TOPOLOGICAL DATA ANALYSIS: FROM SOYBEAN LEAVES AND A CLASSIFICATION BASED ON IMAGES OF FLY WINGS TO EDUCATIONAL DATA.

The project is triggered by ongoing work that is analyzing the appearance of stress in soybean plants that have developed under controlled conditions of temperature and atmospheric concentration of CO_2. Images were taken of soybean leaves (Glycine max) from groups of individuals that developed under 4 controlled climatic conditions:

1 Usual conditions (control);
2 Increased temperature;
3 Increased atmospheric concentration of CO_2;
4 Temperature and atmospheric concentration of CO_2 increased.

From these images we are carrying out an analysis, using TDA tools, to measure any unexpected asymmetries that characterize stress in plant development. The main objective is to verify the effects of predicted climate change on the development of this species.

The completion and discussion of this project will allow discussions to begin on the use of TDA in other databases.

For insect groups, wing morphology is an important character for species identification. In recent decades, software has been developed to digitize morphological characteristics and conduct consequent statistical analysis, such as principal component analysis and canonical correlation analysis, to establish the phylogenic relations of species. Examples of this software include TPSDig. DrawWing and SHAPE , which are based on modern digitalizing methods, such as 'Landmark' and 'Semilandmark'.

Most adult insects have wings that are frequently used in taxonomy. However, the detailed characteristics of veins and spots on wings require recognition and measurement by well-trained experts; thus, these features are difficult for general users to use.

In this part of the project we intend to use methods from TDA (Topological Data Analysis) to obtain a classification based on images of fly wings. Specifically, based on the venation extracted from those images. We will use a database made available by the entomology research group at the São Paulo State University in Rio Claro.

We also intend to use TDA tools on public data on student results in the Brazilian university entrance exams (ENEN). This data, along with a description of the exams, is publicly available at https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem, already anonymized. This analysis will be able to describe student performance over the years (the data covers 1998 to 2023). In this study, the main objective will be to assess the impact of the COVID-19 pandemic on student performance, in particular by evaluating which types of questions in the test affected performance the most. Our hypothesis is that questions involving greater text interpretation and/or more theoretical elaboration had lower success rates in the post-pandemic period.