The Deep Zero-Inflated Latent Position Block Model for the Clustering of Nodes in Graphs

Seydina Ousmane Niang, Charles Bouveyron, Marco Corneli, Pierre Latouche

The Deep Zero-Inflated Latent Position Block Model for the Clustering of Nodes in Graphs

S. NIANG[1] & C. BOUVEYRON[1] & M. CORNELI[1,2] & P. LATOUCHE[3,4]

[1] *Université Côte d'Azur, Inria, CNRS, Laboratoire J.A.Dieudonné, Maasai team, Nice, France*
[2] *Université Côte d'Azur, Laboratoire CEPAM, Nice, France*
[3] *Université Clermont Auvergne, CNRS, Laboratoire LMBP, Aubière, France*
[4] *Institut Universitaire de France (IUF)*

Résumé. L'augmentation des capacités de stockage a entraîné une explosion des données, rendant les réseaux essentiels pour modéliser les relations entre objets (nœuds). Ces réseaux, souvent complexes, nécessitent des méthodes efficaces de clustering et de visualisation pour en résumer l'information. Le deep latent position block model (Deep-LPBM), conçu pour les réseaux binaires, combine clustering par blocs et représentation latente continue pour visualiser les nœuds. Nous proposons ici une extension, le deep zero-inflated latent posistion block model (Deep-ZLPBM), adapté aux réseaux non binaires, où la matrice d'adjacence peut prende des valeurs entières. Ce modèle repose sur un auto-encodeur variationnel intégrant un réseau de convolution de graphes (GCN) et un décodeur utilisant une loi ZIP (zero-inflated Poisson). L'inférence se concentre sur la maximisation de la vraisemblance marginale, et l'optimisation s'effectue par descente de gradient.

Mots-clés. Clustering de nœuds, Auto-encodeur variationnel pour graphe, Modélisation par blocs, Visualisation de graphes, Zero-inflatted Poisson

Abstract. The evolution in storage capacities has led to a data explosion, making networks essential for modeling relationships between objects (nodes). These complex networks require effective clustering and visualization methods to summarize and interpret their information. The deep latent position block model (Deep-LPBM), designed for binary networks, combines partial block-based clustering and continuous latent representation to visualize nodes. Here, we propose an extension, the deep zero-inflated latent position block model (Deep-ZLPBM), designed for non-binary networks, where the entries of the adjacency matrix can take integer values. This model is based on a deep variational autoencoder that integrates a graph convolutional network (GCN) and a decoder leveraging a zero-inflated Poisson (ZIP) distribution. Inference relies on the maximization of the marginal likelihood, and optimization is performed using stochastic gradient descent.

Keywords. Nodes clustering, Graph variational autoencoder, Block modelling, Graph visualization, Zero-inflated Poisson

1 Introduction

Networks represent connections among entities, such as individuals, organizations, or biological elements. Nodes are entities and edges characterize their relationships. Social

networks are a common example, with nodes representing individuals and edges representing friendship or collaborations. As network data grows in scale and complexity, understanding these interconnections becomes increasingly challenging. The interdependence between nodes creates heterogeneous connectivity patterns, requiring models that balance complexity and interpretability. Clustering nodes [1] is a key approach in network analysis, offering a high-level summary of complex networks. While many studies focus on binary networks, where connections are either present or absent, valued networks are equally important. These networks capture rich details, such as interaction frequencies or relationship strengths. This paper explores model-based clustering for valued networks, where connections are modeled probabilistically.

Historical models for binary networks [2, 3] relied on simple probabilistic assumptions, like fixing edge probability constant. Yet, real-world networks exhibit characteristics that defy these simplified approaches. Pioneering works allowed for the introduction of advanced models and among them, latent variable models (LVMs) have gained significant attention. LVMs assume that interconnections between nodes depend on unobserved latent variables. The earliest and most popular LVMs for node clustering include the latent position cluster model (LPCM) [4] and the stochastic block model(SBM) [5]. The SBM efficiently clusters nodes into groups and supposes that interactions between two nodes only depend on the clusters they belong to, but does not allow any direct latent representation (visualisation) of the underlying structure of the graph. The LPCM allows to have a direct latent visualisation of the network but this clustering approach only excels at community detection and lacks the flexibility to capture more diverse connectivity patterns. Recent developments in this area include [6], which uses deep variational inference procedure and extends the LPCM to textual data analysis.

One limitation of both SBMs and LPCMs is the assumption they make that each node belongs to a single group. In reality, individuals often belong to multiple overlapping groups, reflecting the multiple roles they play within several groups. Many works have been proposed to account for such partial memberships [7, 8].
Recently, in [9], the deep latent position block model (Deep-LPBM) was proposed to combine the strengths of SBM as well as LPCM, and to deal with partial memberships. Deep-LPBM enables both latent visualization and partial membership block modeling, providing a more versatile clustering framework than traditional positional clustering methods.

In this work, we introduce the deep zero inflated latent position block model (Deep-ZLPBM). This framework is built within the class of latent position block models, extending [9] to the analysis of valued networks. Several works have also attempted to extend binary networks methods to integer-valued networks. For example in [10], the degree-corrected stochastic block model (DCSBM) is proposed. DCSBM extends the traditional SBM to valued networks by mainly replacing the Bernoulli distribution by a Poisson [11]. Many other works have adopted the same strategy [12, 13].

2 Generative models

In this work, we suppose that the graph is undirected and identified by its adjacency matrix A, a square $N \times N$ symmetric matrix such that its element $A_{ij} = A_{ji} \in \mathbb{N}$.

2.1 Generative model for the graph

We assume that the N nodes belong to Q clusters in a non exclusive manner and that each node node i belongs to cluster q with a probability η_{iq} as a function of the latent variable $Z_i \overset{\text{iid}}{\sim} \mathcal{N}_d(0, \mathbb{I}_{Q-1})$. The set of node embeddings is denoted by $Z := \{Z_i\}_i$ in the rest of the paper.

To link the latent representation of the nodes Z_i with partial membership probabilities $\eta_i \in \Delta_Q := \{x \in [0,1]^Q; \sum_{q=1}^Q x_q = 1\}$, we rely on the bijective softmax transformation $h : \mathbb{R}^{Q-1} \to \Delta_Q$ such that:

$$\eta_{iq} := h(Z_i)_Q := \begin{cases} \frac{\exp(Z_{iq})}{1+\sum_{r=1}^{Q-1}\exp(Z_{ir})} & \text{if} \quad q \neq Q, \\ \frac{1}{1+\sum_{r=1}^{Q-1}\exp(Z_{ir})} & \text{if} \quad q = Q, \end{cases} \tag{1}$$

and we denote $\eta = (\eta_1, \ldots, \eta_N)^\intercal$. The mapping h aims at encoding Z into cluster partial membership probabilities. We assume that the probability of connection between two nodes follows a zero-inflated Poisson distribution [14] with parameters depending on η such that:

$$\begin{aligned} A_{ij}|Z_i, Z_j &\sim \text{ZIP}(\eta_i^\intercal \Pi \eta_j, \eta_i^\intercal \Lambda \eta_j), \\ p(A, Z|\Pi, \Lambda) &= \left(\prod_{i=1}^N p(Z_i)\right) p(A|Z, \Pi, \Lambda), \end{aligned} \tag{2}$$

where, adapting the simplified notation $\prod_{i<j} = \prod_{i=1}^{N-1}\prod_{j=i+1}^{N}$, we have

$$\begin{aligned} p(A|Z, \Lambda, \Pi) &= \prod_{i<j} p(A_{ij}|Z, \Pi, \Lambda) \\ &= \prod_{\substack{i<j \\ A_{i,j}=0}} \left[1 - \eta_i^\intercal \Pi \eta_j (1 - \exp\left(-\eta_i^\intercal \Lambda \eta_j\right))\right] \prod_{\substack{i<j \\ A_{i,j}>0}} \left[\eta_i^\intercal \Pi \eta_j \, \frac{(\eta_i^\intercal \Lambda \eta_j)^{A_{ij}}}{A_{ij}!} \exp(-\eta_i^\intercal \Lambda \eta_j)\right] \end{aligned} \tag{3}$$

where the $Q \times Q$ matrix $\Pi = (\Pi_{qr})_{1 \leq q,r \leq Q}$ refers to the connectivity matrix whose entry (q, r) is the probability that a node in block q is connected to a node in block r (i.e. the corresponding value of the adjacency matrix is not zero). The parameter Λ is also a $Q \times Q$ matrix whose entry (q, r) refers to the expectation of A_{ij} with i and j respectively belonging to clusters q and r.

.

3 Variational inference and optimisation

The marginal log-likelihood of the graph is given by

$$\log p(A|\Pi, \Lambda) = \log \int_Z p(A, Z|\Pi, \Lambda) dZ.$$

Unfortunately, this quantity cannot be computed analytically. Consequently, we propose to rely on a variational inference strategy for approximation purposes. For any distribution $q(\cdot)$ for the latent variables in Z, in force of the Jensen's inequality, the following

decomposition holds:

$$\log p(A|\Pi,\Lambda) = \log \int_Z \frac{p(A,Z|\Pi,\Lambda)}{q(Z)} q(Z)dZ = \log \mathbb{E}_{q(Z)}\left[\frac{p(A,Z|\Pi)}{q(Z)}\right]$$
$$\geq \mathbb{E}_{q(Z)}\left[\log \frac{p(A,Z|\Pi)}{q(Z)}\right] := \mathcal{L}(\Pi,\Lambda,q(\cdot)).$$

In this paper, we refer to $\mathcal{L}(\Pi,\Lambda,q(\cdot))$ as the evidence lower bound (ELBO). Furthermore, the exact difference between $\log p(A|\Pi,\Lambda)$ and $\mathcal{L}(\Pi,\Lambda,q(\cdot))$ is

$$\mathrm{KL}(q(\cdot)||p(Z|A,\Pi)) = \log p(A|\Pi,\Lambda) - \mathcal{L}(\Pi,\Lambda,q(\cdot)), \tag{4}$$

where the Kullback-Leibler (KL) divergence between q and the posterior distribution $p(Z|A,\Pi)$ is always non-negative, indicating that the ELBO is a lower bound of the marginal log-likelihood. Since the marginal log-likelihood does not depend on $q(\cdot)$, maximizing the ELBO with respect to $q(\cdot)$ is equivalent to minimizing the Kullback-Leibler divergence between $q(\cdot)$ and the posterior distribution. This minimisation is satisfied when the variational $q(\cdot)$ is equal to the true posterior distribution $p(Z|A,\Pi)$ which is not tractable here, so we adopt a mean-field approximation and other hypotheses to make the ELBO tractable:

$$q(Z) := q(Z|A) = \prod_{i=1}^{N} q_\phi(Z_i|A) = \prod_{i=1}^{N} \mathcal{N}(Z_i; \mu_\phi(\bar{A})_i, \sigma^2_\phi(\bar{A})_i \mathbb{I}_{Q-1}), \tag{5}$$

with $\left[\mu_\phi(\bar{A}), \sigma_\phi(\bar{A})\right] = f_\phi(\bar{A})$, where $f_\phi : \mathbb{R}^{N\times N} \to \mathbb{R}^Q$ is a graph convolutional network (GCN, [15]) parameterized by ϕ that maps the normalized adjacency matrix $\bar{A} = \tilde{D}^{-\frac{1}{2}}(A + I_N)\tilde{D}^{-\frac{1}{2}}$ to the vector of the variational means and the log standard deviations. Here, $\tilde{D}$ is the diagonal matrix representing the degree of nodes with respect to the matrix $(A+I_N)$, defined as $\tilde{D}_{ii} = 1 + \sum_{j=1}^{N} A_{ij}$. In Equation (5), $\mu_\phi(\bar{A})_i$ denotes the i-th row of $\bar{\mu}_\phi(A)$, and $\bar{\sigma}^2_\phi(A)_i$ denotes the i-th element of $\sigma^2_\phi(A)$. Thus, the ELBO can be decomposed as follows:

$$\begin{aligned}
\mathcal{L}(q,\Lambda,\Pi) &= \mathbb{E}_{q(\cdot)}\left[\log \frac{p(A,Z|\Pi,\Lambda)}{q(Z|A)}\right] \\
&= \mathbb{E}_{q(\cdot)}\left[\log \frac{p(A|Z,\Pi,\Lambda)p(Z)}{q(Z|A)}\right] \\
&= \underbrace{\mathbb{E}_{q(\cdot)}\left[\log p(A|Z,\Pi,\Lambda)\right]}_{\text{Reconstruction term}} - \underbrace{\mathrm{KL}(q(Z|A)||p(Z))}_{\text{Regularisation term}}.
\end{aligned}$$

Finally:

$$\begin{aligned}
\mathcal{L}(q,\Pi,\Lambda) = \mathbb{E}_{q(\cdot)}\Big[&\sum_{\substack{i<j \\ A_{i,j}=0}} \log(1 - \eta_i^\intercal \Pi \eta_j (1 - \exp(-\eta_i^\intercal \Lambda \eta_j))) \\
&+ \sum_{\substack{i<j \\ A_{i,j}>0}} \left[\log(\eta_i^\intercal \Pi \eta_j) + A_{ij}\log(\eta_i^\intercal \Lambda \eta_j) - \log(A_{ij}!) - \eta_i^\intercal \Lambda \eta_j\right]\Big] - \\
&\sum_i \left[(Q-1)\log \sigma_\phi(\bar{A})_i - \frac{Q-1}{2} + \frac{1}{2}\left|\left|\mu_{\phi,\xi}(\bar{A})_i\right|\right|_2^2 + \frac{Q-1}{2}\sigma_\phi(\bar{A})_i^2\right].
\end{aligned} \tag{6}$$

The regularization term, which involves the KL divergence between q and the prior distribution, can be computed exactly. The main challenge lies in calculating the reconstruction term. To address this and optimize the ELBO, we introduce a Monte Carlo gradient descent algorithm, detailed in the next section.

3.1 Monte Carlo variational optimisation

The model and variational parameters cannot be updated with analytical formulas because of the integral involving the variational distribution $q(\cdot)$ in the ELBO. In this section, we aim at deriving estimates $\tilde{\mathcal{L}}(\cdot)$ of the ELBO $\mathcal{L}(\cdot)$ to perform Monte Carlo gradient descent:

$$\begin{aligned}\tilde{\mathcal{L}}(q,\Pi,\Lambda) = \frac{1}{S}\sum_{s=1}^{S}\Bigg[&\sum_{\substack{i<j\\A_{i,j}=0}} \log\Big(1-\eta_i^{\intercal(s)}\Pi\eta_j^{(s)}\big(1-\exp\big(-\eta_i^{\intercal(s)}\Lambda\eta_j^{(s)}\big)\big)\Big)\\ &+\sum_{\substack{i<j\\A_{ij}>0}}\Big[\log\Big(\eta_i^{\intercal(s)}\Pi\eta_j^{(s)}\Big)+A_{ij}\log\Big(\eta_i^{\intercal(s)}\Lambda\eta_j^{(s)}\Big)-\log(A_{ij}!)-\eta_i^{\intercal(s)}\Lambda\eta_j^{(s)}\Big]\Bigg]\\ &-\sum_i\left[(Q-1)\log\sigma_\phi(\bar{A})_i-\frac{Q-1}{2}+\frac{1}{2}\left\|\mu_\phi(\bar{A})_i\right\|_2^2+\frac{Q-1}{2}\sigma_\phi(\bar{A})_i^2\right],\end{aligned} \tag{7}$$

where $\eta_i^{(s)} = h(Z_i^{(s)})$ and $Z_i^{(s)} \sim q(\cdot)$. This usual (naïve) Monte Carlo gradient estimator exhibits very high variance and is then impractical [16]. To deal with this problem, as done in [17], we adopt the reparameterization trick. In particular, if $\epsilon^{(s)} \sim \mathcal{N}(0, \mathbb{I}_{Q-1})$, then $Z_i^{(s)} = \mu_\phi(\bar{A})_i \ + \sigma_\phi(\bar{A})_i \ \epsilon^{(s)} \sim \mathcal{N}(\mu_\phi(\bar{A})_i, \sigma_\phi(\bar{A})_i^2\mathbb{I}_{Q-1}) = q(\cdot|A)$. To simplify the gradient descent, we map the constrained values $(\Pi_{qr})_{q,r}$ (resp $(\Lambda_{qr})_{q,r}$) from the interval $]0,1[$ (resp $]0,+\infty[$) to the unconstrained set $\mathbb{R}$ using the bijective function g (resp f) defined as:

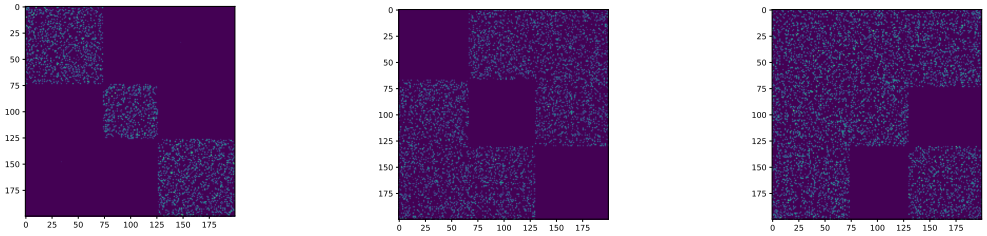
$$g: \begin{cases}\mathbb{R}\to]0,1[\\ x\mapsto 0.5+\frac{1}{\pi}\arctan(x)\end{cases} \qquad \text{resp} \quad f: \begin{cases}\mathbb{R}\to]0,\infty[\\ x\mapsto \exp(x)\end{cases}$$

.

4 Evaluation on synthetic datasets

To evaluate the ability of our method to represent diverse network topologies, we tested it on three different network structures, each consisting of 200 nodes spread into $Q = 3$ clusters. For the connectivity matrix Π, we considered three different network structures each reflecting a kind of social interaction: i) the **community structure**, where we assume that nodes in the same group have a high probability of connection indicated by β and that nodes in different groups have a low probability of connection indicated by δ, ii) the **disassortative structure** where we assume that nodes in different clusters have a high probability of connection denoted by β, and that nodes in the same cluster have a low probability of connection denoted by δ, iii) a **hub structure** where we assume that one of the clusters is highly connected to all the clusters, with a probability β and that the other clusters are communities. For the expectation matrix Λ, we set for all $q < r$,

$\Lambda_{qr} = \alpha$ and $\Lambda_{qq} = \sigma$.
To generate the adjacency A, we fix the labels and one-hot them to build κ such that $\kappa_{iq} = 1$ if node i is in cluster q and 0 otherwise. We then generate $M \sim Bernoulli(\cdot|1; \kappa^\intercal \Pi \kappa)$ and $N \sim Poisson(\cdot|\kappa^\intercal \Lambda \kappa)$ and set $A = M \times N$ (Hadamard product). In Figure 1, we plot some adjacency matrices, reorganized by blocs, of the three different structures presented above with different connectivity levels $\beta = 0.2, \delta = 0.01, \alpha = 5$ and $\sigma = 0.03$.
The adjusted rand index (ARI, [18]) serves as our primary measure of clustering accuracy, reflecting the similarity between the true and inferred node partitions. An ARI of 0 suggests clustering is no better than random, while an ARI close to 1 indicates alignment with true node labels up to label switching. In Table 1, we show the performance of Deep-ZLPBM on three different connectivity levels.



Community Disassortative Hub

Figure 1: Examples of adjacency matrices corresponding to the three structures and connectivity levels $\beta = 0.2, \delta = 0.01, \alpha = 5$ and $\sigma = 0.03$ and reorganised by cluster.

Connectivity levels	ARI		
	Communities	Disassortative	Hub
$\beta = 0.3, \delta = 0.01, \alpha = 5, \sigma = 0.03$	1 ± 0.00	1 ± 0.00	0.98 ± 0.01
$\beta = 0.2, \delta = 0.01, \alpha = 5, \sigma = 0.03$	1 ± 0.00	1 ± 0.00	0.8 ± 0.09
$\beta = 0.1, \delta = 0.01, \alpha = 5, \sigma = 0.03$	0.96 ± 0.08	0.48 ± 0.23	0.2 ± 0.13

Table 1: ARI of Deep–ZLPBM on different connectivity levels.

When the connectivity levels are high ($\beta \in \{0.3, 0.2\}$), in both communities and disassortative structures, Deep-ZLPBM efficiently recovers the true nodes partition (ARI = 1). In the Hub configuration, it is harder for our method to recover all the clusters (ARI $\in \{0.98, 0.8\}$). For $\beta = 0.1$, the problem becomes much harder. Deep-ZLPBM performs well in the community structure but struggle with hub and disassortative.

5 Conclusion

This paper extended the Deep-LPBM, a new methodology combining a block model with latent position model adapted to valued networks. A preliminary study with simulated data illustrated the capacity of our methodology to clusters nodes in valued networks. This is an on going work and an extensive benchmark will be provided if the paper were to be accepted.

References

[1] S. E. Schaeffer, "Graph clustering," *Computer science review*, vol. 1, no. 1, pp. 27–64, 2007.

[2] O. Erdős and A. Renyi, "On random graphs," *Publ. Math*, vol. 6, pp. 290–297, 1959.

[3] P. W. Holland and S. Leinhardt, "An exponential family of probability distributions for directed graphs," *Journal of the american Statistical association*, vol. 76, no. 373, pp. 33–50, 1981.

[4] P. D. Hoff, A. E. Raftery, and M. S. Handcock, "Latent space approaches to social network analysis," *Journal of the american Statistical association*, vol. 97, no. 460, pp. 1090–1098, 2002.

[5] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.

[6] R. Boutin, P. Latouche, and C. Bouveyron, "The deep latent position topic model for clustering and representation of networks with textual edges," *arXiv preprint arXiv:2304.08242*, 2023.

[7] K. A. Heller, S. Williamson, and Z. Ghahramani, "Statistical models for partial membership," in *Proceedings of the 25th International Conference on Machine learning*, pp. 392–399, 2008.

[8] E. M. Airoldi, D. M. Blei, S. E. Fienberg, E. P. Xing, and T. Jaakkola, "Mixed membership stochastic block models for relational data with application to protein-protein interactions," in *Proceedings of the international biometrics society annual meeting*, vol. 15, p. 1, 2006.

[9] R. Boutin, P. Latouche, and C. Bouveyron, "The deep latent position block model for the block clustering and latent representation of networks," *arXiv preprint arXiv:2412.01302*, 2024.

[10] B. Karrer and M. E. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 1, p. 016107, 2011.

[11] P. C. Consul and G. C. Jain, "A generalization of the poisson distribution," *Technometrics*, vol. 15, no. 4, pp. 791–799, 1973.

[12] T. Herlau, M. N. Schmidt, and M. Mørup, "Infinite-degree-corrected stochastic block model," *Physical review E*, vol. 90, no. 3, p. 032819, 2014.

[13] C. Aicher, A. Z. Jacobs, and A. Clauset, "Learning latent block structure in weighted networks," *Journal of Complex Networks*, vol. 3, no. 2, pp. 221–248, 2015.

[14] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.

[15] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[16] J. Paisley, D. Blei, and M. Jordan, "Variational bayesian inference with stochastic search," *arXiv preprint arXiv:1206.6430*, 2012.

[17] D. P. Kingma, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[18] K. Y. YEUNG and W. L. RUZZO, "Details of the adjusted rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data," *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.