|  |
| --- |
| Spherical Histogram Approximation for Maximum Flow of Brain Connectivity  Lanston Hau Man Chu  Department of Computer Science, University of Wisconsin-Madison  Abstract  How neural signals flow within a biological neural network is highly related to how knowledges are being processed in brains. Maximum flow of neural activities can be studied further to understand more about perception of creatures. With the density/intensity data obtained in a rAAV injection experiment on mouse brains, diffusion graphs can be constructed by Fast Matching Method, and this paper would investigate how specific areas with highest density of flows can be efficient obtained by an approach we called Spherical Histogram Approximation, as well as a variation Spherical Clustering. We will see that the maximum flow areas, which is the result of our algorithm, would capture flows that densely squeezing into the corresponding region. We will discuss how the current approach can be modified further in the future study to obtain better result in biological sense.  **Contact:** hchu34@wisc.edu  **Supplementary information:** [Interactive Visualization](https://lanstonchu.github.io/gallery/brain_connectivity/) ; [Source Code and Data](https://github.com/lanstonchu/brain-connectivity) ; Slides ; [Presentation](https://youtu.be/tDp00Rhi7mE) |

# Introduction

A picture containing text

Description automatically generated

Figure 1:(Left) Hypothesis: Maximum flow of pathways is highly related to the embedding size in creature's perception. (Right) Result of a rAAV injection experiment on anterograde prospective. Color reflects the rAAV density detected by the 3D scan.

Series of fluorescent virus (e.g. rAAV: Recombinant Adeno-associated Virus) injection experiments have been continuously taken by the Allen Institute since 2010s, providing researchers tons of detailed connectivity data at different positions of mice’s brains[[1]](#footnote-2). Those injections are anterograde (i.e. from source to destination), and therefore we can determine the direction of neural signal's forward propagation along the axons' connection, by following the direction of the injection spread. By using the density/intensity values of an injection, a **diffusion graph**, i.e. graph of flow, can be constructed by scanning through points in the direction of the gradient of the data via the Fast Matching Method (Sethian, 1999). In this paper, we would discuss how the maximum flow areas of the relative pathways in brains can be efficiently determined by using the graphs constructed by the Fast Matching Method.

There are several reasons why we care about the maximum flow areas of the injection graphs. First, those injection graphs reflect the traffic volume of different position of the brains’ pathways, and it is reasonable to conjecture that the maximum flow of the graph is highly related to the maximum data size required to embed knowledge of specific tasks. The ground of this hypothesis originates from the successful autoencoder structures in the Artificial Intelligence (AI) area of the last decade. In those structures the bottlenecks of those artificial neural networks reflect the size of embedding for the task, while the bottleneck is with respect to the minimum cut location of the graph of network. Due to the Max-flow min-cut theorem in graph theory, the maximum amount of flow from source to sink is equal to the total weight of edges in a minimum cut of the graph. Second, the level of connection integration among neurons can be significantly correlated to the consciousness problem, e.g. the assumption of Integrated information theory (Tononi, Boly, Massimini, &Koch, 2016).Graphical user interface, application

Description automatically generated

Figure 2: (Left) Graph G constructed by Fast Matching Method on retrograde’s prospective, i.e. flows pointing towards a specific destination. Color of lines are determined by the sources’ position, at destination site in LGd. (Middle)Location of LGd in the coronal/sagittal cross sections. (Right) Setting of the source/sink problem of a graph.

# Methods

There are two directions-related prospective to view the tracked injections: **Anterograde** (i.e. from source to destination) vs. **Retrograde** (i.e. from destination to source). In anterograde prospective, we would follow the direction of the forward propagation of neural network, i.e. following the neural signals, or equivalently the dentrite-soma-axon direction of neurons. In retrograde prospective, we would follow the opposite direction. The types of virus selected in the Allen Institute injection experiments are all anterograde. In this paper, we are interested in the retrograde perspective instead, i.e. we would like to study the flows pointing towards a specific **destination site** . Therefore, we are not looking at the graph of a specific experiment . Instead, we would gather all edges of that points towards to form a combined graph. That is, we view as , where which path consists of a set of edges, i.e. and each is an edge in the graph, while is the injection site of experiment . We would then construct a graph where is the union of all relevant paths across various experiments pointing the specific destination site . This data transformation/integration makes sense since the Allen Institute’s 3D brain scans of different experiments are aligned in an accurate manner up to the level of microns (Allen Institute for Brain Science, 2017). The graph can be achieved by using the API of Allen Institute’s database.

Text

Description automatically generated with medium confidence

Figure 3: Construction of .

Diagram

Description automatically generatedFrom now on, the diffusion graph would now be described as , where each line is the path starting from the injection site to the destination . So the values of starting point of each would be different, but the destination of different paths would be the same. Please also note that is a matrix since there are coordinates in the path.

Figure 4: (Left)The path . (Middle) Construction of Histogram Walls. (Right) Searching for Wall Hitting Points.

We would then do the Spherical Coordinate transformation:

where

Please note that the above spherical coordinate transformation is different to the standard format of spherical coordinate, as we have switched y-axis and z-axis to accommodate some coordinate settings in WebGL in our data visualization stage.

Chart

Description automatically generated

Figure 5: Result of Spherical Histogram on wall-0, 1 and 2 at . Red dot refers to the destination site . Please note that the input of the wall-1 histogram would only use the paths which passes through the maximum-flow area of wall-0.

Then we can do some statistics on the radius:

**Table 1.**Radii of Histogram Walls for in LGd. Note that the typical size of the brain-containing cuboid in μm is

|  |  |  |  |
| --- | --- | --- | --- |
| Wall radii |  |  |  |
| Value | 886μm | 1773μm | 2659μm |

We can see that each is for 1 experiment only and are the statistics result of the maxima. We hereby define 3 **Histogram Walls**, namely the inner wall , middle wall and outer wall . We want more than 25% of paths passing through the outer wall so we picked the percentile at 75% and an adjustment ratio 0.9. We can then determine the **Wall Hitting Points**, i.e. the points of each paths that are closest to the walls:

where is the first point of path-i passing through wall-0 (i.e. inner wall), while is the point just before . If does not exist we will take . Apart from we can also construct and for middle wall and outer wall.

We can then apply **spherical histogram** at bins on the three walls. We would take bins for (wall-0, ), bins for (wall-1, ), and bins for (wall-2, ). For paths passing through wall-(u+1), we would like to focus on those passing through wall-u, since we are focusing on diffusion paths. Therefore, the input of the wall-1 histogram would only use the paths which passes through the maximum-flow area of wall-0. And so on so forth for wall-2:

In , we call the bin with maximum points the **maximum-flow area** of wall-u, and is denoted by .

Once we obtained the values of , and , we can in fact do something more than just histogram. For example, we can do **spherical clustering** on the values of . The k-means clustering can be based on 2 to 10 clusters, and we would take the beset result with maximum Silhouette score:

where maximizes the Sihouette score of the clustering output.A screenshot of a video game

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated with medium confidence

Figure 6:(Left) The maximum flow areas, namely (yellow), (orange), and (rose) of the destination site at LGd (i.e. red dot). (Right) Zoomed-in version of another set of maximum flow areas at in IG. Yellow lines refer to all lines passing through , while the orange lines refer to all lines passing through both and . Red lines refer to lines passing through all , and .

# Results

Figure 5 shows the result of spherical histogram at coordinate (7400 μm, 3300 μm, 3300 μm), which is a point at the Dorsal part of the lateral geniculate complex (LGd) and is a key intermediary point of the visual system, i.e. the pathways from the retina to the primary visual cortex (VISp).

In Figure 6, we can see that maximum flow area , are visually in line with the paths. The region with highest density of lines is captured by the colored pieces of partial spheres. For interactive visualization of the result, please refer to <https://lanstonchu.github.io/gallery/brain_connectivity/> of the supplementary information. The destination site of the right image of Figure 6 is at Innduseum Griseum (IG).

As mentioned, we can use k-means clustering as an alternative of spherical histogram for the lines selection. In Figure 7, we demonstrate the result of k-means at wall-u by using obtained by spherical histogram as input. Since spherical clustering is not a focus of this paper, we just simply apply the Euclidean distance to for spherical clustering.

Chart, scatter chart

Description automatically generated

Figure 7: Result of spherical clustering at Wall-0, 1 and 2. The input are filtered by obtained from spherical histogram.

# Discussion

A picture containing polygon

Description automatically generatedWe can see that the spherical histogram can efficiently determine the maximum flow areas of the rAAV injections. The colored maximum flow areas indicate the region that the flows are densest across various experiments at specific destination site of different flow range. For example, for destination site in IG, the maximum flow areas determined by the algorithm locate at the neighboring CC (Corpus Callosum), which is quite reasonable as we know that CC is one of the axons richest regions in brain. The performance of the algorithm is as expected since the paths are derived by the gradients of the fluorescent rAAV injection intensity/density along time.

Figure 8: Animation of the algorithm result at destination site in Innduseum Griseum (IG). Note that the maximum flow areas locate at the neighboring Corpus Callosum (CC), which is known to be an axons richest region in brain.

It is also worth to point out that the complexity of the algorithm is only (and for the spherical clustering approach is used instead; given that the paths lengths are in the same order), as the spherical histogram is fundamentally a histogram. Thus our algorithm is efficient comparing to many spatial searching algorithms.

Locating the densest area will help future research to further determine various properties of capacity and connectivity of brains in depth, e.g. channel capacity of perceptions from the perspective of information theory:

where I() is the mutual information, input and output are random variables of the channel. Currently there are research of psychological experiments in computational cognitive science which tries to determine human brain’s channel capacity of specific task by using Rate Distortion Theory (RDT) approach, e.g. estimating the visual working memory (VWM) (Bates, Lerch, Sims, &Jacobs, 2019).

Diagram

Description automatically generatedA RDT package using deep learning model is also developed to solve for Rate/Distortion/Channel Capacity for various patterns of experimental input (Sims, 2016). Therefore, the maximum flow of pathways defined from bioinformatics prospect can be compare with the RDT result of psychological experiments, and to test various hypothesis connecting the two fields, i.e. whether channel capacity of psychological perception is proportional to the biological maximum flow in brains. In the recent decades, autoencoder had been practically proven that knowledge and perception can be embedded into a reduced dimensional space. Thus, the hypothesis can even be extended to a “golden triangle” to cover the three fields: Maximum flow areas in bioinformatics, perceptional channel capacity in computational cognitive science, as well as embedding analysis in the field of machine learning.

Figure 9: The "golden triangle" of hypothesis among the three fields: Machine Learning, Computational Cognitive Science and Bioinformatics regarding the problem of knowledge processing.

The result of this paper is referring to the retrograde perspective (i.e. from destination to source). Future study can be done on anterograde perspective by using a similar approach and would still make biological sense since the spreads are just diffusion of rAAV along neural connections, and thus we can study for both anterograde and retrograde directions.

The spherical histogram of this paper is just a simple demonstration on what how we can group the paths of injection flow by using spherical perspective. Several extensions can be done:

* Bins merging: We can merge the 1st maximum flow area and the 2nd maximum flow area if they are neighbors to each other. That would allow us to obtain a more representative region of dense flows.
* More sophisticated distance function for clustering approach: Although this paper mainly focuses on spherical histogram, we have also discussed k-means clustering and have shown a primitive result in Figure 7, which we have only applied a vanilla Euclidean distance function for in the k-means. For future study, the clustering can be done by considering the spherical topology. The distance can be even more complicated by considering prior knowledge of neural science. For example, can be more likely than the other angles (or less likely) depending on the future discoveries of inter/intra-layers neural connections of cortex.

# Conclusion

In this paper, the algorithm using spherical histogram efficiently determined the maximum flow areas. An alternative approach of spherical clustering is also demonstrated. Some potential improvement involving automatic bin merging and sophisticated distance function are also introduced. We hope further study will help revealing more connections along Machine Learning, Computational Cognitive Science and Bioinformatics regarding the problem of knowledge processing.

Diagram

Description automatically generatedReferences

Figure 10:(Left) Extensions for bin merging. (Right) Extensions for more complicated distance function in clustering.

Allen Institute for Brain Science. (2017). Allen Mouse Brain Connectivity Atlas : Technical White Paper : Overview. *Technical White Paper*, (March), 1–15.

Bates, C. J., Lerch, R. A., Sims, C. R., &Jacobs, R. A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of Vision*, *19*(2), 1–23. https://doi.org/10.1167/19.2.11

Sethian, J. A. (1999). Fast Marching Methods. *SIAM Review*, *41*(2), 199–235. https://doi.org/10.1137/S0036144598347059

Sims, C. R. (2016). Rate-distortion theory and human perception. *Cognition*, *152*, 181–198. https://doi.org/10.1016/j.cognition.2016.03.020

Tononi, G., Boly, M., Massimini, M., &Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

1. <http://connectivity.brain-map.org/> [↑](#footnote-ref-2)