Reverse Engineering Functional Brain Networks from fMRI Data Using Probabilistic Boolean Networks

Erin Boggess*[1], Tiffany Jann*[2], Michael Stevens[3], Reinhard Laubenbacher[4], Paola Vera-Licona[4]

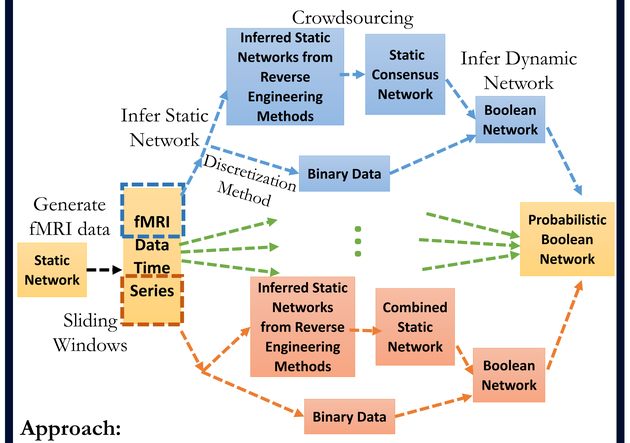
[1]Simpson College, [2]University of California, Berkeley, [3]Olin Neuropsychiatry Research Center, Institute of Living, [4]Center for Quantitative Medicine, UConn Health

*These two authors contributed equally to this work

Introduction

Motivation: The brain functions by communicating information across multiple regions, and neurological diseases can alter the way these brain regions communicate. Modeling the brain as a functional brain connectivity network (fBCN) from **functional Magnetic Resonance Imaging (fMRI)** data can allow clinicians to compare patient and non-patient brain networks and systematically diagnose brain disorders. fMRI is a is a functional neuroimaging procedure that indirectly measures brain activity. We want to model the mechanisms involved in brain disorders through **probabilistic Boolean networks (PBNs)**. PBNs are dynamic and stochastic, capturing moment to moment changes in the brain and making them robust to uncertainty and noise present in biological systems, data collection, and data processing. PBNs are defined as $G(V, \mathcal{F}, \alpha)$, where $V = \{v_1, \dots v_n\}$ represents a set of n brain regions, $\mathcal{F} = \{F_1, \dots Fn\}$ represents a set of n corresponding families of Boolean functions, and $\alpha = \{a_1, \dots an\}$ represents a set of n families of associated network selection probabilities corresponding to each F_i.

Objective: Develop and test a pipeline to reverse engineer a fBCN from fMRI data using Probabilistic Boolean Networks.



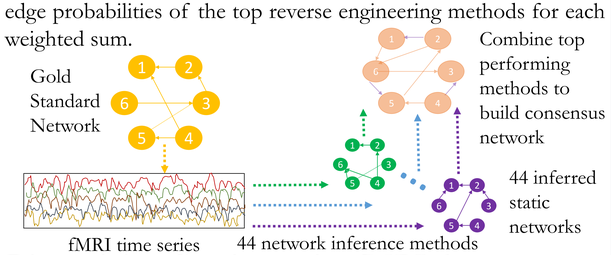
Approach:

To test our pipeline, we generated 10 fMRI data sets with 49 brain regions and 204 time points from 10 static networks (5 Barabasi-Albert scale free networks and 5 Erdos-Renyi random networks). We divided the datasets using sliding windows and computed a dynamic Boolean network for each window. We then generated PBNs by combining the Boolean models.

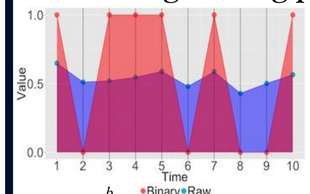
Methods

Reverse engineering static network

We generated static brain networks by applying 44 inference methods encoded in the software MULAN [1]. 42 of the methods came from neuroscience and 2 came from molecular biology [2,3]. To determine which combinations of inference methods would out-perform individual methods, we computed 11 weighted sums of two metrics: accuracy (ACC) and area under the receiving operating characteristic curve (AUC of the ROC) [4]. We then computed 198 consensus networks by taking the mean and median edge probabilities of the top reverse engineering methods for each weighted sum.



Discretizing floating point fMRI data



$$\sum_{nodes\ in\ v} \sum_{a} \left| (f_{raw}) - (f_{bin}) \right| \times dx$$

$$f_{raw} = m_{raw}x + b_{raw}$$

$$f_{bin} = m_{bin}x + b_{bin}$$

We benchmarked 11 discretization methods encoded in GED PRO TOOLS [5]. We proposed a novel benchmarking metric that compares the discretized data with the normalized raw fMRI data by computing the absolute area difference between data sets.

Lower area, less error, better performance

Inferring deterministic Boolean network

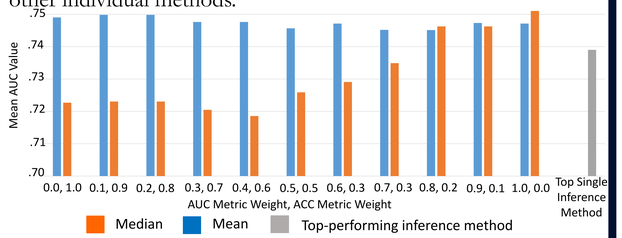
We inferred a dynamic Boolean network by inputting the discretized fMRI data and consensus networks into REACT [6]. We systematically tuned the parameters by running 50 simulations on the 10 data sets.

Inferring probabilistic Boolean network

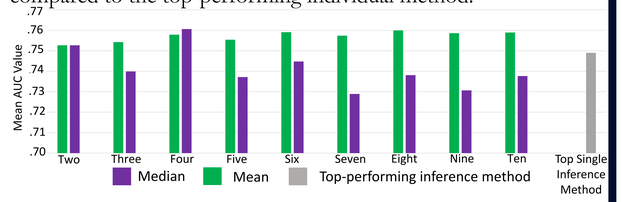
We divided the 5 discretized Barabasi-Albert data sets into two windows of length 150 with overlap of 96 and into four windows of length 101 with overlap of 34 or 35. To run REACT, each of the windows was paired with its static network. We made prototype PBNs by grouping Boolean update functions by brain region and assigning probabilities based on occurrence.

Results

This plot shows the results of the 11 AUC/ACC weighted sums. We compared these results to the top-performing individual method to observe which consensus networks out-perform all other individual methods.



This plot shows the performance of consensus networks with varying number of top-performing constituent methods compared to the top-performing individual method.

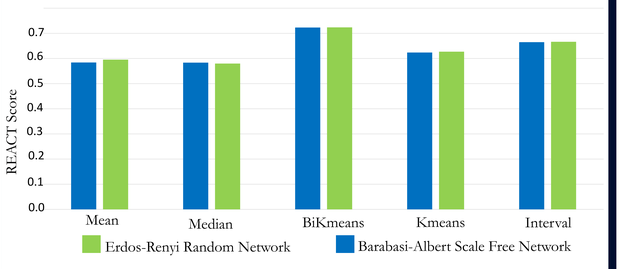


97.97% of the networks constructed by taking the mean of the edge probability of the constituent methods out-performed all 44 network inference methods. 45.5% of the consensus networks constructed using the median of the edge probabilities outperformed all 44 inference methods.

Top Discretization Methods

Rank	1	2	3	4	5
Scale Free	BiKmeans	Interval	Kmeans	Mean	Median
Random	BiKmeans	Interval	Kmeans	Mean	Median

REACT Score for Each Discretization Method



Discussion

- Consensus networks constructed using the mean edge probabilities of the constituent methods consistently out-performed all individual methods.
- Our experimentation with which weighted sum and how many constituent methods to include in the consensus networks only differed to the thousandths digit when constructing networks using the mean edge probability.
- We found that combining inference methods outperforms individual methods, which is consistent with the learnings of molecular biology [7].
- REACT performed best when the fMRI data was discretized using the BiKmeans method. This is consistent with the predictions from our benchmarking metric.
- Similarly, in the preliminary probabilistic Boolean network prototypes, we observed recurring Boolean functions when using BiKmeans output, suggesting that we can use the occurrence of Boolean functions to assign probabilities.

Conclusion and Future Work

- Due to their dynamic, functional, and stochastic nature, PBNs can be used to model the functional connectivity of brain networks.
- In the future, to evaluate the success of our pipeline we will 1) hide part of the fMRI time series when training the model and see how well the model predicts the hidden data 2) compare our method to other methods for computing PBNs.
- We used *in silico* data in order to evaluate our pipeline. To assess its clinical value, we will compare the *in silico* data generated by MULAN to experimental data.

References

[1] Wang, H. E., et al. (2014). Frontiers in Neuroscience
[2] Haury, A.-C., et al. (2012). BMC Systems Biology.
[3] Huynh-Thu, V. A., et al. (2010). PLoS ONE
[4] Bradley, A. P., (1997). Pattern Recognition
[5] Gallo, C.A. et al. (2015). Briefings in Bioinformatics
[6] Vera-Licona, P. et al. (2014). BMC systems biology
[7] Marbach, D., et al. (2012). Nature Methods

Acknowledgements

- Funded by National Science Foundation award #1460967
- Contributions by Christopher Tseng and Shichao Wang