GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT
UNIVERSITÉ DE GENÈVE

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

Davide La Vecchia

joint work with C. Jiang and R. Rastelli

CREST, 7-March -2024

Structure

- Motivation: Food and Agriculture Organization of the United Nations (FAO) dataset
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- Generalised Linear Latent Variable Models (GLLVM) and Graph Laplace-Approximated MLE (GLAMLE)
  - Definition
  - Inference
- Synthetic data
- FAO data (reprise)
- Take home message

Structure

- Motivation: Food and Agriculture Organization of the United Nations (FAO) dataset
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- Generalised Linear Latent Variable Models (GLLVM) and Graph Laplace-Approximated MLE (GLAMLE)
    - Definition
    - Inference
- Synthetic data
- FAO data (reprise)
- Take home message

Structure

- Motivation: Food and Agriculture Organization of the United Nations (FAO) dataset
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- Generalised Linear Latent Variable Models (GLLVM) and Graph Laplace-Approximated MLE (GLAMLE)
  - Definition
  - Inference
- Synthetic data
- FAO data (reprise)
- Take home message

Structure

- Motivation: Food and Agriculture Organization of the United Nations (FAO) dataset
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- Generalised Linear Latent Variable Models (GLLVM) and Graph Laplace-Approximated MLE (GLAMLE)
  - Definition
  - Inference
- Synthetic data
- FAO data (reprise)
- Take home message

Structure

- Motivation: Food and Agriculture Organization of the United Nations (FAO) dataset
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- Generalised Linear Latent Variable Models (GLLVM) and Graph Laplace-Approximated MLE (GLAMLE)
  - Definition
  - Inference
- Synthetic data
- FAO data (reprise)
- Take home message

Structure

- Motivation: Food and Agriculture Organization of the United Nations (FAO) dataset
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- Generalised Linear Latent Variable Models (GLLVM) and Graph Laplace-Approximated MLE (GLAMLE)
  - Definition
  - Inference
- Synthetic data
- FAO data (reprise)
- Take home message

Structure

- Motivation: Food and Agriculture Organization of the United Nations (FAO) dataset
- Formalization of the inference problem
- Related work (quick and partial/incomplete literature review)
- Generalised Linear Latent Variable Models (GLLVM) and Graph Laplace-Approximated MLE (GLAMLE)
  - Definition
  - Inference
- Synthetic data
- FAO data (reprise)
- Take home message

Motivation

In our daily life, we are involved in different types of complex systems, related to economics, social sciences, transportation, energy to name some examples.

The data collected on those systems are of different nature: e.g. they can be binary, count, continuous data or mixture thereof. These complex systems can be thought of as networks; see e.g. Kolaczyk (2009); Lusher et al. (2013); Kolaczyk and Csardi (2014).

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**.

Motivation

In our daily life, we are involved in different types of complex systems, related to economics, social sciences, transportation, energy to name some examples.

The data collected on those systems are of different nature: e.g. they can be binary, count, continuous data or mixture thereof. These complex systems can be thought of as networks; see e.g. Kolaczyk (2009); Lusher et al. (2013); Kolaczyk and Csardi (2014).

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**.

Motivation

In our daily life, we are involved in different types of complex systems, related to economics, social sciences, transportation, energy to name some examples.

The data collected on those systems are of different nature: e.g. they can be binary, count, continuous data or mixture thereof. These complex systems can be thought of as networks; see e.g. Kolaczyk (2009); Lusher et al. (2013); Kolaczyk and Csardi (2014).

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**.

Motivation

- 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).
- The import/export network is an economic network in which 364 layers represent commodities (primarily food products), 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 (i.e. the data is binary).

Motivation

- 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).
- The import/export network is an economic network in which 364 layers represent commodities (primarily food products), 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 (i.e. the data is binary).

Motivation

- 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).
- The import/export network is an economic network in which 364 layers represent commodities (primarily food products), 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 (i.e. the data is binary).

Motivation

- 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).
- The import/export network is an economic network in which 364 layers represent commodities (primarily food products), 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 (i.e. the data is binary).

Motivation

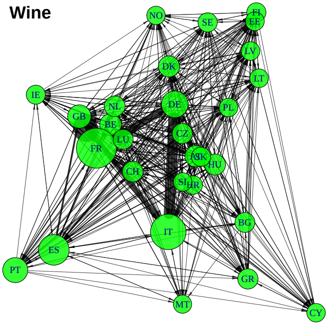
FAO network summary: Number of different products traded by each of the countries

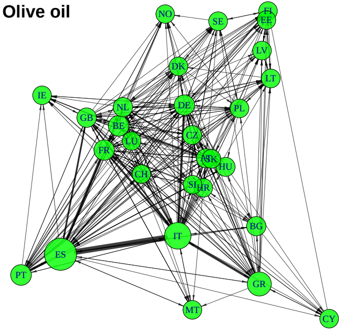
Belgium	343	Austria	327	Sweden	299	Luxembourg	265
Germany	343	Czech Republic	326	Ireland	297	Estonia	251
Netherlands	341	Poland	321	Slovakia	297	Slovenia	248
France	338	Hungary	317	Lithuania	294	Finland	245
Spain	337	Denmark	313	Bulgaria	283	Croatia	175
Italy	336	Portugal	304	Romania	276	Cyprus	173
United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

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

Motivation

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





Motivation

Inferential goal: We aim at modeling the **probability** of a *commercial relationship*. 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 derive a **novel inference approach** able to estimate the latent structure
- we provide the ultimate user with a computationally **efficient inference** procedure and algorithm.

Motivation

Inferential goal: We aim at modeling the **probability** of a *commercial relationship*. 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 derive a **novel inference approach** able to estimate the latent structure
- we provide the ultimate user with a computationally **efficient inference** procedure and algorithm.

Motivation

Inferential goal: We aim at modeling the **probability** of a *commercial relationship*. 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 derive a **novel inference approach** able to estimate the latent structure
- we provide the ultimate user with a computationally **efficient inference** procedure and algorithm.

Formalization of the inference problem

Let $\mathcal{G} = (V, E)$ be a random graph, where V is the set of vertices/nodes, E is the set of edges and $n_V = |V|$.

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

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

Remark

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

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.

  Limitations:

  - intractable likelihood function
  - heavy computational cost

  $\Rightarrow$ they often become impractical and not suitable for the FAO data analysis.

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.

  Limitations:

  - intractable likelihood function
  - heavy computational cost

  $\Rightarrow$ they often become impractical and not suitable for the FAO data analysis.

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, whereby the nodes of the network are characterized by their individual latent coordinates, which in turn determine their connectivity patterns.

Remark

Inference is typically done 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

- 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, whereby the nodes of the network are characterized by their individual latent coordinates, which in turn determine their connectivity patterns.

Remark

Inference is typically done 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 $\mathbf{Z}=\{\mathbf{z}_i\}_{i\in V}$ are parameters and latent factors to be estimated.

Example

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

$$\log\left(\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}\right)=\theta_1+X_{ij}\theta_2-D_{ij}$$

  with $\theta:=(\theta_1,\theta_2)\in\mathbb{R}$, X_{ij} is a univariate observable covariate, and $D_{ij}=|\mathbf{Z}_i-\mathbf{Z}_j|$.

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 $\mathbf{Z} = \{\mathbf{z}_i\}_{i \in V}$ are parameters and latent factors to be estimated.

Example

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

$$\log \left(\frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} \right) = \theta_1 + X_{ij}\theta_2 - D_{ij}$$

with $\boldsymbol{\theta} := (\theta_1, \theta_2) \in \mathbb{R}$, X_{ij} is a univariate observable covariate, and $D_{ij} = |\mathbf{Z}_i - \mathbf{Z}_j|$.

Generalised Linear Latent Variable Models (GLLVM)

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

- The additional latent variable term adds more *flexibility*.
- Inference typically done with **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)
- GLLVM have been widely-applied in psychology, ecology and social-sciences. To apply them to our network data,, we need to adapt the GLLVM to our network/graph setting.

Generalised Linear Latent Variable Models (GLLVM)

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

- The additional latent variable term adds more *flexibility*.
- Inference typically done with **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)
- GLLVM have been widely-applied in psychology, ecology and social-sciences. To apply them to our network data,, we need to adapt the GLLVM to our network/graph setting.

Generalised Linear Latent Variable Models (GLLVM)

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

- The additional latent variable term adds more *flexibility*.
- Inference typically done with **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)
- GLLVM have been widely-applied in psychology, ecology and social-sciences. To apply them to our network data,, we need to adapt the GLLVM to our network/graph setting.

Generalised Linear Latent Variable Models (GLLVM)

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

- The additional latent variable term adds more *flexibility*.
- Inference typically done with **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)
- GLLVM have been widely-applied in psychology, ecology and social-sciences. To apply them to our network data,, we need to adapt the GLLVM to our network/graph setting.

Generalised Linear Latent Variable Models (GLLVM)

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

- The additional latent variable term adds more *flexibility*.
- Inference typically done with **MCMC**, **Variational approx** or **Laplace approx**; see Huber et al. (2004), Niku et al. (2019)
- GLLVM have been widely-applied in psychology, ecology and social-sciences. To apply them to our network data,, we need to adapt the GLLVM to our network/graph setting.

Graph Generalised Linear Latent Variable Models (GGLLVM)

The generic ingredients of our Graph GLLVM (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

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

Graph Generalised Linear Latent Variable Models (GGLLVM)

The generic ingredients of our Graph GLLVM (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

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

Graph Generalised Linear Latent Variable Models (GGLLVM)

The generic ingredients of our Graph GLLVM (GGLLVM) are:

Data is $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(K)}$, 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

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

Graph Generalised Linear Latent Variable Models (GGLLVM)

The generic ingredients of our Graph GLLVM (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

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

  namely,

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

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

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

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 $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\alpha})$

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

$$\ell(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \log \left(\int \left\{ \prod_{i \neq j} g_{ij}(Y_{ij}^{(k)} | \mathbf{z}) \right\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)} \right).$$

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 $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\alpha})$

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

$$\ell(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \log \left(\int \left\{ \prod_{i \neq j} g_{ij}(Y_{ij}^{(k)} | \mathbf{z}) \right\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)} \right).$$

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

$$\ell(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \log \left(\int \left\{ \prod_{i \neq j} g_{ij}(Y_{ij}^{(k)} | \mathbf{z}) \right\} h(\mathbf{z}_{(2)}) d\mathbf{z}_{(2)} \right).$$

Inference (GLAMLE)

The treatment of $\ell(\boldsymbol{\alpha})$ 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

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

where m is the number of dyads and

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

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

Inference (GLAMLE)

The treatment of $\ell(\boldsymbol{\alpha})$ 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

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

where m is the number of dyads and

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

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

$$\tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \left(\frac{2\pi}{m}\right)^{q/2} \det\left\{-U\left(\hat{\mathbf{z}}^{(k)}\right)\right\}^{-1/2} \exp\left\{mQ\left(\boldsymbol{\alpha}, \hat{\mathbf{z}}^{(k)}, \boldsymbol{y}^{(k)}\right)\right\}, \quad (3)$$

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

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

Inference (GLAMLE)

- We derive the Laplace-approximated marginal density function

$$\tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \left(\frac{2\pi}{m}\right)^{q/2} \det\left\{-U\left(\hat{\mathbf{z}}^{(k)}\right)\right\}^{-1/2} \exp\left\{mQ\left(\boldsymbol{\alpha}, \hat{\mathbf{z}}^{(k)}, \boldsymbol{y}^{(k)}\right)\right\}, \tag{3}$$

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

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

Inference (GLAMLE)

- So, we have an approximate density which we apply to approximate the exact intractable likelihood. 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

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

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

$$\boxed{\tilde{\mathcal{S}}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \tilde{\mathcal{S}}^{(k)}(\boldsymbol{\alpha}) = \mathbf{0}.} \tag{4}$$

Inference (GLAMLE)

- So, we have an approximate density which we apply to approximate the exact intractable likelihood. 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

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

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

$$\boxed{\tilde{\mathcal{S}}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \tilde{\mathcal{S}}^{(k)}(\boldsymbol{\alpha}) = \mathbf{0}.} \tag{4}$$

Inference (GLAMLE)

- So, we have an approximate density which we apply to approximate the exact intractable likelihood. 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

$$\boxed{\tilde{\boldsymbol{\mathcal{S}}}(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \tilde{\boldsymbol{\mathcal{S}}}^{(k)}(\boldsymbol{\alpha}) = \mathbf{0}.} \tag{4}$$

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*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$
- *In analogy with (4), 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* $\tilde{\ell}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$, *in the sense of White (1982):*
  - *Asymptotic regime (i): rate of convergence to the pseudo-true value* $K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}}))$, *as* $K \to \infty$
  - *Asymptotic regime (ii): rate of convergence to the true value, as* $K \to \infty$ *and* $m \to \infty$.
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities...*

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*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$
- *In analogy with (4), 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* $\tilde{\ell}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$, *in the sense of White (1982):*
  - *Asymptotic regime (i): rate of convergence to the pseudo-true value* $K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}}))$, *as* $K \to \infty$
  - *Asymptotic regime (ii): rate of convergence to the true value, as* $K \to \infty$ *and* $m \to \infty$.
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities...*

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*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$
- *In analogy with (4), 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* $\tilde{\ell}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$*, in the sense of White (1982):*
  - *Asymptotic regime (i): rate of convergence to the pseudo-true value* $K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}}))$*, as* $K \to \infty$
  - *Asymptotic regime (ii): rate of convergence to the true value, as* $K \to \infty$ *and* $m \to \infty$*.*
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities...*

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*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$
- *In analogy with (4), 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 $\tilde{\ell}$ is a pseudo likelihood associated to the functional $\tilde{\boldsymbol{\alpha}}$, in the sense of White (1982):*
    - *Asymptotic regime (i): rate of convergence to the pseudo-true value $K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}}))$, as $K \to \infty$*
    - *Asymptotic regime (ii): rate of convergence to the true value, as $K \to \infty$ and $m \to \infty$.*
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for $\mathbf{Z}$ which are functions of the observed quantities...*

Inference (GLAMLE)

Remark (Statistical properties/guarantees)

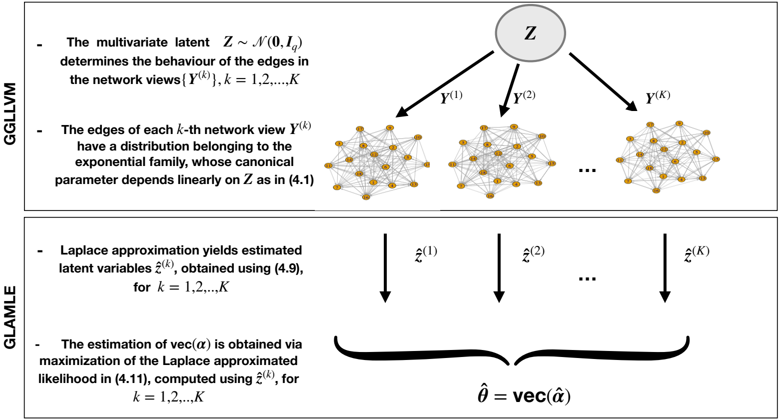
- *The* $\hat{\mathbf{z}}_{(2)}^{(k)}$ *can be formally interpreted as the maximum likelihood estimates of the latent factors in the k-th network view*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$
- *In analogy with (4), 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* $\tilde{\ell}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$, *in the sense of White (1982):*
  - *Asymptotic regime (i): rate of convergence to the pseudo-true value* $K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}}))$, *as* $K \to \infty$
  - *Asymptotic regime (ii): rate of convergence to the true value, as* $K \to \infty$ *and* $m \to \infty$.
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities...*

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*
- $f_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) = \tilde{f}_{\boldsymbol{\alpha}}(\boldsymbol{y}^{(k)}) \left\{1 + O\left(m^{-1}\right)\right\}$
- *In analogy with (4), 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* $\tilde{\ell}$ *is a pseudo likelihood associated to the functional* $\tilde{\boldsymbol{\alpha}}$, *in the sense of White (1982):*
  - *Asymptotic regime (i): rate of convergence to the pseudo-true value* $K^{1/2}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{V}(\tilde{\boldsymbol{\alpha}}))$, *as* $K \to \infty$
  - *Asymptotic regime (ii): rate of convergence to the true value, as* $K \to \infty$ *and* $m \to \infty$.
- *There is a Bayesian interpretation in terms of minimal sufficient statistics for* $\mathbf{Z}$ *which are functions of the observed quantities...*

To summarize via a graphical illustration:



Synthetic data

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

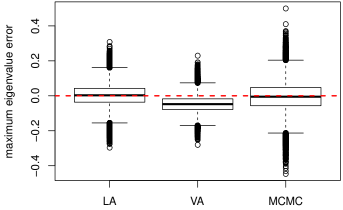
- We focus on Bernoulli random variables for a simulated directed network with $n_V = 18$ and we study the estimation problem of an adjacency matrix
- We consider 1000 Monte Carlo runs, where in each run, we set $K = 100$, namely, we have 100 network layers
- To investigate on the role of the number 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**.

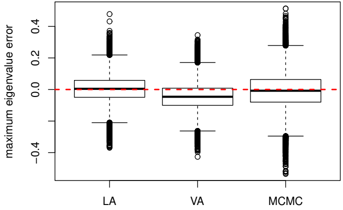
Synthetic data

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

$q = 1$



$q = 2$

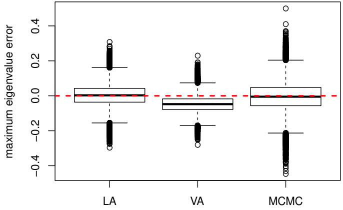


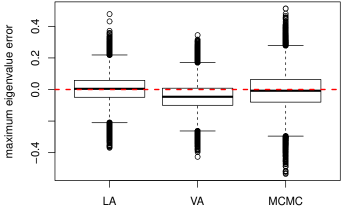
(a) large bias appears with VA (fast); (b) MCMC provides accurate estimates (slow), but it has a slightly larger interquartile range than LA and more outliers; (c) LA has small bias (mid)

Synthetic data

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

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



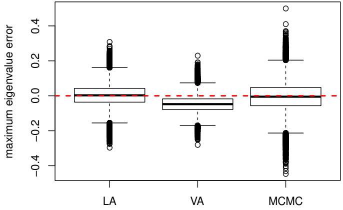


(a) large bias appears with VA (fast); (b) MCMC provides accurate estimates (slow), but it has a slightly larger interquartile range than LA and more outliers; (c) LA has small bias (mid)

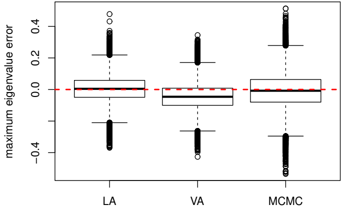
Synthetic data

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

$q = 1$



$q = 2$

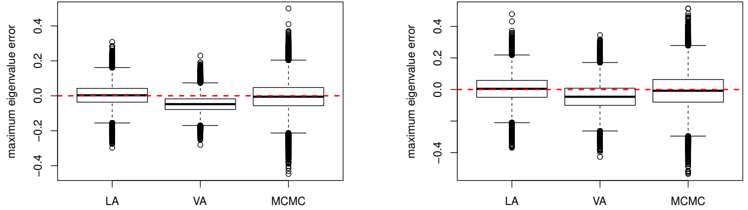


(a) large bias appears with VA (fast); (b) MCMC provides accurate estimates (slow), but it has a slightly larger interquartile range than LA and more outliers; (c) LA has small bias (mid)

Synthetic data

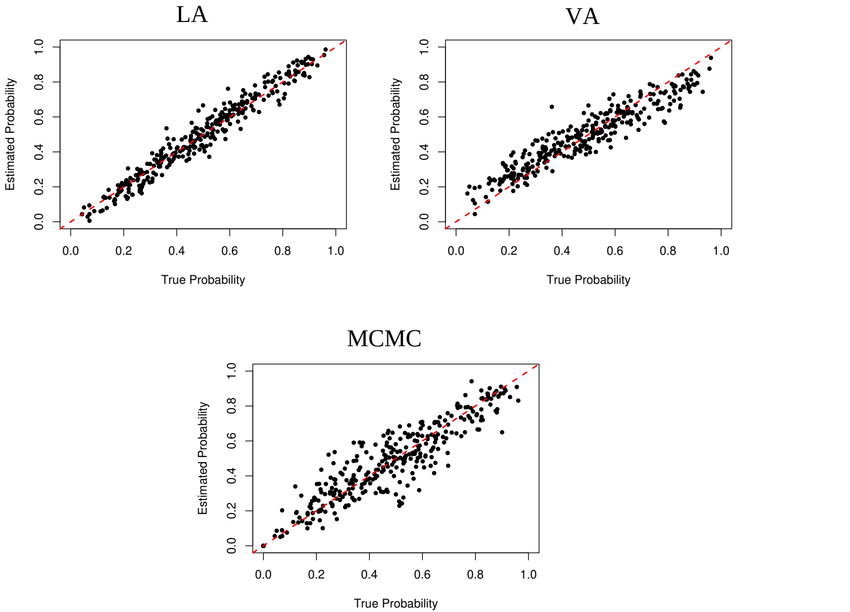
Maximum eigenvalue difference between connection probability matrices ($\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}_0$, for each layer and then averaged):

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



(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 $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{z}}^{(k)}$, for $k = 1, 2..., K$.

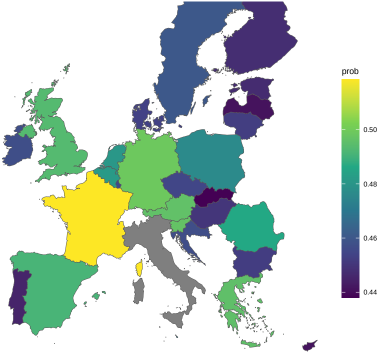
FAO data (reprise)

- 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 $k = 1, 2..., K$.

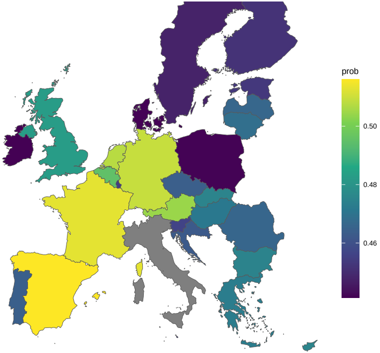
FAO data (reprise)

These parameters yield an estimated probability for import and export for each country. Here an example for Italy:

Export



Import

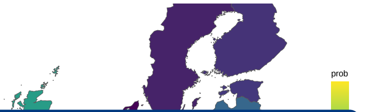


FAO data (reprise)

These parameters yield an estimated probability for import and export for each country. Here an example for Italy:

Export

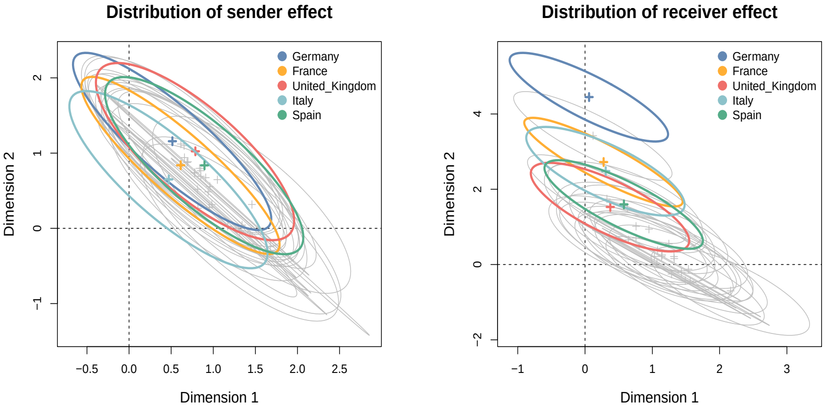
Import



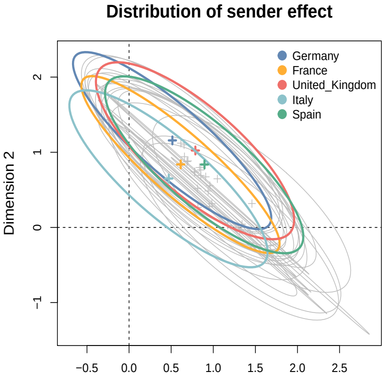
Remark

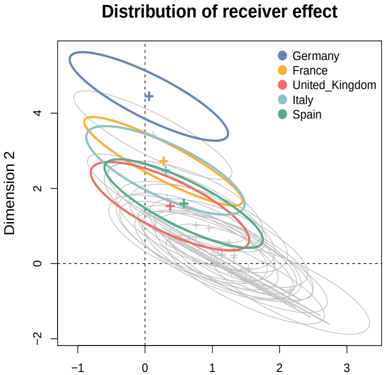
These probability values are obtained averaging across all the layers, so they represent an aggregate view of a country's import/export, and may be interpreted as a prediction on trading partners for a new product: we see a strong tendency of Italy to export products to France, and to import products from a number of countries including Spain, France, Germany, Austria, Netherlands.

FAO data (reprise)



FAO data (reprise)

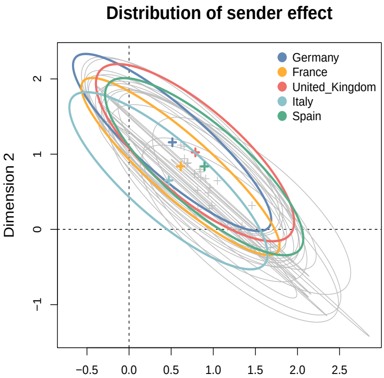


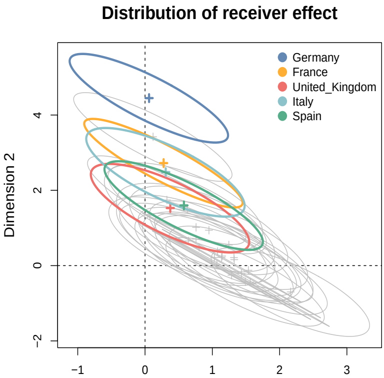


Remark (i)

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).*

Take home message

Pre-print available at: `https://arxiv.org/abs/2107.05915`

- 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.

Take home message

Pre-print available at: `https://arxiv.org/abs/2107.05915`

- 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.

Take home message

Pre-print available at: `https://arxiv.org/abs/2107.05915`

- 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.

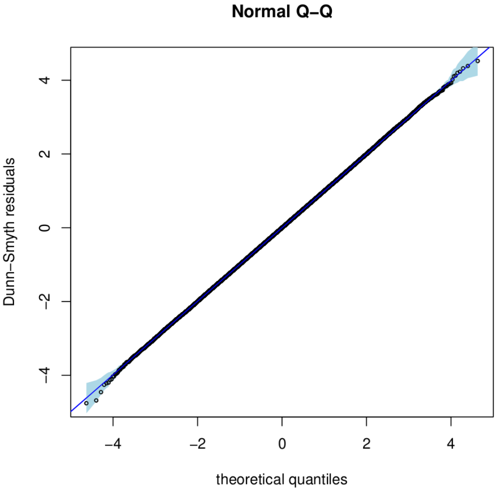
Take home message

Pre-print available at: `https://arxiv.org/abs/2107.05915`

- 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.

Appendix: FAO trade results

Randomised quantile-quantile residuals plot.



Another statistical standpoint: Bayes theorem and posterior distribution inference

Adopting a Bayesian perspective offers great help in the interpretation of the GGLLVM and on the related inference aspects. Indeed, besides facilitating the use of MCMC inference procedures, it allows us to highlight some connections between the latent factors and the dyads on the graph.

Another statistical standpoint: Bayes theorem and posterior distribution inference

To elaborate further, let us assume that all variables are continuous and admit a density. Since only the Y_{ij}s are observable, we have

$$f(\boldsymbol{y}) = g(\boldsymbol{y}|\boldsymbol{z})h(\boldsymbol{z}), \tag{5}$$

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

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

We propose a convenient family of distributions g_{ij}: the one-parameter exponential family, where the canonical parameter $\tilde{\eta}_{ij}$ is expressed as a linear combination of the latent variables, namely

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

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

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

We propose a convenient family of distributions g_{ij}: the one-parameter exponential family, where the canonical parameter $\tilde{\eta}_{ij}$ is expressed as a linear combination of the latent variables, namely

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

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

$$\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \quad \propto \quad h(\boldsymbol{z})W(\tilde{\boldsymbol{\eta}}) \exp\left\{\sum_{i\neq j}^{n_V} \tilde{\eta}_{ij} u_{ij}(y_{ij})\right\}, \tag{6}$$

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

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

This implies that that the posterior distribution of the latent factors in (6) 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

$$\underline{h}(\boldsymbol{z}|\boldsymbol{y}) \quad \propto \quad h(\boldsymbol{z})W(\tilde{\boldsymbol{\eta}})\exp\left\{\sum_{i\neq j}^{n_V}\tilde{\eta}_{ij}u_{ij}(y_{ij})\right\}, \tag{6}$$

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

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

This implies that that the posterior distribution of the latent factors in (6) 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 $\boldsymbol{U}$—of course it should be noted that the functions $\{u_{ij}\}$ involve some unknown parameters (the factor loadings) and then, in speaking of U_ℓ as a statistic, we are abusing of the frequentist usage according to which sufficient statistics are functions of the observed data only.

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 (except for the general requirement that it has to be compatible with (5)). 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 $\boldsymbol{U}$—of course it should be noted that the functions $\{u_{ij}\}$ involve some unknown parameters (the factor loadings) and then, in speaking of U_ℓ as a statistic, we are abusing of the frequentist usage according to which sufficient statistics are functions of the observed data only.
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 (except for the general requirement that it has to be compatible with (5)). 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 $\boldsymbol{U}$—of course it should be noted that the functions $\{u_{ij}\}$ involve some unknown parameters (the factor loadings) and then, in speaking of U_ℓ as a statistic, we are abusing of the frequentist usage according to which sufficient statistics are functions of the observed data only.

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 (except for the general requirement that it has to be compatible with (5)). 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.*