High Dimensional Models

Time-Varying Graphical Lasso

Andrew Boomer & Jacob Pichelmann

Toulouse School of Economics
M2 EEE

March 7, 2021

Network Theory in Practice (?, ?)

- The Google search algorithm, PageRank, is based on network theory, and is an extension of a measure of network importance called eigenvector centrality. This measure assigns higher importance to nodes that are connected to other highly connected nodes. Search results are sorted according to this network importance vector.

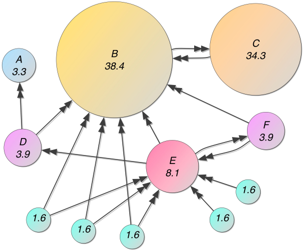


Figure: Google Page Rank

Introduction to Graphical Models

- A graph is represented by a set of vertices $V = \{1, 2, \cdots, p\}$ and by a set of edges $E \subset VxV$
- A graph is undirected when there is no distinction between the edge (s,t) and $(t,s) \quad \forall\ s,t \in E$
- This graphical representation can be extended to a high dimensional set of random variables X_s.
- In this example, s corresponds to a single vertex with the whole vertex set V of the total graph. The connections between each vertex in the Markov transition matrix representation quantify the relationship between this random variables.

The two visualizations in Figure 2 are equivalent, where white spaces indicate no relationship, and grey spaces indicate a 1-to-1 relationship.

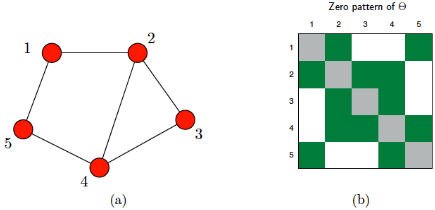


Figure: Graph and Sparsity Matrix

Gaussian Graphical Models

- For a gaussian graphical model, we start from a gaussian distribution with dimension equal to the number of vertices $X \sim \mathcal{N}(\mu, \Sigma)$
- The sparsity matrix in the slide above, for a gaussian graphical model, is given by $\Theta = \Sigma^{-1}$, and is also known as the precision matrix. The goal of a gaussian graphical model is to solve for Θ.
- The gaussian distribution can be rewritten in terms of a maximum likelihood problem to solve for $\hat{\Theta}_{MLE}$, where S is the empirical covariance matrix, rearranged and simplified to:

$$\mathcal{L}(\Theta; X) = log\ det\ \Theta - trace(S\Theta)$$

- This MLE problem converges to the true precision matrix when $N \to \infty$.

Graphical LASSO

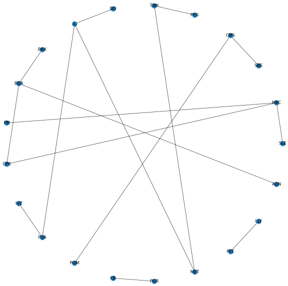
- The MLE problem is **not feasible if** $p > N$, where p is the number of dimensions.
- As with a LASSO in the context of a linear regression, adding a **regularization term can solve the issue** of a non-full rank matrix.
- Additionally, the LASSO term induces low-importance terms to zero. Weak edge connections in the precision matrix will go to zero, increasing the sparsity of the resulting solution. Increased sparsity aids in interpretability by filtering out noisy relationships, and can be used even when N is close to p.
- The graphical lasso estimator is the $\hat{\Theta}$ such that:

$$\hat{\Theta} = \text{argmin}_{\Theta \geq 0} \left(\text{tr}(S\Theta) - \log \det(\Theta) + \lambda \sum_{j \neq k} |\Theta_{jk}| \right)$$

λ in the Context of Networks

- The penalty parameter determines the structure of the network.
- Typical methods can be applied, e.g. cross validation

$\lambda = 0.05$



$\lambda = 0.005$

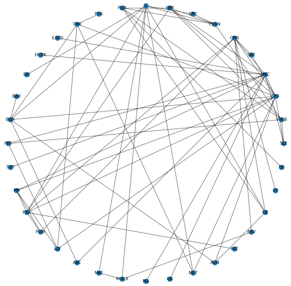
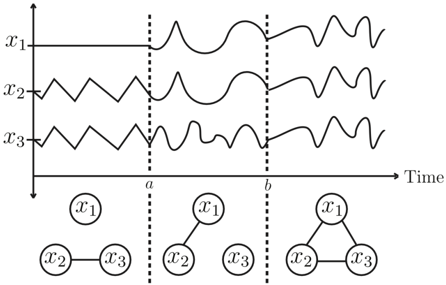


Figure: Tuning λ, own calculations based on data from (?, ?)

Challenge: The Network Structure Can Change Over Time

In many real world settings (e.g. financial markets) the structure of the complex system changes over time.



Figure

Solution: Optimization on a Chain Graph (TVGL)

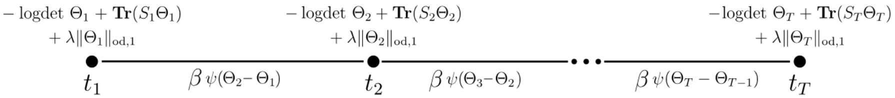


Figure: (?, ?)

The optimization problem becomes

$$\underset{\Theta \in \mathrm{S}^p_{++}}{\text{minimize}} \quad \sum_{i=1}^{T} -l_i(\Theta_i) + \lambda \|\Theta_i\|_{\mathrm{od},1} + \beta \sum_{i=2}^{T} \psi(\Theta_i - \Theta_{i-1})$$

where β determines how strongly correlated neighboring covariance estimations should be. A small β will lead to θ's which fluctuate from estimate-to-estimate, whereas large β's lead to smoother estimates over time.

Choice of ψ

- ψ allows to enforce different behaviors in the evolution of the network structure
- Expectations how the underlying network may change over time can be encoded into ψ

Options:

- **Global restructuring** - $\psi(X) = \sum_j \|[X]_j\|_2$
- **Smoothly varying over time** - $\psi(X) = \sum_{i,j} X_{i,j}^2$
- **Perturbed node** - $\psi(X) = \min_{V:V+V^T=X} \sum_j \|[V]_j\|_2$

Optimization Algorithm

- The authors use the alternating direction method of multipliers (ADMM), a distributed convex optimization approach.
- ADMM allows splitting the problem up into a series of sub-problems.
- They then use a message-passing algorithm to converge on the globally optimal solution.
- Making use of proximal operators they derive closed form solutions for each of the ADMM sub-problems.

GLASSO vs TVGL

Time varying graphical lasso with $\psi(X) = \min_{V : V + V^T = X} \sum_j \|[V]_j\|_2$ (perturbed node)

Importance of ψ

- Choice of ψ relies on knowledge about network behavior
- No a priori decision possible
- ψ is fixed over time

We illustrate the importance of the choice of ψ by replicating the author's case studies with different penalty functions.

Changing ψ

Extension: Cross Validation for ψ

Extension: Understanding Network Importance

- The eigenvector centrality allows additional insight into the network's dynamic.

References