Functional Graphical Models to map Brexit debate on Twitter

Nicola Pronello[a], Emiliano del Gobbo[b], Lara Fontanella[a], Rosaria Ignaccolo[c], Luigi Ippoliti[a], and Sara Fontanella[d]

[a]Università degli Studi "G. d'Annunzio" Chieti - Pescara; nicola.pronello@unich.it, lara.fontanella@unich.it, luigi.ippoliti@unich.it
[b]Università degli Studi di Foggia; emiliano.delgobbo@unifg.it
[c]Università degli Studi di Torino; rosaria.ignaccolo@unito.it
[d]Imperial College London; s.fontanella@imperial.ac.uk

Abstract

In recent years a literature on multivariate functional graph models has been developed. The graphical representation of the conditional dependence among a finite number of random variables is indeed appealing in different applications, such as e.g. the analysis of the brain connectivity. We want to investigate a novel extension of this methodology, considering random functions spatially and temporally correlated. A motivating case study is the analysis of the semantic network that tracks the change of the Brexit debate on Twitter across UK during a particular time frame. By considering the change in time of a word usage as a functional realization, a semantic network on the topic of interest is defined by a graphical representation of the conditional dependence among functional variables.

Keywords: Functional graphical models, Functional data analysis, Kernel Smoothing, Semantic network

1. Introduction

In recent years, literature on graphical models for functional data has been developed (6; 4). Indeed visualizing an estimated graph representing relationships between functional variables can be very effective to represent conditional dependence among them. In this framework, we contribute to extend this methodology by considering functional graphs that are spatially and temporally correlated; in particular they are supposed to vary on a spatio-temporal lattice, and for their estimation only limited measurements are available. This setting is motivated by data representing daily word usage on the Brexit debate on Twitter during 13 months (temporal units) and 41 districts in UK (spatial units) without replicates. Our main goal is to estimate a semantic network representing the connections of words used in such a debate. By means of functional graphical models it is possible to represent the connection between words from their monthly usage trends in Twitter. So doing, we offer a different insight on a public debate, moving beyond classical semantic networks built from co-occurrences of words in a sentence/tweet.

2. Motivating case study: Twitter conversations about Brexit

We dispose of data extracted from Twitter regarding the Brexit debate, as collected and pre-processed by del Gobbo et al. (1), by combining around one million of tweets with hashtag *Brexit* spanning in

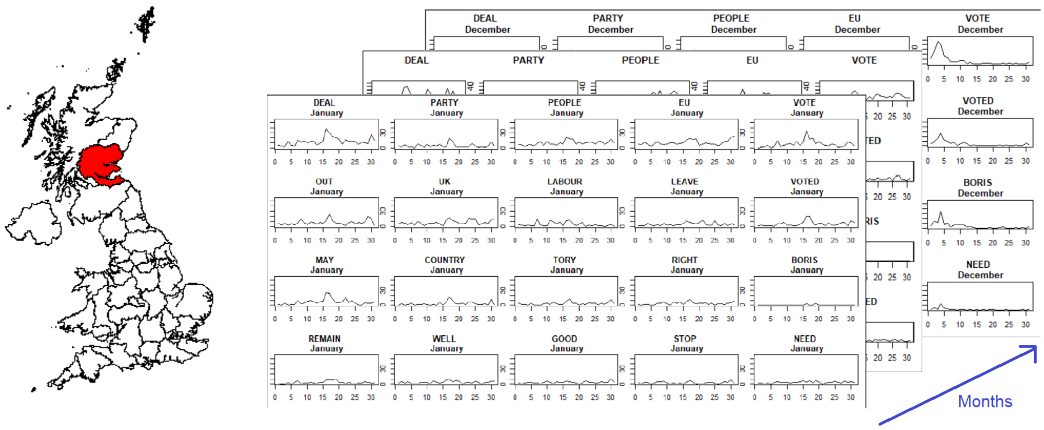

Figure 1: Example for a specific district (Eastern Scotland)

a range of 13 months (from 31 December 2018 to 9 February 2020) and geolocalized in the United Kingdom territory with 41 NUTS-2 (Nomenclature of territorial units for statistics) districts. We consider the daily usage of the first p most common words in each district in a time window of a month. Thus we have p daily time series (one for the daily usage of each word) for each month in each district, as shown in Figure 1 for the particular case of Eastern Scotland, and each time series can be seen as a noisy realization of a functional random variable. Hence the domain of the considered multiple functional data is a spatio-temporal irregular lattice $\mathcal{L} = \mathcal{S} \times \mathcal{M}$ where $\mathcal{S}$ represents the set of 41 districts and $\mathcal{M}$ the set of 13 months.

3. Functional graphical models

Let us consider a spatio-temporal irregular lattice $\mathcal{L} = \mathcal{S} \times \mathcal{M}$ where $\mathcal{S}$ represents a set of spatial units (districts) and $\mathcal{M}$ a set of time frames. Let us denote with $\mathbf{c} = (s, t)$ the spatio-temporal coordinate of a generic element of $\mathcal{L}$ and let $\mathbf{Y^c} = \{Y_1^{\mathbf{c}}, ..., Y_p^{\mathbf{c}}\}$ be a p-dimensional functional random variable taking values in $(L_2(\mathcal{T}))^p$, with $\mathcal{T} = [a, b]$, defined at a specific site (district) and time frame (month). The vector of mean functions is $E[\mathbf{Y^c}(\tau)] = \mu^{\mathbf{c}}(\tau) = \{\mu_j^{\mathbf{c}}(\tau)\}_{j=1,..,p}$, while the matrix of functional covariances is written as $\mathbf{C}(\mathbf{Y^c}(\tau), \mathbf{Y^c}(\tau')) = \mathbf{C}_{\mathbf{Y}}^{\mathbf{c}}(\tau, \tau') = \{C_{j,r}^{\mathbf{c}}(\tau, \tau')\}_{(j,r)=1,..,p}$.

By considering a continuous orthonormal basis functions, $\phi_{j,1}, \phi_{j,2}, \ldots$, the well-known Karhunen-Loéve expansion (KLE, see e.g. (2)) allows us to represent each functional variable $Y_j(\tau)^{\mathbf{c}}$ as

$$Y_j^{\mathbf{c}}(\tau) = \sum_{d=1}^{\infty} a_{j,d}^{\mathbf{c}} \phi_{j,d}(\tau),$$

where the expansion coefficients $a_{j,d}^{\mathbf{c}} = \int_{\mathcal{T}} Y_j^{\mathbf{c}}(\tau)\phi_{j,d}(\tau)d\tau$ are uncorrelated random variables, and then we consider with $D < \infty$ its truncated version

$$\tilde{Y}_j^{\mathbf{c}}(\tau) = \sum_{d=1}^{D} a_{j,d}^{\mathbf{c}} \phi_{j,d}(\tau).$$

Let $\mathbf{A^c} = \{\mathbf{a}_1^{\mathbf{c}}, ..., \mathbf{a}_p^{\mathbf{c}}\} \in \mathcal{R}^{pD}$ be the $(D \times p)$ matrix collecting the expansion coefficients associated with the p functions and let

$$\Sigma^{\mathbf{c}} = \{Cov(\mathbf{a}_j^{\mathbf{c}}, \mathbf{a}_r^{\mathbf{c}})\}_{(j,r)=1,...,p} = \{\Sigma_{j,r}^{\mathbf{c}}\}_{(j,r)=1,...,p}$$

be the $(Dp \times Dp)$ matrix collecting all covariances between expansion coefficients of all p functional variables. For the functional covariances we have the approximation:

$$C_{j,r}^{\mathbf{c}}(\tau, \tau') = C(Y_j^{\mathbf{c}}(\tau), Y_r^{\mathbf{c}}(\tau')) \approx C(\tilde{Y}_j^{\mathbf{c}}(\tau), \tilde{Y}_r^{\mathbf{c}}(\tau')),$$

and since vectors $\mathbf{a}_j^{\mathbf{c}}$ share the same information with $\tilde{Y}_j^{\mathbf{c}}(\tau)$ we can work with $\Sigma_{j,r}^{\mathbf{c}}$ and the matrix of partial correlations $\Theta^{\mathbf{c}} = (\Sigma^{\mathbf{c}})^{-1} = \{\Theta_{j,r}^{\mathbf{c}}\}_{(j,r)=1,...,p}$.

The matrix of partial correlation just defined is a key ingredient to evaluate relationships among functional variables in the framework of the graphical models. To this goal, for each coordinate $\mathbf{c}$, let $G^{\mathbf{c}} = \{V, E^{\mathbf{c}}\}$ be a undirected graph where $V = \{1, ..., p\}$ represents the set of vertices corresponding to the p random functions and $E^{\mathbf{c}} \subseteq \{(j, r) \in V \times V, \ j \neq r\}$ represents the set of edges specified by means of

$$(j, r) \notin E^{\mathbf{c}} \ \text{ if } \ ||\Theta_{j,r}^{\mathbf{c}}||_F = 0$$

where $||\cdot||_F$ denotes the Frobenius norm. The condition below is equivalent to conditional independence between the j-th and r-th variables in the Gaussian case and, in analogy with Qiao et al. (2019, (6)), by considering the p functional random variables as vertices of the graph $G^{\mathbf{c}}$ we can say that $\mathbf{Y}^{\mathbf{c}}$ follows a *Functional Graphical Model (FGM)*.

To retrieve a network, for every $\mathbf{c}$, among p functional random variables in a *FGM* we need to estimate $\Theta^{\mathbf{c}}$. Moreover, since we are interested in underlying just the most important connections among words, we consider a sparse estimator for $\Theta^{\mathbf{c}}$. To this goal, we consider the Functional graphical lasso criterion (`fglasso`) introduced by (6) as a block extension of the classical `glasso` algorithm (8; 3). The sparsity in the precision matrix is achieved by imposing a group lasso penalty, so that the estimated $\widehat{\Theta}^{\mathbf{c}}$ at each $\mathbf{c}$ is obtained by solving an optimization problem:

$$\widehat{\Theta}^{\mathbf{c}} = \underset{\Theta^{\mathbf{c}}}{\operatorname{argmax}} \left(\log\det \Theta^{\mathbf{c}} - \operatorname{trace}(\hat{\Sigma}^{\mathbf{c}}\Theta^{\mathbf{c}}) - \lambda \sum_{j \neq r} ||\Theta_{j,r}^{\mathbf{c}}||_F \right), \tag{1}$$

where $\hat{\Sigma}^{\mathbf{c}}$ is an estimate of $\Sigma^{\mathbf{c}}$ (that needs to be obtained) and λ is a nonnegative tuning parameter. The group lasso penalty $\lambda \sum_{j \neq r} ||\Theta_{j,r}^{\mathbf{c}}||_F$ shrinks all the elements in $\Theta_{j,r}^{\mathbf{c}}$ towards zero (or all nonzero, that is the case of an estimate edge between $Y_j^{\mathbf{c}}$ and $Y_r^{\mathbf{c}}$) leading to a sparser $\widehat{\Theta}^{\mathbf{c}}$ in a blockwise way, and consequently to a sparser graph $\hat{G}^{\mathbf{c}}$, when λ increases. Obviously, if $\lambda = 0$ there is no penalty and the choice of this regularization parameter is crucial: there exist suitable AIC indexes used in (6), but instead we adopt an heuristic choice to ease the interpretability of the resulting semantic network by considering an overall percentage of nonzero links below 10%.

4. Nonparametric estimator of $\Sigma^{\mathbf{c}}$

The reconstruction of sparse network structures by means of Equation 1 is possible if we have an estimation of $\Sigma^{\mathbf{c}}$ varying on the lattice $\mathcal{L}$, that is one for spatio-temporal coordinate $\mathbf{c} = (s, t)$ (indicating district and month in our case study). However, constructing an estimator of $\Sigma^{\mathbf{c}}$ represents a challenge because we deal with the case of extremely sparse data since in each $\mathbf{c}$ we observe only one realization of $\mathbf{Y}^{\mathbf{c}}$. By considering mean-corrected functional data for the sake of simplicity, a naive estimate of $\Sigma^{\mathbf{c}}$ is given by the raw covariance

$$\Sigma^{*\mathbf{c}} = \left\{\mathbf{\Sigma}_{j,r}^{*\mathbf{c}}\right\}_{j,r=1,\ldots,p} = \left\{\mathbf{a^c}_j\mathbf{a^c}_r^T\right\}_{j,r=1,\ldots,p},$$

that exploits only the information in one datum and is not full rank since by construction it holds $\operatorname{rank}(\Sigma_{j,r}^{*\mathbf{c}}) = 1, \forall j, r$. Then to obtain a reliable estimator of $\Sigma^{\mathbf{c}}$ we propose to borrow information from the neighbours of the unit with coordinate $\mathbf{c}$ in the lattice by means of linear smoothing.

Given n units in the spatio-temporal irregular lattice $\mathcal{L}$ with coordinates $\mathbf{c}_i$, with $i = 1, \ldots, n$, we consider the class of linear smoother estimators defined by

$$\widehat{\Sigma}^{\mathbf{c}} = \sum_{i=1}^{n} \omega_i(\mathbf{c})\Sigma^{*\mathbf{c}_i}$$

where the coefficients $\omega_i(\mathbf{c})$ of the linear combination need to be determined; this class includes the Gaussian process regression estimator as well as the Kernel smoother and local polynomial estimator.

Finding $\omega(\mathbf{c}) = (\omega_1(\mathbf{c}), ..., \omega_n(\mathbf{c}))$ is equivalent to solve an optimization problem:

$$\widehat{\Sigma}^{\mathbf{c}} = \text{argmin}_{\Sigma^{\mathbf{c}}} \sum_{i=1}^{n} K(\mathbf{c}_i, \mathbf{c}) d(\Sigma^{\mathbf{c}}, \Sigma^{*\mathbf{c}_i}),$$

where and $K(\cdot,\cdot) : \mathcal{L} \times \mathcal{L} \to R$ is a kernel function and $d(\cdot,\cdot)$ is a suitable distance between covariance matrices. The choice of the distance d is not uniquely identified and different choices may or may not incorporate the constraints of the space of the positive definite matrices.

Determining K from space-time contiguity In this work, we propose to construct kernel weights that are suitable for a spatio-temporal lattice domain as in our case study. Usually a Kernel function is defined as a function of distances, and so in a spatial lattice distances among areas would be among their barycenters while in a temporal lattice distances would be lags.

To consider properly the nature of the spatio-temporal domain where graphs live, we instead take a contiguity point of view both in space and time and define the kernel by means of a Laplacian $\mathbf{L}$. Let $\mathbf{W}^{\text{Space}}$ be the adjacency matrix for the areal units such that entries are equal to 1 if two areal units are neighbours (contiguous regions), and 0 otherwise; and let $\mathbf{W}^{\text{Time}}$ be the adjacency matrix for the time frames (months) with element equal to 1 when a month before or after is considered (like a 3 months moving average window). In order to have a global adjacency matrix for the units in the spatio-temporal lattice $\mathcal{L}$ we take the Kronecker product and define

$$\mathbf{W} = \mathbf{W}^{\text{Space}} \otimes \mathbf{W}^{\text{Time}},$$

and then we consider the Laplacian

$$\mathbf{L} = \mathbf{D} - \mathbf{W}$$

with $\mathbf{D} = diag\{\mathbf{W}\mathbf{1}_n\}$.

Finally, as kernel values in the vector $\omega(\mathbf{c})$ for the linear smoother we use the elements of the matrix

$$\mathbf{K} = (\mathbf{I} + \gamma \mathbf{L})^{-1},$$

where γ is a smoothing parameter (to be fixed) that behaves as a kernel bandwidth. In fact, γ tunes the values of the weights and the extension of the neighborhood of each unit in $\mathcal{L}$ and for $\gamma \to \infty$ one has $\omega_i \to 1/n$.

5. Results about Brexit

The proposed methodology allows to estimate semantic networks, tracking the change of the Brexit debate on Twitter, for each of the 41 districts in UK along months from January 2019 to January 2020, having borrowed information by a spatio-temporal neighborhood. We take the first $p = 20$ most common words in the dataset and consider the daily usage of each word for each month in each district as (spatially and temporally correlated) functional data. Looking at the sequence of estimated semantic networks in a specific district along time (as e.g. those in Figure 2) we can detect which word is really central in the debate month by month. The estimated networks show an interesting pattern: during the first months of 2019 the word `MAY` (the former prime minister surname) is connected with other words in the debate; this holds until the summer when she resigned in favor of Boris Johnson, and then the word `BORIS` becomes connected with others in August and October. Figure 2 shows the estimated networks in Inner West London for the months January, February and March 2019, with two different values of γ, namely $\gamma = 0.2$ and $\gamma = 2$. Even with only three months we can note that the estimated network changes enough with a smaller value of γ (first row in Figure 2) while it persists with a larger value (second row in Figure 2): γ controls the 'width' of the kernel used in the smoothing process, i.e. the magnitude of elements of the matrix $\mathbf{K}$. In the first case, with smaller γ, the only weights different from zero are those related to the spatial neighbours in the same month; while a larger value acts as a large kernel bandwidth and leads to an estimate close to an average.

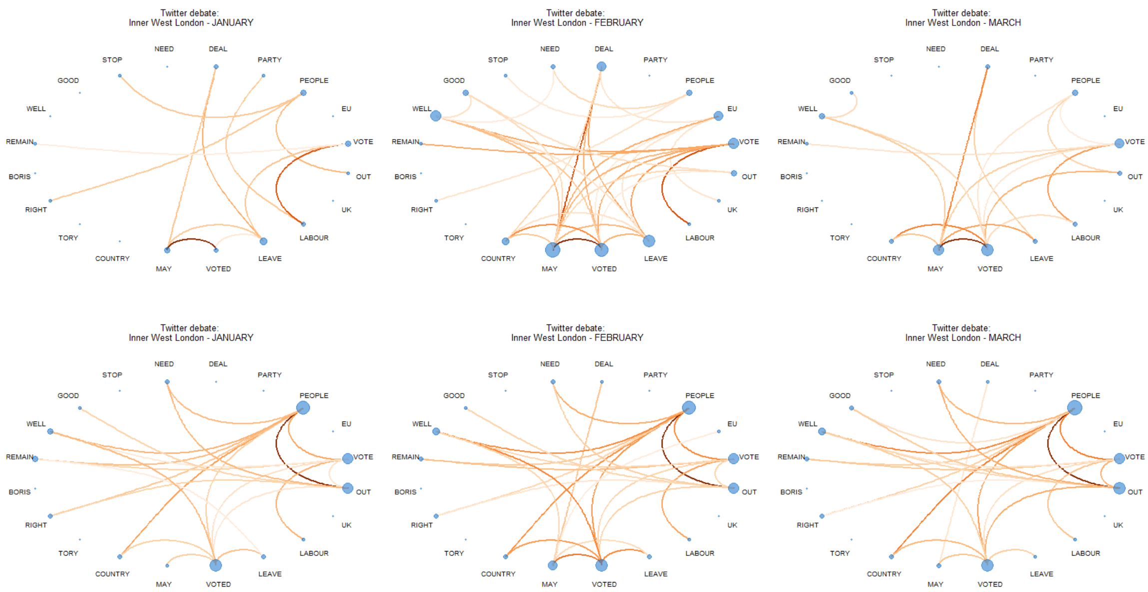

Figure 2: Estimated networks in Inner West London for the months January, February and March 2019, with $\gamma = 0.2$ (*first row*) and with $\gamma = 2$ (*second row*). Each vertex corresponds to a functional variable representing the usage of a word, the area of the blue dots is proportional to how many times that word is connected to the other words. The color and the thickness of the estimated edge change with the norm of $\hat{\Theta}_{j,r}$ for two words j and r.

6. Discussion

By considering the change in time of a word usage as a functional realization, we investigate how to estimate semantic networks via functional graphical models. Actually, by penalizing, we obtain sparse solutions that uncover interesting connections among variables/words.
In future developments, the choice of the smoothing parameter γ needs to be addressed and new strategies to preserve positive definitiveness in the estimates investigated. As for the Brexit debate analysis, the number of words p considered will be substantially increased.

References

[1] del Gobbo, E., Fontanella, S., Sarra, A., Fontanella, L.: Emerging Topics in Brexit Debate on Twitter Around the Deadlines. Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, **156(2)**, 669-688 (2021)

[2] Daw, R., Simpson, M., Wikle, C.K., Holan, S.H., Bradley. J.R.: An Overview of Univariate and Multivariate Karhunen Loéve Expansions in Statistics. J Indian Soc Probab Stat, **23**, 285-326 (2022)

[3] Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical Lasso, Biostatistics, **9(3)**, 432-441 (2008)

[4] Lee, K.Y., Ji, D., Li, L., Constable, T., Zhao, H: Conditional Functional Graphical Models, Journal of the American Statistical Association (2021) doi: 10.1080/01621459.2021.1924178

[5] Petersen, A., Deoni, S., Müller H.G.: Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. The Annals of Applied Statistics, **13(1)**, 393-419 (2019)

[6] Qiao, X., Guo, S., James, G.M.: Functional graphical models. Journal of the American Statistical Association, **114(525)**, 211-222 (2019)

[7] Yin, J., Geng, Z., Li, R., Wang, H.: Nonparametric covariance model. Statistica Sinica, **20(1)**,

469-479 (2010)
[8] Yuan, M., Lin, Y.: Model selection and estimation in the Gaussian graphical model. Biometrika **94(1)**, 19-35 (2007)