High Dimensional Models

Time-Varying Graphical Lasso

Andrew Boomer & Jacob Pichelmann

Toulouse School of Economics
M2 EEE

March 6, 2021

Application of Network Theory (?, ?)

- 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.
- The PageRank measure of importance ends up being the long run probability of a random surfer finding themselves on any given webpage. The search results are sorted according to this network importance vector.

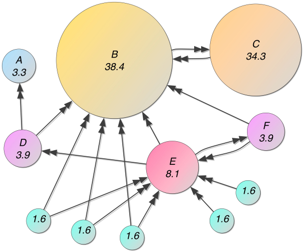
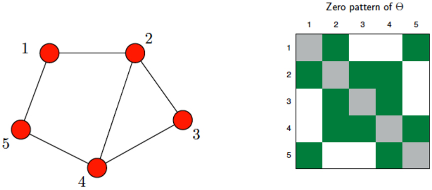


Figure: Google Page Rank

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.



(a) (b)

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 $N \to \infty$. **The MLE problem is, however, not feasible if $p > N$, where p is the number of dimensions.**

Graphical LASSO

- 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.
- In the case of the graphical LASSO, weak edge connections in the precision matrix will go to zero, increasing the sparsity of the resulting solution. This 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:

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

Application

- The penalty parameter determines the structure of the network.

$\alpha = 0.05$ $\alpha = 0.005$

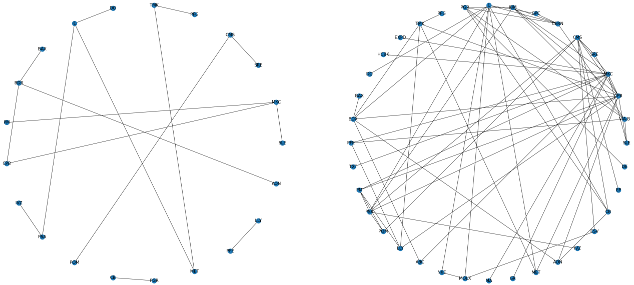
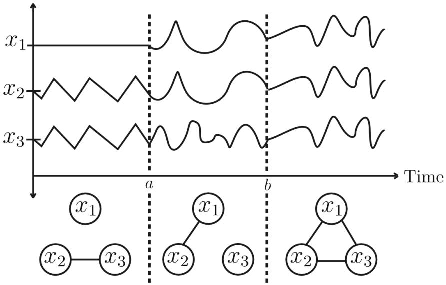


Figure: Tuning α

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)

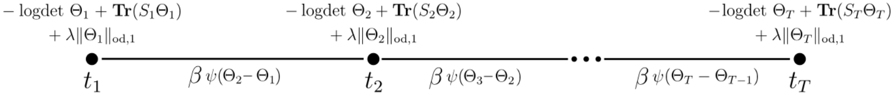


Figure: (?, ?)

The optimization problem becomes

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

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:

- **A few edges changing at a time** - $\psi(X) = \sum_{i,1} |X_{i,j}|$
- **Global restructuring** - $\psi(X) = \sum_{j} ||[X]_j||_2$
- **Smoothly varying over time** - $\psi(X) = \sum_{i,j} X_{i,j}^2$
- **Block-wise restructuring** - $\psi(X) = \sum_{j} (max_i |X_{i,j}|)$
- **Perturbed node** - $\psi(X) = \min_{V:V+V^T=X} \sum_j \|[V]_j\|_2$

Optimization Algorithm

- How do we arrive at closed form solution?

TVGL Application

Limitations

- Does $X \sim \mathcal{N}(\mu, \Sigma)$ hold?
- Choice of ψ relies on knowledge about network behavior. No a priori decision possible.

References