expression level, hence she starts a high-intensity-selection in the *E-Profile* (Figure 5B).

Find connected areas iteratively: By selecting the structural connectivity matrix in the *Query Toolbar* and clicking the *Target Query* button, the user receives the accumulated connectivity instantly as 2D/3D visualization and *C-Profile*. The name of the profile is indented and prefixed with "First-Order" to highlight the iterative procedure (Figure 5C). CA2 and CA3 are the strongest connections (not counting DG, since it is a connection to itself), so they are chosen for the next high-intensity-selection to go further along the circuit (R2). Note that CA1 receives strong input, but this is primarily caused by their spatial closeness and data acquisition technique for structural connectivity [AMB]. The next *C-Profile* shows CA1 as strongest connection (except for the originating CA2/CA3).

By performing a high-intensity-selection on CA1, the results show strong connectivity in the Retrohippocampal Region. After browsing its subregions in the *Parcellation Browser*, the connections to Entorhinal Cortex and Subiculum are revealed (highlighted in Figure 5C bottom)

Visualize the Circuit: Finally, the experts wanted to see the circuit in a 2D-3D *Graph Representation* to give them a spatial context (R4), so that they can use their exploration for presentation and discussion purposes with colleagues (R5). They selected the "Show graph representation" button next to the profiles, which instantly shows them the network graphs (Figure 5D). After filtering for the strongest connections, the graph shows only the DG → CA2/CA3 → CA1 → Entorhinal cortex/Subiculum circuit (and a connection to the lateral septal complex LTX, which is not relevant for this case study). Since the *C-Profiles* were automatically marked with different colors, the colored histograms in the *Filtering Toolbar*, that are ordered according to their iteration, allow association of graph edges with the connectivity-order originating from DG.

In this case study, we showed the iterative exploration of a 90 GB connectivity matrix at voxel-level, enabled by quantitative information in *C-Profile* and high-intensity *VOI* selection. To our best knowledge, this could be done so far only at region-level connectivity. We provide a video of this study as *Supplementary Video 2*.

9. Discussion and Conclusion

In this paper we present a novel integrated workflow for analyzing and fusing heterogeneous neurobiological data of different types, modalities and scale via their spatial context. The workflow is integrated in a framework that combines data from different large-scale brain initiatives with user generated data. We incorporate a hierarchically organized data structure, which enables real-time querying and aggregating of huge brain network connectivity of different scales and resolutions on a common standard brain space. This data structure can be accessed via a web-component, which allows selecting *VOI* on 2D/3D visualizations of various volumetric data. The resulting voxel-level connectivity is rendered again in 2D/3D, where it can directly be compared to spatial gene expression or connectivity of different modality. Importantly, these operations remain interactive, despite operating on matrices with up to hundreds of gigabytes.

To allow a quantitative evaluation on brain-region level, the results are shown as region-wise profile, presenting the mean-connectivity for every brain region, as well as for all its sub-regions in a bar chart. This directly highlights strongly connected subregions, which would otherwise be missed if the user operated at a higher brain region level. The region-level profile of connectivity further allows for *VOI* selection for transitive connectivity. Although it would generally be possible for a user to manually select strongly connected voxels in the 2D slice view in order to start another target/source query, the domain experts preferred to query for transitive connectivity within brain regions of their interest (i.e. they were not interested in retrieving the connectivity of all target/source sites, but only from sites within brain regions of interest). Therefore, the user selects strongly connected brain-regions in the region-level profile, chooses a connectivity threshold and receives a voxel-level selection that can be used for further queries.

Query results can be abstracted as 2D and 3D network graphs, which reduces the complexity of voxel-level while incorporating neuroanatomical context. Multiple graphs (i.e. region-level graph representation of connectivity query results) can be rendered with parallel arrows/lines, or combined by multiplication. Domain experts confirmed that they feel familiar with this type of visualization. Moreover, it is suitable to be used for discussion or as figures in neuroscientific publications.

Finally, the case studies conducted with domain experts showed biological validity and reproducibility by reproducing findings of known microcircuits that are subject to current research.

10. Acknowledgments

References

[AABS*14] AL-AWAMI A. K., BEYER J., STROBELT H., KASTHURI N., LICHTMAN J. W., PFISTER H., HADWIGER M.: Neurolines: A subway map metaphor for visualizing nanoscale neuronal connectivity. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec 2014), 2369–2378. doi:10.1109/TVCG.2014.2346312. 4

[all] Allen institute. URL: https://www.alleninstitute.org/. 1, 4

[AMB] Allen mouse brain atlas whitepapers. URL: http://help.brain-map.org/display/mouseconnectivity/Documentation. 9

[BA15] BARIBEAU D. A., ANAGNOSTOU E.: Oxytocin and vasopressin: Linking pituitary neuropeptides and their receptors to social neurocircuits, 2015. doi:10.3389/fnins.2015.00335. 8, 9

[BAAK*13] BEYER J., AL-AWAMI A., KASTHURI N., LICHTMAN J., PFISTER H., HADWIGER M.: Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. *IEEE Transactions on Visualization and Computer Graphics (Proceedings IEEE SciVis 2013) 19*, 12 (2013), 2868–2877. 4

[BD07] BALZER M., DEUSSEN O.: Level-of-detail visualization of clustered graph layouts. *Asia-Pacific Symposium on Visualisation 2007, Proceedings* (2007), 133–140. 4

[BSL*14] BOETTGER J., SCHAEFER A., LOHMANN G., VILLRINGER A., MARGULIES D. S.: Three-Dimensional Mean-Shift Edge Bundling for the Visualization of Functional Connectivity in the Brain. *IEEE Transactions on Visualization and Computer Graphics 20*, 3 (2014), 471–480. 4

[GKHB18] GANGLBERGER F., KACZANOWSKA J., HAUBENSAK W., BUEHLER K.: A data structure for real-time aggregation queries of big brain networks. *bioRxiv* (2018). doi:10.1101/346338. 2, 5, 8, 9

[GSB*15] GENÇ E., SCHÖLVINCK M. L., BERGMANN J., SINGER W., KOHLER A.: Functional connectivity patterns of visual cortex reflect its anatomical organization. *Cerebral Cortex* (2015), bhv175. 2

[HBG*17] HUNTENBURG J. M., BAZIN P.-L., GOULAS A., TARDIF C. L., VILLRINGER A., MARGULIES D. S.: A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. *Cerebral Cortex* (2017), 1–17. 3

[HKC*10] HAUBENSAK W., KUNWAR P. S., CAI H., CIOCCHI S., WALL N. R., PONNUSAMY R., BIAG J., DONG H.-W., DEISSEROTH K., CALLAWAY E. M., ET AL.: Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature 468*, 7321 (2010), 270–276. 2

[hum] Human brain project. URL: https://www.humanbrainproject.eu. 1

[KZM*17] KIM J., ZHANG X., MURALIDHAR S., LEBLANC S. A., TONEGAWA S.: Basolateral to central amygdala neural circuits for appetitive behaviors. *Neuron 93*, 6 (2017), 1464–1479. 1

[LBD17] LEAL G., BRAMHAM C., DUARTE C.: Chapter eight - bdnf and hippocampal synaptic plasticity. In *Neurotrophins*, Litwack G., (Ed.), vol. 104 of *Vitamins and Hormones*. Academic Press, 2017, pp. 153 – 195. doi:https://doi.org/10.1016/bs.vh.2016.10.004. 9

[LDTS14] LAPLANTE R. A., DOUW L., TANG W., STUFFLEBEAM S. M.: The Connectome Visualization Utility: Software for Visualization of Human Brain Networks. *PLoS ONE 9*, 12 (2014). 4

[LFG*15] LI H., FANG S., GONI J., CONTRERAS J. A., LIANG Y.: Integrated Visualization of Human Brain Connectome Data. *Lecture Notes in Computer 9250* (2015), 295–305. 4

[LHA*07] LEIN E. S., HAWRYLYCZ M. J., AO N., AYRES M., BENSINGER A., BERNARD A., BOE A. F., BOGUSKI M. S., BROCKWAY K. S., BYRNES E. J., ET AL.: Genome-wide atlas of gene expression in the adult mouse brain. *Nature 445*, 7124 (2007), 168–176. 2

[LNT*08] LAU C., NG L., THOMPSON C., PATHAK S., KUAN L., JONES A., HAWRYLYCZ M., CARNINCI P., KASUKAWA T., KATAYAMA S., TOGA A.: Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics 9*, 1 (2008), 153. 4

[LTW*11] LIN C. Y., TSAI K. L., WANG S. C., HSIEH C. H., CHANG H. M., CHIANG A. S.: The neuron navigator: Exploring the information pathway through the neural maze. In *IEEE Pacific Visualization Symposium 2011, PacificVis 2011 - Proceedings* (2011). doi:10.1109/PACIFICVIS.2011.5742370. 4

[MBB*16] MURUGESAN S., BOUCHARD K., BROWN J. A., HAMANN B., SEELEY W. W., TRUJILLO A., WEBER G. H.: Brain Modulyzer : Interactive Visual Analysis of Functional Brain Connectivity. *IEEE/ACM transactions on computational biology and bioinformatics* (2016), 1–14. 4

[mPlDI*16] MING POO M., LIN DU J., IP N. Y., XIONG Z. Q., XU B., TAN T.: China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing. doi:10.1016/j.neuron.2016.10.050. 1

[NSvA*13] NOWKE C., SCHMIDT M., VAN ALBADA S. J., EPPLER J. M., BAKKER R., DIESRNANN M., HENTSCHEL B., KUHLEN T.: VisNEST - Interactive analysis of neural activity data. *Biological Data Visualization (BioVis), 2013 IEEE Symposium on* (2013), 65–72. 3

[OHN*14] OH S. W., HARRIS J. A., NG L., WINSLOW B., CAIN N., MIHALAS S., WANG Q., LAU C., KUAN L., HENRY A. M., ET AL.: A mesoscale connectome of the mouse brain. *Nature 508*, 7495 (2014), 207–214. 2

[PLK*15] POLDRACK R. A., LAUMANN T. O., KOYEJO O., GREGORY B., HOVER A., CHEN M.-Y. E. A.: Long-term neural and physiological phenotyping of a single human. *Nature Communications 6* (2015), 8885. 3, 4

[SAM*05] SHERBONDY A., AKERS D., MACKENZIE R., DOUGHERTY R., WANDELL B.: Exploring connectivity of the brain's white matter with dynamic queries. In *IEEE Transactions on Visualization and Computer Graphics* (2005). doi:10.1109/TVCG.2005.59. 4

[SCHT09] SAALFELD S., CARDONA A., HARTENSTEIN V., TOMANČÁK P.: CATMAID: Collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp266. 4

[SE12] SCHMITT O., EIPERT P.: neuroVIISAS: Approaching Multiscale Simulation of the Rat Connectome. *Neuroinformatics 10*, 3 (2012), 243–267. 4

[Sor13] SORGER J.: neuroMap - Interactive Graph-Visualization of the Fruit Fly ' s Neural Circuit. In *Biological Data Visualization (BioVis), 2013 IEEE Symposium on* (2013). 4

[SSDV14] SCHOLTENS L. H., SCHMIDT R., DE REUS M. A., VAN DEN HEUVEL M. P.: Linking Macroscale Graph Analytical Organization to Microscale Neuroarchitectonics in the Macaque Connectome. *The Journal of Neuroscience* (2014). 3

[VESB*13] VAN ESSEN D. C., SMITH S. M., BARCH D. M., BEHRENS T. E., YACOUB E., UGURBIL K., CONSORTIUM W.-M. H., ET AL.: The wu-minn human connectome project: an overview. *Neuroimage 80* (2013), 62–79. 1

[XWH13] XIA M.-R., WANG J.-H., HE Y.: BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one 8*, 7 (2013). 4

[YW04] YOUNG L. J., WANG Z.: The neurobiology of pair bonding. *Nature neuroscience 7*, 10 (2004), 1048–1054. 2