Modeling complexity through statistical approaches

Complex systems often exhibit emergent behaviors due to non-linear interactions, with the relevant quantities described by probability density functions (PDFs). Emphasis on prevalent distributions—such as Normal, Gamma-like, and Power laws—and associated models is vital for advancing complexity studies. This thesis explores modeling techniques for complex systems in socio-economics and biology, focusing on PDFs that either influence or characterize key model properties.

In the socio-economic part, a new measure termed 'gintropy' is introduced, combining the Gini index with entropy. Together with its generalization (f-gintropy) and a Kullback-Leibler-based divergence metric, it is applied to income data to capture distributional differences—especially those with power-law features—more effectively than traditional metrics. Another study links scientific productivity, impact, and the Hirsch index, deriving limits from gintropy and validating them with citation data that follow Tsallis-Pareto distributions, highlighting research output inequality. We also relate gintropy to the recently proposed Kolkata index and evaluate its and the Gini index's ability to quantify excellence in groups such as Nobel laureates and Olympic medalists. These results show that inequality measures may be used to distinguish a subgroup of high-achieving individuals from the general population.

The biological part introduces the Local Growth and Global Reset (LGGR) model, applied to tree size dynamics and validated using diameter data from Romanian forests, wood-pastures, and Hungarian plantations measured by our team. Incorporating realistic growth and reset (mortality and recruitment) kernels, the model predicts PDFs and reveals ecological patterns, with Gamma-like distributions emerging prominently. A subsequent study addresses discrepancies in axonal length distributions across granularities in structural brain connectome data: neuron level fine-grained data follow the Exponential Distance Rule (EDR), while regional coarse data tend toward Gamma-like forms due to endpoint aggregation. A one-dimensional model captures this coarse-graining effect, with dimensional and curvature-related limitations explored. Quantitative analysis suggests strategies for predicting decay rate parameters in connectomes where exhaustive neuron level data are unavailable.

Overall, this thesis contributes with advanced tools for analyzing complex systems by integrating novel statistical methods with real-world socio-economic and biological applications, providing a framework for revealing hidden patterns and enhancing predictive models.