
Bayesian Logistic Gaussian Process Models for Dynamic Networks

Daniele Durante

Department of Statistical Sciences
University of Padua
durante@stat.unipd.it

David B. Dunson

Department of Statistical Science
Duke University
dunson@duke.edu

Abstract

Time-varying adjacency matrices encoding the presence or absence of a relation among entities are available in many research fields. Motivated by an application to studying dynamic networks among sports teams, we propose a Bayesian nonparametric model. The proposed approach uses a logistic mapping from the probability matrix, encoding link probabilities between each team, to an embedded latent relational space. Within this latent space, we incorporate a dictionary of Gaussian process (GP) latent trajectories characterizing changes over time in each team, while allowing learning of the number of latent dimensions through a specially tailored prior for the GP covariance. The model is provably flexible and borrows strength across the network and over time. We provide simulation experiments and an application to the Italian soccer Championship.

1 INTRODUCTION

There is an increasing availability of dynamic relational data allowing links between pairs of entities to change over time. Examples can be found in social network analysis, where friendship relations are monitored over time, and broader application settings in which interest focuses on forecasts and on inferences on the dynamic relational structure among pairs of units, such as countries, products, assets, and teams. Figure 1 shows an example of time-varying adjacency matrices constructed using Italian soccer Championship results. Specifically, each Championship is divided into

two seasons, and for every season, each team plays against all the others giving rise to a sequence of $V \times V$ time varying symmetric matrices $\{Y_t, t \in \mathcal{T} \subset \mathbb{R}^+\}$, with V the number of teams. The matrix Y_t has entries $y_{ij,t} = y_{ji,t} = 1$ if team i and team j tie the match at time t and $y_{ij,t} = y_{ji,t} = 0$, otherwise. There is a literature on statistical applications in soccer (see e.g., Dixon and Coles, 1997), but analyses via network models are lacking, while providing useful contributions for sport betting and results of interest to fans. We follow this direction of research by treating ties as relational data.

Spatial analysis of choice data (DeSarbo and Hoffman, 1987; DeSarbo et al., 1998) provides a possible framework for exploring binary matrices, with generalizations available also in the dynamic case (Sarkar et al., 2007; Xiong et al., 2010). Beside scalability, theoretical and computational issues, these models are specifically tailored for embedding problems in co-occurrence data encoding links between two different types of entities (i.e. author-words, etc.). Our focus is instead on dynamic modeling of one-mode binary matrices.

Traditional literature in social networks analysis via Exponential Random Graph Models (ERGM) (Erdős and Rényi, 1959; Holland and Leinhardt, 1981; Frank and Strauss, 1986), and generalizations for dynamic inference (Robins and Pattinson, 2001; Guo et al., 2007), represent overly restrictive approaches in modeling of one-mode binary matrices. Solutions can be degenerate (Handcock et al., 2003), and questions remain about tractability, coherence, flexibility and other key issues (Chatterjee and Diaconis, 2013).

Alternative classes of models provide node clustering methods via Stochastic Block Models (SBM) (Nowicki and Snijders, 2001) and Infinite Relational Models (IRM) (Kemp et al., 2006), with dynamic generalizations covering discrete dynamic evolution via hidden Markov models (Ishiguro et al., 2010) and continuous time analysis via extended Kalman filter (Xu and Xero, 2013). We instead embed the nodes in a low-dimensional latent Euclidean space, with coordinate

Appearing in Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS) 2014, Reykjavik, Iceland. JMLR: W&CP volume 33. Copyright 2014 by the authors.

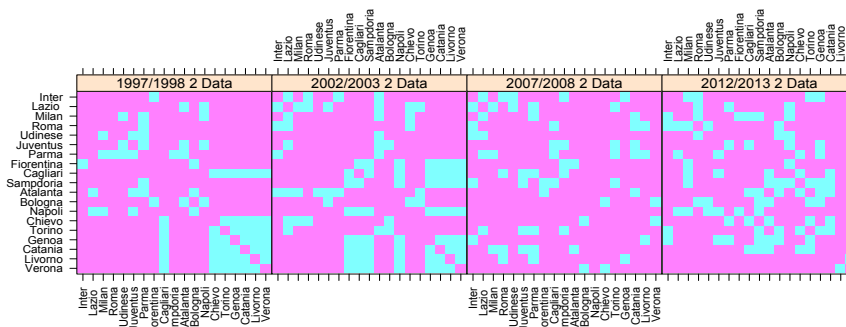


Figure 1: Example of time-varying adjacency matrices built using Italian Championship results. Purple color refers to 0 values and sky blue to 1 values.

trajectories equipped with GP priors, providing a contribution closer to the literature on latent space models (Hoff et al., 2002) and Mixed Membership Stochastic Block models (MMSB) (Airoldi et al., 2008), which allow each node to belong to multiple blocks with fractional membership. Dynamic latent space models (Sarkar and Moore, 2005) and MMSB models (Xing et al., 2010) rely on several layers of approximation without theory available to justify accuracy. Heaukulani and Ghahramani (2013) enforce latent feature propagation via hidden Markov chains, but the time constant assumption for the feature-interaction matrix and the equally spaced time grid evolution are not suitable in many applications. In contrast, we consider a simple Gibbs sampling algorithm for our model, which converges to the exact posterior, online learns the dimension of the latent space and allows an irregular grid of observations.

The paper is organized as follows. In Section 2, we describe the general model structure focusing on prior specification, properties and posterior computation. A simulation study is examined in Section 3, and an application to Italian soccer Championship results is presented in Section 4. The complete Gibbs sampling algorithm and proofs of theorems are provided in Durante and Dunson (2013).

2 DYNAMIC MODEL

2.1 Dynamic Model Formulation

We propose a nonparametric Bayesian dynamic model for relational data, which efficiently exploits a latent space representation of network data while incorporating time dynamics via Gaussian process latent factors. Specifically, letting Y_t be the symmetric adja-

ciency matrix at time $t \in \mathcal{T}$ and $\pi(t)$ be the corresponding symmetric probability matrix having entries $\pi_{ij}(t) = \text{pr}(y_{ij,t} = 1)$ for every $i = 1, \dots, V$ and $j = 1, \dots, V$, the model defines

$$y_{ij,t} | \pi_{ij}(t) \sim \text{Bern}(\pi_{ij}(t)) \quad t \in \mathcal{T}, \quad (1)$$

independently for each $i = 2, \dots, V$ and $j = 1, \dots, i - 1$, with

$$\text{E}[y_{ij,t} | \pi_{ij}(t)] = \pi_{ij}(t) = \frac{1}{1 + e^{-s_{ij}(t)}}, \quad (2)$$

for each $i = 2, \dots, V$ and $j = 1, \dots, i - 1$. To avoid modeling separately $\frac{1}{2}V(V - 1)$ stochastic processes, one for each time-varying similarity measure $s_{ij}(t)$, and to borrow strength across the network, $s_{ij}(t)$ is constructed via a quadratic combination of a set of latent coordinates for unit i and unit j . Specifically

$$s_{ij}(t) = \mu(t) + x_i(t)^T x_j(t), \quad (3)$$

with $x_i(t) = [x_{i1}(t), \dots, x_{iH}(t)]^T$ for $i = 2, \dots, V$ and $x_j(t) = [x_{j1}(t), \dots, x_{jH}(t)]^T$ for $j = 1, \dots, i - 1$, providing the following matrix factorization

$$S(t) = \mu(t)1_V 1_V^T + X(t)X(t)^T, \quad (4)$$

where $S(t)$ is a $V \times V$ real symmetric matrix with latent similarity entries $s_{ij}(t)$ and $X(t) = [x_1(t), x_2(t), \dots, x_V(t)]^T$.

As a result the link probabilities are modeled via a logistic regression, with $\mu(t)$ a baseline process characterizing the overall propensity to form links and $x_i(t)^T x_j(t)$ inducing a higher probability of a link (i.e., $y_{ij,t} = 1$) when i and j have latent coordinates in the same direction. This formulation is also intuitive in practical applications. Recall our motivating example of soccer, and assume for simplicity $\mu(t) = 0$ and only

two latent coordinates representing for example defense and scoring abilities, respectively. Then matches between teams with defense and scoring characteristics in the same directions (i.e. both high defense and scoring, both low defense and high scoring, etc.) will have a higher probability of ending with a tie, while teams with opposite abilities (i.e. high defense and scoring against low defense and scoring) will more likely lead to matches in which one of the two wins.

Such a representation provides also a provably flexible model in which the lower triangular matrix elements of any symmetric probability matrix $\pi(t)$ can be represented as in (2), with latent similarities specified as in (3); see Durante and Dunson (2013) for proofs. Non-uniqueness of matrix factorization in (4) is not an issue, since we are taking a Bayesian approach to inference and are not directly interested in the non-identifiable parameters (Bhattacharya and Dunson, 2011). We obtain identifiability in a Bayesian sense for the induced similarity matrix $S(t)$ and probability matrix $\pi(t)$, which are of interest for inference. By avoiding identifiability restrictions on parameters (Bollen, 1989; Hoff et al., 2002), we substantially improve simplicity and efficiency of computations.

2.2 Prior Specification

Our prior choice is motivated by computational tractability and a desire to obtain a nonparametric Bayesian approach, implying that the formulation is flexible in the sense of ‘large support’. This means that our prior can generate time-varying adjacency matrices within an arbitrarily small neighborhood of the ‘truth’, allowing the truth to fall in a wide class. We choose independent priors, Π_X and Π_μ , for $X_{\mathcal{T}} = \{X(t), t \in \mathcal{T}\}$ and $\mu_{\mathcal{T}} = \{\mu(t), t \in \mathcal{T}\}$ in order to induce a large support prior Π_π for $\pi_{\mathcal{T}} = \{\pi(t), t \in \mathcal{T}\}$ through (2) and (3). Automatic strategies for selecting the number of coordinates in latent space models are typically lacking or computationally sub-optimal (e.g., requiring re-running for different choices of dimension). We propose an extension of Bhattacharya and Dunson’s (2011) latent factor selection method from Gaussian latent factors to Gaussian process latent factors. Specifically, we let

$$x_{ih}(\cdot) \sim \text{GP}(0, \tau_h^{-1} c_X),$$

independently for all $i = 1, \dots, V$ and $h = 1, \dots, H$, with c_X a squared exponential correlation function $c_X(t, t') = \exp(-\kappa_X \|t - t'\|_2^2)$, which favors unequal spacing and continuous time analysis, and τ_h^{-1} a shrinkage parameter defined as

$$\tau_h = \prod_{k=1}^h \vartheta_k, \quad \vartheta_1 \sim \text{Ga}(a_1, 1), \quad \vartheta_k \sim \text{Ga}(a_2, 1), \quad k \geq 2.$$

Note that if $a_2 > 1$ the expected value for ϑ_k is greater than 1. Hence, as h goes to infinity, τ_h tends to infinity, shrinking $x_{ih}(\cdot)$, for every $i = 1, \dots, V$ towards zero. This provides a flexible prior for $x_{ih}(\cdot)$ with a local shrinkage parameter τ_h^{-1} that enforces many latent coordinates being close to 0 as h increases. To conclude prior definition, we choose

$$\mu(\cdot) \sim \text{GP}(0, c_\mu),$$

with $c_\mu(t, t') = \exp(-\kappa_\mu \|t - t'\|_2^2)$.

Such a specification can generate a time-varying symmetric probability matrix that is arbitrarily close to any function $\{\pi(t), t \in \mathcal{T}\}$. Intuitively, large support on continuous symmetric similarity matrix functions $\{S(t), t \in \mathcal{T}\}$ relies on the continuity of the Gaussian process coordinate functions. Since for each fixed $t = t_0$, $x_{ih}(t_0)$ are independently Gaussian distributed, $X(t_0)X(t_0)^T$ is distributed according to a sum of independent Wishart random variables. Combining the large support of the Wishart distribution with the one of the Gaussian for the baseline $\mu(t_0)$, provides large support for the induced prior Π_S . Since $\pi(t)$ is obtained via a one to one continuous increasing function of $S(t)$, we will map non-null probability subsets of the space of $S(t)$ into non-null probability subsets of the space of $\pi(t)$, providing the desired large support for the induced prior Π_π . Refer to Durante and Dunson (2013) for proof of large support property for Π_S , and Π_π . This ensures that our specification is sufficiently flexible to characterize any true generating process, and hence can be viewed as nonparametric given sufficiently flexible Gaussian process priors.

2.3 Posterior Computation

Posterior computation is performed exploiting a recently proposed data-augmentation scheme based on a new class of Pólya-Gamma distributions; see Polson et al. (2013) for a detailed description and Choi and Hobert (2013) for recent results on uniform ergodicity of the resulting algorithm. The approach provides a strategy for fully Bayesian inference exploiting the representation of binomial likelihoods parameterized by log-odds via a mixture of Gaussians with respect to Pólya-Gamma distributions.

Specifically assuming a Bayesian logistic regression setting where $y_i \sim \text{Bern}(1/[1 + e^{-\psi_i}])$, $i = 1, \dots, n$, $\psi_i = x_i^T \beta$ and β having Gaussian prior $\beta \sim N_p(b, B)$, the resulting Gibbs sampler alternates between two full conditional conjugate steps

$$\omega_i | \beta, x_i \sim \text{PG}(1, x_i^T \beta) \quad \text{and} \quad \beta | y, \omega, x \sim N_p(\mu_\beta, \Sigma_\beta),$$

where $\Sigma_\beta = (X^T \Omega X + B^{-1})^{-1}$ and $\mu_\beta = \Sigma_\beta (X^T z + B^{-1} b)$; with $z = [y_1 - 1/2, \dots, y_n - 1/2]^T$ and Ω is the diagonal matrix with ω_i ’s entries.

Recalling model (1), with probabilities defined as in (2) and latent similarities from (3), for $i = 2, \dots, V$, $j = 1, \dots, i - 1$ and $t \in \mathcal{T}_0 = \{t_1, \dots, t_T\}$, and taking a fixed truncation level H^* (see Bhattacharya and Dunson (2011) for a method adaptively choosing the truncation levels), the Gibbs sampler for our model alternatives between:

1. Update each augmented $\omega_{ij,t}$ from the full conditional Pólya-Gamma posterior.
2. Given $\{y_{ij,t}\}$, $X(t)$ and $\{\omega_{ij,t}\}$, the full conditional posterior distribution for $\mu(t)$ with $t \in \mathcal{T}_0 = \{t_1, \dots, t_T\}$ is a T -variate Normal distribution.
3. Update the time-varying latent coordinate vector $\{x_v(t) = [x_{v1}(t), \dots, x_{vH^*}(t)]^T\}_{t=t_1}^{t_T}$ for every unit $v = 1, \dots, V$ from its conditional posterior. Specifically, conditionally on $X^{(-v)} = \{x_j(t) : j \neq v, t \in \mathcal{T}_0\}$, $\mu = [\mu(t_1), \dots, \mu(t_T)]^T$, $\{y_{ij,t}\}$, $\{\omega_{ij,t}\}$ and $\{\tau_h\}$; we can easily rewrite the conditional model as a proper logistic regression with linear predictors, and utilize the standard results from Pólya-Gamma sampling scheme.
4. Conditioned on $X(t)$ and $\{\tau_h\}$, the global shrinkage hyperparameters are sampled from their corresponding full conditional Gamma distributions.
5. Conditionally on $\mu(t)$ and $X(t)$, latent similarities $s_{ij}(t)$ and link probabilities $\pi_{ij}(t)$ are updated from (3) and (2), respectively.

Missing at random values can be easily accommodated by adding a further step imputing the unobserved links from their conditional distribution given the current state of the chain. Specifically:

6. Given $X(t)$ and $\mu(t)$ sample each missing value from its conditional distribution (1) with probabilities specified as in (2) and latent similarity measures obtained via the projection approach defined in (3).

Step 6 provides also a strategy for predicting future networks. Specifically, if we are interested in making inference on $\pi(t_{T+1})$ with $t_{T+1} > t_T$ given the observed adjacency matrices Y_t , $t \in \mathcal{T}_0 = \{t_1, \dots, t_T\}$, then we can simply perform the previous posterior computations adding to the observed dataset $\{Y_t\}_{t \in \mathcal{T}_0}$ a new matrix $Y_{t_{T+1}}$ of missing values and make inference on the posterior predictive distribution using the samples of the Markov chain for $\pi(t_{T+1})$.

3 SIMULATION STUDY

We provide a simulation study to evaluate the performance of the proposed model in analyzing a dataset

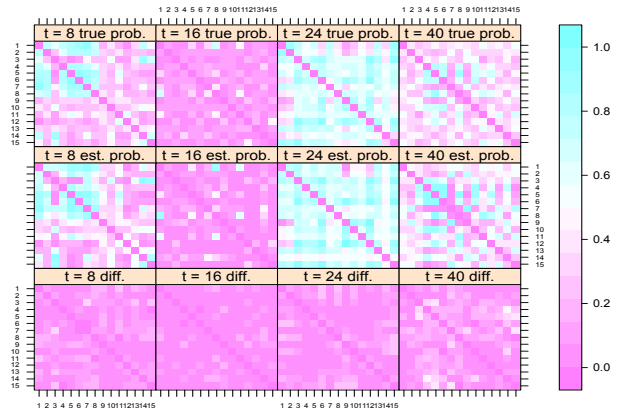


Figure 2: For some selected times t , plot of the true probability matrix $\pi(t)$ (top), posterior mean $\hat{\pi}(t)$ from our model (middle), and $|\pi(t) - \hat{\pi}(t)|$ (bottom).

constructed to mimic a possible generating process with respect to the soccer application. The focus is on the ability in correctly modeling the true underlying probability processes and also on the performance with respect to out of sample predictions. We also provide a comparison between our proposed approach and the estimated probability process for each time-varying binary outcome when using only temporal information without exploiting matrix structure, showing the suboptimality of the latter in terms of bias. Although our model is suitable for analyzing higher V and T settings, we focus for simplicity on a relatively small simulated dataset in which the results can be visualized.

3.1 Estimating Performance

We consider 15×15 time varying adjacency matrices Y_t with t in the discrete set $\mathcal{T}_0 = \{1, 2, \dots, 40\}$. Links $y_{ij,t}$ are simulated according to model (1) with probabilities obtained from (2) and latent similarities specified as in (3). We generate the baseline process $\{\mu(t)\}_{t=1}^{40}$ from a $\text{GP}(0, c_\mu)$ with length scale $\kappa_\mu = 0.01$, and consider 2 time-varying latent coordinates $\{x_{i1}(t)\}_{t=1}^{40}$, $\{x_{i2}(t)\}_{t=1}^{40}$ from Gaussian processes with length scale $\kappa_x = 0.01$, independently for each unit $i = 1, \dots, 15$. Data at the last time Y_{40} are held out in the estimation process to evaluate the out of sample predictive performance, and links between units 10 and 11 and all the others are considered missing at times $t = 20, \dots, 25$ to assess the behavior with respect to missing data. For inference we choose a truncation level $H^* = 10$, length scales $\kappa_\mu = \kappa_x = 0.05$ and set $a_1 = a_2 = 2$ for the shrinkage parameters. We ran 5,000 Gibbs iterations which proved to be enough for reaching con-

Table 1: Summary of the squared differences between true $\mu(t)$, $t = 1, \dots, 40$ and true $\pi_{ij}(t)$, $t = 1, \dots, 40$, $i = 2, \dots, V$, $j = 1, \dots, i - 1$ and their corresponding posterior mean estimated via our model.

	mean	0.95% quantile	max
$\mu(t)$	0.018	0.042	0.077
$\pi(t)$	0.002	0.011	0.023

Table 2: Summary of the squared differences between true $\mu(t)$, $t = 1, \dots, 40$ and true $\pi_{ij}(t)$, $t = 1, \dots, 40$, $i = 2, \dots, V$, $j = 1, \dots, i - 1$ and their corresponding posterior mean estimated via independent approach.

	mean	0.95% quantile	max
$\mu(t)$	0.700	2.744	3.016
$\pi(t)$	0.025	0.074	0.101

vergence and discarded the first 1,000. Mixing was assessed by analyzing the effective sample sizes of the MCMC chains for the quantities of interest (i.e. $\pi_{ij}(t)$, for $i = 2, \dots, V$, $j = 1, \dots, i - 1$ and $t \in \mathcal{T}_0$) after burn-in. Most of these values concentrate around $\approx 1,700$ effective samples on a total of 4,000, providing a good mixing result.

The comparison in Figure 2 between true probability matrices and their corresponding posterior mean for some selected time t , highlights the good performance of our model in correctly estimating the true underline data generating process and making predictions. The latter can be noticed by comparing true and estimated probability matrices at $t = 40$, recalling that in our simulation we allowed Y_{40} to have missing entries and we were interested in analyzing the predictive performance of our model with respect to $\pi(40)$. Good performance is further confirmed by an area underneath the ROC curve of 0.87.

Tables 1 and 2 compares the performance of our model with respect to $\mu(t)$ and $\pi(t)$, and the inferential results when the mean process and probability processes are estimated with the same setting of our model but using only the time series of the corresponding $y_{ij,t}$ without borrowing strength across the network. The sub-optimality of the independent approach is apparent in terms of biased estimates. Appropriately borrowing of information across the network provides substantially improved estimates, while accurately selecting the dimension of the latent space. In particular, we find that the estimated $\hat{\tau}_h^{-1}$ values start at 0.8 and 0.7 for $h = 1$ and 2, respectively, but drop to small values for $h \geq 3$, limiting the influence of later factor

trajectories. Additionally, hyperparameter sensitivity benefits from borrowing of information across the network over time, in particular with respect to the length scale in GP prior. We obtain similar results when instead letting $\kappa_\mu = \kappa_x = 0.03$, $\kappa_\mu = \kappa_x = 0.1$ and $\kappa_\mu = \kappa_x = 0.5$ in sensitivity analyses.

4 DYNAMIC SOCCER NETWORKS

We apply our model to Italian soccer Championships results from 1993 to 2013. Specifically, given the results in each match we focus on the time-varying Y_t matrices having entries $y_{ij,t} = y_{ji,t} = 1$ if team i and team j tie the match at time t (teams are similar), and $y_{ij,t} = y_{ji,t} = 0$ if one of the two wins the match (teams are dissimilar), and we are interested in the underlying dynamic network structure and prediction.

4.1 Data Analysis

Since for each Championship the last 3 teams are relegated in the lower category, and the first 3 teams of the lower category take their place, we choose the 19 teams that played most of the Championships in the time window considered (i.e. $V = 19$). The total number of Championships considered is 20, giving rise to a time grid $t \in \mathcal{T}_0 = \{1, 2, \dots, 41\}$ since each one is divided in two seasons, and $t = 41$ is associated to a matrix Y_{41} of missing values used for predicting the outcomes in the first season of the 2013/2014 Championship. Finally, given the non-negligible number of missing values, and since most of them are not missing at random (i.e. if team i is in the lower category at time t , then $y_{ij,t}$ is missing for all $j \neq i$), we prefer to impute such values according to our model interpretation. Specifically if $y_{ij,t}$ is missing and only one of the two teams is in the lower category, we define $y_{ij,t} = 0$ assuming the team in the lower category is weaker than those playing in the top category, thus a possible match is more likely to end with the top one winning it. If at time t both teams are in the lower category, we impute $y_{ij,t} = 1$ assuming them to be more similar and therefore more likely to tie the match. We could have alternatively defined our model and inference algorithm to allow teams to move in and out of the network, but prefer to avoid the resulting complexity in our descriptions.

4.2 Results

We apply model (1), with probabilities specified as in (2) and latent similarity measures obtained via the projection approach in (3). For posterior computation we run 5,000 Gibbs iterations with a burn-in of 1,000, setting $H^* = 15$, length scales $\kappa_\mu = \kappa_x = 0.03$

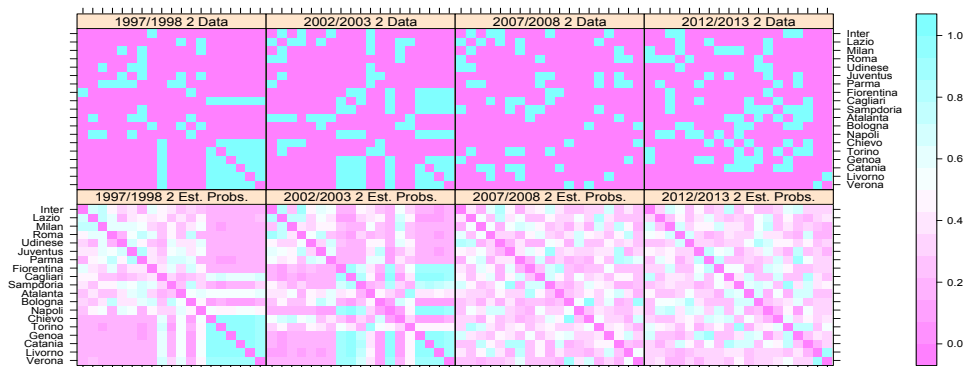


Figure 3: For some selected seasons, plot of the observed data matrices Y_t (top), and posterior mean $\hat{\pi}(t)$ (bottom).

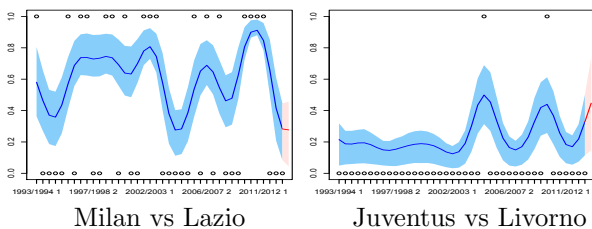


Figure 4: For selected matches, plot of the observed time varying outcomes (black dots), point-wise posterior mean $\hat{\pi}_{ij}(t)$ (blue lines) and 0.95 highest posterior density (hpd) intervals (colored blue areas). Red lines and areas are the same quantities for the posterior predictive distribution.

and $a_1 = a_2 = 2$. Similarly to the simulation study, most of the chains have effective sample sizes around 2,000 on a total of 4,000 after burn-in, showing good mixing. We find that the first six latent factors are the most informative, with the remaining 9 latent processes being concentrated near zero. It was interesting to us that this many latent factors were needed.

Figure 3 shows a graphical comparison between the observed data matrices Y_t and the estimated probability matrices $\hat{\pi}(t)$, for some seasons, showing the good performance of the proposed model in adaptively learning the data structure, confirmed also by a ROC curve having an area underneath of 0.94. Similar results are provided in Figure 4 showing the time-varying point-wise posterior mean and 0.95 hpd intervals for some selected matches. It is worth noticing that the local adaptivity of the estimated trajectories is not due to an over-parameterization of the model since the shrinkage prior on τ_h and the choice of small length scales in the GP covariance functions imply smooth trajectories and a parsimonious model formulation. Thus adaptivity is provided by the information borrowed in the network for each time t .

Note also that, as expected, when similar teams such as Milan vs Lazio (both top teams) play, tying probabilities seem to be in general higher, with cycles due to the fact that in some specific seasons a team may be stronger than the corresponding opponent. On the other hand, matches between a top team (i.e Juventus) and a low team (i.e Livorno) have a higher probability of not ending with a tie, as expected.

Figure 5 provides interesting insight into the network structure among Italian soccer teams. Specifically we represent three different weighted networks, with weights given by (a) the average estimated tying probability $\pi_{ij}^* = \sum_{t=1}^{40} \hat{\pi}_{ij}(t)/40$, (b) the predicted tying probability for the new season $\pi_{ij}^{**} = \hat{\pi}_{ij}(t = 41)$ and (c) the probability π_{ij}^{***} obtained by averaging over Championships 2004/2005 and 2005/2006, related to the betting scandal in Italian soccer, which caused penalties for the teams supposed to be most involved (i.e. Juventus, Lazio, Fiorentina and Milan). A reasonable global network structure with teams having similar strength most closely related among each other and relative positions respecting the overall ranking of the teams is provided in plot (a). Note how the relative position of Naples changes when considering the predicted network in (b) which assigns more weight to the recent Championships when Naples had a top ranking performance, differently from older seasons in which they were also relegated in lower categories. Finally, analyzing the network during Championships 2004/2005 and 2005/2006 in (c), we interestingly learn close relationships among the teams supposedly most involved in the betting scandal in those seasons.

5 DISCUSSION

We proposed a Bayesian nonparametric dynamic model for adjacency matrices, borrowing information

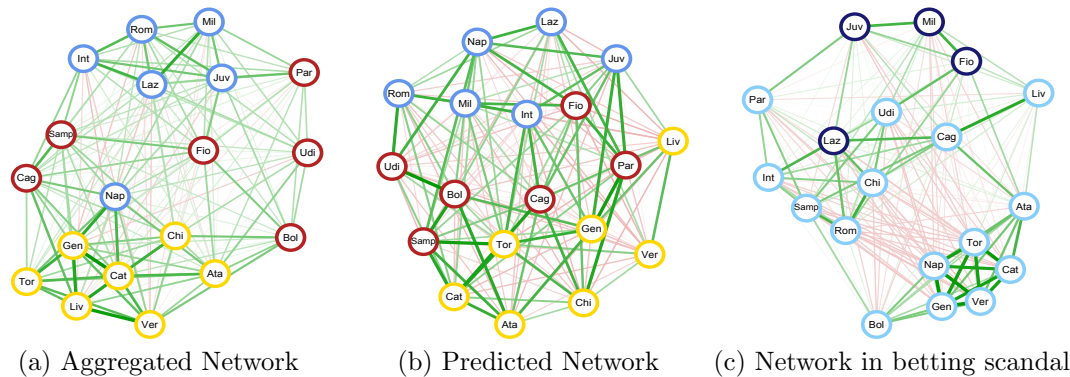


Figure 5: Left: weighted network visualization with weights obtained averaging $\hat{\pi}(t)$ over \mathcal{T}_0 . Middle: Predicted weighted network visualization for the first season of 2013/2014 Italian soccer Championship with weight matrix given by the forecasted $\pi(t)$. Right: weighted network visualization with weights obtained averaging $\hat{\pi}(t)$ over Championships 2004/2005 and 2005/2006. Edge dimensions are proportional to the corresponding value of the averaged probability matrix, with colors going from red to green as the corresponding weight goes from 0 to 1. Nodes are colored with respect to prior knowledge on teams strength (Left and Middle plots), and penalties received after the betting scandal during 2004/2005 and 2005/2006 Championships (Right plot).

across time and the network structure of the data under investigation and allowing for dimensionality reduction. The model has been constructed using latent similarity measures defined by the dot product of latent coordinate vectors, with entries evolving in continuous time via Gaussian process priors. The shrinkage hyperprior allows us to automatically learn the dimension of the latent space and ensures a parsimonious definition of the model, with the risk of overparameterization due to a higher number of latent features avoided. The Pólya-Gamma data augmentation strategy allows us to define a simple and efficient Gibbs sampler for posterior computations based on full conditional conjugate posterior distributions, which is promising in terms of scaling to moderately large V , and easily handling missing values as well as forecasting problems. Scalability to large T could be, instead, improved via stochastic differential equations models approximating the GP prior on the latent coordinate processes (Zhu and Dunson, 2013).

Our model has a broad range of applicability, with dynamic social network analysis and time-varying binary evaluations among units providing two natural fields of application. Further directions of research could be devoted to the definition of similar models for discrete valued dynamic matrices, which could provide useful tools for analyzing the scoring difference in sport applications, as well as edge valued dynamic social networks or datasets with comparison among units expressed on a Likert scale.

References

E.M. Airoldi, D.M. Blei, S.E. Fienberg and E.P. Xing (2008). Mixed Membership Stochastic Blockmodels. *Journal of Machine Learning Research* 9:1981–2014.

A. Bhattacharya and D.B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika* 98: 291–306.

K.A. Bollen (1989). *Structural Equations with Latent Variables*. New York: Wiley.

S. Chatterjee and P. Diaconis (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics* 41: 2263–2702.

H.M. Choi and J.P. Hobert (2013). The Poly-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electronic Journal of Statistics* 7: 2150–2163.

W.S. DeSarbo, Y. Kim and D. Fong (1998). A Bayesian multidimensional scaling procedure for the spatial analysis of revealed choice data. *Journal of Econometrics* 89:79–108.

W.S. DeSarbo and D.L. Hoffman (1987). Constructing MDS joint spaces from binary choice data: A new multidimensional unfolding threshold model for marketing research. *Journal of Marketing Research* 24:40–54.

M.J. Dixon and S.G. Coles (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46:265–280.

D. Durante and D.B. Dunson (2013). Nonparametric Bayes dynamic modeling of relational data. arxiv.org/abs/1311.4669.

- P. Erdős and A. Rényi (1959). On Random Graphs. *Publicationes Mathematicae* 6:290–297.
- O. Frank and D. Strauss (1986). Markov Graphs. *Journal of the American Statistical Association* 81:832–842.
- F. Guo, S. Hanneke, W. Fu and E.P. Xing (2007). Recovering temporally rewiring networks: A model-based approach. In *Proceedings of the International Conference in Machine Learning*.
- M.S. Handcock, G.L. Robins, T.A.B. Snijders, J. Moody and J. Besag (2003). Assessing Degeneracy in Statistical Models of Social Networks. *Journal of the American Statistical Association* 76:33–50.
- C. Heaukulani and W. Ghahramani (2013). Dynamic probabilistic models for latent feature propagation in social networks. In *Proceedings of the International Conference in Machine Learning*.
- P.D. Hoff, A.E. Raftery and M.S. Handcock (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association* 97:1090–1098.
- P.W. Holland and S. Leinhardt (1981). An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association* 76:62–65.
- K. Ishiguro, T. Iwata, N. Ueda and J. Tenenbaum (2010). Dynamic infinite relational model for time-varying relational data analysis. In *Advances in Neural Information Processing Systems*.
- C. Kemp, J.B. Tenenbaum, T.L. Griffiths, T. Yamada and N. Ueda (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the National Conference on Artificial Intelligence*.
- K. Nowicki and T.A.B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96:1077–1087.
- N.G. Polson, J.G. Scott and J. Windle (2013). Bayesian inference for logistic models using Poly-Gamma latent variables. *Journal of the American Statistical Association* (forthcoming).
- G.L. Robins and P.E. Pattison (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology* 25:5–41.
- P. Sarkar, S.M. Siddiqi and G.J. Gordon (2007). A latent space approach to dynamic embedding of co-occurrence data. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*.
- P. Sarkar and A.W. Moore (2005). Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*.
- E.P. Xing, W. Fu, W. and L. Song (2010). A State-Space Mixed Membership Blockmodel for Dynamic Network Tomography. *The Annals of Applied Statistics* 4:535–566.
- L. Xiong, X. Chen, T.Z. Huang, J. Schneider and J.G. Carbonell (2010). Temporal Collaborative Filtering with Bayesian Probabilistic Tensor Factorization. In *proceedings of SIAM Data Mining*.
- S.K. Xu and A.O. Hero III (2013). Dynamic stochastic blockmodels: Statistical models for time-evolving networks. *arxiv.org/abs/1304.5974*.
- B. Zhu and D.B. Dunson (2013). Locally adaptive Bayes nonparametric regression via nested Gaussian processes. *Journal of the American Statistical Association* 108:1445–1456.