M2 Quantitative Finance

Machine Learning

Probabilistic Graphical Models for financial dependence analysis

---

Paolo Bortun & Fátima Lastra Incera

Academic Year: 2025–2026

Index

1 Part I – Bayesian Network Model

In this part, our goal is to model the **conditional dependence structure** of a high-dimensional financial system using a **directed graphical model** that reflects economically meaningful causal relationships. We consider a synthetic dataset of daily returns for a market index (MKT), ten sector indices (SEC01–SEC10), and 100 individual assets (A001–A100), each assigned to a single sector.

From a financial perspective, a **hierarchical dependence structure** is natural: market movements affect sector returns, which in turn affect individual asset returns. This motivates the following causal ordering within each day:

$$\mathrm{MKT}_t \to \mathrm{SEC}_{k,t} \to A_{i,t}, \quad A_i \in \mathrm{SEC}_k \tag{1}$$

A **Bayesian Network (BN)** can be static or dynamic, depending on whether temporal dependencies are modeled. While static BNs capture only contemporaneous relationships, **Dynamic Bayesian Networks (DBNs)** allow dependencies across time slices. In this project, we adopt a **DBN** including self-lag edges of the form:

$$X_{t-1} \to X_t, \tag{2}$$

to account for possible short-term autocorrelation in financial returns. This choice is motivated by the fact that financial returns may exhibit **short-term autocorrelation**, and a DBN provides a principled framework to test for such effects rather than assuming independence across time.

1.1 Structure learning via hill-climbing

Structure learning in Bayesian Networks can be performed using either **constraint-based** or **score-based methods**. While constraint-based approaches rely on conditional independence tests, score-based methods formulate structure learning as a discrete optimization problem, searching for the graph that maximizes a global score balancing **goodness of fit** and **model complexity**.

In this work, we adopt a **score-based hill-climbing algorithm** with the **Bayesian Information Criterion (BIC)** as scoring function. Starting from an empty graph, the algorithm iteratively applies local modifications (edge addition or removal), retaining the change that yields the largest improvement in the BIC score, until no further improvement is possible.

This approach is particularly well suited for our setting for three main reasons. First, the dataset is **high-dimensional** (111 variables), making constraint-based methods unstable due to the large number of conditional independence tests required. Second, financial returns are **noisy**, leading to unreliable independence tests and error propagation in PC-style algorithms. Third, a clear **economic hierarchy** (market → sector → asset) is available and can be exploited to restrict the search space, dramatically reducing the number of candidate graphs.

1.2 Learned edges and consistency with the market/sector hierarchy

Now, we report the main families of learned edges and verify whether they align with the expected **market–sector–asset (M–S–A) hierarchy**.

Intra-temporal edges ($t \to t$). The learned structure contains **110 intra-temporal edges**, which match exactly the hierarchical decomposition of the system: **10 edges of the form $\mathbf{MKT}_t \to \mathbf{SEC}_{k,t}$**, one for each sector, and **100 edges of the form $\mathbf{SEC}_{k,t} \to \mathbf{A}_{i,t}$**, one for each asset. Importantly, no intra-temporal edges violating the hierarchy are found (e.g. $A_{i,t} \to \mathrm{SEC}_{k,t}$, $\mathrm{SEC}_{k,t} \to \mathrm{MKT}_t$, or direct $\mathrm{MKT}_t \to A_{i,t}$ edges), confirming that the learned graph is fully consistent with the intended **M–S–A structure** and remains economically interpretable.

Inter-temporal edges ($t-1 \to t$). The learned structure contains only **two inter-temporal edges**, both corresponding to asset self-dependence:

$$A021_{t-1} \to A021_t, \qquad A047_{t-1} \to A047_t. \tag{3}$$

No lagged edges are selected for the **market index** or for the **sector indices**. This is coherent with the behavior of daily returns, where **autocorrelation at lag 1** is typically **weak** and often **statistically insignificant**. Therefore, the DBN formulation allows for **temporal effects**, but the score-based learning procedure (BIC) retains only the **lagged dependencies** supported by the data.

Overall, the **learned edges** strongly **support the market/sector hierarchy**: the contemporaneous dependence structure is entirely explained by the chain (1) with **each asset connected to its own sector** as expected. The scarcity of inter-temporal edges indicates that, in this synthetic dataset, **most of the explainable dependence is cross-sectional (market and sector factors) rather than temporal**, which is consistent with standard stylized facts of financial returns.

1.3 Parameter learning: conditional distributions (CPDs)

Parameter learning estimates the **conditional probability distributions (CPDs)** associated with each node given its parents. In a BN, parameter learning consists of **specifying a distributional family for each node** and then **estimating its parameters from data**, conditional on the **parent set** $\mathrm{Pa}(X)$ implied by the learned graph. Two standard choices for CPDs are:

- **Discrete CPDs**: used when variables take values in a **finite set** (categories). In this case, the CPD is a conditional probability table, that means that $P(X \mid \mathrm{Pa}(X))$ is represented by probabilities for each configuration of parent states.
- **Linear Gaussian CPDs**: used when variables are **continuous** and **approximately Gaussian**. In this case, each node is modeled as a l**inear regression** on its parents with Gaussian noise:

$$X = \beta_0 + \sum_{p \in \mathrm{Pa}(X)} \beta_p \, p + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2). \tag{4}$$

Since all variables are continuous daily returns, we adopt **linear Gaussian CPDs**, avoiding arbitrary discretization and information loss. For each node X at time t, we estimate the CPD parameters using **ordinary least squares (OLS)** on the **training data**. Concretely, given the learned parent set $\mathrm{Pa}(X)$, we fit:

$$X_t = \beta_0 + \beta^\top \mathrm{Pa}(X_t) + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{5}$$

and store the estimated regression coefficients $\hat{\beta}$ (including the intercept) and the residual variance $\hat{\sigma}^2$. Root nodes (with no parents) reduce to an intercept-only Gaussian model.

A total of **111 CPDs** are estimated (1 market index, 10 sectors, and 100 assets). The estimated parent sets match the learned hierarchical structure. For instance:

- MKT_t is a root node with no parents, modeled as $\mathrm{MKT}_t \sim \mathcal{N}(\mu, \sigma^2)$.
- $\mathrm{SEC01}_t$ has parent MKT_t, consistent with the market-to-sector dependence.
- $A001_t$ has parent $\mathrm{SEC01}_t$, consistent with the sector-to-asset dependence.

Table 1 summarizes representative parameter estimates of the learned CPDs. These estimates confirm that **sector returns load strongly** on the **market factor**, while **asset returns load primarily** on their **sector factor**, which is consistent with the intended economic hierarchy.

Table 1: Illustrative examples of learned linear Gaussian CPDs.

Variable	**Parents**	**Estimated coefficients**	$\hat{\sigma}^2$
MKT_t	$\emptyset$	$\hat{\beta}_0 = -3.79 \times 10^{-4}$	9.52×10^{-5}
$\mathrm{SEC01}_t$	$\{\mathrm{MKT}_t\}$	$\hat{\beta}_0 = 2.66 \times 10^{-4}$, $\hat{\beta}_{\mathrm{MKT}} = 1.085$	6.96×10^{-5}
$A001_t$	$\{\mathrm{SEC01}_t\}$	$\hat{\beta}_0 = 3.10 \times 10^{-4}$, $\hat{\beta}_{\mathrm{SEC01}} = 0.834$	4.77×10^{-5}

1.4 Asset return prediction

Predictive performance of the learned BN is evaluated by a **time-based train/test split**, using the first 70% of the observations for training and the remaining 30% for testing. This preserves the temporal ordering of the data and prevents look-ahead bias, as the model is trained on past observations and evaluated on a held-out future sample using contemporaneous market and sector information.

Predictions are generated for a small subset of assets to illustrate the behavior of the model. Consistent with the learned structure, forecasts follow the hierarchical chain (1), reflecting the information flow encoded in the BN. Forecasting is performed in two steps, consistent with the learned graph and avoids using information that is not causally available at prediction time.

First, **sector returns** are predicted from the **market return** using a **linear Gaussian CPD**:

$$\widehat{\mathrm{SEC}}_{k,t} = \beta_0 + \beta_1 \mathrm{MKT}_t. \tag{6}$$

Second, **asset returns** are predicted using the **predicted sector return** as input:

$$\widehat{A}_{i,t} = \alpha_0 + \alpha_1 \widehat{\mathrm{SEC}}_{k,t}. \tag{7}$$

The results show that the model captures a significant fraction of the variation in asset returns and that the estimated parameters are economically meaningful. In particular, sector returns are strongly explained by the market factor. For **sector** SEC01, the estimated model is:

$$\widehat{\mathrm{SEC01}}_t = \beta_0 + \beta_1 \mathrm{MKT}_t, \qquad \beta_1 \approx 1.07, \tag{8}$$

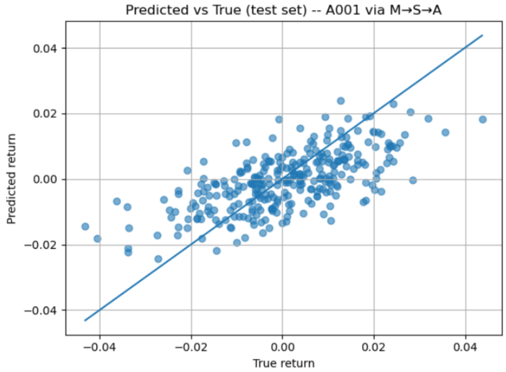
indicating a near one-to-one exposure of the sector to market movements. At the **asset** level, returns are primarily driven by the corresponding sector factor. For example, for asset A001 we obtain:

$$\widehat{A001}_t = \alpha_0 + \alpha_1 \widehat{\mathrm{SEC01}}_t, \qquad \alpha_1 \approx 0.83, \tag{9}$$

showing that sector-level fluctuations explain most of the predictable component of the asset return. Similar relationships are observed for other assets, with sector loadings varying across assets but remaining clearly dominant. Test set predictions are small, with mean squared errors of order 10^{-4}, consistent with the scale of daily returns in the dataset, indicating that the predictive performance of the BN is primarily driven by cross-sectional market and sector dependencies, while temporal effects play a limited role, in line with standard stylized facts of daily returns.

Figure 1 shows the predicted versus realized returns for asset A001 on the test set. The concentration of points around the 45-degree line indicates that the model captures the **systematic component** of returns, while deviations reflect **idiosyncratic noise** typical of daily returns. The absence of strong **bias** or **nonlinear patterns** suggests that the linear Gaussian CPDs provide an adequate approximation of the **conditional relationships** encoded in the BN, leading to reasonable predictive performance on unseen data.

Figure 1: Predicted versus true returns for asset A001 on the test set.



2 Part II – Markov network

In this part, we model the *conditional dependence structure among financial assets* using a **Markov Random Field** (MRF), an **undirected graphical model** (UGM) designed to capture contemporaneous interactions in multivariate systems. Within this framework, each *node* represents an individual asset return (A001–A100), grouped into ten economic sectors (SEC01–SEC10), while *edges* encode **symmetric conditional dependence relationships** that persist after controlling for all other assets in the system. Importantly, since the graph is undirected, edges **do not imply causality** or **directionality**, but instead represent contemporaneous **residual comovements between assets**.

In our setting, the graph structure and the model parameters are **unknown** *a priori*. Consequently, the **conditional dependence relationships** among assets must be **inferred from data**. This requires a **learning phase** prior to inference, based on a **training dataset**, involving both **structure** and **parameter learning**. Under the assumption of Gaussian asset returns, this yields a fully specified **Gaussian Graphical Model** (GGM), which provides a tractable and interpretable framework for modeling and inferring conditional dependence relationships among **continuous financial returns**.

2.1 Structure and Parameter Learning

Structure learning aims at **discovering the graphical structure** itself, that is, the **set of edges** connecting the nodes, which represents **conditional dependence relationships** between variables.

Under the Gaussian assumption, the joint distribution of **standardized asset returns** $X = (X_1, \dots, X_p)^\top \sim \mathcal{N}(0, \Sigma)$, with $p = 100$, is fully characterized by its covariance matrix Σ or, equivalently, by its precision matrix $\Lambda = \Sigma^{-1}$. A fundamental property of GGMs is that conditional independence relationships are directly encoded by the precision matrix:

$$\Lambda_{ij} = 0 \quad \iff \quad X_i \perp X_j \mid X_{-\{i,j\}}. \tag{10}$$

where $X_{-\{i,j\}}$ denotes the collection of all variables except X_i and X_j. Consequently, learning the graphical structure reduces to identifying the **sparsity pattern** of the **precision matrix**, where **non-zero off-diagonal entries** correspond to **edges** in the underlying undirected graph.

Before estimating the GGM, asset return series are standardized using the mean and standard deviation computed on the **training dataset**, defined as the first 70% of the observations, to ensure that all variables are on a **comparable scale** during model estimation.

Structure learning is performed by estimating a sparse precision matrix using a **score based method**, as motivated in Part I: the **Graphical Lasso**, which solves an ℓ_1-regularized maximum likelihood problem, as explained in part I why. The ℓ_1 penalty promotes sparsity in the estimated precision matrix, reflecting the assumption that each asset is directly related to only a limited subset of other assets. In this setting, **structure** and **parameter learning** are **performed jointly**, as both the sparsity pattern and the numerical values of the precision matrix are estimated simultaneously.

For a given regularization parameter $\lambda > 0$, the estimated precision matrix $\hat{\Lambda}(\lambda)$ is obtained as:

$$\hat{\Lambda}(\lambda) = \arg\min_{\Lambda \succ 0} \left\{ -\log\det(\Lambda) + \operatorname{tr}(S\Lambda) + \lambda \sum_{i \neq j} |\Lambda_{ij}| \right\}, \tag{11}$$

where $S = \frac{1}{T} \sum_{t=1}^{T} x^{(t)} x^{(t)\top}$ denotes the empirical covariance matrix estimated from $T = 1000$ daily observations. The regularization parameter $\lambda > 0$ controls the strength of this sparsity-inducing penalty and is selected using the **BIC**, as our goal in this part is to recover a **sparse** and **interpretable conditional dependence structure** rather than to maximize pure predictive performance. The GGM estimated via the Graphical Lasso, using the regularization parameter 0.061, yields a sparse representation of conditional dependencies among the 100 assets.

Figure 2 is the **inferred Markov network**, where nodes represent individual assets and edges encode direct conditional dependence relationships. Then, Figure 3 shows the **sparsity pattern** of the estimated precision matrix, where **most off-diagonal entries** are **zero**, indicating widespread **conditional independence**, while block-like patterns along the diagonal reveal stronger within-group conditional dependencies and weaker cross-group interactions.

Figure 2: GGM.

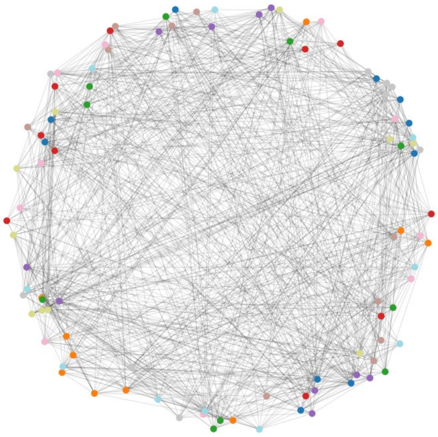


Figure 3: Sparsity pattern of the estimated precision matrix.

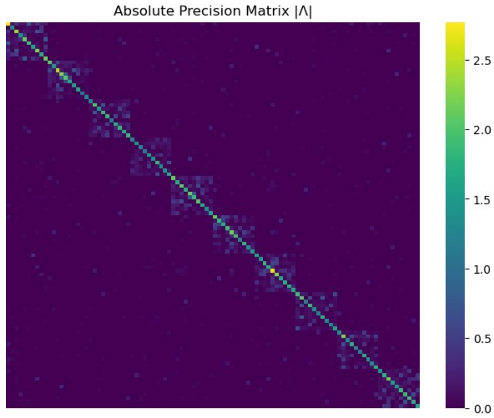
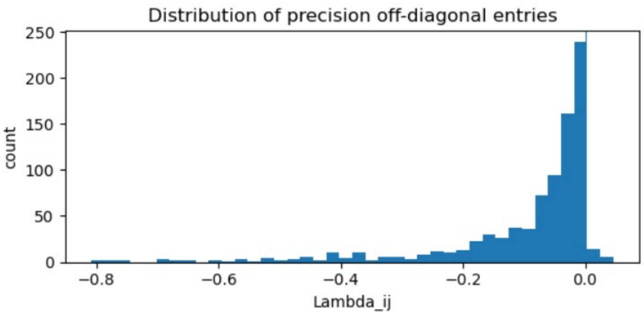


Figure 4 shows that the **off-diagonal entries** of the estimated precision matrix are **strongly concentrated around zero**, indicating predominantly **weak conditional dependencies** and a **sparse network** driven by a **few strong direct interactions**.

Figure 4: Distribution of the off-diagonal entries of the estimated precision matrix.



Once the precision matrix has been estimated, the learned graph structure is obtained by defining an undirected edge between nodes i and j whenever $|\hat{\Lambda}_{ij}| > \varepsilon$, for a small threshold $\varepsilon > 0$. The estimated graph is obtained from $\hat{\Lambda}$ by defining the **edge set**:

$$\hat{\mathcal{E}} = \{(i,j) : i < j,\ |\hat{\Lambda}_{ij}| > \varepsilon\}. \tag{12}$$

The learned network is sparse, with 859 edges corresponding to a **sparsity level** of approximately 83% **in the off-diagonal entries** of the precision matrix, indicating that most asset pairs are **conditionally independent** once the rest of the system is taken into account. In other words, while assets may appear correlated at a marginal level, only a relatively **small subset of relationships** reflects **direct interactions**. Moreover, we identify connected components using **Breadth-First Search (BFS)**, which traverses the network by iteratively visiting neighboring nodes. The estimated network forms a **fully connected dependency structure**. Finally, a **sectoral decomposition** of the estimated edges shows that 45% **of connections** occur within the **same sector**, while the remaining across different sectors. This suggests that, although sectoral structure plays an important role, a substantial fraction of conditional dependencies captures **economically meaningful cross-sector interactions**.

2.2 Model validation

After learning the GGM on the training set, model performance is evaluated on a **held-out test set** as a **robustness check**. Given the estimated precision matrix $\hat{\Lambda}$, the implied covariance matrix $\hat{\Sigma} = \hat{\Lambda}^{-1}$ fully characterizes the joint Gaussian distribution of asset returns. Model validation is performed using the **out-of-sample (OOS) Gaussian log-likelihood**: $\ell_{\text{test}}(\hat{\Lambda}) = \log\det(\hat{\Lambda}) - \text{tr}(S_{\text{test}}\hat{\Lambda})$, where S_{test} denotes the empirical covariance matrix of the standardized test observations. The BIC-selected sparse model achieves an OOS log-likelihood of -55.05, substantially outperforming the dense baseline (-62.50), indicating that **regularization improves generalization**. Although a slightly smaller regularization parameter maximizes the OOS likelihood, the difference is negligible. So, as our goal is structure learning rather than prediction, the **BIC-selected model** is **retained**.

2.3 Inference

Once the graphical structure and model parameters have been learned, inference consists in **using the fitted model** to **compute probabilistic quantities of interest**. Under the **Gaussian assumption**, the joint distribution of asset returns is **fully characterized** by the **estimated covariance matrix** $\hat{\Sigma}$, and conditional inference can be performed exactly in **closed form**.

In particular, we consider **conditional inference**. Let the set of variables be partitioned into two subsets A and B, where the variables in A are **observed** and the variables in B are **unobserved**. The conditional distribution of X_B given $X_A = a$ is Gaussian and is given by:

$$X_B \mid X_A = a \sim \mathcal{N}\left(\Sigma_{BA}\Sigma_{AA}^{-1}a,\ \Sigma_{BB} - \Sigma_{BA}\Sigma_{AA}^{-1}\Sigma_{AB}\right), \tag{13}$$

where the covariance submatrices are extracted from $\hat{\Sigma}$.

Table 2 reports the results of probabilistic inference for the unobserved assets $A002$ and $A003$, conditional on the observed assets $A001$, $A010$ and $A050$, within the GGM framework. Both inferred assets exhibit **positive conditional means**, with a **stronger response observed** for A003. Moreover, the conditional variance analysis highlights substantial **heterogeneity in information transmission across assets**. While the variance of A002 is reduced by approximately 15%, suggesting a largely **idiosyncratic behavior**, the conditional variance of A003 decreases by more than 44%, indicating a **strong conditional dependence** on the **observed asset set**. So the GGM captures differences in systemic exposure and risk explainability across assets through probabilistic inference.

Table 2: Conditional inference for unobserved assets B given given the observed set $A = \{\text{A001} = 0.5,\ \text{A010} = -0.2,\ \text{A050} = 0.1\}$.

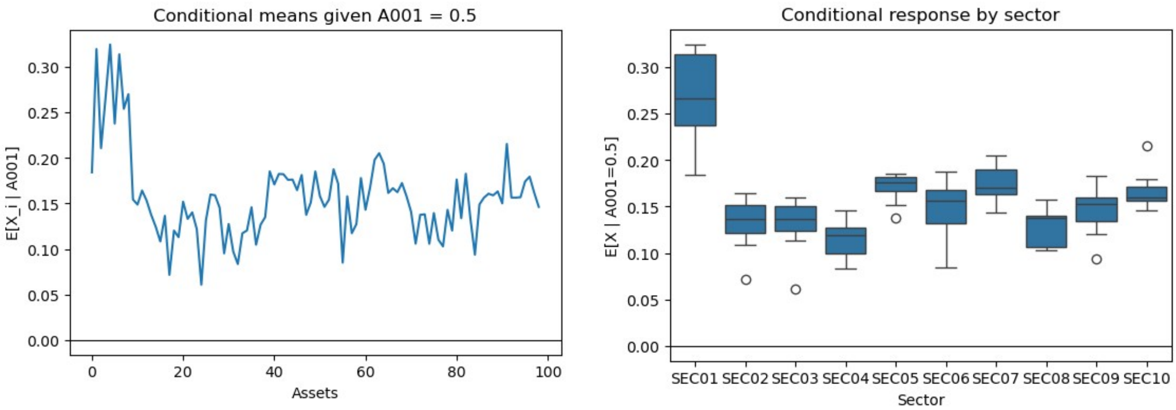
Asset	$\mathbb{E}[X \mid A = a]$	$\text{Var}[X \mid A = a]$	$\text{Var}(X)$	Variance reduction
A002	0.1294	0.8539	1.0014	14.73%
A003	0.2271	0.5582	1.0014	44.26 %

Figure 5 illustrates the conditional response of the GGM model when conditioning on a moderate positive standardized shock of 0.5 standard deviations to asset A001. Panel (a) displays the asset-level conditional **expected returns** for all remaining assets. The responses are **predominantly positive** but **heterogeneous** in magnitude, reflecting differences in conditional connectivity and exposure to the shocked asset within the learned network structure. Panel (b) aggregates these **conditional responses** at the **sector level**. The boxplots clearly show that the strongest **transmission of the shock** occurs **within the sector** of the shocked asset, where both the median response and dispersion are higher. **Other sectors** exhibit **weaker** yet non-negligible **positive responses**, indicating the presence of **cross-sector spillover effects** mediated by the **conditional dependence structure**.

Figure 5: Propagation of a positive asset-level shock in the GGM.

(a) Conditional expected returns at the asset level given a positive shock to A001.

(b) Sector-level distribution of conditional expected returns.



3 Part III – Discussion

3.1 Identification of dependencies

The two models identify **complementary types of dependencies**. The BN explicitly encodes a **hierarchical and directed structure** consistent with the assumed economic organization of the market and therefore captures dependencies that the MRF cannot represent, namely **market-to-sector relations** (MKT $\rightarrow$ SEC$_k$), **sector-to-asset relations** (SEC$_k$ $\rightarrow$ A$_i$), and **occasional temporal self-dependencies** through lagged edges. These dependencies are imposed by the graph structure and validated by data via structure learning.

Conversely, the MRF includes only asset returns and models **symmetric conditional dependencies** after conditioning on all others. As a result, it uncovers dependencies that are invisible to the BN, in particular **residual within-sector dependencies** beyond common sector effects and **cross-sector conditional interactions** reflecting indirect economic linkages. Hence, the BN explains dependence through **explicit common factors**, while the MRF reveals **residual asset-level interactions** once those factors are removed.

3.2 Symmetry versus directionality

A central conceptual distinction between the two approaches lies in the **nature of the graph edges**. The **BN** is **directed** and **acyclic**, allowing edges to be interpreted as **asymmetric influence** or **causal ordering**, conditional on the modeling assumptions. In our context, this directionality aligns naturally with financial intuition: market movements affect sectors, which in turn affect individual assets. This structure enables a **generative interpretation** and supports sequential **prediction** and **scenario analysis**. In contrast, the **MRF** is **undirected**, with **edges** representing **symmetric** conditional dependence relationships without implying causality. This symmetry makes the MRF well suited for **exploratory analysis** and for **identifying tightly connected groups of assets**.

3.3 Inference

Inference also differs substantially across the two frameworks. In **BN**, inference is directional and follows the structure of the graph. Conditional distributions such as $P(A_i \mid \mathrm{SEC}_k)$ or $P(\mathrm{SEC}_k \mid \mathrm{MKT})$ are computed via the learned conditional probability distributions, allowing for hierarchical prediction and decomposition of asset returns into market-driven, sector-driven, and idiosyncratic components.

In the **MRF**, inference relies on the joint **Gaussian distribution** characterized by the estimated covariance matrix. Conditional distributions of the form $P(B \mid A)$ are symmetric and computed via standard Gaussian conditioning formulas. Conditioning on a shock to one asset propagates through the network according to the precision matrix, generating spillover effects whose magnitude depends on the strength of conditional connections.

Empirically, conditioning on a single asset shock in the **MRF** yields **localized effects within the same sector and weaker effects across sectors**. This contrasts with the **BN**, where conditioning at the market or sector level induces **broad and structured responses across many assets**. Thus, **BN** inference emphasizes **top-down transmission mechanisms**, whereas **MRF** inference highlights **local propagation** through conditional dependence links.

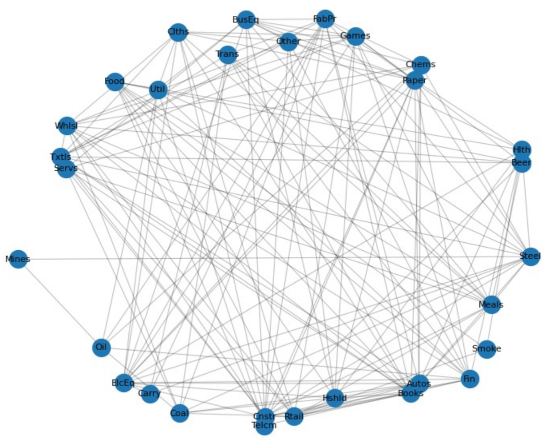
3.4 Application to Real Data: 30 Industry Portfolios

We apply a **Markov Network** to the **30 Industry Portfolios** monthly dataset ($T = 1194$), as it captures **direct conditional dependence relationships among assets** without imposing directional or causal assumptions, which is appropriate for **symmetric contemporaneous interactions** in financial returns. The optimal regularization parameter is $\alpha = 0.013$, indicating a relatively low level of regularization and resulting in a dense conditional dependence structure.

The estimated network contains **255 edges** out of **435 possible connections**, corresponding to a density of approximately 58.6%, reflecting residual dependencies across industries after controlling for all others. This suggests the presence of **global common factors** affecting multiple sectors simultaneously. The **off-diagonal sparsity** of the estimated precision matrix equals 41.4%, indicating that more than half of the potential conditional dependencies across industries are retained.

Figure 6 shows the learned network, where industries exhibit a **strongly interconnected structure** with visible local clustering among economically related sectors, alongside numerous cross-sector links. This pattern highlights the importance of **market-wide shocks and common risk factors**, which generate widespread conditional dependence even after accounting for sector-specific effects. Overall, the results confirm that the Markov Network captures rich cross-industry interactions.

Figure 6: Markov Network estimated on the 30 Industry Portfolios dataset.



In Table 3, conditioning on the observed industries leads to **positive conditional means** for *Beer* and *Smoke*, together with **variance reductions** above 50%. This indicates **strong conditional dependence** within the estimated network.

Table 3: Conditional inference for unobserved industries B given the observed set $A = \{\text{Food} = 0.8,\ \text{Hshld} = -0.4,\ \text{Cnstr} = 0.2\}$.

Industry	$\mathbb{E}[X \mid A = a]$	$\text{Var}[X \mid A = a]$	$\text{Var}(X)$	Variance reduction
Beer	0.2826	0.4379	1.0012	56.26%
Smoke	0.4809	0.4665	1.0012	53.40%

In conclusion, the MRF estimated on the 30 Industry Portfolios data reveals a dense network of residual conditional dependencies across industries. Significant within-sector clustering and cross-sector links persist even after controlling for all others, pointing to strong common risk factors. Overall, the MRF effectively captures rich contemporaneous inter-industry interactions in real financial data.