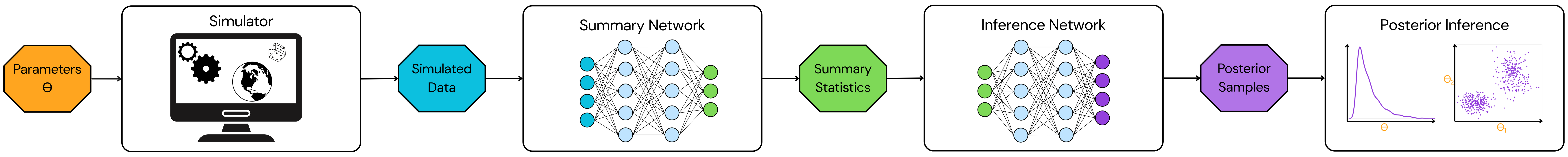
From Mice to Trains: Amortized Bayesian Inference on Graph Data

technische universität dortmund

Svenja Jedhoff Paul-Christian Bürkner



ABI workflow

- Amortized Bayesian Inference uses a neural network trained on many simulated (θ, x) pairs to learn a direct mapping from data x (and the prior) to an approximate posterior $p(\theta \mid x)$.
- Unlike per-dataset MCMC, the training cost is paid once; afterwards, inference for new datasets is near-instant.
- It's **likelihood-free** and **simulation-based**: you only need a simulator for your generative model, so that no closed-form likelihood is required.
- In our setup, working with the python library **BayesFlow**, a summary network is combined with an inference network:
  - The **summary network** compresses the data in (sufficient) summary statistics.
  - The **inference network** maps the summary statistics to an approximate posterior distribution via a generative neural network such as a normalizing flow.

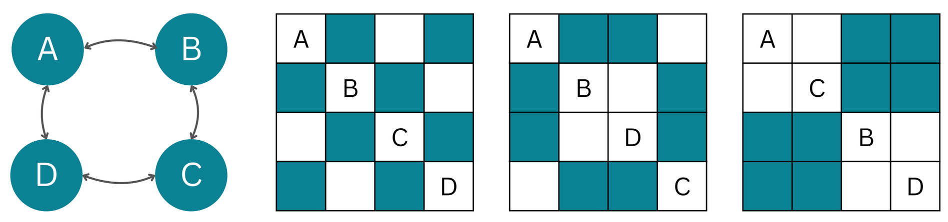
ABI on Graph Data

Why Graph Data?

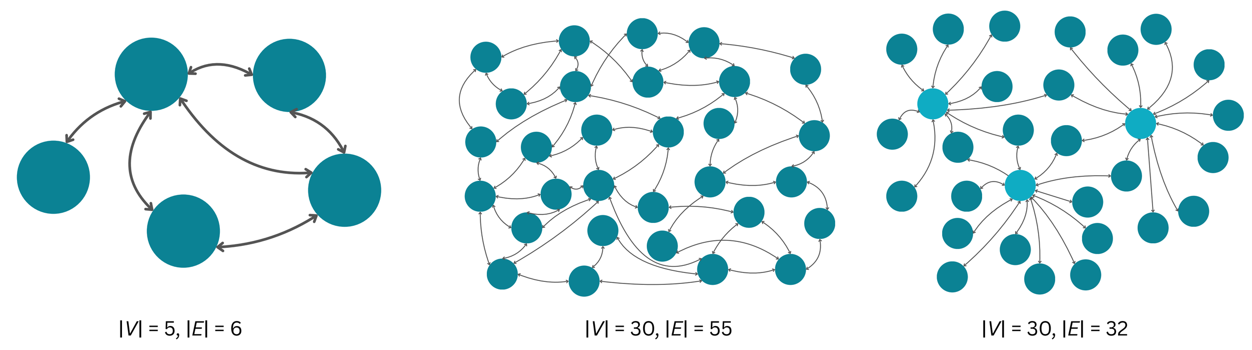
- Graphs are a universal representation for relational systems where entities interact (e.g., people, molecules, devices, documents) and edges encode dependencies, flows, or constraints.
- Graph structure captures non-local effects: local changes (failures, interventions, shocks) can propagate through paths, hubs, and communities, creating system-wide impacts.
- Graph-based ABI enables fast inference and scenario testing: once trained, amortized posteriors can be queried repeatedly for new observations, perturbations, or partial graph changes without re-running expensive simulation-based inference from scratch.
- Uncertainty-aware predictions on graphs improve robustness and trust: posteriors provide calibrated confidence for downstream decisions like ranking, anomaly detection, and risk-sensitive optimization.

Key Challenges

- The summary network in the ABI workflow must be able to handle graph structured data.
- **Permutation symmetry** and **isomorphism**s must be handled to avoid label-induced multi-modality and automorphism-driven non-identifiability.



- **Variable size** and **sparsity** across graphs, including wide ranges in $|V|$ and $|E|$ and heavy-tailed degrees, complicate batching, memory usage, and statistical efficiency.



- **Scalability** and **evaluation** remain challenging since expressive encoders and flows must be balanced with tractable training and validated using SBC, coverage, and posterior predictive checks.

Experiment: Train Scheduling

- Fixed railway network represented as a graph; each node is a track section with default traversal time $t_{\text{default}}(i)$. Each section can accommodate at most one train at a time.
- Four trains move on the graph according to a predefined schedule that may vary between scenarios.
- The simulator introduces a random delay in 10 % of train journeys, which can cause unplanned conflicts for track occupancy and additional delays: $T_{\text{train } j, \text{ section } i} = t_{\text{default}}(i) + \mathbb{I}_{\{x=1\}} \cdot \delta$, $x \sim \mathrm{Ber}(p = 0.1), \quad \delta \sim \Gamma(\alpha = 5, \lambda = 0.5)$
- Quantity of interest: the total travel time of each of the four trains $\Rightarrow$ Neural likelihood estimation
- Summary network: Set Transformer, Inference network: Coupling Flow with spline transformation

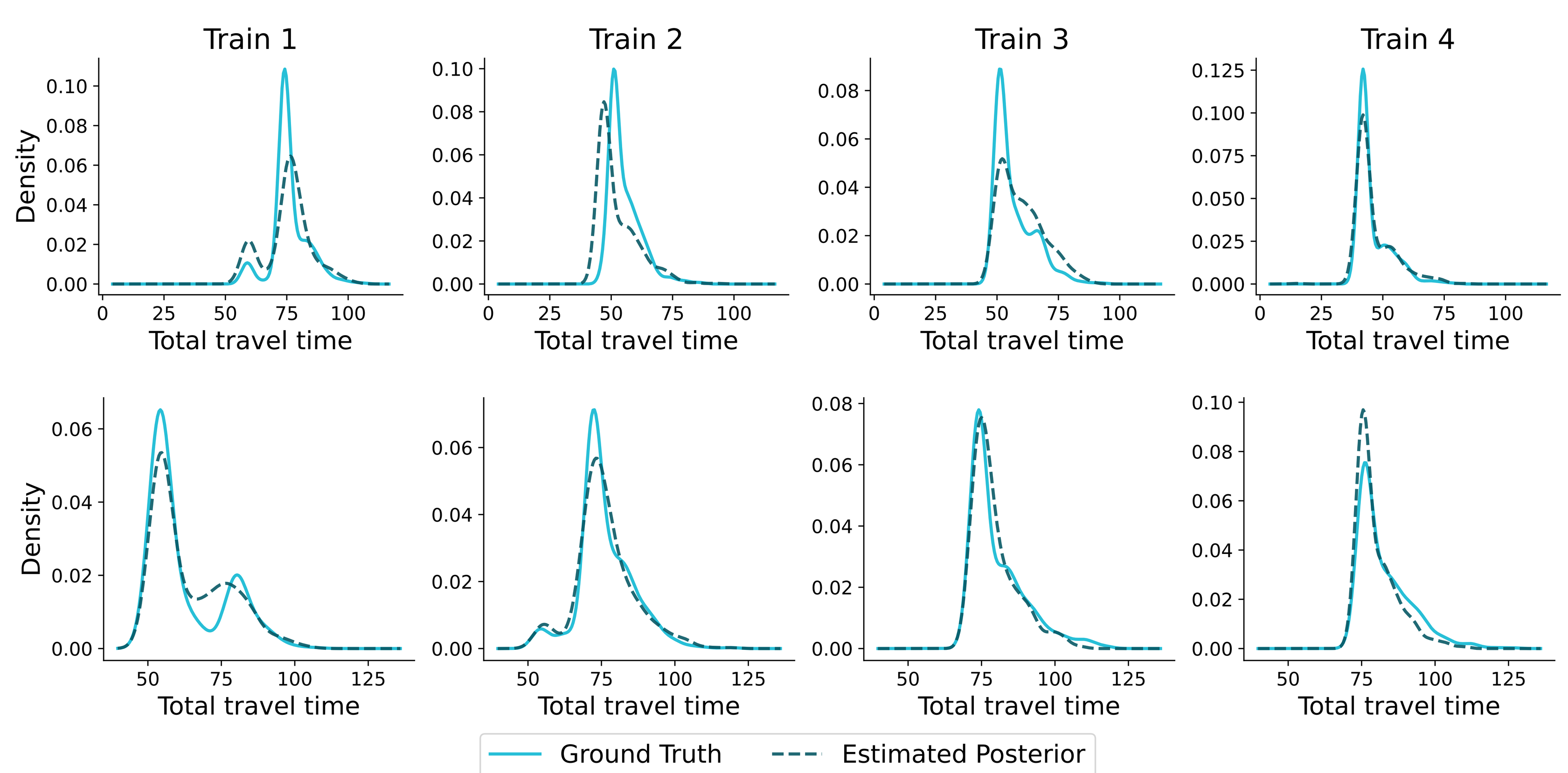


Figure 1. Two example settings (one per row) showing the estimated posterior densities of total travel time for the four trains, alongside the corresponding ground-truth densities. Ground truth is approximated by 500 simulator runs per setting, using the resulting travel-time samples. The estimated posterior densities closely match the ground-truth distributions, even in the more complex settings with the bimodal distributions.

Experiment: Comparison of Summary Networks

- Simulate undirected graphs with $N = 30$ nodes (no self-loops). Each node is labeled A or B with number of A-nodes drawn as $\texttt{num\_a} \sim \mathrm{Unif}\{5, \dots, 25\}$; remaining nodes are type B.
- Type-dependent edge probabilities: Draw $\pi_{AA}, \pi_{BB}, \pi_{AB} \sim \mathrm{Unif}(0.1, 0.9)$.
- Triadic closure parameter γ: Sample $\gamma \sim \mathrm{Unif}(0, 1)$ and compute common neighbors CN_{ij}.

$$p^{\star}_{ij} = \mathrm{clip}\left(p_{ij} + \gamma \frac{\mathrm{CN}_{ij}}{N-2},\, 0,\, 1\right), \quad p_{ij} \in \{\pi_{AA}, \pi_{BB}, \pi_{AB}\}.$$

- Inference target: Parameters $\theta = (\pi_{AA}, \pi_{BB}, \pi_{AB}, \gamma)$; perform neural posterior inference for θ given the simulated graph. We try different suitable neural architectures as summary networks and use a coupling flow with spline transformation

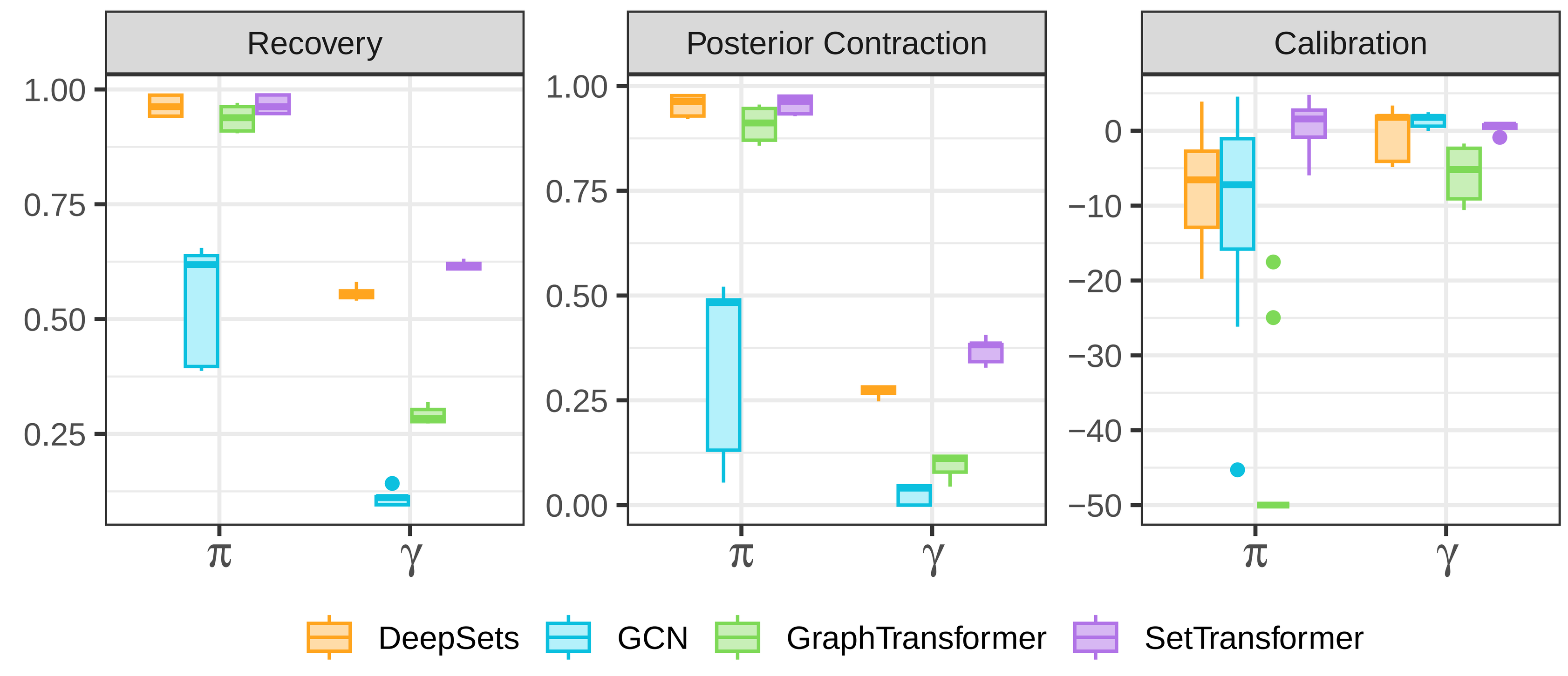


Figure 2. Performance of four different summary networks in terms of recovery (correlation of the posterior median to the true parameter value), posterior contraction ($1 - \sigma^2_{\text{post}}/\sigma^2_{\text{prior}}$) and simulation-based calibration ($\log \gamma$ score, values above 0 indicate good calibration). Triadic closure parameter γ harder to recover than baseline edge probabilities π; GCN performs worst for both parameters; best recovery and contraction from the Set Transformer.

Experiment: Mice Interaction Network

- Research question: How do social interactions among free-ranging mice shape the composition and dynamics of their gut microbiomes?
- Simulator: Cohort represented as a weighted interaction network $G = (V, E)$:
  - Nodes = mice; each mouse carries a subset of microbial taxa stored as node attributes.
  - Edges capture interaction propensity with edge weights $w_{ij} \in [0, 1]$.
  - In discrete daily steps, pairs with $w_{ij} > 0$ can exchange taxa. Exchange magnitude scales with both w_{ij} and an exchange factor $\alpha \in [0, 1]$.
  - Network density parameter δ controls the overall number of ties.
- Inference task: Infer (δ, α) from (i) the interaction network adjacency matrix and (ii) final-day microbiome composition within the mice.

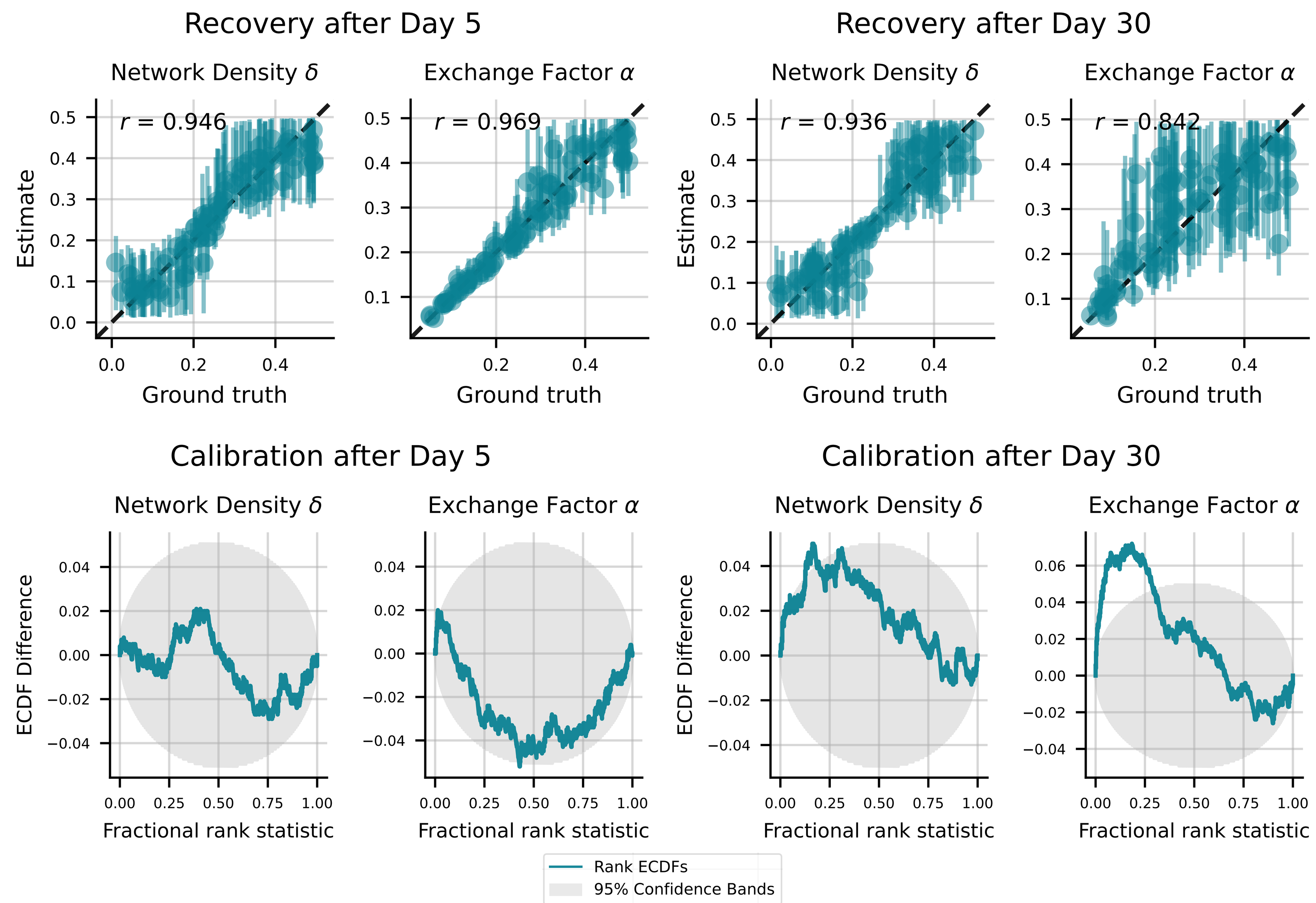


Figure 3. Top: Parameter recovery for (δ, α) after observation horizons of 5 days (left) and 30 days (right). Points indicate posterior medians and error bars denote 95% credible intervals. Overall recovery is good, with reduced precision for α at larger true values, particularly for the longer horizon. Bottom: Calibration ECDFs for both parameters. Calibration is close to excellent for Day 5, and slightly degraded for Day 30 but remains acceptable.

TL;DR

- **Graph-ABI:** Extend amortized Bayesian inference to variable-size graph-structured data by replacing the standard summary network with a graph-capable encoder.
- **Summary Architectures:** Benchmark DeepSets, GCN, Graph Transformer, and Set Transformer as summary encoders in terms of recovery, posterior contraction, and calibration. Set Transformer performs best, despite not being graph-specific.
- **Applications:** Two case studies in logistics and biology demonstrate calibrated, reliable posterior estimates in realistic settings.

sjedhoff jedhoff@statistik.tu-dortmund.de