Modeling Dynamic International Relationship

Yunran Chen

1. Introduction

I am interested in modeling dynamic networks, where each edge between pairs of nodes can be seen as a time series. In this project, I am trying to fit a dynamic international relationship to explore the time-varying relational structure instead of making prediction.

For social networks, the latent space model proposed by Hoff *et al.* (2002 [6], 2005 [5]) is widely used. The basic idea is to introduce a multiplicative random effect to the mix effect model. Such multiplicative random effect can induce a third dependence structure, which accommodates some typical dyadic characteristics, such as homophily and stochastic equivalence ("friends of my friends are my friends", "enemies of enemies are my friends") (Hoff 2008, [4]). However, the latent space model is for static networks. Durante and Dunson(2014 [3]) extended this model to a model for dynamic networks evolving in continuous time by setting Gaussian process for each element. Instead of considering continuous time, I consider discrete time series. Explicitly, I assume each element as a stationary AR(1) process. First, I explored the properties of product of two stationary AR(1) processes. Based on this, I derived Gibbs sampling scheme with Pólya-Gamma augmentation for posterior computation. Then I did extensive simulations on the choice of parameters and applied the proposed model to analyze the real dataset. At the end I provide detailed discussion on the limitations and further directions.
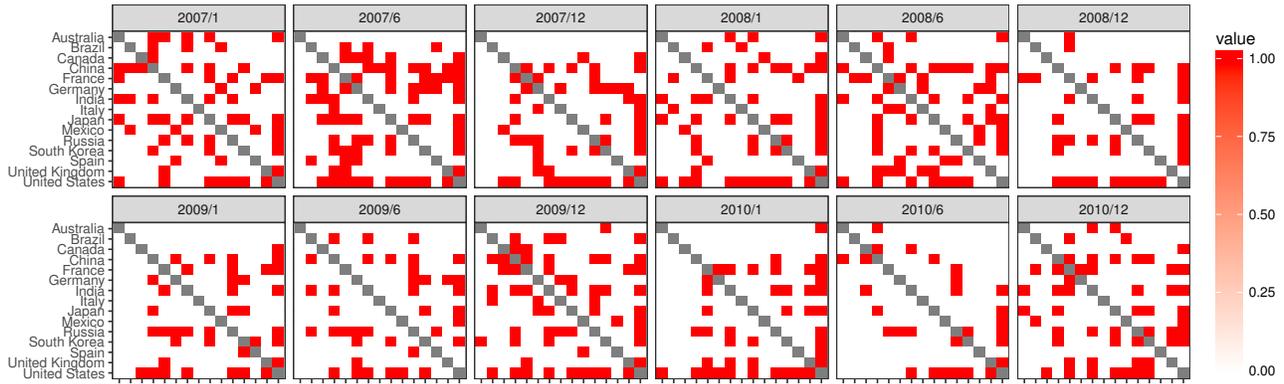
Figure 1. Dynamic international relationship at selected time

2. Modeling

2.1. Model specification

Consider a series of symmetric binary matrices $Y_t, t \in \{1, ..., T\}$, with corresponding probability for each entry denoted as $\pi_{ij,t} = Pr(y_{ij,t} = 1)$ for $i, j = 2, ..., V; i > j$. I consider logistics regression and only include main effect μ_t and multiplicative latent factor $x_{i,t}$.

$$y_{ij,t}|\pi_{ij,t} \sim Bernoulli\{\pi_{ij,t}\}; \qquad \pi_{ij,t} = \{1 + e^{-S_{ij,t}}\}^{-1}; \qquad S_{ij,t} = \mu_t + x_{i,t}^T x_{j,t} \qquad (1)$$

where μ_t is a scalar at time t, indicating the main effect for all the individuals at time t; where $x_{i,t} = \{x_{i1,t}, ..., x_{iH,t}\}^T$ for $i, j = 2, ..., V; i > j$ are multiplicative vectors indicating the location of "latent space" for individual i, with H indicates the dimension of the latent space.

Here I assume each element μ, x_{ih} where $i = 2, ..., V; h = 1, ..., H$ evolves via stationary AR(1) process separately.[1] Since the policy of a country is consistent in real world, I only consider positive value of ϕ, which means $0 < \phi < 1$.

$$\mu_t = \phi_\mu \mu_{t-1} + \epsilon_{\mu,t}, \qquad \epsilon_{\mu,t} \sim N(0, v_\mu). \qquad (2)$$

$$x_{ih,t} = \phi_{x_{ih}} x_{ih,t-1} + \epsilon_{x_{ih},t}, \qquad \epsilon_{x_{ih},t} \sim N(0, v_{x_{ih}}). \qquad (3)$$

The multiplicative effect in equation 1 is a summation of a series of product of two stationary AR(1) processes. In order to design a sampling scheme and have more insights on the model, I explored the properties of product of two AR(1) processes.

2.2. Properties of product of two AR(1) processes

Consider two independent AR(1) processes $x_{1t} \leftarrow AR(1|(\phi_1, v_1))$ and $x_{2t} \leftarrow AR(1|(\phi_2, v_2))$. The marginal distribution of x_{1t} and x_{2t} are normal distribution $N(0, s_1)$ and $N(0, s_2)$ respectively, where $s_1 = v_1/(1 - \phi_1^2)$ and $s_2 = v_2/(1 - \phi_2^2)$. We are interested in the properties of z_t where $z_t = x_{1t}x_{2t}$.

z_t **is stationary.**

$$E(z_t) = 0, \quad Var(z_t) = s_1 s_2, \quad Corr(z_t, z_{t-k}) = (\phi_1 \phi_2)^k. \qquad (4)$$

Since x_{1t} and x_{2t} are stationary processes, the product of two stationary processes will still be stationary. We can calculate expectation $E(z_t)$ using the formula $E(XY) = E(X)E(Y)$ for independent X, Y; calculate $Var(z_t)$ using the formula $Var(XY) = E(X^2Y^2) - (E(XY))^2 = Var(X)Var(Y) + Var(X)(E(Y))^2 + Var(Y)(E(X))^2$ for independent X, Y; calculate $Corr(z_t)$ using law of total expectation $E(X) = E(E(X|Y))$.

Consider $\phi_1 = \phi_2 = 0.9, v_1 = v_2 = 1$, I simulated a z_t series. As Figure 3 shows, z_t is also stationary and centered at 0. However, z_t has higher volatility and we can see more spikes at a high level compared to the x_{1t} and x_{2t}. And the dependence across time of z_t decays much faster than x_{1t} and x_{2t} do.

z_t **is non-Markovian.**

Figure 2 captures the relationship between x_{1t}, x_{2t}, z_t. Since the series x_{1t} is correlated and so is x_{2t}, when we marginalize x_{1t} and x_{2t}, the z_t will be non-Markovian. This property restricts the sampling schemes for posterior computation.
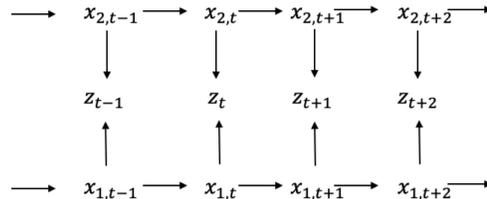
Figure 2. Directed acyclic graph (DAG) for z_t

---

[1]The choice of AR(1) process is for the simplicity of modeling.
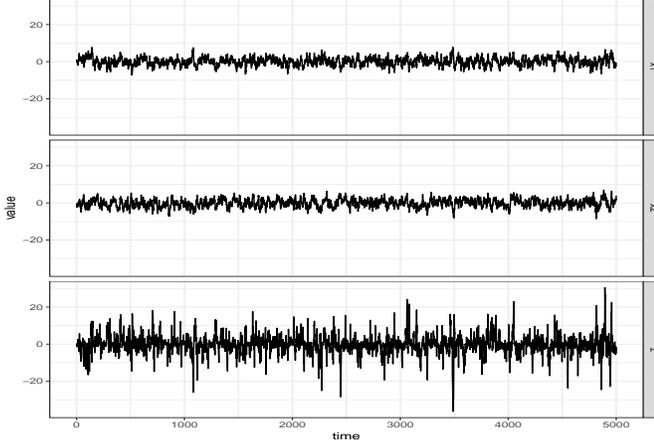
Figure 3. A trajectory of x_{1t}, x_{2t} and corresponding z_t
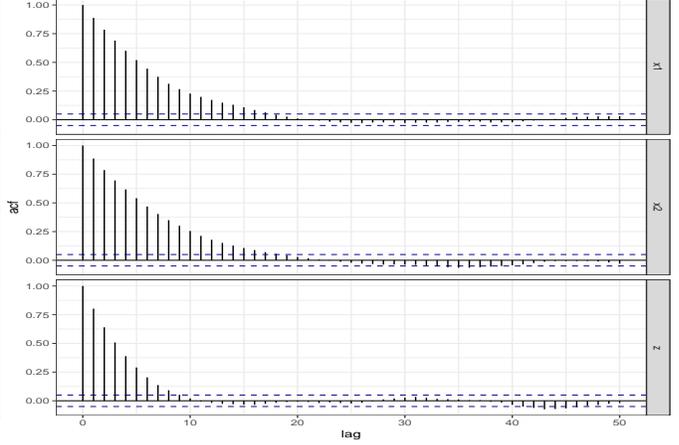
Figure 4. Correlation plot for z_t

We cannot identify x_{1t} and x_{2t}, but z_t is identifiable. For any non-zero constant c, $z_t = (x_{1t} \times c) \times (x_{2t}/c) = x_{1t}x_{2t}$. Therefore, given z_t, x_{1t} and x_{2t} are not identifiable.[2] However, in this project, since I only focus on inference on z_t instead of inference on x_{1t} and x_{2t}, this non-identifiability is not a issue here.

Gibbs sampling is needed. Since the z_t is the product of two normal distributed random variables, we cannot derived posterior of x_{1t} and x_{2t} given z_t directly. But given known x_{1t}, we can directly derive the full conditional distribution of x_{2t} given z_t and x_{1t}. Therefore, I apply a Gibbs sampling. I sample x_{1t} from $p(x_{1t}|x_{2t}, z_t)$ and x_{2t} from $p(x_{1t}|x_{2t}, z_t)$ then get $z_t = x_{1t}x_{2t}$. In section 2.4, I show how to apply Gibbs sampling in logistic regression with Pólya-Gamma data augmentation.

2.3. Choice of ϕ and v

Here I provide some insights on how the values of ϕ and v effect the π_t in the model

$$logit(\pi_t) = x_{1t}x_{2t}$$

where $x_{1t} \leftarrow AR(1|(\phi_1, v_1))$ and $x_{2t} \leftarrow AR(1|(\phi_2, v_2))$. A better understanding on the meaning of ϕ and v will help us to set a reasonable prior for our model.

For AR(1) process x_t, v controls the scale of volatility. Explicitly, a larger value of v lead to larger variance of marginal distribution of x_t and larger covariance of x_t across time. Moreover, such influence is boosted by the value of ϕ, since the marginal distribution of x_t is $N(0, s)$ where $s = v/(1 - \phi^2)$. A larger value of ϕ leads to a larger value of s. Apart from the "boosting" effect on volatility, ϕ also controls the dependence across time. The correlation between x_t and x_{t-k} is ϕ^k. With a larger ϕ, the dependence is larger.

For the product of AR(1) processes, equation 4 shows ϕ and v of each AR(1) process add multiplicative effects on the distribution of z_t. The variance and the covariance of z_t is a function of product of ϕ_1, ϕ_2 and product of s_1, s_2. With larger v_1, v_2, the volatility of z_t is larger. With larger ϕ_1, ϕ_2, the volatility of z_t and the dependence across time are larger.

From z_t to π_t, I use the logistics function. As Figure 5 shows, those points with values close to 0 will be mapped to points centered around 0.5, while those points with values far from 0 will be pushed to the boundary $0, 1$. Notice the logistics mapping is sensitive to area around 0 but flat for area outside 5. Therefore, z_t with a distribution centered at 0 with less volatility will be mapping to π_t with distribution concentrated at 0.5. Otherwise, z_t with a flat distribution spreading within a wider range will be mapping to π_t with higher probability on both 0 and 1. The range of z_t is controlled by s. This means a large s will push the π_t centered at two extreme values with large probability.

---

[2]Consider a matrix case in equation 1, for any orthogonal matrix Q, $Z_t = X_t Q Q^T X_t^T = X_t X_t^T$. Z_t is non-identifiable.
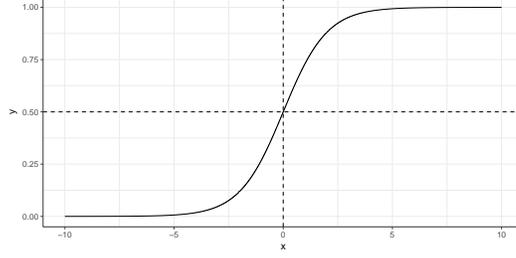
Figure 5. Logistic function
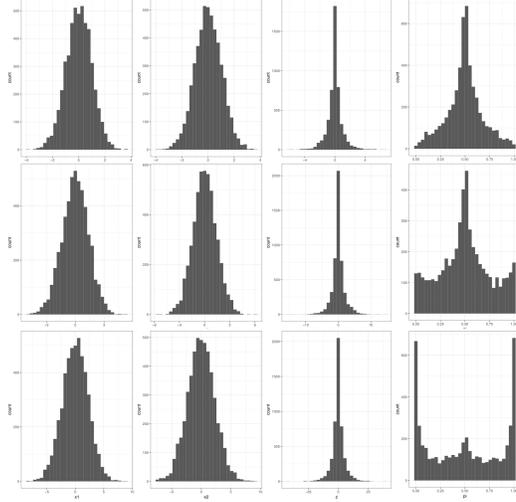
Figure 6. Histograms of marginal distributions of x_{1t}, x_{2t}, z_t and π_t. From left to right, each column indicates histograms for series x_{1t}, x_{2t}, z_t and π_t. From top to bottom, each row indicates the values of $\phi = 0.1, 0.5, 0.7$ and fixed $v = 1$.

Figure 6 shows the marginal distribution of x_{1t}, x_{2t}, z_t, π_t under different values of ϕ and same value of v. We can see as ϕ increases, the variance s increases, leading to the distributions of x_{1t} and x_{2t} spread within a larger range. (Notice that the scale for each histogram is different.) Accordingly, the z_t will be more flat within a much wider range. After the logistic mapping, the points around 0 are mapped to the points around 0.5, while the points with absolute value greater than 5 are pushed to the the boundary $0, 1$. Therefore, with increasing ϕ, we see the shape of distribution of π_t changing from one peak centered at 0.5 to two peaks centered at the boundary. This inspired us to choose s within a relative small range to provide a non-informative prior for π. Also, since the effect of v on the series is always combined with $1 - \phi^2$, we consider (ϕ, s) instead of (ϕ, v) for the distribution specification.

2.4. Gibbs sampling scheme with Pólya-Gamma augmentation

Since we do not focus on the identification on $x_{i,t}$, there is no need to set different ϕ for each component of latent factor. Assume each component of latent factor share the same hyperparameter ϕ_x, v_x, $x_{ih,t} \leftarrow AR(1|(\phi_x, v_x))$, where $i, j = 2, ..., V; i > j; h = 1, ..., H^*$. Notice that H^* denotes the estimation of unknown dimension of latent factor, which is considered as a hyperparameter. For μ_t, assume $\mu_t \leftarrow AR(1|(\phi_\mu, v_\mu))$. Then we have the stationary joint distribution for μ_t and $x_{ih,t}$ as follows:

$$(\mu_1, ..., \mu_N))^T \sim N(0, c_\mu), \qquad \text{where} \quad c_{\mu,ij} = s_\mu \phi_\mu^{|t-t'|}, \quad s_\mu = v_\mu/(1 - \phi_\mu)$$

4

$$(x_{ih,1}, ..., x_{ih,N})^T \sim N(0, c_x), \qquad \text{where} \quad c_{x,ij} = s_x \phi_x^{|t-t'|}, \quad s_x = v_x/(1 - \phi_x)$$

For a logistic regression with normal distribution as a prior, we can apply a Pólya-Gamma data augmentation for posterior computation (Polson *et al.*,2013 [7]). The Gibbs sampling scheme shows as follows.

1. Update $w_{ij,t}$. $w_{ij,t}|- \sim PG\{1, \mu_t + \sum_{h=1}^{H^*} x_{ih}^T x_{jh}\}$

2. Update μ. $\mu|- \sim N(\mu_\mu, \Sigma_\mu)$, where $\Sigma_\mu = \{diag(\sum_i \sum_{j \neq i} w_{ij,1}, ..., \sum_i \sum_{j \neq i} w_{ij,N}) + c_\mu^{-1}\}^{-1}$

$$\mu_\mu = \Sigma_\mu \begin{bmatrix} \sum_i \sum_{j \neq i}\{y_{ij,1} - 0.5 - w_{ij,2}(x_{i,1}^T x_{j,1})\} \\ \vdots \\ \sum_i \sum_{j \neq i}\{y_{ij,N} - 0.5 - w_{ij,N}(x_{i,N}^T x_{j,N})\} \end{bmatrix}$$

3. Update $X^{(v)}$. Consider $X^{(v)} = \{x_{v1,1}, ..., x_{v1,N}, ..., x_{vH^*,1}, ..., x_{vH^*,N}\}^T$, we have
$X^{(v)}|- \sim N(\mu_{X^{(v)}}, \Sigma_{X^{(v)}})$, where $\Sigma_{X^{(v)}} = \{\tilde{X}^{(-v)T} \Omega^{(v)} \tilde{X}^{(-v)} + K_{X^{(v)}}^{-1}\}^{-1}$

$$\mu_{X^{(v)}} = \Sigma_{X^{(v)}}[\tilde{X}^{(-v)T}\{y^{(v)} - 1_{N(V-1)}0.5 - \Omega^{(v)}(1_{2(V-1)} \otimes \mu)\}]$$

where $y^{(v)} = \{y_{v1,1}, \ldots, y_{v1,N}, \ldots, y_{vV,N}\}^T$, $\Omega^{(v)} = diag(\{w_{v1,1}, \ldots, w_{v1,N}, \ldots, w_{vV,N}\}^T)$.

$$\tilde{X}^{(-v)} = \begin{pmatrix} x_{11,1} & x_{12,1} & \cdots & x_{1H^*,1} \\ \ddots & \ddots & \cdots & \ddots \\ x_{11,N} & x_{12,N} & \cdots & x_{1H^*,N} \\ \vdots & \vdots & \cdots & \vdots \\ x_{V1,1} & x_{V2,1} & \cdots & x_{VH^*,1} \\ \ddots & \ddots & \cdots & \ddots \\ x_{V1,N} & x_{V1,N} & \cdots & x_{VH^*,N} \end{pmatrix}$$

$$K_{X^{(v)}}^{-1} = \{diag(1, \ldots, 1) \otimes c_x^{-1}\}$$

3. Simulations

I conduct simulations to evaluate the model's performance on fitting and prediction based on ROC-AUC criterion. I also include sensitivity analysis on the choice of the dimension of latent space H^*, and $\phi_x, \phi_\mu, v_x, v_\mu$ for the stationary AR(1) process. And more discussion on learning of these parameters can be found in section 5.

I generated dynamic matrices $Y_t, t \in \{1, ..., 40\}; i, j = 2, ..., 15; i > j$ as equation 1 describes. In addition, I set the whole matrix Y_{40} as missing values to evaluate the performance for prediction. I consider $\phi_x = \phi_\mu = 0.9$, $s_x = s_\mu = 1$ and $H = 2$ when generating the data. I consider the prior with $\phi_x' = \phi_\mu' = 0.8$, $s_x = s_\mu = 1$ and $H^* = 10$ for posterior sampling.

I ran 5000 iterations with 1000 burn-in. The effective sample size of $\pi_{ij,t}$ is around 2500, showing a good mixing. Figure 7 and figure 8 suggest good performance of model fitting. Figure 7 shows the comparison between $\pi_{ij,t}$ and $\hat{\pi}_{ij,t}$. We can see the estimators capture the structural characteristics well and the difference is relative small. From figure 8, we can see most of the points are gathering around the $y = x$ line. The AUC for estimation and prediction is 0.9885 and 0.6822 respectively. Figure 9 presents graphical summaries for trajectories of sampled μ_t and $\pi_{58,t}$. We can see that the posterior mean nearly overlaps with the truth and captures the trend well across

time. And most of the time the truth is covered by the 95% credible interval. Notice that the performance of prediction of proposed model is poor. I set the $H^* = 10$ which may result in overfitting [3]. One way to avoid overfitting is to add a shrinkage prior to restrict the effect of the latent factors (Durante and Dunson, 2014 [3]).
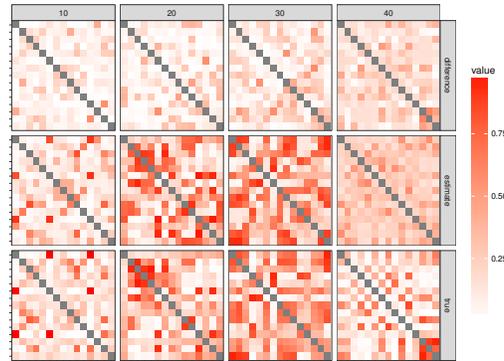
Figure 7. Model performance at selected time. From top to bottom, each panel indicates the absolute difference $|\pi_t - \hat{\pi}_t|$, estimate values $\hat{\pi}_t$ and true values π_t respectively. From left to right, each panel indicate matrices at the different time point $t = 10, 20, 30, 40$. Notice the Y_{40} is held out for prediction.
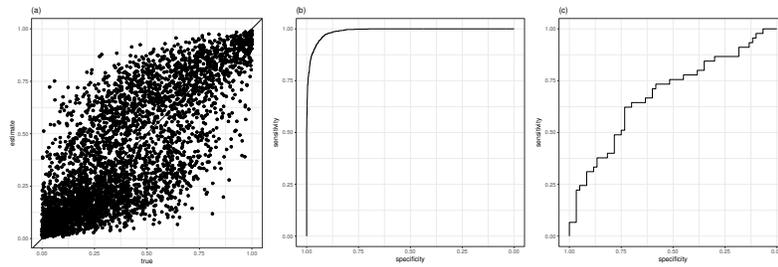
Figure 8. Model performance on fitting and prediction. (a)Scatter plot for estimator $\hat{\pi}_{ij}$ versus true π_{ij} including all data. (b)ROC curve only for fitting values. (c)ROC curve only for prediction values.
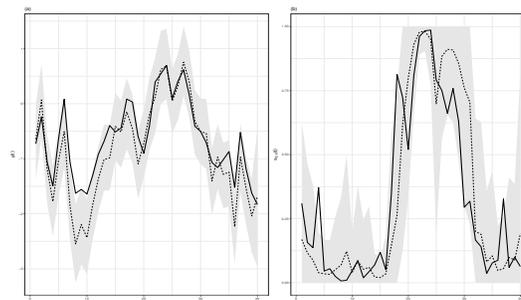
Figure 9. Graphical summary of trajectories of sampled $\hat{\mu}$ and $\hat{\pi}_{5,8}$. The solid line and dash line represent true and estimated values respectively. The grey ribbon represents 95% highest posterior density intervals.

Furthermore, I conducted extensive sensitivity analysis as table 1 shows. The simulations with ID 1-5 provide baselines for the AUC under correctly specified parameters, from which we can see the AUC still fluctuate from

---

[3]In the sensitivity analysis, for correctly specified H, the AUC for prediction is around 0.86.

0.78 to 0.95. Under the well specified parameters, a larger value of ϕ brings better performance on fitting but worse performance on prediction. Since a larger value of ϕ brings more dependency across time, the model will learn from past and show less flexibility for fitting. In contrast, making use of past information helps the model to predict more accurately. Therefore, the choice of ϕ is a trade-off between the accuracy of estimation and prediction. For the s, a relative small value indicate a large number of $\pi_{ij,t}$ gathering around 0.5, which contains little information. If the truth contains little information, it is hard to learn from the data even if the parameters are correctly specified. Therefore, we see the poor performance for both model fitting and prediction. Notice that if the s is too large, trying to fit the data is meaningless since most of $\pi_{ij,t}$ will be around 0 or 1.

ID	ϕ_x,ϕ_μ	s_x,s_μ	ϕ'_x,ϕ'_μ	s'_x,s'_μ	H^*	AUC(fitting)	AUC(prediction)
1	0.9	1	0.9	1	2	0.8692	0.6395
2	0.8	1	0.8	1	2	0.8911	0.6919
3	0.99	1	0.99	1	2	0.8615	0.8541
4	0.9	2	0.9	2	2	0.9463	0.842
5	0.9	0.5	0.9	0.5	2	0.776	0.5385
6	0.9	1	0.9	1	5	0.92	0.5279
7	0.9	1	0.9	1	10	0.958	0.6682
8	0.9	1	0.8	1	10	0.9885	0.6822
9	0.9	1	0.99	1	10	0.8352	0.6661
10	0.9	1	0.8	0.5	10	0.949	0.5322
11	0.9	1	0.99	0.5	10	0.8048	0.5364
12	0.9	1	0.99	5	10	0.9218	0.7046

Table 1. Sensitivity of parameter setting. ID 1-5 act as a baseline to see the fluctuation of the AUC under correctly specified parameters. ID 1,6,7 are for comparison of the choice of H^*. ID 7,8,9 are for comparison of the mis-specified ϕ. ID 8,10 and ID 9,11,12 are for comparison of the mis-specified s.

Generally, larger H^* brings more flexibility to the model, leading to better performances of model fitting. As a trade-off, the model with large H^* may take a risk of overfitting, leading to a poor performance for prediction. Similar trade-off for the choice of ϕ. With a smaller value of ϕ, the model relies less on the past and thus is more flexible, leading to a better performance for fitting. In contrast, the less information learned from the past leading to a worse performance for prediction. For the choice of s, small value brings non-informative prior, which may be inefficient for parameter estimation. Large value of s may lead to wrong estimation with higher confidence.

For a good performance of model fitting, I suggest to choose a relative large H^*, a moderate ϕ for flexibility and a relative small s for a non-informative prior.

4. Real Data Analysis

My goal is to fit a social network of international cooperative actions between the world's top 15 largest economies ranked by nominal GDP from 2007 to 2010. I extracted the event "Sign formal agreement"(CAMEO code 57) from Integrated Crisis Early Warning System (ICEWS) Dataverse [1], and constructed dynamic symmetric binary matrices $\{Y_{ij,t} : t \in \{1, ..., 48\}, i, j : \{2, ..., 15\}, i > j\}$ using monthly interactions. Specifically, $Y_{ij,t} = 1$ indicates country i signed formal agreements with country j at least once in month t, while $Y_{ij,t} = 0$ indicates country i did not sign formal agreements with country j at all in month t.

I set $H^* = 10$, $\phi_x = \phi_\mu = 0.9$, $s = 1$. This is the best parameter setting I found for good performance of both model fitting and prediction. Notice that although I focus on model fitting, I also consider the performance of prediction to avoid overfitting. Specifically, I set the whole matrix Y_{48} as missing values to evaluate the prediction performance.
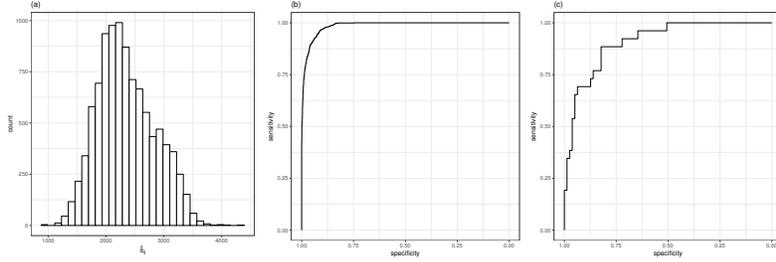
Figure 10. Model performance on fitting and prediction. (a)Scatter plot for estimator $\hat{\pi}_{ij}$ versus true π_{ij} including all data. (b)ROC curve only for fitting values. (c)ROC curve only for prediction values.
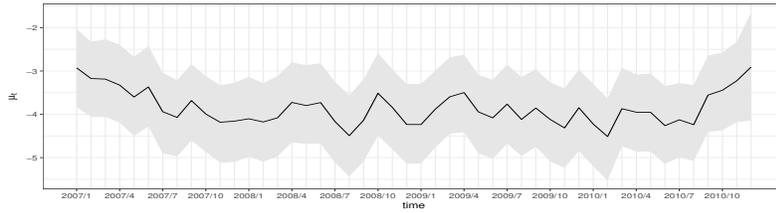
Figure 11. Graphical summary of trajectories of sampled $\hat{\mu}$. The solid line represent posterior mean. The grey ribbon represents 95% highest posterior density intervals.

I ran 5000 iterations with 1000 burn-in. Most of the effective sample size of $\pi_{ij,t}$ is around 2300 as shown in Figure 10, suggesting good mixing. The AUC for fitting and prediction are 0.9854 and 0.8092 respectively, showing good performance. Figure 11 shows the baseline trends μ_t. The flat μ_t indicates the relationships between the 15 countries are quite stable and do not have significant trends or dramatic changes. However, μ_t slightly goes down during 2007, and has higher amplitude during 2008, and has higher frequency during the second half year of 2009, then goes up at the August 2010. This trend may correspond to the 2008 financial crisis.

Figure 12 shows the correlation matrices between 15 countries at selected time points. The darker color indicates the higher probability of the existence of a link. The model captures the pattern of the relation matrices well for the dark parts in the top panel almost match with the black parts in the bottom panel.
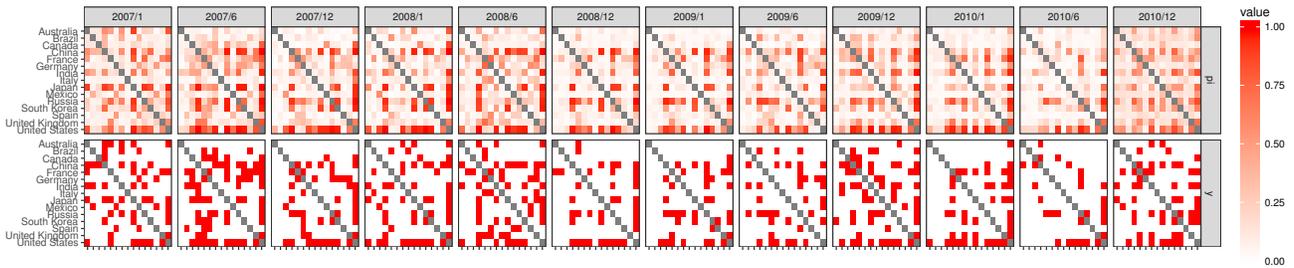
Figure 12. Model performance at selected time. The top panel indicates posterior mean $\hat{\pi}(t)$ and the bottom panel indicates the true relation data.

Figure 13 shows the probability of the existences of dynamic bilateral relations. We can refer which country maintained close relationship with China, and the corresponding evolving trend. From Figure 13(a), Japan, Russia, South Korea and the United States maintained more close relationship compared to the other countries. Detailed trajectory plots unveil the bilateral relations. From 2009/7 to 2010/7, Japan had less connections with China,

8

which is a little weird, since Yukio Hatoyama, who took office during this period, is pro-China politician. Maybe a large number of agreements were signed, which is hided behind the binary data. From July 2009 to July 2010, less agreements were assigned between China and Russia, which may due to the thaw between Russia and the United States took the main focus. Since October 2009, the relationship between China and South Korea deteriorated. This may be connected to the deteriorating relationship between South Korea and North Korea. The government of South Korea showed their dissatisfaction to the attitude of China towards the dispute between South Korea and North Korea. From Figure 13 (e), we can also see large fluctuations between China and the United States since January 2010, which may be connected to the tensions due to the tightened censorship from China. Generally, from Figure 13(a), we can see some significant light color on 2009/7 and 2010/4 to 2010/7. Such decrease in the relations may be connected to the worst ethnic violence in Xinjiang region.[4] The light color from 2010/4 to 2010/7 may be connected to the political change.(Vice-President Xi Jinping named vice-chairman of powerful Central Military Commission in October 2010.)
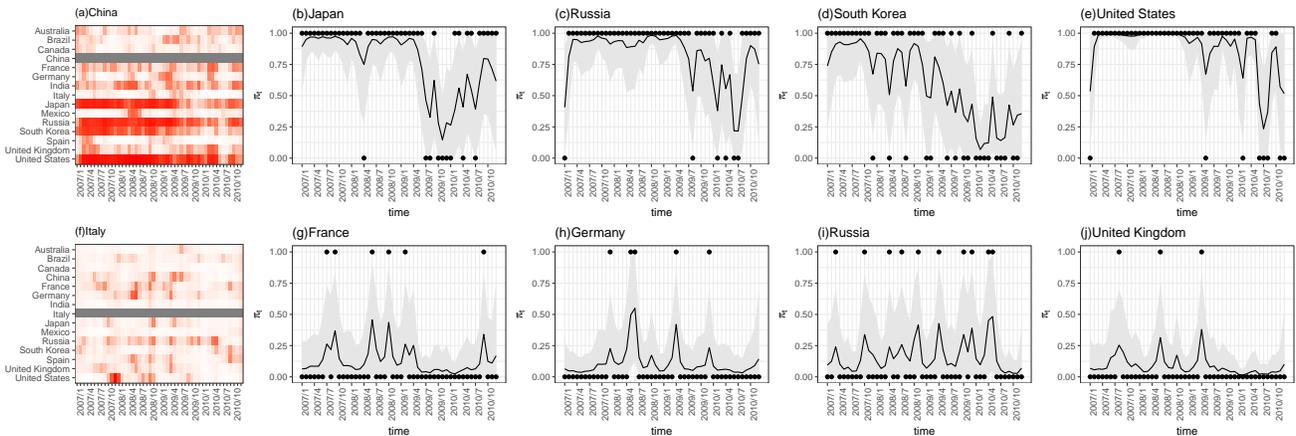
Figure 13. Parameter estimation of $\pi_{ij,t}$ for China and Italy, and the bilateral relations. The left tiles plots show the dynamic relationship between China(or Italy) and other countries. The grey plots indicate the dynamic bilateral relations. The solid lines represent posterior mean. The grey ribbons represent 95% highest posterior density intervals. The black points represent the true relation data.

Compared to China, Italy did not actively take part into the international relationships. From Figure 13(f), most tiles are in light color, and there are several dark colored ribbon at around 2008/4, 2008/10 and 2009/4. This active period may be correlated to diplomatic policy from a new government (Berlusconi wins the general elections in April 2008). These active periods may also be correlated to assistance corresponding to the recession (officially declared in November 2008) and the earthquake (in the mountainous Abruzzo region in April 2009).

Generally, the trends of the parameters are close related to natural disasters, financial crisis, new administrations (diplomatic policies), and other events. Since I only analyzed on four years, it is hard to delve into more insights. However, this model can serve as a useful tool for capturing the underlying pattern of the bilateral relations well.

5. Discussion

The first limitation comes from the data itself. Binary data contains very little information. From Figure 13, the credible interval is relative wide, showing great uncertainty on the estimation. From Figure 9, even under reasonable specified hyperparameters, the credible interval is still wide. Actually, my data comes from count data then I transformed it to binary data. The reason is Gibbs sampling is unavailable for Poisson regression. Also

---

[4]In July 2009, leaders of China and Taiwan exchange direct messages for the first time in more than 60 years. I am not sure whether this would take the focus of China, leading to decrease for the activities with other countries.

the characteristics of the dataset suggest it is inappropriate to use normal distribution to approximate the Poisson distribution. This is because around 77.2% of the entries are zero, and over 96.8% of the entries are less than 20. Modeling the count data directly and applying other sampling methods, such as Metropolis-Hastings, for posterior computation may be a better idea.

The second limitation is that too many parameters may bring overfitting of the model. As I shows in the table 1, a large value of H^* may introduce too many parameters leading to overfitting. To avoid overfitting, I leave one matrix out to take the performance of prediction into consideration. Another method to avoid overfitting is to set a shrinkage prior to allow learning from the dimension of the latent factor (Durante and Dunson, 2014 [3]). Specifically, for x_{ih}, a shrinkage parameter τ_h is added to learn the dimension of the latent space. $(x_{ih}(t_1), ..., x_{ih}(t_N))^T \sim N(0, \tau_h^{-1} c_x)$, where $\tau_h = \prod_{l=1}^h \nu_l, \nu_l \sim Ga(a_1, 1), \nu_l \sim Ga(a_2, 1), l \leq 2$. I conducted several simulations showing the shrinkage parameter works well.

The third limitation is computation complexity. Compared to FFBS sampling we learned from class, the Gibbs sampling is time consuming, which restricts the application of the model in large-scale networks or networks evolving in a relative long time period. Other sampling methods or algorithms are desired to ease the computation. For example, maybe we could consider using a series with Markovian properties to approximate the series, or approximating the posterior distributions directly. In addition, we should take computation into consideration if we want to extend it to a more complex model, such as allowing learning of the parameters ϕ, s, considering AR(p) and TVAR models, and adding node-specific addictive effect.

The fourth limitation is the identification of multiplicative effect. If we focus on inference on π_t, we do not need to worry about this issue. However, evolving patterns of latent factors will provide us more insights on the dynamic networks. Also, learning from the parameters ϕ, s and other model extensions are based on the identification of the multiplicative effects. In latent space model for a static network, Hoff *et al.* (2002) proposed Procrustean transformations. I need to explore whether I could extend this method for dynamic evolving network. And the computation is still a issue here. Another method is to add restriction and use informative priors. How to set the prior is an interesting topic worth exploring.

In addition, it will be interesting to adopt the "emulation" idea. In this project, I consider model the dynamic network from a node's perspective. Specifically, I model the characteristics of nodes in the network and consider the probability of an edge to be a function of paired nodes. Another modeling process is to fit the edges directly (Chen *et al.*, 2018 [2]). For example, consider the time series of each edge between paired nodes separately, and fit each time series using DGLM model or others, then apply FFBS sampling for parameter estimation and map the model back to latent space model [5]. This could be my direction in following updates of the project.

References

[1] E. Boschee, J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz, and M. Ward. Icews coded event data. 2015.

[2] X. Chen, K. Irie, D. Banks, R. Haslinger, J. Thomas, and M. West. Scalable bayesian modeling, monitoring, and analysis of dynamic network flow data. *Journal of the American Statistical Association*, 113(522):519–533, 2018.

[3] D. Durante and D. B. Dunson. Nonparametric bayes dynamic modelling of relational data. *Biometrika*, 101(4):883–898, 2014.

[4] P. Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in neural information processing systems*, pages 657–664, 2008.

[5] P. D. Hoff. Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association*, 100(469):286–295, 2005.

[6] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098, 2002.

[7] N. G. Polson, J. G. Scott, and J. Windle. Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349, 2013.

---

[5]Thanks for the advice from Professor Mike West