The paper introduces a comprehensive, open-source dataset of over 210 Boolean network models. This dataset addresses issues in current model benchmarks by offering a large, validated collection of models from various databases and literature sources.

The paper starts with a background about Boolean networks; they give definitions of relevant concepts, explain what a signed influence graph is and go over some the important Properties of Boolean functions: essentiality and monotonicity that are necessary requirements who “map to the biological assumptions about the modelled system.”

The dataset, called Biodivine Boolean Models (BBM), includes models from databases like CellCollective, Biomodels, and GINsim. It offers several formats (bnet, aeon, sbml) and includes a validation pipeline that ensures the integrity and logical consistency of each model.

Models are curated from all known databases and publications, with the only requirement being that they represent real biological systems. The validation pipeline checks for, regulation monotonicity, and essentiality, performs input node normalization, removing unused components to make all influence graphs weakly connected, correcting errors and maintaining model accuracy. Multi-valued networks are also included, converted to Boolean representations for compatibility.

BBM is managed through a versioned git repository, regularly updated and officially published editions at intervals. Each model is accompanied by metadata, including publication references and descriptions of any modifications.