Title of the research project:
Disentangling multiple components correlations: methods and techniques for hypergraph or simplicial complex representation of data

Keywords (up to five)
Complex networks, Data Mining, Big Data, Machine Learning, Hypergraphs and Simplicial Complexes

Supervisors (at least two from two different areas):

Supervisor 1 (name, contact details, homepage, area of expertise)
Prof. Dr. Regino Criado, Universidad Rey Juan Carlos de Madrid, Madrid, Spain
Mail: regino.criado@urjc.es Area of expertise: Graph Theory, Applied Mathematics

Supervisor 2 (name, contact details, homepage, area of expertise)
Prof. Dr. Stefano Boccaletti, CNR, Institute for Complex Systems, Florence, Italy
Mail: stefano.boccaletti@gmail.com Area of expertise: Statistical and nonlinear physics

Project description (max 5000 characters)

The last two decades have witnessed the development of the multidisciplinary field of Network Science, wherein a variety of systems have been modelled as networks of coupled units, in the attempt to unveil the mechanisms underneath their observed functionality. When one adopts a network representation of a complex system the implicit assumption is made that the overall interplay among the unitary components of the system is exhaustively described by combinations of pairwise interactions.

Such a hypothesis may indeed be justified when one studies certain types of processes, but it is very short in representing faithfully many other circumstances. There are a lot of practical situations which simply cannot be factorized in terms of pairwise interactions and where, instead, higher-order interactions must be taken into account for a proper description and representation of the system.

In these cases, one must adopt hypergraphs and/or simplicial complexes representations.

The main objective of this project is to construct general tools and techniques for representing real-world data as hypergraphs or simplicial complexes. We will use skills from data mining (and more in general from big-data science) together with methods from machine learning and deep learning [1].

If the analysis is limited to only pairwise interactions, two main methods are available: functional networks and parenclitic networks. Functional networks allow to represent a multivariate time series as graphs, where each node is one of the signals, and each pairwise interaction is the value of a given time-independent function of the corresponding two signals [2]. Parenclitic networks, introduced in Ref. [3], furnish instead a reliable representation of static data and provide powerful classifiers by identifying differences of single subjects from a given class. Initially applied to plant genomics, parenclitic networks have later found many other applications, for instance in epigenetics of tumors [4] and in cybersecurity [5].

The objective is to extend these two representation methods to the detection of multiple-body correlations (or functions) in real data. We plan to move along a very ambitious path leading to the introduction and full characterization of a set of general methods able to reconstruct functional and/or parenclitic hypergraphs and/or simplices from data of any discipline, source or experimental observations. Specific, customized, applications will be made for brain data, financial market analysis, and for other systems where risk management is of fundamental relevance.

**Relevance to the MERC PhD Program (max 2000 characters)**

The project is openly multidisciplinary, as it involves bridging methods and results of statistical physics with those of applied mathematics and data analysis.

The accomplishment of our objectives will allow elucidating many fundamental questions such as: What is a truly genuine higher-order interaction in real world system? How do we tell it apart from low-order ones in data? And if it is possible, what type of data do we need to tell the difference between low and higher-order interactions?

The novel methods and tools that we will develop will therefore be of great value and large use in a variety of fields and circumstances, far beyond the practical cases to which we will apply them.

Potential applications and beneficiaries are therefore numerous, and it is easy to forecast that the results obtained will be of essential value in a wealth of circumstances along the following years.

Moreover, the project has implications in the areas of “systems and control theory”, “data driven mathematical modelling and simulation of complex systems”, thereby squarely fitting within the MERC program.

**Joint supervision arrangements**

The first supervisor is based at the Department of Applied Mathematics, Material Science and Engineering, and Electronics of the University Rey Juan Carlos of Madrid. During the initial period of stay abroad of the candidate, the supervisor will discuss and schedule the entire program of activities which will form part of the final PhD thesis. The second supervisor is working at the CNR, Institute of Complex Systems, in Florence. He is a member of the Board of MERC. During the period of stay in Italy, weekly meetings will be organized with him, and at least two visits to Naples will be made to ensure the best possible supervision of the candidate. Regular online meetings (at least on a monthly basis) will be organized involving the PhD student the two supervisors during the entire duration of the project.

**Location and length of the study period abroad (min 12 months)**

The project foresees a total of 12 months where the candidate will collaborate with the group of scientists led by Prof. Regino Criado in the Department of Applied Mathematics, Material Science and Engineering, and Electronics of the University Rey Juan Carlos of Madrid, located in the campus of Mostoles, in the south area of Madrid.

This period will be realized during 2024, because during this period the Main Supervisor of the project (Prof. Criado) will have the opportunity to discuss with the candidate (and to schedule) the entire program of activities which will form part of the final PhD thesis.

The URJC has already agreed to host the candidate.