DRAFT

LAPO SANTI

Contents

1. Introduction

Introduction:

The purpose of this paper is to introduce a novel methodology to infer the latent ordering of blocks that exhibit a transitivity property in directed networks. Specifically, we develop a method to infer the community structure of binary interactions, where the groups are ordered in a one-dimensional hierarchy that determines the preferred edge direction. The paper builds on a class of models closely related to stochastic blockmodels, where only non-negative entries are considered. However, in this model, we are ignoring the zero entries, which can be debatable whether we are still dealing with a network as we are ignoring the sparsity that is one of the staples of networks.

Research Question:

The research question of this paper is to infer the block structure underlying a set of binary interactions with transitivity properties in directed networks. Our approach advances the existing literature by considering a method to learn the ordering of the blocks that reflects better an ordering of the block themselves.

Importance of the Study:

This paper innovates the field of statistics by providing a novel methodology to infer the latent ordering of blocks, which can be very useful under certain circumstances. The proposed method has practical applications in analyzing tennis tournament data, where we can learn how competitive a tournament is by inferring the blocks and their ordering, from the strongest to the weakest.

Objectives:

The objective of this paper is to develop a method to infer the community structure of binary interactions where the groups are ordered in a one-dimensional hierarchy that determines the preferred edge direction. Specifically, we aim to learn the ordering of the blocks that exhibit a transitivity property in directed networks.

Methodology:

We start by introducing the theory behind Partially Ordered Markov Models (POMMs) and the definition of a Strong Stochastic Transitivity matrix (SST matrix). Then we combine the two ideas into one, showing that we can represent the SST matrix with a collection of level sets. We impose a very specific dependence structure between the blocks, a Markovian one. Next, we specify the full model, where outcomes of tennis matches are represented as a binomial distribution, the POMM ordering is imposed on the entries of the winning probability matrix, and a dirichlet multinomial distribution is imposed on the assignment of players to blocks. This distribution is expanded into a Chinese Restaurant Process (CRP), which can be used to avoid assumptions on the initial number of blocks. Initially, we assume that the number of blocks follows a Poisson distribution. Finally, we simulate data and test our methodology with simulated data, followed by applying it to a real tennis dataset.

Outline of the Paper:

This paper is organized as follows. Section 2 introduces the theory behind Partially Ordered Markov Models (POMMs) and the definition of a Strong Stochastic Transitivity

matrix (SST matrix). In Section 3, we combine the two ideas into one, showing that we can represent the SST matrix with a collection of level sets. Section 4 describes the dependence structure imposed on the blocks, a Markovian one, and how we specify the full model. In Section 5, we explain how we expand the dirichlet multinomial distribution into a Chinese Restaurant Process (CRP) and the assumptions we make on the number of blocks. Section 6 presents our simulation study, where we test our methodology with simulated data. Finally, in Section 7, we apply our methodology to a real tennis dataset. We conclude the paper with a discussion of our findings and potential future research directions.

2. Posets and Dags

Definition 1. *[Poset]To define poset, a partially ordered set we start from D, a set of elements. The binary relation $\prec$ on D is said to be a partial order if:*

(1) *For any $x \in D, x \prec x$(reflexivity)*

(2) *For any $x, y, z \in D, x \prec y$ and $y \prec z \implies x \prec z$(transitivity)*

(3) *For any $x, y \in D, x \prec y$ and $y \prec x \implies x = y$(antisymmetry).*

Then we call $(D, \prec)$ a partially ordered set, or a poset.

A *finite poset* $(D, \prec)$ is a poset where D has a finite number of distinct elements.

Example: let D be the finite set defined by representing the $M \times N$ array of probabilities. Let (u, v) and (q,r) be any two elements of D and define the binary relation on D by

$$(q, r) \prec (u, v) \iff q \prec u \text{ and } r \prec v \tag{4}$$

There exists a correspondence between posets and directed cyclic graphs.

Let (D, F) be a directed acyclic graph, where $D = \{y_1, \ldots, y_n\}$, a finite set. To construct a poset to which this digraph corresponds, we define the binary relation $\prec$ on D by

(5) $y_i \prec y_i$ for $i = 1 \ldots n$

(6) $y_i \prec y_j$ if there exists a directed path from y_i to $y_j \in (D, F)$

We saw above that the correspondence is many-to-one. Given a finite poset, one may construct a class of directed acyclic graphs; the correspondence described above is in a sense the minimal directed acyclic graph since it has the smallest possible directed edge set Pomms definitions

Definition 2 (Cone)**.** *For any $y \in D$, the cone of y is the set*

$$cone\ y = \{x \in D : x \prec y; x \neq y\}$$

Definition 3 (Adjl)**.** *For any $y \in D$ the adjacent lower neighbourhood of y is the set*

$$adjl\ y = \{x \in D : (x, y) is\ a\ directed\ edge\ in (D, F)\}$$

Definition 4 (Dilation)**.** *For any $y \in D$, the dilation of y is the set*

$$dil\ y = \bigcup \{\overline{adjl}\ x : y \in \overline{adjl}\ \}$$

Definition 5 (Excluded dilation)**.** *For any $y \in D$, the excluded dilation of y is the set*

$$dil\ ^* y = dil\ y \setminus \{y\} \tag{7}$$

Definition 6 (Minimal element)**.** *In general, an element $y \in D$ is called minimal element if there is no other element x satisfying $x \prec y$ where adjl s is the set of adjacent lower neighbors of $s \in D$.*

Definition 7 (Cover of a Subset)**.** *The cover of a subset B is a set of all elements x in D such that x is adjacent to an element in B and x is not in B. Formally, the cover of B is defined as follows:*

$$covr\ B = x \in D : adjl\ x \subset B \text{ and } x \notin B$$

where adjl x is the set of all adjacent elements of x in D.

Intuitively, the cover of a subset B represents all the elements in D that are outside of B but are adjacent to at least one element in B. In other words, the cover of B captures the neighborhood of B in D.

Definition 8 (Level Sets)**.** *The level sets of a poset D are a sequence of nonempty cover sets defined recursively as follows:*

$$L^0 = D_{min}; \quad L^i = covr\left(\bigcup_{k=0}^{i-1} L^k\right)$$

where D_{min} is the set of all minimal elements in D.

The first level set L^0 is simply the set of all minimal elements in D. The subsequent level sets are defined by taking the union of all the previous level sets and taking the cover of this union. Intuitively, each level set captures the neighborhood of the previous level sets in D.

2.1. **Partially ordered Markov models:** Consider a finite set of random variables $\{Z(s_1), \ldots, Z(s_n)\}$ indexed by location or "points"

$$D = \{s_1, \ldots, s_n\} : n \in \{1, 2, \ldots\}$$

That is, we assume the existence of a directed acyclic graph (D, F) and its corresponding poset $(D, \prec)$. Let (D, F) be a finite, directed acyclic graph and its corresponding poset $(D, \prec)$. Consider $s \in D$ and recall the definition of cone s. Also, let the quantity U_s denote any subsets of points not related to s. Formally:

$$U_s \subset \{u \in D : u \text{ and } s \text{ are not related } \}$$

Definition 9 (POMM)**.** *Then $\{Z(s) : s \in D\}$ is said to be a partially ordered Markov model (POMM) if , for all $s \in D$ and any U_s*

$$P(Z(s)|Z(\ cone\ s), Z(U_s)) = P(Z(s)|Z(\ adjl\ s) \tag{8}$$

Proposition 1. *[Joint Distribution] Let (D, F) be a directed acyclic graph with no singleton points and let $(D, \prec)$, be its associated poset. Suppose that $\{Z(s) : s \in D\}$ is a POMM. Then*

$$P(Z(D)) = P(Z(L^0)) \prod_{k=1}^{m} \prod \{P(Z(u))|Z(\ adjl\ u) : u \in L^k\} \tag{9}$$

$$= P(Z(L^0)) \prod \{P(Z(u))|Z(\ adjl\ u) : u \in D \setminus L^0\} \tag{10}$$

where $L^0, L^1, \ldots L^m$ are the level sets as defined previously.

Result 1 relates the probability of a random variable defined on a poset to the probabilities of its restrictions to the lower level sets of the poset.

The result states that the probability of Z on the entire poset D can be expressed as a product of the probabilities of Z restricted to the level sets $L^0, L^1, \ldots, L^m$ of the poset, where $L^0 = D_{\min}$ is the set of minimal elements of D, and L^k is the set of elements of D that are not in any of the previous level sets $L^0, L^1, \ldots, L^{k-1}$ and whose immediate predecessors are all in the union of the previous level sets $\bigcup_{i=0}^{k-1} L^i$.

The first part of the result states that the probability of Z on D is equal to the product of the probability of Z on L^0 and the conditional probabilities of Z on the elements of each subsequent level set L^k, given the values of Z on their immediate predecessors. This can be seen as a form of the chain rule of probability, where the joint probability of Z on D is decomposed into a product of conditional probabilities.

The second part of the result simplifies the product by noting that the conditional probabilities of Z on the elements of $D \setminus L^0$ are determined by the values of Z on their immediate predecessors, which are all in L^0 or $D \setminus L^0$. Therefore, the product can be simplified to the product of the probability of Z on L^0 and the conditional probabilities of Z on the elements of $D \setminus L^0$ given the values of Z on their immediate predecessors in $D \setminus L^0$. This simplification reduces the number of terms in the product and makes the computation of the joint probability of Z on D more efficient.

3. Application to the SST matrix

In this section, I will introduce a matrix that displays the Strong Stochastic Transitivity (SST) property. The aim of this section is to model the probability of a player winning a match in a tournament against another one. To achieve this goal, we need to ensure that the probabilities are arranged in a consistent manner. Specifically, we want to ensure that if Player A is stronger than Player B, and Player B is stronger than Player C, then Player A must be stronger than Player C. This is a well-known mathematical concept known as transitivity, and it is essential to impose it on the probabilities of victory between the players. For instance, if we know that Djoković is stronger than Medvedev and Medvedev is stronger than Kyrgios, then we can infer that Djoković must be stronger than Kyrgios. By using the SST property, we can guarantee that the probabilities of victory reflect this logical relationship, which enhances the clarity and coherence of our model. After having introduced the SST property, we want to build upon the definitions of Section(1), and re-express the SST matrix within the POMMs' framework. This new definition will allow us to have a coherent and tractable framework to express the joint probability distribution of such ordered probabilities and, ultimately, to perform inference.

3.1. **Defining matrix P (SST matrix).** The matrix under consideration, that is P, is a collection of victory probabilities among K entities, which could represent players or also groups of players. The matrix is denoted by $\underset{K\times K}{P}$, where K is the number of players/ group of players taken into consideration.

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,K} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,K} \\ \vdots & \vdots & \vdots & \vdots \\ p_{K,1} & p_{K,2} & \cdots & p_{K,K} \end{pmatrix}$$

Each element $p_{i,j}$ in the matrix P represents the probability of player i winning over player j in a tennis match, where draws are not permitted. Therefore, it must be the case that $p_{i,j} + p_{j,i} = 1$ in order to satisfy the requirements for a valid probability. It follows that $p_{j,i}$ can be expressed as $1 - p_{i,j}$. The lower triangular entries of matrix P can be determined from the upper triangular entries. Consequently, our focus is on modelling the upper triangular part of P.

Without loss of generality, we can assume that player/team 1 is the strongest, and player/team K is the weakest.

From this assumption, it follows that the elements in the upper triangular part of the matrix must remain above 0.5 to maintain the assumption of monotonicity in the probabilities. Violation of this assumption could lead to a contradiction where a weaker player/team has a higher probability of winning than a stronger one. For instance, if $p_{1,2} = 0.4 \leq 0.5$, then $p_{2,1} = 0.6$ would imply that player 1 is weaker than player 2, which is contradictory with our baseline assumption. Furthermore, we set the main diagonal of matrix P to 0.5 for teams and 0 for individual players.

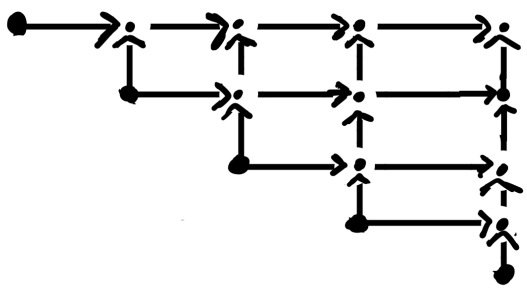


FIGURE 1. Dag representation of the Poset imposed onto the SST matrix

Therefore, we must constrain the probabilities as follows:

- The probabilities must increase monotonically as the index of the columns j increases;
- The probabilities must decrease monotonically as the index of the rows i increases.

The matrix P with the described modification will look like this:

$$\begin{pmatrix} 0.5 & \leq & p_{1,2} & \leq & ... & \leq & p_{1,K} \\ & & \vee| & & ... & & \vee| \\ 1-p_{1,2} & \leq & 0.5 & \leq & ... & \leq & p_{2,K} \\ \vdots & \vdots & \vdots & \vdots & ... & \vdots & \vdots \\ 1-p_{1,K} & \leq & 1-p_{2,K} & \leq & ... & \leq & 0.5 \end{pmatrix}$$

3.2. **Defining a Poset over P.** Having defined P, now we want to re-define it within a finite Poset framework to obtain $(P, \leq)$. Let (i, j) and (p, q) be two elements of P and define the binary relation $\leq$ on P as

$$(p, q) \leq (i, j) \iff p \leq i \text{ and } q \leq j \tag{11}$$

Now, $(P, \leq)$ is clearly a Poset since it satisfies the three properties of Definition (**??**). We represent the corresponding directed acyclic graph (P, F), where F is the set of directed edges between vertices in Figure (**??**) by using the definition of Adjacent Lower Neighborhood that we introduced before.

Definition 10 (Cone over P)**.** *In this case, for any $(i, j) \in D$ the cone of (i, j) is the set:*

$$\textit{cone}\ (i, j) = \{(i, j-1), \dots (i, i), (i+1, j), \dots, (K, j)\} \tag{12}$$

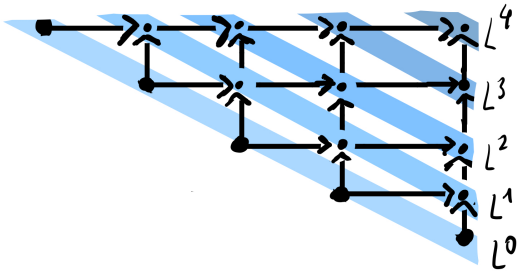


FIGURE 2. Visual Display of the level sets $L^k : k = 0, \ldots, 4$

Definition 11 ($\overline{\text{cone}}\,(i,j)$)**.** *In this case, for any $(i,j) \in P$ the closure of the cone of (i,j) is the set:*

$$\overline{cone}\,(i,j) = \{(i,j), \ldots (i,i), (i,j), \ldots, (K,j)\} \tag{13}$$

Definition 12 (The adjacent lower neighborhood of (i,j))**.** *In this case, for any $(i,j) \in P$ the adjacent lower neighborhood of (i,j) is the set:*

$$adjl\ (i,j) = \{(i, j-i), (i+1, j)\} \tag{14}$$

Definition 13 (closure of adjl (i,j))**.** *In this case, for any $(i,j) \in P$ the closure of adjl (i,j) is the set:*

$$\overline{adjl}\,(i,j) = \{(i, j-i), (i+1, j), (i,j)\} \tag{15}$$

Definition 14 (D^P_{min})**.** *In this case, the minimal element denoted by D_{min} is such that*

$$P_{min} = \{(i,j) \in P : i = j\} \tag{16}$$

namely the main diagonal of the P matrix.

The *level sets* of P, in this case, corresponds to the diagonals above the main one.

3.3. **Partially ordered Markov Models applied to P.** Considering the finite set of random variables $\{P_{1,1}, \ldots, P_{K,K}\}$ indexed by locations where

$$D \equiv \{(1,1), \ldots (K,K)\}$$

Having showed the existence of a directly acyclic graph (P,F) and its corresponding poset $(P, \leq)$, we can write down

$$P(P_{ij}|Z(\text{cone } i,j)) = P(P_{ij}|P(\text{adl } i,j)) \tag{17}$$

$$P(P_{ij}|P_{i+1,j}, P_{i,j-1}) \tag{18}$$

Now, exploiting Proposition (**??**), we can write:

$$P(P(D)) = \prod_{i=1}^{K} \prod_{j=i}^{K} P(P_{ij}|P_{i+1,j}, P_{i,j-1}) \tag{19}$$

4. Power Law prior process

The Power Law process is used to generate a prior process that can produce a SST matrix with a desired degree of flexibility. The probability values in the SST matrix should increase in the columns and decrease in the rows. To achieve this effect, the Power Law process uses level sets, which correspond to the diagonals in the upper triangular part of the matrix. The entries within each level set are modelled as truncated beta distributions with parameters (1,1).

More specifically, let us denote the set of upper triangular entries of the matrix as $U = (i,j) \in \mathbb{N}^2 : i \leq j$. Then, we can use the notation $\mathbf{P}_{(i,j)\in U}$ to represent the upper triangular entries of the matrix $\mathbf{P}$. The probability density function of the entries in the upper triangular part of the matrix can be written as:

$$p(\mathbf{P}(i,j) \in U) = \prod k = 1^K \prod_{j-i=k}^{|L^{(k)}|} \frac{1}{y^{(k+1)} - y^{(k)}} = \prod_{k=1}^{K} \left(\frac{1}{y^{(k+1)} - y^{(k)}} \right)^{|L^{(k)}|} \tag{20}$$

where $|L^{(k)}|$ is the cardinality of the level set $L^{(k)}$, and $y^{(k)}$ is the truncation point of the beta distribution that corresponds to level set k.

The truncation generating process starts with the function $f(x) = x^\alpha + 0.5 \in [0.5, \beta^0_{\max}]$, which is a monotonically increasing function that is linear for $\alpha = 1$, concave for $\alpha \in (0,1)$, and convex for $\alpha > 1$. The function is used to provide truncation points for the beta distributions.

To generate truncations, we take the inverse function of $f(x)$, denoted as $f(y)^{\leftarrow} = (y - 0.5)^{(1/\alpha)}$. The support of $f(x)$ is then found to be $[f(0.5)^{\leftarrow}, f(\beta^0_{\max})^{\leftarrow}] = [0, (\beta^0_{\max} - 0.5)^{(1/\alpha)}]$, which is then divided into K segments of equal length $\Delta = ((\beta^0_{\max} - 0.5)^{(1/\alpha)} - 0)/K$. The segment endpoints are defined as $x_k = \Delta \times (k-1)$ for $k = 1, \ldots, K+1$. The cumulative sum of the segment endpoints is then mapped back to $f(x)$, resulting in K truncation points $y^{(k)}$, which are defined as:

$$y^{(k)} = f(x_k) = \left(\frac{(\beta^0_{\max} - 0.5)^{(1/\alpha)}}{K} \times (k-1) \right)^{\alpha} + 0.5 \tag{21}$$

Let us define once again $Y = L(\alpha)^k$. Then let us denote with p^Y the pdf of Y and with p^α the pdf of α. We have that:

(22)

$$
\begin{aligned}
p_Y(y^{(k)}|\mu,\sigma^2) &= \left|\frac{dL^{-1}(y^{(k)})}{d(y^{(k)})}\right| \cdot p^\alpha\left(L^{-1}(y^{(k)})\right) \\
&= \left|\frac{2}{\ln\left(\frac{(k-1)}{K-1}\right)\left(2y^{(k)}-1\right)}\right| \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{\ln\left(y^{(k)}-0.5\right)-\ln\left((\beta^0_{\max}-0.5)\right)}{\ln\left(\frac{(k-1)}{K-1}\right)}-\mu\right)^2\right) \\
&\mathbb{I}\left(\frac{\ln\left(y-0.5\right)-\ln\left((\beta^0_{\max}-0.5)\right)}{\ln\left(\frac{(k-1)}{K}\right)} > 0\right)
\end{aligned}
$$

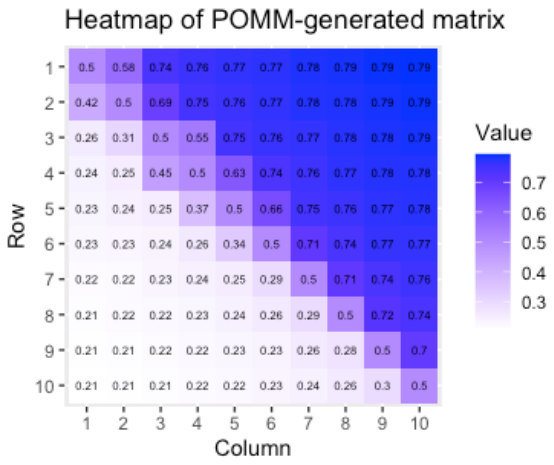


FIGURE 3. alpha1_10.png

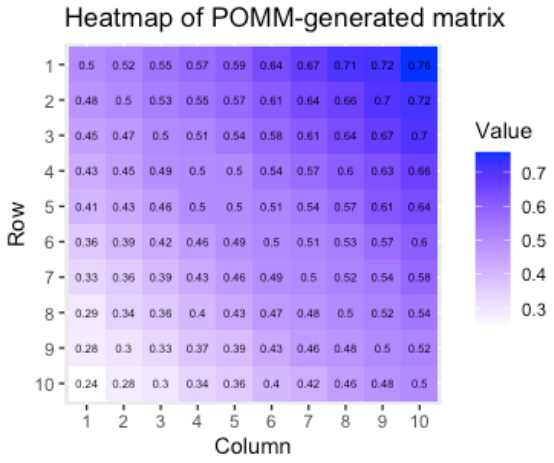


FIGURE 4. alpha1.png

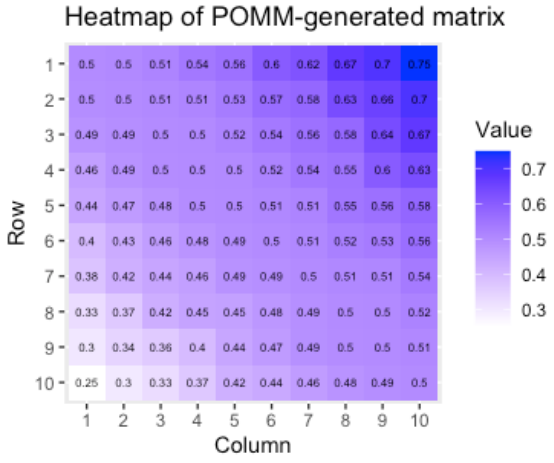


FIGURE 5. alpha1_5.png

5. Full model specification

The following Bayesian model is used to describe and analyze pairwise data, with the specific aim to identify clusters of points with similar connectivity patterns.

The model uses a Poisson distribution to model the number of blocks, a Dirichlet-multinomial distribution to model the distribution of nodes' assignment across blocks, and a binomial distribution to model the distribution of edges within blocks. Additionally, the model includes a POMM process to model the probability of edge formation between nodes within blocks.

The goal of the model is to estimate the number of blocks, the distribution of nodes across blocks, and the probability of edge formation between nodes within blocks, given observed network data. The Bayesian approach allows for uncertainty in these estimates and provides a framework for incorporating prior knowledge and updating beliefs as new data becomes available.

5.1. **Model specification.** We denote the nodes in this setting as $1, \dots, n$. We define the model as follows:

- Let K follow a Poisson distribution Poisson($\lambda = 1$), subject to the condition $K > 0$.
- Let $\theta_1, \dots, \theta_K$ be drawn from a Dirichlet distribution with parameter vector $\boldsymbol{\gamma}$.
- Let z_i be independently and identically drawn from a multinomial distribution with one trial and probability vector $(\theta_1, \dots, \theta_K)$.
- Let $P = \{p_{i,j}\} \sim \text{POMM}(\alpha, \beta^0_{\max}, K)$.
- Let $y_{i,j}$ follow a binomial distribution with parameters n_{ij} and p_{z_i,z_j}, subject to the condition $n_{ij} > 0$.

Here, z_i takes values in $1, \dots, K$. The likelihood function can be specified as:

$$p(\mathbf{y}|p, z) = \prod_{i<j}^{n-1} \binom{n_{ij}}{y_{ij}} p_{z_i,z_j}^{y_{ij}} (1 - p_{z_i,z_j})^{n_{ij}-y_{ij}} \tag{23}$$

To specify the prior distribution, we start with the prior on z following a multinomial distribution. We have n players to assign to K different labels, by replacing the extracted labels after each draw. We denote the variable which is the extracted label $i \in \{1, ..., K\}$ as Z_i, and denote as θ_i the probability that a given extraction will be of label i. The joint probability of $Z_1, \dots, Z_n$ will be:

$$f(z_1, \dots, z_n, \theta_1, \dots, \theta_K) = \Pr(Z_1 = z_1, \dots, Z_n = z_n) \tag{24}$$

$$= \begin{cases} \dfrac{n!}{n_1! \cdots n_K!} \theta_1^{n_1} \times \cdots \times \theta_K^{n_K}, & \text{when } \sum_{i=1}^K n_i = n \\ 0 & \text{otherwise,} \end{cases} \tag{25}$$

where $n_k = \sum_{i=1}^N \mathbb{I}(z_i = k) \quad$ for $k = 1, \dots, K$ for non-negative integers $z_1, ..., z_K$.
The probability mass function can be expressed using the gamma function as:

$$p(z_1, \ldots, z_n | \theta_1, \ldots, \theta_K) = \frac{\Gamma(\sum_i n_i + 1)}{\prod_i \Gamma(n_i + 1)} \prod_{i=1}^{K} \theta_i^{n_i} \tag{26}$$

Here, n_i represents the number of nodes/players allocated to block i. On θ, the Dirichlet prior has the following form:

$$f(\theta_1, \ldots, \theta_K; \gamma_1, \ldots, \gamma_K) = \frac{1}{\mathrm{B}(\boldsymbol{\gamma})} \prod_{i=1}^{K} \theta_i^{\gamma_i - 1} \tag{27}$$

By marginalizing out θ, following the common practice in the literature, we can express the marginal distribution of z as:

$$p(\mathbf{y}|p, z) = \prod_{i<j}^{n-1} \binom{n_{ij}}{y_{ij}} p_{z_i, z_j}^{y_{ij}} (1 - p_{z_i, z_j})^{n_{ij} - y_{ij}} \tag{28}$$

$$p(\mathbf{z}|\boldsymbol{\gamma}) = \frac{\Gamma(\sum_{i=1}^{k} \gamma_i)}{\prod_{i=1}^{k} \Gamma(\gamma_i)} \frac{\prod_{i=1}^{k} \Gamma(n_i + \gamma_i)}{\Gamma(\sum_{i=1}^{k} (n_i + \gamma_i))} \tag{29}$$

where $\mathbf{z} = (z_1, z_2, \ldots, z_k)$ is a vector of labels for k different categories, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_k)$ is a vector of hyperparameters, and $\Gamma(\cdot)$ is the Gamma function. For the distribution of P we refer to the previous section in which we have detailed the specification.

Finally, we specify the prior distribution for the parameter K as:

$$p(K|\lambda = 1) = \frac{1^K e^{-1}}{K!} = \frac{1}{e \cdot K!} \propto \frac{1}{K!} \tag{30}$$

6. Estimation

6.1. **Updating z.**

$$r = \frac{\prod_{i<j} \binom{n_{ij}}{y_{ij}} p_{z_i' z_j'}^{y_{ij}} \cdot (1 - p_{z_i' z_j'})^{n_{ij} - y_{ij}} \cdot \frac{\Gamma(\gamma_0)\Gamma(n+1)}{\Gamma(n+\gamma_0)} \cdot \prod_{k=1}^{K} \frac{\Gamma(n_k' + \gamma_k)}{\Gamma(\gamma_k)\Gamma(n_k' + 1)}}{\prod_{i<j} \binom{n_{ij}}{y_{ij}} p_{z_i z_j}^{y_{ij}} \cdot (1 - p_{z_i z_j})^{n_{ij} - y_{ij}} \cdot \frac{\Gamma(\gamma_0)\Gamma(n+1)}{\Gamma(n+\gamma_0)} \cdot \prod_{k=1}^{K} \frac{\Gamma(n_k + \gamma_k)}{\Gamma(\gamma_k)\Gamma(n_k + 1)}} \tag{31}$$

$$= \frac{\prod_{i<j} p_{z_i' z_j'}^{y_{ij}} \cdot (1 - p_{z_i' z_j'})^{n_{ij} - y_{ij}} \cdot \prod_{k=1}^{K} \frac{\Gamma(n_k' + \gamma_k)}{\Gamma(\gamma_k)\Gamma(n_k' + 1)}}{\prod_{i<j} p_{z_i z_j}^{y_{ij}} \cdot (1 - p_{z_i z_j})^{n_{ij} - y_{ij}} \cdot \prod_{k=1}^{K} \frac{\Gamma(n_k + \gamma_k)}{\Gamma(\gamma_k)\Gamma(n_k + 1)}} \tag{32}$$

Passing to the log:

$$
\begin{aligned}
log(r) &= \log\left(\prod_{i<j} p_{z'_i z'_j}^{y_{ij}} \cdot (1-p_{z'_i z'_j})^{n_{ij}-y_{ij}} \cdot \prod_{k=1}^{K} \frac{\Gamma(n'_k+\gamma_k)}{\Gamma(\gamma_k)\Gamma(n'_k+1)}\right) \\
&\quad - \log\left(\prod_{i<j} p_{z_i z_j}^{y_{ij}} \cdot (1-p_{z_i z_j})^{n_{ij}-y_{ij}} \cdot \prod_{k=1}^{K} \frac{\Gamma(n_k+\gamma_k)}{\Gamma(\gamma_k)\Gamma(n_k+1)}\right) \\
&= \sum_{i<j} \left(y_{ij} \cdot \log p_{z'_i z'_j} + (n_{ij}-y_{ij}) \cdot \log\left(1-p_{z'_i z'_j}\right)\right) \\
&\quad + \sum_{k=1}^{K} \left(\log\left(\Gamma(n'_k+\gamma_k)\right) - \log\left(\Gamma(\gamma_k)\right) - \log\left(\Gamma\left(n'_k+1\right)\right)\right) \\
&\quad - \sum_{i<j} \left(y_{ij} \cdot \log p_{z_i z_j} + (n_{ij}-y_{ij}) \cdot \log\left(1-p_{z_i z_j}\right)\right) \\
&\quad - \sum_{k=1}^{K} \left(\log\left(\Gamma(n_k+\gamma_k)\right) - \log\left(\Gamma(\gamma_k)\right) - \log\left(\Gamma\left(n_k+1\right)\right)\right)
\end{aligned}
\tag{33}
$$

Algorithm 1 Updating z step

for $i \leftarrow 1$ to N **do**
  Sample `new_label` from $1, ..., K$
  Set $z' \leftarrow z$ with the i-th element replaced by `new_label`
  Compute new victory probabilities $p_{z'_i z'_j}$ using z'
  Compute probability ratio $log(r)$ using $p_{z'_i z'_j}$ and $p_{z_i z_j}$
  Set $\alpha_r \leftarrow \min(1, r)$
  Sample u from a uniform distribution on $(0, 1)$
  if $u < \alpha_r$ **then**
    Update z to z'
    Update $p_{z_i z_j}$ to $p_{z'_i z'_j}$
    Increment $acc.count_z$
  end if
  Store $z_{current}$ in $z.container$
end for

6.2. **Updating P.**

$$r = \frac{\prod_{i<j} \binom{n_{ij}}{y_{ij}} p'^{y_{ij}}_{z_i z_j} \cdot (1 - p'_{z_i z_j})^{n_{ij} - y_{ij}} \cdot \prod_{k=1}^{K} \left(\frac{1}{y'^{(k+1)} - y'^{(k)}} \right)^{|L'^{(k)}|}}{\prod_{i<j} \binom{n_{ij}}{y_{ij}} p^{y_{ij}}_{z_i z_j} \cdot (1 - p_{z_i z_j})^{n_{ij} - y_{ij}} \cdot \prod_{k=1}^{K} \left(\frac{1}{y^{(k+1)} - y^{(k)}} \right)^{|L^{(k)}|}} \tag{34}$$

(35)

Passing to the log:

(36)

$$log(r) = \sum_{i<j} \left(y_{ij} \cdot \log p'_{z_i z_j} + (n_{ij} - y_{ij}) \cdot \log \left(1 - p'_{z_i z_j}\right) \right) - \sum_{k=1}^{K} |L'^{(k)}| \cdot \log \left(y'^{(k+1)} - y'^{(k)} \right)$$

(37)

$$- \sum_{i<j} \left(y_{ij} \cdot \log p_{z_i z_j} + (n_{ij} - y_{ij}) \cdot \log \left(1 - p_{z_i z_j}\right) \right) + \sum_{k=1}^{K} |L^{(k)}| \cdot \log \left(y^{(k+1)} - y^{(k)} \right)$$

Algorithm 2 Updating P step

$j \leftarrow 1$
while $j \leq N_{iter}$ **do**
    Sample α' from a truncated normal distribution
    Generate a new proposal matrix P'
    Compute new victory probabilities $p'_{z_i z_j}$ using P' and $z_{current}$
    Compute probability ratio $log(r)$ using $p'_{z_i z_j}$ and $p_{z_i z_j}$
    Set $\alpha_r \leftarrow \min(1, r)$
    Sample u from a uniform distribution on $(0, 1)$
    if $u < \alpha_r$ **then**
        Update α to α'
        Update P to P'
        Update $p_{z_i z_j}$ to $p'_{z_i z_j}$
        Increment $acc.count_p$
    end if
    Store P in $P.container$
    Store α in $\alpha.container$
    $j \leftarrow j + 1$
end while

7. Point Estimate, Model Selection, and inference

While algorithmic methods produce a single estimated partition, our model offers the entire posterior distribution across different node partitions.

- Variation of Inforamation (VI): fully utilize this posterior and engage in inference directly within the partition space, we adopt the decision-theoretic approach introduced by Wade and Ghahramani (2018) for block modeling. This involves summarizing posterior distributions using the variation of information (vi) metric, developed by Meilă (2007), which measures the distance between two clusterings by comparing their individual and joint entropies. The vi metric ranges from 0 to log2 V, where V represents the number of nodes. Intuitively, the vi metric quantifies the amount of information contained in two clusterings relative to the shared information between them. As a result, it decreases towards 0 as the overlap between two partitions increases. Refer to Wade and Ghahramani (2018) for a detailed exploration of the key properties of the vi metric. Within this framework, a formal Bayesian point estimate for z is obtained by selecting the partition with the lowest averaged vi distance from the other clusterings
- WAIC
- Posterior predictive

Posterior predictive distribution. Once we have obtained the MCMC samples, we can retrieve the posterior predictive distribution.

While the WAIC yields practical and theoretical advantages and has direct connections with Bayesian leave-one-out cross-validation, thus providing a measure of edge predictive accuracy, the calculation of the WAIC only requires posterior samples of the log-likelihoods for the edges:

(38)
$$\log p(y_{ij}|z,\Theta) = y_{ij}\log\theta_{z_i,z_j} + (n_{ij}-y_{ij})\log(1-\theta_{z_i,z_j}), \quad i=2,\dots,N, \quad j=1,\dots,i-1.$$

Let us assume $i^\star$ is a new player . The posterior predictive is:

$$p(y_{i^\star j}|\mathbf{Y}) = \int p(z|\{\mathbf{Y}\setminus y_{i^\star}\})\cdot p(y_{i^\star j}|z)dz \tag{39}$$

$$\approx \frac{1}{T}\sum_{t=1}^{T}\sum_{z_i=1}^{K} p(y_{i^\star j}|z_j^{(t)}) \tag{40}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\sum_{z_{i^\star}=1}^{K} p_{z_{i^\star},z_j^{(t)}}^{y_{i^\star j}}\cdot(1-p_{z_{i^\star},z_j^{(t)}})^{n_{i^\star j}-y_{i^\star j}} \qquad \text{for } y_{i^\star j}=0,\dots,n_{i^\star j} \tag{41}$$

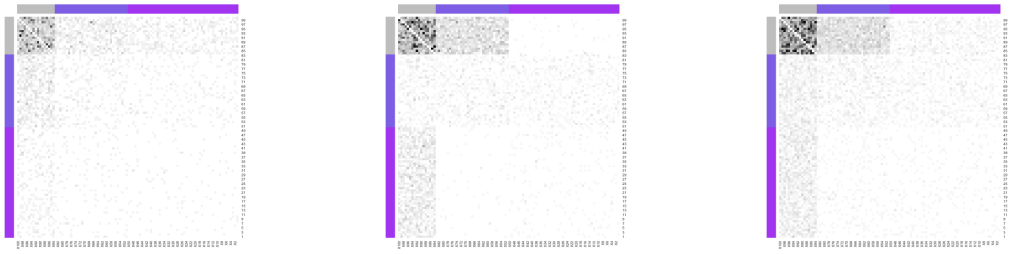
(42)

where $z_j^{(t)} \sim p(z|Y)$

8. Simulation Study

In order to evaluate how well our model performs in a situation similar to our intended use, and measure its advantages compared to the best existing alternatives, we generated three simulated tournaments with 100 players. Each tournament had a different number of games played among the participants. This quantity, denoted as M, essentially represents our sample size. We divided the players into three groups (K = 3) and used semi-symmetric

block probabilities in the simulations. In Figure (), we display the three simulated tournaments, where the difficulty of accurately determining the group membership increases as the number of games decreases.



(a) M4000.png (b) M7000.png (c) M10000.png

FIGURE 6. Images

TABLE 1. Y_{ij} drawn from a simple model

Method	Macro Columns			Row Names		
	$M = 3000$	$M = 10000$	$M = 30000$	Waic	VI distance	Predictive performance
Simple model	-	-	-	-	-	-
	-	-	-	-	-	-
	-	-	-	-	-	-
POMM model	-	-	-	-	-	-
	-	-	-	-	-	-
	-	-	-	-	-	-
Spectral Clustering	-	-	-	-	-	-
	-	-	-	-	-	-
	-	-	-	-	-	-
K-means	-	-	-	-	-	-
	-	-	-	-	-	-
	-	-	-	-	-	-
DBscan algorithm	-	-	-	-	-	-
	-	-	-	-	-	-
	-	-	-	-	-	-

TABLE 2. Y_{ij} drawn from a simple model

Method	Macro Columns			
	Measures	$M = 3000$	$M = 10000$	$M = 30000$
Simple model	Waic	-	-	-
	VI distance	-	-	-
	Predictive performance	-	-	-
POMM model	Waic	-	-	-
	VI distance	-	-	-
	Predictive performance	-	-	-
Spectral Clustering	Waic	-	-	-
	VI distance	-	-	-
	Predictive performance	-	-	-
K-means	Waic	-	-	-
	VI distance	-	-	-
	Predictive performance	-	-	-
DBscan algorithm	Waic	-	-	-
	VI distance	-	-	-
	Predictive performance	-	-	-

TABLE 3. Y_{ij} drawn from a simple model

Method	$M = 3000$			$M = 10000$			$M = 30000$		
	Waic	VI distance	Pred. perf.	Waic	VI distance	Pred. perf.	Waic	VI distance	Pred. perf.
Simple model	-	-	-	-	-	-	-	-	-
POMM model	-	-	-	-	-	-	-	-	-
Spectral Clustering	-	-	-	-	-	-	-	-	-
K-means	-	-	-	-	-	-	-	-	-
DBscan algorithm	-	-	-	-	-	-	-	-	-

DRAFT