A Joint Graph Inference Case Study: the *Caenorhabditis elegans* Neural Networks

Li Chen

Applied Mathematics and Statistics
Johns Hopkins University,
Baltimore, MD 21218

lchen87@jhu.edu

Joint work with Joshua Vogelstein and Carey Priebe

August 6, 2014

Overview

The *Caenorhabditis* Neural Network

The *C. elegans* neural network

- The *C.elegans* is a non-parasitic and transparent roundworm.
- 253 neurons. Each neuron belongs to exactly one neuron type: motor(43.5%), interneurons (30%), and sensory(26.5%).
- Two types of synaptic connections: chemical A_c and electrical A_g. They result in a pair of neural networks.

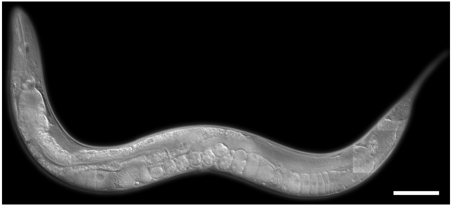


Figure : An image of the *Caenorhabditis elegans* (*C.elegans*) roundworm.

Graphs

Making inferences about graphs

- Graph $G = (V, E)$ consists of vertices and edges. Adjacency matrix A for undirected graph G.
- Vertex based inferences: clustering, classification, nomination, matching,...

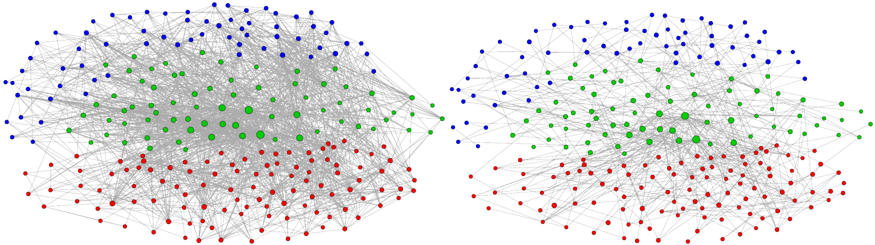


Figure : (Left) Chemical synaptic neural network A_c. (Right) Electrical synaptic neural network A_g. Red: motor. Green: inter. Blue: sensory.

Statistical Models for Graphs

Random Graph Models

- Erdos-Renyi Graph: each edge is present independently with equal probabilities.
- Stochastic Blockmodel: each vertex is a member of one block. The block membership of a pair of vertices determine their edge presence probabilities.
- Latent Position Models: each vertex has latent attributes. Edge probabilities are based on a link function.

Stochastic Blockmodel (SBM)

Definition

- K: number of blocks.
- B: $K \times K$ symmetric matrix specifying the probability of block connectivities.
- π: a length K block membership probability vector.
- Y: block membership of each vertex, given by $Y : \pi \to [K]$.

Then $A \sim SBM([n], B, \pi)$ if the edge probabilities are conditionally independent given the block memberships, and determined by entries of B given the memberships.

Model Dimension

The SBM is d-dimensional if $\text{rank}(B) = d$.

The block-structure of the neural network

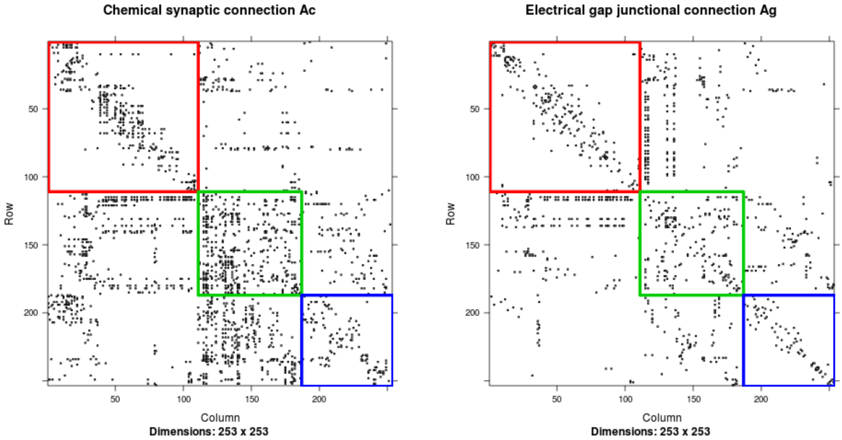


Figure : The adjacency matrices of the *C. elegans* neural networks A_c and A_g.

Low rank eigen-structure of the neural network

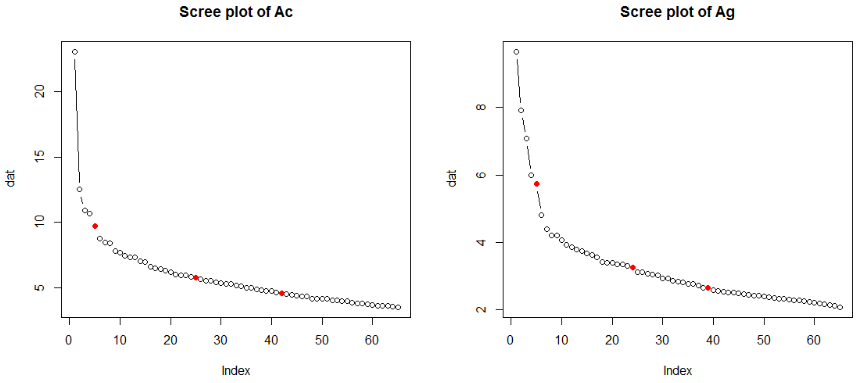
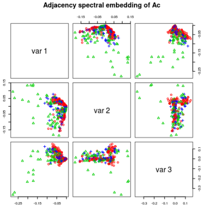


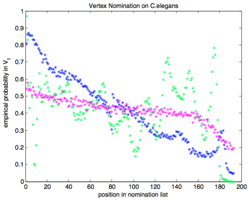
Figure : The low-rank eigen-structure of the *C. elegans* neural network.

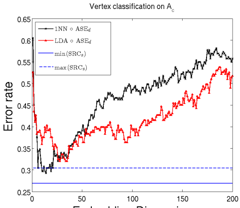
Joint Graph Inference Investigation

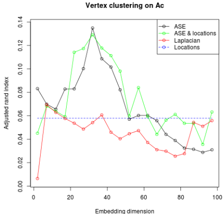
Inference on single networks

Vertex classification [1], [5] , vertex clustering [4] , vertex nomination [3].









Joint Graph Inference Framework

Joint inference on a pair of networks

We focus on two aspects of joint graph inference:

- Seeded graph matching - finding the correspondence of vertices across the pair of *C.elegans* neural network.
- Joint vertex classification - predicting the class membership of a vertex using information from the joint graph space.

Seeded Graph Matching

The Problem of Graph Matching

- Given two adjacency matrices A, B.
- Objective: minimize the number of edge disagreements.

$$\arg\min_{P \in \mathcal{P}} f(P) = \arg\min_{P \in \mathcal{P}} \|A - PBP^T\|_F = \arg\max_{P \in \mathcal{P}} \operatorname{tr}(APBP^T). \tag{1}$$

- Tool: Frank-Wolfe Algorithm.

The Problem of Seeded Graph Matching

- Seeds: vertices whose true alignments are known.
- Addition of seeds improves accuracy.
- Small change to graph matching algorithm.

Neurological Motivation for applying SGM on *C.elegans* neural network

- In neuroscience, it is interest to compare brains both within and across species.
- The extent of graph heritability with a species remains an open question.
- Compare graphs across species to enable comparative connectomics.

All of these basic science questions benefit from graph matching methods!

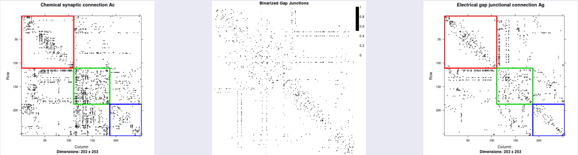
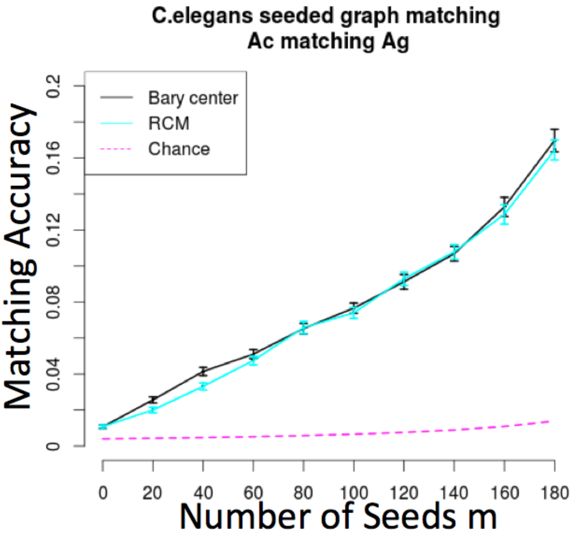
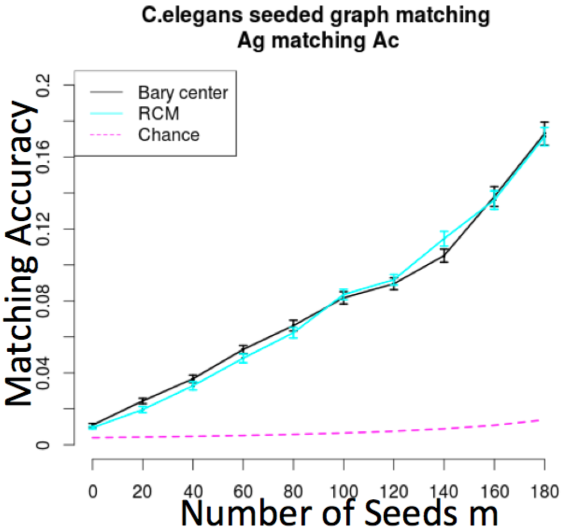


Figure : An visualization of the SGM procedure.

Finding the Correspondence between the Chemical and the Electrical Synapses



Finding the Correspondence between the Chemical and the Electrical Synapses



Motivation of Joint Vertex Classification on the *C.elegans* Neural Network

Statistical Motivation

The result of SGM on the *C.elegans* Neural Network suggests that inference must proceed in the joint space.

Neurological Motivation

We intend to understand the significance of the coexistence of the chemical and electrical connections.

Joint Classification on the Pair of Neural Networks

Algorithm 1 Joint Vertex Classification [2]

Goal: Classify the neuron v in G_1 whose neuron type is Y.

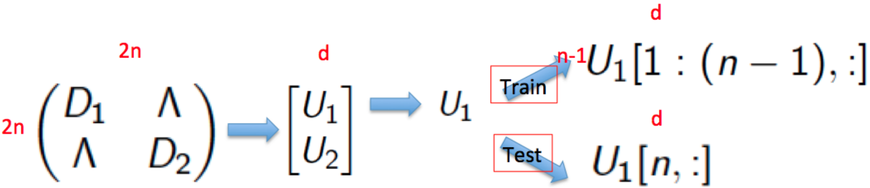
Input: A pair of the neural networks, $\{G_1, G_2\}$. A specified dissimilarity measure D.

1. **Compute the dissimilarities of** G_1 **and** G_2 **using** D
2. **Compute the Omnibus matrix** M

$$M = \begin{pmatrix} D_1 & \Lambda \\ \Lambda & D_2 \end{pmatrix} \in \mathbb{R}^{2n \times 2n}. \tag{2}$$

3. **Embed** The ominibus matrix M into d-dimensional Euclidean space using classical multidimensional scaling. $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \in \mathbb{R}^{2n \times d}$. $U_1 \in \mathbb{R}^{n \times d}$ is the joint embedding corresponding to G_1, and $U_2 \in \mathbb{R}^{n \times d}$ to G_2.
4. **Train** on $\mathcal{T}_{n-1} = U_1[1 : (n-1), :] \in \mathbb{R}^{(n-1) \times d}$ and **classify** v

Algorithm 1 Flow Chart



Joint Classification on the Pair of Neural Networks

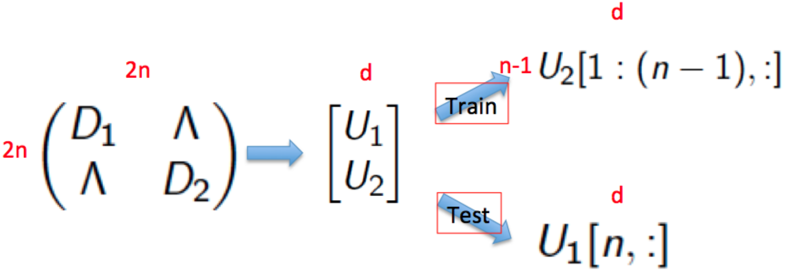
Algorithm 2 Joint Vertex Classification [2]

Goal: Classify the neuron v in G_1 whose neuron type is Y.
Input: A pair of the neural networks, $\{G_1, G_2\}$.
1. **Compute the dissimilarities of G_1 and G_2.**
2. **Compute the Omnibus matrix M**

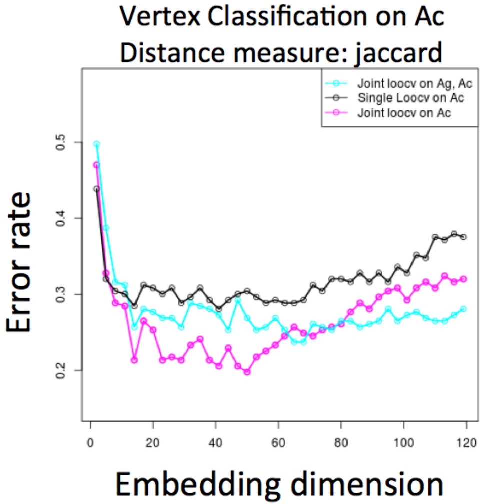
$$M = \begin{pmatrix} D_1 & \Lambda \\ \Lambda & D_2 \end{pmatrix} \in \mathbb{R}^{2n \times 2n}. \qquad (3)$$

3. **Embed** The ominibus matrix M into d-dimensional Euclidean space using classical multidimensional scaling. $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \in \mathbb{R}^{2n \times d}$. $U_1 \in \mathbb{R}^{n \times d}$ is the joint embedding corresponding to G_1, and $U_2 \in \mathbb{R}^{n \times d}$ to G_2.
4. **Train** on $\mathcal{T}_{n-1} = U_2[1 : (n-1), :] \in \mathbb{R}^{(n-1) \times d}$ and **classify** v

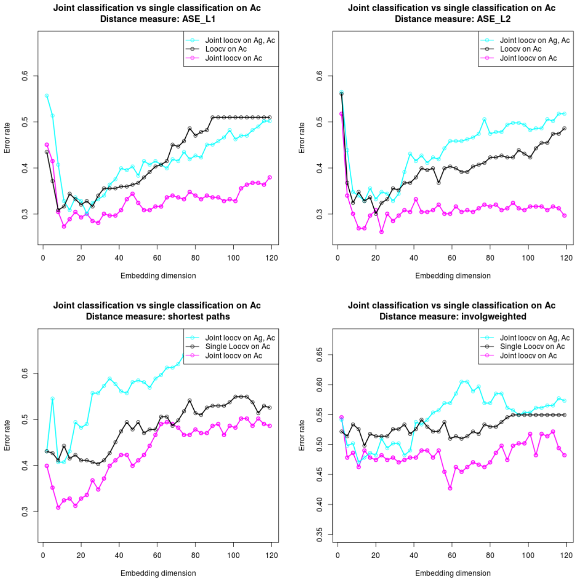
Algorithm 2 Flow Chart



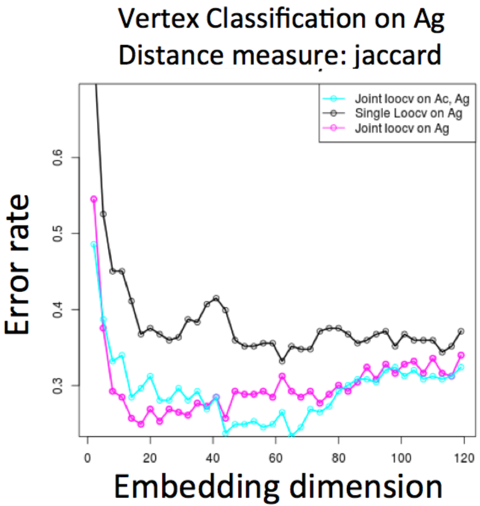
Compare the performance of joint vertex classification and separate vertex classification: Classification on A_c



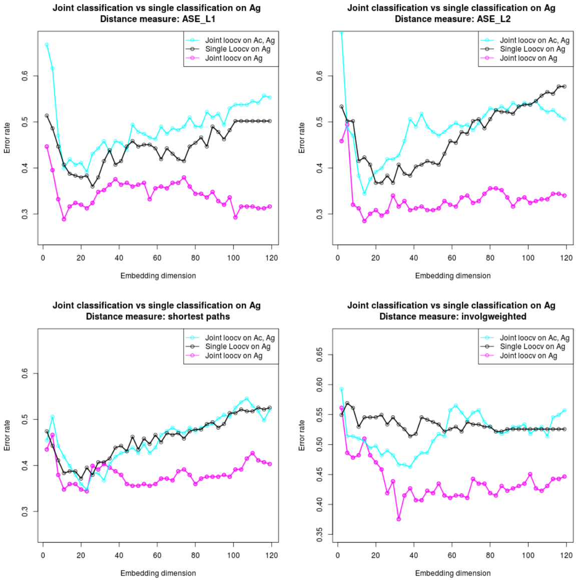
More distances



Compare the performance of joint vertex classification and separate vertex classification: Classification on A_g



More distances



Understanding the Coexistence of the Chemical and the Electrical Synapses

Implication of the Joint Classifier

- The classifier using the joint information from both networks performs better than the classifier using the information from the network separately.
- The improvement in classification indicates significance of the coexistence of the chemical and the electrical synapses.
- This discovery deserves further investigation in both the neuroscience and the statistics fields.

References

Li Chen, Joshua Vogelstein, and Carey Priebe.
Robust vertex classification.
arXiv preprint arXiv:1311.5954, 2013.

Li Chen, Joshua Vogelstein, and Carey Priebe.
A graph inference case study: the caenorhabditis elegans neural networks.
Working Manuscript, 2014.

Donniell Fishkind, Vince Lyzinski, Henry Pao, Li Chen, and Carey Priebe.
Vertex nomination schemes for membership prediction.
arXiv preprint arXiv:1312.2638, 2013.

Vince Lyzinski, Daniel Sussman, Minh Tang, Avanti Athreya, and Carey Priebe.
Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding.
arXiv preprint arXiv:1310.0532, 2013.

Daniel L Sussman, Minh Tang, Donniell E Fishkind, and Carey E Priebe.
A consistent adjacency spectral embedding for stochastic blockmodel graphs.
Journal of the American Statistical Association, 107(499):1119–1128, 2012.

Lav R Varshney, Beth L Chen, Eric Paniagua, David H Hall, and Dmitri B Chklovskii.
Structural properties of the caenorhabditis elegans neuronal network.
PLoS computational biology, 7(2):e1001066, 2011.

Thank you!