GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT
UNIVERSITÉ DE GENÈVE

GLAMLE: a novel inference procedure for networks, in the presence of latent variables

Davide La Vecchia
joint work with C. Jiang and R. Rastelli

University of Bern, 17-Feb -2025

Structure

- FAO commodities trading data among 28 EU Countries: conduct inference on complex/connected system
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- The model (GGLLVM) and the estimator (GLAMLE)
  - Definition
  - Inference procedure and key properties
- Synthetic data and FAO data (reprise)
- Take home message: GLAMLE is
  - More accurate than VA and faster than MCMC
  - It offers novel: (a) options for data anaylsis (insights into network structure) and (b) inference on the propensity to trade of each Country
  - Connections to factor models for random fields over graphs

Structure

- FAO commodities trading data among 28 EU Countries: conduct inference on complex/connected system
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- The model (GGLLVM) and the estimator (GLAMLE)
  - Definition
  - Inference procedure and key properties
- Synthetic data and FAO data (reprise)
- Take home message: GLAMLE is
  - More accurate than VA and faster than MCMC
  - It offers novel: (a) options for data anaylsis (insights into network structure) and (b) inference on the propensity to trade of each Country
  - Connections to factor models for random fields over graphs

Structure

- FAO commodities trading data among 28 EU Countries: conduct inference on complex/connected system
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- The model (GGLLVM) and the estimator (GLAMLE)
  - Definition
  - Inference procedure and key properties
- Synthetic data and FAO data (reprise)
- Take home message: GLAMLE is
  - More accurate than VA and faster than MCMC
  - It offers novel: (a) options for data anaylsis (insights into network structure) and (b) inference on the propensity to trade of each Country
  - Connections to factor models for random fields over graphs

Structure

- FAO commodities trading data among 28 EU Countries: conduct inference on complex/connected system
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- The model (GGLLVM) and the estimator (GLAMLE)
  - Definition
  - Inference procedure and key properties
- Synthetic data and FAO data (reprise)
- Take home message: GLAMLE is
  - More accurate than VA and faster than MCMC
  - It offers novel: (a) options for data anaylsis (insights into network structure) and (b) inference on the propensity to trade of each Country
  - Connections to factor models for random fields over graphs

Structure

- FAO commodities trading data among 28 EU Countries: conduct inference on complex/connected system
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- The model (GGLLVM) and the estimator (GLAMLE)
  - Definition
  - Inference procedure and key properties
- Synthetic data and FAO data (reprise)
- Take home message: GLAMLE is
  - More accurate than VA and faster than MCMC
  - It offers novel: (a) options for data anaylsis (insights into network structure) and (b) inference on the propensity to trade of each Country
  - Connections to factor models for random fields over graphs

Structure

- FAO commodities trading data among 28 EU Countries: conduct inference on complex/connected system
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- The model (GGLLVM) and the estimator (GLAMLE)
  - Definition
  - Inference procedure and key properties
- Synthetic data and FAO data (reprise)
- Take home message: GLAMLE is
  - More accurate than VA and faster than MCMC
  - It offers novel: (a) options for data anaylsis (insights into network structure) and (b) inference on the propensity to trade of each Country
  - Connections to factor models for random fields over graphs

Structure

- FAO commodities trading data among 28 EU Countries: conduct inference on complex/connected system
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- The model (GGLLVM) and the estimator (GLAMLE)
  - Definition
  - Inference procedure and key properties
- Synthetic data and FAO data (reprise)
- Take home message: GLAMLE is
  - More accurate than VA and faster than MCMC
  - It offers novel: (a) options for data anaylsis (insights into network structure) and (b) inference on the propensity to trade of each Country
  - Connections to factor models for random fields over graphs

Related published papers

Econometrics and Statistics
Available online 25 September 2024
In Press, Corrected Proof What's this?

GLAMLE: inference for multiview network data in the presence of latent variables, with an application to commodities trading

Chaonan Jiang[a], Davide La Vecchia[b], Riccardo Rastelli[c]

Theory and Methods

Saddlepoint Approximations for Spatial Panel Data Models

Chaonan Jiang, Davide La Vecchia, Elvezio Ronchetti & Olivier Scaillet

Received 01 Jul 2020, Accepted 09 Sep 2021, Accepted author version posted online: 20 Sep 2021, Published online: 17 Nov 2021
Download citation https://doi.org/10.1080/01621459.2021.1981913 Check for updates

Statistics > Methodology

[Submitted on 5 Dec 2023]

General Spatio-Temporal Factor Models for High-Dimensional Random Fields on a Lattice

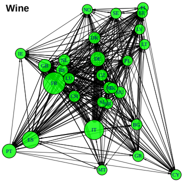
Matteo Barigozzi, Davide La Vecchia, Hang Liu

Motivated by the need for analysing large spatio-temporal panel data, we introduce a novel dimensionality reduction methodology for n-dimensional random fields observed across a number S spatial locations and T time periods. We call it General Spatio-Temporal Factor Model (GSTFM). First, we provide the probabilistic and mathematical underpinning needed for the representation of a random field as the sum of two components: the common component (driven by a small number q of latent factors) and the idiosyncratic component (mildly cross-correlated). We show that the two components are identified as $n \to \infty$. Second, we propose an estimator of the common component and derive its statistical guarantees (consistency and rate of convergence) as $\min(n, S, T) \to \infty$. Third, we propose an information criterion to determine the number of factors. Estimation makes use of Fourier analysis in the frequency domain and thus we fully exploit the information on the spatio-temporal covariance structure of the whole panel. Synthetic data examples illustrate the applicability of GSTFM and its advantages over the extant generalized dynamic factor model that ignores the spatial correlations.

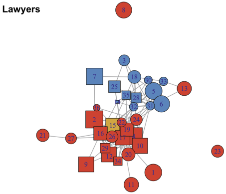
Challenge

Inference on **network data** is a very active research area; see e.g. Kolaczyk (2009); Lusher et al. (2013); Kolaczyk and Csardi (2014), just to mention book-length introductions.

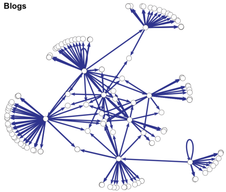
Economics



Social Sciences



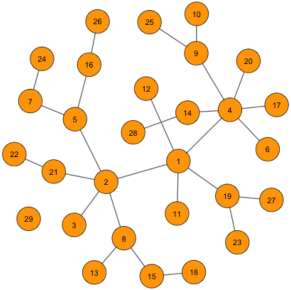
Internet



The data collected on those systems are of different nature: e.g. they can be binary, count, continuous data or mixture thereof.

Challenge

Networks are represented by means of **graphs**. Let $\mathcal{G} = (V, E)$ where V is the set of vertices/nodes and E is the set of edges (connections between nodes)

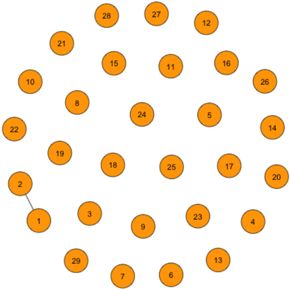


Remark (Research question)

What is the underlying probabilistic mechanism/model that explains the edges? How to define a network probability model that embeds (observable and) latent factors? Put it in another way: how can we learn the mechanism generating the graph structure?

Challenge

Networks are represented by means of **graphs**. Let $\mathcal{G} = (V, E)$ where V is the set of vertices/nodes and E is the set of edges (connections between nodes)



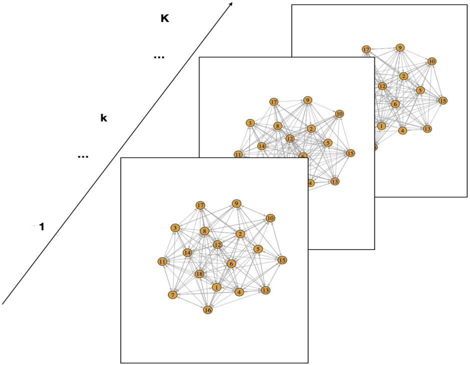
Remark (Research question)

What is the underlying probabilistic mechanism/model that explains the edges? How to define a network probability model that embeds (observable and) latent factors? Put it in another way: how can we learn the mechanism generating the graph structure?

Challenge

Among the different types of networks, **multiview networks** (Gollini and Murphy, 2016; Salter-Townshend and McCormick, 2017; Sosa and Betancourt, 2022) have been recently attracting the attention of the research community. **This type of networks consists of multiple layers of interactions among the same nodes**.

We have K random samples $Y^{(1)}, Y^{(2)}, \cdots, Y^{(K)}$.



FAO commodities trading data

The dataset contains records on different types of trade relationships among 28 European countries, obtained from the FAO in 2010. This network dataset has been attracting the attention of both the empirical and theoretical research communities, becoming a benchmark for the analysis of multiplex network data; see, among the others, Rahmede et al. (2018) and Yuan and Qu (2021).

Some details:

- The import/export network is an economic network in which 364 layers represent commodities, nodes are countries, and the edges at each layer represent import/export relationships of a specific commodity among countries. The edges of each of the networks are directed.
- We define the edge weight being equal to one when there exists a commercial exchange among two nodes, or zero otherwise: the data is a Bernoulli. An alternative option we are currently considering is the use of a Poisson to model the quantity traded (gravity model)

FAO commodities trading data

The dataset contains records on different types of trade relationships among 28 European countries, obtained from the FAO in 2010. This network dataset has been attracting the attention of both the empirical and theoretical research communities, becoming a benchmark for the analysis of multiplex network data; see, among the others, Rahmede et al. (2018) and Yuan and Qu (2021).

Some details:

- The import/export network is an economic network in which 364 layers represent commodities, nodes are countries, and the edges at each layer represent import/export relationships of a specific commodity among countries. The edges of each of the networks are directed.
- We define the edge weight being equal to one when there exists a commercial exchange among two nodes, or zero otherwise: the data is a Bernoulli. An alternative option we are currently considering is the use of a Poisson to model the quantity traded (gravity model)

FAO commodities trading data

The dataset contains records on different types of trade relationships among 28 European countries, obtained from the FAO in 2010. This network dataset has been attracting the attention of both the empirical and theoretical research communities, becoming a benchmark for the analysis of multiplex network data; see, among the others, Rahmede et al. (2018) and Yuan and Qu (2021).

Some details:

- The import/export network is an economic network in which 364 layers represent commodities, nodes are countries, and the edges at each layer represent import/export relationships of a specific commodity among countries. The edges of each of the networks are directed.
- We define the edge weight being equal to one when there exists a commercial exchange among two nodes, or zero otherwise: the data is a Bernoulli. An alternative option we are currently considering is the use of a Poisson to model the quantity traded (gravity model)

FAO commodities trading data

The dataset contains records on different types of trade relationships among 28 European countries, obtained from the FAO in 2010. This network dataset has been attracting the attention of both the empirical and theoretical research communities, becoming a benchmark for the analysis of multiplex network data; see, among the others, Rahmede et al. (2018) and Yuan and Qu (2021).

Some details:

- The import/export network is an economic network in which 364 layers represent commodities, nodes are countries, and the edges at each layer represent import/export relationships of a specific commodity among countries. The edges of each of the networks are directed.
- We define the edge weight being equal to one when there exists a commercial exchange among two nodes, or zero otherwise: the data is a Bernoulli. An alternative option we are currently considering is the use of a Poisson to model the quantity traded (gravity model)

FAO commodities trading data

The dataset contains records on different types of trade relationships among 28 European countries, obtained from the FAO in 2010. This network dataset has been attracting the attention of both the empirical and theoretical research communities, becoming a benchmark for the analysis of multiplex network data; see, among the others, Rahmede et al. (2018) and Yuan and Qu (2021).

Some details:

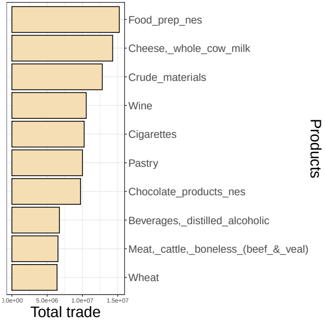
- The import/export network is an economic network in which 364 layers represent commodities, nodes are countries, and the edges at each layer represent import/export relationships of a specific commodity among countries. The edges of each of the networks are directed.
- We define the edge weight being equal to one when there exists a commercial exchange among two nodes, or zero otherwise: the data is a Bernoulli. An alternative option we are currently considering is the use of a Poisson to model the quantity traded (gravity model)

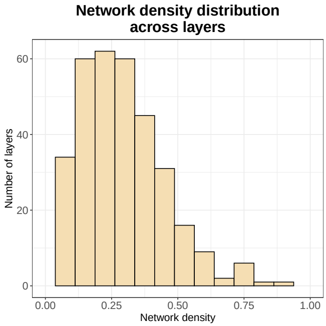
FAO commodities trading data (overview)

Different products traded by each Country...

Belgium	343	Austria	327	Sweden	299	Luxembourg	265
Germany	343	Czech Republic	326	Ireland	297	Estonia	251
Netherlands	341	Poland	321	Slovakia	297	Slovenia	248
France	338	Hungary	317	Lithuania	294	Finland	245
Spain	337	Denmark	313	Bulgaria	283	Croatia	175
Italy	336	Portugal	304	Romania	276	Cyprus	173
United Kingdom	335	Greece	299	Latvia	275	Malta	81

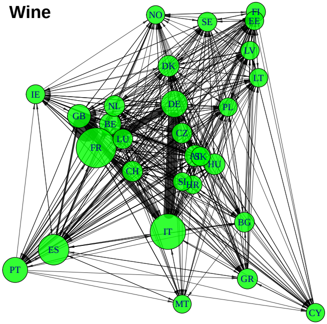
Top 10 products

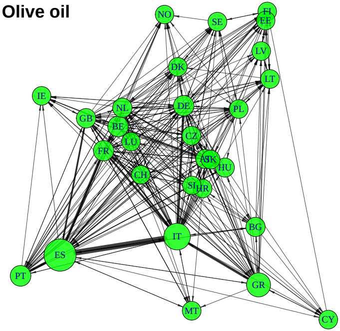




... and we visualize two layers as directed graphs...

Trades volumes (in \$) between EU countries wine and olive oil:





FAO commodities trading data

Inferential goal: We aim at modeling the **probability** governing *commercial relationships*. The statistical issue is that, often, this probability depends on some **unobservable factors**, including:

- socio-economical conditions;
- political views;
- level of the infrastructures.

Our solution:

- we formulate a statistical model, in which a **latent variable framework** allows to capture the *latent structure*: we propose a statistical model that relates a small number of unobserved variables to observed variables in some way
- we derive a **novel inference approach** able to estimate the latent variables and the related parameters
- we provide the ultimate user with a computationally **efficient inference** procedure and algorithm.

FAO commodities trading data

Inferential goal: We aim at modeling the **probability** governing *commercial relationships*. The statistical issue is that, often, this probability depends on some **unobservable factors**, including:

- socio-economical conditions;
- political views;
- level of the infrastructures.

Our solution:

- we formulate a statistical model, in which a **latent variable framework** allows to capture the *latent structure*: we propose a statistical model that relates a small number of unobserved variables to observed variables in some way
- we derive a **novel inference approach** able to estimate the latent variables and the related parameters
- we provide the ultimate user with a computationally **efficient inference** procedure and algorithm.

FAO commodities trading data

Inferential goal: We aim at modeling the **probability** governing *commercial relationships*. The statistical issue is that, often, this probability depends on some **unobservable factors**, including:

- socio-economical conditions;
- political views;
- level of the infrastructures.

Our solution:

- we formulate a statistical model, in which a **latent variable framework** allows to capture the *latent structure*: we propose a statistical model that relates a small number of unobserved variables to observed variables in some way
- we derive a **novel inference approach** able to estimate the latent variables and the related parameters
- we provide the ultimate user with a computationally **efficient inference** procedure and algorithm.

Gravity model, Factor models

To learn the mechanism generating the graph structure in the FAO data, our focus is on the adjacency matrix (self-edges are not allowed) contains entries

$$\mathbf{Y} = \{Y_{ij},\ (ij) \in (V \times V)\},$$

where

$$Y_{ij} = \begin{cases} 1 & \text{if an edge from } i \text{ to } j \text{ appears in the graph} \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

Remark (label=foo)

The use of a Bernoulli Y_{ij} is dictated by the considered real-data motivation. However, other distributions belonging to the exponential family can be considered:

- *Using **covariates** for Y_{ij} and **unobservable** variables, one can find connections to the **gravity model** (Anderson and Van Wincoop, AER, 2003) working on **Poisson** pmf for volume of trades* Gravity
- *In the case of Gaussian rvs, we can find several analogies with **Bai's static factor model** (Econometrica, 2009)*

Both of them need further exploration...

Related work

One popular class of network models is the so-called Exponential Random Graph Model (ERGM). See, e.g., Frank and Strauss (1986), Wasserman and Pattison (1996), Snijders (2002), Robins et al. (2007) and Lusher et al. (2013): available for either undirected or directed graphs and some network statistics (like e.g. the number of edges, triangles, stars) in combination with some observable variables explain the dependence among network data.

More precisely, for the **Bernoulli rvs** in the random adjacency matrix, an ERGM takes the form

$$P_{\boldsymbol{\theta}}(\boldsymbol{Y} = \boldsymbol{y}) = \exp\left\{\sum_{H} \theta_H g_H(\boldsymbol{y}) - \ln \kappa(\boldsymbol{\theta})\right\} \tag{2}$$

where

i each H is a configuration, which is defined to be a set of possible edges among a subset of the vertices belonging to V.

One popular class of network models is the so-called Exponential Random Graph Model (ERGM). See, e.g., Frank and Strauss (1986), Wasserman and Pattison (1996), Snijders (2002), Robins et al. (2007) and Lusher et al. (2013): available for either undirected or directed graphs and some network statistics (like e.g. the number of edges, triangles, stars) in combination with some observable variables explain the dependence among network data.

More precisely, for the **Bernoulli rvs** in the random adjacency matrix, an ERGM takes the form

$$P_{\boldsymbol{\theta}}(\boldsymbol{Y} = \boldsymbol{y}) = \exp\left\{\sum_{H} \theta_H g_H(\boldsymbol{y}) - \ln \kappa(\boldsymbol{\theta})\right\} \tag{2}$$

where

- i each H is a configuration, which is defined to be a set of possible edges among a subset of the vertices belonging to V.

Related work

ii $g_H(\boldsymbol{y}) = \prod_{y_{ij} \in H} y_{ij}$ and is there for either 1 if the configuration H occurs in $\boldsymbol{y}$, or 0 otherwise

iii a non-zero value for θ_H means that the Y_{ij} are dependent for all pairs of vertices $\{v_i, v_j\}$ in H, conditional upon the rest of the graph

iii $\kappa(\boldsymbol{\theta})$ is a normalization constant $\kappa(\boldsymbol{\theta}) = \sum_{\boldsymbol{y}} \exp\{\sum_H \theta_H g_H(\boldsymbol{y})\}$.

Limitations:

- often the likelihood function is numerically intractable for most distributions belonging to the exponential family
- when the likelihood function is numerically tractable, it entails heavy computational cost

$\Rightarrow$ ERGMs often become impractical and they are not suitable for the FAO data analysis (latent factors treatment is usually problematic).

Related work

ii $g_H(\boldsymbol{y}) = \prod_{y_{ij} \in H} y_{ij}$ and is there for either 1 if the configuration H occurs in $\boldsymbol{y}$, or 0 otherwise

iii a non-zero value for θ_H means that the Y_{ij} are dependent for all pairs of vertices $\{v_i, v_j\}$ in H, conditional upon the rest of the graph

iii $\kappa(\boldsymbol{\theta})$ is a normalization constant $\kappa(\boldsymbol{\theta}) = \sum_{\boldsymbol{y}} \exp\{\sum_H \theta_H g_H(\boldsymbol{y})\}$.

Limitations:

- often the likelihood function is numerically intractable for most distributions belonging to the exponential family
- when the likelihood function is numerically tractable, it entails heavy computational cost

$\Rightarrow$ ERGMs often become impractical and they are not suitable for the FAO data analysis (latent factors treatment is usually problematic).

Related work

ii $g_H(\boldsymbol{y}) = \prod_{y_{ij} \in H} y_{ij}$ and is there for either 1 if the configuration H occurs in $\boldsymbol{y}$, or 0 otherwise

iii a non-zero value for θ_H means that the Y_{ij} are dependent for all pairs of vertices $\{v_i, v_j\}$ in H, conditional upon the rest of the graph

iii $\kappa(\boldsymbol{\theta})$ is a normalization constant $\kappa(\boldsymbol{\theta}) = \sum_{\boldsymbol{y}} \exp\{\sum_H \theta_H g_H(\boldsymbol{y})\}$.

Limitations:

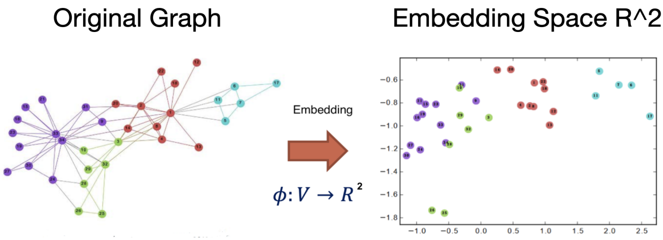
- often the likelihood function is numerically intractable for most distributions belonging to the exponential family
- when the likelihood function is numerically tractable, it entails heavy computational cost

$\Rightarrow$ ERGMs often become impractical and they are not suitable for the FAO data analysis (latent factors treatment is usually problematic).

Related work

Working on the limitations of ERGMs, Hoff et al. (2002) propose the Latent Position Model (LPM): a framework which relies on a latent space representation, namely an embedding in an Euclidean space, whereby the nodes of the network are characterized by their individual latent coordinates which in turn determine their connectivity patterns.

Essentially, one represents each individual node in an Euclidean space via a function $\phi : V \to \mathbb{R}^q$, for some $q \in \mathbb{N}$



$\Rightarrow$ nodes which are close in the original graph should be close also in $\mathbb{R}^2$

Related work

The presence or absence of an edge is **independent** of all other edges, **given** the unobserved positions/factors in a latent space:

$$P(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z},\mathbf{X}=\mathbf{x},\boldsymbol{\theta})=\prod_{i\neq j}P(Y_{ij}=y_{ij}|\mathbf{z}_i,\mathbf{z}_j,\mathbf{x}_{ij},\boldsymbol{\theta})$$

where the entries $\mathbf{x}_{ij}$ are observable characteristics, while $\boldsymbol{\theta}$ and $\mathbf{Z}=\{\mathbf{z}_i\}_{i\in V}$, with $\mathbf{z}_i\in\mathbb{R}^q$, are parameters and latent factors to be estimated.

$\Rightarrow$ Inference is typically conducted using MCMC methods; see Hoff et al. (2002); Handcock et al., 2007; Krivitsky et al. (2009), and Rastelli et al. (2016)—the last two papers contain connections to the literature on random effects.

Related work

The presence or absence of an edge is **independent** of all other edges, **given** the unobserved positions/factors in a latent space:

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}, \mathbf{X} = \mathbf{x}, \boldsymbol{\theta}) = \prod_{i \neq j} P(Y_{ij} = y_{ij} | \mathbf{z}_i, \mathbf{z}_j, \mathbf{x}_{ij}, \boldsymbol{\theta})$$

where the entries $\mathbf{x}_{ij}$ are observable characteristics, while $\boldsymbol{\theta}$ and $\mathbf{Z} = \{\mathbf{z}_i\}_{i \in V}$, with $\mathbf{z}_i \in \mathbb{R}^q$, are parameters and latent factors to be estimated.

$\Rightarrow$ Inference is typically conducted using MCMC methods; see Hoff et al. (2002); Handcock et al., 2007; Krivitsky et al. (2009), and Rastelli et al. (2016)—the last two papers contain connections to the literature on random effects.

Related work

Example (Latent Position Model (Distance))

The statistician associates a latent position in Euclidean space with each node, then postulates that nodes that are closer are more likely to be linked, with the probability of connection depending on the distance:

- $\mathbf{Z}_i, \mathbf{Z}_j \in \mathbb{R}^q$, denotes the latent factors of node i and j, respectively
- **Distance model**: edge probability (mass/density) satisfies

$$\ln\left(\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}\right) = \beta_1 + X_{ij}\beta_2 - D_{ij}$$

  with $\boldsymbol{\beta} := (\beta_1, \beta_2) \in \mathbb{R}^2$, X_{ij} is a univariate observable covariate, and $D_{ij} = |\mathbf{Z}_i - \mathbf{Z}_j|$.

Related work

Example (Latent Position Model (Distance))

The statistician associates a latent position in Euclidean space with each node, then postulates that nodes that are closer are more likely to be linked, with the probability of connection depending on the distance:

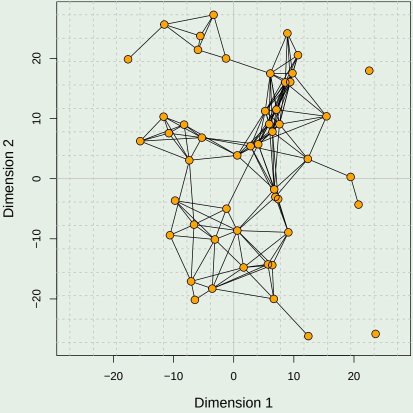
- $\mathbf{Z}_i, \mathbf{Z}_j \in \mathbb{R}^q$, denotes the latent factors of node i and j, respectively
- **Distance model**: edge probability (mass/density) satisfies

$$\ln\left(\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}\right) = \beta_1 + X_{ij}\beta_2 - D_{ij}$$

  with $\boldsymbol{\beta} := (\beta_1, \beta_2) \in \mathbb{R}^2$, X_{ij} is a univariate observable covariate, and $D_{ij} = |\mathbf{Z}_i - \mathbf{Z}_j|$.

Related work

Example (cont'd)



A randomly generated LPM: $\mathbb{R}^2$-embedded version of the graph.

Related work

Example (Projection mode)

Edge probabilities satisfy

$$\ln\left(\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}\right) = \beta + \langle \mathbf{Z}_i, \mathbf{Z}_j \rangle$$

with $\beta \in \mathbb{R}$, and $\langle \cdot, \cdot \rangle$ indicates the dot product.

Since

$$\langle \mathbf{z}_i, \mathbf{z}_j \rangle = ||\mathbf{z}_i|| ||\mathbf{z}_j|| \cos\theta$$

we need to look at the **angle** θ that the two nodes form in the centre of the space. Also, the distance from the centre $|| \cdot ||$ plays a role (sociality).

Let's illustrate the role of the angle θ.

Related work

Example (Projection mode)

Edge probabilities satisfy

$$\ln\left(\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}\right) = \beta + \langle \mathbf{Z}_i, \mathbf{Z}_j \rangle$$

with $\beta \in \mathbb{R}$, and $\langle \cdot, \cdot \rangle$ indicates the dot product.

Since

$$\langle \mathbf{z}_i, \mathbf{z}_j \rangle = \|\mathbf{z}_i\| \|\mathbf{z}_j\| \cos\theta$$

we need to look at the **angle** θ that the two nodes form in the centre of the space. Also, the distance from the centre $\|\cdot\|$ plays a role (sociality).

Let's illustrate the role of the angle θ.

Related work

Example (Projection mode)

Edge probabilities satisfy

$$\ln\left(\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}\right) = \beta + \langle \mathbf{Z}_i, \mathbf{Z}_j \rangle$$

with $\beta \in \mathbb{R}$, and $\langle \cdot, \cdot \rangle$ indicates the dot product.

Since

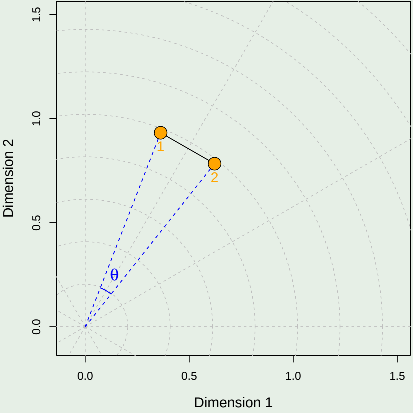
$$\langle \mathbf{z}_i, \mathbf{z}_j \rangle = \|\mathbf{z}_i\| \|\mathbf{z}_j\| \cos\theta$$

we need to look at the **angle** θ that the two nodes form in the centre of the space. Also, the distance from the centre $\| \cdot \|$ plays a role (sociality).

Let's illustrate the role of the angle θ.

Related work

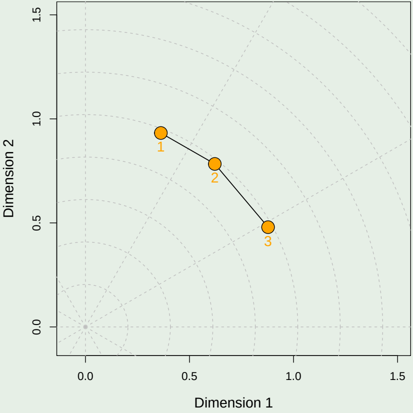
Example (cont'd)



Small angle $\Rightarrow$ connection.

Related work

Example (cont'd)



No connection between 1 and 3 due to large angle.

GLLVM

In a similar spirit, we propose to build on Generalised Linear Latent Variable Models (GLLVM), which combine GLMs and factor models.

GLLVMs extend GLMs by defining the **canonical parameter** as a *linear combination* of **latent variables**; see Skrondal and Rabe-Hesketh, 2004; Bartholomew et al., 2011

- GLLVM have been widely-applied in psychology, ecology and social-sciences.
- The additional latent variable term adds more *flexibility*.
- Inference typically conducted via **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)

Remark (Goal)

To apply them to our FAO network data, we need to adapt the GLLVM to the graph setting: this yields the Graph Generalised Linear Latent Variable Model (GGLLVM)

GLLVM

In a similar spirit, we propose to build on Generalised Linear Latent Variable Models (GLLVM), which combine GLMs and factor models.

GLLVMs extend GLMs by defining the **canonical parameter** as a *linear combination* of **latent variables**; see Skrondal and Rabe-Hesketh, 2004; Bartholomew et al., 2011

- GLLVM have been widely-applied in psychology, ecology and social-sciences.
- The additional latent variable term adds more *flexibility*.
- Inference typically conducted via **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)

Remark (Goal)

To apply them to our FAO network data, we need to adapt the GLLVM to the graph setting: this yields the Graph Generalised Linear Latent Variable Model (GGLLVM)

GLLVM

In a similar spirit, we propose to build on Generalised Linear Latent Variable Models (GLLVM), which combine GLMs and factor models.

GLLVMs extend GLMs by defining the **canonical parameter** as a *linear combination* of **latent variables**; see Skrondal and Rabe-Hesketh, 2004; Bartholomew et al., 2011

- GLLVM have been widely-applied in psychology, ecology and social-sciences.
- The additional latent variable term adds more *flexibility*.
- Inference typically conducted via **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)

Remark (Goal)

To apply them to our FAO network data, we need to adapt the GLLVM to the graph setting: this yields the Graph Generalised Linear Latent Variable Model (GGLLVM)

GLLVM

In a similar spirit, we propose to build on Generalised Linear Latent Variable Models (GLLVM), which combine GLMs and factor models.

GLLVMs extend GLMs by defining the **canonical parameter** as a *linear combination* of **latent variables**; see Skrondal and Rabe-Hesketh, 2004; Bartholomew et al., 2011

- GLLVM have been widely-applied in psychology, ecology and social-sciences.
- The additional latent variable term adds more *flexibility*.
- Inference typically conducted via **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)

Remark (Goal)

To apply them to our FAO network data, we need to adapt the GLLVM to the graph setting: this yields the Graph Generalised Linear Latent Variable Model (GGLLVM)

GLLVM

In a similar spirit, we propose to build on Generalised Linear Latent Variable Models (GLLVM), which combine GLMs and factor models.

GLLVMs extend GLMs by defining the **canonical parameter** as a *linear combination* of **latent variables**; see Skrondal and Rabe-Hesketh, 2004; Bartholomew et al., 2011

- GLLVM have been widely-applied in psychology, ecology and social-sciences.
- The additional latent variable term adds more *flexibility*.
- Inference typically conducted via **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)

Remark (Goal)

To apply them to our FAO network data, we need to adapt the GLLVM to the graph setting: this yields the Graph Generalised Linear Latent Variable Model (GGLLVM)

GLLVM

In a similar spirit, we propose to build on Generalised Linear Latent Variable Models (GLLVM), which combine GLMs and factor models.

GLLVMs extend GLMs by defining the **canonical parameter** as a *linear combination* of **latent variables**; see Skrondal and Rabe-Hesketh, 2004; Bartholomew et al., 2011

- GLLVM have been widely-applied in psychology, ecology and social-sciences.
- The additional latent variable term adds more *flexibility*.
- Inference typically conducted via **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)

Remark (Goal)

To apply them to our FAO network data, we need to adapt the GLLVM to the graph setting: this yields the Graph Generalised Linear Latent Variable Model (GGLLVM)

GGLLVM

The generic ingredients of our GGLLVM are:

Data is $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, where $\mathbf{Y}^{(k)}$ represents the k-th layer

The **conditional density function of** $Y_{ij}^{(k)}$, for each k, belongs to the exponential family

$$\begin{aligned} P(y_{ij}|\boldsymbol{\alpha}_{ij}, \mathbf{z}) &= g_{ij}(y_{ij}|\mathbf{z}) \\ &= \exp\left\{ [y_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}} - b_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}}]/\varphi_{ij} + c_{ij}(y_{ij}, \varphi_{ij}) \right\}. \end{aligned}$$

Latent variables: $\{\boldsymbol{\alpha}'_{ij}\}$ (factor loadings) and the (factor) $\mathbf{Z} = (1, Z_1, \cdots, Z_q)' = (1, \mathbf{Z}'_{(2)})' \in \mathbb{R}^{q+1}$, for $\mathbf{Z}'_{(2)} \in \mathbb{R}^q$, with $q \ll n_V$ and $q \in \mathbb{N}$. From now on, I assume that

$$\mathbf{Z}'_{(2)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q).$$

GGLLVM

The generic ingredients of our GGLLVM are:

Data is $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, where $\mathbf{Y}^{(k)}$ represents the k-th layer

The **conditional density function of** $Y_{ij}^{(k)}$, for each k, belongs to the exponential family

$$\begin{aligned} P(y_{ij}|\boldsymbol{\alpha}_{ij}, \mathbf{z}) &= g_{ij}(y_{ij}|\mathbf{z}) \\ &= \exp\left\{ [y_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}} - b_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}}]/\varphi_{ij} + c_{ij}(y_{ij}, \varphi_{ij}) \right\}. \end{aligned}$$

Latent variables: $\{\boldsymbol{\alpha}'_{ij}\}$ (factor loadings) and the (factor) $\mathbf{Z} = (1, Z_1, \cdots, Z_q)' = (1, \mathbf{Z}'_{(2)})' \in \mathbb{R}^{q+1}$, for $\mathbf{Z}'_{(2)} \in \mathbb{R}^q$, with $q \ll n_V$ and $q \in \mathbb{N}$. From now on, I assume that

$$\mathbf{Z}'_{(2)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q).$$

GGLLVM

The generic ingredients of our GGLLVM are:

Data is $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, where $\mathbf{Y}^{(k)}$ represents the k-th layer

The **conditional density function of** $Y_{ij}^{(k)}$, for each k, belongs to the exponential family

$$\begin{aligned} P(y_{ij}|\boldsymbol{\alpha}_{ij}, \mathbf{z}) &= g_{ij}(y_{ij}|\mathbf{z}) \\ &= \exp\left\{ [y_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}} - b_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}}]/\varphi_{ij} + c_{ij}(y_{ij}, \varphi_{ij}) \right\}. \end{aligned}$$

Latent variables: $\{\boldsymbol{\alpha}'_{ij}\}$ (factor loadings) and the (factor) $\mathbf{Z} = (1, Z_1, \cdots, Z_q)' = (1, \mathbf{Z}'_{(2)})' \in \mathbb{R}^{q+1}$, for $\mathbf{Z}'_{(2)} \in \mathbb{R}^q$, with $q \ll n_V$ and $q \in \mathbb{N}$. From now on, I assume that

$$\mathbf{Z}'_{(2)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q).$$

The generic ingredients of our GGLLVM are:

Data is $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, where $\mathbf{Y}^{(k)}$ represents the k-th layer

The **conditional density function of** $Y_{ij}^{(k)}$, for each k, belongs to the exponential family

$$\begin{aligned} P(y_{ij}|\boldsymbol{\alpha}_{ij}, \mathbf{z}) &= g_{ij}(y_{ij}|\mathbf{z}) \\ &= \exp\left\{ [y_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}} - b_{ij} \underbrace{(\boldsymbol{\alpha}'_{ij}\mathbf{z})}_{\text{can. par.}}]/\varphi_{ij} + c_{ij}(y_{ij}, \varphi_{ij}) \right\}. \end{aligned}$$

Latent variables: $\{\boldsymbol{\alpha}'_{ij}\}$ (factor loadings) and the (factor) $\mathbf{Z} = (1, Z_1, \cdots, Z_q)' = (1, \mathbf{Z}'_{(2)})' \in \mathbb{R}^{q+1}$, for $\mathbf{Z}'_{(2)} \in \mathbb{R}^q$, with $q \ll n_V$ and $q \in \mathbb{N}$. From now on, I assume that

$$\mathbf{Z}'_{(2)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q).$$

Example (Binary data: Adjacency matrix)

In each layer, let $Y_{ij}|\mathbf{Z}$ be independent Bernoulli r.v.s with mean π_{ij}, so $\boldsymbol{\pi} = \{\pi_{ij}\}$.

- using the canonical link function we have

$$P(Y_{ij} = 1|\mathbf{Z} = \mathbf{z}, \boldsymbol{\alpha}_{ij}) = \pi_{ij} = \frac{\exp(\boldsymbol{\alpha}'_{ij}\mathbf{z})}{1 + \exp(\boldsymbol{\alpha}'_{ij}\mathbf{z})}, \tag{3}$$

  namely,

$$\ln(\pi_{ij}/(1 - \pi_{ij})) = \boldsymbol{\alpha}'_{ij}\mathbf{z}.$$

- $P(y_{ij}|\mathbf{z}, \boldsymbol{\alpha}_{ij})$ becomes

$$\begin{aligned} g_{ij}(y_{ij}|\mathbf{z}) &= \exp\{y_{ij}\ln(\pi_{ij}/(1 - \pi_{ij})) + \ln(1 - \pi_{ij})\} \\ &= \exp\left\{y_{ij}(\boldsymbol{\alpha}'_{ij}\mathbf{z}) - \ln\left[1 + \exp\left(\boldsymbol{\alpha}'_{ij}\mathbf{z}\right)\right]\right\}. \end{aligned}$$

Inference (GLAMLE)

- The complete data likelihood for the k-th layer $\boldsymbol{Y}^{(k)} = \boldsymbol{y}^{(k)}$ is

$$\prod_{i \neq j} g_{ij}(y_{ij}^{(k)} | \mathbf{z}) h(\mathbf{z}_{(2)}),$$

where $h(\mathbf{z}_{(2)})$ is the density function of latent variables (pdf of a standard q-dim normal)

- Integrating out the latent variables, we get the marginal density function

$$f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \int \left\{ \prod_{i \neq j} g_{ij}(y_{ij}^{(k)} | \mathbf{z}) \right\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)},$$

where $\boldsymbol{\alpha}$ is the $m \times (q+1)$ matrix containing all the $\boldsymbol{\alpha}_{ij}$, **for m representing the number of dyads** and $\boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{\alpha})$

- Given K random samples $\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \cdots, \boldsymbol{Y}^{(K)}$, the **exact** log-likelihood is

$$S(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \;\Rightarrow\; \nabla_{\boldsymbol{\alpha}} S(\boldsymbol{\alpha}) = \mathbf{0}.$$

Inference (GLAMLE)

- The complete data likelihood for the k-th layer $\boldsymbol{Y}^{(k)} = \boldsymbol{y}^{(k)}$ is

$$\prod_{i \neq j} g_{ij}(y_{ij}^{(k)} | \mathbf{z}) h(\mathbf{z}_{(2)}),$$

  where $h(\mathbf{z}_{(2)})$ is the density function of latent variables (pdf of a standard q-dim normal)

- Integrating out the latent variables, we get the marginal density function

$$f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \int \left\{ \prod_{i \neq j} g_{ij}(y_{ij}^{(k)} | \mathbf{z}) \right\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)},$$

  where $\boldsymbol{\alpha}$ is the $m \times (q+1)$ matrix containing all the $\boldsymbol{\alpha}_{ij}$, **for** m **representing the number of dyads** and $\boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{\alpha})$

- Given K random samples $\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \cdots, \boldsymbol{Y}^{(K)}$, the **exact** log-likelihood is

$$S(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \;\Rightarrow\; \nabla_{\boldsymbol{\alpha}} S(\boldsymbol{\alpha}) = \mathbf{0}.$$

Inference (GLAMLE)

- The complete data likelihood for the k-th layer $\boldsymbol{Y}^{(k)} = \boldsymbol{y}^{(k)}$ is

$$\prod_{i \neq j} g_{ij}(y_{ij}^{(k)} | \mathbf{z}) h(\mathbf{z}_{(2)}),$$

  where $h(\mathbf{z}_{(2)})$ is the density function of latent variables (pdf of a standard q-dim normal)

- Integrating out the latent variables, we get the marginal density function

$$f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \int \left\{ \prod_{i \neq j} g_{ij}(y_{ij}^{(k)} | \mathbf{z}) \right\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)},$$

  where $\boldsymbol{\alpha}$ is the $m \times (q+1)$ matrix containing all the $\boldsymbol{\alpha}_{ij}$, **for m representing the number of dyads** and $\boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{\alpha})$

- Given K random samples $\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \cdots, \boldsymbol{Y}^{(K)}$, the **exact** log-likelihood is

$$S(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \;\Rightarrow\; \nabla_{\boldsymbol{\alpha}} S(\boldsymbol{\alpha}) = \mathbf{0}.$$

Inference (GLAMLE)

The treatment of $S(\boldsymbol{\alpha})$ is numerically complex. Thus, some approximations need to be considered (e.g. MCMC and VA). Differently from other common approaches, we use a **Laplace approximation**:

- latent variables $\mathbf{z}_{(2)}$ have standard normal distributions and that they are independent: this allows to rewrite

$$f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \int \exp\left\{mQ\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right)\right\} d\mathbf{z}_{(2)}, \tag{4}$$

where m is the number of dyads and

$$mQ\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right) = \left[\sum_{i \neq j}^{n_V} \left\{y_{ij}^{(k)} \boldsymbol{\alpha}'_{ij}\mathbf{z} - \ln\left(1 + \exp\left(\boldsymbol{\alpha}'_{ij}\mathbf{z}\right)\right)\right\} - \frac{\mathbf{z}'_{(2)}\mathbf{z}_{(2)}}{2} - \frac{q}{2}\ln(2\pi)\right],$$

where $\sum_{i \neq j}^{n_V}$ is a double sum having different expressions for undirected and directed relations.

Inference (GLAMLE)

The treatment of $S(\boldsymbol{\alpha})$ is numerically complex. Thus, some approximations need to be considered (e.g. MCMC and VA). Differently from other common approaches, we use a **Laplace approximation**:

- latent variables $\mathbf{z}_{(2)}$ have standard normal distributions and that they are independent: this allows to rewrite

$$f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \int \exp\left\{mQ\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right)\right\} d\mathbf{z}_{(2)}, \tag{4}$$

where m is the number of dyads and

$$mQ\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right) = \left[\sum_{i\neq j}^{n_V} \left\{y_{ij}^{(k)} \boldsymbol{\alpha}'_{ij}\mathbf{z} - \ln\left(1 + \exp\left(\boldsymbol{\alpha}'_{ij}\mathbf{z}\right)\right)\right\} - \frac{\mathbf{z}'_{(2)}\mathbf{z}_{(2)}}{2} - \frac{q}{2}\ln(2\pi)\right],$$

where $\sum_{i\neq j}^{n_V}$ is a double sum having different expressions for undirected and directed relations.

Inference (GLAMLE)

- We derive the Laplace-approximated marginal density function

$$\tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \left(\frac{2\pi}{m}\right)^{q/2} \det\left\{-U\left(\hat{\mathbf{z}}^{(k)}\right)\right\}^{-1/2} \exp\left\{mQ\left(\boldsymbol{\alpha}, \hat{\mathbf{z}}^{(k)}, \boldsymbol{y}^{(k)}\right)\right\}, \tag{5}$$

where

$$U\left(\boldsymbol{\alpha}, \hat{\mathbf{z}}^{(k)}\right) = \left.\frac{\partial^2 Q\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right)}{\partial \mathbf{z}' \partial \mathbf{z}}\right|_{\mathbf{z}=\hat{\mathbf{z}}^{(k)}}$$

with $\boldsymbol{\alpha}_{ij} = (\alpha_{ij,0}, \boldsymbol{\alpha}'_{ij(2)})'$ and $\hat{\mathbf{z}}^{(k)} = (1, (\hat{\mathbf{z}}^{(k)}_{(2)})')'$ maximizing $Q\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right)$, therefore being the solution to

$$\partial Q\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right) / \partial \mathbf{z} = \mathbf{0}.$$

This defines a **concentrated** approximated likelihood, the exact likelihood being intractable, which yields the estimated factors as functions of $\boldsymbol{\alpha}$.

Inference (GLAMLE)

- We derive the Laplace-approximated marginal density function

$$\tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \left(\frac{2\pi}{m}\right)^{q/2} \det\left\{-U\left(\hat{\mathbf{z}}^{(k)}\right)\right\}^{-1/2} \exp\left\{mQ\left(\boldsymbol{\alpha}, \hat{\mathbf{z}}^{(k)}, \boldsymbol{y}^{(k)}\right)\right\}, \tag{5}$$

where

$$U\left(\boldsymbol{\alpha}, \hat{\mathbf{z}}^{(k)}\right) = \left.\frac{\partial^2 Q\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right)}{\partial \mathbf{z}' \partial \mathbf{z}}\right|_{\mathbf{z}=\hat{\mathbf{z}}^{(k)}}$$

with $\boldsymbol{\alpha}_{ij} = (\alpha_{ij,0}, \boldsymbol{\alpha}'_{ij(2)})'$ and $\hat{\mathbf{z}}^{(k)} = (1, (\hat{\mathbf{z}}^{(k)}_{(2)})')'$ maximizing $Q\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right)$, therefore being the solution to

$$\partial Q\left(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{y}^{(k)}\right) / \partial \mathbf{z} = \mathbf{0}.$$

This defines a **concentrated** approximated likelihood, the exact likelihood being intractable, which yields the estimated factors as functions of $\boldsymbol{\alpha}$.

Inference (GLAMLE)

Then, optimizing the concentrated likelihood w.r.t. $\boldsymbol{\alpha}$, we obtain the estimates of the factor loadings. We call the resulting M-estimator GLAMLE $\hat{\theta}$: it is implied by setting to zero the equations derived from the first order conditions of the Laplace-approximated likelihood.

In formulae: for the k-th network view we have

$$\tilde{\boldsymbol{\mathcal{S}}}^{(k)}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} \ln \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})$$

so, by definition of M-estimator, the GLAMLE $\hat{\boldsymbol{\theta}}$ solves

$$\boxed{\tilde{\boldsymbol{\mathcal{S}}}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \tilde{\boldsymbol{\mathcal{S}}}^{(k)}(\boldsymbol{\alpha}) = \mathbf{0}.} \tag{6}$$

Inference (GLAMLE)

Then, optimizing the concentrated likelihood w.r.t. $\boldsymbol{\alpha}$, we obtain the estimates of the factor loadings. We call the resulting M-estimator GLAMLE $\hat{\theta}$: it is implied by setting to zero the equations derived from the first order conditions of the Laplace-approximated likelihood.

In formulae: for the k-th network view we have

$$\tilde{\boldsymbol{\mathcal{S}}}^{(k)}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} \ln \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})$$

so, by definition of M-estimator, the GLAMLE $\hat{\boldsymbol{\theta}}$ solves

$$\boxed{\tilde{\boldsymbol{\mathcal{S}}}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \tilde{\boldsymbol{\mathcal{S}}}^{(k)}(\boldsymbol{\alpha}) = \mathbf{0}.} \tag{6}$$

Inference (GLAMLE)

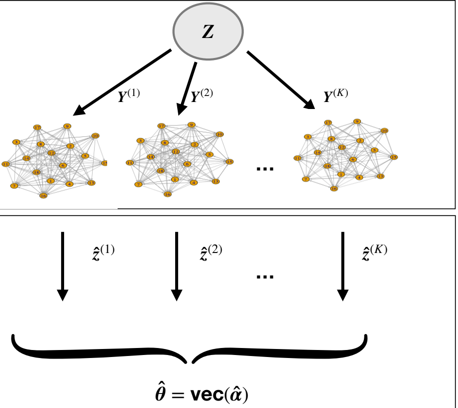
To summarize via a graphical illustration:

GGLLVM

- **The multivariate latent** $Z \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q)$ **determines the behaviour of the edges in the network views** $\{\boldsymbol{Y}^{(k)}\}, k = 1,2,...,K$
- **The edges of each** k**-th network view** $\boldsymbol{Y}^{(k)}$ **have a distribution belonging to the exponential family, whose canonical parameter depends linearly on** Z **as in (4.1)**

GLAMLE

- **Laplace approximation yields estimated latent variables** $\hat{z}^{(k)}$**, obtained using (4.9), for** $k = 1,2,..,K$
- **The estimation of vec(**$\boldsymbol{\alpha}$**) is obtained via maximization of the Laplace approximated likelihood in (4.11), computed using** $\hat{z}^{(k)}$**, for** $k = 1,2,..,K$



Inference (GLAMLE)

Remark (Statistical properties/guarantees)

- *The* $\hat{\mathbf{z}}^{(k)}_{(2)}$ *can be formally interpreted as the maximum likelihood estimates of the latent factors in the k-th network view*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})\left\{1 + O\left(m^{-1}\right)\right\}$
- *The contribution of the k-th view to the exact likelihood score is* $\boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})$, *so* $\boldsymbol{S}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha})$. *Under* $\boldsymbol{\alpha}_0$, *the Fisher consistency of the MLE holds for any* k, *thus* $E_0[\boldsymbol{\mathcal{S}}(\boldsymbol{\alpha}_0)] = \mathbf{0}$. *Due to its approximate nature, the function* $\tilde{S}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$, *in the sense of White (1982):*

$$K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}})), \quad K \to \infty,$$

  and, of course, identifiability restrictions are needed
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities (they depend on the observed graph topology)...* Jump to Bayes

Inference (GLAMLE)

Remark (Statistical properties/guarantees)

- *The* $\hat{\mathbf{z}}_{(2)}^{(k)}$ *can be formally interpreted as the maximum likelihood estimates of the latent factors in the* k*-th network view*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$
- *The contribution of the* k*-th view to the exact likelihood score is* $\boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})$*, so* $\boldsymbol{S}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha})$*. Under* $\boldsymbol{\alpha}_0$*, the Fisher consistency of the MLE holds for any* k*, thus* $E_0[\boldsymbol{\mathcal{S}}(\boldsymbol{\alpha}_0)] = \mathbf{0}$*. Due to its approximate nature, the function* $\tilde{S}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$*, in the sense of White (1982):*

$$K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}})), \quad K \to \infty,$$

  and, of course, identifiability restrictions are needed
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities (they depend on the observed graph topology)...* Jump to Bayes

Inference (GLAMLE)

Remark (Statistical properties/guarantees)

- *The* $\hat{\mathbf{z}}^{(k)}_{(2)}$ *can be formally interpreted as the maximum likelihood estimates of the latent factors in the* k*-th network view*

- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$

- *The contribution of the* k*-th view to the exact likelihood score is* $\boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})$*, so* $\boldsymbol{S}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha})$*. Under* $\boldsymbol{\alpha}_0$*, the Fisher consistency of the MLE holds for any* k*, thus* $E_0[\boldsymbol{\mathcal{S}}(\boldsymbol{\alpha}_0)] = \mathbf{0}$*. Due to its approximate nature, the function* $\tilde{S}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$*, in the sense of White (1982):*

$$K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}})), \quad K \to \infty,$$

  and, of course, identifiability restrictions are needed

- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities (they depend on the observed graph topology)...*

Jump to Bayes

Inference (GLAMLE)

Remark (Statistical properties/guarantees)

- *The* $\hat{\mathbf{z}}_{(2)}^{(k)}$ *can be formally interpreted as the maximum likelihood estimates of the latent factors in the k-th network view*

- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$

- *The contribution of the k-th view to the exact likelihood score is* $\boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})$, *so* $\boldsymbol{S}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha})$. *Under* $\boldsymbol{\alpha}_0$, *the Fisher consistency of the MLE holds for any k, thus* $E_0[\boldsymbol{\mathcal{S}}(\boldsymbol{\alpha}_0)] = \mathbf{0}$. *Due to its approximate nature, the function* $\tilde{S}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$, *in the sense of White (1982):*

$$K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}})), \quad K \to \infty,$$

  and, of course, identifiability restrictions are needed

- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities (they depend on the observed graph topology)...* Jump to Bayes

Inference (GLAMLE)

Remark (Statistical properties/guarantees)

- *The* $\hat{\mathbf{z}}^{(k)}_{(2)}$ *can be formally interpreted as the maximum likelihood estimates of the latent factors in the k-th network view*

- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})\left\{1 + O\left(m^{-1}\right)\right\}$

- *The contribution of the k-th view to the exact likelihood score is* $\boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} \ln f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)})$*, so* $\boldsymbol{S}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \boldsymbol{\mathcal{S}}^{(k)}(\boldsymbol{\alpha})$*. Under* $\boldsymbol{\alpha}_0$*, the Fisher consistency of the MLE holds for any* k*, thus* $E_0[\boldsymbol{\mathcal{S}}(\boldsymbol{\alpha}_0)] = \mathbf{0}$*. Due to its approximate nature, the function* $\tilde{S}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$*, in the sense of White (1982):*

$$K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}})), \quad K \to \infty,$$

  and, of course, identifiability restrictions are needed

- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities (they depend on the observed graph topology)...* Jump to Bayes

Synthetic data

To illustrate numerically the performance of the GLAMLE, we consider several MC experiments.

- We focus on Bernoulli random variables for a simulated directed network with $n_V = 18$ and we study the estimation problem of an adjacency matrix
- We consider 1000 MC runs, where in each run, we set $K = 100$, namely, we have 100 network layers
- To investigate on the role of the number of latent variables, we consider the cases of one ($q = 1$) and two ($q = 2$) latent variable/s $\mathbf{Z}$.

The goal of the numerical exercises is to study the bias due to the likelihood approximation and the variability of the estimates. Moreover, we compare our method to the state-of-the art methods (MCMC and VA), **suitably adapted to our setting**.

Synthetic data

To illustrate numerically the performance of the GLAMLE, we consider several MC experiments.

- We focus on Bernoulli random variables for a simulated directed network with $n_V = 18$ and we study the estimation problem of an adjacency matrix
- We consider 1000 MC runs, where in each run, we set $K = 100$, namely, we have 100 network layers
- To investigate on the role of the number of latent variables, we consider the cases of one ($q = 1$) and two ($q = 2$) latent variable/s $\mathbf{Z}$.

The goal of the numerical exercises is to study the bias due to the likelihood approximation and the variability of the estimates. Moreover, we compare our method to the state-of-the art methods (MCMC and VA), **suitably adapted to our setting**.

Synthetic data

To illustrate numerically the performance of the GLAMLE, we consider several MC experiments.

- We focus on Bernoulli random variables for a simulated directed network with $n_V = 18$ and we study the estimation problem of an adjacency matrix
- We consider 1000 MC runs, where in each run, we set $K = 100$, namely, we have 100 network layers
- To investigate on the role of the number of latent variables, we consider the cases of one $(q = 1)$ and two $(q = 2)$ latent variable/s $\mathbf{Z}$.

The goal of the numerical exercises is to study the bias due to the likelihood approximation and the variability of the estimates. Moreover, we compare our method to the state-of-the art methods (MCMC and VA), **suitably adapted to our setting**.

Synthetic data

To illustrate numerically the performance of the GLAMLE, we consider several MC experiments.

- We focus on Bernoulli random variables for a simulated directed network with $n_V = 18$ and we study the estimation problem of an adjacency matrix
- We consider 1000 MC runs, where in each run, we set $K = 100$, namely, we have 100 network layers
- To investigate on the role of the number of latent variables, we consider the cases of one $(q = 1)$ and two $(q = 2)$ latent variable/s $\mathbf{Z}$.

The goal of the numerical exercises is to study the bias due to the likelihood approximation and the variability of the estimates. Moreover, we compare our method to the state-of-the art methods (MCMC and VA), **suitably adapted to our setting**.

Synthetic data

To illustrate numerically the performance of the GLAMLE, we consider several MC experiments.

- We focus on Bernoulli random variables for a simulated directed network with $n_V = 18$ and we study the estimation problem of an adjacency matrix
- We consider 1000 MC runs, where in each run, we set $K = 100$, namely, we have 100 network layers
- To investigate on the role of the number of latent variables, we consider the cases of one ($q = 1$) and two ($q = 2$) latent variable/s $\mathbf{Z}$.

The goal of the numerical exercises is to study the bias due to the likelihood approximation and the variability of the estimates. Moreover, we compare our method to the state-of-the art methods (MCMC and VA), **suitably adapted to our setting**.

Synthetic data

To illustrate numerically the performance of the GLAMLE, we consider several MC experiments.

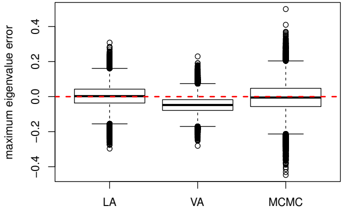
- We focus on Bernoulli random variables for a simulated directed network with $n_V = 18$ and we study the estimation problem of an adjacency matrix
- We consider 1000 MC runs, where in each run, we set $K = 100$, namely, we have 100 network layers
- To investigate on the role of the number of latent variables, we consider the cases of one ($q = 1$) and two ($q = 2$) latent variable/s $\mathbf{Z}$.

The goal of the numerical exercises is to study the bias due to the likelihood approximation and the variability of the estimates. Moreover, we compare our method to the state-of-the art methods (MCMC and VA), **suitably adapted to our setting**.

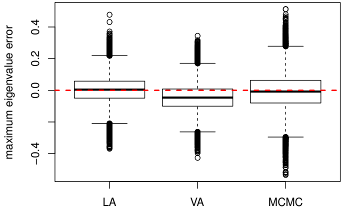
Synthetic data

Maximum eigenvalue difference between connection probability matrices ($\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}_0$, for each layer and then averaged):

$q = 1$



$q = 2$

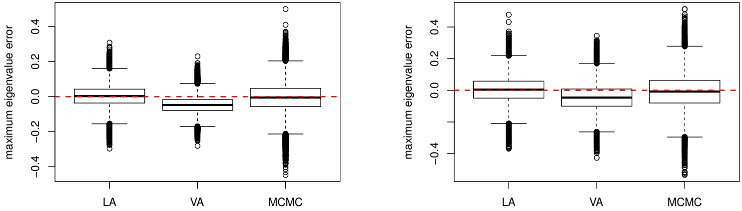


(a) large bias appears with VA (fast); (b) MCMC provides accurate estimates (slow), but it has a slightly larger interquartile range than LA and more outliers (c) LA has small bias (mid)

Synthetic data

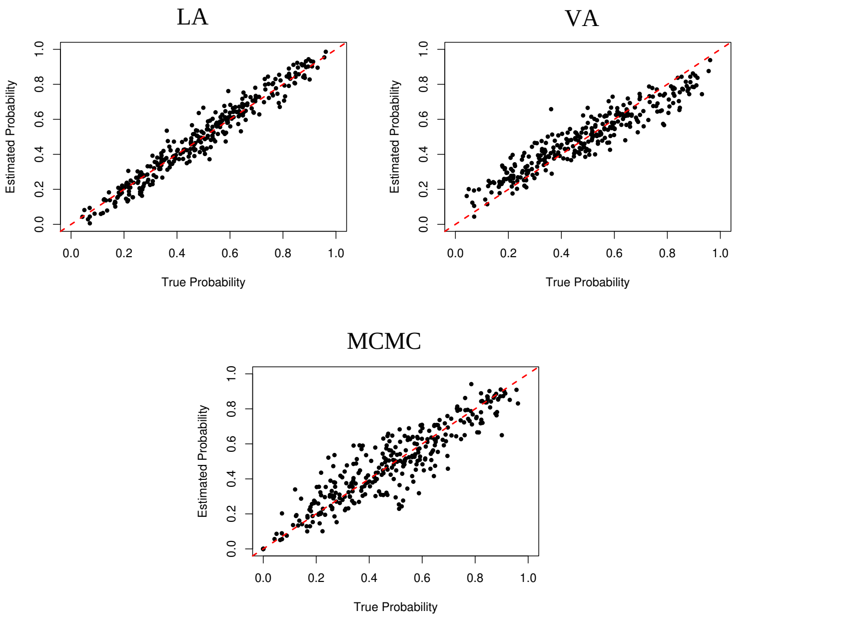
Maximum eigenvalue difference between connection probability matrices ($\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}_0$, for each layer and then averaged):

$q = 1$ $\qquad\qquad$ $q = 2$



(a) large bias appears with VA (fast); (b) MCMC provides accurate estimates (slow), but it has a slightly larger interquartile range than LA and more outliers (c) LA has small bias (mid)

Synthetic data



FAO data (reprise)

- Modeling: GGLLVM with $q = 2$. In this setting, $\mathbf{Z}$ and $\boldsymbol{\alpha}_{ij}$ are both two-dimensional vectors. We choose this particular setup for visualization purposes, and because it returns a good model fit. An arbitrary country i of the network is characterized by:
  (1) a set of bivariate vectors $\boldsymbol{\alpha}_{i1}, \ldots, \boldsymbol{\alpha}_{in_V}$, which determine the tendency of node i to send a connection to any other node in the network, respectively
  (2) a set of bivariate vectors $\boldsymbol{\alpha}_{1i}, \ldots, \boldsymbol{\alpha}_{n_V i}$, indicating the tendency of i to receive a connection from any other node in the network, respectively

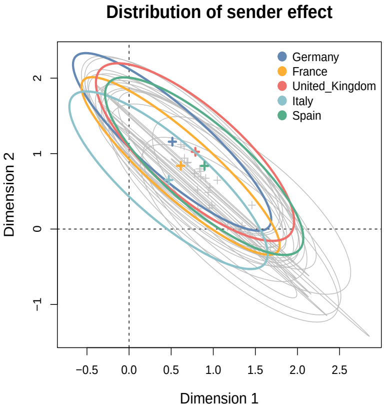
- Inference: Laplace approximation to get the GLAMLE. The output consists of the estimates $\hat{\alpha}$ and $\hat{\mathbf{z}}^{(k)}$, for $k = 1, 2..., 365$.

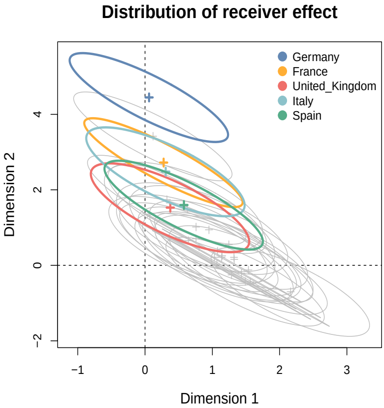
FAO data (reprise)

- Modeling: GGLLVM with $q = 2$. In this setting, $\mathbf{Z}$ and $\boldsymbol{\alpha}_{ij}$ are both two-dimensional vectors. We choose this particular setup for visualization purposes, and because it returns a good model fit. An arbitrary country i of the network is characterized by:
  1. a set of bivariate vectors $\boldsymbol{\alpha}_{i1}, \ldots, \boldsymbol{\alpha}_{in_V}$, which determine the tendency of node i to send a connection to any other node in the network, respectively
  2. a set of bivariate vectors $\boldsymbol{\alpha}_{1i}, \ldots, \boldsymbol{\alpha}_{n_V i}$, indicating the tendency of i to receive a connection from any other node in the network, respectively

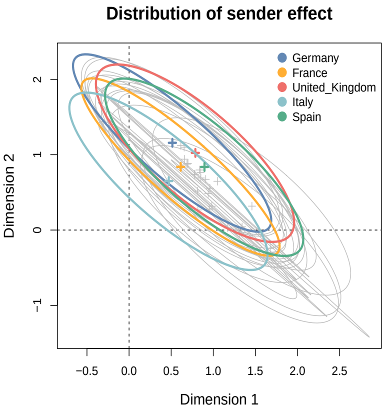
- Inference: Laplace approximation to get the GLAMLE. The output consists of the estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{z}}^{(k)}$, for $k = 1, 2..., 365$.

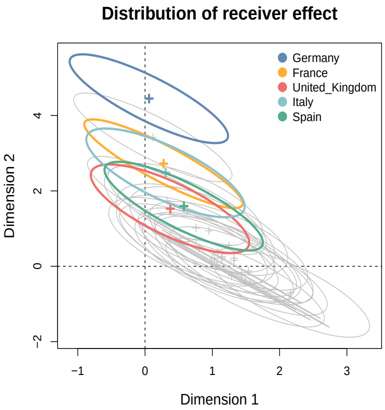
FAO data (reprise)





FAO data (reprise)

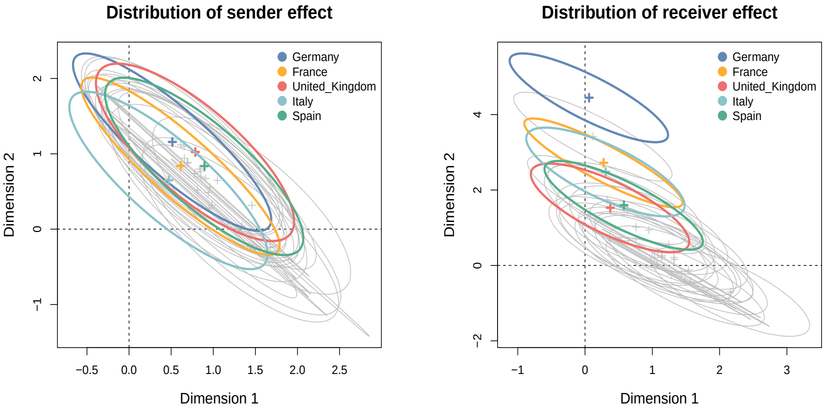




Remark (i)

In the left panel, the ellipses represent the dispersion of the estimated $\boldsymbol{\alpha}_{ij}$ sender values as j varies (resp. receiver values as i varies, on the right panel). The center of each ellipse, represented by a cross, corresponds to the median value.

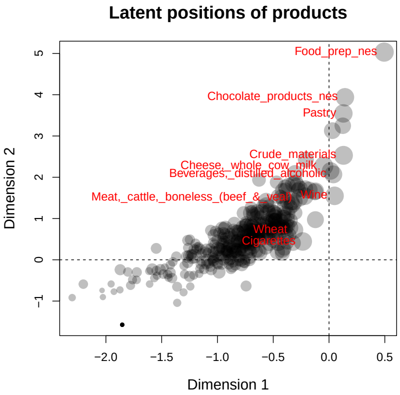
FAO data (reprise)



Remark (ii)

Info from ***concentration and shape of ellipsis*** *(high concentration about (0,0) implies low* π_{ij} *values, hence fewer connections; narrow shape indicates that the country primarily specializes in importing/exporting with some specific countries trading on some specific products).*

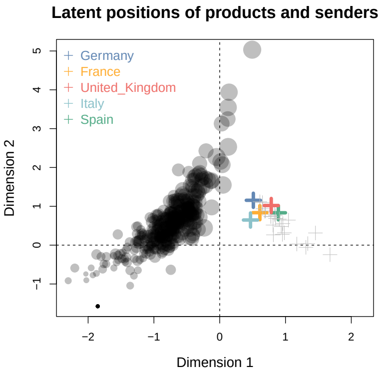
FAO trade results

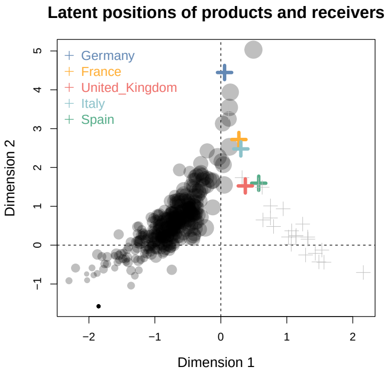


Each dot corresponds to a different product, and the size of each dot indicates its importance (total volume that is traded between all countries). We see that the products are positioned roughly along a curve: where the two ends correspond to the least and most important products. There is an appreciable variability along this line, to indicate that products of same importance can be traded in different ways.

FAO trade results

Predictors are $\boldsymbol{\alpha}_{ij}^{\top}\mathbf{z}^{(k)}$, so the **angle formed with** $(0,0)$ determines how likely an edge is to appear.

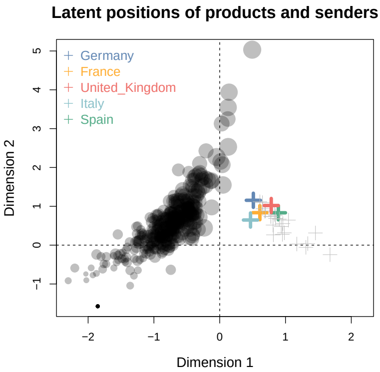


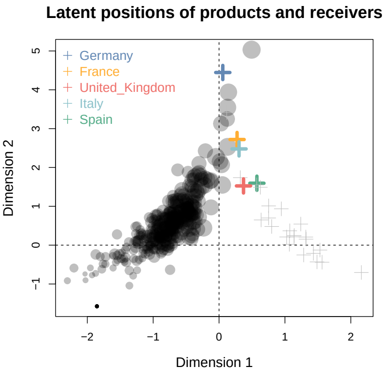


- Germany strongly aligns with the top-traded products: it imports/exports a large variety of them, acting as a sort of hub also for new products
- Some EU countries (on the right side) are located literally opposite to the least traded products: these countries do not trade those products at all and the corresponding predictors would be the lowest.

FAO trade results

Predictors are $\boldsymbol{\alpha}_{ij}^{\top}\mathbf{z}^{(k)}$, so the **angle formed with** $(0,0)$ determines how likely an edge is to appear.

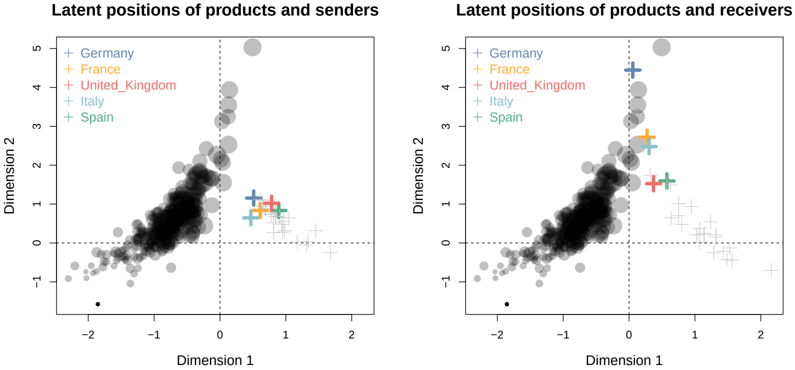




- Germany strongly aligns with the top-traded products: it imports/exports a large variety of them, acting as a sort of hub also for new products
- Some EU countries (on the right side) are located literally opposite to the least traded products: these countries do not trade those products at all and the corresponding predictors would be the lowest.

FAO trade results

Predictors are $\boldsymbol{\alpha}_{ij}^\top \mathbf{z}^{(k)}$, so the **angle formed with** $(0,0)$ determines how likely an edge is to appear.



- Germany strongly aligns with the top-traded products: it imports/exports a large variety of them, acting as a sort of hub also for new products
- Some EU countries (on the right side) are located literally opposite to the least traded products: these countries do not trade those products at all and the corresponding predictors would be the lowest.

Take home message

- New model-based **visualisations** of multi-layer networks.
- **Laplace approximation** is very convenient in this framework.
- Interesting connections to theory on M-estimators, **asymptotic results** on consistency.
- Possibility to use a Poisson distribution and **connect GGLLVM to the Gravity Model for trading flows**.

Take home message

- New model-based **visualisations** of multi-layer networks.
- **Laplace approximation** is very convenient in this framework.
- Interesting connections to theory on M-estimators, **asymptotic results** on consistency.
- Possibility to use a Poisson distribution and **connect GGLLVM to the Gravity Model for trading flows**.

Take home message

- New model-based **visualisations** of multi-layer networks.
- **Laplace approximation** is very convenient in this framework.
- Interesting connections to theory on M-estimators, **asymptotic results** on consistency.
- Possibility to use a Poisson distribution and **connect GGLLVM to the Gravity Model for trading flows**.

Take home message

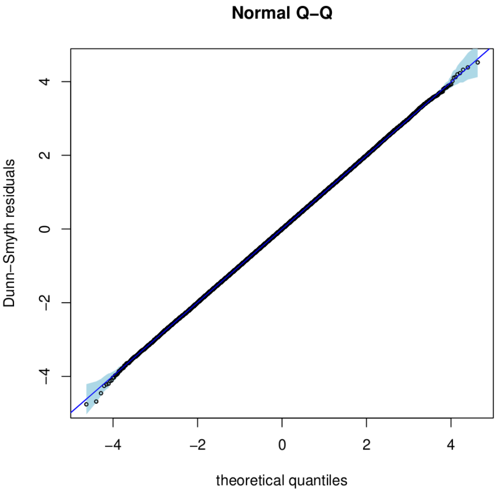
- New model-based **visualisations** of multi-layer networks.
- **Laplace approximation** is very convenient in this framework.
- Interesting connections to theory on M-estimators, **asymptotic results** on consistency.
- Possibility to use a Poisson distribution and **connect GGLLVM to the Gravity Model for trading flows**.

Take home message

- New model-based **visualisations** of multi-layer networks.
- **Laplace approximation** is very convenient in this framework.
- Interesting connections to theory on M-estimators, **asymptotic results** on consistency.
- Possibility to use a Poisson distribution and **connect GGLLVM to the Gravity Model for trading flows**.

Appendix: FAO trade results

Randomised quantile-quantile residuals plot.



Another statistical standpoint: Bayes theorem and posterior distribution inference

A Bayesian perspective offers great help in the interpretation of the GGLLVM: it allows us to highlight some connections between the latent factors and the observed dyads.

Further elaboration: since the Y_{ij}s are observable, we have

$$\boxed{f(\boldsymbol{y}) = g(\boldsymbol{y}|\boldsymbol{z})h(\boldsymbol{z})},$$

where h is the prior distribution of the latent variables and g is the conditional density of $\mathbf{Y}|\mathbf{Z}$. We are interested in what can be known about $\mathbf{Z}$ after that $\mathbf{Y}$ has been observed. This is expressed by the conditional density deduced from Bayes theorem:

$$\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \propto h(\boldsymbol{z})g(\boldsymbol{y}|\boldsymbol{z}),$$

namely the posterior distribution.

Another statistical standpoint: Bayes theorem and posterior distribution inference

A Bayesian perspective offers great help in the interpretation of the GGLLVM: it allows us to highlight some connections between the latent factors and the observed dyads.

Further elaboration: since the Y_{ij}s are observable, we have

$$\boxed{f(\boldsymbol{y}) = g(\boldsymbol{y}|\boldsymbol{z})h(\boldsymbol{z})},$$

where h is the prior distribution of the latent variables and g is the conditional density of $\mathbf{Y}|\mathbf{Z}$. We are interested in what can be known about $\mathbf{Z}$ after that $\mathbf{Y}$ has been observed. This is expressed by the conditional density deduced from Bayes theorem:

$$\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \propto h(\boldsymbol{z})g(\boldsymbol{y}|\boldsymbol{z}),$$

namely the posterior distribution.

Another statistical standpoint: Bayes theorem and posterior distribution inference

In our construction, for some h and g_{ij}, we have

$$f(\boldsymbol{y}) = g(\boldsymbol{y}|\boldsymbol{z})h(\boldsymbol{z}) = \left\{\prod_{i \neq j}^{n_V} g_{ij}(y_{ij}|\boldsymbol{z})\right\} h(\boldsymbol{z}),$$

and the Bayes theorem yields

$$\boxed{\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \propto \prod_{i \neq j}^{n_V} g_{ij}(y_{ij}|\boldsymbol{z})h(\boldsymbol{z})}.$$

We propose a convenient family of distributions g_{ij}:

$$\boxed{\tilde{\eta}_{ij} = \boldsymbol{\alpha}'_{ij}\mathbf{z} = \sum_{\ell=1}^{q} \alpha^{\ell}_{ij} z_{\ell}},$$

where α^{ℓ}_{ij} is the ℓ-th element of $\boldsymbol{\alpha}_{ij}$.

Another statistical standpoint: Bayes theorem and posterior distribution inference

In our construction, for some h and g_{ij}, we have

$$f(\boldsymbol{y}) = g(\boldsymbol{y}|\boldsymbol{z})h(\boldsymbol{z}) = \left\{ \prod_{i \neq j}^{n_V} g_{ij}(y_{ij}|\boldsymbol{z}) \right\} h(\boldsymbol{z}),$$

and the Bayes theorem yields

$$\boxed{\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \propto \prod_{i \neq j}^{n_V} g_{ij}(y_{ij}|\boldsymbol{z}) h(\boldsymbol{z})}.$$

We propose a convenient family of distributions g_{ij}:

$$\boxed{\tilde{\eta}_{ij} = \boldsymbol{\alpha}'_{ij}\mathbf{z} = \sum_{\ell=1}^{q} \alpha^{\ell}_{ij} z_{\ell}},$$

where α^{ℓ}_{ij} is the ℓ-th element of $\boldsymbol{\alpha}_{ij}$.

Another statistical standpoint: Bayes theorem and posterior distribution inference

Setting $\tilde{\boldsymbol{\eta}}$ as the vector obtained staking the canonical parameters $\tilde{\eta}_{ij}$ for all i, j, simple algebra yields

$$\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \quad \propto \quad h(\boldsymbol{z})W(\tilde{\boldsymbol{\eta}})\exp\left\{\sum_{i\neq j}^{n_V}\tilde{\eta}_{ij}u_{ij}(y_{ij})\right\}, \tag{7}$$

where $W(\cdot)$ is real-valued function of $\tilde{\boldsymbol{\eta}}$ and each $u_{ij}(\cdot)$ is a real-valued function that transforms the observed dyadic values (it can change with every pair ij). Now, let us notice that, setting
$U_\ell = \sum_{i\neq j}^{n_V}\alpha_{ij}^{\ell}u_{ij}(y_{ij})$, we have

$$\exp\left\{\sum_{i\neq j}^{n_V}\tilde{\eta}_{ij}u_{ij}(y_{ij})\right\} = \exp\left\{\sum_{\ell=1}^{q} z_\ell U_\ell\right\}.$$

$\Rightarrow$ the posterior distribution of the latent factors in (7) depends on the observable dyads only through the q-dimensional vector $\boldsymbol{U} = (U_1, ..., U_\ell, ..., U_q)$, which has dimension equal to that of $\mathbf{Z}$.

Another statistical standpoint: Bayes theorem and posterior distribution inference

Setting $\tilde{\boldsymbol{\eta}}$ as the vector obtained staking the canonical parameters $\tilde{\eta}_{ij}$ for all i, j, simple algebra yields

$$\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \quad \propto \quad h(\boldsymbol{z})W(\tilde{\boldsymbol{\eta}}) \exp\left\{\sum_{i \neq j}^{n_V} \tilde{\eta}_{ij} u_{ij}(y_{ij})\right\}, \tag{7}$$

where $W(\cdot)$ is real-valued function of $\tilde{\boldsymbol{\eta}}$ and each $u_{ij}(\cdot)$ is a real-valued function that transforms the observed dyadic values (it can change with every pair ij). Now, let us notice that, setting $U_\ell = \sum_{i \neq j}^{n_V} \alpha_{ij}^{\ell} u_{ij}(y_{ij})$, we have

$$\exp\left\{\sum_{i \neq j}^{n_V} \tilde{\eta}_{ij} u_{ij}(y_{ij})\right\} = \exp\left\{\sum_{\ell=1}^{q} z_\ell U_\ell\right\}.$$

$\Rightarrow$ the posterior distribution of the latent factors in (7) depends on the observable dyads only through the q-dimensional vector $\boldsymbol{U} = (U_1, ..., U_\ell, ..., U_q)$, which has dimension equal to that of $\mathbf{Z}$.

Another statistical standpoint: Bayes theorem and posterior distribution inference

In the Bayesian sense, $\boldsymbol{U}$ is a minimal sufficient statistic for $\mathbf{Z}$: it yields a dimensionality reduction from the observable dyads to the q-dimensional vector $\boldsymbol{U}$.

The above arguments imply that the dimensionality reduction does not depend on h: the use of minimal sufficient statistics does not entail any information loss for the calculation of $\underline{h}(\boldsymbol{z}|\boldsymbol{y})$ and they can be obtained without any specific reference to h. This is reassuring due to the arbitrariness in the choice of the prior distribution.

Another statistical standpoint: Bayes theorem and posterior distribution inference

In the Bayesian sense, $\boldsymbol{U}$ is a minimal sufficient statistic for $\mathbf{Z}$: it yields a dimensionality reduction from the observable dyads to the q-dimensional vector $\boldsymbol{U}$.

The above arguments imply that the dimensionality reduction does not depend on h: the use of minimal sufficient statistics does not entail any information loss for the calculation of $\underline{h}(\boldsymbol{z}|\boldsymbol{y})$ and they can be obtained without any specific reference to h. This is reassuring due to the arbitrariness in the choice of the prior distribution.

Another statistical standpoint: Bayes theorem and posterior distribution inference

In the Bayesian sense, $\boldsymbol{U}$ is a minimal sufficient statistic for $\mathbf{Z}$: it yields a dimensionality reduction from the observable dyads to the q-dimensional vector $\boldsymbol{U}$.

The above arguments imply that the dimensionality reduction does not depend on h: the use of minimal sufficient statistics does not entail any information loss for the calculation of $\underline{h}(\boldsymbol{z}|\boldsymbol{y})$ and they can be obtained without any specific reference to h. This is reassuring due to the arbitrariness in the choice of the prior distribution.

Another statistical standpoint: Bayes theorem and posterior distribution inference

Remark

When $\boldsymbol{Y}$ is the adjacency matrix, we have

$$U_\ell = \sum_{i \neq j}^{n_V} \alpha_{ij}^\ell y_{ij}, \text{ for } \ell = 1, ..., q.$$

This implies that ***all that we can learn about the latent variables given the observed adjacency matrix can be, without loss of information, summed up in the linear combinations of the Y_{ij}s****. Since each Y_{ij} can be either zero or one, the q minimal sufficient statistics for $\mathbf{Z}$ contain linear combinations of the factor loadings, where the different $\{\alpha_{ij}^\ell,\ \ell = 1, .., q\}$ imply that each U_ℓ is obtained giving different weights to the non zero edges: each minimal sufficient statistic for the latent variables is obtained assigning different weights to the observed network topology.*

Jump Back

The GGLLVM including observable covariates can capture the gravity trade model and possibly extend it by incorporating latent variables to account for unobserved effects. To illustrate how, consider the **structural disaggregated gravity model** :

$$Y_{ij}^{(k)} = \exp\left\{\boldsymbol{\beta}^{(k)\prime}\mathbf{X}_{ij}\right\}\epsilon_{ij}^{(k)} \tag{8}$$

- $\mathbf{X}_{ij}$: dyadic (e.g. dummy for: commercial agreements, sharing a border, were in colonial relationship) or node specific (GDP) covariates to model trades
- ϵ_{ij}^{k} is an error term with positive support (e.g. Poisson)

To my knowledge, the model parameters are estimated using either OLS on $\ln Y_{ij}^{(k)}$ or Poisson Pseudo-Maximum Likelihood (**PPML**) method, in which one assumes that ϵ_{ij}^{k} is a Poisson random variable.

Poisson case: connection to the gravity model

For the GGLLVM, we model Y_{ij} with a Poisson distribution. The marginal log-likelihood with fixed effects has the following form:

$$\sum_{k=1}^{K}\left(-\frac{1}{2}\ln\left[\det\left\{\Gamma\left(\boldsymbol{\alpha},\boldsymbol{\beta}^{(k)},\mathbf{z}^{(k)}\right)\right\}\right]+\sum_{i\neq j}^{n_V}\left\{Y_{ij}^{(k)}(\boldsymbol{\alpha}_{ij}'\mathbf{z}^{(k)}+\boldsymbol{\beta}^{(k)\prime}\mathbf{x}_{ij})\right.\right.$$
$$\left.\left.-\exp\left(\boldsymbol{\alpha}_{ij}'\mathbf{z}^{(k)}+\boldsymbol{\beta}^{(k)\prime}\mathbf{x}_{ij}\right)-\ln\left(y_{ij}^{(k)}!\right)\right\}-\frac{(\mathbf{z}_{(2)}^{(k)})'\mathbf{z}_{(2)}^{(k)}}{2}\right) \tag{9}$$

For the gravity model, the PPML objective function is given by

$$\sum_{k=1}^{K}\sum_{i\neq j}^{n_V}\left\{y_{ij}^{(k)}(\boldsymbol{\beta}^{(k)\prime}\mathbf{x}_{ij})-\exp\left(\boldsymbol{\beta}^{(k)\prime}\mathbf{x}_{ij}\right)-\log\left(y_{ij}^{(k)}!\right)\right\}. \tag{10}$$

Remark

Setting $\boldsymbol{\alpha}=\mathbf{0}$, the approximated log-likelihood of the GGLLVM becomes the same as the one for the gravity trade equation.

Jump back to Eq. (1)

The GGLLVM and its connection to Bai's factor model

Bai *(Econometrica, 2009)* introduced a panel data model with fixed effects, which has the following structure (in his notation):

$$
\begin{aligned}
Y_{it} &= X'_{it}\beta + u_{it}, \\
u_{it} &= \lambda'_i F_t + \epsilon_{it}, \quad (i = 1, 2, ..., N, \ t = 1, 2, ..., T)
\end{aligned}
$$

- X'_{it} are observable covariates, β are unknwon parameters of interest
- λ'_i is a vector of factor loadings and F_t are common factors, unobservable
- $\hat{F}_t, \hat{\lambda}_i, \hat{\beta}$ are estimated via a concentrated approach

The key ingredients of this model are similar to the ones of the GLLVM and GLAMLE, with changes:

$$\underbrace{i}_{\text{Bai's notation}} \mapsto \underbrace{ij}_{\text{GLLLVM}} \quad \text{and} \quad \underbrace{t}_{\text{Bai's notation}} \mapsto \underbrace{k}_{\text{GLLLVM}}$$

and with Gaussian errors.

Remark

Bai develops an asymptotic theory with $N, T \to \infty$ and the ϵ_{it}s can be weakly cross-sectionally correlated. How about the GLAMLE? $N, K \to \infty$? Graphs asymptotic behaviour?