Law of Large Graphs

1 Introduction

OCP/HCP/FCP connectome setting
examples of $\bar{A}$ being used
Looking for better estimators of mean graph by making assumption that many vertices will behave similarly
[1] [3] [4] [5]
[8]

2 Model

2.1 MLE

Bernoulli says use ABar

2.2 RDPG

2.3 Model Assumption: SBM

2.4 ASE: PHat

dimension reduction, helps reduce variance, explanation that it is asymp unbiased.

2.5 Performance Evaluation: ARE

Defn and importance of asymp unbiased
[7] [9]

3 Results

3.1 Main Result

RE Eqn from var(PHat) eqn

3.2 Validation with simulated data

showing variance and RE simulations against expectations.

3.3 Performance on Real Data

cross-validation different dimensions of PHat vs. ABar
Perhaps permutation tests

4 Discussion

Given the popularity of large N connectome datasets that have known vertex correspondence, perhaps ASE should be chosen over ABar when estimating a group averaged graph. NMF may be even better estimate than ASE. [6]

5 Methods

5.1 Algorithm

5.2 Choosing Dimension

Zhu and Ghodsi

5.3 Dealing with Diagonal

Scheinerman Iterations

5.4 Dataset Description

CoRR

5.5 Source code and data

5.6 Proof of Var(PHat)

We are assuming that it should be condensed. Perhaps reference full proof uploaded to arxiv?

References

[1] Avanti Athreya, Vince Lyzinski, David J Marchette, Carey E Priebe, Daniel L Sussman, and Minh Tang. A central limit theorem for scaled eigenvectors of random dot product graphs arXiv : 1305 . 7388v2 [math . ST] 23 Dec 2013. pages 1–24, 2013.

[2] G. G. Brown and H. C. Rutemiller. Means and Variances of Stochastic Vector Products with Applications to Random Linear Models. *Management Science*, 24(2):210–216, 1977.

[3] Andressa Cerqueira, Daniel Fraiman, Claudia D Vargas, and Florencia Leonardi. A test of hypotheses for random graph distributions built from EEG data. pages 1–17, 2015.

[4] Donniell E Fishkind, Daniel L Sussman, Minh Tang, Joshua T Vogelstein, and Carey E Priebe. Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. 2012.

[5] Charles Freer. Functional Neuroimaging. *Journal of neurology, neurosurgery, and psychiatry*, 59:220, 2014.

[6] N D Ho. Nonnegative matrix factorization algorithms and applications. *Thesis*, (June):185, 2008.

[7] By Vince Lyzinski, Daniel L Sussman, Minh Tang, Avanti Athreya, and Carey E Priebe. PERFECT CLUSTERING FOR STOCHASTIC. pages 1–13, 2000.

[8] Roberto Imbuzeiro Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. page 46, 2009.

[9] Edward R. Scheinerman and Kimberly Tucker. Modeling graphs using dot product representations. *Computational Statistics*, 25:1–16, 2010.