On Spreading Phenomena and Influence Maximisation in Multilayer Networks

Abstract

mgr inż. Michał Jerzy Czuba

Networks offer a convenient framework for representing many complex systems. Social bonds between individuals, interactions among proteins in living organisms, transportation routes, or water supply systems: all of these can be modelled using graph-based formalisms. Over the years, numerous approaches to network modelling have been developed. In particular, the notion of multilayer networks has become successful, as it enables the simultaneous representation of multiple types of relations. Conceptually, a multilayer network can be seen as a stacked collection of graphs, where each one is spanned on the same node set (or on a subset thereof). Contextually, this structure can be obtained, e.g., from users aggregated across several social platforms together with the friendship relations between them.

One of the most important aspects of networks concerns their dynamics. This may take various forms, including structural changes in topology or fluctuations in the states of agents. The latter has typically been examined within the framework of spreading processes and underpins many domain-specific research areas such as epidemic modelling, opinion dynamics, and information diffusion. A key issue in this context is how to propagate a given phenomenon across a network as effectively as possible. This is traditionally referred to as the influence maximisation problem, which has a broad practical impact. In its canonical formulation, it seeks to identify a set of agents that, when chosen to initiate the spread, maximise the eventual reach of the diffusion process. Over the years, however, this problem has been reformulated in numerous ways, and various methodological paradigms have been leveraged to approach it.

This doctoral dissertation aims to address the problem of modelling spreading phenomena in multilayer networks and to analyse various approaches to solving the influence maximisation problem in such structures. As this work has been carried out within the domain of Information and Communication Technology, a complementary challenge was undertaken: to contribute to reliable and open research by publishing data, algorithms, methods, and code alongside the scientific findings. Seven research gaps have been identified within the scope of this dissertation. These gaps were subsequently transformed into corresponding objectives and prerequisites leading towards the completion of the overarching goal. The goal itself has been formulated as to *develop and evaluate approaches for maximising influence in multilayer networks by leveraging selected paradigms that provide complementary perspectives on complex systems: a notion of actor centrality, controllability means, graph-based machine learning, and structural interventions on the analysed system.* The research presented here approaches this agenda through a collection of seven (P1-P7), thematically related publications, together with the binding part that demonstrates how each of them contributes to the problem under consideration.

The first research prerequisite: *identify various methodologies for modelling spreading phenomena in multilayer networks, and select a single mechanism under which all research objectives will be evaluated* has been addressed through the development of a framework (`network-diffusion`) that represents any configuration of coexisting phenomena by means of an n-dimensional grid. Furthermore, a recently established approach to modelling actor-based spread — protocol functions that govern how layer-level activations are aggregated — has been adapted to the multilayer setting and extended from the Linear Threshold Model to its counterpart, the Independent Cascade Model.

The subsequent one — to *develop metrics for quantifying spreading effectiveness in multilayer networks, enabling comparison of spreads triggered under varying conditions* — has resulted in a set of new measures designed to assess influence maximisation methods from both static

and dynamic perspectives. In addition, an evaluation approach capable of abstracting over specific spreading conditions has been proposed. Finally, a methodology for estimating the spreading capabilities of individual actors has been coined and employed accordingly.

The third prerequisite was formulated as to *provide multirelational data that can serve as a structural backbone for experiments on influence maximisation and develop methods capable of real-time augmentation.* It has been addressed by acquiring and publishing a new longitudinal dataset that captures the physical interactions within an isolated group of people. A representative collection of networks was also compiled and used throughout all simulations covering this doctoral research, thereby ensuring coherence. Moreover, a novel synthetic generator of power-law multilayer networks, `mABCD`, has been introduced, distinguished by its efficiency and its ability to model nuanced interlayer dependencies.

After implementing the three prerequisites, works aimed at investigating various perspectives on maximising influence were carried out. Within the following objective: *evaluate various rank refining methods for influence maximisation in multilayer networks and assess how they differ in terms of spreading effectiveness*, a comprehensive study of the most basic methods considered in this research was conducted. As a result, sixteen heuristics were evaluated, amongst them three (based on centrality metrics originally defined for monoplex networks) were extended to the multilayer scenario. Moreover, a new approach, *Neighbourhood Size Discount*, was proposed. The main conclusion was that all evaluated parameters: those pertaining to the spreading model, the network type, the seed budget, and the seed selection method, collectively determine the overall effectiveness of influence maximisation.

The consecutive objective was formulated as: *propose a seed-selection approach that leverages the minimal dominating set for the influence maximisation problem in multilayer networks and verify its effectiveness.* The work addressing this objective resulted in extending the concept of the minimal dominating set — a collection of agents through which one can control the entire system — to the actor-based standpoint. It was subsequently used to enhance the most effective rank-refining seeding methods to test whether triggering diffusion from the dominating set leads to greater network coverage. The results revealed that this method is significantly advantageous, although only under particular spreading conditions.

Another paradigm employed was machine learning. The study concerning this objective aimed to *develop an inductive graph neural network that predicts which actors in a given multilayer network are the most effective spreading seeds, and validate its performance against other state-of-the-art methods.* To achieve this, a dataset comprising the spreading potential of every actor within more than two hundred multilayer networks was created. Then, a graph neural network, *TopSpreadersNetwork*, was developed to perform the task of predicting these values. The main design goal was to make it capable of inferring from unseen network data, a feature uncommon in the domain of machine learning on graphs. The comparative study showed it is able to generalise and yields better results than competing methods.

The final objective realised within the scope of this dissertation was to *design a method to manipulate the structure of a multilayer network at a macro scale, enabling comparison of spreading effectiveness between the original and modified networks, and identify the structural properties that facilitate the influence maximisation process.* To this end, a conceptual study examining how the parameters of the `mABCD` model affect the network's spreading capabilities was performed. Moreover, the research resulted in a method to embed a real network into the configuration space of `mABCD`, which may be particularly relevant in the context of digital twins of complex networked systems. The results demonstrate, i.a., a strong impact of the intercommunity noise level on spreading effectiveness.

In conclusion, this dissertation presents multiple perspectives on the influence maximisation problem in multilayer networks. It also addresses several preliminary gaps, such as the modelling of spreading phenomena, whose resolution enables subsequent domain-specific studies. The results demonstrate that the problem in the considered network type exhibits greater complexity than its canonical form. The contributions of this work encompass advances in multilayer network research, including novel spreading and centrality metrics, new modelling approaches, methods for identifying the most influential actors, algorithms, datasets, and a synthetic graph generator. The central goal has been realised, albeit with some caveats arising from the limitations of the proposed methods.