Complex Networks. FINAL PROJECT.

General specifications:

- Select/Construct a network of your interest. The choice of the network should be justified (please, include this justification in the document) and should be of medium size (1000 to 3000 nodes approx.). The use of networks already available (for example, in the links provided on Aula digital or available by googling) will penalise the final score of the project.
- If the network is obtained from data processing (web link tracking, database queries, plain text analysis, etc.), it will report an extra on the final score of the project.
- The use of different softwares used to study and visualize properties of the network is allowed.
- The document must be generated using the LATEX template of the TFM of the MUSI available on Aula digital (except for the cover page, which will be deleted).
- The document must include a bibliography section and (very important) all the sources consulted with a relevance in the presented project must be included on it. Included a declaration of usage of generate AI.
- The maximum allowed length of the document is 10 pages (including figures). This means that you should focus on presenting your best and most relevant results, rather than showing everything you have computed. For example, this may encourage you to merge information into more complete and expressive figures, instead of displaying many separate ones.
- A link (like a Colab link) must be shared to a repository with the code generated during the project, duly commented.
- The project must be carried out individually.
- The evidence of any type of academic misconduct (like plagiarism, share the code with others, or not to include in the bibliography the sources consulted, to name a few) will result in a final score of 0 in the whole subject and the corresponding report to local and EDISS coordinators.

The following two sections constitute the **minimum** content that is required to be analyzed in the project.

First part: The aim of this part is to study, interpret and compare different centrality indices and the study of the degree distribution of nodes.

1. A section must be included with the properties and basic structure of the network: number of nodes, number of edges, number of connected components, average degree, diameter, etc. The method of obtaining the network must also be included.
2. At least 2 centrality indices should be considered. Each one of them must be interpreted individually, and a section of comparison between the different indices must be included. That is, to study if there is correlation between some of them, to determine if they follow the same trend, to know if their maximum value is established over the same node, etc. Could these results have been anticipated in the context of the chosen network?
3. The degree distribution of the nodes must be obtained and represented accordingly, also identifying whether it fits any known model and, if so, calculating its parameters.
4. Study of homophily and assortative mixing.
5. At least 3 graphical representations of significant parts of the network shall be obtained (significant node environments, prominent clusters, etc.).

Second part (this part depends on the execution of the final chapters during the course): The aim of this part is to study, interpret and compare random networks with networks obtained from real data.

6. In addition to the (real) network under study, a set of networks obtained by random methods, using the Erdős-Renyi and configuration models, with appropriate parameters, will be considered.
7. Centrality measures, clustering coefficients, degree (or others) distribution, size of giant components, etc should be analyzed for random networks and compared to the real one.
8. Distribution algorithms will be applied to communities in both the reference and random networks, comparing and interpreting the results.
9. Random simulations of node percolation will be carried out, following a uniform model, studying how this affects the components (if a giant component is maintained), both for the reference network and the random ones. It will be represented, on average, how the occupancy probability affects the size of the largest component.
10. The previous point will be repeated considering non-uniform percolation, where all nodes with degree greater than a given value are unoccupied. The analysis will be done according to this degree.

What is expected from an outstanding final project?

A high-level project should go beyond technical computation and show critical understanding of complex networks in a real context. In particular, an excellent project should:

1. Clearly motivate the network and its context. Explain why this system can be studied as a network, what its nodes and edges represent, and why its structure matters (scientifically, socially, economically, etc.).
2. Formulate meaningful and non-trivial research questions. Rather than computing metrics mechanically, identify questions that can be answered using the tools learned in the course. Show how each question motivates a methodological choice.
3. Compare results with networks of similar nature. If the network belongs to a well-known domain (e.g., transportation, friendship, biological interaction, citation graphs), the analysis should not stop at obvious conclusions ("airports have hubs"). Instead, compare with known benchmarks or previous findings in similar systems.
4. Interpret results critically, not just report them. Discuss whether the observations confirm or contradict expectations. Consider the limitations of your data, sampling, or assumptions.
5. Use appropriate visualizations to support arguments.

Presentations

The exact day of the presentations will be decided at a later date.

- First oral presentation of 5 minutes (1 or 2 slides) that expose the selection of the network and its main properties (december).
- Final oral presentation of 20 minutes with extra time for questions (january). Again, do not explain all your work. Only your best and most relevant results.
- Presentation of the document (january): send it via Aula digital.