



Importance weighted directed graph variational auto-encoder for block modelling of complex networks

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

► To cite this version:

Seydina Ousmane Niang, Charles Bouveyron, Marco Corneli, Pierre Latouche. Importance weighted directed graph variational auto-encoder for block modelling of complex networks. 2025. hal-05077099

HAL Id: hal-05077099

<https://hal.science/hal-05077099v1>

Preprint submitted on 21 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Importance weighted directed graph variational auto-encoder for block modelling of complex networks

NIANG S. Ousmane*

University Côte d’Azur, Inria, CNRS,
Laboratoire J.A.Dieudonné,
Maasai team, Nice, France
seydina-ousmane.niang@inria.fr

BOUVEYRON Charles

University Côte d’Azur, Inria, CNRS,
Laboratoire J.A.Dieudonné,
Maasai team, Nice, France
charles.bouveyron@inria.fr

LATOUCHE Pierre

Université Clermont Auvergne, CNRS,
Laboratoire LMBP, Aubière, France
Institut Universitaire de France (IUF)
pierre.latouche@inria.fr

CORNELI Marco

University Côte d’Azur, Inria, CNRS,
Laboratoire CEPAM,
Maasai team, Nice, France
marco.corneli@inria.fr

Abstract

This paper addresses the fundamental challenges of jointly performing node clustering and representation learning in directed and valued graphs, which need both global and local network structures to be captured. While these two tasks are highly interdependent, they are often treated separately in existing works. We propose the deep zero-inflated latent position block model (Deep-ZLPBM) in the context of directed and valued networks characterized by non-symmetric adjacency matrices with positive integer entries. Our approach leverages a variational autoencoder (VAE) framework, combining a directed graph neural network (DirGNN) encoder designed to handle directed edges and a zero-inflated Poisson (ZIP) block modelling decoder to model sparse, integer-weighted interactions. Recognizing the limitations of the standard evidence lower bound (ELBO) in VAEs, we explore the importance weighted ELBO (iw-ELBO), a tighter bound on the marginal log-likelihood optimized via gradient ascent, to enhance inference. Extensive experiments on synthetic datasets demonstrate that iw-ELBO optimization yields significant performance gains. Moreover, our results validate that Deep-ZLPBM effectively models complex network structures, providing interpretable partial memberships and insightful visualizations for directed, valued graphs.

1 Introduction

Node clustering is a fundamental task in unsupervised network analysis. The goal is to partition nodes into groups (clusters) based on shared connectivity patterns, providing a macroscopic summary of the network structure. This is valuable for tasks like community detection in social networks [Newman and Girvan, 2004] or when looking for functional groupings of proteins [Krogan et al., 2006]. Clustering methods for networks range from algorithmic approaches [Ahn et al., 2010; Derényi et al., 2005], probabilistic models [Goldenberg et al., 2010] to spectral clustering [White and Smyth, 2005]. Among probabilistic approaches for node clustering, the stochastic block model (SBM, Holland et al. [1983]) remains a central tool. In SBM, each of the N nodes is assigned to one of Q blocks, and the probability of an edge between two nodes depends solely on their block memberships. Traditional clustering methods like SBM assume that each node belongs exclusively

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

to a single group. However, real-world systems often exhibit partial or mixed memberships, where nodes can simultaneously belong to multiple groups. Considering, for example, a publication network where a researcher who specializes in both biology and computer science, it would be reasonable to represent this individual as partly belonging to two different fields or clusters, as his expertise overlaps both disciplines. Models like the mixed membership stochastic block model (MMSBM, Airoldi et al. [2006]) and the Bayesian partial membership model (BPM, Heller et al. [2008]) allow nodes to hold fractional cluster memberships.

Node representation learning focuses on embedding nodes into low-dimensional spaces, capturing local structures. These embeddings are useful for tasks like node classification, link prediction, and graph visualization [Hamilton et al., 2017]. Popular methods include DeepWalk [Perozzi et al., 2014], node2vec [Grover and Leskovec, 2016], and LINE [Tang et al., 2015]. Variational autoencoders (VAEs) [Kingma et al., 2019, Rezende and Mohamed, 2015] offer a flexible generative framework combining deep learning and probabilistic modeling. Their adaptation to graphs, is referred to as the variational graph autoencoders (VGAE, Kipf and Welling [2016b]).

Although clustering (global structure) and representation learning (local structure) are often studied separately, they are inherently interconnected. Thus, the latent position cluster model (LPCM, Handcock et al. [2007]) extends the latent position model (LPM [Hoff et al. [2002]]) to the clustering framework allowing simultaneous clustering and visualization. LPCM supposes that the latent positions arise from a mixture distribution, with components corresponding to node clusters, akin to standard model-based clustering [Bouveyron et al., 2019]. This approach excels at detecting assortative structures since the probability of linkage increases with proximity or similarity between node positions. However, it lacks the flexibility to capture more diverse connectivity patterns. Contributions such as vGraph [Sun et al., 2019] perform simultaneous partial clustering and node embedding, but like LPCM, they primarily detect assortative structures and fail at detecting more complex structures. Addressing all these limitations, Boutin et al. [2024] introduced the deep latent position block model (Deep-LPBM), unifying ideas from SBM, LPCM, and mixed membership models. Deep-LPBM jointly performs partial node clustering and visualization in undirected binary networks, capturing a broad spectrum of structures including communities, disassortative structures as well as hubs. We emphasize that all the methods mentioned so far deal only with undirected binary networks.

However, in many real-world applications, connections are valued rather than binary, and discarding edge weights by binarization can result in significant information loss [Tylianakis et al., 2007]. The degree-corrected SBM (DCSBM) introduced by Karrer and Newman [2011] replaced the Bernoulli distribution in SBM with a Poisson to account for valued edges and to deal with degree heterogeneity, mitigating the tendency of SBMs to cluster nodes by degree alone. Parameters are estimated via maximum likelihood, with closed-form solutions enabling efficient computation. Building on this, Herlau et al. [2014] proposed the infinite DCSBM (IDCSBM), a non-parametric Bayesian variant that infers the necessity of degree correction and number of groups using the Chinese restaurant process [Kemp et al., 2006, Xu et al., 2006] and MCMC methods. It can revert to a standard SBM when degree correction is unnecessary. Aicher et al. [2015] extended this idea further with the weighted-SBM (WSBM), which models edge weights using the family of exponential distributions. A Bayesian variational approach was used to approximate the posterior distribution over latent block structures. For a broader overview of block model adaptations to complex networks, see the survey by Liu et al. [2025]. An important modeling consideration for valued networks is sparsity, i.e., the abundance of zero entries. While binary networks naturally accommodate sparsity through Bernoulli modeling, valued networks require special treatment. In the WSBM, an accommodation for sparsity is handled by modeling existence of an edge with a Bernoulli distribution and then, conditionally on the existence of an edge, its weight is modeled with an exponential distribution. However, the sparsity parameter is fixed and not dependent on the data. In Lu et al. [2025], a zero-inflated latent position cluster model (Zip-LPCM) is introduced as an extension of the LPCM. In this model a zero-inflated Poisson (ZIP) distribution [Lambert, 1992] is used for the edges and inference is performed within a fully Bayesian framework using a partially collapsed Gibbs sampler, coupled with a mixture of finite prior to automatically infer the number of clusters, and a novel truncated absorb-eject move to improve mixing across cluster partitions.

In all these approaches, the inference of the model parameters is performed either via approximation inference (e.g. variational) or via computational demanding sampling schemes. In this paper, we introduce the first importance weighted graph variational autoencoder for weighted and directed networks: the **deep zero-inflated latent position block model (Deep-ZLPBM)**. Our framework

extends Deep-LPBM [Boutin et al., 2024], adapting its strengths unifying probabilistic clustering, latent space visualization, and partial memberships to the directed, valued setting. Moreover, in traditional variational autoencoders (VAEs) framework, the evidence lower bound (ELBO) is maximized to approximate the intractable log-likelihood of the data. However, the ELBO can be a loose bound, especially when the approximate posterior is far from the true posterior distribution, leading to non-optimal latent representations. Importance-weighted autoencoders (IWAE, Burda et al. [2015]) address this by introducing multiple samples from the variational distribution, effectively tightening the lower bound as the number of samples increases. By optimizing the importance-weighted ELBO (iw-ELBO), we obtain more accurate approximations of the posterior distribution, resulting in better latent embeddings and improved clustering and visualization performance. In summary, our contributions are the following:

1. We propose Deep-ZLPBM, a VAE-based generative model tailored for directed graphs with positive integer edge weights.
2. We employ a novel encoder architecture incorporating a directed graph neural network (DirGNN) to capture directional dependencies effectively.
3. We utilize a block-wise ZIP distribution in the decoder to model valued (count-based) and sparse edge data appropriately.
4. We investigate the use of the iw-ELBO, a tighter bound than the standard ELBO, for potentially improved inference and richer latent representation learning.

Through experiments, we demonstrate the effectiveness of Deep-ZLPBM in analyzing complex directed sparse, valued networks and show that optimizing the iw-ELBO can yield significant performance enhancements compared to the standard ELBO objective.

2 Generative model for the graph

We consider a direct graph characterized by its adjacency matrix A , a square $N \times N$ matrix such that each elements $A_{ij} \in \mathbb{N}$ denotes the number of interactions between nodes i and j and $A_{ii} = 0$ (no self-loops) for all i . 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} which is a function of a latent variable $Z_i \stackrel{\text{iid}}{\sim} \mathcal{N}_d(0, \gamma^2 \mathbb{I}_d)$ with $d = Q - 1$. The set of node embeddings is denoted by $Z := \{Z_i\}_i$ in the rest of the paper. To link the latent representations of the nodes Z with partial membership probabilities η , we rely on a bijective softmax transformation $h : [0, 1]^{Q-1} \rightarrow \Delta_Q := \{x \in \mathbb{R}^Q; \sum_{q=1}^Q x_q = 1\}$ such that:

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

and we denote $\eta = (\eta_1, \dots, \eta_N)^\top$. The mapping h encodes Z into cluster partial membership probabilities. Eventually, we suppose that the probability of connection between two nodes follows a zero-inflated Poisson distribution (ZIP, Lambert [1992]) with parameters depending on η such that:

$$A_{ij} \mid Z_i, Z_j \sim \text{ZIP}(\eta_i^\top \Pi \eta_j, \eta_i^\top \Lambda \eta_j).$$

Thus given Z_i and Z_j , the probability of having an edge between nodes i and j is $\eta_i^\top \Pi \eta_j$ and if an edge is created, its weight is sampled from a Poisson distribution with intensity $\eta_i^\top \Lambda \eta_j$. By construction, if η_i and η_j are extremal points in Δ_Q , say $\eta_{iq} = 1$ and $\eta_{jr} = 1$, then the corresponding probability and intensity are given by Π_{qr} and Λ_{qr} , as in block modelling perspective. Thus, assuming conditional independence, we have:

$$\begin{aligned} p(A, Z \mid \Pi, \Lambda) &= \left(\prod_{i=1}^N p(Z_i) \right) \prod_{i \neq j} p(A_{ij} \mid Z, \Pi, \Lambda), \quad \text{with} \\ p(A \mid Z, \Lambda, \Pi) &= \prod_{\substack{i \neq j \\ A_{i,j}=0}} \left[1 - \eta_i^\top \Pi \eta_j (1 - \exp(-\eta_i^\top \Lambda \eta_j)) \right] \prod_{\substack{i \neq j \\ A_{i,j}>0}} \left[\eta_i^\top \Pi \eta_j \frac{(\eta_i^\top \Lambda \eta_j)^{A_{ij}}}{A_{ij}!} \exp(-\eta_i^\top \Lambda \eta_j) \right], \end{aligned} \quad (2)$$

where the $Q \times Q$ matrix $\Pi = (\Pi_{qr})_{1 \leq q, r \leq Q}$ is 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 0). The parameter Λ is also a $Q \times Q$ matrix whose entry (q, r) refers to the expectation of A_{ij} conditional on $A_{ij} > 0$ ($A_{ij}|A_{ij} > 0$) with i and j respectively belonging to clusters q and r . We also used the simplified notation $\prod_{i \neq j} = \prod_{i=1}^N \prod_{j=1, j \neq i}^N$.

3 Inference and optimisation

In order to estimate the model parameters Π and Λ , we aim at relying on the marginal log-likelihood of the graph, after integrating away the latent variable Z :

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

Unfortunately, this quantity cannot be computed analytically. In addition, an expectation-maximisation (EM) algorithm cannot be employed directly since the posterior distribution $p(Z|A, \Pi, \Lambda)$ is not tractable. Consequently, we choose to rely on a variational inference strategy for approximation purposes.

3.1 Variational decomposition of the marginal log-likelihood

For any distribution $q(\cdot)$ for the latent variable Z , in force of the Jensen's inequality, the following decomposition holds:

$$\log p(A|\Pi, \Lambda) = \log \mathbb{E}_{Z \sim q} \frac{p(A, Z|\Pi, \Lambda)}{q(Z)} \geq \mathbb{E}_{q(Z)} \left[\log \frac{p(A, Z|\Pi, \Lambda)}{q(Z)} \right] =: \mathcal{L}(\Pi, \Lambda, q). \quad (3)$$

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

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

where the Kullback-Leibler (KL) divergence between q and the posterior distribution $p(Z|A, \Pi, \Lambda)$ 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. Since this minimisation is satisfied when the variational $q(\cdot)$ is equal to the true posterior distribution $p(Z|A, \Pi, \Lambda)$ which is not tractable here, we restrict the family of variational distributions by assuming a mean-field approximation and other hypotheses to make the ELBO tractable:

$$q(Z) := q(Z|A) := \prod_{i=1}^N q_\phi(Z_i|A) = \prod_{i=1}^N \mathcal{N}(Z_i; \mu_\phi(\bar{A})_i, \sigma_\phi^2(\bar{A})_i \mathbb{I}_{Q-1}), \quad (4)$$

with $[\mu_\phi(\bar{A}), \log \sigma_\phi(\bar{A})] = f_\phi(\bar{A})$, where $f_\phi : \mathbb{R}^{N \times 2N} \rightarrow \mathbb{R}^{N \times Q}$ is a directed graph convolutional network (DirGCN) that maps the concatenated normalized adjacency matrix $\bar{A} = [\overrightarrow{A}, \overleftarrow{A}]$ to the vector of the variational means and the log standard deviations:

$$\overrightarrow{A} = \overrightarrow{D}^{-\frac{1}{2}}(A_{\text{up}} + A_{\text{up}}^\top + I_N) \overrightarrow{D}^{-\frac{1}{2}}, \quad \overleftarrow{A} = \overleftarrow{D}^{-\frac{1}{2}}(A_{\text{low}} + A_{\text{low}}^\top + I_N) \overleftarrow{D}^{-\frac{1}{2}}. \quad (5)$$

Here A_{up} and A_{low} respectively represent the upper and lower triangular matrices obtained from A , \overrightarrow{D} and \overleftarrow{D} are the diagonal matrices representing the degree of nodes with respect to the matrices $A_{\text{up}} + I_N$ and $A_{\text{low}} + I_N$, defined as $\overrightarrow{D}_{ii} = 1 + \sum_{j=1}^N A_{\text{up}_{ij}}$, and $\overleftarrow{D}_{ii} = 1 + \sum_{j=1}^N A_{\text{low}_{ij}}$. DirGCN f_ϕ extends graph convolutional networks (GCNs, Kipf and Welling [2016a]) to directed graphs with two separate aggregations over the in-and-out neighbours of a node at each layer: $X^k = \sigma(\overrightarrow{A} X^{k-1} \overrightarrow{W}^k + \overleftarrow{A} X^{k-1} \overleftarrow{W}^k)$, where \overrightarrow{W}^k and \overleftarrow{W}^k represent in-and-out learnable weights.

Finally, the ELBO can be decomposed as follows:

$$\begin{aligned} \mathcal{L}(q, \Lambda, \Pi) &= \mathbb{E}_{Z \sim q} \left[\log \frac{p(A, Z|\Pi, \Lambda)}{q(Z|A)} \right] = \mathbb{E}_{Z \sim q} \left[\log \frac{p(A|Z, \Pi, \Lambda)p(Z)}{q(Z|A)} \right] \\ &= \underbrace{\mathbb{E}_{Z \sim q} [\log p(A|Z, \Pi, \Lambda)]}_{\text{Reconstruction term}} - \underbrace{\text{KL}(q(Z|A)||p(Z))}_{\text{Regularisation term}}. \end{aligned} \quad (6)$$

The regularization term, which involves the KL divergence between q and the posterior, can be computed exactly:

$$\text{KL}(q(Z|A)||p(Z)) = \sum_i \left[-(Q-1) \log \frac{\sigma_\phi(\bar{A})_i}{\gamma^2} - \frac{Q-1}{2} + \frac{1}{2\gamma^2} \|\mu_\phi(\bar{A})_i\|_2^2 + \frac{Q-1}{2\gamma^2} \sigma_\phi(\bar{A})_i^2 \right].$$

As in the β -VAE framework [Burgess et al., 2018], the hyperparameter γ controls the trade-off between reconstruction fidelity and the strength of the KL divergence regularization. When $\gamma < 1$ (i.e., $\frac{1}{\gamma} > 1$), greater weight is placed on the KL term, encouraging the variational distribution $q(Z | A)$ to align closely with the prior $\mathcal{N}(0, I)$. This promotes disentangled representations, where individual latent dimensions capture distinct factors of variation. Conversely, in our setting, where the latent variables are bijective transformations of cluster membership probabilities, we focus on $\gamma > 1$ (i.e., $\frac{1}{\gamma} < 1$), which reduces the regularization pressure and emphasizes reconstruction. This typically yields more faithful reconstructions, better preserving the input data. Throughout the paper, we fix $\gamma = N^{\frac{1}{2}}$.

3.2 Importance weighted ELBO

While computationally convenient, the standard ELBO (Equation 6) can sometimes provide a loose lower bound on the marginal log-likelihood $\log p(A|\Pi, \Lambda)$, especially if the chosen variational family q is not flexible enough to closely match the true posterior. This looseness can potentially hinder model training and evaluation.

To address this, a tighter lower bound can be constructed using multiple samples from the same proposal distribution $q(Z|A)$. This leads to the importance weighting ELBO (iw-ELBO), originally proposed by [Burda et al., 2015]. The core idea relies on importance sampling principles. Recall that the marginal likelihood can be expressed as an expectation over the variational distribution:

$$p(A|\Pi, \Lambda) = \mathbb{E}_{Z \sim q} \left[\frac{p(A, Z|\Pi, \Lambda)}{q(Z|A)} \right]$$

Instead of taking the logarithm and directly applying Jensen's inequality to derive the ELBO, we can first estimate the expectation using K independent samples $Z^1, \dots, Z^K \sim q$:

$$\begin{aligned} \log p(A | \Pi, \Lambda) &= \log E_{Z^1, \dots, Z^K \sim q} \left[\frac{1}{K} \sum_{l=1}^K \frac{p(A, Z^l | \Pi, \Lambda)}{q(Z^l | A)} \right] \\ &\geq \mathbb{E}_{Z^1, \dots, Z^K \sim q(\cdot | A)} \left[\log \left(\frac{1}{K} \sum_{l=1}^K \frac{p(A, Z^l | \Pi, \Lambda)}{q(Z^l | A)} \right) \right]. \end{aligned} \quad (7)$$

The resulting lower bound, denoted \mathcal{L}_K , is known as the importance weighting ELBO (iw-ELBO) denoted \mathcal{L}_K .

Theorem 1 *For all $K \geq 1$, the iw-ELBO satisfies*

$$\mathcal{L}(q, \Lambda, \Pi) = \mathcal{L}_1 \leq \mathcal{L}_K \leq \mathcal{L}_{K+1} \leq \log p(A|\Pi, \Lambda).$$

Moreover, if $p(A, Z)/q(Z|A)$ is bounded, $(\mathcal{L}_K)_K$ converges uniformly to $\log p(A|\Pi, \Lambda)$.

This theorem have been proved in [Burda et al., 2015] in the context of continuous data. The proof in our case is detailed in Appendix A

Reinterpreting iw-ELBO This iw-ELBO can be considered as an ELBO but with a more versatile and flexible variational distribution. Given a batch of samples drawn from the variational distribution introduced in Equation 4, $Z^2, \dots, Z^K \sim q(\cdot | A)$, let us consider this following unnormalized distribution:

$$q_{\text{iw}}(Z|A, Z^{2:K}) = \frac{p(A, Z)}{\frac{1}{K} \left(\frac{p(A, Z)}{q(Z|A)} + \sum_{\ell=2}^K \frac{p(A, Z^\ell)}{q(Z^\ell|A)} \right)}.$$

Theorem 2 $q_{\text{iw}}(Z|A) = \mathbb{E}_{Z^{2:K} \sim q(\cdot | A)} [q_{\text{iw}}(Z|A, Z^{2:K})]$ is a probability distribution and $\mathbb{E}_{Z \sim q_{\text{iw}}} \log \left[\frac{p(A, Z)}{q_{\text{iw}}(Z|A)} \right] \geq \mathcal{L}_K$.

This theorem have also been proved in [Cremer et al., 2017] in the context of continuous data. The proof in our case is detailed in Appendix A.

Hence, iw-ELBO can be reinterpreted to the ELBO with a different variational distribution. So when \mathcal{L}_K converges to the log-likelihood, q_{EW} approximates exactly the true posterior distribution. A procedure to sample from q_{EW} is detailed in Appendix C.

In all the experiments, we fixed $K = 10$.

3.3 Parameters optimisation and model selection

The model parameters Π and Λ , and variational parameters ϕ cannot be updated with analytical formulas because of the integral involving the variational distribution $q(\cdot)$ in both the ELBO and the iw-ELBO. We thus perform stochastic gradient descent based on the following estimates $\tilde{\mathcal{L}}$ of the ELBO and $\tilde{\mathcal{L}}_K$ of the iw-ELBO to perform stochastic gradient descent:

$$\begin{aligned} \tilde{\mathcal{L}}(q, \Pi, \Lambda) &= \frac{1}{S} \sum_{s=1}^S \log p(A|Z^{(s)}, \Pi, \Lambda) \\ &\quad - \sum_i \left[-(Q-1) \log \sigma_\phi(\bar{A})_i - \frac{Q-1}{2} + \frac{1}{2} \|\mu_\phi(\bar{A})_i\|_2^2 + \frac{Q-1}{2} \sigma_\phi(\bar{A})_i^2 \right], \end{aligned} \quad (8)$$

$$\tilde{\mathcal{L}}_K(q, \Pi, \Lambda) = \frac{1}{S} \sum_{s=1}^S \log \frac{1}{K} \sum_{\ell=1}^K \frac{p(A, Z^{(s,\ell)})}{q(Z^{(s,\ell)}|A)}, \quad \text{where } S = 1 \quad \text{in our experiments and} \quad (9)$$

where $Z^{(s,1)}, \dots, Z^{(s,K)} \sim q(\cdot)$, independently. Usual (naïve) Monte Carlo gradient estimator exhibits very high variance and is then impractical [Paisley et al., 2012]. To deal with this problem, as done in Kingma [2013], we adopt the reparameterization trick. In particular, if $\epsilon^{(s)} \sim \mathcal{N}(0, \mathbb{I}_{Q-1})$, then $Z_i^{(s)} = \mu_\phi(\bar{A})_i + \sigma_\phi(\bar{A})_i \epsilon^{(s)} \stackrel{iid}{\sim} \mathcal{N}(\mu_\phi(\bar{A})_i, \sigma_\phi^2(\bar{A})_i \mathbb{I}_{Q-1}) = q(\cdot|A)$.

Additionally, to simplify the gradient descent, we map the constrained parameters $(\Pi_{qr})_{q,r}$ (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:

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

Letting $\Pi = g(\tilde{\Pi})$ and $\Lambda = f(\tilde{\Lambda})$, we can optimize the lower bound w.r.t. $\tilde{\Pi} = (\tilde{\Pi}_{qr})_{q,r}$, $\Lambda = (\tilde{\Lambda}_{qr})_{q,r}$, unconstrained $Q \times Q$ matrices, using gradient descent. For clarity, we denote $\Pi = g(\tilde{\Pi})$ and $\Lambda = g(\tilde{\Lambda})$ as the element-wise mapping of $\tilde{\Pi}$ by g and $\tilde{\Lambda}$ by f .

An effective initialization is essential for clustering algorithms such as expectation-maximization (EM) or variational EM (V-EM). Due to the non-convex nature of the related optimization problems, there is a risk of getting trapped in local stationary points, which can lead to sub-optimal solutions. In Appendix C the optimisation procedure and the initialisation are better detailed.

3.4 Model selection

To estimate the number of clusters Q , we compare the performance of Akaike's information criterion [Akaike, 1974], the Bayesian information criterion [Schwarz, 1978] and the integrated classification likelihood [Biernacki et al., 1998]. The map estimate \hat{Z} is considered as given and treated as a parameter, as well as η . Hence, denoting \mathcal{M} the generative model, and Q the fixed number of clusters considered, the three criteria can be computed as:

$$\begin{aligned} AIC(Q, \mathcal{M}) &:= \log p(A|\hat{Z}, \hat{\Pi}, \hat{\Lambda}) - 2(Q-1)^2 - N(Q-1), \\ BIC(Q, \mathcal{M}) &:= \log p(A|\hat{Z}, \hat{\Pi}, \hat{\Lambda}) - \frac{(2(Q-1)^2 + N(Q-1))}{2} \log(N(N-1)), \\ ICL(Q, \mathcal{M}) &:= \log p(A, \hat{Z}|\hat{\Pi}, \hat{\Lambda}) - \frac{(2(Q-1)^2 + N(Q-1))}{2} \log(N(N-1)), \end{aligned}$$

where $\hat{\Pi}$ and $\hat{\Lambda}$ respectively denote the estimated Π and Λ after maximizing the ELBO and $\hat{Z} = \mu_\phi(\bar{A})$ with ϕ the variational estimates after optimizing the lower bound. The comparison of these criteria is assessed on synthetic data in Appendix B.

4 Evaluation on synthetic datasets

As we consider an unsupervised problem, to assess the quality of the network representations and node partitions produced by our approach, we evaluate the performances on synthetic datasets with known ground truth. We first describe the network structures, simulation scenarios, and evaluation procedure, then benchmark our method against state-of-the-art baselines.

4.1 Experimental setup

We test our method on three synthetic networks, each with 200 nodes with $Q = 3$ clusters, designed to represent distinct social structures.

Network structures The three network structures we considered are: i) the **community structure**, where we assume that connexions between nodes in the same group have a high probability of occurrence and a high expected intensity denoted respectively by β and α and that connexion between nodes in different groups have a low probability of occurrence and low expected value indicated by δ and ω , ii) the **disassortative structure** where we assume that connexions between nodes in different clusters have a high probability of occurrence and high expected values denoted by β and α , and that connexion between nodes in the same cluster have low probability of occurrence and low expected values denoted by δ and ω , iii) a **hub structure** where we assume a cluster of central nodes (stars) that have relatively high in-connection and relatively missing out-connexions to the rest of the network and the other clusters are communities.

Since both Π (binary connectivity) and Λ (interaction intensity) influence edge generation, the contribution of Λ is often suppressed when Π is low. As a result, clustering performance heavily depends on the density of Π . Because community structures typically have lower connection densities than hub or disassortative structures, methods may appear more effective on the latter simply due to more observed edges. To ensure fair comparisons across configurations, we normalize Π to match the average density of the community setting: Π : $\bar{\Pi} = \frac{d}{\text{norm}} \Pi$, where norm is the actual density of Π and d is the density connection in the community structure.

For the generation of the matrix 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 $X \sim \text{Bernoulli}(\cdot|1; \kappa^\top \bar{\Pi} \kappa)$ and $Y \sim \text{Poisson}(\cdot|\kappa^\top \Lambda \kappa)$ and set $A = X \circ Y$ (Hadamard product).

Simulated scenarios We test three cases: **scenario 1** where both Π and Λ share the same structure (community, disassortative, or hub), **scenario 2** where only Λ has structure while Π is uniform (Erdős–Rényi), and **scenario 3** where Π has structure and all entries of Λ have the same value. In Appendix D we plot examples of adjacency matrices from these three scenarios.

Evaluation The adjusted rand index (ARI, Yeung and Ruzzo [2001]) 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 a perfect alignment with true node labels, up to label switching.

4.2 Benchmarking

In this section, we evaluate Deep-ZLPBM for node clustering across the three synthetic network structures described above. Recall that Deep-ZLPBM is designed to infer partial node memberships rather than hard cluster assignments, as in classical SBMs. For evaluation purposes, we assign each node to the cluster with the highest estimated membership. We compare the clustering performance of Deep-ZLPBM against several baselines: the Deep-LPBM, the Poisson stochastic block model (PSBM; Leger [2016]) and the embedding methods LINE, DeepWalk, and Node2vec, all followed by k-means clustering. Since LINE, DeepWalk, and Deep-LPBM are designed for unweighted and undirected graphs, we binarize and symmetrize the adjacency matrix by setting all positive entries to 1. Results for scenarios 1 and 2 are described in Table II.

Scenario 1 In the community structure, Deep-ZLPBM and Deep-LPBM consistently recover true node partitions across all difficulty levels, demonstrating strong clustering ability. Node2vec and DeepWalk also perform well, while PSBM and LINE struggle throughout. In disassortative structures, only Deep-ZLPBM and Deep-LPBM yield good results, as other methods fail to capture anti-community patterns, resulting in near-zero ARI scores. In the hub structure with $\beta = 0.1$, Deep-ZLPBM outperforms all models, while Deep-LPBM, Node2vec, DeepWalk, and PSBM perform

Scenario 1				Scenario 2			
$\beta = 0.1, \delta = 0.01, \alpha = 4, \omega = 2$				$\beta = 0.1, \delta = 0.1, \alpha = 4, \omega = 2$			
Method	Communities	Disassortative	Hub	Method	Communities	Disassortative	Hub
Deep-LPBM	1 ± 0.00	0.97 ± 0.02	0.41 ± 0.02	Deep-LPBM	0 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Node2vec	0.94 ± 0.15	0.009 ± 0.007	0.40 ± 0.05	Node2vec	0.06 ± 0.08	0.002 ± 0.003	0.08 ± 0.05
Deepwalk	1 ± 0.00	0.0 ± 0.00	0.44 ± 0.06	Deepwalk	0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
LINE	0.49 ± 0.08	0.0 ± 0.00	0 ± 0.00	LINE	0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
PSBM	0.48 ± 0.17	0.41 ± 0.19	0.43 ± 0.19	PSBM	0.21 ± 0.26	0.12 ± 0.29	0.13 ± 0.28
Deep-ZLPBM ELBO	1 ± 0.00	0.97 ± 0.03	0.98 ± 0.01	Deep-ZLPBM ELBO	0.24 ± 0.15	0.27 ± 0.13	0.15 ± 0.09
Deep-ZLPBM iw-ELBO	1 ± 0.00	0.99 ± 0.02	0.99 ± 0.03	Deep-ZLPBM iw-ELBO	0.89 ± 0.13	0.91 ± 0.03	0.62 ± 0.16
$\beta = 0.15, \delta = 0.01, \alpha = 4, \omega = 2$				$\beta = 0.15, \delta = 0.15, \alpha = 4, \omega = 2$			
Method	Communities	Disassortative	Hub	Method	Communities	Disassortative	Hub
Deep-LPBM	1 ± 0.00	0.98 ± 0.03	0.97 ± 0.01	Deep-LPBM	0 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Node2vec	1 ± 0.00	0.009 ± 0.007	0.44 ± 0.06	Node2vec	0.26 ± 0.17	0.009 ± 0.007	0.01 ± 0.009
Deepwalk	1 ± 0.00	0.0 ± 0.00	0.46 ± 0.07	Deepwalk	0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
LINE	0.55 ± 0.20	0.0 ± 0.00	0.25 ± 0.13	LINE	0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
PSBM	0.51 ± 0.16	0.43 ± 0.18	0.433 ± 0.19	PSBM	0.38 ± 0.2	0.21 ± 0.24	0.21 ± 0.26
Deep-ZLPBM ELBO	1 ± 0.00	0.99 ± 0.04	0.98 ± 0.02	Deep-ZLPBM ELBO	0.81 ± 0.11	0.41 ± 0.11	0.39 ± 0.16
Deep-ZLPBM iw-ELBO	1 ± 0.00	1 ± 0.00	1 ± 0.00	Deep-ZLPBM iw-ELBO	0.98 ± 0.01	1 ± 0.00	0.75 ± 0.06
$\beta = 0.3, \delta = 0.01, \alpha = 4, \omega = 0.01$				$\beta = 0.3, \delta = 0.3, \alpha = 4, \omega = 2$			
Method	Communities	Disassortative	Hub	Method	Communities	Disassortative	Hub
Deep-LPBM	1 ± 0.00	1 ± 0.00	1 ± 0.00	Deep-LPBM	0 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Node2vec	1 ± 0.00	0.0 ± 0.0	0.68 ± 0.22	Node2vec	0.89 ± 0.15	0.04 ± 0.03	0.33 ± 0.11
Deepwalk	1 ± 0.00	0.0 ± 0.00	0.86 ± 0.23	Deepwalk	0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
LINE	0.90 ± 0.03	0.21 ± 0.06	0.29 ± 0.04	LINE	0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
PSBM	0.91 ± 0.08	0.44 ± 0.05	0.52 ± 0.06	PSBM	0.51 ± 0.16	0.46 ± 0.18	0.41 ± 0.19
Deep-ZLPBM ELBO	1 ± 0.00	1 ± 0.00	1 ± 0.00	Deep-ZLPBM ELBO	1 ± 0.00	1 ± 0.00	0.73 ± 0.14
Deep-ZLPBM iw-ELBO	1 ± 0.00	1 ± 0.00	1 ± 0.00	Deep-ZLPBM iw-ELBO	1 ± 0.00	1 ± 0.00	0.91 ± 0.07

Table 1: Mean ARIs obtained in scenarios 1 and 2 by Deep-ZLPBM and its competitors. Standard deviations are reported. Results are averaged over 9 different networks in each configuration. All experiments were conducted on a MacBook Air with an M2 chip and 24598GB of RAM. For ELBO inference, the median time per epoch is 0.002s and for iw-ELBO inference, the median time per epoch was 0.0025s.

similarly ($ARI \simeq 0.4$); LINE fails ($ARI \simeq 0$). At $\beta = 0.15$, Deep-ZLPBM with iw-ELBO achieves perfect clustering ($ARI = 1$), outperforming the standard ELBO variant ($ARI \simeq 0.98$). At $\beta = 0.3$, both Deep-ZLPBM and Deep-LPBM achieve perfect clustering across all inference strategies.

Scenario 2 In this setting, the link probability matrix Π is homogeneous (Erdős-Rényi), and consequently structure arises only from interaction intensities in Λ . Binary-only models (LINE, Node2vec, DeepWalk) fail entirely ($ARI \simeq 0$), as they ignore valued information, highlighting a key limitation of binary encoding. In contrast, models handling real-valued data, like Deep-ZLPBM and PBM, perform much better. Notably, Deep-ZLPBM with iw-ELBO inference consistently outperforms its standard ELBO counterpart, yielding adjusted rand indices that are more than twice better, proving the significant advantage of importance-weighted inference in capturing complex valued network structures.

Results for Scenario 3 are reported in Appendix E. In this scenario, since only the connection probabilities Π play a role in the structure, Deep-LPBM which use an adjacency matrix with higher connectivity is better than Deep-ZLPBM for $\beta = 0.1$.

5 Real dataset analysis: The London bike sharing network

In this section, we analyze the public London Bike Sharing dataset² which records bike trips between $N = 779$ stations across London. The resulting graph comprises 224070 directed edges, where each entry (i, j) of the adjacency matrix indicates the number of trips from station i to station j . In addition to trip data, the dataset includes several station-level attributes, such as geographic coordinates (longitude and latitude) that we use only for interpretability of the results.

Deep-ZLPBM analysis We begin this analysis by estimating the number of clusters. Deep-ZLPBM was applied 10 times, for different values of Q ranging from 2 to 40 and the optimal value found was $Q = 17$ clusters. In Figure 1, the stations are plotted on a London map where different colors correspond to different clusters identified by the model. Nearby stations are placed in the same cluster and the geographical distance between them plays a key role. In addition the estimated $\hat{\Pi}$ and $\hat{\Lambda}$ matrix are displayed in Figure 2. We observe that with the exception of clusters 7, 13, 14, and 16, all clusters exhibit high internal connection probabilities and high probabilities to connect

²data

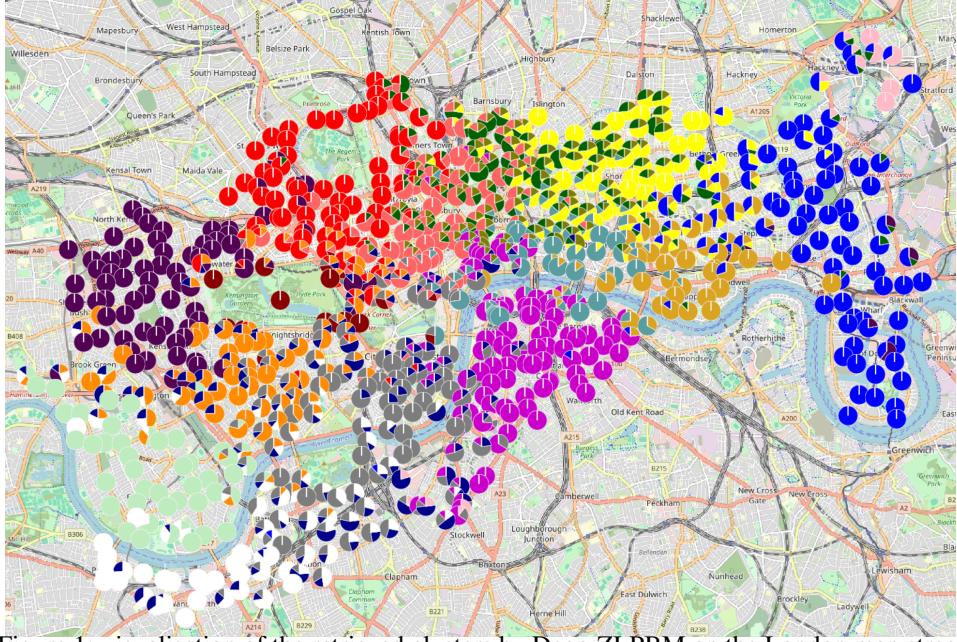


Figure 1: visualisation of the retrieved clusters by Deep-ZLPBM on the London map street.

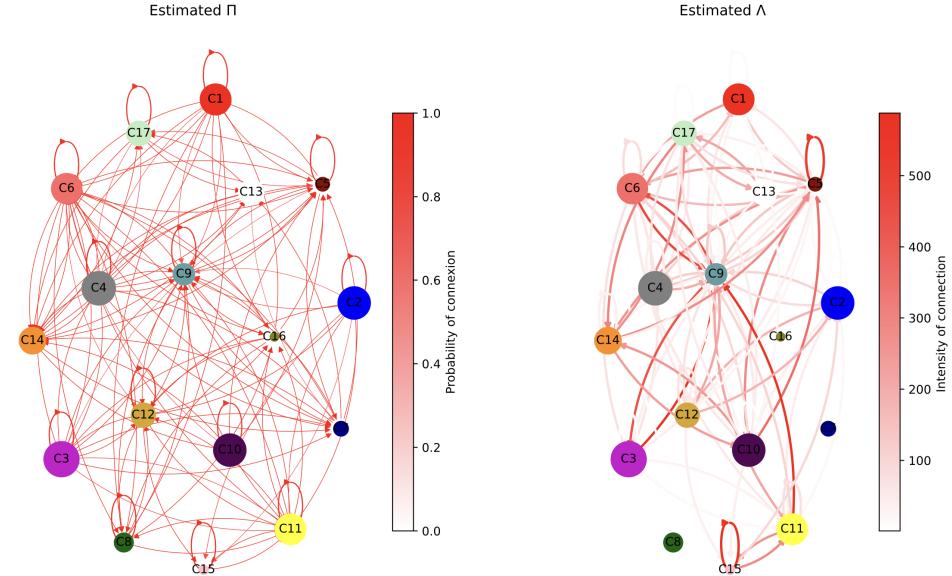


Figure 2: visualisation of the estimated parameters $\hat{\Pi}$ and $\hat{\Lambda}$ by Deep-ZLPBM. The size of nodes is proportional to the size of the clusters in term of hard clustering. For the sake of visibility, we removed edge probabilities lower than 0.3 and edge intensities lower than 1.

with stations in nearby clusters. This is expected, as short bike trips are generally more frequent than long-distance ones. Moreover clusters in the extremities of London also have high probability to connect with clusters in the center of London. this suggests trips from residential area to work-place. Interestingly high probability of connexion does not always align with high intensity of connexion. Indeed, stations in clusters 7, 8, 16 have high probabilities to connect to the neighbor clusters but do not really connect with them (very low intensities). This suggest that stations in these clusters are ghost-stations. In term of hard-clustering, these clusters have very small size with no station belonging to them with more than 90% certainty. Stations from other clusters and partially belonging to these ghost clusters also are lowly connected to the rest of the stations. Stations in the center of London have also many partial membership to the different clusters in the center and high intensity connexion between them.

6 Conclusion

This paper introduced Deep-ZLPBM, the first importance-weighted graph variational autoencoder specifically designed for valued and directed networks. Our model jointly performs partial node clustering and representation learning, addressing key challenges in analyzing complex network. The proposed architecture features a novel directed graph convolutional network (DirGCN) encoder and a ZIP block modelling decoder. Through extensive experiments on synthetic datasets We demonstrated that optimizing the importance-weighted ELBO (iw-ELBO) provides significant performance gains compared to the standard ELBO optimization and that Deep-ZLPBM is effective modeling complex network structures like disassortative of hubs. Furthermore, our analysis of the London bike sharing network showcased the model’s ability to provide interpretable insights into real-world valued graphs. Exciting future directions include incorporating temporal dynamics and extending the model to handle multidimensional network data.

References

- Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *nature*, 466(7307):761–764, 2010.
- Christopher Aicher, Abigail Z Jacobs, and Aaron Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 2015.
- Edoardo M Airoldi, David M Blei, Stephen E Fienberg, Eric P Xing, and Tommi Jaakkola. Mixed membership stochastic block models for relational data with application to protein-protein interactions. In *Proceedings of the international biometrics society annual meeting*, volume 15, page 1, 2006.
- Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- Christophe Biernacki, Gilles Celeux, and Gérard Govaert. *Assessing a mixture model for clustering with the integrated classification likelihood*. PhD thesis, INRIA, 1998.
- Rémi Boutin, Pierre Latouche, and Charles Bouveyron. The deep latent position block model for the block clustering and latent representation of networks. *arXiv preprint arXiv:2412.01302*, 2024.
- Charles Bouveyron, Gilles Celeux, T Brendan Murphy, and Adrian E Raftery. *Model-based clustering and classification for data science: with applications in R*, volume 50. Cambridge University Press, 2019.
- Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders. *arXiv preprint arXiv:1509.00519*, 2015.
- Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in β -vae. *arXiv preprint arXiv:1804.03599*, 2018.
- Chris Cremer, Quaid Morris, and David Duvenaud. Reinterpreting importance-weighted autoencoders. *arXiv preprint arXiv:1704.02916*, 2017.
- Jean-Jacques Daudin, Laurent Pierre, and Corinne Vacher. Model for heterogeneous random networks using continuous latent variables and an application to a tree-fungus network. *Biometrics*, 66(4): 1043–1051, 2010.
- Imre Derényi, Gergely Palla, and Tamás Vicsek. Clique percolation in random networks. *Physical review letters*, 94(16):160202, 2005.
- Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, Edoardo M Airoldi, et al. A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2):129–233, 2010.
- Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

- William L Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.
- Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 170(2):301–354, 2007.
- Katherine A Heller, Sinead Williamson, and Zoubin Ghahramani. Statistical models for partial membership. In *Proceedings of the 25th International Conference on Machine learning*, pages 392–399, 2008.
- Tue Herlau, Mikkel N Schmidt, and Morten Mørup. Infinite-degree-corrected stochastic block model. *Physical review E*, 90(3):032819, 2014.
- Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098, 2002.
- Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
- Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107, 2011.
- Charles Kemp, Joshua B Tenenbaum, Thomas L Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, volume 3, page 5, 2006.
- Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.
- T N Kipf and M Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016a.
- Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016b.
- Nevan J Krogan, Gerard Cagney, Haiyuan Yu, Gouqing Zhong, Xinghua Guo, Alexandre Ignatchenko, Joyce Li, Shuye Pu, Nira Datta, Aaron P Tikuisis, et al. Global landscape of protein complexes in the yeast *saccharomyces cerevisiae*. *Nature*, 440(7084):637–643, 2006.
- Diane Lambert. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992.
- Jean-Benoist Leger. Blockmodels: A r-package for estimating in latent block model and stochastic block model, with various probability functions, with or without covariates. *arXiv preprint arXiv:1602.07587*, 2016.
- Xueyan Liu, Wenzhuo Song, Katarzyna Musial, Yang Li, Xuehua Zhao, and Bo Yang. Stochastic block models for complex network analysis: A survey. *ACM Transactions on Knowledge Discovery from Data*, 2025.
- Chaoyi Lu, Riccardo Rastelli, and Nial Friel. A zero-inflated poisson latent position cluster model. *arXiv preprint arXiv:2502.13790*, 2025.
- Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- John Paisley, David Blei, and Michael Jordan. Variational bayesian inference with stochastic search. *arXiv preprint arXiv:1206.6430*, 2012.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710, 2014.

- Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.
- Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.
- Fan-Yun Sun, Meng Qu, Jordan Hoffmann, Chin-Wei Huang, and Jian Tang. vgraph: A generative model for joint community detection and node representation learning. *Advances in Neural Information Processing Systems*, 32, 2019.
- Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, pages 1067–1077, 2015.
- Jason M Tylianakis, Teja Tscharntke, and Owen T Lewis. Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature*, 445(7124):202–205, 2007.
- Scott White and Padhraic Smyth. A spectral clustering approach to finding communities in graphs. In *Proceedings of the 2005 SIAM international conference on data mining*, pages 274–285. SIAM, 2005.
- Zhao Xu, Volker Tresp, Kai Yu, Hans-Peter Kriegel, et al. Learning infinite hidden relational models. *Uncertainty in Artificial Intelligence (UAI2006)*, 2, 2006.
- Ka Yee Yeung and Walter 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*, 17(9):763–774, 2001.