Latent Space Models for Climate-related News Interaction

Antonio Peruzzi and Roberto Casarin

Ca' Foscari University of Venice

GRINS Workshop 2024, Venice,
December 2nd, 2024

This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 - CUP H73C22000930001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Università Ca'Foscari Venezia

Introduction

Motivation:

- **Climate risks management** has become of prominent importance in the agenda of the EU;
- Having information about the **public awareness** concerning **climate risks** is **instrumental** to the decision maker (e.g. Eurobarometer Surveys).
- **Online Social Media** offer an interesting, although approximate, perspective on public opinion.

Our Project:

- We aim at providing a set of **measures of climate risk awareness** based on **Social Media interaction data** together with **uncertainty quantification**;
- We would like to leverage on our experience on **dynamic Latent Space models** which allow to extract a dynamic set of latent features from network data.

Introduction

Motivation:

- **Climate risks management** has become of prominent importance in the agenda of the EU;
- Having information about the **public awareness** concerning **climate risks** is **instrumental** to the decision maker (e.g. Eurobarometer Surveys).
- **Online Social Media** offer an interesting, although approximate, perspective on public opinion.

Our Project:

- We aim at providing a set of **measures of climate risk awareness** based on **Social Media interaction data** together with **uncertainty quantification**;
- We would like to leverage on our experience on **dynamic Latent Space models** which allow to extract a dynamic set of latent features from network data.

About Latent Space models...

Latent Space (LS) models are generative models for **networks** in which:

- the edge-specific **weights** are treated as **random variables**;
- the value of the weights depends on **node-specific** d-dimensional **latent factors** (latent coordinates);
- **the higher the similarity** of two nodes in terms of latent factors **the higher the edge weight**.

LS models project the nodes of a network onto a d-dimensional latent space and the relative location of the nodes on this space reveals interesting characteristics of the whole network.

This family of models has been **introduced by Hoff, Raftery, and Handcock, 2002** and finds **application** in several fields: e.g. **political science** (Barberá, 2015; Park and Sohn, 2020; Yu and Rodriguez, 2021), **finance** (Linardi et al., 2020; Casarin and Peruzzi, 2024), and **neuroscience** (Durante, Dunson, and Vogelstein, 2017; Wilson, Cranmer, and Lu, 2020).

Università Ca'Foscari Venezia

About Latent Space models...

Latent Space (LS) models are generative models for **networks** in which:

- the edge-specific **weights** are treated as **random variables**;
- the value of the weights depends on **node-specific** d-dimensional **latent factors** (latent coordinates);
- **the higher the similarity** of two nodes in terms of latent factors **the higher the edge weight**.

LS models project the nodes of a network onto a d-dimensional latent space and the relative location of the nodes on this space reveals interesting characteristics of the whole network.

This family of models has been **introduced by Hoff, Raftery, and Handcock, 2002** and finds **application** in several fields: e.g. **political science** (Barberá, 2015; Park and Sohn, 2020; Yu and Rodriguez, 2021), **finance** (Linardi et al., 2020; Casarin and Peruzzi, 2024), and **neuroscience** (Durante, Dunson, and Vogelstein, 2017; Wilson, Cranmer, and Lu, 2020).

Università Ca'Foscari Venezia

About Latent Space models...

Latent Space (LS) models are generative models for **networks** in which:

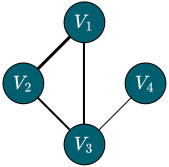
- the edge-specific **weights** are treated as **random variables**;
- the value of the weights depends on **node-specific** d-dimensional **latent factors** (latent coordinates);
- **the higher the similarity** of two nodes in terms of latent factors **the higher the edge weight**.

LS models project the nodes of a network onto a d**-dimensional latent space** and the relative location of the nodes on this space reveals interesting characteristics of the whole network.

This family of models has been **introduced by Hoff, Raftery, and Handcock, 2002** and finds **application** in several fields: e.g. **political science** (Barberá, 2015; Park and Sohn, 2020; Yu and Rodriguez, 2021), **finance** (Linardi et al., 2020; Casarin and Peruzzi, 2024), and **neuroscience** (Durante, Dunson, and Vogelstein, 2017; Wilson, Cranmer, and Lu, 2020).

Introduction

Input data

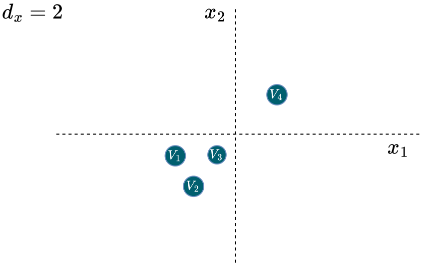


$$Y = [y_{ij}] = \begin{bmatrix} - & 3 & 2 & 0 \\ 3 & - & 2 & 0 \\ 2 & 2 & - & 1 \\ 0 & 0 & 1 & - \end{bmatrix}$$

Latent Space Model

$$y_{ij} \sim \mathcal{P}oi(\lambda_{ij})$$
$$\lambda_{ij} = g(\mu - ||\mathbf{x}_i - \mathbf{x}_j||)$$

Output



Dynamic Latent Space models

Let $\mathcal{G}_t = (V, E_t, Y_t)$ for $t = 1, \ldots, T$ be a weighted temporal network where $V \subset \mathbb{N}$ denotes the vertex set, $E_t \subset V \times V$ denotes the edge set, and Y_t is a $N \times N$ weighted adjacency matrix.

The LS model assumes:

$$y_{ijt} \sim f(y_{ij} | g(\eta_{ijt}), \kappa),$$

for $i, j = 1, \ldots, N$, $i \neq j$ with parameters $g(\eta_{ijt})$ and κ.

The parameter $g(\eta_{ijt})$ is driven by node-specific dynamic latent features:

$$\eta_{ijt} = \mu_t - ||\mathbf{x}_{it} - \mathbf{x}_{jt}||,$$

where $||x - y||$ is a distance between the d-dimensional node-specific set of coordinates $\mathbf{x}_{it}$ for $i = 1, \ldots, N$.

Dynamic component:

Time independence:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_d)$.

Random Walk:

- $\mu_t \sim \mathcal{N}(\mu_{t-1}, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{x}_{it-1}, \sigma^2 I_d)$.

Markov Switching:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} = \tilde{\mathbf{x}}_{i,S_t=k}$.
- $\tilde{\mathbf{x}}_{i,k} \sim \mathcal{N}(\mathbf{0}$

Dynamic Latent Space models

Let $\mathcal{G}_t = (V, E_t, Y_t)$ for $t = 1, \dots, T$ be a weighted temporal network where $V \subset \mathbb{N}$ denotes the vertex set, $E_t \subset V \times V$ denotes the edge set, and Y_t is a $N \times N$ weighted adjacency matrix.

The LS model assumes:

$$y_{ijt} \sim f(y_{ij} | g(\eta_{ijt}), \kappa),$$

for $i, j = 1, \dots, N$, $i \neq j$ with parameters $g(\eta_{ijt})$ and κ.

The parameter $g(\eta_{ijt})$ is driven by node-specific dynamic latent features:

$$\eta_{ijt} = \mu_t - ||\mathbf{x}_{it} - \mathbf{x}_{jt}||,$$

where $||x - y||$ is a distance between the d-dimensional node-specific set of coordinates $\mathbf{x}_{it}$ for $i = 1, \dots, N$.

Dynamic component:

Time independence:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_d)$.

Random Walk:

- $\mu_t \sim \mathcal{N}(\mu_{t-1}, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{x}_{it-1}, \sigma^2 I_d)$.

Markov Switching:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} = \tilde{\mathbf{x}}_{i,S_t=k}$.
- $\tilde{\mathbf{x}}_{i,k} \sim \mathcal{N}(\mathbf{0}$

Dynamic Latent Space models

Let $\mathcal{G}_t = (V, E_t, Y_t)$ for $t = 1, \dots, T$ be a weighted temporal network where $V \subset \mathbb{N}$ denotes the vertex set, $E_t \subset V \times V$ denotes the edge set, and Y_t is a $N \times N$ weighted adjacency matrix.

The LS model assumes:

$$y_{ijt} \sim f(y_{ij} | g(\eta_{ijt}), \kappa),$$

for $i, j = 1, \dots, N$, $i \neq j$ with parameters $g(\eta_{ijt})$ and κ.

The parameter $g(\eta_{ijt})$ is driven by node-specific dynamic latent features:

$$\eta_{ijt} = \mu_t - ||\mathbf{x}_{it} - \mathbf{x}_{jt}||,$$

where $||x - y||$ is a distance between the d-dimensional node-specific set of coordinates $\mathbf{x}_{it}$ for $i = 1, \dots, N$.

Dynamic component:

Time independence:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_d)$.

Random Walk:

- $\mu_t \sim \mathcal{N}(\mu_{t-1}, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{x}_{it-1}, \sigma^2 I_d)$.

Markov Switching:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} = \tilde{\mathbf{x}}_{i,S_t=k}$.
- $\tilde{\mathbf{x}}_{i,k} \sim \mathcal{N}(\mathbf{0},$ [illegible]

Dynamic Latent Space models

Let $\mathcal{G}_t = (V, E_t, Y_t)$ for $t = 1, \ldots, T$ be a weighted temporal network where $V \subset \mathbb{N}$ denotes the vertex set, $E_t \subset V \times V$ denotes the edge set, and Y_t is a $N \times N$ weighted adjacency matrix.

The LS model assumes:

$$y_{ijt} \sim f(y_{ij} | g(\eta_{ijt}), \kappa),$$

for $i, j = 1, \ldots, N$, $i \neq j$ with parameters $g(\eta_{ijt})$ and κ.

The parameter $g(\eta_{ijt})$ is driven by node-specific dynamic latent features:

$$\eta_{ijt} = \mu_t - ||\mathbf{x}_{it} - \mathbf{x}_{jt}||,$$

where $||x - y||$ is a distance between the d-dimensional node-specific set of coordinates $\mathbf{x}_{it}$ for $i = 1, \ldots, N$.

Dynamic component:

Time independence:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_d)$.

Random Walk:

- $\mu_t \sim \mathcal{N}(\mu_{t-1}, \sigma^2_\mu)$;
- $\mathbf{x}_{it} \sim \mathcal{N}(\mathbf{x}_{it-1}, \sigma^2 I_d)$.

Markov Switching:

- $\mu_t \sim \mathcal{N}(0, \sigma^2_\mu)$;
- $\mathbf{x}_{it} = \tilde{\mathbf{x}}_{i,S_t=k}$.
- $\tilde{\mathbf{x}}_{i,k} \sim \mathcal{N}(\mathbf{0}, \sigma^2_k I_d)$

Università Ca'Foscari Venezia

Estimation

For LS models, the joint posterior distribution is not tractable. It is common practice to use **Metropolis-within-Gibbs** to approximate the posterior distribution.

For a simple time independent LS model, a systematic Gibbs sampling algorithm (GS) iterates the following steps for each h and t:

- Draw $\mu_t^{(h)}$ from $\pi(\mu_t|\ldots)$ via Metropolis-Hastings (MH);
- Draw $\boldsymbol{x}_{it}^{(h)}$ from $\pi(\boldsymbol{x}_{it}|\ldots)$ $i = 1, \ldots, N$ and via **Adaptive MH** (AMH).

Identification Issues:

- The latent coordinates are identified up to **translation, rotation and reflection**;
- These issues are solved with a combination of **on-the-fly re-centering**, $\sum_i x_{itd} = 0$ for each d, and post-processing (**Procrustes transformation**).

Estimation

For LS models, the joint posterior distribution is not tractable. It is common practice to use **Metropolis-within-Gibbs** to approximate the posterior distribution.

For a simple time independent LS model, a systematic Gibbs sampling algorithm (GS) iterates the following steps for each h and t:

- Draw $\mu_t^{(h)}$ from $\pi(\mu_t | \dots)$ via Metropolis-Hastings (MH);
- Draw $\boldsymbol{x}_{it}^{(h)}$ from $\pi(\boldsymbol{x}_{it} | \dots)$ $i = 1, \dots, N$ and via **Adaptive MH** (AMH).

Identification Issues:

- The latent coordinates are identified up to **translation, rotation and reflection**;
- These issues are solved with a combination of **on-the-fly re-centering**, $\sum_i x_{itd} = 0$ for each d, and post-processing (**Procrustes transformation**).

The Raw Data

Facebook data gathered by (Schmidt et al., 2018):

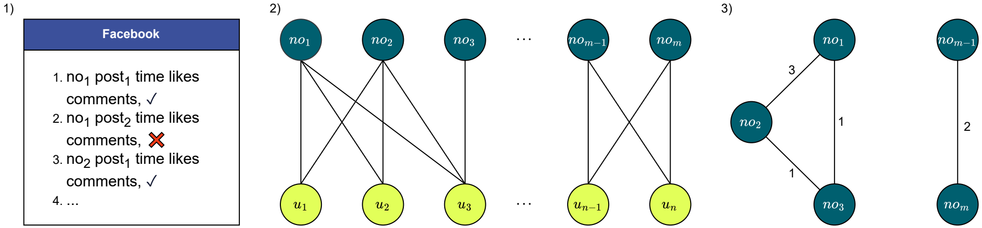
- on **National** and **Local news outlets** from:
  - France (65 outlets);
  - Germany (49);
  - Italy (54);
  - Spain (47).
- Time lapse: Jan-2015 - Dec-2016;
- Data include: Posts, Likes, Comments and Shares;
- Users anonymized IDs and interactions with news outlets.

Country	Pages	Posts	Comments	Commenters	Likes	Likers
France	65	1,008,018	47,225,675	5,755,268	419, 371, 366	21, 647, 888
Germany	49	749,805	31,881,407	5,338,195	183, 599, 003	14, 367, 445
Italy	54	1,554,817	51,515,121	4,086,351	409, 243, 176	14, 012, 658
Spain	57	1,372,805	34,336,356	6,494,725	333, 698, 985	32, 812, 007

Climate-related News Dataset

Pipeline:

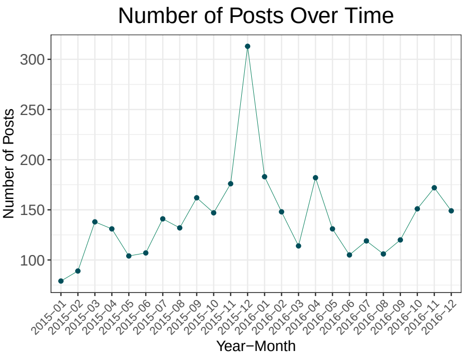
- Posts filtering according to Climate-related keywords;
    - Keywords related to: *zero emissions*, *carbon neutrality*, *greenhouse effect*, *floods*, *deforestation*, *wildfires*, *CO2 emissions*, *COP21*, *climate change*.
- Creation of Bipartite networks of interaction Users-News Outlets;
- Creation of Bipartite projection networks:
    - News outlet - News outlet;
    - User - User.

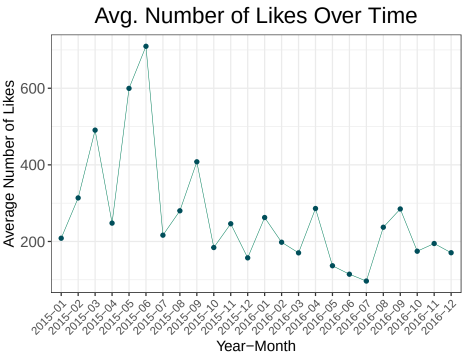


Climate-related News Dataset - Italy

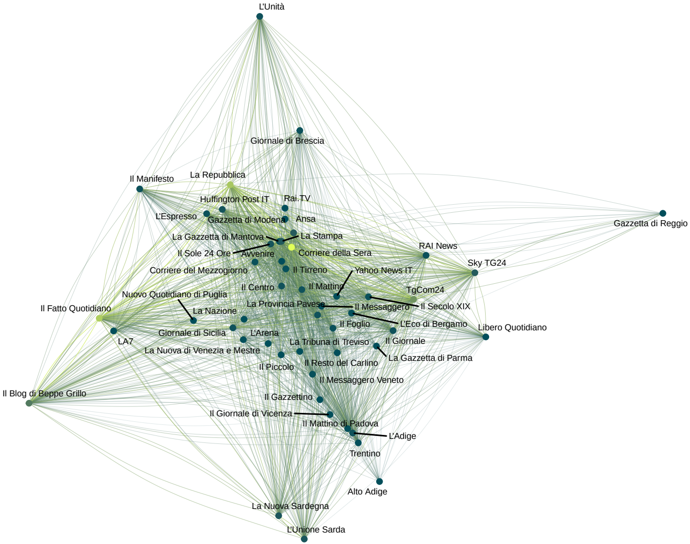
Resulting Dataset:

- Number of Posts: 3,399, 0.22% of total posts;
- Number of News Outlets involved: 54;
- Unique likers: 309,715, 2.2% of total unique likers.

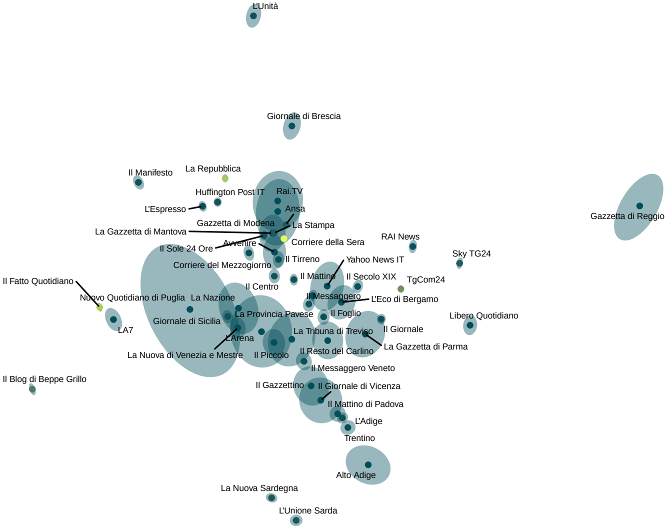




News Outlet Projection Network - Italy

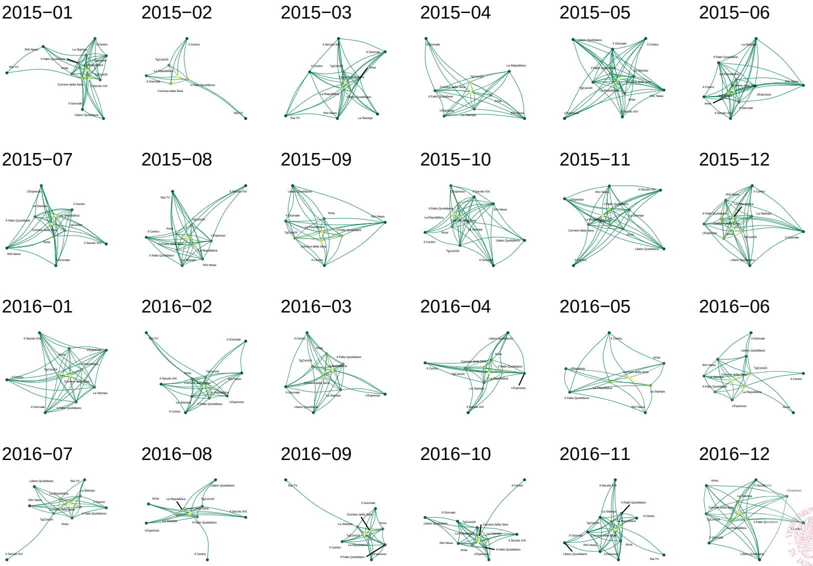


News Outlet Latent Space - Italy



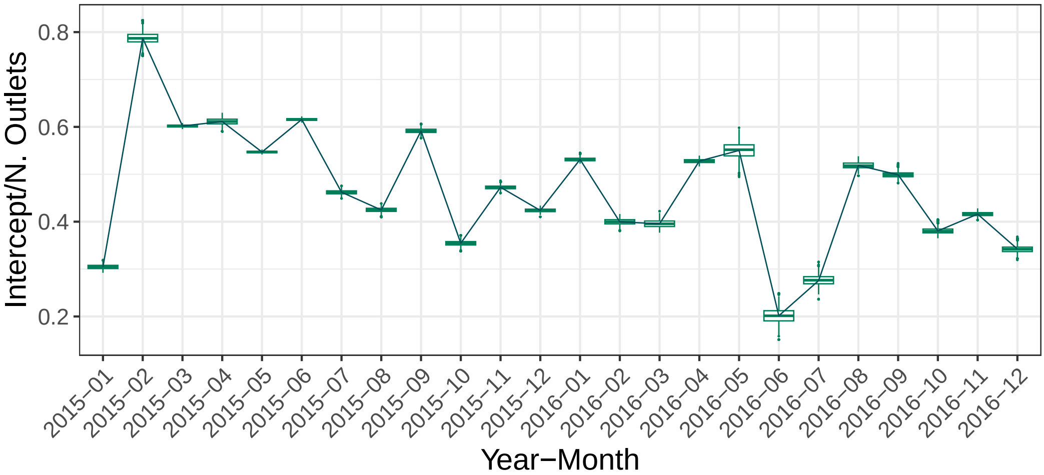
Università Ca'Foscari Venezia

News Outlet Temporal Network - Italy - National News Outlets



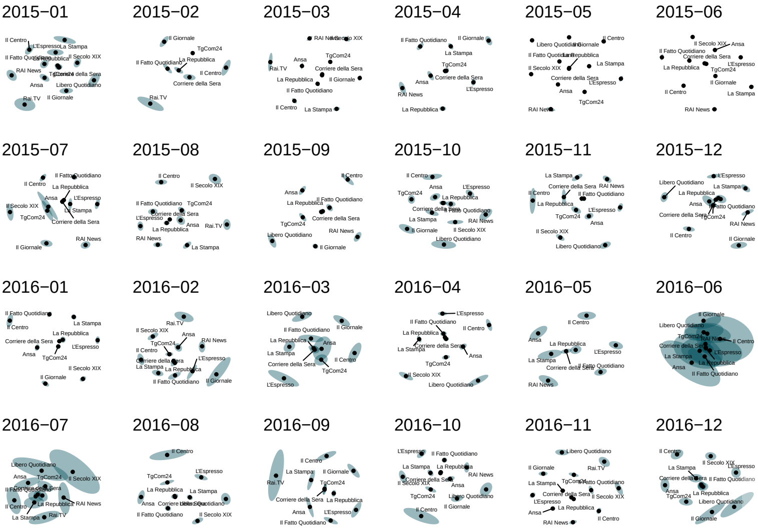
Università Ca'Foscari Venezia

News Outlet Latent Space Intercept - Italy - National News Outlets



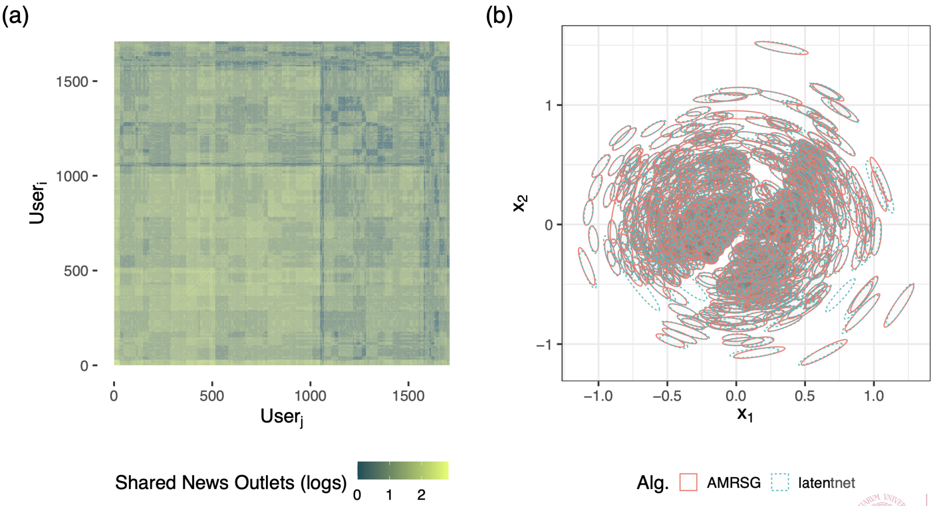
Università Ca'Foscari Venezia

News Outlet Latent Space - Italy



Università Ca'Foscari Venezia

Users Projection Network - Italy



Università Ca'Foscari Venezia

What as been done and further steps

What has been done:

- We constructed a dataset of Climate-related news on Facebook based on a set of keywords;
- We obtained a set of time-varying networks of climate-related news interaction based on bipartite projections;
- We provided a static and dynamic latent space representation of the network of news outlets and we provided a static latent space representation on the network of users.

What to do:

- Improve posts selection and look for more recent data (maybe use bluesky API?);
- Devise a suitable dynamic specification for the LS model based on the problem at hand;
- Devise and validate the set of climate risk awareness measures based on the LS estimation output.

Thank you for your time!

antonio.peruzzi@unive.it

References I

- Barberá, Pablo (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation using Twitter Data". In: *Political Analysis* 23.1, pp. 76–91.
- Casarin, Roberto and Antonio Peruzzi (2024). "A Dynamic Latent-Space Model for Asset Clustering". In: *Studies in Nonlinear Dynamics & Econometrics* 28.2, pp. 379–402.
- Durante, Daniele, David B Dunson, and Joshua T Vogelstein (2017). "Nonparametric Bayes Modeling of Populations of Networks". In: *Journal of the American Statistical Association* 112.520, pp. 1516–1530.
- Hoff, Peter D, Adrian E Raftery, and Mark S Handcock (2002). "Latent Space Approaches to Social Network Analysis". In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098.
- Linardi, Fernando et al. (2020). "Dynamic Interbank Network Analysis using Latent Space Models". In: *Journal of Economic Dynamics and Control* 112, p. 103792.
- Park, Jong Hee and Yunkyu Sohn (2020). "Detecting Structural Changes in Longitudinal Network Data". In: *Bayesian Analysis* 15.1, pp. 133 –157. doi: `10.1214/19-BA1147`. url: `https://doi.org/10.1214/19-BA1147`.

References II

- Schmidt, Ana Lucía et al. (2018). “Polarization Rank: A Study on European News Consumption on Facebook”. arXiv preprint arXiv:1805.08030.
- Wilson, James D, Skyler Cranmer, and Zhong-Lin Lu (2020). “A Hierarchical Latent Space Network Model for Population Studies of Functional Connectivity”. In: *Computational Brain & Behavior* 3, pp. 384–399.
- Yu, Xingchen and Abel Rodriguez (2021). “Spatial Voting Models in Circular Spaces: A Case Study of the US House of Representatives”. In: *The Annals of Applied Statistics* 15.4, pp. 1897–1922.