Discovering latent association structure via Bayesian one-mode projection of temporal bipartite graphs

Ioannis Psorakis*1,2, Iead Rezek1, Zach Frankel1, and Stephen J. Roberts1

1 Pattern Analysis and Machine Learning Research Group, University of Oxford
2 Edward Grey Institute, University of Oxford

Abstract. We propose an extension to the notion of one-mode projection, for the case of temporal bipartite graphs. Through a Bayesian iterative update scheme, our method produces an estimate of the one-mode network at each step, by describing each link via probability distributions over i) its presence/absence and ii) weight. Our approach models the statistical significance of each link in the projected network, avoids overfitting and naturally handles noise and missing observations via probabilistic outputs.

1 Introduction

One-mode projection [2] is the operation by which a bipartite network $G = \{U, V, E\}$ described by the $N \times K$ incidence matrix $B$ is mapped to a graph with only one class of nodes, $\hat{G}_U = \{U, E_U\}$ via $B : N \times K \rightarrow W : N \times N$. The new connections are now placed between nodes of set $U$ based on the way they linked to nodes of the “vanished” set $V$. For example, consider a set $U$ of customers that performs purchases over a set $V$ of books and we seek to extract the similarity network between the individuals based on their reading preferences. In the present work, we seek the one-mode projection of a temporal bipartite network $G(t) = \{U, V(t), E(t)\}$, $t \in \{1, \ldots, T\}$ that is described by a sequence of incidence matrices $\{B(t)\}_{t=1}^T$. The one-mode projection at any given time point $t$ captures the associations between nodes $i, j \in U$, by taking into account past and present link information from all steps 1 to $t$. We require that all projected connections between nodes $i, j$ are appropriately weighted so that we take into account both the strength and the statistical significance of the association. Finally, we seek to model the uncertainty over the resulting topology, by placing probability distributions over the presence of each link.

2 Bayesian one-mode projection

Consider a bipartite graph sequence $\{B(t)\}_{t=1}^T$, where we isolate one particular timestamp $t$ so that $B = B(t)$. Each element $b_{ik}$ is 1 if agent $i$ links to target $k$ and zero otherwise. Let us now define an additional variable $x_{ij}$ that we will call opportunities, which is the number of target nodes that either $i$ or $j$ link to; that is obtained by performing an element-by-element logical disjunction on the rows of $B$ and summing the elements of the resulting vector $x_{ij} = \sum_{k=1}^K \text{OR}(b_{ik}, b_{jk})$. 

* ioannis.psorakis@eng.ox.ac.uk
From the observed $N \times K$ incidence matrix $B$, we extract the co-occurrence matrix via $W = BB^T$. Each $w_{ij} = \sum_{k=1}^{K} b_{ik}b_{jk}$ represents integer-valued counts that we can model as a draw from a binomial distribution $w_{ij} \sim \text{Binom}(\pi_{ij}x_{ij})$ with two parameters; the number of opportunities $x_{ij}$ and a bias term $\pi_{ij} \in [0, 1]$. This term is a latent attraction coefficient between all pairs $i,j$ that controls the extent to which opportunities $x_{ij}$ are manifested as co-occurrences $w_{ij}$ and can be interpreted as the connection probability for $i,j$ under a Bernoulli test. Our goal is to infer the attraction coefficients $\pi_{ij}$ for all dyads $i,j$, encoded in the matrix $\Pi \in \mathbb{R}^{N \times N}$.

Based on the binomial noise model, the probability of observing a particular number of co-occurrences, or link weight, $w_{ij}$ is given by:

$$P(w_{ij}|\pi_{ij}, x_{ij}) = \left(\frac{x_{ij}}{w_{ij}}\right)^{w_{ij}} \left(1 - \frac{x_{ij}}{w_{ij}}\right)^{x_{ij} - w_{ij}},$$

which is the likelihood function of the observed weights $w_{ij}$. As our inference task is to describe the attraction coefficient $\pi_{ij}$ given the known $w_{ij}, x_{ij}$, we employ a Bayesian approach by working with the probability distribution over $\pi_{ij}$:

$$P(\pi_{ij}|w_{ij}, x_{ij}) = \frac{P(w_{ij}|\pi_{ij}, x_{ij})P(\pi_{ij})}{\int_0^1 P(w_{ij}|\pi_{ij}, x_{ij})d\pi_{ij}},$$

where $P(\pi_{ij})$ is the prior and expresses our belief on how the attraction coefficient for $i,j$ varies before observing $w_{ij}$ and $x_{ij}$. On the other hand, the posterior $P(\pi_{ij}|w_{ij}, x_{ij})$ is the revised belief on $\pi_{ij}$ under the light of these observations. Because $\pi_{ij} \in [0, 1]$ we can model $P(\pi_{ij})$ as a Beta distribution $\pi_{ij} \sim \text{Beta}(\alpha_{ij}, \beta_{ij})$ parameterised by $\alpha_{ij}$ and $\beta_{ij}$, so that:

$$P(\pi_{ij}) = \frac{\pi_{ij}^{\alpha_{ij}-1}(1 - \pi_{ij})^{\beta_{ij}-1}}{\int_0^1 u^{\alpha_{ij}-1}(1 - u)^{\beta_{ij}-1}du}.$$ 

Based on Eq. (4) we combine our prior in Eq. (3) with the likelihood from Eq. (1) to get the posterior:

$$P(\pi_{ij}|w_{ij}, x_{ij}) = \text{Beta}(\alpha_{ij} + w_{ij}, \beta_{ij} + x_{ij} - w_{ij})$$

which is a revised Beta distribution over $\pi_{ij}$, with updated parameters:

$$\alpha'_{ij} = \alpha_{ij} + w_{ij}$$

$$\beta'_{ij} = \beta_{ij} + x_{ij} - w_{ij}$$

The posterior distribution $P(\pi_{ij}|w_{ij}, x_{ij}) = \text{Beta}(\alpha'_{ij}, \beta'_{ij})$ provides all the information we need to describe the attraction coefficient $\pi_{ij}$, capturing the uncertainty over each possible value in $[0, 1]$, while all dependencies between links are encoded in the $w_{ij}$ and $x_{ij}$ terms. To define the projection matrix $\Pi \in \mathbb{R}^{N \times N}$, fixed-point estimates can be directly derived from the posterior. For this particular study we have used the expected value $E(\pi_{ij}) = \frac{\alpha_{ij}}{\alpha_{ij} + \beta_{ij}}$ for each element of $\Pi \in \mathbb{R}^{N \times N}$.

Having a fully probabilistic formulation for the attraction coefficient, we can proceed one more step further and “integrate out” $\pi_{ij}$ from the likelihood function in Eq. (4) in order to obtain the probability distribution over the connection weight $w_{ij}$:
\[
P(w_{ij}|x_{ij}, \alpha'_{ij}, \beta'_i) = \binom{x_{ij}}{w_{ij}} \frac{B(w_{ij} + \alpha'_{ij}, x_{ij} - w_{ij} + \beta'_i)}{B(\alpha'_{ij}, \beta'_i)}, \tag{7}
\]

which is a Beta-binomial probability density function and \(B(.,.)\) is the standard beta function. This distribution captures the variability of co-occurrences \(w_{ij}\) given our noise model and past observations. From the above equation we can estimate the expected value for the weights as \(E(w_{ij}) = \frac{x_{ij}\alpha_{ij}}{\alpha_{ij} + \beta_{ij}}\).

We have described the theoretical foundation of our model along with the one-mode projection scheme for a single learning step \(t\). The full process involves starting with initial values for \(\alpha^{(0)}_{ij}, \beta^{(0)}_{ij}\) and then cycling through the update equations Eq. (5), (6) for each time step, as described in Algorithm 1.

**Algorithm 1 Bayesian One-Mode Projection**

**Require:** bipartite sequence \(\{B^{(t)}\}_{t=1}^T\)

1. Initialise \(\alpha^{(0)}_{ij}, \beta^{(0)}_{ij}, \forall i, j \in \{1, ..., N\}\)
2. for \(t = t_0\) to \(T\) do
3. Set \(B = B^{(t)}\)
4. Get opportunities \(x^{(t)}_{ij}\) via \(x_{ij} = \sum_{h=1}^K \text{OR}(b_{ih}, b_{jh})\)
5. Get co-occurrences via \(W^{(t)} = BB^T\)
6. for \(i, j \in \{1, ..., N\}\) do
7. update \(\alpha^{(t)}_{ij}\) from Eq. (5)
8. update \(\beta^{(t)}_{ij}\) from Eq. (6)
9. \(E^{(t)}(\pi_{ij}) = \frac{\alpha^{(t)}_{ij}}{\alpha^{(t)}_{ij} + \beta^{(t)}_{ij}}\)
10. \(E^{(t)}(w_{ij}) = \frac{x^{(t)}_{ij}\alpha^{(t)}_{ij}}{\alpha^{(t)}_{ij} + \beta^{(t)}_{ij}}\)
11. end for
12. end for
13. return \(\Pi^{(t)} = [E^{(t)}(\pi_{ij})]_{i,j \in N}, \{\alpha^{(t)}_{ij}, \beta^{(t)}_{ij}, E^{(t)}(w_{ij})\}, \forall t \in \{1, ..., T\}\)

### 3 Example and discussion

Our main application area for this methodological development is the study of animal social networks. By exploiting field data, where individuals are observed at various sites across time, we seek to infer the latent affiliation graph based on animal co-occurrences. For example, consider the following simulated dataset. We generate a bipartite sequence \(\{B^{(t)}\}_{t=1}^{100}\) where each \(B^{(t)}\) is a draw from a template matrix \(B^{\text{seed}}\), so that \(b^{(t)}_{ik} \sim \text{Bernoulli}(i_{ik}^{\text{seed}})\).
\[ B^{\text{seed}} = \begin{bmatrix}
0.80 & 0.90 & \epsilon & \epsilon \\
0.90 & 0.70 & \epsilon & \epsilon \\
0.90 & 0.80 & 0.90 & 0.90 \\
\epsilon & \epsilon & 0.80 & 0.90 \\
\epsilon & \epsilon & 0.60 & 0.90
\end{bmatrix} \]

draw sample at \( t \rightarrow B^{(t)} = \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1
\end{bmatrix}\]

Such dataset can represent \( N = 5 \) animals that may visit every day \( K = 4 \) locations, across a period of \( T = 100 \) days while \( \epsilon \) is a very small number. Pairs 1 & 2 and 3 & 4 tend to co-appear quite frequently, while 3 tends to visit all sites, perhaps due to its gregariousness. For each day \( t \), our goal is to learn the associations between each pair \( i, j \), by exploiting current and past observations.

We start at \( t = 0 \) by assuming no prior knowledge on any link structure; we set \( \alpha_{ij}^{(0)} = \beta_{ij}^{(0)} = 10 \) that gives \( \mathbb{E}[\pi_{ij}^{(0)}] = 0.5 \ \forall i, j \), denoting that we are unable to tell if there should be a link between \( i, j \) before seeing any data. We proceed based on Algorithm 1 by calculating the opportunities \( x_{ij}^{(t)} \) and co-occurrences \( w_{ij}^{(t)} \) from \( B^{(t)} \) and update \( \alpha_{ij}^{(t)}, \beta_{ij}^{(t)} \) accordingly. The distribution \( P(\pi_{ij}^{(t)} | w_{ij}^{(t)}, x_{ij}^{(t)}, \alpha_{ij}^{(t)}, \beta_{ij}^{(t)}) \) over \( \pi_{ij}^{(t)} \) is revised accordingly at every step, effectively learning the attraction coefficient between \( i \) and \( j \).

In Fig. 1(a) we plot how the posterior distribution \( P(\pi_{12} | w_{12}, x_{12}, \alpha_{12}, \beta_{12}) \) of the attraction coefficient \( \pi_{12}^{(t)} \) progresses during each iteration. We can see that for \( t = 0 \) the distribution is our flat prior centered around 0.5, because we have no evidence to strongly support either the presence or absence of a link between nodes 1 and 2. As we start observing co-occurrences \( w_{12} \neq 0 \) this prior belief is updated in order to explain the incoming data, effectively shifting the distribution around larger values of \( \pi_{12} \). It is important to notice that the increase of \( \mathbb{E}[\pi_{12}^{(t)}] \) is less steep at later \( t \), as the impact of new co-occurrences \( i, j \) is not so strong as in the beginning of data collection. Such important saturation or “diminishing returns” property arises naturally in Bayesian learning models, without the need to explicitly induce it via additional machinery such as hyperbolic tangent functions [1].

On the other hand, although node 2 tends to co-occur with 3 as it does with node 1, their number of opportunities \( x_{23} \) are much higher due to the participation of node 3 across all \( K \). For that reason the model estimates a lower value of the attraction coefficient \( \pi_{23} \), effectively penalising such lack of exclusivity in the co-appearances of 2 with 3. This is a very attractive property of the model, which not only regulates the link weights between “gregarious” nodes (that tend to link to everywhere) and “selective” ones (with a small set of targets they point to), but also allows the model to downplay the effect of purely coincidental co-appearances, which would otherwise introduce “junk” associations in the projection network.

The probabilistic outputs, noise-handling, incorporation of past data, saturation effect and regulation of gregariousness, make the method a promising step forward towards analysing systems such as ecological, citation, co-purchasing and recommendation networks.
Posterior density over link presence between nodes 1 & 2, across time steps
$P(\pi_{12} | w_{12}, x_{12}, \_12, `12)$

Posterior density over link presence between nodes 2 & 3, across time steps
$P(\pi_{23} | w_{23}, x_{23}, \_23, `23)$

Fig. 1. Based on the example dataset, we demonstrate how our method models two different kinds of relationships; in Fig. 1(a) we have two individuals that consistently point to the same target nodes (sites) in the bipartite graph $B(t)$, thus our belief over their attraction coefficient $\pi_{12}$ increases as we observe more data. On the other hand, in Fig. 1(b) although the visitations of 2 & 3 overlap, individual 3 participates across all sites. Such lack of exclusivity in co-appearances prevents the model from increasing the probability of link between nodes 2 and 3. For both cases we have started with $\pi_{ij}$ centered around 0.5, denoting a vague (unbiased) prior.

References