RESEARCH



Discovering a change point and piecewise linear structure in a time series of organoid networks via the iso-mirror

Tianyi Chen¹, Youngser Park¹, Ali Saad-Eldin¹, Zachary Lubberts¹, Avanti Athreya¹, Benjamin D. Pedigo¹, Joshua T. Vogelstein¹, Francesca Puppo², Gabriel A. Silva³, Alysson R. Muotri², Weiwei Yang⁴, Christopher M. White⁴ and Carey E. Priebe^{1*}

*Correspondence: cep@jhu.edu

 ¹ Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, USA
² Department of Pediatrics, University of California, San Diego, USA
³ Department of Neurosciences, University of California, San Diego, USA
⁴ Microsoft Research, Seattle, USA

Abstract

Recent advancements have been made in the development of cell-based in-vitro neuronal networks, or organoids. In order to better understand the network structure of these organoids, a super-selective algorithm has been proposed for inferring the effective connectivity networks from multi-electrode array data. In this paper, we apply a novel statistical method called spectral mirror estimation to the time series of inferred effective connectivity organoid networks. This method produces a one-dimensional iso-mirror representation of the dynamics of the time series of the networks which exhibits a piecewise linear structure. A classical change point algorithm is then applied to this representation, which successfully detects a change point coinciding with the neuroscientifically significant time inhibitory neurons start appearing and the percentage of astrocytes increases dramatically. This finding demonstrates the potential utility of applying the iso-mirror dynamic structure discovery method to inferred effective connectivity time series of organoid networks.

Keywords: Organoid, Time series network analysis, Multidimensional scaling, Manifold learning, Change point detection

Introduction

Detecting structural changes in time series of networks is central to many modern network science applications. However, due to the complexity of temporal network data and the myriad possible aspects for potential structural change, this problem can be daunting. For discovering underlying dynamics in time series of networks, Athreya et al. (2022) proposes theory and methods for representing temporal network structure with a curve, or 'mirror', in low dimensional Euclidean space, enabling the use of classical change point detection algorithms. In this paper, we estimate the mirror for a time series of brain organoid connectivity networks and subsequently identify change points. Because the mirror estimation method requires a 1-1 vertex correspondence for the networks across time, we first demonstrate that the putative 1-1 correspondence obtained directly from data collection is sufficiently accurate by comparing it to the vertex



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativeCommons.org/licenses/by/4.0/.

correspondence obtained via graph matching (Vogelstein et al. 2015). Thence, mirror estimation and manifold learning recovers a 1-dimensional piecewise linear 'iso-mirror' representation, with an evident slope change. By using the change point detection algorithm from (Bücher et al. 2021) and break point estimation for piecewise linear models from (Muggeo 2017), we identify a change points neuroscientific significance coinciding with development stages.

We organize this paper as follows. In section "Organoids" we introduce our brain organoids data and the extraction of effective connectivity networks based on extracellular electrophysiology recordings. In section "Graph Matching" we define the graph matching problem and present the fast approximate assignment algorithm. In section "Discovering underlying dynamics in times series of networks" we introduce the mirror estimation method and relevant model assumptions. In the "Results" section we apply these methods to the time series of organoid networks and present the graph matching results and the change point detection results. We conclude the paper with a discussion.

Organoids

The brain organoids we consider are self-organizing structures composed of roughly 2.5 million neural cells. They are generated from human induced pluripotent stem cells (hiPSCs) (Muguruma et al. 2015). After growing for 6 weeks, they are plated in 8 wells of a multi-electrode array (MEA) plate (Axion Biosystem, Atlanta, GA, USA). MEA contains 64 low-impedance (0.04 MU) platinum microelectrodes with 30 µm of diameter spaced by 200 µm. Each well contains two or three organoids. Then, to characterize the functional development of the organoids, extracellular spontaneous electrical activity is recorded weekly using Maestro MEA system and AxIS Software Spontaneous Neural Configuration (Axion Biosystems) with a 0.1-Hz to 5-kHz band-pass filter. Then spikes are detected with an adaptive threshold crossing set to 5.5 times the standard deviation of the estimated noise for each electrode. Each time series consists of five minutes of recorded neural activity across 10 months and the data are recorded irregularly-not exactly once a week. Cortical organoids show low and sparse activity during the first 2 months with an average firing frequency of 0.5-Hz, then they start exhibiting highly synchronized and stereotypical network activity which transitions into 2-Hz and 3-Hz rhythmic activity by 4-6 months. At later stages (6 to 9 months), the activity includes high-rate spiking with peak of activity reaching a 20-Hz pace and highly complex bursting behaviors with cross-frequency coupling (Puppo et al. 2021). A combination of principal component analysis and k-means clustering is used to spike sort the multi-unit activity from the 64 electrodes of each well. The average number of neurons detected in each well increases with the maturation of the organoids onto the electrodes, sometime reaching saturation after 6 months. Similar behavior is observed in all MEA wells. For example, the number of spike-sorted neurons in one well is detected as 122, 160, 189, 174, 190 at 2, 4, 6, 8 and 10 months. The other well has 81, 160, 174, 171, 170 neurons detected at 2, 4, 6, 8, and 10 months.

To infer the effective connections between neurons—i.e., the adjacency matrices with neurons as vertices across time from the spike activity data—the algorithm proposed in Puppo et al. (2021) is applied. This algorithm uses a super-selection rule to individuate and discard correlation peaks corresponding to apparent and indirect interactions, and

reconstructs the effective connectivity of the network considering the remaining correlation delays. In the end, 45 effective connectivity networks on 127 vertices across 244 days are obtained. This is our time series of organoid networks.

Graph matching

Given a time series of networks G_1, \dots, G_T on the same set of vertices V, a 1-1 vertex correspondence across all the networks facilitates joint spectral embedding, which is a key step in the mirror estimation method we will present in the next section, "Discovering underlying dynamics in time series of networks". Such a vertex correspondence may be available a priori in labeled networks, or it can be inferred from unlabeled networks. This inference problem—the so-called graph matching problem—is to find an alignment of vertices between two graphs such that the corresponding edge differences are minimized. We denote A, B as two $n \times n$ adjacency matrices for two graphs with n vertices each. Then the graph matching problem is to find the permutation matrix P that maximizes the objective function

 $f(A, B; P) = -||A - P^{\top}BP||_{F}^{2}$

for $P \in \mathcal{P}$ where $\mathcal{P} = \{P \in \{0, 1\}^{n \times n} : P^\top \mathbf{1} = P\mathbf{1} = \mathbf{1}\}, \mathbf{1} = (1, 1, ...1)^\top$, and $|| \cdot ||_F$ denotes the Frobenius norm. This formulation is equivalent to maximizing

 $f(A, B; P) = trace(APB^T P^T).$

Because solving this optimization problem is combinatorically difficult, approximation algorithms have been proposed. We use the Fast Approximate Quadratic (FAQ) assignment algorithm (Vogelstein et al. 2015) to obtain an approximate solution. FAQ iteratively finds a local solution to the relaxed problem—expanding the constraint set to the convex hull of \mathcal{P} —and then projects the solution back to \mathcal{P} , and has been shown empirically to be competitive with or superior to other state-of-art methods. In Section Results we use FAQ to demonstrate that the given putative 1-1 correspondence for each pair of networks is close to the solution to the corresponding graph matching problem.

Discovering underlying dynamics in time series of networks

To discover the underlying dynamics in time series of networks, we use the model and method proposed in Athreya et al. (2022). First we introduce the generative joint model for time series of networks. We consider *T* networks, each containing *n* vertices, with adjacency matrices $A_t, t \in \{1, 2, ..., T\}$. In the model, each vertex is associated with a time varying *d*-dimensional latent vector. These vectors X_t are each one realization from a stochastic process, called the latent position process (LPP)—each X_t is a *d*-dimensional random variable. For *n* vertices, we generate *n* i.i.d. samples from the LPP, which collection then forms the latent position matrices $\{X_t\}$, where $X_t \in \mathbb{R}^{n \times d}$ for $t \in \{1, 2, ..., T\}$. The connection probability between vertex *i* and vertex *j* at time *t* is the inner product of the associated latent vectors at time *t*. That is $\mathbb{E}(A_t) = \mathbf{X}_t \mathbf{X}_t^T$. Note that each network corresponds to a latent position random variable. Thus we can capture the distance between graphs using the corresponding random variables. We define the distance

$$d_{MV}(X_t, X_{t'}) = \min_{W \in \mathcal{O}^{d \times d}} \|\mathbb{E}[(X_t - WX_{t'})(X_t - WX_{t'})^{\top}]\|_2^{1/2},$$

where $\mathcal{O}^{d \times d}$ is the set of orthogonal transformation matrices with dimension d. When X_t and $X_{t'}$ are centered, the d_{MV} distance can be interpreted as the maximum directional variation for the random vector $X_t - WX_{t'}$, where W is an orthogonal transformation used to align X_t and $X_{t'}$. We evaluate this distance for every pair of random variables in the LPP and obtain a $T \times T$ distance matrix \mathcal{D} . Then we apply classical multidimensional scaling (MDS) (Torgerson 1952) to \mathcal{D} to get a low-dimensional Euclidean representation of underlying network structure, called the *mirror*, { $\psi(t)$ }. In practice, the LPP is unknown and only network realizations { A_t } are observed. For the $n \times n$ symmetrized adjacency matrix A_t , we use adjacency spectral embedding (ASE) (Athreya et al. 2018) to obtain $\hat{\mathbf{X}}_t = U \Sigma^{1/2}$, where the diagonal matrix Σ contains the top d eigenvalues of A_t and U contains the associated eigenvectors. Then we use

$$\hat{d}_{MV}(\hat{X}_t, \hat{X}_{t'}) = \min_{W} \frac{1}{\sqrt{n}} \|\hat{X}_t - \hat{X}_{t'}W\|_2^{1/2}$$

to estimate the pairwise distance between networks, yielding \hat{D} . Applying MDS to \hat{D} yields the mirror estimate $\{\hat{\psi}(t)\}$. When the mirror exhibits a manifold structure, we can further simplify the change point detection problem by applying the manifold learning method isometric mapping (ISOMAP) (Tenenbaum et al. 2000) to $\{\hat{\psi}(t)\}$. This yields the *iso-mirror*, which captures the geodesic distance along the mirror and preserves it in lower dimensions with Euclidean distance. Subsequent inference is then performed using the iso-mirror representation. For convenience, the iso-mirror representation will also be denoted as $\{\hat{\psi}(t)\}$.

Results

Time series of organoid networks data

All results in this section are based on data collected from well 8. For analogous results from well 5, please refer to the Appendix. For well 8, the time series of organoid networks consists of 45 time stamps $\{1, 2, ..., 45\}$; each time stamp corresponds an effective connectivity graph G_t with adjacency matrix A_t . Each graph is directed, weighted, and hollow. We symmetrize the directed graphs, and use ranks in place of the raw edge weights. All graphs have the same vertex set $V = \{1, 2, ..., n\}$ with n = |V| = 127, although some of the graphs contain isolated vertices. See Fig. 1.

Putative 1-1 correspondence

For this time series of organoid networks, inferred neuron location gives a putative 1-1 vertex correspondence across the graphs. We assess this correspondence via graph matching using the objective function value (OFV) f(A, B; P). The OFV for the putative 1-1 correspondence is given by f(A, B; I) where I is the identity matrix. We denote the FAQ output for matching adjacency matrices A and B initialized at C as $P_{A,B;C}$. Typically we choose the barycenter $b = \frac{11^T}{n}$ as the initial point. For all times $i \in \{1, 2, ..., 44\}$, we consider A_i , A_{i+1} and FAQ yields $P_{A_i,A_{i+1};b}$ (denoted $P_{i,i+1;b}$ for short). In Fig. 2 we



Fig. 1 The number of non-isolated vertices and the number of edges for the graphs at each of the 45 time stamps. The number of edges are counted after symmetrizing the directed graph

compare $f(I) = f(A_i, A_{i+1}; I)$ and $f(P_{i,i+1;b}) = f(A_i, A_{i+1}; P_{i,i+1;b})$. Although $f(P_{i,i+1;b})$ is always larger than f(I), the two OFVs are close to each other for all time stamps, indicating that the putative 1-1 correspondence is close to FAQ's solution.

To further asses the putative 1-1 correspondence, we consider a specific pair of adjacency matrices: A_{28} , A_{29} . We uniformly generate 100,000 random permutation matrices R and evaluate $f(R) = f(A_{28}, A_{29}; R)$. We plot the histogram in Fig. 3. We also indicate $f(A_{28}, A_{29}; I)$, $f(A_{28}, A_{29}; P_{28,29;I})$, $f(A_{28}, A_{29}; P_{28,29;B})$ and $f(A_{28}, A_{29}; P_{28,29;R})$, where R are 100 randomly drawn permutation matrices. We see that the putative 1-1 correspondence performs better than all 100,000 instantiations of f(R) and is close to $P_{28,29;b}$, $P_{28,29;I}$ and $P_{28,29;R}$. Thus we conclude that the putative 1-1 correspondence is sufficiently accurate, and we will proceed apace for mirror estimation and change point detection.

Change point detection

For the 45 graphs, time stamps are from 1 to 244, in days. We choose time stamps in [150,230] to avoid growth and death regimes.

For these graphs, we find the largest common connected component, which contains 112 vertices. The average number of edges for the largest common connected component is approximately 6130. We use the putative 1-1 vertex correspondence across time. We apply our mirror estimation method to this time series of networks, and ISOMAP manifold learning yields the 1-dimensional representation of the dynamics $\{\hat{\psi}_t\}$. We choose 10 time stamps shown in Fig. 4. As we see, the representation is approximately piecewise linear with an evident change of slope at t = 4, day 188.



Fig. 2 Comparison of OFV using *I* and $P_{i,i+1;b}$ for 44 pairs of graphs, demonstrating that FAQ increases the OFV only slightly. **a** f(l) and $f(P_{i,i+1;b})$. **b** $\frac{f(P_{i,i+1;b})}{f(l)}$

Assuming the true underlying 1-dimensional representation $\psi(t)$, $t \in [0, T]$, is piecewise linear and continuous, it is natural to define the change point t^* as the point when the slope changes. If we assume there is only one change point, then we can write

$$\psi(t) = \beta_0 + \beta_1 t + \beta(t - t^*)I(t > t^*).$$

Both the change point detection algorithm from Bücher et al. (2021) and the break point estimation for piecewise linear models from Muggeo (2017) yield an estimated change point $\hat{t^*} = 4$, day 188, which coincides with neuroscientifically significant developmental changes—inhibitory neurons start appearing and the percentage of astrocytes increases dramatically—as described in Trujillo et al. (2019). Note that the emergence



Fig. 3 Matching A_{28} and A_{29} . Comparison of OFV for using I_{R} , $P_{28,29;b}$, $P_{28,29;b}$, $P_{28,29;l}$, and $P_{28,29;R}$. **a** Histogram demonstrating that OFV for R is not nearly as good as for the others. **b** Enlargement of the far right portion of the top figure, demonstrating that FAQ output at different initial points I, b and R improves the OFV only slightly compared to the putative 1-1, I

of astrocytes does not happen at once but builds over time, so there is no one precise date for the change point and detection of a time coinciding with this change can be but suggestive.

Conclusion

Reconstruction of effective connectivity networks of electrophysiologically active brain organoids reflect their structural (increasing number of neurons (nodes) and connections (edges)) and electrical development over time, as previously demonstrated in Trujillo et al. (2019).



Fig. 4 Discovering a change point in time series of inferred effective connectivity organoid networks via the iso-mirror. The x-axis is time stamp, in days. The y-axis is the ISOMAP representation of the estimated mirror $\{\hat{\psi}_l\}$. This indicates a change in the network dynamics at day 188

By applying the spectral mirror estimation method to the time series of organoid networks, we obtain a 1-dimensional iso-mirror representation of dynamic inferred effective connectivity organoid networks. Two change point detection algorithms successfully detect a change at day 188. At approximately 188 days (~6 months), cortical organoids start showing inhibitory neurons and the percentage of astrocytes increases from 5% to 30–40% (Trujillo et al. 2019) resulting in added complexity in the activity and network distribution of brain organoids.

There are several change point detection algorithms available for analyzing time series of graphs, including the one proposed in Wang et al. (2021). However, it is important to note that the spectral mirror estimation method used in our study is not restricted to change point detection alone. Instead, it provides a low-dimensional (in our case, one-dimensional Euclidean) representation of network dynamics, enabling us to visualize network evolution. As illustrated in Fig. 4, we observe piecewise linear structure and apply segment regression to identify a significant increase in slope after day 188. This suggests that organoid graphs drift continuously over time, but their rate of drift accelerates significantly after day 188. Our method is preferred in this regard as it provides us with more than just one change point. Additionally, since the iso-mirror represents the underlying LPP, the detected change point in the iso-mirror reflects a fundamental change in the underlying generative LPP of the time series of networks, which may not necessarily correspond to any specific network measure. Furthermore, the spectral mirror method can be applied to other dynamic graphs that meet our model assumptions, namely that the time series of networks are generated from a time series of latent



Fig. 5 The number of non-isolated vertices and the number of edges for the graphs at each of the 40 time stamps for well 5. The number of edges are counted after symmetrizing the directed graph

position random networks whose vertices have independent, identically distributed latent positions given by an LPP.

Future work includes addressing two major theoretical issues of note, to make the change point inference formally principled. First, the theory in Athreya et al. (2022) requires a known 1-1 vertex correspondence across time. It remains to study the effect of errors in this correspondence, such as those inherent in our putative 1-1 correspondence deemed sufficiently accurate for practical purposes. In addition, it remains to study the entry-wise behavior of the error term $\epsilon(t) = \hat{\psi}(t) - \psi(t)$. For example: in Bücher et al. (2021) the proof of consistency of the change point estimator requires the error process to be stationary; in Muggeo (2017) the $\epsilon(t)$ are assumed to be i.i.d. normal to construct a confidence interval for \hat{t}^* . For now, Athreya et al. (2022) has shown that $\hat{\psi}(t)$ converges to $\psi(t)$ in Frobenius norm, that is $\sum_{t=1}^{T} (\epsilon(t))^2 \rightarrow 0$ with high probability, which is insufficient to conclude normality or stationarity. What's more, this convergence result is for the mirror rather than the iso-mirror. As for whether the same conclusion can be extended to the iso-mirror, further investigation is needed.

Appendix

Time series of organoid networks data for well 5

For well 5, there are 40 graphs with time stamps [1229] and all of them have the same vertex set with |V| = 128, although some of the graphs contain isolated vertices. See Fig. 5. Each graph is directed, weighted, and hollow. We symmetrize the directed graphs, and use ranks in place of the raw edge weights.

Putative 1-1 correspondence for well 5

We assess the putative 1-1 correspondence exactly the same way as in the paper and Fig. 6 indicates that the putative 1-1 correspondence is close to FAQ's solution.



Fig. 6 Comparison of OFV using *I* and $P_{j,j+1;b}$ for 39 pairs of graphs, demonstrating that FAQ increases the OFV only slightly. **a** f(l) and $f(P_{i,j+1;b})$. **b** $\frac{f(P_{i,j+1;b})}{f(l)}$



Fig. 7 Matching A_{28} and A_{29} . Comparison of OFV for using $I_{,R}$, $P_{28,29;b}$, $P_{28,29;l}$, and $P_{28,29;R}$. Left: Histogram demonstrating that OFV for R is not nearly as good as for the others. Right: Enlargement of the far right portion of the left figure, demonstrating that FAQ output at different initial points I, b and R improves the OFV only slightly compared to the putative 1-1, I

We also consider a specific pair of adjacency matrices: A_{28} , A_{29} to assess the 1-1 putative correspondence. See Fig. 7. The result is similar to well 8 and we conclude the 1-1 putative correspondence is accurate enough.

Change point detection for well 5

For the 40 graphs, time stamps are from 1 to 244, in days. We choose the same time stamps in [150,230] as in the paper. We find the largest common connected component for these graphs. We use the putative 1-1 vertex correspondence across time. We apply our mirror estimation method to this time series of networks, and ISOMAP manifold



Fig. 8 For well 5, iso-mirror result

learning yields the 1-dimensional representation of the dynamics $\{\hat{\psi}_t\}$. We choose the same 10 time stamps as in the paper shown in Fig. 8. As we see, the representation is approximately piecewise linear. The break point estimation for piecewise linear models from Muggeo (2017) yield an estimated change point $\hat{t}^* = 4$, day 188.

Abbreviations

MEA	Multi-electrode array
PCA	Principal component analysis
FAQ	Fast approximate quadratic
LPP	Latent position process
MDS	Multidimensional scaling
ISOMAP	lsometric mapping
ASE	Adjacency spectral embedding
OFV	Objective function value

Author contributions

TC, ZL, AA, CEP developed the theory. ARM, FP, GAS provided the data. TC, YP, AS, BPD designed & implemented the methods. TC, YP conducted the experiments. TC, YP, CEP wrote the manuscript. WY, CW, JTV, CEP guided the whole process. All authors read and approved the manuscript.

Funding

TC's work is partially supported by the Johns Hopkins Mathematical Institute for Data Science (MINDS) Data Science Fellowship and Azure sponsorship credits granted by Microsoft's AI for Good Research Lab.

Availability of data and materials

The organoid data and code are available at https://github.com/youngser/organoid.

Declarations

Ethics approval and consent to participate Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 9 March 2023 Accepted: 17 June 2023 Published online: 21 July 2023

References

Athreya A, Fishkind DE, Tang M, Priebe CE, Park Y, Vogelstein JT, Levin K, Lyzinski V, Qin Y, Sussman DL (2018) Statistical inference on random dot product graphs: a survey. J Mach Learn Res 18(226):1–92

Athreya A, Lubberts Z, Park Y, Priebe CE (2022) Discovering underlying dynamics in time series of networks. arXiv preprint arXiv:2205.06877

Bücher A, Dette H, Heinrichs F (2021) Are deviations in a gradually varying mean relevant? A testing approach based on sup-norm estimators. Ann Stat 49(6):3583–3617

Muggeo VM (2017) Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach. Aust N Z J Stat 59(3):311–322

Muguruma K, Nishiyama A, Kawakami H, Hashimoto K, Sasai Y (2015) Self-organization of polarized cerebellar tissue in 3d culture of human pluripotent stem cells. Cell Rep 10(4):537–550

Puppo F, Pré D, Bang AG, Silva GA (2021) Super-selective reconstruction of causal and direct connectivity with application to in vitro ipsc neuronal networks. Front Neurosci 15:647877

Tenenbaum JB, Silva Vd, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323

Torgerson WS (1952) Multidimensional scaling: I. theory and method. Psychometrika 17(4):401–419

Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, Wang A, Wu W, Haddad GG, Chaim IA et al (2019) Complex oscillatory waves emerging from cortical organoids model early human brain network development. Cell Stem Cell 25(4):558–569

Vogelstein JT, Conroy JM, Lyzinski V, Podrazik LJ, Kratzer SG, Harley ET, Fishkind DE, Vogelstein RJ, Priebe CE (2015) Fast approximate quadratic programming for graph matching. PLoS ONE 10(4):0121002

Wang D, Yu Y, Rinaldo A (2021) Optimal change point detection and localization in sparse dynamic networks

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[™] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com