Media Bias and Polarization through the Lens of a Markov Switching Latent Space Network Model

Roberto Casarin [*]

Department of Economics, Ca' Foscari University of Venice

`r.casarin@unive.it`

Antonio Peruzzi

Department of Economics, Ca' Foscari University of Venice

`antonio.peruzzi@unive.it`

and

Mark F.J. Steel

Department of Statistics, University of Warwick

`m.steel@warwick.ac.uk`

August 5, 2023

Abstract

News outlets are now more than ever incentivized to provide their audience with slanted news, while the intrinsic homophilic nature of online social media may exacerbate polarized opinions. Here, we propose a new dynamic latent space model for time-varying online audience-duplication networks, which exploits social media content to conduct inference on media bias and polarization of news outlets. Our model contributes to the literature in several directions: 1) we provide a model-embedded data-driven interpretation for the latent leaning of news outlets in terms of media bias; 2) we endow our model with Markov-switching dynamics to capture polarization regimes while maintaining a parsimonious specification; 3) we contribute to the literature on the statistical properties of latent space network models. The proposed model is applied to a set of data on the online activity of national and local news outlets from four European countries in the years 2015 and 2016. We find evidence of a strong positive correlation between our media slant measure and a well-grounded external source of media bias. In addition, we provide insight into the polarization regimes across the four countries considered.

Keywords: Bayesian Inference, Latent Variables, Political Leaning, News Outlets.

1 Introduction

We propose a new statistical model able to offer meaningful insights into the perceived media bias and regime changes in polarization within online social media. The risks of being unintentionally exposed to biased news and polarized opinions gained awareness both in the public debate (WEF 2022) and in the academic sphere (see Puglisi and Snyder Jr 2015; Gentzkow et al. 2015; Cinelli et al. 2021) due to the rapid changes in the news consumption landscape (Newman et al. 2017).
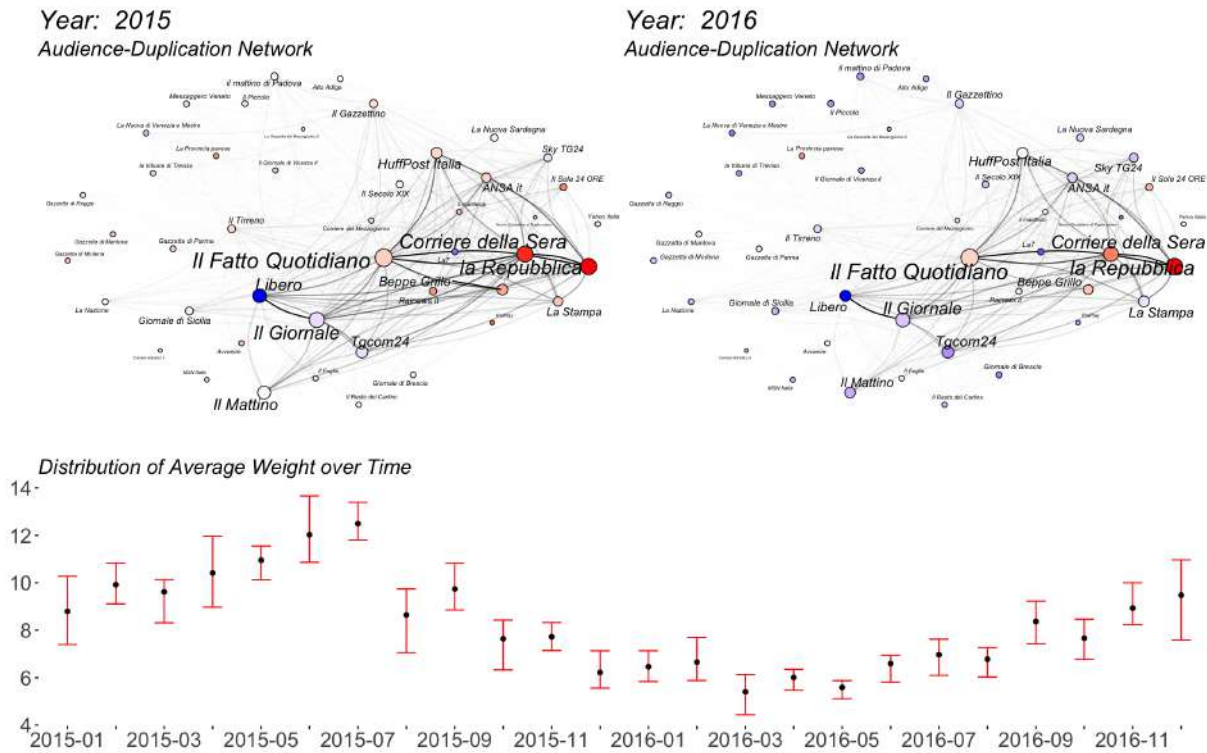

Figure 1: **Italian Audience-Duplication Networks** obtained from the bipartite network of Italian news outlets and their Facebook commenters in 2015 (top left) and 2016 (top right). Node size is proportional to the news outlets' engagement in terms of comments. Nodes are colored from red (left) to blue (right) according to the text-analysis political-leaning score computed following Gentzkow and Shapiro (2010) and Garz et al. (2020). Edge width is proportional to the number of commenters in common between any two news outlets. The bottom panel indicates networks' average edge weight - average number of commenters in common between any two outlets - over time, with red error bars displaying the intra-month inter-quartile range, while the black dot denotes the median.

Figure 1 provides an illustrative example of both media bias and polarization starting from a preliminary analysis of the dataset described in Section 4. The figure displays a network of Italian news outlets in which the edges' thickness is proportional to the number of Facebook commenters in common between any two outlets in the years 2015 (left) and 2016 (right). While media bias, in terms of political leaning, can be inferred indirectly from the structure of the network or directly by analyzing news outlets' content production

(top panel), an increase in polarization can be detected when the average number of users interacting with various sources decreases.

Media bias entails the propagation of biased news pieces, the aim of which is often to support the interest of some individuals or groups, such as political parties. The phenomenon is considered detrimental to consumer welfare (Gentzkow et al. 2015) as it entails a reduction in the informativeness of news pieces, while some argue that biased news outlets jointly driven by ideological interests and profits could even affect political outcomes (Anderson and McLaren 2012). Recent developments in measuring media bias include the implementation of text-analysis techniques to take into account the similarity between news articles and political content (see Gentzkow and Shapiro 2010; Garz et al. 2020) and the use of latent-space models (Hoff et al. 2002; Friel et al. 2016; Sewell and Chen 2016) to obtain a political leaning measure from social-media data (Barberá 2015).

Polarization refers to the radicalization of people's opinions in the sense that they are further apart. Some fear that this change in attitudes may be reflected in more partisan positions of people's representatives even though there is no obvious evidence of this (Prior 2013). Others claim that online social media exacerbate polarization by offering incentives for homophilous behavior, i.e. the tendency to interact with similar individuals (Dandekar et al. 2013). However, while a predisposition toward homophily has been observed on several social platforms (see Hanusch and Nölleke 2019; Cinelli et al. 2021), evidence of an exacerbation of polarization in social media environments is mixed (Kubin and von Sikorski 2021). Several different methodologies have been adopted for measuring polarization (see Esteban and Ray 1994; Yarchi et al. 2021), including in the field of network science (see Garimella et al. 2018; Cinelli et al. 2021). Two common objects of investigation are bipartite networks, relating social media users to online pages, and audience duplication networks, in which nodes represent pages and weighted edges denote the number of users in common between any pair of pages. Figure 2 illustrates the two concepts.

Previous studies about media polarization use heuristics to detect communities and informal sequential analysis for time variation, while we propose a formal statistical framework for dynamic polarization analysis. In particular, we introduce a novel dynamic Latent-Space (LS) network model which exploits both time-varying online audience-duplication network data and textual content to characterize a set of news outlets both in terms of a dynamic latent political-leaning dimension and in terms of popularity via individual effects.

LS models (Hoff et al., 2002) project the nodes of a network on a lower d-dimensional latent space. Extensions of the original model include e.g. a dynamic component for the latent coordinates (Friel et al., 2016; Sewell and Chen, 2016; Kim et al., 2018), mixtures of latent coordinates (see Handcock et al., 2007) or extensions to multi-layer networks (see Sosa and Betancourt, 2022). The statistical properties of LS models have been addressed by Rastelli et al. (2016) for binary networks and Barberá (2015) presents an early application of LS modeling for the estimation of latent ideology on social media.

Our paper adds to the methodological literature in several respects:

- We propose an LS model with Markov-Switching dynamics (MS-LS) which allows capturing polarization regimes using a parsimonious specification.

- We provide a model-embedded data-driven interpretation for the latent space in terms of media bias of news outlets exploiting text-based indicators (e.g. Gentzkow et al., 2015; Garz et al., 2020).
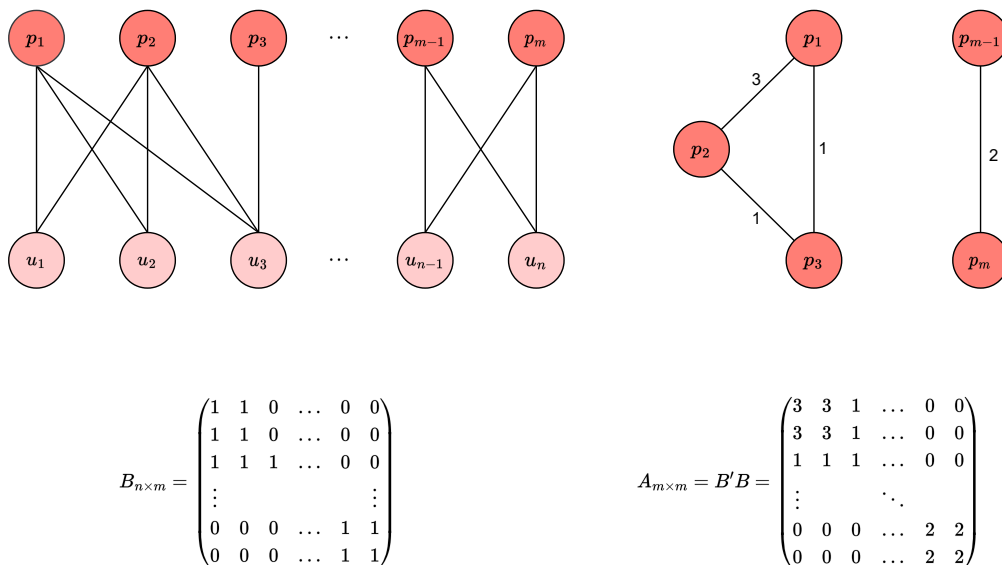
Figure 2: **Media Networks:** An example of a bipartite network of n readers, $u_1, \ldots, u_n$ and m news outlets, $p_1, \ldots, p_m$ (top left) and the corresponding audience duplication network obtained using one-mode projection (top right). At the bottom, the adjacency-matrix representation of both the bipartite network, $B_{n \times m}$, and the audience-duplication network $A_{m \times m}$. The matrix A is obtained as $A = B'B$.

- We extend the results of Rastelli et al. (2016) on statistical properties of LS models to weighted temporal networks with MS dynamics. These are general results that apply outside of the particular model we have implemented here.

Our model is applied to a novel time-varying network dataset we obtained from the Facebook daily online activity of a broad set of news outlets from four European countries (France, Germany, Italy, and Spain) in the years 2015 and 2016. This provides estimates of media bias that are coherent with the PEW Research survey (Mitchell et al., 2018). We also shed light on the in-platform (i.e. within Facebook) polarization regimes across the four countries. We do not find evidence of a common shift from a low to a high polarization regime in these data, in line with Prior (2013). The newly constructed dataset is freely available (see Appendix J).

Section 2 will be dedicated to the overall description of the model within a Bayesian setup and to the discussion of the statistical properties of the model. In Section 3, we discuss posterior inference along with the constraints used in our model and present a simulation exercise. Finally, Section 4 describes the dataset of European news outlets and applies our model in both a static and a dynamic setup.

2 The Longitudinal Markov-Switching Latent Space Model

2.1 The Model

Let $\mathcal{G} = \{\mathcal{G}_{rt}, r = 1, 2, \ldots, R, \, t = 1, 2, \ldots, T_r\}$ be an undirected and weighted temporal (in our application audience-duplication) network with multiple layers $\mathcal{G}_{rt} = (V_{rt}, E_{rt}, Y_{rt})$. For each layer (country) r, the vertex set (collection of news outlets) is constant, i.e. $V_{rt} = V_r \subset \mathbb{N}$, and the edge set $E_{rt} \subset V_r \times V_r$ (common commenters between any two outlets) is time-varying. For each edge $(i, j) \in E_{tr}$ we assume the (i, j)-th element of the weighted adjacency matrix Y_{rt}, i.e. Y_{ijrt}, is observed and denotes the number of connections (commenters in common) between news outlets i and j at time t in country r. We adopt a Poisson model for the connections:

$$Y_{ijrt} \overset{ind}{\sim} \mathcal{P}oi(\lambda_{ijrt}), \tag{1}$$

for $i, j = 1, \ldots, N_r$, $t = 1, \ldots, T_r$ and $r = 1, \ldots, R$, where $N_r = Card(V_r)$ is the number of nodes in layer r and $\mathcal{P}oi(\lambda)$ denotes a Poisson distribution with intensity parameter $\lambda > 0$. In our LS model, the intensity is driven by a set of static (α_{ir}) and d-dimensional dynamic node-specific latent features ($\mathbf{x}_{irt} \in \mathcal{X} \subset \mathbb{R}^d$):

$$\log \lambda_{ijrt} = \alpha_{ir} + \alpha_{jr} - \beta_r ||\mathbf{x}_{irt} - \mathbf{x}_{jrt}||^2. \tag{2}$$

The parameters α_{ir}, $i = 1, \ldots, N_r$ have the natural interpretation of individual effects which are news-outlet specific and can be considered a proxy of the popularity of the outlet (the engagement of the audience with the newspaper). The latent variables $\mathbf{x}_{irt}$, $i = 1, \ldots, N_r$ enter the log intensity through the squared Euclidean distance. This way, we account for the proximity of the news outlets on a given manifold. The more similar the news outlets (the closer the nodes), the higher the number of commenters they tend to have in common (if $\beta_r > 0$).

We also employ an observable political leaning proxy, ℓ_{irt}, to provide additional information on the location of news outlets within the latent space. Our modeling choice is to not include ℓ_{irt} in the log-intensity equation to preserve the tractability of the random graph model properties. Rather we assume that the political-leaning proxy ℓ_{irt} is driven by the same latent variable $\mathbf{x}_{irt}$ as in the network log-intensity:

$$\ell_{irt} \overset{ind}{\sim} \mathcal{B}e(\varphi(\gamma_{0r} + \boldsymbol{\gamma}'_{1r}\mathbf{x}_{irt})\phi_r, (1 - \varphi(\gamma_{0r} + \boldsymbol{\gamma}'_{1r}\mathbf{x}_{irt}))\phi_r), \tag{3}$$

where $\varphi(x) = 1/(1 + \exp(-x))$ is the logistic function, so that $\varphi(\gamma_{0r} + \boldsymbol{\gamma}'_{1r}\mathbf{x}_{irt})$ is the expectation of ℓ_{irt} and $\phi_r > 0$ is a precision parameter. This modeling choice allows endowing the latent feature $\mathbf{x}_{i,rt}$ with a media-bias interpretation while ensuring y_{ijrt} is independent of ℓ_{irt} conditional on $\mathbf{x}_{irt}$. Thus, the latent space can be interpreted as the political spectrum of news outlets.

To make our model dynamic, we assume that a Markov-Switching (MS) process drives the dynamic latent features. For this reason, we assume the existence of a Hidden Markov Chain with $K_r < \infty$ possible states of the world. The MS process allows our latent features $\mathbf{x}_{irt}$ to vary jointly through time across the different states. In the media environment, one

can e.g. think about a state of low polarization, where news outlets are perceived, on average, closer within the political spectrum, and a state of high polarization, in which news outlets are perceived politically further apart (e.g., see Macy et al. (2021); Leonard et al. (2021) on polarization dynamics). We can reparameterize our dynamic latent features $\mathbf{x}_{it}$ to account for this MS dynamic:

$$\mathbf{x}_{irt} = \sum_{k=1}^{K_r} \mathbb{I}(s_{rt} = k)\boldsymbol{\zeta}_{ikr}, \tag{4}$$

where $\boldsymbol{\zeta}_{ikr}$ denotes the latent position of news outlet i of country r in state k, while $\mathbb{I}(s_{rt} = k)$ is an indicator function which is 1 if the observed state $s_{rt} = k$ for country r at time t, and 0 otherwise. Moreover, we characterize the transition between states through:

$$\mathbb{P}(s_{rt} = k | s_{r,t-1} = l) = q_{lkr}, \ l, k = 1, \ldots, K_r, \tag{5}$$

which can be grouped in the transition probability matrix $\mathbf{Q}_r = \{\boldsymbol{q}_{1r}, \ldots, \boldsymbol{q}_{lr}, \ldots, \boldsymbol{q}_{K_r r}\}$, where $\boldsymbol{q}_{lr} = (q_{l1r}, \ldots, q_{lK_r r})$ denotes a column vector such that $\boldsymbol{q}'_{lr}\mathbf{1} = 1$ for each state l.

The use of a large number of parameters can produce over-fitting. Thus we take a Bayesian approach to inference and choose the following priors (as explained in Section 3.2 we fix β_r in our empirical application):

$$\alpha_{ir} \sim \mathcal{N}(0, \sigma_\alpha^2), \ \boldsymbol{\zeta}_{ikr} \sim \mathcal{N}(0, \sigma_{kr}^2 I_d), \sigma_{kr}^2 \sim IG(a_{\sigma^2}, b_{\sigma^2}), \tag{6}$$

$$\gamma_{0r} \sim \mathcal{N}(0, b_{\gamma_0}), \boldsymbol{\gamma}_{1r} \sim \mathcal{N}(0, b_{\gamma_1} I_d), \ \phi_r \sim \mathcal{G}(a_\phi, b_\phi), \ \boldsymbol{q}_{lr} \sim \mathcal{D}ir(\omega_1, \ldots, \omega_{K_r}), \tag{7}$$

where $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, $\mathcal{G}(a, b)$, $IG(a, b)$ and $\mathcal{D}ir(c, d)$ denote the Normal, the Gamma, the Inverse Gamma and the Dirichlet distribution respectively. The directed acyclic graph in Figure 3 summarizes our Bayesian MS-LS model. In our implementation of the MS-LS model with $d = 1$ and $K = 2$, we have little prior information at our disposal and we opt for the use of relatively vague priors to let the data speak: $\alpha_{ir} \sim \mathcal{N}(0, 15^2)$, $\zeta_{ikr} \sim \mathcal{N}(0, \sigma_{kr}^2)$, $\sigma_{kr}^2 \sim IG(0.1, 0.1)$, $\gamma_{0r} \sim \mathcal{N}(0, 15^2)$, $\gamma_{1r} \sim \mathcal{N}(0, 15^2)$, $\phi_r \sim \mathcal{G}(0.01, 0.01)$ and $\boldsymbol{q}_{lr} \sim \mathcal{D}ir(2, 2)$. Our results are robust to substantial changes in these priors.

2.2 Model Properties

We now present some of the properties of the MS-LS Model in (1), (2) and (4). With Assumption 2.1, we extend the scope of the Latent Variable Model provided in Rastelli et al. (2016) to weighted temporal networks. To enhance readability, we drop the country index r, and we drop the political-leaning equation (3) since this is quite specific to our application and we aim to present the properties of a general network model.

Assumption 2.1. *Given an undirected temporal network, $\mathcal{G}_t = (V, E_t)$, for $t = 1, 2, \ldots$ having vertex set $V \subset \mathbb{N}$ and weighted edge sets $E_t \subset V \times V$ with characteristic weight Y_{ijt}, we assume a sequence of latent coordinates $\{X_{1t}, \ldots, X_{Nt}\}$ for $t = 1, 2, \ldots$ with $X_{it} \in \mathcal{X} \subset \mathbb{R}^d$ for each node $i \in V$ and time index $t \in \mathbb{N}$.*

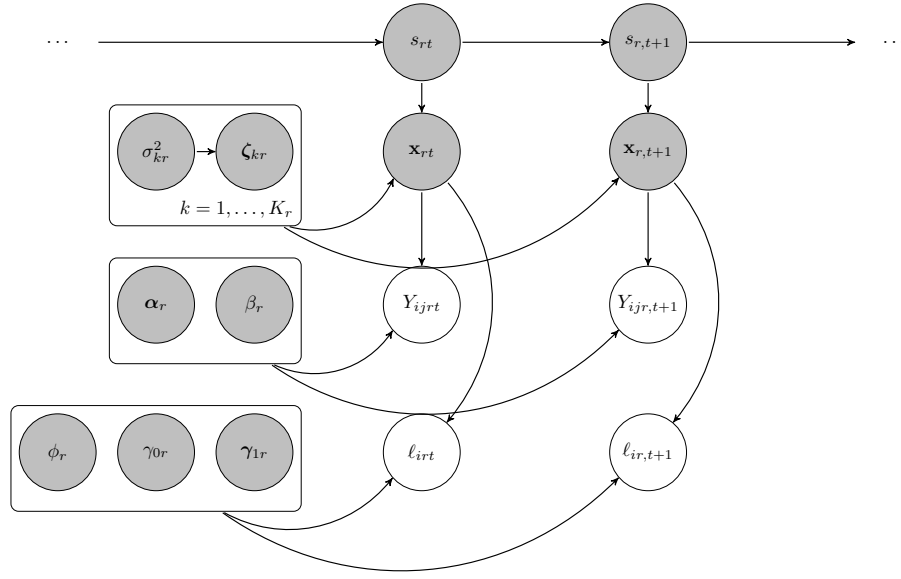With Assumption 2.2, we introduce the Markov-Switching dynamics.

Figure 3: **Directed Acyclic Graph** of the Markov-switching latent-space model. The graph exhibits the conditional independence structure of the observation model for y_{ijrt} and ℓ_{irt} within white circles with parameters, $\boldsymbol{\alpha}_r = \{\alpha_1, \ldots, \alpha_{N_r}\}$, β_r, γ_{0r}, $\boldsymbol{\gamma}_{1r}$, ϕ_r, latent coordinates $\mathbf{x}_{rt} = \{\mathbf{x}_{1,rt}, \ldots, \mathbf{x}_{N_r,rt}\}$, their state-dependent counterparts $\boldsymbol{\zeta}_{kr} = \{\boldsymbol{\zeta}_{1,kr}, \ldots, \boldsymbol{\zeta}_{N_r,kr}\}$ with variance σ^2_{kr}, and latent states, s_{rt}, within grey circles.

Assumption 2.2. *Given a K-state latent Markov-chain process $s_t \in \{1, 2, \ldots, K\}$ for $K < \infty$ and $t = 1, 2, \ldots$ with transition probabilities $q_{lk} = \mathbb{P}(s_t = k | s_{t-1} = l)$, we assume the latent variables $X_{it} = \sum_{k=1}^{K} \mathbb{I}(s_t = k) Z_{ik}$. We also define the set $\mathscr{Z}_k = \{\zeta_{1k}, \ldots, \zeta_{Nk}\}$ consisting of the i.i.d. realizations of the latent random variables $\{Z_{1k}, \ldots, Z_{Nk}\}$ with $k \in \{1, \ldots, K\}$, where each Z_{ik} is distributed according to $\pi_k(\cdot)$, a given probability measure.*

Assumption 2.3 introduces the conditional independence between any two edges given the latent variables and the current state of the world.

Assumption 2.3. *We assume conditional independence between any two edges given the latent variables for a given state s_t. Hence, $\forall (j, i) \in E_t$, $Y_{ijt} | X_{it}, s_t \sim \mathcal{P}oi(\lambda_{ijt})$ is a Poisson random variable with intensity parameter λ_{ijt}.*

Moreover, we assume that our set of latent variables is jointly normally distributed.

Assumption 2.4. *The realized latent variables belonging to $\mathscr{Z}_k$ in Assumption 2.2 are points in the Euclidean d-dimensional space, for a fixed d, and they are normally distributed:*

$$p(\mathscr{Z}_k \mid \sigma^2_k) = \prod_{i=1}^{N} f_d\left(\zeta_{ik}; \underline{0}, \sigma^2_k I_d\right)$$

Finally, we specify the form of the intensity parameter λ_{ijt}.

Assumption 2.5. *Given the individual effects α_i and α_j and the latent variables, we assume the Poisson rate parameter:*

$$\lambda_{ijt} = \exp\left\{\alpha_i + \alpha_j - \sum_{k=1}^{K} \mathbb{I}(s_t = k)\beta||\zeta_{ik} - \zeta_{jk}||^2\right\}$$

Under Assumptions 2.1-2.5 our model is a time-varying MS-LS model with Poisson weights and normally distributed latent variables.

The nodal strength, defined as $Y_{it} = \sum_{j\neq i} Y_{ijt}$, is a quantity of particular interest when dealing with weighted networks as it provides information on how strongly connected a node is with its neighbors. We can derive the following properties of the probability generating function (pgf) of the nodal strength, where in the sequel we will focus on the strength of a random node (and, thus, omit the index i):

Proposition 2.1. *The m-th derivative of the conditional pgf of the nodal strength evaluated in $x = 1$ given $s_{t-1} = l$ for the MS-LS model can be written as:*

$$\left.\frac{\partial^m G_l(x)}{\partial x^m}\right|_{x=1} = \sum_{k=1}^{K} q_{lk} \left.\frac{\partial^m \widetilde{G}_k(x)}{\partial x^m}\right|_{x=1} = \sum_{k=1}^{K} \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} e^{\sum_{j\neq i}(\alpha_i+\alpha_j)h_j} (\sigma_k^2)^{-\frac{d}{2}} b_{\underline{h},k} q_{lk}$$

where $\widetilde{G}_k(x)$ is the conditional pgf given $s_t = k$ and $s_{t-1} = l$ and

$$b_{\underline{h},k} = \left(\frac{1}{\sigma_k^2} + \sum_{j\in\mathcal{N}_i} \frac{2\beta h_j}{2\beta h_j \sigma_k^2 + 1}\right)^{-\frac{d}{2}} \prod_{j\in\mathcal{N}_i} \left(2\beta h_j \sigma_k^2 + 1\right)^{\frac{d}{2}}$$

with multi-index $\underline{h}_i = \{h_1, \ldots, h_{i-1}, h_{i+1}, \ldots, h_N\}$ and index set $\mathcal{H}_i = \{h_j \in \{0, \ldots, m\}, \quad j \neq i| \sum_{j\neq i} h_j = m\}$, $\mathcal{N}_i = \{j|j \neq i, h_j > 0\}$ and $\beta > 0$.

The first derivative of the pgf returns the conditional expectation of the strength for a random node, $\mathbb{E}(Y_t|s_{t-1} = l)$.

Corollary 2.1. *Defining $\alpha = \alpha_i + \alpha_j$ for each i and each j, the expected nodal strength of the underlying network $\mathcal{G}_t$ can be expressed as*

$$\mathbb{E}(Y_t|s_{t-1} = l) = G'_l(x)|_{x=1} = \sum_{k=1}^{K} q_{lk} \left.\widetilde{G}'_k(x)\right|_{x=1} = (N-1)e^{\alpha} \sum_{k=1}^{K} q_{lk} \left(4\sigma_k^2\beta + 1\right)^{-\frac{d}{2}}.$$

Note that $\mathbb{E}(Y_t|s_{t-1} = l)$ turns out to be a weighted sum of the expected nodal strength obtained by conditioning on each possible state of the world. In this sense, the result in Rastelli et al. (2016) can be obtained as a special case imposing all but one of the conditional probabilities q_{lk} for $k \in \{1, \ldots, K\}$ to be zero.

The expected strength for each regime increases linearly with the number of nodes, N, and exponentially with the intercept parameter α.

Corollary 2.2. *The analytical expression of the variance of the strength distribution uses the first and the second factorial moment of the pgf, resulting in*

$$Var(Y_t|s_{t-1} = l) = \sum_{k=1}^{K} q_{lk} Var(Y_t|s_t = k) + \sum_{k=1}^{K} q_{lk} \left(\left.\widetilde{G}'_k(x)\right|_{x=1} - G'_l(x)|_{x=1}\right)^2,$$

8

where $Var(Y_t|s_t = k) = \widetilde{G}_k''(x)\Big|_{x=1} + \widetilde{G}_k'(x)\Big|_{x=1} - \widetilde{G}_k'^2(x)\Big|_{x=1}.$

Similarly, an analytical expression for the Dispersion Index can be obtained.

$$\mathfrak{D}(Y_t|s_{t-1} = l) = \sum_{k=1}^{K} q_{lk}\mathfrak{D}(Y_t|s_t = k) + v,$$

where

$$\mathfrak{D}(Y_t|s_t = k) = 1 + \frac{\widetilde{G}_k''(x)\Big|_{x=1}}{\widetilde{G}_k'(x)\Big|_{x=1}} - \widetilde{G}_k'(x)\Big|_{x=1}$$

$$v = \frac{\sum_{k=1}^{K} q_{lk}\, \widetilde{G}_k''(x)\Big|_{x=1}}{\sum_{k=1}^{K} q_{lk}\, \widetilde{G}_k'(x)\Big|_{x=1}} - \sum_{k=1}^{K} q_{lk}\frac{\widetilde{G}_k''(x)\Big|_{x=1}}{\widetilde{G}_k'(x)\Big|_{x=1}}.$$

The result for the second factorial moment differs from the results in Rastelli et al. (2016) as our derivation reflects the existing heterogeneity in weighted edges.

Figure 4 displays the contour plots of the expected value (Top Panels), standard deviation (Middle Panels), and dispersion index (Bottom Panels) of the Strength Distribution of a two-state MS Poisson LS. Labels "L" and "H" denote low and high polarization states. We consider $d = 1$ and different values of α, $\sigma_L^2\beta$ and p_L, $p_H = 1 - p_L$. As one can expect, due to the equal-dispersion property of the Poisson distribution, both the expected strength and the spread increase with the intercept value α. The lower $\sigma_L^2\beta$, the larger the similarity across the features of the nodes in that state and, in turn, the higher the expected strength. Expected strength also increases with the probability of state L.

3   Inference

3.1   Posterior sampling algorithm

In this section, we will go back to the setup in (1)-(7) and assume $d = 1$ in line with the particular application we focus on, so that $\mathbf{x}_{irt}$ and $\boldsymbol{\gamma}_1$ are scalars x_{irt} and γ_1. Let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_R)$ where $\mathbf{Y}_r = (\mathbf{Y}_{1r}, \ldots, \mathbf{Y}_{T_r r})$ be the collection of observed network weights with characteristic element Y_{ijrt}, $\boldsymbol{\ell} = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_R)$ where $\boldsymbol{\ell}_r = (\boldsymbol{\ell}_{1r}, \ldots, \boldsymbol{\ell}_{T_r r})$ are the observable political-leaning proxies with characteristic element ℓ_{irt} and $\mathbf{s}_r = (s_{r1}, \ldots, s_{rT_r})$ are the latent states for each country r. Consider the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_R)$ where $\boldsymbol{\theta}_r = (\boldsymbol{\alpha}_r, \boldsymbol{\zeta}_r, \boldsymbol{\sigma}_r^2, \gamma_{0r}, \gamma_{1r}, \phi_r, \boldsymbol{Q}_r)$, while β_r is fixed in our empirical implementation (see Section 3.2). Here $\boldsymbol{\zeta}_r = (\boldsymbol{\zeta}_{1r}, \ldots, \boldsymbol{\zeta}_{ir}, \ldots, \boldsymbol{\zeta}_{N_r r})$ denotes the latent parameters, where $\boldsymbol{\zeta}_{ir} = (\zeta_{i1r}, \ldots, \zeta_{iK_r r})$, a row vector with K_r elements, while $\boldsymbol{\sigma}_r^2 = (\sigma_{1r}^2, \ldots, \sigma_{K_r r}^2)$. The joint posterior $\pi(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\ell}) \propto f(\mathbf{Y}, \boldsymbol{\ell}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ is not tractable. Thus we follow a data augmentation approach and apply a Gibbs sampler for posterior inference (see Appendix C). Let us denote with $\boldsymbol{\xi}_r = (\boldsymbol{\xi}_{1r}, \ldots, \boldsymbol{\xi}_{T_r r})$ the collection of state indicator variables, where $\boldsymbol{\xi}_{rt} = (\xi_{1,rt}, \ldots, \xi_{K_r,rt})$ and $\xi_{k,rt} = \mathbb{I}(s_{rt} = k)$. Then, the complete-data likelihood function for
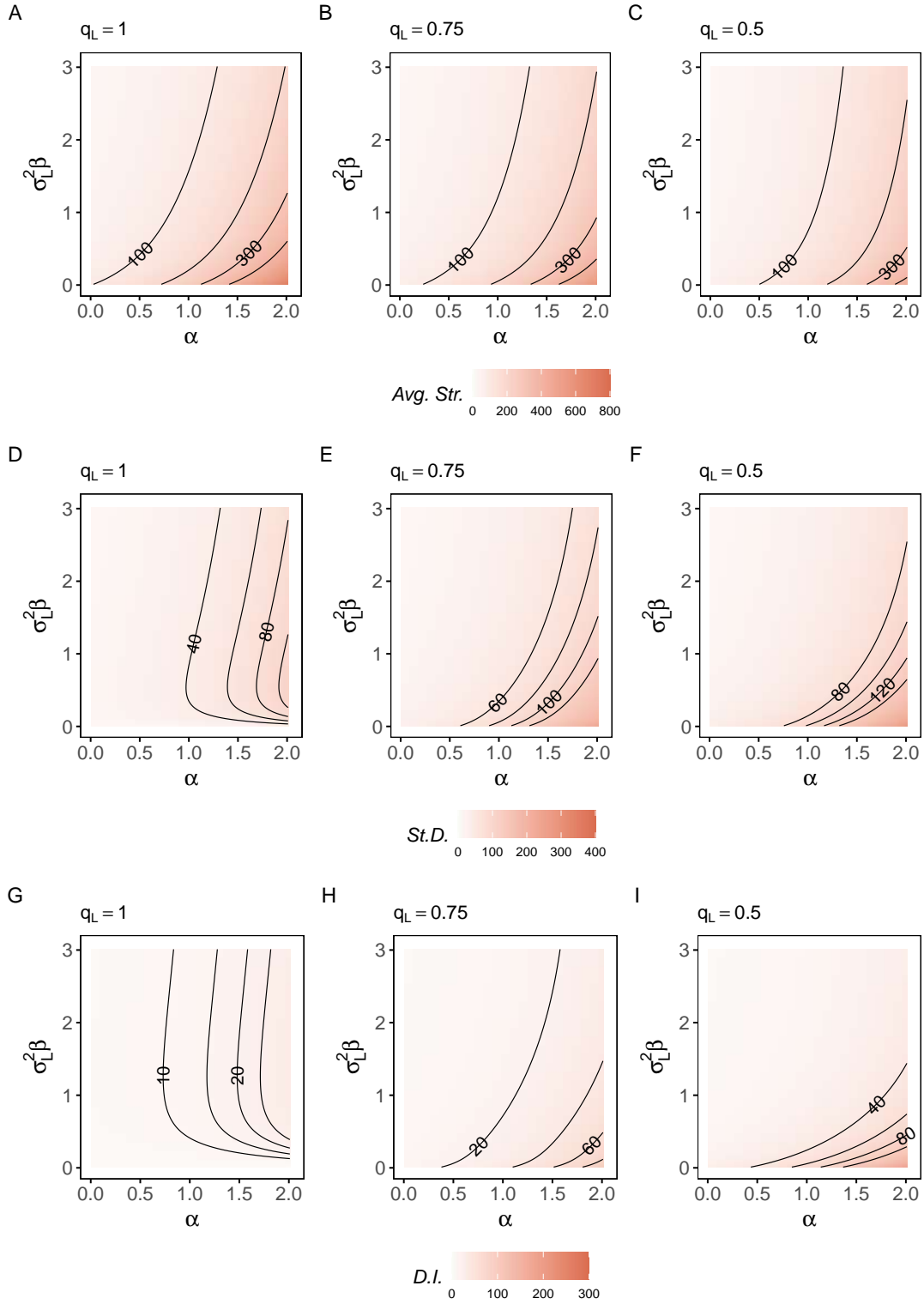
Figure 4: **Strength Distribution of an MS Poisson LS model:** Contour plots of the expected value (Top Panels), standard deviation (Middle Panels), and dispersion index (Bottom Panels) of the Strength Distribution of a two-state MS Poisson LS model for different values of α and $\sigma_L^2\beta$, and for different transition probabilities p_L and $p_H = 1 - p_L$. We assume $d = 1$, $N = 100$, and $\sigma_H^2\beta = 4$.

$r = 1, \ldots, R$ is the product of the following:

$$f(\mathbf{Y}_r, \boldsymbol{\ell}_r, \boldsymbol{\xi}_r | \boldsymbol{\theta}_r) = \prod_{t=1}^{T_r} \prod_{i=1}^{N_r} \left(f_B(\ell_{irt}|s_{rt}, \boldsymbol{\theta}_r) \prod_{j=i+1}^{N_r} f_P(Y_{ijrt}|s_{rt}, \boldsymbol{\theta}_r) \right) \prod_{l=1}^{K_r} \prod_{k=1}^{K_r} q_{lkr}^{\xi_{lr,t-1}\xi_{krt}}, \qquad (8)$$

where $f_P(Y_{ijrt}|s_{rt}, \boldsymbol{\theta}_r)$ is the Poisson pmf in (1) with dynamic intensity given in (2), $f_B(\ell_{irt}|s_{rt}, \boldsymbol{\theta}_r)$ the Beta pdf given in (3). With this notation, x_{irt} can be written as $\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt}$, in (1) and (3).

We approximate the joint posterior distribution by Markov-chain Monte Carlo (MCMC) sampling. Our Gibbs sampling algorithm iterates the following steps for each r:

1. Draw α_{ir} from $\pi(\alpha_r|\ldots)$, $i = 1, \ldots, N_r$ via Adaptive Metropolis-Hastings (MH);

2. Draw ϕ_r from $\pi(\phi_r|\ldots)$ via MH with truncated normal proposal;

3. Draw γ_{0r} and γ_{1r} from $\pi(\gamma_{0r}, \gamma_{1r}|\ldots)$ via MH;

4. Draw ζ_{ikr} from $\pi(\zeta_{ikr}|\ldots)$, $i = 1, \ldots, N_r$ and for $k = 1, \ldots, K_r$ via Adaptive MH;

5. Draw σ_{kr}^2 from $\pi(\sigma_{kr}^2|\boldsymbol{\zeta}_{kr})$ for $k = 1, \ldots, K_r$;

6. Draw $\boldsymbol{q}_{lr}$ from $\pi(\boldsymbol{q}_{lr}|\boldsymbol{\xi}_r)$ for $l = 1, \ldots, K_r$.

7. Draw $\mathbf{s}_r$ via the forward-filtering and backward-sampling algorithm (see Frühwirth-Schnatter, 2006).

Further details on the algorithmic design can be found in Appendix C.

3.2 Identifying Restrictions

The model presents well-known identification challenges. The first issue is related to the multiplication of the squared Euclidean distance $||\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt} - \boldsymbol{\zeta}_{jr}\boldsymbol{\xi}_{rt}||^2$ by the parameter β_r. As there is a clear scale indeterminacy between β_r and the variance of the latent variables in terms of σ_k^2, we choose to set $\beta_r = 1$. In addition, latent coordinates enter the parameter λ_{ijrt} only through the squared distance. This makes – in principle – positions that differ just by means of reflection, translation, and rotation equally likely (see Hoff et al., 2002 and Friel et al., 2016). Nonetheless, the introduction of (3) helps prevent the emergence of many equivalent latent-space representations. Still, translation and reflection along the y-coordinate remain possible. To overcome translation issues, we center the latent coordinates to the origin of the axes at each Gibbs-sampling iteration. To overcome reflection, we assume that the position of a single outlet is known in terms of left and right political leaning (e.g. $\zeta_{i^*kr} < 0$ for a left-leaning outlet i^* and for each state k), and we apply a reflection transformation to the latent leaning coordinates every time the latent leaning of i^* is in the wrong orthant, similarly to Barberá (2015). Finally, as pointed out in Frühwirth-Schnatter (2006), the joint posterior in Markov-Switching models is invariant with respect to a re-labeling of the hidden states. We tackle this issue by imposing an ordering restriction on the latent regimes across states. In particular, we label latent regimes in increasing order of median distance, $\tilde{D}_{rk} = med_{j>i}(D_{ijrk})$, where $D_{ijrk} = ||\zeta_{ikr} - \zeta_{jkr}||$.

3.3 Analysis of Simulated Data

We assess the performance of our inference method by running the algorithm on simulated data. Our simulation consists of 20 fictitious news outlets observed for 100 periods. Each period may belong to one of two polarization states $k \in \{L, H\}$, where State L is characterized by a lower average distance in the political-leaning dimension across news outlets, and State H has a higher average distance. So news outlets jointly undergo periods of high polarization and low polarization.

The latent leaning positions in State H, ζ_{iH}, are drawn from a normal distribution centered at -0.75 for $i \in \{1 : 10\}$ and in 0.75 for $i \in \{11 : 20\}$ with $\sigma_H = 0.15$, while the latent leaning positions in State L, ζ_{iL}, are centered around the positions -0.25 for $i \in \{1 : 10\}$ and 0.25 for $i \in \{11 : 20\}$ with $\sigma_L = 0.15$; the individual effect parameters α_i are randomly drawn from a normal distribution with $\mu_\alpha = 0$ and $\sigma_\alpha = 2$; the transition probability matrix is $Q = (q_{1.}, q_{2.})'$ where $q_{1.} = (0.95, 0.05)$ and $q_{2.} = (0.05, 0.95)$, and the sequence of states is randomly drawn from the Markov Chain initialized at State L; finally, we set $\phi = 200$, $\gamma_0 = -0.1$, and $\gamma_1 = 0.5$. We sample Y_{ijt} and ℓ_{it} from the data generating process (1)-(4). Our simulation represents a situation in which news outlets diverge in magnitude – via α_i – and in terms of political leaning – via ζ_{ik}.

We run our MCMC algorithm for 50,000 iterations and we discard the first 30,000 iterations as burn-in while applying thinning by a factor of 10 to reduce auto-correlation in the draws increasing the effective sample size. To correctly identify left and right-leaning, we consider the leaning position of news outlet 14 as known to be left.

Figure 5 reports a summary of the simulation results. From a comparison between the true values and the marginal posterior distributions, the model performs very well in terms of identification of the individual effects and latent variables (Panel A), the latent states (Panel B), as well as the other parameters in the simulation (Panel C). Credible regions for the pairs (α_i, ζ_{ik}) estimated with our model (red solid ellipses in Panel A) are narrower than those obtained disregarding the political-leaning proxy ℓ_{it}, i.e. dropping (3) from the model (dashed black ellipses). This suggests that the information-borrowing strategy is effective in improving the estimation accuracy of the latent variables. Properties of the MCMC chains are reported in the Supplementary Material (Appendix D). Our MCMC algorithm is implemented in R and $C++$ and we make the scripts freely available (see Appendix J.2).

4 Political Leaning and Polarization of News Outlets

We provide an application of our model to a dataset of daily Facebook activities related to 225 national and local news outlets in France, Germany, Italy and Spain. We provide both static and dynamic analyses which allow us to assess the media slant in local and national European news outlets as well as the polarization levels and regimes across countries.

4.1 Dataset Description and Construction

For our application, we construct and exploit a novel time-varying set of media networks, which we will call the *network dataset*. We build the networks from the *source dataset* collected by Schmidt et al. (2018) containing tick-by-tick information on the Facebook activity of national and local news outlets from four European countries (France, Germany,
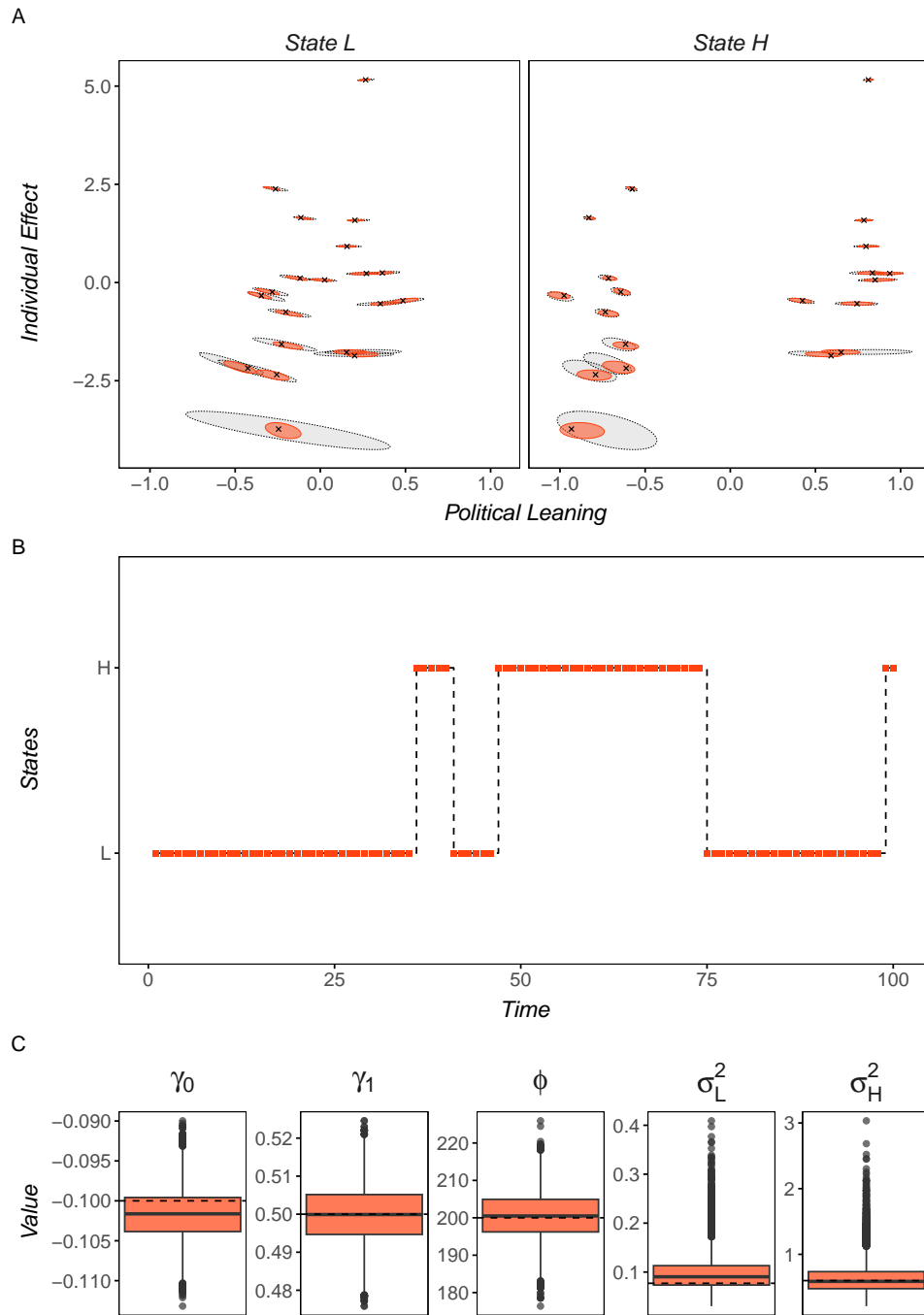
Figure 5: **Simulated data:** Panel A: Estimated Coordinates in the Latent Leaning - Individual Effect Plane with 99% credible ellipses for the mode with (3) (solid red), without (3) (dashed black) and true values (crosses). Panel B: estimated latent states (red dots) and true states (dashed lines). Panel C: boxplots of the marginal posteriors for the parameters γ_0, γ_1, ϕ, σ_L^2 and σ_H^2. The dashed lines indicate the true values.

Italy, and Spain) in a time-span entirely covering the years 2015 and 2016. We aggregate this activity to daily data. The news outlet list is reported in the Reuters Digital News Report (2017) (Newman et al., 2017). The source dataset contains all posts published by the news outlets in those years along with the associated metadata and also all the data on anonymized user interactions with these posts in the form of comments. Table 1 reports a summary description of the source dataset, which includes for each country the set of news outlets' pages, their posts and users' comments.

Country	Pages	Posts	Comments	Commenters
France	65	1,008,018	47,225,675	5,755,268
Germany	49	749,805	31,881,407	5,338,195
Italy	54	1,554,817	51,515,121	4,086,351
Spain	57	1,372,805	34,336,356	6,494,725

Table 1: **Source Dataset Description:** Description of the Facebook dataset on national and local news outlets of the four European countries (France, Germany, Italy and Spain) gathered by Schmidt et al. (2018). The dataset entirely covers the years 2015 and 2016. The news outlet list is the one reported in the Reuters Digital News Report (2017).

After constructing the set of daily bipartite networks of interaction between news outlets and *commenters* - those Facebook users commenting on news outlets' posts - for each country $r \in \{France, Germany, Italy, Spain\}$ at time t, we obtain the set of audience-duplication networks $\mathcal{G}_{tr}$ presented in Fig. 6 by performing the one-mode projection on the side of news outlets (as in Fig. 2).

We complement our network dataset with data from Crowdtangle (CrowdTangle Team, 2022) and Chapel Hill Survey Data (CHES) (Polk et al., 2017). Crowdtangle allows retrieving Facebook posts for public pages and provides additional metadata for each post. In particular, the fields *Link Text* and *Description* contain information on the text of linked pages, such as the texts of news articles published on the Facebook walls of the news outlets. There is not a perfect match between all the pages available in the source dataset of Schmidt et al. (2018) and those available in Crowdtangle. Some news outlets may have changed account or ceased to exist. In this case, information about these news outlets may no longer be available on Facebook at the time of writing. The CHES questions political scientists on different aspects related to politics and European integration. The *CHES dataset* contains all the information at an aggregate level about scientists' opinions on the ideological position of political parties in Europe. Here we will make use of the *lrgen* variable, which provides the ideological stance of a political party from 0 (extreme left) to 10 (extreme right). The information retrieved from Crowdtangle and CHES allows us to construct a text-analysis proxy for media slant. In particular, we obtain our observed proxy for daily media slant ℓ_{itr} by computing the index proposed by Gentzkow and Shapiro (2010) and adapted to online media outlets by Garz et al. (2020). Such a media slant index relies on text analysis techniques to assess the similarity between pieces by news outlets and texts published by politicians. We then associate a political leaning to each news outlet as a function of this similarity and the parties' political leaning. Further information on the adopted methodology can be found in Appendix F.

The network dataset and the media slant index are publicly available as described in
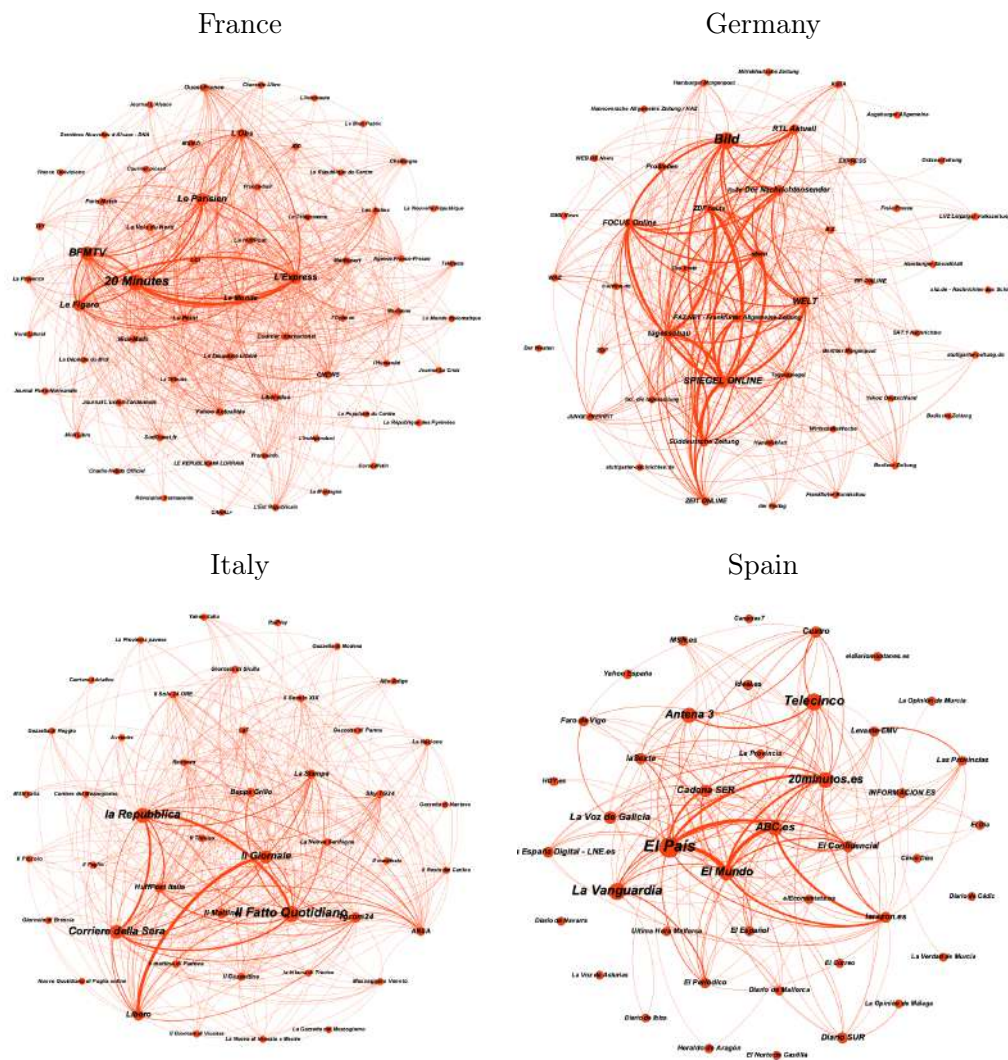
14

Appendix J.1.

France

Germany

Italy

Spain

Figure 6: **Audience-Duplication Networks:** Cumulative audience duplication network for France (top-left panel, 62 outlets), Germany (top-right panel, 47 outlets), Italy (bottom-left panel, 45 outlets) and Spain (bottom-right panel, 43 outlets) from January 1st 2015 to December 31st 2016. Node sizes are proportional to the cumulative number of comments received in the time interval, while edge thickness is proportional to the number of common commenters between each pair of outlets. Relevant quantities are normalized for each country.

4.2   Results from a Static Analysis

First, we implement a static version of our model on the whole 2-year time period, without the MS dynamic component. For this, we use an overall audience duplication network $\tilde{\mathcal{G}}_{tr}$, where the weighted edge for each pair of outlets is $\tilde{Y}_{ijr} = \sum_{t=1}^{T} Y_{ijrt}$, and the overall observable leaning-feature is constructed as $\tilde{\ell}_{ir} = T^{-1} \sum_{t=1}^{T} \ell_{itr}$. Panels A, B, C, and D in Fig. 7 report the estimated latent coordinates in the latent leaning-individual effect space.

We notice how the individual effect parameter α_{ir} associated with each news outlet and country may be interpreted in terms of news outlet's engagement, as major national news outlets are concentrated at the top in each graph.

We correlate our estimated (posterior mean) media slant with the results obtained by the PEW Research Survey (Mitchell et al., 2018). In this survey, participants were asked to assess the left-right leaning of national news outlets on a 0-6 scale with 0 indicating far left and 6 indicating far right. We will refer to the left-right ranking obtained by PEW Research as the PEW Research index. We find that the PEW Research index has a 0.73 correlation with our estimated latent leaning, see Panel E in Fig. 7. Moreover, we notice the presence of both a left-leaning cluster (bottom-left) and a right-leaning cluster (top-right).
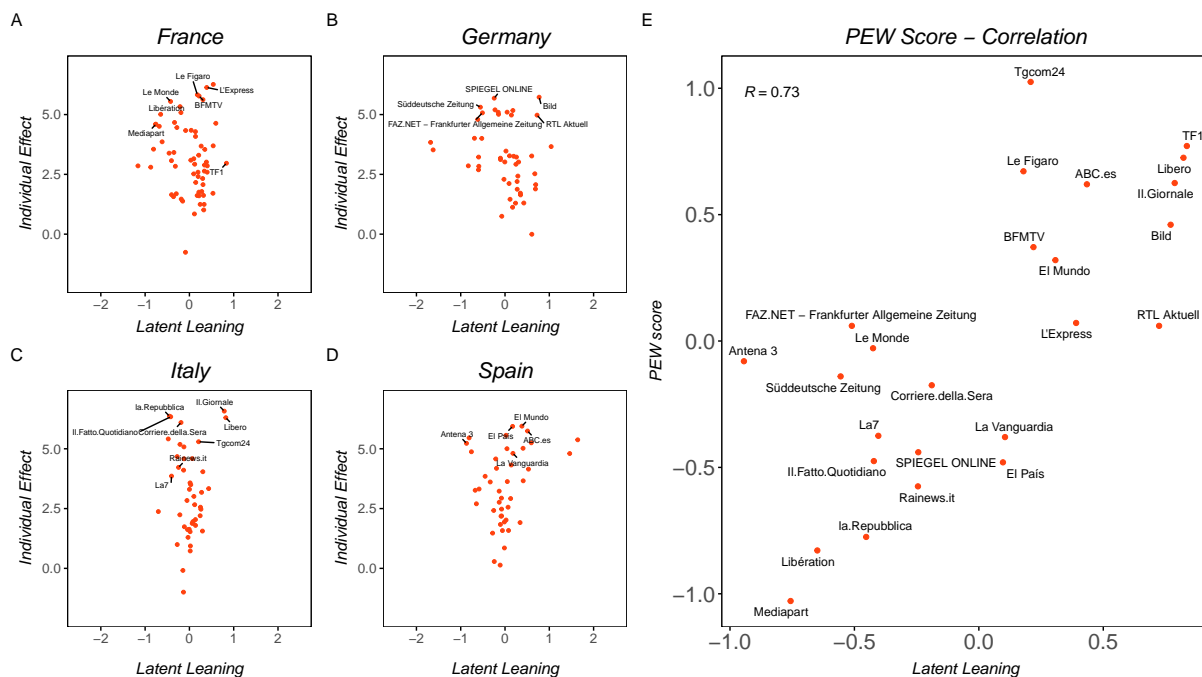

Figure 7: **PEW Index - Latent Leaning Static Comparison:** Panels A, B, C, and D present the posterior mean of the latent coordinates of our news outlets, while Panel E displays the scatter plot comparing the PEW survey results with our estimated latent leaning proxy (country-specific mean has been subtracted from PEW Score to improve readability).

Fig. I.1 in Appendix I presents the marginal posterior distribution of the parameters γ_{0r}, γ_{1r} and ϕ_r. The parameter γ_{1r} conveys information on the relationship between the latent variable x_{ir} and the observable leaning proxy ℓ_{ir}. Latent leaning appears to be a strong driver for the observable proxy only in the case of Italy, as the posterior mass of γ_{1r} is located far away from zero, while it seems a weak driver for France and mostly irrelevant for both Germany and Spain. Nonetheless, the strong correlation with the PEW Research index suggests that having information on online users' interactions with news outlets may still be sufficient to provide an effective classification on the political spectrum.

4.3 Results from a Dynamic Analysis

We now estimate the model described in Subsection 2.1 in its dynamic specification, i.e. including the MS specification of (4). The dynamic analysis uses daily data, and we deleted from our dataset those outlets that remained inactive – i.e. did not receive any comment – for more than 15 consecutive days. Overall, we removed 13 news outlets (DE: 4 outlets, FR: 5, IT: 2, SP: 2, see Appendix G), who displayed unusual behavior. Posterior results on the parameters are presented in Figure 8, where it is clear that the large number of observations leads to more precise inference than in the static case. Inference on γ_{1r} indicates a clear link between x_{ir} and ℓ_{ir} for both France and Italy. Also, as expected, values of σ_H^2 tend to be larger than those for σ_L^2. In Fig. 9, we report the posterior means of the latent positions for the four countries in both states. The individual-effect values are coherent with the engagement interpretation in both states: well-known national newspapers appear in the upper part of the graph, while local newspapers are most prevalent at the bottom. Moreover, our latent variable also positively correlates with the PEW Research Survey Index in this setting. Figure 10 illustrates the correlation of 0.68 in the lower polarisation state and 0.69 in the state of higher polarisation.
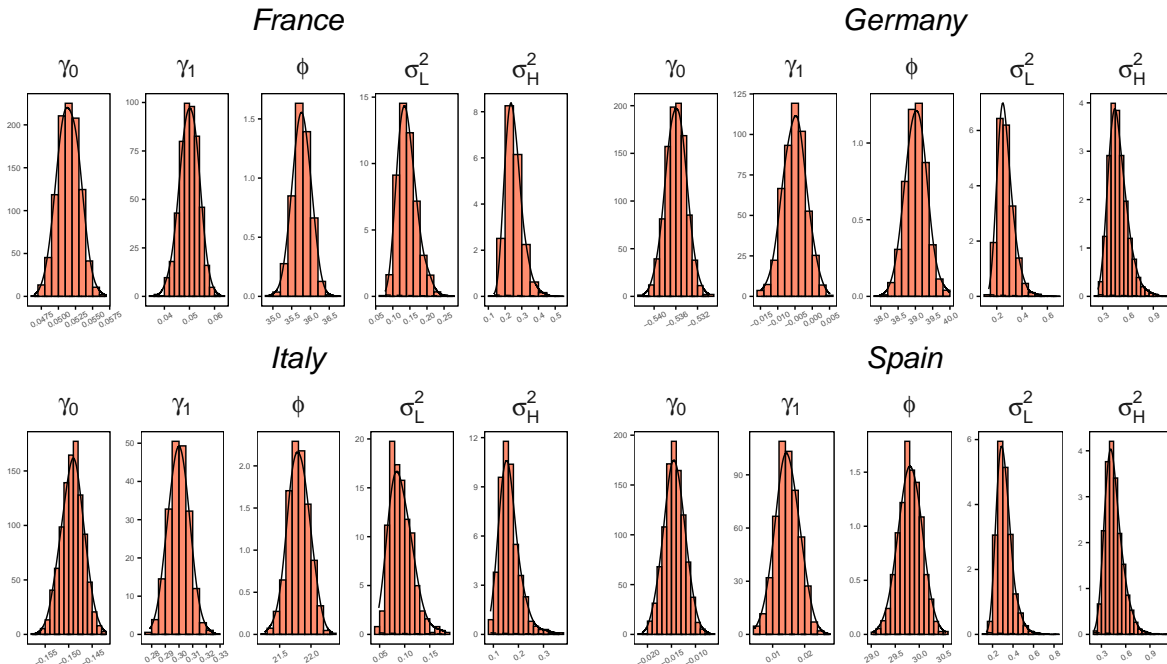

Figure 8: **Marginal Posteriors - Dynamic Analysis:** Histograms for the MCMC draws for the parameters γ_{0r}, γ_{1r}, ϕ_r, σ_{Lr}^2 and σ_{Hr}^2 for France, Germany, Italy and Spain with prior pdfs indicated by dashed lines (mostly indistinguishable from the horizontal axis).

Table 2 reports a posterior predictive check (see Gelman et al., 2014) in which the posterior predictive expected nodal strength and its variance and dispersion index are compared with the empirical values for the standard Poisson random graph model, the Poisson random graph model with individual effects and observed leaning distances and the MS-LS model. We notice how all models are able to mimic the first moment of the strength distribution, while the two more elaborate models are able to capture the observed
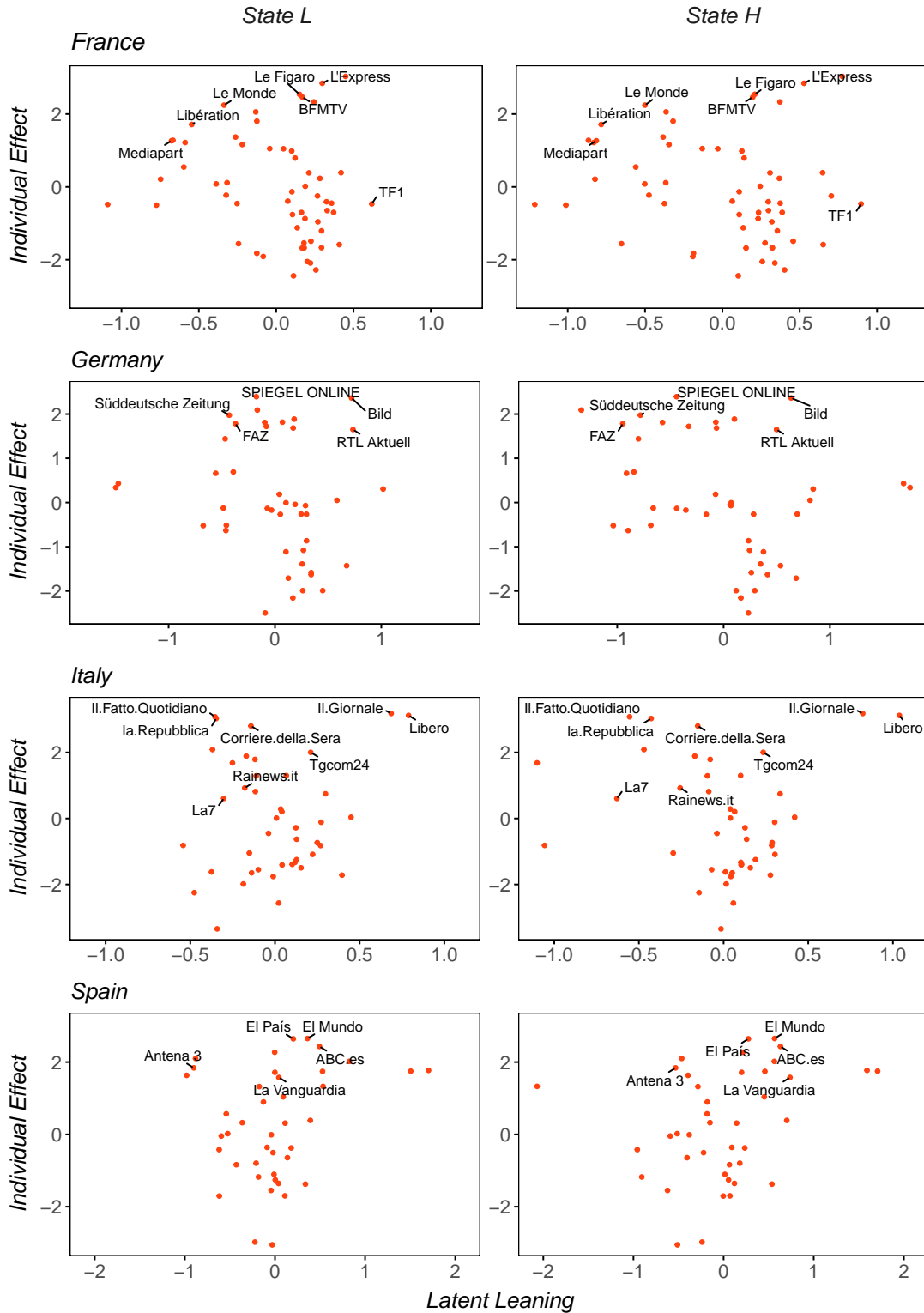
Figure 9: **Latent Positions - Dynamic Analysis:** Estimated latent coordinates of the news outlets for France (Panel A), Germany (Panel B), Italy (Panel C) and Spain (Panel D) in State L and in State H.
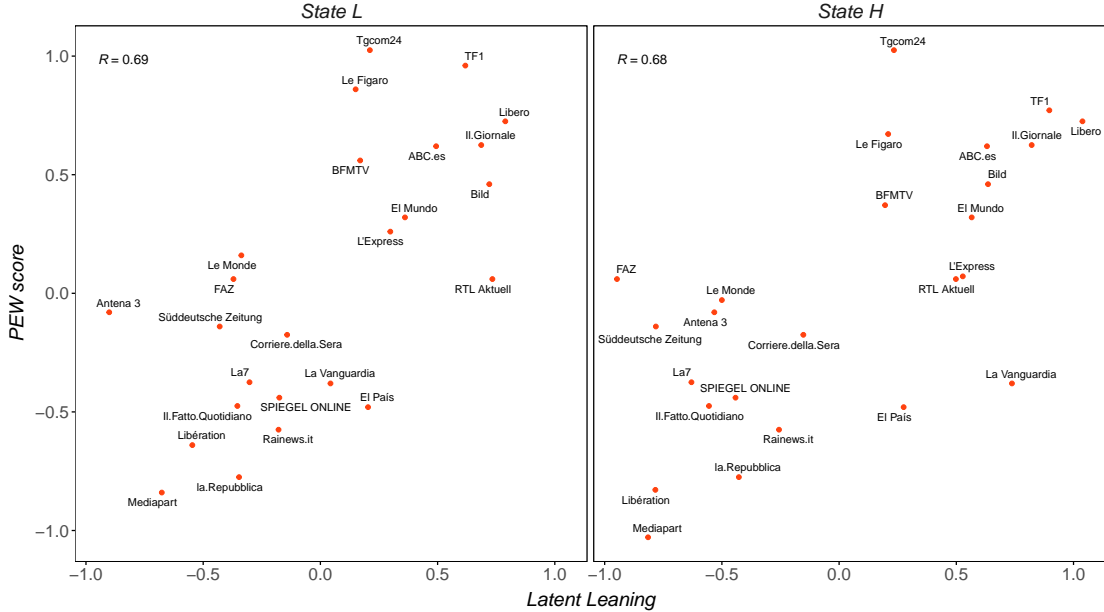
Figure 10: **PEW Index - Latent Leaning Dynamic Comparison:** Scatter plot comparing the PEW survey results against our latent leaning proxy in State L and H. The country-specific mean has been subtracted from the PEW Score to improve readability.

dispersion in the strength distribution (with the MS-LS model giving a slightly better fit).

Latent states signal the presence of lower or higher in-platform polarization regimes. In state L the average distance between outlets is lower in terms of political leaning than in state H, making Facebook users more prone to interact with news outlets with different political tendencies. Panels A-D of Fig. 11 report the estimated posterior probabilities for State H through time for the four countries. We notice a tendency to move from a state of high in-platform polarization to a low polarization for Germany and Spain. Italy instead moves in the opposite direction, while France does not show a clear pattern, rather it alternates between the two states.

Overall, our findings contradict the hypothesis of a common shift toward a high polarization regime on social media in this time frame. Panels E-H of Fig. 11 report the estimated transition probabilities. Spain is characterized by high persistence whereas France switches most frequently between the two states.

4.4 Model Selection

We perform model selection considering three alternative models: $\mathcal{M}_{1r}$ is the unrestricted model described in Subsection 2.1, $\mathcal{M}_{2r}$ omits the text-analysis interpretation in (3) (equivalent to imposing $\gamma_{1r} = 0$) for each country r, while the static model $\mathcal{M}_{3r}$ omits the Markov-switching dynamics described in (4). This comparison highlights the contributions of the observable leaning *a-là* Gentzkow et al. (2015) and the dynamic component.

Model selection is carried out via two popular predictive measures, the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) and the log pointwise predictive density (lppd) of Gelman et al. (2014) (see Appendix E for further details). Tables 3 and 4 report both criteria and indicate that the model without the dynamic component ($\mathcal{M}_{3r}$) is dom-

	Average Empirical Network Metrics			
	France	Germany	Italy	Spain
Expected Strength	280.83	161.82	380.98	239.32
S.D. Strength	450.53	209.39	600.58	340.76
Dispersion Index	814.20	306.11	1037.82	562.56
	R.G. $\lambda_{ijt} = \exp\{\alpha\}$			
Expected Strength	280.82	161.82	380.99	239.30
	(280.55, 281.10)	(161.56, 162.04)	(380.62, 381.37)	(239.00, 239.59)
S.D. Strength	16.53	12.49	19.17	15.18
	(16.43, 16.62)	(12.40, 12.57)	(19.04, 19.29)	(15.07, 15.28)
Dispersion Index	0.98	0.98	0.98	0.98
	(0.97, 0.99)	(0.96, 0.99)	(0.96, 0.99)	(0.96, 0.99)
	R.G. $\lambda_{ijt} = \exp\{\alpha_i + \alpha_j - \beta\|\ell_{it} - \ell_{jt}\|^2\}$			
Expected Strength	276.64	162.73	375.68	239.50
	(276.36, 276.92)	(162.47, 163.00)	(375.26, 376.09)	(239.19, 239.80)
S.D. Strength	431.33	201.97	577.62	312.12
	(430.84, 431.85)	(201.60, 202.30)	(576.89, 578.30)	(311.67, 312.54)
Dispersion Index	672.57	250.68	888.15	406.79
	(671.57, 673.75)	(250.10, 251.23)	(886.72, 889.50)	(405.96, 407.55)
	Dynamic MS-LS Model			
Expected Strength	280.84	161.81	380.99	239.32
	(280.60, 281.09)	(161.57, 162.03)	(380.64, 381.36)	(239.00, 239.64)
S.D. Strength	437.96	201.90	587.77	313.64
	(437.52, 438.44)	(201.59, 202.24)	(587.12, 588.43)	(313.16, 314.13)
Dispersion Index	683.13	251.94	906.83	411.12
	(682.10, 684.22)	(251.33, 252.54)	(905.48, 908.02)	(410.20, 412.04)

Table 2: **Network Posterior Predictive Check:** The table compares the network characteristics in terms of expected strength, strength's standard deviation, and dispersion index for the standard Poisson random graph model (second panel), the Poisson random graph model with individual effects and observed leaning distances as a covariate (third panel) and the dynamic latent space network model (bottom panel) with the observed network metrics averaged over time (top panel). The 95% credible intervals are in parenthesis.

inated by the other two specifications for each country r. Except for Spain and to some extent Italy, very similar scores are obtained for $\mathcal{M}_{1r}$ and $\mathcal{M}_{2r}$. This is to be expected as $\mathcal{M}_{1r}$ aims to offer more informative latent coordinates rather than an improved fit for the network. Table 4 also includes results for the random graph models, which are clearly performing worse than our models.
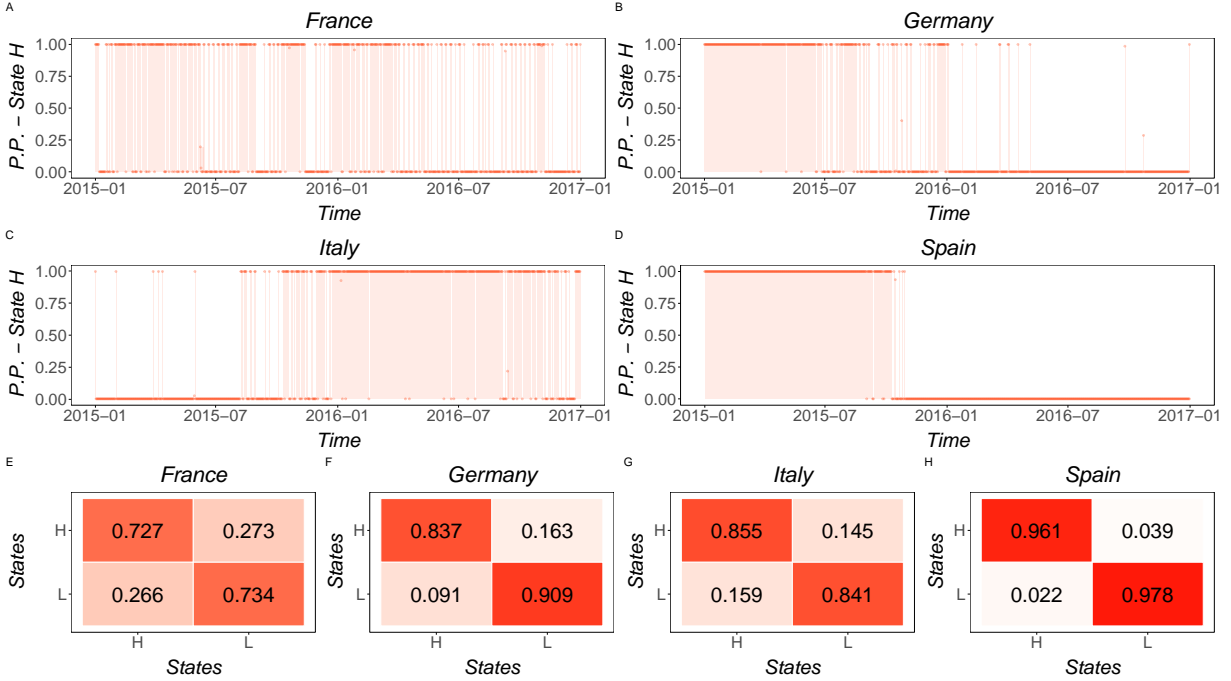
Figure 11: **Latent States Comparison:** Posterior Probabilities for State H through time for France, Germany, Italy and Spain (Panels A-D). Posterior mean transition matrices for France, Germany, Italy and Spain (Panels E-H) with red shades proportional to the magnitude of the transition probability.

	DIC$\times 10^{-6}$			
Model	France	Germany	Italy	Spain
$\mathcal{M}_1$	**4.4696**	**2.3669**	**3.3049**	**4.6139**
$\mathcal{M}_2$	4.4698	2.3669	3.3066	4.6390
$\mathcal{M}_3$	4.6434	2.4766	3.4825	4.9049

Table 3: **Model Selection:** DIC scores for the three nested models considered here.

	lppd $f_P \times 10^{-6}$			
Model	France	Germany	Italy	Spain
$\mathcal{M}_1$	$\mathbf{-2.2784}$	-1.2191	-1.6776	$\mathbf{-2.3347}$
$\mathcal{M}_2$	-2.2784	$\mathbf{-1.2190}$	$\mathbf{-1.6771}$	-2.3471
$\mathcal{M}_3$	-2.3597	-1.2702	-1.7657	-2.4511
R.G. $\lambda_{ijt} = \exp\{\alpha\}$	-12.5047	-4.8341	-13.1011	-7.6049
R.G. $\lambda_{ijt} = \exp\{\alpha_i + \alpha_j - \beta\|\ell_{it} - \ell_{jt}\|^2\}$	-2.6153	-1.4121	-2.2424	-2.7350

Table 4: **Model Selection:** Log pointwise predictive density (lppd) as in Gelman et al. (2014) for the models introduced here and two random graph models.

5 Conclusion

We propose a dynamic Markov-Switching Latent Space model through which insightful information can be extracted concerning media ideology and in-platform polarization regimes. The model projects the audience duplication network of news outlets on a one-dimensional Euclidean space where the latent positions can be interpreted in terms of political leaning through the use of a suitable proxy. Inference is carried out within a Bayesian framework which allows for reliable results with relatively standard MCMC methods. We derive the theoretical model properties and assess the efficacy of the proposed methodology on simulated data. Our model is applied to a Facebook dataset of news outlets in four European countries, covering the years 2015 and 2016. We carry out both a static and a dynamic analysis and in both settings we found that the inferred latent leaning variable strongly correlates with the independent PEW Research Survey Index and correctly clusters news outlets in terms of left and right leaning. Moreover, we find that inference on the latent states does not support the hypothesis of a unidirectional shift toward high polarization on Facebook. Finally, model selection suggests that the dynamic specification should be preferred to the static model and that a text-analysis index helps estimation for three of the four countries.

References

Anderson, S. P. and J. McLaren (2012). Media Mergers and Media Bias with Rational Consumers. *Journal of the European Economic Association 10*(4), 831–859.

Andrieu, C. and J. Thoms (2008). A Tutorial on Adaptive MCMC. *Statistics and Computing 18*(4), 343–373.

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation using Twitter Data. *Political Analysis 23*(1), 76–91.

Barrat, A., M. Barthelemy, R. Pastor-Satorras, and A. Vespignani (2004). The Architecture of Complex Weighted Networks. *Proceedings of the National Academy of Sciences 101*(11), 3747–3752.

Cinelli, M., G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini (2021). The Echo Chamber Effect on Social Media. *Proceedings of the National Academy of Sciences 118*(9), e2023301118.

CrowdTangle Team (2022). CrowdTangle. Menlo Park, CA: Meta.

Dandekar, P., A. Goel, and D. T. Lee (2013). Biased Assimilation, Homophily, and the Dynamics of Polarization. *Proceedings of the National Academy of Sciences 110*(15), 5791–5796.

Esteban, J.-M. and D. Ray (1994). On the Measurement of Polarization. *Econometrica 62*(4), 819–851.

Friel, N., R. Rastelli, J. Wyse, and A. E. Raftery (2016). Interlocking Directorates in Irish Companies using a Latent Space Model for Bipartite Networks. *Proceedings of the National Academy of Sciences 113*(24), 6629–6634.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Science & Business Media.

Garimella, K., G. D. F. Morales, A. Gionis, and M. Mathioudakis (2018). Quantifying Controversy on Social Media. *ACM Transactions on Social Computing 1*(1), 1–27.

Garz, M., J. Sörensen, and D. F. Stone (2020). Partisan Selective Engagement: Evidence from Facebook. *Journal of Economic Behavior & Organization 177*, 91–108.

Gelman, A., J. Hwang, and A. Vehtari (2014). Understanding Predictive Information Criteria for Bayesian Models. *Statistics and Computing 24*(6), 997–1016.

Gentzkow, M. and J. M. Shapiro (2010). What Drives Media Slant? Evidence from US Daily Newspapers. *Econometrica 78*(1), 35–71.

Gentzkow, M., J. M. Shapiro, and D. F. Stone (2015). Media Bias in the Marketplace: Theory. In *Handbook of Media Economics*, Volume 1, pp. 623–645. Elsevier.

Geweke, J. F. et al. (1991). Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments. Technical report, Federal Reserve Bank of Minneapolis.

Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007). Model-based Clustering for Social Networks. *Journal of the Royal Statistical Society: Series A 170*(2), 301–354.

Hanusch, F. and D. Nölleke (2019). Journalistic Homophily on Social Media: Exploring Journalists' Interactions with Each Other on Twitter. *Digital Journalism 7*(1), 22–44.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association 97*(460), 1090–1098.

Kim, B., K. H. Lee, L. Xue, and X. Niu (2018). A Review of Dynamic Network Models with Latent Variables. *Statistics Surveys 12*, 105.

Kubin, E. and C. von Sikorski (2021). The Role of (Social) Media in Political Polarization: A Systematic Review. *Annals of the International Communication Association 45*(3), 188–206.

Leonard, N. E., K. Lipsitz, A. Bizyaeva, A. Franci, and Y. Lelkes (2021). The Nonlinear Feedback Dynamics of Asymmetric Political Polarization. *Proceedings of the National Academy of Sciences 118*(50), e2102149118.

Macy, M. W., M. Ma, D. R. Tabin, J. Gao, and B. K. Szymanski (2021). Polarization and Tipping Points. *Proceedings of the National Academy of Sciences 118*(50), e2102144118.

Mitchell, A., K. Simmons, K. E. Matsa, L. Silver, E. Shearer, C. Johnson, M. Walker, and K. Taylor (2018). In Western Europe, Public Attitudes toward News Media more Divided by Populist Views than Left-Right Ideology. *PEW Research Center*.

Newman, N., R. Fletcher, D. Levy, and R. K. Nielsen (2017). Reuters Institute Digital News Report 2017.

Polk, J., J. Rovny, R. Bakker, E. Edwards, L. Hooghe, S. Jolly, J. Koedam, F. Kostelka, G. Marks, G. Schumacher, et al. (2017). Explaining the Salience of Anti-elitism and Reducing Political Corruption for Political Parties in Europe with the 2014 Chapel Hill Expert Survey Data. *Research & Politics 4*(1), 2053168016686915.

Prior, M. (2013). Media and Political Polarization. *Annual Review of Political Science 16*, 101–127.

Puglisi, R. and J. M. Snyder Jr (2015). Empirical Studies of Media Bias. In *Handbook of Media Economics*, Volume 1, pp. 647–667. Elsevier.

Rastelli, R., N. Friel, and A. E. Raftery (2016). Properties of Latent Variable Network Models. *Network Science 4*(4), 407–432.

Schmidt, A. L., F. Zollo, A. Scala, and W. Quattrociocchi (2018). Polarization Rank: A Study on European News Consumption on Facebook. arXiv preprint arXiv:1805.08030.

Sewell, D. K. and Y. Chen (2016). Latent Space Models for Dynamic Networks with Weighted Edges. *Social Networks 44*, 105–116.

Sosa, J. and B. Betancourt (2022). A Latent Space Model for Multilayer Network Data. *Computational Statistics & Data Analysis 169*, 107432.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B 64*(4), 583–639.

WEF (2022). The Global Risks Report 2022 17th edition. Technical report, World Economic Forum.

Yarchi, M., C. Baden, and N. Kligler-Vilenchik (2021). Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication 38*(1-2), 98–139.

A Proof of the Latent Space Model Properties

In this section, we provide proof of the propositions reported in Subsection 2.2 for what concerns the properties of a Latent Space model applied to a weighted multi-layer network $\mathcal{G}_r$. For ease of exposition and without loss of generality we drop the index r.

A.1 Relevant Results

We report here below some results that will turn out useful in the derivation of our results.

Proposition A.1. *Multinomial Theorem*

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{\underline{k} \in \mathcal{K}} \binom{n}{\underline{k}} \prod_{i=1}^{m} x^m, \tag{A.1}$$

where $\underline{k} = \{k_1, \ldots, k_m\}$, $\mathcal{K} = \{k_i \in \mathbb{N} | \sum_i k_i = \ell\}$.

Proposition A.2. *Integral of the product of N zero-mean independent MVNs*

$$\int_{\mathcal{X}} \prod_{i=1}^{N} f(x; \underline{0}, \sigma_i^2 I_d) dx = (2\pi)^{-\frac{(N-1)d}{2}} \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)^{-\frac{d}{2}} \prod_{i=1}^{N} (\sigma_i^2)^{-\frac{d}{2}}, \tag{A.2}$$

where $f(x; \underline{0}, \sigma_i^2 I_d)$ is the pdf of a d-dimensional multivariate normal distribution with mean $\underline{0}$ and variance $\sigma_i^2 I_d$ for each $i = 1, \ldots, N$.

The proofs are straightforward hence they are omitted.

Proposition A.3. *Convolution of Normal Distributions*

$$G_1 * G_2(z) = \int G_1(x) G_2(z - x) dx \tag{A.3}$$

$$= \int \frac{1}{(2\pi)^{p/2}|A|^{1/2}} e^{-\frac{1}{2}(x-a)'A^{-1}(x-a)} \frac{1}{(2\pi)^{p/2}|B|^{1/2}} e^{-\frac{1}{2}(z-x-b)'B^{-1}(z-x-b)} dx \tag{A.4}$$

$$= \int g_p(x; a, A) \cdot g_p(x; z - b, B) dx \tag{A.5}$$

$$= \frac{1}{(2\pi)^{p/2}|A + B|^{1/2}} e^{-\frac{1}{2}(z-(a+b))'(A+B)^{-1}(z-(a+b))} \tag{A.6}$$

$$= g_p(z; a + b, A + B). \tag{A.7}$$

A.2 Probability Generating Function

For a general LS model, following the conditional independence and HMM assumptions (Assumption 2.1 and 2.2) and from the law of iterated expectation, the probability gener-

ating function (pgf) for the weighted degree can be written as:

$$G_l(x) = \sum_{v=0}^{\infty} x^v p_v = \sum_{k=1}^{K} \widetilde{G}_k(x) q_{lk}, \tag{A.8}$$

where

$$\widetilde{G}_k(x) = \sum_{v=0}^{\infty} x^v \int_{\mathcal{Z}_k^N} \mathbb{P}\left(S_{it} = v \mid s_t = k, \underline{\zeta}_k\right) \left(\prod_{j=1}^{N} f\left(\zeta_{jk}\right)\right) d\zeta_{1k} \cdots d\zeta_{Nk} \tag{A.9}$$

$$= \int_{\mathcal{Z}_k^N} \left(\sum_{v=0}^{\infty} x^v \mathbb{P}\left(S_{it} = v \mid s_t = k, \underline{\zeta}_k\right)\right) \left(\prod_{j=1}^{N} f\left(\zeta_{jk}\right)\right) d\zeta_{1k} \cdots d\zeta_{Nk} \tag{A.10}$$

$$= \int_{\mathcal{Z}_k} f(\zeta_{ik}) \left(\int_{\mathcal{Z}_k^{(N-1)}} \left(\prod_{j\neq i} \varphi_{ijk}\left(x; \zeta_{ik}, \zeta_{jk}\right)\right) \left(\prod_{j\neq i} f\left(\zeta_{jk}\right)\right) d\zeta_{jk}\right) d\zeta_{ik} \tag{A.11}$$

$$= \int_{\mathcal{Z}_k} f\left(\zeta_{ik}\right) \left(\prod_{j\neq i} \int_{\mathcal{Z}_k^{(N-1)}} \varphi_{ijk}\left(x; \zeta_{ik}, \zeta_{jk}\right) f(\zeta_{jk}) d\zeta_{jk}\right) d\zeta_{ik} \tag{A.12}$$

$$= \int_{\mathcal{Z}_k} f\left(\zeta_{ik}\right) \prod_{j\neq i} \theta_{ik}\left(x; \zeta_{ik}\right) d\zeta_{ik}, \tag{A.13}$$

with latent coordinates $\underline{\zeta}_k = \{\zeta_{1k}, \ldots, \zeta_{Nk}\}$, nodal strength $Y_{it} = Y_{i1t} + \ldots + Y_{ij-1t} + Y_{ij+1t} + \ldots + Y_{iNt}$, strength's pgf $\varphi_{ijk}(x; \zeta_{ik}, \zeta_{jk}) = \mathbb{E}\left(x^{Y_{ijt}} | s_t = k, \zeta_{ik}, \zeta_{jk}\right) = e^{\lambda_{ijk}(\zeta_{ik}, \zeta_{jk})(x-1)}$. The $\theta_{ik}(x; \zeta_{ik}) = \int_{\mathcal{Z}_k} \varphi_{ijk}(x; \zeta_{ik}, \zeta_{jk}) f(\zeta_{jk}) d\zeta_{jk}$ is the pgf of the weight of the edge between a node chosen at random and a node with latent information ζ_{ik}.

A.3   Derivatives of an LS model with individual effects

Consider an LS model with Poisson likelihood for the edges $Y_{ijt} \sim \mathcal{P}oi\left(\lambda_{ijk}\right)$ with intensity parameter $\lambda_{ijk} = \exp\left\{\alpha_i + \alpha_j - \beta||\zeta_{ik} - \zeta_{jk}||^2\right\}$ (Assumption 2.3). From the independence assumption in Assumption 2.1, and normal assumption for the latent features (Assumption 2.4), the m-th derivative of the corresponding pgf can be written as:

$$\left.\frac{\partial^m}{\partial x^m} \prod_{j\neq i} \theta_{ik}(x, \zeta_{ik})\right|_{x=1} = \left.\frac{\partial^m}{\partial x^m} \int_{\mathcal{Z}_k^{N-1}} \prod_{j\neq i} e^{(x-1)\lambda_{ijk}} f(\zeta_{jk}) d\zeta_{jk}\right|_{x=1} \tag{A.14}$$

$$= \left.\int_{\mathcal{Z}_k^{N-1}} \frac{\partial^m}{\partial x^m} e^{(x-1)\sum_{j\neq i} \lambda_{ijk}} \prod_{j\neq i} f(\zeta_{jk}) d\zeta_{jk}\right|_{x=1} \tag{A.15}$$

$$= \left.\int_{\mathcal{Z}_k^{N-1}} \left(\sum_{j\neq i} \lambda_{ijk}\right)^m e^{(x-1)\sum_{j\neq i} \lambda_{ijk}} \prod_{j\neq i} f(\zeta_{jk}) d\zeta_{jk}\right|_{x=1} \tag{A.16}$$

$$= \int_{\mathcal{Z}_k^{N-1}} \left(\sum_{j\neq i} \lambda_{ijk}\right)^m \prod_{j\neq i} f(\zeta_{jk}) d\zeta_{jk} \tag{A.17}$$

$$= \int_{\mathcal{Z}_k^{N-1}} \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} \prod_{j \neq i} \lambda_{ijh}^{h_j} f(\zeta_{jk}) d\zeta_{jk} \tag{A.18}$$

$$= \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} \prod_{j \neq i} \int_{\mathcal{Z}} \lambda_{ijk}^{h_j} f(\zeta_{jk}) d\zeta_{jk} \tag{A.19}$$

$$= \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} \prod_{j \in \mathcal{N}^*} e^{(\alpha_i+\alpha_j)h_j} \int_{\mathcal{Z}_h} e^{-\beta ||\zeta_{ik}-\zeta_{jk}||^2 h_j} f(\zeta_{jk}) d\zeta_{jk} \tag{A.20}$$

$$= \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} \prod_{j \in \mathcal{N}^*} e^{(\alpha_i+\alpha_j)h_j} \int_{\mathcal{Z}_k} e^{-\beta ||\zeta_{ik}-\zeta_{jk}||^2 k_j} f(\zeta_{jk}) \frac{(2\pi)^{d/2}}{(2\pi)^{d/2}} \frac{(1/2\beta h_j)^{d/2}}{(1/2\beta h_j)^{d/2}} d\zeta_{jk} \tag{A.21}$$

$$= \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} \prod_{j \in \mathcal{N}_i^*} (1/2\beta h_j)^{\frac{d}{2}} (2\pi)^{\frac{d}{2}} e^{(\alpha_i+\alpha_j)h_j} f\left(\zeta_{ik}; \underline{0}\left(\sigma_k^2 + \frac{1}{2\beta h_j}\right) I_d\right) \tag{A.22}$$

where $\underline{h}_i = \{h_1, \ldots, h_{i-1}, h_{i+1}, \ldots, h_N\}$, $\mathcal{H}_i = \{h_j \in \mathbb{N}_i \quad j \neq i | \sum_{j \neq i} h_j = m\}$, $\mathcal{N}_i^* = \{j | j \neq i, h_j > 0\}$, $\beta > 0$, and where the fifth equation follows from Proposition A.1 and the last equation from Proposition A.3. Thus we obtain the following:

$$\left. \frac{\partial^m \widetilde{G}_k(x)}{\partial x^m} \right|_{x=1} = \int_{\mathcal{Z}_k} f(\zeta_{ik}) \left(\sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} \prod_{j \in \mathcal{N}_i^*} (1/2\beta h_j)^{\frac{d}{2}} (2\pi)^{\frac{d}{2}} e^{(\alpha_i+\alpha_j)h_j} f\left(\zeta_{ik}; \underline{0}, \tau_{kj} I_d\right) \right) d\zeta_{ik}, \tag{A.23}$$

where $\tau_{kj} = \left(\sigma_k^2 + \frac{1}{2\beta h_j}\right)$. Set $\alpha_i + \alpha_j = \alpha$ for each i and j, then from Proposition A.2:

$$\left. \frac{\partial^m \widetilde{G}_k(x)}{\partial x^m} \right|_{x=1} = \int_{\mathcal{Z}_k} f(\zeta_{ik}) \left(\sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} \prod_{j \in \mathcal{N}^*} (1/2\beta h_j)^{\frac{d}{2}} (2\pi)^{\frac{d}{2}} e^{\alpha h_j} f\left(\zeta_{ik}; \underline{0}, \tau_{kj} I_d\right) \right) d\zeta_{ik} \tag{A.24}$$

$$= \left(\sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} e^{\alpha m} \prod_{j \in \mathcal{N}_i^*} (1/2\beta h_j)^{\frac{d}{2}} (2\pi)^{\frac{d}{2}} \int_{\mathcal{Z}_k} f(\zeta_{ik}) \prod_{j \in \mathcal{N}^*} f\left(\zeta_{ik}; \underline{0}, \tau_{kj} I_d\right) d\zeta_{ik} \right) \tag{A.25}$$

$$= \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} e^{\alpha m} (\sigma_k^2)^{-\frac{d}{2}} \left(\frac{1}{\sigma_k^2} + \sum_{j \in \mathcal{N}^*} \tau_{kj}^{-1} \right)^{-\frac{d}{2}} \prod_{j \in \mathcal{N}^*} \left((1/2\beta h_j)^{\frac{d}{2}} (2\pi)^{\frac{d}{2}} (2\pi)^{\frac{-d}{2}} (\tau_{kj})^{-\frac{d}{2}} \right) \tag{A.26}$$

$$= \sum_{\underline{h}_i \in \mathcal{H}_i} \binom{m}{\underline{h}_i} e^{\alpha m} (\sigma_k^2)^{-\frac{d}{2}} \left(\frac{1}{\sigma_k^2} + \sum_{j \in \mathcal{N}_i^*} \tau_{kj}^{-1} \right)^{-\frac{d}{2}} \prod_{j \in \mathcal{N}^*} \left((1/2\beta h_j)^{\frac{d}{2}} \tau_{kj}^{-\frac{d}{2}} \right) \tag{A.27}$$

A.3.1 First Conditional Factorial Moment

If we solve for $m = 1$, we obtain:

$$\left. \frac{\partial \widetilde{G}_k(x)}{\partial x} \right|_{x=1} = e^{\alpha} (\pi)^{\frac{d}{2}} \left(\sum_{\underline{h}_i \in \mathcal{H}_i} \int_{\mathcal{Z}_k} f(\zeta_{ik}) f\left(\zeta_{ik}; \underline{0}\left(\sigma_k^2 + \frac{1}{2\beta}\right) I_d\right) d\zeta_{ik} \right) \tag{A.28}$$

$$= e^{\alpha} (\pi)^{\frac{d}{2}} (N-1) \left(2\sigma_k^2 + \frac{1}{2\beta} \right)^{-\frac{d}{2}} (2\pi\beta)^{-\frac{d}{2}} \tag{A.29}$$

27

$$= (N-1)e^{\alpha} \left(4\sigma_k^2 \beta + 1 \right)^{-\frac{d}{2}} \tag{A.30}$$

A.3.2 Second Conditional Factorial Moment

If we solve for m we obtain:

$$\widetilde{G}_k''(x)\Big|_{x=1} = (N-1)e^{2\alpha} \left(\sigma_k^2 + \frac{1}{4\beta} \right)^{-\frac{d}{2}} \left(\frac{1}{\sigma_k^2 + \frac{1}{4\beta}} + \frac{1}{\sigma_k^2} \right)^{-\frac{d}{2}} \left(\sigma_k^2 \right)^{-\frac{d}{2}} \left(4\beta \right)^{\frac{d}{2}} \tag{A.31}$$

$$+ (2\beta)^{-d}(N-1)(N-2)e^{2\alpha} \left(\sigma_k^2 + \frac{1}{2\beta} \right)^{-d} \left(\frac{2}{\sigma_k^2 + \frac{1}{2\beta}} + \frac{1}{\sigma_k^2} \right)^{-\frac{d}{2}} \left(\sigma_k^2 \right)^{-\frac{d}{2}} \tag{A.32}$$

$$= e^{2\alpha}(N-1) \left(8\sigma_k^2 \beta + 1 \right)^{-\frac{d}{2}} + (N-1)(N-2)e^{2\alpha} \left(2\sigma_k^2 \beta + 1 \right)^{-\frac{d}{2}} \left(6\sigma_k^2 \beta + 1 \right)^{-\frac{d}{2}} \tag{A.33}$$

A.3.3 First Moment

Employing the result in A.3.1 along with the linearity property of differentiation, we can write the expected strength for our MS-LS model as:

$$\mathbb{E}(Y_t|s_{t-1} = l) = G_l'(x)\big|_{x=1} = \sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'(x)\Big|_{x=1} = (N-1)e^{\alpha} \sum_{k=1}^{K} q_{lk} \left(4\sigma_k^2 \beta + 1 \right)^{-\frac{d}{2}} \tag{A.34}$$

A.3.4 Second Central Moment

We notice that in general for a given discrete random variable X with pgf $G(x)$, the variance can be computed as:

$$Var(X) = G''(x)\big|_{x=1} + G'(x)\big|_{x=1} - \left(G'(x)\big|_{x=1} \right)^2 \tag{A.35}$$

We thus obtain the following expression for the variance of the strength distribution:

$$\mathbb{V}ar(Y_t|s_{t-1} = l) = \sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k''(x)\Big|_{x=1} + \sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'(x)\Big|_{x=1} + \left(\sum_{k \in \{1,\dots,K\}} q_{lk}\widetilde{G}_k'(x)\big|_{x=1} \right)^2 \tag{A.36}$$

$$= \sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k''(x)\Big|_{x=1} + \sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'(x)\Big|_{x=1} + \left(\sum_{k=1}^{K} q_{lk}\widetilde{G}_k'(x)\big|_{x=1} \right)^2 \tag{A.37}$$

$$+ \sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'^2(x)\Big|_{x=1} - \sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'^2(x)\Big|_{x=1} \tag{A.38}$$

$$= \sum_{k=1}^{K} q_{lk} \left(\widetilde{G}_k''(x)\Big|_{x=1} + \widetilde{G}_k'(x)\Big|_{x=1} - \widetilde{G}_k'^2(x)\Big|_{x=1} \right) \tag{A.39}$$

$$+ \sum_{k=1}^{K} q_{lk} \left(\widetilde{G}_k'(x)\Big|_{x=1} - G_l'(x)\big|_{x=1} \right)^2 \tag{A.40}$$

$$= \sum_{k=1}^{K} q_{lk} Var(Y_t|s_t = k) + \sum_{k=1}^{K} q_{lk} \left(\widetilde{G}_k'(x)\Big|_{x=1} - G_l'(x)\big|_{x=1} \right)^2 \tag{A.41}$$

A.3.5 Dispersion Index

We provide a formula for the Dispersion Index ($\mathfrak{D}$) similar to the one suggested by Rastelli et al. (2016):

$$\mathfrak{D} = 1 + \frac{G''(x)|_{x=1}}{G'(x)|_{x=1}} - G'(x)|_{x=1} \tag{A.42}$$

For our model, the dispersion index is the following:

$$\mathfrak{D}(Y_t|s_{t-1} = l) = 1 + \frac{\sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k''(x)\Big|_{x=1}}{\sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'(x)\Big|_{x=1}} - \sum_{k=1}^{K} q_{lk}\widetilde{G}_k'(x)|_{x=1} \tag{A.43}$$

$$= 1 + \frac{\sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k''(x)\Big|_{x=1}}{\sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'(x)\Big|_{x=1}} - \sum_{k=1}^{K} q_{lk}\widetilde{G}_k'(x)|_{x=1} + \sum_{k=1}^{K} q_{lk} \frac{\widetilde{G}_k''(x)\Big|_{x=1}}{\widetilde{G}_k'(x)\Big|_{x=1}} - \sum_{k=1}^{K} q_{lk} \frac{\widetilde{G}_k''(x)\Big|_{x=1}}{\widetilde{G}_k'(x)\Big|_{x=1}} \tag{A.44}$$

$$= \sum_{k=1}^{K} q_{lk} \left(1 + \frac{\widetilde{G}_k''(x)\Big|_{x=1}}{\widetilde{G}_k'(x)\Big|_{x=1}} - \widetilde{G}_k'(x)\Big|_{x=1} \right) + v \tag{A.45}$$

$$= \sum_{k=1}^{K} q_{lk}\mathfrak{D}_k + v, \tag{A.46}$$

where $v = \frac{\sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k''(x)\Big|_{x=1}}{\sum_{k=1}^{K} q_{lk} \, \widetilde{G}_k'(x)\Big|_{x=1}} - \sum_{k=1}^{K} q_{lk} \frac{\widetilde{G}_k''(x)\Big|_{x=1}}{\widetilde{G}_k'(x)\Big|_{x=1}}.$

B   DGP Simulations

As a robustness check, we study the statistical properties of a weighted network generated by the LS model in simulation. We draw the latent coordinates $x_i \sim \mathcal{N}(0, \sigma^2 I_d)$ for $i = 1, \ldots, 100$, $d = 1$ and $K = 1$ and generate the networks weights $Y_{ij} \sim \mathcal{P}oi(\lambda_{ij})$ with intensity parameter $\log \lambda_{ij} = \alpha - \beta ||x_i - x_j||^2$.

As is common in network theory, we define the strength of node i as $S_i = \sum_{j \neq i} Y_{ij}$. We proceed studying the properties of the empirical distribution of $\mathbf{S} = \{S_1, \ldots, S_N\}$. In particular we cast out attention on the following statistics: the sample mean $\overline{S}$, the sample standard deviation $SD(S)$, the divergence index $\overline{\mathfrak{D}}(S)$ and the clustering coefficient CC as defined in Barrat et al. (2004).

Figures B.1 and B.2 report the sensitivity analysis for the aforementioned statistics as α and σ^2 vary. In this simulation we assume $\beta = 1$. Notice how the theoretical quantities obtained in Appendix A (dashed lines) match with empirical quantities (solid lines).
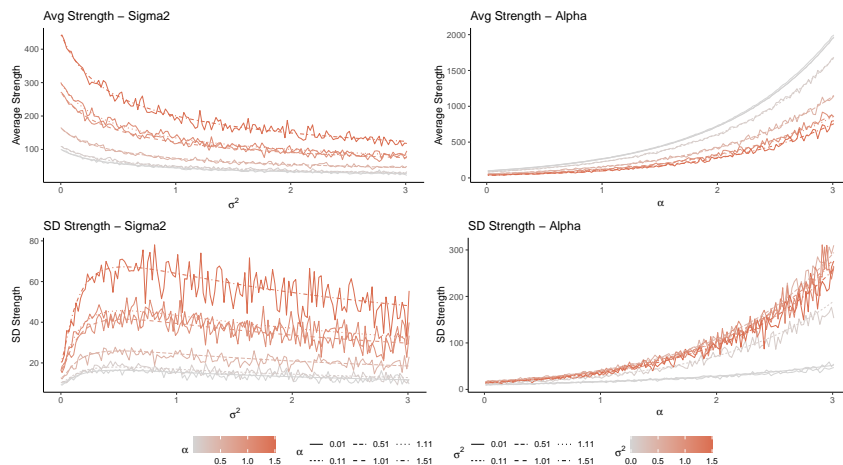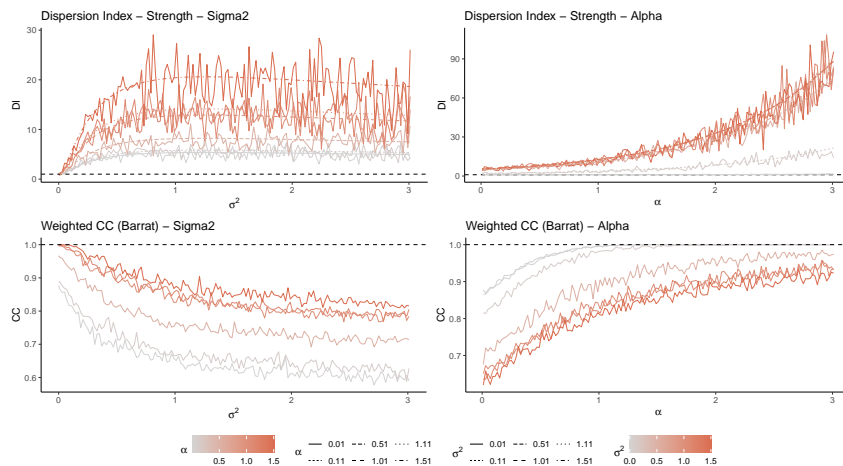

Figure B.1: **Simulation Result - 1**

Figure B.2: **Simulation Result - 2**

C Details of the MCMC Sampler

The joint likelihood function $f(\mathbf{Y}, \boldsymbol{\ell}|\boldsymbol{\theta})$ is written assuming cross-country independence as the product of the country-specific likelihoods for $r \in 1, \ldots, R$:

$$f(\mathbf{Y}_r, \boldsymbol{\ell}_r|\boldsymbol{\theta}_r) = \sum_{\mathbf{s}_r \in \{1,\ldots,K_r\}^{T_r}} \prod_{t=1}^{T_r} \prod_{i=1}^{N_r} \left(f_B(\ell_{irt}|s_{rt}, \boldsymbol{\theta}_r) \prod_{j=i+1}^{N_r} f_P(Y_{ijrt}|s_{tr}, \boldsymbol{\theta}_r) \right) h(s_{rt}|s_{r,t-1}),$$

(C.1)

where $f_P(Y_{ijrt}|s_{rt})$ is the Poisson distribution with dynamic intensity given in Eq. 2, $f_B(\ell_{irt}|s_{rt})$ the Beta distribution given in Eq. 3 and

$$h(s_{rt}|s_{r,t-1}) = \prod_{l=1}^{K_r} \prod_{k=1}^{K_r} p_{lkr}^{\mathbb{I}(s_{r,t-1}=l)\mathbb{I}(s_{rt}=k)},$$

(C.2)

the hidden Markov chain transition distribution. The joint posterior distribution is not tractable thus a data augmentation has been followed. In addition, given the cross-country independence assumption for both the likelihood and the joint prior, the Gibbs sampler iterates independent chains over the R countries. We report below the derivation of the full conditional distributions for the components of $\boldsymbol{\theta}_r$ and briefly discuss the sampling method.

C.1 Full conditional distribution of α_{ir}

$$\pi(\alpha_{ir}|\ldots) \propto \pi(\alpha_{ir}|\mu_\alpha, \sigma_\alpha^2) f(\mathbf{Y}_r, \boldsymbol{\ell}_r|\boldsymbol{\theta}_r) \propto \mathcal{N}(\alpha_{ir}; \mu_\alpha, \sigma_\alpha^2) \prod_{t=1}^{T_r} \prod_{i=1}^{N_r} \prod_{j=i+1}^{N_r} f_P(Y_{ijrt}|\lambda_{ijrt})$$

$$\propto \mathcal{N}(\alpha_r; \mu_\alpha, \sigma_\alpha^2 I) \prod_{t=1}^{T_r} \prod_{i=1}^{N_r} \prod_{j=i+1}^{N_r} \lambda_{ijrt}^{Y_{ijrt}} e^{-\lambda_{ijrt}},$$

where $\log \lambda_{ijrt} = \alpha_{ir} + \alpha_{jr} - ||\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt} - \boldsymbol{\zeta}_{jr}\boldsymbol{\xi}_{rt}||^2$. We sample from $\pi(\alpha_{ir}|\ldots)$ via Adaptive RW-MH.

C.2 Full conditional distribution of ϕ_r

$$\pi(\phi_r|\ldots) \propto \pi(\phi_r) f(\mathbf{Y}_r, \boldsymbol{\ell}_r|\boldsymbol{\theta}_r)$$

$$\propto \phi_r^{a_\phi-1} e^{-b_\phi \phi_r} \prod_{t=1}^{T_r} \prod_{i=1}^{N_r} \frac{\ell_{irt}^{a_{it}-1}(1-\ell_{irt})^{b_{it}-1}}{B(a_{it}, b_{it})},$$

where $a_{irt} = \varphi(\gamma_{0r} + \gamma_{1r}\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt})\phi_r$, $b_{irt} = (1 - \varphi(\gamma_{0r} + \gamma_{1r}\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt}))\phi_r$. We sample from $\pi(\phi_r|\ldots)$ via RW-MH with a truncated Gaussian proposal.

C.3 Full conditional distribution of γ_{0r} and γ_{1r}

$$\pi(\gamma_{0r}, \gamma_{1r}|\ldots) \propto \pi(\gamma_{0r}, \gamma_{1r}) f(\mathbf{Y}_r, \boldsymbol{\ell}_r|\boldsymbol{\theta}_r)$$

$$\propto \mathcal{N}(a_{\gamma_0}, b_{\gamma_0})\mathcal{N}(a_{\gamma_1}, b_{\gamma_1})\prod_{t=1}^{T_r}\prod_{i=1}^{N_r}\frac{\ell_{irt}^{a,rt-1}(1-\ell_{irt})^{b_{irt}-1}}{B(a_{irt}, b_{irt})},$$

where $a_{irt} = \varphi(\gamma_{0r} + \gamma_{1r}\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt})\phi_r$, $b_{irt} = (1 - \varphi(\gamma_{0r} + \gamma_{1r}\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt}))\phi_r$. We sample from $\pi(\gamma_{0r}, \gamma_{1r}|\ldots)$ via RW-MH.

C.4 Full conditional distribution of ζ_{ikr}

$$\pi(\zeta_{ikr}|\ldots) \propto \pi(\zeta_{ikr})f(\mathbf{Y}_r, \boldsymbol{\ell}_r|\boldsymbol{\theta}_r)$$

$$\propto \mathcal{N}(\zeta_{ikr}; 0, \sigma_{kr}^2)\prod_{t\in\mathcal{T}_{kr}}\frac{\ell_{irt}^{a_{it}-1}(1-\ell_{irt})^{b_{it}-1}}{B(a_{irt}, b_{irt})}\prod_{j\in\mathcal{S}_r^{-i}}\lambda_{ijrt}^{Y_{ijrt}}e^{-\lambda_{ijrt}},$$

where $\mathcal{T}_{kr} = \{t : \xi_{k,rt} = 1\}$, $\mathcal{S}_r^{-i} = \{j \in 1, \ldots N_r : j \neq i\}$, $\log\lambda_{ijrt} = \alpha_{ir} + \alpha_{jr} - ||\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt} - \boldsymbol{\zeta}_{jr}\boldsymbol{\xi}_{rt}||^2$, $a_{it} = \varphi(\gamma_{0r} + \gamma_{1r}\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt})\phi_r$, $b_{it} = (1 - \varphi(\gamma_{0r} + \gamma_{1r}\boldsymbol{\zeta}_{ir}\boldsymbol{\xi}_{rt}))\phi_r$. We sample from $\pi(\zeta_{ikr}|\ldots)$ via Adaptive RW-MH (see Subsection C.7).

C.5 Full conditional distribution of σ_{kr}^2

$$\pi(\sigma_{kr}^2|\boldsymbol{\zeta}_{kr}, \boldsymbol{\xi}_r) \propto \pi(\sigma_{kr}^2)\pi(\boldsymbol{\zeta}_{kr}|\sigma_{kr}^2, \boldsymbol{\xi}_r)$$

$$\propto (1/\sigma_{kr}^2)^{a_{\sigma^2}+1}e^{-b_{\sigma^2}\frac{1}{\sigma_{kr}^2}}\prod_{i=1}^{N_r}\left(1/\sigma_{kr}^2\right)^{\frac{1}{2}}e^{-\frac{1}{2\sigma_{kr}^2}\zeta_{ikr}^2}$$

$$\propto (1/\sigma_{kr}^2)^{a_{\sigma^2}+1}e^{-b_{\sigma^2}\frac{1}{\sigma_{kr}^2}}\left(1/\sigma_{kr}^2\right)^{\frac{N_r}{2}}e^{-\frac{1}{2\sigma_{kr}^2}\sum_{i=1}^{N_r}\zeta_{ikr}^2}$$

$$\propto (1/\sigma_{kr}^2)^{a_{\sigma^2}+\frac{N_r}{2}+1}e^{-\left(b_{\sigma^2}+\frac{\sum_{i=1}^{N_r}\zeta_{ikr}^2}{2}\right)\frac{1}{\sigma_{kr}^2}} \propto IG(a_{\sigma^2kr}^*, b_{\sigma^2kr}^*)$$

where $a_{\sigma^2kr}^* = a_{\sigma^2} + \frac{N_r}{2}$ and $b_{\sigma^2kr}^* = b_{\sigma^2} + \frac{\sum_{i=1}^{N_r}\zeta_{ikr}^2}{2}$.

C.6 Full conditional distribution of $\mathbf{q}_{lr}$

$$\pi(\boldsymbol{q}_{lr}|\boldsymbol{\xi}) \propto \left(\prod_{t=1}^{T_r}\prod_{k=1}^{K_r}q_{lkr}^{\xi_{l,r,t-1}\xi_{k,rt}}\right)\left(\prod_{k=1}^{K_r}q_{lk}^{\omega_k-1}\right)$$

$$\propto \prod_{k=1}^{K_r}q_{lk}^{\sum_{t=1}^{T_r}(\xi_{l,r,t-1}\xi_{k,rt})+\omega_k-1}$$

$$\propto Dir(\bar{\omega}_{1r}, \ldots, \bar{\omega}_{Kr}),$$

where $\bar{\omega}_{kr} = \sum_{t=1}^{T_r}(\xi_{l,r,t-1}\xi_{k,rt}) + \omega_k - 1$.

C.7 Adaptive MCMC

Sampling from the full conditional of $\boldsymbol{\alpha}_r$ and $\boldsymbol{\zeta}_{kr}$ is obtained via Adaptive MH algorithm with global adaptive scaling proposed in Andrieu and Thoms (2008). Adaptive MH generates samples from the distribution of $\boldsymbol{\theta}$ by iterating the following steps:

1. Starting values for the parameter of interest $\boldsymbol{\theta}_0$ and for $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are chosen.

2. For each iteration h, given $\boldsymbol{\theta}_h$, $\boldsymbol{\mu}_h$, $\boldsymbol{\Sigma}_h$ and δ_h:

   (a) $\tilde{\boldsymbol{\theta}}_h \sim \mathcal{N}(\boldsymbol{\theta}_{h-1}, \delta_{h-1}\boldsymbol{\Sigma}_{h-1})$ is sampled and $\boldsymbol{\theta}_h = \tilde{\boldsymbol{\theta}}_h$ with probability $\alpha(\boldsymbol{\theta}_{h-1}, \tilde{\boldsymbol{\theta}}_h)$, otherwise $\tilde{\boldsymbol{\theta}}_h = \boldsymbol{\theta}_{h-1}$;

   (b) Update $\log(\delta_h) = \log(\delta_{h-1}) + \gamma_h[\alpha(\boldsymbol{\theta}_{h-1}, \tilde{\boldsymbol{\theta}}_h) - \alpha^*]$;

   (c) Update $\boldsymbol{\mu}_h = \boldsymbol{\mu}_{h-1} + \gamma_h(\boldsymbol{\theta}_h - \boldsymbol{\mu}_{h-1})$;

   (d) Update $\boldsymbol{\Sigma}_h = \boldsymbol{\Sigma}_{h-1} + \gamma_h[(\boldsymbol{\theta}_h - \boldsymbol{\mu}_{h-1})(\boldsymbol{\theta}_h - \boldsymbol{\mu}_{h-1})' - \boldsymbol{\Sigma}_{h-1}]$;

   where $\gamma_h = \frac{1}{h^\psi}$ for $\psi \in (0,1)$ and α^* is a target acceptance rate, here chosen to be 25%.
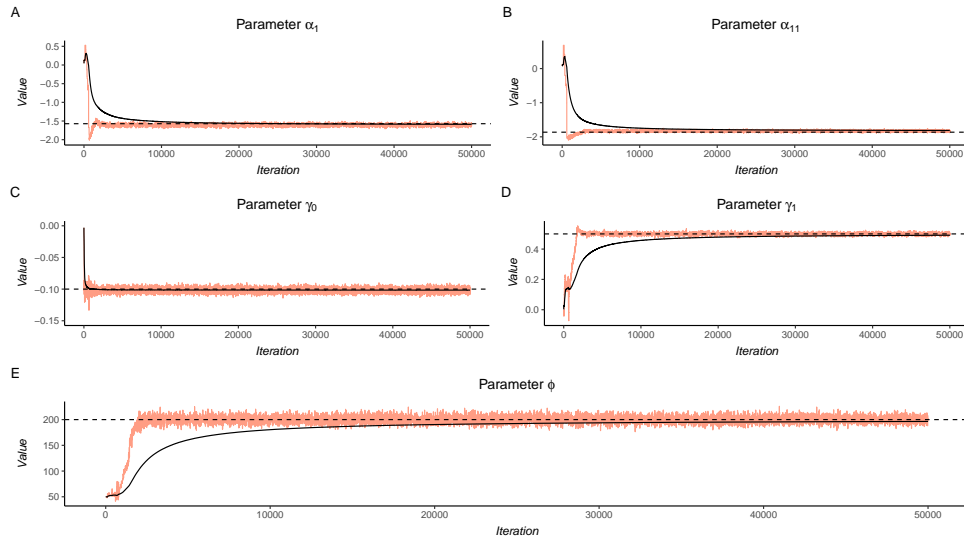
D MCMC Properties for the Simulated Data

Figure D.1: **Simulation Output:** Trace plots for the parameters α_1 (Panel A), α_{11} (Panel B), γ_0 (Panel C), γ_1 (Panel D) and ϕ (Panel E). The solid black line represents the cumulative average while the dashed black line represents the true parameter in the simulation.
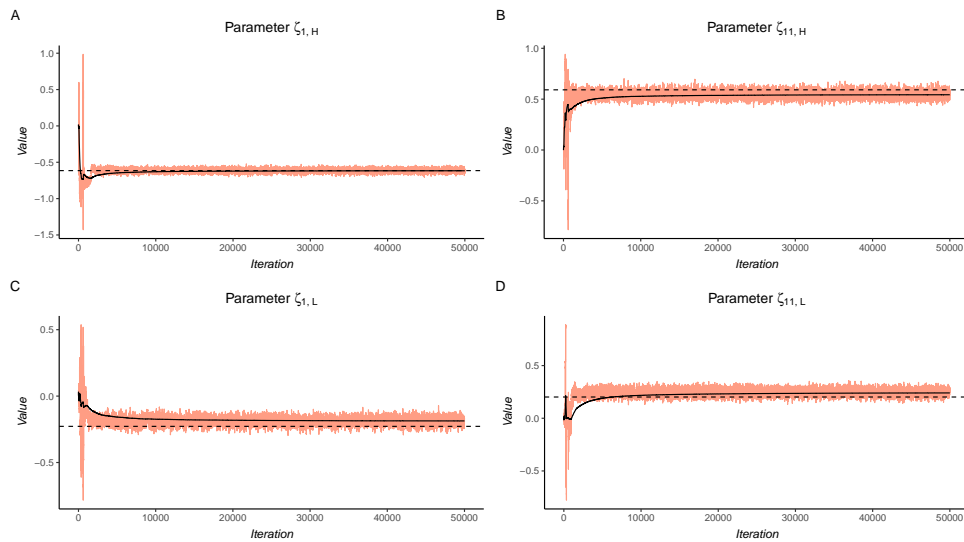

Figure D.2: **Simulation Output:** Trace plots for the latent leaning for outlet $1 - \zeta_{1,L}$ (Panel A), $\zeta_{1,H}$ (Panel B) – an for outlet $11 - \zeta_{11,L}$ (Panel C), $\zeta_{11,H}$ (Panel D). The solid black line represents the cumulative average while the dashed black line represents the true parameter in the simulation.
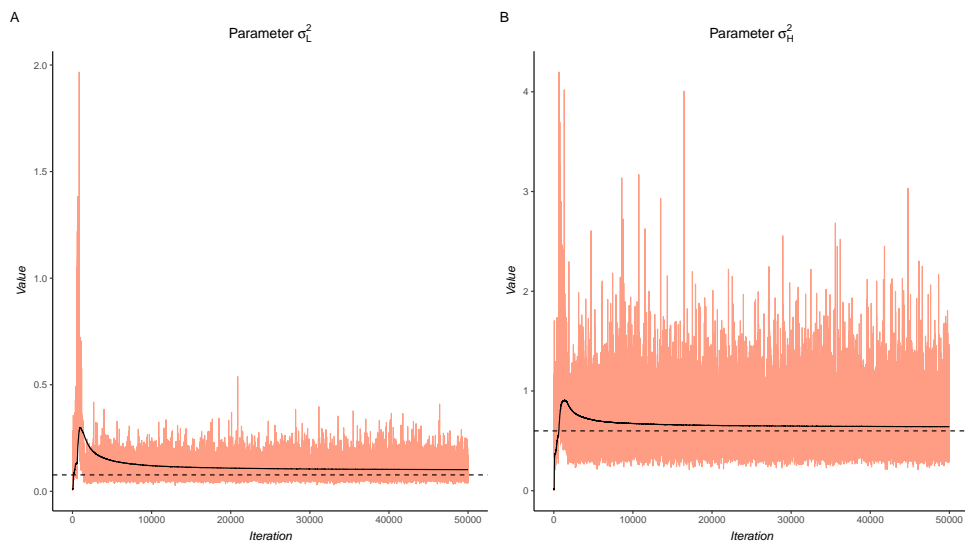
Figure D.3: **Simulation Output:** Trace plots for the parameters σ_L^2 and σ_H^2. The solid black line represents the cumulative average while the dashed black line represents the true parameter in the simulation.

Raw Series (50,000 obs.)						
Parameter	$\bar{\alpha}_i$	γ_0	γ_1	ϕ	$\bar{\zeta}_L$	$\bar{\zeta}_H$
ACF(1)	0.996	0.788	0.997	0.980	0.924	0.956
ACF(10)	0.972	0.190	0.978	0.925	0.713	0.803
ACF(30)	0.930	0.009	0.947	0.911	0.487	0.641
Acc.	25%	26%	26%	21%	25%	25%
ESS	0%	8%	0%	0%	2%	1%
CD p-val	0.050	0.175	0.000	0.000	0.109	0.142
With Burn-in (20,000 obs.)						
Parameter	$\bar{\alpha}_i$	γ_0	γ_1	ϕ	$\bar{\zeta}_L$	$\bar{\zeta}_H$
ACF(1)	0.891	0.730	0.898	0.750	0.849	0.766
ACF(10)	0.529	0.034	0.398	0.067	0.374	0.149
ACF(30)	0.318	-0.017	0.145	-0.011	0.174	0.048
Acc.	25%	25%	25%	22%	25%	25%
ESS	3%	16%	4%	14%	6%	10%
CD p-val	0.213	0.399	0.040	0.146	0.300	0.256
With Burn-in and Thinning every 10 (2000 obs.)						
Parameter	$\bar{\alpha}_i$	γ_0	γ_1	ϕ	$\bar{\zeta}_L$	$\bar{\zeta}_H$
ACF(1)	0.527	0.017	0.401	0.040	0.367	0.149
ACF(10)	0.135	-0.005	0.054	0.028	0.057	0.011
ACF(30)	0.022	0.022	0.003	0.004	0.005	0.008
Acc.	-	-	-	-	-	-
ESS	23%	100%	35%	93%	44%	66%
CD p-val	0.209	0.402	0.033	0.013	0.275	0.259

Table D.1: **Simulated data:** Auto-correlation (ACF) at lag 1, 10, 30, Acceptance Rate, Effective Sample Size over the number of draws (ESS) and Convergence Diagnostic p-value (CD) as defined in Geweke et al. (1991) for the MCMC sampling of $\bar{\alpha}$, γ_0, γ_1, ϕ, $\bar{\zeta}_L$, $\bar{\zeta}_H$ respectively on the raw series, on the series collected after burn-in, and on the series applying burn-in and thinning.

E Model Selection

We compute DIC to compare models as in Gelman et al. (2014): $DIC_r(\mathcal{M}_r) = -2\overline{\log f}(\mathbf{Y}_r, \boldsymbol{\ell}_r | \boldsymbol{\theta}, \mathcal{M}_r) + p_D$, where $\overline{\log f}(\mathbf{Y}_r, \boldsymbol{\ell}_r | \boldsymbol{\theta}, \mathcal{M}_r) = (1/H) \sum_{h=1}^{H} \log f(\mathbf{Y}_r, \boldsymbol{\ell}_r | \boldsymbol{\theta}_r^h, \mathcal{M}_r)$ is the sample average of the loglikelihood computed at each iteration h of model $\mathcal{M}$ while $p_D = 2Var\left(\log f(\mathbf{Y}_r, \boldsymbol{\ell}_r | \boldsymbol{\theta}_r^{1:H}, \mathcal{M}_r)\right)$.

We also compute the Log Pointwise Predictive Density (lppd) as in Gelman et al. (2014):

$$lppd_r(\mathcal{M}_r) = \sum_{i>j,t} \log\left(\frac{1}{H} \sum_{h=1}^{H} f_P\left(Y_{ijrt} \mid \boldsymbol{\theta}_r^h, \mathcal{M}_r\right)\right).$$

Model	France	Germany	Italy	Spain
$\overline{\log f}_P(\mathbf{Y}_r | \boldsymbol{\theta}_r, \mathbf{s}_r, \mathcal{M}) \times 10^{-6}$				
$\mathcal{M}_1$	$-\mathbf{2.2796}$	$-\mathbf{1.2198}$	-1.6789	$-\mathbf{2.3365}$
$\mathcal{M}_2$	-2.2796	-1.2198	$-\mathbf{1.6788}$	-2.3491
$\mathcal{M}_3$	-2.3602	-1.2706	-1.7664	-2.4536
$\overline{\log f}(\mathbf{Y}_r, \boldsymbol{\ell}_r | \boldsymbol{\theta}_r, \mathbf{s}_r, \mathcal{M}) \times 10^{-6}$				
$\mathcal{M}_1$	$-\mathbf{2.2347}$	$-\mathbf{1.1834}$	$-\mathbf{1.6524}$	$-\mathbf{2.3069}$
$\mathcal{M}_2$	-2.2348	-1.1834	-1.6532	-2.3194
$\mathcal{M}_3$	-2.3153	-1.2342	-1.7397	-2.4240

Table E.1: **Model Selection:** DIC-score components for the three nested models considered for each country.

F Details of the Observable Media Slant Index

To construct an observable proxy for media slant we rely on the methodology proposed by Gentzkow and Shapiro (2010) and extended to online textual data by Garz et al. (2020). Textual processing is carried out with the use of the R package `quanteda`. The underlying intuition is that of computing the distance between the language used by news outlets in their posts and the language used by political parties. To do so we compose two corpora of textual data: the *Parties Corpus* consisting of the textual content of the posts published by the major parties in the years 2015-2016 on their Facebook wall, the *Outlets Corpus* consisting of the same information related to the posts published by the Italian news outlets considered in this work.

On both corpora, textual pre-processing is carried out (lower case transformation, punctuation removal, stopwords removal and n-gram tokenization) and tokens not present in the Outlets corpus are filtered out from the Parties Corpus. By means of the TF-IDF score applied on the Parties Corpus, we retrieve the top 100 tokens with highest TF-IDF score for each party. We proceed assessing the cosine similarity between the vector of tokens $\mathbf{x}_{t,o}$ obtained from the set of posts published by outlet o at time t and the set of tokens $\mathbf{y}_p$ characteristic of each party p.

$$sim_{pot} = \frac{\sum_{k=1}^{K} x_{kot} y_{kp}}{\sqrt{\sum_{k=1}^{K} x_{kot}^2}\sqrt{\sum_{k=1}^{K} y_{kp}^2}} \tag{F.1}$$

To take into account the fact that the style of posting adopted by some parties is closer to the one of news outlets and vice-versa, Garz et al. (2020) suggest regressing the similarity on a constant and both outlet and party fixed-effects to extract the residuals ϵ_{pot}, which can be interpreted as a proxy of unexplained similarity. Finally the media slant for outlet o at time t is computed as:

$$slant_{ot} = \sum_{p=1}^{P} \epsilon_{pot} score_p \tag{F.2}$$

where $score_p$ is the political leaning assigned to party p by the 2014 Chapel Hill Experts Survey classification provided in Polk et al. (2017), see Fig. F.1.

Figure F.2 reports the average media slant for the available set of news outlets for France, Germany, Italy, and Spain, trough the whole time lapse.
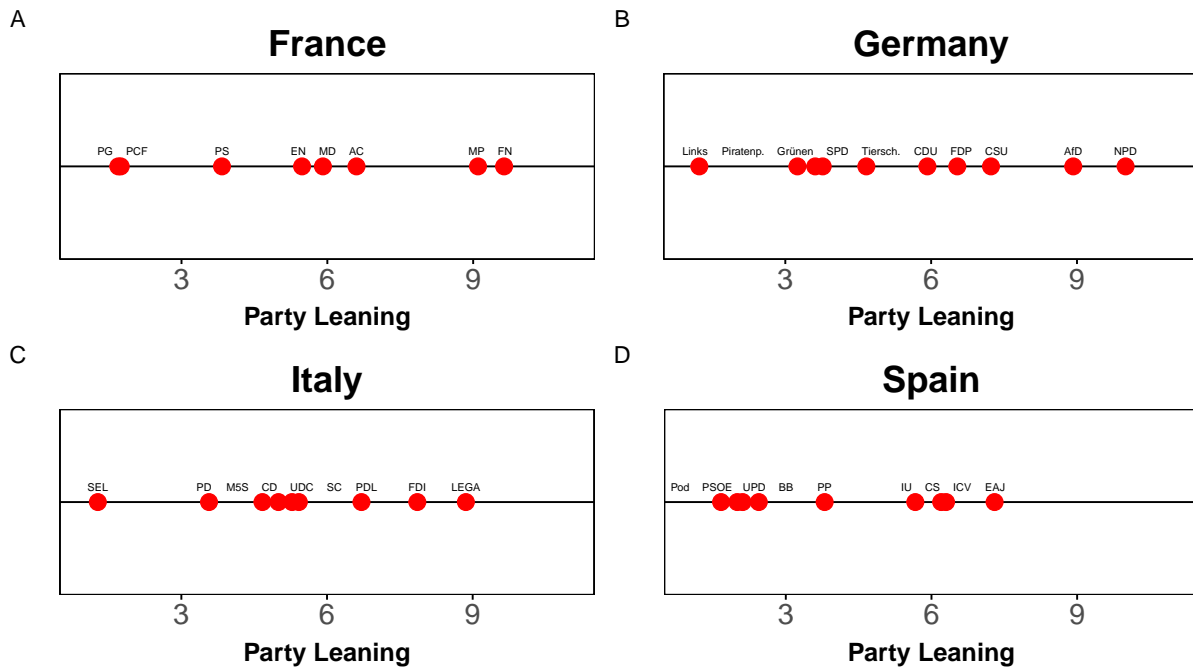
Figure F.1: **Ideology of Political Parties:** The index is computed according to the 2014 Chapel Hill Experts Survey (Polk et al., 2017).
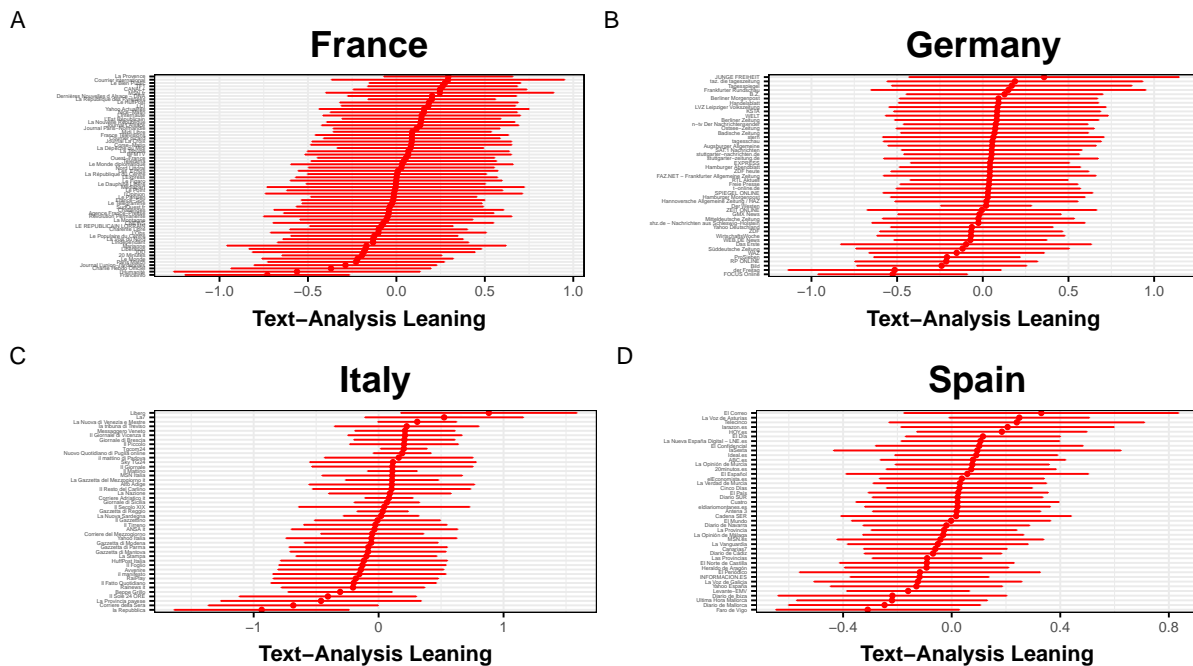

Figure F.2: **Media Slant of News Outlets:** The index is computed following the methodology proposed by Gentzkow and Shapiro (2010) and Garz et al. (2020). Full dots represent the average leaning and error bars represent the $\mu \pm \sigma$ interval over the time-lapse.

G Identifiers and List of Removed Outlets

The table below lists the news outlets used as identifiers of the political orientation.

News Outlet	Country	Assumed Media Bias	Sign	PEW score (0-6)
l'Humanité	France	Left	<0	NA
Bild	Germany	Center-Right	>0	3.1
Libero	Italy	Center-Right	>0	3.6
ABC	Spain	Center-Right	>0	3.3

Table G.1: **Identifiers:** Set of outlets used to identify the political orientation of the news outlets. The column sign reports the assumed political orientation: either < (left) or > (right). The column PEW score reports the estimated leaning as per Mitchell et al., 2018.

A few news outlets were removed in the dynamic analysis because of their prolonged inactivity, i.e. 15 days without any comment. They are DE: *Der Westen, GMX News, WEB.DE, News ZDF*, FR: *Charlie Hebdo Officiel, France Télévisions, Franceinfo, LCI, Révolution Permanente*, IT: *La Gazzetta del Mezzogiorno.it, MSN Italia*, SP: *La Voz de Asturias, Yahoo España*.

H   Controlling for Exposure

Here we discuss the inclusion of an additional variable, $TotCom_{rt}$, accounting for the exposure to the total number of comments at time t. This is to compare polarisation across periods with different engagement levels. To maintain parameter interpretation, we add to the log-intensity the de-meaned logarithm of the total number of comments, $\log TotCom_{rt}$:

$$\log \lambda_{ijrt} = \alpha_{ir} + \alpha_{jr} + \delta \log TotCom_{rt} - \beta_r ||x_{irt} - x_{jrt}||^2. \tag{H.1}$$

We run our modified LS model for the Italian data. Figure H.1 reports the posterior estimates of the latent space and states through time when we include this control. Panel A reports the latent space of Italian news outlets in the two states. State identification is not trivial anymore. In fact, we notice heterogeneous behavior between local and national outlets. We achieve state identification by considering the average distance computed across national news outlets (triangles), as they involve a larger number of commenters. Panel B reports the latent states through time. The states are coherent with those found in Sec. 4. Figure H.2 reports the marginal posteriors draws for the parameters δ, γ_{0r}, γ_{1r} and ϕ_r. The sign of the parameter δ is as expected. Higher comments overall in the network implies also a higher number of comments in common between the pairs of pages.
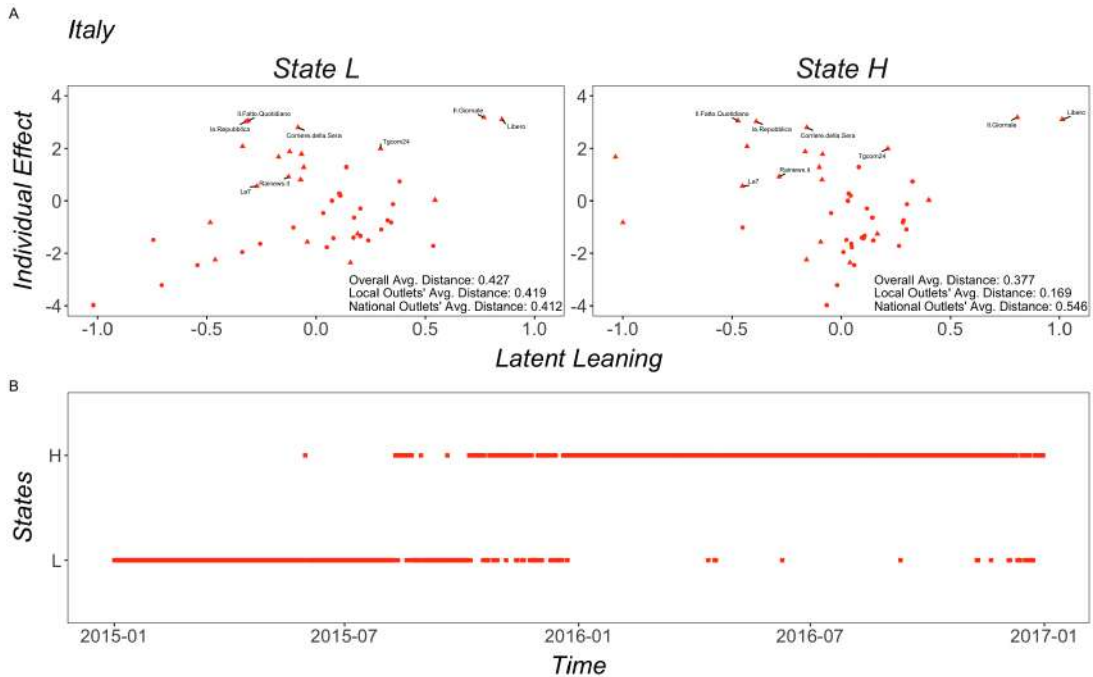

Figure H.1: **Posterior Mean Latent Space and States with Exposure Control:** Panel A reports the latent space of Italian news outlets in the two states. State identification is achieved considering the average distance computed across national news outlets (triangles). Panel B reports the latent states through time. The states are coherent with those found in Sec. 4.
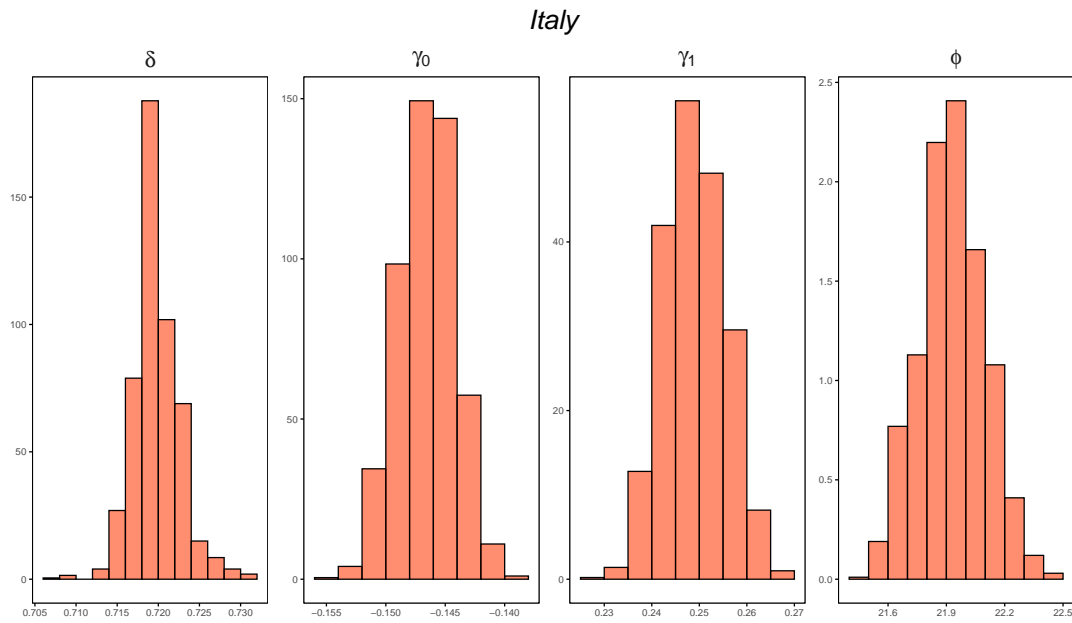
41

Figure H.2: **Marginal Posterior Distributions with Exposure cControl:** Histograms for the MCMC draws for the parameters δ, γ_{0r}, γ_{1r} and ϕ_r for Italy.

I   Further Results for the Facebook Data

Figure I.1 presents the posterior distributions of the parameters of the static LS model for the Facebook data.
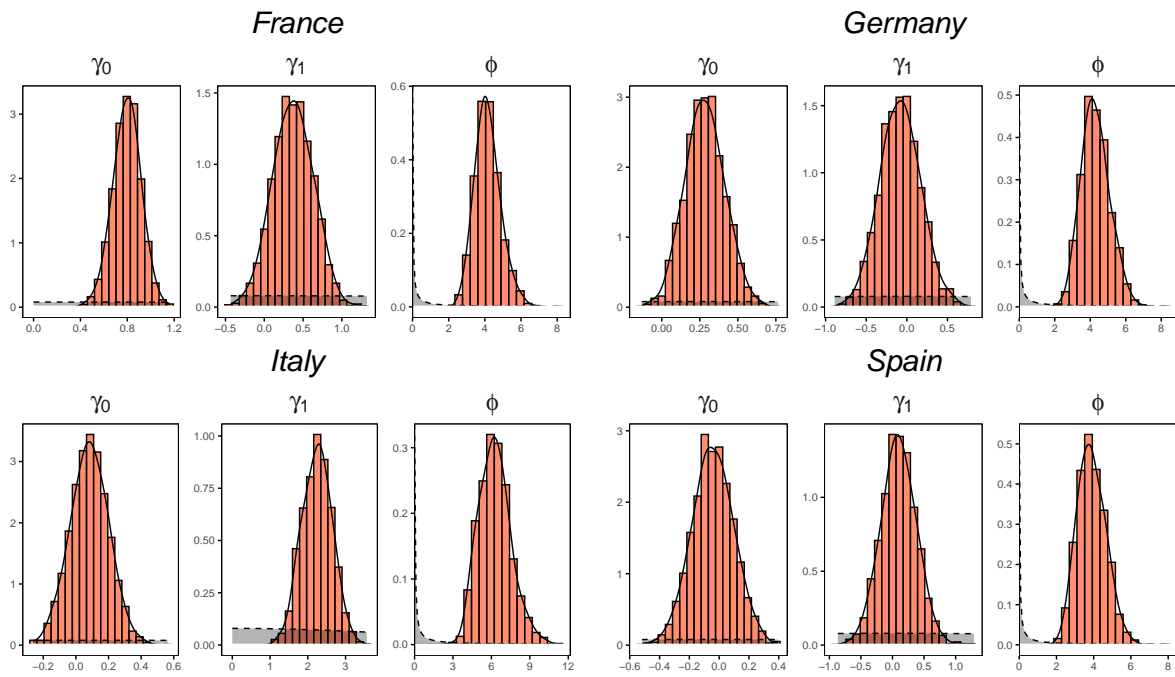
Figure I.1: **Marginal Posteriors - Static Analysis:** Histograms for the MCMC draws for the parameters γ_{0r}, γ_{1r} and ϕ_r for France, Germany, Italy and Spain with prior pdfs indicated by dashed lines.
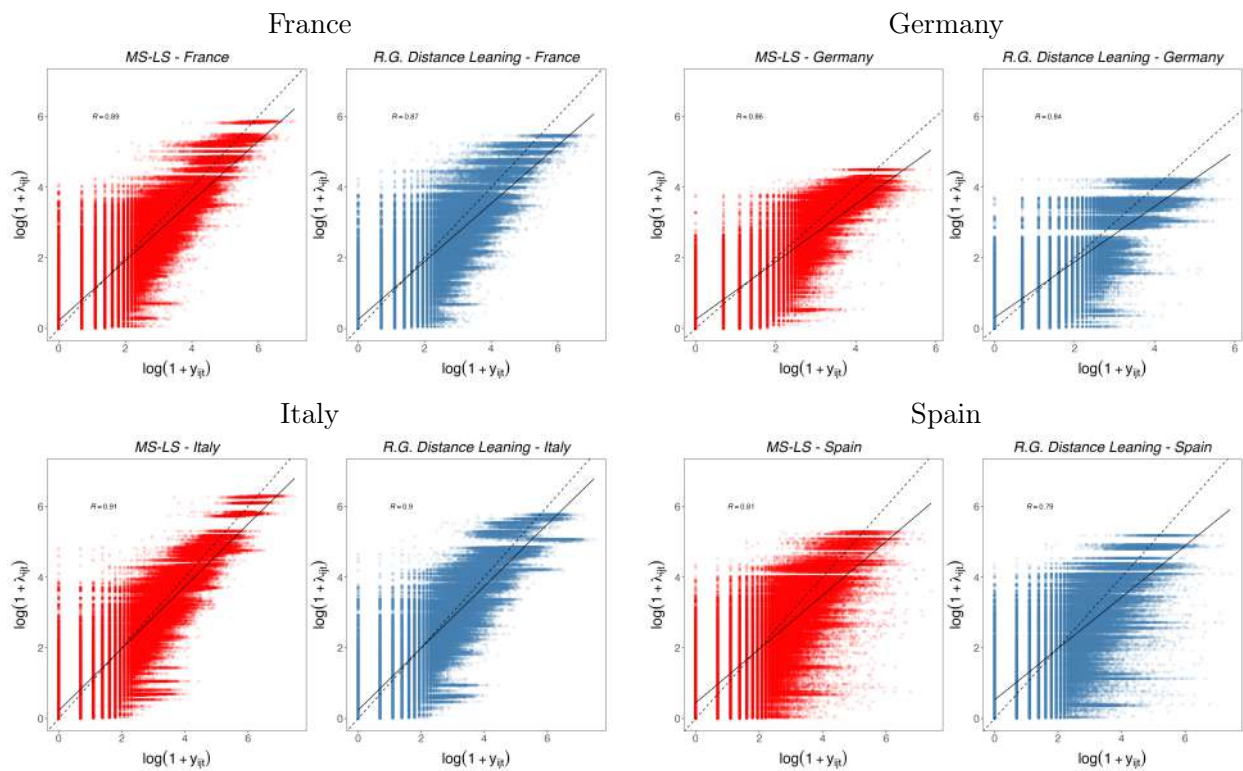
Figure I.2: **Predictive-Check Comparison:** Scatter plot of between y_{ijrt} and the posterior λ_{ijt} for the MS-LS model (left panels) and for the R.G. model with the observed-leaning distance covariate in log scales. We also plot the 45-degree line (dashed) and the set of regression lines for each iteration (solid lines, very close to each other)

J Data and Scripts Repository

Data and Scripts are stored in the following Repository: `https://github.com/BayesianEcon/Dyn-MS-LS-Media`. Refer to the README file in the repository for a complete description of each file.

J.1 Data

The data entirely covers 729 days from "2015-01-01" to "2016-12-31". The number of news outlets included in each dataset is the following: Germany 47 news outlets, France 62, Italy 45 and Spain 43.

Network Dataset: Data set used in the illustration of the MS-LS Network Model in Section 4. The data set is an edge-list representation of the media networks where columns "i" and "j" refer to the nodes, column "t" refers to the day and column "w" refers to the number of unique Facebook commenters in common between "i" and "j" at time "t". The static-version of the Network Dataset of each country is contained within the file ("Data_Env_single_(country).RData"), while the dynamic version is in the file ("DataEnv_(country)_all.RData").

Slant-Index Dataset: Data set used in the illustration of the MS-LS Network Model in Section 4. Refer to Appendix F and the README file for an illustration of how the index has been obtained. The data set represents a nodal feature of the media networks where column "i" refers to the nodes, column "t" refers to the day and column "leaning" refers to the Slant Index.

The static version of the Slant-Index Dataset of each country is contained within the file ("Data_Env_single_(country).RData"), while the dynamic version is included in the file ("DataEnv_(country)_all.RData").

J.2 Scripts

We report here a brief description of the main scripts used to estimate the Bayesian MS-LS network model on the datasets studied in the main paper (Sections 3 and 4) and the supplementary material. Our MCMC algorithm is entirely implemented in C++, enabling faster execution speed compared to interpreted languages like R or Python. However, we still rely on R for data manipulation and plotting. The smooth integration of the two languages has been made possible through the utilization of the Rcpp package, which offers a convenient interface for invoking C++ scripts within R. Refer to the README.txt file in the repository for a complete description of each script.

JASA_Simulation_02_results.R:
Estimates the MS-LS model on the simulated network dataset.
Running time $\sim$ 12 mins (50,000 iterations, Apple M2, 8 GB Memory)

JASA_Static_01_Results_(Country).R:
Estimates the MS-LS model on the static network dataset.
Running time $\sim$ 45 mins. (15,000 iterations, Apple M2, 8 GB Memory)

JASA_Dynamic_01_Results_(country).R:

Estimates the MS-LS model on the dynamic network dataset.

Running time > 20 hrs. (35,000 iterations, Apple M2, 8 GB Memory)

JASA_MS_LS_FE.cpp:

The script contains the function to generate MCMC draws for the dynamic Bayesian MS-LS network model.