Abstract

Every single person, animal, or thing we can see in the world around us is part of a broader collection of components that can spontaneously self-organize to exhibit non-trivial global structures and behaviors at larger scales, often without external intervention, central authorities, or leaders. The properties of the collection these components give life cannot be understood or predicted from the full knowledge of its elements alone. Each collection is an example of complex systems whose behavior is intrinsically challenging to model due to the high non-linearity of the interactions between its constituents.

Traditionally, complex systems have been successfully studied through graphs abstracting the underlying network with vertices and edges connecting pairs of interacting components. Over the years, the scientific community has enriched the graph modeling framework for better capturing the richness of the interactions among such units. However, graphs have a substantial limitation encoded in their nature: they exclusively capture pairwise interactions. Yet, many complex systems are characterized by group interactions that cannot be described simply in terms of dyads. Studying such systems hence require new mathematical frameworks and scientific methodologies for its investigation.

Hypergraphs are the perfect candidates to tackle this task. A hypergraph is a generalization of a graph, where a (hyper)edge allows the connection of an arbitrary number of vertices. However, the powerful expressiveness of hypergraphs has a few drawbacks: dealing with the complexity of such data structures and the lack of appropriate tools and algorithms for their study. For this reason, hypergraphs have been little used in literature in favor of their graph-counterpart. Recently, this trend

has been drifting, thanks to an increasing number of systematic studies demonstrating that considering the higher-order structure of complex systems can enhance our modeling capacities and help us understand and predict their dynamical behavior.

This dissertation fits in this broad context of modeling complex systems with the general objective of formalizing and implementing more expressive network models. Specifically, the whole work is rooted in understanding how much and when we need high-order information conveyed by hypergraphs. The contribution described in this thesis can be grouped according to three macro topics.

I. *Tools for hypergraphs.* Motivated by the lack of a comprehensive and efficient hypergraph-specific library and the need for software libraries designed to perform operations directly on hypergraphs, we developed SimpleHypergraphs.jl, a software library to model, analyze, and visualize hypergraphs, written in Julia and designed for high-performance computing. This dissertation describes the main motivations behind creating SimpleHypergraphs.jl, the library's design choices, and its memory model. We further illustrate the functionalities offered by the software, including graph transformations and hypergraph visualization methods. We also present two case studies with the twofold objective of demonstrating how it is possible to exploit the proposed library and comparing hypergraphs with their corresponding graph counterpart to explore whether high-order structures convey more information in addressing specific tasks. Contextually, we also describe a generalized version of the label propagation algorithm for community detection suitable for hypergraphs.

The second and third topics addressed instantiate the initial broad research question into two principal research directions, both tied to the concept of diffusion. Informally, the term *diffusion* means the process according to whom an entity (e.g., information) spreads within a network, moving from node to node or group of nodes, through their interactions. Under this umbrella, this thesis focuses on studying *social influence propagation* and *epidemic spreading processes* within a population.

II. *Social influence on high-order networks.* In the context of social influence diffusion, we propose the formal definition of a high-order diffusion process, the generalization of a well-known graph problem to hypergraphs, and a set of heuristics to tackle it. Specifically, in this thesis, we first introduce the motivation behind this line of study and discuss a new high-order diffusion model with linear thresholds that mimics real-world social dynamics, where individuals influence the group they belong to, but - in turn - the group itself influences their choices. We further introduce the formal definition of the Target Set Selection problem on hypergraphs (TSSH), a key algorithmic question in information diffusion research, whose goal is to find the smaller set of vertices that can influence the whole network according to the diffusion model defined. Since the TSSH problem is NP-hard, we propose four heuristics to address it and extensively evaluate these algorithms on random and real-world networks.

III. *Epidemic dynamics on temporal high-order networks.* From the perspective of epidemic dynamics, we propose the formal definition of time-varying hypergraphs (TVHs), the introduction of direct and indirect interactions when studying an epidemic spreading via a TVH contact network, and an epidemic diffusion algorithm built on top of TVHs and direct and indirect contagion pathways. This dissertation motivates why one should use (temporal) hypergraphs rather than (temporal) graphs to analyze epidemic spreading processes. We then introduce the formal definition of temporal hypergraphs, describe a high-order SIS compartmental equation model suitable for TVHs, and discuss how we assembled these elements into an agent-based framework. We further present a sensitivity analysis of the TVH model to the epidemic parameters and different discretization of the time intervals when direct or indirect contacts may happen.

Built on top of the TVH model, we also propose a fine-grain modeling methodology for Non-Pharmaceutical Interventions (NPIs). We first motivate why one should evaluate such epidemic control strategies in the framework of agent-based models (ABMs) and high-order interactions. We then delve into reviewing personal protective, environmental, and social distancing measures and how they can be embedded into an epidemiological model based on high-order networks, ABMs, and the SIS equation-based model. We further describe how we formally enriched the TVH modeling framework to support the evaluation of NPIs. After assessing the ability of each intervention in controlling an epidemic propagation, we discuss a multi-objective optimization framework, which, based on the epidemiological data, calculates the NPI combination that should be implemented to minimize the spread of an epidemic as well as the damage due to the intervention.