# Successfully learning networks from undersampled neuroimaging data

D. Danks, S.M. Plis, C. Freeman, A. Caprihan, E. Damaraju, V. D. Calhoun

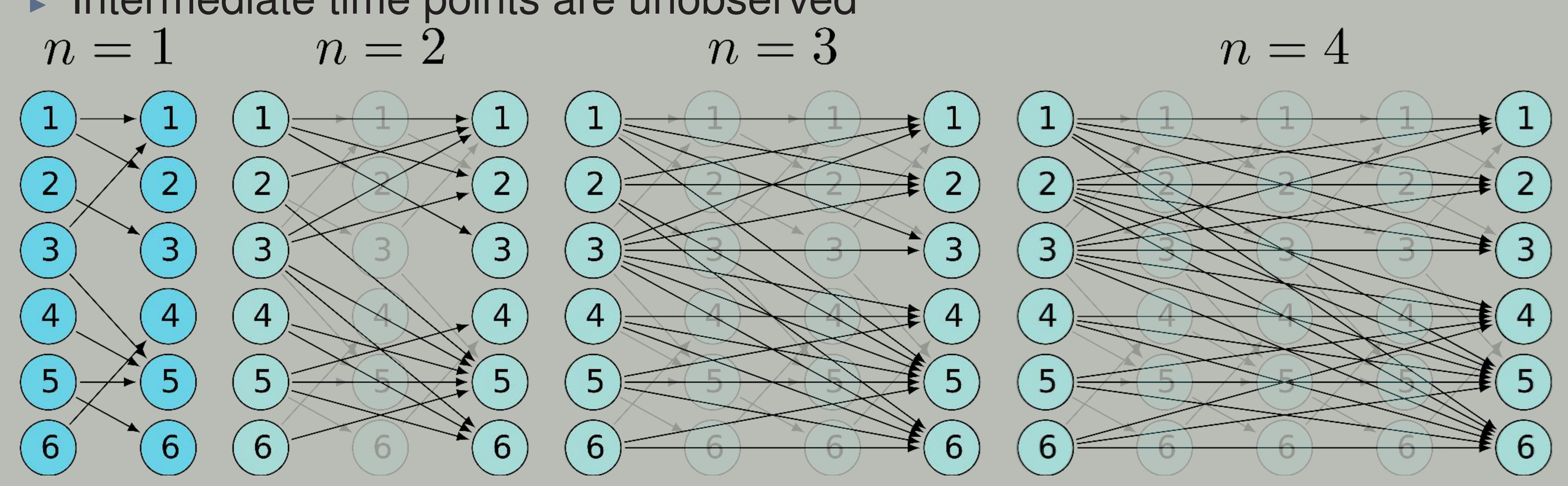
Carnegie Mellon University, The Mind Research Network


## Abstract

- Many structure learning algorithms are based on Granger causality
- Granger causality is unreliable given undersampled time series data
- We developed RASL (rate-agnostic structure learning) to learn structure from undersampled data
- In simulated data, RASL algorithms reveal causal timescale structure and improved measurement timescale learning
- RASL algorithms provide additional insight on fMRI data


## Problems with “Granger Causality”

- **Granger causality**: $\boldsymbol{X}$ Granger-causes $\boldsymbol{Y}$ $\equiv$ $\boldsymbol{X}$'s history provides information about $\boldsymbol{Y}$'s current state (beyond $\boldsymbol{Y}$'s history)
- Mathematically: $\boldsymbol{Y}^t = \sum_{i=1}^{k} \left[\alpha_i \boldsymbol{Y}^{t-i} + \beta_i \boldsymbol{X}^{t-i}\right]$ is a significantly better predictor of $\boldsymbol{Y}^t$ than $\boldsymbol{Y}^t = \sum_{i=1}^{k} \alpha_i \boldsymbol{Y}^{t-i}$ (perhaps with covariates)
- Granger causality only reliable if key assumptions hold:
  - *Linearity* (but hemodynamic convolution does not create problems)
  - *Causal sufficiency* (but becoming less of a problem)
  - *Equal timescales* for both measurement and underlying causation
- **Undersampling**: Measurement timescale significantly slower than causal or communication timescale
  - Intermediate time points are unobserved



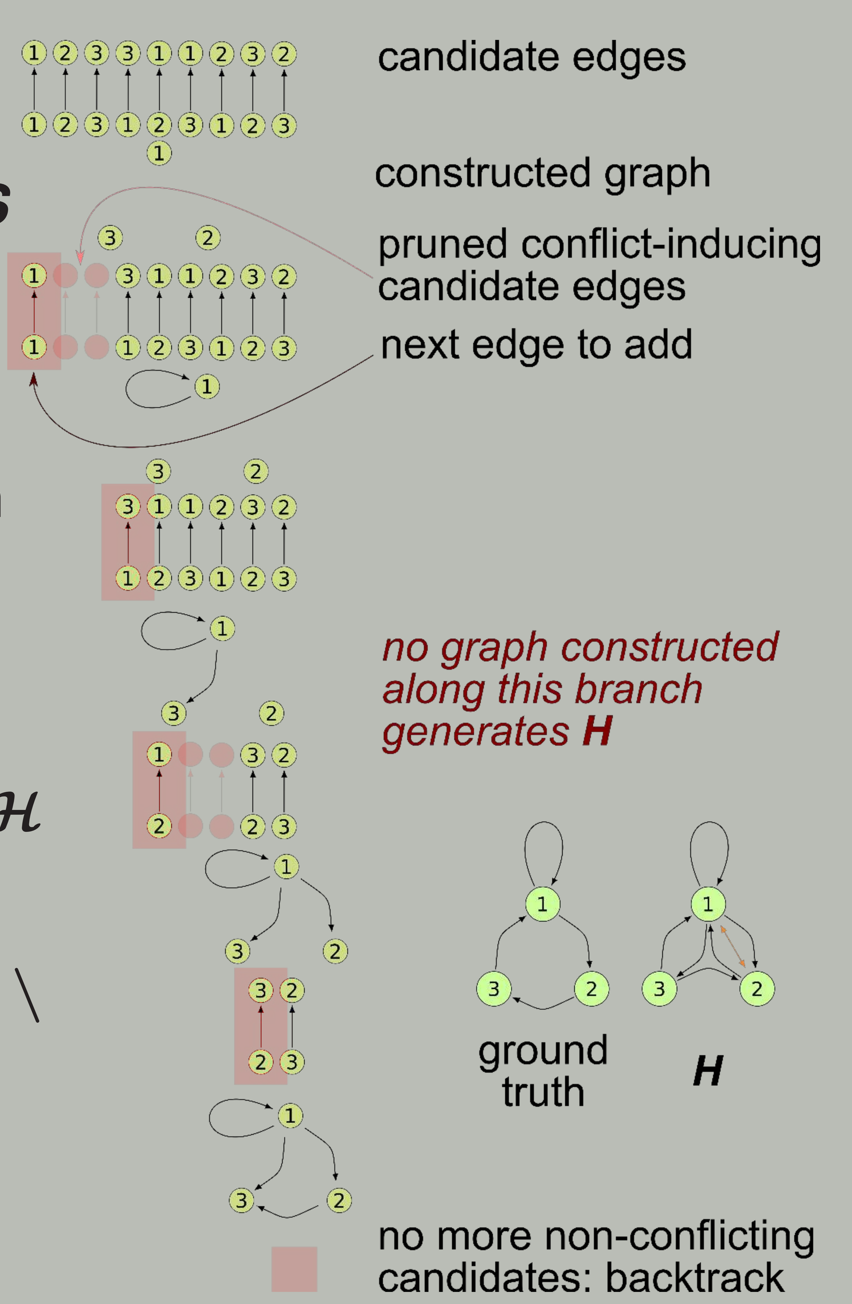

- Granger causality can be arbitrarily wrong given undersampling
  - $\boldsymbol{X}$ GC $\boldsymbol{Y}$ even though $\boldsymbol{Y}$ actually causes $\boldsymbol{X}$
  - $\boldsymbol{X}$ GC $\boldsymbol{Y}$ even though no direct causal connection
  - $\boldsymbol{X}$ doesn't GC $\boldsymbol{Y}$ even though $\boldsymbol{X}$ actually causes $\boldsymbol{Y}$
- Undersampling is a ubiquitous, persistent feature of fMRI data
- **Conclusion**: Structure learning algorithms based on Granger causality are likely unreliable given fMRI data

## How to Overcome the Problems

**Algorithm** *RecursiveEqClass*
**Input**: $\mathcal{H}$
**Output**: $[\![\mathcal{H}]\!]$
initialize empty graph $\mathcal{G}$ and set $\boldsymbol{S}$

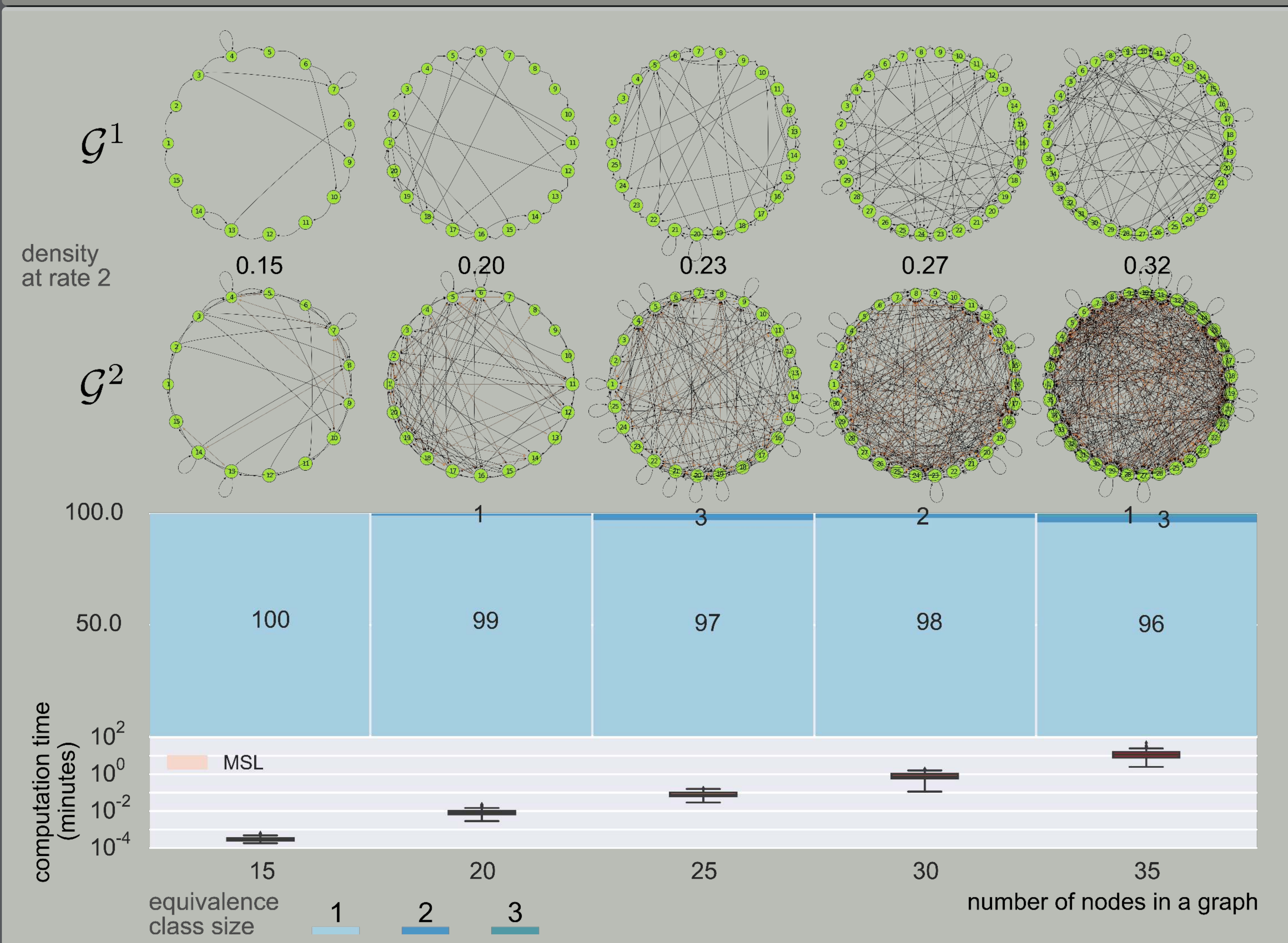
**begin** ***EdgeAdder*** $\mathcal{G}^*, \mathcal{H}, \boldsymbol{L}$
  **if** $\boldsymbol{L}$ *has elements* **then**
    **forall the** *edges in* $\boldsymbol{L}$ **do**
      **if** *edge creates a conflict* **then**
        remove it from $\boldsymbol{L}$
    **if** $\boldsymbol{L}$ *has elements* **then**
      **forall the** *edges in* $\boldsymbol{L}$ **do**
        add the edge to $\mathcal{G}^*$
        **if** $\exists \mathcal{G} \in \{(\mathcal{G}^*)^u\}$ *s.t.* $\mathcal{G} = \mathcal{H}$ **then**
          add $\mathcal{G}^*$ to $\boldsymbol{S}$
        ***EdgeAdder*** $\mathcal{G}^*, \mathcal{H}, \boldsymbol{L}$ $\setminus$ the edge
        remove the edge from $\mathcal{G}^*$

put all $\boldsymbol{n}^2$ edges into list $\boldsymbol{L}$
***EdgeAdder***$(\mathcal{G}, \mathcal{H}, \boldsymbol{L})$
**return** $\boldsymbol{S}$

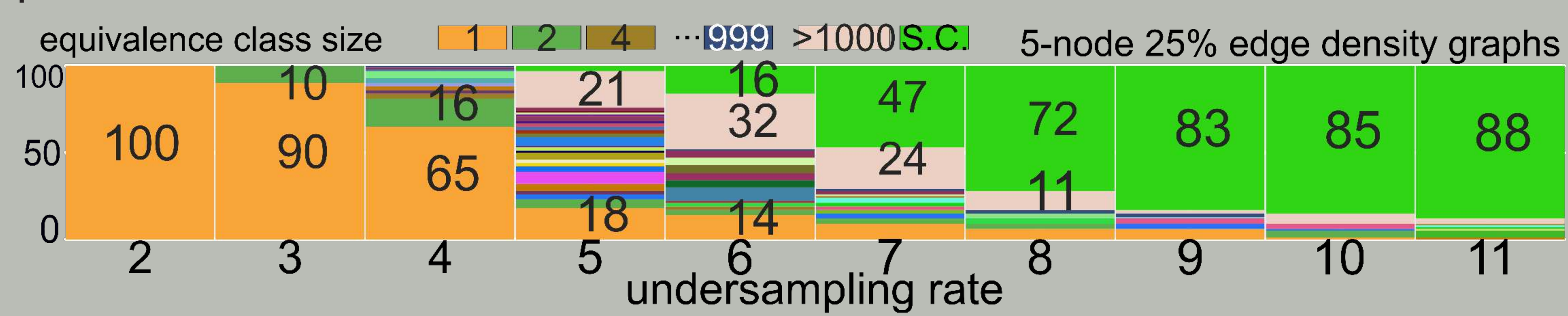



Start with an empty graph and $\boldsymbol{n}^2$ possible edges. For each edge $\boldsymbol{e}$, construct a graph $\mathcal{G}$ containing only $\boldsymbol{e}$. If $\mathcal{G}^u \not\subseteq \mathcal{H}$ for all $\boldsymbol{u}$, then reject; else if $\mathcal{G}^u = \mathcal{H}$ for some $\boldsymbol{u}$, then add $\mathcal{G}$ to $[\![\mathcal{H}]\!]$; else, recurse into non-conflicting graphs in order.

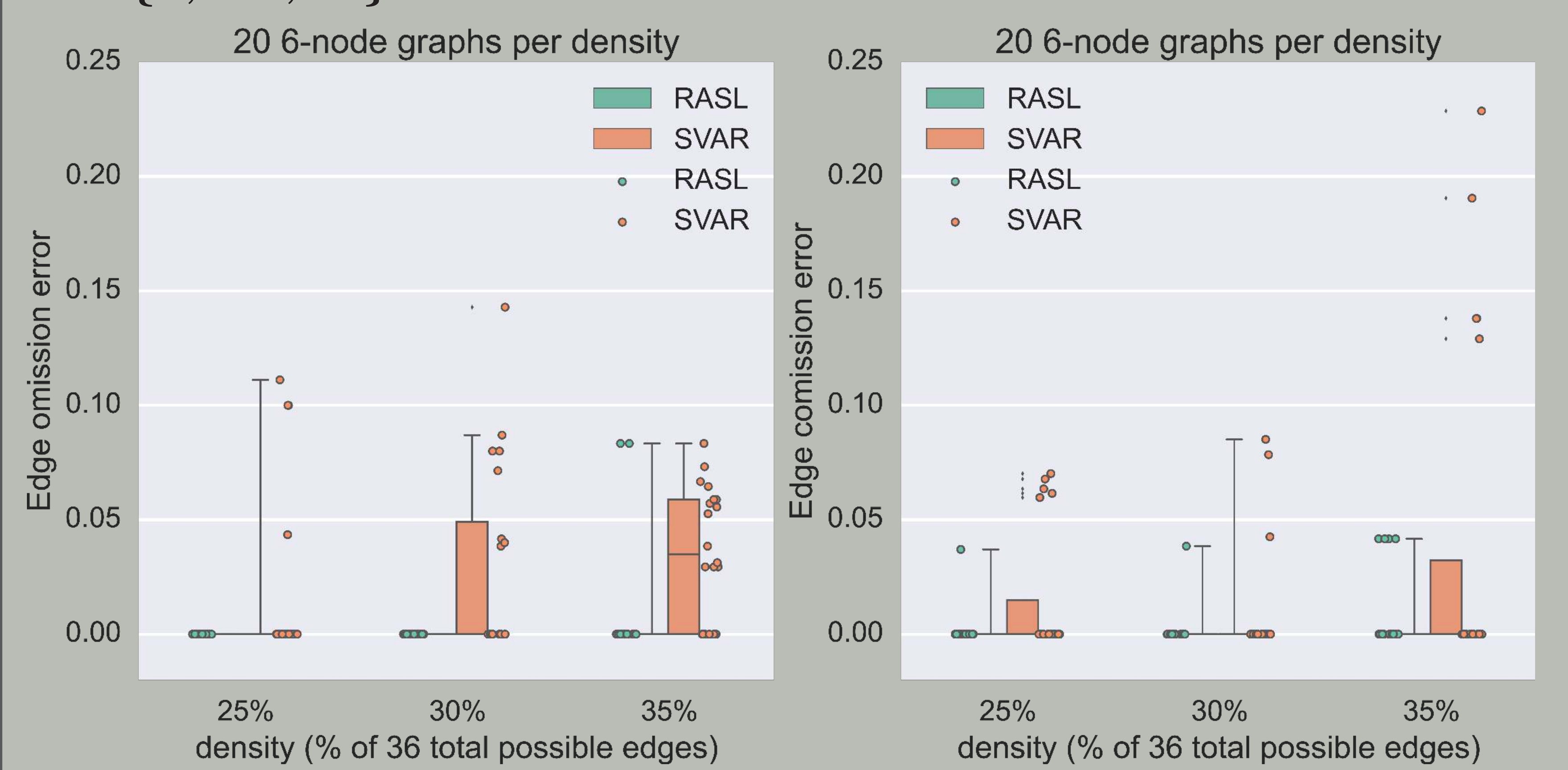
## Synthetic Data




Graphs with **15**, **20** ,**25**, **30**, and **35** nodes at the density of **10%**, their corresponding $\mathcal{G}^2$ and equivalence class size distribution as well as the running time summarizing the computation of **100** random SCCs per node size




The above plots the size of $[\![\mathcal{H}]\!]$ as a function of $\boldsymbol{u}$ for **100** random **5**-node SCCs with **25%** density, each of which was undersampled for $\boldsymbol{u} \in \{\boldsymbol{2}, \ldots, \boldsymbol{11}\}$.




The estimation and search errors on synthetic data undersampled at rate **2**.

## Conclusions

We have shown that undersampling leads to incorrect estimation of causal interactions. We have solved the forward problem and provided methods to generate undersampled graph from the ground truth. More importantly, we have solved the inverse problem and provided efficient algorithms both for known and unknown undersampling rates. We discovered that very often equivalence classes of estimated graphs often are singletons - a useful property for post-hoc results analysis in practical applications. We have evidence that our method can further be used to correct statistical estimation errors in standard within time-scale algorithms.

## References


[1] Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. In *Econometrica*, 37(3), 424-438,1969.

[2] Plis S.M., Danks D., Yang J. Mesochronal structure learning In *Proceedings of the Thirty-First Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-15)*, 2015.


http://www.mrn.org NIH R01EB005846 & NSF IIS-1318759 and NSF IIS-1318815 & NIH U54HG008540 http://www.lrri.org