

## ABSTRACT

---

Complex systems, including natural, social, and artificial structures such as traffic control, weather forecasting, policy-making, and epidemic dynamics, consist of interconnected components that exhibit unintelligible behavior when examined in isolation. To analyze these systems, researchers often collaborate across fields to develop models that replicate real-world behavior. The study of complex systems is marked by challenges such as data access and complexity, difficulties in modeling nonlinear and dynamic relationships, high computational demands, and the need for interdisciplinary cooperation. Researchers typically adopt either a bottom-up or a top-down approach, or a mix of both, to unravel human behavior within these systems. Agent-based models (ABMs) can serve as powerful tools for simulating interactions at the component level when adopting a bottom-up approach. Similarly, hypergraphs can accurately model systems to investigate higher-order interactions. Recent technological advancements have further enhanced this analytical capacity, notably with large language models (LLMs) that, after extensive training on large datasets, exhibit context-awareness and superior semantic understanding. These LLMs have proven particularly effective in tasks such as stance detection, showing adaptability across different datasets and outperforming traditional supervised methods with fewer resources. All these topics naturally give rise to a wide range of theoretical, methodological, and practical challenges. Working at the intersection of complex systems, online social dynamics, ABMs, hypergraph-based modeling, and LLM-driven analysis requires navigating heterogeneous data sources, integrating multiple formalisms, and balancing computational scalability with conceptual expressiveness. These challenges arise from both the intrinsic complexity of the systems under study and the limitations of existing analytical tools. This dissertation fits within this broader context and adopts an integrated perspective. Rather than pursuing two distinct research pathways, it brings together complementary approaches to advance the analysis of complex systems. On the one hand, the work contributes to improving

ABM frameworks, with particular attention to challenges such as scalability and modeling large, heterogeneous populations.

On the other hand, it investigates online social interactions through a combined methodological lens: group interactions are represented and analyzed using hypergraphs, while LLMs are employed to capture their semantic dimension. These components are not treated independently; instead, they try to form a unified framework that leverages structural and semantic information simultaneously.

The primary objective of this thesis is to investigate and analyze complex systems by leveraging ABMs, hypergraph-based network models, and LLMs. The contribution described in this thesis can be grouped into these three macro areas.

*Agent-based modelling.* In the context of ABMs, our goal is to enhance their functionality to advance the state of the art in ABM engines, particularly by addressing their computational limits, as the most widely used ABMs prioritize either performance or usability. Hence, my primary focus is to develop an approach that effectively integrates both features. The main research question this thesis aims to answer is:

*How can advances in parallel and distributed computing technologies be harnessed to improve the scalability and efficiency of ABMs for simulating large-scale, real-time, and highly dynamic systems?*

This question led us to: (i) conduct an in-depth review of existing ABM tools to identify current trends, challenges, and opportunities for innovation; (ii) develop a stable version of krABMaga, an ABM simulation engine, and various simulation models; (iii) extend the features and capabilities of krABMaga to include distributed computation and GIS data integration. Additionally, we focused on improving the system's usability by creating a user-friendly GUI.

*Hypergraphs.* Another topic in my Ph.D. project is analyzing group interactions and the emergence of collective behavior in online social media through the lens of hypergraphs. In this regard, the main research questions my project aims to answer are:

*Are specific structural patterns in hypergraphs dependent on the communities involved, or do they highlight specific*

*user behaviors? Can we predict the formation of these patterns?*

Motivated by this question we: (i) developed a community-driven open-source platform to store hypergraph datasets; (ii) examined interactions in online social media groups through the lenses of hypergraphs (contributing to  $H_{HG}^1$ ); (iii) investigated the application of modern techniques, such as graph neural networks, to analyze high-order interactions (contributing to  $H_{HG}^2$ ).

Large Language Models. Regarding LLMs, my project's goal is to assess the feasibility of using these tools to extract meaningful insights from text and enhance our understanding of how user interact and exchange their opinions in online social contexts. Specifically, the main research question my project aims to answer is:

*Can LLMs improve our understanding of online social interactions and opinion dynamics?*

The investigation on this topic led us to: (i) analyze the semantic information within the text linked to each node in a hypergraph using LLMs; (ii) develop a model to embed LLM-generated information into a stance detection task.

## ITALIAN ABSTRACT

---

I sistemi complessi, che comprendono strutture naturali, sociali e artificiali come il controllo del traffico, le previsioni meteorologiche, l'elaborazione di politiche pubbliche e le dinamiche epidemiche, sono costituiti da componenti interconnesse che mostrano comportamenti di difficile interpretazione se esaminate in isolamento. Per analizzare questi sistemi, i ricercatori spesso collaborano tra diverse discipline per sviluppare modelli in grado di replicare il comportamento osservato nel mondo reale. Lo studio dei sistemi complessi è caratterizzato da sfide quali l'accesso ai dati e la loro complessità, le difficoltà nel modellare relazioni non lineari e dinamiche, gli elevati requisiti computazionali e la necessità di una cooperazione interdisciplinare. In genere, i ricercatori adottano un approccio bottom-up o top-down, oppure

una combinazione di entrambi, per comprendere il comportamento umano all'interno di questi sistemi. I modelli ad agenti (Agent-Based Models, ABM) possono costituire strumenti potenti per simulare le interazioni a livello di singolo componente in un'ottica bottom-up. Analogamente, gli ipergrafi permettono di modellare con precisione i sistemi per indagare le interazioni di ordine superiore. I recenti progressi tecnologici hanno ulteriormente potenziato questa capacità di analisi, in particolare grazie ai Large Language Models (LLM) che, dopo un esteso addestramento su grandi insiemi di dati, mostrano consapevolezza del contesto e una comprensione semantica superiore. Questi LLM si sono dimostrati particolarmente efficaci in compiti come la stance detection, mostrando adattabilità a diversi dataset e superando i metodi supervisionati tradizionali con un minore impiego di risorse. Tutti questi temi danno naturalmente origine a un'ampia gamma di sfide teoriche, metodologiche e pratiche. Lavorare all'intersezione tra sistemi complessi, dinamiche sociali online, ABM, modellazione basata su ipergrafi e analisi guidata da LLM implica affrontare fonti di dati eterogenee, integrare molteplici formalismi e bilanciare la scalabilità computazionale con l'espressività concettuale. Queste sfide derivano sia dall'intrinseca complessità dei sistemi oggetto di studio, sia dai limiti degli attuali strumenti di analisi. Questa tesi si colloca all'interno di questo quadro più ampio e adotta una prospettiva integrata. Invece di perseguire due direttrici di ricerca distinte, essa riunisce approcci complementari per avviare l'analisi dei sistemi complessi. Da un lato, il lavoro contribuisce al miglioramento dei framework ABM, con particolare attenzione a sfide quali la scalabilità e la modellazione di popolazioni ampie ed eterogenee.

Dall'altro lato, la tesi indaga le interazioni sociali online attraverso una lente metodologica combinata: le interazioni di gruppo sono rappresentate e analizzate tramite ipergrafi, mentre gli LLM vengono impiegati per catturarne la dimensione semantica. Questi componenti non sono trattati in modo indipendente; al contrario, mirano a formare un quadro unitario che sfrutti simultaneamente le informazioni strutturali e semantiche.

L'obiettivo principale di questa tesi è investigare e analizzare i sistemi complessi sfruttando ABM, modelli di rete basati su ipergrafi e LLM. I contributi descritti in questa tesi possono essere raggruppati in tre macro-aree.

Agent-based modelling. Nel contesto dei ABM, il nostro obiettivo è potenziarne le funzionalità per avanzare lo stato dell'arte dei motori ABM, in particolare affrontandone i limiti computazionali, dal momento che gli ABM più diffusi privilegiano o le prestazioni o l'usabilità. Di conseguenza, il mio obiettivo principale è sviluppare un approccio capace di integrare efficacemente entrambe le caratteristiche. La principale domanda di ricerca a cui questa tesi mira a rispondere è:

*In che modo i progressi nelle tecnologie di calcolo parallelo e distribuito possono essere sfruttati per migliorare la scalabilità e l'efficienza degli ABM nella simulazione di sistemi su larga scala, in tempo reale e altamente dinamici?*

Questa domanda ci ha portato a: (i) condurre una revisione approfondita degli strumenti ABM esistenti per identificare i trend attuali, le sfide e le opportunità di innovazione; (ii) sviluppare una versione stabile di krABMaga, un motore di simulazione ABM, e diversi modelli di simulazione; (iii) estendere le funzionalità e le capacità di krABMaga per includere il calcolo distribuito e l'integrazione con dati GIS. Inoltre, ci siamo concentrati sul miglioramento dell'usabilità del sistema creando una GUI intuitiva per l'utente.

Hypergraphs. Un altro tema del mio progetto di Ph.D. è l'analisi delle interazioni di gruppo e dell'emergere di comportamenti collettivi nei social media online attraverso la lente degli ipergrafi. A questo proposito, le principali domande di ricerca a cui il mio progetto mira a rispondere sono:

*Determinati pattern strutturali negli ipergrafi dipendono dalle comunità coinvolte oppure mettono in evidenza specifici comportamenti degli utenti? Possiamo prevedere la formazione di questi pattern?*

Motivati da questa domanda, abbiamo: (i) sviluppato una piattaforma open-source guidata dalla comunità per archiviare dataset di ipergrafi; (ii) esaminato le interazioni nei gruppi dei social media online attraverso la lente degli ipergrafi (contribuendo a  $H_{HG}^1$ ); (iii) investigato l'applicazione di tecniche moderne, come le graph neural networks, per analizzare interazioni di ordine superiore (contribuendo a  $H_{HG}^2$ ).

Large Language Models. Per quanto riguarda gli LLM, l'obiettivo del mio progetto è valutare la fattibilità di utilizzare questi stru-

menti per estrarre informazioni significative dal testo e migliorare la nostra comprensione di come gli utenti interagiscono e scambiano opinioni nei contesti sociali online. In particolare, la principale domanda di ricerca a cui il mio progetto mira a rispondere è:

*Gli LLM possono migliorare la nostra comprensione delle interazioni sociali online e delle dinamiche delle opinioni?*

L'indagine su questo tema ci ha portato a: (i) analizzare le informazioni semantiche contenute nel testo associato a ciascun nodo in un ipergrafo utilizzando gli LLM; (ii) sviluppare un modello per incorporare le informazioni generate dagli LLM in un compito di stance detection.