Visualising the Transition of Large Networks via Dimensionality Reduction to Illustrate the Evolution of the Human Brain

Florian Ganglberger[1], Joanna Kaczanowska[2], Wulf Haubensak[2] and Katja Bühler[1]

[1]VRVIS Research Center, Vienna
[2]Institute of Molecular Pathology, Vienna

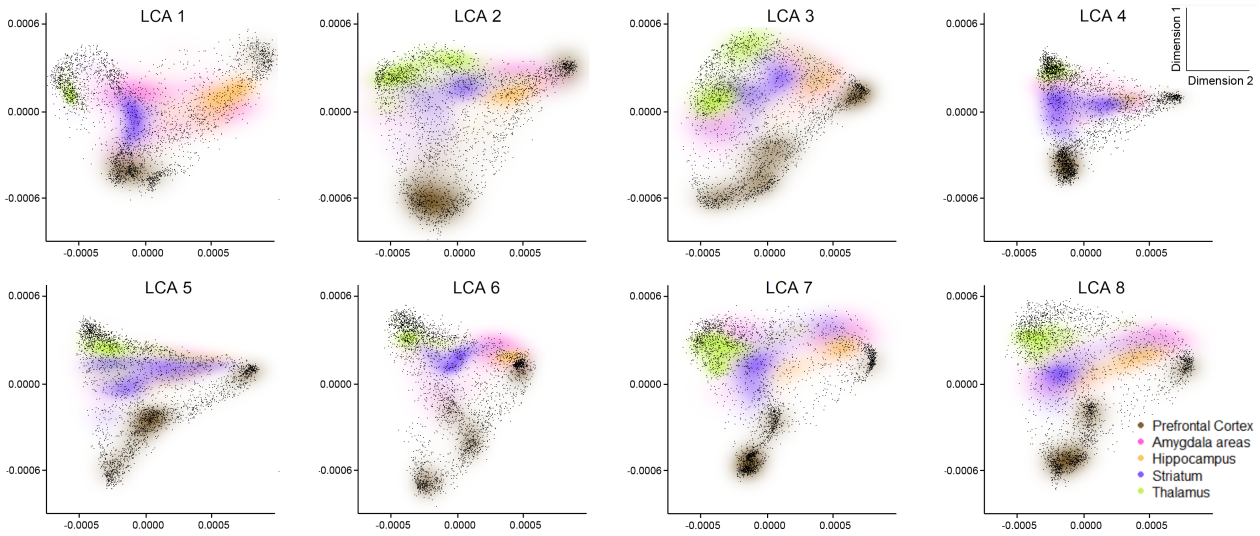

Figure 1: *Co-evolution networks across different evolutionary timepoints, i.e. the last common ancestor (LCA) of two mammalian species ordered chronologically (mouse, bushbaby, marmoset, macacque, gibbon, chimp, denisovan, neanderthal, and human). LCA 1 (LCA of mouse and bushbaby) is the first evolutionary timepoint, LCA 2 (LCA of bushbaby and marmoset) its successor and so on. The maps show the spatial co-evolution in human brain space as distance in a common low-dimensional embedding space. Each dot represents a position in the human brain, where positions with high co-evolution at the respective evolutionary timepoint are close together. The densities of positions belonging to five brain regions are colour-coded which enables visually tracing them. Hence, an interpretation of which regions have undergone evolutionary changes at the respective time is possible.*

Abstract

Advances in high-throughput imaging techniques enable the creation of networks depicting spatio-temporal biological and neurophysiological processes with unprecedented size and magnitude. These networks involve thousands of nodes, which can not be compared over time by traditional methods due to complexity and clutter. When investigating networks over multiple time steps, a crucial question for the visualisation research community becomes apparent: How to visually trace changes of the connectivity over several transitions? Therefore, we developed an easy-to-use method that maps multiple networks to a common embedding space. Visualising the distribution of node-clusters of interest (e.g. brain regions) enables their tracing over time. We demonstrate this approach by visualizing spatial co-evolution networks of different evolutionary timepoints as small multiples to investigate how the human brain genetically and functionally evolved over the mammalian lineage.

1. Introduction

Recent brain initiatives, such as the Allen Institute [OHN*14], the Human Brain Project [MML*11], and the WU-Minn Human Connectome Project [VESB*13] created a wealth of large scale brain networks for neurscience research. These networks represent the various (e.g. functional or structural) relations between different spatial locations. To trace changes in these networks over time, it is necessary to visualise their similarities/differences [MKF*15]. For example, Fair et al. showed differences in functional networks between children and adults via color coding in a node-link diagram overlaying brain anatomy [FDC*07]. A design study by Alper et al. [ABH*13] suggested that matrix visualisation in combination with glyphs outperforms superimposed node-link diagrams of two different modalities. Ma et al. [MKF*15] proposed the tracing of

dynamic functional networks via animations, where changes from previous transition are visualised via glyphs. To do this without animation, small multiples [BHRD*15] can be used to compare a series of similar graphs with the same scale. Different approaches have been proposed by in VIOLA [SCH*18], with rendering networks over time in a 3D volume (i.e., the third dimension is time), and by Van Den Elzen [VHB*14] with *Massive Sequence Views* in a circular graph.

With increasing amount of nodes, traditional approaches become more complex and cluttered. Dimensionality reduction methods have been shown to be promising tools to visualise large networks [KRM*17, HVP*19, HSS*20], especially in neuroscience [LGG15, MGG*16]. Here, the nodes, representing brain regions, are mapped to a low-dimensional embedding space where regions with a high connectivity are close to each other and regions with a low connectivity are farther apart. Non-linear approaches, such as Diffusion Maps [LGG15] are widely used for mapping brain networks, for they are more robust than linear methods in modelling the complex structure of the brain [VdWBP*20]. However, they are only defined for the mapped data points [HN03] (i.e. they do not provide a projection for other data points) and therefore can not be used to create a common space where the individual time points are mapped to for comparison [VdWBP*20]. To solve this problem, Vos de Wael et al. [VdWBP*20] used Procrustes analysis to align the embedded data to a template via scaling, shifting, and rotation. Similar results have been achieved by joint manifold alignment [XNS*19] for cross-species comparison.

These methods depend on the quality of the alignment, hence, Locality Preserving Projection (*LPP*) [HN03] (a linear approximation of non-linear Laplacian Eigenmaps) represents a promising alternative, as it provides also a projection matrix of the data to the embedding space. In this paper, we utilize *LPP* to create a common (mean) embedding space out of individual time points [FNS*18, FSG*18]. Here, we take advantage of the provided projection matrix to map individual networks onto the same space for comparison. Furthermore, we propose a visual representation based on combining density plots of different subgroups (i.e. anatomical brain regions) so that they are visually traceable and interpretable over time using small multiples (Figure 1). The combination of these methods can be seen as the main contributions of this paper and as a first step towards more advanced, interactive tools. We demonstrate the superiority of our approach over standard methods such as heatmaps and node-link diagrams on co-evolution networks derived from a recent neuroscientific study about the evolution of the human brain [KGG*19].

2. Method

Comparing networks over time becomes increasingly complex and cluttered with traditional methods, such as heatmaps and node-link diagrams (Figure 3 (a) and (b)). Instead, we use dimensionality reduction to map these networks, given as affinity matrices, to a low-dimensional common embedding space. Here, the nodes' affinities are encoded in their distance on the embedding space instead of links or edges. On this low-dimensional space, networks can be traced over time visually.

For every time point, one affinity matrix is given, where each entry represents a certain relation (e.g. a connectivity or correlation) between nodes (e.g. spatial positions in the brain). The basic principle to visualize these data in a common 2D space is the generation of a mean embedding space from a mean affinity matrix. Therefore, *LPP*, a linear approximation of non-linear dimensionality reduction, was chosen since it produces a projection matrix [HN03], while other non-linear dimensionality reduction methods, such as Diffusion Maps (as used by Margulies et al. [MGG*16]), are only defined for the data points they have been computed for (i.e. no projection matrix can be obtained). The projection matrix can then be used to map the individual affinity matrices (i.e. from each timepoint) onto this space.

In a first step, a mean affinity matrix is computed, where the mean affinity between two nodes is the mean affinity of these two nodes over all timepoints (Figure 2 (1)). In a second step, all affinity matrices keep only the top affinities (strongest 10%) of each

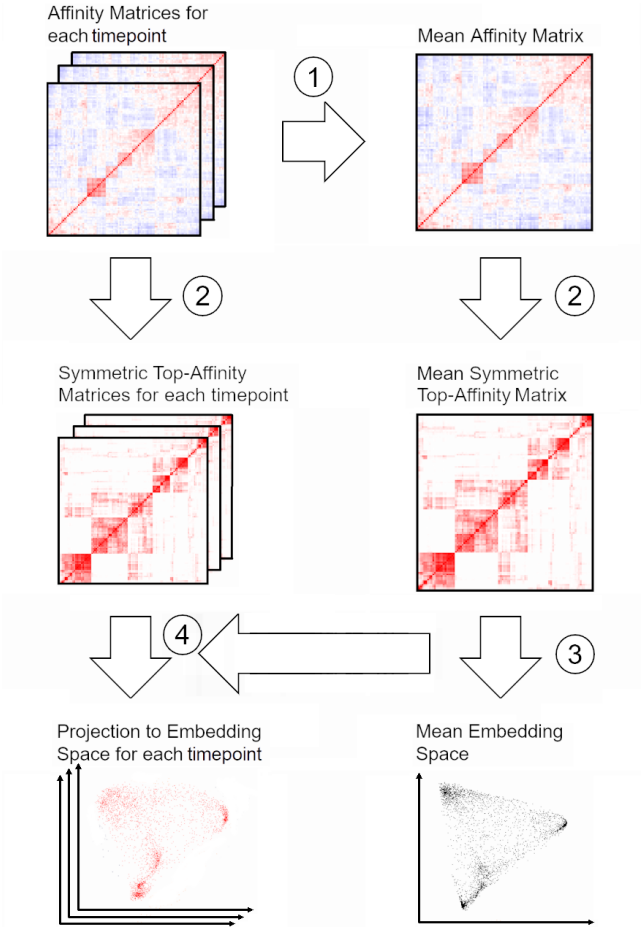

Figure 2: *Schematic description of the procedure to map affinity matrices to a common embedding space. 1) Computing the mean affinity matrix. 2) Creating symmetric top-affinity matrices for each affinity matrix. 3) Applying LPP on the mean symmetric top-affinity matrix. 4) Projecting symmetric top-affinity matrices of each timepoint to the mean embedding space by using the projection matrix obtained in the previous step.*

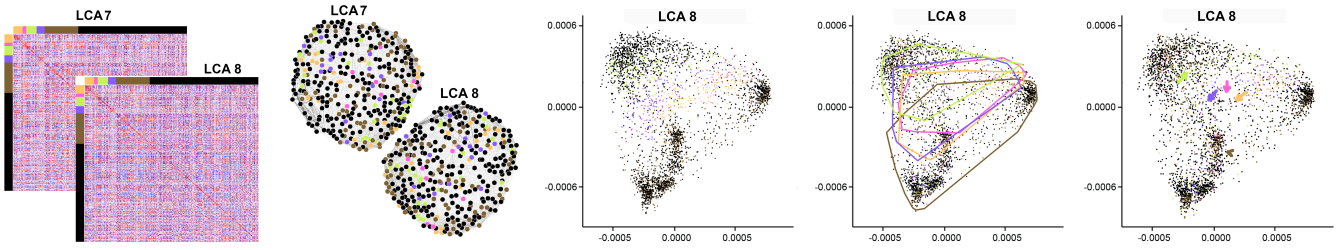

Figure 3: *Different design iterations of the affinity matrices. Brain-regions are colour coded. (a) Heatmap visualization (b) Force-directed node-link diagram of a random sample, otherwise it would be too complex (c) Colour-coded dots. (d) Contours were added to span around the brain regions (filtered for outliers based on the Mahalanobis distance to the region center, in this case,* 50%*). (e) Added arrows, representing a vector between the center of gravity of a brain region from one time point to another.*

row to lay the focus in the embedding space on local affinity rather than global. Since this creates an asymmetric matrix, cosine similarity is applied to the rows to generate a symmetric matrix (*LPP* requires a positive, symmetric matrix) (Figure 2 (2)). A mean embedding space is created by applying *LPP* on the mean symmetric top-affinity matrix, where eigenvalue decomposition produces a projection matrix (Figure 2 (3)). Multiplication of the individual affinity matrices with the projection matrix map them to a common space (Figure 2 (4)). An easy-to-use short code snippet in *R* to perform this process is available at `http://github.com/NeuroscienceTools/LPP_density_plots`.

3. Visual Encoding

To trace brain regions of interest over time, we iteratively developed a visual encoding in informal discussions with our domain experts (both co-authors of this paper). First, we colour-coded the nodes according to five major brain regions in the Allen Human Brain Atlas [HLGB*12]. Figure 3 (c) gives an idea how the regions are distributed, but it is still not feasible to track them over several time points or make out clusters. Therefore, we added a convex hull (Figure 3 (d)), spanning around brain regions (filtered for outliers based on the Mahalanobis distance to the region center). As the regions do not form distinct clusters, and are therefore distributed over the whole embedding space, these hulls are still too coarse to trace changes of regions over time. To improve this perception, we added arrows, representing a vector between the center of gravity of a brain region from one time point to another (Figure 3 (e)). We performed this first on a node level, which led to cluttered results. However, our domain experts were not able to infer biological meaning from these movements without taking the distribution within regions into account.

As a consequence, we overlaid the local density of the region's associated nodes on the mapping (Figure 1). We removed the colour coding of individual nodes as it would interfere with the overlayed density. To compute the local density, we used a two-dimensional kernel density estimation evaluated on a square grid. These "clouds" or "density maps" of major brain regions are then combined via alpha blending in a single figure per timepoint. The transitions are visualized by juxtaposing individual timepoints (Figure 1) as small multiples to allow the visual tracing of the development of individual brain regions and identifying subclusters.

4. Case-Study

For the creation of the co-evolution networks we used genomic signatures, the dN/dS ratios of chronologically ordered mammalian species – mouse, bushbaby, marmoset, macacque, gibbon, chimp, denisovan, neanderthal, and human – from previous work, Kasczanowska et al. [KGG*19]. Evolutionary timepoints represent the genetic split from two species' evolutionary last common ancestor (*LCA*), where *LCA 1* is the last common ancestor of mouse and bushbaby, *LCA 2* of bushbaby and marmoset and so on. We extracted the genes with the highest (top 10%) dN/dS ratios (i.e. highest selection pressure) for all evolutionary timepoints (i.e. eight gene sets). To map these genes to brain regions and ultimately form networks, we used spatial gene expression data retrieved from the Allen Human Brain Atlas [HLGB*12]. The data assemble gene expression for 3702 biopsy-sites (i.e. positions in the brain) from microarray data. The spatial gene expression correlation between biopsy sites was then computed for each gene set/time point. These correlations represent eight co-evolution networks, where the positive correlations indicate the co-evolution (i.e. affinity) between biopsy sites (i.e. network nodes).

We investigated these eight affinity matrices first as heatmaps, spatially ordered by brain regions (Figure 3 (a)) and as force-directed node-link diagrams (Figure 3 (b)) with our domain experts. Due to the extent of the data, as well as relatively small variations of the affinity, it was not possible to identify major changes at a glance. Furthermore, the node-link diagram lacked spatial alignment. When using our proposed visualization in the form of small multiples (Figure 1), it was possible to focus on the deformation of, and shifting between the colored point clouds representing the brain regions over multiple timepoints. Several subcortical areas showed intra-regional deformations, i.e. splitting into subclusters (*Thalamus* at *LCA 2 and 3*, *Striatum* at *LCA 3,4 and 5*), inter-regional overlaps (*Striatum*, *Hippocampus* and *Amygdala* at *LCA 4 and 5* at early timepoints, and were getting increasingly fixed at later timepoints *LCA 6, 7 and 8*. In contrast to subcortial regions, cortical co-evolution (*Prefrontal Cortex*) spanned more variation across the time period and remained continuously flexible as indicated in the literature by chimp-human comparisons [GRHSS15].

5. Discussion and Conclusion

Section 4 showed the potential and relevance of our approach in neurobiological research. By comparing the mapping of different timepoints onto a common embedding space using small multiples, affinity changes within and between brain regions were clearly visible. Given the size of the matrices, this would not have been possible with heatmaps or node-link diagrams. Our approach provides alignment over time, and therefore traceability for direct comparison, even without an interactive visualization. Although not explicitly covered in this study, the generic nature of affinity matrices does not limit the methods application in other fields than spatial brain networks.

Nevertheless, there is still room for improvements. First of all, the common embedding space where the individual time points are mapped to, is based on the mean affinity matrix. Therefore, the quality of the individual mappings decreases with the variance, especially with more outliers in the data. One way to counteract this is a robust estimation of the mean (trimmed mean, median), or in extreme cases Procrustes analysis as suggested by Vos de Wael et al. [VdWBP*20]. Secondly, the amount of clusters of interest to be traced are limited by the finite amount of colours that can be used. This is especially crucial, since the outcome of the analysis depends on which clusters have been selected and are thereby visually traceable. A thorough pre-selection is advisable, which could be supported by time-depended clustering methods. This could result in more distinct, compact clusters, where previously rejected approaches in Section 3 (contours and arrows) have the potential to lead to meaningful results.

Ultimately, the proposed method can be seen as a first step towards a more detailed analysis. Although it allows a quick assessment of changes in large networks, using this approach in a visual analytics tool could enable interactive selection of clusters/brain regions, linked views with brain anatomy, direct investigation of individual nodes via brushing, and the use of animations/scrolling instead of small multiples.

6. Acknowledgments

VRVis is funded by BMVIT, BMDW, Styria, SFG and Vienna Business Agency in the scope of COMET (854174) which is managed by FFG. Wulf Haubensak was supported by the IMP, Boehringer Ingelheim, the FFG, and a grant from the European Community's Seventh Framework Programme (FP/2007-2013) / ERC grant agreement no. 311701. We want to thank Lisa Frauenstein, who has laid the basis for this publication with her master's thesis [Fra17].

References

[ABH*13] Alper B., Bach B., Henry Riche N., et al.: Weighted Graph Comparison Techniques for Brain Connectivity Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), CHI '13, pp. 483–492. 1

[BHRD*15] Bach B., Henry-Riche N., Dwyer T., et al.: Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks. *Computer Graphics Forum 34* (2015), 31–40. 2

[FDC*07] Fair D. A., Dosenbach N. U. F., Church J. A., et al.: Development of distinct control networks through segregation and integration. *Proceedings of the National Academy of Sciences 104* (2007), 13507–13512. 1

[FNS*18] Frauenstein L., Nenning K.-H., Schwartz E., et al.: Dimensionality reduction for analysis of functional connectivity in the developing human brain. In *Forum of Neuroscience of the Federation of European Neuroscience Societies* (2018). 2

[Fra17] Frauenstein L.: *Dimensionality Reduction for Analysis and Visualization of Functional Connectivity in the Developing Human Brain*. Master's thesis, Fakultät für Informatik, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria, 2017. 4

[FSG*18] Frauenstein L., Swoboda N., Ganglberger F., et al.: A prototypic visual analytics framework for interactive exploration of functional connectivity development. In *Forum of Neuroscience of the Federation of European Neuroscience Societies* (2018). 2

[GRHSS15] Gomez-Robles A., Hopkins W., Schapiro S., Sherwood C.: Relaxed genetic control of cortical organization in human brains compared with chimpanzees. *Proceedings of the National Academy of Sciences 112* (2015). 3

[HLGB*12] Hawrylycz M. J., Lein E. S., Guillozet-Bongaarts A. L., et al.: An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature 489* (2012), 391–399. 3

[HN03] He X., Niyogi P.: Locality preserving projections. In *NIPS* (2003), vol. 16. 2

[HSS*20] Hinterreiter A., Steinparz C., Schöfl M., et al.: Exploring visual patterns in projected human and machine decision-making paths. 2

[HVP*19] Höllt T., Vilanova A., Pezzotti N., et al.: Focus+context exploration of hierarchical embeddings. *Computer Graphics Forum 38* (2019). 2

[KGG*19] Kaczanowska J., Ganglberger F., Galik B., et al.: Molecular archaeology of the human brain. *bioRxiv* (2019). 2, 3

[KRM*17] Kruiger J. F., Rauber P. E., Martins R. M., et al.: Graph Layouts by t-SNE. *Computer Graphics Forum 36* (2017). 2

[LGG15] Langs G., Golland P., Ghosh S. S.: Predicting activation across individuals with resting-state functional connectivity based multi-atlas label fusion. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015), pp. 313–320. 2

[MGG*16] Margulies D. S., Ghosh S. S., Goulas A., et al.: Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences 113* (2016), 12574–12579. 2

[MKF*15] Ma C., Kenyon R. V., Forbes A. G., et al.: Visualizing Dynamic Brain Networks Using an Animated Dual-Representation. In *Eurographics Conference on Visualization (EuroVis) - Short Papers* (2015). 1

[MML*11] Markram H., Meier K., Lippert T., et al.: Introducing the Human Brain Project. In *Procedia Computer Science* (2011). 1

[OHN*14] Oh S. W., Harris J. A., Ng L., et al.: A mesoscale connectome of the mouse brain. *Nature 508* (2014), 207–214. 1

[SCH*18] Senk J., Carde C., Hagen E., et al.: VIOLA—A Multi-Purpose and Web-Based Visualization Tool for Neuronal-Network Simulation Output. *Frontiers in Neuroinformatics 12* (2018), 75. 2

[VdWBP*20] Vos de Wael R., Benkarim O., Paquola C., et al.: Brainspace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Communications Biology 3* (2020). 2, 4

[VESB*13] Van Essen D. C., Smith S. M., Barch D. M., et al.: The wu-minn human connectome project: an overview. *Neuroimage 80* (2013), 62–79. 1

[VHB*14] Van Den Elzen S., Holten D., Blaas J., et al.: Dynamic network visualization with extended massive sequence views. *IEEE Transactions on Visualization and Computer Graphics 20* (2014). 2

[XNS*19] Xu T., Nenning K.-H., Schwartz E., et al.: Cross-species functional alignment reveals evolutionary hierarchy within the connectome. *bioRxiv* (2019). 2