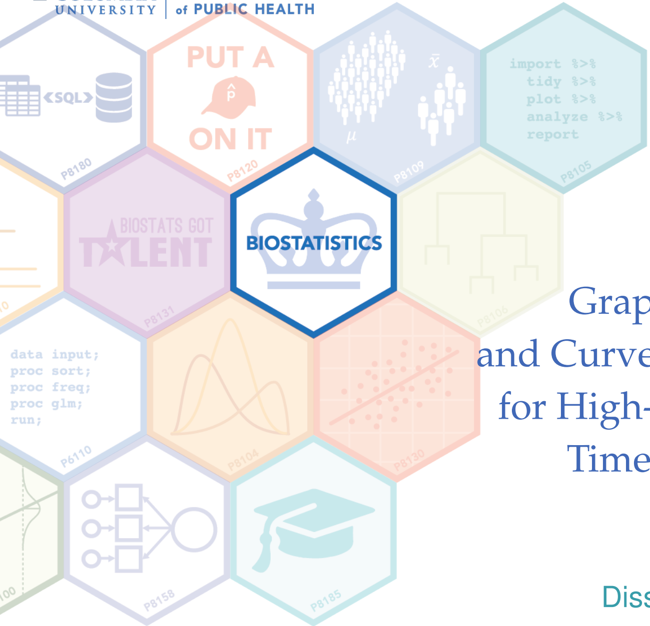
COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH



Dynamic Graphical Models and Curve Registration for High-Dimensional Time Course Data

Erin McDonnell

Dissertation Defense

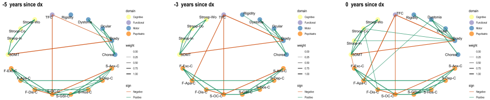
July 1, 2021

Dissertation themes

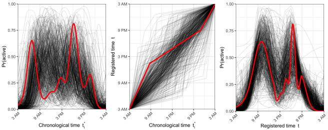
- Data sources in medical research are growing/diversifying.
- More data sources = more nuanced stratification of patients for precision medicine.
- One important layer of data: Longitudinal markers.
- This dissertation: Exploring subgroups' temporal patterns in a wide array of data:
  - Clinical evaluations
  - Repeated imaging
  - Wearable device data
- Motivating examples in neurology and chronobiology.

Dissertation themes

- Undirected graphical models:
  1. Clinical/imaging markers of Huntington’s disease.
  2. Alzheimer’s disease vs. Lewy body dementia.



- Functional data analysis / curve registration:
  3. Circadian rhythms phenotypes from accelerometer data.



- All involve high-dimensional, correlated temporal data.
  - Penalization: l_1 (Lasso), l_2 (Ridge), l_0 (Best subset).
  - (Functional) Principal Components Analysis (FPCA).

Project 1. Smooth time-varying Gaussian graphical models[1]

[1] First prize, ASA Section on Statistics in Imaging 2021 Student Paper Award.

Huntington's disease (HD)

- Fatal, hereditary, neurodegenerative disorder.
- Caused by a CAG repeat expansion in the HTT gene.
- Average age of onset: 36-45 years.
- Symptoms: Motor, functional, cognitive, and psychiatric.
- Manifest HD (phenoconversion) is usually diagnosed from motor symptoms alone.
- Changes in non-motor symptoms occur years prior.

PREDICT-HD study[2]

instead of using year as the time variable, using the year to diagnosis as time variable reduces heterogeneity

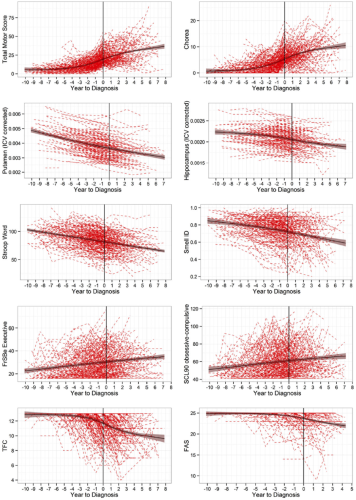
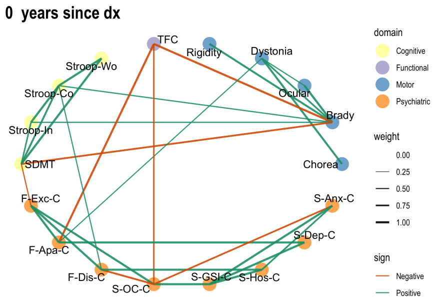


Figure 3. Trajectories of motor, imaging, cognitive, functional, and psychiatric variables for *N*=225 converters

[2] Jane S Paulsen et al. In: *Lancet Neurology* 13.12 (2014), pp. 1193–1201.

Our goal

- Cross-domain symptoms may be connected through a partial correlation matrix.
- These partial correlations may change over time.
- We want to understand this changing network of symptom associations leading up to HD diagnosis.



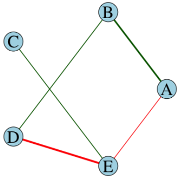
1. UG may leads to DAG
2. Some nodes with high connectivity

Overview: Gaussian graphical models

- Let $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$.
- The **precision matrix** $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ reveals associations:
  - $\theta_{jk} \neq 0$ implies non-zero partial correlation, $\rho_{jk} \neq 0$.
  - $\theta_{jk} = 0$ implies conditional independence, $\rho_{jk} = 0$.
- Connect nodes on a graph with **edges** weighted by ρ_{jk}.

Partial correlation matrix

	A	B	C	D	E
A	1	0.4	0	0	-0.05
B	0.4	1	0	0.1	0
C	0	0	1	0	0.1
D	0	0.1	0	1	-0.5
E	-0.05	0	0.1	-0.5	1



- Gaussian graphical models estimate $\boldsymbol{\Theta}$ via **regularization.**

partial correlation gives a sparser graph than marginal corelation

Estimating Gaussian graphical models

1. **Global approach**[3] (graphical lasso[4])
   - l_1-**penalized** full multivariate Gaussian log-likelihood:

$$\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} \left\{ \log\det \mathbf{\Theta} - \mathrm{trace}(\mathbf{\Theta} \boldsymbol{S}) - \lambda ||\mathbf{\Theta}||_1 \right\},$$

2. **Neighborhood selection**[5]

Recall the definition of partial correlation

   - Fit p **lasso models**, regressing each node on the rest.
   - For the j^{th} model, we estimate:

$$\hat{\boldsymbol{\beta}}_j = \arg\min_{\beta_{jk:k\neq j}} \frac{1}{n} \sum_{i=1}^{n} \left(x_{ij} - \sum_{k\neq j}^{p} x_{ik}\beta_{jk} \right)^2 + \lambda \sum_{k\neq j}^{p} |\beta_{jk}|,$$

   - Derive $\hat{\mathbf{\Theta}}$ from the coefficients of all p models.

[3] Ming Yuan and Yi Lin. In: *Biometrika* 94.1 (2007), pp. 19–35.

[4] Jerome Friedman et al. In: *Biostatistics* 9.3 (2008), pp. 432–441.

[5] Nicolai Meinshausen and Peter Bühlmann. In: *Annals of Statistics* 34.3 (2006), pp. 1436–1462.

Dynamic Gaussian Graphical Models

- Now let the nodes and their joint distribution vary with time:

$$x_i(t) = \left[x_{i1}(t), \ldots, x_{ip}(t)\right]^T \sim MVN\left(\mathbf{0}, \mathbf{\Sigma}(t)\right).$$

- Extensions to graphical lasso and neighborhood selection have been proposed to estimate $\mathbf{\Theta}(t) = \mathbf{\Sigma}^{-1}(t)$.

Extension type	Estimation	Approach	Likelihood/Pseuo-likelihood	Penalty
Kernel weights	Neighborhood	KELLER	$\sum_{l=1}^{M} K_h(t_l - t)\left(x_{lj} - \sum_{k=1}^{p} x_{lk}\beta_k\right)^2$	$\lambda \sum_{k=1}^{p} \lvert\beta_k\rvert$
Kernel weights	Graphical lasso	Zhou 2010	$\text{trace}\left(\mathbf{\Theta}\hat{S}(t)\right) - \log\det\mathbf{\Theta}$	$\lambda\lvert\lvert\mathbf{\Theta}\rvert\rvert_1$
Fused lasso	Neighborhood	TESLA	$\sum_{l=1}^{M}\left(x_{lj} - \sum_{k=1}^{p} x_{lk}\beta_k(t_l)\right)^2$	$2\lambda_1 \sum_{l=1}^{M}\sum_{k=1}^{p}\lvert\beta_k(t_l)\rvert + 2\lambda_2\sum_{k=1}^{p}\sum_{l=2}^{M}\lvert\beta_k(t_l) - \beta_k(t_{l-1})\rvert$
Group fused lasso	Neighborhood	TD-lasso	$\sum_{l=1}^{M}\left(x_{lj} - \sum_{k=1}^{p} x_{lk}\beta_k(t_l)\right)^2$	$2\lambda_1 \sum_{l=1}^{M}\sum_{k=1}^{p}\lvert\beta_k(t_l)\rvert + 2\lambda_2\sum_{l=2}^{M}\sqrt{\sum_{k=1}^{p}\left(\beta_k(t_l) - \beta_k(t_{l-1})\right)^2}$
Kernel + Group fused	Graphical lasso	GFGL	$\sum_{l=1}^{M}\left[\text{trace}\left(\mathbf{\Theta}(t_l)\hat{S}(t_l)\right) - \log\det\mathbf{\Theta}(t_l)\right]$	$\lambda_1\sum_{l=1}^{M}\lvert\lvert\mathbf{\Theta}(t_l)\rvert\rvert_1 + \lambda_2\sum_{l=2}^{M}\sqrt{\sum_{j,k}^{p}\left(\theta_{jk}(t_l) - \theta_{jk}(t_l)\right)^2}$
Kernel + Fused	Graphical lasso	SINGLE[6]	$\sum_{l=1}^{M}\left[\text{trace}\left(\mathbf{\Theta}(t_l)\hat{S}(t_l)\right) - \log\det\mathbf{\Theta}(t_l)\right]$	$\lambda_1\sum_{l=1}^{M}\lvert\lvert\mathbf{\Theta}(t_l)\rvert\rvert_1 + \lambda_2\sum_{l=2}^{M}\lvert\lvert\mathbf{\Theta}(t_l) - \mathbf{\Theta}(t_{l-1})\rvert\rvert_1$
Group lasso	Both	LOGGLE[7]	$\frac{1}{\sqrt{\lvert\mathcal{N}_{l,d}\rvert}}\sum_{q\in\mathcal{N}_{l,d}}\left[\text{trace}\left(\mathbf{\Theta}(t_q)\hat{S}(t_q)\right) - \log\det\mathbf{\Theta}(t_q)\right]$	$\lambda\sum_{j\neq k}\sqrt{\sum_{q\in\mathcal{N}_{l,d}}\theta_{jk}(t_q)^2}$

(Table layout inspired by[8])

[6] Ricardo Pio Monti et al. In: *NeuroImage* 103 (2014), pp. 427–443.

[7] Jilei Yang and Jie Peng. In: *Journal of Computational and Graphical Statistics* 021707 (2019).

[8] Alexander J. Gibberd and James D.B. Nelson. In: *Journal of Computational and Graphical Statistics* 26.3 (2017), pp. 623–634.

Gaps to be filled

- We want to anchor our time scale on diagnosis time.
- We need to accommodate multiple observations per time point and random visit times.
- l_1 penalization produces biased estimates.
- l_0 penalization outperforms l_1 penalization (but non-convex).

Proposed method

An l_0, l_1, and l_2 penalized dynamic Gaussian graphical model:

Proposed feature	Contribution
Revival process	Realign data on diagnosis time.
Kernel weights	Borrow from neighboring observations
Adaptive lasso	Reduce bias; borrow from time-adjacent estimates.
Novel "Elastic Fuse L_0" penalty	Hard-threshold and smooth edges across time.

$$\hat{\boldsymbol{\beta}}_j(s) = \underset{\beta_{jk:k\neq j}}{\arg\min} \frac{1}{N} \sum_{i=1}^{n} \sum_{l=1}^{n_i} \left[K_h(s_{il} - s) \left(z_{ij}(s_{il}) - \sum_{k\neq j}^{p} z_{ik}(s_{il}) \beta_{jk} \right)^2 \right] +$$

kernel weight, distance between two time point

$$\lambda_1 \sum_{k\neq j}^{p} w_{jk} |\beta_{jk}| +$$

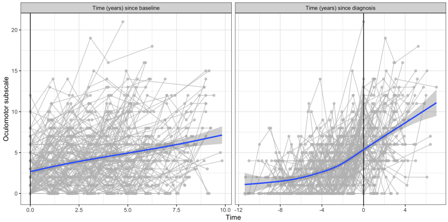
lasso bias comes from much shrinkage on large coefficients

$$\rho \sum_{k\neq j}^{p} I(\beta_{jk} \neq 0) + \lambda_2 \sum_{k\neq j}^{p} (\beta_{jk} - C)^2,$$

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Revival process

- Anchor time on a critical landmark event (e.g. diagnosis).[9]
- Let T_i = time from baseline to event for subject i.
  - Assuming everyone has an event.
- Let $\{t_{i1}, \ldots, t_{in_i}\}$ = visit times (from baseline) for subject i.
- Realign the data by defining: $z_{ij}(s_{il}) = x_{ij}(t_{il})$.
- $s_{il} = t_{il} - T_i$ are observed **gap times** (years since event).



[9] Walter Dempsey and Peter McCullagh. In: *Lifetime Data Analysis* 24.4 (2018), pp. 550–584.

Kernel-weighted neighborhood selection

- Goal: Estimate network at $s \in \{0, \pm 1, \pm 2\}$ years since dx.
- Add kernel weights to neighborhood selection to borrow information from neighboring observations.[10,11,12]

$$\hat{\boldsymbol{\beta}}_j(s) = \underset{\beta_{jk:k\neq j}}{\arg\min} \overbrace{\sum_{i=1}^{n}\sum_{l=1}^{n_i}\left[K_h(s_{il}-s)\Big(z_{ij}(s_{il}) - \sum_{k\neq j}^{p} z_{ik}(s_{il})\beta_{jk}\Big)^2\right]}^{\mathcal{L}(\boldsymbol{\beta}_j, \boldsymbol{Z}, s)} + \lambda \sum_{k\neq j}^{p} |\beta_{jk}|,$$

where $K_h(s_{il} - s) \propto \exp\left\{ -\frac{1}{2}\left(\frac{s_{il}-s}{h}\right)^2\right\}$ with bandwidth h.

---

[10] Donald R. Hoover et al. In: *Biometrika* 85.4 (1998), pp. 809–822.

[11] Le Song et al. In: *Bioinformatics* 25.12 (2009), pp. 128–136.

[12] Mladen Kolar and Eric P. Xing. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (2011).

Adaptive lasso to encourage temporal smoothness

- Coefficient-specific weights distribute the l_1 penalty better.[13]
- Weights are constructed from a consistent estimator of $\boldsymbol{\beta}_j$.
- Proposed method: Construct weights at a **previously-estimated gap time** $q_s =$, e.g. $q_s = s - 1$.
- First estimate **ridge regression** coefficients at time q_s:

$$\tilde{\boldsymbol{\beta}}_j(q_s) = \underset{\beta_{jk:k\neq j}}{\arg\min}\, \mathcal{L}(\boldsymbol{\beta}_j, \mathbf{Z}, q_s) + \gamma \sum_{k\neq j}^{p} \beta_{jk}^2,$$

- Then plug in adaptive lasso weights $w_{jk}(q_s) = \frac{1}{|\tilde{\beta}_{jk}(q_s)|}$:

$$\hat{\boldsymbol{\beta}}_j(s) = \underset{\beta_{jk:k\neq j}}{\arg\min}\, \mathcal{L}(\boldsymbol{\beta}_j, \mathbf{Z}, s) + \lambda \sum_{k\neq j}^{p} w_{jk}(q_s)|\beta_{jk}|.$$

[13] Hui Zou. In: *Journal of the American Statistical Association* 101.476 (2006), pp. 1418–1429.

Elastic Fuse L_0 for sparsity and temporal smoothness

- Swap l_1 with **APM-L_0**: l_0 norm on a **surrogate parameter** ψ, and bound difference by a smooth convex function.[14]
- Add l_2 norm to also pull ψ toward previous estimate:

$$\{\hat{\boldsymbol{\beta}}_j(s), \hat{\boldsymbol{\psi}}_j(s)\} = \underset{\beta_{jk}, \psi_{jk:k\neq j}}{\arg\min} \mathcal{L}(\boldsymbol{\beta}_j, \mathbf{Z}, s) + \lambda_1 \sum_{k\neq j}^{p} w_{jk}(q_s)|\beta_{jk} - \psi_{jk}| + \rho \sum_{k\neq j}^{p} I(\psi_{jk} \neq 0) + \lambda_2 \sum_{k\neq j}^{p} \left(\psi_{jk} - C_{jk}(q_s)\right)^2,$$

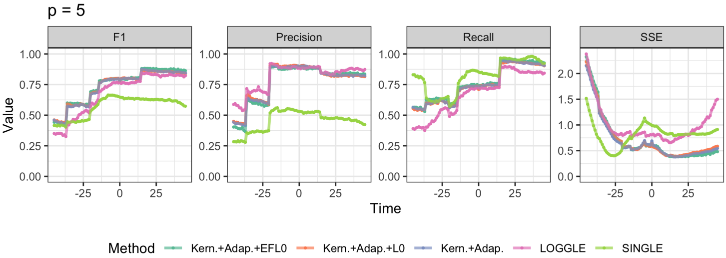
- where $C_{jk}(q_s) = \hat{\beta}_{jk}(q_s)$ is treated as a constant.
- The EF-L_0 penalty encourages temporal smoothness in network strength and structure.

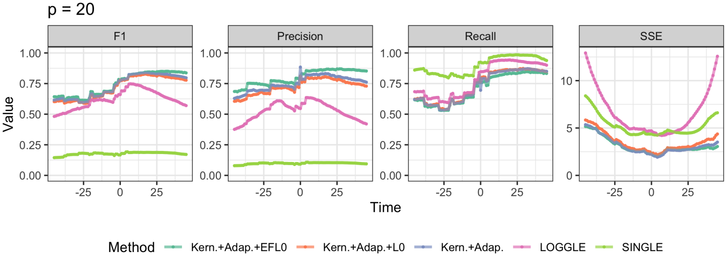
[14] Xiang Li et al. In: *Statistics in Medicine* 37.3 (2018), pp. 473–486.

Simulations

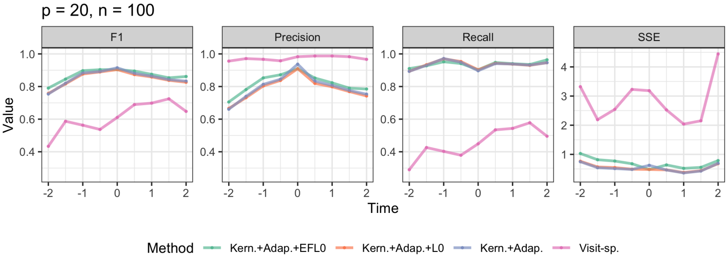
Metric	Definition	Comments
F1 score	$2\frac{(Precision * Recall)}{Precision + Recall}$	Harmonic mean of Precision and Recall
Precision	$\frac{TP}{TP+FP}$	Positive predictive value
Recall	$\frac{TP}{TP+FN}$	Sensitivity
SSE	$\sum_{j=1}^{p} \sum_{k<j} (\hat{\theta}_{jk} - \theta_{jk})^2$	Sum of squared error in edge estimates

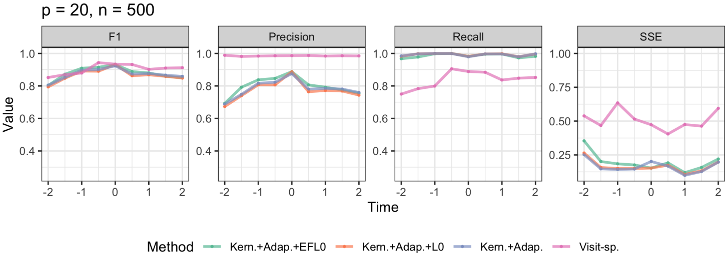
Simulation setting 1. $N = 1$, evenly spaced time points





Simulation setting 1. $N > 1$, random time points





Application to PREDICT-HD

We tried two ways to define HD diagnosis:

1. Diagnostic confidence level (DCL) of 4, estimated with a Hidden Markov Model.[15]
   - $N = 197$ in clinical model.
   - Gap times of interest: $s =$ annually from -5 to 3 years.
2. CAG Age Product (CAP)[16] of 100:

$$CAP = 100 * Age * \frac{CAG - 30}{627}.$$

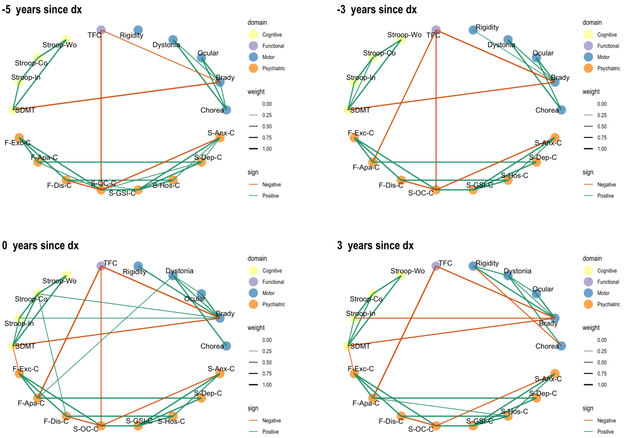
   - Observed diagnosis not required.
   - Increases N AND allows adjustment for age, CAG repeats.
   - $N = 882$ in imaging model.
   - Gap times of interest: $s =$ from CAP 50 to 130, by 10.

[15] Dawei Liu et al. In: *Journal of Neurology* 262.12 (2015), pp. 2691–2698.

[16] Christopher A. Ross et al. In: *Nature Reviews Neurology* 10.4 (2014), pp. 204–216.

Application to PREDICT-HD

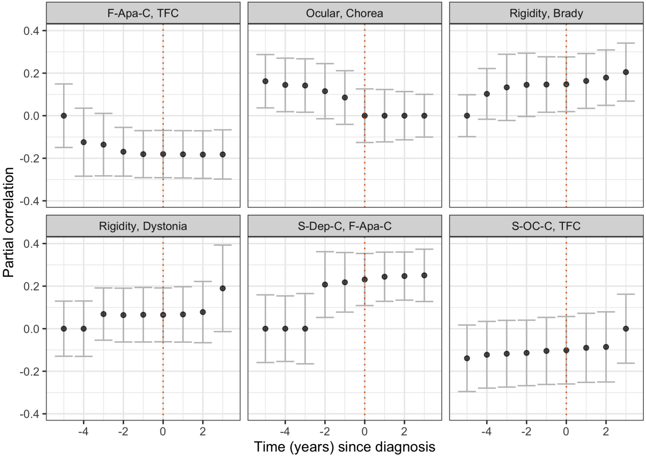
Clinical symptom network, using DCL to define diagnosis



1. Far from diagnoses, less connections2. More connectic

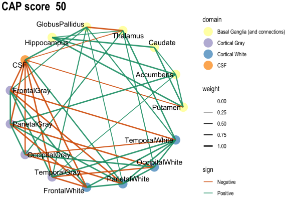
Application to PREDICT-HD

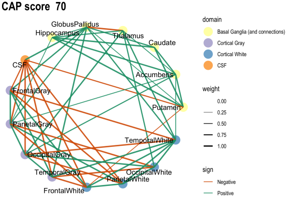
Clinical symptom network, using DCL to define diagnosis



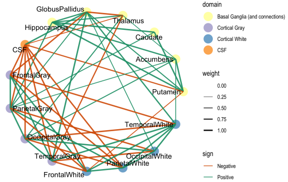
Application to PREDICT-HD

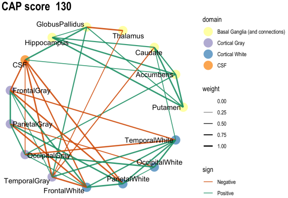
Imaging marker network, using CAP to define diagnosis





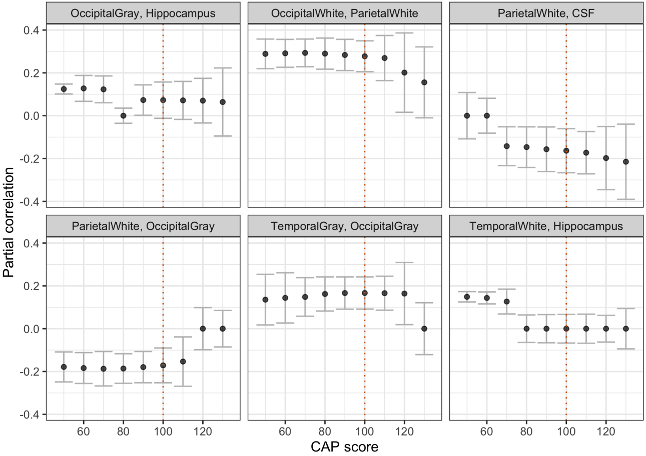
Similar to diagnosis time





Application to PREDICT-HD

Imaging marker network, using CAP to define diagnosis



Summary and future directions

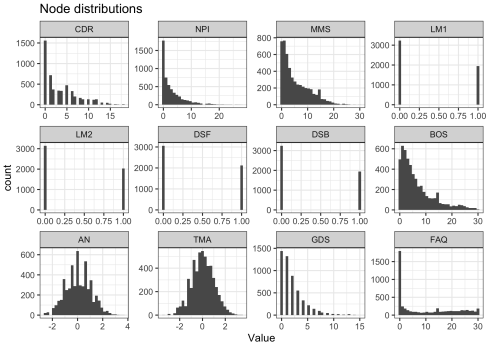
- Partial correlations between symptoms from different domains change over time.
- We observed a loss of connection in structural MRI imaging markers.
- Future work is needed to create a covariate-dependent time-varying Gaussian graphical model
  - Adjust for baseline age, gender, education, CAG repeat length, other genetic factors, medications.
  - Extend the work of Xie et al.[17]

[17] Shanghong Xie et al. In: *Biometrics* (2019).

Project 2. Smooth time-varying mixed graphical models for heterogeneous variable types

When normality is not met

- Gaussian graphical models assume $x_i \sim MVN(\mathbf{0}, \Sigma)$.
- In practice, x_i may be defined on several domains:
  - Continuous, count-valued, binary/categorical.
- Motivating example: Time-varying symptom networks in Alzheimer's disease (AD) vs. Lewy body dementia (LBD).



AD vs. LBD

- AD: Most common type of dementia.
  - Characterized by amyloid plaques, neurofibrillary tangles.
- LBD: Next leading type of dementia.
  - Protein deposits (Lewy bodies) develop in parts of the brain that control cognitive/motor function.
- Clinically similar; often misdiagnosed.
- Confirmatory diagnosis only possible post-mortem.
- We explore differences in time-varying symptom networks of **neuropathically-defined** AD and LBD.

Mixed graphical models[18]

- Recall exponential family form:

$$P(X = x|\boldsymbol{\theta}) = \exp\left\{\boldsymbol{\theta}\phi(x) + B(x) - \Phi(\boldsymbol{\theta})\right\},$$

  where:
  - θ is a natural parameter,
  - $\phi(x)$ is a vector of sufficient statistics,
  - $B(x)$ is a base measure,
  - $\Phi(\boldsymbol{\theta})$ is a log-normalizing constant.
- Let $\{X_1, ..., X_p\}$ follow (possibly different) exponential family distributions.
- Conditional distribution of X_j given all other nodes:

$$P(X_j|\boldsymbol{X}_{V\setminus j}) = \exp\left\{E_j(\boldsymbol{X}_{\setminus j})\boldsymbol{\phi}_j(X_j) + B_j(X_j) - \Phi_j(\boldsymbol{X}_{V\setminus j})\right\},$$

[18] Eunho Yang et al. In: *Proceedings of Machine Learning Research*. Vol. 33. 2014, pp. 1042–1050.

Mixed graphical models[19]

- For simplicity we assume:
  1. Only Gaussian, Poisson, or Bernoulli distributions.
  2. Corresponding graph $G = (V, E)$ has pairwise edges only.
- Then the joint distribution is:

$$P(\boldsymbol{X}; \theta) = \exp\Big\{ \sum_{j \in V} \theta_j \phi_j(X_j) + \sum_{(j,k) \in E} \theta_{jk} \phi_j(X_j) \phi_k(X_k) + \sum_{j \in V} B_j(X_j) - \Phi(\boldsymbol{\theta}) \Big\}.$$

- Markov factorization property holds (distribution respects conditional dependence statements made by the graph)
- Is the precision matrix $\boldsymbol{\Theta}$ graph-structured?

[19] Eunho Yang et al. In: *Proceedings of Machine Learning Research.* Vol. 33. 2014, pp. 1042–1050.

Estimating mixed graphical models

- The inverse **generalized** (augmented) covariance matrix is graph-structured.[20,21,22]
- Neighborhood selection algorithm: For the j^{th} node:
  1. Build design matrix with augmented interactions.
  2. Fit l_1-penalized GLM model for node X_j with appropriate link function, with tuning parameter λ.
  3. Hard-threshold at level τ, a multiple of $\sqrt{\frac{\log p}{n}}$.
- Combine edge weights $\hat{\theta}_{jk}$ and $\hat{\theta}_{kj}$ with AND or OR rule.
- Choose λ with Extended Bayesian Information Criterion.[23]

[20] Jonas M. B. Haslbeck and Lourens J. Waldorp. In: *arXiv* (2015).

[21] Jonas M. B. Haslbeck and Lourens J. Waldorp. In: *Journal of Statistical Software* 98.8 (2020), pp. 1–46.

[22] Po Ling Loh and Martin J. Wainwright. In: *Annals of Statistics* 41.6 (2013), pp. 3022–3049.

[23] Rina Foygel and Mathias Drton. In: *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems* (2010), pp. 1–9.

Time-varying mixed graphical models[25]

- Kernel weight for observed time t and estimation time t_q^e:

$$K_h(t - t_q^e) = \exp\left\{\frac{(t - t_q^e)^2}{2h^2}\right\}.$$

- Sum of weights = effective sample size $N_{eff}(t_q^e)$.
- We use Silverman's rule of thumb for bandwidth:[24]

$$h = 0.9 * \min\left\{\hat{\sigma}, \frac{I\hat{Q}R}{1.34}\right\} * n^{-1/5},$$

[24] B.W. Silverman. London: Chapman & Hall/CRC, 1986, p. 45.

[25] Jonas M. B. Haslbeck and Lourens J. Waldorp. In: *Journal of Statistical Software* 98.8 (2020), pp. 1–46.

Proposed method

- We use ideas from Project 1 to propose an **adaptive lasso** extension to the time-varying mixed graphical model.
- For all $q > 1$:

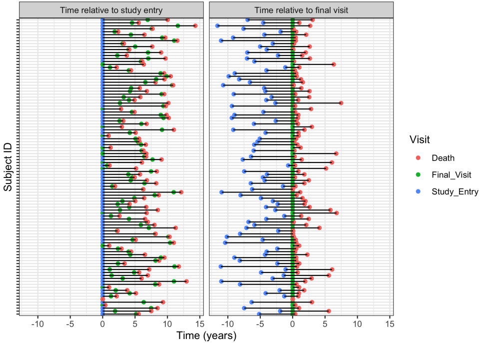
$$\hat{\boldsymbol{\theta}}_j(t_q^e) = \underset{\theta_{jk:k\neq j}}{\arg\min} \frac{-1}{N_{eff}(t_q^e)} \sum_{i=1}^{n} \sum_{l=1}^{n_i} K_h(t_{il} - t_q^e)\mathcal{L}\big(\boldsymbol{\theta}_j, \boldsymbol{X}_i(t_{il})\big) + \lambda \sum_{k\neq j}^{p^*} w_{jk}(t_{q-1}^e)|\theta_{jk}|,$$

- where $w_{jk}(t_{q-1}^e) = 1/|\tilde{\theta}_{jk}(t_{q-1}^e)|$, and:

$$\tilde{\boldsymbol{\theta}}_j(t_{q-1}^e) = \underset{\theta_{jk:k\neq j}}{\arg\min} \frac{-1}{N_{eff}(t_{q-1}^e)} \sum_{i=1}^{n} \sum_{l=1}^{n_i} K_h(t_{il} - t_{q-1}^e)\mathcal{L}\big(\boldsymbol{\theta}_j, \boldsymbol{X}_i(t_{il})\big).$$

Reverse time scale

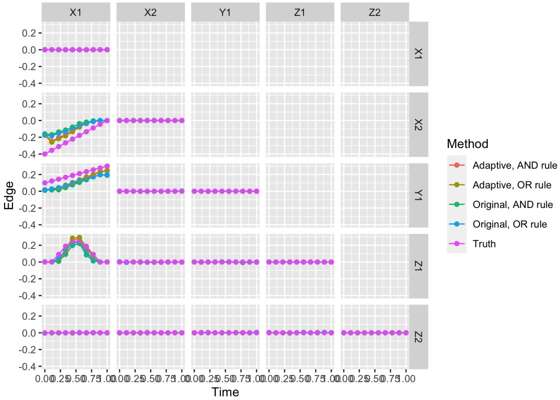
- To treat neuropathological diagnosis as a baseline covariate, model time in reverse (years before final visit).[26]
- Network estimation works backwards from final visit.



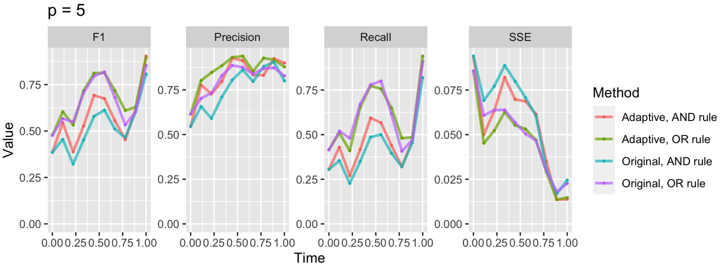
[26] Jing Qian et al. In: *JAMA Neurology* 74.5 (2017), pp. 540–548.

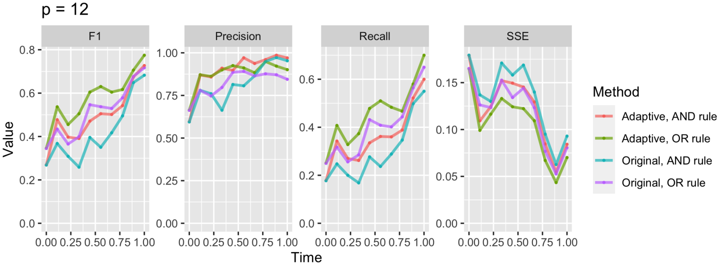
Simulation studies

- 100 simulations, with $n = 100$ subjects, and $n_i = 10$ visits.
- Poisson: $\{X_1, X_2\}$, Gaussian $\{Y_1\}$, Bernoulli $\{Z_1, Z_2\}$.



Simulation studies





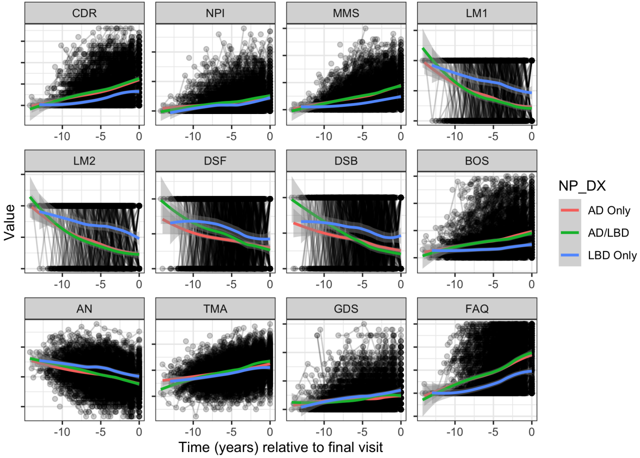
Application to NACC database

- National Alzheimer's Coordinating Center database.
- 3 subgroups: AD only, AD/LBD combo, and LBD only.
- Approximately annual evaluations.
- $p = 12$ nodes:

Node	Label for figures	Data type (transformed)
Clinical Dementia Rating Scale (CDR) Sum of Boxes	CDR	Poisson
Neuropsychiatric Inventory Questionnaire (NPI-Q) severity	NPI	Poisson
Mini Mental State Examination (MMSE)	MMS	Poisson
Logical Memory part IA (immediate recall)	LM1	Bernoulli (> 10 vs. ≤ 10)
Logical Memory part IIA (delayed recall)	LM2	Bernoulli (> 8 vs. ≤ 8)
Digit Span Forward length	DSF	Bernoulli (> 6 vs. ≤ 6)
Digit Span Forward length	DSB	Bernoulli (> 4 vs. ≤ 4)
Boston Naming Test total score	BOS	Poisson
Animal Naming verbal fluency test	AN	Gaussian
Trail Making part A	TMA	Gaussian
Geriatric Depression Scale (GDS)	GDS	Poisson
Functional Activities Questionnaire (FAQ)	FAQ	Poisson

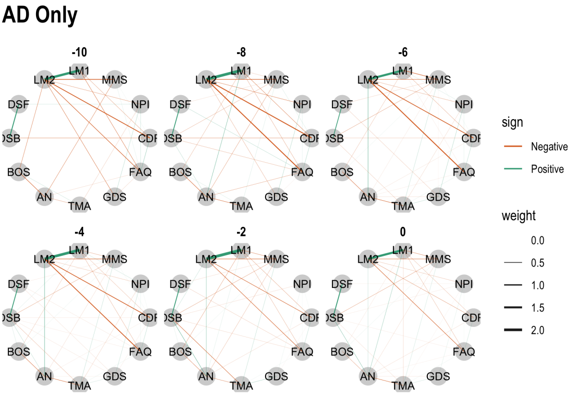
COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Results



Results

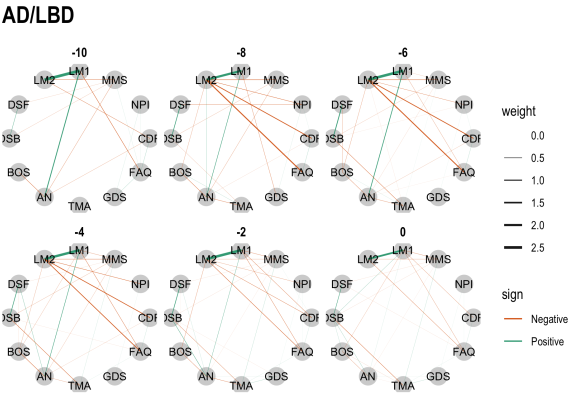
- $N = 1,052$ with AD only.



COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

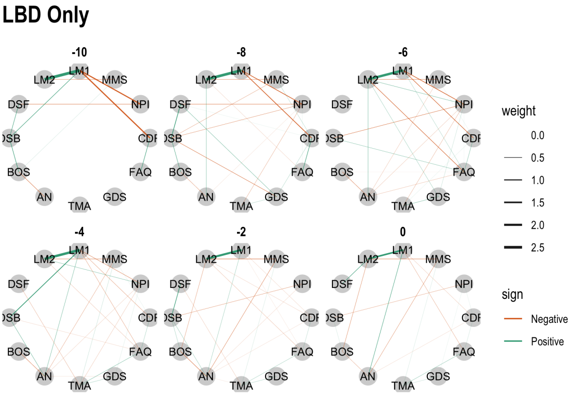
Results

- $N = 380$ with AD/LBD.

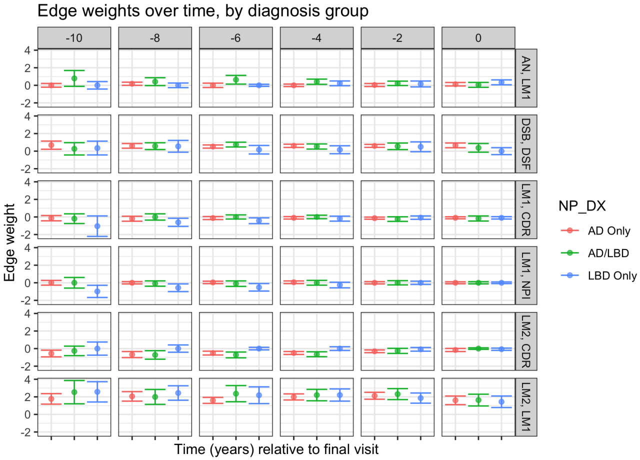


Results

- $N = 139$ with LBD only.



Results


Edge weights over time, by diagnosis group
-10
-8
-6
-4
-2
0
AN, LM1
DSB, DSF
LM1, CDR
LM1, NPI
LM2, CDR
LM2, LM1
Edge weight
Time (years) relative to final visit
NP_DX
AD Only
AD/LBD
LBD Only

Summary and future directions

- Association between Logical Memory IA and IIA may be a valuable discriminant predictor of AD vs. LBD.
- Association between Logical Memory IA and Animal Naming may be an informative marker of AD/LBD.
- Future work: Covariate-adjusted mixed graphical model.
- Limitation of mixed graphical model: Scaling.
  - Magnitude of associations may mean different things for different variable type pairs.
  - Consider different tuning parameters for each data type.[27]

[27] Jonas M. B. Haslbeck and Lourens J. Waldorp. In: *Journal of Statistical Software* 98.8 (2020), pp. 1–46.