Report from project SAD2

1 Reconstruction of Synthetic Boolean Networks

1.1 Datasets and Ground Truth

Datasets (simulated trajectories) and their corresponding ground-truth Boolean networks are generated by sweeping the following parameter grid:

Parameter	**Values**
Network parameters	
`n_nodes`	{5, 8, 11, 16}
`network_seed`	{0, 1}
Trajectory parameters	
`n_trajectories`	{1, 5, 20}
`sync transition`	{True, False}
`trajectory_len`	{5, 20, 100}
`sampling_frequency`	{1, 3}

Table 1: Parameter grid used for generating datasets and ground-truth networks.

Larger networks (e.g., `n_nodes` > 12) significantly increased runtime for both dataset generation and BNFinder inference, so instead of exhaustively testing all sizes we used a representative set spanning the range $[5, 16]$. In total, the sweep contains $4 \times 2 \times 2 \times 3 \times 3 \times 2 = 288$ parameter combinations, which already requires substantial computational effort during evaluation. In addition, we tracked the `attractor_state_percentage`, which varies between 0 and 1 depending on the trajectory. Overall, the chosen grid allowed us to assess which conditions favor accurate graph-structure inference.

1.2 Evaluation

- **Reconstruction setup.** We used the simplest BNFinder invocation: `bnf -e input1.txt -n output1.sif -l 3`. The key adjustment was `-l 3`, which limits each Boolean function to at most three parents; this greatly reduces the search space and speeds up reconstruction.
- **Scoring functions.** We evaluated candidate network structures using two scores recommended for BNFinder: *Minimal Description Length (MDL)* and *Bayesian–Dirichlet equivalence (BDe)*.
- **Loss functions (accuracy metrics).** We measured structural prediction error using `edge jaccard distance` (one minus the Jaccard similarity between graph edge sets) and `graph edit distance (GED)` (the minimum cost of transforming one graph into another through a sequence

	Jaccard		GED	
	BDe	**MDL**	**BDe**	**MDL**
Shapiro–Wilk statistic	0.830	0.870	0.882	0.876
p-value	6.718×10^{-16}	8.570×10^{-14}	5.050×10^{-13}	1.960×10^{-13}

Table 2: Shapiro–Wilk normality test results for Jaccard distance and graph edit distance (GED) under BDe and MDL scoring.

	Jaccard		GED	
	BDe	**MDL**	**BDe**	**MDL**
n (samples)	252	252	252	252
Median	0.875	0.848	11.0	11.0
Median difference	0.027		0.0	
p-value	0.107		0.798	
U statistic	34339.0		32171.0	

Table 3: Summary statistics and Mann–Whitney U test results for Jaccard distance and graph edit distance (GED) under BDe and MDL scoring.

of edit operations): together they capture both *local edge overlap* (exact wiring agreement) and *global topological discrepancy* (how many edits are needed to transform one graph into the other). We chose these metrics because they are widely used in graph-structure evaluation and provide complementary views of reconstruction quality.

1.3 Scoring functions vs. reconstruction quality

First, we assessed whether the obtained accuracy metric values followed a normal distribution within each scoring–loss function subset using the Shapiro–Wilk test. As all resulting p-values were below the significance level $\alpha = 0.05$ (Table 2), the null hypothesis of normality was rejected for all subsets, indicating that the accuracy metric values were not normally distributed.

Given this result, we evaluated whether there were statistically significant differences in accuracy metric values between the two scoring function groups (BDe and MDL). Since the normality assumption was violated, we employed the non-parametric Mann–Whitney U test. For both the `edge jaccard distance` ($U = 34339.0$, $p = 0.107$) and the `graph edit distance` (GED; $U = 32171.0$, $p = 0.798$), the results at $\alpha = 0.05$ indicated no statistically significant differences between the BDe and MDL scoring functions (Table 3). We therefore conclude that, under the evaluated conditions, the choice of scoring function does not significantly affect graph reconstruction accuracy.

1.4 Dataset parameters vs. reconstruction quality

To assess whether parameter values were significantly associated with the final reconstruction accuracy, Spearman's rank correlation coefficients and their corresponding p-values were computed between the parameter values and the loss function values, given the assumed non-normality of the latter. The results are presented in Table 4.

Important note. For both loss functions, lower values correspond to improved reconstruction accuracy. Consequently, a negative correlation coefficient indicates that as the parameter values increase, reconstruction accuracy improves.

Loss function	Jaccard		GED	
Scoring function	**BDe**	**MDL**	**BDe**	**MDL**
sync	0.01	0.02	-0.12	**-0.13**
sampling_frequency	**0.34**	**0.33**	0.09	0.07
n_trajectories	**-0.40**	**-0.37**	-0.10	-0.07
trajectory_len	**-0.31**	**-0.29**	-0.04	- 0.03
n_nodes	0.00	**0.13**	**0.76**	**0.79**
attractor_state_percentage	**-0.24**	**-0.26**	**-0.31**	**-0.33**

Table 4: Spearman's rank correlation coefficient values between loss function and parameter values. Coefficient values which had their corresponding p-value $< \alpha = 0.05$ (observed correlation is unlikely due to random chance) were marked in **bold**.

1.5 Analysis

Correlation analysis between loss function values and parameter values revealed several noteworthy patterns.

- **Scoring functions.** There are no notable differences in correlation trends between `MDL` and `BDe` BNFinder scoring functions.
- **Loss functions.** As `edge jaccard distance` and `graph edit distance` capture different aspects of graph reconstruction accuracy, their correlations with parameter values varied. Nonetheless, employing both metrics allowed us to obtain complementary perspectives on reconstruction quality, as each parameter exhibited statistically significant correlations with one or both measures.
- **Parameters:**
  - **sync** (0 for False, 1 for True) - slightly negative correlation with `GED` suggests a positive effect of synchronously generated trajectories on reconstruction accuracy.
  - **sampling_frequency** - a relatively strong positive correlation with `edge jaccard distance` indicates that as the number of time steps between consecutive sampled states increases, reconstructing the graph becomes more difficult.
  - **attractor_state_percentage** - a consistent negative correlation with both `edge jaccard distance` and `graph edit distance` suggests that a higher proportion of attractor states relative to transient states in the dataset facilitates graph structure inference.
  - **n_trajectories** and **trajectory_len** - a relatively strong negative correlation with `edge jaccard distance` (Figure 3) indicates that larger datasets facilitate more accurate graph reconstruction.
  - **n_nodes** - network size exhibited a very strong positive correlation with `GED` (Figure 4), indicating that larger networks are more difficult to infer accurately.

Final conclusions. We evaluated a wide range of parameter configurations; however, presenting all resulting plots would reduce the readability of this report. Therefore, we selected only the most informative visualizations for inclusion here. All plots are available in the accompanying repository under the `plots/` directory.

The results varied substantially, ranging from near-perfect reconstructions to cases with almost no correctly inferred edges. For networks of fixed size, the most influential factor was dataset size:

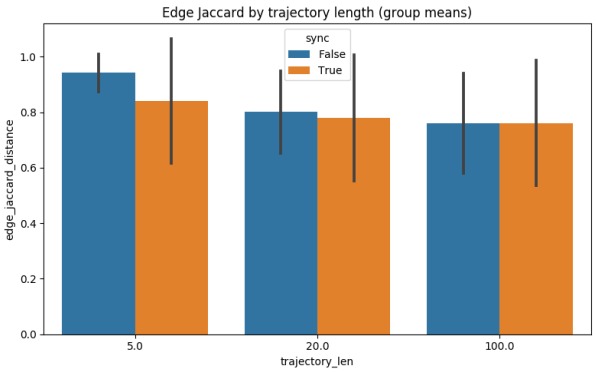

Figure 1: *
Jaccard vs. trajectory length (MDL)

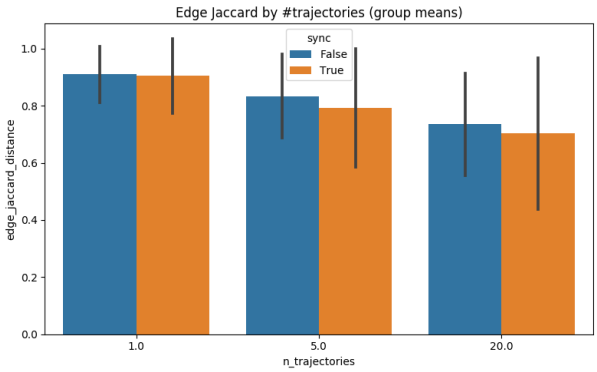

Figure 2: *
Jaccard vs. number of trajectories (MDL)

Figure 3: Edge Jaccard distance under MDL for two dataset axes.

increasing both the number of trajectories and their length consistently improved reconstruction performance (Figure 3). This observation aligns with intuition, as larger datasets provide more information for accurate inference. Conversely, reducing the number of states in the dataset by increasing the sampling frequency had a detrimental effect on performance.

We also observed a slight advantage for synchronous transitions, which is expected since synchronous updates are deterministic and, at each step, constrain the Boolean transition functions of all nodes simultaneously. The number of attractor states was another important factor, with a higher proportion of attractor states leading to improved reconstruction quality. We hypothesize this effect may arise because the algorithm can analyze the same set of transitions multiple times, thereby establishing stronger and more reliable connections between nodes.

The evaluation metrics (`edge_jaccard_distance` and `graph_edit_distance`) provided complementary insights into the relationships between model parameters and graph reconstruction quality. In contrast, we did not observe a consistent difference in performance between the MDL and BDe scoring methods. Finally, a strong negative correlation between network size and reconstruction accuracy was observed (Figure 4), which is expected given the increased difficulty of accurately inferring larger networks.

Insights for the next task. Based on our analysis, we gained several important insights that guided the design of the subsequent task. For the inference of a real biological model, we therefore propose generating a dataset using the following parameter settings: `sync` = True, `sampling_frequency` = 1, `attractor_state_percentage` - best to choose a random seed that allows us to generate a dataset with a high percentage of attractor states, `n_trajectories` = 20 (or higher), `trajectory_len` = 100 (or higher), `n_nodes` - the smaller the model the easier it will probably be to infer.

2 Reconstruction of Real-Life Biological Networks

In the second part of the project, we moved from synthetically generated Boolean networks to validated models describing real biological processes. The goal of this step was to assess whether the insights obtained from controlled simulations translate to realistic network structures.

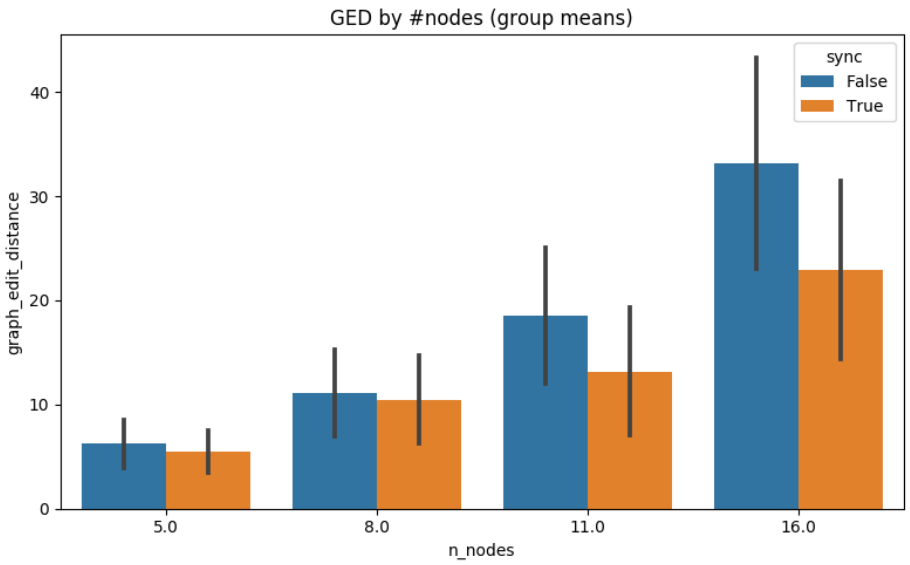

Figure 4: GED under MDL as a function of network size (`n_nodes`).

2.1 Selection of biological models

We selected two Boolean network models from the Biodivine repository that differ in size and regulatory complexity, while remaining within the computational limits identified in Part I. Using two models allowed us to qualitatively compare reconstruction behavior across biological systems of different scales under otherwise identical inference conditions.

The first model represents the Toll signaling pathway in *Drosophila* [2]. This pathway is a classical example of a regulatory cascade governing early developmental processes. Due to its relatively compact structure and moderate number of regulatory interactions, it serves as a tractable example of a small biological network.

The second model describes a neurotransmitter signaling pathway [3], capturing interactions between multiple molecular components involved in neuronal signal transduction. Compared to the Toll pathway, this model contains a larger number of variables and regulatory links, making it a more challenging target for network inference.

Model	**Variables**	**Inputs**	**Regulations**
Toll pathway	9	2	11
Neurotransmitter signaling pathway	14	2	20

Table 5: Overview of the biological Boolean network models used in the study. Inputs (also referred to as source nodes) are variables that have no incoming regulatory interactions. These nodes represent external signals or upstream components whose activity is not governed by other elements within the modeled system. They are treated as free variables, meaning their update function is omitted. Regulations correspond to directed interactions between variables in the network.

2.2 Analysis

Since previous part revealed no significant differences in performance between the MDL and BDe scoring functions, we selected the default BDe scoring for all evaluations in this section. To validate and extend the insights obtained from synthetic networks, we performed a similar parameter sweep for both biological models, using the same evaluation metrics: edge jaccard distance and graph edit distance.

Building on the findings from Part I that larger datasets improve reconstruction accuracy, we extended the parameter grid to include even larger dataset sizes: `n_trajectories` = {1, 5, 20, 25, 30} and `trajectory_len` = {5, 20, 100, 150, 200}. We wanted to determine whether further increasing these parameters would continue to improve accuracy, or whether reconstruction quality reaches a plateau beyond a certain dataset size.

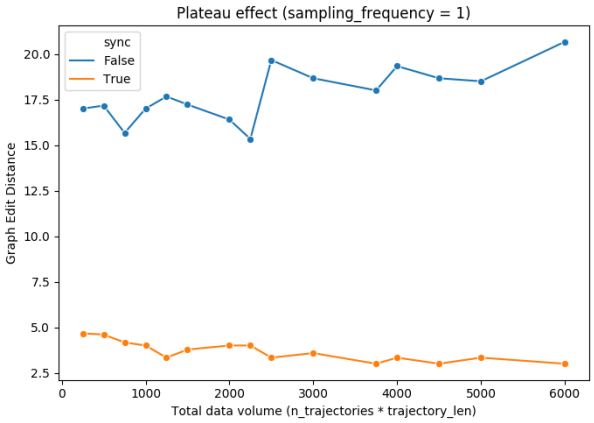

Figure 5: *
GED vs. Total data volume (Toll)

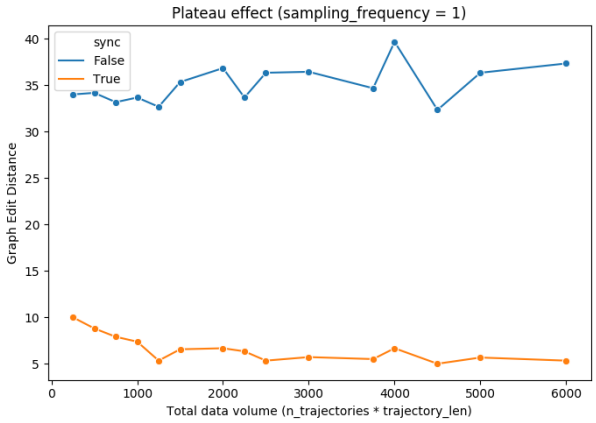

Figure 6: *
GED vs. Total data volume (Neurotransmitter)

Figure 7: Graph edit distance as a function of total data volume

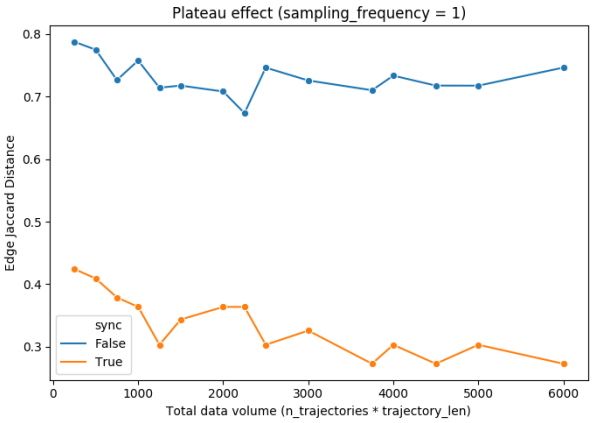

Figure 8: *
Jaccard vs. Total data volume (Toll)

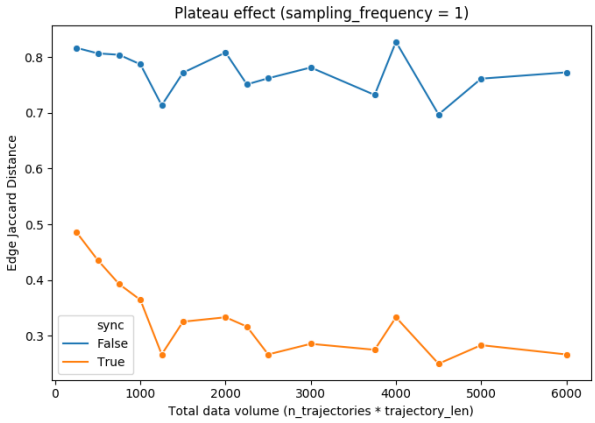

Figure 9: *
Jaccard vs. Total data volume (Neurotransmitter)

Figure 10: Edge jaccard distance as a function of total data volume

The results reveal a plateau effect: beyond a certain total data volume (n_trajectories × trajectory_len), both Jaccard distance and GED stabilize. This indicates that further increases in dataset size do not necessarily translate into proportional improvements in reconstruction quality.

It is also worth noting the substantial difference in reconstruction quality between synchronous and asynchronous update modes, with synchronous transitions consistently outperforming asynchronous ones across all dataset sizes, which is consistent with the findings from Part I.

2.3 Reconstruction visualization

The parameter sweep confirmed that the key findings from previous part hold for real biological networks: synchronous transitions (`sync=1`), minimal sampling frequency (`sampling_frequency=1`), higher values of `n_trajectories` and `trajectory_len`, and higher attractor state percentages consistently resulted in the higher reconstruction accuracy.

For visual comparison with the reference ground-truth networks, we selected reconstructions with the following optimal parameters and corresponding accuracy metrics:

Model	`n_traj`	`traj_len`	`sync`	`samp_freq`	`attr_%`	**Jaccard**	**GED**
Toll pathway	30	200	1	1	0.98	0.27	3.0
Neurotransmitter pathway	30	200	1	1	0.98	0.25	5.0

Table 6: Parameters and accuracy metrics for the visualized reconstructions

The reconstructed networks were visualized using Cytoscape (https://cytoscape.org) and compared side-by-side with the ground-truth network structures from the published reference models.

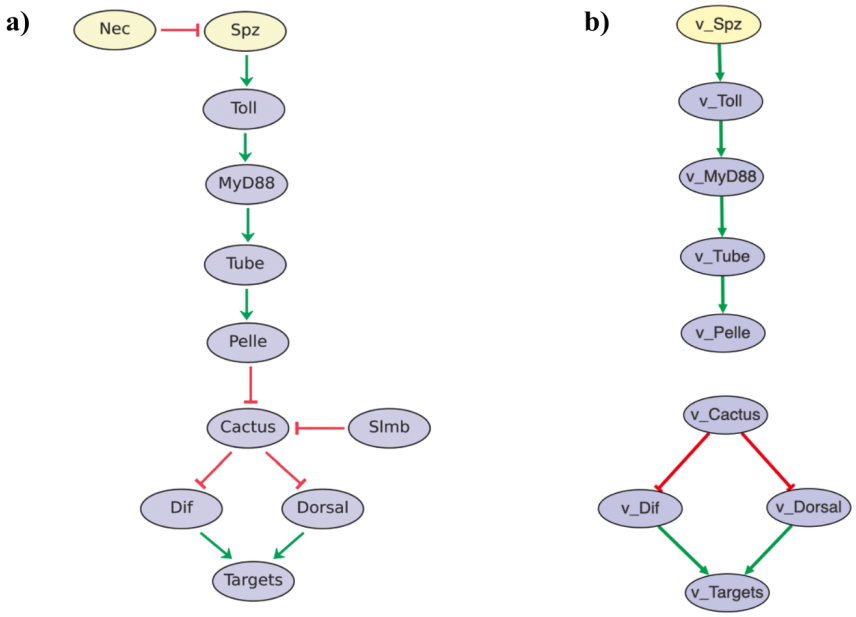

Figure 11: Reconstruction of the Toll pathway of *Drosophila*. (a) Ground-truth network structure from the published model [2] (b) and BNFinder2 reconstruction with BDe scoring.

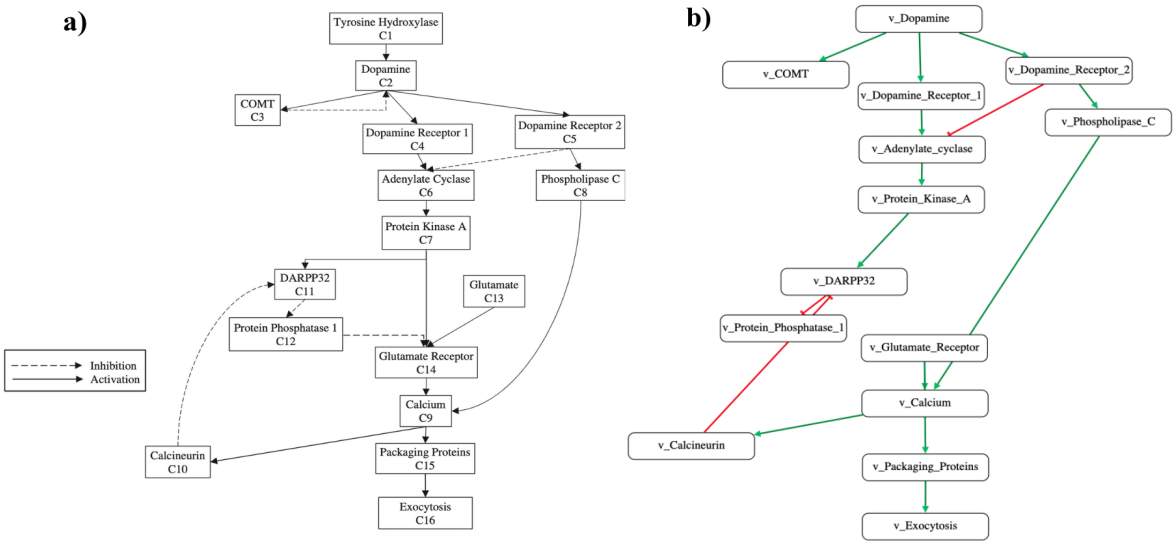

Figure 12: Reconstruction of the neurotransmitter signaling pathway. (a) Ground-truth network structure from the published model [3] and (b) BNFinder2 reconstruction with BDe scoring.

Visual inspection of the reconstructed networks demonstrates that BNFinder2 captured the majority of regulatory interactions in both biological models, while a limited number of edges remained missing or incorrectly inferred.

The neurotransmitter signaling pathway reconstruction exhibited more discrepancies than the Toll pathway, as expected given its larger size and greater regulatory complexity. Several edges were either missing or incorrectly classified, but overall the reconstruction still captured the main regulatory structure of the pathway.

It is important to note that even for the largest datasets and optimal inference parameters, the graph edit distance does not converge to zero, indicating that some structural discrepancies persist regardless of data volume. A contributing factor is the presence of input variables in both biological models. These variables act as external signals and do not have regulators themselves. As a consequence, their influence may be weakly reflected in the observed state transitions, and the corresponding regulatory edges are often underrepresented or absent in the generated trajectories. This makes such interactions inherently difficult to recover using data-driven structure inference.

However, the observed discrepancies cannot be fully attributed to missing input variables alone. In particular, the higher GED observed for the neurotransmitter signaling pathway, despite having the same number of inputs as the Toll model, suggests that increasing network size and regulatory complexity further limit the achievable reconstruction accuracy.

Nevertheless, examining both models provided valuable insights into how reconstruction quality scales with biological network complexity and validated the general applicability of our findings from synthetic networks.

3 Final conclusions

This project systematically investigated how dataset characteristics and network properties influence the accuracy of Boolean network structure inference using dynamic Bayesian networks.

Through comprehensive evaluation of 288 parameter combinations on synthetic networks, we identified dataset size, sampling frequency and update mode as the most critical factors affecting reconstruction quality. On the other hand, the choice between MDL and BDe scoring functions had minimal impact on performance.

Validation on two biological models: the Toll pathway and neurotransmitter signaling pathway confirmed these findings. Reconstructions using optimized parameters successfully captured the majority of regulatory interactions, with accuracy inversely proportional to network complexity. Notably, we observed a plateau effect suggesting diminishing returns beyond a certain data volume.

4 Challenges encountered

- **Legacy environment constraints.** One of the biggest challenges was setting up a fully compatible Python 2.7 environment (libraries, type checking, and tooling), since BNFinder does not support Python 3+. This turned out to be a valuable lesson in reproducibility: some of us managed the setup via `pyenv`, while others had to rely on Docker due to limited compatibility of their local machines with that early of a Python version.

- **Computational cost of the sweep.** To meaningfully assess how inference quality depends on key parameters (e.g., transition mode, network size etc.), we needed to run evaluations for a few hours, which limited how many additional combinations we could explore. To reduce runtime,

we parallelized runs where possible and used BNFinder's `-l` option to cap the maximum number of parents per node, substantially decreasing computational burden.

5 Code and data availability

All code used for trajectory dataset generation, analysis, and visualization, together with the generated datasets, the biological models employed, and the resulting plots, is available in our GitHub repository: https://github.com/Dziewiat/SAD2_project/tree/main.

6 Contributions

- **Authors:** Agata Paluch, Michał Zmyślony, Mikołaj Dziewiatowski, Norbert Szala
- **Contributions:** In Part 1, **Mikołaj** implemented the Boolean Network model and trajectory dataset generation using both single parameter sets and parameter grids. **Michał** inferred dynamic Bayesian network structures from the grid-generated datasets using BNFinder and evaluated inference performance with respect to dataset characteristics, scoring functions, and two structure-based graph distance measures. In Part 2, **Norbert** adapted the software pipeline to process real biological data, conducted an iterative parameter sweep to identify optimal parameter sets, and produced comparative plots illustrating the relationship between parameter values and reconstruction quality. **Agata** reconstructed real biological model graphs using BNFinder with the identified optimal parameters and summarized the biological analyses in a report. Collectively, these contributions supported both the methodological development and the biological validation presented in this study.

References

[1] N. Dojer, P. Bednarz, A. Podsiadło, B. Wilczyński, *BNFinder2: Faster Bayesian network learning and Bayesian classification*, Bioinformatics, Volume 29, Issue 16, August 2013, Pages 2068–2070, https://doi.org/10.1093/bioinformatics/btt323

[2] A. Mbodj, G. Junion, C. Brun, E. E. M. Furlong, D. Thieffry, *Logical modelling of Drosophila signalling pathways*, Molecular BioSystems, Volume 9, Issue 9, 2013, Pages 2248–2258.

[3] S. Gupta, S. S. Bisht, R. Kukreti, S. Jain, S. K. Brahmachari, *Boolean network analysis of a neurotransmitter signaling pathway*, Journal of Theoretical Biology, Volume 244, Issue 3, 2007, Pages 463–469.