TEXAS
The University of Texas at Austin

Stay on path: PCA along graph paths

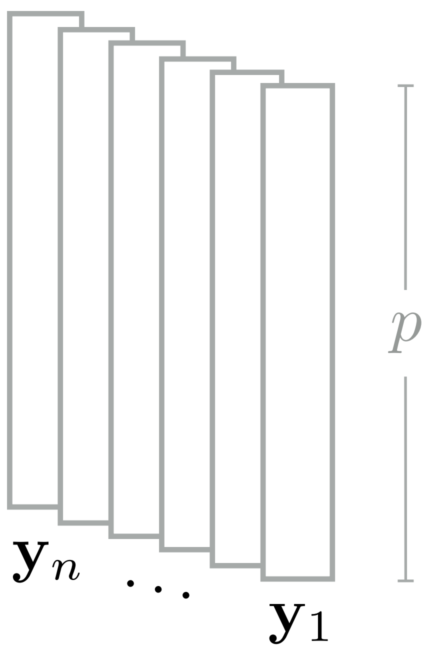
Megasthenis Asteris
Anastasios Kyrillidis
Alexandros Dimakis

Electrical and Computer Engineering

Han - Gyol Yi
Bharath Chandrasekaran

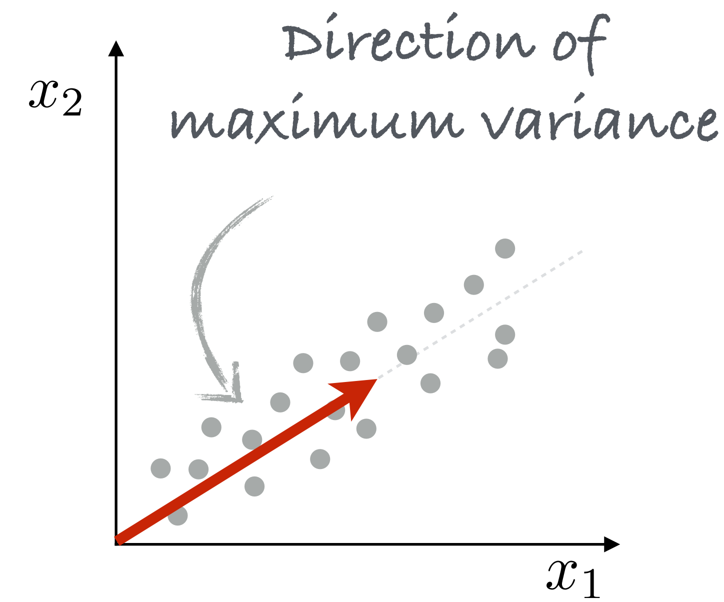
Communication Sciences and Disorders

Sparse PCA

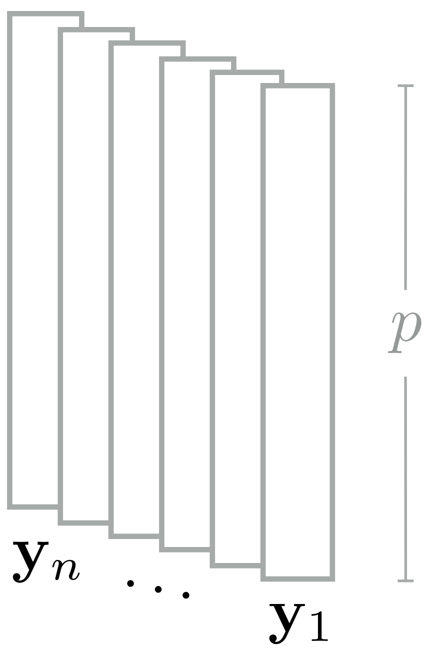


n observations / datapoints
p variables

Find new variable (feature) that captures most of the variance.

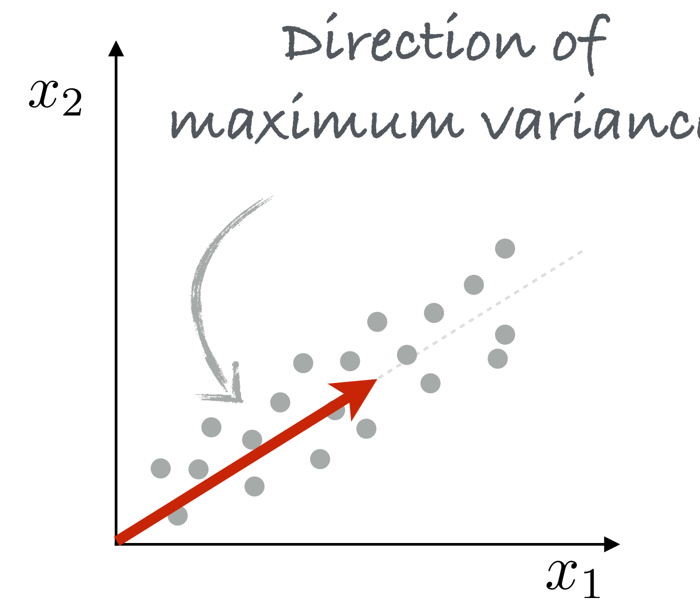


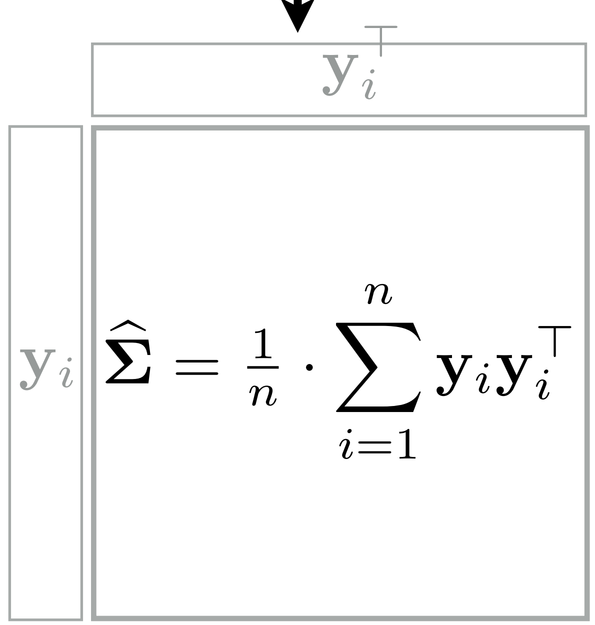
Sparse PCA



n observations / datapoints
p variables

Find new variable (feature) that captures most of the variance.

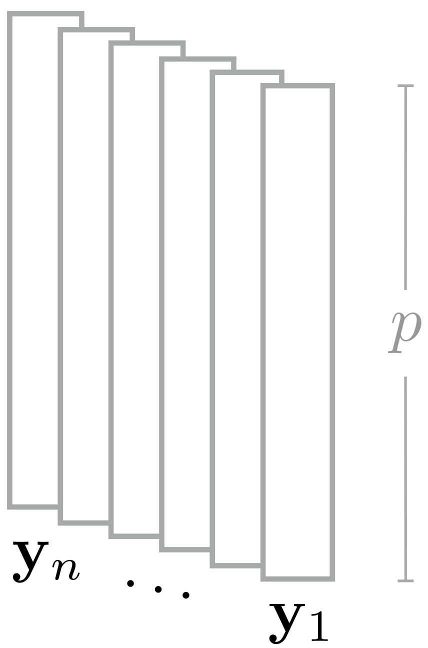




Empirical cov. matrix

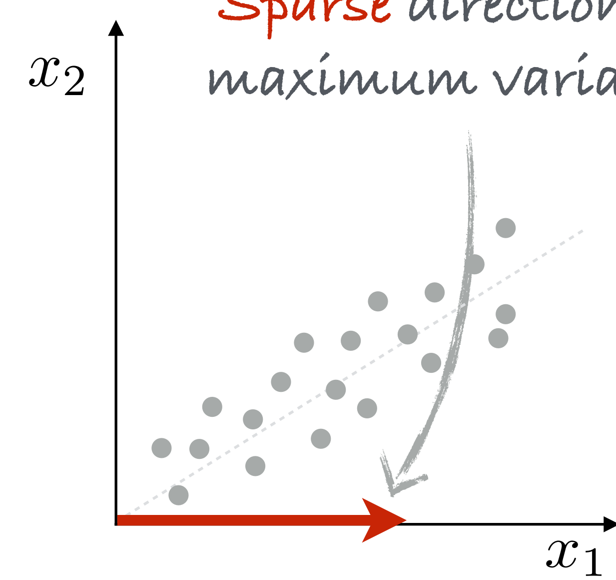
$$\begin{array}{ll} \max\limits_{\mathbf{x}} & \mathbf{x}^\top \widehat{\mathbf{\Sigma}}\, \mathbf{x} \\ \text{subject to} & \|\mathbf{x}\|_2 = 1 \end{array}$$

Sparse PCA

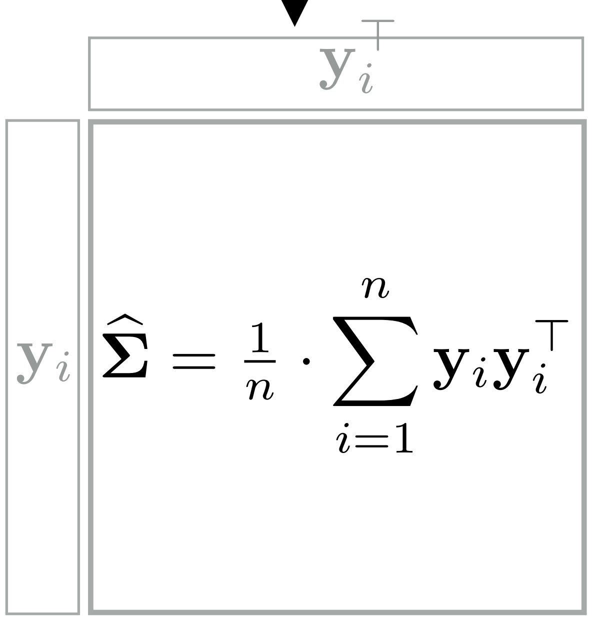


n observations / datapoints
p variables

Find new variable (feature) that captures most of the variance.



Empirical cov. matrix



$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i^\top$$

NP-Hard

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^\top \widehat{\boldsymbol{\Sigma}}\, \mathbf{x} \\ \text{subject to} \quad & \|\mathbf{x}\|_2 = 1 \\ & \|\mathbf{x}\|_0 = k \end{aligned}$$

Sparse PCA

Why sparsity?

- [Engineer] Extracted feature is more ***interpretable***; it depends on only a few original variables.
- [Statistician] Recovery of “true” PC in high dimensions; # observations << # variables.

Sparse PCA

Why sparsity?

- [Engineer] Extracted feature is more *interpretable*; it depends on only a few original variables.
- [Statistician] Recovery of “true” PC in high dimensions; # observations << # variables.

More structure…?

- More *interpretable*.
- Better sample complexity.

E.g. wavelets of natural images, block structures, periodical neuronal spikes, …

[Baraniuk et al., 2008; Kyrillidis et al., 2014, Friedman et al., 2010, …]

Sparse PCA

Why sparsity?

→ [Engineer] Extracted feature is more *interpretable*; it depends on only a few original variables.

→ [Statistician] Recovery of “true” PC in high dimensions; # observations << # variables.

More structure…?

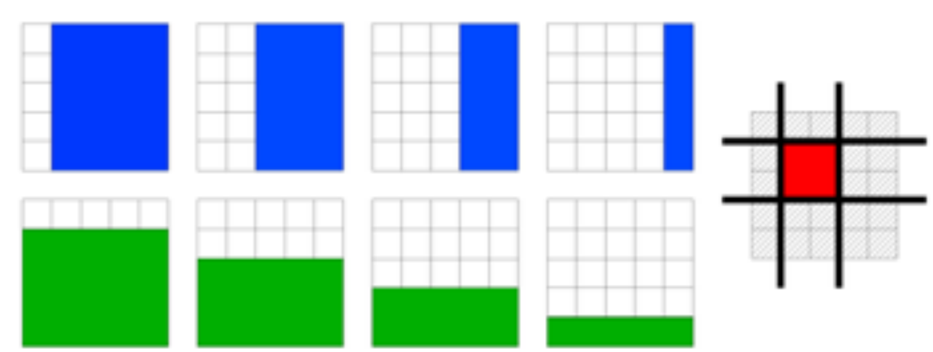
→ More *interpretable*.

→ Better sample complexity.

E.g. wavelets of natural images, block structures, periodical neuronal spikes, …

[Baraniuk et al., 2008; Kyrillidis et al., 2014, Friedman et al., 2010, …]

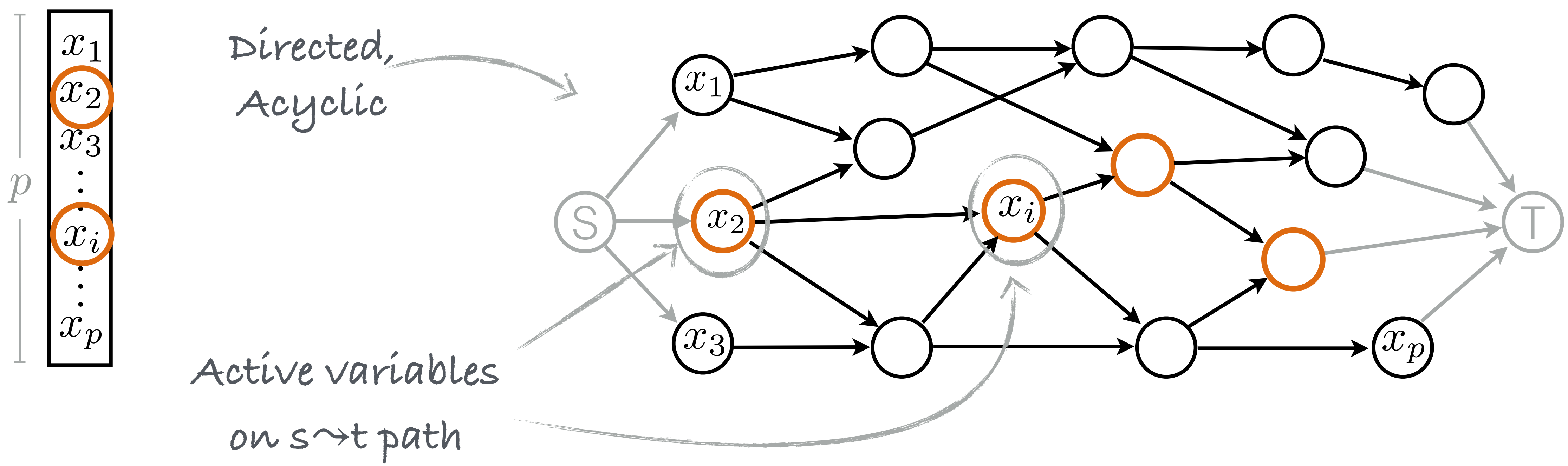
- Structured sparse PCA [Jenatton et al., 2010]
  - Sparsity-inducing norm
  - 2D grid, rectangular nonzero patterns



[PCA On Graph Paths]

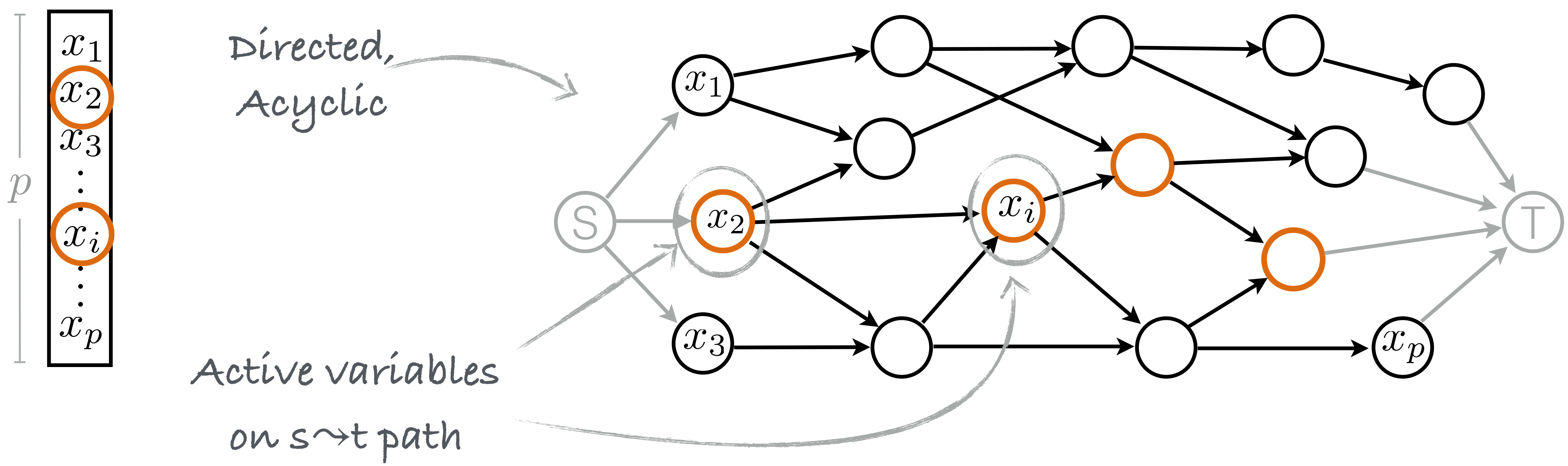
Problem Definition

- **Structure captured by an underlying graph.**



Problem Definition

- **Structure captured by an underlying graph.**



Graph Path PCA

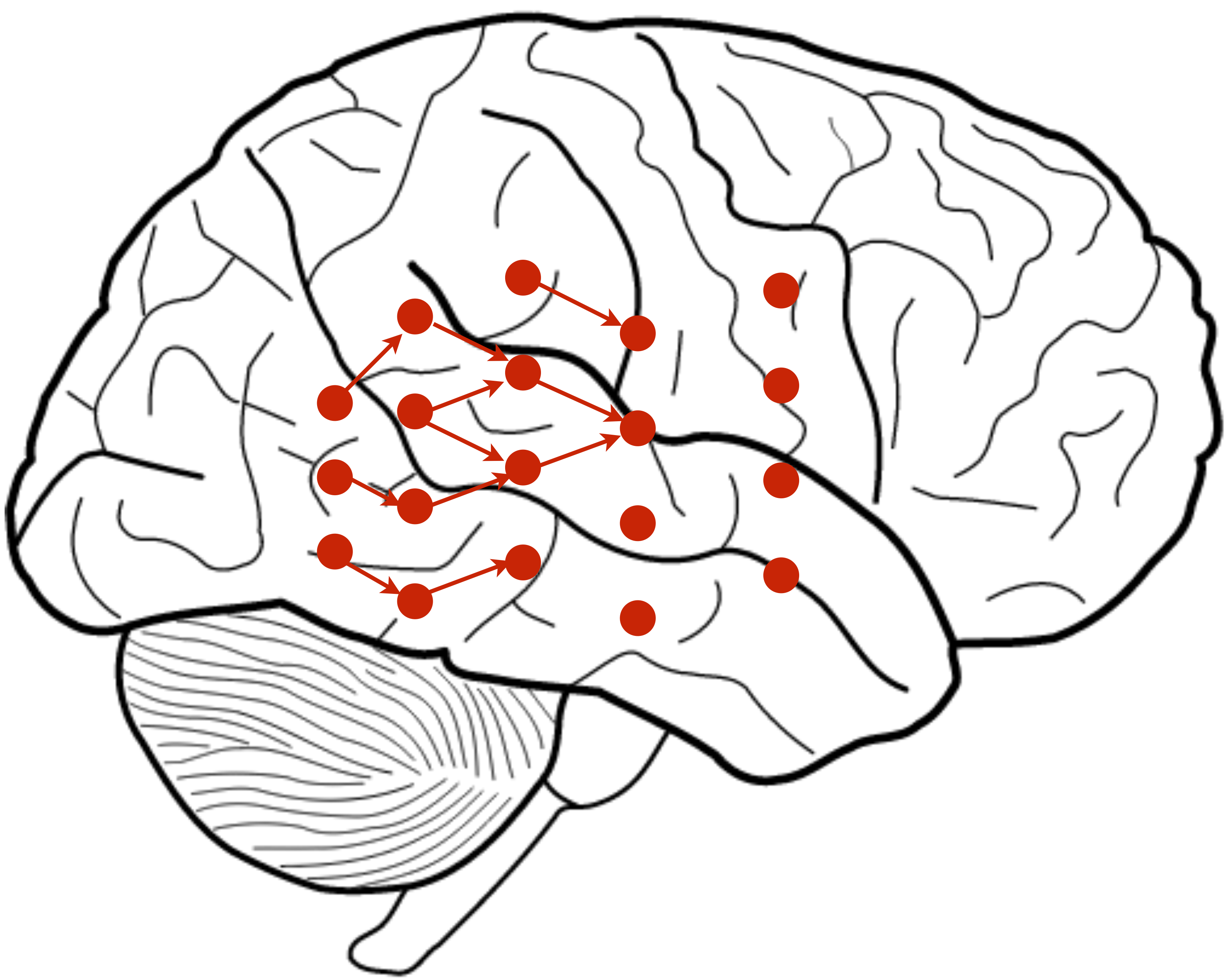
$$\max_{\mathbf{x}} \quad \mathbf{x}^\top \widehat{\boldsymbol{\Sigma}}\, \mathbf{x}$$

$$\text{subject to} \quad \mathbf{x} \in \mathcal{X}(G)$$

$$\mathcal{X}(G) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_2 = 1, \ \text{supp}(\mathbf{x}) \subset \text{path of } G\}$$

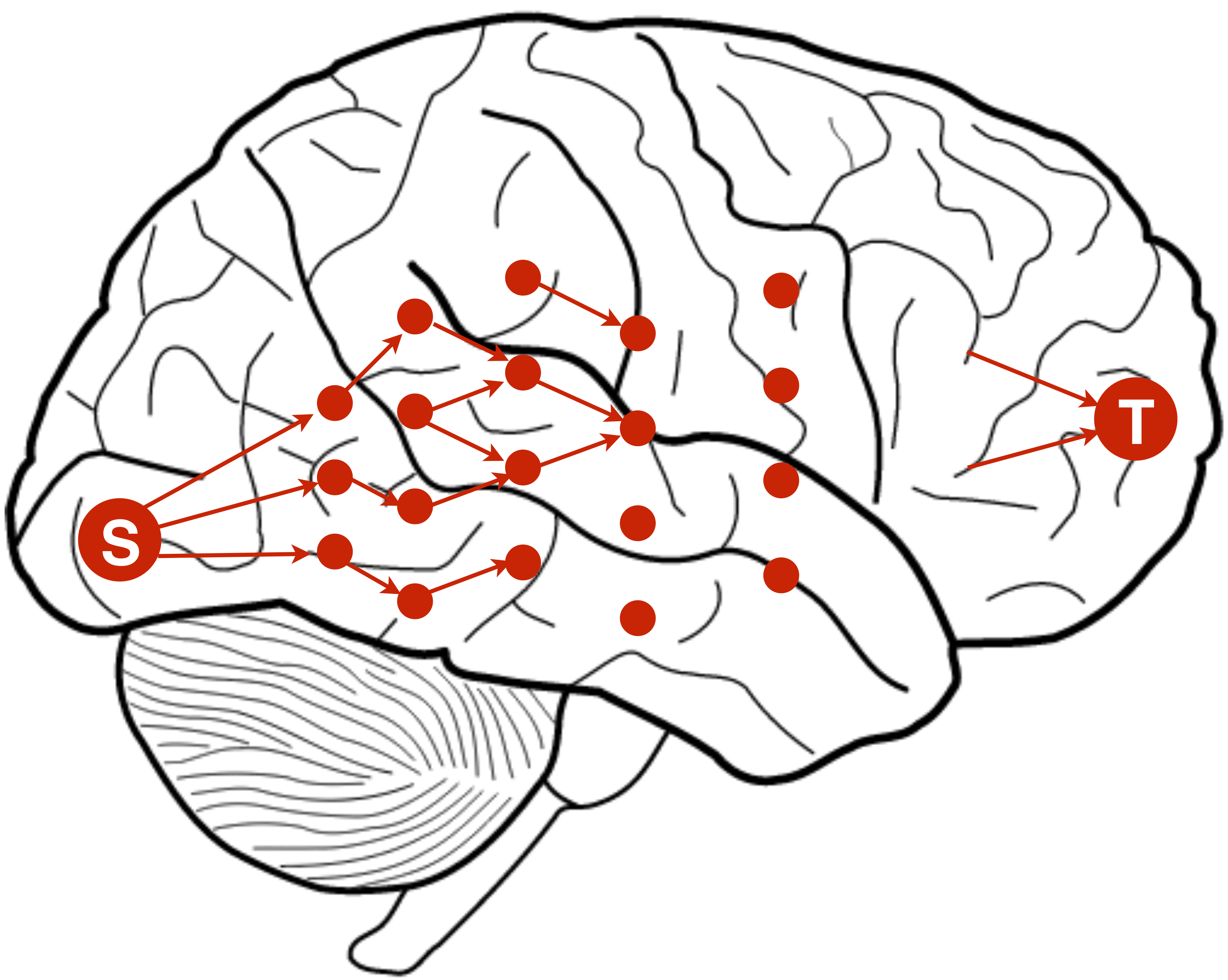
Motivation 1: Neuroscience

- Variables: “***voxels***” (points in the brain)
- Measurements: blood-oxygen levels



Motivation 1: Neuroscience

- Variables: “***voxels***” (points in the brain)
- Measurements: blood-oxygen levels



Motivation 2: Finance

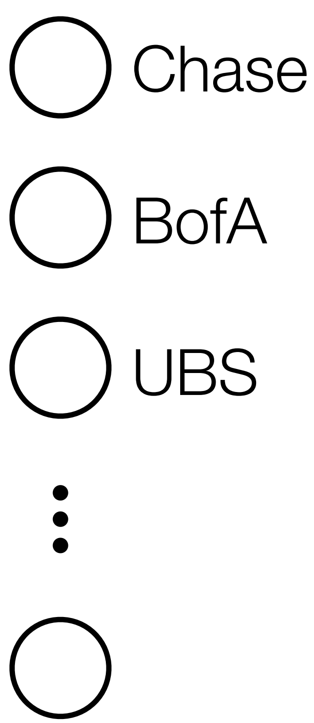
- Variables: **stocks**
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance

Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector

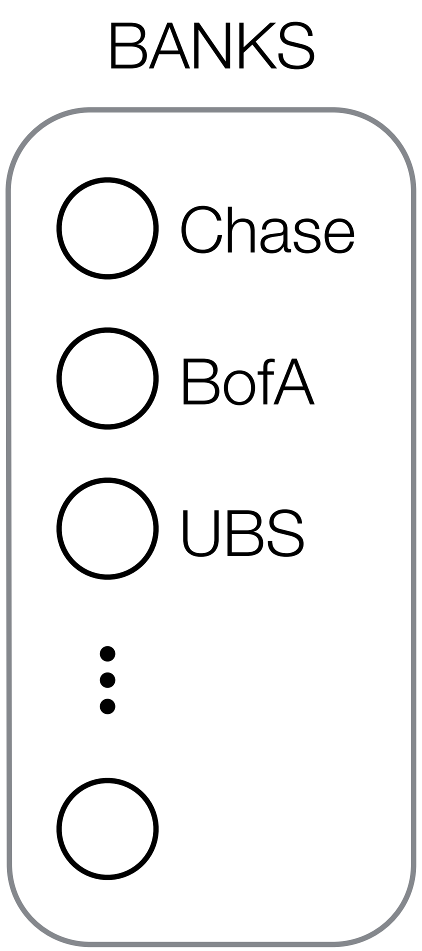
Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



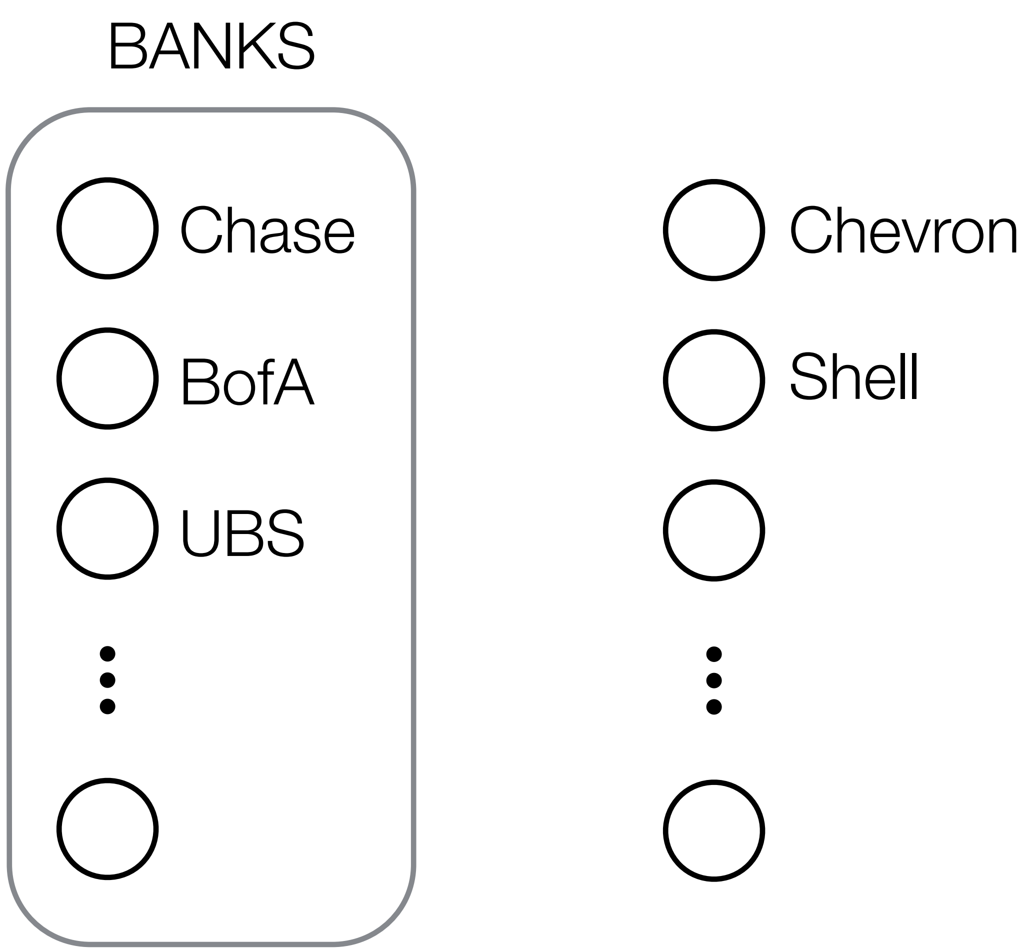
Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



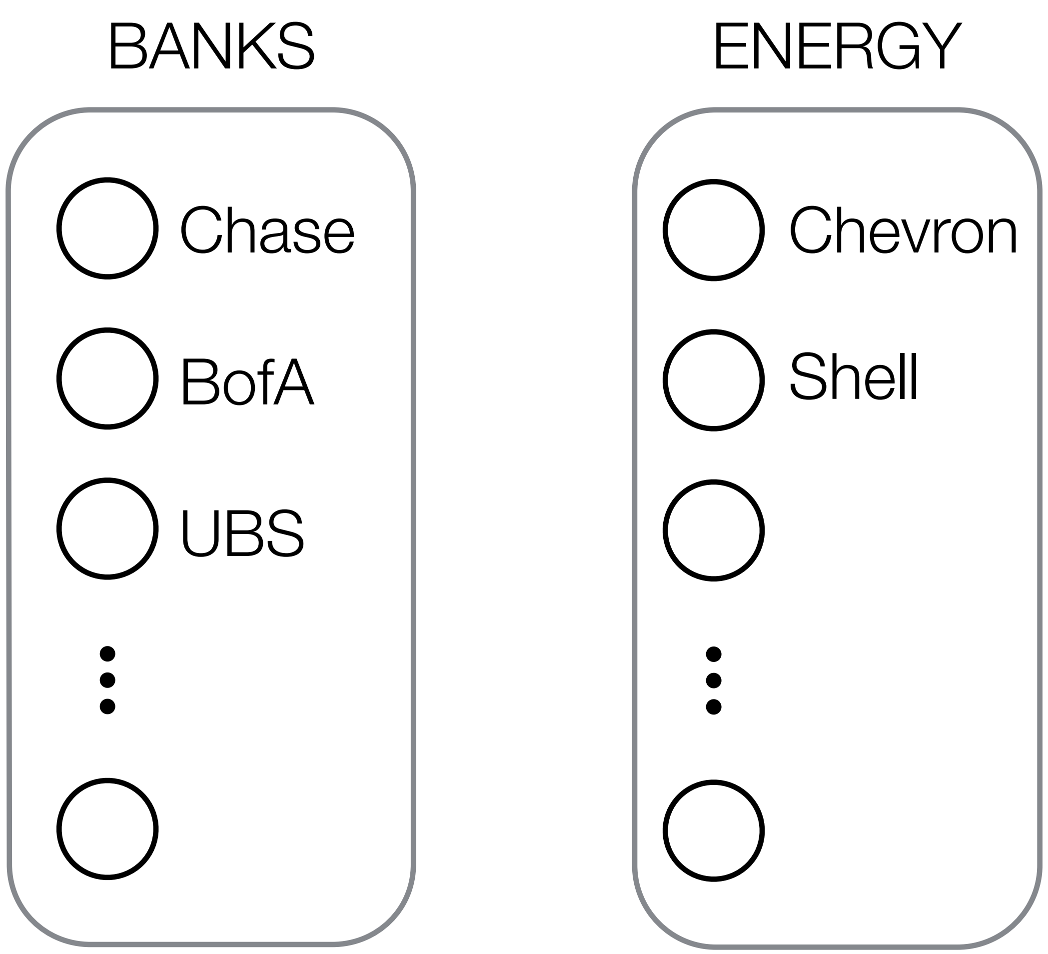
Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



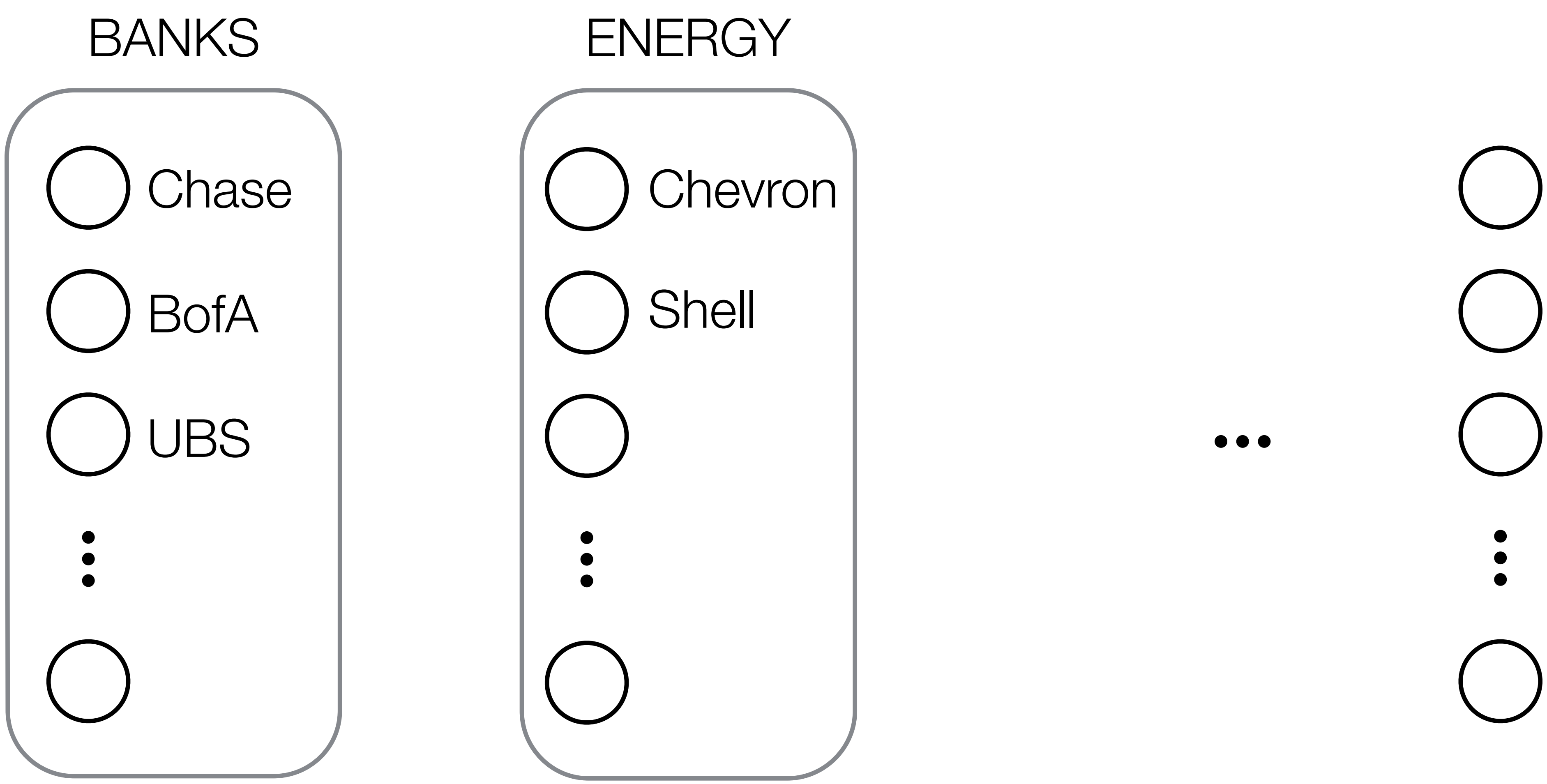
Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



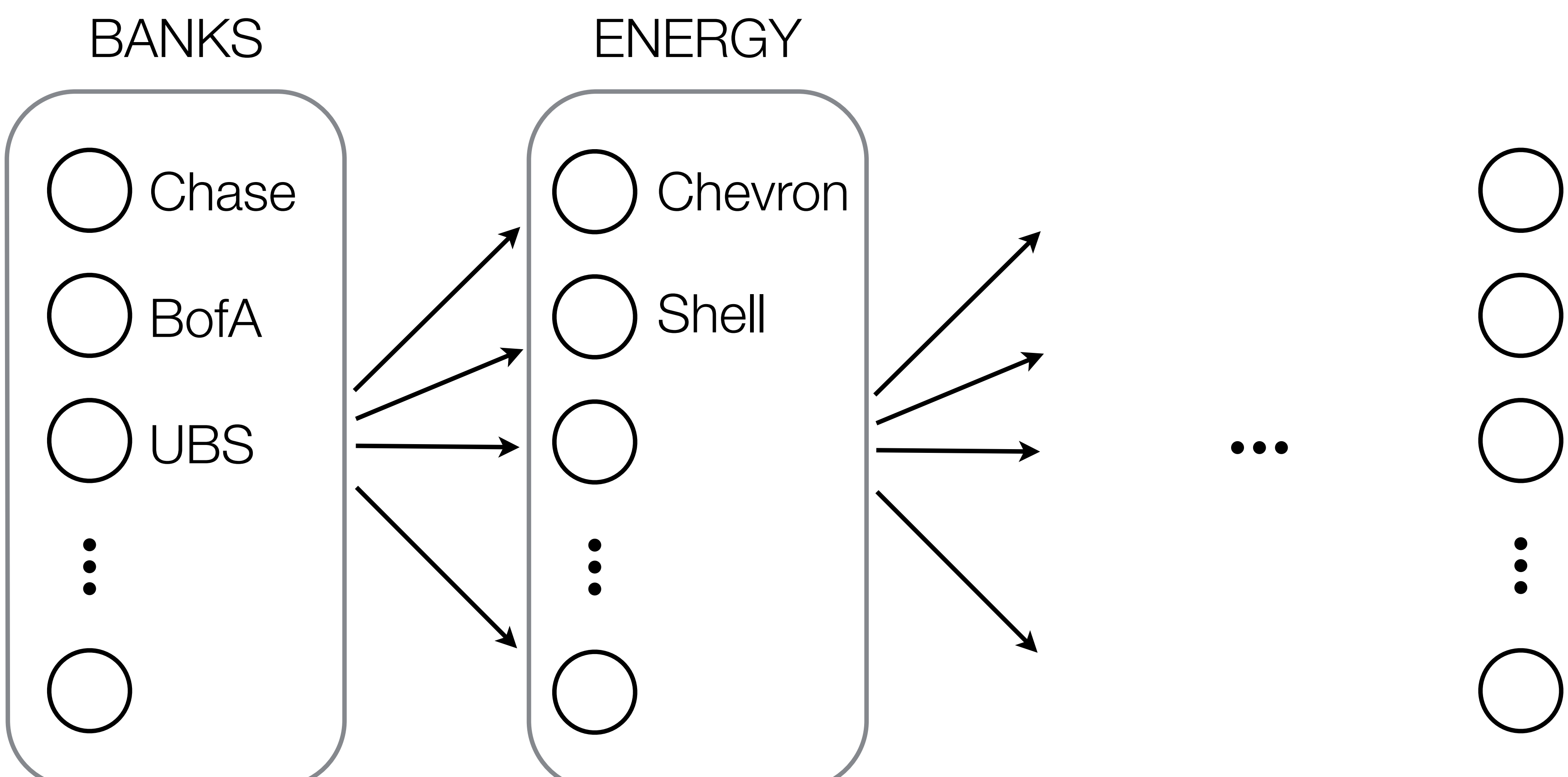
Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



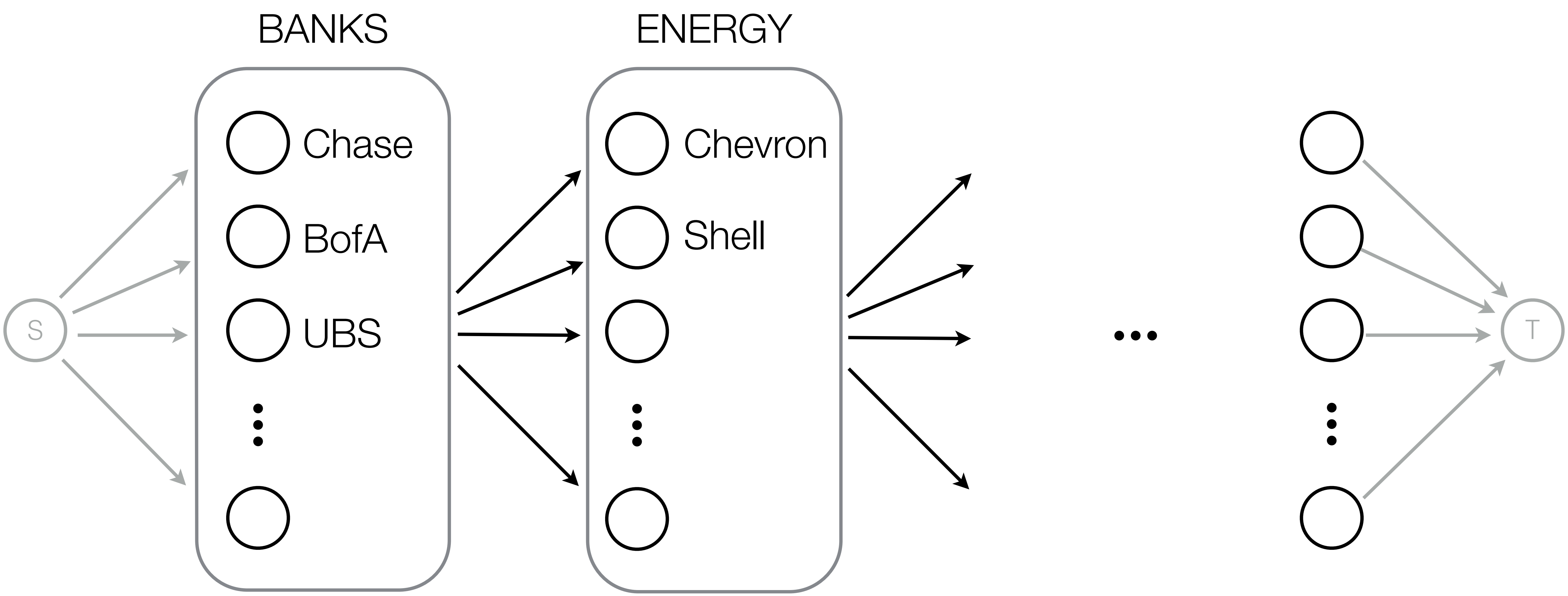
Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



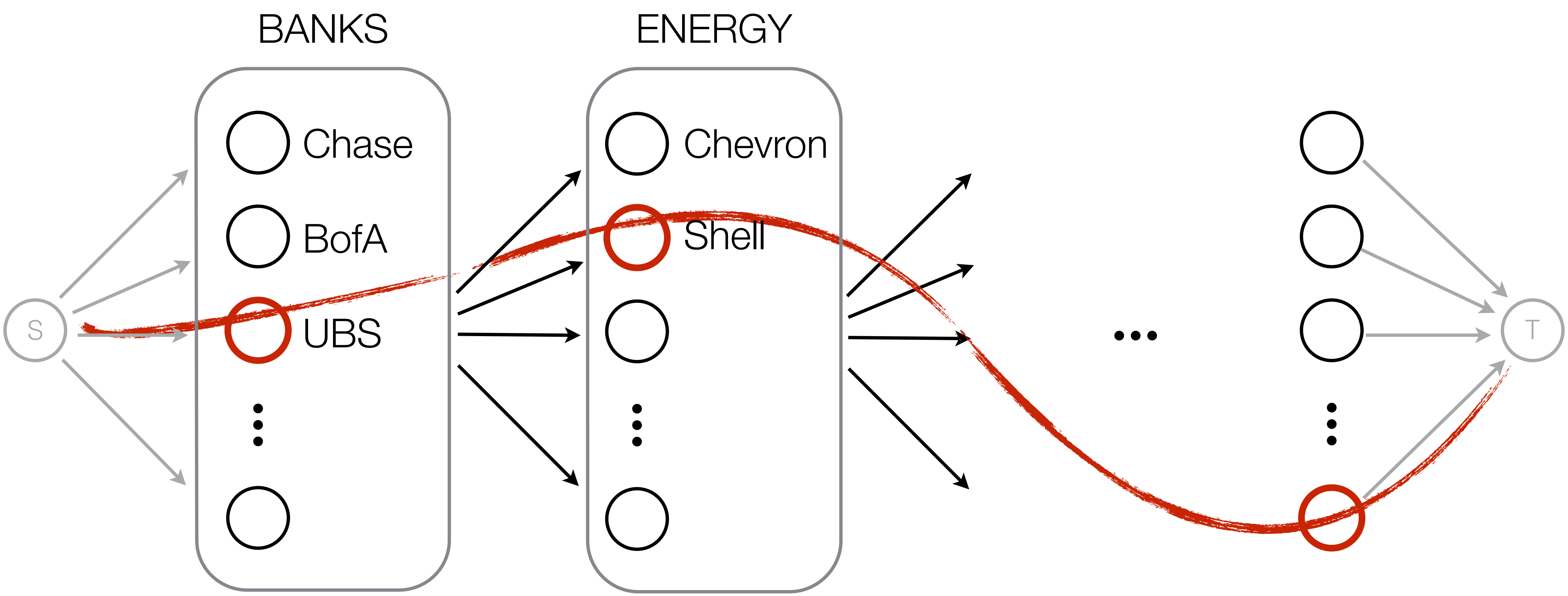
Motivation 2: Finance

- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



Motivation 2: Finance

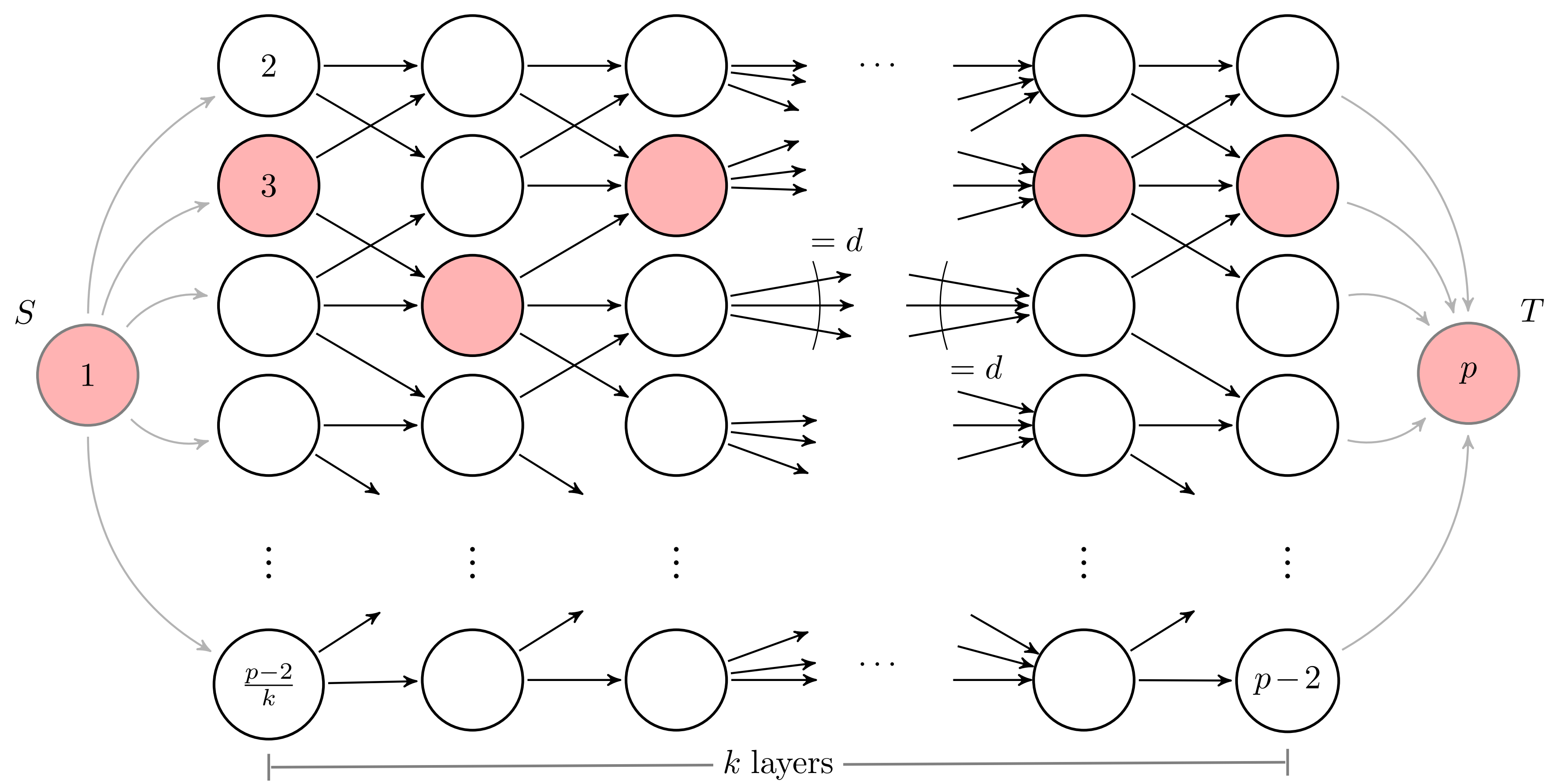
- Variables: **stocks** divided in sectors
- Measurements: **prices over time**
- **Goal**: Find subset that explains variance 1 stock/ sector



[Statistical Analysis]

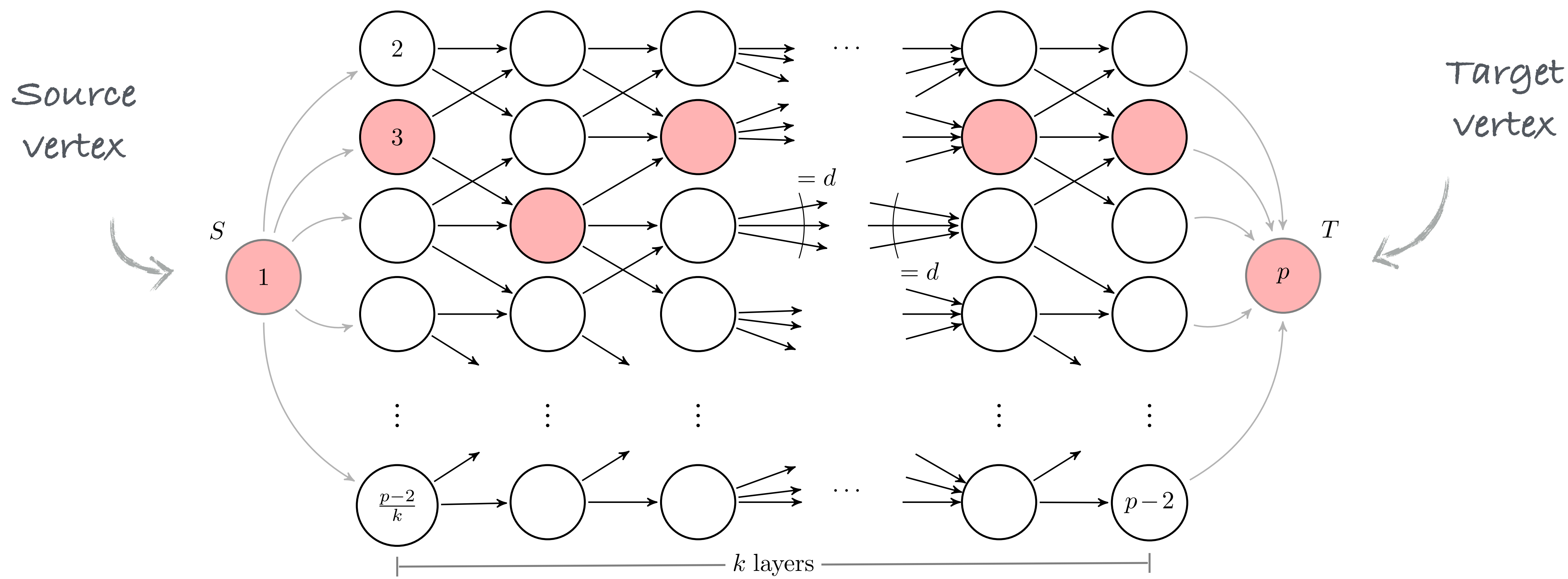
Data model

(p,k,d)-layer graph



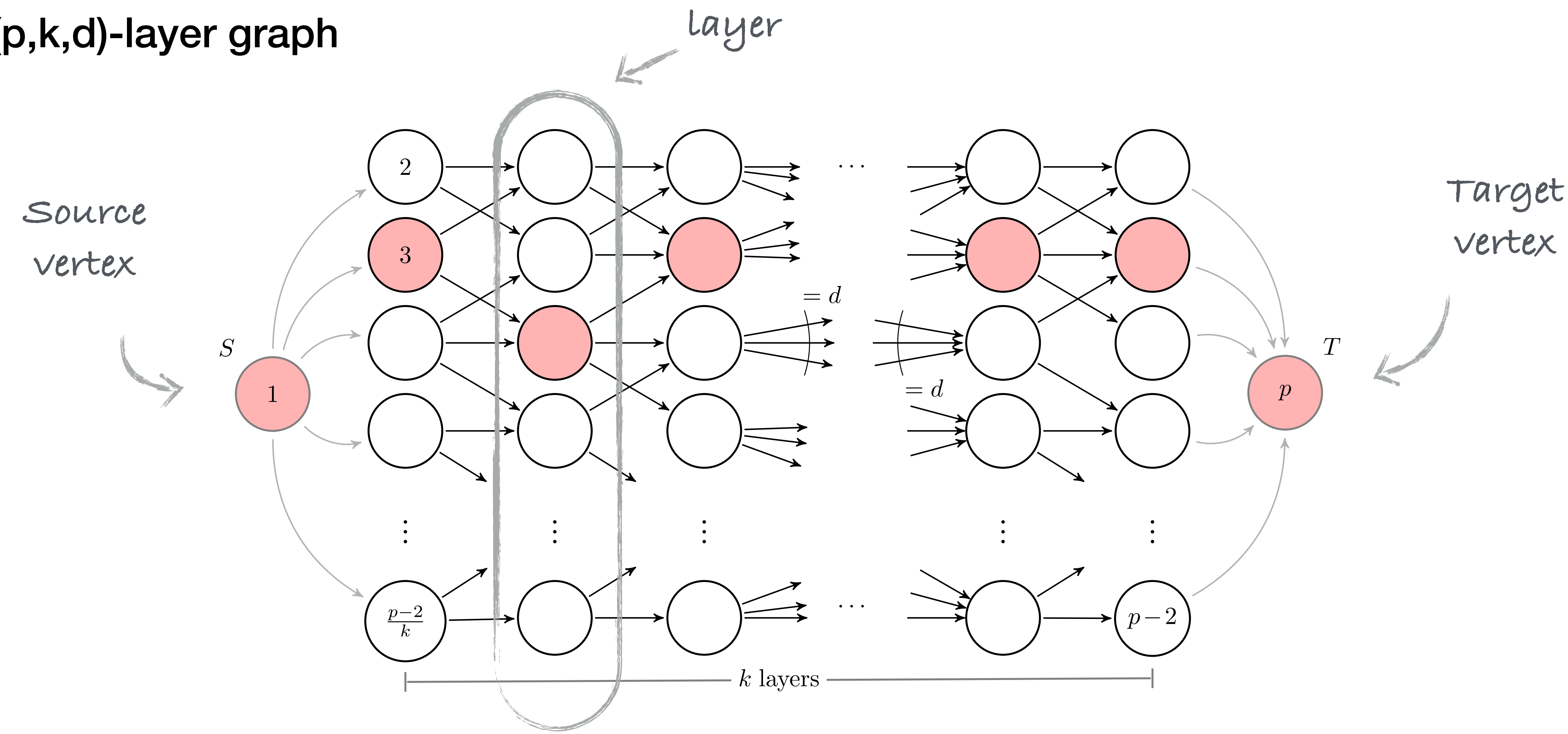
Data model

(p,k,d)-layer graph



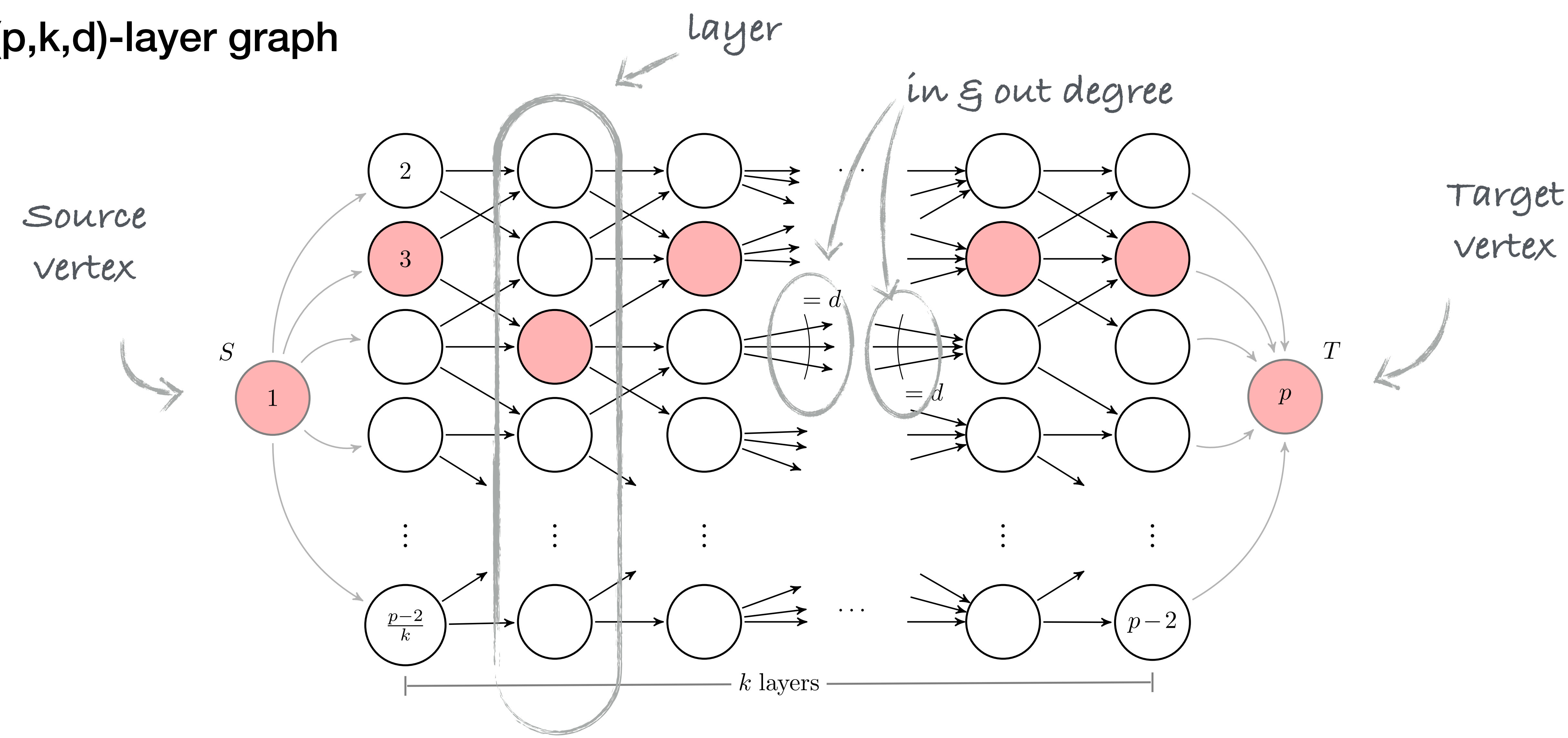
Data model

(p,k,d)-layer graph



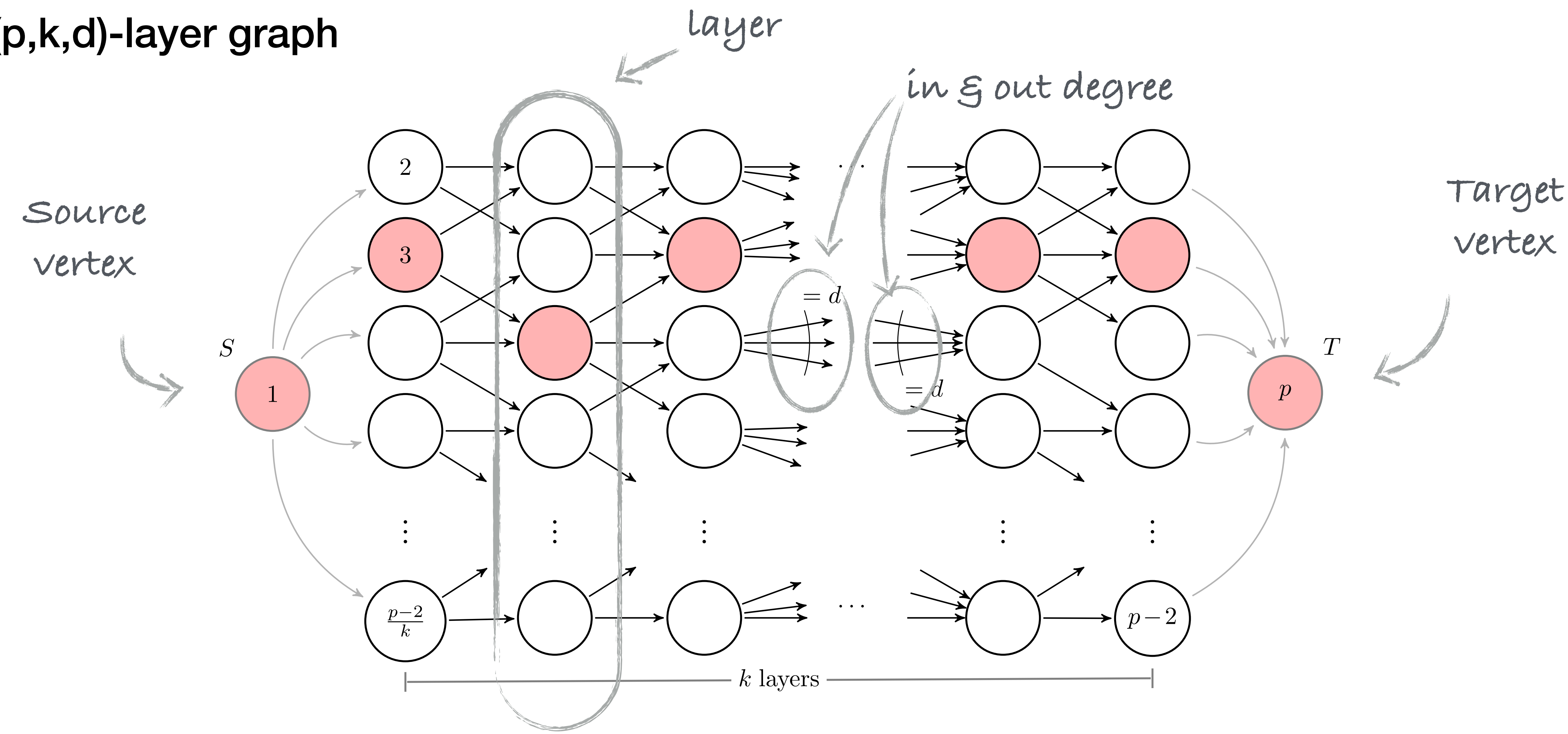
Data model

(p,k,d)-layer graph



Data model

(p,k,d)-layer graph



Spike along a path

Samples

$$\mathbf{y}_i = \sqrt{\beta} \cdot u_i \cdot \mathbf{x}_\star + \mathbf{z}_i$$

Gaussian noise (i.i.d)

Signal, supported on path of G.

Bounds

[Theorem 1]

G : (p, k, d)-layer graph (known). $\mathbf{x}_\star$: signal support on st-path of G. (unknown)

Observe sequence $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of i.i.d. samples from $\mathcal{N}(\mathbf{0},\ \beta \cdot \mathbf{x}_\star \mathbf{x}_\star^\top + \mathbf{I})$.

$$\widehat{\boldsymbol{\Sigma}} \longrightarrow \boxed{\begin{array}{ll} \max\limits_{\mathbf{x}} & \mathbf{x}^\top \widehat{\boldsymbol{\Sigma}}\, \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}(G) \end{array}} \longrightarrow \widehat{\mathbf{x}}$$

Then, $n = O\left(\log \frac{p}{k} + k \log d\right)$ samples **suffice** for recovery.

Bounds

[Theorem 1]

G : (p, k, d)-layer graph (known). $\mathbf{x}_\star$: signal support on st-path of G. (unknown)

Observe sequence $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of i.i.d. samples from $\mathcal{N}(\mathbf{0},\ \beta \cdot \mathbf{x}_\star \mathbf{x}_\star^\top + \mathbf{I})$.

$$\widehat{\boldsymbol{\Sigma}} \longrightarrow \boxed{\begin{array}{rl} \max\limits_{\mathbf{x}} & \mathbf{x}^\top \widehat{\boldsymbol{\Sigma}}\, \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}(G) \end{array}} \longrightarrow \widehat{\mathbf{x}}$$

Then, $n = O\left(\log \frac{p}{k} + k \log d\right)$ samples **suffice** for recovery.

vs $\Omega\left(k \log \frac{p}{k}\right)$

for sparse PCA.

Bounds

[Theorem 1]

G : (p, k, d)-layer graph (known). $\mathbf{x}_\star$: signal support on st-path of G. (unknown)

Observe sequence $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of i.i.d. samples from $\mathcal{N}(\mathbf{0},\ \beta \cdot \mathbf{x}_\star \mathbf{x}_\star^\top + \mathbf{I})$.

$$\widehat{\mathbf{\Sigma}} \longrightarrow \boxed{\begin{array}{ll} \max\limits_{\mathbf{x}} & \mathbf{x}^\top \widehat{\mathbf{\Sigma}}\, \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}(G) \end{array}} \longrightarrow \widehat{\mathbf{x}}$$

Then, $n = O\left(\log \frac{p}{k} + k \log d\right)$ samples **suffice** for recovery.

vs $\Omega\left(k \log \frac{p}{k}\right)$

for sparse PCA.

[Theorem 2]

That many samples are also **necessary**.

Bounds

[Theorem 1]

G : (p, k, d)-layer graph (known). $\mathbf{x}_\star$: signal support on st-path of G. (unknown)

Observe sequence $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of i.i.d. samples from $\mathcal{N}(\mathbf{0}, \beta \cdot \mathbf{x}_\star \mathbf{x}_\star^\top + \mathbf{I})$.

$$\widehat{\mathbf{\Sigma}} \longrightarrow \boxed{\begin{array}{ll} \max\limits_{\mathbf{x}} & \mathbf{x}^\top \widehat{\mathbf{\Sigma}}\, \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}(G) \end{array}} \longrightarrow \widehat{\mathbf{x}}$$

NP-HARD

Then, $n = O\left(\log \frac{p}{k} + k \log d\right)$ samples **suffice** for recovery.

vs $\Omega\left(k \log \frac{p}{k}\right)$ for sparse PCA.

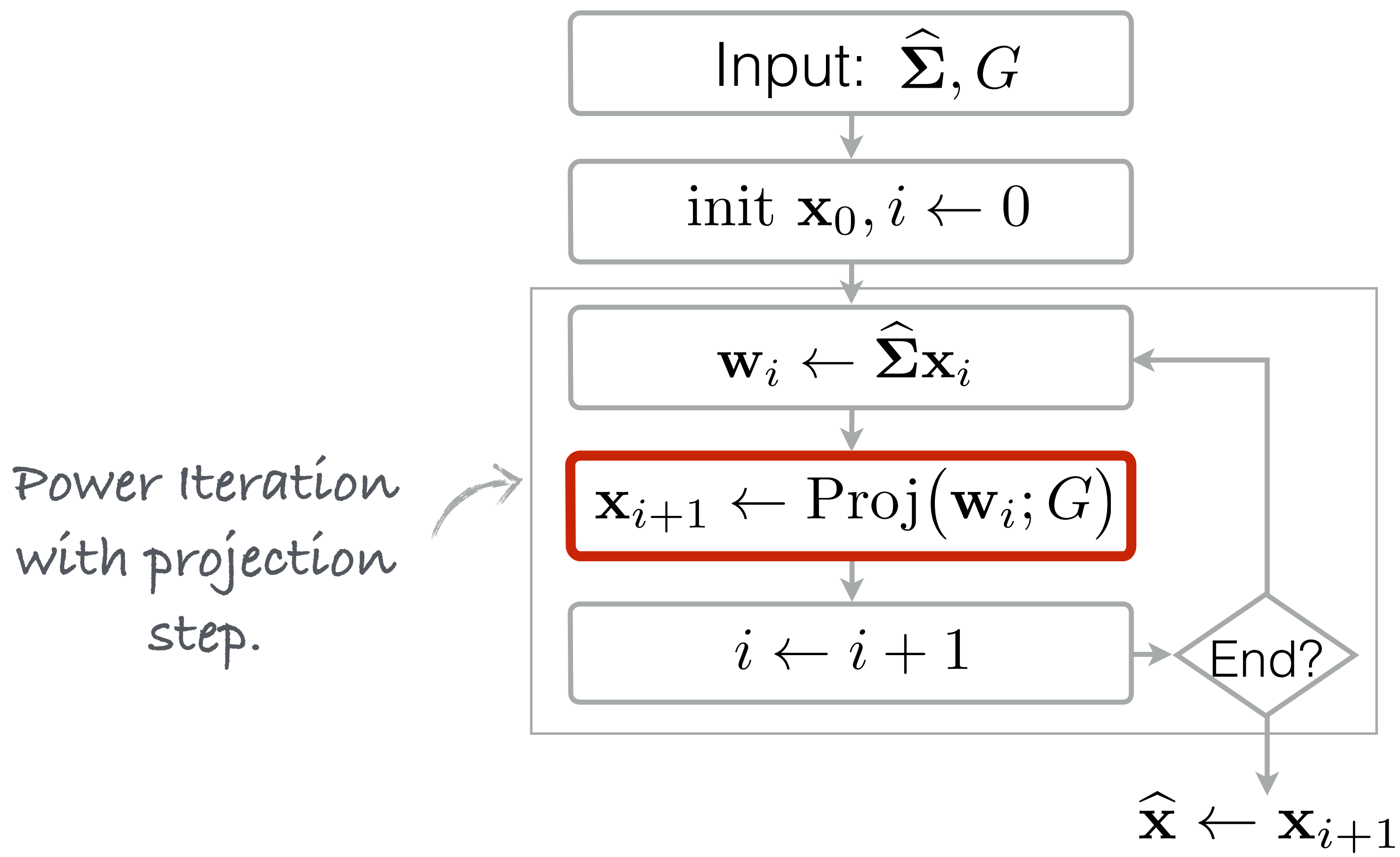
[Theorem 2]

That many samples are also **necessary**.

[Algorithms]

Algorithm 1

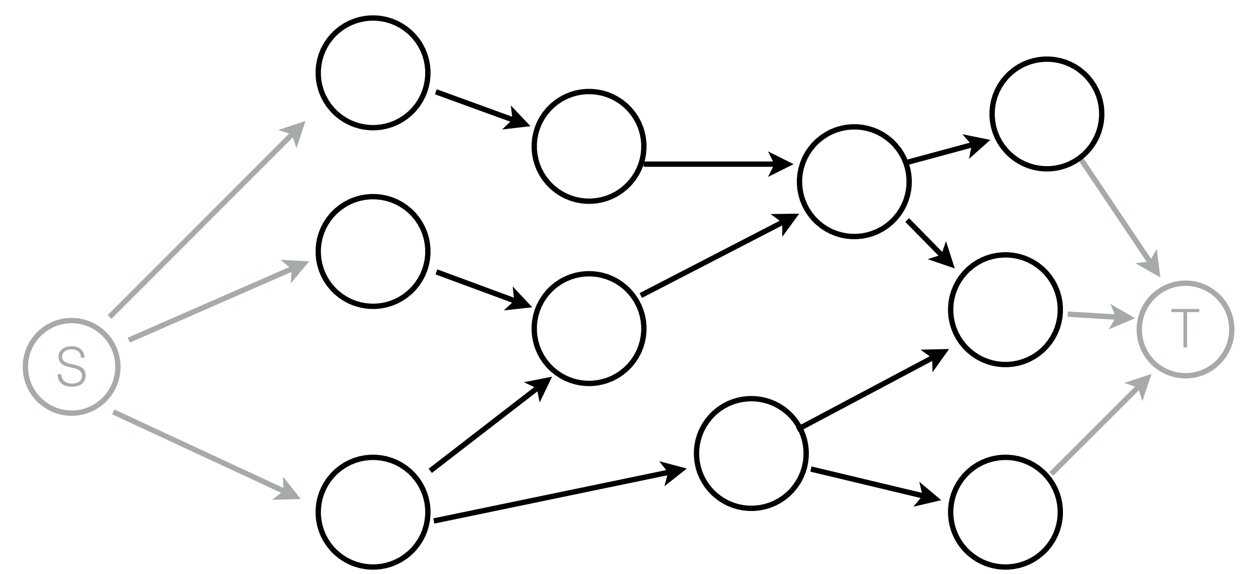
A Power Method-based approach.



[Projection Step]

Project a p-dimensional $\mathbf{w}$ on $\mathcal{X}(G) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_2 = 1,\ \mathrm{supp}(\mathbf{x}) \subset \text{path of } G\}$

$$\mathrm{Proj}\big(\mathbf{w}; G\big) = \arg\min_{\mathbf{x}\in\mathcal{X}(G)} \|\mathbf{x} - \mathbf{w}\|_2$$



[Projection Step]

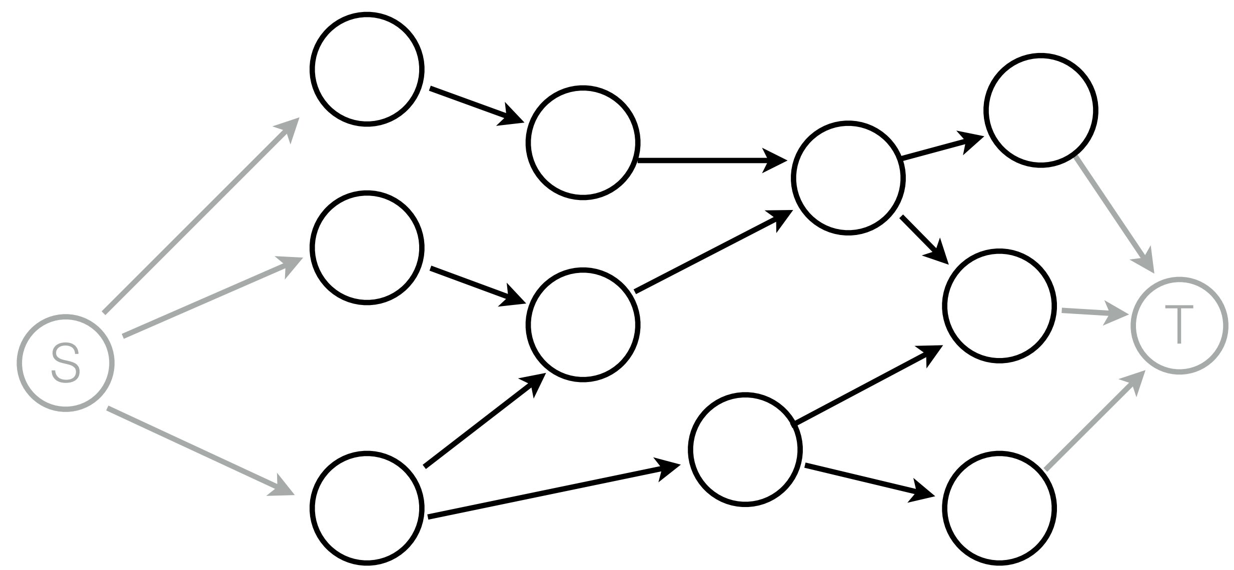
Project a p-dimensional $\mathbf{w}$ on $\mathcal{X}(G) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_2 = 1,\ \mathrm{supp}(\mathbf{x}) \subset \text{path of } G\}$

$$\mathrm{Proj}\big(\mathbf{w}; G\big) = \arg\min_{\mathbf{x} \in \mathcal{X}(G)} \|\mathbf{x} - \mathbf{w}\|_2$$

$$\Updownarrow$$

Due to the constraints.

$$\arg\max_{\mathbf{x} \in \mathcal{X}(G)} \big(\mathbf{x}^\top \mathbf{w}\big)^2$$



[Projection Step]

Project a p-dimensional $\mathbf{w}$ on $\mathcal{X}(G) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_2 = 1,\ \mathrm{supp}(\mathbf{x}) \subset \text{path of } G\}$

$$\mathrm{Proj}\big(\mathbf{w}; G\big) = \arg\min_{\mathbf{x} \in \mathcal{X}(G)} \|\mathbf{x} - \mathbf{w}\|_2$$

Due to the constraints.

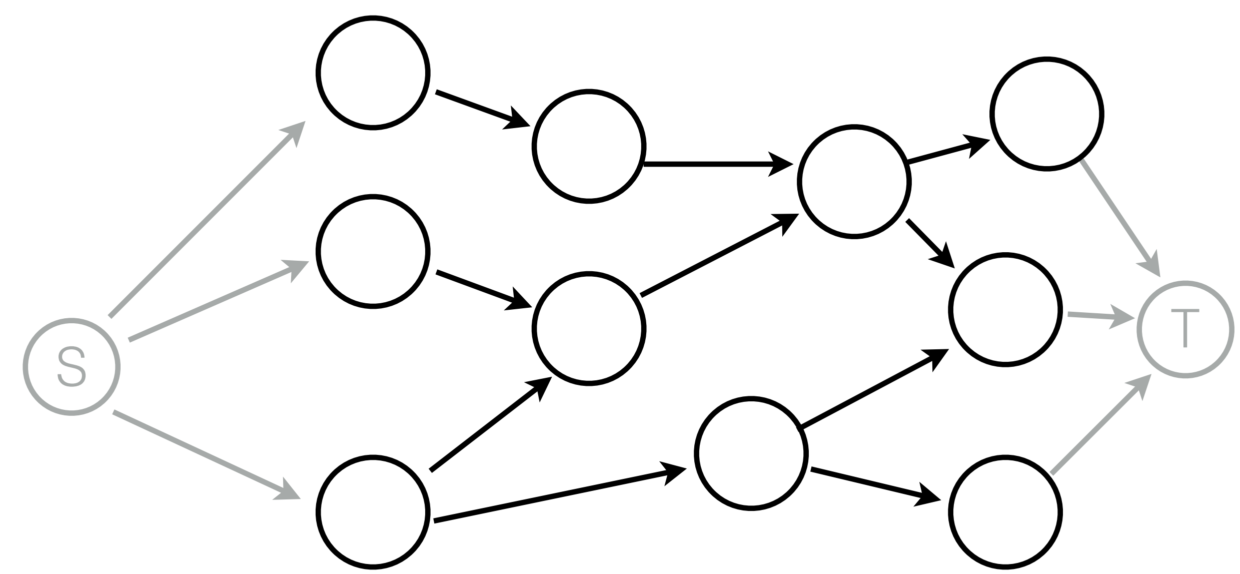
$$\Updownarrow$$

$$\arg\max_{\mathbf{x} \in \mathcal{X}(G)} \big(\mathbf{x}^\top \mathbf{w}\big)^2$$

Due to Cauchy -Schwarz

$$\Updownarrow$$

$$\arg\max_{\text{path } \pi} \sum_{i \text{ on } \pi} w_i^2$$



Projection Step

Project a p-dimensional $\mathbf{w}$ on $\mathcal{X}(G) = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_2 = 1,\ \mathrm{supp}(\mathbf{x}) \subset \text{path of } G\}$

$$\mathrm{Proj}(\mathbf{w}; G) = \arg\min_{\mathbf{x} \in \mathcal{X}(G)} \|\mathbf{x} - \mathbf{w}\|_2$$

Due to the constraints.

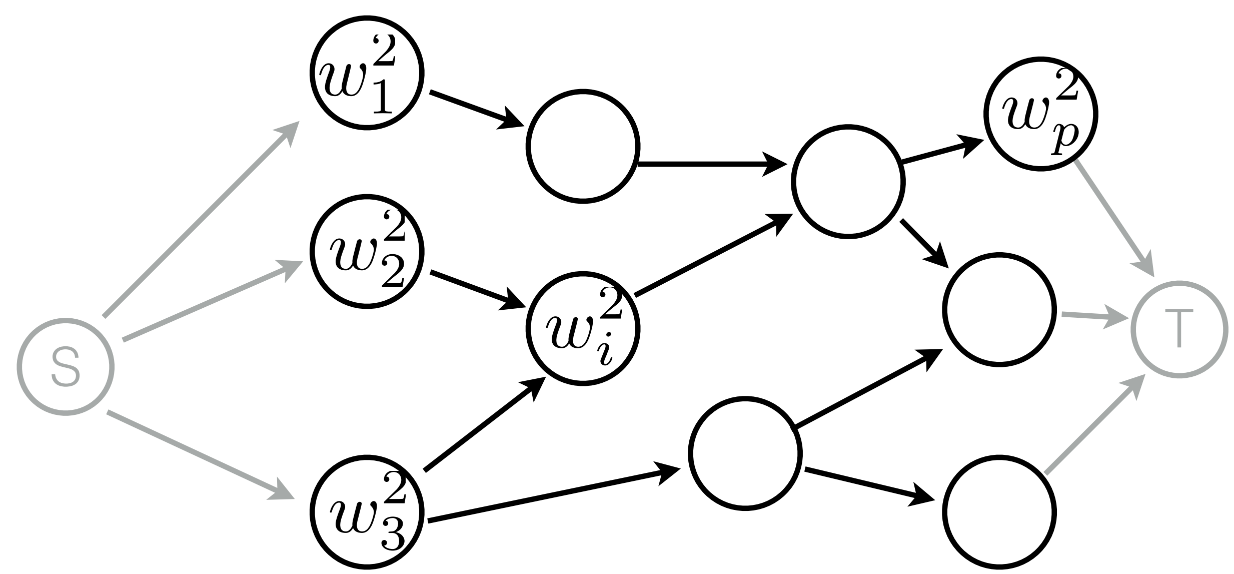
$$\Updownarrow$$

$$\arg\max_{\mathbf{x} \in \mathcal{X}(G)} \left(\mathbf{x}^\top \mathbf{w}\right)^2$$

Due to Cauchy -Schwarz

$$\Updownarrow$$

$$\arg\max_{\text{path } \pi} \sum_{i \text{ on } \pi} w_i^2$$



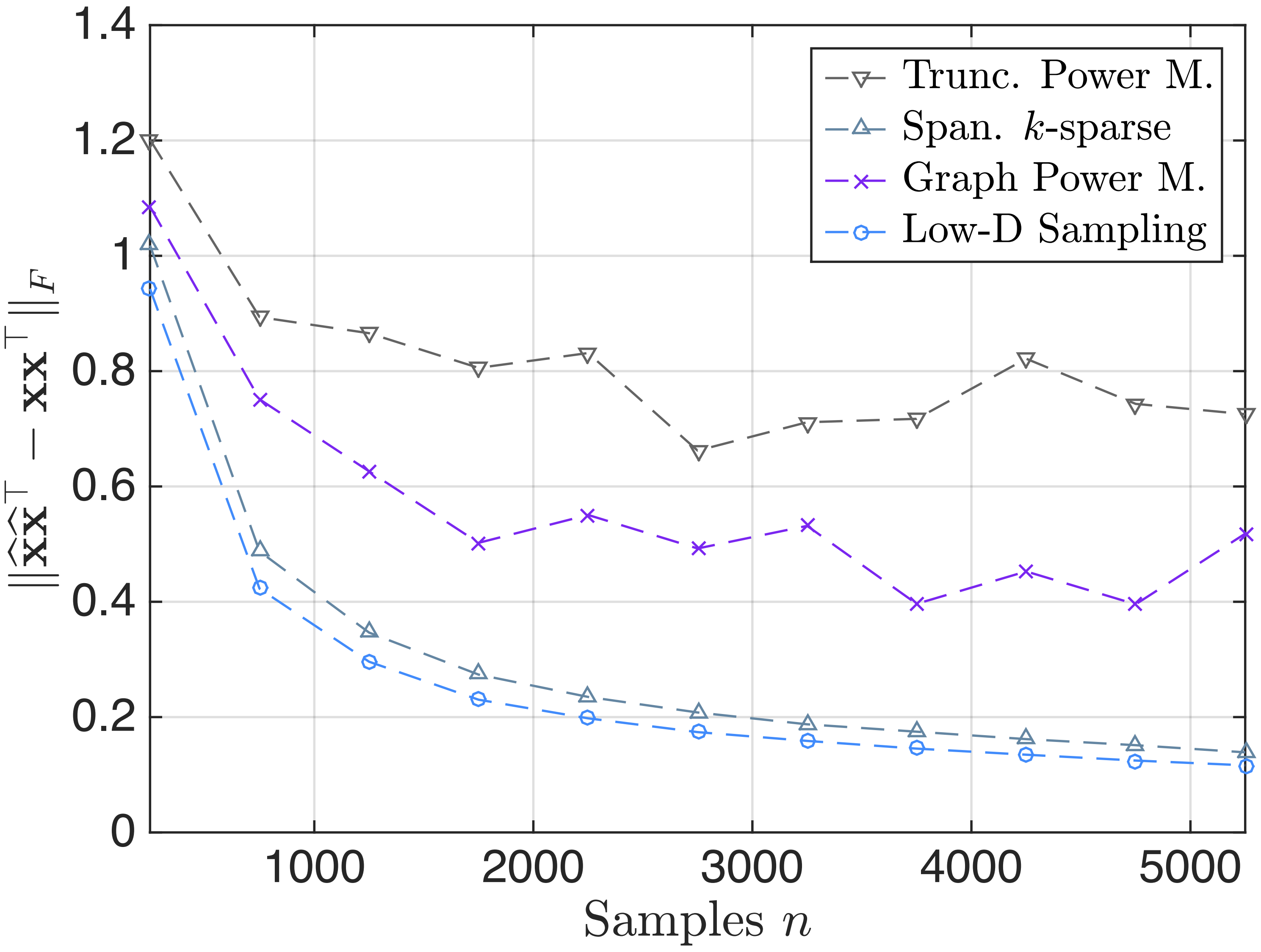
Longest (weighted) path problem on G, with special weights!

G acyclic; $O(p + |E|)$

[Experiments]

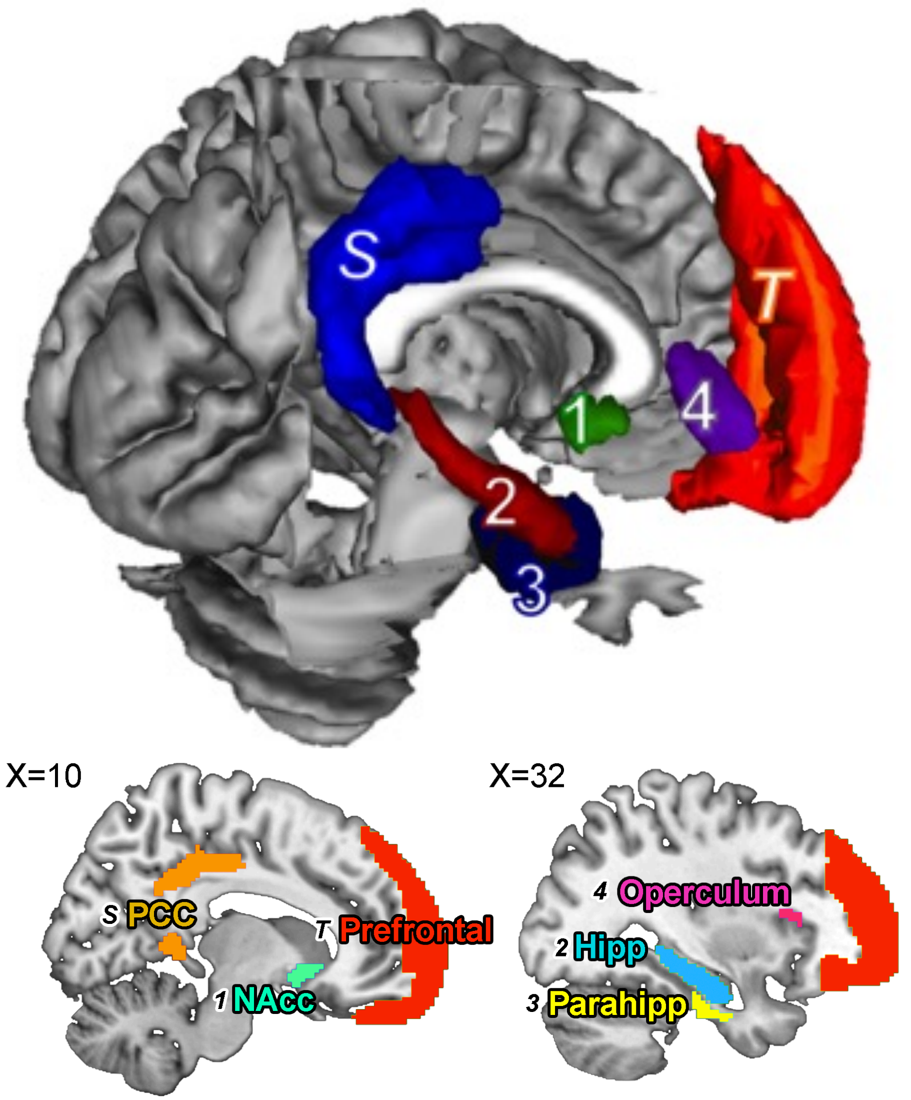
Synthetic

Data generated according to the (p,k,d)-layer graph model.
(p=1000, k=50, d=10 , 100 MC iterations)



Neuroscience

- Resting state fMRI dataset.*
- 111 regions of interest (ROIs) (variables), extracted based on Harvard-Oxford Atlas [Desikan et al., 2006].
- Graph extracted based on Euclidean distances between center of mass of ROIs.



Identified core neural components of the **brain's memory network.**

*[Human Connectome Project, WU-Minn Consortium]

Summary

- New problem: sparse PCA with support restricted on paths of DAGs.
- Statistical analysis
  - Introduced a simple graph model.
  - Side information (underlying graph) reduces statistical complexity.
- Approximation algorithms
  - Projection step → Longest path on weighted graph.

[Future]

- Other combinatorial structures?
- Algorithm guarantees
- Neuroscience applications