POLITECNICO DI TORINO

MASTER's Degree in COMPUTER ENGINEERING

MASTER's Degree Thesis

A Bayesian Framework for Distributed Mean Estimation with Adaptive Network Topology

Supervisors

Prof. Emilio LEONARDI

Dr. Franco GALANTE

Candidate

Valerio XEFTERIS

December 2025

Abstract

This thesis introduces **Dynamic Bayesian Mean Estimation (DBME)**, a decentralized probabilistic framework for distributed mean estimation in networked multi-agent systems. The framework is designed for scenarios where a set of agents, each collecting local observations, cooperatively infer a shared latent quantity while communicating only with their neighbors. DBME offers an alternative to consensus and PAC-based methods by expressing distributed inference through a probabilistic formulation that explicitly represents uncertainty. Each agent maintains a parametric belief whose dynamics follow Bayesian principles of evidence accumulation and variance contraction.

Nodes start from a bootstrap prior encoding an empirical or subjective belief about their local mean. As observations are collected, these beliefs are updated incrementally and fused with neighborhood information via the product of Gaussian likelihoods. Each update is non-recursive and depends on past information only through cumulative local statistics, ensuring analytical tractability and statistical coherence in a fully decentralized setting.

A core contribution of DBME is the coupling between inference and adaptive topology control. A *pruning* stage removes statistically inconsistent links, while a *rewiring* stage restores connectivity through randomized neighbor selection and compatibility tests. Together, these mechanisms let the network self-organize, while adapting the network topology to the evolving statistical relationships among agents.

The theoretical analysis discusses the independence assumptions and the posterior calibration underlying the model. Coverage and Z-score diagnostics are used to assess the agreement between inferred variances and empirical dispersion, providing a measure of uncertainty reliability. Simulations across different regimes illustrate the behavior of DBME in terms of convergence, uncertainty calibration, and cluster formation under varying noise and structural conditions.

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to Dr. Franco Galante and Professor Emilio Leonardi for their availability and invaluable help during the preparation of this document, as well as for the interest they expressed in the research I conducted. My heartfelt thanks go to my family, and in particular my parents, for investing their time and financial resources in my education. I thank my friends for standing by me despite the endless hours of study, which at times made me irritable and anxious. Finally, a last thank you is extended to my brother, a continuous source of inspiration for me, and to my grandmother, whose tenacity and humility will always continue to guide me.

Table of Contents

1 Introduction **1**

1.1 Introduction 1

1.2 Related Work 1

1.2.1 B-ColME overview 2

1.2.2 C-ColME overview 2

2 Background **4**

2.1 Normal distributions - linear combinations 4

2.2 Central Limit Theorem and asymptotic normality 5

2.3 Confidence intervals and concentration 5

2.4 Bayesian updating and conjugate models 6

2.4.1 Normal–Normal model 6

2.4.2 Normal–Inverse–Gamma model 6

2.5 Credibility intervals 7

2.6 Precision weighting and heteroscedastic observations 8

2.6.1 Fusion of Gaussian estimates 8

2.7 Conditional Independence 9

2.8 Sufficient statistics for Gaussian models 9

3 DBME **11**

3.1 Problem Statement 12

3.1.1 How DBME solves this problem 13

3.2 Network Initialization 16

3.2.1 Random network initialization 16

3.3 Bayesian Update 17

3.3.1 Setup after bootstrap 17

3.3.2 Step 1: Local evidence and cumulative update 18

3.3.3 Step 2: Social aggregation of local beliefs 21

3.3.4 Step 3: Bayesian fusion — Gaussian product 24

4 Decentralized pruning: an evidence–driven heuristic **27**

4.0.1 Motivation and rationale 27

4.0.2 Decision statistic and test 28

4.0.3 Automatic adaptivity from evidence 28

4.0.3.1 Complexity, decentralization, and invariances 29
4.0.4 Pseudocode (node–local execution) 30
4.0.5 Error trade–offs and design guidance 30
4.0.6 Limitations and scope 31

5 Rewiring – Decentralized Restoration of Connectivity 32
5.1 Heuristic principle and statistical rationale 32
5.2 Operational procedure 33
5.2.1 Pseudocode (node–local view) 34
5.2.2 Connectivity preservation 34
5.3 Algorithmic properties and limitations 35
5.4 Emergent connectivity and dynamic equilibrium 35
5.5 Computational Complexity of DBME 37

6 Theoretical Justification: Credibility Intervals and Conditional Independence 40
6.1 Local and social posteriors under conditional independence 40
6.2 Experimental configuration and parameter regime 42
6.3 Uncertainty propagation and empirical calibration 43
6.3.1 Transient underestimation of uncertainty 44
6.3.2 Heterogeneity among agents 45
6.3.3 Calibration curves 46
6.3.4 Width-Error analysis 47
6.3.5 Z-score calibration 49
6.4 Adaptive weights and expected coverage degradation 51
6.5 Agent-wise coverage analysis 52

7 Experiments 55
7.1 Variants of DBME 55
7.1.1 Error probability dynamics 56
7.1.2 On the Posterior Variance in Collaborative Bayesian Estimation 58
7.1.3 Wrong Link Fraction 60
7.2 Ablation Study: Effect of Removing Rewiring 60
7.3 Ablation Study: Effect of Social Weights without Topological Adaptation 61
7.4 Non-identical distributions 63
7.5 Comparison with RRR-ColME 65
7.5.1 Restricted Round-Robin ColME 65
7.5.2 Experimental setting 66

8 Conclusions, Limitations and Future Directions 69
8.1 Limitations of the Current Framework 69
8.2 Future Directions 70
8.3 Concluding Remarks 71

A Bayesian Update **72**

A.1 Pairwise divergence statistic 72

A.2 Independence of local and social beliefs (three-node toy case) 72

A.3 Correlation between fused posterior estimates 73

A.4 Asymptotic consistency of DBME 73

Bibliography **78**

List of Figures

3.1 Schematic representation of DBME inference flow. 15

5.1 Compact cyclic representation of pruning and rewiring in DBME. . . 36

6.1 Information flow for agent i 41
6.2 Calibration curves: empirical versus nominal coverage. The near-diagonal alignment demonstrates that posterior uncertainties τ_i^{-1} are well calibrated under dynamic topology updates (pruning/rewiring every 150 steps, $k_{ij} = 1$). 47
6.3 Posterior standard deviation versus absolute estimation error. The red dashed line denotes the theoretical identity $|\text{error}| = \sigma_{\text{post}}$. 48
6.4 Z-score calibration histograms comparing empirical and ideal $\mathcal{N}(0, 1)$. 50
6.5 Distribution of the average coverage per agent across inference modes. The red dashed line marks the nominal 95% target. All four regimes cluster tightly around the theoretical value, confirming overall calibration at the agent level. 53
6.6 Agent-wise relationship between average coverage and inter-run variability. Each hexagon represents the density of agents with similar (coverage_mean, coverage_std). The monotonic decline highlights that unstable agents (higher variability) systematically exhibit lower coverage. 54

7.1 Averaged statistics over 50 independent runs with confidence bands. The initial 50 iterations corresponding to the bootstrap phase are omitted 57
7.2 Error probability in absence of rewiring 61
7.3 Error probability in absence of pruning and rewiring. 62
7.4 Evolution of global inference metrics for heterogeneous agent distributions (NIG mode, 50 runs). The initial 50 iterations corresponding to the bootstrap phase are omitted for clarity. 64
7.5 Monte Carlo evolution of the error probability $P(|\hat{\mu}_a(t) - \mu_a| > \varepsilon)$ with 95% confidence intervals for DBME and ColME. 67

List of Tables

6.1 Global coverage and calibration summary under the filtered regime ($T = 2000$, $N_{\text{runs}} = 150$). All inference modes exhibit nearly Gaussian residuals with zero bias and unit variance. The Bayesian fusion modes show a mild ($\sim$1%) undercoverage due to temporal correlations in posterior updates, while the `local_only` baseline remains perfectly calibrated. 44
6.2 Static topology with adaptive weights coverage. 51
6.3 Dynamic topology with pruning and rewiring coverage drop 52

7.1 Final metrics across estimation modes (mean $\pm$ 1.96·std/$\sqrt{n_{\text{runs}}}$). . . 59

Chapter 1

Introduction

1.1 Introduction

The proliferation of interconnected devices, from smartphones and wearables to distributed sensors has amplified the demand for decentralized and collaborative learning. In many of these settings, agents continuously acquire local observations that may be too sensitive or voluminous to be transmitted. Performing learning directly at the device is often the only feasible option, but when data streams arrive slowly, isolated estimation can lead to delayed convergence. In such scenarios, collaboration among agents provides a concrete way to accelerate this process. A clear example of this paradigm is federated learning, where multiple devices contribute to the training of a shared model. In this work statistical estimation can be regarded as a simplified instance of federated learning, where instead of training a full model, agents collaborate to estimate basic parameters (such as means and variances) from distributed data. However, when local data distributions differ, naive pooling can introduce unwanted bias and the estimates could be skewed. In addition, sharing only point estimates is often insufficient, as it ignores uncertainty associated with local statistics and may lead to erroneous interactions. A more robust approach requires agents to exchange not just values, but indicators of the distributions that encode both central tendencies and confidence levels, allowing the collaborative mechanism to adapt dynamically to the evolving quality of the communication.

1.2 Related Work

The problem of collaborative mean estimation has recently attracted attention in the context of multi-agent learning , as a first reference we cite the algorithm presented in Collaborative Algorithms for Online Personalized Mean Estimation ,called ColME [1]. It addresses the setting where agents sequentially collect samples from heterogeneous distributions and aim to estimate their local means more efficiently by leveraging collaboration. Particularly, in ColME, when one agent queries another, it receives the current local average computed over all the samples observed so far, together with the corresponding number of observations. Each agent maintains these summaries

for the neighbors it has queried . To decide whether two agents can match the same latent class (or whether their means are close enough to justify their collaboration) ,ColME applies PAC-based confidence intervals and once the classes are identified , the agent's local mean is aggregated with social information leveraging specific weighting schemes.

From an empirical point of view ColME performs excellently, but the main limitations emerge when analyzing its computational complexity. Let $\mathcal{A}$ denote the set of agents. In the fully collaborative formulation, each agent evaluates its compatibility with all others at every iteration, resulting in a space and time complexity of $\mathcal{O}(|\mathcal{A}|^2)$, which rapidly becomes impractical in large-scale settings. The Restricted Round-Robin variant reduces the per-iteration burden by allowing only one agent to perform the compatibility evaluation at each time step, but this evaluation is still carried out with respect to the entire population, preserving a quadratic complexity over a full communication cycle of $|\mathcal{A}|$ iterations. To mitigate these issues, recent scalable algorithms have been designed to reduce the per-agent complexity to a fixed quantity $O(r)$, while preserving accuracy and stability of the estimate [2]. Specifically , B-ColME and C-ColME provide optimized, linearly scaling variants of the original collaborative approach. Both B-ColME and C-ColME build upon the ColME framework but rely explicitly on a network structure represented as a graph $G = (\mathcal{A}, E)$, rather than assuming full connectivity among all agents. Below is a quick overview of the mechanism implemented by the two variants:

1.2.1 B-ColME overview

B-ColME extends ColME by allowing information to propagate along the edges of G up to a fixed hop distance d . This mechanism implies that, after a transient phase in which similarity classes are identified, each agent can indirectly access data from more distant parts of the network without evaluating every other agent. Aggregation is performed by averaging all received messages within distance d, leading to a fast diffusion of information. In this sense the distinctive feature of B-ColME is its ability to reduce the overall complexity from $O(|\mathcal{A}|^2)$ to $O(r\,|\mathcal{A}|\log|\mathcal{A}|)$, while still achieving convergence guarantees on large random graphs.

1.2.2 C-ColME overview

On the other hand , C-ColME follows a different approach, inspired by consensus protocols , where Each agent maintains two quantities: (i) the empirical mean of its local samples, and (ii) a consensus estimate $\hat{\mu}_a^t$. At each iteration, the agent updates its estimate as a convex combination:

$$\hat{\mu}_a^{t+1} = (1-\alpha_t)\,\bar{x}_a^{t+1} \;+\; \alpha_t \sum_{a' \in C_a} W_{a,a'}\,\hat{\mu}_{a'}^t,$$

where $\alpha_t \in (0, 1)$ is a memory parameter and W is a doubly-stochastic weight matrix restricted to the local neighborhood C_a. In practice, this means that each node blends its current local evidence with the consensus state of its neighbors, producing smoother dynamics than direct averaging. It is intuitive how the peculiarity of C-ColME is its stability: the convex combination mitigates oscillations during the transient phase, ensures robustness once similarity classes have been detected, and scales with complexity $O(r\,|\mathcal{A}|)$.

Chapter 2

Background

In this work, we take inspiration from scalable ColME architectures, more specifically from C-ColME because we try to emulate a consensus mechanism, but replace PAC-style confidence intervals with a Bayesian treatment of uncertainty. In order to clarify our approach and justify the methodology adopted for our algorithm in this chapter we summarize the main probabilistic tools and results used throughout the work, based on standard references such as Casella and Berger [3], Vershynin [4], and Gelman [5].

2.1 Normal distributions - linear combinations

A random variable X is said to follow a normal distribution with mean μ and precision $\tau = 1/\sigma^2$, written $X \sim \mathcal{N}(\mu, \tau^{-1})$, if its probability density function is

$$f(x) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau(x-\mu)^2}{2}\right).$$

The normal distribution is closed under linear combinations of independent random variables, meaning that if $X \sim \mathcal{N}(\mu_X, \tau_X^{-1})$ and $Y \sim \mathcal{N}(\mu_Y, \tau_Y^{-1})$ are independent, then for any constants $a, b \in \mathbb{R}$

$$aX + bY \ \sim \ \mathcal{N}\left(a\mu_X + b\mu_Y, \ \frac{1}{a^2\tau_X + b^2\tau_Y}\right).$$

In particular, the difference between two independent Gaussian estimators $\hat{\mu}_i \sim \mathcal{N}(\mu^\star, \tau_i^{-1})$ and $\hat{\mu}_j \sim \mathcal{N}(\mu^\star, \tau_j^{-1})$ is itself Gaussian:

$$\hat{\mu}_i - \hat{\mu}_j \ \sim \ \mathcal{N}\Big(0, \ (\tau_i + \tau_j)^{-1}\Big),$$

a result that will later justify divergence tests among agents.

2.2 Central Limit Theorem and asymptotic normality

Let $\{X_1, \ldots, X_n\}$ be independent and identically distributed random variables with mean μ and precision $\tau = 1/\sigma^2$, the Central Limit Theorem (CLT) states that:

$$Z_n = \sqrt{n\tau}\,\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\Big)$$

which is the standardized sample mean, converges in distribution to a standard normal variable as $n \to \infty$:

$$Z_n \xrightarrow{d} \mathcal{N}(0,1).$$

Equivalently, for sufficiently large n the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is approximately Gaussian:

$$\bar{X}_n \;\approx\; \mathcal{N}\Big(\mu,\; \frac{1}{n\tau}\Big), \qquad \text{with precision } \tau_n = n\tau.$$

Consequentially, the precision of the sample mean grows linearly with n, indicating that uncertainty decreases proportionally to $1/n$.

2.3 Confidence intervals and concentration

The precision τ_i is constructed as the inverse of the variance associated with an estimator $\hat{\mu}_i$. Under Gaussian assumptions, the probability that $\hat{\mu}_i$ deviates from the true mean $\mu^\star$ by more than $z_{1-\alpha/2}\,\tau_i^{-1/2}$ is bounded by α:

$$\mathbb{P}\Big(|\hat{\mu}_i - \mu^\star| \le z_{1-\alpha/2}\,\tau_i^{-1/2}\Big) \ge 1-\alpha,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. These confidence intervals contract as $\tau_i \to \infty$, and when the number of samples increases.

Beyond exact normality, similar results hold for sub-Gaussian random variables. A centered variable X is called *sub-Gaussian* with proxy precision $\tau = 1/\sigma^2$ if

$$\mathbb{E}\Big[e^{\lambda X}\Big] \le \exp\left(\frac{\lambda^2}{2\tau^{-1}}\right), \quad \forall\,\lambda \in \mathbb{R}.$$

This condition guarantees Gaussian-like tail decay:

$$\mathbb{P}(|X| \ge t) \le 2\exp\left(-\frac{\tau t^2}{2}\right),$$

ensuring exponential concentration of X around its mean.

2.4 Bayesian updating and conjugate models

Bayesian inference treats unknown parameters as random variables informed by a prior distribution $p(\theta)$, which represents an a priori hypothesis on the nature of the data. Given data $X = (X_1, \dots, X_n)$ generated according to a likelihood $p(X \mid \theta)$, the posterior is obtained by Bayes' rule:

$$p(\theta \mid X) = \frac{p(X \mid \theta)p(\theta)}{p(X)}, \quad p(X) = \int p(X \mid \theta)p(\theta)\, d\theta.$$

In most practical settings, the marginal likelihood $p(X)$ has no closed–form expression, and the evaluation of this integral requires numerical approximation or sampling–based techniques. However, when both the likelihood and the prior belong to a conjugate family, as in the Gaussian case, the posterior retains the same functional form as the prior, producing an analytically tractable function. Below we present the two models of interest.

2.4.1 Normal–Normal model

When the parameter of interest μ represents the mean of a Gaussian with known precision $\tau = 1/\sigma^2$, and the prior on μ is Gaussian $\mathcal{N}(\mu_0, \tau_0^{-1})$, the posterior remains Gaussian:

$$\mu \mid X_{1:n} \sim \mathcal{N}(\mu_n, \tau_n^{-1}),$$

with updated parameters

$$\tau_n = \tau_0 + n\tau, \qquad \mu_n = \frac{\tau_0\mu_0 + n\tau\,\bar{X}}{\tau_n}.$$

The posterior mean μ_n is a precision-weighted average of prior and data, and the posterior precision τ_n increases linearly with n.

$$\tau_{\text{post}} = \tau_{\text{prior}} + \tau_{\text{data}}$$

2.4.2 Normal–Inverse–Gamma model

When both μ and σ^2 are unknown, conjugacy is obtained with a Normal–Inverse–Gamma (NIG) prior:

$$\sigma^2 \sim \text{Inv}-\text{Gamma}(\alpha_0, \beta_0), \qquad \mu \mid \sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right).$$

Equivalently, defining the precision $\tau = 1/\sigma^2$, one can write

$$\tau \sim \text{Gamma}(\alpha_0, \beta_0^{-1}), \qquad \mu \mid \tau \sim \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0\tau}\right).$$

The posterior remains in the same family with parameters

$$\kappa_n = \kappa_0 + n, \quad \mu_n = \frac{\kappa_0 \mu_0 + n\bar{X}}{\kappa_0 + n}, \quad \alpha_n = \alpha_0 + \frac{n}{2},$$

$$\beta_n = \beta_0 + \frac{1}{2}\mathrm{SSD} + \frac{\kappa_0 n}{2(\kappa_0 + n)}(\bar{X} - \mu_0)^2,$$

where $\mathrm{SSD} = \sum_{i=1}^{n}(X_i - \bar{X})^2$. The conditional precision of μ given τ is

$$\tau_\mu = \kappa_n \tau.$$

Although the joint posterior over (μ, τ) remains in the same Normal–Gamma family, the marginal distribution of μ is a Student-t.

By integrating out τ from the joint posterior,

$$p(\mu \mid X) = \int p(\mu, \tau \mid X)\, d\tau,$$

one obtains a Student-t distribution with $2\alpha_n$ degrees of freedom, mean μ_n, and scale parameter

$$s_\mu^2 = \frac{\beta_n}{\alpha_n \kappa_n}, \qquad \mu \mid X \sim t_{2\alpha_n}(\mu_n, s_\mu^2).$$

For a sufficiently large number of observations, the Student-t distribution can be accurately approximated by a Normal law, since the degrees of freedom $2\alpha_n$ grow linearly with n, the heavy tails of the Student distribution vanish asymptotically, and the posterior concentrates around μ_n. By matching the first two moments, the marginal posterior can be approximated as

$$\mu \mid X \;\approx\; \mathcal{N}\!\left(\mu_n,\, \frac{\beta_n}{\alpha_n \kappa_n}\right),$$

where the corresponding precision is

$$\tau_\mu = \frac{\alpha_n \kappa_n}{\beta_n}.$$

This moment–matching approximation becomes increasingly accurate as the sample size grows, effectively recovering the Normal–Normal model as a limiting case of the conjugate Normal–Inverse–Gamma formulation.

2.5 Credibility intervals

In the Bayesian framework, uncertainty is expressed through the posterior distribution. Given a posterior

$$p(\theta \mid \mathcal{D}) = \mathcal{N}(\hat{\theta},\, \tau^{-1}),$$

a credibility interval is defined as the region containing a fixed portion (e.g. 95%) of the posterior probability mass:

$$\hat{\theta} \pm 1.96\,\tau^{-1/2}.$$

Unlike confidence intervals, credibility intervals provide a direct probabilistic statement about the parameter itself:

$$\mathbb{P}\big(\theta \in [\hat{\theta} \pm 1.96\,\tau^{-1/2}] \mid \mathcal{D}\big) = 0.95.$$

As a consequence as the amount of data increases or the likelihood becomes more informative, the posterior precision τ increases, and the interval becomes narrower, reflecting higher belief confidence.

2.6 Precision weighting and heteroscedastic observations

When independent measurements Y_j of a common quantity μ have unequal precisions τ_j, the best linear unbiased estimator (BLUE) of μ is the precision-weighted mean:

$$\hat{\mu}_{\mathrm{GLS}} = \frac{\sum_j \tau_j\, Y_j}{\sum_j \tau_j}, \qquad \mathrm{Prec}(\hat{\mu}_{\mathrm{GLS}}) = \sum_j \tau_j.$$

Equivalently, the variance of this estimator is

$$\mathrm{Var}(\hat{\mu}_{\mathrm{GLS}}) = \Big(\sum_j \tau_j\Big)^{-1}.$$

proportionally to τ_j the aggregated estimate assign greater importance to more reliable local estimates, ensuring that the resulting estimator achieves minimal posterior variance or equivalently, maximal precision, among all unbiased linear estimators.

2.6.1 Fusion of Gaussian estimates

A gaussian fusion is the combination of multiple Gaussian beliefs referring to the same latent variable $\mu^\star$. Suppose two independent agents hold local posteriors of the form

$$\mu^\star \sim \mathcal{N}(\hat{\mu}_1,\, \tau_1^{-1}), \qquad \mu^\star \sim \mathcal{N}(\hat{\mu}_2,\, \tau_2^{-1}),$$

each representing an independent Gaussian belief centered on its local estimate $\hat{\mu}_i$ with precision τ_i. The fused belief about $\mu^\star$ is obtained by multiplying the two densities and renormalizing:

$$p(\mu^\star \mid \hat{\mu}_1, \hat{\mu}_2) \propto \exp\Big(-\tfrac{1}{2}\tau_1(\mu^\star - \hat{\mu}_1)^2 - \tfrac{1}{2}\tau_2(\mu^\star - \hat{\mu}_2)^2\Big)\,.$$

The result is again Gaussian,

$$p(\mu^\star \mid \hat{\mu}_1, \hat{\mu}_2) = \mathcal{N}\Big(\hat{\mu}_f,\, \tau_f^{-1}\Big)\,,$$

with fused parameters

$$\tau_f = \tau_1 + \tau_2, \qquad \hat{\mu}_f = \frac{\tau_1 \hat{\mu}_1 + \tau_2 \hat{\mu}_2}{\tau_1 + \tau_2}.$$

overall it is always a weighted estimation for precision.

This expression generalizes naturally to K independent agents, each providing a Gaussian belief $\mu^\star \sim \mathcal{N}(\hat{\mu}_k, \tau_k^{-1})$:

$$\tau_f = \sum_{k=1}^{K} \tau_k, \qquad \hat{\mu}_f = \frac{\sum_{k=1}^{K} \tau_k \, \hat{\mu}_k}{\sum_{k=1}^{K} \tau_k}.$$

The fusion process can therefore be interpreted as a precision-weighted averaging of the means, where each agent contributes in proportion to its confidence τ_k.

2.7 Conditional Independence

Two random variables X and Y are said to be *conditionally independent given* Z, denoted as

$$X \perp\!\!\!\perp Y \mid Z,$$

if, once Z is known, knowledge of Y provides no additional information about X.

Formally,

$$p(X, Y \mid Z) = p(X \mid Z)\, p(Y \mid Z).$$

Equivalently, this implies that

$$p(X \mid Y, Z) = p(X \mid Z) \quad \text{and} \quad p(Y \mid X, Z) = p(Y \mid Z),$$

that is, conditioning on Z renders X and Y statistically independent.

Intuitively, once the conditioning variable Z is known, any dependence between X and Y disappears. Conditional independence is the key assumption that enables probabilistic models to separate sources of information and to perform modular inference: each component can update its beliefs based only on the relevant subset of variables, without redundancy or double counting of evidence.

2.8 Sufficient statistics for Gaussian models

For samples $x_1, \ldots, x_n$ drawn from a normal distribution, all information relevant to estimating the parameters (μ, σ^2) is contained in the sufficient statistics

$$S = \sum_{t=1}^{n} x_t, \qquad N = n, \qquad Q = \sum_{t=1}^{n} x_t^2.$$

From these quantities one recovers the sample mean and variance:

$$m = \frac{S}{N}, \qquad v = \frac{Q - Nm^2}{N - 1}.$$

This set of statistics fully characterizes the Gaussian model, referring to the Bayesian setting, the concept of sufficiency assumes the same formal definition as in classical statistics, but with a slightly different operational meaning. Formally a statistic $T(x_{1:n})$ is said to be sufficient for a parameter θ if

$$p(\theta \mid x_{1:n}) = p(\theta \mid T(x_{1:n})).$$

This means that T contains all the information needed to update the prior $p(\theta)$ into the posterior $p(\theta \mid x_{1:n})$; as an example, in the case of a Gaussian model with Normal–Inverse–Gamma prior (see Section 2.4.2),

$$x_t \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2), \qquad (\mu, \sigma^2) \sim \mathrm{NIG}(\mu_0, \kappa_0, \alpha_0, \beta_0),$$

the posterior depends on the data only through the sufficient statistics from which the updated hyperparameters are obtained as

$$\begin{aligned} \kappa_n &= \kappa_0 + N, \quad \mu_n = \frac{\kappa_0 \mu_0 + S}{\kappa_n}, \\ \alpha_n &= \alpha_0 + \tfrac{N}{2}, \quad \beta_n = \beta_0 + \tfrac{1}{2}(Q - Nm^2) + \frac{\kappa_0 N (\mu_0 - m)^2}{2(\kappa_0 + N)}. \end{aligned}$$

In this way the triplet (S, Q, N) fully characterizes the Bayesian state for the Gaussian model, and serves as the minimal memory representation.

Chapter 3

DBME

The *Dynamic Bayesian Mean Estimation* (DBME) framework is introduced as an alternative to the scalable mean–estimation algorithms of the ColME family. Rather than reproducing their consensus-based mechanics, DBME reinterprets the estimation process through the lens of Bayesian inference, using probabilistic reasoning and heuristic approximations instead of concentration inequalities. The core intuition is straightforward: instead of operating in the space of data, DBME operates in the space of estimates. Each node maintains not only a local mean estimate but also an assessment of its own precision. If an agent can quantify the uncertainty of its local belief, it can in turn improve the precision of its neighbors' beliefs. This mechanism is framed in Bayesian terms: at each iteration, the local information of node i is treated as the marginal likelihood over the latent parameter, whereas the information received from its neighbors acts as an informative prior, formally independent from the local component. In this formulation, the conditional independence between local and social evidence is a structural property of the model. Each component provides a distinct probabilistic view of the same latent parameter, allowing the aggregation of sources of information to remain analytically valid and internally consistent. The validity of these structural assumptions — conditional independence, Gaussianity, and variance aggregation — holds under specific regimes that are explicitly examined throughout this work. In particular, subsequent chapters analyze the core structure of the algorithm , the empirical calibration of posterior variances, the statistical consistency of the fusion step, and the topological conditions under which these approximations remain reliable. Under such regimes, the DBME framework proves to be theoretically coherent yielding a model that is interpretable, scalable, and consistent to the spirit of distributed Bayesian inference.

3.1 Problem Statement

We now introduce the formalization of the distributed estimation problem that motivates the DBME framework. Consider a network of A agents, where each agent $i \in \{1, \dots, A\}$ observes over time a sequence of random variables

$$x_{i,t} \sim F_i, \qquad \mathbb{E}[x_{i,t}] = \mu_i \in \mathbb{R}, \qquad \mathrm{Var}(x_{i,t}) = \sigma_i^2 < \infty,$$

with F_i of unknown form. Unlike the original ColME formulation, which assumes the data–generating process to be σ–sub–Gaussian, our framework does not impose such tail conditions. Instead, relying on the Central Limit Theorem (CLT), the sampling distribution of the empirical mean can be asymptotically approximated by a Gaussian regardless of the specific form of the underlying distribution, provided it admits finite mean and variance.

We now analyze the type of relationship that holds among the data at different levels of aggregation. The sequence $\{x_{i,t}\}_{t\geq 1}$ is ideally i.i.d. or at least stationary with finite variance, meaning that the local data generated by each agent come from the same distribution and are independent of those of other agents. By the CLT, the empirical mean of the observations collected by each agent up to time n,

$$\bar{x}_i(n) = \frac{1}{n} \sum_{t=1}^{n} x_{i,t},$$

admits, already for moderate n, the approximation

$$\bar{x}_i(n) \approx \mathcal{N}(\mu_i,\ \sigma_i^2/n).$$

The *Central Limit Theorem* ensures that, under mild regularity conditions, the standardized sample mean of independent random variables with finite variance converges in distribution to a normal law [6]. This asymptotic normality holds not only for strictly i.i.d. samples but also under more general weak–dependence or non–identically distributed settings, as formalized in the extended forms of the CLT discussed by Feller [7].

In general, the vector $\{x_{i,t}\}_{i=1}^{A}$ is not identically distributed, since each agent may have distinct parameters (μ_i, σ_i^2) and even a different family $F_{i,t}$. If homogeneous classes of agents exist, then within each class the variables can be i.i.d. The interaction over the network should not alter the generation of raw data $x_{i,t}$, which remain conditionally independent across agents. It can, however, introduce dependence among the evolving estimates, since in a collaborative context each $\mu_i(t)$ integrates information from neighboring agents. This interdependence gives rise to a critical issue: if the same information is inadvertently propagated multiple times through the network, agents may perform double counting of evidence. Avoiding such redundant

information flow is therefore essential to preserve the statistical "purity" of the distributed estimation process.

As previously mentioned, our algorithm builds upon the data structure introduced by the scalable variants of ColME [2]. In this setting, the agents interact over a time–varying communication graph G_t, where each node exchanges with its neighborhood $\mathcal{N}_t(i)$ a compact summary of the collected statistical evidence. The objective is to estimate the mean μ_i for every agent i, while accelerating convergence through the information provided by statistically compatible neighbors, minimizing redundant information recirculation, and without requiring any prior knowledge of classes or clusters (unsupervised setting).

3.1.1 How DBME solves this problem

In practical terms, DBME operates as a distributed Bayesian estimator that evolves over time through a sequence of *local* and *social* inference phases. Each agent maintains only the minimal information required to update its own belief about the unknown parameter, without any centralized coordination , except for a global list of agent identifiers, used solely to establish random connections during the rewiring phase (see Chapter 5).

Each agent starts from an individual prior belief over its latent mean:

$$p_i^{(0)}(\mu) = \mathcal{N}(\mu_{i,0},\, \tau_{i,0}^{-1}),$$

where both the prior mean $\mu_{i,0}$ and the precision $\tau_{i,0}$ are initialized independently based on bootstrap statistics from the agent's own data. This design allows each node to begin the inference process from a personalized, data-driven initialization.

Each node i then maintains a compact set of sufficient statistics constructed from its local data stream $\{x_{i,t}\}$. These cumulative quantities , the running sum, squared sum, and sample count $(S_i^{\mathrm{loc}}, Q_i^{\mathrm{loc}}, N_i^{\mathrm{loc}})$, define the empirical basis upon which all subsequent inferences are built. From them, the agent derives the parameters of its *local posterior*, obtained by fusing the individual prior $p_i^{(0)}(\mu)$ with the likelihood induced by its own observations:

$$p_i^{\mathrm{loc}}(\mu) \;\propto\; p_i^{(0)}(\mu)\, p(x_{i,1:n} \mid \mu) \;=\; \mathcal{N}(\mu_i^{\mathrm{loc}}, \tau_{i,\mathrm{loc}}^{-1}),$$

with

$$\tau_{i,\mathrm{loc}} = \tau_{i,0} + \frac{N_i^{\mathrm{loc}}}{\sigma_i^2}, \qquad \mu_i^{\mathrm{loc}} = \frac{\tau_{i,0}\mu_{i,0} + \frac{N_i^{\mathrm{loc}}}{\sigma_i^2}\,\bar{x}_i}{\tau_{i,0} + \frac{N_i^{\mathrm{loc}}}{\sigma_i^2}}.$$

At the same time, each agent collects from its neighbors $j \in \mathcal{N}(i)$ their respective local parameters $(\mu_j^{\mathrm{loc}}, \tau_{j,\mathrm{loc}})$. This minimal exchange of statistics allows information to circulate efficiently through the network without transmitting raw data or previously

fused posteriors.

Optionally, the model can incorporate an adaptive weighting mechanism that quantifies the statistical compatibility between agents. In this case, each link (i, j) is associated with a coefficient $k_{ij} \in [0, 1]$ that modulates the contribution of neighbor j to the belief of node i according to the agreement between their current posteriors.

At every iteration, node i updates its local sufficient statistics with the new observation $x_{i,t}$, thereby refining its local precision $\tau_{i,\mathrm{loc}}$ and mean μ_i^{loc}. Concurrently, it aggregates the local beliefs received from its neighbors into a *social prior* characterized by

$$\tau_{i,\mathrm{soc}} = \sum_{j \in \mathcal{N}(i)} k_{ij}\, \tau_{j,\mathrm{loc}}, \qquad \mu_{i,\mathrm{soc}} = \frac{\sum_{j \in \mathcal{N}(i)} k_{ij}\, \tau_{j,\mathrm{loc}}\, \mu_j^{\mathrm{loc}}}{\sum_{j \in \mathcal{N}(i)} k_{ij}\, \tau_{j,\mathrm{loc}}}.$$

This neighborhood summary is then interpreted as an informative prior and analytically fused with the agent's own local belief through the product of Gaussian densities:

$$\tau_i^{\mathrm{post}} = \tau_{i,\mathrm{loc}} + \tau_{i,\mathrm{soc}}, \qquad \mu_i^{\mathrm{post}} = \frac{\tau_{i,\mathrm{loc}}\, \mu_i^{\mathrm{loc}} + \tau_{i,\mathrm{soc}}\, \mu_{i,\mathrm{soc}}}{\tau_{i,\mathrm{loc}} + \tau_{i,\mathrm{soc}}}.$$

Conceptually, DBME treats these two sources of information as distinct and complementary with the local evidence providing direct and unbiased data about the latent parameter, while the social evidence acting as an informative prior shaped by the neighborhood context, inferred exclusively from the independent local beliefs of adjacent nodes. In this way, each agent communicates only the parameters $(\mu_i^{\mathrm{loc}}, \tau_{i,\mathrm{loc}})$ derived from its personal data stream.

As a result, all social updates are constructed without access to remote fused posteriors or recursively aggregated beliefs. They rely solely on locally updated summaries, ensuring that local and social information sources remain statistically disjoint and conditionally independent given the latent parameter and the network structure. Unlike standard linear Bayesian filters such as the Kalman filter [8], DBME does not require an explicit temporal transition in which the posterior at time t becomes the prior for time $t{+}1$. Each agent's posterior distribution depends only on new data and the current local statistics, which are cumulatively updated over time. This mechanism implicitly embeds temporal dependence within the sufficient statistics $(S_i^{\mathrm{loc}}, Q_i^{\mathrm{loc}}, N_i^{\mathrm{loc}})$, ensuring that each Bayesian update naturally reflects the information accumulated so far.

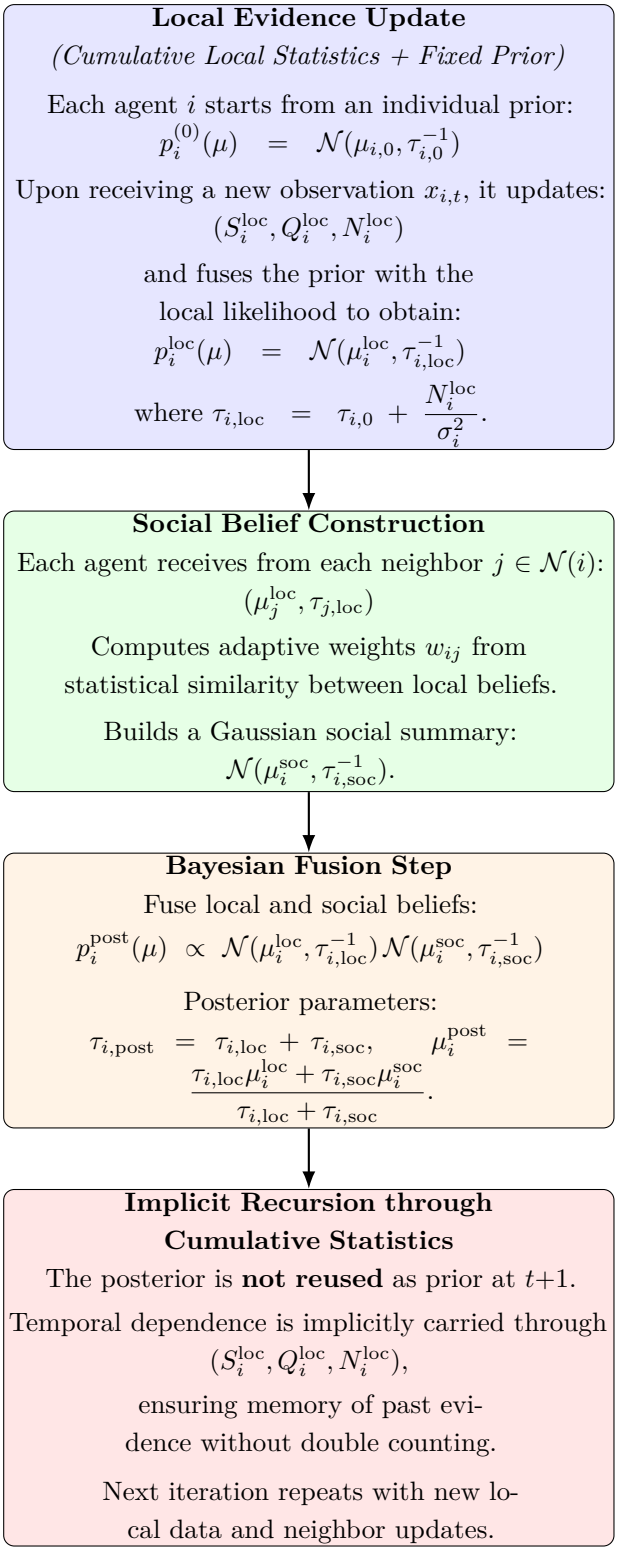


Figure 3.1: Schematic representation of DBME inference flow.

3.2 Network Initialization

The initialization of the network determines the transient regime in which the agents begin to exchange information: on it depend the initial connectivity, robustness to spurious links, and the speed at which latent clusters emerge. In the current DBME formulation, the initialization is intentionally *neutral*: it does not exploit any information about the data, but generates a random graph whose connectivity is controlled solely by the desired average degree r. This approach provides a reproducible and unbiased baseline, avoiding any a priori structural assumptions while ensuring that information diffusion can start effectively from the very first iterations.

3.2.1 Random network initialization

The network is initialized according to the classical model proposed by Erdős and Rényi, in which a graph $G(N, p)$ is obtained by independently adding an edge between each pair of nodes with probability p [9]. Given a target average degree r, one sets

$$p \;=\; \frac{r}{N-1}.$$

For each unordered pair $\{i, j\}$, with $i \neq j$, the edge (i, j) is drawn with probability p, independently of all others. The degree K_i of each node follows a binomial distribution as shown in the monograph by Bollobás [10]:

$$K_i \sim \mathrm{Bin}\big(N-1,\, p\big), \qquad \mathbb{E}[K_i] = r, \quad \mathrm{Var}(K_i) = (N-1)p(1-p),$$

which, for large N and small p, is well approximated by a Poisson distribution with parameter r [7]. The expected fraction of isolated nodes is therefore approximately e^{-r}, and the expected number of edges in the graph is

$$\mathbb{E}[\,|E|\,] \;=\; \binom{N}{2} p \;=\; \frac{N\,r}{2}.$$

From a connectivity perspective, the emergence of the giant component occurs around $r \simeq 1$, while full connectivity becomes likely when r exceeds approximately $\log N$. These critical thresholds correspond to the classical phase–transition phenomena identified by Erdős and Rényi [11].

This neutral initialization strategy defines the baseline structure upon which the subsequent pruning and rewiring mechanisms act. By adjusting r, one can directly control the initial level of connectivity and thus the density of information exchange during the early transient. In all experiments presented in this work, the condition $r > \log N$ was enforced to guarantee a highly connected starting configuration, while leaving the topology free to reorganize adaptively through the subsequent steps.

3.3 Bayesian Update

This section provides a formal explanation of the inference mechanism illustrated in Figure 3.1, detailing how each agent in the Dynamic Bayesian Mean Estimation (DBME) framework performs its local and social Bayesian updates. The figure schematically represents the recursive process by which agents collect local evidence, synthesize neighborhood information, and fuse these two sources through the Bayesian product of Gaussian distributions. We now introduce the *bootstrap phase*, which defines the initial conditions of the DBME algorithm. This phase establishes the first local beliefs for each agent, providing the statistical baseline upon which the subsequent iterative inference process will build.

3.3.1 Setup after bootstrap

To properly initialize the algorithm, each node performs a short empirical bootstrap phase that provides a local statistical foundation for estimating its mean and, when necessary, its variance. During this initialization stage, each agent constructs a *data–driven prior* directly from its own observations. The number of bootstrap samples T_0 determines an information trade–off: too few samples yield unstable and slow convergence, whereas too many make the need for collaboration negligible.

After this preliminary phase of length T_0, each agent i forms an *empirical prior* on its latent mean μ_i using the sample mean and variance estimated from its first T_0 local observations:

$$p_i^{(0)}(\mu) = \mathcal{N}(\mu_{i,0},\, \tau_{i,0}^{-1}), \qquad \mu_{i,0} = \frac{1}{T_0}\sum_{t=1}^{T_0} x_{i,t}, \quad \tau_{i,0} = \frac{T_0}{s_{i,0}^2},$$

where the local empirical variance is

$$s_{i,0}^2 = \frac{1}{T_0 - 1}\sum_{t=1}^{T_0}\big(x_{i,t} - \mu_{i,0}\big)^2.$$

This construction provides each agent with a *Gaussian prior* centered on its empirical evidence, with precision $\tau_{i,0}$ increasing linearly with the number of bootstrap samples T_0. Consequently, the Bayesian process begins from statistically grounded and locally informed beliefs.

This *empirical prior* is subsequently employed in the updating of local beliefs, as detailed in the following section. In the present formulation, it encapsulates information extracted directly from the agent's own bootstrap data; however, in a real application, it may also represent an *exogenous belief* or any other form of prior domain knowledge. Importantly, this prior remains fixed at time T_0, and its information is deliberately excluded from the cumulative sufficient statistics used in subsequent updates.

3.3.2 Step 1: Local evidence and cumulative update

Once the bootstrap phase is completed, each agent continues to refine its knowledge through the incremental accumulation of new local data. At each iteration $t > T_0$, node i receives a fresh observation $x_{i,t}$ from its environment and updates its cumulative sufficient statistics:

$$S_i^{\text{loc}}(t) = S_i^{\text{loc}}(t-1)+x_{i,t}, \qquad Q_i^{\text{loc}}(t) = Q_i^{\text{loc}}(t-1)+x_{i,t}^2, \qquad N_i^{\text{loc}}(t) = N_i^{\text{loc}}(t-1)+1.$$

These quantities represent, respectively, the running sum of samples, the sum of squared values, and the total number of observations collected up to time t.

Unlike in the bootstrap phase, however, the Bayesian update is now performed only on the incremental data block acquired after T_0, while the prior obtained during the bootstrap remains fixed and external to these statistics as previously mentioned.

Letting

$$n_i^{\Delta}(t) = N_i^{\text{loc}}(t)-T_0, \qquad S_i^{\Delta}(t) = S_i^{\text{loc}}(t)-S_i^{\text{loc}}(T_0), \qquad Q_i^{\Delta}(t) = Q_i^{\text{loc}}(t)-Q_i^{\text{loc}}(T_0),$$

the empirical mean and variance of the incremental observations are

$$m_i^{\Delta}(t) = \frac{S_i^{\Delta}(t)}{n_i^{\Delta}(t)}, \qquad s_i^{2,\Delta}(t) = \frac{Q_i^{\Delta}(t) - \frac{1}{n_i^{\Delta}(t)}\left(S_i^{\Delta}(t)\right)^2}{n_i^{\Delta}(t) - 1}.$$

These statistics summarize only the data acquired after the prior was fixed and constitute the informative component of the local Bayesian update.

By denoting the local precision contribution from the new data as

$$\tau_{i,\text{data}}(t) = \frac{n_i^{\Delta}(t)}{s_i^{2,\Delta}(t)},$$

the resulting local posterior after incorporating the prior $p_i^{(0)}(\mu) = \mathcal{N}(\mu_{i,0}, \tau_{i,0}^{-1})$ takes the form

$$p_i^{\text{loc}}(\mu) = \mathcal{N}(\mu_i^{\text{loc}}(t),\, \tau_{i,\text{loc}}(t)^{-1}),$$

with

$$\tau_{i,\text{loc}}(t) = \tau_{i,0} + \tau_{i,\text{data}}(t), \qquad \mu_i^{\text{loc}}(t) = \frac{\tau_{i,0}\mu_{i,0} + \tau_{i,\text{data}}(t)\, m_i^{\Delta}(t)}{\tau_{i,0} + \tau_{i,\text{data}}(t)}.$$

This update follows the standard precision–weighted Bayesian fusion rule, introduced in Section 2.6.1, where the prior and the newly observed data contribute additively to the total precision. Depending on how the observation noise is modeled, DBME supports three local inference modes, described below:

Known noise variance.

If the local observation noise σ^2 is assumed known and constant across agents, the update of node i follows the standard Normal–Normal conjugate form, introduced in Section 2.6.1, with prior parameters $(\mu_{i,0}, \tau_{i,0})$ fixed at T_0:

$$\tau_{i,\mathrm{loc}}(t) = \tau_{i,0} + \frac{n_i^{\Delta}(t)}{\sigma^2}, \qquad \mu_i^{\mathrm{loc}}(t) = \frac{\tau_{i,0}\mu_{i,0} + \frac{n_i^{\Delta}(t)}{\sigma^2}\, m_i^{\Delta}(t)}{\tau_{i,0} + \frac{n_i^{\Delta}(t)}{\sigma^2}}.$$

As the number of new samples $n_i^{\Delta}(t)$ increases, the influence of the prior naturally diminishes, and the posterior mean converges to the empirical average of the newly observed data. This configuration represents the most optimistic regime, in which each sample contributes a constant amount of precision.

Unknown noise variance (plug–in).

When the local observation noise is unknown and may differ across agents, each node estimates it empirically from its own incremental data and substitutes this value into the same Gaussian update:

$$\hat{\sigma}_i^{2,\Delta}(t) = \frac{Q_i^{\Delta}(t) - \frac{1}{n_i^{\Delta}(t)}\big(S_i^{\Delta}(t)\big)^2}{n_i^{\Delta}(t) - 1}, \qquad \hat{\tau}_{i,\mathrm{data}}(t) = \frac{n_i^{\Delta}(t)}{\hat{\sigma}_i^{2,\Delta}(t)}.$$

The resulting local posterior is then given by

$$\tau_{i,\mathrm{loc}}(t) = \tau_{i,0} + \hat{\tau}_{i,\mathrm{data}}(t), \qquad \mu_i^{\mathrm{loc}}(t) = \frac{\tau_{i,0}\mu_{i,0} + \hat{\tau}_{i,\mathrm{data}}(t)\, m_i^{\Delta}(t)}{\tau_{i,0} + \hat{\tau}_{i,\mathrm{data}}(t)}.$$

This "plug–in" variant corresponds to an empirical Bayes approximation: the local variance is inferred directly from the agent's own incremental data and treated as fixed within each update. Although it slightly underestimates uncertainty by neglecting the sampling variability of $\hat{\sigma}_i^{2,\Delta}(t)$, it remains a simple and robust mechanism that automatically adapts each agent's confidence to the stability of its recent observations.

Normal–Inverse–Gamma (full conjugacy)

The most coherent with the problem setting local inference mode adopts a joint Normal–Inverse–Gamma prior over (μ_i, σ_i^2), as introduced in Section 2.4.2:

$$(\mu_i, \sigma_i^2) \sim \mathcal{N}\text{–}\mathcal{IG}(\mu_{i,0},\, \kappa_{i,0},\, \alpha_{i,0},\, \beta_{i,0}),$$

with hyperparameters $(\mu_{i,0}, \kappa_{i,0}, \alpha_{i,0}, \beta_{i,0})$ fixed at the bootstrap time T_0. After each incremental batch of local data, these parameters are updated analytically following the conjugate rules of the Normal–Inverse–Gamma model:

$$\kappa_i(t) = \kappa_{i,0} + n_i^{\Delta}(t), \qquad \mu_i^{\mathrm{loc}}(t) = \frac{\kappa_{i,0}\mu_{i,0} + n_i^{\Delta}(t) m_i^{\Delta}(t)}{\kappa_i(t)},$$

$$\alpha_i(t) = \alpha_{i,0} + \tfrac{1}{2} n_i^{\Delta}(t), \qquad \beta_i(t) = \beta_{i,0} + \tfrac{1}{2} n_i^{\Delta}(t) s_i^{2,\Delta}(t) + \frac{\kappa_{i,0} n_i^{\Delta}(t)}{2\kappa_i(t)} \big(m_i^{\Delta}(t) - \mu_{i,0}\big)^2.$$

The marginal posterior of μ_i follows a Student–t distribution [see 5, 6] with mean $\mu_i^{\mathrm{loc}}(t)$ and variance

$$\mathrm{Var}[\mu_i] = \frac{\beta_i(t)}{(\alpha_i(t) - 1)\,\kappa_i(t)}, \qquad (\alpha_i(t) > 1).$$

Equivalently, the corresponding precision can be written as

$$\tau_{i,\mathrm{loc}}(t) = \frac{(\alpha_i(t) - 1)\,\kappa_i(t)}{\beta_i(t)}.$$

For analytical consistency with the Gaussian fusion rule (Section 2.6.1), this Student–t distribution is projected onto the Gaussian family via moment matching:

$$p_i^{\mathrm{loc}}(\mu) \approx \mathcal{N}\big(\mu_i^{\mathrm{loc}}(t),\, \tau_{i,\mathrm{loc}}^{-1}(t)\big), \qquad \mu_i^{\mathrm{loc}}(t) = \mathbb{E}[\mu_i], \quad \tau_{i,\mathrm{loc}}^{-1}(t) = \mathrm{Var}[\mu_i].$$

This projection preserves the first two moments of the exact Student–t posterior, ensuring compatibility with the Gaussian fusion step while retaining the correct belief uncertainty. The moment–matching approximation becomes increasingly accurate as the number of samples grows as already stated in 2. For distributions within the exponential family, minimizing the reverse Kullback–Leibler divergence $D_{\mathrm{KL}}(q\|p)$ is equivalent to matching the sufficient statistics of the target distribution [12, 13]. In the Gaussian case, this reduces precisely to matching the mean and variance of the exact Student–t posterior, yielding a consistent Normal approximation that retains the correct first and second moments of the true belief.

Overall, these three modes correspond to increasing levels of Bayesian expressivity. The *known–variance* and *plug–in* settings remain close to their frequentist counterparts, while introducing an explicit prior fixed at T_0. The *Normal–Inverse–Gamma* configuration achieves a formally richer Bayesian description by jointly inferring both the mean and the variance from the incremental data, treating σ_i^2 as a random variable. However, this increased expressivity remains effectively confined to the estimation of the mean. Although the variance is locally inferred within the NIG model, it is not structurally exploited in the interaction between agents: only the marginal belief over μ_i is communicated and fused across the network. As such, even in the NIG case, the collaborative inference process operates over the one-dimensional parameter space of the mean, and the additional uncertainty captured on σ_i^2 influences the dynamics only indirectly through the local precision.
In all cases, the resulting local belief distribution

$$p_i^{\mathrm{loc}}(\mu) = \mathcal{N}\big(\mu_i^{\mathrm{loc}}(t), \tau_{i,\mathrm{loc}}^{-1}(t)\big)$$

represents the agent's private uncertainty about its latent mean.

From a theoretical standpoint, all these local Bayesian updates exhibit asymptotic

consistency with their frequentist analogues. Following standard results on the asymptotic behavior of conjugate Bayesian models and the Bernstein–von Mises theorem [14], the posterior distribution of the mean parameter asymptotically concentrates around the corresponding frequentist estimator. Under regularity conditions and as the number of incremental local observations $n_i^{\Delta}(t)$ increases, the posterior of μ_i becomes approximately Gaussian and centered at the maximum– likelihood estimator, with variance shrinking at rate $O(1/n_i^{\Delta}(t))$.

$$\mu_i^{\mathrm{loc}}(t) \xrightarrow{p} m_i^{\Delta}(t), \qquad \tau_{i,\mathrm{loc}}^{-1}(t) \xrightarrow{p} \frac{\sigma_i^2}{n_i^{\Delta}(t)},$$

so that

$$p_i^{\mathrm{loc}}(\mu) \Rightarrow \mathcal{N}(m_i^{\Delta}(t), \sigma_i^2/n_i^{\Delta}(t)).$$

In the large–sample limit, therefore, all three local inference modes collapse to the frequentist estimator of the mean distribution, and their posteriors become indistinguishable from the sampling distribution of the empirical average.

3.3.3 Step 2: Social aggregation of local beliefs

During the social phase, each agent refines its estimate by aggregating the current *local beliefs* of its neighbors, as illustrated in Figure 3.1. At every iteration, each node constructs a social estimate derived from the most recent beliefs of its neighbors. This mechanism produces a dynamic and fully decentralized process in which social influence reflects the instantaneous state of the network.

Let $\mathcal{N}_t(i)$ denote the set of neighbors of agent i at time t. Each neighbor $j \in \mathcal{N}_t(i)$ communicates its local Gaussian belief:

$$p_j^{\mathrm{loc}}(\mu) = \mathcal{N}(\mu_j^{\mathrm{loc}}(t), \tau_{j,\mathrm{loc}}^{-1}(t)),$$

where $\mu_j^{\mathrm{loc}}(t)$ and $\tau_{j,\mathrm{loc}}(t)$ denote the local mean and precision maintained by agent j. The precision $\tau_{j,\mathrm{loc}}$ quantifies the confidence of node j in its estimate: higher values correspond to sharper, more reliable beliefs.

Agent i interprets these neighboring beliefs as indirect, noisy sources of information about its own latent variable μ_i. This perspective treats inter–agent connections not as direct observations, but as probabilistic channels through which compatible beliefs reinforce each other, while inconsistent ones exert diminishing influence.

In the following, we describe how this adaptive modulation can be implemented through a pairwise coherence mechanism based on the relative precisions and means of neighboring beliefs, and we also present a simplified alternative that assumes uniform interaction strengths across all neighbors.

Pairwise Social Coherence Heuristic

To adaptively regulate the mutual influence among agents, DBME introduces the *Pairwise Social Coherence Heuristic*, which quantifies the statistical compatibility between two connected nodes in terms of their local posterior means and precisions.

The pairwise dissimilarity between agents i and j is defined as

$$\ell_{ij}(t) = \frac{\left(\mu_i^{\mathrm{loc}}(t) - \mu_j^{\mathrm{loc}}(t)\right)^2}{2\left[\tau_{i,\mathrm{loc}}^{-1}(t) + \tau_{j,\mathrm{loc}}^{-1}(t)\right]}.$$

This quantity grows with both the squared difference between local means and the precisions of the two beliefs: when agents are highly confident (large τ_i, τ_j), even small deviations lead to large dissimilarity values. Conversely, when both beliefs are uncertain (small τ_i, τ_j), the denominator dominates and $\ell_{ij}(t)$ remains small, indicating weak evidence of disagreement.

The dissimilarity $\ell_{ij}(t)$ is symmetric, non–negative, and vanishes only when the two local posteriors are in perfect agreement, $\mu_i^{\mathrm{loc}}(t) = \mu_j^{\mathrm{loc}}(t)$. Its frequentist interpretation, detailed in Appendix A.1, shows that $\ell_{ij}(t)$ corresponds to a continuous likelihood–ratio test on the equality of local means under independent Gaussian posteriors.

Each edge (i, j) is then assigned an adaptive interaction weight that translates this dissimilarity into a continuous modulation of social influence:

$$k_{ij}(t) = k_{\max}\,\exp\!\big(-\ell_{ij}(t)\big), \tag{3.1}$$

where $k_{\max} \in [0, 1]$ denotes the maximal admissible interaction strength. The exponential decay ensures that statistically coherent pairs (low ℓ_{ij}, high agreement) interact more strongly, while divergent pairs (high ℓ_{ij}) exert progressively weaker influence. At this stage, the weighting mechanism effectively acts as a *smoothing operator* over the information produced by the neighbors, attenuating the impact of inconsistent or noisy beliefs while preserving the contribution of statistically compatible ones.

Since $\ell_{ij}(t)$ depends explicitly on both the means and precisions of the local posteriors, the resulting weights $k_{ij}(t)$ adapt in real time to the evolving statistical coherence between agents. This mechanism allows the network to self–adjust its effective topology, reinforcing connections between consistent beliefs while attenuating those between statistically incompatible nodes. As will be discussed in detail in Chapters 6 and 7, this adaptive weighting scheme has been specifically investigated in the experimental phase to enhance the algorithm's performance under static or semi–static network conditions, where the topology does not evolve dynamically but edges can still be selectively pruned based on statistical incompatibility.

Social belief aggregation

Given the pairwise weights $k_{ij}(t)$, each agent i forms a *social belief* by aggregating the local beliefs of its neighbors through precision–weighted averaging:

$$\mu_i^{\mathrm{soc}}(t) = \frac{\sum_{j\in\mathcal{N}_t(i)} k_{ij}(t)\,\tau_{j,\mathrm{loc}}(t)\,\mu_j^{\mathrm{loc}}(t)}{\sum_{j\in\mathcal{N}_t(i)} k_{ij}(t)\,\tau_{j,\mathrm{loc}}(t)}, \qquad \tau_{i,\mathrm{soc}}(t) = \sum_{j\in\mathcal{N}_t(i)} k_{ij}(t)\,\tau_{j,\mathrm{loc}}(t).$$

The resulting pair $(\mu_i^{\mathrm{soc}}(t),\, \tau_{i,\mathrm{soc}}(t))$ defines an *instantaneous social belief* that summarizes the statistical information currently available to agent i through its neighborhood. This aggregation is recomputed at every iteration and is not accumulated over time, making the social influence inherently *state–aware* and directly responsive to the network's instantaneous configuration.

Because the interaction weights $k_{ij}(t)$ depend explicitly on the local posterior means and precisions, each agent constructs a personalized social belief that generally differs from those of its neighbors. Even two agents connected to the same set of peers will compute distinct aggregations, since each edge weight reflects the specific statistical relationship between that particular pair. This formulation enables a fully decentralized and adaptive form of social learning, in which mutual influence is determined locally and symmetrically across edges, with the precision terms ensuring that more confident neighbors contribute proportionally more to the social estimate.

Optional use of adaptive weights.

The use of the pairwise coefficients $k_{ij}(t)$ is optional and mainly intended for static networks, in which neighboring nodes may encode substantially different local posteriors and the topology does not evolve dynamically. In such cases, adaptive weights act as a compatibility filter that attenuates the contribution of statistically distant or inconsistent neighbors, thereby preventing multimodal or biased aggregations. When the local beliefs are already homogeneous, as in well-connected or stationary regimes, setting $k_{ij}(t)=1$ for all edges yields an equivalent and often more stable formulation.

In this simplified configuration, the social belief reduces to the analytically exact precision–weighted fusion of local posteriors:

$$\mu_i^{\mathrm{soc}}(t) = \frac{\sum_{j\in\mathcal{N}_t(i)} \tau_{j,\mathrm{loc}}(t)\,\mu_j^{\mathrm{loc}}(t)}{\sum_{j\in\mathcal{N}_t(i)} \tau_{j,\mathrm{loc}}(t)}, \qquad \tau_{i,\mathrm{soc}}(t) = \sum_{j\in\mathcal{N}_t(i)} \tau_{j,\mathrm{loc}}(t).$$

Here, the influence of each neighbor depends solely on its local precision, yielding an unbiased and fully calibrated aggregation when the neighborhood is statistically coherent.

In particular,as demonstrated in A.2, when $k_{ij}(t)=1$, the social and local estimates become statistically independent, since the social mean $\mu_i^{\mathrm{soc}}(t)$ no longer depends on the local estimate $\mu_i^{\mathrm{loc}}(t)$ itself but only on the neighbors' beliefs. As a consequence,

the subsequent fusion between the local and social components becomes an exact Bayesian aggregation between independent Gaussian sources.

From a broader design perspective, the DBME framework was conceived to replace the role of the adaptive coefficients $k_{ij}(t)$ by acting directly on the network topology itself , through two decentralized structural mechanisms: *pruning*, which removes incoherent or statistically incompatible links (see Chapter 4), and *rewiring*, which restores connectivity by establishing new edges between statistically compatible agents (see Chapter 5). These topological interventions achieve the same goal of adaptive weighting, therefore promoting coherent interactions and attenuating spurious correlations, ensuring long–term stability even in large evolving networks.

3.3.4 Step 3: Bayesian fusion — Gaussian product

At this stage, each agent i possesses two distinct probability distributions representing complementary sources of information: a *local belief*, summarizing the evidence extracted from its own data stream, and a *social belief*, synthesized from the most recent beliefs of its neighbors. Since these two components are constructed independently at each time step t, their combination produces the posterior belief of agent i about its latent mean μ_i. This posterior encodes the only quantity that effectively guides the evolution of the algorithm, concentrating all the information available to node i at that instant.

Formally, the two beliefs are expressed as

$$p_i^{\mathrm{loc}}(\mu) = \mathcal{N}\big(\mu_i^{\mathrm{loc}}(t),\, \tau_{i,\mathrm{loc}}^{-1}(t)\big), \qquad p_i^{\mathrm{soc}}(\mu) = \mathcal{N}\big(\mu_i^{\mathrm{soc}}(t),\, \tau_{i,\mathrm{soc}}^{-1}(t)\big).$$

The first term represents the cumulative information derived from the local observations, whereas the second acts as an *informative prior* dynamically reconstructed from the current neighborhood configuration. The social belief is recomputed at every iteration based on the instantaneous statistical coherence of the network as a result, it evolves synchronously with the local evidence while remaining statistically independent of the agent's i previous local posterior.

Given conditional independence between p_i^{loc} and p_i^{soc}, the posterior belief follows directly from their product:

$$p_i^{\mathrm{post}}(\mu) \propto p_i^{\mathrm{loc}}(\mu)\, p_i^{\mathrm{soc}}(\mu) = \mathcal{N}\big(\mu_i^{\mathrm{post}}(t),\, \tau_{i,\mathrm{post}}^{-1}(t)\big),$$

with parameters obtained through precision–weighted fusion:

$$\tau_{i,\mathrm{post}}(t) = \tau_{i,\mathrm{loc}}(t) + \tau_{i,\mathrm{soc}}(t), \qquad \mu_i^{\mathrm{post}}(t) = \frac{\tau_{i,\mathrm{loc}}(t)\,\mu_i^{\mathrm{loc}}(t) + \tau_{i,\mathrm{soc}}(t)\,\mu_i^{\mathrm{soc}}(t)}{\tau_{i,\mathrm{loc}}(t) + \tau_{i,\mathrm{soc}}(t)}.$$

The resulting posterior $\mathcal{N}\big(\mu_i^{\mathrm{post}}(t),\, \tau_{i,\mathrm{post}}^{-1}(t)\big)$ constitutes the final belief of agent i at iteration t. It integrates both the individual contribution of self–acquired data and

the collective information gathered from the surrounding neighborhood, acting as a probabilistic equilibrium between private and social evidence.

From a dynamical standpoint, this update is not memoryless because the posterior at time t depends on the evolving local sufficient statistics, which implicitly encode the history of the observations, and on the social prior reconstructed from the current neighbors state , a further discussion on this topic can be found in A.3. No fused posterior belief is explicitly propagated through time; instead, information flows forward solely through cumulative evidence and instantaneous neighborhood interaction. This mechanism prevents any recursive reuse of past posteriors and guarantees that convergence arises from real information fusion rather than iterative reusing. Within this structure, the only factor that proportionally amplifies the effective precision of an agent is the size of its neighborhood: since the social precision satisfies $\tau_{i,\text{soc}} = \sum_{j \in N(i)} \tau_{j,\text{loc}}$, the number of neighbors directly scales the amount of independent information aggregated at each iteration. In particular, the target average degree r acts as a structural control parameter on the global sharpness of the posterior distributions as exhaustively covered in A.4 , which also shows that this estimation pipeline under certain topology conditions is proven to be approximately consistent.

Conclusion on the Bayesian Update Process and Structural Adaptation

As discussed throughout the previous sections, the DBME algorithm performs a sequential Bayesian update that progressively integrates information originating from two distinct sources: the agent's own data and the beliefs circulating in its neighborhood. The validity of this inference process relies on a crucial assumption: that neighboring agents are approximately homogeneous with respect to the latent variable they are estimating. In practice, this means that the local distributions exchanged within a neighborhood should correspond to nodes belonging to the same latent class (basically their true means must coincides), so that their fusion remains meaningful under the Gaussian approximation of conditional independence. However, since the initial topology of the network is random, this assumption does not hold *a priori*. It must instead *emerge dynamically* through the progressive restructuring of the graph itself.

In ColME, a related consistency requirement is enforced through PAC-based compatibility tests: two agents are considered to belong to the same latent class if their confidence intervals overlap with high probability. This mechanism operates as an external decision rule guiding collaboration, and not as an intrinsic structural property of the network. In DBME, by contrast, statistical compatibility becomes embedded within the topology itself. The network is not simply filtered according to compatibility rules, but reshaped by them: the communication structure evolves in response to parametric intervals derived from local posterior beliefs, progressively aligning connectivity with the underlying statistical heterogeneity of the agents.

To ensure that the assumptions underpinning Bayesian fusion remain valid over time, DBME incorporates two complementary topological mechanisms: **pruning** and **rewiring**. Pruning removes edges connecting statistically incompatible nodes, suppressing information pathways that would otherwise propagate inconsistent evidence and destabilize inference. Rewiring counterbalances this process by restoring connectivity where statistical evidence supports compatibility, ensuring that coherence is preserved without reintroducing heterogeneity-induced noise.

Together, these mechanisms allow DBME to satisfy the assumptions required for its own inference rules to remain valid. As the topology adapts, the network converges toward a configuration in which conditional independence and neighborhood homogeneity become increasingly accurate approximations. Only under these structural conditions does the Bayesian fusion process retain its theoretical coherence and empirical reliability. The following chapters analyze in detail how pruning and rewiring operate as coupled forces.

Chapter 4

Decentralized pruning: an evidence–driven heuristic

The pruning rule in DBME is a *decentralized, evidence–driven heuristic* that removes edges whose endpoints are statistically incompatible given the local evidence available at the current iteration. It is called "heuristic" in the operational sense because it relies on local independence and edge–wise testing without a global optimality proof, yet it remains *principled*, as its decisions are grounded on standard frequentist hypothesis testing applied to the same sufficient summaries that govern the Bayesian updates.

4.0.1 Motivation and rationale

At time t, each node i holds a local Gaussian summary

$$\mathcal{N}\big(\mu_i^{\mathrm{loc}}(t),\, \tau_{i,\mathrm{loc}}^{-1}(t)\big),$$

which consolidates the information extracted from its own data stream (up to t) under the fixed bootstrap prior. When two neighboring agents (i, j) strongly disagree relative to their joint uncertainty, the edge likely bridges heterogeneous regions of the latent space and becomes a candidate for removal. Conversely, if the discrepancy is small when expressed in units of the pooled standard deviation, the edge is statistically compatible and should be preserved.

The rule is *evidence–driven* because its selectivity naturally adapts to the amount of information: as learning progresses, local posterior precisions $\tau_{i,\mathrm{loc}}(t)$ increase, and the same absolute mean difference becomes more significant. Hence pruning tightens where evidence is strong and remains lenient where it is weak, without the need for global coordination.

4.0.2 Decision statistic and test

For each existing edge (i, j), the node–local decision is based on the two–sided z–test:

$$z_{ij}(t) = \frac{|\mu_i^{\text{loc}}(t) - \mu_j^{\text{loc}}(t)|}{\sqrt{\tau_{i,\text{loc}}^{-1}(t) + \tau_{j,\text{loc}}^{-1}(t)}}, \qquad p_{ij}(t) = 2\big[1 - \Phi\big(z_{ij}(t)\big)\big],$$

which tests $\mathcal{H}_0 : \mu_i = \mu_j$ versus $\mathcal{H}_1 : \mu_i \neq \mu_j$.

Here $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) of the standard normal variable:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt,$$

representing the probability that a standard normal variable does not exceed z. Hence, $1 - \Phi(|z|)$ corresponds to the probability of observing a value more extreme than $|z|$ under $\mathcal{H}_0$, and the factor of 2 accounts for the two–sided nature of the test. Under $\mathcal{H}_0$, $z_{ij}^2(t) \sim \chi_1^2$ (see Appendix A.1). An edge is pruned if $p_{ij}(t) < \alpha_{\text{prune}}$, where $\alpha_{\text{prune}} \in (0, 1)$ is a fixed significance level.

Why heuristic yet principled. The test is *principled* (standard likelihood–ratio test under Gaussianity), but its application is *heuristic* because the network ignores cross–edge dependencies and does not enforce a global multiple–testing correction. Furthermore apart from the empirical studies executed in 7, we have not formally demonstrated any kind of convergence to an optimal topology. However the mechanism preserves full decentralization and maintains empirically the desired bias–variance trade–off . If needed, a node–local conservative correction (e.g., Bonferroni [15] over $\deg(i)$) can be applied without altering decentralization.

4.0.3 Automatic adaptivity from evidence

A key property is the *evidence–driven adaptivity*:

- **Early stage (low precision).** Small $\tau_{i,\text{loc}}(t)$ inflate the denominator of z_{ij}, so p_{ij} remains high unless means are grossly different. Pruning is naturally conservative, preserving connectivity while beliefs are still diffuse.

- **Late stage (high precision).** As $\tau_{i,\text{loc}}(t)$ grow, even moderate mean gaps generate large z_{ij} and small p_{ij}; pruning becomes selective and exposes latent boundaries, carving out homogeneous clusters.

This behavior mimics an annealing schedule driven entirely by the data: no additional hyperparameter is required beyond α_{prune}. To ensure robust behavior under finite precision and heterogeneous uncertainty, two safeguards are introduced. First, each local precision $\tau_{i,\text{loc}}(t)$ is replaced by $\tilde{\tau}_i(t) = \max\{\tau_{i,\text{loc}}(t), \varepsilon^{-1}\}$, with $\varepsilon > 0$ small,

in order to prevent numerical instabilities caused by near–zero denominators in the computation of z_{ij}. Second, edge removals are applied symmetrically to the adjacency lists of both i and j, ensuring structural consistency of the graph throughout the pruning process.

4.0.3.1 Complexity, decentralization, and invariances

The pruning mechanism operates in a fully decentralized and edge–local fashion and is activated periodically every Δt_{prune} iterations. At each pruning round, every decision involves only a single pair of neighboring nodes and depends exclusively on their current local posterior means and precisions (μ_i, τ_i) and (μ_j, τ_j). Because these quantities are already available as part of each node's belief state, no global coordination or shared memory is required. This independence across edges makes the procedure embarrassingly parallel: every node can evaluate its neighborhood concurrently, without synchronization barriers or message ordering constraints.

For a given node i, the computational effort grows linearly with its degree, requiring $O(\deg(i))$ pairwise evaluations per pruning cycle. Each comparison entails only constant–time operations, reading the neighbor's mean and precision, computing a standardized distance, and checking a threshold, so the total cost of a full pruning round across the graph is $O(|E_t^-|)$, where $|E_t^-|$ denotes the number of edges present before the pruning step. Since this operation is performed only every Δt iterations, the amortized computational cost per time step remains low, further reinforcing the scalability of the method in large networks.

Communication overhead is equally modest. During each pruning phase, nodes exchange only their current Gaussian parameters, which constitute fixed–size messages. No additional statistics or normalization factors are needed, ensuring that bandwidth usage scales linearly with the number of active edges and remains bounded per node.

A key property of the pruning rule is its invariance to affine transformations that preserve Gaussian scale. If all local summaries undergo a reparameterization of the form $x' = ax + b$ with $a > 0$, the standardized pairwise distances remain unchanged, since the scaling factor cancels in the ratio between mean differences and combined precisions. As a result, the logical outcome of the pruning stage is invariant to the units or offsets used to represent the variables.

Finally, because the pruning criterion depends solely on the state of the system at time t, the outcome of a pruning round is unaffected by the order in which edges are evaluated. Whether edges are processed synchronously or asynchronously, the same set of connections will be retained or removed.

4.0.4 Pseudocode (node–local execution)

Require: Node i local summary $(\mu_i^{\mathrm{loc}}(t),\, \tau_{i,\mathrm{loc}}(t))$
Require: Neighbor set $\mathcal{N}_t(i)$ with summaries $(\mu_j^{\mathrm{loc}}(t),\, \tau_{j,\mathrm{loc}}(t))$
Require: Significance $\alpha_{\mathrm{prune}} \in (0,1)$, precision floor $\varepsilon > 0$

$\tilde{\tau}_i \leftarrow \max\{\tau_{i,\mathrm{loc}}(t),\, \varepsilon^{-1}\}$
$\mathcal{R} \leftarrow \varnothing$ ▷ edges to remove
for all $j \in \mathcal{N}_t(i)$ **do**
    $\tilde{\tau}_j \leftarrow \max\{\tau_{j,\mathrm{loc}}(t),\, \varepsilon^{-1}\}$
    $d \leftarrow |\mu_i^{\mathrm{loc}}(t) - \mu_j^{\mathrm{loc}}(t)|$
    $s \leftarrow \sqrt{\tilde{\tau}_i^{-1} + \tilde{\tau}_j^{-1}}$
    $z \leftarrow d/s$
    $p \leftarrow 2\big[1 - \Phi(z)\big]$
    if $p \le \alpha_{\mathrm{prune}}$ **then**
        $\mathcal{R} \leftarrow \mathcal{R} \cup \{(i,j)\}$
    end if
end for
Remove all $(i,j) \in \mathcal{R}$ from i's adjacency; notify j for symmetry

4.0.5 Error trade–offs and design guidance

The significance level α_{prune} controls the balance between Type I and Type II statistical errors in local pruning decisions. Each test on an edge (i,j) assesses whether two neighboring nodes remain consistent with belonging to the same latent cluster. A wrong pruning decision can therefore arise in two distinct ways.

A *Type I error* (false removal) occurs when $\mathcal{H}_0 : \mu_i = \mu_j$ is rejected even though it is true, leading to the premature deletion of a correct edge. Such errors cause fragmentation: clusters may break apart before local beliefs have converged, reducing information flow and slowing consensus.

Conversely, a *Type II error* (false retention) occurs when the test fails to reject $\mathcal{H}_0$ despite a real difference between μ_i and μ_j, causing persistent connections between heterogeneous regions and delaying structural adaptation.

The parameter α_{prune} explicitly tunes this trade–off: smaller values make the test more conservative (fewer Type I errors), preserving connectivity but delaying boundary detection; larger values have the opposite effect, promoting faster cluster separation but at the risk of over–pruning in high–uncertainty phases. In practice, the optimal setting depends on the noise regime and on whether the application prioritizes early separation or robust connectivity during the transient phase.

4.0.6 Limitations and scope

The pruning rule inherits its behavior and limitations from the assumptions underlying the local z–test. Two main conditions are implicit in its validity: (i) statistical independence of local data streams across nodes, ensuring that deviations in posteriors genuinely reflect heterogeneity; and (ii) approximate normality and unimodality of belief distributions within each neighborhood. In this sense, the pruning criterion is rooted in a frequentist hypothesis test: the z–statistic is used to evaluate the null hypothesis of equal means, relying on asymptotic Gaussian assumptions , that we already justified in 3.3.2.
Although this test is formally frequentist, it admits a close Bayesian interpretation under the Gaussian assumptions adopted in DBME. Consider two neighboring local beliefs,

$$p_i^{\mathrm{loc}}(\mu) = \mathcal{N}(\mu_i^{\mathrm{loc}},\, \tau_{i,\mathrm{loc}}^{-1}), \qquad p_j^{\mathrm{loc}}(\mu) = \mathcal{N}(\mu_j^{\mathrm{loc}},\, \tau_{j,\mathrm{loc}}^{-1}).$$

The posterior distribution of their difference $\delta_{ij} = \mu_i - \mu_j$ is therefore

$$\delta_{ij} \sim \mathcal{N}\Big(\mu_i^{\mathrm{loc}} - \mu_j^{\mathrm{loc}},\; \tau_{i,\mathrm{loc}}^{-1} + \tau_{j,\mathrm{loc}}^{-1}\Big).$$

A $(1-\alpha)$ Bayesian credible interval for δ_{ij} excludes zero if and only if

$$|\mu_i^{\mathrm{loc}} - \mu_j^{\mathrm{loc}}| > z_{1-\alpha/2}\sqrt{\tau_{i,\mathrm{loc}}^{-1} + \tau_{j,\mathrm{loc}}^{-1}},$$

which coincides exactly with the rejection rule of the two–sided z–test at significance level α. Hence, under Gaussian local beliefs, rejecting the null hypothesis $H_0 : \mu_i = \mu_j$ via the z–test is equivalent to verifying that zero does not belong to the Bayesian credible interval of the difference δ_{ij}. This equivalence should not be confused with the simpler heuristic based on the overlap of the marginal intervals for μ_i and μ_j, which is more conservative and does not coincide exactly with the z–test threshold. During early transients, when estimates remain uncertain, the test may produce temporary misclassifications: large uncertainties blur cross–cluster boundaries, increasing Type II errors, whereas strong early fluctuations can induce isolated false removals (Type I errors). These effects are largely self–correcting as learning progresses: posterior precisions grow, cluster means stabilize, and the statistical power of the test improves. However, it is important to note that each pruning operation inevitably reduces the amount of independent information available within the graph. Removing an edge does not only suppress potentially inconsistent influence, but also eliminates a source of statistically independent evidence that could have contributed to variance reduction and faster convergence.

Remark. Connectivity maintenance after pruning is delegated to the decentralized reconnection routine described in Chapter 5. This complementary mechanism compensates for local Type I removals by restoring statistically plausible links, ensuring that information flow remains uninterrupted even under conservative significance settings.

Chapter 5

Rewiring – Decentralized Restoration of Connectivity

Pruning, while essential to remove misleading edges, inevitably reduces the overall connectivity of the network, this means that over successive iterations, entire regions of agents may become weakly linked or even fully isolated, causing a temporary degradation in global information flow. The purpose of the rewiring phase is therefore to restore a minimal level of connectivity by creating new edges between agents that are statistically compatible according to their current local evidence.

Rewiring in DBME is a fully decentralized *evidence-driven heuristic* where each node autonomously decides whether to form new connections based solely on the same pairwise statistical compatibility criterion employed in pruning. As a consequence rewiring does not rely on globally informed metrics, on the contrary it emerges spontaneously from the local interactions of agents that recognize one another as statistically equivalent.

5.1 Heuristic principle and statistical rationale

The guiding intuition is simple: after pruning has eliminated inconsistent neighbors, each node may be left with too few social connections to sustain effective information diffusion. To prevent this, rewiring restores a small number of links by testing potential partners for statistical compatibility using the same frequentist z–statistic introduced in Chapter 4:

$$z_{ij}(t) = \frac{|\mu_i^{\mathrm{loc}}(t) - \mu_j^{\mathrm{loc}}(t)|}{\sqrt{\tau_{i,\mathrm{loc}}^{-1}(t) + \tau_{j,\mathrm{loc}}^{-1}(t)}}, \qquad p_{ij}(t) = 2[1 - \Phi(z_{ij}(t))].$$

A new link (i, j) is added only if $p_{ij}(t) \geq \alpha_{\mathrm{rewire}}$, that is, if the null hypothesis of equal means cannot be rejected at the chosen confidence level.

The decision is again purely local: node i evaluates compatibility with a small

set of randomly sampled candidate nodes and accepts only those that satisfy the statistical coherence condition. Such as in pruning phase the outcome depends on the uncertainty: when local precisions are low, many pairs may appear compatible, promoting mild reconnection; as precisions grow, the test becomes stricter and new edges form only within well-defined homogeneous regions. Overall the effect is of a progressive reinforcement of statistically consistent clusters.

5.2 Operational procedure

Such as Pruning , Rewiring is executed periodically, at intervals Δt_{rewire} , following each pruning cycle. At these times, each node i evaluates whether its current degree $\deg(i,t)$ falls below a target average r (corresponds to the mean degree of the initial random graph). If so, node i enters a random search phase where it scans a limited set of randomly sampled candidate nodes $j \notin \mathcal{N}_t(i)$ across the entire graph and applies the compatibility test defined above.

Decentralized heuristic.

Each node executes the rewiring procedure independently, without any form of global synchronization consequently neither knowledge of disconnected components or global connectivity is required. The only minimal global element is the availability of an identifier space $V = \{1, \dots, N\}$ representing the set of all node indices in the population. This set carries no structural or state information: it merely defines the addressable name space from which each node can independently draw random candidates during the sampling phase. Having access to this global list of identifiers allows fully decentralized random sampling without any need for coordination or global topology awareness.

At each rewiring round, node i checks whether $\deg(i,t) < r$. If so, it samples M random candidates, $j \sim \text{Unif}(V \setminus \mathcal{N}_t(i))$. This bounded sampling keeps the computational and communication load per event at $O(M)$, independent of the total network size.

For each candidate, node i computes a standardized compatibility score:

$$z_{ij}(t) = \frac{|\mu_i^{\text{loc}}(t) - \mu_j^{\text{loc}}(t)|}{\sqrt{\tau_{i,\text{loc}}^{-1}(t) + \tau_{j,\text{loc}}^{-1}(t)}}, \qquad p_{ij}(t) = 2[1 - \Phi(z_{ij}(t))],$$

which measures the probability that the two local estimates originate from the same underlying mean. The candidate is considered statistically compatible if $p_{ij}(t) \geq \alpha_{\text{rewire}}$.

After passing this compatibility test, node i performs a lightweight peer–to–peer

handshake in which each candidate j reports its current degree together with its mean and precision, via the message:

$$\text{msg}_j = (\mu_j^{\text{loc}}(t),\, \tau_{j,\text{loc}}(t),\, \deg(j,t)).$$

If j is not yet saturated ($\deg(j,t) < r$), a new undirected edge (i,j) is created symmetrically in both adjacency lists. The process terminates once node i reaches $\deg(i,t) \geq r$.

This mechanism ensures that rewiring decisions depend only on locally available beliefs and one-hop peer information, preserving full decentralization. Side by side with pruning, rewiring form a self-balancing mechanism that alternates spurious edge deletion with evidence-based reconnection, maintaining network connectivity while allowing the topology to adapt dynamically to the heterogeneity of local beliefs.

5.2.1 Pseudocode (node–local view)

Require: Node i local summary $(\mu_i^{\text{loc}}(t),\, \tau_{i,\text{loc}}(t))$
Require: Current neighbor set $\mathcal{N}_t(i)$ and total population $V = \{1, \ldots, N\}$
Require: Target degree r; significance level α_{rewire}; precision floor ε

if $\deg(i,t) < r$ **then**
  Randomly sample M candidates $j \in V \setminus \mathcal{N}_t(i)$
  $\tilde{\tau}_i \leftarrow \max\{\tau_{i,\text{loc}}(t),\, \varepsilon^{-1}\}$
  for all j in candidate set **do**
    Request from j: $\text{msg}_j = (\mu_j^{\text{loc}}(t),\, \tau_{j,\text{loc}}(t),\, \deg(j,t))$
    $\tilde{\tau}_j \leftarrow \max\{\tau_{j,\text{loc}}(t),\, \varepsilon^{-1}\}$
    $d \leftarrow |\mu_i^{\text{loc}}(t) - \mu_j^{\text{loc}}(t)|$
    $s \leftarrow \sqrt{\tilde{\tau}_i^{-1} + \tilde{\tau}_j^{-1}}$
    $z \leftarrow d/s$
    $p \leftarrow 2[1 - \Phi(z)]$
    if $p \geq \alpha_{\text{rewire}}$ **and** $\deg(j,t) < r$ **then**
      Add symmetric edge (i,j)
      break if $\deg(i,t) \geq r$
    end if
  end for
end if

5.2.2 Connectivity preservation

From a theoretical perspective, rewiring acts as a local mechanism of statistical connectivity control: each node preserves its participation in the global communication graph only by establishing links with peers whose current empirical distributions are

statistically compatible with its own. Because the compatibility test is symmetric and bounded by $p_{ij}(t) \geq \alpha_{\text{rewire}}$, the resulting connections naturally emerge within regions of low divergence in the belief space.

The process produces two complementary effects:

- **Restoration of information flow.** Sparsely connected nodes regain access to collective information streams, preventing stagnation of their posteriors.
- **Topological self–regularization.** Edges reappear preferentially within homogeneous regions, preventing cross–cluster contamination.

5.3 Algorithmic properties and limitations

The decentralized rewiring mechanism is heuristic but consistent with the statistical structure of DBME. However, of great importance remains the setting of the threshold parameter for the rewiring test , which we remember must be chosen a priori on the basis of the type of behavior that the rewiring must implement:

The parameter α_{rewire} controls the trade–off between reconnection accuracy and reconnection speed:

- Low α_{rewire} (lenient test) faster reconnection but may temporarily include weakly compatible nodes.
- High α_{rewire} (strict test) selective reconnection, allowing only strongly compatible agents to form new links.

Anyway, empirically, the system is robust to moderate variations: since pruning and rewiring share the same underlying test, their effects naturally balance over time, indeed edges erroneously reintroduced during rewiring are likely to be re-pruned in subsequent cycles.

5.4 Emergent connectivity and dynamic equilibrium

Over multiple pruning–rewiring cycles, the network empirically converges toward a dynamic equilibrium characterized by statistically coherent clusters interconnected through a minimal set of stable edges, as we will see in 7. Within each cluster, local consensus is rapidly achieved; between clusters, information diffusion stops, preserving global coherence.

This alternating sequence of selective removal and cautious reconnection allows the DBME topology to remain simultaneously sparse, modular, and functionally

connected. The network self-organizes its own connectivity based on the compatibility of the beliefs emerging from the inference process itself.

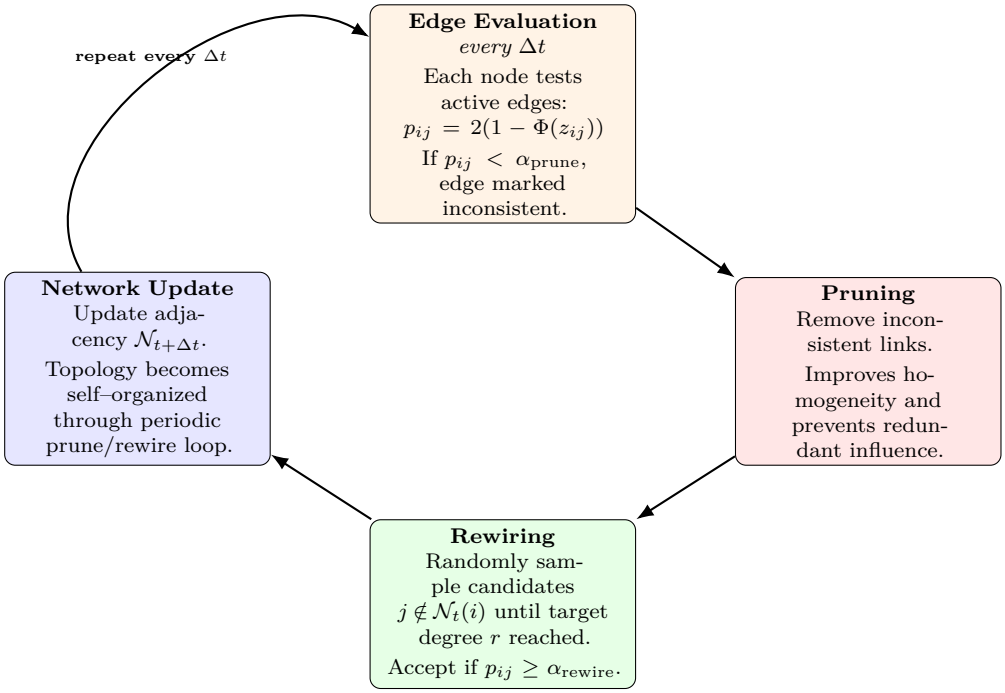


Figure 5.1: Compact cyclic representation of pruning and rewiring in DBME.

Choice of significance levels

Although pruning and rewiring rely on the same statistical test, their thresholds serve opposite purposes : pruning removes edges with $p_{ij} < \alpha_{\text{prune}}$, while rewiring creates new ones only when $p_{ij} \geq \alpha_{\text{rewire}}$. Since the p–value quantifies statistical compatibility, large p_{ij} correspond to pairs of agents with highly similar beliefs, whereas small p_{ij} indicate disagreement.

Keeping α_{rewire} larger than α_{prune} makes the reconnection process more conservative and selective because an higher α_{rewire} imposes a stricter acceptance rule, requiring very high statistical similarity before establishing a new link, whereas a lower α_{prune} allows weakly compatible edges to persist until their incompatibility becomes statistically undeniable.

This asymmetry prevents rapid oscillations between pruning and reconnection, ensuring that new edges emerge only when the associated local posteriors have converged. In practice, settings around $\alpha_{\text{prune}} \approx 10^{-3}$ and $\alpha_{\text{rewire}} \approx 10^{-1}$ yield stable and inter-

pretable network dynamics: incompatible edges are quickly removed, while new links form only among agents whose beliefs are both precise and statistically aligned.

5.5 Computational Complexity of DBME

From an algorithmic perspective, DBME is designed to maintain a minimal computational footprint at the level of each individual agent. Every node executes a fixed sequence of scalar operations per iteration, whose cost scales only with its local degree d_i and remains independent of the global network size N.

Local processing. Each agent updates its sufficient statistics $\{S_i(t), Q_i(t), N_i(t)\}$ and computes the corresponding local posterior parameters $(\mu_i^{\mathrm{loc}}(t),\ \tau_{i,\mathrm{loc}}^{-2}(t))$,so the update involves only scalar additions and multiplications, ensuring a constant-time complexity:

$$O(1).$$

Even under the Normal-Inverse-Gamma formulation, where four hyperparameters $(\mu_i, \kappa_i, \alpha_i, \beta_i)$ are updated recursively, the computation remains closed–form and independent of the number of past samples.

Social aggregation. Each node synthesizes a *social belief* $(\mu_i^{\mathrm{soc}},\ \tau_{i,\mathrm{soc}}^{-2})$ by aggregating the local posteriors of its neighbors through precision–weighted averaging:

$$\mu_i^{\mathrm{soc}}(t) = \frac{\sum_{j\in\mathcal{N}(i)} k_{ij}(t)\,\tau_{j,\mathrm{loc}}^{-2}(t)\,\mu_j^{\mathrm{loc}}(t)}{\sum_{j\in\mathcal{N}(i)} k_{ij}(t)\,\tau_{j,\mathrm{loc}}^{-2}(t)}, \qquad \tau_{i,\mathrm{soc}}^{-2}(t) = \sum_{j\in\mathcal{N}(i)} k_{ij}(t)\,\tau_{j,\mathrm{loc}}^{-2}(t).$$

This operation requires one precision-weighted summation per neighbor, resulting in a linear complexity with respect to the local degree:

$$O(d_i).$$

If adaptive coefficients $k_{ij}(t)$ are used, each requires computing a pairwise divergence $\ell_{ij}(t)$ (a scalar Mahalanobis distance in precision space), which adds only a constant factor but it would also mean eliminating pruning and rewiring.

Fusion step. The Bayesian fusion of local and social posteriors is performed through the additive rule in precision space:

$$\tau_{i,\mathrm{post}}^{-2}(t) = \tau_{i,\mathrm{loc}}^{-2}(t) + \tau_{i,\mathrm{soc}}^{-2}(t), \qquad \mu_i^{\mathrm{post}}(t) = \tau_{i,\mathrm{post}}^{2}(t)\left(\frac{\mu_i^{\mathrm{loc}}(t)}{\tau_{i,\mathrm{loc}}^{2}(t)} + \frac{\mu_i^{\mathrm{soc}}(t)}{\tau_{i,\mathrm{soc}}^{2}(t)}\right).$$

Because this step involves only scalar arithmetic, its computational cost per iteration is constant:

$$O(1).$$

Pruning and rewiring. Graph adaptation occurs once every Δt iterations. During pruning, each node computes the statistical compatibility

$$p_{ij}(t) = 2\left[1 - \Phi\left(\frac{|\mu_i^{\mathrm{loc}}(t) - \mu_j^{\mathrm{loc}}(t)|}{\sqrt{\tau_{i,\mathrm{loc}}^{-2}(t) + \tau_{j,\mathrm{loc}}^{-2}(t)}}\right)\right]$$

for all neighbors $j \in \mathcal{N}(i)$, an operation that scales linearly with d_i:

$$O(d_i).$$

Rewiring, on the other hand, involves sampling a bounded number M of candidate nodes and testing the same compatibility condition, giving a cost $O(M)$ per event. Since both pruning and rewiring occur only every Δt steps, the total number of adaptation events per node over the full time horizon T is $T/\Delta t$. The cumulative topological adjustment cost per agent is therefore

$$O\left(\frac{T}{\Delta t}\,(d_i + M)\right),$$

and the amortized per–iteration overhead becomes

$$O\left(\frac{d_i + M}{\Delta t}\right),$$

which is negligible compared with the $O(d_i)$ cost of the aggregation phase.

Overall per–agent complexity. Summing all contributions, the instantaneous computational cost per node per iteration is

$$O(1 + d_i) = O(d_i),$$

where the constant term corresponds to the local Bayesian updates and the linear term to social aggregation and compatibility evaluation. Over the entire time horizon T, the cumulative computational effort per node is

$$O(T\, d_i),$$

and for a network with bounded average degree r the total cumulative cost across all nodes in time is

$$O(T\, N\, r).$$

Comparison with COLME. An important consequence of enforcing a fixed target degree r through the pruning–rewiring mechanism is that the network remains sparse by construction. When r is tuned a priori to a reasonable value, as done in our experiments, the global number of edges scales as $|E_t| \sim Nr$, and all social and topological operations remain linear in the number of active links. As a result, the

overall computational complexity per step of DBME scales as

$$O(Nr),$$

This represents an improvement over the worst-case complexity of COLME, where each agent may test compatibility against all other agents, leading to a quadratic cost $O(N^2)$.

Chapter 6

Theoretical Justification: Credibility Intervals and Conditional Independence

6.1 Local and social posteriors under conditional independence

The conditional independence assumption that underlies DBME follows directly from the probabilistic nature of the data, indeed, as we have already seen in 3.1 each agent i collects its own observations

$$x_{i,t} \sim \mathcal{F}_c, \qquad \text{for } i \in \mathcal{C}_c,$$

where $\mathcal{C}_c$ denotes the subset of agents belonging to the same latent class c, and $\mathcal{F}_c$ is the corresponding generative distribution.

Within each homogeneous cluster, all agents draw independent realizations from the same process, this means that the samples $\{x_{i,t}\}_{i \in \mathcal{C}_c}$ are independent and identically distributed across agents of the same class, not only within each local sequence. This property guarantees that, in expectation, the local posteriors produced by different agents represent independent estimators of the same latent mean μ_c^*.

Because each local dataset $\mathcal{D}_i$ originates from an independent sampling process, the two main sources of information available to node i, its own data and the posteriors received from its neighbors, can be treated as conditionally independent given the latent parameter μ_i:

$$p(p_i^{\text{loc}}, p_i^{\text{soc}} \mid \mu_i) \simeq p(p_i^{\text{loc}} \mid \mu_i)\, p(p_i^{\text{soc}} \mid \mu_i).$$

The Gaussian fusion rule used in DBME,

$$p_i^{\text{post}}(\mu) \propto p_i^{\text{loc}}(\mu)\, p_i^{\text{soc}}(\mu),$$

then is a meaningful Bayesian operation, combining independent probabilistic estimates of the same underlying quantity.

When the network reaches a stationary regime , that is, when local inverse precisions $\tau_{i,\mathrm{loc}}^{-1}(t)$ stabilize , the assumption becomes self-consistent: all nodes belonging to the same class produce posteriors centered on the same mean, and the mutual influence among them reflects statistical aggregation.

Independence without identical distribution

During the early transient phase of network evolution, connections may temporarily form between agents belonging to different latent classes, so that their respective local posteriors $p_i^{\mathrm{loc}}(\mu)$ and $p_j^{\mathrm{loc}}(\mu)$ represent independent estimators of distinct true means. However, this does not represent a problem from an inferential point of view since the datasets $\mathcal{D}_i$ and $\mathcal{D}_j$ are disjoint, the corresponding beliefs remain statistically independent even though they are *not identically distributed*. In this regime, the Gaussian fusion

$$p_i^{\mathrm{post}}(\mu) \propto p_i^{\mathrm{loc}}(\mu)\, p_j^{\mathrm{loc}}(\mu)$$

is still a formally valid Bayesian operation , it aggregates independent pieces of information , but it is not optimal in terms of statistical compatibility, since the fused posteriors refer to heterogeneous generative processes.

As pruning and rewiring gradually remove incompatible edges, the network self-organizes into homogeneous clusters where all agents share the same generative distribution $\mathcal{F}_c$. At this point both independence and identical distribution are recovered, and the Gaussian fusion becomes fully coherent.

To achieve this result, DBME enforces an explicitly **acyclic flow of information**, where at each iteration the structure ensures that information propagates forward without recursive reuse:

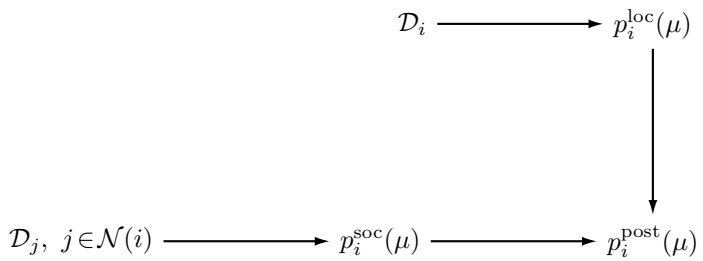


Figure 6.1: Information flow for agent i

As a result, every update remains a correct Bayesian combination of independent evidence sources , a formal illustration of this independence principle, including a toy proof in the three-node case, is provided in Appendix A.2.

Equivalently, one can view the network at each iteration as a static snapshot (as illustrated in Fig. 6.1) in which all operations occur over mutually independent quantities. At time t, every social update is computed using the neighbors' *local* posteriors $p_j^{\text{loc}}(\mu)$, which depend solely on disjoint datasets $\mathcal{D}_j$ collected up to that step.

Within this instantaneous configuration, the exchange of information across edges involves no statistical overlap: all messages circulating in the network are independent realizations conditioned on their respective latent means. Although posterior means evolve recursively in an implicit way, the computations performed at a fixed iteration do not reuse any previously shared information. The asymptotic implications of this structure, including consistency and the equivalence with centralized Gaussian fusion in the limit of stable topology, are discussed in Appendix A.4.

6.2 Experimental configuration and parameter regime

Theoretical and empirical analyzes presented in this chapter refer to the configuration adopted for large-scale DBME experiments. All simulations were conducted with $N = 10^4$ agents and $T = 2000$ iterations on an undirected random graph with average degree 10.

Each node observes an independent Gaussian process generated from one of two latent classes, respectively distributed as

$$x_{i,t} \sim \begin{cases} \mathcal{N}(\mu = 0,\, \sigma^2 = 4), & i \in \mathcal{C}_0, \\ \mathcal{N}(\mu = 1,\, \sigma^2 = 4), & i \in \mathcal{C}_1, \end{cases}$$

where $\mathcal{C}_0$ and $\mathcal{C}_1$ denote the subsets of agents belonging to each class.

The empirical prior for each agent is estimated from $T_0 = 50$ bootstrap samples drawn from its local data stream before any social interaction occurs, providing an initial data-driven estimate of both local mean and variance. The tolerance on the estimation error is fixed to $\epsilon = 0.1$, such that

$$|\hat{\mu}_i(t) - \mu_i^*| < \epsilon = 0.1.$$

The probability of violating this condition,

$$P(|\hat{\mu}_i(t) - \mu_i^*| > \epsilon)\,,$$

represents the empirical estimation error rate, which the algorithm seeks to minimize through iterative local-social fusion.

Similarity parameters

Pruning and rewiring operations are performed every 150 iterations. At each pruning step, edges connecting statistically dissimilar agents (based on the mean-similarity test with significance level $\alpha_{\text{prune}} = 0.001$) are removed, while during rewiring, new edges are created between statistically compatible nodes (selected using $\alpha_{\text{rewire}} = 0.1$).

As described in Chapters 4 and 5:

- α_{prune} defines the rejection threshold for pairwise divergence in posterior means, determining how different two connected agents must become before their link is removed;
- α_{rewire} sets the compatibility threshold for new link formation between previously disconnected nodes.

Together, these parameters define the statistical sensitivity of the topological adaptation process:

$$\mathcal{H}_0 : \ \mu_i \approx \mu_j \quad \text{vs.} \quad \mathcal{H}_1 : \ \mu_i \neq \mu_j,$$

where smaller α values correspond to stricter, more conservative rules, favoring topological stability and intra-cluster homogeneity.

Inference modes and simulation protocol

Four inference configurations were tested under identical dynamics: the *known-fixed* mode assumes a constant and known observation variance σ^2; the *plugin* mode updates this variance empirically at each iteration; the *NIG* mode performs joint inference of mean and variance via a Normal–Inverse–Gamma prior; and the *local-only* mode excludes social interaction, serving as a calibration baseline.

Each configuration was repeated for $N_{\text{runs}} = 150$ independent Monte Carlo runs with distinct random seeds ($113 \times i$, $i = 1, \ldots, 150$). Unless otherwise specified, the social fusion weights were fixed to $k_{ij} = 1$ to isolate the intrinsic Bayesian calibration properties of each inference mode, avoiding the variability introduced by dynamic weighting schemes.

6.3 Uncertainty propagation and empirical calibration

Given the conditional independence between local and social posteriors, the posterior precision of each agent evolves as the additive combination of two statistically independent contributions:

$$\tau_{i,\text{post}}(t) \approx \tau_{i,\text{loc}}(t) + \sum_{j \in \mathcal{N}_t(i)} \tau_{j,\text{loc}}(t),$$

where $\tau_{i,\text{loc}}(t)$ denotes the precision (inverse variance) of agent i's local posterior at time t, and the sum aggregates the precision contributions from its neighbors.

Uncertainty for this reason contracts smoothly over time, with posterior inverse precisions (variances) decreasing monotonically as evidence accumulates across the network.

To assess the empirical calibration of this process, the *coverage* was computed as the fraction of times the true latent mean μ_i^* falls within the 95% posterior credibility interval of each agent. Formally,

$$\mathbf{Coverage_{95}} = \frac{1}{N_{\text{runs}} N_{\text{agents}} T} \sum_{r=1}^{N_{\text{runs}}} \sum_{i=1}^{N_{\text{agents}}} \sum_{t=1}^{T} \mathbb{I}\Big\{ \mu_i^* \in (\mu_{i,r}^{\text{post}}(t) - z\, \tau_{i,r,\text{post}}^{-1/2}(t),\ \mu_{i,r}^{\text{post}}(t) + z\, \tau_{i,r,\text{post}}^{-1/2}(t)) \Big\}$$

where $z = 1.96$ corresponds to the 95% quantile of the standard normal distribution, and $\sqrt{\tau_{i,r,\text{post}}^{-1}(t)}$ denotes the posterior standard deviation.

Here $\mathbb{I}[\cdot]$ denotes the indicator function, equal to 1 when the true value lies within the confidence interval and 0 otherwise. All results reported below correspond to averages over $N_{\text{runs}} = 150$ Monte Carlo repetitions, each involving $N = 10^4$ agents and $T = 2000$ iterations on a random graph of average degree 10, with pruning and rewiring events triggered every 150 steps.

Mode	**Coverage**$_{95}$	z**-mean**	z**-std**	**Within 1**σ **(ideal** 0.6827**)**
Known-fixed	0.9444	0.0021	1.0242	0.6731
Plugin	0.9434	0.0021	1.0273	0.6720
NIG	0.9440	0.0021	1.0253	0.6726
Local-only	**0.9496**	**0.0008**	**1.0013**	**0.6817**

Table 6.1: Global coverage and calibration summary under the filtered regime ($T = 2000$, $N_{\text{runs}} = 150$). All inference modes exhibit nearly Gaussian residuals with zero bias and unit variance. The Bayesian fusion modes show a mild ($\sim$1%) undercoverage due to temporal correlations in posterior updates, while the `local_only` baseline remains perfectly calibrated.

The results in Table 6.1 show that all inference modes maintain coverage extremely close to the nominal 0.95 level. The three Bayesian schemes (known-fixed, plugin, NIG) stabilize around 0.944, whereas the purely local mode (local-only) attains 0.9496, effectively matching the theoretical target. The slight undercoverage observed in the Bayesian modes originates primarily from the transient phase of the network, when local posteriors are still adapting and the precision terms are not yet fully stabilized. Once the system reaches equilibrium, all methods exhibit nearly Gaussian residuals with zero bias and unit variance.

6.3.1 Transient underestimation of uncertainty

During the early iterations, each agent has access to few local samples, and its inverse precision estimate $\tau_{i,\text{loc}}^{-1}(t)$ (variance) remains highly unstable. At this stage, the social posteriors $p_i^{\text{soc}}(\mu)$ are built from neighbors that are themselves only partially informed.

Consequently, the initial fusion steps tend to slightly underestimate uncertainty: the aggregated posteriors appear overconfident relative to the available evidence. Once this bias is introduced, it propagates forward through the recursive update rule, causing the network to converge to a mildly overconfident equilibrium state, with credible intervals roughly 1–0.5% narrower than optimal.

A second contributing factor is the intrinsic temporal and spatial dependence of posterior estimates, discussed in A.3. Because $p_i^{\text{post}}(\mu, t)$ reuses the entire cumulative dataset $\mathcal{D}_{i,1:t}$, consecutive posteriors are not statistically independent. This dependency effectively reduces the number of independent realizations and induces a small compression of empirical uncertainty.

The resulting Z-score statistics ($z_{\text{mean}} \approx 0$, $z_{\text{std}} \approx 1.02$) confirm this mild overconfidence. This structural autocorrelation, typical of any recursive Bayesian estimator, fully explains the $\sim 1\%$ undercoverage observed in the fusion modes.

In addition, this transient bias is amplified by small subgroups of agents whose posterior beliefs remain uncertain between the two latent classes $\mathcal{N}(0, 4)$ and $\mathcal{N}(1, 4)$. These boundary agents require more iterations to reach a stationary regime, as their local evidence alternates between the two hypotheses.

During this period, their inverse precision estimates (variances) fluctuate significantly, temporarily perturbing the global calibration. As later shown in Section 6.5, these agents represent the primary source of residual heterogeneity in the empirical coverage distribution.

6.3.2 Heterogeneity among agents

Even though the network eventually forms coherent clusters of statistically similar nodes, the local uncertainty levels remain heterogeneous across agents. Nodes in stable, low-variance regions produce tighter posteriors, while those with noisier observations maintain broader ones.

When these distinct uncertainty profiles are fused, coherent clusters contribute slightly overconfident beliefs, while peripheral or noisy regions compensate with conservative ones. The global coverage

$$\text{Coverage}_{95} \approx 0.944$$

thus represents the population-weighted equilibrium between these local regimes.

This pattern is consistent across all Bayesian inference modes, demonstrating that the minor undercoverage is a structural property of decentralized Bayesian aggregation under finite-sample and asynchronous exchange.

Overall, these results confirm that DBME produces accurately calibrated posterior variances and credible intervals. The small deviation from the theoretical target reflects the inherent stochasticity of sequential Bayesian inference and the finite-time adaptation of a large-scale, evolving network.

6.3.3 Calibration curves

Beyond the global 95% coverage, the calibration of the posterior variances was assessed across multiple confidence levels. For each inference mode, a set of nominal coverage levels $\mathcal{C} = \{0.5, 0.55, \ldots, 0.99\}$ was considered, and the corresponding quantile of the standard normal distribution was computed as

$$z_c = \Phi^{-1}\left(\tfrac{1}{2} + \tfrac{c}{2}\right),$$

where Φ^{-1} denotes the inverse of the cumulative normal distribution.

Each agent was then tested for whether the true latent mean μ_i^* lay within the posterior interval:

$$\mu_i^{\text{post}} - z_c \sqrt{\tau_i^{\text{post},-1}} \leq \mu_i^* \leq \mu_i^{\text{post}} + z_c \sqrt{\tau_i^{\text{post},-1}},$$

where $\sqrt{\tau_i^{\text{post},-1}}$ denotes the posterior standard deviation, yielding the empirical coverage

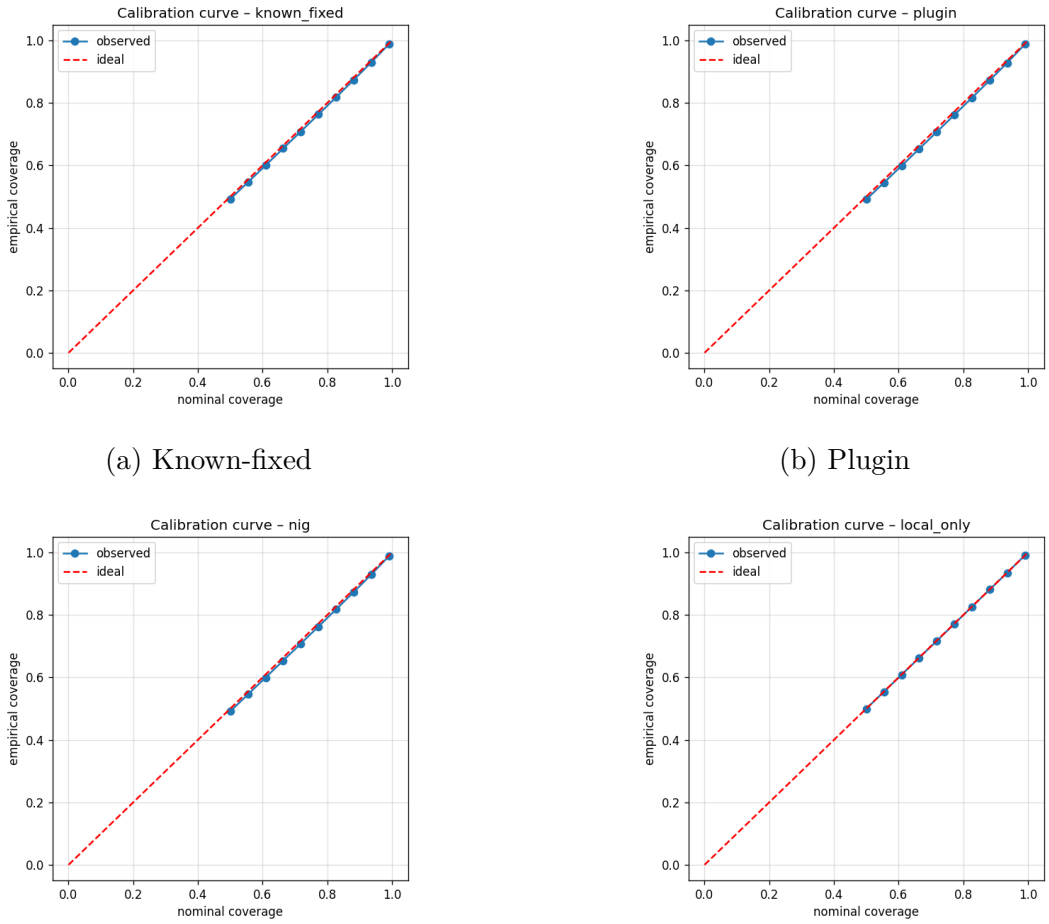
$$\widehat{C}(c) = \frac{1}{N_{\text{agents}}} \sum_{i=1}^{N_{\text{agents}}} \mathbb{I}\left[\left|\mu_i^{\text{post}} - \mu_i^*\right| \leq z_c \sqrt{\tau_i^{\text{post},-1}}\right].$$

A perfectly calibrated model satisfies $\widehat{C}(c) = c$ for all $c \in [0, 1]$.

Figure 6.2 shows the empirical calibration curves (blue) against the ideal diagonal (red dashed line). All inference modes exhibit a near-perfect alignment, indicating that the posterior inverse precisions (variances) τ_i^{-1} are statistically consistent with the true dispersion of the latent means.

The slopes of all curves remain extremely close to unity, confirming that the empirical coverage matches the nominal target across the entire confidence spectrum. Bayesian fusion modes show a minimal downward deviation from the diagonal in the mid-range ($0.5 \leq c \leq 0.8$), consistent with the slight undercoverage quantified in Table 6.1.

The local-only baseline instead lies almost exactly on the diagonal, demonstrating that purely local inference produces fully calibrated confidence intervals when social fusion is disabled.



(a) Known-fixed

(b) Plugin

(c) NIG

(d) Local-only

Figure 6.2: Calibration curves: empirical versus nominal coverage. The near-diagonal alignment demonstrates that posterior uncertainties τ_i^{-1} are well calibrated under dynamic topology updates (pruning/rewiring every 150 steps, $k_{ij} = 1$).

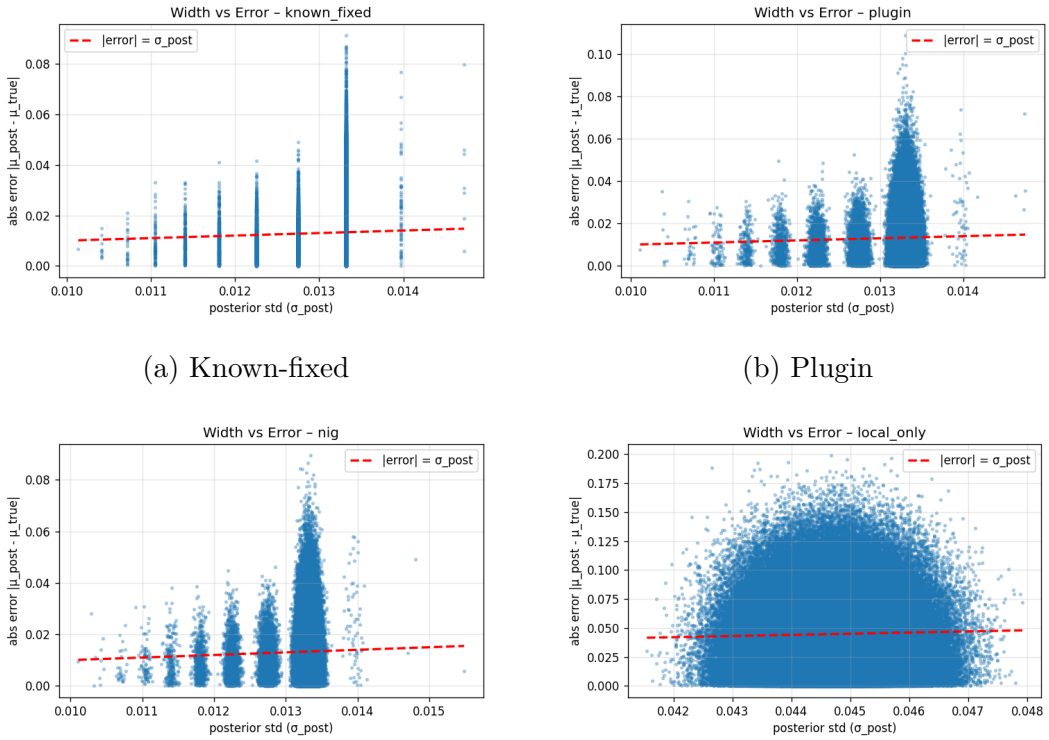
6.3.4 Width-Error analysis

To further assess the coherence between inferred uncertainty and empirical accuracy, the absolute estimation error of each agent was compared to its posterior standard deviation:

$$e_i = |\mu_i^{\text{post}} - \mu_i^*|, \qquad \sigma_i = \sqrt{\tau_i^{\text{post},-1}}.$$

In a perfectly calibrated Bayesian system, agents with low posterior inverse precision (i.e., high precision) are expected to exhibit smaller estimation errors on average, yielding a monotonic correspondence between σ_i and e_i. Equivalently, the scatter points should cluster around the identity relation $|\text{error}| = \sigma_{\text{post}}$.

Figure 6.3 reports this relation for the four inference modes. The empirical trend aligns with the theoretical expectation: agents with lower posterior standard deviation systematically achieve smaller absolute errors, while higher σ_{post} values correspond to broader estimation variability.



(a) Known-fixed

(b) Plugin

(c) NIG

(d) Local-only

Figure 6.3: Posterior standard deviation versus absolute estimation error. The red dashed line denotes the theoretical identity $|\text{error}| = \sigma_{\text{post}}$.

Known-fixed. The known-fixed configuration produces discrete vertical bands of σ_{post}. This effect directly stems from the assumption of homoscedasticity, which enforces a shared noise level across all agents and leads to identical precision update patterns for equal sample counts.

Plugin and NIG. Compared to the known-fixed mode, both adaptive formulations produce a slightly more irregular distribution of points, without the discrete band structure induced by fixed variance assumptions. Consequently the dispersion of $(\sigma_{\text{post}}, |\mu_i^{\text{post}} - \mu_i^*|)$ pairs is smoother and less discrete, reflecting the continuous adaptation of posterior variances over time. Agents with higher precision (smaller σ_{post}) still achieve proportionally lower mean errors, but with a more diffuse alignment due to the heterogeneity introduced by variance estimation.

Local-only. In contrast, the local-only baseline shows a nearly flat distribution of σ_{post} centered around a constant value ($\sim$0.045). Because all agents collect the same number of samples from identically distributed local processes slowly, the expected inverse correlation between uncertainty and error cannot fully emerge: the model lacks the diversity in σ_{post} needed to reveal the trend visible in the Bayesian cases. This explains the nearly uniform "dome-shaped" scatter in Fig. 6.3(d).

6.3.5 Z-score calibration

To assess the global statistical validity of the inferred uncertainties, the posterior estimates were standardized into Z-scores:

$$z_i = \frac{\mu_i^{\text{post}} - \mu_i^*}{\sqrt{\tau_i^{\text{post},-1}}}.$$

Under perfect calibration, these normalized residuals should follow a standard normal distribution:

$$z_i \sim \mathcal{N}(0, 1),$$

meaning that the mean of z_i is zero and its standard deviation equals one. This test is complementary to coverage analysis: while coverage evaluates the probability of credible intervals containing the truth, Z-score calibration assesses the *shape* and *scale* of the entire posterior uncertainty.

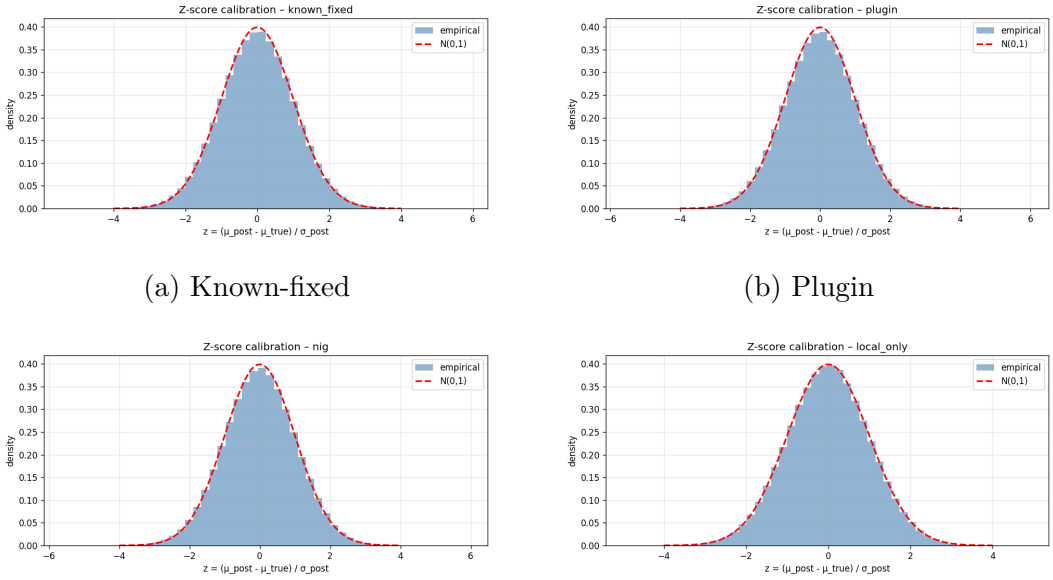
Since Z-scores inherently assume independence of the underlying estimates it should be further emphasized that consecutive posterior estimates are not statistically independent in time, but this is an unavoidable structural property of this type of inference on graphs (as discussed in Appendix A.3).

However, when averaged over a large population of agents operating on distinct local datasets, these correlations largely cancel out: the aggregated effect remains weak, and the resulting Z-score distributions are nearly indistinguishable from the ideal $\mathcal{N}(0, 1)$. The undercalibration observed in the global analysis therefore represents a residual signature of this structural dependence.

Known-fixed, Plugin, and NIG. All Bayesian fusion modes produce Z-score distributions that are essentially Gaussian, centered at zero and with unit scale. A very slight flattening of the central peak compared to the ideal $\mathcal{N}(0, 1)$ suggests marginally heavier tails: posterior inverse precisions are just below the empirical dispersion of the errors.

This means that the fused posteriors tend to be a bit overconfident, assigning slightly narrower uncertainty bands than would be optimal. Such behavior is entirely consistent with the $\sim$0.5-1% undercoverage observed in the global metrics.

Local-only. The baseline case displays a nearly perfect match with the standard normal curve, including a slightly sharper peak at $z = 0$. This indicates that, in the absence of social fusion, the empirical residuals align exactly with the theoretical Gaussian shape, confirming that local Bayesian updates remain perfectly calibrated when based on i.i.d. data and stable noise statistics.



(a) Known-fixed

(b) Plugin

(c) NIG

(d) Local-only

Figure 6.4: Z-score calibration histograms comparing empirical and ideal $\mathcal{N}(0,1)$.

Interpretation and theoretical implications

Overall, the close agreement between empirical and nominal coverage indicates that the posterior inverse precisions τ_i^{-1} (variances) provide an accurate quantification of the **true uncertainty** surrounding the latent parameters μ_i^*. In other words, the spread of each Gaussian posterior $\mathcal{N}(\mu_i^{\text{post}}, \tau_i^{-1})$ coincides with the actual dispersion of the true means across repeated simulations.

This alignment confirms that DBME, under uniform weights $k_{ij} = 1$, acts as an **empirically calibrated marginal estimator**: the inferred variances correctly represent the probability that the true value lies within a given credibility interval.

As the system evolves, each node's posterior uncertainty reflects the effective amount of information available from both its local observations and the independent contributions of its neighbors.

6.4 Adaptive weights and expected coverage degradation

The introduction of adaptive coefficients k_{ij} leads to two qualitatively different behaviors depending on whether the communication topology is fixed or dynamically updated through pruning and rewiring.

Static topology

When the graph remains fixed, each weight k_{ij} depends only on the current local posteriors of nodes i and j:

$$k_{ij} \propto \exp\left[-\frac{1}{2}\frac{(\mu_i^{\text{loc}} - \mu_j^{\text{loc}})^2}{\tau_{i,\text{loc}}^{-1} + \tau_{j,\text{loc}}^{-1} + \epsilon}\right],$$

where $\tau_{i,\text{loc}}^{-1}$ and $\tau_{j,\text{loc}}^{-1}$ denote the inverse precisions (variances) of the local posteriors. The network structure $\mathcal{E}_{\text{fixed}}$ is constant, so no feedback occurs between posterior updates and connectivity. As a consequence the joint likelihood of local and social evidence factorizes as

$$p(x_{\text{local}}, x_{\text{social}} \mid \mu, \mathcal{E}_{\text{fixed}}) = p(x_{\text{local}} \mid \mu)\, p(x_{\text{social}} \mid \mu, \mathcal{E}_{\text{fixed}}),$$

preserving conditional independence.

Under this regime, the Gaussian fusion rule remains analytically valid and the resulting posteriors are only marginally overconfident, as shown in Table 6.2.

Mode	**Coverage**$_{95}$	z**-mean**	z**-std**	**Within** 1σ
Known-fixed	0.9341	0.0012	1.0773	0.6333
Plugin	0.9335	0.0012	1.0794	0.6322
NIG	0.9339	0.0012	1.0777	0.6331
Local-only	**0.9496**	**0.0008**	**1.0013**	**0.6817**

Table 6.2: Static topology with adaptive weights coverage.

Dynamic topology

When pruning and rewiring are active, the edge set $\mathcal{E}(t)$ evolves as a deterministic function of the agents' posteriors. Each adaptive weight $k_{ij}(t)$ depends on local quantities that are simultaneously used to update the network itself:

$$k_{ij}(t) = f(\mu_i^{\text{loc}}(t-1), \mu_j^{\text{loc}}(t-1), \tau_{i,\text{loc}}^{-1}(t-1), \tau_{j,\text{loc}}^{-1}(t-1)).$$

As a consequence, the topology becomes a random variable driven by past posterior statistics. Therefore the joint likelihood must be marginalized over possible graphs:

$$p(x_{\text{local}}, x_{\text{social}} \mid \mu) = \int p(x_{\text{local}} \mid \mu)\, p(x_{\text{social}} \mid \mu, \mathcal{E})\, p(\mathcal{E} \mid x_{\text{local}}, \mu)\, d\mathcal{E},$$

breaking the independence assumption at the core of Gaussian fusion.

This recursive coupling between beliefs and topology produces systematic undercoverage, as summarized in Table 6.3.

Mode	**Coverage**$_{95}$	z**-mean**	z**-std**	**Within** 1σ
Known-fixed	0.9029	0.0017	1.1942	0.5726
Plugin	0.9016	0.0018	1.1987	0.5703
NIG	0.9024	0.0021	1.1956	0.5717
Local-only	**0.9496**	**0.0008**	**1.0013**	**0.6817**

Table 6.3: Dynamic topology with pruning and rewiring coverage drop

Overall, the contrast between Tables 6.2 and 6.3 confirms that the degradation of coverage is not inherent to the adaptive weighting itself, but to the feedback loop between posterior estimates and evolving topology. In static networks, calibration remains nearly ideal; in dynamic ones, the interplay between local evidence and structural adaptation induces a systematic overconfidence in the fused posteriors.

6.5 Agent-wise coverage analysis

While the global coverage curves in previous sections demonstrate the average calibration of DBME over the full population, they conceal the fine-grained variability that emerges at the level of individual agents. To expose this microscopic structure, the coverage statistics were computed separately for each node across the $N_{\text{runs}} = 150$ Monte Carlo repetitions, yielding an empirical coverage value per agent:

$$\widehat{C}_i = \frac{1}{N_{\text{runs}}} \sum_{r=1}^{N_{\text{runs}}} \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}\left[\mu_i^* \in (\mu_{i,r}^{\text{post}}(t) - z\sqrt{\tau_{i,r,\text{post}}^{-1}(t)}, \mu_{i,r}^{\text{post}}(t) + z\sqrt{\tau_{i,r,\text{post}}^{-1}(t)})\right]$$

where $z = 1.96$.

This procedure gives us a coverage distribution $\{\widehat{C}_i\}_{i=1}^{N=10^4}$ for each inference mode, providing a direct measure of how uncertainty is calibrated across agents.

Because the empirical coverage $\widehat{C}_i$ depends on the stabilization of each agent's posterior inverse precision $\tau_{i,\text{post}}^{-1}(t)$ (variance), the number of time steps T plays a critical role in the asymptotic calibration. Preliminary analysis suggest that increasing T beyond 2000 would likely improve coverage uniformity across agents, especially for peripheral nodes with fewer effective interactions.

As shown in Fig. 6.5, the four inference schemes produce nearly overlapping distributions centered near but not precisely on the nominal value 0.95. The local-only mode is marginally shifted upward, indicating a more stable tendency, that reflects the results analyzed in the other sections .

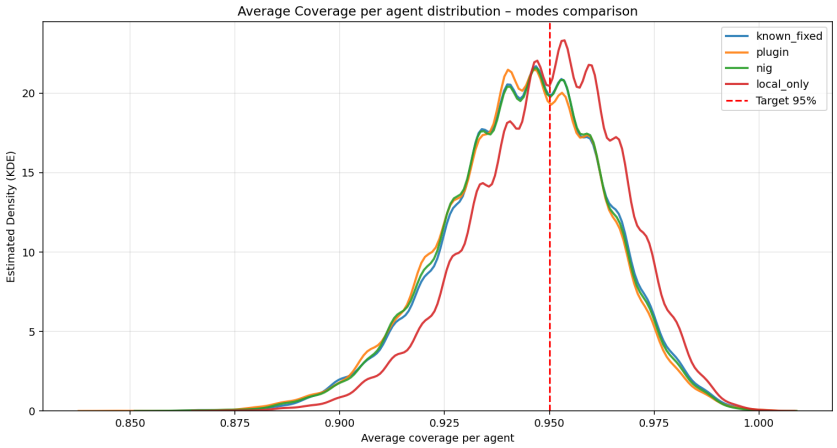


Figure 6.5: Distribution of the average coverage per agent across inference modes. The red dashed line marks the nominal 95% target. All four regimes cluster tightly around the theoretical value, confirming overall calibration at the agent level.

Coverage stability and inter-run variability

To assess the relationship between coverage accuracy and stability, the standard deviation of coverage across runs ($\mathrm{std}_r[\widehat{C}_{i,r}]$) was computed for each agent. Figures 6.6 display the corresponding hexbin maps, where each hexagon represents the density of agents sharing similar coverage mean and inter-run variability.

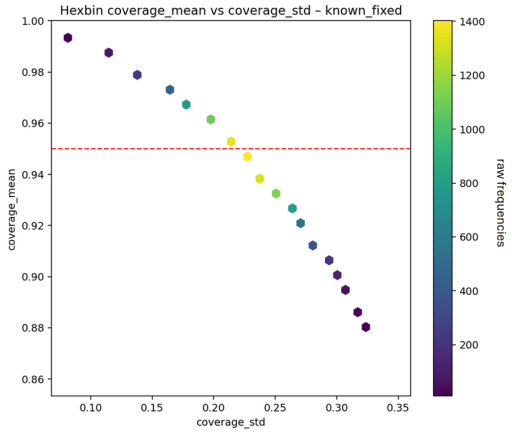
All models exhibit a clear negative correlation: as the inter-run variability increases, the mean coverage decreases from ~ 1.0 to ~ 0.85. The pattern is almost identical across inference modes, indicating that this relationship arises from sampling effects.

Agents with stable posteriors (low standard deviation across runs) achieve coverage near the nominal 0.95-1.0, while those with high variability tend to under-cover due to overconfident inverse precision estimates (underestimated variances).

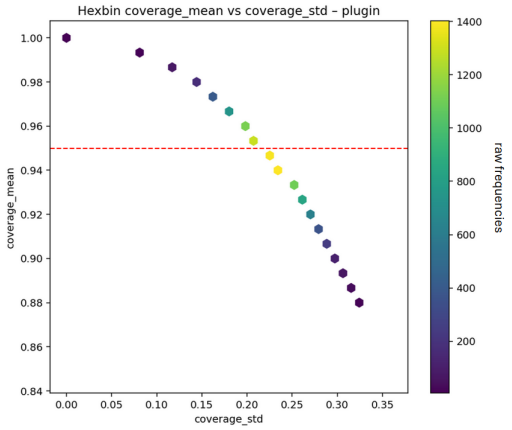
Interpretation

The agent-wise perspective reveals that DBME's calibration is not uniform in a trivial sense, but self-consistently distributed around the theoretical target. Each agent's coverage reflects the effective reliability of its local evidence: heterogeneous noise and finite-sample effects generate a continuum of confidence levels across the network.

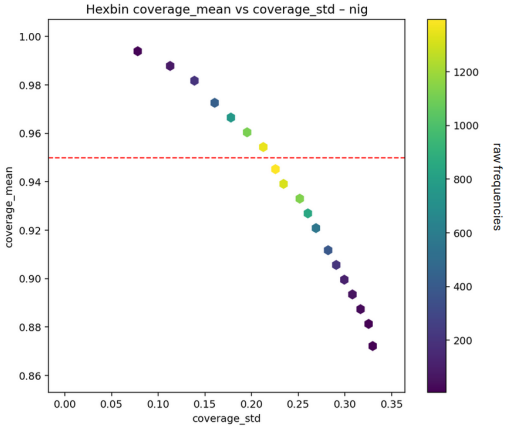
Overall the mean coverage of ~ 0.94 emerges as a population-level average of diverse microscopic behaviors.



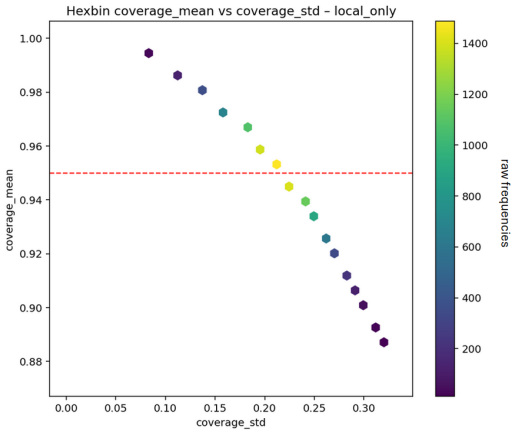
(a) Known-fixed



(b) Plugin



(c) NIG



(d) Local-only

Figure 6.6: Agent-wise relationship between average coverage and inter-run variability. Each hexagon represents the density of agents with similar (coverage_mean, coverage_std). The monotonic decline highlights that unstable agents (higher variability) systematically exhibit lower coverage.

Chapter 7

Experiments

7.1 Variants of DBME

To assess the behaviour of the Dynamic Bayesian Mean Estimation framework, we implemented four inference variants differing only in how local uncertainty is modeled (known-fixed, plugin, NIG, and local-only). All DBME modes follow the same Bayesian fusion rule and the same topological adaptation mechanism.

Experimental setting. The full configuration mirrors the parameter regime commonly adopted in the Scalable ColME literature [2]. In particular, the number of agents, the time horizon, and the admissible estimation tolerance have been chosen to match the scale and statistical regime used in those experiments:

$$\begin{aligned} N = 10^4, \qquad & T = 2000, \\ r = 10, \qquad & \varepsilon = 0.1, \\ T_0 = 50, \qquad & \sigma^2_{\text{true}} = 4.0, \\ \alpha_{\text{prune}} = 10^{-3}, \qquad & \alpha_{\text{rewire}} = 10^{-1}, \\ \Delta t_{\text{prune}} = 50, \qquad & \Delta t_{\text{rewire}} = 50, \\ k_{ij}\text{–weights}: \ \text{disabled}. & \end{aligned}$$

In particular the time horizon was fixed at $T = 2000$. Likewise, the tolerance $\varepsilon = 0.1$ and the network size $N = 10^4$ follow the same rationale, reproducing a large-scale operating regime where fluctuations and calibration properties can be observed at population level.

Parameter calibration. The baseline connectivity is determined by the average degree $r = 10$ of the initial Erdős–Rényi graph $G(N, p)$. This level of sparsity ensures effective information diffusion without introducing redundant communication paths.

The bootstrap horizon is limited to $T_0 = 50$, providing only a short local warm–up phase before social interaction begins. This exposes the system to initially noisy posteriors, making the emergence of consistent estimates primarily a consequence of network–level self-organization rather than extensive pre-observation.

Adaptive interaction coefficients k_{ij} are disabled in this experimental stage so that all neighbors influence social aggregation uniformly. This isolates the intrinsic Bayesian fusion behaviour from additional sources of adaptivity; later experiments show that enabling k_{ij} becomes beneficial when topological adaptation is unavailable.

Topological adaptation. Pruning and rewiring are both activated and executed every $\Delta t = 50$ iterations.

Pruning uses a conservative significance level,

$$\alpha_{\text{prune}} = 10^{-3},$$

so that an edge is removed only when the observed disagreement between two local posteriors is strong enough for the associated p–value to fall below 0.001.

Rewiring adopts a more permissive acceptance rule,

$$\alpha_{\text{rewire}} = 10^{-1},$$

which provides an effective compromise: it is selective enough to avoid introducing incompatible links too early, yet sufficiently loose to prevent communication deadlocks and allow low–degree or isolated nodes to recover connectivity.

Evaluation metrics. To monitor network learning dynamics we tracked: (i) error probability, (ii) fraction of incoherent links, (iii) average degree per node, (iv) posterior standard deviation.
These jointly capture convergence speed, topological coherence, and uncertainty reduction. We simulated a controlled scenario with two classes of data coming from normal distributions of parameters : $(\mu_1 = 1,\ \mu_2 = 0,\ \sigma_1^2 = \sigma_2^2 = 4)$. Each configuration was run independently 50 times, and results are reported with 95% confidence intervals computed under a Normal approximation.

7.1.1 Error probability dynamics

The panel of Figure 7.1 shows the evolution of the error probability, defined as

$$\text{Error probability}(t) \ = \ \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\Big\{\, |\mu_i(t) - \mu_i^{\text{true}}| \ \geq \ \epsilon \,\Big\}, \qquad \epsilon = 0.1.$$

This quantity measures the proportion of agents whose posterior mean differs from the true latent value by more than the specified tolerance, thus providing a direct indicator of global estimation accuracy. All collaborative variants display a clear exponential decay of the error probability, confirming the strong collective effect of Bayesian fusion over the network. The overall trend is stable, smooth, and consistent across multiple runs, demonstrating that the conservative topological configuration yields a highly regular convergence regime. Among the three collaborative models, the *NIG* configuration converges the fastest. By maintaining a joint posterior over

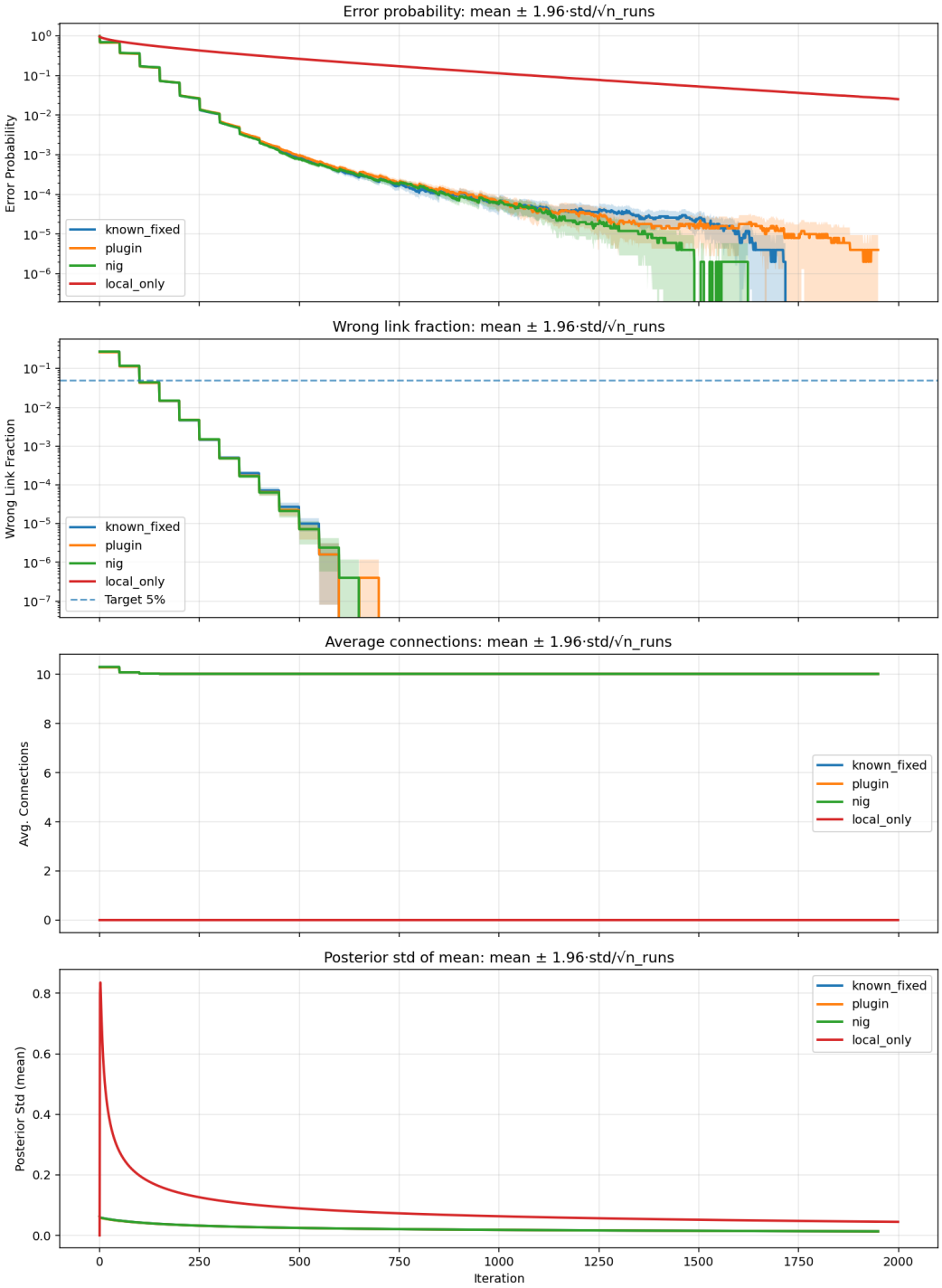


Figure 7.1: Averaged statistics over 50 independent runs with confidence bands. The initial 50 iterations corresponding to the bootstrap phase are omitted

both the mean and the variance, each agent automatically adapts its confidence to the observed noise level. This adaptive adjustment produces more decisive updates in high–certainty conditions and mitigates overconfidence in noisier regions, leading to a steeper decline of the error probability and earlier stabilization. However , the *Normal–Inverse–Gamma (NIG)* mode initially follows the same decay rate as the others but exhibits small oscillations in the late stage of convergence. This behaviour

arises because the locally estimated variance incorporates uncertainty on the noise parameter itself. When the network enters the rewiring phase, the corresponding hypothesis tests rely on these estimated variances, which fluctuate across agents. As a consequence, the statistical tests of compatibility occasionally misjudge the similarity between neighboring posteriors, causing transient reconnections or short cycles of link reactivation. These events appear in the curve as minor peaks and drops near the end of convergence, reflecting small topological readjustments. The *known-fixed* mode follows a slightly slower but more regular trajectory. Its updates are fully homogeneous across the network since all agents share the same fixed noise level. This uniformity produces a monotonic and noise-free decay curve, with minimal fluctuations in both the mean trend and the confidence band. The model is therefore statistically conservative but extremely stable, exhibiting smooth convergence toward the asymptotic error floor. The *plug–in* variant achieves similar overall accuracy but shows a slightly higher and more oscillatory error profile. Because the local empirical variance fluctuates during estimation, the corresponding aggregation weights vary over time, producing mild irregularities and broader confidence intervals. This leads to transient phases of under– and over–confidence that slow the final convergence slightly, even though the overall decay remains comparable to the other modes. The *local–only* configuration remains clearly separated from the collaborative ones, with an error probability nearly two orders of magnitude higher at steady state. This result emphasizes the necessity of information exchange: collective Bayesian fusion dramatically improves accuracy, while isolated learning remains dominated by sampling noise. Overall, all Bayesian models achieve extremely low error probabilities ($< 10^{-5}$ after 2000 iterations).

7.1.2 On the Posterior Variance in Collaborative Bayesian Estimation

The lower panel of Figure 7.1 shows the evolution of the posterior standard deviation of the mean for the four inference modes. The three collaborative configurations display closely aligned trajectories, converging toward a common asymptotic variance. The *local–only* configuration remains consistently higher, highlighting the limitations of independent learning. Here, updates follow a purely frequentist rule, without any bootstrap initialization, and the variance is computed directly from the first samples. The sharp initial peak arises because the estimator begins from an uninformative state: at $t = 0$ the information mass is effectively null, so the first variance estimate momentarily diverges before stabilizing as data accumulate. After a few iterations, the frequentist process reaches equilibrium, but its residual variance remains larger since no information is exchanged across agents.

In the collaborative cases, the brief bootstrap phase ($T_0 = 50$) regularizes the early updates, preventing any transient inflation and ensuring a monotonic contraction of uncertainty. The posterior variance here quantifies the true dispersion of the estimators around the latent means as demonstrated in Chapter 6, the empirical coverage of

the resulting confidence intervals aligns almost perfectly with the theoretical nominal levels, and the standardized residuals follow a $N(0,1)$ distribution. These results prove that the posterior standard deviation behaves as an unbiased and calibrated estimator of real uncertainty, validating its interpretation as the "true" posterior variance.

Overall, the variance dynamics reveal that collaborative Bayesian estimation achieves statistical efficiency and empirical calibration simultaneously. The posterior variance evolves smoothly, encodes the sampling variability of the process, and yields credibility intervals that are equivalent with their frequentist counterparts. The table 7.1 shows the metrics at the final iteration , its values show how the collaborative modes, apart from the plugin, generate estimates with almost zero frequentist uncertainty.

Table 7.1: Final metrics across estimation modes (mean $\pm$ 1.96·std/$\sqrt{n_{\text{runs}}}$).

Mode	**Error probability**	**Avg connections**	**Posterior std (mean)**
known_fixed	0.0 ± 0.0	10.021 ± 0.001	0.01330 ± 10^{-6}
plugin	$(4.0 \pm 5.0) \times 10^{-6}$	10.021 ± 0.001	0.01329 ± 10^{-6}
nig	0.0 ± 0.0	10.021 ± 0.001	0.01330 ± 10^{-6}
local_only	0.0253 ± 0.0004	0 ± 0	$0.0447 \pm 2 \times 10^{-6}$

In Appendix A.4 we show that the ratio between the local posterior variance and the variance produced by DBME coincides with the reduction that would be achieved by an equivalent centralized model with r agents. In particular, since DBME performs iterative Gaussian fusion across a connected cluster, its stationary posterior variance matches that of a hypothetical "super–agent" that aggregates all independent local posteriors at once. Thus, the ratio

$$\frac{\sigma^2_{\text{loc}}}{\sigma^2_{\text{DBME}}}$$

directly estimates the effective number of independent sources contributing to each agent's belief. In our experiments this value is approximately $K \approx 11$, consistent with the average degree of the graph and with the amount of information actually circulating through the network.

7.1.3 Wrong Link Fraction

The second panel of Figure 7.1 illustrates the evolution of the wrong link fraction, defined as

$$\text{Wrong link fraction}(t) \;=\; \frac{\#\{(i,j) \in E(t) :\; \mu_i^{\text{true}} \neq \mu_j^{\text{true}}\}}{|E(t)|},$$

where $E(t)$ denotes the set of active edges at iteration t. This metric quantifies the structural purity of the communication graph, measuring the proportion of links that erroneously connect agents belonging to different latent classes.
All collaborative inference modes display an almost identical and monotonically decreasing profile. The wrong link fraction drops exponentially by several orders of magnitude, quickly approaching zero within the first few hundred iterations. This behaviour confirms that the pruning mechanism acts in a stable and selective manner, removing only statistically incompatible edges while preserving intra-class connectivity.
The *known fixed*, *plug–in*, and *NIG* formulations are indistinguishable throughout the process. Despite relying on different local uncertainty models, all three variants produce posterior variances that evolve coherently across agents. As a result, the test that triggers pruning remain aligned across the different inference schemes, leading to nearly identical topological adjustments at every pruning-rewiring cycle.
After convergence, the active graph is composed exclusively of intra-class connections, yielding a sparse topology that mirrors the latent class organization. At the same time, as shown in the third panel of Figure 7.1, the average degree remains tightly concentrated around the target value $r = 10$, with occasional values slightly above ten, especially in the first iterations. This behavior is expected. Pruning removes only statistically inconsistent edges, while rewiring may introduce new compatible links to nodes that already satisfy $\deg(i,t) \geq r$. Crucially, the degree constraint is evaluated only for the initiating node i during the rewiring phase, and not for its selected neighbor j. Consequently, even if node j already exceeds the target degree, it can still receive additional connections. As a result, the network stabilizes near, but not exactly at,the prescribed degree.

7.2 Ablation Study: Effect of Removing Rewiring

To quantify the structural impact of the rewiring mechanism, we performed an ablation experiment in which adaptive reconnection was disabled. In this configuration, once a link is removed by the pruning rule, it is permanently deleted and no new edges are created. The objective is to isolate the intrinsic role of pruning and to verify whether the network can maintain connectivity without active link regeneration.
For clarity and reproducibility, the experiment was conducted exclusively under the *known-fixed* inference mode and the test was performed under the standard parameter setting defined before in section 7.1 , but disabling rewiring process.

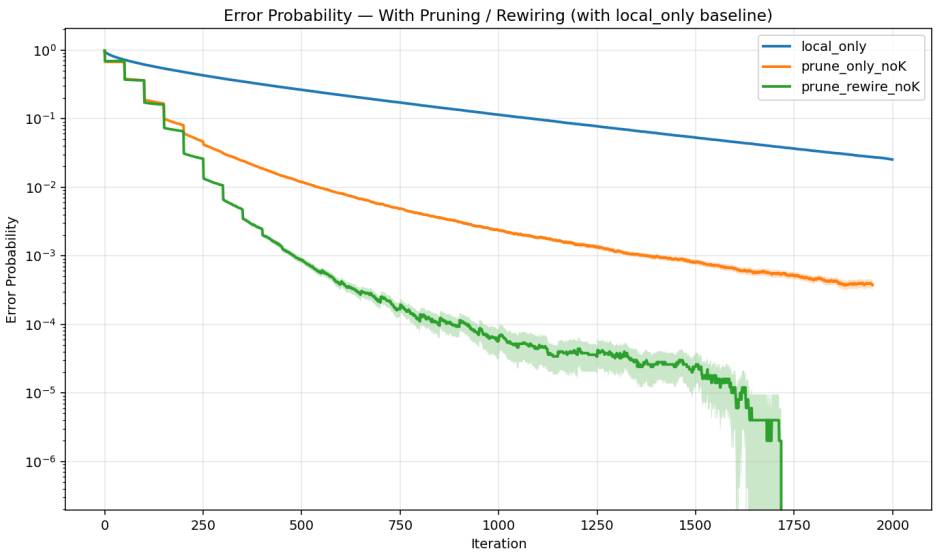


Figure 7.2: Error probability in absence of rewiring

As shown in Figure 7.2, the absence of rewiring produces a significant slow-down in convergence. While the *local-only* (blue curve) baseline exhibits a monotonic but very gradual decay, the *prune-only* (orange curve) configuration improves the estimation accuracy thanks to the selective elimination of incompatible links, yet its convergence saturates around 10^{-3}–10^{-4}. In contrast, the combined *prune+rewire* mechanism maintains a high level of information diffusion across the network, resulting in an exponential decrease of the error probability down to 10^{-6} within 2000 iterations. Structurally, the behaviour of the *prune–only* case confirms that pruning alone drives the average degree toward a stable equilibrium at approximately half of the nominal connectivity $r = 10$. This value coincides with the mean fraction of nodes belonging to one of the two latent clusters, meaning that each agent retains only intra-class edges while systematically discarding inter-class connections. The resulting network thus remains well divided but sparser, with an average degree of about 5, sufficient to preserve local consensus within clusters but insufficient to sustain global information flow across the entire population.

The inclusion of rewiring restores this balance: new edges are formed between statistically compatible agents, compensating for those lost to pruning and maintaining the nominal degree $r = 10$. Consequently, the network remains densely connected, and the collective estimation process continues to refine until full convergence is achieved.

7.3 Ablation Study: Effect of Social Weights without Topological Adaptation

To isolate the specific contribution of the adaptive social weights, we conducted a second ablation experiment in which both pruning and rewiring were disabled. In this setting, the communication graph remains fixed for the entire time horizon and

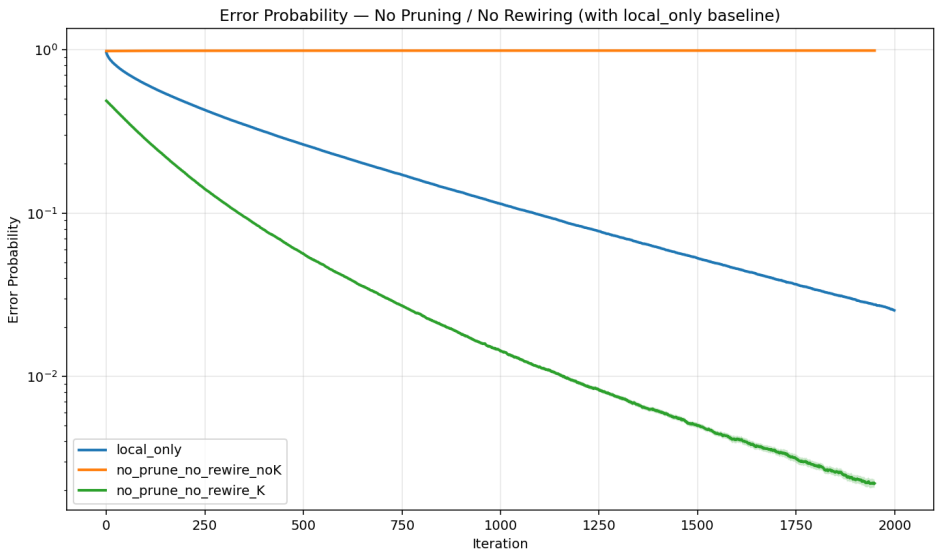


Figure 7.3: Error probability in absence of pruning and rewiring.

no structural filtering is available to remove incoherent edges, the only mechanism capable of attenuating the influence of statistically dissimilar neighbors is therefore the exponential weighting based on the pairwise divergence score defined in 3.3.3, i.e. the coefficients k_{ij}. We compared three configurations:

- **Local-only mode (blue curve)**: agents update exclusively from their own observations, with no social interaction;
- **Fixed graph with uniform social weights (orange curve)**: all agents communicate over an unchanged topology, but social influence is uniform ($k_{ij} = 1$ for all (i, j)), with no pruning or rewiring;
- **Fixed graph with adaptive social weights (green curve)**: all agents communicate over an unchanged topology, and social influence is modulated by adaptive coefficients k_{ij} based on statistical similarity.

The result, reported in Figure 7.3 shows interesting behaviors, first of all as always the *local-only* configuration (blue curve) shows the expected slow but monotone decrease, driven solely by the accumulation of local samples. When communication is enabled without any adaptive mechanism (fixed graph with uniform social weights, orange curve), the error probability does not improve and remains close to 10^0 throughout the entire time horizon. This indicates that mere information exchange is insufficient in a heterogeneous two-class scenario: incompatible agents continuously influence one another with equal strength, and the social layer systematically reintroduces bias, preventing class separation. By contrast, activating adaptive social weights (fixed graph with adaptive social weights, green curve) produces a qualitatively different behavior. Although the topology remains frozen, the fusion process becomes selective: connections between statistically dissimilar agents are not removed, but their contribution is exponentially suppressed at each update. Operationally, k_{ij}

acts as a form of *soft pruning* in parameter space: edges persist in the graph, but their effective statistical influence vanishes as soon as local beliefs diverge. As a consequence, the error probability resumes a clear decay and reaches values several orders of magnitude lower than both the local-only mode and the uniform-weight configuration ,performing very close to the pruning-only model discussed in the previous section.
To conclude pruning and rewiring operate in the structural domain, whereas k_{ij} acts directly on the estimates; the two mechanisms are therefore scenario dependent. When the graph cannot change (i.e. because of structural constraints), adaptive weights are necessary to suppress cross-class interference; when the graph is adaptive, structural pruning performs this role better, and the k_{ij} introduce only unwanted overconfidence as seen in 6.4 .

Consequences of the Adaptive Weights on the probabilistic model

This mechanism of soft-pruning makes the algorithm robust to partial failures, communication dropouts, or to the absence of a clear adjacency structure, allowing DBME to function in environments that are not strictly graph-based.
However, the use of adaptive weights also introduces a fundamental theoretical limitation. Since k_{ij} depends explicitly on the current local posterior of node i, the social contribution incorporated in the Bayesian update is no longer independent from the local evidence. As a result, the standard fusion step, which assumes that local and social components can be combined as conditionally independent terms , is no longer formally valid. The aggregation remains numerically stable and often accelerates convergence, but it ceases to represent a valid Bayesian combination of independent information sources.
From a modelling perspective, the update becomes a self–consistent but not strictly Bayesian process: each agent reweights the incoming information based on its own belief, creating a feedback loop between estimation and interaction strength. While this mechanism can sustain learning in the absence of structural adaptation, it alters the probabilistic interpretation of the posterior.

As discussed in Appendix A.2, this loss of independence breaks the formal validity of the fusion rule, and its consequences become evident in Chapter 6 where we can see how the posterior variances obtained under adaptive weighting lose their ability, even if only slightly , of calibration.

7.4 Non-identical distributions

To evaluate the robustness of DBME under heterogeneous observation models, we conducted a series of simulations in which the population was divided into four groups of agents, each following a distinct likelihood model: two Gaussian sources with means $\mu = \{1, 2\}$ and variances $\sigma^2 = \{3, 5\}$, an exponential source with mean 3, and a Poisson source with mean 4. The number of agents per class was identical (2500

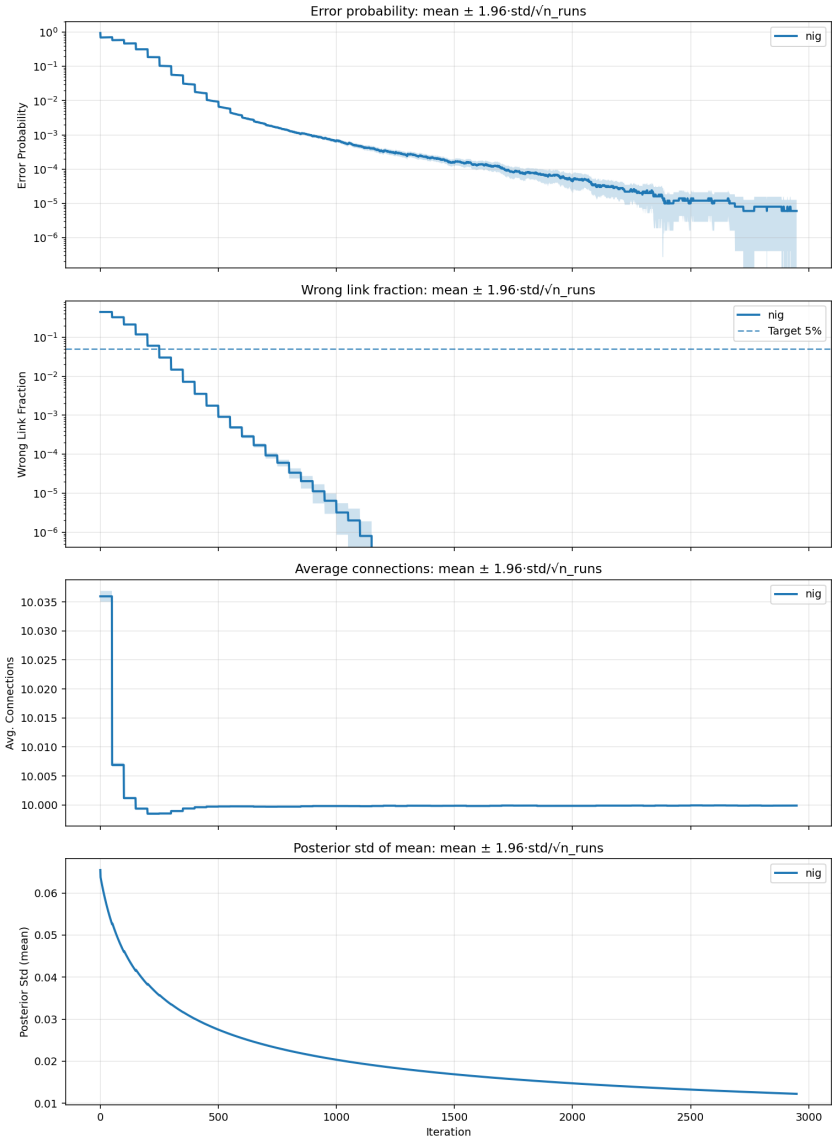


Figure 7.4: Evolution of global inference metrics for heterogeneous agent distributions (NIG mode, 50 runs). The initial 50 iterations corresponding to the bootstrap phase are omitted for clarity.

each), ensuring balanced representation. For the exponential and Poisson cases, the effective variance was automatically determined by the properties of the distribution, while for the Gaussian classes it was explicitly controlled. This configuration allowed the investigation of a mixed population where both the shape and the dispersion of the generating distributions differ.

All runs were executed under the *NIG* inference mode for consistency, with parameters identical to the baseline configuration: $N = 10^4$, $r = 10$, $\epsilon = 0.1$, $\alpha_{\text{pruning}} = 0.001$, $\alpha_{\text{rewiring}} = 0.1$ and $T = 3000$ iterations, selected higher than before to remedy for the slower convergence typical of heterogeneous generative processes. Each curve in Figure 7.4 represents the average over 50 independent runs with identical initial conditions.

As shown in Figure 7.4, the network preserves full statistical stability even under markedly non-identical distributions. The global error probability decreases exponentially over time, reaching values below 10^{-5} after approximately 2500 iterations. The wrong-link fraction follows a similar trend, collapsing to zero after the early pruning cycles. Meanwhile, the average number of connections remains constant at the nominal degree $r = 10$, confirming that rewiring effectively compensates for the links removed by pruning and maintains a balanced network topology.
The posterior standard deviation of the mean shows a smooth and monotonic contraction, indicative of proper uncertainty propagation within each class. Compared to homogeneous Gaussian setups, the convergence rate is slightly slower, reflecting the increased heterogeneity of the likelihood functions. Nevertheless, the posterior precision continues to improve steadily until full stabilization. From a structural perspective, the resulting graph splits into four perfectly balanced connected components of approximately equal size (2500 agents each), as confirmed by the final connectivity statistics:

$$\text{components} = 4, \quad \text{Largest} = 2500, \quad \text{Avg size} = 2500.0.$$

Each component corresponds to a distinct generative model, showing that the Bayesian pruning mechanism successfully enforces distributional coherence, while rewiring restores local connectivity within each homogeneous subpopulation.

7.5 Comparison with RRR-ColME

In this part of the experimental analysis, we evaluate the behaviour of DBME in direct comparison with the **Restricted Round-Robin ColME (RRR-ColME)** procedure , treated in [1] as a variant of the standard ColME algorithm. It is important to clarify that RRR-ColME is not a scalable variant in a strict sense, as it still exhibits worst-case complexity $\mathcal{O}(N^2)$ due to the global construction of compatibility relations over the entire agent set , we therefore adjusted the experiment setting to account for possible long simulation times. Below is a quick reminder of how RRR-ColME works.

7.5.1 Restricted Round-Robin ColME

Let $\mathcal{A} = 1, \ldots, N$ denote the set of agents. Each agent a collects observations $x_{a,t}$ and maintains the empirical mean

$$\hat{\mu} * a(t) = \frac{1}{n_a(t)} \sum *s = 1^{n_a(t)} x_{a,s}, \tag{7.1}$$

where $n_a(t)$ is the number of samples observed up to time t.
RRR-ColME relies on a PAC-based compatibility test. The confidence radius is

defined as

$$\beta_\gamma(n) = \sigma\sqrt{\frac{2}{n}\left(1+\frac{1}{n}\right)\log\left(\frac{\sqrt{n+1}}{\gamma}\right)}, \tag{7.2}$$

where $\gamma \in (0,1)$ controls the confidence level and σ^2 is the known sub-Gaussian variance. The optimistic distance between two agents a and b is then

$$d_t(a,b) = \left|\hat{\mu} * a(t) - \hat{\mu} * b(t)\right| - \beta * \gamma(n_a(t)) - \beta * \gamma(n_b(t)). \tag{7.3}$$

The compatibility set of agent a is defined as

$$\mathcal{C}_t(a) = b \in \mathcal{A} : d_t(a,b) \leq 0. \tag{7.4}$$

Under the restricted round-robin policy, at each iteration a single active agent a_t updates its estimate by aggregating information from its compatibility set:

$$\hat{\mu} * a_t(t+1) = \frac{\sum * b \in \mathcal{C}_t(a_t) n_b(t), \hat{\mu} * b(t)}{\sum * b \in \mathcal{C}_t(a_t) n_b(t)}. \tag{7.5}$$

All remaining agents retain their previous estimates.

7.5.2 Experimental setting

The comparison is conducted in a synthetic scenario with $N = 1000$ agents divided into two latent clusters. Each cluster generates observations according to

$$x_{a,t} \sim \mathcal{N}(\mu_c, \sigma^2),$$

with means $\mu_1 = 0$, $\mu_2 = 1$ and variance $\sigma^2 = 4$. At each iteration, every agent observes one new sample and updates its local statistics.

DBME configuration

DBME is executed with a bootstrap phase of $T_0 = 50$ iterations. The graph has average degree equal to $r = 15$. Pruning and rewiring are enabled every 20 iterations, with significance levels $\alpha_{\text{prune}} = 0.001$ and $\alpha_{\text{rewire}} = 0.1$, respectively. The inference mode is set to known-fixed, ensuring coherence with the fixed-variance assumption used in RRR-ColME.

RRR–ColME configuration

For RRR-ColME, we set $\gamma = 0.01$ to enforce a strict PAC confidence level, reducing the probability of early wrong-connection formation and limiting spurious cross-class aggregation, thus ensuring a cleaner and more interpretable comparison with DBME. At each iteration, each agent selects one peer from its compatibility set $\mathcal{C}_t(a)$ according to a restricted round-robin schedule and updates its estimate using the PAC-based aggregation rule described above. Figure 7.5 reports the Monte Carlo evolution of

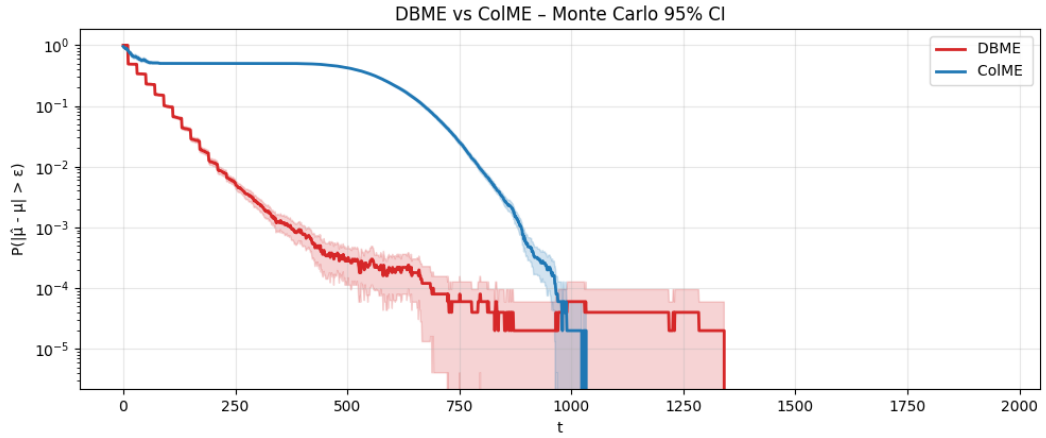


Figure 7.5: Monte Carlo evolution of the error probability $P(|\hat{\mu}_a(t) - \mu_a| > \varepsilon)$ with 95% confidence intervals for DBME and ColME.

the error probability $P(|\hat{\mu}_a(t) - \mu_a| > \varepsilon)$ with 95% confidence intervals for DBME and ColME under identical statistical conditions and across 50 runs.

A clear qualitative difference emerges in the transient regime where DBME exhibits a faster initial decay of the error probability, falling below 10^{-3} within approximately the first 400 iterations. This behavior can be attributed to its topology-adaptive information exchange, which enables multiple agents to update concurrently through social fusion.

After this initial phase, RRR-ColME presents a sharper asymptotic decrease and reaches very low error values earlier than DBME, this can be considered consistent with its globally informed aggregation mechanism, where each update leverages compatibility relations computed over the entire agent population, resulting in a more synchronized convergence once sufficient statistical confidence has been accumulated.

Beyond convergence speed, the two approaches differ in their structural and interpretative characteristics. While DBME operates in a fully decentralized manner, based exclusively on local neighborhood interactions and the evolving network topology, achieving an effective computational complexity of $\mathcal{O}(Nr)$ with $r \ll N$, RRR-ColME maintains a global compatibility evaluation, leading to a worst-case complexity of $\mathcal{O}(N^2)$, although updates are restricted to a round-robin schedule.

A further qualitative difference concerns the interpretation of uncertainty. In ColME, the uncertainty bound takes the form "with probability at least $1 - \delta$, the estimation error is bounded by ε", whereas DBME maintains a Bayesian posterior distribution over the unknown parameter μ_a, allowing credible intervals to be associated with each estimate.

This difference is reflected in the width of the Monte Carlo confidence bands. DBME displays wider intervals, particularly in the low-error regime, indicating a higher variability across independent runs. This behavior does not indicate instability, it reflects the dependence of DBME on the stochastic evolution of the interaction topology and adaptive link dynamics. In contrast, the narrower bands observed for RRR-ColME stem from its more deterministic convergence pattern.

Overall, these results suggest that DBME can be viewed as a viable alternative to ColME-based strategies, offering a decentralized and topology-adaptive framework with comparable asymptotic accuracy.

Chapter 8

Conclusions, Limitations and Future Directions

As the thesis has progressively illustrated, DBME offers a Bayesian based approach to distributed mean estimation, combining local inference with socially exchanged beliefs while keeping communication efficient on sparse graphs. The framework performs well in terms of calibration and adaptability, and provides a good alternative to PAC-based mechanisms such as ColME. Nevertheless, the clarity of its formulation also allows us to identify a number of structural limitations that become particularly evident once the model is analysed in depth.

8.1 Limitations of the Current Framework

A first limitation concerns the type of information used to drive pruning and rewiring. Compatibility is assessed solely on the basis of local posterior means: whenever two agents have statistically overlapping means, the system treats them as originating from the same latent process. This behavior persists even when their original class distributions variances differ substantially, with the result that DBME effectively performs clustering in the space of expectations. Agents generated by processes that share the same mean but differ in noise level are therefore merged into a single connected region, and the network loses the ability to detect heterogeneity.

As partially introduced in 3.3.2 this limitation persists even under the Normal–Inverse –Gamma formulation. Indeed , although the NIG model explicitly treats both the mean and the variance of the underlying population as random variables and provides a principled posterior over (μ_i, σ_i^2), only the marginal Gaussian belief on the mean, summarized by the pair $(\hat{\mu}_i, \tau_i)$, is actually exchanged during social communication. As consequence, the uncertainty associated with the variance parameter, despite being locally inferred and statistically meaningful, does not enter the mechanisms that govern pruning, rewiring, or social weighting. Decisions about structural compatibility are therefore driven solely by discrepancies in posterior means, while differences in dispersion remain invisible at the network level.

A second limitation is the use of fixed significance thresholds for the compatibility tests. The statistical regime of the network changes markedly over time: estimates are highly variable at the beginning and become increasingly concentrated as more observations accumulate. Static thresholds may therefore be too aggressive early on, leading to premature pruning or too permissive near convergence, causing unnecessary rewiring and topological instability.

Another structural constraint is that DBME, in its current form, lacks a hierarchical structure linking the priors of different agents. Each node initializes its prior independently, via local bootstrap statistics, and there is no mechanism for sharing higher-level information across groups of agents. This limits the model's ability to express population-level structure or latent clustering beyond simple mean-based grouping.

Finally, DBME propagates only Gaussian summaries. This assumption ensures analytical tractability and efficient computation, but it prevents the model from capturing more complex forms of uncertainty, such as heavy tails, skewness, or multimodality arising from heterogeneous sources or small-sample effects. The decentralized nature of the algorithm would still support richer distributions in principle, but their use would require revisiting the fusion and compatibility rules.

8.2 Future Directions

Several extensions follow naturally from the limitations discussed above. A first direction is to develop variance-aware compatibility rules, propagating the full NIG hyperparameters $(\mu_i, \kappa_i, \alpha_i, \beta_i)$ and comparing agents in the joint mean–variance space. This would allow DBME to detect heterogeneity that is invisible to its current mean-based mechanism.
For instance, variance–aware compatibility can be defined through joint tests on both expectation and dispersion. We already experimented with the following method, applicable specifically to the NIG case, which avoids heavy distributional assumptions while allowing the network to detect similarity in both posterior means and posterior variances of local populations:

- **Compatibility via overlapping posterior intervals:** two agents i and j are considered statistically compatible if both the posterior means and the posterior variances are mutually consistent. Concretely:

$$\underbrace{\frac{|\hat{\mu}_i - \hat{\mu}_j|}{\sqrt{\tau_i^{-1} + \tau_j^{-1}}} \leq z_\alpha^{(\mu)}}_{\text{Bayesian } z\text{-test on local posterior means}} \qquad \text{and} \qquad \underbrace{\mathcal{I}_i^{(\sigma^2)} \cap \mathcal{I}_j^{(\sigma^2)} \neq \emptyset}_{\text{overlap of credible intervals for variance}}$$

  where:

  - $\mathcal{I}_i^{(\sigma^2)}$ and $\mathcal{I}_j^{(\sigma^2)}$ are the $(1-\alpha)$ credible intervals obtained from the Inverse–Gamma posterior on each local variance.

These tests would allow the topology to react not only to discrepancies in expectation but also to differences in uncertainty, enabling DBME to distinguish agents with identical means but heterogeneous noise levels.
A second direction involves adaptive significance thresholds. A more flexible alternative would be to introduce a sort of statistical annealing, letting the thresholds evolve dynamically with time. For example, the pruning level could follow:

$$\alpha_{\text{prune}}(t) = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\, e^{-t/T_s},$$

becoming gradually stricter as the system stabilizes. Similarly, the rewiring threshold could increase according to:

$$\alpha_{\text{rewire}}(t) = \alpha_{\max}^{(\text{rew})} - (\alpha_{\max}^{(\text{rew})} - \alpha_{\min}^{(\text{rew})})\, e^{-t/T_r}$$

Such schedules would balance early exploration with late stability, though their explicit time dependence would require separate analysis to establish convergence and calibration guarantees.

A third extension involves the propagation of richer belief distributions. Replacing Gaussian summaries with Student-t, Gaussian mixture, or variational approximations would allow DBME to represent more complex uncertainty structures.

Finally, in large-scale systems or embedded environments, it may be useful to incorporate energy- and communication-aware constraints. Budgeted communication, selective neighbor querying, or cost-aware rewiring strategies would make DBME suitable for real-world sensor networks, robotic platforms, or IoT infrastructures.

8.3 Concluding Remarks

DBME provides a decentralized and statistically supported framework for distributed mean estimation. The limitations highlighted in this section point toward a broader research direction: the development of self-organizing Bayesian networks capable of reasoning jointly over means, variances, and structural dynamics, while preserving the decentralization and tractability that motivate the design of DBME.

All simulations, experimental protocols, and implementations discussed throughout the thesis are publicly available in the accompanying repository, which provides full reproducibility and serves as a reference platform for future extensions of the method. [1]

[1] `https://github.com/ValeXpoli/DBME-Dynamic----Bayesian-Mean-Estimation-`

Appendix A

Bayesian Update

A.1 Pairwise divergence statistic

Consider two agents i, j with local Gaussian posteriors

$$\mu_i^{\text{loc}}(t) \sim \mathcal{N}(\mu_i, \tau_{i,\text{loc}}^{-1}(t)), \qquad \mu_j^{\text{loc}}(t) \sim \mathcal{N}(\mu_j, \tau_{j,\text{loc}}^{-1}(t)).$$

The difference is distributed as

$$\mu_i^{\text{loc}}(t) - \mu_j^{\text{loc}}(t) \sim \mathcal{N}\Big(\mu_i - \mu_j, \tau_{i,\text{loc}}^{-1}(t) + \tau_{j,\text{loc}}^{-1}(t)\Big).$$

The standardized squared distance is

$$z_{ij}^2(t) = \frac{\left(\mu_i^{\text{loc}}(t) - \mu_j^{\text{loc}}(t)\right)^2}{\tau_{i,\text{loc}}^{-1}(t) + \tau_{j,\text{loc}}^{-1}(t)},$$

which under $\mathcal{H}_0 : \mu_i = \mu_j$ follows a χ_1^2 distribution.
We define the continuous divergence

$$\ell_{ij}(t) = \frac{1}{2} z_{ij}^2(t) = \frac{\left(\mu_i^{\text{loc}}(t) - \mu_j^{\text{loc}}(t)\right)^2}{2\left[\tau_{i,\text{loc}}^{-1}(t) + \tau_{j,\text{loc}}^{-1}(t)\right]}.$$

This statistic acts as a likelihood–ratio–equivalent distance and is used to define adaptive interaction weights $k_{ij}(t) = \exp(-\ell_{ij}(t))$.

A.2 Independence of local and social beliefs (three-node toy case)

Let agents $1, 2, 3$ possess independent local posteriors

$$\mu_k^{\text{loc}} \sim \mathcal{N}(m_k, \tau_k^{-1}), \qquad k \in \{1, 2, 3\}.$$

Agent 1 forms its social belief as

$$\mu_1^{\text{soc}} = \frac{\tau_2 \mu_2^{\text{loc}} + \tau_3 \mu_3^{\text{loc}}}{\tau_2 + \tau_3}.$$

Since μ_1^{loc} does not enter this expression and is independent from $\mu_2^{\text{loc}}, \mu_3^{\text{loc}}$, we have

$$\mu_1^{\text{loc}} \perp\!\!\!\perp \mu_1^{\text{soc}}.$$

If weights depend on μ_1^{loc} (e.g. through adaptive k_{ij}), this independence no longer holds, explaining the coupling introduced by the adaptive mechanism in heterogeneous settings.

A.3 Correlation between fused posterior estimates

In DBME, each agent maintains a local posterior belief $\mu_i^{\text{loc}}(t)$, constructed exclusively from its private data stream. Under the standing assumption that local observations are independent across nodes, these local posteriors remain statistically independent across agents at every time step:

$$\text{Cov}(\mu_i^{\text{loc}}(t),\, \mu_j^{\text{loc}}(t)) = 0, \qquad i \neq j.$$

The situation changes once social aggregation and repeated Bayesian fusion are introduced. The fused posterior beliefs $\mu_i^{\text{post}}(t)$ are no longer independent across the network. Through iterative exchange and feedback, information propagates along multiple paths, so that agents progressively incorporate evidence that is partially shared with their neighbors. As a result, their posterior estimators become correlated:

$$\text{Cov}(\mu_i^{\text{post}}(t),\, \mu_j^{\text{post}}(t)) \neq 0, \qquad j \in \mathcal{N}(i).$$

This correlation is not a defect of the algorithm but an unavoidable structural property of decentralized inference on graphs. Since each posterior is recursively built from previously aggregated beliefs, any two agents that share information paths (direct or indirect) will eventually exhibit statistically dependent posterior estimates. The dependence therefore persists even after the topology has converged and the clustering has stabilized.

A.4 Asymptotic consistency of DBME

Under ideal conditions, the DBME estimator is asymptotically consistent. Assume that, for a given agent i, the pruning–rewiring mechanism converges to a stable neighborhood $\mathcal{N}^*(i)$ composed exclusively of agents sharing the same latent mean, i.e.

$$\mu_j^* = \mu_i^* \quad \forall j \in \mathcal{N}^*(i),$$

and that all local data streams are independent with finite variance. In this regime, the network has effectively decomposed into homogeneous clusters and no wrong links persist in the limit.

ATTENTION: It is important to emphasize, however, that such convergence of the topology is not guaranteed in a strict theoretical sense. Pruning and rewiring are heuristic, data–driven mechanisms, and the stabilization toward perfectly homogeneous neighborhoods should therefore be interpreted as a plausible structural outcome supported by empirical evidence shown in 7 rather than as a formally proven property of the algorithm. The consistency analysis that follows is thus conditional on this stability assumption, which provides a meaningful but idealized framework for understanding the asymptotic behavior of DBME.

Let $\mu_i^{\text{post}}(t)$ and $\tau_{i,\text{post}}(t)$ denote the posterior mean and precision produced by DBME at time t. As the number of local observations grows,

$$N_i(t) \longrightarrow \infty,$$

and the network structure stabilizes, we show that

$$\mu_i^{\text{post}}(t) \xrightarrow{P} \mu_i^*, \qquad \text{Var}(\mu_i^{\text{post}}(t)) = \tau_{i,\text{post}}^{-1}(t) \longrightarrow 0.$$

Step 1: local Bayesian consistency

Consider first the known–variance mode (the other local modes converge to the same limit by standard Bernstein–von Mises arguments). After the bootstrap phase, agent i collects $N_i(t)$ incremental observations with true mean μ_i^* and variance σ^2, and maintains the sufficient statistics

$$S_i(t) = \sum_{s>T_0} x_{i,s}, \qquad N_i(t) = t - T_0.$$

With a Gaussian prior fixed at T_0, the local posterior at time t is

$$p_i^{\text{loc}}(\mu) = \mathcal{N}(\mu_i^{\text{loc}}(t),\, \tau_{i,\text{loc}}^{-1}(t)),$$

where, up to negligible prior terms,

$$\mu_i^{\text{loc}}(t) = \frac{S_i(t)}{N_i(t)}, \qquad \tau_{i,\text{loc}}(t) = \frac{N_i(t)}{\sigma^2}, \qquad \tau_{i,\text{loc}}^{-1}(t) = \frac{\sigma^2}{N_i(t)}.$$

By the law of large numbers and the central limit theorem,

$$\mu_i^{\text{loc}}(t) \xrightarrow{P} \mu_i^*, \qquad \text{Var}(\mu_i^{\text{loc}}(t)) = \frac{\sigma^2}{N_i(t)} \longrightarrow 0,$$

so each local posterior is consistent and asymptotically concentrated around μ_i^*.

Step 2: homogeneous cluster and independent local posteriors

Fix an agent i and its limiting homogeneous neighborhood $\mathcal{N}^*(i)$. By assumption,

$$\mu_j^* = \mu_i^* \quad \forall j \in \mathcal{N}^*(i),$$

and local data streams are independent across nodes. Hence, for each $j \in \mathcal{N}^*(i)$, the local posterior

$$p_j^{\text{loc}}(\mu) = \mathcal{N}(\mu_j^{\text{loc}}(t),\, \tau_{j,\text{loc}}^{-1}(t))$$

is independent of all others at fixed t, and satisfies

$$\mu_j^{\text{loc}}(t) \xrightarrow{P} \mu_i^*, \qquad \text{Var}(\mu_j^{\text{loc}}(t)) = \tau_{j,\text{loc}}^{-1}(t) \approx \frac{\sigma^2}{N_j(t)}.$$

In the symmetric setting used in the main experiments, all agents in a cluster accumulate approximately the same number of samples, so that $N_j(t) \approx N_i(t)$ and therefore

$$\tau_{j,\text{loc}}(t) \approx \frac{N_i(t)}{\sigma^2} \quad \text{for all } j \in \mathcal{N}^*(i) \cup \{i\}.$$

Step 3: Gaussian fusion within a homogeneous cluster

Once the topology has stabilized, the social belief at node i is computed as precision–weighted fusion of the local posteriors in its neighborhood (see Section 2.6.1):

$$p_i^{\text{soc}}(\mu) \propto \prod_{j \in \mathcal{N}^*(i)} \mathcal{N}(\mu;\, \mu_j^{\text{loc}}(t),\, \tau_{j,\text{loc}}^{-1}(t)) = \mathcal{N}(\mu;\, \mu_i^{\text{soc}}(t),\, \tau_{i,\text{soc}}^{-1}(t)),$$

with

$$\tau_{i,\text{soc}}(t) = \sum_{j \in \mathcal{N}^*(i)} \tau_{j,\text{loc}}(t), \qquad \mu_i^{\text{soc}}(t) = \frac{\sum_{j \in \mathcal{N}^*(i)} \tau_{j,\text{loc}}(t)\, \mu_j^{\text{loc}}(t)}{\sum_{j \in \mathcal{N}^*(i)} \tau_{j,\text{loc}}(t)}.$$

Because all $\mu_j^{\text{loc}}(t)$ are unbiased estimators of the same mean μ_i^*, their precision–weighted average $\mu_i^{\text{soc}}(t)$ is also unbiased and satisfies

$$\mu_i^{\text{soc}}(t) \xrightarrow{P} \mu_i^*.$$

Moreover, using independence and the variance formula for a weighted mean,

$$\text{Var}(\mu_i^{\text{soc}}(t)) = \Big(\sum_{j \in \mathcal{N}^*(i)} \tau_{j,\text{loc}}(t) \Big)^{-1}.$$

In the symmetric case $\tau_{j,\text{loc}}(t) \approx N_i(t)/\sigma^2$, if we denote $K_i = |\mathcal{N}^*(i)|$ the number of neighbors in the limiting cluster of i, we obtain

$$\tau_{i,\text{soc}}(t) \approx K_i\, \frac{N_i(t)}{\sigma^2}, \qquad \text{Var}(\mu_i^{\text{soc}}(t)) \approx \frac{\sigma^2}{K_i\, N_i(t)}.$$

Step 4: Final fusion and effective centralized rate

The final posterior at node i is obtained by fusing the local and social beliefs, again via the Gaussian product rule:

$$p_i^{\text{post}}(\mu) \propto p_i^{\text{loc}}(\mu)\, p_i^{\text{soc}}(\mu) = \mathcal{N}(\mu;\, \mu_i^{\text{post}}(t),\, \tau_{i,\text{post}}^{-1}(t)),$$

with

$$\tau_{i,\text{post}}(t) = \tau_{i,\text{loc}}(t) + \tau_{i,\text{soc}}(t), \qquad \mu_i^{\text{post}}(t) = \frac{\tau_{i,\text{loc}}(t)\,\mu_i^{\text{loc}}(t) + \tau_{i,\text{soc}}(t)\,\mu_i^{\text{soc}}(t)}{\tau_{i,\text{post}}(t)}.$$

Both contributions are asymptotically unbiased with respect to μ_i^*, hence $\mu_i^{\text{post}}(t)$ is a consistent estimator as well:

$$\mu_i^{\text{post}}(t) \xrightarrow{P} \mu_i^*.$$

For the variance, in the symmetric configuration we have

$$\tau_{i,\text{loc}}(t) \approx \frac{N_i(t)}{\sigma^2}, \qquad \tau_{i,\text{soc}}(t) \approx K_i\,\frac{N_i(t)}{\sigma^2},$$

which yields

$$\tau_{i,\text{post}}(t) \approx (K_i+1)\,\frac{N_i(t)}{\sigma^2}, \qquad \tau_{i,\text{post}}^{-1}(t) \approx \frac{\sigma^2}{(K_i+1)\,N_i(t)}.$$

In other words,

$$\text{Var}(\mu_i^{\text{post}}(t)) \sim \frac{\sigma^2}{(|\mathcal{N}^*(i)|+1)\,N_i(t)} \;\longrightarrow\; 0 \qquad \text{as } t \to \infty.$$

Conclusion

Therefore, under stable clustering and sustained information flow, DBME achieves consistency in the classical sense: the posterior mean $\mu_i^{\text{post}}(t)$ converges in probability to the true parameter μ_i^*, and the posterior uncertainty shrinks asymptotically at the same rate as an ideal centralized estimator that would fuse $|\mathcal{N}^*(i)|+1$ independent local posteriors.

Centralized benchmark and effective cluster size

While the asymptotic consistency result relies on the assumption that the pruning–rewiring dynamics converge to stable and homogeneous neighborhoods, this hypothesis is not proven formally and must be regarded as an idealized structural condition. It is therefore important to verify whether the effective behavior of DBME observed in experiments is compatible with this assumption.

To this end, we compare the empirical posterior variance of DBME with the one

obtained by an ideal centralized estimator operating on the same data. Assume K agents observe independent samples from a common Gaussian source

$$x_{i,t} \sim \mathcal{N}(\mu^*, \sigma^2),$$

and each forms a local posterior

$$p_i^{\text{loc}}(\mu) = \mathcal{N}(\mu_i^{\text{loc}}, \tau_{i,\text{loc}}^{-1}), \qquad \tau_{i,\text{loc}} = \frac{T}{\sigma^2}.$$

This yields a local posterior standard deviation

$$\sigma_{\text{loc}} = \frac{\sigma}{\sqrt{T}}.$$

If a centralized super-agent fuses all K local posteriors as independent Gaussian factors, the resulting posterior has precision

$$\tau_{\text{glob}} = \sum_{i=1}^{K} \tau_{i,\text{loc}} = \frac{KT}{\sigma^2},$$

and standard deviation

$$\sigma_{\text{glob}} = \frac{\sigma}{\sqrt{KT}} = \frac{\sigma_{\text{loc}}}{\sqrt{K}}.$$

Hence, fusing K independent sources reduces uncertainty by a factor $\sqrt{K}$.

This relation can be inverted to estimate the effective number of statistically independent contributors to the posterior of each agent:

$$K = \left(\frac{\sigma_{\text{loc}}}{\sigma_{\text{post}}} \right)^2.$$

In our experiments we observe

$$\sigma_{\text{loc}} \approx 0.0447, \qquad \sigma_{\text{post}} \approx 0.0133,$$

which gives

$$K \approx 11.$$

This value closely matches the target degree $r = 10$ and the result from the consistency proof. Although this does not constitute a proof of topological convergence, it provides empirical evidence that, in practice, DBME behaves as if each agent were aggregating information from a stable cluster of approximately r statistically homogeneous neighbors. In practical terms, this means that once the network stabilizes, each agent behaves as if it were receiving information from an effective set of K independent peers, lending empirical credibility to the idealized asymptotic framework discussed above.

Bibliography

[1] Mahsa Asadi, Aurélien Bellet, Odalric-Ambrym Maillard, and Marc Tommasi. "Collaborative algorithms for online personalized mean estimation". In: *arXiv preprint arXiv:2208.11530* (2022) (cit. on pp. 1, 65).

[2] Franco Galante, Giovanni Neglia, and Emilio Leonardi. "Scalable Decentralized Algorithms for Online Personalized Mean Estimation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 39. 16. 2025, pp. 16699–16707 (cit. on pp. 2, 13, 55).

[3] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, 2002 (cit. on p. 4).

[4] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018 (cit. on p. 4).

[5] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. 3rd. Boca Raton, FL: CRC Press, 2013. ISBN: 9781439840955 (cit. on pp. 4, 20).

[6] Patrick Billingsley. *Probability and Measure*. 3rd. New York: John Wiley & Sons, 1995. ISBN: 9780471007104 (cit. on pp. 12, 20).

[7] William Feller. *An Introduction to Probability Theory and Its Applications, Volume I*. 3rd. New York: John Wiley & Sons, 1968 (cit. on pp. 12, 16).

[8] Rudolf E. Kalman. *A new approach to linear filtering and prediction problems*. Vol. 82. 1. ASME, 1960, pp. 35–45 (cit. on p. 14).

[9] Paul Erdős and Alfréd Rényi. "On Random Graphs I". In: *Publicationes Mathematicae* 6 (1959), pp. 290–297 (cit. on p. 16).

[10] Béla Bollobás. *Random Graphs*. 2nd. Cambridge, UK: Cambridge University Press, 2001. DOI: 10.1017/CBO9780511814068 (cit. on p. 16).

[11] Paul Erdős and Alfréd Rényi. "On the Evolution of Random Graphs". In: *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960), pp. 17–60 (cit. on p. 16).

[12] Thomas P. Minka. "Expectation Propagation for Approximate Bayesian Inference". In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI)*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 2001, pp. 362–369. URL: https://tminka.github.io/papers/ep/minka-ep-uai.pdf (cit. on p. 20).

[13] Manfred Opper and Cédric Archambeau. "The Variational Gaussian Approximation Revisited". In: *Neural Computation* 21.3 (2009), pp. 786–792. DOI: 10.1162/neco.2008.08-07-592 (cit. on p. 20).

[14] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1998. ISBN: 978-0-521-78450-4. DOI: 10.1017/CBO9780511802256 (cit. on p. 21).

[15] Carlo Emilio Bonferroni. *Teoria statistica delle classi e calcolo delle probabilità*. Studi in Onore di G. Morera. Firenze: Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936, pp. 1–62 (cit. on p. 28).