

Biological Computation – Final Project

Shlomo Ezra 205359938

Yonatan Golan 208980888

Ex1:

Github link:

https://github.com/YonatanGolan/Biological_Computation/

Our output:

[illegible]

Ex2.1: Summary of the Paper:

"Repository of Logically Consistent Real-World Boolean Network Models"

S. Pastva, D. Šafránek, N. Beneš, L. Brim, T. Henzinger Affiliated with the Institute of Science and Technology Austria and Masaryk University, Czechia

Abstract:

The paper presents a comprehensive, open-source dataset of 210+ Boolean network models collected from various databases and literature. It aims to address the limitations of small, manually curated model sets commonly used for tool validation in systems biology. The dataset is accompanied by a validation pipeline ensuring the integrity and logical consistency of each model, identifying and fixing over 400 potential issues in existing models.

Introduction:

- **Background:** Boolean networks are fundamental in systems biology for modeling complex biological systems. Traditionally, these networks are designed and analyzed manually, which is labor-intensive and prone to errors.
- **Problem Statement:** Current validation practices rely on small, manually curated datasets that lack comprehensiveness and are often inconsistent. This paper introduces the Biodivine Boolean Models (BBM) dataset to provide a large, validated collection of models for better benchmarking and validation of Boolean network tools.

Methods:

- **Boolean Networks:** Defined formally, where each network comprises a set of variables and Boolean update functions. The networks include metadata such as human-readable names and biological relevance.
- **Model Acquisition:** Models were sourced from databases like GINsim, CellCollective, Biomodels, and the COVID-19 Disease Map, among others. Only models based on real biological systems were included, and multi-valued networks were converted to Boolean representations for consistency.
- **Validation and Repair:** An automated pipeline was developed to ensure logical consistency. This includes checking input node normalizations, regulation monotonicity, and essentiality, and removing unused components. Errors were systematically identified and corrected.

Results:

- **Dataset Composition:** The BBM dataset includes 212 models from various sources. Each model is available in multiple formats (bnet, aeon, sbml) and accompanied by metadata for easy reference.
- **Validation Pipeline:** The pipeline detected and repaired numerous issues in the models, ensuring a high-quality dataset. It normalized input nodes, validated regulation monotonicity and essentiality, and eliminated disconnected components to retain only the largest weakly connected component.

Conclusion:

The BBM dataset provides a significant advancement in the field of systems biology by offering a large, validated collection of Boolean network models. This dataset facilitates more extensive and reproducible validation of Boolean network tools and represents the largest publicly available set of real-world Boolean network models.

Significance:

- **Quality Control:** Ensures high standards in model integrity and logical consistency, addressing common issues in existing models.
- **Comprehensive Resource:** Provides a wide-ranging dataset that supports various Boolean network tools and methods.
- **Reproducibility:** Promotes reproducible research by offering a well-documented and accessible dataset for the community.

Availability:

The dataset is managed through a versioned git repository, with official editions published periodically. It includes machine-readable and human-readable metadata for each model, detailing sources, modifications, and relevant keywords.

Ex2.2:

Summary of "A Boolean Model of the Gene Regulatory Network Underlying Mammalian Cortical Area Development" by Giacomantonio, Goodhill (2010)

i. An Overview of the Article's Main Contributions

The article makes several significant contributions to the understanding of the gene regulatory network involved in mammalian cortical area development:

1. **Computational Modeling of Gene Interactions:** The study employs a computational model to simulate the interactions between five critical genes (Fgf8, Emx2, Pax6, Coup-tfi, and Sp8) that are essential for specifying areal identity along the anterior-posterior axis of the cerebral cortex.
2. **Exhaustive Network Simulation:** By simulating gene expression patterns for all possible networks containing these five genes, the study evaluates which interactions and combinations of interactions can reproduce the experimentally observed expression patterns. This exhaustive approach highlights the specificity and complexity of the gene regulatory network.
3. **Identification of Key Interactions:** The research finds that only 0.1% of the simulated networks can replicate the observed expression patterns. These successful networks lack certain interactions, such as auto-regulation and inductive loops, and many higher-order combinations of interactions.
4. **Repressive vs. Inductive Interactions:** The study reveals that repressive interactions are generally more likely to be present in the gene network than inductive ones. However, it also finds that mutually repressive loops are not critical for correct network functioning.
5. **Design Principles and Predictions:** The model illuminates the design principles of the gene network regulating cortical area development and makes novel predictions that can be tested experimentally. This provides a framework for future research to validate and expand upon these findings.

ii. Identify a Shortcoming in the Article (or How It Can Be Improved)

A shortcoming in the article is the reliance on Boolean expression levels to model gene interactions. The Boolean model with only two expression levels and two spatial compartments is a simplification that cannot capture the full complexity of the system, including mutant phenotypes. While this approach reflects the qualitative nature of the current expression data, it may oversimplify the actual dynamics of gene expression. This simplification could limit the model's applicability to more nuanced biological scenarios where gene expression levels are not strictly binary but can vary continuously. Improving the model by incorporating more detailed quantitative data on gene expression levels could enhance its accuracy and predictive power.

Other limitations we found:

- The asynchronous updating of nodes is more realistic than synchronous updating, but can lead to multiple steady states that are difficult to analyze.
- The lack of signaling between the two spatial compartments limits the model's ability to capture spatial dynamics.

- The lack of quantitative experimental data on gene and protein expression levels over time constrains the model's ability to make accurate predictions.

iii. Describe Your Impressions of the Article - Your Review May Be Positive or Negative (or Mixed)

The article presents a computational model to understand the gene regulatory network involved in mammalian cortical area development. The study focuses on the interactions between five key genes: Fgf8, Emx2, Pax6, Coup-tfi, and Sp8, which are crucial for specifying areal identity along the anterior-posterior axis of the cerebral cortex.

We think that one of the strengths of the paper is its comprehensive approach to simulating gene expression patterns. By evaluating all possible networks containing the five genes, the study provides a thorough analysis of which interactions are necessary for reproducing experimentally observed expression patterns. The finding that only 0.1% of the networks could replicate these patterns underscores the complexity and specificity of the gene regulatory network.

In addition, the paper also makes significant contributions by identifying that repressive interactions are generally more likely than inductive ones and that mutually repressive loops are not essential for correct network functioning. These insights offer valuable predictions that can guide future experimental research.

On the other hand, the model's reliance on Boolean expression levels, while reflective of the qualitative nature of current data, may oversimplify the actual dynamics of gene expression. This simplification could limit the model's applicability to more nuanced biological scenarios.

Overall, the article provides a robust framework for understanding the gene regulatory network in cortical area development, with clear implications for future experimental validation. To our opinion the study's methodological rigor and novel predictions make it a valuable contribution to the field, despite some limitations in the model's complexity.

iiii. Explain in Your Own Words the Definitions and Results

The paper does not specify the exact tools or software used for the simulations of the gene regulatory networks. The focus is on the methodology and the theoretical framework rather than the specific computational tools employed. Therefore, the paper does not provide details on the simulation tools used.

Important point in the theoretical framework and definitions:

1. **Boolean Model:** The model treats gene expression levels as Boolean variables, meaning each gene is either "on" or "off." This simplification is used because the current experimental data on gene expression is qualitative rather than quantitative. Repressive interactions were incorporated with a negation (NOT operator). The authors assumed that if a gene has multiple regulators, all regulatory conditions must be met, and so they combined their action with a logical conjunction (AND operator).
2. **Simulation of Networks:** The researchers simulated all possible networks containing these five critical genes: Fgf8, Emx2, Pax6, Coup-tfi, and Sp8. This

involved evaluating 1.68×10^7 possible networks to determine which configurations could reproduce the experimentally observed expression patterns.

3. **Criteria for Good Performance:** Networks were assessed based on their ability to replicate the spatial expression patterns of the five genes as observed in experimental data. Only networks that met these criteria were considered successful. Also, a network is considered to reliably reach a desired steady state if it does so with a probability greater than 50%. Networks that achieve both anterior and posterior steady states from their respective starting states are defined as “good.”
4. **Analysis of Interactions:** The study analyzed the probability of each interaction occurring in the successful networks. This helped identify which interactions were most likely to be present in the gene network regulating cortical area development.

Results:

1. **Successful Networks:** Out of the 1.68×10^7 possible networks, only 0.1% were able to reproduce the experimentally observed expression patterns. This indicates a high level of specificity and complexity in the gene regulatory network.
2. **Lack of Certain Interactions:** The successful networks lacked certain interactions, such as auto-regulation and inductive loops. This suggests that these types of interactions are not necessary for the correct functioning of the network.
3. **Repressive Interactions:** The study found that repressive interactions were generally more likely to be present in the successful networks than inductive ones. This highlights the importance of repression in the gene regulatory network.
4. **Non-Essential Mutually Repressive Loops:** The analysis revealed that mutually repressive loops were not critical for correct network functioning. This challenges the assumption that such loops are essential for gene regulation.

In summary, this article provides a theoretical framework for understanding the gene regulatory network in cortical area development, offering insights and predictions that can guide future experimental research.