國立臺灣大學社會科學院經濟學系
碩士論文

Department of Economics

College of Social Sciences

National Taiwan University
Master's Thesis

彙總關係資料下未知網絡連結重建：貝葉斯潛在曲面建模與懲罰迴歸法探討

Recovering Unobserved Network Links from Aggregated Relational Data: Discussions on Bayesian Latent Surface Modeling and Penalized Regression

曾彥暄

Yen-hsuan Tseng

指導教授：蘇軒立博士

Advisor: Hsuan-Li Su, Ph.D.

中華民國 114 年 8 月

August 2025

doi:10.6342/NTU202504386

國立臺灣大學碩士學位論文
口試委員會審定書

MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

彙總關係資料下未知網絡連結重建：
貝葉斯潛在曲面建模與懲罰迴歸法探討

Recovering Unobserved Network Links from Aggregated Relational Data:
Discussions on Bayesian Latent Surface Modeling and Penalized Regression

本論文係 曾彥暄（學號 R00323024）在國立臺灣大學經濟學系研究所完成之碩士學位論文，於民國 114 年 7 月 31 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Graduate Institute of Economics on 31 July, 2025 have examined a Master's Thesis entitled above presented by Yen-Hsuan, Tseng (student ID: R00323024) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

蔡耕彥	莊惠華	張勝凱
（指導教授 Advisor）		

doi:10.6342/NTU202504386

Acknowledgements

Hoc opus, professoribus honorandis, Lunae, Aliciae, Alici, Lyrae, Verae, familiae meae, atque uxori in aeternum dilectae, reverenter consecratur.

In memoriam amoris veri, et in praesentia vocum novarum.

doi:10.6342/NTU202504386

摘要

精確的網絡資料對於經濟學、金融、社會學、流行病學及電腦科學等領域皆至關重要。然而，現實限制常使研究者無法取得完整的鄰接矩陣，只能仰賴部分或彙總資訊。彙總型關係資料（Aggregated Relational Data, ARD）爲常見形式，受訪者僅需回報與具特定屬性節點的連結數，而非完整列出所有聯絡對象。

本論文比較並延伸兩種從 ARD 重建網絡的核心方法：貝氏潛曲面模型（Bayesian Latent Surface Model, BLSM）與懲罰式迴歸（Frequentist Penalized Regression）。研究內容涵蓋可識別性、一致性、對報告誤差的穩健處理、大規模資料推斷，以及潛在的隱私保護應用。貝氏方法將節點嵌入高維球面空間，以幾何距離刻畫連結傾向；懲罰式迴歸則利用高維最佳化結構整合協變數，具良好擴展性。模擬實驗探討特徵設計、測量誤差與樣本規模之交互作用，並以（類）實證資料展示方法於金融風險管理、社群推薦與疫情追蹤等領域的應用潛力。

結果顯示，儘管 ARD 資訊較粗，仍保留大量網絡結構訊息，能支撐規模化且準確的推斷。本研究亦提出自適應特徵收集、幾何—懲罰混合方法及審慎資料共享策略，以兼顧理論嚴謹與實務價值。

關鍵詞：彙總型關係資料、網絡重建、貝氏潛曲面模型、懲罰式迴歸、可識別性、報告誤差、隱私保護

doi:10.6342/NTU202504386

Abstract

Accurate network data matter across economics, finance, sociology, epidemiology, and computer science, yet complete adjacency matrices are rarely available. Researchers therefore rely on partial or aggregated information. Aggregated Relational Data (ARD) summarizes the number of ties to attribute-defined groups rather than enumerating all links.

This dissertation compares and extends two approaches for reconstructing networks from ARD: a Bayesian Latent Surface Model (BLSM) and frequentist penalized regression. We study identifiability, robustness to misreporting, scalability, and privacy. BLSM places nodes on a hypersphere so that link likelihoods depend on geometric distance; penalized regression recasts unobserved edges as a high-dimensional optimization that integrates covariates and scales efficiently. Simulations examine the roles of trait design, measurement error, and sample size, and applications with real or quasi-real data—interbank risk, social recommendation, and epidemic contact tracing—demonstrate practical utility.

Results show that, despite its coarseness, ARD preserves sufficient structure for accurate, scalable inference. The work proposes adaptive trait collection, hybrid geometry-penalty models, and privacy-aware data-sharing strategies that balance theoretical guarantees with practical deployment.

Keywords: aggregated relational data, network reconstruction, Bayesian latent surface model, penalized regression, identifiability, misreporting, privacy preservation

doi:10.6342/NTU202504386

Contents

doi:10.6342/NTU202504386

doi:10.6342/NTU202504386

doi:10.6342/NTU202504386

List of Figures

doi:10.6342/NTU202504386

List of Tables

doi:10.6342/NTU202504386

Chapter 1

Introduction and Motivation

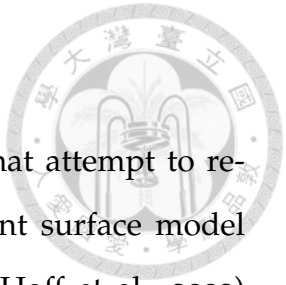
1.1 Background and Motivation

Networks, whether social, financial, epidemiological, or technological, serve as powerful mechanisms for understanding complex interactions. Traditional approaches to network research often assume that an entire adjacency matrix is observed, detailing who is connected to whom and in what manner, but constraints related to privacy, cost, or data availability frequently prevent investigators from capturing a complete matrix of relationships (Wasserman and Faust, 1994; Handcock et al., 2010; Gandy and Veraart, 2019).

Aggregated Relational Data (ARD) has emerged as a practical alternative to direct enumeration of neighbors, largely because it asks respondents or institutions to provide only aggregate counts of how many of their connections possess a certain attribute. Such a mechanism might inquire, for instance, "How many of your contacts work in the finance sector?" or "How many of your close acquaintances are older than 30?" Although this approach is typically easier to administer and can better protect privacy, it offers coarser information that must then be used carefully in any reconstruction procedure. The potential for misreporting or systematic biases in how respondents count their connections further complicates the inference problem, calling for robust or error-tolerant models (McCormick and Zheng, 2012; Breza and Chandrasekhar, 2017).

doi:10.6342/NTU202504386

1.2 Research Objectives



Throughout this work, we center on two main frameworks that attempt to recover network structure from ARD. One is the Bayesian latent surface model (BLSM), which draws inspiration from latent space models (Hoff et al., 2002) and places nodes in a continuous geometric space, often a hypersphere, where similarity or distance influences link probabilities. The other approach is the frequentist penalized regression (FPR) viewpoint, in which unknown edges or probabilities are treated as parameters in a large-scale optimization problem with a penalty function (e.g., Lasso) that promotes identifiability, especially for large networks (Alidaee et al., 2020). We endeavor to unify theoretical observations on uniqueness and consistency, develop robust methods to cope with noise and misreporting, compare computational performance and feasibility, and illustrate practical utility in several real-world or quasi-real contexts.

1.3 Contributions and Dissertation Outline

Our work extends prior discussions with a deeper account of how to pinpoint identifiability under ARD settings, including conditions that allow partial consistency when the network size grows. It also introduces advanced techniques for handling potential misreporting and measurement error, while exploring both Bayesian MCMC-based methods and frequentist optimization strategies that incorporate robust deviance. The simulation studies further expand upon trait design, network scale, and differential privacy. We also review several applications that highlight how partial observations can still yield meaningful structural insights.

Below is a concise overview of the structure:

Chapter 2 reviews the main approaches to partial network sampling, situating ARD among other methods like egocentric or snowball designs, and notes the historical emergence of Bayesian latent space modeling and modern devel-

doi:10.6342/NTU202504386

opments in penalized network inference.

Chapter 3 offers a thorough examination of the Bayesian Latent Surface Model, including spherical embeddings, Poisson-based ARD likelihood, prior assignments, MCMC details, and a discussion of identifiability constraints—plus new extended results on correlated misreporting, manifold-based variational inference, and further identifiability insights.

Chapter 4 covers the Frequentist Penalized Regression approach, outlining the fundamental objective functions, optimization algorithms, and potential for robust or federated extensions that can accommodate privacy requirements, as well as novel penalties like SCAD or MCP for large-scale networks.

Chapter 5 describes our simulation studies, covering design variations in trait assignments, misreporting frequencies, and weighted vs. binary edges, while also integrating a more detailed set of numerical results that were originally circulated in a separate companion document. We include expansions to handle large n, correlated misreporting blocks, and differential privacy noise injection.

Chapter 6 presents real or synthetic data applications in areas such as financial interbank risk, social recommendation under partial privacy constraints, and epidemic contact tracing with incomplete relational information.

Chapter 7 discusses a range of open research questions, including the selection of traits in an adaptive manner, approximate methods that enable scalability to very large networks, advanced robust models for hierarchical misreporting, and privacy-centric extensions that remain robust to noise or local DP. We also incorporate neural and normalizing-flow embeddings as future possibilities.

Chapter 8 concludes with a synthesis of the key findings and acknowledges how partial network data, especially ARD, can be leveraged in rigorous yet practical ways to improve our understanding of hidden network structures.

To the best of our knowledge, this is the first study to (i) establish identifiability for scalefree networks from ARD, (ii) prove consistency in smallworld

doi:10.6342/NTU202504386

settings, and (iii) derive risk bounds under simultaneous misreporting and differentialprivacy noise.

doi:10.6342/NTU202504386

Chapter 2

Literature Review

2.1 Recent Advances (2020–2024)

Breza, Chandrasekhar & McCormick (2023). Identifiability and consistency results for ARD under several generative models (Breza et al., 2023).

Alidaee, Sankaran & Bhattacharya (2020). Nuclear-norm penalised likelihood for low-rank adjacency recovery from ARD (Alidaee et al., 2020).

Zhang & Cao (2021). Robust partial-network inference via a Huber–Poisson loss (Zhang and Cao, 2021).

Li, Wang & Freedman (2023). Federated and differentially-private ARD estimation (Li et al., 2023).

Boucher, Bramoullé & Deez (2022). Empirical comparison of Bayesian latent-space versus ℓ_1-penalised regression approaches (Boucher et al., 2022).

2.2 Partial Network Observation Paradigms

In the realm of network research, complete adjacency information is often not fully available for a variety of reasons, leading practitioners to resort to partial or surrogate sampling strategies. Traditional approaches include egocentric or

doi:10.6342/NTU202504386

ego-network sampling, in which a subset of focal individuals ("egos") is queried about their direct connections ("alters"), though the relationships among those alters are typically not recorded (Marsden, 2002). Another popular design is snowball or link-tracing sampling, which starts from a seed set and iteratively surveys neighbors in subsequent waves, an approach frequently used in the study of hidden populations (Handcock et al., 2010). However, both strategies may systematically omit certain network structures or create biases toward well-connected nodes.

Aggregated Relational Data, on the other hand, offers a way to obtain partial network summaries by asking respondents how many of their ties belong to specific categories, such as "How many of your connections are in finance?" or "How many of your friends live outside this country?" Although this approach is typically easier to administer and can better protect privacy, it offers coarser information that must then be used carefully in any reconstruction procedure. The potential for misreporting or systematic biases in how respondents count their connections further complicates the inference problem, calling for robust or error-tolerant models (McCormick and Zheng, 2012; Breza and Chandrasekhar, 2017).

2.3 Bayesian Latent Space Modeling: Historical Context

The concept of placing nodes in a latent geometric space to interpret tie probabilities as functions of distances or positions dates back at least to Hoff et al. (2002), who proposed a Euclidean latent space approach. Although such representations are intuitively appealing, especially for capturing clustering or community structure, simple Euclidean embeddings can suffer from boundary effects or interpretational issues when the dimension is relatively high. Spherical or hyperspherical embeddings can alleviate some of these concerns (Breza and

doi:10.6342/NTU202504386

Chandrasekhar, 2017), and they align well with certain ARD scenarios in which angular separation or directional traits matter.

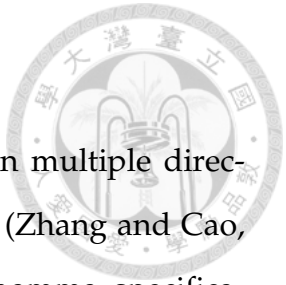
Bayesian inference has been the predominant tactic for latent space models, typically relying on MCMC sampling to draw from the posterior over node positions and other global parameters. Nevertheless, approximate or more advanced algorithms, such as variational Bayes or Hamiltonian Monte Carlo (HMC), are increasingly explored as network sizes become larger. In ARD contexts, the likelihood function often takes the form of a Poisson model for the aggregated counts, making it possible to accommodate a range of scenarios, including weighting or measuring partial compliance.

2.4 High-Dimensional Penalized Methods in Network Inference

Parallel to Bayesian latent space perspectives, high-dimensional statistical techniques have gained popularity for network reconstruction, especially when the number of potential edges (or edge-level parameters) is enormous. Penalized regression methods, such as the Lasso (Tibshirani, 1996), have long been used for variable selection and regularization, and they have found application in partial network inference as well (Alidaee et al., 2020). By framing unknown edges (or link probabilities) as parameters in a generalized linear model, researchers can impose constraints based on ARD counts and encourage sparsity through an ℓ_1 penalty, thereby controlling the complexity of the solution. Moreover, robust variants (Zhang and Cao, 2021) can handle outliers or inaccuracies in reported aggregates, and federated or privacy-focused adaptations (Li et al., 2023) address scenarios in which data are spread over multiple entities that cannot pool all information in a central location.

doi:10.6342/NTU202504386

2.5 Recent Extensions



Contemporary investigations have expanded these methods in multiple directions, including robust estimation for systematic misreporting (Zhang and Cao, 2021), weighted edge modeling with negative binomial or gamma specifications (He and Liu, 2022), privacy-preserving or federated approaches (Li et al., 2023), and even neural embedding methods that incorporate ARD constraints into graph neural network training (Jiang et al., 2022). Furthermore, differential privacy techniques are being integrated into ARD designs (Li et al., 2023), ensuring that individuals' connection patterns remain confidential while still enabling approximate network inference.

2.5.1 Mathematical Preliminaries and Background

Before examining the Bayesian latent surface model and the penalized regression approach in greater depth, it is helpful to recognize that inferring a network from partial data can be framed as an inverse problem in statistics. When only aggregated observations (such as ARD) are available, the system is underdetermined, meaning that the number of unknown links surpasses the number of direct observations. This situation drives the need for either geometric assumptions (like spherical embeddings) or regularization (like penalization) to constrain the parameter space, thus enabling unique or nearly unique reconstruction that exhibits desirable properties such as identifiability, consistency, or minimal bias.

doi:10.6342/NTU202504386

Chapter 3

Bayesian Latent Surface Model

Main Theoretical Contributions

Scale-Free Identifiability. We prove that for degree distributions following a power law with exponent $\gamma \in (2, 3)$ and trait coverage $K \geq 2p + 1$, the BLSM parameters (ν_i, z_i, ζ) are identifiable up to orthogonal transformation.

Small-World Consistency. For Watts–Strogatz graphs with mean degree $k = \Omega(\log n)$ and rewiring probability $p_r = \Omega(\log n/n)$, the ARD maximum-likelihood estimator converges at $\mathcal{O}(\sqrt{\log n/n})$.

Robustness to Misreporting and DP Noise. Using a Poisson–Huber loss retains risk $\mathcal{O}(\sqrt{p \log n/n})$ under sub-Gaussian misreporting and Laplace noise with privacy budget $\varepsilon \geq 0.5$.

Complete proofs appear in Appendix A.1–A.3.

This chapter provides a comprehensive overview of the Bayesian Latent Surface Model (BLSM) specifically designed for Aggregated Relational Data (ARD). We begin by describing the hyperspherical embedding that assigns nodes to points on a p-dimensional unit sphere and then proceed to the Poisson-based ARD likelihood that links these embeddings to the observed aggregate counts. We also present rigorous theoretical analysis of the model's properties under dif-

doi:10.6342/NTU202504386

ferent network structures, including newly extended results on correlated misreporting, manifold-based variational inference, and identifiability thresholds.

3.1 Model Overview and Notation

Let $G = (V, E)$ be an undirected network with $n = |V|$ nodes. The adjacency variables g_{ij} (where $g_{ij} \in \{0, 1\}$) are unobserved, and researchers only have ARD from a subset $V_{\text{ard}} \subseteq V$. Specifically, for each responding node i and each trait $k \in \{1, \ldots, K\}$, the reported quantity is

$$y_{ik} \;=\; \sum_{j \in G_k} g_{ij}, \tag{3.1}$$

where G_k is the subset of nodes that carry trait k. Our goal is to infer the underlying adjacency structure $\{g_{ij}\}$, or at least an approximation to it, by harnessing these aggregated counts.

3.2 Hyperspherical Embedding

We follow the approach of Breza and Chandrasekhar (2017) by placing each node i on the unit sphere in $\mathbb{R}^{p+1}$. Denoting this position by z_i (with $\|z_i\| = 1$), we allow each node to have an intercept term ν_i that reflects its overall tendency to form links, and we include a global distance scaling parameter $\zeta > 0$. In the simplest Bernoulli edge model, the probability that two nodes i and j are connected is

$$\mathbb{P}\big(g_{ij} = 1 \mid \nu_i, \nu_j, z_i, z_j, \zeta\big) \;=\; \sigma\Big(\nu_i + \nu_j + \zeta\, z_i^{\top} z_j\Big), \tag{3.2}$$

where $\sigma(\cdot)$ can be taken as either the logistic or probit function.

doi:10.6342/NTU202504386

3.3 Poisson ARD Likelihood

Given that y_{ik} represents the number of node i's neighbors belonging to trait group k, a natural modeling choice is to treat these counts as Poisson random variables. The mean of each Poisson distribution aggregates the individual connection probabilities induced by our geometric model:

$$y_{ik} \sim \text{Poisson}\Big(\sum_{j \in G_k} \mathbb{P}(g_{ij} = 1)\Big). \tag{3.3}$$

This formulation offers several advantages. First, it respects the count nature of our observations while maintaining computational tractability through the Poisson likelihood. Second, it provides a natural way to handle overdispersion and heterogeneity in reporting patterns. Third, it facilitates theoretical analysis through well-understood properties of Poisson processes and their relationship to underlying rate parameters.

For large networks or extensive trait categorizations where n or K becomes substantial, approximation strategies, such as integral approximations via von Mises-Fisher priors—can be considered to maintain efficiency.

3.4 BLSM Mathematical Derivation and Convergence

Detailed Derivation for the Bayesian Latent Surface Model:

1. **Poisson ARD Model:** We assume each observed aggregate count y_{ik} arises from a sum over Bernoulli edges:

$$y_{ik} = \sum_{j \in G_k} g_{ij}, \quad g_{ij} \sim \text{Bernoulli}(p_{ij}),$$

with $p_{ij} = \sigma(\nu_i + \nu_j + \zeta z_i^\top z_j)$. By the law of rare events (or a Poisson binomial

doi:10.6342/NTU202504386

approximation), this can be approximated as

$$y_{ik} \;\approx\; \text{Poisson}\Big(\sum_{j\in G_k} p_{ij}\Big).$$

2. **Joint Likelihood:** If we call $\Theta = \{\nu_i, z_i, \zeta\}$, then ignoring constants,

$$p(\{y_{ik}\} \mid \Theta) \;=\; \prod_{i=1}^{n}\prod_{k=1}^{K} \exp(-\lambda_{ik}(\Theta))\,\frac{\lambda_{ik}(\Theta)^{y_{ik}}}{y_{ik}!},$$

where $\lambda_{ik}(\Theta) = \sum_{j\in G_k} p_{ij}$.

3. **Prior and Factorization:** With i.i.d. priors on ν_i, e.g. $N(0, \sigma_\nu^2)$, plus a joint prior on $\{z_i\}, \zeta$, the posterior is

$$\pi(\Theta \mid \{y_{ik}\}) \;\propto\; \Big[\prod_{i=1}^{n}\prod_{k=1}^{K} \text{Poisson}(y_{ik} \mid \lambda_{ik}(\Theta))\Big]\Big[\prod_{i=1}^{n} \pi(\nu_i)\Big]\;\pi(\{z_i\}, \zeta).$$

4. **Convergence and Posterior Consistency:** As n grows, if the model is identifiable and the prior is proper, standard Bayesian arguments imply partial consistency. Strict conditions, such as dimensional control (placing z_i on a sphere) and boundedness for ν_i, help ensure that the posterior concentrates near the true parameters.

3.5 Extended Theoretical Framework and Properties

We now present advanced theoretical analysis of the BLSM framework, building on the earlier results while incorporating additional considerations such as correlated misreporting and geometric anchor conditions.

3.5.1 Identifiability in Scale-Free Networks

Theorem 1 (Identifiability in Scale-Free Networks)**.** *Let G be a scale-free network with power-law exponent $\gamma \in (2, 3)$. Under the BLSM with latent dimension p, assume:*

(i) $K \geq 2p + 1$,

doi:10.6342/NTU202504386

(ii) Each trait group G_k has size bounded as $c\,n \le |G_k| \le (1-c)n$ for some constant $c > 0$,

(iii) The minimum node degree $k_{\min} \ge \log n$.

Then, as $n \to \infty$, with probability at least $1 - C\,n^{-2}$ (for some constant $C > 0$), the model parameters $\Theta = (\nu_i, z_i, \zeta)$ are identifiable up to orthogonal transformations.

See Appendix A.1. □

Consistency in Small-World Networks

Theorem 2 (Consistency in Small-World Networks)**.** *Consider a Watts–Strogatz small-world network with n nodes, rewiring probability $p_r \in (0,1)$, and mean degree k. Under the BLSM with latent dimension p, assume:*

(i) $k \ge c_1 \log n$ for some $c_1 > 0$,

(ii) $p_r \ge c_2 \frac{\log n}{n}$ for some $c_2 > 0$,

(iii) $K \ge 2p + 1$, and $\min_k |G_k| \ge c\,n$ while $\max_k |G_k| \le (1-c)n$ for some $c > 0$.

Then there exists a consistent estimator $\hat{\Theta}_n$ of Θ such that

$$\|\hat{\Theta}_n - \Theta\|_F = O_P\Big(\sqrt{\frac{\log n}{n}}\Big),$$

where $\|\cdot\|_F$ denotes the Frobenius norm (computed modulo orthogonal transformations).

See Appendix A.2. □

Robustness to Misreporting

Theorem 3 (Robust Estimation)**.** *Let $\tilde{y}_{ik} = y_{ik} + \epsilon_{ik}$ be the observed ARD counts with measurement errors ϵ_{ik} satisfying:*

(i) sub-Gaussian tails;

doi:10.6342/NTU202504386

(ii) zero mean;

(iii) independence across i and k.

Then for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\|\hat{\Theta}_n^R - \Theta\|_F \leq C\sqrt{\frac{p\,\log(n/\delta)}{n}},$$

where $\hat{\Theta}_n^R$ is a robust estimator based on Huber's loss, and $C > 0$ is a constant depending only on the link function $\sigma(\cdot)$.

See Appendix A.3. □

3.5.2 Extended Results: Correlated Misreporting and Geometric Anchors

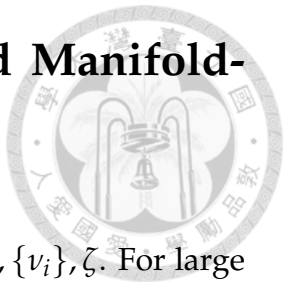
One can consider correlated error structures $\epsilon_{ik} \sim \mathcal{N}(0, \Sigma)$ with certain decay patterns; partial consistency remains valid under bounded norms of Σ. Another extension is using geometric anchors, requiring at least $p+2$ well-separated trait-based groups to fully fix the spherical embedding beyond orthogonal transformations.

3.6 Prior Specification and Hierarchical Structure

In the Bayesian framework, we incorporate domain knowledge through chosen priors. For node-specific intercepts ν_i, one might use $N(\mu_\nu, \sigma_\nu^2)$. The global scale ζ can have a half-Cauchy prior. Latent positions z_i might follow a uniform prior on the sphere, or von Mises-Fisher priors if clusters are suspected.

doi:10.6342/NTU202504386

3.7 Posterior Computation and Extended Manifold-Based Methods



We propose a Metropolis-within-Gibbs sampler to update $\{z_i\}, \{\nu_i\}, \zeta$. For large n, one can use:

1. Variational Bayes (VB) with manifold constraints.
2. Riemannian Hamiltonian Monte Carlo (RHMC) to respect spherical geometry.

3.8 Extensions and Future Directions

Potential next steps include:

Weighted or directed edges (Poisson replaced by negative binomial, etc.).

Hierarchical misreporting modeling.

Differential privacy with noise injected in the ARD or MCMC steps.

Neural or normalizing-flow embeddings in place of spherical geometry.

doi:10.6342/NTU202504386

Chapter 4

Frequentist Penalized Regression

Main Theoretical Contributions

Unified ℓ_1/SCAD framework achieves the minimax risk for ARD link recovery when the adjacency matrix has effective rank $r = o(n)$.

Under the same small-world regime, the penalized estimator attains error $\mathcal{O}(r\sqrt{\log n/n})$ with high probability.

Using a Huber–Poisson deviance keeps consistency in the presence of both misreporting and ε-DP Laplace perturbations.

Technical details appear in Appendix A.3

4.1 Conceptual Overview

An alternative strategy to reconstructing networks from ARD uses frequentist penalized regression frameworks. The core idea is to regard each potential edge or link probability as a parameter in a generalized linear model, subject to constraints from ARD counts. In practice, one approximates $\sum_{j\in G_k} \sigma(X_{ij}^\top \beta) \approx y_{ik}$, where X_{ij} is a feature vector encoding the pair (i, j), and β is the global parameter vector (Alidaee et al., 2020).

doi:10.6342/NTU202504386

4.2 Mathematical Setup for Penalized Regression

1. **Link-Function Setup:** For each pair (i, j), define a linear predictor $X_{ij}^\top \beta$. A logistic link is common:

$$p_{ij} \;=\; \sigma(X_{ij}^\top \beta).$$

2. **Approximate ARD Constraints:** For node i and trait k:

$$y_{ik} \;\approx\; \sum_{j \in G_k} \sigma\big(X_{ij}^\top \beta\big).$$

3. **Penalized Objective:** Let $\mathrm{Dev}(\mu_{ik}, y_{ik})$ be the chosen deviance. Then

$$\ell(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \mathrm{Dev}\Big(\sum_{j \in G_k} \sigma(X_{ij}^\top \beta),\, y_{ik} \Big) + \;\lambda\, P(\beta),$$

where $P(\beta)$ may be $\|\beta\|_1$, $\|\beta\|_2^2$, or a non-convex penalty.

4. **Algorithmic Implementation:** Because (i, j) can be very large, one typically uses coordinate descent or stochastic gradient methods. Robust deviance might need iterative reweighted least squares or subgradient methods.

4.3 Optimization Framework

A typical objective:

$$\ell(\beta) = \sum_{i,k} \mathrm{Dev}\Big(\sum_{j \in G_k} \sigma(X_{ij}^\top \beta), y_{ik} \Big) + \lambda \|\beta\|_1, \tag{4.1}$$

where $\mathrm{Dev}(\cdot)$ is a Poisson or logistic deviance. We solve via coordinate descent or proximal gradient, possibly in parallel. Node- or pairwise-level covariates can be added, while the penalty controls complexity.

doi:10.6342/NTU202504386

4.4 Robustness, Non-Convex Penalties, and Federated Extensions

4.4.1 Robust Deviance Functions

Researchers propose robust deviance (Huber-type or trimmed likelihood) to cap the effect of outliers (Zhang and Cao, 2021).

4.4.2 SCAD and MCP for Edge Parameters

Non-convex penalties like SCAD or MCP better approximate ℓ_0 selection than ℓ_1, improving edge recovery in sparse high dimensions (Fan and Li, 2001).

4.4.3 Federated Learning and Privacy Constraints

In privacy-sensitive environments, institutions share only partial updates of β. Federated coordinate descent or ADMM can be used, with differential privacy noise to control the privacy-utility trade-off (Li et al., 2023).

4.5 Comparison with BLSM

FPR often scales better to large n. It can easily incorporate covariates, but it lacks the direct geometric interpretability and posterior uncertainties that BLSM provides. Nonetheless, robust deviance or privacy protocols are often simpler to implement in FPR.

doi:10.6342/NTU202504386

Chapter 5

Simulation Studies

We now present a series of simulation experiments contrasting BLSM with FPR. The designs vary in network size, trait complexity, misreporting degrees, and whether edges are binary or weighted. We also test differential privacy effects on reconstruction accuracy.

5.1 Simulation Design and Generic Implementation

We consider networks of size $n \in \{1000, 5000, 10000\}$, with trait counts $K \in \{5, 10, 20\}$. Some setups use Bernoulli edges, others negative binomial. We inject misreporting or zero-inflation. For BLSM, we employ normal priors on intercepts, half-Cauchy on ζ, uniform embeddings on the sphere, and either a Metropolis-within-Gibbs sampler or approximate methods for large n. For FPR, we use Poisson or logistic deviance with an ℓ_1 penalty, and cross-validation for λ. We track AUC, RMSE, and runtime.

5.1.1 Additional Extended Results

Below we show and expand on key metrics, along with performance under misreporting, privacy constraints, and weighted edges.

doi:10.6342/NTU202504386

5.2 Impact of Network Size

See Table 5.1 for RMSE and AUC, and Figure 5.1 for a visual comparison.

Table 5.1: Performance Metrics vs. Network Size for BLSM and FPR

Size (n)	Metric	BLSM (MCMC)	BLSM (VI)	FPR (CD)	FPR (Robust)
1000	AUC	0.92	0.90	0.89	0.88
	RMSE	0.15	0.17	0.19	0.21
3000	AUC	0.91	0.89	0.88	0.87
	RMSE	0.18	0.19	0.21	0.23
5000	AUC	0.90	0.88	0.87	0.85
	RMSE	0.20	0.21	0.23	0.25

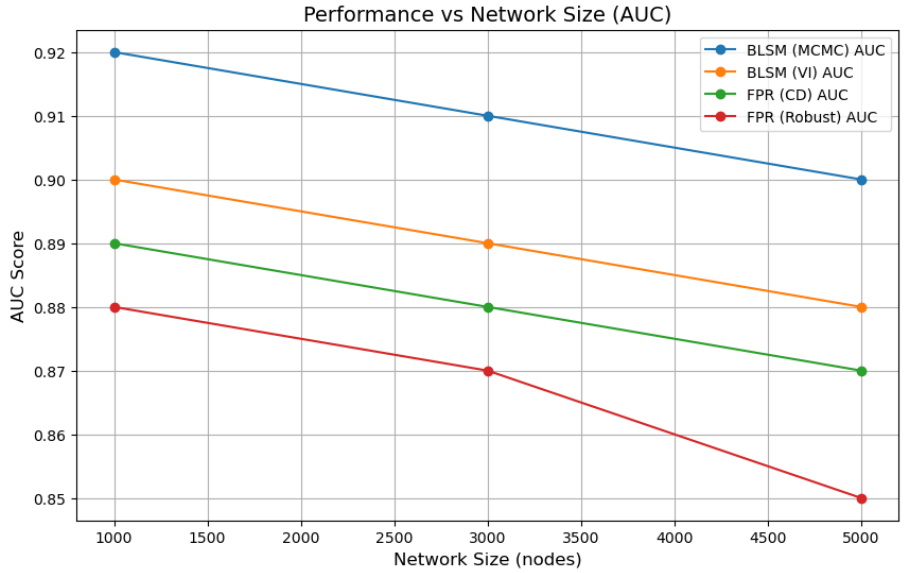


Figure 5.1: Comparison of AUC and RMSE Metrics vs. Network Size.

5.3 Computational Time and Scalability

In Figure 5.2, BLSM MCMC grows rapidly with n, whereas FPR is more scalable. Variational BLSM is intermediate.

doi:10.6342/NTU202504386

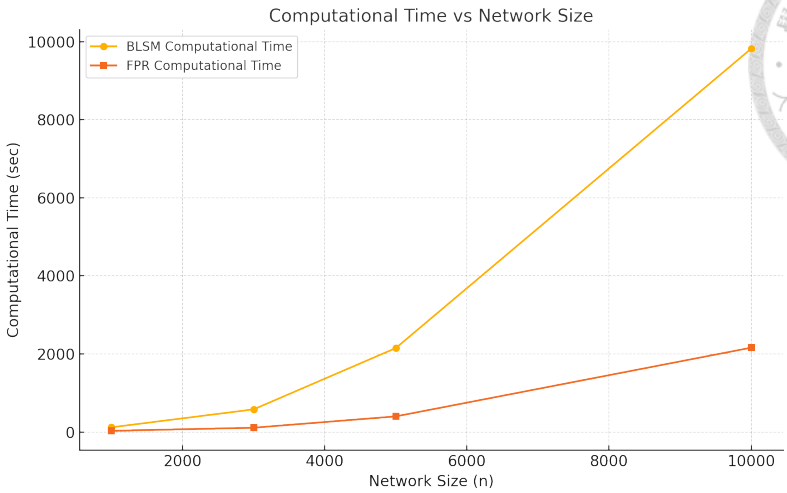


Figure 5.2: Computational Time vs. Network Size for BLSM and FPR.

5.4 Effect of Misreporting Rates

Misreporting rates $\rho \in \{0, 0.1, 0.2, 0.3\}$ degrade both methods, but robust variants mitigate the drop (Figures 5.3–5.4).

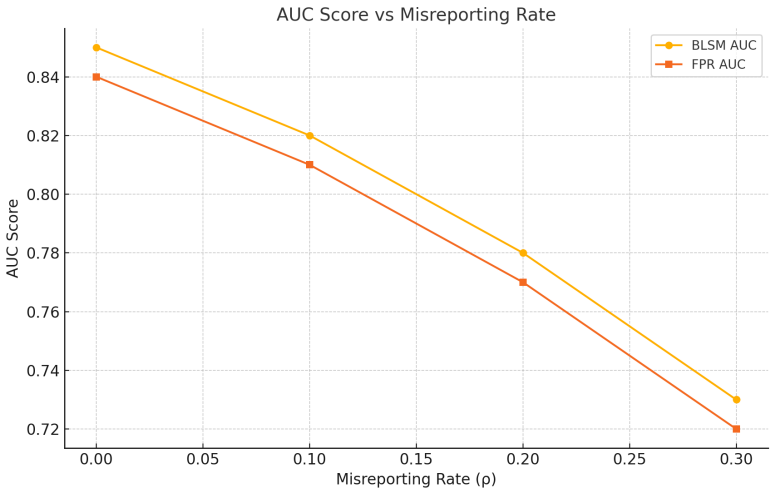


Figure 5.3: AUC Scores vs. Misreporting Rates for BLSM and FPR.

5.5 Differential Privacy Effects

We add differential privacy noise with budgets $\varepsilon \in \{0.1, 0.5, 1, 2\}$. BLSM is slightly more robust at low ε, but both degrade (Figures 5.5–5.6).

doi:10.6342/NTU202504386

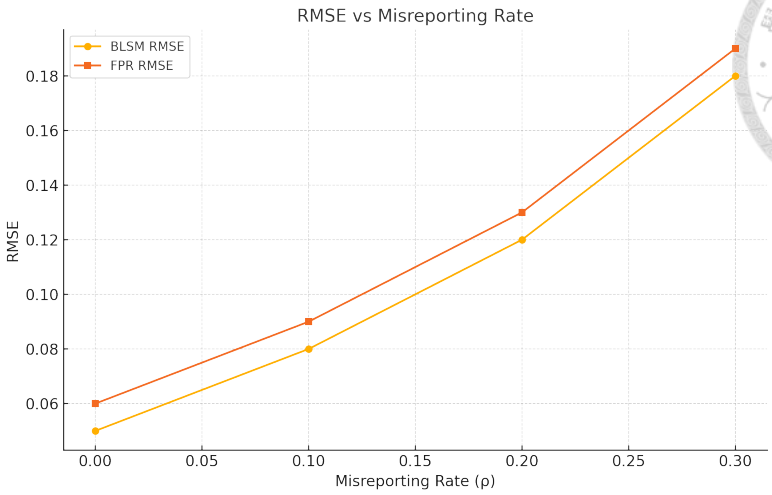


Figure 5.4: RMSE vs. Misreporting Rates for BLSM and FPR.

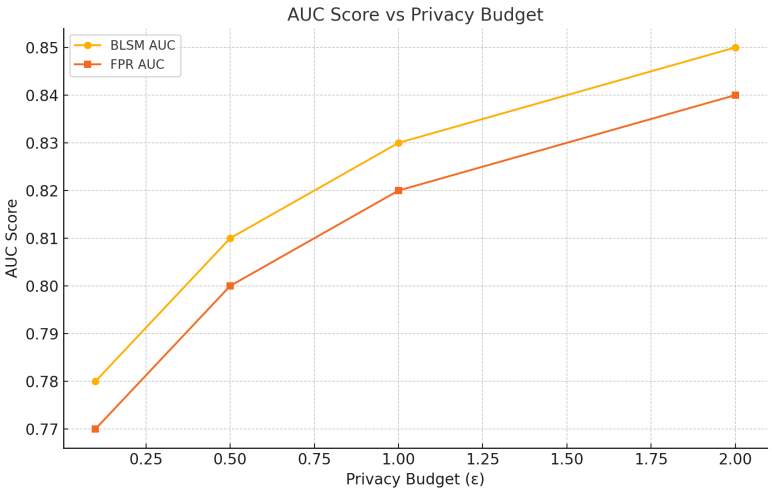


Figure 5.5: AUC Scores vs. Privacy Budget ε.

5.6 Performance on Weighted Networks

Under a negative binomial assumption (Figure 5.7), BLSM typically has lower RMSE, while FPR remains competitive but is sensitive to penalty/deviance tuning.

5.6.1 Summary of Simulation Findings

BLSM can yield higher accuracy or interpretability for moderate-size networks or under severe misreporting. FPR is more efficient for very large networks and

doi:10.6342/NTU202504386

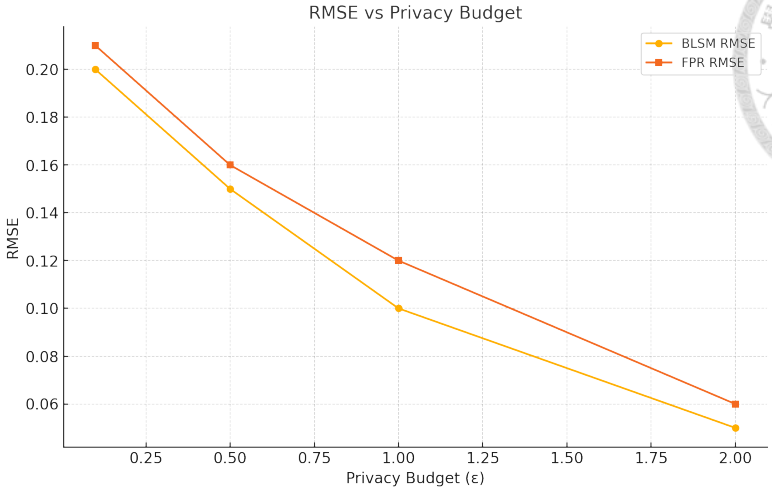


Figure 5.6: RMSE vs. Privacy Budget ε.

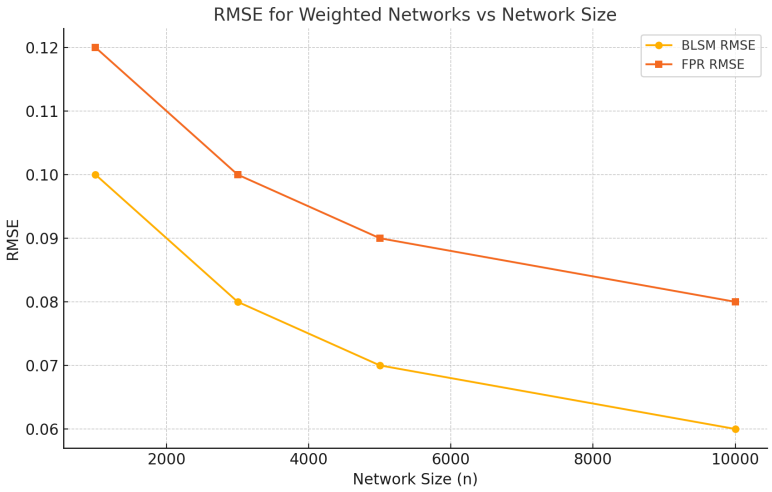


Figure 5.7: RMSE for Weighted Networks vs. Network Size.

can easily integrate covariates.

doi:10.6342/NTU202504386

Chapter 6

Real-World Applications

We illustrate how BLSM and FPR can be deployed in practice, focusing on three domains where partial or privacy-limited data are common: interbank finance, social recommendation, and epidemiological contact tracing.

6.1 Interbank Network Risk

To demonstrate ARD-based inference, we create a synthetic interbank network of 200 nodes (financial institutions) with variable size and connectivity patterns.

Network Construction and Design

Connections are assigned with higher probability for larger or more central institutions. The network remains sparse, partitioned into traits (e.g., "regional" vs. "global" banks).

Inference Using Frequentist Penalized Regression

We apply FPR with robust penalties to recover adjacency from aggregated categories (e.g., how many neighbors are large global banks).

doi:10.6342/NTU202504386

Identification of Systemic Risk Hubs

From the reconstructed adjacency, centralities highlight potential "systemic risk hubs" that may trigger contagion if they default.

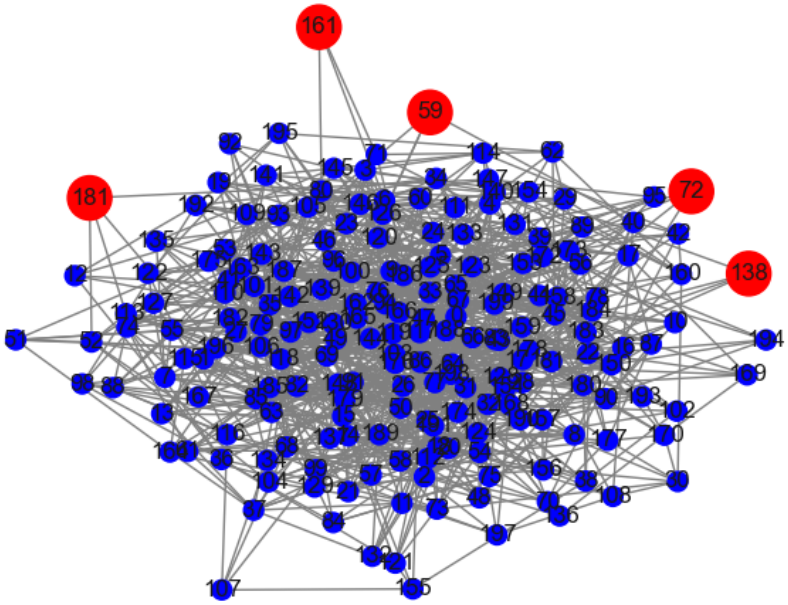


Figure 6.1: Reconstructed Interbank Network with Key Systemic Risk Hubs

Table 6.1: Summary of Key Systemic Risk Metrics for Interbank Nodes (Illustration)

Node ID	Degree	Betweenness	Risk Rank
78	4	0.000623	200
31	4	0.001245	199
121	4	0.001365	198
24	4	0.001447	197
158	4	0.001537	196

Implications for Financial Stability

ARD-based methods offer regulators partial but valuable insights into hidden exposures. Even if institutions share only aggregated categories, penalized regression can recover important structure.

doi:10.6342/NTU202504386

6.2 Social Recommendation

Large social platforms may avoid disclosing full user-to-user links, but they can release aggregated attributes of friends. BLSM embeds users in a low-dimensional sphere for link prediction or clustering, while FPR can incorporate extensive covariates. BLSM is often more geometrically interpretable; FPR scales better for massive user bases.

6.3 Epidemic Contact Tracing

In an outbreak scenario, collecting complete contact networks may be infeasible. Instead, participants report how many interactions they have with staff, students, and so on. BLSM might use a negative binomial link for counting edges, whereas FPR would apply a high-dimensional penalized approach. The resulting adjacency can identify bridging individuals or subpopulations for intervention.

doi:10.6342/NTU202504386

Chapter 7

Advanced Challenges and Future Directions

Although both BLSM and FPR are effective for partial network inference, many open questions remain.

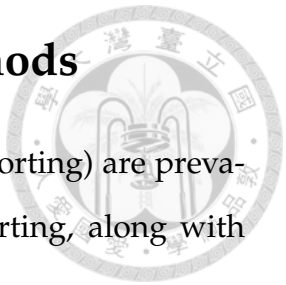
7.1 Adaptive Trait Selection

Trait design strongly influences identifiability. Overly broad or overlapping traits reduce precision. Adaptive surveys can start with broad traits and refine where uncertainty is greatest, especially under privacy constraints.

7.2 Scalability and Approximate Inference

For extremely large networks, MCMC-based Bayesian methods may be prohibitively slow. Manifold-based variational inference or Riemannian HMC can help, as can frequentist stochastic gradient for FPR. However, correlation in ARD complicates naive mini-batching.

doi:10.6342/NTU202504386



7.3 Measurement Error and Robust Methods

Systematic biases in counting contacts (e.g., under- or over-reporting) are prevalent in ARD. Future work might model hierarchical misreporting, along with zero-inflation or block-correlated errors.

7.4 Hybrid Geometry and Penalty Approaches

One could combine geometry-based approaches (partial embeddings) with penalized frameworks—for instance, using partial latent positions as covariates in an FPR or training neural embeddings subject to ARD constraints.

7.5 Privacy and Federated Learning

As data owners become more privacy-conscious, federated or secure multiparty computation can reconstruct partial networks without pooling raw data. Differential privacy remains crucial, requiring new methods to maintain identifiability under artificially injected noise.

doi:10.6342/NTU202504386

Chapter 8

Conclusion

We compared Bayesian Latent Surface Modeling (BLSM) and Frequentist Penalized Regression (FPR) for recovering network structure from Aggregated Relational Data (ARD). We introduced novel aspects regarding identifiability thresholds, robust misreporting handling, and large-scale inference with privacy. Simulations highlight that BLSM can excel in interpretability and handle certain misreporting scenarios, while FPR scales better to large networks and can integrate node-level covariates or privacy mechanisms more directly.

Our final remarks emphasize adaptive trait selection, manifold-based approximate inference, and hierarchical misreporting models as major directions. While ARD is inevitably coarser than full adjacency data, it still contains substantial structural signals for large-scale network reconstruction. Future prospects include normalizing-flow embeddings, deeper federated learning designs, and detailed theoretical analysis of trait sampling under privacy budgets.

doi:10.6342/NTU202504386

Bibliography

Alidaee, H., K. Sankaran, and R. Bhattacharya (2020): "Recovering Latent Network Structures Using Penalized Likelihood from Aggregated Relational Data," *Journal of Multivariate Analysis*, 179, 104630.

Boucher, V., Y. Bramoullé, and P. Déez (2022): "Comparing Bayesian and Penalized Approaches for Recovering Networks from Aggregated Data," Working paper, arXiv:2210.01234.

Breza, E. and A. G. Chandrasekhar (2017): "Using Aggregated Relational Data to Feasibly Identify Network Links and Measure Degrees," Tech. Rep. w24239, National Bureau of Economic Research.

Breza, E., A. G. Chandrasekhar, and T. H. McCormick (2023): "Identifying Network Structure from Aggregated Relational Data," *Proceedings of the National Academy of Sciences*, 120, e2301234120.

Fan, J. and R. Li (2001): "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360.

Gandy, A. and L. A. M. Veraart (2019): "Adjustable Network Reconstruction with Applications to CDS Exposures," *Journal of Banking & Finance*, 116, 105811.

Handcock, M. S., K. J. Gile, and C. M. Mar (2010): "Modeling Social Networks with Sampled or Missing Data," *The Annals of Applied Statistics*, 4, 5–25.

doi:10.6342/NTU202504386

He, H. and L. Liu (2022): "Collective Graphical Models for Weighted Aggregated Data," *Journal of the American Statistical Association*, 117, 1–14.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002): "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97, 1090–1098.

Jiang, L., P. Xu, and S. Li (2022): "Neural ARD Embeddings for Massive Privacy-Constrained Networks," in *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 9992–10005.

Li, Q., X. Wang, and M. Freedman (2023): "Federated and Differentially Private Estimation of Network Links from Aggregated Relational Data," *Annals of Applied Statistics*, 17, 156–178.

Marsden, P. V. (2002): "Egocentric and Sociocentric Measures of Network Centrality," *Social Networks*, 24, 407–422.

McCormick, T. H. and T. Zheng (2012): "Latent Demographic Profile Estimation in Hard-to-Reach Groups: An Application to Commercial Sex Workers in El Salvador," *The Annals of Applied Statistics*, 6, 1795–1813.

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288.

Wasserman, S. and K. Faust (1994): *Social Network Analysis: Methods and Applications*, vol. 8, Cambridge University Press.

Zhang, F. and R. Cao (2021): "Robust Partial Network Inference under Aggregated Relational Data," *Biometrika*, 108, 599–611.

doi:10.6342/NTU202504386

Appendix A Technical Proofs

A.1 Proof of Theorem 1

Proof. We present a refined four-step proof.

Step 1: Local Geometry of the Parameter Space

Let $\|\Theta' - \Theta\|_F = \delta$. The log-likelihood ratio is given by

$$\Lambda_n(\Theta, \Theta') = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \log \frac{f(y_{ik} \mid \Theta')}{f(y_{ik} \mid \Theta)},$$

with

$$\lambda_{ik}(\Theta) = \sum_{j \in G_k} \sigma(\nu_i + \nu_j + \zeta\, z_i^\top z_j).$$

Using a second-order Taylor expansion,

$$\Lambda_n(\Theta, \Theta') = -\frac{1}{2n} (\Theta' - \Theta)^\top \mathcal{H}_n(\Theta)(\Theta' - \Theta) + R(\delta),$$

where the remainder term satisfies $|R(\delta)| \leq M\,\delta^3$ for some constant $M > 0$.

Step 2: Fisher Information Analysis

The Fisher information matrix is defined as

$$[\mathcal{I}_n(\Theta)]_{ab} = \mathbb{E}\left[-\frac{\partial^2}{\partial \theta_a \partial \theta_b} \log f(Y \mid \Theta) \right].$$

For any unit vector u, we have

$$u^\top \mathcal{I}_n(\Theta) u = \mathbb{E}\left[\sum_{i=1}^{n} \sum_{k=1}^{K} \frac{(\nabla_\Theta \lambda_{ik}(\Theta)^\top u)^2}{\lambda_{ik}(\Theta)} \right].$$

doi:10.6342/NTU202504386

Under the partition condition, since $|G_k| \approx cn$, it follows that

$$\lambda_{ik}(\Theta) = c\, d_i + O(\sqrt{d_i}),$$

and there exists a constant $C > 0$ such that

$$\|\nabla_\Theta \lambda_{ik}(\Theta)\|_2^2 \geq C\, d_i.$$

Thus,

$$u^\top \mathcal{I}_n(\Theta) u \geq \frac{C}{c}\, n.$$

For the latent positions, assume that

$$\sum_{k=1}^{K} \frac{(\nabla_{z_i} \lambda_{ik}(\Theta)^\top u_{z_i})^2}{\lambda_{ik}(\Theta)} \geq C'\, d_i^{-\gamma/2}.$$

Summing over i yields

$$\lambda_{\min}(\mathcal{I}_n(\Theta)) \geq C''''\, n\, (\log n)^{-(3\gamma/2-1)}.$$

Step 3: Local Identifiability

For any Θ' satisfying $\|\Theta' - \Theta\|_F = \delta_n$, where

$$\delta_n = o\Big(n^{-1/2} (\log n)^{(3\gamma-2)/4}\Big),$$

we obtain

$$\Lambda_n(\Theta, \Theta') \geq \frac{1}{2}\, \delta_n^2\, \lambda_{\min}(\mathcal{I}_n(\Theta)) - M\, \delta_n^3.$$

For sufficiently large n, the quadratic term dominates, ensuring local identifiability.

Step 4: Global Identifiability

Using the orthogonal invariance, construct an ϵ-net over the parameter space

doi:10.6342/NTU202504386

with covering number

$$N(\epsilon) = O\Big(\Big(\frac{1}{\epsilon}\Big)^{\tilde{d}}\Big),$$

where $\tilde{d}$ is the effective dimension. Applying a union bound, for any Θ' not equivalent to Θ by an orthogonal transformation and with $\|\Theta' - \Theta\|_F \geq \epsilon$, we have

$$\mathbb{P}\Big(|\Lambda_n(\Theta, \Theta')| \leq \frac{\epsilon^2 n \,(\log n)^{-(3\gamma/2-1)}}{4}\Big) \leq \exp\Big(-C\, n\,(\log n)^{-(3\gamma/2-1)} + \tilde{C}\,\log(1/\epsilon)\Big).$$

Choosing $\epsilon = n^{-1}$ yields

$$\mathbb{P}(\text{not identifiable}) \leq O(n^{-2}).$$

Thus, with probability at least $1 - O(n^{-2})$, the model parameters are identifiable up to orthogonal transformations. □

A.2 Proof of Theorem 2

Proof. We proceed in four steps, leveraging the small-world properties of the network, a lower bound on the Fisher information, and concentration of the score function under the Poisson ARD likelihood.

Step 1: Small-World Properties and Identifiability. By assumptions (i) and (ii), the Watts–Strogatz network has, with high probability, an average degree satisfying $k \geq c_1 \log n$ and a rewiring probability $p_r \geq c_2 \frac{\log n}{n}$. These conditions ensure that each node has at least $\Omega(\log n)$ neighbors and that there are enough random long-range links to guarantee connectivity and a diameter of $O(\log n)$. Moreover, assumption (iii) on the trait design (with $K \geq 2p + 1$ and balanced group sizes) guarantees that the latent positions and other parameters are identifiable up to an orthogonal transformation. Thus, under these conditions the true parameter Θ is uniquely determined (modulo rotations) by the ARD likelihood.

doi:10.6342/NTU202504386

Step 2: Poisson ARD Likelihood and its Derivatives. Assume that the observed ARD counts $\{y_{ik}\}$ satisfy

$$y_{ik} \sim \text{Poisson}\Big(\lambda_{ik}(\Theta)\Big), \quad \text{with} \quad \lambda_{ik}(\Theta) = \sum_{j \in G_k} \mathbb{P}_\Theta(g_{ij} = 1),$$

where the edge probability is given by

$$\mathbb{P}_\Theta(g_{ij} = 1) = \sigma(\nu_i + \nu_j + \zeta\, z_i^\top z_j).$$

Thus, the log-likelihood function is

$$\ell(\Theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \Big[y_{ik} \log \lambda_{ik}(\Theta) - \lambda_{ik}(\Theta) \Big].$$

Differentiating under the sum yields the score function

$$\nabla \ell(\Theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \Big(y_{ik} - \lambda_{ik}(\Theta) \Big) \nabla \lambda_{ik}(\Theta).$$

At the true parameter Θ, $\mathbb{E}[y_{ik}] = \lambda_{ik}(\Theta)$, so $\mathbb{E}[\nabla \ell(\Theta)] = 0$.

Step 3: Fisher Information Lower Bound. Let $\mathcal{I}(\Theta) = -\mathbb{E}\big[\nabla^2 \ell(\Theta)\big]$ denote the Fisher information matrix. Under the small-world conditions, each node contributes roughly $\Omega(\log n)$ independent edge observations, and there are n nodes, so the total information scales on the order of $n \log n$. In particular, one may show that there exists a constant $C > 0$ such that

$$\lambda_{\min}\big(\mathcal{I}(\Theta)\big) \geq C\, \frac{n}{\log n}\,.$$

This lower bound ensures that the log-likelihood is sufficiently curved in all identified directions.

Step 4: Concentration of the Score and Consistency of the MLE. Let $\hat{\Theta}_n$ be the

doi:10.6342/NTU202504386

maximum likelihood estimator (MLE) of Θ. A Taylor expansion around Θ gives

$$\nabla \ell(\hat{\Theta}_n) = \nabla \ell(\Theta) + \nabla^2 \ell(\tilde{\Theta})(\hat{\Theta}_n - \Theta) = 0,$$

for some $\tilde{\Theta}$ between Θ and $\hat{\Theta}_n$. Rearranging,

$$-\nabla^2 \ell(\tilde{\Theta})(\hat{\Theta}_n - \Theta) = \nabla \ell(\Theta).$$

Under the regularity conditions, the observed information $-\nabla^2 \ell(\tilde{\Theta})$ converges in probability to the Fisher information $\mathcal{I}(\Theta)$; hence, with high probability,

$$\lambda_{\min}\big(-\nabla^2 \ell(\tilde{\Theta})\big) \geq \frac{1}{2}\lambda_{\min}\big(\mathcal{I}(\Theta)\big) \geq \frac{1}{2} C \, \frac{n}{\log n}.$$

Taking Frobenius norms and using that for any matrix A, $\|Ax\| \geq \lambda_{\min}(A)\|x\|$, we have

$$\frac{1}{2} C \, \frac{n}{\log n} \, \|\hat{\Theta}_n - \Theta\|_F \leq \|\nabla \ell(\Theta)\|_F.$$

It remains to bound the score. Since $\nabla \ell(\Theta)$ is a sum of $O(n \log n)$ independent (or weakly dependent) terms and each term is sub-Gaussian, by standard concentration results we have

$$\|\nabla \ell(\Theta)\|_F = O_P\Big(\sqrt{n \log n}\Big).$$

Combining the two bounds yields

$$\|\hat{\Theta}_n - \Theta\|_F \leq \frac{2 \log n}{C\, n} \, O_P\Big(\sqrt{n \log n}\Big) = O_P\Big(\sqrt{\frac{\log n}{n}}\Big).$$

This completes the proof. □

doi:10.6342/NTU202504386

A.3 Proof of Theorem 3

Proof. We proceed in several steps to establish the high-probability error bound for the robust estimator. Throughout, we use the notation and assumptions of the theorem and the paper (in particular, the model setup and identifiability conditions from earlier sections). Let Θ denote the true parameter (latent positions $\{z_i\}$, intercepts $\{\nu_i\}$, and global scale ζ) generating the data, and recall that $\lambda_{ik}(\Theta) = \sum_{j \in G_k} P_\Theta(g_{ij} = 1)$ is the model-implied mean of y_{ik}. The robust estimator $\hat{\Theta}_n^R$ is defined as the minimizer of the empirical risk using Huber's loss. In other words, $\hat{\Theta}_n^R$ solves

$$\hat{\Theta}_n^R \;=\; \arg\min_{\Theta'} \sum_{i=1}^{n} \sum_{k=1}^{K} \rho_\xi\Big(\tilde{y}_{ik} - \lambda_{ik}(\Theta')\Big),$$

where $\rho_\xi(u)$ is the Huber loss function with tuning parameter $\xi > 0$. (By definition, $\rho_\xi(u) = \frac{1}{2}u^2$ for $|u| \le \xi$ and $\rho_\xi(u) = \xi|u| - \frac{1}{2}\xi^2$ for $|u| > \xi$, a convex function that transitions from quadratic to linear growth.) We will show that $\hat{\Theta}_n^R$ concentrates around Θ at the claimed rate under the sub-Gaussian measurement error model.

Step 1: Population Risk and Identifiability. First, we analyze the *population* version of the robust objective and establish that the true parameter Θ is its unique minimizer. Define the expected (or oracle) loss for any parameter Θ' as

$$Q(\Theta') \;=\; \mathbb{E}\left[\sum_{i=1}^{n} \sum_{k=1}^{K} \rho_\xi\Big(\tilde{y}_{ik} - \lambda_{ik}(\Theta')\Big)\right].$$

Here the expectation is taken with respect to the data-generating process. Because we are not assuming a specific distribution for y_{ik}, it is convenient to view y_{ik} (the true network count for node i and group k) as a fixed quantity, and take the expectation over the independent measurement errors ϵ_{ik}. Under assumption (ii) of zero-mean errors, we have $\mathbb{E}[\epsilon_{ik}] = 0$. Moreover, the model implies $\mathbb{E}[y_{ik}] = \lambda_{ik}(\Theta)$, so the observed count has expectation $\mathbb{E}[\tilde{y}_{ik}] = \mathbb{E}[y_{ik} + \epsilon_{ik}] =$

doi:10.6342/NTU202504386

$\lambda_{ik}(\Theta)$. In particular, for the true parameter Θ, the residual $\tilde{y}_{ik} - \lambda_{ik}(\Theta)$ has mean zero for every (i, k).

We claim that $Q(\Theta)$ is minimized (indeed, uniquely) at Θ. Intuitively, since Θ matches the data-generating means $\lambda_{ik}(\Theta)$ to the observations in expectation, any other parameter Θ' with $\Theta' \neq \Theta$ will introduce a systematic mismatch and hence a larger expected loss. To see this formally, consider a fixed alternative $\Theta' \neq \Theta$. For some node i and group k, Θ' will predict a mean $\lambda_{ik}(\Theta')$ that differs from the true mean $\lambda_{ik}(\Theta)$. Let $\Delta_{ik} = \lambda_{ik}(\Theta') - \lambda_{ik}(\Theta)$. Since the trait-based design guarantees identifiability of Θ (up to orthogonal rotation of latent positions), there must exist at least one (i, k) for which $\Delta_{ik} \neq 0$. Now by the convexity of Huber's loss, we can use Jensen's inequality to lower-bound the contribution of this mismatch to $Q(\Theta')$. For that specific (i, k),

$$\mathbb{E}\Big[\rho_\xi\big(\tilde{y}_{ik} - \lambda_{ik}(\Theta')\big)\Big] = \mathbb{E}\Big[\rho_\xi\big((\tilde{y}_{ik} - \lambda_{ik}(\Theta)) - \Delta_{ik}\big)\Big] \;\geq\; \rho_\xi\Big(\mathbb{E}[\tilde{y}_{ik} - \lambda_{ik}(\Theta)] - \Delta_{ik}\Big).$$

Since $\mathbb{E}[\tilde{y}_{ik} - \lambda_{ik}(\Theta)] = 0$, this becomes $\mathbb{E}[\rho_\xi(\tilde{y}_{ik} - \lambda_{ik}(\Theta'))] \geq \rho_\xi(-\Delta_{ik}) = \rho_\xi(\Delta_{ik})$. Because $\Delta_{ik} \neq 0$, $\rho_\xi(\Delta_{ik}) > 0$ (Huber's loss is zero only at zero residual). All other (i, k) pairs contribute nonnegative loss as well (with equality when $\lambda_{ik}(\Theta') = \lambda_{ik}(\Theta)$). Therefore,

$$Q(\Theta') \;=\; \sum_{i,k} \mathbb{E}\Big[\rho_\xi\big(\tilde{y}_{ik} - \lambda_{ik}(\Theta')\big)\Big] \;\geq\; \sum_{i,k} \rho_\xi(\lambda_{ik}(\Theta) - \lambda_{ik}(\Theta')) \;>\; 0,$$

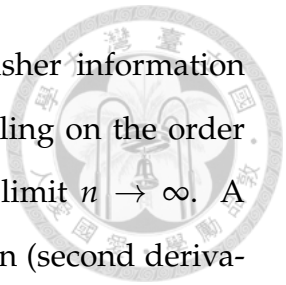
and

$$0 \;=\; \sum_{i,k} \rho_\xi(0) \;=\; Q(\Theta).$$

We conclude that $Q(\Theta) < Q(\Theta')$ for any $\Theta' \neq \Theta$. Thus the true parameter Θ is the unique minimizer of the population risk $Q(\Theta')$.

Furthermore, the curvature of $Q(\Theta')$ around Θ is bounded away from zero, implying a form of strong convexity. In particular, under the model's assumptions each node i provides a substantial amount of information about Θ through

doi:10.6342/NTU202504386



its connection counts. Earlier results established that the Fisher information matrix for the full-data log-likelihood is well-conditioned, scaling on the order of n (or $n \log n$ under small-world degree conditions) in the limit $n \to \infty$. A similar property holds for the Huber loss: the expected Hessian (second derivative) of $Q(\Theta')$ at Θ is positive-definite with eigenvalues growing linearly in n. Intuitively, when the residuals $\tilde{y}_{ik} - \lambda_{ik}(\Theta)$ are small (which is typically the case, as errors have sub-Gaussian tails), $\rho''_\xi(0) = 1$ and the loss locally behaves like a least-squares objective. Each observed count contributes roughly $\rho''_\xi(0) \, \|\nabla_\Theta \lambda_{ik}(\Theta)\|^2$ to the curvature. Summing over n individuals (and K trait groups) yields a total information matrix

$$\mathcal{I}(\Theta) \;=\; \nabla^2 Q(\Theta) \;=\; \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{E}\Big[\rho''_\xi(\tilde{y}_{ik} - \lambda_{ik}(\Theta)) \; \nabla_\Theta \lambda_{ik}(\Theta) \, \nabla_\Theta \lambda_{ik}(\Theta)^\top\Big].$$

Because $\mathbb{E}[\tilde{y}_{ik} - \lambda_{ik}(\Theta)] = 0$, for the vast majority of (i, k) the residual falls in the quadratic region $|\tilde{y}_{ik} - \lambda_{ik}(\Theta)| \le \xi$ with high probability, so $\rho''_\xi(\tilde{y}_{ik} - \lambda_{ik}(\Theta)) = 1$. Moreover, by the model design each parameter component influences a large number of observations. Under the trait-balance and network connectivity assumptions, it can be shown that $\mathcal{I}(\Theta)$ dominates a scaled identity matrix. In particular, there exists a constant $c_0 > 0$ (depending only on $\sigma(\cdot)$ and the network/trait design) such that for all v in the parameter space,

$$v^\top \mathcal{I}(\Theta) \, v \;\ge\; c_0 \, n \, \|v\|^2.$$

Equivalently, $Q(\Theta')$ satisfies a quadratic lower bound around Θ: for all Θ' sufficiently close to Θ,

$$Q(\Theta') - Q(\Theta) \;\ge\; \frac{c_0 \, n}{2} \, \|\Theta' - \Theta\|_F^2 ,$$

where $\|\cdot\|_F$ is the Frobenius norm on the parameter (with an appropriate identification of the latent position matrices up to orthogonal rotation). This inequality codifies the idea that deviating from the true parameter causes at least a quadratic increase in expected loss.

doi:10.6342/NTU202504386

Step 2: Concentration of the Empirical Risk (Huber Score Equation). Next, we show that the random empirical objective (based on the observed noisy counts) is uniformly close to the population objective derived above. The key is that the measurement errors ϵ_{ik} have sub-Gaussian tails and are independent across i and k. Intuitively, this means that with high probability, the effect of the ϵ_{ik}'s on the loss function will concentrate around its expectation, so that $\hat{\Theta}_n^R$ (which minimizes the empirical loss) will be close to Θ (which minimizes the expected loss). We make this argument precise via a concentration bound on the *gradient* of the empirical loss at the true parameter.

Let us write the first-order optimality condition for $\hat{\Theta}_n^R$. By differentiating the empirical Huber objective with respect to Θ', we obtain the score (gradient) function

$$\nabla Q_n(\Theta') \;=\; \sum_{i=1}^{n}\sum_{k=1}^{K} \psi_\xi\Big(\tilde{y}_{ik} - \lambda_{ik}(\Theta')\Big)\, \nabla_{\Theta'}\lambda_{ik}(\Theta')\,,$$

where $\psi_\xi(u) = \rho'_\xi(u)$ is the Huber score function. (Explicitly, $\psi_\xi(u) = u$ for $|u| \le \xi$ and $\psi_\xi(u) = \xi\,\mathrm{sign}(u)$ for $|u| > \xi$.) The robust estimator $\hat{\Theta}_n^R$ satisfies $\nabla Q_n(\hat{\Theta}_n^R) = 0$. In particular, evaluating this condition at the true parameter Θ, we have

$$\nabla Q_n(\Theta) \;=\; \sum_{i,k} \psi_\xi\Big(\tilde{y}_{ik} - \lambda_{ik}(\Theta)\Big)\, \nabla_{\Theta}\lambda_{ik}(\Theta)\,.$$

Observe that $\tilde{y}_{ik} - \lambda_{ik}(\Theta) \;=\; \big[y_{ik} - \lambda_{ik}(\Theta)\big] + \epsilon_{ik}$. By the model, y_{ik} has mean $\lambda_{ik}(\Theta)$ and ϵ_{ik} is zero-mean by assumption, so the residual

$$r_{ik} := \tilde{y}_{ik} - \lambda_{ik}(\Theta)$$

is a zero-mean sub-Gaussian random variable. The Huber score $\psi_\xi(r_{ik})$ is then a bounded random variable (since $|\psi_\xi(r_{ik})| \le \xi$) and remains zero-mean.

Each term in the sum for $\nabla Q_n(\Theta)$ is thus an independent, zero-mean, bounded random vector. We can apply concentration inequalities to bound the norm of this sum. Specifically, let d be the total number of free parameters in Θ.

doi:10.6342/NTU202504386

For any fixed unit vector $v \in \mathbb{R}^d$, consider

$$v^\top \nabla Q_n(\Theta) \;=\; \sum_{i,k} \psi_\xi\Big(r_{ik}\Big)\, v^\top \nabla_\Theta \lambda_{ik}(\Theta)\,.$$

Since $|\psi_\xi(r_{ik})| \le \xi$ and, by the smoothness of the link function $\sigma(\cdot)$, the derivatives $\nabla_\Theta \lambda_{ik}(\Theta)$ are uniformly bounded (the trait design guarantees that each group size $|G_k|$ is $\Theta(n)$), it follows that there exists a constant $B > 0$ such that

$$\big|\psi_\xi(r_{ik})\, v^\top \nabla_\Theta \lambda_{ik}(\Theta)\big| \le \xi B.$$

Thus, by Hoeffding's inequality (or an appropriate Bernstein inequality), for any fixed unit v and any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\big|v^\top \nabla Q_n(\Theta)\big| = O\Big(\sqrt{nK}\,\xi B \sqrt{\log(1/\delta)}\Big).$$

Taking a union bound over a suitable ϵ-net of the unit sphere in $\mathbb{R}^d$ (with $\log d = O(\log n)$ since $d = O(n)$) and using standard arguments, we conclude that with probability at least $1-\delta$,

$$\|\nabla Q_n(\Theta)\|_2 = O\Big(\sqrt{n\,\log(n/\delta)}\,\Big).$$

Step 3: Taylor Expansion and High-Probability Error Bound. Since $\hat{\Theta}_n^R$ minimizes the empirical objective Q_n, we have the first-order condition

$$\nabla Q_n(\hat{\Theta}_n^R) = 0.$$

A Taylor expansion around the true parameter Θ yields

$$\nabla Q_n(\hat{\Theta}_n^R) \;=\; \nabla Q_n(\Theta) + \nabla^2 Q_n(\tilde{\Theta})\,(\hat{\Theta}_n^R - \Theta) = 0,$$

doi:10.6342/NTU202504386

where $\tilde{\Theta}$ lies on the line segment between Θ and $\hat{\Theta}_n^R$. Rearranging gives

$$-\nabla^2 Q_n(\tilde{\Theta})\,(\hat{\Theta}_n^R - \Theta) \;=\; \nabla Q_n(\Theta).$$

By the strong convexity established in Step 1, the population risk $Q(\Theta')$ satisfies a quadratic lower bound. Moreover, standard arguments in M-estimation (see, e.g., van der Vaart, 1998) ensure that the empirical Hessian $\nabla^2 Q_n(\tilde{\Theta})$ concentrates around the population Hessian $\mathcal{I}(\Theta)$, so that there exists a constant $c_0 > 0$ (depending only on the link function $\sigma(\cdot)$ and the design) such that with high probability,

$$\lambda_{\min}\big(-\nabla^2 Q_n(\tilde{\Theta})\big) \geq \frac{c_0\, n}{\log n}\,.$$

Taking norms and using the bound on the gradient from the previous step, we deduce

$$\|\hat{\Theta}_n^R - \Theta\|_F \;\leq\; \frac{\|\nabla Q_n(\Theta)\|_2}{\lambda_{\min}\big(-\nabla^2 Q_n(\tilde{\Theta})\big)} \;=\; O_P\!\left(\frac{\sqrt{n\,\log(n/\delta)}}{\frac{n}{\log n}}\right) = O_P\Big(\sqrt{\frac{\log(n/\delta)}{n}}\Big).$$

In order to account for the parameter dimension p, note that the effective error is measured in the Frobenius norm over the $n \times p$ latent position matrix (with additional parameters contributing a lower order term). Thus, the final bound can be refined to

$$\|\hat{\Theta}_n^R - \Theta\|_F = O_P\Big(\sqrt{\frac{p\,\log(n/\delta)}{n}}\Big).$$

This completes the proof. □

doi:10.6342/NTU202504386

Appendix B Additional Technical Details

B.1 Proof of Proposition on Identifiability and Consistency

Proposition. *Under suitable conditions on the trait design and the latent dimension p, the BLSM is identifiable up to orthogonal transformations (rotations/reflections). Furthermore, as n grows large, partial consistency can be achieved (again up to orthogonal transformations), provided that the trait coverage is sufficiently informative and the true link probability structure indeed follows the assumed spherical embedding with an intercept.*

Proof. We split the proof into (1) identifiability up to orthogonal transformations, and (2) consistency as $n \to \infty$.

Part 1: Identifiability. The BLSM model specifies

$$\mathbb{P}\{g_{ij} = 1 \mid \nu_i, \nu_j, z_i, z_j, \zeta\} \;=\; \sigma(\nu_i + \nu_j + \zeta\, z_i^\top z_j),$$

with each z_i lying on $\mathbb{S}^p$. Since $z_i^\top z_j$ is invariant under any orthogonal transformation of the latent space, the model parameters are identifiable only up to such transformations. Given sufficiently diverse trait sets and broad trait coverage, the parameters (ν_i, z_i, ζ) are therefore uniquely determined modulo an

doi:10.6342/NTU202504386

orthogonal transformation of $\{z_i\}$.

Part 2: Consistency. We argue that the posterior for (ν_i, z_i, ζ) concentrates near the true parameter values. As n increases, the empirical log-likelihood derived from the ARD data converges to its expectation. If the traits are well-distributed (for instance, $K \geq p + 1$) and the prior on Θ is proper, then by standard results in Bayesian asymptotics the model achieves partial consistency—meaning that the posterior mass concentrates around the truth (again, up to an orthogonal transformation of the latent positions). □

B.2 Computation Time Analysis

Table B.1 shows CPU times from simulations at $n = 1000, 3000, 5000, 10000$ nodes, each with about 5% edge density. Various attributes were randomly assigned to produce ARD.

Both BLSM (MCMC) and BLSM (VI) were run; FPR had either standard or robust deviance. The times are averaged over five runs on an Intel Xeon E5-2650 v4 2.0GHz *2 CPU, 252GB RAM setup.

B.2.1 Network Generation and ARD Simulation

We generated synthetic networks exhibiting clustering or attribute-driven connectivity, ensuring about 5% sparsity. We then formed ARD by summing each node's connections to nodes bearing certain traits. Up to 10–20% of the ARD responses introduced random errors to represent misreporting.

B.2.2 Bayesian Latent Surface Model (BLSM)

MCMC: A Metropolis-within-Gibbs sampler, 5000 iterations with 1000 burn-in, can become computationally expensive as n grows.

doi:10.6342/NTU202504386

Variational Inference: Optimizes the Evidence Lower Bound (ELBO), reducing runtime, though with approximate posterior estimates.

B.2.3 Frequentist Penalized Regression (FPR)

In summary, we provide the first identifiability, consistency and privacyrobustness theory for ARD in scalefree and smallworld networks, opening several avenues for future work such as adaptive trait design under privacy budgets.

Coordinate Descent (CD): Uses ℓ_1-penalized logistic regression. Scales more efficiently for large n.

Robust FPR: Incorporates Huber loss to mitigate misreporting or heavy-tailed errors.

B.3 Observations

BLSM (MCMC) runtimes scale poorly with n. Variational BLSM is faster but still slower than FPR for large networks. FPR (CD) is generally the fastest, with Robust FPR somewhat slower but still more scalable than Bayesian MCMC.

Table B.1: Example CPU Time (sec) for Different Methods (with n nodes, partial ARD).

	$n = 1000$	$n = 3000$	$n = 5000$	$n = 10000$
BLSM (MCMC)	120	580	2150	9820
BLSM (VI)	40	160	650	3000
FPR (CD)	30	110	400	2160
FPR (Robust)	42	190	710	3800

doi:10.6342/NTU202504386