Final Project: Graphical Models

FISE 2025-2026

January 2026

Beforehand, please form a group of 2 students and fill in the following Google sheet with members' full names. If you have any contraints with your schedule, please fill in the column **Remarks**.

Overview You will model a synthetic high-dimensional financial dataset using **(i) a Bayesian Network (directed)** and **(ii) a Markov Network (undirected)**. You will perform structure learning, parameter learning, and probabilistic inference, then compare the modeling assumptions and the economic interpretations. undirected graphical models when applied to the same financial system ?

Objectives :

— Translate financial dependence assumptions into a directed (BN) and an undirected (MRF/GGM) graphical model.

— Perform structure learning in high dimension with appropriate regularization and model selection.

— Perform parameters learning (CPDs/regression coefficients ; precision matrices).

— Interpret the results.

Dataset You are given daily returns (synthetic) for :

— **MKT** market index return ;

— **SEC01-SEC10** : sector index proxy returns ;

— **A001-A100** : 100 asset returns grouped into 10 sectors (given in `map.csv`)

Read `readme.txt` for details.

Guidelines :

Part I : Bayesian Network Model a directed acyclic graph over variables within each day (time slice), optionally including lag edges from day $t-1$ to day t.

$$\text{MKT} \longrightarrow \text{Sector}_k \longrightarrow \text{Assetreturns}_i, \quad Asset_i \in SEC_k.$$

1. Structure learning
   — Decide whethter to learn a static BN or a dynamic BN (allow some edges from $A(t-1)$ to $A(t)$).
   — Learn structure using hill climb method or PC algorithm.
   — Report learned edges and discuss whether edges align with the market/sector hierarchy.
2. Parameters learning
   — Fit conditional distributions given the learned structure (eg, descrete CPDs or linear Gaussian CPDs).
   — Use a time-based split (eg, 70% train + 30% test)
   — Give predictions for a few assets.

Part II : Markov network

Model undirected conditional dependence among assets using a sparse precision matrix (Gaussian graphical model).

Markov Network (MRF) : Nodes are assets only. Edges represent conditional dependence between assets after controlling for all others.

1. Structure learning
   — Fit a sparse Gaussian graphical model on asset returns (A001-A100)
   — Use Graphical Lasso and choose the regularization parameter using cross-validation, BIC for example.
   — Report the learned graph, sparsity level, connected component, whether edges are mostly within-sector.
2. Parameter learning
   — Estimate the precision matrix and implied covariance.
   — Evaluate on the test set
3. Inference : Choose a set of assets A and unobserved assets B, compute $P(B|A)$.

Part III : Discussion

- Which dependencies appear in BN but not in MRF, and vice versa ?
- Discuss symmetry (MRF) vs directionality (BN) and what that means for interpretation.
- Compare inference outputs $P(B|A)$ of the two approaches.
- Apply to real dataset, for example the `30 Industry Portfolios` found here.

Submissions : A project report **maximum 8 pages** including implementation results, figures and interpretations + the supported materials (implemeted codes) need to be compressed into a `NAME1_NAME2.zip` and submit to `exam.ensiie.fr`
If you don't have ensIIE account, send the `NAME1_NAME2.zip` file to email : thiphuongthuy.vo@ensiie.fr

Deadline : Submit the completed project by midnight, **Febuary, the 1st 2026**.

Final evaluation : 20% code + 30 % report + 50% presentation (15 minutes).

Remarks : Please refer and cite properly the additional references that you used in the project.

Feel free to reach out for any clarifications or guidance.