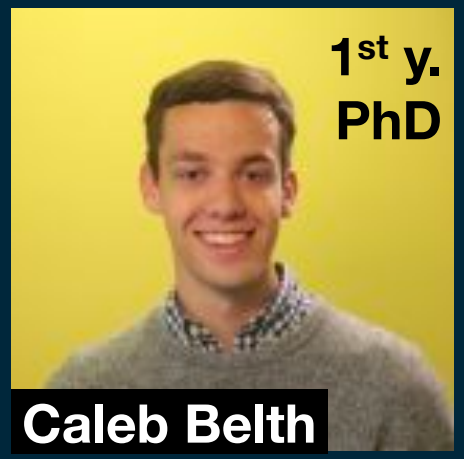
Inference, summarization, and interpretation of noisy network data

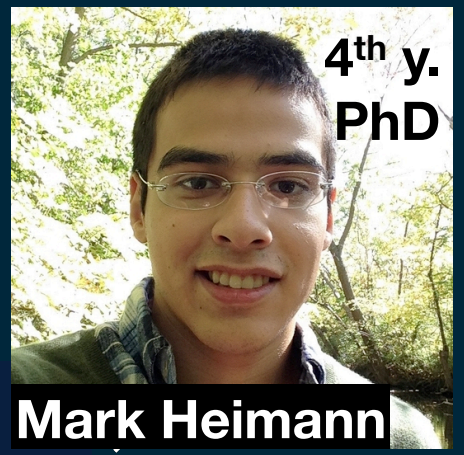
Danai Koutra

Assistant Professor, CSE

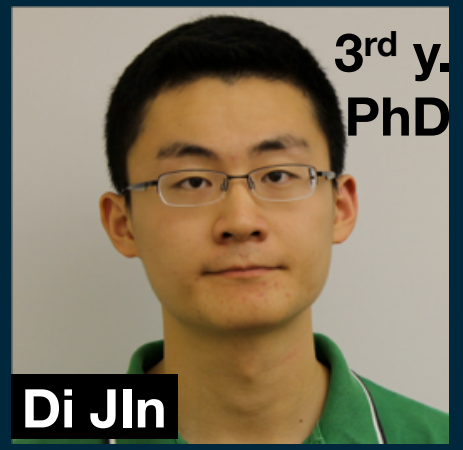
Machine Learning in Network Science – May 27, 2019

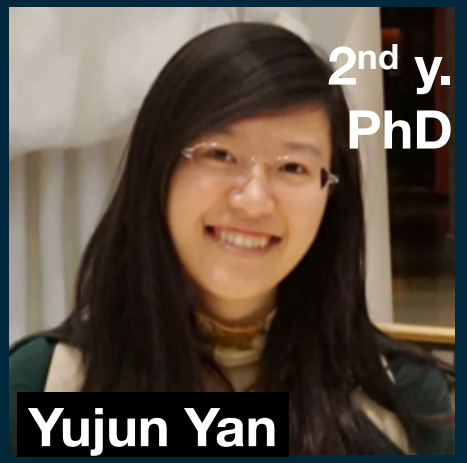
Slides at: https://bit.ly/2K87LTI





+++





GEMS Lab @ University of Michigan

Home | People | News | Research | Data and code | Lab photos

GEMS LAB

Welcome!

We are the **Graph Exploration and Mining at Scale (GEMS)** lab at the University of Michigan, founded and led by Danai Koutra. Our team researches important data mining and machine learning problems involving interconnected data: in other words, *graphs or networks*.

From airline flights to traffic routing to neuronal interactions in the brain, graphs are ubiquitous in the real world. Their properties and complexities have long been studied in fields ranging from mathematics to the social sciences. However, many pressing problems involving graph data are still open. One well-known problem is *scalability*. With continual advances in data generation and storage capabilities, the size of graph datasets has dramatically increased, making scalable graph methods indispensible. Another is the changing nature of data. Real graphs are almost always *dynamic*, evolving over time. Finally, many important problems in the social and biological sciences involve analyzing not one but *multiple* networks.

So, what do we do?

The problems described above call for **principled, practical, and highly scalable graph mining methods**, both theoretical and application-oriented. As such, our work connects to fields like linear algebra, distributed systems, deep learning, and even neuroscience. Some of our ongoing projects include:

- Algorithms for multi-network tasks, like matching nodes across networks
- Learning low-dimensional representations of networks in metric spaces
- Abstracting or "summarizing" a graph with a smaller network
- Analyzing network models of the brain derived from fMRI scans
- Distributed graph methods for iteratively solving linear systems
- Network-theoretical user modeling for various data science applications

We're grateful for funding from Adobe, Amazon, the Army Research Lab, the Michigan Institute for Data Science (MIDAS), Microsoft Azure, the National Science Foundation (NSF), and Trove.

News

May 2019
Welcome new PhDs!

May 2019
Tara passes her prelim

April 2019
3 papers accepted to KDD 2019

March 2019
Danai receives NSF CAREER award

January 2019
Danai awarded an Amazon research grant

December 2018
Yujun and Marlena selected for CRA-W Grad Cohort 2019

December 2018
1 paper accepted at SDM

September 2018
Welcome new PhDs!

August 2018
1 paper accepted at CIKM

August 2018
New website

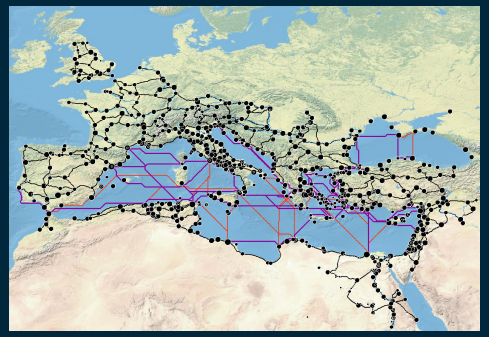
May 2018
1 paper accepted at KDD

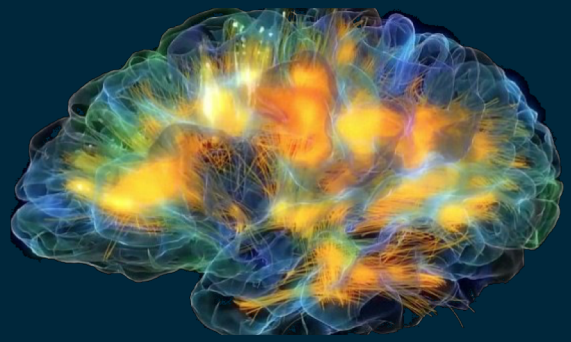
May 2018
Grant for music + big data

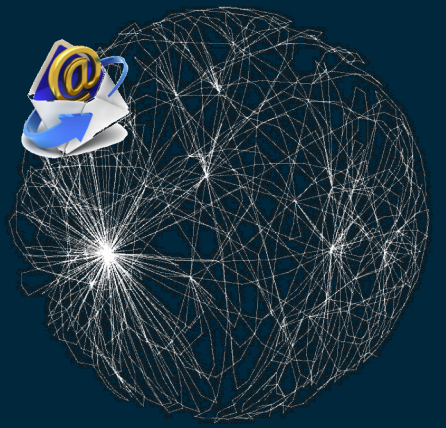
April 2018
Danai awarded Adobe Digital

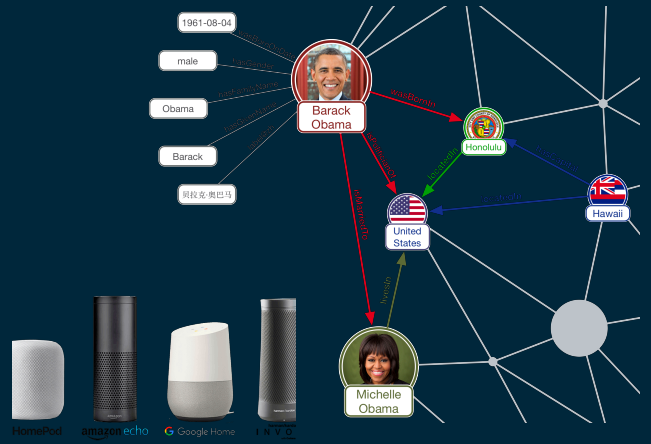
NSF
ARL
TROVE
MIDAS
Adobe
Microsoft Azure
amazon

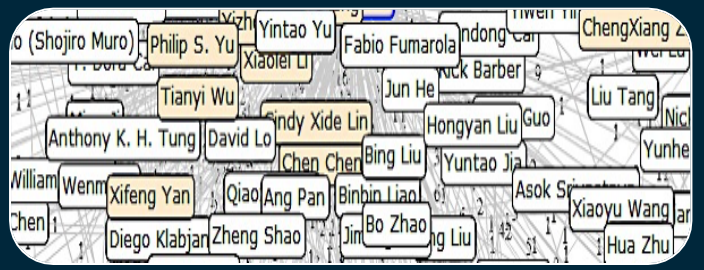
Networks are everywhere!

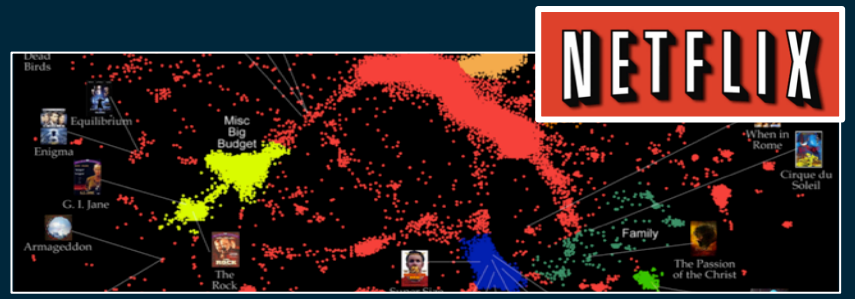




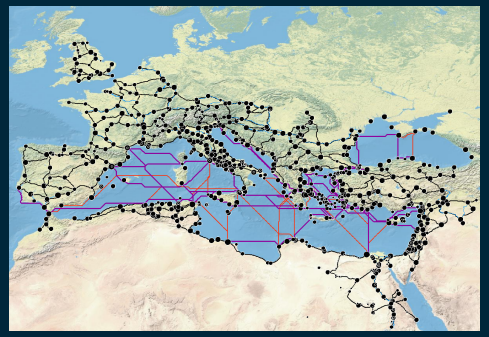


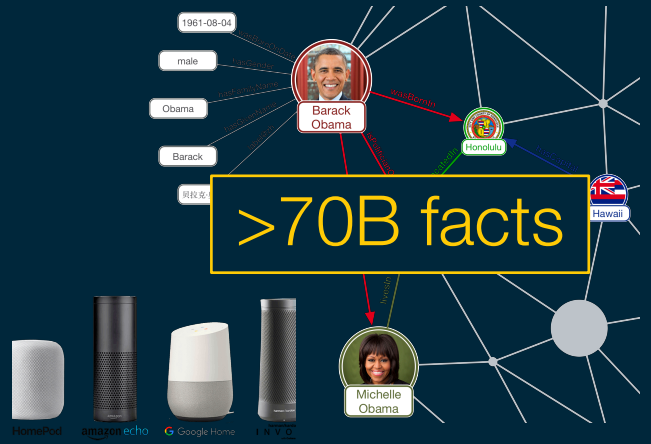


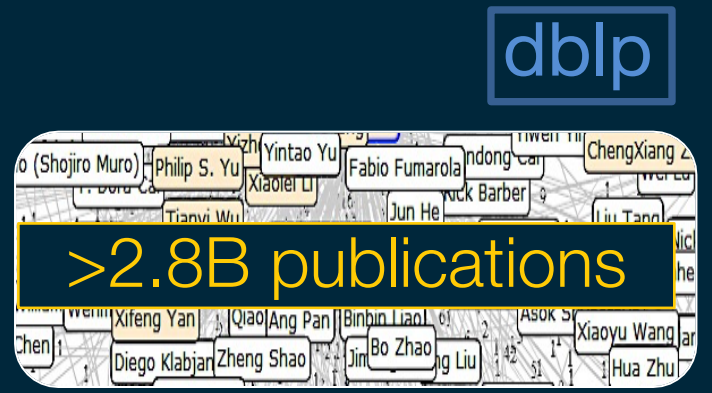




… and they're growing quickly!

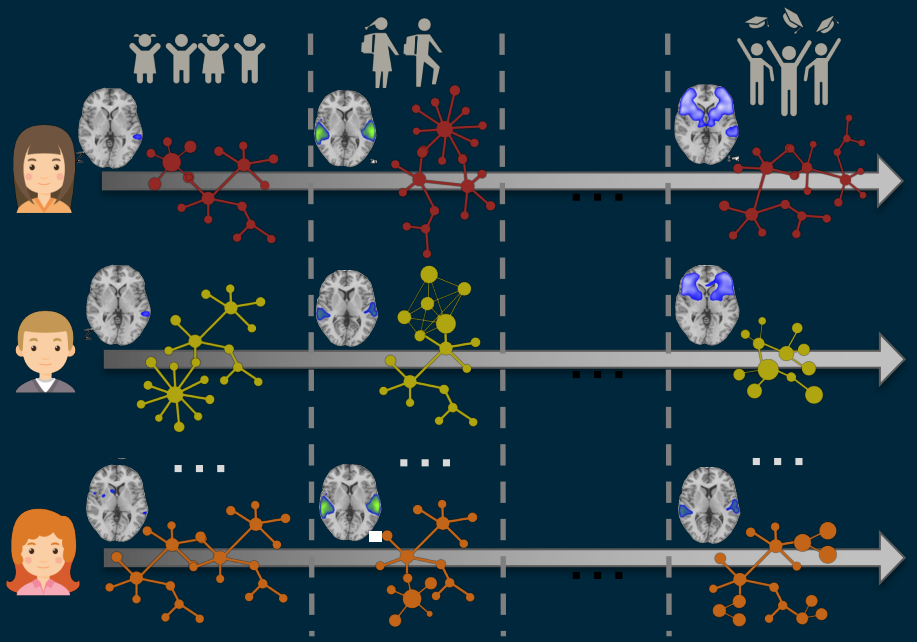




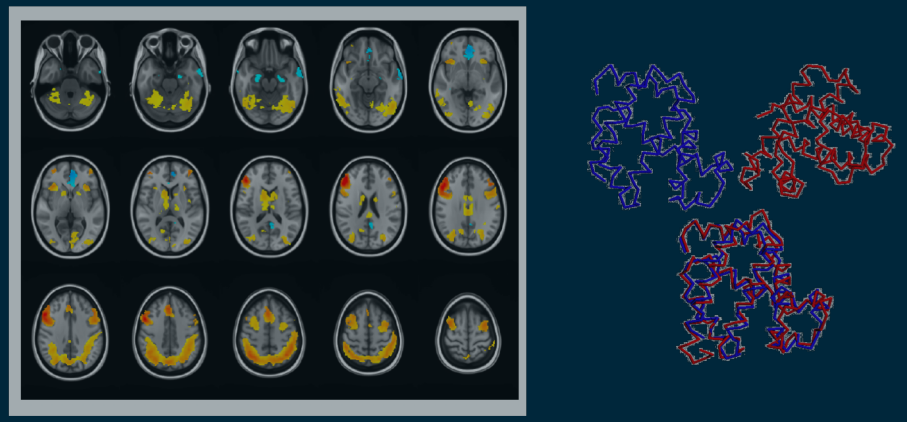


Many (small- or medium-sized) networks

Pattern mining in many temporal networks

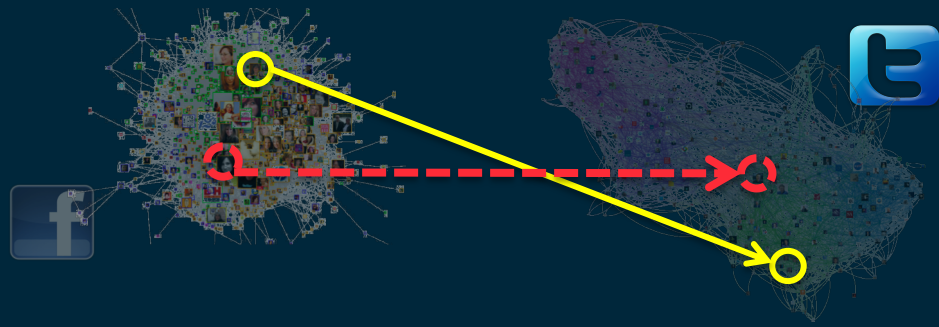


Graph similarity / classification



Pattern mining / search against a DB

Alignment or matching



Many (small- or medium-sized) networks

Networks are everywhere

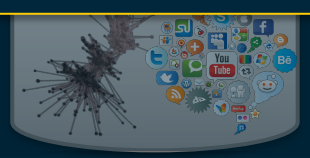
…but are not always directly observed!

How can we

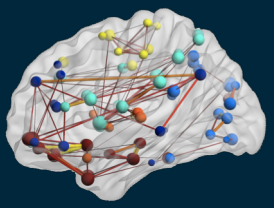
(1) infer networks from other data,

(2) summarize large collections of networks and

(3) interpret the underlying phenomena efficiently?



Talk Outline



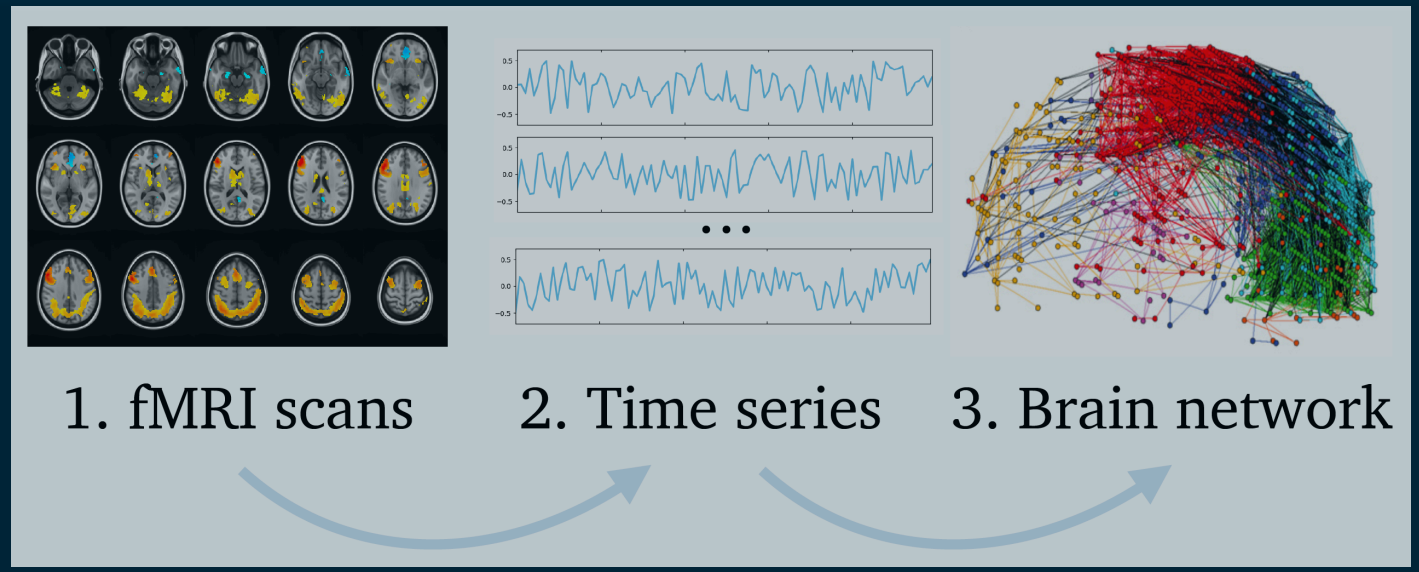
- Network inference from non-network data [ICDM'17, KAIS'18]
- Domain-aware summarization [ICDM'17]
- NN-based summarization for interpretation [KDD'19]

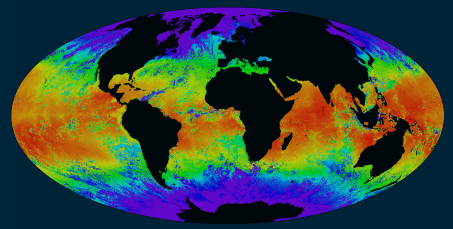
Based on:

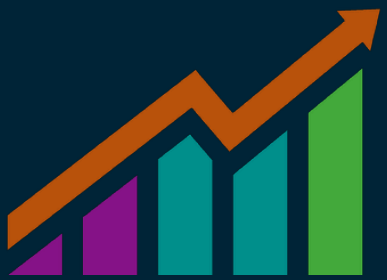
- T. Safavi, C. Sripada, D. Koutra. Fast Network Discovery on Sequence Data via Time-Aware Hashing. KAIS'18.
- T. Safavi, C. Sripada and D. Koutra. Scalable Hashing-Based Network Discovery. IEEE ICDM'17
- Di Jin and Danai Koutra. Exploratory Analysis of Graph Data by Leveraging Domain Knowledge. IEEE ICDM'17
- Y. Liu, T. Safavi, A. Dighe, D. Koutra. Graph Summarization Methods and Applications: A Survey. ACM Computing Surveys '18.
- Y. Yan, J. Zhu, M. Duda, E. Solarz, C. Sripada, D. Koutra. GroupINN: Grouping-based Interpretable Neural Network-based Classification of Limited, Noisy Brain Data. ACM KDD'19.

Network Inference

- Given: indirect, possibly noisy measurements with unobserved interactions
- Reconstruct: a network



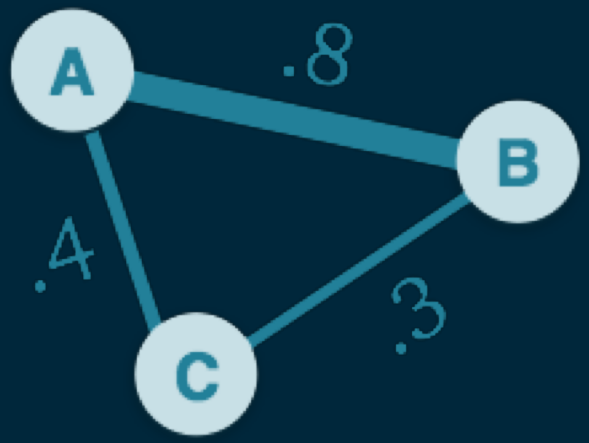




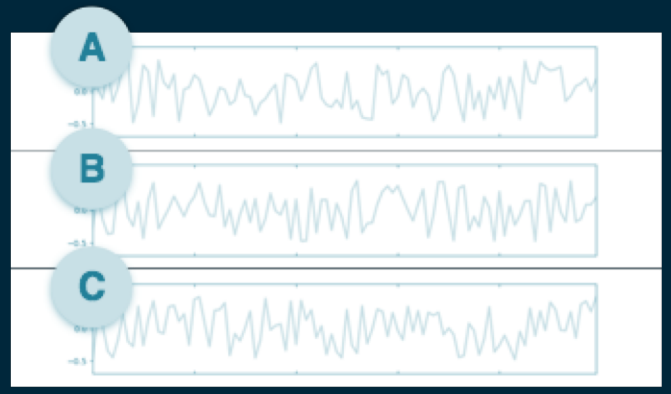
[Survey by Brugere, Gallagher, Berger-Wolf. ACM Computing Surveys 2018.]

Traditional Approach

2. Fully-connected weighted network

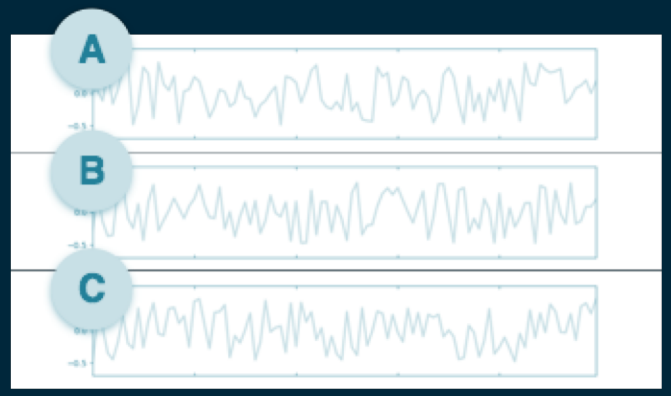


All-pairs correlation



1. *N* time series

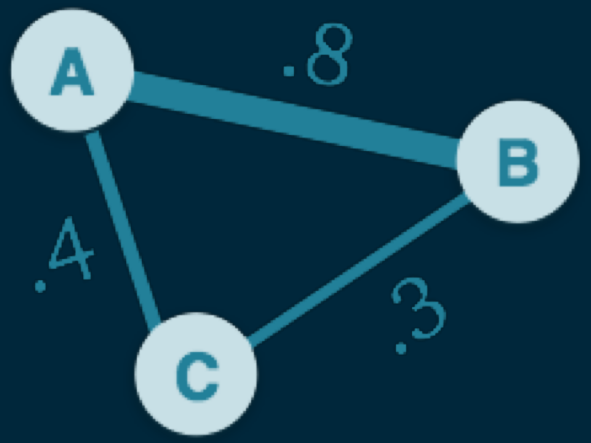
Traditional Approach



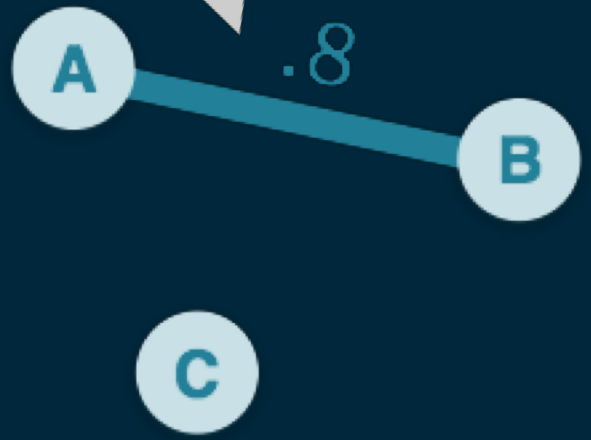
1. *N* time series

All-pairs correlation

2. Fully-connected weighted network



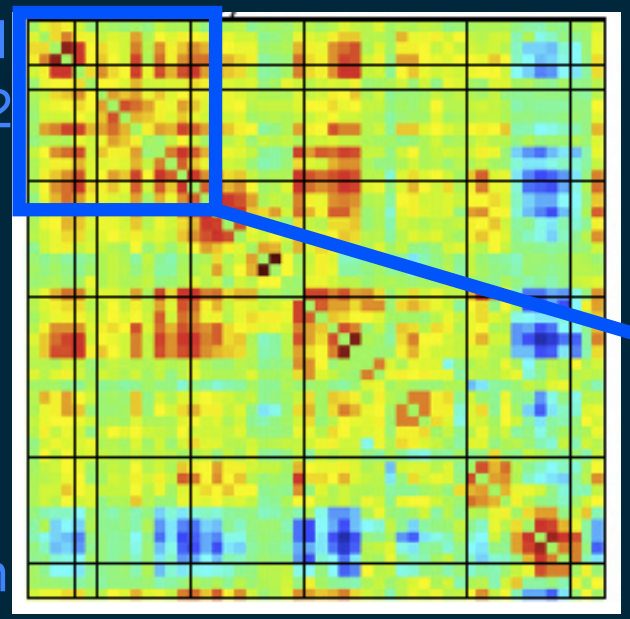
Drop edges below threshold θ



3. Sparse graph

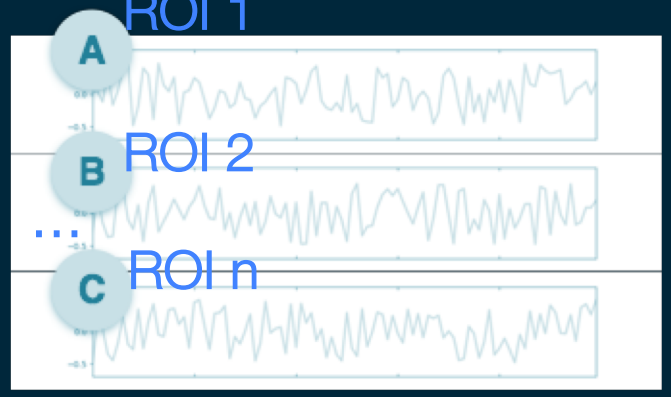
Traditional Approach

2. Fully-connected weighted network



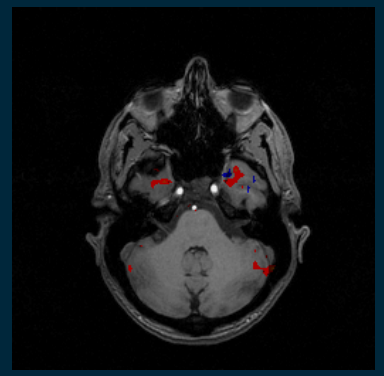
All-pairs correlation

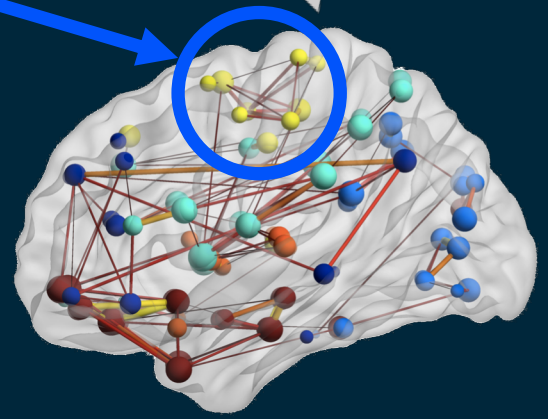
Drop edges below threshold θ



1. *N* time series

neural activity via blood oxygen level dependent (BOLD) signal



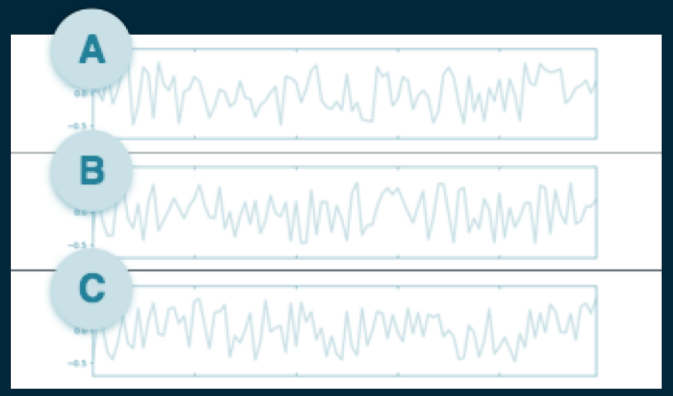


3. Sparse graph

[Smith (2001) (https://users.fmrib.ox.ac.uk/)] [BrainNetViewer, Beijing Normal University]

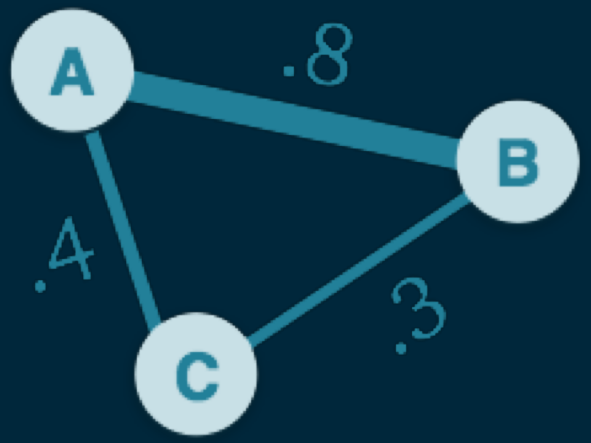
Traditional Approach

1. *N* time series



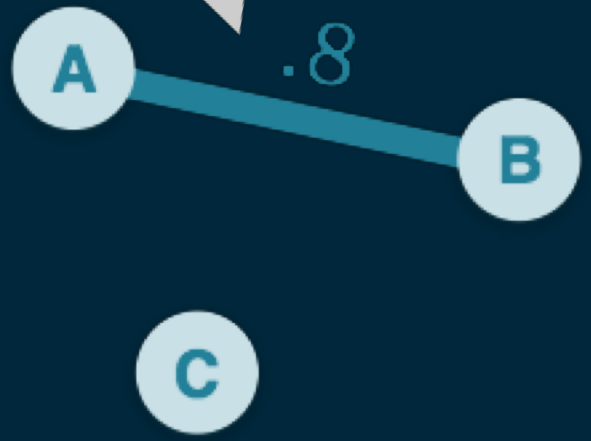
All-pairs correlation

2. Fully-connected weighted network



Widely used in many domains, **interpretable**, but…

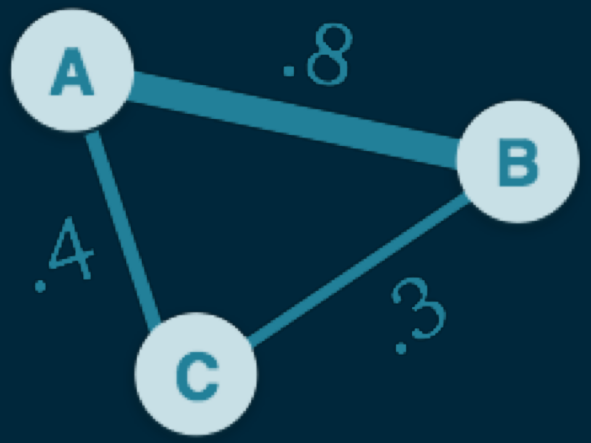
Drop edges below threshold θ



3. Sparse graph

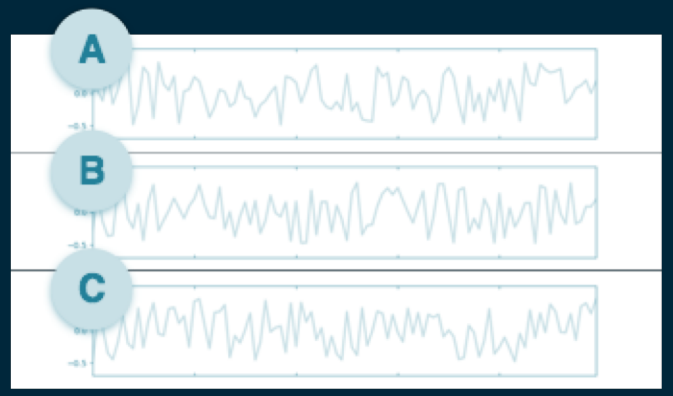
Traditional Approach

2. Fully-connected weighted network



All-pairs correlation

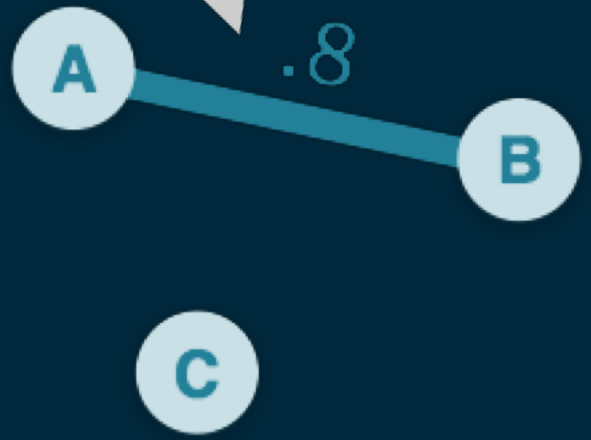
Drop edges below threshold θ



1. *N* time series

$O(N^2)$ comparisons

How to set?



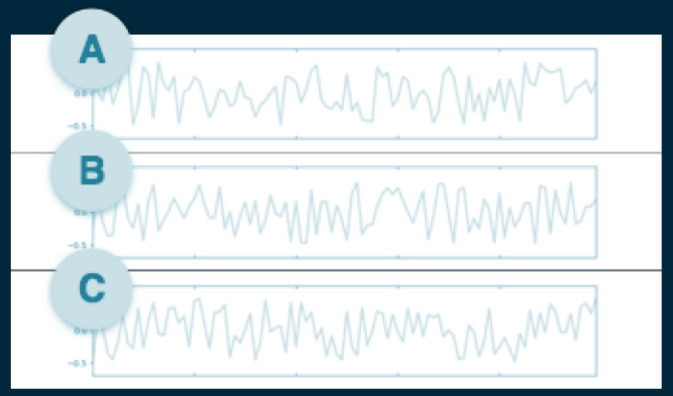
3. Sparse graph

Proposed Approach

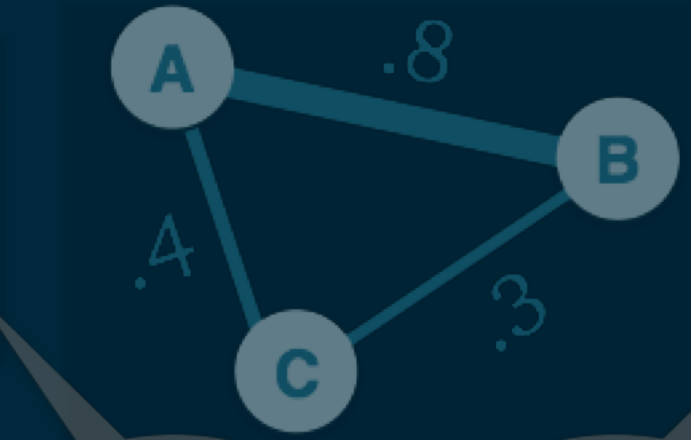
2. Fully-connected weighted network

All-pairs correlation

Drop edges below threshold θ

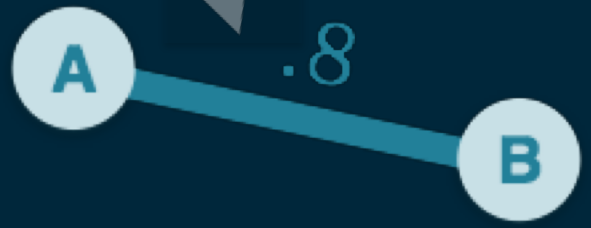


1. *N* time series

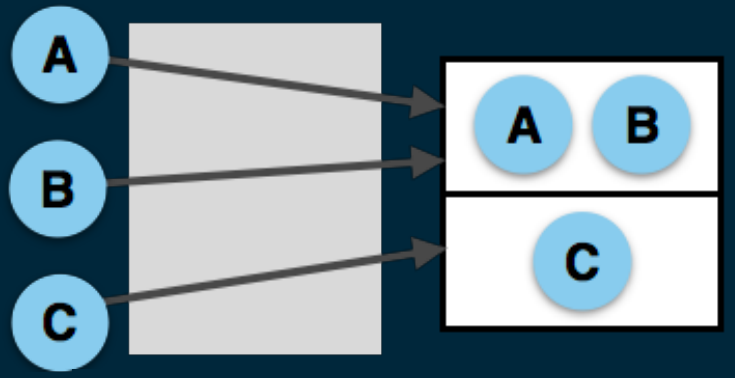


$O(N^2)$ comparisons

How to set?



2. Hash sequences



Hash function Buckets

binarize

Bucket pairwise similarity

C

3. Sparse graph

Proposed ABC measure
or Approximate Binarized Correlation

- Emphasize consecutive similarity between sequences
  - over pointwise comparison
  - Capture variable-length consecutive runs between series
- Similarity score *s:* sum of *p* geometric series, each of length k_i

$$s(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{p}\sum_{b=0}^{k_i}(1+\alpha)^b = \frac{\sum_{i=1}^{p}(1+\alpha)^{k_i} - p}{\alpha}$$

$0 < \alpha << 1$
consecutiveness
weighting factor

x:	1	1	0	1	0	0	0
y:	1	1	1	1	0	0	1

$$(1+\alpha)^0 + (1+\alpha)^1 \quad + \quad (1+\alpha)^0 + (1+\alpha)^1 + (1+\alpha)^2$$

[Tara Safavi, et al. IEEE ICDM'17 and KAIS'18]

Proposed ABC measure

or Approximate Binarized Correlation

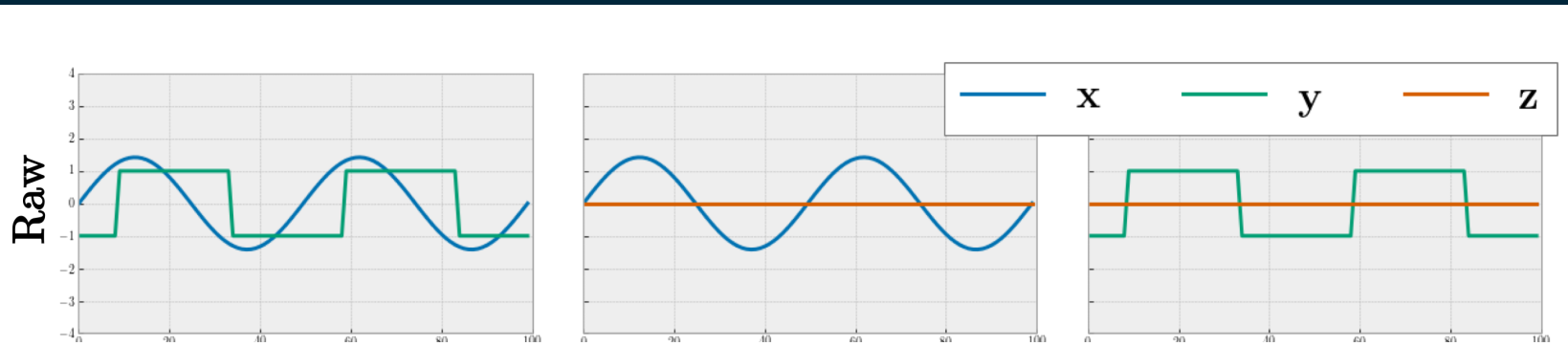
- Emphasize consecutive similarity between sequences
- Similarity score *s:* sum of *p* geometric series, each of length k_i

$$s(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{p}\sum_{b=0}^{k_i}(1+a)^b = \frac{\sum_{i=1}^{p}(1+\alpha)^{k_i} - p}{\alpha}$$

x: 1 1 0 1 0 0 0

y: 1 1 1 1 0 0 1

$(1+\alpha)^0 + (1+\alpha)^1$ + $(1+\alpha)^0 + (1+\alpha)^1 + (1+\alpha)^2$



Proposed ABC measure

or Approximate Binarized Correlation

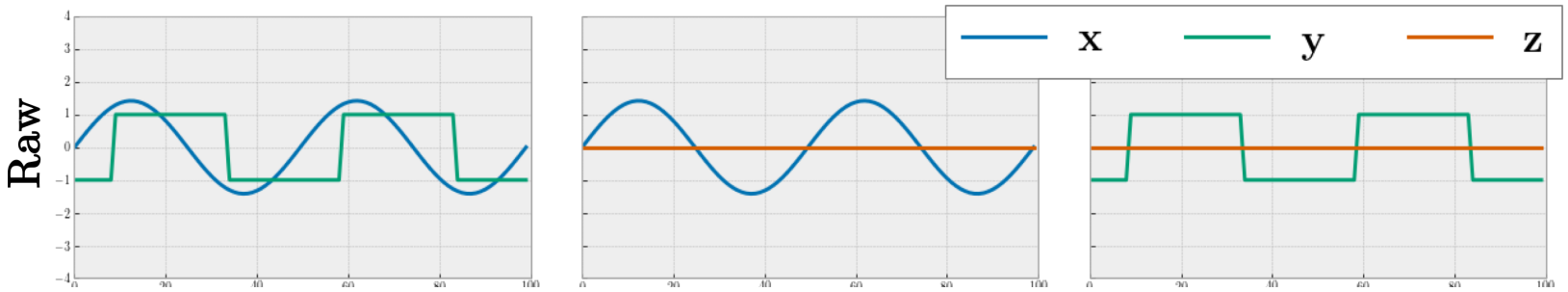
- Emphasize consecutive similarity between sequences
- Similarity score *s:* sum of *p* geometric series, each of length k_i

$$s(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{p}\sum_{b=0}^{k_i}(1+a)^b = \frac{\sum_{i=1}^{p}(1+\alpha)^{k_i} - p}{\alpha}$$

x: 1 1 0 1 0 0 0

y: 1 1 1 1 0 0 1

$(1+\alpha)^0 + (1+\alpha)^1$ + $(1+\alpha)^0 + (1+\alpha)^1 + (1+\alpha)^2$



$\mathrm{ED}(\mathbf{x}, \mathbf{y}) = 11.1$
$\mathrm{ABC}(\mathbf{b}(\mathbf{x}), \mathbf{b}(\mathbf{y})) = 20.4$

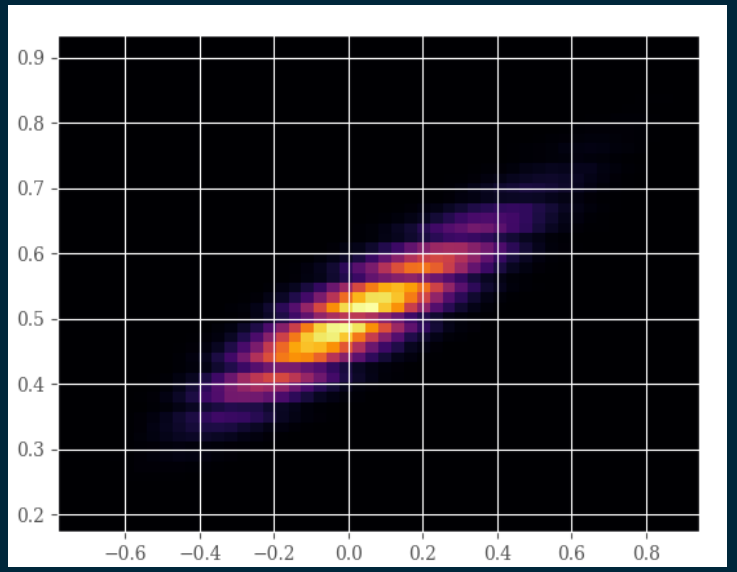
$\mathrm{ED}(\mathbf{x}, \mathbf{z}) = 10.0$
$\mathrm{ABC}(\mathbf{b}(\mathbf{x}), \mathbf{b}(\mathbf{z})) = 50.4$

$\mathrm{ED}(\mathbf{y}, \mathbf{z}) = 10.0$
$\mathrm{ABC}(\mathbf{b}(\mathbf{y}), \mathbf{b}(\mathbf{z})) = 50.4$

Proposed ABC measure

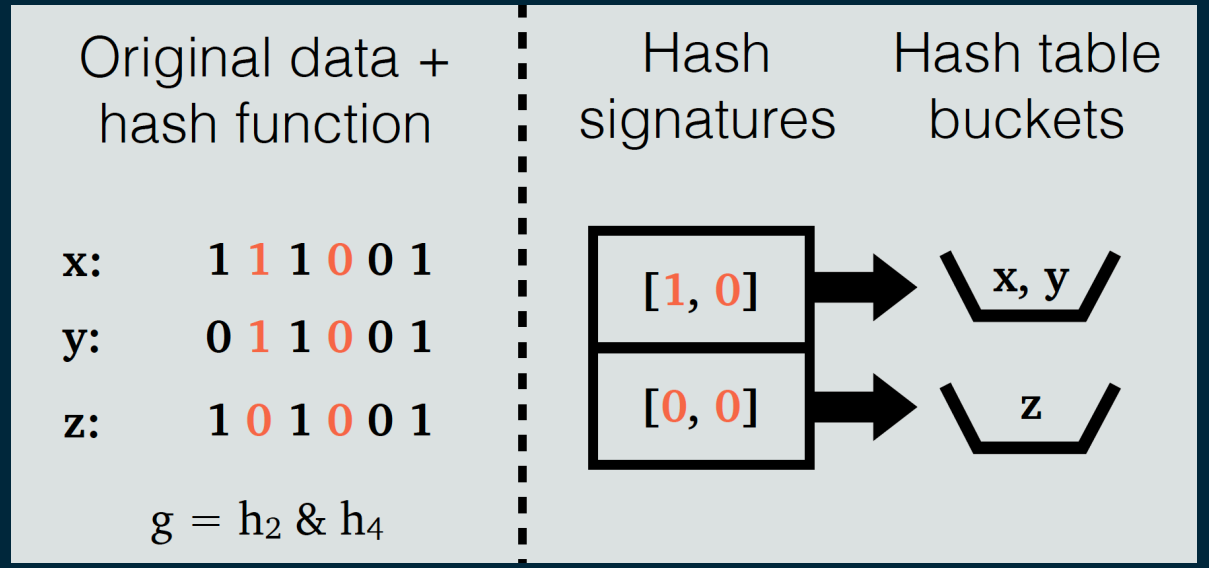
or Approximate Binarized Correlation

- Empirically, a good estimator of correlation coefficient *r*
  - Similarity scores *s* correlate well with *r*
- Added benefit of time-aware hashing
  - LSH requires a *metric*: satisfies triangle inequality
  - ABC distance is a metric (critical)



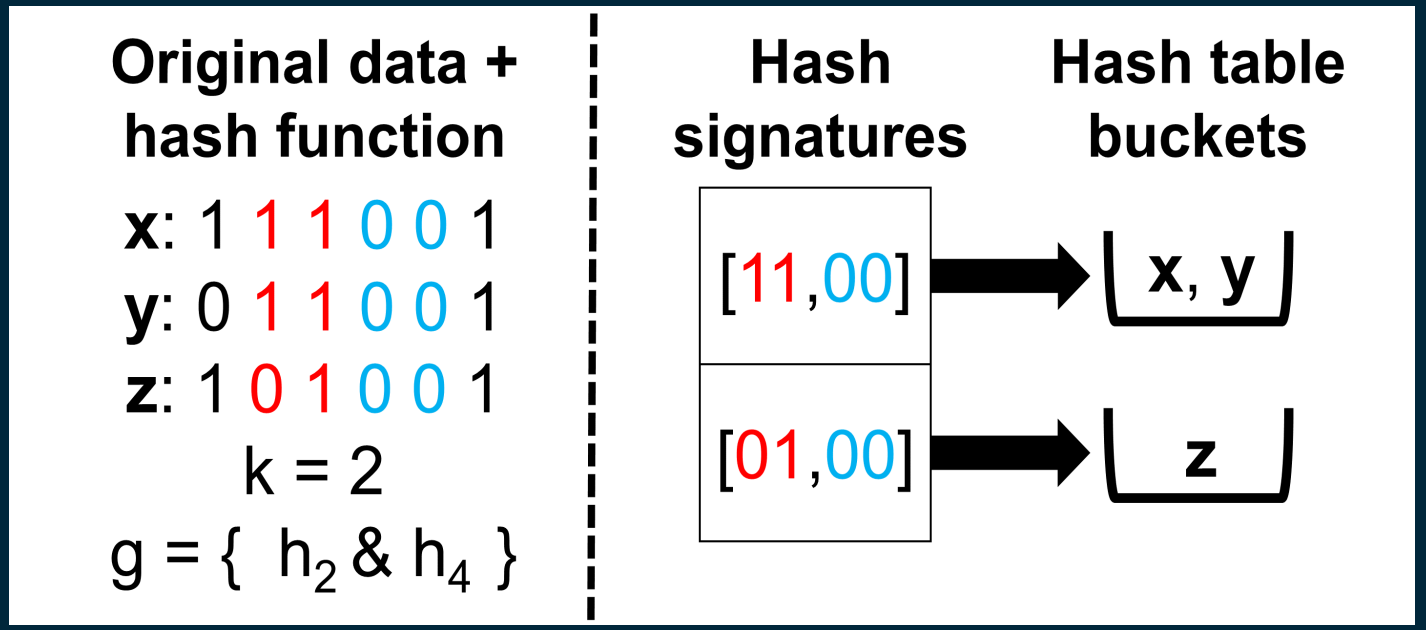
Locality Sensitive Hashing

- Hash data s.t. similar items likely to collide
- Family of hash fns **F**: (d_1, d_2, p_1, p_2)-sensitive
  - Control false negative/positive rates
- Parameters
  - b: number of hash tables, increases p_1
  - r: number of hash functions to concatenate, lowers p_2



Proposed LSH Family

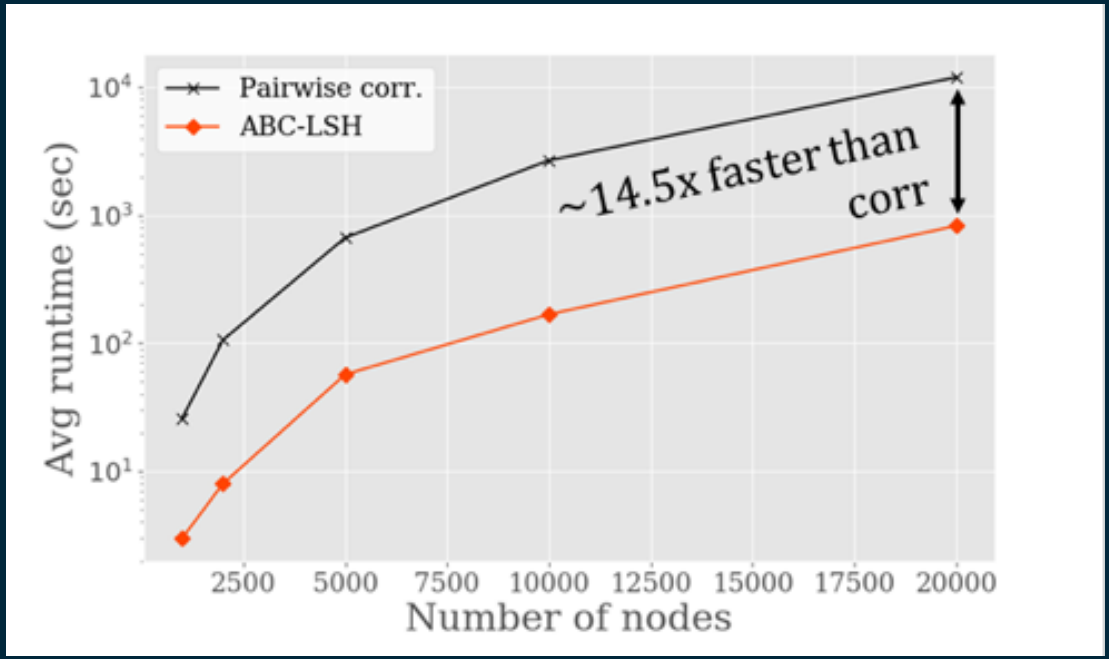
- "Window" sampling LSH family



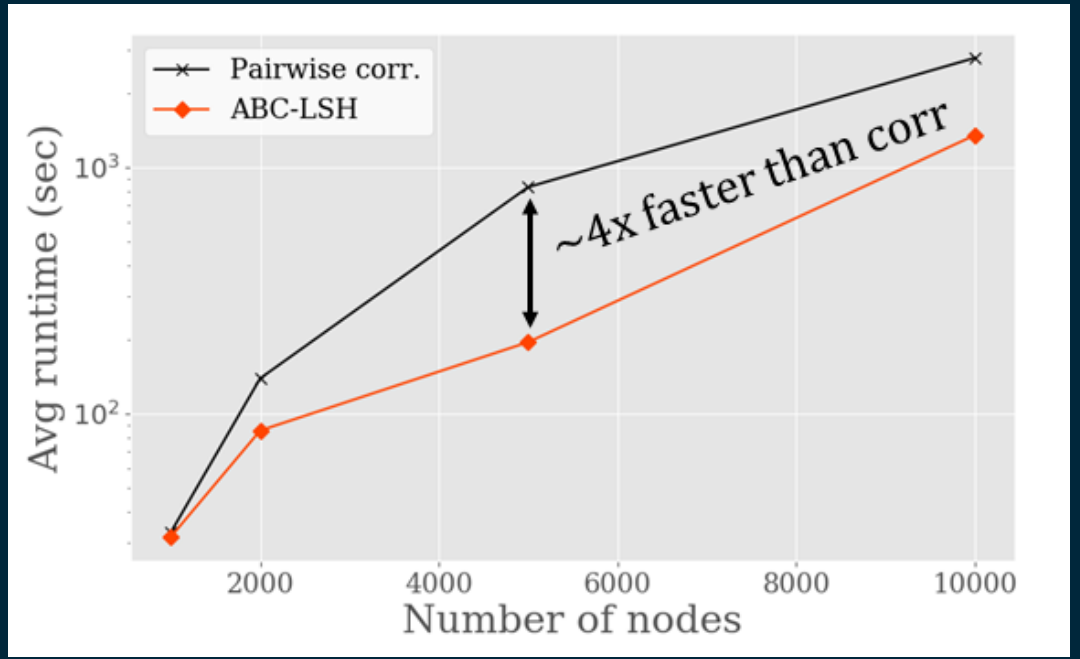
The new family is $(d_1, d_2, 1 - a\frac{d_1}{(1+a)^n-1}, 1 - a\frac{d_2}{(1+a)^n-1})$ - sensitive

Question 1: Scalability

Penn (synthetic)



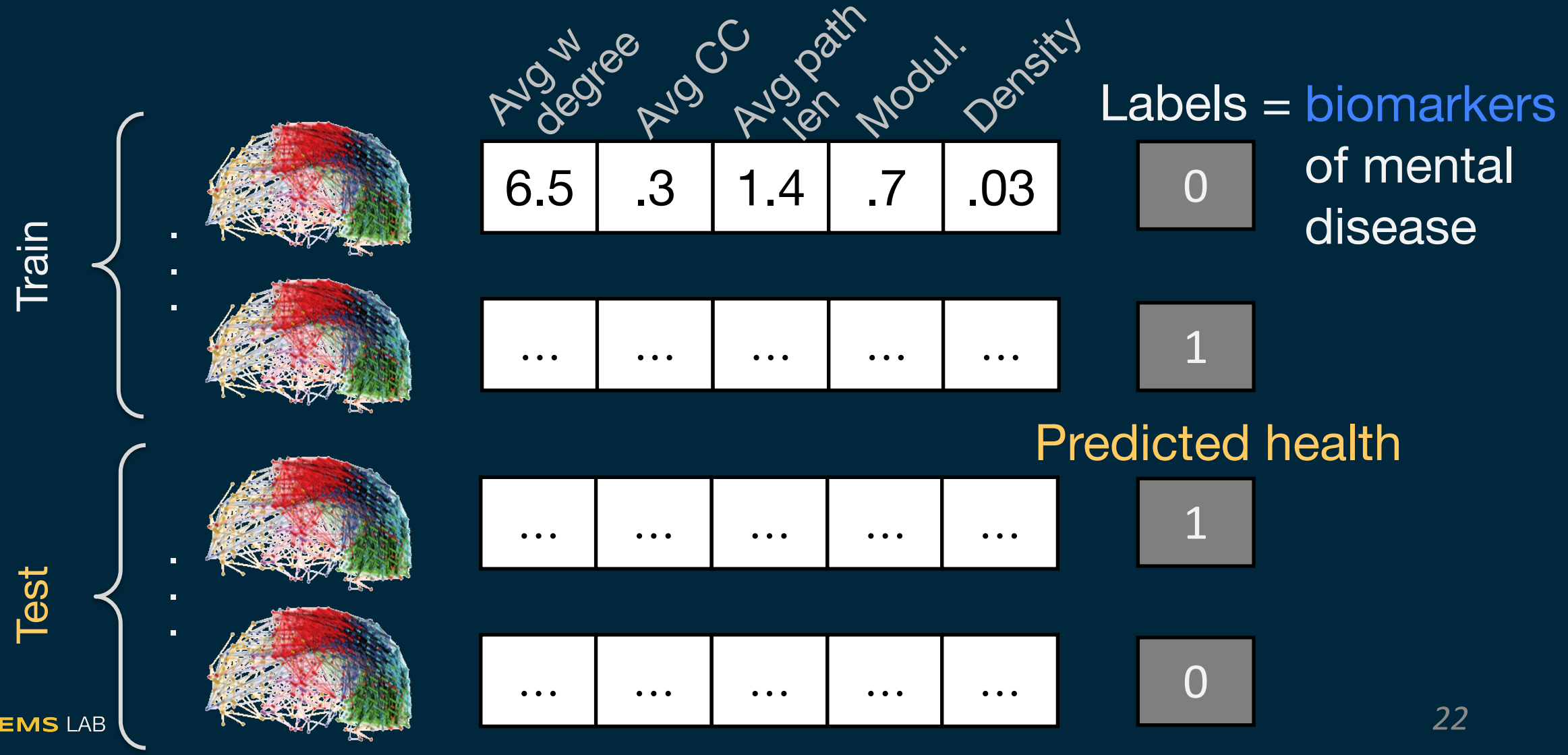
StarLightCurves



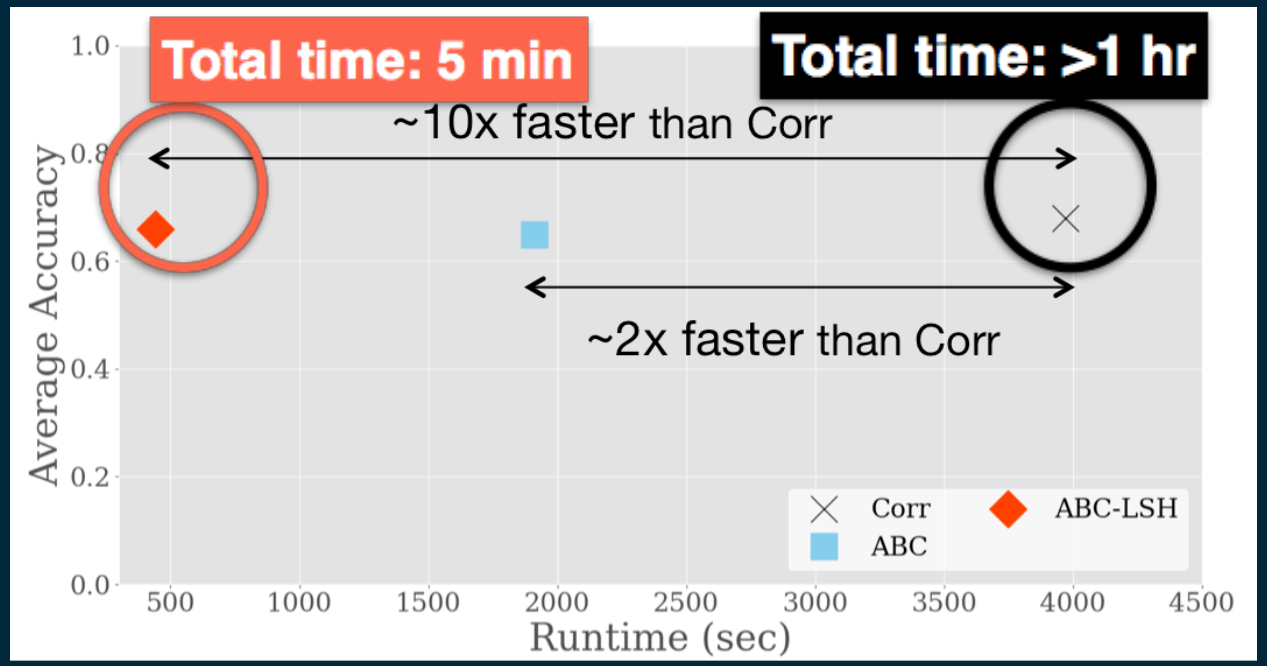
ABC-LSH is up to 2-15x faster than pairwise correlation.

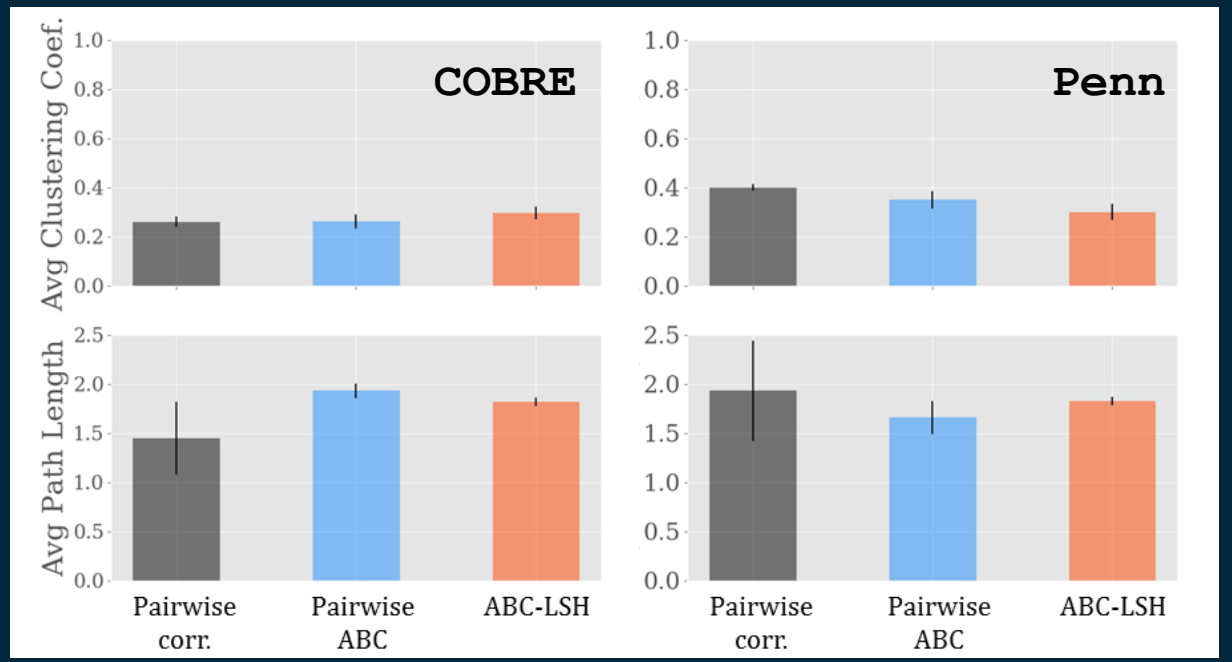
Question 2: Task-based Evaluation

- Logistic regression classifier, 10-fold stratified CV



Question 2: Task-based Evaluation





ABC-LSH approximates the average properties of graphs well, while being >=10x faster than correlation.

ABC-LSH: More Applications

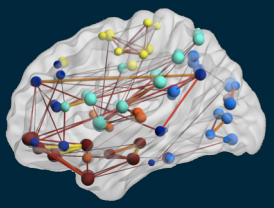
- Identify users with similar behaviors,
- Identify regions with similar traffic,
- Identify anomalous patterns in computer networks
- …

Google

https://github.com/tsafavi/hashing-based-network-discovery

[Tara Safavi, et al. IEEE ICDM'17 and KAIS'18]

Talk Outline



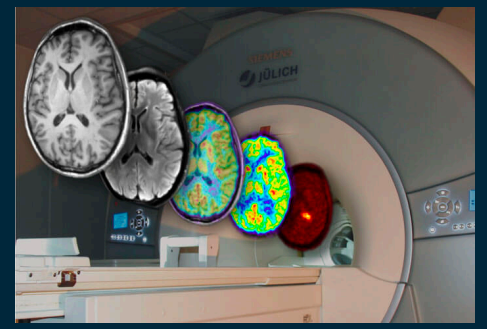
- Network inference from non-network data [ICDM'17, KAIS'18]
- Domain-aware summarization [ICDM'17]
- NN-based summarization for interpretation [KDD'19]

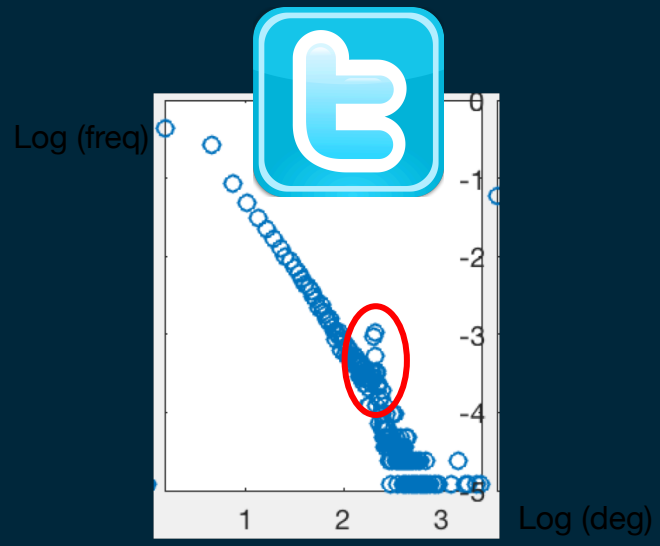
Based on:

- T. Safavi, C. Sripada, D. Koutra. Fast Network Discovery on Sequence Data via Time-Aware Hashing. KAIS'18.
- T. Safavi, C. Sripada and D. Koutra. Scalable Hashing-Based Network Discovery. IEEE ICDM'17
- Di Jin and Danai Koutra. Exploratory Analysis of Graph Data by Leveraging Domain Knowledge. IEEE ICDM'17
- Y. Liu, T. Safavi, A. Dighe, D. Koutra. Graph Summarization Methods and Applications: A Survey. ACM Computing Surveys '18.
- Y. Yan, J. Zhu, M. Duda, E. Solarz, C. Sripada, D. Koutra. GroupINN: Grouping-based Interpretable Neural Network-based Classification of Limited, Noisy Brain Data. ACM KDD'19.

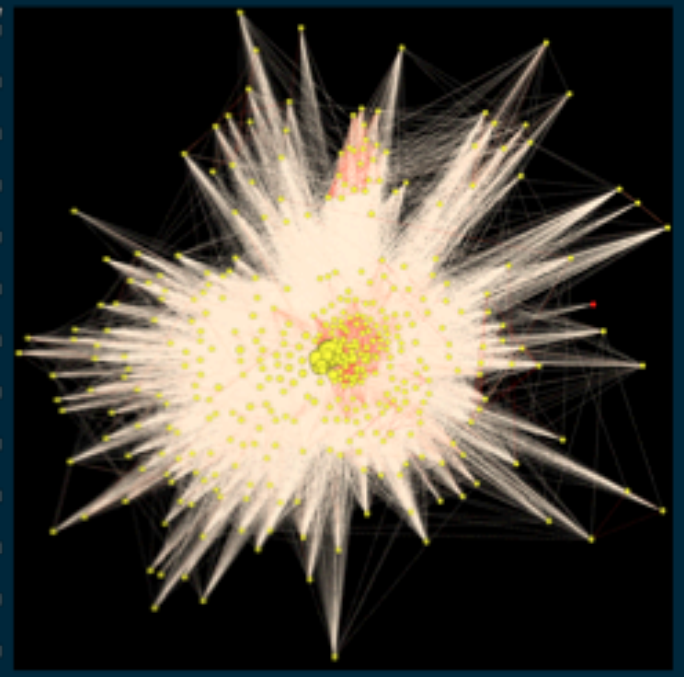
Applications based on “summaries” of features

- Healthy and unhealthy subjects in neuroscience
  - Degree
  - Clustering coefficient
  - Average path length
  - ...
- Anomaly detection in Youtube graph
  - Power laws (degree etc)
  - 6-degree of separation
  - ...

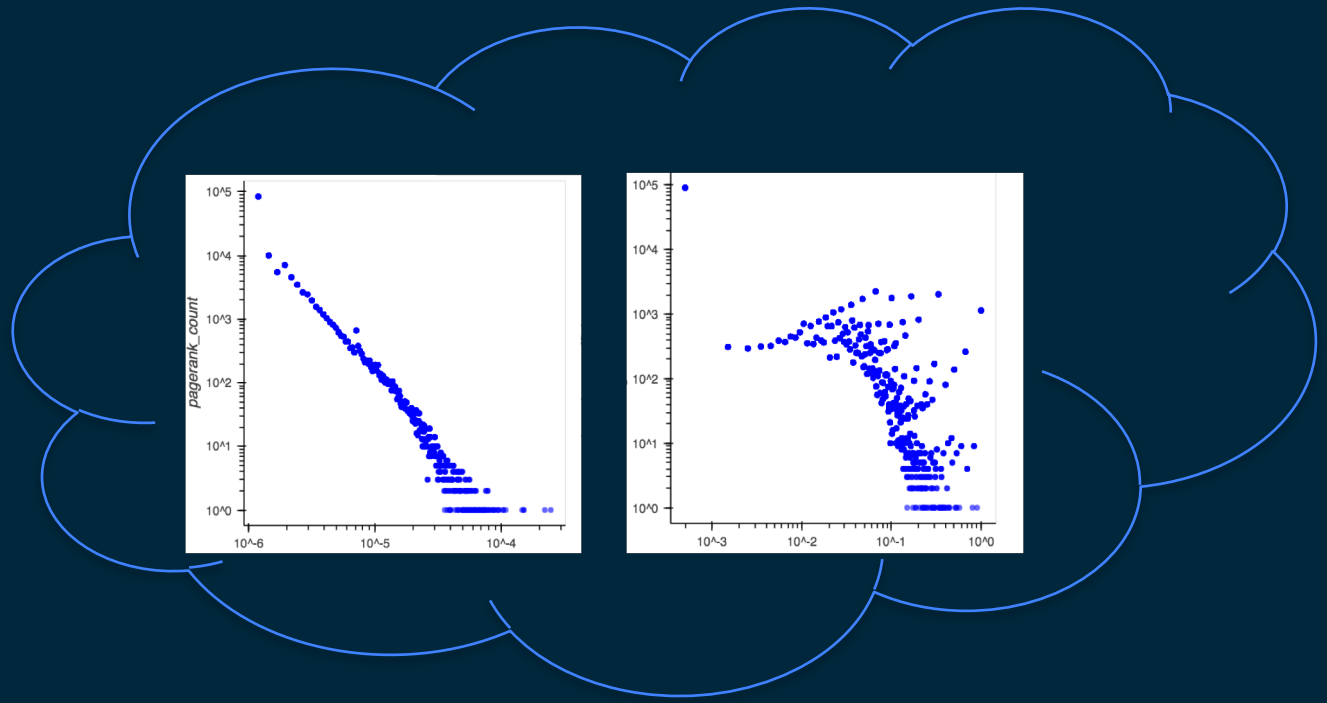




One summary does not fit all



?



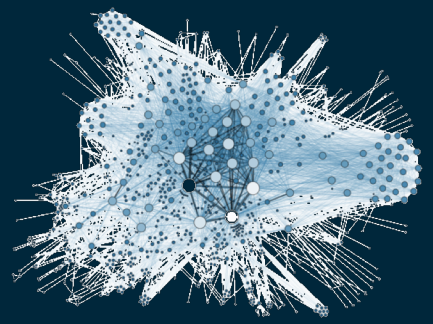
Citation: DBPL, Arxiv

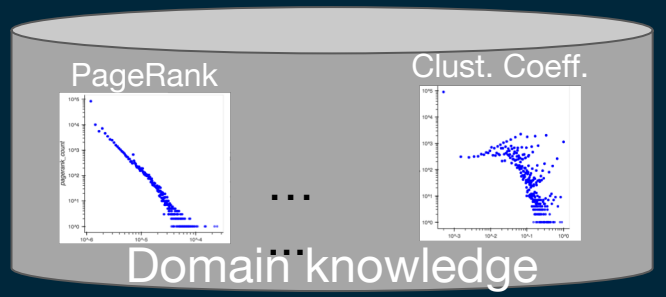
Social: Twitter, Epinions

Neural: Brain, Proteins

EAGLE: Domain-specific Summarization

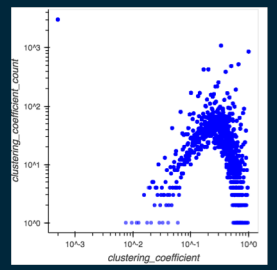
Given: an input graph & domain knowledge





a collection of graphs with all their features

Find: representative features with desired properties (e.g., diversity)



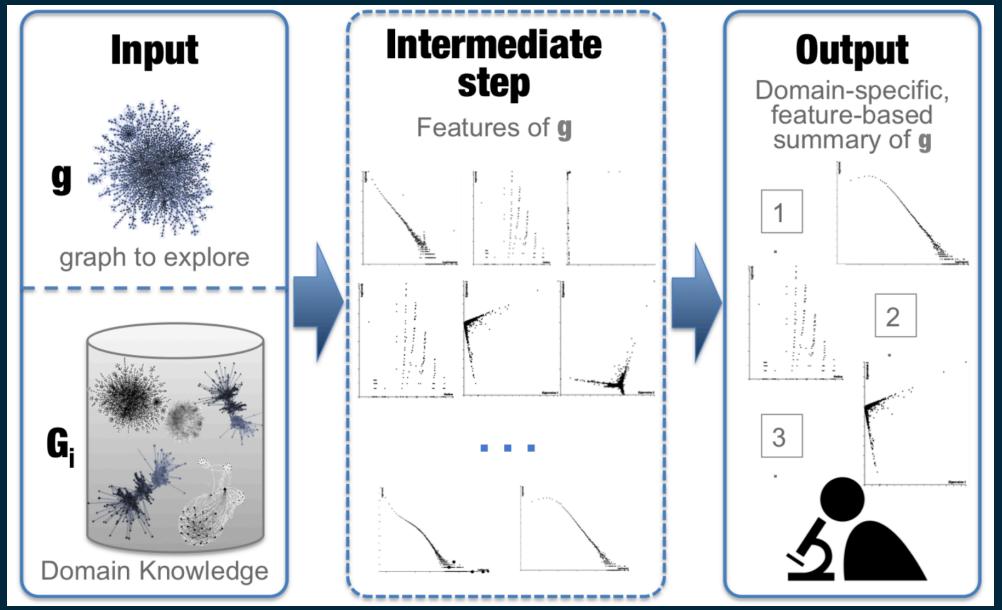
graph invariant distributions (PDF)

[Di Jin, Danai Koutra. IEEE ICDM '17.]

Domain-specific Summarization

Requirements for summary:

- Diverse
- Concise
- Domain-specific
- Interpretable
- Efficient to compute

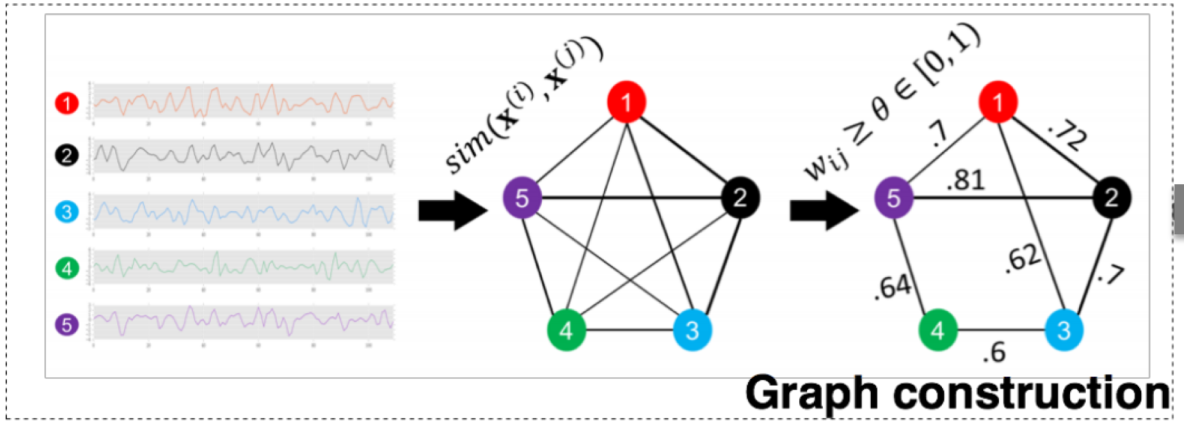


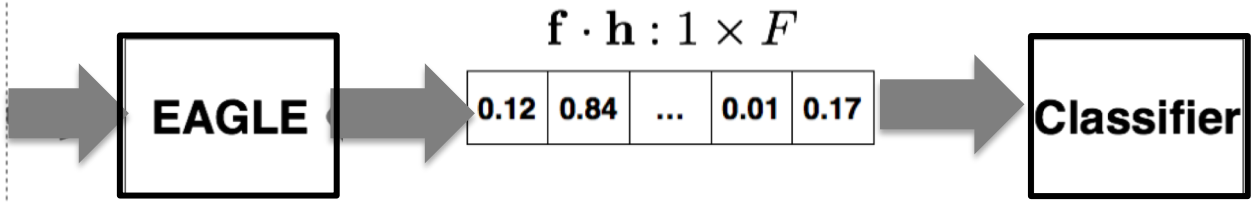
$$\underset{f}{\operatorname{argmin}}\ \lambda_1 f^T S_F f + \lambda_2 \|f\|_0 + \lambda_3 \varphi(g, G_1, G_2, \ldots, G_K)$$

diversity conciseness domain specificity

https://github.com/DerekDiJin/Domain_Knowledge

EAGLE for Graph Classification





Methods	AUC
Avg. feat. values	0.7028
Flattened adj. mat.	0.1099
Full	0.7147
EAGLE-Fix (6 feat.)	0.7371

Although not designed explicitly for this,
features selected by EAGLE
can be applied to specific tasks,
such as classification,
with promising performance.

[Di Jin, Danai Koutra. IEEE ICDM '17.]

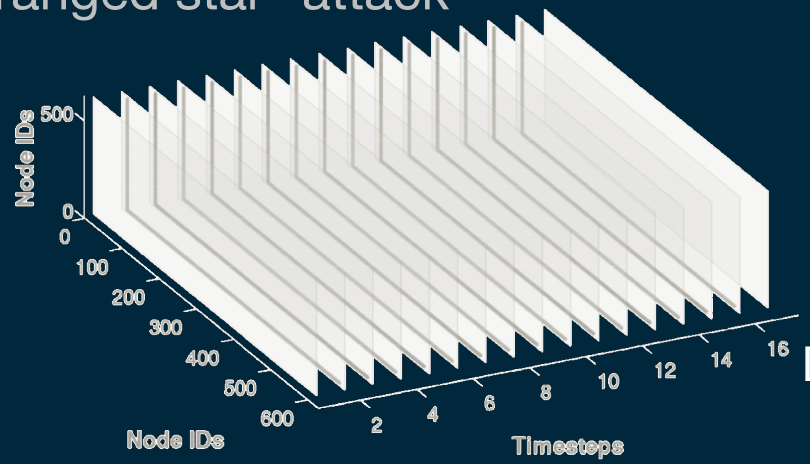
Summarizing large networks: Overview

Survey: [CSUR'18]

Structural Summaries

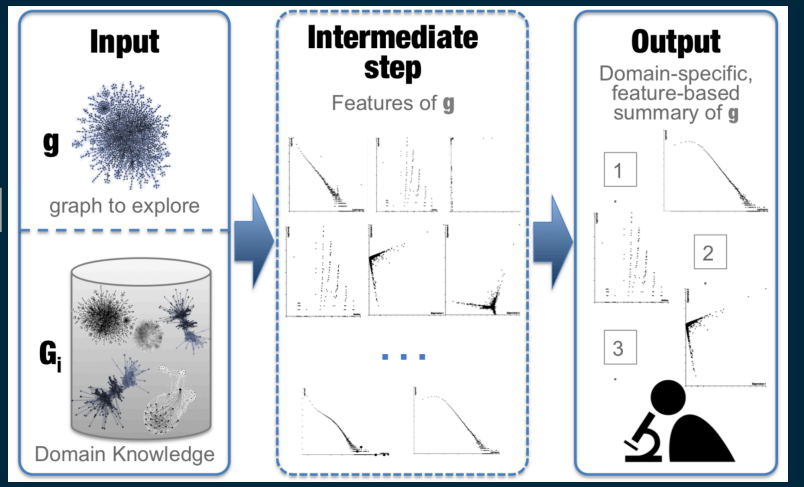
[SDM'14, KDD'15, Dat Bull Eng'17, SNAM'18, SDM'19, …]

"ranged star" attack



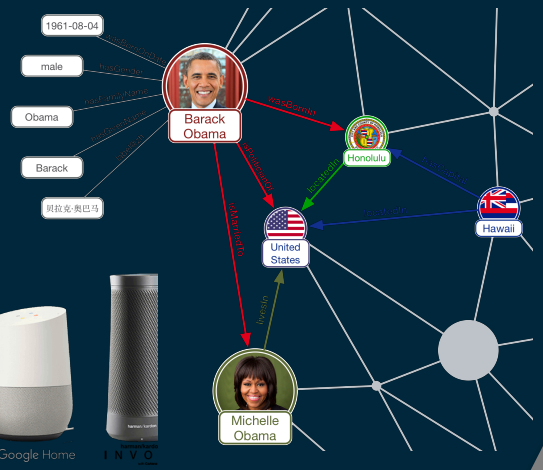
Domain-specific Summaries

[ICDM'17, KDD'19]



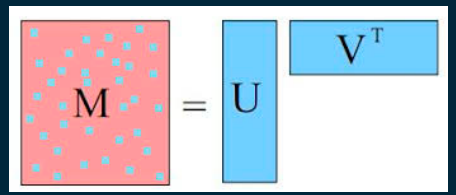
Query-on-the-edge Summaries

[ongoing]



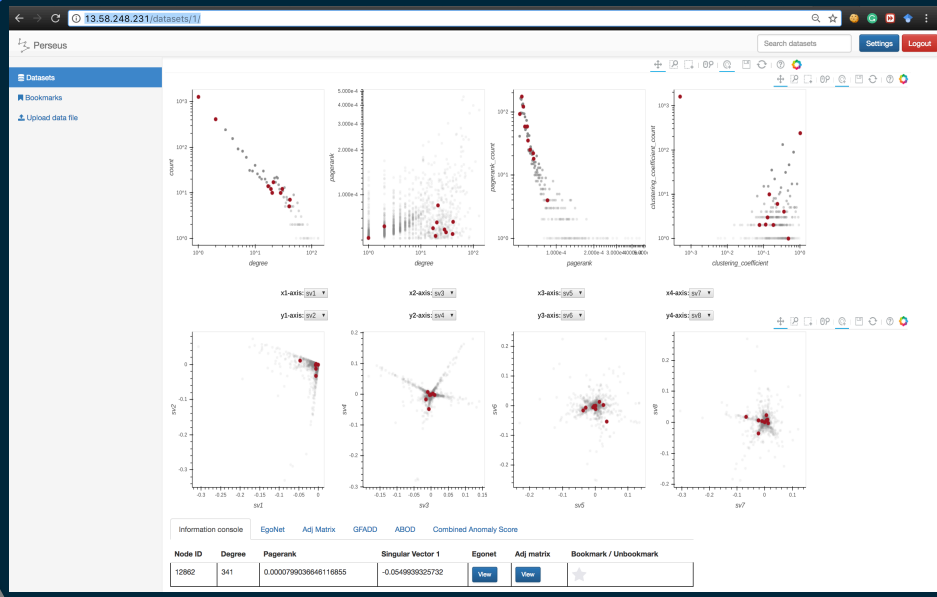
Latent Summaries

[KDD'19]

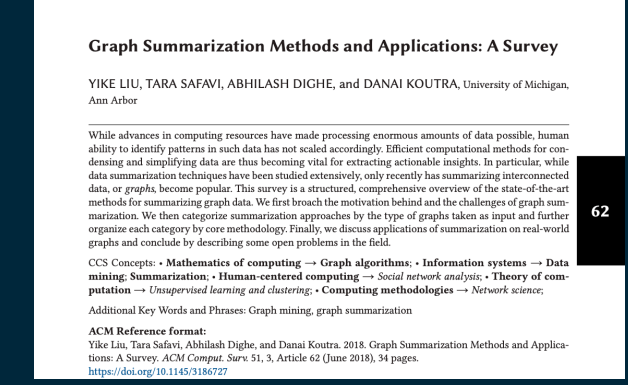


Interactive Summaries

[VLDB'15, Informatics'17]



Graph Summarization Survey



Graph Summarization Methods and Applications: A Survey

YIKE LIU, TARA SAFAVI, ABHILASH DIGHE, and DANAI KOUTRA, University of Michigan, Ann Arbor

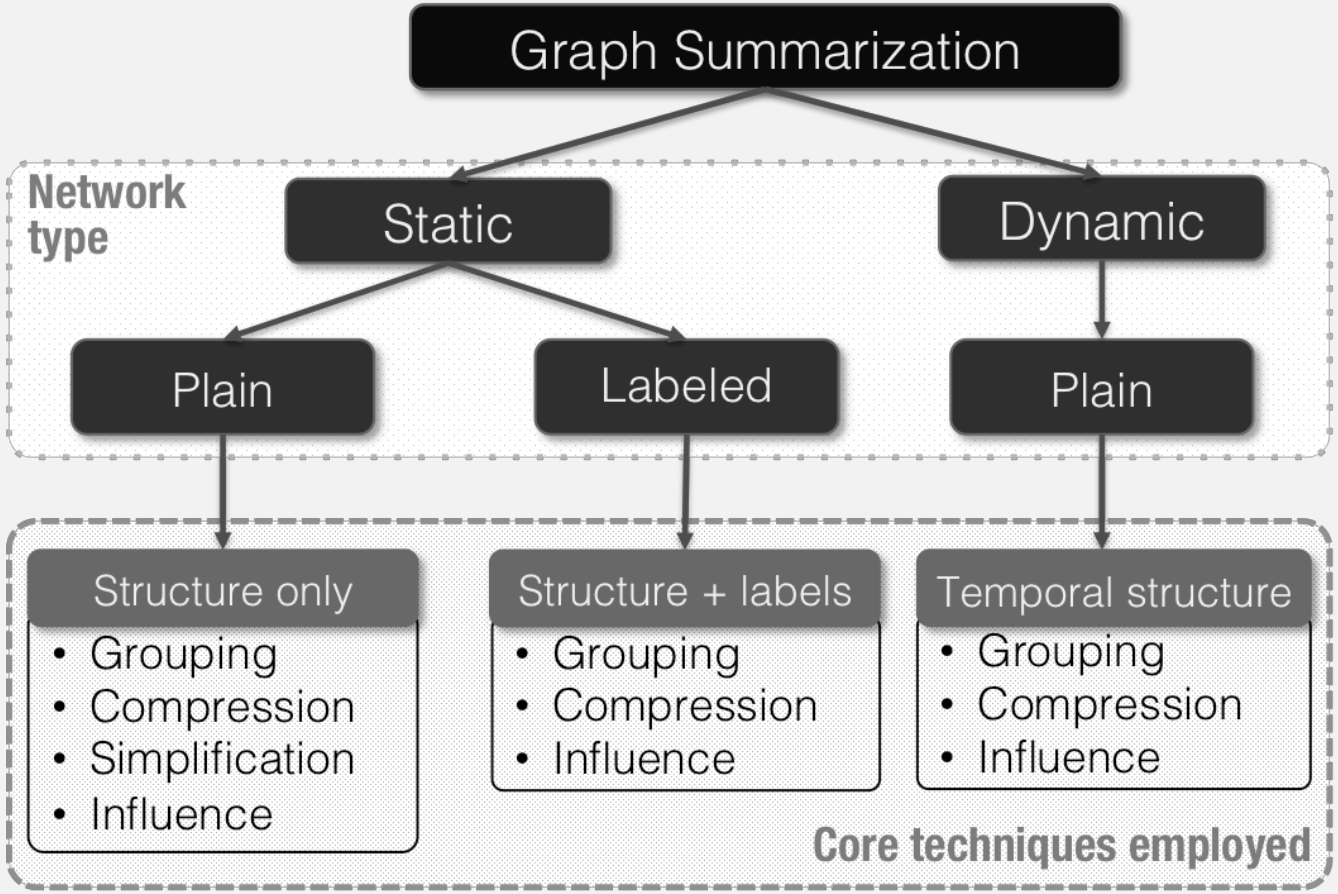
While advances in computing resources have made processing enormous amounts of data possible, human ability to identify patterns in such data has not scaled accordingly. Efficient computational methods for condensing and simplifying data are thus becoming vital for extracting actionable insights. In particular, while data summarization techniques have been studied extensively, only recently has summarizing interconnected data, or *graphs*, become popular. This survey is a structured, comprehensive overview of the state-of-the-art methods for summarizing graph data. We first broach the motivation behind and the challenges of graph summarization. We then categorize summarization approaches by the type of graphs taken as input and further organize each category by core methodology. Finally, we discuss applications of summarization on real-world graphs and conclude by describing some open problems in the field.

CCS Concepts: • **Mathematics of computing** → **Graph algorithms**; • **Information systems** → **Data mining**; **Summarization**; • **Human-centered computing** → *Social network analysis*; • **Theory of computation** → *Unsupervised learning and clustering*; • **Computing methodologies** → *Network science*;

Additional Key Words and Phrases: Graph mining, graph summarization

ACM Reference format:
Yike Liu, Tara Safavi, Abhilash Dighe, and Danai Koutra. 2018. Graph Summarization Methods and Applications: A Survey. *ACM Comput. Surv.* 51, 3, Article 62 (June 2018), 34 pages.
https://doi.org/10.1145/3186727

62

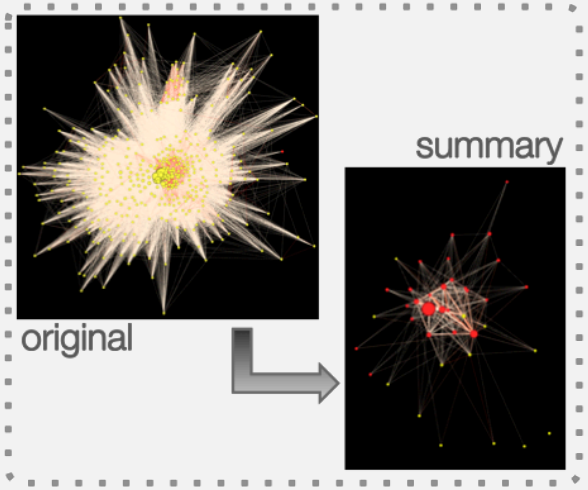


Applications

- Query Efficiency
- Compression
- Visualization
- Pattern Discovery
- Influence Analysis

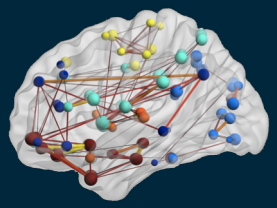
Open Problems

- Richer Data
- Streaming / Incremental Summarization
- Automated Insight Extraction
- Evaluation



Related Research Areas:
graph clustering, partitioning, community detection, sampling, sparsification, sketches, compression

Talk Outline



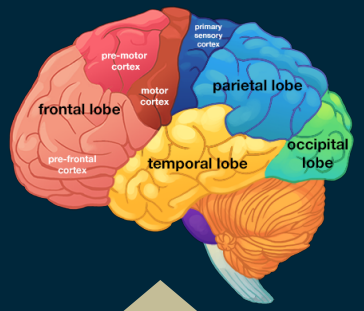
- Network inference from non-network data [ICDM'17, KAIS'18]
- Domain-aware summarization [ICDM'17]
- NN-based summarization for interpretation [KDD'19]

Based on:

- T. Safavi, C. Sripada, D. Koutra. Fast Network Discovery on Sequence Data via Time-Aware Hashing. KAIS'18.
- T. Safavi, C. Sripada and D. Koutra. Scalable Hashing-Based Network Discovery. IEEE ICDM'17
- Di Jin and Danai Koutra. Exploratory Analysis of Graph Data by Leveraging Domain Knowledge. IEEE ICDM'17
- Y. Liu, T. Safavi, A. Dighe, D. Koutra. Graph Summarization Methods and Applications: A Survey. ACM Computing Surveys '18.
- Y. Yan, J. Zhu, M. Duda, E. Solarz, C. Sripada, D. Koutra. GroupINN: Grouping-based Interpretable Neural Network-based Classification of Limited, Noisy Brain Data. ACM KDD'19.

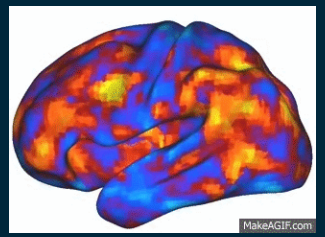
Beyond hand-crafted features

- Can we gain further insights into
  - ✧ which brain regions and interactions between regions are related to the phenotype of interest (e.g diseases, traits)?



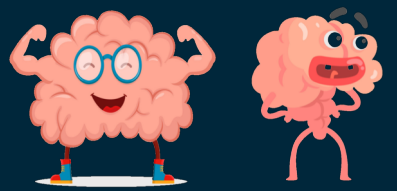
Mapping

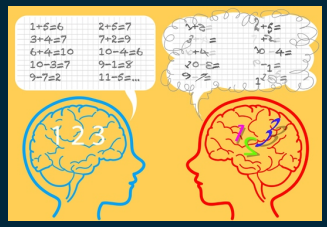
fMRI data



?

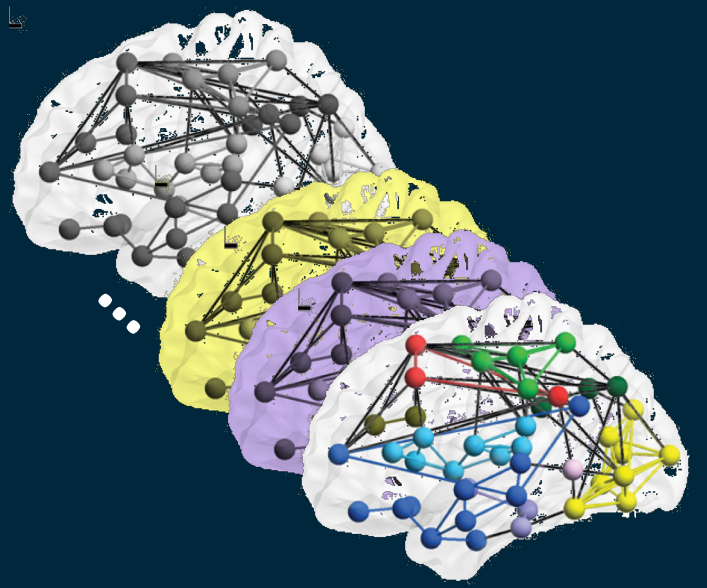
Interactions patterns





https://hopes.stanford.edu/the-hopes-brain-tutorial-text-version/brain-lobes/
https://mappingignorance.org/2017/10/30/numerical-cognition-numbers-brain-plasticity/
https://news.psu.edu/story/349747/2015/03/24/research/more-school-more-challenging-assignments-add-higher-iq-scores

Grouping-based Interpretable NN-based Classification



- Given a set of subjects
  - each with its corresponding brain graph and
  - a label associated with a certain phenotype

- we seek to devise an efficient, interpretable, and parsimonious *neural network* model
  - that can accurately predict each phenotype




M CSE GEMS LAB

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Related Work

- Linear models (PCA, ICA, matrix factorization)
  - \+ Denoising
  - – Fail to capture non-linear interactions between ROIs
- Neural-network models (different variants of GCN)
  - \+ Able to model non-linear interactions between ROIs
  - – Need many training samples
  - – Need many parameters
  - – Long time for training
  - – "Black" box

	Fast	Parsimonious	Interpretable
CNN (KDD'17), GraphCNN (NIPS'16)	✗	✗	✗
GCN (ICLR'17), DGCNN (AAAI'18)	✓	✗	✗
Diffpool (NIPS'18)	✓	✗	inadequate
GroupINN (proposed)	✓	✓	✓

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Related Work

- Linear models (PCA, ICA, matrix factorization)
  - \+ Denoising
  - – Fa
- Neu
  - + Ab
  - – Ne
  - – Ne
  - – Long time for training
  - – “Black” box

Can we build an interpretable NN-based model that is insensitive to noise, parsimonious and able to capture nonlinearities in the prediction task?

			nterpretable
GraphCNN (NIPS’16)	✓	✓	✗
GCN (ICLR’17), DGCNN (AAAI’18)	✓	✗	✗
Diffpool (NIPS’18)	✓	✗	inadequate
GroupINN (proposed)	✓	✓	✓

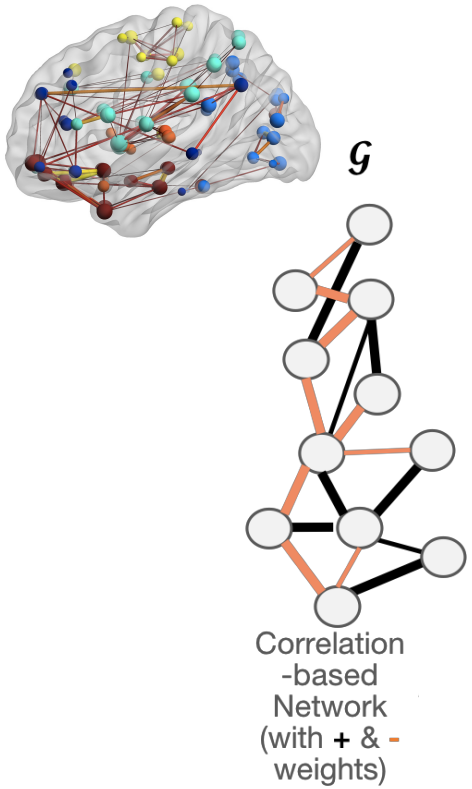
[Yujun Yan, Jiong Zhu, et al. ACM KDD ’19]

Challenges	Proposed Solutions
Noisy fMRI-based brain graphs • Correlation matrices of the same person at the same day differ	Use coarsened brain graphs instead of the (noisier) original graphs
Small samples of high-dim data • A few hundred subjects • 10^4 -10^6 non-0s in the correlation mat.	Dimensionality reduction with supervision
Need for interpretability • Important for driving scientific discoveries (e.g., relation of activation & cognition)	Pinpoint which connections between ROIs are indicative to a specific phenotype

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

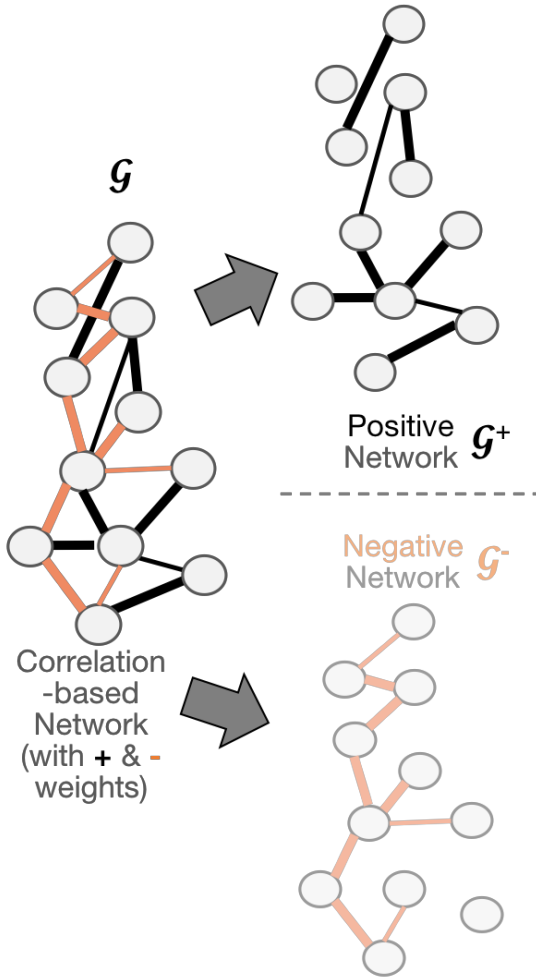
GroupINN Architecture





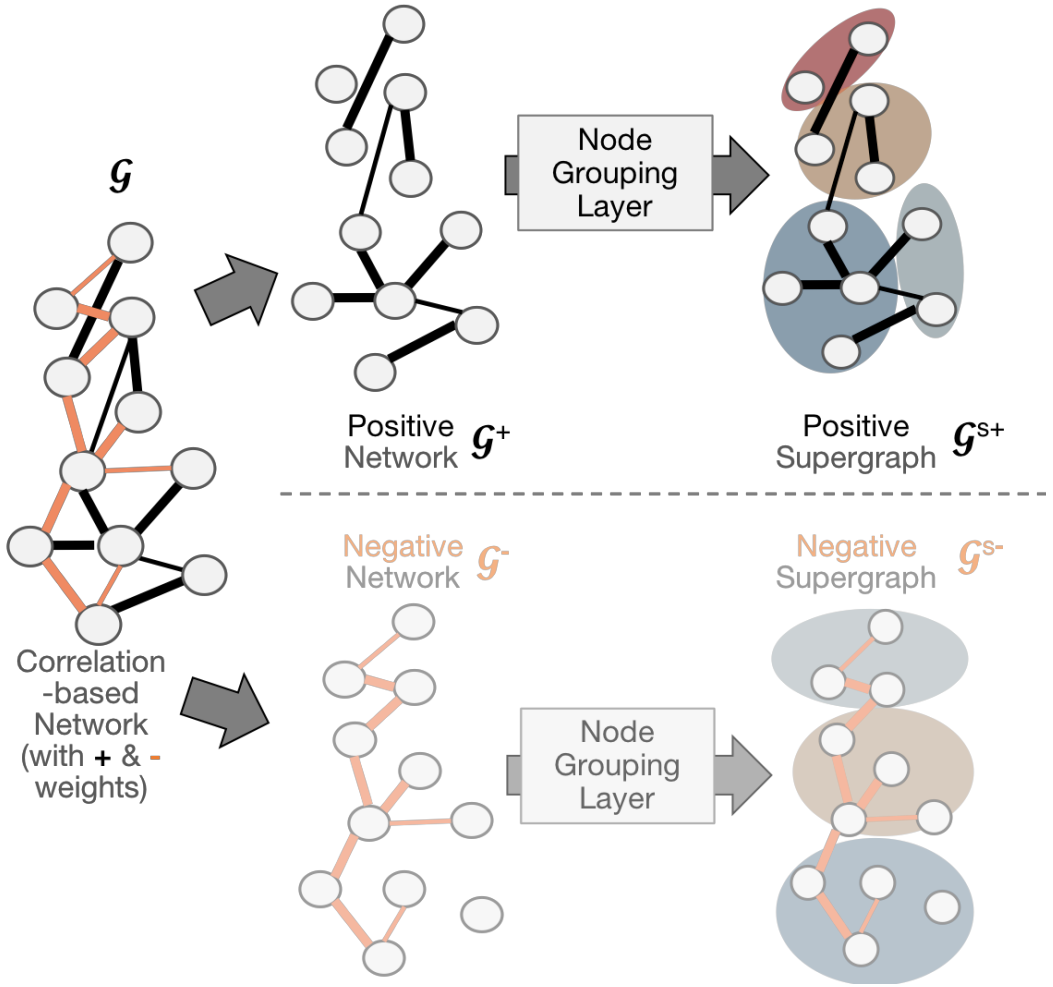
[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

GroupINN Architecture



[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

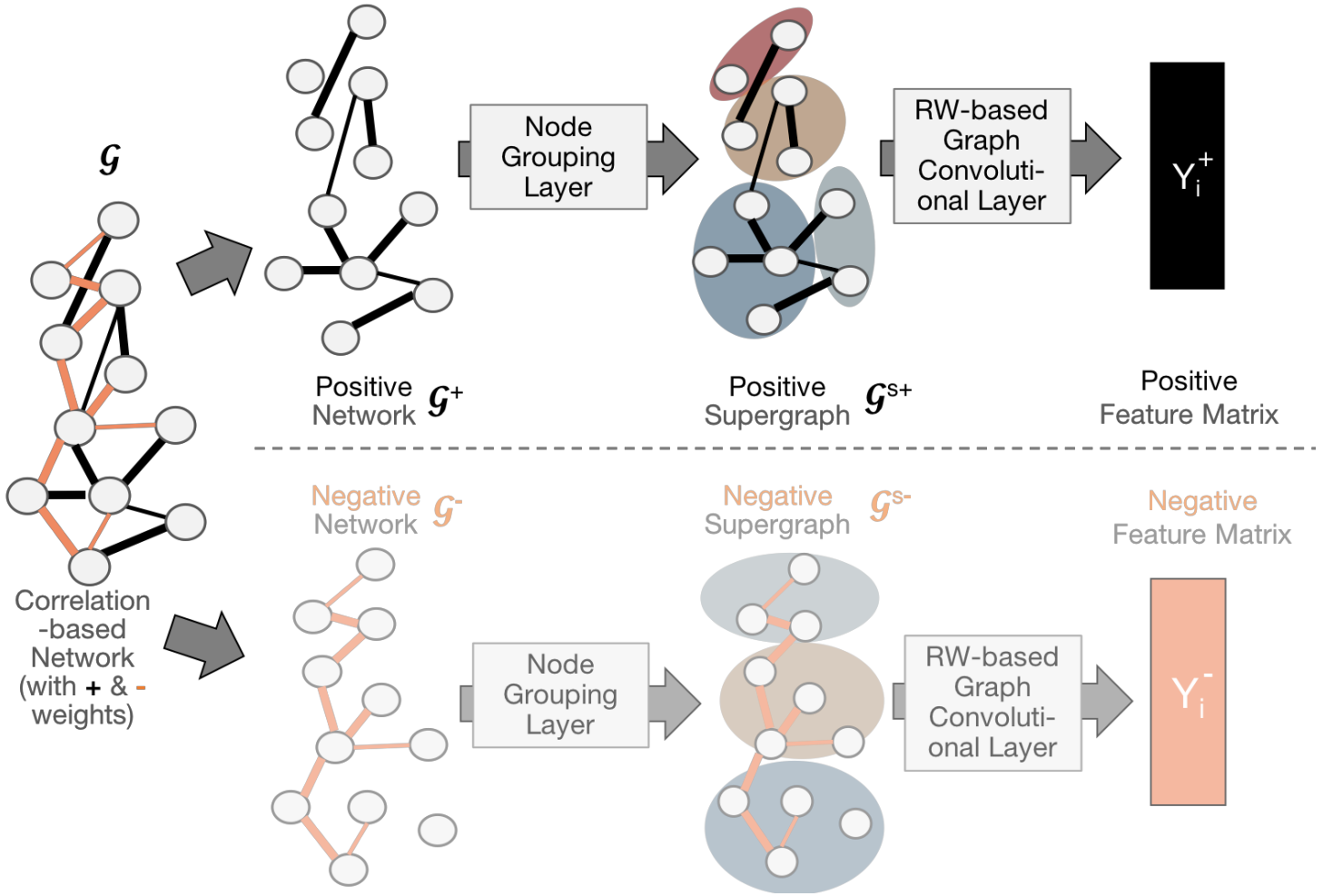
GroupINN Architecture



[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

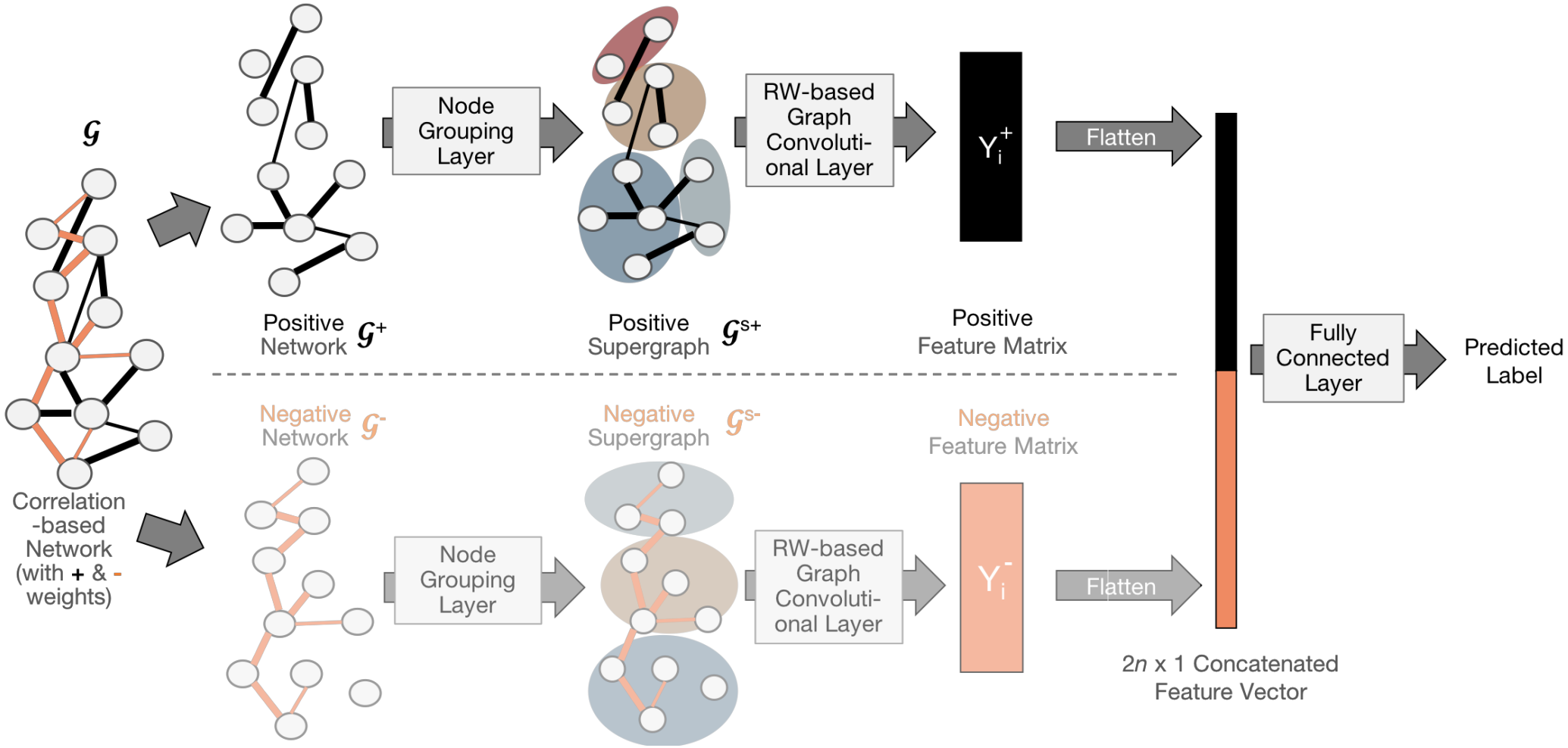
GroupINN Architecture





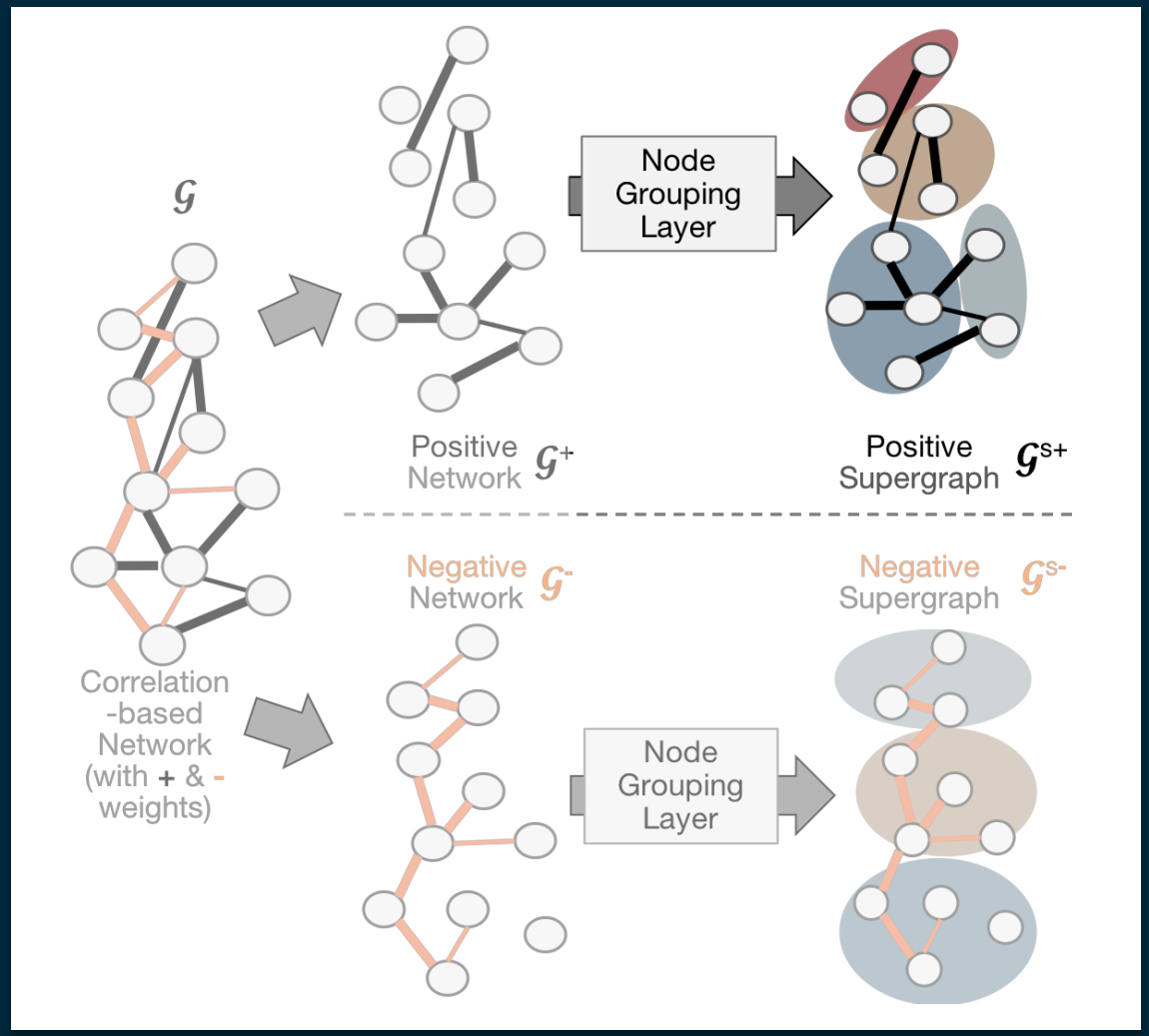
[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

GroupINN Architecture



[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

1. Node Grouping Layer: Intuition



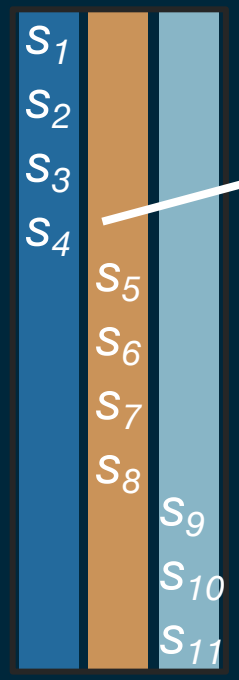
- Recent findings have shown that some ROIs are most related to the phenotype of interest
  → some edges are expected to be more indicative
- Node grouping layer:
  - "hides" the non-indicative edges into a *supernode* and
  - highlights the indicative edges

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

[Cohen, et al. J Neurosci '16]
[Cole, et al. NeuroImage '07]

1. Node Grouping Layer

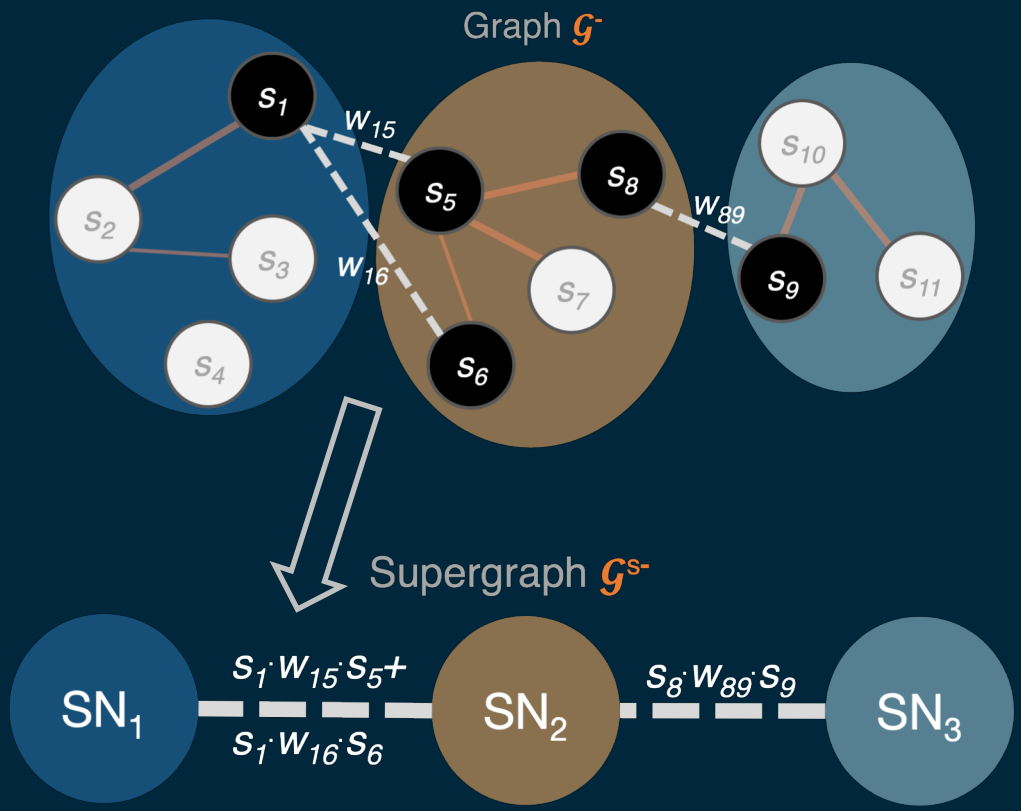
- $\mathbf{F}$: learnable common membership matrix



Real valued importance score of node i in the prediction task

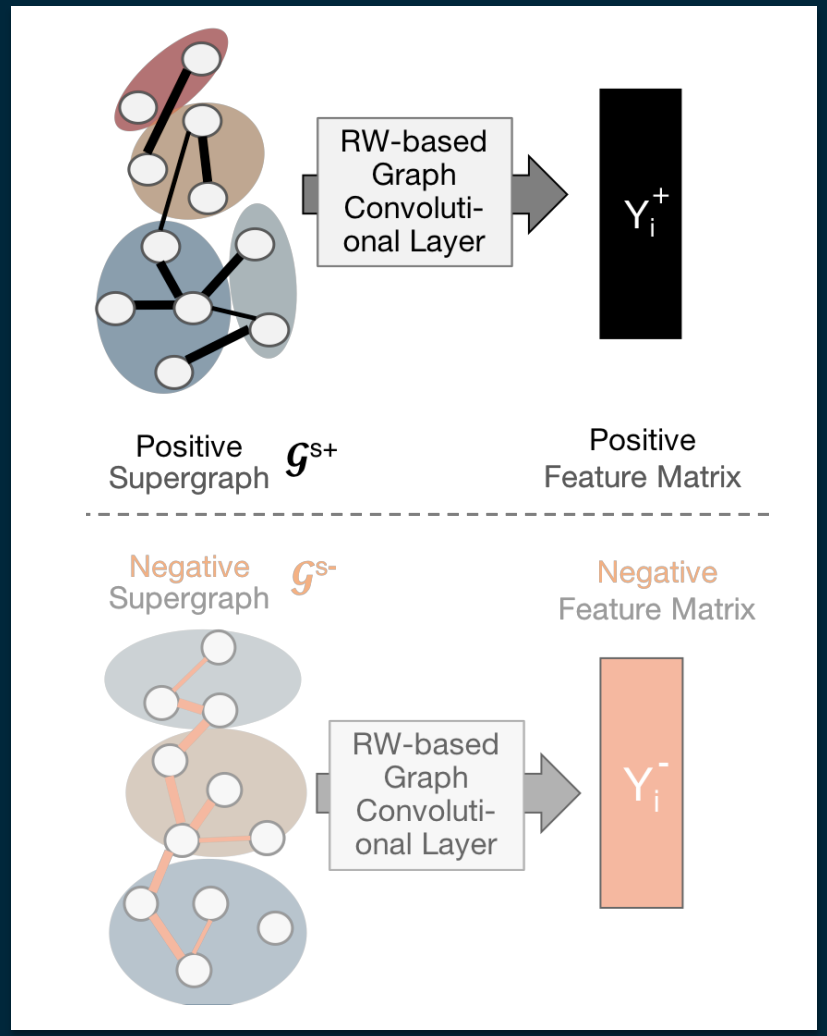
Interpretability

- Nonnegative
- Orthogonal (ideally)
- Nodes in supernode **not** required to be similar/well connected



Adj. of supergraph $\mathbf{W}^S = \mathbf{F}^T\mathbf{W}\mathbf{F}$

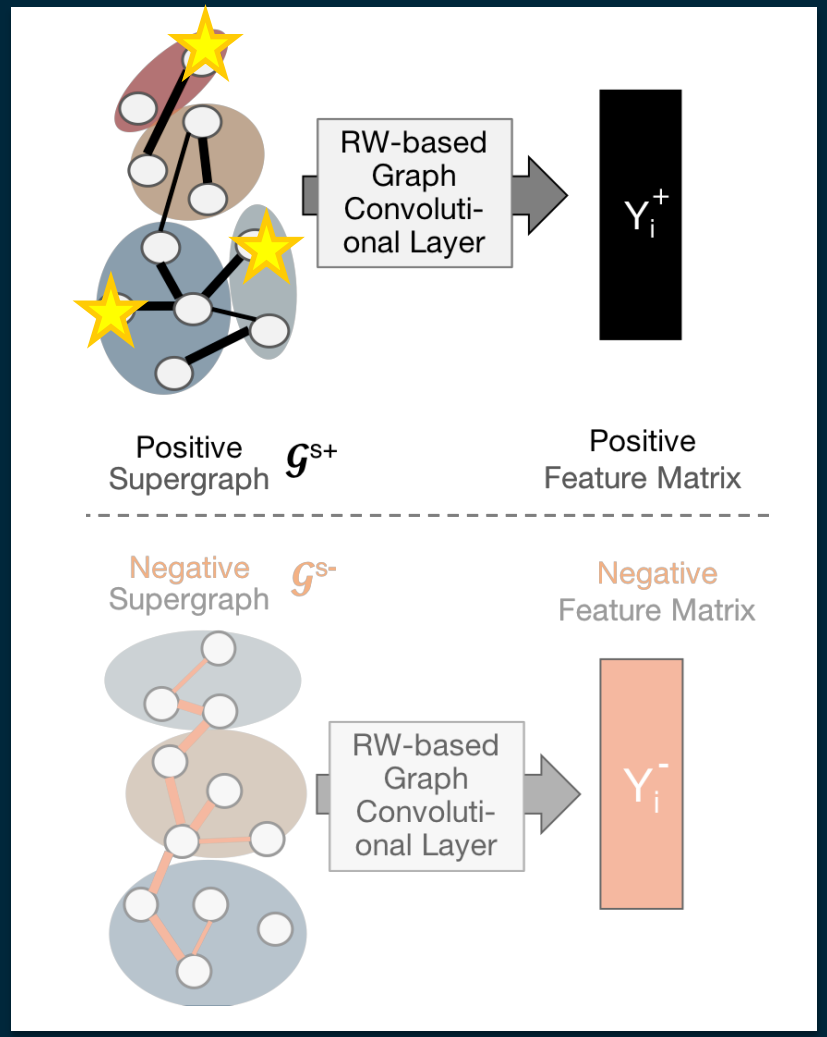
2. RWR-based Graph Conv Layer: Intuition



- Random walks:
  - useful tool to sample graph structure
  - the RWR scores quantify the similarities of other nodes to the selected ones

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

2. RWR-based Graph Conv Layer



Given the seed nodes $\mathbf{q}$, the RWR scores are given by $\mathbf{r} = (1-c)(\mathbf{I} - c\tilde{\mathbf{W}})^{-1}\mathbf{q}$

Restart prob. — Column norm. adj. matrix

The largest eigenvalue of $c\tilde{\mathbf{W}} < 1$, thus:

$$\mathbf{r} = (1-c)(\mathbf{q} + c\tilde{\mathbf{W}}\mathbf{q} + c^2\tilde{\mathbf{W}}^2\mathbf{q} + \cdots)$$

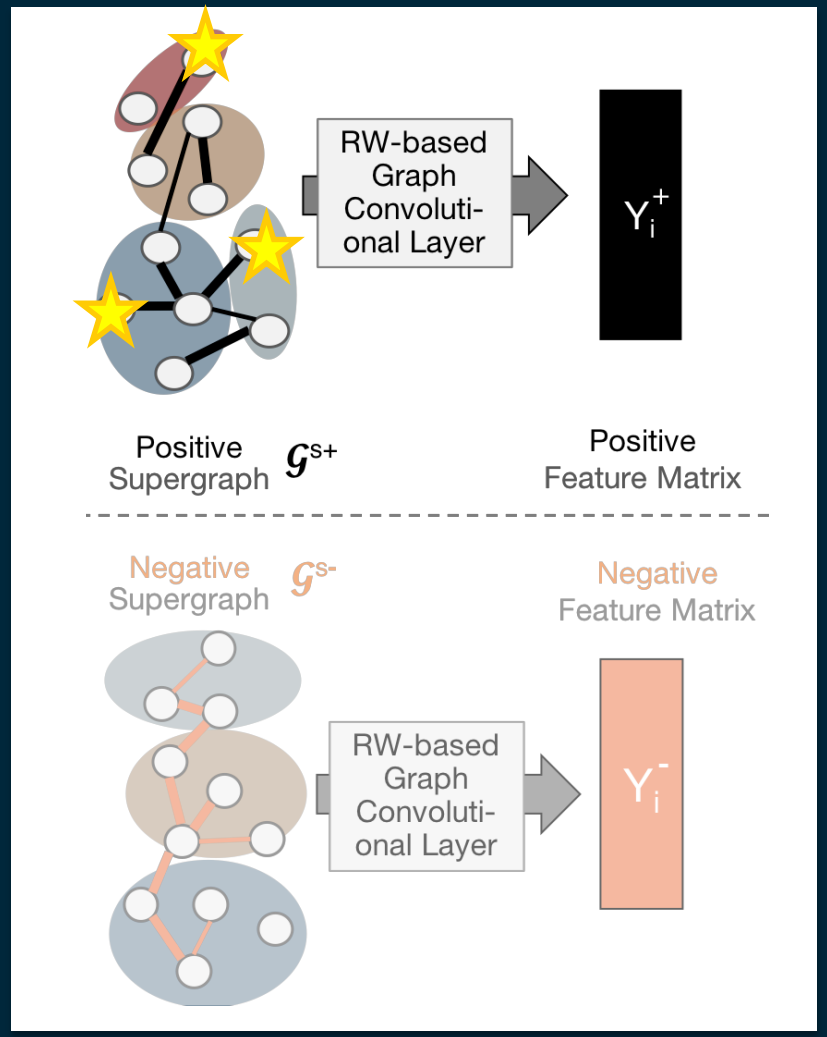
For more structure, multiple $\mathbf{q}$

$$\mathbf{R} = (1-c)(\tilde{\mathbf{Q}} + c\tilde{\mathbf{W}}\tilde{\mathbf{Q}} + c^2\tilde{\mathbf{W}}^2\tilde{\mathbf{Q}} + \cdots)$$

For multiple $\tilde{\mathbf{Q}}$ at different distances

$$\mathbf{M} = (1-c)(\widetilde{\mathbf{Q}_0} + c\tilde{\mathbf{W}}\widetilde{\mathbf{Q}_1} + c^2\tilde{\mathbf{W}}^2\widetilde{\mathbf{Q}_2} + \cdots)$$

2. RWR-based Graph Conv Layer



If the output $\mathbf{Y}_i$ of layer i is: $\mathbf{Y}_i = c\mathbf{W}^s\mathbf{Y}_{i-1}\mathbf{Q}_i + \mathbf{I}$

Supergraph adj.

$$\mathbf{Y}_i = \mathbf{I} + c\mathbf{W}^s\mathbf{Q}_i + c^2\mathbf{W}^{s^2}\mathbf{Q}_i\mathbf{Q}_{i-1} + \cdots$$

$$\sim \mathbf{M} = (1-c)\left(\widetilde{\mathbf{Q}_0} + c\widetilde{\mathbf{W}}\widetilde{\mathbf{Q}_1} + c^2\widetilde{\mathbf{W}}^2\widetilde{\mathbf{Q}_2} + \cdots\right)$$

- Design:

Adding nonlinearity, the output $\mathbf{Y}_i$ of layer i is:

$$\mathbf{Y}_i = \sigma(c\mathbf{W}^s\mathbf{Y}_{i-1}\mathbf{Q}_i + \mathbf{I})$$

Nonlinear function

e.g. relu

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

https://github.com/GemsLab/GroupINN

Constraint	Loss

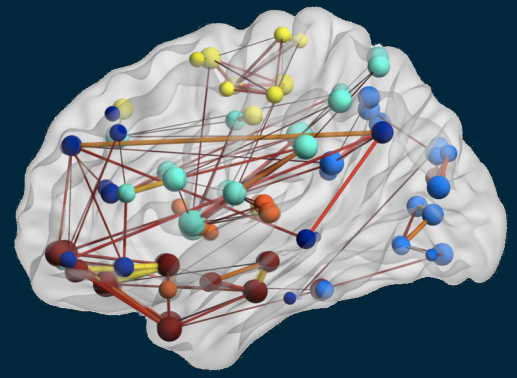
https://github.com/GemsLab/GroupINN

Constraint	Loss	
Membership matrix $\mathbf{F}$ is non-negative	Thresholding real matrix $\tilde{\mathbf{F}}$	$\mathrm{Relu}(\tilde{\mathbf{F}})$
Less overlap between supernodes	Orthogonal penalty on matrix $\mathbf{F}$	$\|\|\mathbf{F}^T\mathbf{F} - diag(diag(\mathbf{F}^T\mathbf{F})\|\|_F$
Balanced clusters	Variance penalty on the diagonal elements of $\mathbf{F}^T\mathbf{F}$	$\mathrm{Var}(\mathrm{diag\_part}(\mathbf{F}^T\mathbf{F}))$
Seed nodes should have positive weights	Negative penalty on real matrix $\mathbf{Q}_i$	$\mathrm{Sum}(\mathrm{Relu}(-\mathbf{Q}_i))$
Prevent overfitting	L_2 penalty on dense layer	L_2

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Data: HCP 1200 release

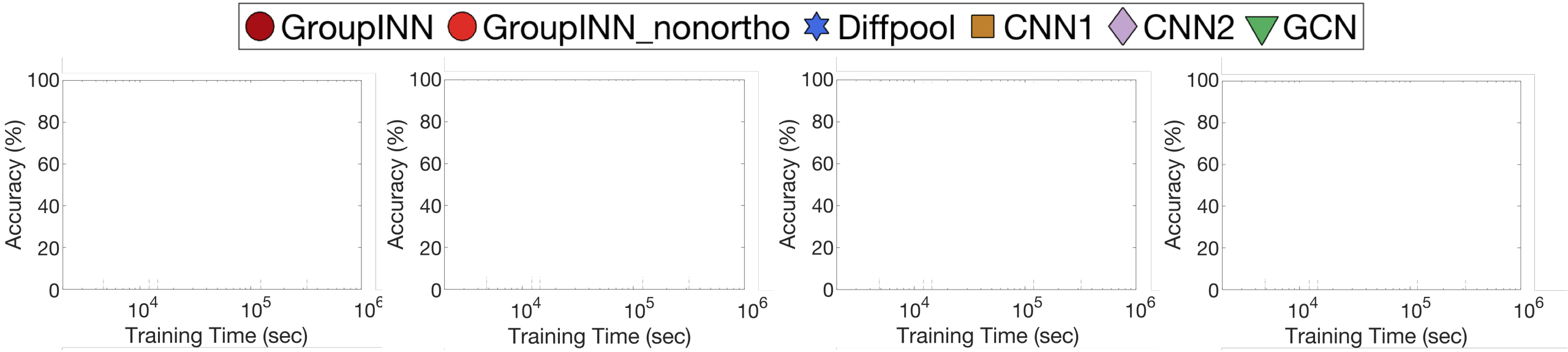
- 446-448 subjects
- 264 ROIs per subject
  - 176-405 time points (depending on the task)
- Tasks
  - Working memory (WM) – 0-back and 2-back
  - Emotion
  - Gambling
  - Social
- Prediction (using SVM with RBF kernel)
  - General Executive Factor (GenExec), a measure of general intellectual ability



[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Q1. Comparison with *NN*-based methods

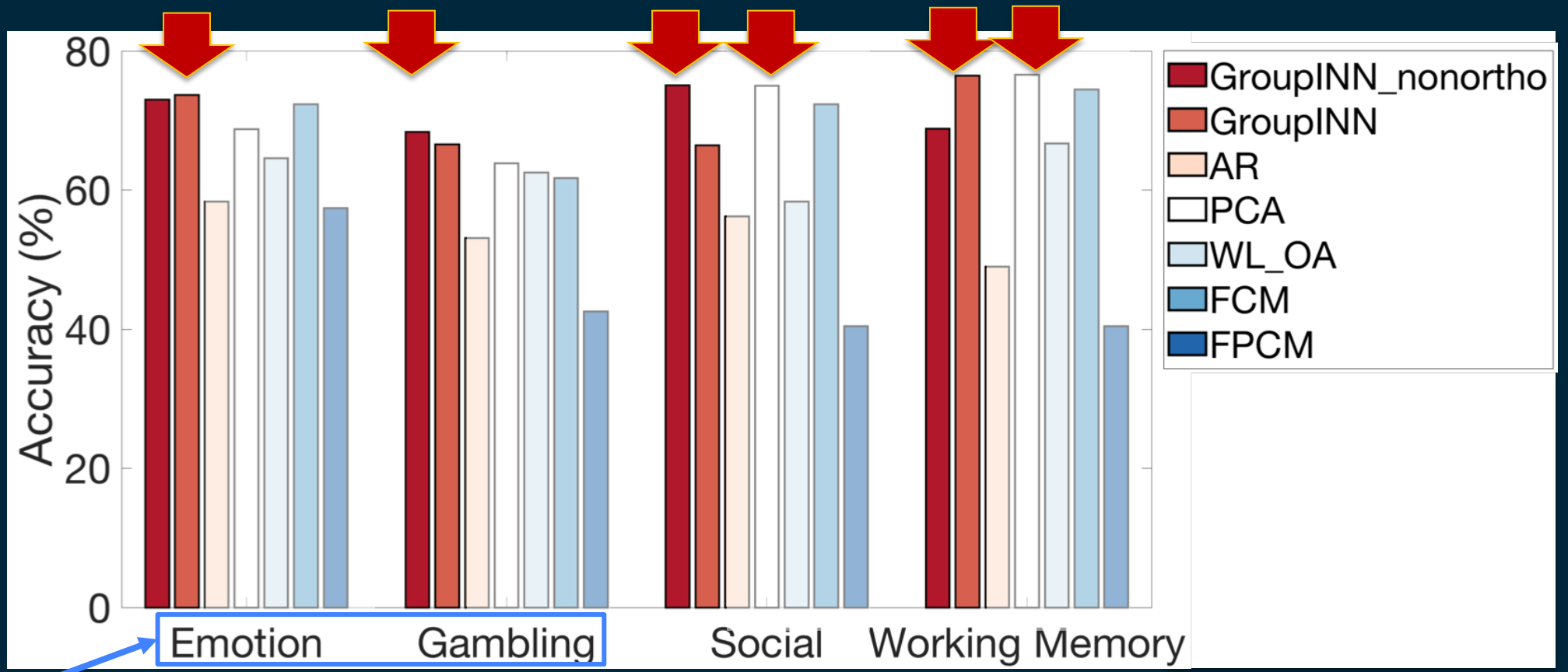
w/o orthogonality Ying, NeurIPS' 18 Wang, KDD' 17 Kipf, ICLR' 17



GroupINN models are up to 69× faster at training than all the baseline methods, while achieving same or higher accuracy in a variety of prediction tasks.

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Q1. Comparison with *non*-NN methods



Not directly related to GenExec

Our two variants (with and without orthogonality) have better or comparable accuracy.

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Q2. Parsimony of GroupINN

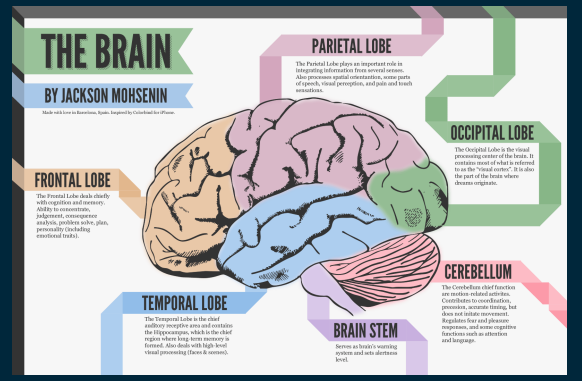
Less is better!

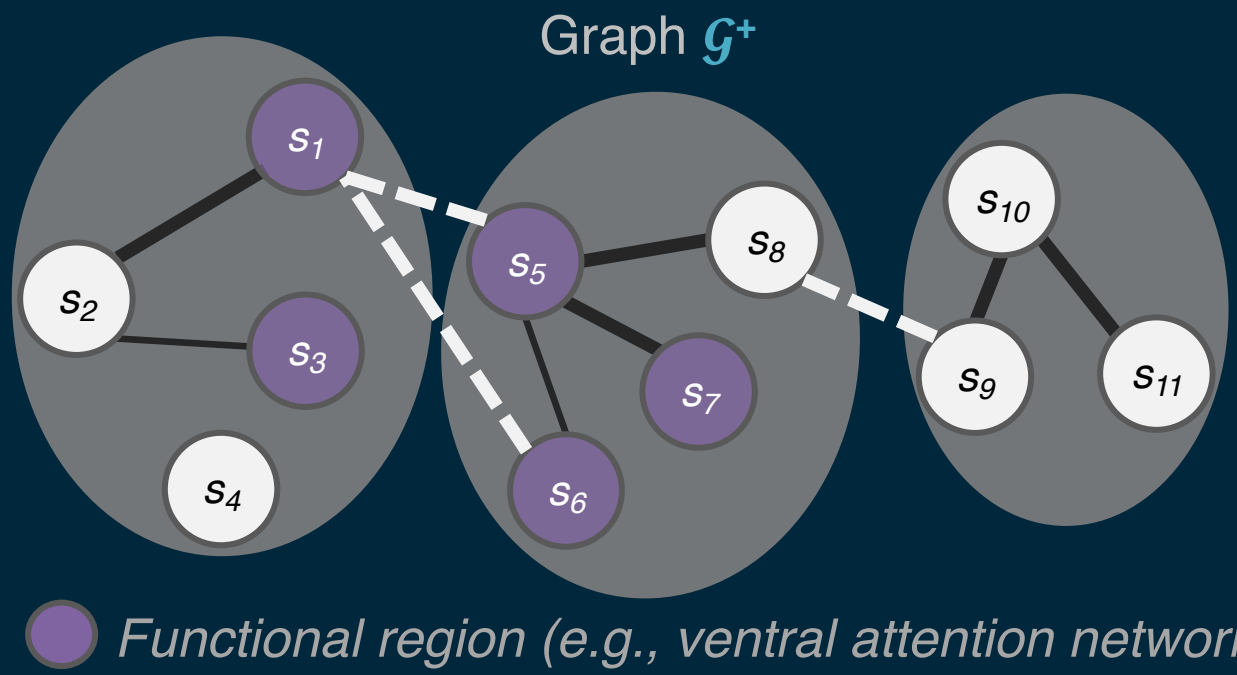
Methods	# parameters	Normalized wrt GroupINN
CNN-1		
CNN-2		
GCN		
Diffpool		
GroupINN	2,892	1×

GroupINN can use 15% or much fewer model parameters to achieve comparable or better performance of the baseline methods.

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Q3. Interpretability





Importance score of subnetwork $\mathcal{R}$:

$$S_{\mathcal{R}} = \frac{2}{|\mathcal{R}|^2} \sum_{\substack{\text{i,j} \in \mathcal{R} \text{ and} \\ c(i) \neq c(j)}} s_i s_j$$

- PCA: average weights in the first principal component
- Diffpool: similar to GroupINN, but the scores are averaged across different subjects

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Q3. Interpretability

Acronyms of brain subnetworks. AN: auditory; CBLN: cerebellar; CON: cingulo-opercular; DAN: dorsal attention; FPN: frontoparietal; MRN: memory retrieval; SN: salience ; VAN: ventral attention ; VN: vision; SM.M: sensory/somatomotor mouth; SM.H: sensory/somatomotor hand

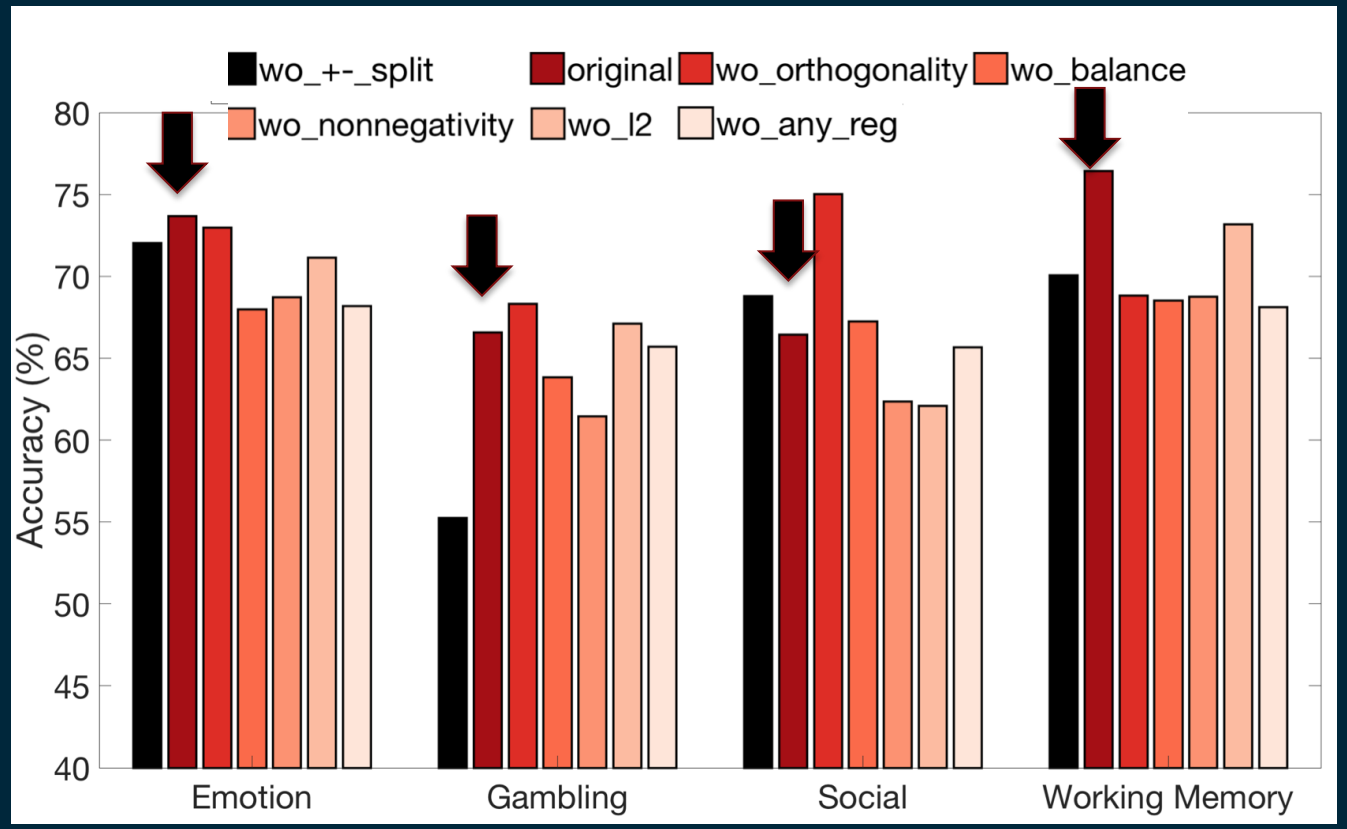
Tasks	Within subnetworks								
	GroupINN			PCA			Diffpool		
Working Memory	MRN	FPN	SN	SM.M	SM.H	AN	SM.M	MRN	CBLN
Gambling	VAN	VN	DAN	SM.H	AN	SM.M	SM.M	FPN	MRN
Emotion	SN	CON	VAN	SM.M	SM.H	AN	DMN	MRN	SM.M
Social	FPN	SN	VAN	SM.M	CBLN	AN	DAN	SM.M	FPN

- GroupINN find the most task-positive sub-networks.
- PCA and Diffpool are misled by strong noisy signals from SM.M and SM.H.

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

[Cohen, et al. , J Neurosci 2016];
[Cole, et al. Neuron 2014];
[Davison, et al. PLOS Comp Bio 2015]

Q4. Impact of network splitting and regularization terms



- Splitting the network into positive and negative sub-networks helps.
- The various loss functions contribute to higher accuracy.

[Yujun Yan, Jiong Zhu, et al. ACM KDD '19]

Conclusion: inference, summarization, and interpretation

- Pipeline for network discovery on time series [ICDM'17, KAIS'18]
  - ABC: time-consecutive similarity measure + metric on binary sequences
  - Associated LSH family
  - Modular & applicable in other settings, fast + accurate
  - Impact: integrated into production systems

Google

- Domain-aware summarization: one summary does not fit all
  - Summarize one graph wrt multiple baseline graphs

- NN-based approach to interpret real NNs [KDD'19]
  - Fast, parsimonious, interpretable
  - Up to 69x less training time
  - Impact: Insights into brain subnetworks

Explore Graduate Studies in CSE, University of Michigan

- Participants learn about the grad school and the application process
  - how to prepare their applications
  - MS vs. PhD
  - Career paths

- Saturday, October 12, 2019
- Deadline: July 15

M CSE GEMS LAB

2019 WORKSHOP

Students from all institutions are invited to apply to attend this one-day workshop. Participants will learn about the graduate school application process and the opportunities that exist for those who pursue graduate work in this impactful discipline. Travel awards are available for non-local students.

WORKSHOP HIGHLIGHTS

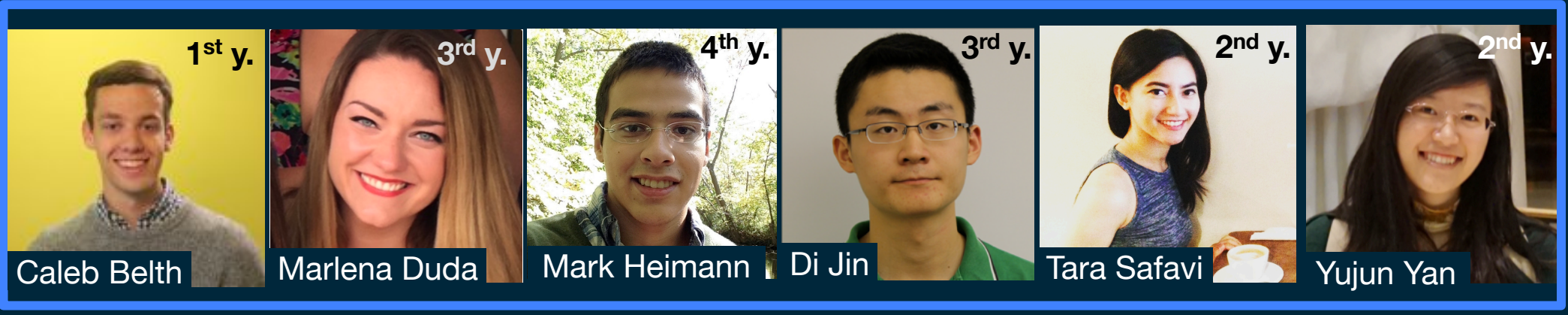
- "Life as a PhD student" – Q&A panel of current graduate students
- Live demos of current and ongoing research
- Lunch with faculty and graduate students
- "Academia vs. Industry" – Q&A panel with current faculty and industry partners
- A writing clinic on "How to Write a Strong Statement of Purpose"
- The opportunity to receive one-on-one feedback from faculty regarding your application

Apply today at: bit.ly/ExpGrad2019

M | CSE COMPUTER SCIENCE AND ENGINEERING UNIVERSITY OF MICHIGAN

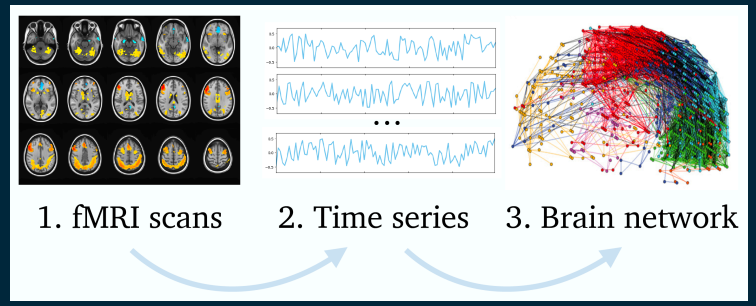
cse.umich.edu/egs

GEMS LAB

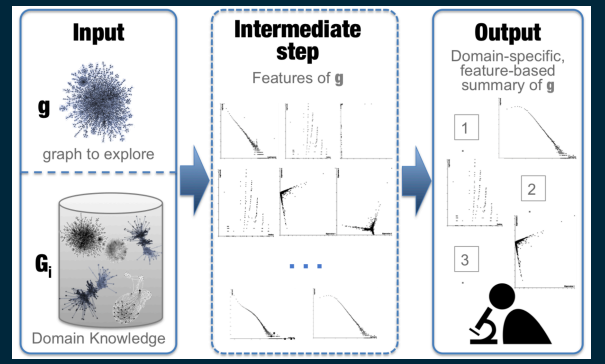


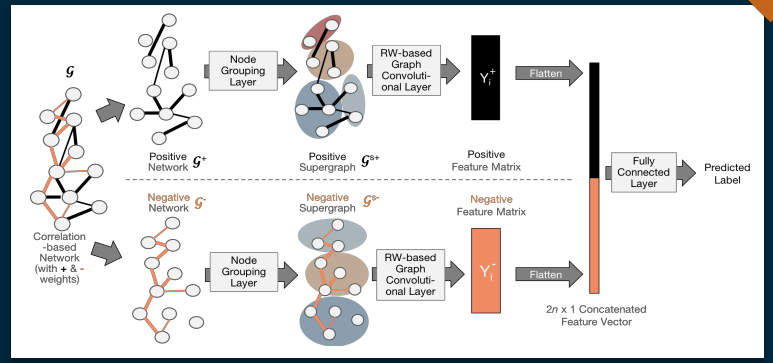
Thank you!
Questions?

http://danaikoutra.com
dkoutra@umich.edu



Inference,
Summarization,
Interpretation





https://github.com/tsafavi/hashing-based-network-discovery
https://github.com/DerekDiJin/Domain_Knowledge

https://github.com/GemsLab/GroupINN

NSF ARL TROVE MIDAS Adobe Microsoft Azure amazon