Multi-View Weighted Network

by

Xi Yang

Program in Statistical and Economic Modeling

Date:____________________

Approved:

____

Sayan Mukherjee, Supervisor

____

Surya Tokdar

____

Edward Tower

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science in the Program in Statistical and Economic Modeling in the
Graduate School of Duke University

2016

ABSTRACT

Multi-View Weighted Network

by

Xi Yang

Program in Statistical and Economic Modeling

Date:_____________________
Approved:

____
Sayan Mukherjee, Supervisor

____

Surya Tokdar

____

Edward Tower

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Program in Statistical and Economic Modeling in the
Graduate School of Duke University

2016

Abstract

Extensive investigation has been conducted on network data, especially weighted network in the form of symmetric matrices with discrete count entries. Motivated by statistical inference on multi-view weighted network structure, this paper proposes a Poisson-Gamma latent factor model, not only separating view-shared and view-specific spaces but also achieving reduced dimensionality. A multiplicative gamma process shrinkage prior is implemented to avoid over parameterization and efficient full conditional conjugate posterior for Gibbs sampling is accomplished. By the accommodating of view-shared and view-specific parameters, flexible adaptability is provided according to the extents of similarity across view-specific space. Accuracy and efficiency are tested by simulated experiment. An application on real soccer network data is also proposed to illustrate the model.

Contents

List of Figures

Acknowledgements

First, I would like to give a special gratitude to Professor David Dunson, my advisor for introducing me to the field of Bayesian statistics and machine learning, and Daniele Durante for guidance, help and support me during the whole process. My thesis would not have been possible without the support of them.

Second, I would like to express my gratitude to my committee numbers, Professor Sayan Mukherjee, Professor Kent Kimbrough and Professor Surya Tokdar for all their time and patience.

Finally, I would like to thanks my family and all my friends for their encouragement and support.

1. Introduction

Exploration of weighted network (Dehmer & Basak, 2012) data in the form of symmetric matrices with count entries plays an essential part in extensive fields ranging from biological, engineering and economic systems. To be specific, data of the co-movement in financial markets, occurrence of passing events in soccer matches and individual's friendship or social connections on social software all take the form of weighted network.

The majority of network models are built under the concept of binary entries. For example, the multidimensional unfolding methodology (DeSarbo & Hoffan, 1987), spatial analysis of binary choice data (DeSarbo et al., 1999), probabilistic model for time-varying relational data (Ishiguro et al., 2010), etc., are all structured for binary network. Current approaches for weighted network includes factorizing into adjacency matrix and weight matrix (Dehmer & Basak, 2012), mapping from a weighted network to an unweighted multigraph (Newman, 2004), approximate log-normal model (hattacharya, et al., 2008), studying topological properties and network statistics (Fagiolo, et al., 2008) from the weight matrix, etc. Subject to the previous analysis of binary network, these models lack the over-all interpretation of the weighted network structure, as well as the analysis of interaction between adjacency matrix and weight matrix. Weighted stochastic block model (Aicher, et al., 2015), weighted infinite relational model with Dirichlet process prior (Jiang & Zhang,2015) overcome these problems and gives analysis on the specific properties of weighted network. However, due to lack of full conditional conjugacy and absence of efficient algorithm, computational intractability is still an essential issue. Algorithms like weighted MAX-SAT problem and the MaxWalkSat local search (Cussens,2012), Clique Percolation Method with weights (Farkas, et al., 2007), on-demand distributed clustering algorithm for multi-hop packet radio networks (Chatterjee, et al.,2002) are proposed in a rich literature. Other approaches from genetic fields like optimation of the weighted connection (Whitley, et al.,1990), weighted voxel coactivation network analysis (Mumford, et al., 2010), etc., provide practical algorithm and simulation of weighted network. But few focus on the inference of view specific structure of multi-view weighted network.

Based on the previous analysis of binary network (Durante & Dunson, 2014), the motivation is to conduct more informative analysis of network data by avoiding the process of transforming count to binary entries. The interests of investigators lie in (1) identification of similarities and dissimilarities across views; (2) efficient estimation of sparse matrices (3) adaptive model to the extents of similarity across views; (4) providing general network structure in each view.

A Poisson-Gamma latent factor model is proposed in this paper. With the factorization of the Poisson mean into view-specific factor and view-shared factor, it provides an approach to identify similarities and dissimilarities across views. The both factors are formed as sum of 2 different sets of quadratic combination of latent coordinates, which not only achieves the dimension reduction, but also contributes to full conditional conjugate posteriors for efficient Gibbs sampling by the introducing of multiplicative gamma process shrinkage prior (Bhattacharya & Dunson,2011) to each latent coordinates. The proposed model is adaptive to the level of dissimilarities across views by controlling the sum of quadratic combination. In addition, over parameterization is avoided by the multiplicative gamma shrinkage prior.

In this paper, a basic model allowing single network in each view is first proposed, then a more general model for multi-network structure in each view is further developed under similar prior specification, giving a complete approach for statistical inference of multi-view networks under different structures.

An increasing amount of literatures from the scientific community on the analysis of soccer network occurs in recent years. For example, complex soccer network (Onody & de Castro,2004), regarding the soccer players and the clubs as two kinds of nodes, semantic analysis system to identify the special events in soccer games (Huang, et al.,2006). In this paper, 2 kinds of network are constructed to test the effects of passing events within one team and across teams. The first network regards soccer players as nodes, the passing events happened between 2 players as edges and the number of passing events as the weight, with one network each match.

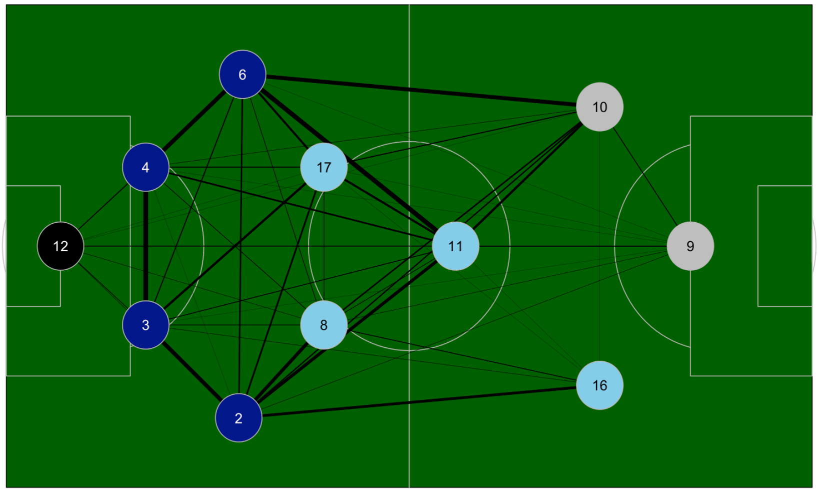

Figure 1: Player network: Brazil- Mexico 0: 0 (0: 0)

Figure 1 gives an example of passing network among Brazilian players of the match Brazil - Mexico in the 2014 FIFA World Cup. The numbers on nodes indicate the players' numbers and colors of nodes indicates the position of players (black node is goalkeeper; dark blue nodes are defenders; light blue nodes are midfield players; gray nodes are forwards). The width of edges indicates the number of passing events happened between the 2 players, i.e. weight. The location of players is identified by the tactical line-up of each match.

Figure 2 gives an example of network constructed by position. Each network has 18 nodes and each node represents a partition of the soccer field. The 18 partitions of the soccer field are identified in figure 2. This is the same match as Figure 1 (2014 World Cup match with Mexico), the order of partition is given by the numbers in figure 2 and the width of lines indicates the number of passing events happened between the 2 partition, i.e. weight.

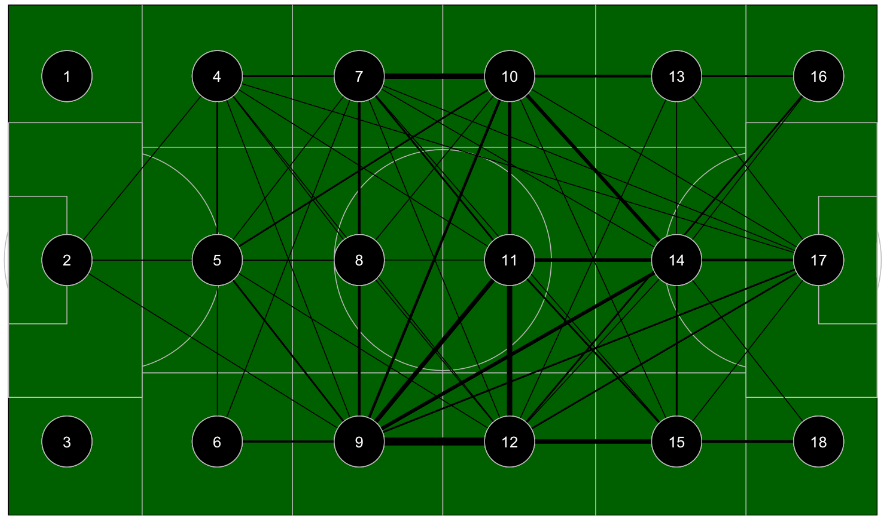

Figure 2: Soccer field network: Brazil-Mexico 0: 0 (0 : 0)

From the soccer networks constructed by these 2 methods, information in different aspects can be obtained by conducting statistical inference.

The model specification and properties are given in section 2. Posterior computation and Markov chain Monte Carlo algorithm are described in section 3. In section 4, a more general model is given to be adaptive to multi-network structure in each view. The model is illustrated and compared with current approaches by a simulation study in section 5. Section 6 gives 2 approaches to construct the soccer network and applies the model to 2014 FIFA World Cup data. Section 7 provides a further discussion over the results and model properties.

2. Poisson-Gamma latent factor model

Besides giving statistical inference on multi-view weighted network adjacency matrix, this model also provides techniques for further development of dynamic multi-view weighted network. This section mainly illustrates the the general notation and structure of the proposed model.

2.1 Model specification

Y_l is a $V \times V$ symmetric count matrix at view $l = 1, \dots, K$ with empty and meaningless diagonal elements. Λ_l is the mean matrix of Y_l with entries $\lambda_{ij,l} = E[y_{ij,l}], \forall\, i = 2, \dots, V; j = 1, \dots, i-1$. Let,

$$y_{ij,l} | \lambda_{ij,l} \sim Poisson(\lambda_{ij,l}) \quad (1)$$

independently for all i, j, l. The motivation is to define a prior for $\{\lambda_{ij,l}\}$, s.t.: (i) efficient computations (ii) dimensionality reduction for large V (iii) allow inference on both sparse and dense matrices (iv) able to distinguish view-shared and view-specific spaces.

Due to symmetric entries and empty diagonal elements, the dimension of Y_l (or . Λ_l) is $\frac{V(V-1)}{2}$.

2.2 Latent space

Motivated by separating view-shared and view-specific spaces, 2 latent variable matrices U, V_l s.t. $Y_l = U + V_l$ are introduced for each Y_l, with entries satisfying:

$$u_{ij} + v_{ij,l} = y_{ij,l} \sim Poisson(\lambda_{ij,l}) \quad (2)$$

i.e. $E[u_{ij}] + E[v_{ij,l}] = \lambda_{ij,l}$ and $u_{ij} \sim Poisson(.)$, $v_{ij,l} \sim Poisson(.)$.

Under this specification, view-shared space matrix U of multi-view network with high similarity across views will be sparse; network with low similarity across views will have dense U matrix. In the extreme condition, when $U = 0$, the network data is totally different independent matrix sequence; when $V_l = 0, \forall l = 1, \ldots, K$, matrices of each view are the same, the data becomes one count symmetric matrix.

With the motivation of the dimensionality reduction, quadratic combination of latent coordinates for is used to construct view-shared factor u_{ij} and view-specific factor $\{v_{ij,l}\}_{l=1}^{K}$,

$$\lambda_{ij,l} = \sum_{h=1}^{H} x_{ih} x_{jh} + \sum_{h=1}^{H*} x_{ih,l} x_{jh,l} \tag{3}$$

$$u_{ij} \sim Poisson\left(\sum_{h=1}^{H} x_{ih} x_{jh}\right) \tag{4}$$

$$v_{ij,l} \sim Poisson\left(\sum_{h=1}^{H*} x_{ih,l} x_{jh,l}\right) \tag{5}$$

In matrix notation, X is $V \times H$ matrix with entries $\{x_{ih}\}_{i=1,\ldots,V;h=1,\ldots,H}$ and $\{X_l\}_l^K$ are $V \times H*$ matrices with entries $\{x_{ih,l}\}_{i=1,\ldots,V;h=1,\ldots,H*,l=1,\ldots,K}$, then $\Lambda = XX' + X_l X_l'$.

The dimensionality is reduced from $\frac{V(V-1)}{2}$ to $V \times H + V \times H*$ $(H, H* \ll V)$ by the latent space representation. $H$and $H*$ are adaptive to the level of similarity across views. Relatively large $H*$ and small H should be used for multi-view network with high similarity, small $H*$ and large H should be used for multi-view network with low similarity. Different methods could be implemented to test the values of H and $H*$, for example, including another prior for H and $H*$ or introducing a loss function.

Multiplicative gamma process shrinkage priors enable X and X_l to shrink toward zero as the column index increases, allowing flexible choices of $H, H*$, while avoiding the over parameterization.

$$x_{ih} = \prod_{t=1}^{h} \delta_{it}, x_{ih,l} = \prod_{t=1}^{h} \delta_{it,l} \quad (6)$$

where $\delta_{it} \sim G(1, \alpha), \delta_{it,l} \sim G(1, \alpha*)$, with $\alpha, \alpha* > 1$, i.e. $E[\delta_{it}] < 1$.

As H or $H*$ goes to infinity, $\{x_{ih}\}$ or $\{x_{ih,l}\}$ tends to zero, shrinking $\{\lambda_{ih,l}\}$ towards zero.

3. Bayesian inference

Focus on constructing efficient full conditional conjugate posterior for Gibbs sampling, latent variables $\{\boldsymbol{z_{ijh}}\}_{\boldsymbol{h=1}}^{\boldsymbol{H*}}$ and $\{\boldsymbol{z_{ijh}}\}_{\boldsymbol{h=1}}^{\boldsymbol{H}}$ are introduced as following:

$$u_{ij} = \sum_{h=1}^{H} z_{ijh} \quad (7)$$

$$v_{ij,l} = \sum_{h=1}^{H*} z_{ijh,l} \quad (8)$$

where $z_{ijh} \sim Poisson(x_{ih}x_{jh}), h = 1, \dots, H;\ z_{ijh,l} \sim Poisson(x_{ih,l}x_{jh,l}), h = 1, \dots, H*$.

The full conditional conjugate posterior and Gibbs sampling steps are the following:

1. Update each view-shared factor u_{ij} from multinomial posteriors conditional on $\{y_{ij,l}\}_{l=1}^{K}$, $\{x_{ih}\}_{h=1}^{H}$ and $\{x_{ih,l}\}_{h=1;l=1}^{h=H*;l=K}$ for every $i = 2, \dots, V; j = 1, \dots, i-1$:

$$A_{ij} = \prod_{l}^{K} \frac{(\sum_{h=1}^{H*} x_{ih,l} x_{jh,l})^{y_{ij,l}-w}}{(y_{ij,l} - w)!} \tag{9}$$

$$B_{ij} = \frac{(\sum_{h=1}^{H} x_{ih} x_{jh})^{w}}{w!} \tag{10}$$

Let $n_{ij} = min_l[y_{ij,l}]$, then

$$u_{ij} | \{y_{ij,l}\}, - \sim multi\big([1, \dots, n_{ij}], [p_{ij}(1), \dots, p_{ij}(n_{ij})]\big) \tag{11}$$

i.e. $P(u_{ij} = w) = p_{ij}(w), w = 1, \dots, n_{ij}$, with

$$P_{ij}(w) = A_{ij} \times B_{ij} \tag{12}$$

$$P_{ij}(w) = \frac{P_{ij}(w)}{\sum_{w=1}^{n_{ij}} P_{ij}(w)} \tag{13}$$

Then, view-specific factor $v_{ij,l}$ can be updated by: $v_{ij,l} = y_{ij,l} - u_{ij}$, for every $i = 2, \dots, V; j = 1, \dots, i-1, l = 1, \dots, K$.

2. Update the latent factor z_{ijh} from multinomial posteriors conditional on $\{x_{ih}\}_{h=1}^{H}$ and u_{ij}, update $z_{ijh,l}$ from multinomial posteriors conditional on $\{x_{ih,l}\}_{h=1}^{H*}$ and $v_{ij,l}$ for every $i = 2, \dots, V; j = 1, \dots, i-1, l = 1, \dots, K$. In this case, $\{z_{ijh}\}$ is independent with $\{x_{ih,l}\}$ and $\{z_{ijh,l}\}$ is independent with $\{x_{ih}\}$. Let $Z_{ij,l} = [z_{ij,l}, \dots, z_{ijH*,l}]$, $Z_{ij} = [z_{ijh}, \dots, z_{ijH}]$.

Then,

$$p\left(Z_{ij}\middle|-\right) \propto p\left(u_{ij}\middle|\left\{z_{ijh}\right\}_{h=1}^{H}\right) p\left(\left\{z_{ijh}\right\}_{h=1}^{H}\right) \tag{14}$$

$$\propto 1_{\{u_{ij}=\sum_{h}^{H} z_{ijh}\}} \prod_{h}^{H} \frac{(x_{ih}x_{jh})^{z_{ijh}}}{z_{ijh}!} \tag{15}$$

$$Z_{ij}| - \sim Multi\left(n = u_{ij}, p_{ij}\right) \tag{16}$$

where $p_{ij} = [\frac{x_{ih}x_{jh}}{\sum_{h=1}^{H} x_{ih}x_{jh}}, \dots, \frac{x_{iH}x_{jH}}{\sum_{h=1}^{H} x_{ih}x_{jh}}]$.

Similarity,

$$Z_{ij,l}| - \sim Multi\left(n_l = v_{ij,l}, p_{ij,l}\right) \tag{17}$$

where $p_{ij,l} = [\frac{x_{ih,l}x_{jh,l}}{\sum_{h=1}^{H*} x_{ih,l}x_{jh,l}}, \dots, \frac{x_{iH*,l}x_{jH*,l}}{\sum_{h=1}^{H*} x_{ih,l}x_{jh,l}}]$.

3. Update $\delta_{it}, \delta_{it,l}$ from gamma posterior:

$$p(\delta_{it}|-) \propto p(\delta_{it}) \prod_{j \neq i} \prod_{h=t}^{H} p\left(z_{ijh}\middle|\delta_{it}\right) \tag{18}$$

$$\propto G(\delta_{it}|1, \alpha) \times C \tag{19}$$

$$C = \prod_{j=i} \prod_{h=t}^{H} \frac{(x_{ih}x_{jh})^{z_{ijh}}}{z_{ijh}!} e^{-x_{ih}x_{jh}} \tag{20}$$

Then $\delta_{it}| - \sim G(\delta_{it}|a_{it}, b_{it})$

$$a_{it} = 1 + \sum_{j \neq i} \sum_{h=t}^{H} z_{ijh} \tag{21}$$

$$b_{it} = \alpha + \sum_{j \neq i} \sum_{h=t}^{H*} \frac{x_{ih}}{\delta_{it}} x_{jh} \tag{22}$$

where $\frac{x_{ih}}{\delta_{it}} = \delta_{i1}\delta_{i2} \dots \delta_{it-1}\delta_{it+1} \dots \delta_{ih}$.

Similarly, $\delta_{it,l}| - \sim G(\delta_{it,l}|a_{it,l}, b_{it,l})$

$$a_{it,l} = 1 + \sum_{j \neq i} \sum_{h=t}^{H*} z_{ijh,l} \tag{23}$$

$$b_{it,l} = \alpha * + \sum_{j \neq i} \sum_{h=t}^{H*} \frac{x_{ih,l}}{\delta_{it,l}} x_{jh,l} \tag{24}$$

where $\frac{x_{ih,l}}{\delta_{it,l}} = \delta_{i1,l}\delta_{i2,l} \dots \delta_{it-1,l}\delta_{it+1,l} \dots \delta_{ih,l}$.

Besides, if the view Y_{l*} is missing, a method to deal with missing data can be generated as following. First, $\alpha *$ can be estimated by the inverse of the sample mean of $\delta_{it,l}$ from the results of MCMC. Then $\{\delta_{it,l*}\}$ can be sample from $G(1, \hat{\alpha} *)$, so the corresponding $\{x_{ih,l*}\}$ and $\{v_{ij,l*}\}$ can be generated by (5) and (6). Then the missing Y_{l*} can be estimated as $\hat{y}_{ij,l*} = \hat{u}_{ij} + \hat{v}_{ij,l*}$.

4. Multi-multi-view network

The Poisson-Gamma latent factor model provides good and efficient estimation on the view-shared properties and view-specific properties. However, it has drawbacks when estimating the structure of each view, because only one network matrix is available in each view, which is not enough for the inference on the view structure. In this section, we further develop a more general model, allowing for different amounts of network matrices in each view. The multi-multi-network has K views and each view includes N_l network $V \times V$ matrices.

4.1 Model Specification

$\{Y_l^{(n)}\}_{n=1}^{N_l}$ are symmetric count adjacency matrix with entries $y_{ij,l}^{(n)}$, for every $n = 1, \dots, N_l, l = 1, \dots, K$. With similar notations as the Poisson-Gamma latent factor model, let

$$y_{ij,l}^{(n)}|\lambda_{ij,l} \sim Poisson(\lambda_{ij,l}) \tag{25}$$

The the latent space will be

$$u_{ij} + v_{ij,l}^{(n)} = y_{ij,l}^{(n)} \sim Poisson(\lambda_{ij,l}) \tag{26}$$

$$v_{ij,l}^{(n)} = \sum_{h=1}^{H*} z_{ijh,l}^{(n)} \tag{27}$$

$$v_{ij,l}^{(n)} \sim Poisson\left(\sum_{h=1}^{H*} x_{ih,l}x_{jh,l}\right) \tag{28}$$

$$z_{ij,l}^{(n)} \sim Poisson(x_{ih,l}x_{jh,l}) \tag{29}$$

Other notations stay the same as section 2.

4.2 Bayesian inference

1. Update u_{ij} from multinomial posteriors conditional on $\{y_{ij,l}^{(n)}\}_{l=1}^{K}, \{x_{ih}\}_{h=1}^{H}$and $\{x_{ih,l}\}_{h=1;l=1}^{h=H*;l=K}$ for every $i = 2, \dots, V; j = 1, \dots, i-1$:

$$A'_{ij} = \prod_{l} \frac{(\sum_{h=1}^{H*} x_{ih,l}x_{jh,l})^{\sum_{n=1}^{N_l}(y_{ij,l}^{(n)}-w)}}{\prod_{n=1}^{N_l}(y_{ij,l}^{(n)} - w)!} \tag{30}$$

$$B_{ij} = \frac{(\sum_{h=1}^{H} x_{ih}x_{jh})^{w}}{w!} \tag{31}$$

Let $n_{ij}' = min_l[y_{ij,l}{}^{(n)}]$, with fixed $i, j, w = 1, \dots, n'_{ij}$, then

$$u_{ij}|\{y_{ij,l}{}^{(n)}\} \sim multi([1, \dots, n'_{ij}], [p'_{ij}(1), \dots, p'_{ij}(n'_{ij})]) \tag{32}$$

i.e. $P(u_{ij} = w) = p'_{ij}(w)$, with

$$P'_{ij}(w) = A'_{ij} \times B_{ij} \tag{33}$$

$$P'_{ij}(w) = \frac{P'_{ij}(w)}{\sum_{w=1}^{n'_{ij}} P'_{ij}(w)} \tag{34}$$

Then, view-specific factor $v_{ij,l}^{(n)}$ can be updated by: $v_{ij,l}^{(n)} = y_{ij,l}^{(n)} - u_{ij}$, for every $i = 2, \dots, V; j = 1, \dots, i-1, l = 1, \dots, K, n = 1, \dots, N_l$.

2. Update $z_{ijh}, z_{ijh,l}^{(n)}$ from multinomial posteriors, let ${Z_{ij,l}}^{(n)} = [{z_{ij,l}}^{(n)}, \dots, {z_{ijH*,l}}^{(n)}]$, $Z_{ij} = [z_{ijh}, \dots, z_{ijH}]$.

$$Z_{ij} | - \sim Multi(n = u_{ij}, p_{ij}) \tag{35}$$

where $p_{ij} = [\frac{x_{ih}x_{jh}}{\sum_{h=1}^{H} x_{ih}x_{jh}}, \dots, \frac{x_{iH}x_{jH}}{\sum_{h=1}^{H} x_{ih}x_{jh}}]$.

Similarity,

$$Z_{ij,l}^{(n)} | - \sim Multi\left(n_l = v_{ij,l}^{(n_l)}, p_{ij,l}\right) \tag{36}$$

where $p_{ij,l} = [\frac{x_{ih,l}x_{jh,l}}{\sum_{h=1}^{H*} x_{ih,l}x_{jh,l}}, \dots, \frac{x_{iH*,l}x_{jH*,l}}{\sum_{h=1}^{H*} x_{ih,l}x_{jh,l}}]$.

3. Update $\delta_{it}, \delta_{it,l}$ from gamma posterior:

Then $\delta_{it} | - \sim G(\delta_{it} | a_{it}, b_{it})$

$$a_{it} = 1 + \sum_{j \neq i} \sum_{h=t}^{H} z_{ijh} \tag{37}$$

$$b_{it} = \alpha + \sum_{j \neq i} \sum_{h=t}^{H} \frac{x_{ih}}{\delta_{it}} x_{jh} \tag{38}$$

where $\frac{x_{ih}}{\delta_{it}} = \delta_{i1}\delta_{i2} \dots \delta_{it-1}\delta_{it+1} \dots \delta_{ih}$.

Similarly, $\delta_{it,l} | - \sim G(\delta_{it,l} | {a'}_{it,l}, {b'}_{it,l})$

$${a'}_{it,l} = 1 + \sum_{j \neq i} \sum_{h=t}^{H*} \sum_{n=1}^{N_l} z_{ijh,l}^{(n)} \tag{39}$$

$${b'}_{it,l} = \alpha * + N_l \sum_{j \neq i} \sum_{h=t}^{H*} \frac{x_{ih,l}}{\delta_{it,l}} x_{jh,l} \tag{40}$$

where $\frac{x_{ih,l}}{\delta_{it,l}} = \delta_{i1,l}\delta_{i2,l} \dots \delta_{it-1,l}\delta_{it+1,l} \dots \delta_{ih,l}$.

The above proposed model gives flexible amount of network in each view, allow for highly accurate estimation when more network data is available in each view. The previous model is the special case when N_l, for all $l = 1, \dots, K$.

An alternative method to deal with missing data is provided. If a network in view l is missing, we can use the estimated $\widehat{\Lambda_l}$ to generated the network by $y_{ij,l}^{(n)} \sim Poissson(\hat{\lambda}_{ij,l})$.

5. Cluster model

In this section, a cluster model is given to cluster the views. The data includes K views, each view l includes n_l network matrices. The motivation is to separate the L views into K clusters, with $K \ll L$.

5.1 Model specification

$$y_{ij,l}^{(n)} | \lambda_{ij,l} \sim Poisson(\lambda_{ij,l}) \tag{41}$$

$$\lambda_{ij,l} \sim Multi\left(\left[\lambda_{ij}^{(k)}\right]_{k=1}^{K}, [p_k]_{k=1}^{K}\right) \tag{42}$$

$$(p_1, \dots, p_K) \sim Dir\left(\frac{1}{K}, \dots, \frac{1}{K}\right) \tag{43}$$

i.e. $P\left(\lambda_{ij,l} = \lambda_{ij}^{(k)}\right) = p_k$. In the matrix notation, $P\left(\Lambda_l = \Lambda^{(k)}\right) = p_k, k = 1, \dots, K$.

With the similar prior specification as the previous model,

$$\lambda_{ij}^{(k)} = \sum_{h=1}^{H} x_{ih}x_{jh} + \sum_{h=1}^{H} x_{ih,k}x_{jh,k} \tag{44}$$

$$x_{ih}=\prod_{t=1}^{h}\delta_{it}\,,x_{ih,k}=\prod_{t=1}^{h}\delta_{it,k} \quad (45)$$

where $\delta_{it}\sim G(1,\alpha),\delta_{it,k}\sim G(1,\alpha_k *)$ and $y_{ij,l}^{(n)}=u_{ij}+$

$v_{ij,l}^{(n)},u_{ij}\sim Po\left(\sum_{h=1}^{H}x_{ih}x_{jh}\right),u_{ij}=\sum_{h=1}^{H}z_{ijh}\,,v_{ij,l}^{n}=\sum_{h=1}^{H}z_{ijh,l}^{(n)}\,,z_{ijh}\sim Po(x_{ih}x_{jh})$.

A new latent variable is introduced as following:

$$c_l\sim Multi([1,\dots,K],[p_1,\dots,p_K]) \quad (46)$$

with $p\left(\Lambda_l=\Lambda^{(k)}\middle|c_l=k\right)=1$, i.e. $\Lambda_l=\Lambda^{(c_l)}$.

Then

$$z_{ij,l}^{(n)}\sim Po\left(x_{ih,c_l}x_{jh,c_l}\right),v_{ij,l}^{n}\sim Po\left(\sum_{h=1}^{H}x_{ih,c_l}x_{jh,c_l}\right)$$

5.2 Bayesian inference

1. Update $\{p_k\}_{k=1}^{K}$ from Dirichlet distribution:

$$(p_1,\dots,p_k)\sim Dir(r_1,\dots,r_K) \quad (47)$$

with $r_k=\frac{1}{K}+\sum_{l=1}^{L}1_{\{c_l=k\}}$

2. Update $[c_1,\dots,c_L]$ from multinomial distribution:

$$P(c_l=k|-)\propto p_k\prod_{i=1}^{V}\prod_{j<i}\left[\left(\lambda_{ij}^{(k)}\right)^{\sum_{n=1}^{N_l}y_{ij,l}^{(n)}}e^{-N_l\lambda_{ij}^{(k)}}\right] \quad (48)$$

3. Update u_{ij}, from multinomial distribution:

$$A'_{ij}=\prod_{l=1}^{L}\frac{\left(\sum_{h=1}^{H*}x_{ih,c_l}x_{jh,c_l}\right)^{\sum_{n=1}^{N_l}(y_{ij,l}^{(n)}-w)}}{\sum_{n=1}^{N_l}(y_{ij,l}^{(n)}-w)\,!} \quad (49)$$

$$B_{ij} = \frac{(\sum_{h=1}^{H} x_{ih}x_{jh})^w}{w!} \tag{50}$$

Let $n'_{ij} = \min\,[y_{ij,l}^{(n)}]$, with fixed $i, j, w = 1, \ldots, n'_{ij}$, then

$$u_{ij}|\{y_{ij,l}^{(n)}\} \sim multi\left([1, \ldots, n'_{ij}], \left[p'_{ij}(1), \ldots, p'_{ij}(n'_{ij})\right]\right) \tag{51}$$

i.e. $P(u_{ij} = w) = p'_{ij}(w)$, with

$$P'_{ij}(w) = A'_{ij} \times B_{ij} \tag{52}$$

$$p'_{ij}(w) = \frac{P'_{ij}(w)}{\sum_{w=1}^{n'_{ij}} P'_{ij}(w)} \tag{53}$$

Then, view-specific factor $v_{ij,l}^{(n)}$ can be updated by: $v_{ij,l}^{(n)} = y_{ij,l}^{(n)} - u_{ij}$, for every $i = 2, \ldots, V; j = 1, \ldots, i-1, l = 1, \ldots, K, n = 1, \ldots, N_l$.

4. Update z_{ijh} and $z_{ijh,l}^{(n)}$ from multinomial distribution:

$$Z_{ij}| - \sim Multi(n = u_{ij}, p_{ij}) \tag{54}$$

$$Z_{ij,l}^{(n)}| - \sim Multi\left(n_l = v_{ij,l}^{(n)}, p_{ij,c_l}\right) \tag{55}$$

where $p_{ij} = \left[\frac{x_{ih}x_{jh}}{\sum_{h=1}^{H} x_{ih}x_{jh}}, \ldots, \frac{x_{iH}x_{jH}}{\sum_{h=1}^{H} x_{ih}x_{jh}}\right], p_{ij,c_l} = \left[\frac{x_{ih,c_l}x_{jh,c_l}}{\sum_{h=1}^{H} x_{ih,c_l}x_{jh,c_l}}, \ldots, \frac{x_{iH,c_l}x_{jH,c_l}}{\sum_{h=1}^{H} x_{ih,c_l}x_{jh,c_l}}\right]$.

5. Update $\delta_{it}, \delta_{it,k}$ from gamma distribution:

Then $\delta_{it}| - \sim G(\delta_{it}|a_{it}, b_{it})$

$$a_{it} = 1 + \sum_{j \neq i} \sum_{h=t}^{H} z_{ijh} \tag{56}$$

$$b_{it} = \alpha + \sum_{j \neq i} \sum_{h=t}^{H} \frac{x_{ih}}{\delta_{it}} x_{jh} \tag{57}$$

where $\frac{x_{ih}}{\delta_{it}} = \delta_{i1}\delta_{i2} \ldots \delta_{it-1}\delta_{it+1} \ldots \delta_{ih}$.

Similarly, $\delta_{it,k}| - \sim G(\delta_{it,k}|a'_{it,k}, b'_{it,k})$

$$a'_{it,k} = 1 + \sum_{l:c_l=k} \sum_{j \neq i} \sum_{h=t}^{H} \sum_{n=1}^{N_l} z_{ijh,l}^{(n)} \tag{58}$$

$$b'_{it,k} = \alpha * + \sum_{l:c_l=k} N_l \sum_{j \neq i} \sum_{h=t}^{H} \frac{x_{ih,k}}{\delta_{it,k}} x_{jh,k} \tag{59}$$

6. Simulation study

In this section, 2 multi-view networks are generated to evaluate the performance of the proposed models. The first one is the basic model in section 2, the second is the more general model in section 4. Comparison with other current approaches is also provided to test the accuracy and bias.

6.1 Poisson-Gamma Latent Factor Model

The aim is to generate a series of $V \times V$ matrices Y_l, with $V = 15, K = 10, l = 1, \dots, K$. Each view l is a $V \times V$ symmetric matrix Y_l. Entries of Y_l is generated by (1), with $\lambda_{ij,l}$ simulated by (3); $H = 10, H *= 5. x_{ih,l}$ and x_{ih} are generated by (6), with $\delta_{it} \sim G(1, \alpha)$, $\delta_{it,l} \sim G(1, \alpha *)$,$\alpha = 3, \alpha *= 2, i, j = 1, \dots, V, l = 1, \dots K$. We run 5000 Gibbs iterations to allow enough draws for convergence and use the first 1000 draws as burn-in. The initial value of $\{u_{ij}\}_{i,j}$ is 0; initial value of $\{v_{ij,l}\}_{i,j,l}$ is $\{y_{ij,l}\}_{i,j,l}$. The initial value of δ_{it} and $\delta_{it,l}$ are separately sampled from $Gamma(1, \alpha + 2)$ and $Gamma(1, \alpha * + 2)$.

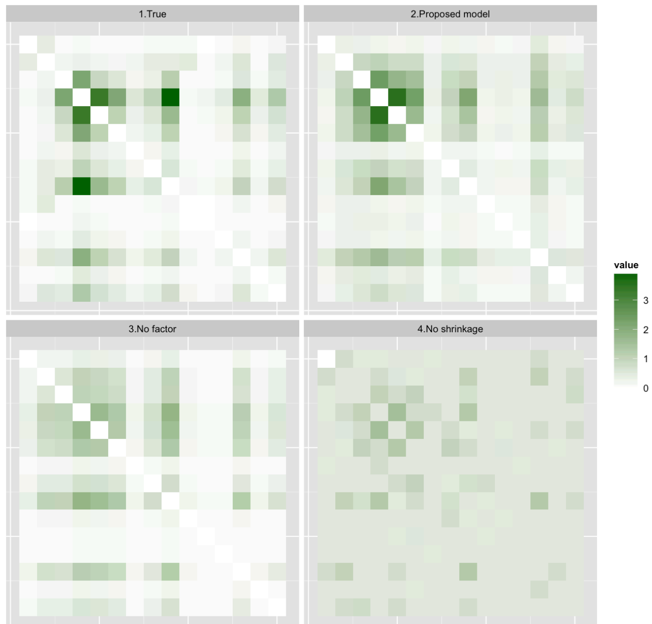

Figure 3: heat map

For elected Λ_l, top left is the true value; top right is the results of proposed Poisson- Gamma latent factor model; bottom left is the results of model without using the view-shared factor and view-specific factor; bottom right is the result of model without using the multiplicative gamma process shrinkage prior.

From figure 3, the shrinkage prior plays an essential part when estimate the mean of sparse matrices. The multiplicative gamma process shrinkage prior increases the accuracy of estimation, avoids the over parameterization with larger $H, H*$, and gives full conjugate posterior for efficient Gibbs sampling. The model without the separating of view-shared factor $\{\boldsymbol{u_{ij}}\}_{i,j}$ and view- specific factor $\{\boldsymbol{v_{ij,l}}\}_{i,j,l}$ performs well in capturing the trends of $\boldsymbol{\lambda}$. When the shared factor is small (i.e. networks across views are very different and has few shared properties), it will have similar results as the proposed model. However, the proposed model can separate the shared properties even if views are very different and it is adaptive to the extents of similarity across views.

6.2 Multi-Multi-View Network

Networks from 5 views are generated in this part. Each view l includes N_l networks in the form of 15×15 symmetric matrices, with $l = 1, \ldots, 5$. Let $N_l = 2l$, so the data generated includes $(1 + 5) \times 5 = 30$ networks of 15×15 matrices within 5 views. Each $y_{ij,l}^{(n)}$ is generated by equation (25) and each Λ_l is generated by equation (3),(6). Let $\alpha = 3$ and $\alpha *= 2$, choose the truncated level $H = 10, H *= 5$.

In this case, each view has more data than the previous model and it is expected to give more accurate estimation. Figure 4 gives a graph comparison between the true value and estimation, which indicates the good performance of the improved model.

When more data is available within each view, this model not only gives the accurate estimation both on the view-shared properties and view-specific properties, but also gives reliable structure information of each views.

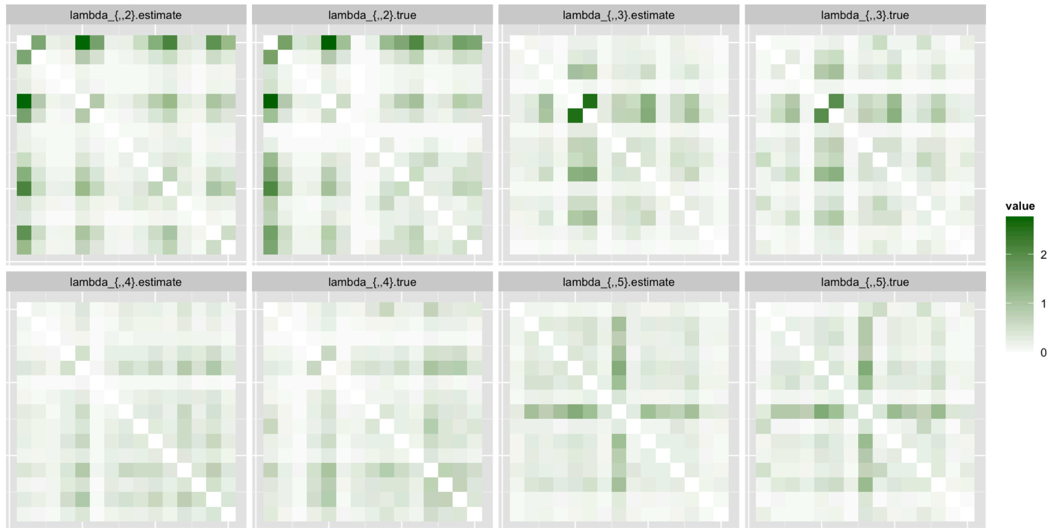

Figure 4: Heat plot of estimated(left) and true(right) $\{\boldsymbol{\Lambda}_l\}$, with $\boldsymbol{l} = \mathbf{1}, \ldots, \mathbf{5}$

7. Soccer network data

In this section, 2 approaches are proposed to construct the soccer networks. The first is to construct the passing network regarding players as nodes. The entries of network indicate the number of passing events happened between the 2 corresponding players. The second method is to construct the passing network by position regarding partitions of soccer field as nodes. The entries of network indicate the number of passing events happened between 2 partitions. Thus, both within team and across team passing events structure are learned from the 2 kinds of network. The data is from the 2014 FIFA World Cup website.

7.1 Soccer player network

The motivation is to understand the passing strategy, find the over all shared passing behavior of Brazil National team, give suggestions to teams who will fight against Brazil and analysis the reasons of Brazil's lost/win/draw.

4-3-3 attacking formation is mainly used by Brazil national team, which indicates the offensive game plan and can be easily transform to the more defensive 4-5-1 formation (the match with Netherlands) and 4-4-2 formation (the match with Mexico). Although Brazil national team changed the formation in the 2 matches with Mexico and Netherlands, the tactical line-up keeps the relatively same position as in Figure 1, with 4 defenders in the back, 3 players in the middle field and 3 players in the forward field, i.e. relative position in tactical line-up keeps unchanged.

We reorder the players by their relative position in the tactical line-up of each match. The No.1 player is always the goal keeper; No.2 to No. 5 are always defenders; No. 6 to No.11 are midfielders and forward players (in most of the case of Brazil national team, No. 6 to No. 8 are midfielders and No.9 to No.11 are forward players), in spite of their names and other characters, i.e. players of same numbers are regarded as identity. The detailed order is given in Figure 6. Let the nodes be the reordered players and edges be the passing events happened between 2 players, we first construct a series of weighted network matrices, with the weight indicates the amount of passing events happened between the corresponding 2 players. Then the series of network matrices are divided into 3 views. The first view includes 3 networks of winning matches, the second view includes 2 draw matches, and the third view includes the rest 2 matches that Brazil lost. Each network matrix is constructed by the passing events of Brazil National team happened in a match.

From Figure 5, we can see the apparent difference in network structure from the 3 different views.

The top left picture is the network structure of all matches that Brazil won. Interaction between players are relative uniformly distributed, with slightly more mess put on the interaction between defenders, which indicates the appropriate teamwork and effective defense strategy. The second row of the network matrix has larger entries than other rows, which indicates he No. 2 player (defender) played a dominant role in the over-all passing distribution.

The top right picture is the network structure of all matches of Brazil that ended up in a draw. Interaction among defenders are relatively high. The 6th row of the network matrix has the least weight comparing with other rows, while the 5th and 4th row has relatively large weight. The No.6 player (midfielder) seems to be isolated, i.e. seldom conducted passing events with other players and the 2 players around him (No. 4 and No.5) did most of the passing events.

The bottom left picture is the network structure of all lost matches of Brazil. The 9th row of the network matrix has very low weight, while 10th and 8th row have relatively large weight. The No.9 player (midfielder in the Netherlands match and the forward of Germany match) only conducted few passing events and the No.8 and No.10 player played the dominant roles in the over-all passing events.

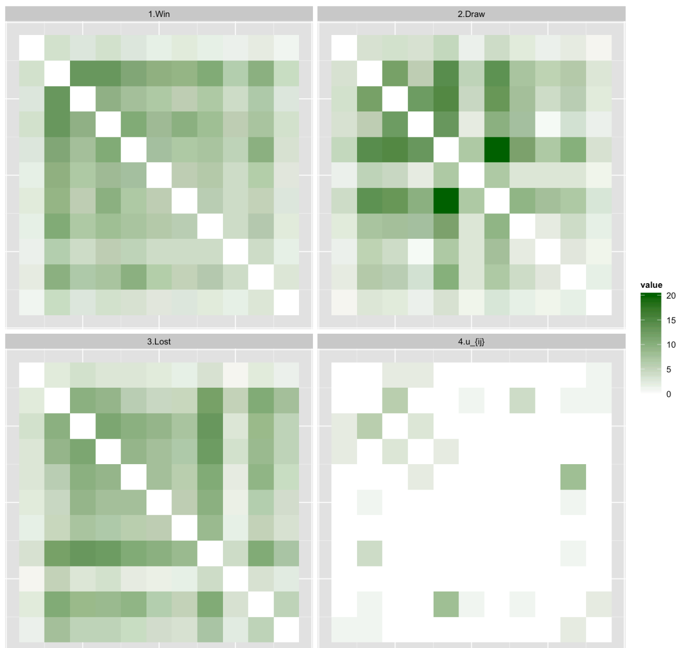

Figure 5: heat map

The top left picture is the network structure of all winning matches; The top right is the network structure of all draw matches; the bottom left is network structure of all matches that Brazil lost; the bottom right is the shared structure among matches.

The structure of draw matches and lost matches indicates the problematic isolation of certain players. Such behaviors have more negative influence to the results of the match when the player is in the forward position than in the midfield position.

The last picture is the shared structure across all views (matches). However, the heat map is not intuitive enough. Figure 6 is the shared passing structures of Brazil in the form of connected nodes, with the width of lines indicates the weight. From Figure 6, the interaction between defenders are relatively high, which implies strong defense skills. However, the interaction between node 5 and other defenders is slightly lower, which may be a breakthrough for Brazilian's opponents when attacking. In addition, No.7 and No. 9 player are isolated, combining with the results from heat map, they have relatively low influence to the results of the match. Due to large amounts of passing between No. 10 player (forward in most of the case) and other players, Brazilian's opponents should pay close attention to him when defending.

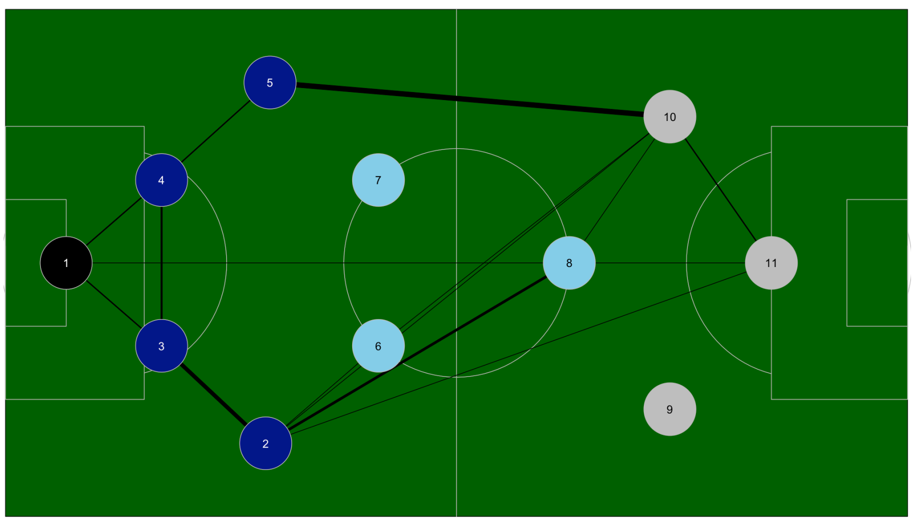

Figure 6: Shared structure among all Brazil matches

7.2 Soccer field network

Motivated by finding connection between the number of goals and passing events and the over-all passing structures, a new network is generated by position. The soccer field is first divided into 18 partitions and then we reorder the partitions as in

Figure 3. The passing distribution of players is transformed into passing distribution of position by multiplying the passing probability and receiving probability of each partition(calculated form the Player Statistics in the 2014 FIFA World Cup Website). Then each partition is a node, the network weight is defined as the amounts of passing events happened between the corresponding partitions. In this setting, we only consider the long passing between 2 partitions, i.e. short pass within one partition are taken no account of.

The data contains 7 views, each view includes the passing network of a team in a match with number of goals equal to 0,1,2,3,4,5,7. For example, view 1 includes 4 networks, with number of goals equal to 0. Each network is generated by one team in a single match. The data only includes competitive teams who had won at least one championship in the previous FIFA World Cup, and all interested matches of these teams are at least draw, i.e. regardless of lost matches. In this case, we have 7 views and 27 networks, with each network take the form of 18×18 symmetric matrices.

The results are shown in figure 7, when goal equal to 0, the over all number of passing events is low, with most passing events happened at the midfield. As the number of goal increase, the passing events happened at the back is stable, with the over-all number of passing increases and most passing events tends to happened at the forward position. Intuitively, as number of goals increase, passing distribution will be more concentrated on the forward position. However, when number of goal equals to 4 or 5, there is a decreasing in the over-all number of passing events.

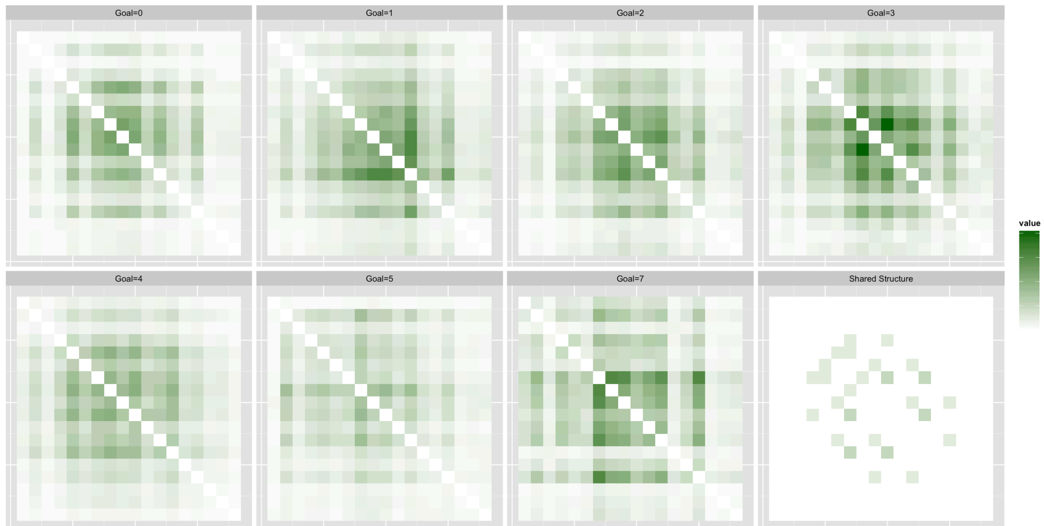

Figure 7: Each heat map represents the network structure of all matches with goal=0,1,2,3,4,5,7. The last heat map is the view-shared structure.

Figure 8 gives the shared structure across view. From figure 8, the network in each view shares very little similarity and all happened at the midfield, which indicates the behavior in the midfield has relatively low influence on the number of goals. Instead, the behaviors at both the back and front position would have essential effects on the number of goals.

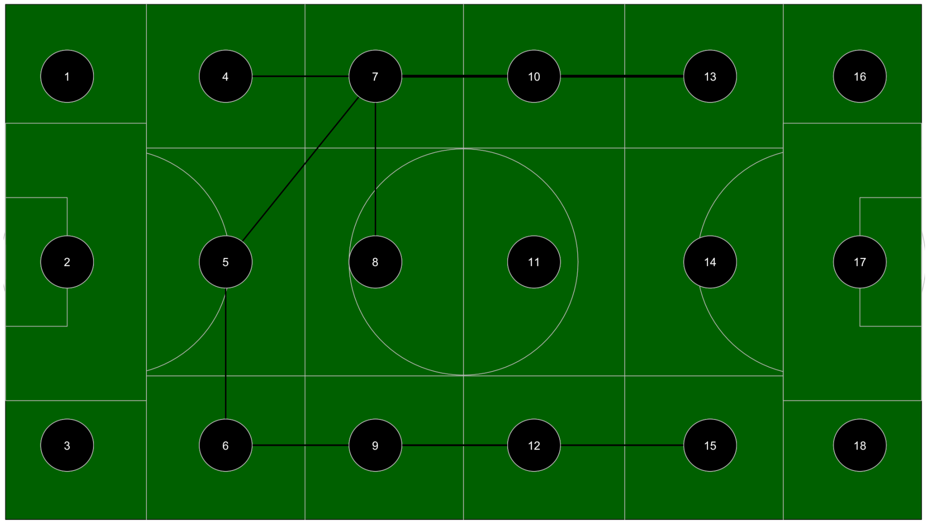

Figure 8: Shared structure among all views

8. Discussion

The Poisson structure allows for the estimation of count entries matrices, avoiding the lost of information while transforming weighted network to binary network. The factorization of Poisson mean provides approach to estimate the similarity and dissimilarity across views. By control the amount of quadratic combination of latent coordinates, the model is flexible to multi-view networks with difference level of similarity. The introducing of quadratic combination also gives an approach for the dimension reduction. With the use of multiplicative gamma process shrinkage prior, over parameterization is avoided. A more general multi-multi-view network model is developed to allow for multi-networks in each view. Methods to deal with missing data are proposed and efficient Gibbs sampling is ensured by the full conditional conjugate posterior. The model can be further developed into dynamic multi-view network by adding the time-varying parameters and applied to extensive fields. However, effective methods are needed to decide the truncated level H and $H*$. In addition, the proposed model can also be applied to network with continuous entries by rounding the entries to integer.

References

1. DeSarbo, W.S., & Hoffman, D.L. (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.

2. DeSarbo, W.S., Kim, Y., & Fong D. (1999). *A Bayesian multidimensional scaling procedure 410 for the spatial analysis of revealed choice data*. Journal of Econometrics 89, 79–108.

3. Ishiguro, K., Iwata, T., Ueda, N. & Tenenbaum, J. (2010). *Dynamic infinite relational model for time-varying relational data analysis*. In Advances in Neural Information Processing Systems (NIPS).

4. Dehmer, M., & Basak, S. C. (Eds.). (2012). *Statistical and machine learning approaches for network analysis*. Wiley.

5. Durante, D., & Dunson, D. B. (2014). *Nonparametric Bayes dynamic modelling of relational data*. Biometrika, asu040.

6. Bhattacharya, K., Mukherjee, G., Saramäki, J., Kaski, K., & Manna, S. S. (2008). *The international trade network: weighted network analysis and modelling. Journal of Statistical Mechanics: Theory and Experiment*, 2008(02), P02002.

7. Newman, M. E. (2004). *Analysis of weighted networks*. Physical Review E, 70(5), 056131.

8. Aicher, C., Jacobs, A. Z., & Clauset, A. (2015). *Learning latent block structure in weighted networks*. Journal of Complex Networks, 3(2), 221-248.

9. Fagiolo, G., Reyes, J., & Schiavo, S. (2008). *On the topological properties of the world trade web: A weighted network analysis*. Physica A: Statistical Mechanics and its Applications, 387(15), 3868-3873.

10. Cussens, J. (2012). *Bayesian network learning by compiling to weighted MAX-SAT*. arXiv preprint arXiv:1206.3244.

11. Farkas, I., Ábel, D., Palla, G., \& Vicsek, T. (2007). *Weighted network modules*. New Journal of Physics, 9(6), 180.

12. Chatterjee, M., Das, S. K., & Turgut, D. (2002). WCA: A weighted clustering algorithm for mobile ad hoc networks. Cluster Computing, 5(2), 193-204.

13. Bhattacharya, A., & Dunson, D. B. (2011). *Sparse Bayesian infinite factor models*. Biometrika, 98(2), 291.

14. Onody, R. N., & de Castro, P. A. (2004). *Complex network study of Brazilian soccer players*. Physical Review E, 70(3), 037103.

15. Huang, C. L., Shih, H. C., & Chao, C. Y. (2006). *Semantic analysis of soccer video using dynamic Bayesian network*. Multimedia, IEEE Transactions on, 8(4), 749-760.

16. Jiang, X., & Zhang, W. (2015). *Weighted Infinite Relational Model for Network Data*. Journal of Communications, 10(6).

17. Mumford, J. A., Horvath, S., Oldham, M. C., Langfelder, P., Geschwind, D. H., & Poldrack, R. A. (2010). *Detecting network modules in fMRI time series: a weighted network analysis approach*. Neuroimage, 52(4), 1465-1476.

18. Whitley, D., Starkweather, T., & Bogart, C. (1990). *Genetic algorithms and neural networks: Optimizing connections and connectivity*. Parallel computing, 14(3), 347-361.