Role Extraction, Dynamics, and Optimisation on Networks

Yu Tian

Mansfield College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Acknowledgements

Doing a DPhil has certainly been the biggest professional challenge that I have faced so far, and I would like to express my deepest thanks to many people for providing such an enjoyable environment in which to learn and discover.

First and foremost, I would like to thank my supervisor, Renaud Lambiotte, for his boundless encouragement and support. His enthusiasm and excitement for network science has inspired me in each of our weekly meetings. I am immensely grateful for his mentorship throughout my DPhil and invaluable influences over the work in this thesis.

This doctoral research is supported by the EPSRC centre for Doctoral Training in Industrially Focused Mathematical Modelling (EP/L015803/1), and I have greatly benefited from an excellent collaboration with Tesco. Particular thanks to my industrial supervisors, Alisdair Wallis and Sebastian Lautz, for actively engaging with their industrial insights in our biweekly meetings and their support throughout my DPhil. I am also grateful to the entire Tesco Data Science team, including Trevor Sidery and Stephen Logan who have provided valuable support in the time series analysis, for warmly hosting me during my visits at the Tesco offices.

I would like to thank Yong-Yoel Ahn for providing the flavour compound and recipe data to validate the product relationships, and also Karel Devrient and Lindon Roberts for useful discussions and comments on my research in dynamics and optimisation on networks. Generally, my thanks go to the entire research group of Networks. I thank my officemates Karel Devrient, Rodrigo Leal Cervantes and John Pougué-Biyong for many casual but fruitful discussions that have been full of interesting ideas. I thank Gesine Reinert, Alice Schwarze, Marya Bazzi, and Roxana Pamfil for their invaluable advice and support in shaping my career as a researcher. I also appreciate the friendship and moral support of other network scientists at Oxford, particularly Mary Sanford, Mattie Landman, Samira Barzin,

Alexandre Bovet, Erik Hörmann, Shazia'Ayn Babul, Sofia Medina, Timothy LaRock, Nicola Pedreschi and Eric Keiji Tokuda.

It has been a privilege working in the Mathematical Institute, as well as being part of InFoMM CDT. I am very grateful for the welcoming community of people that I met here. I would especially like to acknowledge Chris Breward and Colin Please for their hard work in running the InFoMM CDT, as well as Amanda Guthland Elmore, Sarah Howle, and Sophie Hurden for their kind support throughout the years.

Last but never the least, my loving family and friends are an inevitable part of this journey. I am especially grateful to my parents, Shirong Yun and Kai Tian, and my grandparents, Yuying Wang and Baoguo Yun, for their endless love and trust in me. I would not be where I am today without their care, patience, guidance and encouragement throughout every aspects of my life. I am also very grateful to my friend, Huining Yang, who is full of inspirational ideas either of research or of life, and I really appreciate her cherished time with me over the past four years. My final acknowledgement goes to my team 103, i.e. Yanan Yu, Yajie Chen, and particularly Jingyu Xu, for being a constant source of encouragement from the very beginning till the very end. You all make a whole world of difference to me.

Abstract

Understanding the relationships between products is an important problem in industry, where two central concepts are *complements* and *substitutes*. Complementary products are sold separately but used together, each creating a demand for the other, while substitute products can be used in place of each other, each taking a demand from the other. Hence, such two competing product relationships are also essential in further understanding how the demand change of some products transforms into the change of the whole range of products.

To start with, we capture the information in the sales transaction data through a bipartite network representation, then products of the two competing relationships can be characterised as certain *roles* in the network. For this reason, the identification of the nature of a relationship between products is directly related to the more general problem of role extraction in network science. Targeting at extracting the two specific roles, we propose a novel customised data-driven role extraction method. With the product relationships, we can consider unipartite networks only of products, and then the change of demand can be modelled as *dynamics* on top of such networks. Further, the associated decision making can be formulated as *optimisation* problems with appropriate objectives. In this thesis, we consider the one that is closely related to viral marketing, where we aim to maximise the overall revenue through putting a small range of products on promotion. We model the dynamic of demand changes as information propagating through the network, and the overall change as the influence on each node, thus the problem is then a variant of the *influence maximisation*, where our objective is a (price-)weighted sum of the influence. Accordingly, we propose a novel class of information propagation model for demand dynamics, unifying the mechanisms underlying the classic models, and a general framework for the associated influence maximisation problem, from simple networks of a single relation to signed networks of two competing relations.

Although inspired by the retail industry, the proposed role extraction method can be applied to other contexts, including trading networks, ecological systems and social networks, where both the identification of cooperative and competitive relations are of interest. Further, the proposed information propagation model and influence maximisation framework are also applicable to more general dynamical processes on either simple or signed networks, including social networks with trust-distrust relationships and brain networks with positive-negative functional connections.

Contents

Chapter 1

Introduction

1.1 Motivation

There are many fundamental problems in industry that are also of great research value. With our industrial partner Tesco, we are particularly interested in the retail industry. This sector is characterised by huge turnover but relatively small margins which together make the profitability very sensitive to small changes of products. If a potential customer visits a store and finds that it has a very limited range, they are likely to shop elsewhere, which leads to the belief that stores should stock the greatest diversity of products possible. However, stock has to be sourced, distributed, and handled in stores, which increases costs. In order to find the optimal range of products to stock in a store, one primary problem would be to understand the intrinsic relationships between products, including *complements* and *substitutes*. Complementary products are sold separately but used together, each creating a demand for the other, such as hot dog and hot dog bun, while substitute products serve the same purpose and can be used in place of each other, such as apples of different varieties. Once such relationships have been estimated, a follow-up problem would be to understand the *interrelated demand changes* between products, which directly affects revenue or profits. After this, retailers could try to *make decisions* based on such understandings, e.g. which products should they put on promotion so as to maximise the overall revenue at the end of the promotional period.

To answer both questions, we resort to a topological structure emphasising the connection between entities, *networks* [122]. Networks are mathematical representations of systems in which entities, called nodes, interact with each other via edges, typically in a pairwise fashion. Examples of networks range from human brains to

social networks, from the World Wide Web to Earth's climate [44,60,129,159]. Specifically, sales transaction data can be modelled as a *bipartite product-purchase network*, where transactions (or baskets) are one type of nodes, products are another, and they are connected if a product appears in a transaction. One advantage of using a network representation, compared with simply tabular data, is that it considers not only pairwise relations but also indirect connectivity. Another outstanding advantage of networks is the inherited high-dimensional non-Euclidean information, which cannot be incorporated by embedding-based machine learning methods. In addition, in order to analyse demand change between products, it is necessary to extract the inter-relationships between products from the bipartite network, and to model the cascading effect of demand changes on the corresponding *product networks*.

Starting from the bipartite network representation, products of the two competing relationships can be characterised by product nodes exhibiting specific connectivity patterns, thus having certain *roles* in the network. For this reason, the identification of the nature of a particular relationship between product nodes is directly related to the more general problem of role extraction in network science [26,80,104,155]. Once the product relationships have been extracted, the change of demand can be modelled as *dynamics* on top of the product networks where the products are connected by the corresponding relationships, and the associated decision making processes can be formulated as *optimisation* problems with appropriate objectives. In this thesis, we consider a formulation of the problem that is closely related to the one in viral marketing, specifically to select a small portion of products to put on promotion in order to maximise overall revenue. Specifically, we consider the dynamic of demand changes as the *propagation* of information through the network [85,86,113,125,126], and the overall change as the influence on each node, thus the problem to maximise the sales or revenue can be modelled as a variant of the classic *influence maximisation* problem [85,86,95,96,142].

1.2 Role extraction

Roles in network science represent node-level connectivity patterns, such as star-centre nodes, star-edge nodes and bridge nodes. Different roles can also correspond to different functions in certain contexts, and one of the examples is the role of cities in air transportation networks [70]. There are many equivalence conditions proposed in the literature in order to define when two nodes belong to the same role, including the very first and restrictive one, structural equivalence [104], and later relaxations,

regular equivalence [155] and stochastic equivalence [80]. Furthermore, there is also another line of research in structural similarities where, instead of making the binary decision whether two nodes are in the same role or not, researchers aim to measure how close the nodes are to being in the same role according to certain equivalence relations [26]. The latter is relatively more flexible thus more applicable when the network is constructed from noisy real data, hence is the path we will follow in this thesis.

Extracting roles in networks cannot only be defined at the node level, where one aims to determine whether two nodes are in the same role or not, but also at the mesoscale level, where groups of nodes are considered. Hence, the role extraction problem generalises the classic problem of community detection [62,63,128], since the latter specifies the patterns to be densely connected internally but loosely connected externally or clique-like, while roles could correspond to any connectivity patterns. Characterising the mesoscale structure of networks, in terms of either roles in general or communities as a specific example, is an essential part of network science, and finds applications in various domains [121].

Specifically, numerous efficient algorithms exist for detecting communities, including greedy methods, statistical inference, approaches based on dynamics or information theory, and machine learning algorithms [91,119,127,132,149]. Therefore, most state-of-the-art methods solve the role extraction problem indirectly by transforming general roles to be communities or clusters. Among them, Browet and Van Dooren proposed a structural similarity measure based on flow patterns of all possible lengths [26], while Figueiredo et al. developed a node structural embedding from the auxiliary context graphs constructed by the similarity between nodes at different distances [131]. However, the primary focus of such work is on extracting groups of nodes sharing generally similar connectivity patterns, rather than the specific ones corresponding to the complementary and substitute relationships. Hence, as a first contribution, we propose a novel customised role extraction method targeted at extracting the two competing roles at the mesoscale level. Then, with the sales transaction data provided by Tesco, we are able to estimate the two competing relationships between the products, complements or substitutes. Finally, we validate the extracted product relationships with several external datasets characterising products from other perspectives.

Although inspired by the retail industry, the proposed method can also be applied to other contexts. For example, it can in principle be applied to semantic networks in order to identify synonyms, and also flavour networks, so as to explore ingredients

that go well with each other or are effectively interchangeable. Furthermore, the problem can be generalised to extracting the collaborative and competitive relationships between entities, which commonly exist in trading networks, ecological networks, and social networks among others.

1.3 Dynamics and optimisation

Understanding how information spreads in social networks is a central theme in social, behavioural, and economic sciences, with important theoretical and practical implications, such as the influence maximisation (IM) problem for viral marketing. Information propagation is thus a vast research domain, attracting expertise from various fields including mathematics, physics, and biology [113,125,126]. Two widely adopted models are the independent cascade (IC) model where nodes adopt their behaviour from each neighbour independently, and the linear threshold (LT) model where collective effort from the whole neighbourhood is needed to influence a node. However, both models suffer from certain drawbacks, including a binary state space, where nodes are either active or not, and absence of feedback, as nodes cannot be influenced after having been activated. These limits call for more general models for explaining information propagation, and also particularly for them to be applicable to describing the dynamic of demand changes. Hence, our first contribution here is to introduce a non-linear, deterministic model for information propagation which relaxes the constraints of binary variables while allowing feedback between states, and it also has the additional advantage of unifying the two classic models, as the IC model and the LT model after appropriate extensions are recovered by setting the corresponding model parameters.

For the associated IM problem, it is well-known to have potential applications in many domains [95, 96, 142]. Given a network and an associated information propagation process, the classic IM problem consists in selecting a small set of nodes to activate initially with the aim of maximising the overall *influence spread*, commonly defined as the number of activated nodes at the end of the process [67, 85, 86, 96]. However, the theoretical performance guarantee of most approximation algorithms are obtained under certain assumptions on the influence spread function, which could be strict in practice [101]. Hence, our second contribution here is to propose a general framework for the IM problem, where we formulate the problem in terms of a mixed integer nonlinear programming and adopt derivative-free methods as general

solutions. Furthermore, we show that the problem can be exactly solved in one special case of the proposed information propagation model where the selection criterion is closely related to the Katz centrality, and then propose a customised algorithm specifically for the proposed model, with local convergence.

As the first step, we only take into account a single relationship between nodes, i.e. simple networks with positive edges. However, negative connections can be as significant as the positive ones in shaping the dynamics. We also note that our interest in information propagation and influence maximisation originates from analysing interrelated demand changes, where both the complementary and substitute relationships between products are important. Hence in the next step, we incorporate the negative connections and analyse the dynamics subject to the two competing relationships simultaneously, via *signed networks*.

1.4 Signed networks

Using a positive or negative sign to represent two competing types of node interactions, the structure of these systems can be modelled as signed networks. Such competing types of interactions often play an important part in shaping the system behaviour, e.g., activatory or inhibitory functions in biological systems, trustworthy and untrustworthy connections in social or political networks, and cooperative or antagonistic relationships in the economic world [4, 58, 109]. In contrast to the rich literature in simple networks, more can be explored in the context of signed networks, in terms of both structural properties and dynamical implications. Hence, our first contribution here is to extend the results in characterising either simple networks or specific types of signed networks to signed networks in general.

For the problem of information propagation and influence maximisation, most of the existing studies only consider simple networks with positive connections, and the problem in signed networks is relatively unknown. Among such work, most researchers follow the path to extend the two classic models, the IC model and the LT model, to signed networks while maintaining the desired properties of the associated IM problem [38, 77]. However, the expected influence spread in signed networks becomes increasingly difficult to estimate, while the classic models suffer from certain drawbacks as discussed before. Hence as a following contribution, we extend the previously developed general class of information propagation model to signed networks, where the model still possesses the classic models as limiting cases, and then analyse the associated IM problem. We also propose a customised method when there is only

one choice of initial state for each node, which is closely related to product range optimisation. Finally, to complete the story of the whole thesis, we demonstrate how the three main chapters are connected with each other on the one hand, and also how the proposed methods could be useful in practice on the other hand, through an illustrative example. Although we use the retail context as a specific example throughout the thesis, we believe the set of methods we have developed in this thesis are generally applicable to the problems of extracting the two competing types of relationships from networks and also characterising dynamics on simple or signed networks.

1.5 Organisation of the thesis

The rest of the thesis is organised as follows. In Chapter 2, we discuss in detail the customised role extraction method to extract the two competing types of roles in bipartite networks, corresponding to the complementary and substitute relationships between products. This chapter is accompanied by real data, where we describe the different data sources we use, how we connect them, and the exact process of validation. Chapter 2 focuses on the structural properties of networks, while Chapters 3 and 4 emphasise the dynamical patterns. In Chapter 3, we start with simple networks solely composed of positive connections, and propose a general class of information propagation model, unifying the widely accepted mechanisms underlying the classic models, the IC model and the LT model, and a general framework for the IM problem. Based on the features of the proposed model, we also develop a customised algorithm for the IM problem, with local convergence. In Chapter 4, we proceed to signed networks of both positive and negative connections and start by characterising their structural and dynamical properties. We then extend the general class of information propagation model we have proposed for simple networks to signed networks, and analyse the associated IM problem. More importantly, we demonstrate how the techniques we have developed in this thesis connect with each other through an illustrative example at the end of this chapter. Finally, we summarise the work in this thesis and suggest a few topics for future research in Chapter 5. In the Appendix, we further discuss bipartite configuration models, Poisson binomial distribution and real data from Chapter 2, and explore the time complexity of the proposed method and stochastic block models from Chapter 3.

Chapter 2

Extracting two competing roles in bipartite networks

This chapter is based on the paper Tian, Lautz, Wallis & Lambiotte (2021) [148].

2.1 Introduction

2.1.1 Two competing relationships between products

Understanding the hidden relations existing between products is fundamental in both economics and marketing research as well as in retail [55]. This question lies at the core of market structure analysis and finds numerous applications. Retailers must regularly make decisions taking product relationships into account [108], for instance to design their product catalogue and to determine the number of products to offer in each category [87]. Brick-and-mortar retailers seek to identify the best way to arrange the product layout in aisles and stock their shelves [151], and online retailers also strive to optimise the grouping of products in their online shops [24]. Furthermore, they must decide which products to bundle or promote together. These assortment-related decisions have significant influence on customers' choices, sales of products, and finally, profits [25, 87, 108].

Complements and *substitutes* are two central concepts to characterise two competing types of relationships between products, with well-established definitions in economics [124]. Complementary products are sold separately but used together, each creating a demand for the other, such as hot dogs and hot dog buns. Substitute products serve the same purpose and can be used in place of one another, such as Brand A tomatoes and Brand B tomatoes. In the economics literature, the degree

of complementarity (substitutability) is formally defined through the negative (positive) cross-price elasticity of demand, where the rise in the demand of one product is recorded after the price of the other product is reduced (increased) by a unit. The mechanisms of complements and substitutes are also referred to as the *halo effect* and *demand transfer*, respectively, in the retail context [3].

Despite its practical importance, the algorithmic problem of identifying the relationships between products in retail is not well known. For a long time, researchers and practitioners have selected the set of possible complementary or substitute products by means of, for instance, field expertise and simple statistics, and the analysis has usually been restricted to a fairly small number of products [17, 141]. Recent development and application of natural language processing and machine learning (ML) algorithms, especially those based on word embedding, bring in new visions and opportunities, which makes it possible to analyse thousands of products [35, 64, 133]. However, the interpretation of the selected features in the related ML algorithms is difficult, and these methods lack specific criteria to determine whether two products are complements, substitutes, or just independent. Furthermore, these methods are based on co-purchase patterns, but other valuable information in the sales data remains unused.

2.1.2 A network perspective: role structure

While there is intrinsic network structure underlying the sales transaction data, network modelling has not yet been considered for understanding the relationships between products. Hence in this chapter, we propose an alternative to the classic approach based on cross-price elasticity, and take instead a network perspective in order to define the two competing relationships between products. As a starting point, we model the sales data as a bipartite *product-purchase network*, with both transactions (or baskets) and products as nodes. Then the analysis can be directly performed on the network, without having to rely on low-dimensional embeddings which may lead to uncontrolled loss of information. Specifically, we formulate the problem of identifying the complementary and substitute relationships between products as extracting product nodes with specific connectivity patterns from this bipartite network, i.e. the role structure in one part of the nodes, both at the local level in terms of pairwise relationships and at the mesoscale level in terms of node groups.

Roles correspond to specific connectivity patterns, such as star-centre nodes, star-edge nodes and bridge nodes. Two nodes belong to the same role if they are equivalent

according to the predefined equivalence based on their structure. The first such equivalence is the structural balance introduced by Lorrain and White [104], where two nodes are structurally equivalent if they have exactly the same neighbours. Many relaxations have been considered since then, including regular equivalence [155] and stochastic equivalence [80]. For practical applications, structural similarity measures have been proposed accordingly to quantify how close two nodes are with respect to certain equivalence relations [26]. This is the path we will follow in this chapter, but instead of characterising general equivalent nodes as in most literature, we consider the specific roles concerning the complementary and substitute relationships between the nodes in one part of bipartite networks. Specifically, we define null models on bipartite networks to determine significant relationships between the nodes in one part, and also propose similarity measures induced by random walks on networks in order to quantify the intensity of these relationships.

The problem at the local level also lies in the broad family of *bipartite network projection* where one projects a bipartite network onto one part of the nodes to form unipartite ones [161]. Different strategies exist depending on the nature of the relationship that one wants to infer [94, 99, 115, 116, 118, 161]. A majority of works look for assortative relations, in the sense that two nodes are connected in the unipartite network if they tend to share many neighbours in the bipartite one. However, more general types of projections can be defined, which are associated with the role played by the nodes in the original bipartite network. Hence, to extract the two competing relationships can be considered as an extension to the classic bipartite network projection, and a first step to explore general relationships among nodes.

To extract roles at the mesoscale level is a general version of the community detection problem, where communities specifically refer to groups of nodes that are densely connected internally but sparsely connected externally [62, 63, 128]. Various efficient algorithms exist by virtue of interdisciplinary expertise [91, 119, 127, 132, 149], generally aiming to optimise a quality function with respect to different partitions of the network. Therefore, most state-of-the-art methods solve the role extraction problem indirectly by transforming general roles to be communities. For example, Browet and Van Dooren proposed structural similarities based on flow patterns of all possible lengths in [26], while Figueiredo et al. developed a node structural embedding from the auxiliary context graphs in [131]. However, the aforementioned general methods are not applicable to extracting the roles of complements versus the roles of substitutes, where they have similar connectivity patterns globally but subtle differences

locally. Nevertheless, with the previously developed null models and similarity measures, the problem of extracting the two competing roles can be effectively solved by community detection algorithms. Hence, the whole framework we propose establishes a customised approach for this specific role extraction problem at the mesoscale level.

Even though inspired by the competing relationships between products for retailers or general firms with broad product categories, the set of methods we propose can also be applied to other contexts. For instance, one can consider recipe data, and extract the ingredients that go well with each other and those that are effectively substitutes. Furthermore, the problem can also be generalised to the identification of both cooperative and competitive relations which commonly exist in trading networks, ecological systems and social networks among others.

2.2 Data

In this section, we present the different datasets that we use, the sales data in Sec. 2.2.1 to construct the bipartite network for the task of extracting the two competing roles, the product hierarchy data in Sec. 2.2.2 and the flavour compound and recipe data in Sec. 2.2.3 to validate the results.

2.2.1 Sales data

We used anonymised grocery sales data from Tesco, the UK's largest supermarket chain. The data consists of timestamped transactions of stores, and it has been anonymised for general research purposes, i.e. each customer's personal identifiable information has been removed. For each store, the transaction data comprises a transaction ID, which gives a unique code to each shopping trip, the date when the transaction was made, the product IDs, and their purchased quantities; see the top of Fig. 2.1.

The data used for this chapter is from a generic convenience store in an urban area, and spans a three-month period avoiding major holidays such as Christmas and Easter. The time window is chosen to be long enough to be representative of the underlying customer population's product purchase patterns, but also sufficiently short to avoid seasonal effects as well as change of behaviour over time. Furthermore, to facilitate the interpretation of the results, we restrict the analysis to fresh fruit, vegetables and salads where we believe complementary and substitute products commonly exist. We also exclude products that are purchased less than once a month,

and those in almost every transaction. These result in the final dataset of 43,837 transactions and 253 products.

2.2.2 Product hierarchy data

In retail, it is common to organise products in a hierarchy, where similar products are grouped into increasingly generic categories. Products that are close together in the hierarchy are typically sold next to each other in a store. At the lowest hierarchical level, each unique code corresponds to a different product, including the same products of different sizes or flavours. Overall, we have 4 levels, from `L1` to `L4` (excluding the product level). The higher a level is, the more generic the corresponding category. For example, "apple" is a category in the `L1` hierarchy, and "fruit" is a category in the `L3` hierarchy. Hence, a natural way to validate the extracted product relationships and to explore their features would be to compare them to the corresponding product hierarchy.

2.2.3 Flavour compound and recipe data

Ahn et al. [2] provided a systematic list of 1107 flavour compounds and their natural occurrences in terms of 1,525 ingredients overall from Fenaroli's handbook of flavour ingredients [27]. They also provided 56,498 recipes belonging to geographically distinct cuisines (North American, Western European, Southern European, Latin American and East Asian), which were obtained from *epicurious.com*, *allrecipes.com* and *menupan.com*; see the bottom of Figure 2.1. Hence, to validate the extracted product relationships from their features in terms of both flavour compounds and recipes, we match the products in the sales data to the ingredients that the flavour compound and recipe data is based on.

To construct the correspondence between the products and the flavour compounds, we match each product to as many ingredients as possible. For example, "Loose Peppers" is matched to all possibly equivalent peppers including "bell pepper" and "green bell pepper"; see the middle left of Fig. 2.1. This results in the relevant ingredients to be 140, with their corresponding flavour compounds being 865, and each ingredient is linked to 57 flavour compounds on average. Note that there are 11 products which do not have exactly matched ingredients, hence we match them to generic ones[1]. For example, we match the product "Single Pomegranate" to the ingredient "fruit". There are also 44 complex products whose ingredients cannot be

[1]In this data, there are both specific ingredients, e.g. "apple", and generic ones, e.g. "fruit".

directly inferred from their names, thus we match them to their main ingredients on the website. For example, we match the product "Cheddar Coleslaw" to the ingredients "cheddar cheese", "cabbage", "carrot" and "onion".

For the recipe data, we match each product to as few and simple ingredients as possible. For example, "Loose Peppers" is now only matched to "bell pepper"; see the middle right of Fig. 2.1. We then restrict to the products only corresponding to one ingredient, and also remove products that are matched to unrepresentative generic ingredients (e.g. "vegetable"). We take a (generic) ingredient to be unrepresentative, if it shares less than half of its flavour compounds with the ingredients in the same category. As an example, if the product "Loose Aubergine" were only matched to the generic ingredient "vegetable" which shared less than half of its flavour compounds with all other vegetable ingredients, such as "asparagus", "lettuce" and "onion", we would exclude this product. This further reduces the number of relevant products and ingredients to be 175 and 69, respectively, with $47,222$ corresponding recipes, and each recipe being expected to contain 3 such ingredients.

2.3 Two competing roles in bipartite networks

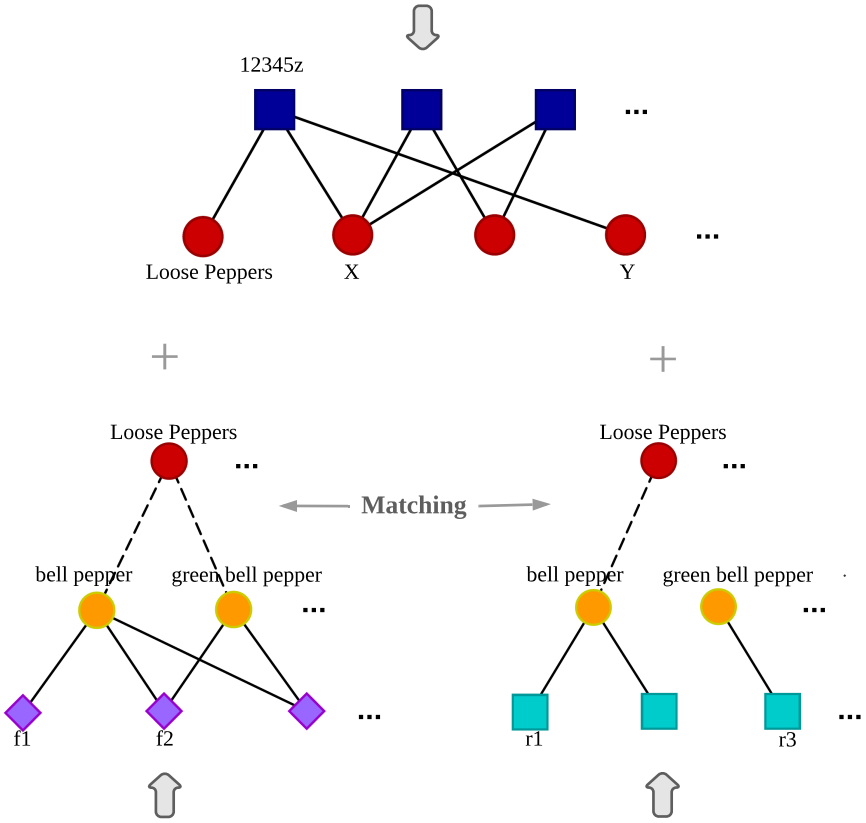
2.3.1 Bipartite product-purchase network

We model the structure in the sales transaction data, after appropriate preprocessing as discussed in Sec. 2.2.1, as an undirected unweighted bipartite network $G = (V, E)$, where $V = \{u_1, \ldots, u_{n_t}, v_1, \ldots, v_{n_p}\}$ is the node set with the l-th transaction as node u_l and the i-th product as node v_i, $\forall l \in \{1, \ldots, n_t\}$ and $i \in \{1, \ldots, n_p\}$, n_t, n_p are the numbers of transactions and products, respectively, and $E = \{(u_l, v_i) : \text{the } i\text{-th product is purchased in the } l\text{-th transaction}\}$ is the edge set; see Figs. 2.1 (top) and 2.2. The network is bipartite because we can find a bipartition $V_1 = \{u_1, \ldots, u_{n_t}\}$ consisting of all transaction nodes, and $V_2 = \{v_1, \ldots, v_{n_p}\}$ consisting of all product nodes, s.t. $V_1 \cup V_2 = V$, $V_1 \cap V_2 = \emptyset$, and $E \subset V_1 \times V_2$ where $\times$ represents Cartesian product. We refer to it as the *product-purchase network*, and aim to extract the two competing roles corresponding to the complementary and substitute relationships between the nodes in the same subset, here V_2 of the product nodes, from this network.

Throughout this chapter, we consider the biadjacency matrix $\mathbf{A}^{(b)} = (A^{(b)}_{li}) \in \{0, 1\}^{n_t \times n_p}$ for the structure of product-purchase network, where $A^{(b)}_{li} = 1$ if transaction node $u_l \in V_1$ is connected with product node $v_i \in V_2$ and 0 otherwise.

Store ID	Date	Transaction ID	Product	Quantity
123	01/05/18	12345z	Loose Peppers	1
123	01/05/18	12345z	X	1
123	01/05/18	12345z	Y	2
…	…	…	…	…

2.1 Sales data



Ingredient	Flavour compound
bell pepper	f1
…	…
green bell pepper	f2
…	…

Ingredient	Recipe
bell pepper	r1
…	…
green bell pepper	r3
…	…

2.3 Flavour compound and recipe data

Figure 2.1: Schematic diagram showing the data structure of the sales data in Sec. 2.2.1 and the flavour compound and recipe data in Sec. 2.2.3, and their corresponding bipartite networks, together with the matching between the products and the ingredients. The "Matching" step is required because of the different names that can appear in different datasets.

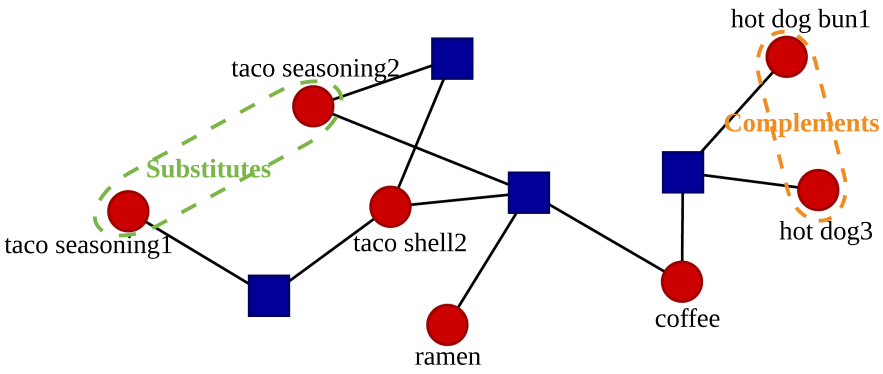


Figure 2.2: Example of a bipartite product-purchase network, where blue squares are transaction nodes, red circles are products nodes and these two sets of nodes are connected if the product is purchased in the corresponding transaction. The underlying sales data contains both complements (e.g. hot dog3 and hot dog bun1) and substitutes (e.g. taco seasoning1 and taco seasoning2).

2.3.2 Role characteristics

To characterise the two competing roles in bipartite networks, we start from interpreting the complementary and substitute relationships between products in terms of their purchase patterns. Specifically, in the context when prices change frequently, complements can be identified through sufficient co-purchases [9], while substitutes have almost no co-purchases. The feature of substitutes that they have similar interactions with other products is commonly used in practice [35, 133], and combined with the almost-no-co-purchase characteristics, it can be used to determine the substitute relationship. Note that the formal definition through cross-price elasticity is expected to emerge from such purchase patterns, where, for example, two products always purchased together implies that the decrease in one's price will result in an increase of the other's demand[2].

We then characterise the complementary and substitute relationships between the nodes in the same subset, $v_i, v_j \in V_2$, through the number of common neighbours they share in the product-purchase network, $cn_{ij} = \left|\{u_l : A^{(b)}_{li} A^{(b)}_{lj} = 1\}\right|$, as follows.

(1) Complements are products (v_i, v_j) that are in the same transactions (cn_{ij}) significantly more frequently than expected.

(2) The degree of complementarity between complements (v_i, v_j satisfying (1)) is positively correlated with how frequently they are in the same

[2]The demand of a product is generally a decreasing function of its own price. This statement is true for the products analysed here.

transactions.

(3) Substitutes are products (v_i, v_j) that share the same complements but are in the same transactions (cn_{ij}) significantly less frequently.

(4) The degree of substitutability between substitutes (v_i, v_j satisfying (3)) is positively correlated with how similar their complements are.

In addition, we define *noise* to be the purchase patterns that are caused by other, often unknown, factors and cannot be explained by complementarity and substitutability. Hence to capture the intrinsic relationships between the nodes and their degrees, it is essential to explain and control the noise appropriately. In networks, *local structure* usually refers to the information around a node, and *global structure* characterises the whole network. For intermediate scales, one often refers to the notion of *mesoscale structure*, which is associated to groups of nodes. Within our context, we consider particularly the community structure, where groups of nodes are densely connected internally but sparsely connected externally, and we exploit the fact that the mesoscale structure is much more robust to noise than the local information [53]. Hence, we further propose the following assumptions **(5)** and **(6)** to restrict the noise effect.

(5) Noise will not change the community structure, i.e. groups of nodes that are mostly complementary or substitute to each other.

(6) Noise can be explained by some random models so that its effect can be removed accordingly.

With these characteristics, the two competing roles at the mesoscale can be specified in the product-purchase network as follows. The roles corresponding to the complementary relationship, referred to as the *complement roles*, are groups of nodes that share the particular similar connectivity patterns as described in characteristics **(1)** and **(2)**. Similarly, the roles corresponding to the substitute relationship, referred to as the *substitute roles*, are groups of nodes that share again similar but another connectivity patterns as described in characteristics **(3)** and **(4)**.

2.4 Customised role extraction method

In this section, we propose a customised method for extracting the complement roles and substitute roles, based on the characteristics in Sec. 2.3.2. Specifically, we propose

several null models in Sec. 2.4.1 to determine the significant relations, and measures based on random walks on networks to quantify the degrees in Sec. 2.4.2. Finally, we discuss how we can extract the two competing types of roles through classic community detection algorithms in Sec. 2.4.3.

2.4.1 Null models

We propose the following two null models on the bipartite product-purchase network, to determine the significance of the number of common neighbours, cn_{ij}, between each pair of nodes v_i, v_j. Then, we can construct two unweighted unipartite networks only consisting of the nodes in V_2: (i) $\mathbf{A}^{(m)} = (A_{ij}^{(m)}) \in \{0,1\}^{n_p \times n_p}$ where $A_{ij}^{(m)} = 1$ if and only if cn_{ij} is significantly more than expected; (ii) $\mathbf{A}^{(l)} = (A_{ij}^{(l)}) \in \{0,1\}^{n_p \times n_p}$ where $A_{ij}^{(l)} = 1$ if and only if cn_{ij} is significantly less than expected. Finally, by characteristics **(1)** and **(3)** in Sec. 2.3.2, we can obtain the networks indicating the existence of the two competing relationships: (i) $\mathbf{A}^{(c)} = (A_{ij}^{(c)}) \in \{0,1\}^{n_p \times n_p}$ where $A_{ij}^{(c)} = 1$ if and only if nodes v_i, v_j are complementary to each other; (ii) $\mathbf{A}^{(s)} = (A_{ij}^{(s)}) \in \{0,1\}^{n_p \times n_p}$ where $A_{ij}^{(s)} = 1$ if and only if nodes v_i, v_j are substitute to each other.

Variant of Bipartite Erdős-Rényi (ER) Models. The ER model assumes a fixed probability for each edge to appear, independently of the others [56], while bipartite ER models only allow edges between the two subsets of nodes.

In our variant, we assign a different connecting probability p_i for each node $v_i \in V_2$. Then, the probability that a node in V_1 is connected with both nodes $v_i, v_j \in V_2$ is $p_i p_j$; the number of their common neighbours, cn_{ij}, is a random variable X_{ij}, s.t. $X_{ij} \sim B(n_t, p_i p_j)$. We further assume n_t is sufficiently large, and approximate the distribution by the normal distribution $N(\mu_{ij}, \sigma_{ij}^2)$, where $\mu_{ij} = n_t p_i p_j$, $\sigma_{ij}^2 = n_t p_i p_j (1 - p_i p_j)$, from the Central Limit Theorem [68]. Hence, cn_{ij} is significantly more than expected if

$$cn_{ij} > \; n_t p_i p_j + \Phi^{-1}(1 - \alpha_m)\sqrt{n_t p_i p_j (1 - p_i p_j)}, \tag{2.1}$$

and is significantly less than expected if

$$cn_{ij} < \; n_t p_i p_j - \Phi^{-1}(1 - \alpha_l)\sqrt{n_t p_i p_j (1 - p_i p_j)}, \tag{2.2}$$

where α_m, α_l are the corresponding significance levels, $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of $N(0,1)$ with $\Phi^{-1}(x) = \sqrt{2}\,\mathrm{erf}^{-1}(2x - 1)$ and $\mathrm{erf}(\cdot)$ being the

error function, and the maximum likelihood estimate for each p_i is

$$\hat{p}_i = \frac{d_i^{(2)}}{n_t}.$$

where $d_i^{(2)}$ is the degree of product node i.

Bipartite Configuration Models (BiCMs). The configuration model creates a network with a given degree sequence $\{d_i\}$, by assigning d_i half-edges (or stubs) to each node v_i and joining two chosen stubs uniformly at random until no more stubs are left [120, 123]. The BiCM takes the bipartite features into account, where two degree sequences are given, dividing the nodes into two subsets, and edges are only allowed between the two subsets of nodes. Note that multi-edges are allowed here, but since we assume finite variance in both degree distributions, their portion is negligible in large networks (see Appendix A.1.1 for details).

We assume a cutoff on the node degree, thus the maximum degree is bounded. Then the probability of nodes $v_i, v_j \in V_2$ sharing a transaction node l can be approximated by

$$p_{ilj} = \frac{d_i^{(2)} d_l^{(1)} d_j^{(2)} (d_l^{(1)} - 1)}{m^2},$$

where the superscripts $1, 2$ stand for the nodes in V_1, V_2, respectively, $d_h^{(\cdot)}$ is the degree of node u_h or v_h, and $m = \sum_{l=1}^{n_t} d_l^{(1)} = \sum_{i=1}^{n_p} d_i^{(2)}$ is the number of edges (see Appendix A.1.1 for details). The variant of bipartite ER models can be seen as an approximation of this model, where we assume that the degree of each transaction node is constant.

The number of common neighbours between nodes $v_i, v_j \in V_2$, cn_{ij}, is the sum of $Bernoulli(p_{ilj})$ over l, where p_{ilj} possibly varies for different transaction node $v_l \in V_1$. We assume independence between different transaction nodes to connect with them both. Hence, cn_{ij} is a Poisson binomial random variable, X_{ij}, with the mean value approximately

$$\mu_{ij} = \sum_{l=1}^{n_t} \frac{d_i^{(2)} d_l^{(1)} d_j^{(2)} (d_l^{(1)} - 1)}{m^2} = \frac{d_i^{(2)} d_j^{(2)}}{m} \frac{\langle d^{(1)2} \rangle - \langle d^{(1)} \rangle}{\langle d^{(1)} \rangle},$$

where $\langle d^{(1)} \rangle = (\sum_{l=1}^{n_t} d_l^{(1)})/n_t$, and $\langle d^{(1)2} \rangle = (\sum_{l=1}^{n_t} d_l^{(1)2})/n_t$.

If the composing Bernoulli probabilities, p_{ilj}, are sufficiently small, the Poisson binomial distribution can be well approximated, with an exact error bound, by a

Poisson distribution with the same mean [13] (see our particular case in Appendix A.2). Following the Poisson approximation, and suppose $Y_{ij} \sim Poisson(\mu_{ij})$, we determine cn_{ij} to be significantly more than expected if

$$1 - F_{ij}(cn_{ij}) < \alpha_m,$$

and to be significantly less than expected if

$$F_{ij}(cn_{ij}) < \alpha_l,$$

where $F_{ij}(y) = e^{-\mu_{ij}} \sum_{k=0}^{\lfloor y \rfloor} \mu_{ij}^k / k!$ is the cumulative distribution function of Y_{ij}, and $\lfloor y \rfloor$ gives the largest integer that is no greater than y. Furthermore, if the composing probabilities, p_{ilj}, are close to each other for different transaction node v_l, or equivalently, the degree of transaction nodes is approximately constant, the Poisson binomial distribution can be well approximated by the normal distribution with the same mean and variance [68]. As we mentioned before, the bipartite configuration model is close to the bipartite ER model in this case (cf. Eqs. (2.1) and (2.2)).

The two null models are proposed to explain the connections in bipartite networks purely from noise; with more information about the noise factors, one can propose more customised null models accordingly. Currently, our null models are only based on difference in popularity of nodes in V_2, and the BiCM also uses the heterogeneity of nodes in V_1: both are sufficiently general to incorporate additional noise factors, but could possibly not be sufficient in their current form, as hidden factors, e.g. correlated preference, could cause more common neighbours between nodes in the product-purchase networks. Hence, by characteristic **(5)** in section 2.3.2, we accompany these null models with extra rules of significance-level selection: (i) α_m is chosen to be the smallest value that maintains the same community structure as that obtained from a baseline significance level, to exclude the above spurious signal; (ii) α_l is chosen to be the largest such value, in order not to accidentally filter out genuine patterns.

Finally, we can obtain the unweighted network of complementary relationship, $\mathbf{A}^{(c)}$, and that of substitute relationship, $\mathbf{A}^{(s)}$. By characteristic **(1)**,

$$\mathbf{A}^{(c)} = \mathbf{A}^{(m)}; \tag{2.3}$$

by characteristic **(3)**,

$$\mathbf{A}^{(s)} = \mathbf{I}_{\{\mathbf{A}^{(m)T}\mathbf{A}^{(m)}>0\}} \odot \mathbf{A}^{(l)}, \tag{2.4}$$

where $\mathbf{I}_{\{\cdot\}}$ is the element-wise indicator matrix, and $\odot$ represents element-wise (Hadamard) matrix product.

2.4.2 Measures

The degree of complementarity and substitutability matters. A significant relationship is not necessarily a strong relationship, and stronger relationships should be given higher weights to be more dominant in the networks. By feature **(2)** in Sec. 2.3.2, the degree of complementarity between each pair of nodes $v_i, v_j \in V_2$ is not directly correlated with how significant cn_{ij} is; by feature **(4)** in Sec. 2.3.2, neither is the degree of substitutability, which causes the results in Sec. 2.4.1 not to be applicable here. Hence, in this section, we further propose measures to quantify both degrees, in order to convert the unweighted unipartite networks, $\mathbf{A}^{(c)}$ and $\mathbf{A}^{(s)}$, to weighted ones, $\mathbf{W}^{(c)}$ and $\mathbf{W}^{(s)}$, respectively, where $\mathbf{W}^{(c)}, \mathbf{W}^{(s)} \in [0,1]^{n_p \times n_p}$.

Measures for complementarity. We propose several measures for the degree of complementarity by interpreting feature **(2)**: for each pair of nodes $v_i, v_j \in V_2$, the more similar their neighbours are in the product-purchase network, the more complementary they are.

We start from an *enhanced version* of feature **(6)** in Sec. 2.3.2: the noise factors change frequently and erratically so that their bias on the relative number of common neighbours between nodes can be neglected. We then propose the *original measure*, motivated by the weighted cosine similarity between random walkers on networks,

$$sim_o(i,j) = \sum_{l=1}^{n_t} \frac{A_{li}^{(b)} A_{lj}^{(b)}}{d_l^{(1)} \sqrt{(\sum_{h=1}^{n_t} \frac{A_{hi}^{(b)}}{d_h^{(1)}})(\sum_{h=1}^{n_t} \frac{A_{hj}^{(b)}}{d_h^{(1)}})}}, \tag{2.5}$$

where $\mathbf{A}^{(b)} = (A_{li}^{(b)})$, $\{d_l^{(1)}\}_{l=1}^{n_t}$ and n_t are the same as before (see Appendix A.3 for the detailed derivation). Note that we exclude the case when v_i or v_j or both are isolated, where the denominator is 0, and set the value to be 0 if so. Hence, each common neighbour $u_l \in V_1$ with $A_{li}^{(b)} A_{lj}^{(b)} = 1$ between each pair of nodes $v_i, v_j \in V_2$ in the product-purchase network is discounted by the degree of u_l, and this quantity is further scaled so that each node is at the maximum level of complementarity to itself, i.e. value 1, in a symmetric manner. A higher value means relatively more common

neighbours of lower degrees. We also propose the *original directed measure*,

$$sim_{od}(i,j) = \sum_{l=1}^{n_t} \frac{A_{li}^{(b)} A_{lj}^{(b)}}{d_l^{(1)} (\sum_{h=1}^{n_t} \frac{A_{hj}^{(b)}}{d_h^{(1)}})}, \tag{2.6}$$

where the normalisation is only w.r.t. node v_j, and each (i,j) entry measures the degree of complementarity of node v_i to node v_j in V_2. We maintain the convention that the value is 0 if the denominator is 0. Compared with those in the literature, the newly proposed measures are globally comparable, especially that node pairs with no common node can also be compared.

Remark. *The above enhanced version of feature* ***(6)*** *is reasonable for the current choice of fresh food in the data, since the price has been changed frequently and erratically, as required, during the chosen time period.*

The original version of feature **(6)** provides a relatively more general approach to remove the noise effect from the proposed measures. However, most literature followed the direction of filtering out insignificant edges [65, 136], rather than removing noise from network measures. Hence in this section, we take an initial step in the latter direction by deducting the mean value from some noise models.

First, we should determine the quantity of interest whose mean value will be subtracted. If we interpret the original measure as the geometric mean,

$$sim_o(i,j) = \sqrt{\frac{\sum_{l=1}^{n_t} \frac{A_{li}^{(b)} A_{lj}^{(b)}}{d_l^{(1)}}}{\sum_{h=1}^{n_t} \frac{A_{hi}^{(b)}}{d_h^{(1)}}} \frac{\sum_{l=1}^{n_t} \frac{A_{li}^{(b)} A_{lj}^{(b)}}{d_l^{(1)}}}{\sum_{h=1}^{n_t} \frac{A_{hj}^{(b)}}{d_h^{(1)}}}} = \sqrt{\frac{\sum_{l\in\Gamma(i)\cap\Gamma(j)} \frac{A_{li}^{(b)}}{d_l^{(1)}}}{\sum_{h\in\Gamma(i)} \frac{A_{hi}^{(b)}}{d_h^{(1)}}} \frac{\sum_{l\in\Gamma(j)\cap\Gamma(i)} \frac{A_{lj}^{(b)}}{d_l^{(1)}}}{\sum_{h\in\Gamma(j)} \frac{A_{hj}^{(b)}}{d_h^{(1)}}}},$$

where $\Gamma(i) = \{l : A_{li}^{(b)} = 1\}$ is the index set of node v_i's neighbours in the product-purchase network, then we find a common quantity $A_{li}^{(b)}/d_l^{(1)}$. Hence, we can propose the corresponding *randomised measure*,

$$sim_r(i,j) = \sqrt{\frac{\sum_{l\in\Gamma(i)\cap\Gamma(j)} \left(\frac{A_{li}^{(b)}}{d_l^{(1)}} - \mathbb{E}\left[\frac{A_{li}^{(r)}}{d_l^{(r)}}\right] \right)}{\sum_{h\in\Gamma(i)} \left(\frac{A_{hi}^{(b)}}{d_h^{(1)}} - \mathbb{E}\left[\frac{A_{hi}^{(r)}}{d_h^{(r)}}\right] \right)} \frac{\sum_{l\in\Gamma(j)\cap\Gamma(i)} \left(\frac{A_{lj}^{(b)}}{d_l^{(1)}} - \mathbb{E}\left[\frac{A_{lj}^{(r)}}{d_l^{(r)}}\right] \right)}{\sum_{h\in\Gamma(j)} \left(\frac{A_{hj}^{(b)}}{d_h^{(1)}} - \mathbb{E}\left[\frac{A_{hj}^{(r)}}{d_h^{(r)}}\right] \right)}}, \tag{2.7}$$

where $\mathbf{A}^{(b)} = (A_{li}^{(b)})$ is still the biadjacency matrix, $\mathbf{A}^{(r)} = (A_{li}^{(r)})$ is the random biadjacency matrix with each $A_{li}^{(r)}$ being a random variable, obtained from some

random bipartite network, and $d_l^{(r)} = \sum_{i=1}^{n_p} A_{li}^{(r)}$, with the convention that the value is 0 if the denominator is 0. We also change the normalisation to be only w.r.t. node v_j and propose the *randomised directed measure*

$$sim_{rd}(i,j) = \frac{\sum_{l \in \Gamma(j) \cap \Gamma(i)} \left(\frac{A_{lj}^{(b)}}{d_l^{(1)}} - \mathbb{E}\left[\frac{A_{lj}^{(r)}}{d_l^{(r)}} \right] \right)}{\sum_{h \in \Gamma(j)} \left(\frac{A_{hj}^{(b)}}{d_h^{(1)}} - \mathbb{E}\left[\frac{A_{hj}^{(r)}}{d_h^{(r)}} \right] \right)}, \tag{2.8}$$

with the same convention when the denominator is 0.

Next, we should determine the noise model. For example, if all nodes are expected to have their actual degrees, then BiCM (cf. Sec. 2.4.1) could be a natural choice for the noise model. In this particular case,

$$\mathbb{E}\left[\frac{A_{li}^{(r)}}{d_l^{(r)}} \right] = \frac{d_i^{(2)}}{m}, \tag{2.9}$$

where $d_i^{(2)}$ is the degree of node $v_i \in V_2$, and m is the number of edges in the product-purchase network. With Eqs. (2.7), (2.8) and (2.9), we accordingly introduce the *randomised configuration measure* and the *randomised configuration directed measure*.

We note that the measures can be computed for any pairs of nodes in V_2, but by characteristic **(2)**, only the pairs that satisfy characteristic **(1)** can be assigned positive degrees. Hence, the degree of complementarity between each pair of nodes in V_2, or the weighted adjacency matrix of the unipartite network for complementary relationship, can be estimated by

$$\mathbf{W}^{(c)} = \mathbf{A}^{(c)} \odot \mathbf{sim}_{\dagger}, \tag{2.10}$$

where $\mathbf{A}^{(c)}$ is the unweighted adjacency matrix in Eq. (2.3), the subscript $\dagger$ can be o, r, od or rd, and $\mathbf{sim}_{\dagger} = (sim_{\dagger}(i,j)) \in [0,1]^{n_p \times n_p}$. We refer to the values in $\mathbf{W}^{(c)}$ as the *complementarity scores*, and determine that a node v_i is complementary to another node v_j if $W_{ij}^{(c)} > 0$.

Measures for substitutability. We then propose measures for the degree of substitutability by feature **(4)**, where the more similar their complementary patterns with other nodes in V_2 are, the more substitutable they are. Here, we characterise each node in V_2 by a vector of its complementarity scores with the other nodes in V_2, and use the (unweighted) cosine similarity between these feature vectors to indicate

the degree of substitutability between pairs of nodes. Specifically, for a pair of nodes $v_i, v_j \in V_2$,

$$sim_s(i,j) = \sum_{k=1}^{n_p} \frac{W_{ik}^{(c)} W_{jk}^{(c)}}{\sqrt{(\sum_{p=1}^{n_p} W_{ip}^{(c)2})(\sum_{p=1}^{n_p} W_{jp}^{(c)2})}}, \tag{2.11}$$

where $\mathbf{W}^{(c)} = (W_{ij}^{(c)})$ is the weight matrix of the unipartite network for complementary relationship in Eq. (2.10), and n_p is the number of (product) nodes in V_2. The substitutability measures are named after the complementarity measure applied in $\mathbf{W}^{(c)}$. For example, with the original measure, we have the *original substitutability measure*; with the randomised configuration measure, we have the *randomised configuration substitutability measure*. Naturally, we also propose the directed version,

$$sim_{sd}(i,j) = \sum_{k=1}^{n_p} \frac{\min(W_{ik}^{(c)}, W_{jk}^{(c)}) W_{jk}^{(c)}}{\sum_{p=1}^{n_p} W_{jp}^{(c)2}}, \tag{2.12}$$

where each (i,j) entry measures the degree of substitutability of node v_i to node v_j in V_2, and the minimum function is applied to guarantee that the measure reaches its maximum value when the complementarity degrees of a node v_i to others in V_2 are at least the reference degrees of node v_j.

We again note that these measures can be computed for any pairs of nodes in V_2, but by characteristic **(4)**, only the pairs that satisfy characteristic **(3)** can be assigned positive degrees. Hence, the degree of substitutability between each pair of nodes in V_2, or the weighted adjacency matrix of the unipartite network for the substitute relationship, can be estimated by

$$\mathbf{W}^{(s)} = \mathbf{A}^{(s)} \odot \mathbf{sim}_{\dagger},$$

where $\mathbf{A}^{(s)}$ is the unweighted adjacency matrix in Eq. (2.4), the subscript † stands for s or sd, and $\mathbf{sim}_{\dagger} = (sim_{\dagger}(i,j)) \in [0,1]^{n_p \times n_p}$. We refer to the values in $\mathbf{W}^{(s)}$ as the *substitutability scores*, and define that a node v_i is substitute to another node v_j if $W_{ij}^{(s)} > 0$.

Remark. *Note the measures of substitutability are based on those of complementarity and we do not apply extra noise removing strategies here, thus it is critical that the complementarity measures are thresholded appropriately so that the substitutability degree is not biased by the nodes of low-complementarity-degree. Hence, by feature*

(5) *in Sec. 2.3.2, we accompany these measures with the following rules of threshold selection in analysing real data: (i) the threshold of the complementarity measures, θ_c, is chosen to be the largest value where the weighted unipartite network $\mathbf{W}^{(c)}$ after thresholding can maintain the same community structure as that obtained from a baseline threshold value; (ii) the threshold of the substitutability measures, θ_s, is chosen to be the smallest such value, for general noise removing purposes.*

2.4.3 Community detection

With the help of the two unipartite networks only of nodes in V_2, one for the complementary relationship with weight matrix $\mathbf{W}^{(c)}$ and other other for the substitute relationship with weight matrix $\mathbf{W}^{(s)}$, the problem to extract the two competing roles in the bipartite product-purchase network, with the particular connectivity patterns described in Sec. 2.3.2, can be effectively transformed to the classic problem of community detection in the unipartite networks. Specifically, the complement roles, i.e. groups of nodes (in V_2) that are mostly complementary to each other in the product-purchase network, are now groups of nodes that are densely connected inside while loosely connected outside in the unipartite network $\mathbf{W}^{(c)}$, i.e. communities. Similarly, the substitute roles are now communities in the unipartite network $\mathbf{W}^{(s)}$. Therefore, various efficient community detection algorithms can be considered [91, 119, 127, 132, 149].

Here, we apply the information-theoretic (hierarchical) map equation [132], which aims to describe the trajectory of random walkers on the network most efficiently, and utilises the fact that random walkers will stay in some communities for a long time if there is certain community structure, thus encoding the community structure appropriately can lead to the optimal description length of the trajectory. This method is known for being not affected by a common problem of community detection algorithms, the resolution limit [84]. Note that we have also applied other methods, including the Leiden algorithm [149], and the results are similar. Furthermore, from the detected structure, we will also examine the underlying assumption that the node groups are clique-like.

Overall, the set of methods we propose establishes an indirect solution to extracting the two competing roles, corresponding to the complementary and substitute relationships between nodes, in bipartite networks. The whole framework starts from various effort to transform the specific connectivity patterns characterising the target roles (in bipartite networks) into direct edges between the nodes (in the corresponding

unipartite networks), e.g. here the proposed null models to determine the significant relations and the proposed measures to quantify their degrees. Then the role extraction problem can be solved with standard community detection tools.

2.4.4 Validation

Since role extraction is part of the broad family of unsupervised learning problems where the role membership of each node is unknown (see Appendix A.6), it is important to validate the results with external datasets, in order to see whether the extracted roles capture some common features of the nodes involved. In this section, we consider specifically the complementary and substitute relationships between products, and then the product hierarchy, flavour compound and recipe datasets for validation as in Sec. 2.2. For the competing relationships in other contexts, one can seek the corresponding external datasets capturing similar information to what we have here.

We start from the product hierarchy information to characterise both complement roles and substitute roles, and then check if the characteristics are consistent with the common understanding of complements and substitutes. Specifically, we exclusively use the `L3` product hierarchy, consisting of fruit (F), organic produce (OP), prepared produce (PP), salad (S), and vegetable (V).

Next, we use the correspondence between flavour compounds and products to compute the Jaccard index, i.e. the relative number of shared flavour compounds, rf_{ij}, between each pair of product nodes v_i and v_j,

$$rf_{ij} = \frac{|C(i) \cap C(j)|}{|C(i) \cup C(j)|},$$

where $C(i)$ is the set of all flavour compounds in the i-th product. We then consider the cases in which $rf_{ij} = 0$ and $rf_{ij} = 1$, and check if the complementary pairs have a higher probability to share no flavour compounds and if the substitute pairs have a higher probability to share all their flavour compounds with each other. Furthermore, we examine the relationship between rf_{ij} and $W_{ij}^{(c)}, W_{ij}^{(s)}$, in terms of the *Pearson* correlation, as well as the *Spearman* correlation. We also compute the p-value for the correlations by assuming that the two list of values are drawn from independent normal distributions.

Subsequently, we use the recipe data to evaluate the relative number of shared

recipes, rr_{ij}, between each pair of product nodes v_i and v_j,

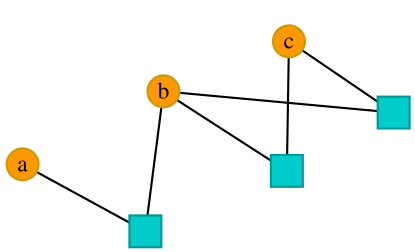
$$rr_{ij} = \frac{|R(i) \cap R(j)|}{|R(i) \cup R(j)|},$$

where $R(i)$ is the set of all recipes including the i-th product, and we set $rr_{ij} = 0$ if product nodes v_i and v_j are matched to the same ingredient. We then assess if the complementary pairs and substitute pairs have significantly higher and lower probabilities to co-appear in relatively more recipes, respectively. This is achieved by the *Mann-Whitney-Wilcoxon* (MWW) tests, where the null hypothesis H_0 is $P(X > Y) = P(Y > X)$, and different versions of the alternative hypothesis H_1 exist: (i) $P(X > Y) \neq P(Y > X)$, (ii) $P(X > Y) > P(Y > X)$, or (iii) $P(X > Y) < P(Y > X)$, with X, Y being two independent random variables [61,107]. Specifically, if X is the relative number of shared recipes from all product pairs $\{rr_{ij}\}$, and Y is that from only complementary pairs $\{rr_{ij} : W_{ij}^{(c)} > 0\}$, we will apply the alternative hypothesis (iii) $H_1 : \ P(X > Y) < P(Y > X)$. While for Y corresponding to the substitute pairs, we will use (ii) $H_1 : \ P(X > Y) > P(Y > X)$. Similarly, we also explore the relationship between rr_{ij} and $W_{ij}^{(c)}, W_{ij}^{(s)}$, in terms of the two correlations.

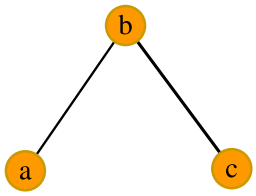
Finally, we apply the whole customised role extraction method to the recipe data, where we treat recipes as transactions and ingredients as products. This stems from the hypothesis that customers purchase products to cook dishes following recipes, thus the recipe data should be a restriction of the sales data. We compare the values of complementarity scores from the recipe data, the *recipe complementarity scores* $\mathbf{W}^{(cr)} = (W_{ij}^{(cr)})$, with those from the sales data, $\mathbf{W}^{(c)}$, and similarly, the *recipe substitutability scores* $\mathbf{W}^{(sr)} = (W_{ij}^{(sr)})$, with $\mathbf{W}^{(s)}$. Note that we set $W_{ij}^{(cr)} = 0$ and $W_{ij}^{(sr)} = 1$ if product nodes v_i and v_j are matched to the same ingredient. We finish the validation stage by comparing the role assignments (of product nodes) from both datasets, where l complement roles and l_1 substitute roles (from the recipe data) are obtained from applying the community detection algorithm on $\mathbf{W}^{(cr)}$ and $\mathbf{W}^{(sr)}$, respectively. We construct extra l_0 substitute roles by grouping together products that are matched to the same ingredients, for reference; see Fig. 2.3.

2.5 Results

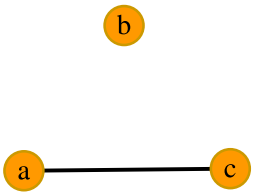
In this section, we consider the problem of extracting two competing roles in the context of product relationships. We first demonstrate the desirable performance of the customised role extraction method we propose through an illustrative example



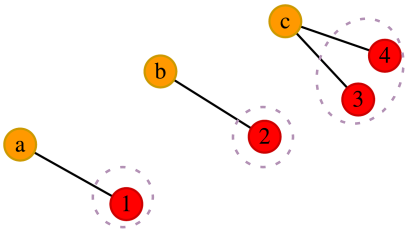
(a) An ingredient-recipe network.



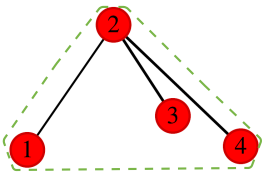
(b) The weighted complement network of ingredients.



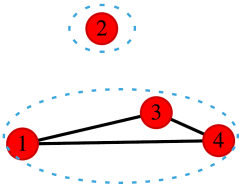
(c) The weighted substitute network of ingredients.



(d) The matching between ingredients and products.



(e) $\mathbf{W}^{(cr)}$.



(f) $\mathbf{W}^{(sr)}$.

Figure 2.3: Illustration of the process to compute the roles from the recipe data, where cyan squares are recipe nodes, orange circles are ingredient nodes, red circles are product nodes, the line thickness corresponds to how high the corresponding scores are, and l_0 substitute roles, l complement role(s) and l_1 substitute roles are shown as groups of product nodes in the purple dashed circles, green dashed polygon(s) and blue dashed circles, respectively.

with simulated sales data. We then show the results from the real data as described in Sec. 2.2.1, in terms of both the pairwise relationships and the roles at the mesoscale. Finally, the extracted relationships and roles are validated via external datasets, where the observed features are consistent with products being complements or substitutes.

2.5.1 Illustrative example

Before investigating noisy real data, we first validate our overall framework in a controlled "ideal world" where the relationship between products is known. Specifically, we simulate a consumer population characterised by a set of rules in this world, and ask whether our null models can capture the right relationship between each pair of products, whether our measures can give the right degree between them, and finally, whether our complement and substitute roles can provide insights into the groups of complements, and the groups of substitutes, respectively.

The simulated world is summarised as follows, similar to the one in [133].

- There are 13 different products: *coffee, wipes, ramen, candy, hot dog1, hot dog2,*

hot dog3, hot dog bun1, hot dog bun2, taco shell1, taco shell2, taco seasoning1, taco seasoning2.

- *coffee, wipes, ramen, candy* are independent products, but are popular with the customers, so are bought frequently. This corresponds to one possible source of noise, correlated preference, where the items are preferred by some customers but purchase decisions are made independently from one another, based on their features, e.g. price.

- The other products form substitute groups and complementary pairs. Products of the same names ignoring the number at the end are groups of substitutes; pairs in $\{$*hot dog1, hot dog2, hot dog3*$\}\times\{$*hot dog bun1, hot dog bun2*$\}$ and $\{$*taco shell1, taco shell2*$\}\times\{$*taco seasoning1, taco seasoning2*$\}$ are complementary pairs. In this world, customers never buy just one item in a complementary pair, and they always buy at most one of all such pairs.

- Customers are sensitive to price. When the price of a popular product is low, they buy it with probability 0.8; otherwise, they buy it with probability 0.2. Each customer purchases each preferred product independently.

  Sensitivity to the price of complementary pairs is different, since the probability to purchase a pair will decrease even if only one item in the pair has a high price. Hence, each pair is treated as a whole here. When all complementary pairs are of low price, customers buy one of them evenly; the case when all pairs are of high price is similar, except that customers have a 0.5 chance not to buy any of them; when one of the pairs has a lower price than the others, they buy this one with probability 0.85, and have 0.15 probability to buy others evenly; see Figure 2.4 for details.

With these specifications, we simulate $1,000$ transactions from this customer population. For a single transaction, each independent product has an 80% chance of being marked up to a high price; there is a 50% chance that all complementary pairs are of low price, a 10% chance that all are of high price, and accordingly a 40% chance that some are marked up, where the one of a lower price is chosen uniformly at random.

We provide the complementarity scores, $\mathbf{W}^{(c)}$, induced by the original measure, $\mathbf{sim}_o$, and by the randomised configuration measure, $\mathbf{sim}_r$, together with the number

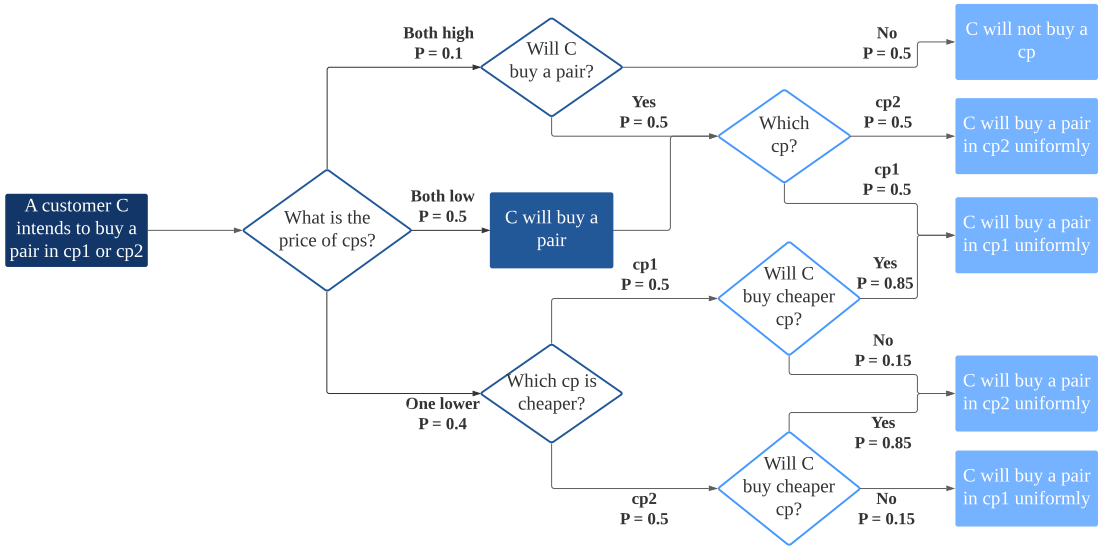


Figure 2.4: Illustration of how each customer chooses a complementary pair (cp), where cp1 and cp2 correspond to the *hot-dog*-and-*hot-dog-bun* and *taco-shell*-and-*taco-seasoning* complementary pairs, respectively.

of co-purchases, (cn_{ij}), in Fig. 2.5. We choose the variant of ER model as the underlying null model, since it better explains the noise here[3]. Note that independent products are bought more frequently, and their numbers of co-purchases with other products are fairly similar to those within complementary pairs. However, the extracted complementary pairs $\{(v_i, v_j) : W_{ij}^{(c)} > 0\}$ successfully retrieve the ground-truth complementary pairs. Accordingly, the extracted substitute pairs $\{(v_i, v_j) : W_{ij}^{(s)} > 0\}$ successfully retrieve the ground-truth substitute pairs. Furthermore, the complementarity scores of the *hot-dog*-and-*hot-dog-bun* complementary pairs is between 0.3 and 0.5, and those of the *taco-shell*-and-*taco-seasoning* complementary pairs is around 0.5. These values are approaching the inverse of the number of products in the corresponding substitute groups, which is consistent with the assuming complete substitution.

Finally, the extracted substitute roles exactly agree with the ground-truth substitute groups; see Fig. 2.6. The extracted complement roles reproduce the ground-truth complementary pairs including their corresponding groups of substitutes. Note that there are no groups of complements beyond the pairwise relationship.

This example demonstrates the ability of our overall framework to determine both product relationships and their corresponding degrees, which paves the way for

[3]Hence, the significance level for the significantly more co-purchases, α_c, for the variant of ER model can be chosen as high as 0.9 with the same results. While for the BiCM, a much lower significance level (e.g. 10^{-4} for α_c) is needed in order to extract the true product relationships.

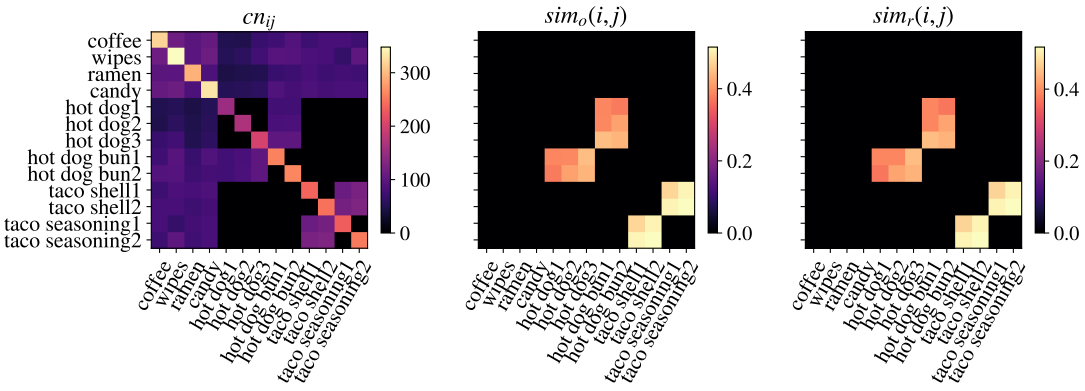


Figure 2.5: Measures on the products, from co-purchases (cn_{ij}) whose diagonal shows the purchase frequency (**left**), the complementarity scores $\mathbf{W}^{(c)}$ induced by the original measure ($sim_o(i,j)$) (**middle**) and by the randomised configuration measure ($sim_r(i,j)$) (**right**), where x-axis, y-axis are products in the same order as being listed in the simulated world assumptions.

us to continue the analysis on real-world data. From a mesoscale perspective, the extracted complement roles and substitute roles have much overlap with the groups of complements and those of substitutes, respectively. Furthermore, the fact that we already have complement roles involving substitutes indicates that the interaction between the two relationships is not negligible. For instance, it is entirely possible that we may find substitute roles including complements in real data.

2.5.2 Sales data

Hereafter, we use the variant of ER model as the underlying null model, since its assumptions are generally applicable in real-world purchases, and we only show the results from the original measure, because both have very similar behaviour; see Appendix A.4 for the parameter calibration and the results from the randomised measure. We first examine the ranking power of the scores we propose, $\mathbf{W}^{(c)}$ and $\mathbf{W}^{(s)}$, by checking the top complementary pairs and substitute pairs for each product. This is done by choosing several query products v_j at random, and output the products of the three highest complementarity scores $W^{(c)}_{ij}$ and the ones of the three highest substitutability scores $W^{(s)}_{ij}$; see Table 2.1 for one run. The substitute pairs of scores greater than 0.1 largely agree with common sense[4]. For example in Table 2.1, the top substitute of organic blueberries is blueberries, and the top substitutes of

[4]Here we refer to the notion that products that are essentially the same are substitutes, for example, Brand A apples and Brand B apples.

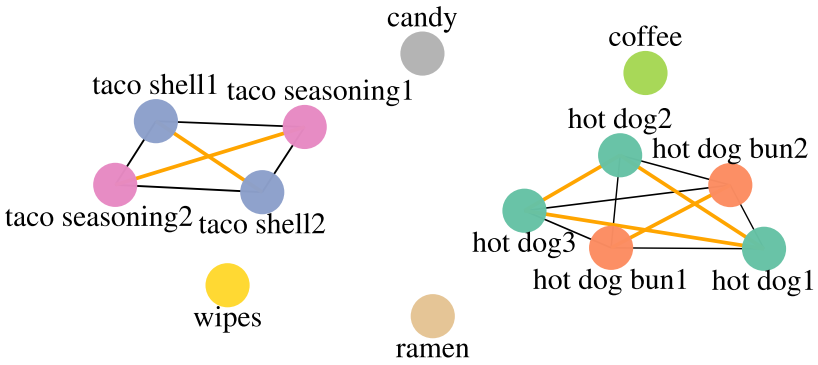


Figure 2.6: The unipartite network in the illustrative example, with product nodes connected by both the complementarity scores ($W_{ij}^{(c)}$) (in black) and the substitutability scores ($W_{ij}^{(s)}$) (in orange) induced by the original measure, where the line thickness corresponds to how high the scores are, and products in the same substitute role are shown in the same colour.

salad tomatoes are other types of tomatoes. Additionally, the ranking indicates that common-sense substitutes have high complementarity scores with the same products. For example salad tomatoes, baby plum tomatoes and tomatoes on the vine are the top three complements of loose cucumbers. These findings justify the feature **(3)** in Sec. 2.3.2. There are also some nontrivial substitutes, of lower score values, from general understanding, which we will discuss in Sec. 2.6.

Table 2.1: Products of the three highest complementarity scores and substitutability scores with the query products.

Query product	Complement		Substitute	
Organic Blueberries	0.14	Organic Raspberries	0.50	Blueberries
	0.059	Organic Strawberries	0.13	Green Seedless Grapes
	0.048	Organic Cherry Tomatoes	0.070	Tomatoes on the Vine
Loose Cucumbers	0.098	Salad Tomatoes	0.54	Organic Loose Cucumbers
	0.089	Baby Plum Tomatoes	0.21	Courgette Spaghetti
	0.079	Tomatoes on the Vine	0.18	Sliced Runner Beans
Salad Tomatoes	0.098	Loose Cucumbers	0.83	Tomatoes on the Vine
	0.063	Iceberg Lettuce	0.79	Baby Plum Tomatoes
	0.046	Mixed Peppers	0.74	Cherry Tomatoes

We proceed for the mesoscale structure, i.e. the complement roles and substitute roles. From an averaged perspective, the complement roles and the substitute roles constitute classic (assortative) communities in the unipartite networks $\mathbf{W}^{(c)}$ and $\mathbf{W}^{(s)}$,

respectively; substitute roles form disassortative communities, where nodes are loosely connected internally while densely connected externally, in $\mathbf{W}^{(c)}$; see Fig. 2.7. The latter observation also justifies the feature **(3)**. Furthermore, the overlap between the two roles is not negligible, with the normalised mutual information (NMI [153]) 0.49. Hence, as mentioned in Sec. 2.5.1, substitutes may appear in the same complement role by their strong complements, and complements may be assigned to the same substitute role for their strong substitutes.

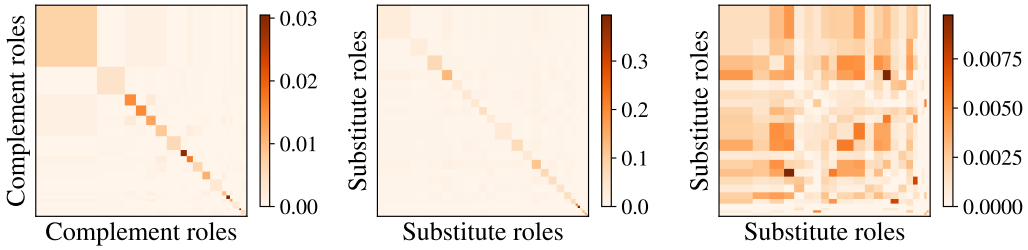


Figure 2.7: Scores averaged over the nodes in the same role, from the complement roles on the complement unipartite network $\mathbf{W}^{(c)}$ (**left**), and from the substitute roles on the substitute unipartite network $\mathbf{W}^{(s)}$ (**middle**) and on the complement unipartite network $\mathbf{W}^{(c)}$ (**right**), where isolated product nodes have been removed.

Finally, we explore the internal structure of complement roles and substitute roles. Generally, strong complements[5] do not tend to form complete graphs in the complement unipartite network $\mathbf{W}^{(c)}$, where there are many products that are complements of the same products but are not complements of each other. For example, blueberries (Blueb) and organic blueberries (Or Blueb) in the complement role of berries (3) are substitutes, but both are complements of raspberries (Raspb), stawberries (Strawb), etc; see Fig. 2.8. There are also cases in which they constitute some complete graph, and further exploration indicates that these products are highly likely to be consumed together. For example, mushroom stir fry (Mushroom SF), vegetable and beansprout stir fry (V Beansprout SF), and egg noodles form a triangle in the complement role of stir-fry (9); see the blue polygon in Fig. 2.8.

Strong substitutes are expected to form complete graphs in the substitute unipartite network $\mathbf{W}^{(s)}$, and our results are largely consistent with the expectation. For example, loose Braeburn apples (LB Apples), loose Pink Lady apples (LPL Apples), and bagged organic Gala apples (BOrG Apples) constitute a triangle in the substitute

[5]Note we determine two product nodes v_i, v_j to be complements if $W^{(c)}_{ij} > 0$, and measure their degree of complementarity (from weak to strong) by the value of $W^{(c)}_{ij}$; substitutes are treated similarly.

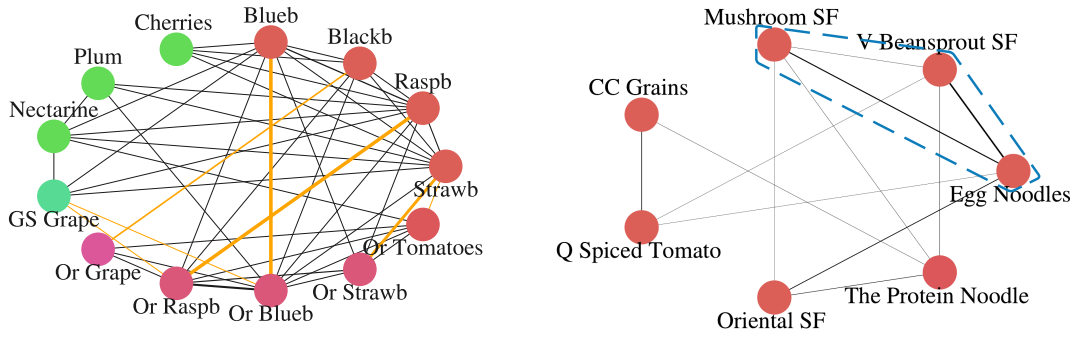


Figure 2.8: Internal structure of the complement roles of berries (3) (**left**) and of stir-fry (9) (**right**), with product nodes connected by both the complementarity scores ($W_{ij}^{(c)}$) (in black) and the substitutability scores ($W_{ij}^{(s)}$) (in orange), where the line thickness corresponds to how high the scores are, and products in the same `L1` category are shown in the same colour.

role of apples (23); see the blue polygon in Fig. 2.9. Note this expectation is only valid if the substitutes are consumed for the same purpose; if this assumption is violated, seemingly substitute products may end up being complements. For example, loose brown onions (LBr Onions) and loose red onions (LR Onions) in the substitute role of onions (4) are both substitutes of products such as bagged red onions (BR Onions) and bagged organic brown onions (BOrBr Onions), but are complements of each other; see Fig. 2.9. The difference between their quantities and with their common substitutes may be the key factor here. Likewise, even with the common substitute bagged organic Gala apples (BOrG Apples), loose ripe pears (LR Pears) is a complement of loose Pink Lady apples (LPL Apples), loose Braeburn apples (LB Apples) and loose Gala apples (LG Apples). The above observations confirm the complexity of the interaction between complements and substitutes.

2.5.3 Validation

Product hierarchy. The distribution of `L3` categories in each complement role is consistent with products being complements; see Fig. 2.10. Most complement roles involve more than one category, which could be explained by the complementarity across categories. For example, the complement role of nuts & fruits (2) contains both fruit and prepared produce, the complement roles of berries & grapes (3) and of grapes & oranges (5) consist of both fruit and organic produce, and the complement role of potatoes, beans & carrots (6) includes both prepared produce and vegetables. There are also complement roles only involving one category, and the related products

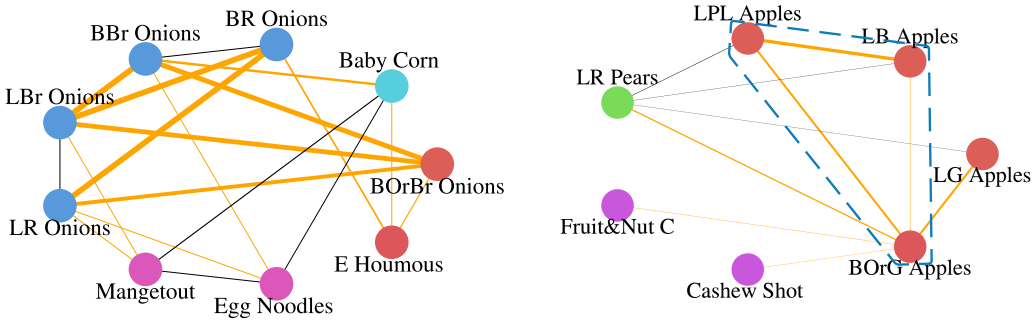


Figure 2.9: Internal structure of the substitute roles of onions (4) (**left**) and of apples (23) (**right**), with product nodes connected by both the complementarity scores ($W_{ij}^{(c)}$) (in black) and the substitutability scores ($W_{ij}^{(s)}$) (in orange), where the line thickness corresponds to how high the scores are, and products in the same `L1` category are shown in the same colour.

are either in fruit or in the prepared produce category. Further, this is in agreement with the notion that products in prepared produce, for instance prepared vegetables and vegetable dips, go well together; similar for products in fruit.

The proportion of `L3` categories in each substitute role also accords with products being substitutes; see Fig. 2.10. Some of them only or mostly involve prepared produce, and some others largely consist of fruit, such as the substitute role of apples (23). This agrees with the tendency of grouping products into categories based on shared characteristics. Other substitute roles contain more than one category, with one of them being prepared produce. For example, the substitute role of grapes (5) includes both fruit and prepared produce, the substitute role of carrots (13) comprises both prepared produce and vegetables, the substitute role of peppers (19) involves prepared produce, salad[6] and vegetables, and the substitute role of avocado salad (25) is composed of fruit, prepared produce and salad. Further investigation shows that products in prepared produce include fresh-cut fruits, prepared salads and prepared vegetables, i.e. prepared versions of products in fruit, salad and vegetable categories.

Flavour compounds and recipes. We observe that the substitute pairs have a significantly higher probability to share all their flavour compounds with each other, i.e. $rf_{ij} = 1$, than all product pairs, while complementary pairs have a significantly higher probability to share no flavour compounds with each other, i.e. $rf_{ij} = 0$;

[6]Here the salad category contains products like cucumber, pepper, lettuce and tomatoes.

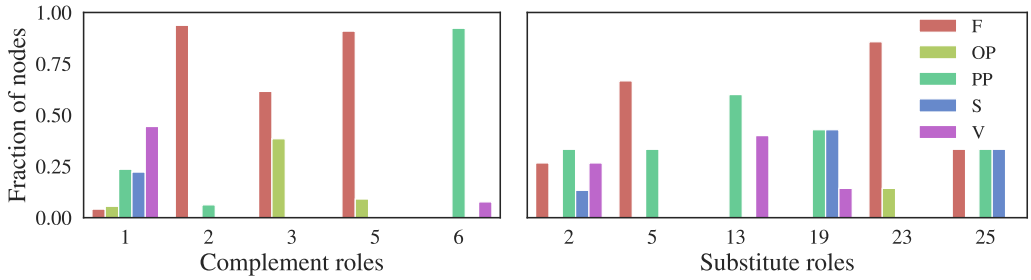


Figure 2.10: Proportion of the products in typical complement roles (**left**) and typical substitute roles (**right**) that fall in each L3 category, fruit (F), organic produce (OP), prepared produce (PP), salad (S) and vegetable (V).

see Fig. 2.11. These characteristics are consistent with the functional definition of complements and substitutes: complements are consumed together, thus tend to have different flavours in order to accompany each other; while substitutes can replace each other, thus tend to have the same flavours. The distributions of rf_{ij} values between 0 and 1 have roughly the same shapes for the three types of pairs.

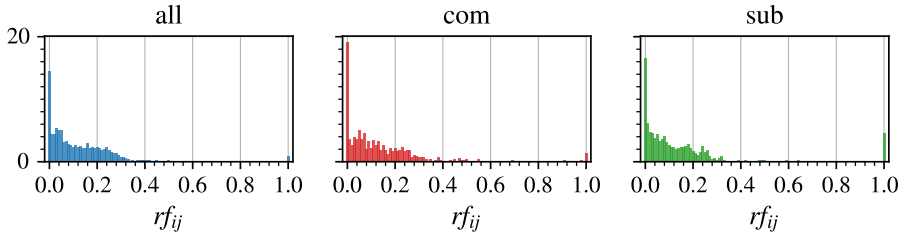


Figure 2.11: Distributions of the relative number of shared flavour compounds of all product pairs $\{rf_{ij}\}$ ("all"), of complementary pairs $\{rf_{ij} : W_{ij}^{(c)} > 0\}$ ("com") and of substitute pairs $\{rf_{ij} : W_{ij}^{(s)} > 0\}$ ("sub"), where probabilities of $rf_{ij} = 0$ and $rf_{ij} = 1$ are of interest.

Further, we investigate the correlations between the relative number of shared flavour compounds (rf_{ij}) and the score values ($W_{ij}^{(c)}$), ($W_{ij}^{(s)}$); see Table 2.2. The Pearson correlation indicates that the product pairs of higher substitutability scores have a significant tendency to share larger portions of their flavour compounds, while the patterns when changing the complementarity scores is more heterogeneous, with a mild negative correlation between the ranking of the complementarity scores and that of the relative number of shared flavour compounds.

We then discern that the complementary pairs have higher probability to co-appear in relatively more recipes, $\{rr_{ij} : W_{ij}^{(c)} > 0\}$, than all product pairs, $\{rr_{ij}\}$, while the substitute pairs have lower probability to co-appear in relatively more

recipes, $\{rr_{ij} : W_{ij}^{(s)} > 0\}$; see Fig. 2.12. Both trends are significant by the MWW test (p-values: 2.8×10^{-123} and 3.1×10^{-4} for the complementary pairs and the substitute pairs, respectively). These features also accord with the interpretation of complements or substitutes from the cooking perspective: complements go well with one another, thus are more likely to appear in the same recipe; while substitutes can be used in place of each other, thus tend to be cooked together with some others but not each other.

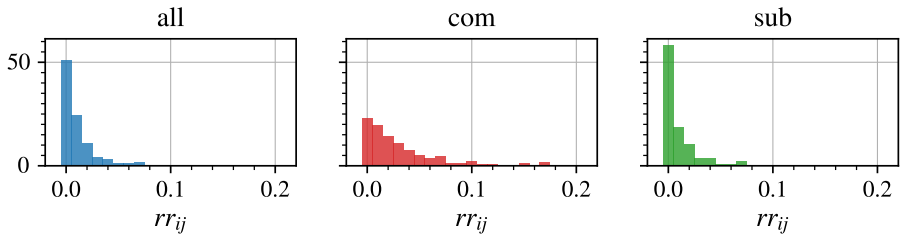


Figure 2.12: Distributions of the relative number of shared recipes of all product pairs $\{rr_{ij}\}$ ("all"), of complementary pairs $\{rr_{ij} : W_{ij}^{(c)} > 0\}$ ("com") and of substitute pairs $\{rr_{ij} : W_{ij}^{(s)} > 0\}$ ("sub"), where the range is chosen for visualisation purpose, while complementary pairs also have positive probabilities at values greater than 0.2.

Moreover, we examine the correlations between the relative number of shared recipes (rr_{ij}) and the score values; see Table 2.2. The Spearman correlation suggests that product pairs of higher rankings in the complementarity scores tend to co-appear in relatively more recipes, which agrees with the Pearson correlation. The trend when increasing the substitutability ranking of product pairs is a mild propensity towards co-appearing in relatively less recipes.

Additionally, we also explore the correlations between the complementarity scores (from the sales data) ($W_{ij}^{(c)}$) and the recipe complementarity scores (from the recipe data) ($W_{ij}^{(cr)}$), and between the substitutability scores ($W_{ij}^{(s)}$) and the recipe substitutability scores ($W_{ij}^{(sr)}$); see Table 2.2. The Spearman correlations of both score pairs indicate significant positive relationships between the scores within each pair, consistent with the information suggested by the Pearson correlations.

Finally, we compare the complement and substitute role assignments from different data sources, in particular the sales data versus the recipe data, where we use the NMI and the adjusted mutual information (AMI [153]) to measure the consistency between role assignments; see Table 2.3. The substitute roles from the sales data are very similar to the complete substitution, i.e. l_0 substitute roles obtained from the recipe data. Although the complement roles from the sales data are relatively more in

Table 2.2: The correlations between the scores from the sales data, $(W_{ij}^{(c)}), (W_{ij}^{(s)})$, and the measures from the flavour compound and recipe data, $(rf_{ij}), (rr_{ij})$ (where the superscripts (c) and (s) below denote the values restricted to $\{(i,j) : W_{ij}^{(c)} > 0\}$ and $\{(i,j) : W_{ij}^{(s)} > 0\}$, respectively), $(W_{ij}^{(cr)}), (W_{ij}^{(sr)})$. Note that the p-value is only for reference since the distributions are not very normal-like.

Score pair	Pearson	(p-value)	Spearman	(p-value)
$(W_{ij}^{(c)}) - (rf_{ij}^{(c)})$	0.085	(4.8×10^{-4})	-0.019	(4.3×10^{-1})
$(W_{ij}^{(s)}) - (rf_{ij}^{(s)})$	**0.50**	(2.4×10^{-79})	0.030	(3.1×10^{-1})
$(W_{ij}^{(c)}) - (rr_{ij}^{(c)})$	**0.16**	(7.2×10^{-8})	**0.24**	(1.1×10^{-16})
$(W_{ij}^{(s)}) - (rr_{ij}^{(s)})$	-0.040	(2.7×10^{-1})	-0.062	(8.8×10^{-2})
$(W_{ij}^{(c)}) - (W_{ij}^{(cr)})$	**0.16**	(3.8×10^{-8})	**0.23**	(2.5×10^{-15})
$(W_{ij}^{(s)}) - (W_{ij}^{(sr)})$	**0.60**	(4.5×10^{-76})	**0.23**	(2.5×10^{-15})

agreement with the l_0 substitute roles than the l complement roles by NMI, this may be caused by the number of l_0 substitute roles being larger than that of l complement roles, since the AMI indicates the opposite. To conclude, the relatively large NMI and AMI values demonstrate the consistency between the extracted product relationships from these two different sources, and also provide evidence that customers could buy products corresponding to ingredients in some recipes.

Table 2.3: NMI and AMI between the partitions by the roles from the sales data (columns) and those from the recipe data (rows), where "com" and "sub" correspond to the complement roles and the substitute roles, respectively, where "/" corresponds to comparing complement roles from one data to the substitute roles from the other, which is not considered here.

NMI/AMI	l_0 sub	l_1 sub	l com
com	0.54/0.16	/	0.36/0.28
sub	0.71/0.21	0.54/0.18	/

2.6 Conclusions

Role structure is important in understanding networks, since different roles normally correspond to different structural and dynamical functions in the networks. In the literature, most role extraction methods aim to extract groups of nodes sharing generally similar connectivity patterns, directly or indirectly. However, there could be roles with specific characteristics that are particularly interesting in certain contexts,

such as the roles related to the complementary and substitute relationships between nodes in one part of bipartite networks, as we analysed here. Therefore, it is necessary to extend the current role extraction methods for the roles with specific connectivity patterns.

In this chapter, targeting two competing roles in bipartite networks, we propose a customised role extraction method accordingly. Starting from the characteristics of the specific roles corresponding to the complementary and substitute relationships between the nodes in one part of bipartite networks, we first propose several null models to determine the significant relations between the nodes. We then propose measures to quantify the degrees of the two competing types of relationships, motivated by random walks on networks. Finally, we apply community detection algorithms to obtain groups of nodes that are mostly complementary or substitute to each other, i.e. the two competing roles at the mesoscale. The set of methods provides an indirect solution to both the specific problem of extracting two competing roles in bipartite networks and also the general problem with respect to roles with predetermined characteristics. We believe that the general framework we propose, consisting of significance tests with null models, measures based on random walks and community detection algorithms, could also provide insights into the role extraction problem in general.

Specifically, the proposed method, together with the characteristic patterns, establishes a network-based solution to the problem of understanding the hidden relations existing between products from sales data (see Appendix A.6 for the formulation as an unsupervised learning problem). It is our belief that a network approach opens up a promising new angle on this problem due to its flexibility, e.g. in determining significant relationships, and the vast network science toolbox. Finally, let us emphasise that we only use basket data to extract the product relationships, without additional information such as the customer profile and the price change, information that are typically required for existing methods and may cause privacy issues [47]. The insights derived from our methods have applications in assortment-related decision making, not only for retailers but also general firms with broad product categories.

In the future, it would be interesting to design a method that directly uncovers the degrees of complementarity and substitutability from bipartite networks, without any intermediary steps, such as extracting the significant relations in the methods we propose here, and also to explore more of the directed scores, since our focus is on the symmetric ones in this chapter. Another direction is to characterise the nodes by their centrality in the projected networks, for instance by their average complementarity and substitutability scores with others. For the practical problem of

understanding the product relationships, one important direction would be to consider the bipartite network from a temporal perspective, in order to explore further the connection between structure and cross-elasticity (see Appendix A.5). Moreover, our current analysis focused on fresh food where prices change frequently throughout the period. Yet, we did not explicitly include price as a factor, but either ignored its bias or removed it by some random models [9]. In order to analyse a more general range of products in the future, it would be necessary to incorporate price information in the current framework in a meaningful way.

Chapter 3

Unifying dynamics on networks and optimisation

This chapter is on the paper Tian & Lambiotte (2021) [147]

3.1 Introduction

3.1.1 Demand change of inter-related products

Understanding how the demand of products changes is an important topic in both economics and marketing research, as well as in most businesses, with direct consequences for the profits. This question is closely related to demand forecasting [71, 83, 97], and finds various applications in the further decision making including shelf stocking, dynamic pricing, and promotions [42, 52, 150]. For instance, firms need a reasonable prediction for the demand change of perishable goods, since the freshness of such products decreases rapidly and they can only be held for at most a few days [83]. Meanwhile, promotions have been widely accepted as one of the most effective marketing instruments in boosting sales, thus it is then interesting to understand the phenomena from a more quantitative perspective [19].

In demand estimation, there are generally two approaches, based on either the product space or the characteristic space. The former is relatively more natural, where consumers have preferences over products, and those preferences lead to demand at the product level. Classic models include the Rotterdam model [14, 146], the translog model [41] and the almost ideal demand system [48]. While the latter view each product as bundles of characteristics, and consumers have preferences over such characteristics rather than the bundles. The characteristic-space approach normally

follows the path from utility functions to obtain the probabilities of purchases [35,133]. Both types of methods generally lead to linear, or log-linear, models in explaining the demand via other factors including the product price. However, few of them consider the relationships between products, and among the work that does, they only consider the direct complements or substitutes but not any further [42].

Here, equipped with a data-driven approach to extract the product relationships developed in Chapter 2, we can consider the network structure underlying the products, where each product is a node and they are connected by both the complementary and substitute relationships. Then we can analyse the *cascading effect* of the demand changes from some products to the whole system through this network. This is the first work of this type, to the best of our knowledge. Specifically, we model the dynamic of demand changes as information propagating through the network, where linear models are widely adopted [85, 137], consistent with the development of demand estimation. Then the problem of further decision making, e.g. to choose a set of products to impose the demand change so as to maximise the overall revenue, can be considered as a variant of the associated influence maximisation problem. Hence, in the following, we contextualise the problem in social networks, and consider specifically the information propagation and influence maximisation, where we propose a general class of information propagation model in order to explain more propagation phenomena including the dynamic of demand changes, and then develop a general framework for the influence maximisation problem. We start from simple networks made only of positive edges in this chapter, and will consider the problem on signed networks with both positive and negative connections, corresponding to the two competing types of relationships between products, in the following chapter.

3.1.2 Information propagation and influence maximisation

The rapid growth of online social networks, such as Facebook and Twitter, allows hundreds of millions of people worldwide to interact with each other, providing access to a vast source of information on an unprecedented scale. The *propagation* of information, opinion, innovation, rumour, etc., is a critical component to explain, e.g., how a piece of information could quickly become pervasive through a network via "word-of-mouth" [11,33,114]. Accordingly, understanding how information spreads in social networks is a central theme in social, behavioural, and economic sciences, with theoretical and practical implications, such as the adoption of political viewpoints

in presidential elections and the *influence maximisation* problem for viral marketing [23,39,95], attracting expertise from various fields including mathematics, physics, and biology [113, 125, 126].

Specifically in the context of influence maximisation, one can distinguish two main classes of information propagation models: the independent cascade (IC) model, and the linear threshold (LT) model, where nodes adapt their behaviour from each neighbour independently, or from the collective influence of the whole neighbourhood, respectively [85, 137]. It is well known that both models, which we refer to as "classic models", suffer from limitations. First of all, the state space of both models is binary, as nodes can only have states either active or not, while various levels of influence and of confidence could co-exist among agents in real scenarios. Moreover, there is no feedback in both processes, as nodes can only stay active after being activated, thus may not influence back nodes that influenced them as would be the case in real life. These limits have called, and still call for more general models allowing us to consider dynamics with feedback between states and more heterogeneity in the agents' behaviour, which is important to explain more propagation phenomena, e.g. the dynamic of demand changes we consider here.

In parallel to this line of work, simple and complex contagions have attracted much research interests in mathematical sociology and physics [34, 69]. Essentially, complex contagion considers situations when the reinforcement of a signal favours its future adoption, which can be modelled via deterministic threshold models. However, their focus is usually on understanding the effects of specific structures, the presence of shortcuts or the density of triangles for example, and they also tend to consider binary state variables. For models with continuous variables, more has been done within the field of opinion dynamics, where linear models build on the heat equation [106], such as the DeGroot model [50], and non-linear models include the bounded confidence model for example [49]. A first contribution of this chapter is to introduce a non-linear, deterministic model for information propagation which relaxes the constraints of binary variables while allowing feedback between nodes, and also possesses the classic models after appropriate extensions as limiting cases, thus providing a unifying framework for information propagation.

As a second contribution, we consider the important problem of influence maximisation (IM), known to have potential applications in various domains [95, 96, 142]. Given a network and an associated information propagation process, the classic IM problem consists in selecting a small set of nodes to activate initially with the aim of maximising the overall *influence spread*, commonly defined as the number of activated

nodes at the end of the process. Kempe et al. [85] formulated it as a stochastic combinatorial optimisation under the classic models, and proposed a greedy algorithm with theoretical approximation guarantees. The near-optimal asymptotic bounds of this seminal work have triggered a vast amount of research in this direction, mostly to further reduce the running time [67, 96]. There are also heuristic solutions, such as centrality-based methods and genetic algorithms, but without theoretical performance guarantees [12].

The aforementioned theoretical guarantees are obtained under certain assumptions on the influence spread function, which could be strict in practice [101], and even for the LT model, occur only for specific choice of the thresholds' distribution. The IM problem when the thresholds are given by distributions other than the uniform distribution is relatively unknown. To consider a general information propagation model, we have thus developed a new framework, where we formulate the IM problem as a *mixed integer nonlinear programming* (MINLP). The exact methods for MINLP are mostly based on relaxing the problem to be linear in an appropriate manner, and require first-order information [16, 22, 28], which is not generally available in the IM problem. For this reason, we treat the objective function as a black box and adopt *derivative-free methods* [22], with a *mesh adaptive direct search* method as a general solution [1]. Furthermore, we propose a customised method with local convergence, specifically designed for the proposed information propagation model.

3.2 Literature review

We first describe the two classic information propagation models, the IC model and the LT model, in more detail [85, 137]. In both models, each individual node has two state values, either 0 for being inactive or 1 for being active, and the primary focus is on progressive processes, where nodes can switch from being inactive to being active but not vice versa. Importantly, there is no feedback between nodes, even on undirected networks, as a node cannot be influenced by another node that it influenced before. Specifically, in the LT model, a node v_j is influenced by each neighbour v_i according to a weight b_{ij} s.t. $\sum_i b_{ij} \leq 1$, where $b_{ij} = 0$ if node v_i is not v_j's neighbour; each node v_j chooses a threshold θ_j uniformly at random from the range $[0, 1]$, which represents the critical influence weight necessary for node v_j to be activated. Given a random choice of the thresholds, and an initial set of active nodes $\mathcal{A}_0$, the propagation process unfolds deterministically in discrete time steps, where at time step $t > 0$, all nodes that are active at $t - 1$ remain active, and an inactive node

v_j will be activated if the total weight of its active neighbours is at least its threshold θ_j, i.e.

$$\sum_{v_i \in \mathcal{A}_{t-1}} b_{ij} \geq \theta_j, \tag{3.1}$$

where set $\mathcal{A}_{t-1}$ contains the active nodes in step $t-1$. However, in the IC model, each activated node v_i has a single chance to activate each currently inactive neighbour v_j when it first becomes active, with success probability p_{ij} independently of the history thus far. If v_j is successfully activated, it will have value 1 in the next time step, but whether or not v_i succeeds, it cannot further attempt to activate its neighbours in the subsequent rounds.

However, nodes do not have to stay active in real systems, since people may become more convinced of a piece of information if more friends believe it, or less convinced, even change their minds and become inactive, if there are fewer friends believing it. More general cascade and (submodular) threshold models have also been considered [85,113]. In particular, Kempe at al. [85] considered variants of both models with feedback by constructing a multilayer network with each layer for each time step, but they required a predetermined depth of the propagation. We refer the reader to [37] for a comprehensive survey on information propagation.

In the two classic models, the influence spread is defined as the number of active nodes at the end of the process, and the IM problem is then to maximise it subject to limited number of nodes that one can activate at the beginning of the process, $|\mathcal{A}_0|$. The IM problem under the two classic models is NP hard, and the key algorithmic breakthrough lies in the approximation guarantees for the greedy hill-climbing algorithms [85]. Subsequently, several methods have been proposed to further improve the efficiency of the greedy algorithms, maintaining the same approximation guarantees [67,96], or not exactly [21,40,154]. One vital assumption is that the information spread is *submodular*. Specifically, a function $f : P(U) \to \mathbb{R}^+ \cup \{0\}$, where $P(U)$ is the power set of a finite set U, is submodular, if

$$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T),$$

for all element $v \in U$ and $S \subseteq T \subseteq U$, i.e., the marginal gain from activating one more node initially is larger if the original set is smaller. However, this is not necessarily true when there are certain threshold effects. For the LT model, the key correspondence lies in the uniform distribution of thresholds, and we can show that the influence spread under the LT model with deterministic thresholds is not

submodular. Furthermore, with deterministic thresholds, the IM problem has been shown to be NP-hard to approximate within certain factor [36,86]. We refer the reader to [12, 101] for more detailed description of the development of the IM problem.

Finally, we note that there are models with continuous variables and feedback between nodes within the field of opinion dynamics [49, 50], but the associated IM problem is relatively unknown there. Meanwhile, there are some continuous models analysed within the context of IM, such as the fully linear models [57], but they do not have well-established connections with the classic models and their intuitive mechanisms. We also note that there could be variants of the constraint in the IM problem when nodes take continuous values, e.g. on the sum of initial state values rather than the number of activated nodes [51]. However in this paper, we are interested in the case when, e.g., companies have limited resources to convince more people to buy the products, thus maintain the original constraint.

3.3 General class of information propagation model

In this section, we propose a novel class of information propagation models, with continuous state variables while allowing the feedback between nodes. The full description of the proposed model is in Sec. 3.3.1. We show that it can be reduced to the IC model and the LT model after the corresponding extensions in Secs. 3.3.2.1 and 3.3.3.1, respectively, as well as its general properties via the differences with the two classic models in Secs. 3.3.2.2 and 3.3.3.2, respectively. Furthermore, we show that with the proposed model, both the locally linear-dynamics-like and the locally linear-threshold-like propagation can co-exist in a single network in Sec. 3.3.4. We conclude with a discussion of the derivative information in Sec. 3.3.5, which plays an important role in the IM problem.

3.3.1 Model description

We consider a social network $G(V, E)$ that is connected, weighted, and undirected[1], where $V = \{v_1, v_2, \ldots, v_n\}$ is the node set, and $E = \{(v_i, v_j) : v_i \text{ is connected with } v_j\}$ is the edge set. Each edge (v_i, v_j) is considered as a channel connecting nodes v_i and v_j from which the information flows, and can be associated with a weight W_{ij}, for example, indicating the strength of the interaction or the level of trust between the

[1]We consider undirected networks in the following, but the results can also be extended to directed networks.

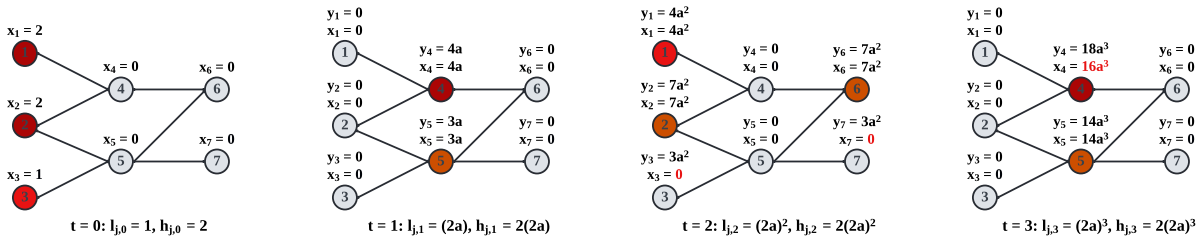


Figure 3.1: Illustration of the information propagation in the first few steps following the GIP model, where $y_j(t) = \sum_i W_{ij} x_i(t-1)$ is the linear product, $x_j(t) = f_{j,t}(y_j(t))$ is the state value, the network has uniform weight α, and the bounds are set to be $l_{j,t} = (2\alpha)^t, h_{j,t} = 2(2\alpha)^t$, $\forall v_j \in V$, with the node colour indicating the level of influence at each step.

agents. Here, we consider a continuous variable $x_j(t) \in \mathbb{R}$ to represent the state value of node v_j at each discrete time step $t \geq 0$, which can be interpreted as the influence on node v_j at t. By using continuous variables, we assume that the influence is *additive*, where, e.g., people could become more convinced of a piece of news if more friends believe it, or buy more products if more friends make a purchase, either at each time or over time. We denote the vector consisting of $x_i(t)$ by $\mathbf{x}(t) = (x_i(t))$.

We now propose a general class of information propagation (GIP) models unifying the mechanisms underlying the two classic models. (i) Each node v_i can independently attempt to influence its neighbours, proportional to the edge weight and its own state value $x_i(t)$, which is consistent with the IC model. (ii) The actual influence on each node v_j is based on the *collective* behaviour of the whole neighbourhood, by applying a non-linear transformation to $y_j(t) = \sum_i W_{ij} x_i(t-1)$, in order to capture how the accumulated attempts from all neighbours transforms into a change for the state of v_j, which is reminiscent of the mechanisms of the LT model, but also of non-linear models for opinion dynamics [143]. Specifically, we assume that at each $t > 0$, there is a lower bound $l_{j,t}$ in the model, corresponding to the critical mass to trigger the propagation s.t. $x_j(t) = 0$ if $y_j(t) < l_{j,t}$, and also an upper bound $h_{j,t}$ for the saturation effect s.t. $x_j(t) = h_{j,t}$ if $y_j(t) \geq h_{j,t}$ [7, 59] (see Fig. 3.1 for a step-by-step illustration of the underlying process which we will discuss later). Explicitly, the GIP model is a

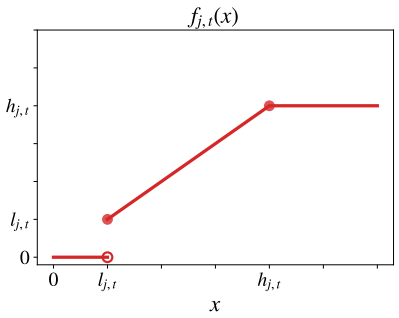


Figure 3.2: Example bound function $f_{j,t}$ of node v_j at time step t where $h_{j,t} = 4l_{j,t} > 0$.

bounded-linear dynamics,

$$x_j(t) = f_{j,t}(\sum_i W_{ij} x_i(t-1)), \quad \forall t > 0,\ v_j \in V, \tag{3.2}$$

$$\text{where } f_{j,t}(x) = \begin{cases} 0, & 0 \le x < l_{j,t}, \\ x, & l_{j,t} \le x < h_{j,t}, \\ h_{j,t}, & x \ge h_{j,t}, \end{cases}$$

is the time-dependent bound function of each node v_j (see Fig. 3.2 for an example), $\mathbf{W} = (W_{ij})$ with $W_{ij} \ge 0$ is the (weighted) adjacency matrix of the underlying network, $\{l_{j,t}\}$ and $\{h_{j,t}\}$ are the time-dependent lower and upper bounds of each node v_j, respectively, with $0 \le l_{j,t} \le h_{j,t}$. The bound values leaves extra freedom to characterize the underlying population, and we will show later that the GIP model can recover the classic models by setting specific bound values. The initial states $\mathbf{x}(0)$ are given, with $x_j(0) \in \{0\} \cup [l_{j,0}, h_{j,0}]$ and $l_{j,0} > 0$.

In the GIP model, Eq. (3.2) determines the time evolution of influence at time step t, and a node v_j is *influenced*, or *active*, at t if $x_j(t) > 0$. We represent the *overall influence* on each node v_j as

$$s_j = \sum_{t=1}^{\infty} (1-\gamma)^t x_j(t), \tag{3.3}$$

where $\gamma \in [0, 1)$ is a time-discounting factor which guarantees convergence.

In order to interpret the GIP model and the underlying process more intuitively, we construct a small undirected social network with seven agents and of a uniform weight $\alpha = 0.4$; see Fig. 3.1. For illustrative purposes, we apply the bounds $l_{j,t} = 0.8^t$ and $h_{j,t} = 2 \times 0.8^t$, $\forall t \ge 0,\ v_j \in V$, set $\gamma = 0$, and activate nodes v_1, v_2 with value 2

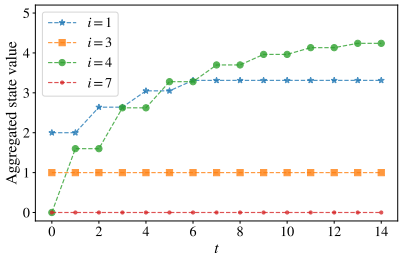


Figure 3.3: The change of the sum of influence on selected nodes along time t, from the network in Fig. 3.1 with uniform weight $\alpha = 0.4$.

and node v_3 with value 1 at $t = 0$. Then at $t = 2$, v_4 can independently influence v_1, consistent with the IC model, as a result of its high state value, while collective effort is needed to influence v_7, consistent with the LT model; see Fig. 3.1. The coexistence of the features in both models is necessary since, e.g., a social network can have people with heterogeneous levels of activity, where people of high activity are more likely to activate others. Moreover, there is a positive feedback among the nodes v_2, v_4, v_5, v_6, as they reinforce their states over time; see Fig. 3.3. This corresponds to the fact that groups of close friends keep receiving positive feedback from each other, thus reinforcing the information.

3.3.2 Limiting case I: the linear dynamics

When the upper bounds are effectively infinity and the lower bounds are effectively 0, the GIP model reduces to the *linear dynamics* of unbounded state variables, where

$$\mathbf{x}(t) = \mathbf{W}^T \mathbf{x}(t-1), \quad \forall t > 0, \tag{3.4}$$

where $\mathbf{W}$ is the (weighted) adjacency matrix of the network. Here, the following condition on the time-discounting factor γ and the spectral radius $\rho(\mathbf{W})$ is required to guarantee the convergence of the overall influence,

$$\gamma > 1 - 1/\rho(\mathbf{W}). \tag{3.5}$$

The linear dynamics can be considered as an extension to the classic IC model for continuous state variables in the deterministic case. In the classic *probabilistic* IC model, each edge weight W_{ij} corresponds to the probability that node v_i can influence node v_j in a Bernoulli trial, thus the expected amount of contribution to the state value of v_j from v_i in one time step. Then for the expected or *deterministic* case, if

we assume that (i) the expected value is the actual influence on each node at each time step, and (ii) the state values have the *no-memory* property where the ability either to be influenced or to influence others at the current time step t is independent of its previous states, then $x_j(t) = \sum_i W_{ij}x_i(t-1),\ \forall v_j \in V, t > 0$, i.e. it has the same updating function as the linear dynamics in Eq. (3.4). We refer to this model as the extended IC model hereafter.

3.3.2.1 Relation between the models

Here, we give the detailed correspondence between the GIP model and the linear dynamics by specifying "effectively infinity" for the upper bounds and "effectively 0" for the lower bounds.

Lemma 1. *If $l_{j,t} \leq l_{\min,0}w^t \leq h_{j,t},\ \forall t > 0,\ v_j \in V$, where $l_{\min,0} = \min_j l_{j,0}$ and $w = \min_{ij:W_{ij}>0} W_{ij}$, in the GIP model, then there is no threshold effect from the lower bounds, i.e. $\forall t > 0,\ v_j \in V$ s.t. $\sum_i W_{ij}x_i(t-1) > 0$,*

$$\sum_i W_{ij}x_i(t-1) \geq l_{j,t}. \tag{3.6}$$

Proof. When a node v_j has $\sum_i W_{ij}x_i(t-1) > 0$, $\exists v_i \in V$ s.t. $x_i(t-1)W_{ij} > 0$. Then

$$\sum_i W_{ij}x_i(t-1) \geq wx_*(t-1),$$

where $x_*(t) = \min\{x_j(t) : x_j(t) > 0\},\ \forall t$. Hence, we can show that statement (3.6) is true by proving

$$wx_*(t-1) \geq l_{\min,0}w^t \Leftrightarrow x_*(t-1) \geq l_{\min,0}w^{t-1}, \quad \forall t > 0,$$

by induction. (i) When $t = 1$, $x_*(0) \geq l_{\min,0}$, since $x_j(0) \in \{0\} \cup [l_{j,0}, h_{j,0}],\ \forall v_j \in V$ and $x_*(0) = \min\{x_j(0) : x_j(0) > 0\} \geq \min_j l_{j,0} = l_{\min,0}$. (ii) Suppose that $x_*(t-1) \geq l_{\min,0}w^{t-1}$ is true $\forall t \leq t'$. Then when $t = t' + 1$, from the updating function of the GIP model, for each node v_j with $x_j(t') > 0$,

$$x_j(t') = f_{j,t'}(\sum_i W_{ij}x_i(t'-1)) \geq f_{j,t'}(wx_*(t'-1)) \geq f_{j,t'}(l_{\min,0}w^{t'}) = l_{\min,0}w^{t'},$$

where the first inequality is obtained by $f_{j,t'}(\cdot)$ being a non-decreasing function, the second one is obtained together with the induction hypothesis, and the equality at the end is because $l_{j,t} \leq l_{\min,0} w^t \leq h_{j,t}$, $\forall t > 0,\ v_j \in V$. □

Theorem 2. *If $l_{j,t} \leq l_{\min,0} w^t \leq \mathbf{h}_0^T \mathbf{W}_{:,j}^t \leq h_{j,t}$, $\forall t > 0, v_j \in V$, where we let $\mathbf{W}_{:,j}^t$ denote the j-th column of matrix $\mathbf{W}^t$, $\mathbf{h}_0 = (h_{j,0})$, and $l_{\min,0}, w$ are the same as in Lemma 1, the GIP model is equivalent to the linear dynamics (3.4).*

Proof. We show that given such bounds in the GIP model,

$$\mathbf{x}(t)^T = \mathbf{x}(0)^T \mathbf{W}^t, \quad \forall t > 0, \tag{3.7}$$

by induction. By Lemma 1, there is no threshold effect from the current lower bounds, thus we will only check the saturation effect from the upper bounds in the following.

(i) When $t = 1$, for each node v_j,

$$\mathbf{x}(0)^T \mathbf{W}_{:,j} \leq \mathbf{h}_0^T \mathbf{W}_{:,j} \leq h_{j,1},$$

since $x_i(0) \leq h_{j,0}$ and $W_{ij} \geq 0$, $\forall i, j$. Hence,

$$x_j(1) = f_{j,1}(\mathbf{x}(0)^T \mathbf{W}_{:,j}) = \mathbf{x}(0)^T \mathbf{W}_{:,j},$$

and $\mathbf{x}(1)^T = \mathbf{x}(0)^T \mathbf{W}$.

(ii) Suppose $\mathbf{x}(t)^T = \mathbf{x}(0)^T \mathbf{W}^t$, $\forall t \leq t'$. Then, for each node v_j

$$\mathbf{x}(t')^T \mathbf{W}_{:,j} = \mathbf{x}(0)^T \mathbf{W}_{:,j}^{t'+1} \leq \mathbf{h}_0^T \mathbf{W}_{:,j}^{t'+1} \leq h_{j,t'+1},$$

where the equality is by the induction hypothesis, and the first inequality is again by $x_i(0) \leq h_{j,0}$ and $W_{ij} \geq 0$, $\forall i, j$. Hence,

$$x_j(t'+1) = f_{j,t'+1}(\mathbf{x}(0)^T \mathbf{W}_{:,j}^{t'+1}) = \mathbf{x}(0)^T \mathbf{W}_{:,j}^{t'+1},$$

and $\mathbf{x}(t'+1)^T = \mathbf{x}(0)^T \mathbf{W}^{t'+1}$. It is then straightforward to show that the dynamics characterised by (3.7) has updating function (3.4) of the linear dynamics. □

3.3.2.2 Differences between the models

We now illustrate how the general properties of the GIP model differ from those of the linear dynamics. Specifically, we analyse the imposed threshold effect when increasing

the lower bounds, where nodes with more overlap in their neighbourhoods of various distances will be able to achieve a higher influence on all nodes, through the lens of stochastic block models (SBMs). In Claim 1, we consider specifically a two-block planted SBM, $SBM(p_{in}, p_{out})$, where it has two communities, and the probabilities for an edge to occur inside each community and between the two communities are p_{in} and p_{out},, respectively, while allowing self-loops which characterise the tendency to maintain ones' states, in contrast to the classical SBMs.

Claim 1. *With $l_{j,0} = h_{j,0} = l_0,\ \forall v_j \in V$, and $\{h_{j,t}\}$ as in Theorem 2 in the GIP model, $SBM(p_{in}, p_{out})$ with two equally sized communities, $\mathcal{B}_1, \mathcal{B}_2$, and uniform edge weight α, while allowing self-loops, has the following properties at $t = 1$:*

1. *when $l_{j,1} \leq l_1^* = l_0\alpha,\ \forall v_j \in V$, the expected influence $\mathbb{E}[\sum_j (1-\gamma)x_j(1)]$ from the initially activated node set (i) $\mathcal{A}_0 = \{v_{i_1}, v_{i_2}\} \subset \mathcal{B}_1$, is the same as that from (ii) $\mathcal{A}_0 = \{v_{j_1}, v_{j_2}\}$ with $v_{j_1} \in \mathcal{B}_1,\ v_{j_2} \in \mathcal{B}_2$;*

2. *when $l_1^* < l_{j,1} \leq 2l_1^*,\ \forall v_j \in V$[2], if*

$$p_{in} \neq p_{out}, \tag{3.8}$$

the expected influence $\mathbb{E}[\sum_j (1-\gamma)x_j(1)]$ from set (i) is larger than that from set (ii).

Proof. For the SBM, $\mathbf{W} = \alpha\mathbf{A}$, where $\mathbf{A}$ is the (unweighted) adjacency matrix, and for each pair of nodes v_i, v_j, $A_{ij} \sim Bernoulli(p_{ij})$, with

$$p_{ij} = p_{in}\delta(\sigma_i, \sigma_j) + p_{out}(1 - \delta(\sigma_i, \sigma_j)),$$

where $\sigma_i \in \{1, 2\}$ indicates the block membership of each node v_i, and $\delta(i, j)$ is the delta function where $\delta(i, j) = 1$ if and only if $i = j$ and 0 otherwise. Further, we denote the linear part of the state vector by $\mathbf{y}(t+1) = \mathbf{W}^T\mathbf{x}(t)$, then for each node v_j at each time step $t > 0$,

$$\begin{cases} x_j(t) &= f_{j,t}(y_j(t)), \\ y_j(t) &= \sum_i W_{ij}x_i(t-1) = \alpha\sum_i A_{ij}x_i(t-1). \end{cases}$$

[2]In the specific case here, the upper bound $2l_1^*$ is equivalent to require that at most two initially activated neighbours are needed to influence a node at $t = 1$.

Then at $t = 1$[3],

$$\begin{aligned} y_j(1) &= \alpha \sum_i A_{ij} x_i(0) = \alpha l_0 \sum_{v_i \in \mathcal{A}_0} A_{ij} \\ &= \alpha l_0 \left(\sum_{v_i \in \mathcal{A}_0 \cap \mathcal{B}_{\sigma_j}} \xi_{in} + \sum_{v_i \in \mathcal{A}_0 \setminus \mathcal{B}_{\sigma_j}} \xi_{out} \right) = \alpha l_0 \left(\zeta_{k_\sigma, p_{in}} + \zeta_{k-k_\sigma, p_{out}} \right), \end{aligned}$$

where $\mathcal{A}_0 = \{v_i : x_i(0) > 0\}$ is the (given) set of initially activated nodes, $k = |\mathcal{A}_0|$, $k_i = |\mathcal{A}_0 \cap \mathcal{B}_i|$, $i = 1, 2$, $\xi_{in} \sim Bernoulli(p_{in})$, $\xi_{out} \sim Bernoulli(p_{out})$, and $\zeta_{n_m, p} \sim Bin(n_m, p)$.

Hence, for set (i),

$$y_j(1) = \alpha l_0 (\zeta_{2,p_{in}} \delta(\sigma_j, 1) + \zeta_{2,p_{out}} \delta(\sigma_j, 2)),$$

while for set (ii),

$$y_j(1) = \alpha l_0 (\zeta_{1,p_{in}} + \zeta_{1,p_{out}}).$$

When $l_{j,1} \leq l_1^*$, $x_j(1) = y_j(1)$, $\forall v_j \in V$. Then, for set (i),

$$\mathbb{E}\left[\sum_j (1-\gamma) x_j(1)\right] = (1-\gamma) n_b \times \alpha l_0 \times (2p_{in} + 2p_{out}),$$

where n_b is the size of each community, and for set (ii),

$$\mathbb{E}\left[\sum_j (1-\gamma) x_j(1)\right] = 2(1-\gamma) n_b \times \alpha l_0 \times (p_{in} + p_{out}).$$

Hence, the expected influence values at $t = 1$ are the same in the two cases.

However, when $\alpha l_0 = l_1^* < l_{j,1} \leq 2l_1^* = 2\alpha l_0$, for set (i),

$$\begin{aligned} P(y_j(1) \geq l_{j,1}) &= P(y_j(1) = 2\alpha l_0) \\ &= P(\zeta_{2,p_{in}} \delta(\sigma_j, 1) + \zeta_{2,p_{out}} \delta(\sigma_j, 2) = 2) \\ &= p_{in}^2 \delta(\sigma_j, 1) + p_{out}^2 \delta(\sigma_j, 2), \end{aligned}$$

[3]Note that in the classical SBM, the expressions of initially activated nodes can be different from others in the same community, due to the common assumption of no self-edges. However, we allow self-loops in our analysis, thus also include $v_j \in \mathcal{A}_0$ in the sum for $y_j(1)$.

thus,

$$\mathbb{E}\left[\sum_j (1-\gamma)x_j(1)\right] = (1-\gamma)\left(\sum_j 0P(y_j(1) < l_{j,1}) + 2\alpha l_0 P(y_j(1) = 2\alpha l_0)\right) \\ = (1-\gamma)n_b \times 2\alpha l_0 \times (p_{in}^2 + p_{out}^2). \tag{3.9}$$

While, for set (ii),

$$\begin{aligned} P(y_j(1) \geq l_{j,1}) &= P(y_j(1) = 2\alpha l_0) \\ &= P(\zeta_{1,p_{in}} + \zeta_{1,p_{out}} = 2) \\ &= p_{in}p_{out}, \end{aligned}$$

thus,

$$\mathbb{E}\left[\sum_j (1-\gamma)x_j(1)\right] = (1-\gamma)\left(\sum_j 0P(y_i(1) < l_{j,1}) + 2\alpha l_0 P(y_j(1) = 2\alpha l_0)\right) \\ = 2(1-\gamma)n_b \times 2\alpha l_0 \times p_{in}p_{out}. \tag{3.10}$$

Hence, the expectation from (i) as in (3.9) is larger than the one from (ii) as in (3.10) by condition (3.8). □

3.3.3 Limiting case II: the linear threshold model

With all upper bounds and the corresponding lower bounds equal to some thresholds that could vary for different nodes or time steps, all the bound functions in the GIP model only have *threshold* effect on the *linear* product, where for each node v_j at each time step $t > 0$,

$$x_j(t) = \begin{cases} \theta_{j,t}, & \sum_i W_{ij}x_i(t-1) \geq \theta_{j,t}, \\ 0, & \text{otherwise}, \end{cases} \tag{3.11}$$

with $\theta_{j,t}$ being the corresponding threshold, and $\mathbf{W}$ being the (weighted) adjacency matrix of the network.

The GIP model in this case can be treated as an extension to the classic LT model for continuous variables and deterministic thresholds. Firstly, we maintain the linear activation strategy as in (3.1), where the linear product of each node's neighbours'

state values and the edge weights is computed. Secondly, with continuous variables, we can set the activated state value to be the threshold value, thus control the source of nonlinearity to be only the activation. Note that the state values can then change magnitude along time, thus we impose time-dependent thresholds $\{\theta_{j,t}\}$. Thirdly, we also assume the no-memory property as in Sec. 3.3.2. These jointly give the updating function (3.11), with $h_{j,t} = l_{j,t} = \theta_{j,t}$, and we refer to this extension as the extended LT model hereafter.

From another perspective in extending the LT model, we can maintain the magnitude of the state value at each time step, but instead of a single value of 1 for being active, each node v_j can take values in a range $[1, m_j]$ depending on how strong the influence attempts from its neighbours are, i.e. $\sum_i W_{ij}x_i(t-1)$. Specifically, as in (3.1), a node v_j starts to take positive state value if the sum is at least a threshold, denoted l'_j here, but further, the state value increases from 1 to the highest possible state value m_j as the sum increases from l'_j to a higher value h'_j. This gives a more direct extension to the LT model, and we name it the multi-valued linear threshold (MLT) model. Explicitly, it has the updating function,

$$x_j(t) = f_j(\sum_i W_{ij}x_i(t-1)), \quad \forall t > 0,\ v_j \in V, \tag{3.12}$$

$$\text{where, } f_j(x) = \begin{cases} 0, & x < l'_j, \\ \frac{m_j - 1}{h'_j - l'_j}(x - l'_j) + 1, & l'_j \le x < h'_j, \\ m_j, & x \ge h'_j, \end{cases}$$

is the time-independent bound function, and $x_j(0) \in \{0\} \cup [1, h'_{j,0}]$ with $h'_{j,0}$ being the upper bound of node v_j's initial state value. We will show later that the two extensions to the LT model are equivalent through their relationships with the GIP model.

3.3.3.1 Relation between the models

To further disentangle the relationship between the GIP model and the threshold models, we consider the following *threshold-type* bounds, where for each node v_j at each time step $t > 0$,

$$\begin{aligned} l_{j,t} &= (\theta_{l,j}\alpha)^t l_{j,0}, \\ h_{j,t} &= \theta_{h,j}\theta_{l,j}^{t-1}\alpha^t h_{j,0}, \end{aligned} \tag{3.13}$$

$\alpha = \sum_{(v_i,v_j)\in E} W_{ij}/|E|$ is the mean weight[4], and $\theta_{l,j}, \theta_{h,j}$ are the thresholds for the lower and upper bounds, respectively, with $0 \le \theta_{l,j} \le \theta_{h,j}$. With such bounds, the GIP model reduces to the extended LT model if $l_{j,0} = h_{j,0}$ and $\theta_{l,j} = \theta_{h,j},\ \forall v_j \in V$. Furthermore, we will show that the GIP model also includes the MLT model as a special case in Theorem 3.

Theorem 3. *If the threshold-type bounds have uniform thresholds s.t.*

$$\theta_{l,j} = \theta_l, \theta_{h,j} = \theta_h, \quad \forall v_j \in V \tag{3.14}$$

the GIP model with such bounds and $l_{j,0} = 1,\ \forall v_j \in V$, is equivalent to the MLT model with $l'_j = \theta_l\alpha,\ h'_j = \theta_h\alpha h_{j,0},\ m_j = (\theta_h h_{j,0})/\theta_l$, and $h'_{j,0} = h_{j,0},\ \forall v_j \in V$, in terms of the overall influence where if we denote the time-discounting factors for the GIP model and the MLT model as γ, γ', respectively, we set $\gamma' = 1 - (1-\gamma)\theta_l\alpha$. Specifically, if we denote the state values from the GIP model as $x_j(t)$ and those from the MLT model as $x'_j(t)$,

$$x_j(t) = (\theta_l\alpha)^t x'_j(t), \quad \forall t \ge 0,\ v_j \in V. \tag{3.15}$$

Proof. We first note that if (3.15) is true, then for each node $v_j \in V$,

$$\sum_{t=0}^{\infty}(1-\gamma)^t x_j(t) = \sum_{t=0}^{\infty}(1-\gamma)^t(\theta_l\alpha)^t x'_j(t) = \sum_{t=0}^{\infty}(1-\gamma')^t x'_j(t),$$

thus the network has the same overall influence from the two models.

We then show that (3.15) is true by induction on the time step t. (i) At $t = 0$, $x_j(0) = (\theta_l\alpha)^0 x'_j(0)$, since $x_j(0) = x'_j(0),\ \forall v_j \in V$. (ii) Suppose $x_j(t) = (\theta_l\alpha)^t x'_j(t),\ \forall v_j \in V$, is true for all $t \le t'$, then for each node v_j at $t = t' + 1$, if we denote $y_j(t'+1) = \sum_i W_{ij}x_i(t')$,

$$x_j(t'+1) = \begin{cases} 0, & y_j(t'+1) < l_{j,t'+1}, \\ y_j(t'+1), & l_{j,t'+1} \le y_j(t'+1) < h_{j,t'+1}, \\ h_{j,t'+1}, & y_j(t'+1) \ge h_{j,t'+1}, \end{cases} \tag{3.16}$$

[4]With the threshold-type bounds, the condition $\sum_i W_{ij}x_i(0) \ge l_{j,1}$ at $t = 1$ is equivalent to $\sum_i (W_{ij}/\alpha)(x_i(0)/l_{j,0}) \ge \theta_{l,j}$. Hence, α will not affect the activation so long as the relative weight W_{ij}/α does not change (e.g., $W_{ij}/\alpha = 1$ if the network has uniform edge weight).

where $l_{j,t'+1} = (\theta_l\alpha)^{t'+1}$ and $h_{j,t'+1} = \theta_h\theta_l^{t'}\alpha^{t'+1}h_{j,0}$, by the GIP model. We now consider the state value $x'_j(t'+1)$ from the MLT model in the three different cases, and compare it with the state value $x_j(t'+1)$ in (3.16). (1) When

$$\sum_i W_{ij}x_j(t') < l_{j,t'+1} = (\theta_l\alpha)^{t'+1} \Leftrightarrow \sum_i W_{ij}(\theta_l\alpha)^{t'}x'_i(t') < (\theta_l\alpha)^{t'+1}$$
$$\Leftrightarrow \sum_i W_{ij}x'_j(t') < \theta_l\alpha = l'_j,$$

we have $(\theta_l\alpha)^{t'+1}x'_j(t'+1) = 0 = x_j(t'+1)$.
(2) When

$$\sum_i W_{ij}x_j(t') \geq h_{j,t'+1} = \theta_h\theta_l^{t'}\alpha^{t'+1}h_{j,0} \Leftrightarrow \sum_i W_{ij}(\theta_l\alpha)^{t'}x'_i(t') \geq \theta_h\theta_l^{t'}\alpha^{t'+1}h_{j,0}$$
$$\Leftrightarrow \sum_i W_{ij}x'_j(t') \geq \theta_h\alpha h_{j,0} = h'_j,$$

we then have $(\theta_l\alpha)^{t'+1}x'_j(t'+1) = (\theta_l\alpha)^{t'+1}m_j = (\theta_l\alpha)^{t'+1}(\theta_h h_{j,0})/\theta_l = h_{j,t'+1} = x_j(t'+1)$.
(3) Finally, in the remaining case when

$$l_{j,t'+1} \leq \sum_i W_{ij}x_j(t') < h_{j,t'+1} \Leftrightarrow\ l'_j \leq \sum_i W_{ij}x'_j(t') < h'_j,$$

the state value is in the linear regime where

$$\begin{aligned}
x'_j(t'+1) &= \frac{m_j - 1}{h'_j - l'_j}(\sum_i W_{ij}x'_j(t') - l'_j) + 1 \\
&= \frac{(\theta_h h_{j,0})/\theta_l - 1}{\theta_h\alpha h_{j,0} - \theta_l\alpha}(\sum_i W_{ij}x'_j(t') - \theta_l\alpha) + 1 \\
&= \frac{1}{\theta_l\alpha}\sum_i W_{ij}x'_j(t') = \frac{1}{\theta_l\alpha}\sum_i W_{ij}\frac{1}{(\theta_l\alpha)^{t'}}x_j(t') \\
&= \frac{1}{(\theta_l\alpha)^{t'+1}}\sum_i W_{ij}x_j(t') = \frac{1}{(\theta_l\alpha)^{t'+1}}x_j(t'+1).
\end{aligned}$$

Hence, we have shown that $x_j(t'+1) = (\theta_l\alpha)^{t'+1}x'_j(t'+1), \forall v_j \in V$. □

Further, if the network has uniform weight α, then setting $h'_j = l'_j = \theta_l\alpha$ in the MLT model is equivalent to requiring θ_l neighbours to have positive state values for activations, which is reminiscent of the constant threshold model [12]. The equivalence can pass to the GIP model and the extended LT model. Hence, $\theta_l = \theta_h = 1$

corresponds to simple contagions, where a node can be influenced by a single active neighbour, and $\theta_l = \theta_h > 1$ corresponds to complex contagions, where collective effort from the neighbourhood is required to influence a node. For general weighted networks, we also need to compare the actual edge weights with the mean value α. Due to the desired correspondence illustrated here, we will consider exclusively the threshold-type bounds hereafter.

3.3.3.2 Differences between the models

We now proceed to the general properties of the GIP model via the deviation from the extended LT model. Particularly, we consider the locally linear effect when increasing the upper bounds through θ_h in (3.14) for the threshold-type bounds (3.13), where a single active node can influence its neighbours, while it can scarcely occur when $\theta_h = \theta_l > 1$.

Claim 2. *With $l_{j,0} = h_{j,0} = l_0$, $\forall v_j \in V$, and the threshold-type bounds (3.13) satisfying (3.14) applied in the GIP model, suppose at a particular time $t' \geq 0$, there is a tree-like structure of the active nodes at t', $\mathcal{A}_{t'} = \{v_i : x_i(t') > 0\}$, and some currently inactive node v_{j_*} s.t.*

$$\exists! v_{j_0} \in \mathcal{A}_{t'} \ s.t. \ W_{j_0 j_*} > 0,$$

where $\mathbf{W}$ is the (weighted) adjacency matrix. Then if the network has uniform weight α,

1. *when $\theta_h = \theta_l > 1$, node v_{j_*} can never have positive state value at $t = t' + 1$;*
2. *when $\theta_h > \theta_l > 1$, node v_{j_*} can have positive state value at $t = t' + 1$ given a sufficiently large θ_h.*

Proof. Since the underlying network has uniform weight α, $\mathbf{W} = \alpha\mathbf{A}$. (1) When $\theta_h = \theta_l > 1$, $x_i(t') = (\theta_l\alpha)^{t'} l_0$, $\forall v_i \in \mathcal{A}_{t'}$. Then we have

$$\sum_i W_{ij_*} x_i(t') = \alpha(\theta_l\alpha)^{t'} l_0 < \theta_l\alpha(\theta_l\alpha)^{t'} l_0 = l_{j_*,t'+1},$$

thus node v_{j_*} cannot have positive state value at $t' + 1$.

(2) When $\theta_h > \theta_l > 1$, the highest possible value of node j_0 at time t' is $\theta_h \theta_l^{t'-1} \alpha^{t'} l_0$. Hence, node v_{j_*} can have positive state value at $t'+1$ if

$$\begin{aligned} W_{j_0 j_*} \theta_h \theta_l^{t'-1} \alpha^{t'} l_0 \geq l_{j_*,t'+1} &\Leftrightarrow \alpha \theta_h \theta_l^{t'-1} \alpha^{t'} l_0 \geq (\theta_l \alpha)^{t'+1} l_0 \\ &\Leftrightarrow \theta_h \geq \theta_l^2, \end{aligned}$$

which can be achieved given that θ_h is sufficiently large. □

We can take the two-block planted SBM for example again as in Sec. 3.3.2.2, and consider the case when increasing θ_h from 2 to 4 while maintaining $\theta_l = 2$. Accordingly, we analyse the performance of the following two sets of four initially activated nodes: (i) $\mathcal{A}_0 = \{v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}\} \subset \mathcal{B}_1$; (ii) $\mathcal{A}_0 = \{v_{j_1}, v_{j_2}, v_{j_3}, v_{j_4}\}$ with $v_{j_1}, v_{j_2} \in \mathcal{B}_1, v_{j_3}$, and $v_{j_4} \in \mathcal{B}_2$. We show that the increase in the expected influence (in one time step) from (i) is more than that from (ii) in Appendix B.2, which is consistent with Claim 2 since nodes in $\mathcal{B}_1$ have a higher probability to reach a higher or the highest state value in the propagation from (i). This corresponds to the feature that people with high activity are more likely to influence their friends in social networks.

3.3.4 Coexistence of regimes

As discussed in the previous sections, an important feature of the GIP model is that it can be reduced to the linear dynamics at one end and the extended LT model at the other. Here, we show that both types of propagation can coexist in a single network given that the underlying propagation process follows the GIP model, which further illustrates its generality.

Specifically, we construct a network as in Fig. 3.4, where we connect a tree, composed by nodes $\cup_{i=1,i\neq 7}^{11}\{v_i\} \cup \{v_{12}, v_{13}, v_{15}, v_{16}, v_{17}, v_{21}\}$, and part of a regular lattice, composed by nodes $\{v_6, v_7, v_8, v_{13}, v_{14}, v_{15}, v_{18}, v_{19}, v_{20}\}$. Suppose we activate the nodes in red initially. Then (1) if the underlying propagation is perfectly linear, every node will have positive state values as soon as possible, while (2) if the propagation follows the extended LT model (requiring more than one active neighbours), some nodes in the tree substructure, e.g. leaf nodes $v_3, v_{12}, v_{16}, v_{17}, v_{21}$, will never have positive state values. However, with the GIP model, we can have both characteristics in the propagation, where (1) nodes such as v_2 and v_3 are influenced immediately after they have active neighbours, while (2) nodes such as v_{12} and v_{17} do not take positive state values when there is certain amount of influence in their neighbourhood for the first

time; see Table 3.1. This is true despite the fact that the model is relatively more restricting than its general formulation since we consider the specific choice of the bounds (3.13) satisfying (3.14) for the GIP model. We choose $\theta_h = 8$ to be slightly larger than $\theta_l = 2$, and $l_{j,0} = h_{j,0} = 1,\ \forall v_j \in V$.

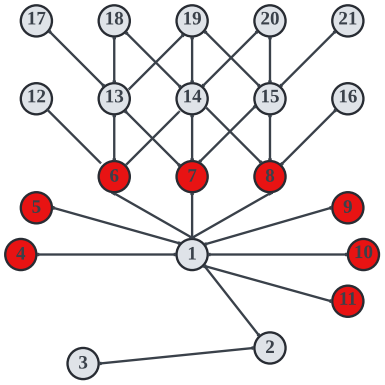


Figure 3.4: Example network with both tree substructure and regular-lattice-like substructure.

Table 3.1: The state values of representative nodes in Fig. 3.4, where the underlying propagation follows the GIP model with the threshold-type bounds (3.13) satisfying (3.14), $l_{j,0} = h_{j,0} = 1,\ \forall v_j \in V$, and $\theta_l = 2, \theta_h = 8$, and the initially activated nodes are in red.

Node	1	2	3	12	13	17	18
$t = 1$	8α	0	0	0	2α	0	0
$t = 2$	0	$8\alpha^2$	0	0	0	0	$4\alpha^2$
$t = 3$	$32\alpha^3$	0	$8\alpha^3$	$12\alpha^3$	$32\alpha^3$	0	0

Although increasing the upper bounds can make the propagation governed by the GIP model perform as the linear dynamics in some tree structures, another feature of the GIP model is that this can only happen within a limited number of time steps. Specifically with the threshold-type bounds (3.13) satisfying (3.14) and $l_{j,0} = h_{j,0} = 1,\ \forall v_j \in V$, suppose a node v_{j_0} hits the upper bounds at $t = t' > 0$, i.e. $x_{j_0}(t') = \theta_h \theta_l^{t'-1} \alpha^{t'}$, and it is the only source of influence for the nodes within certain distance r_0. Then the nodes of distance $r \leq r_0$ can have positive state values if

$$\begin{aligned} \alpha^r (\theta_h \theta_l^{t'-1} \alpha^{t'}) &\geq l_{j,t'+r} = (\theta_l \alpha)^{t'+r} \\ \Leftrightarrow \theta_h &\geq \theta_l^{r+1}, \end{aligned}$$

given that $\theta_h \geq \theta_l \geq 1$[5]. Hence, for the network in Fig. 3.4, if there are extra nodes connecting with node v_3 only or further tree structures with node v_3 as the root node, they will not have positive state values. In real networks, we can consider the case where a highly active person is very likely to influence their friends and further friends, but reinforcement among friends will be required at some point to convince people even further.

3.3.5 Derivative and backpropagation

Understanding how the overall influence change w.r.t. the initial state values is important in many applications, e.g. the IM problem we will consider in the following section. With a functional form for the GIP model at each time step, we can then analyse the derivative information though *backpropagation* or *chain rule* given that the function encoding the overall influence is differentiable, for this purpose.

Now, we slice the overall influence along the dimension of time, and consider the influence of all nodes at each time step t,

$$s_t(\mathbf{x}(0)) = \sum_j (1-\gamma)^t x_j(t). \tag{3.17}$$

Following the GIP model, $s_t(\mathbf{x}(0))$ can be considered as the output of a *neural network*, with t hidden layers, $\mathbf{x}(0)$ as the input layer, $\mathbf{W}^T$ as the weight matrix and $\{f_{j,t'}\}$ as the activation functions for nodes $\{v_j\}$ in the layers corresponding to $t' = 1, 2, \ldots, t$. The output layer only consists of one node, and is computed by summing over the elements in the previous layer corresponding to $\mathbf{x}(t) = (x_j(t))$.

We first note that s_t is not always differentiable, and even discontinuous in the general form, because each bound function $f_{j,r}$ can have a jump discontinuity at $l_{j,r}$ and be non-differentiable at $h_{j,r}$ at each time step $r \leq t$. However, $f_{j,r}$ is always semi-differentiable, specifically right-differentiable, with the right derivative,

$$\partial_+ f_{j,r}(x) = \begin{cases} 1, & l_{j,r} \leq x < h_{j,r}, \\ 0, & x < l_{j,r},\ x \geq h_{j,r}, \end{cases} \tag{3.18}$$

thus so is the function s_t, by the chain rule of semi-differentiability and $W_{ij} > 0$, $\forall i, j$. Therefore, we can obtain the right derivative of s_t with respect to $\mathbf{x}(0)$, $\partial_+ \mathbf{s}_t =$

[5]Note that if $\theta_l < 1$, there is no threshold effect from the lower bounds by Lemma 1, and then the propagation can continue until the time-discounted state values become negligible

$(\partial_+ s_t(x_j(0)))$, where if we denote $\mathbf{y}(r) = \mathbf{W}^T\mathbf{x}(r-1)$, and $\partial_+\mathbf{f}_r = (\partial_+ f_{j,r}(y_j(r)))$, $\forall r \leq t$, we have

$$\frac{1}{(1-\gamma)^t}\partial_+\mathbf{s}_t^T = \partial_+\mathbf{f}_t^T \frac{\partial\mathbf{y}(t)}{\partial\mathbf{x}(t-1)} \prod_{r=1}^{t-1} \mathbf{Diag}(\partial_+\mathbf{f}_r)\frac{\partial\mathbf{y}(r)}{\partial\mathbf{x}(r-1)},$$

where ∂ is the (bidirectional) derivative, $\partial\mathbf{y}/\partial\mathbf{x} = (\partial y_i/\partial x_j)$ with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and $\mathbf{Diag}(\cdot)$ is the corresponding diagonal matrix. Here, $\partial\mathbf{y}(t)/\partial\mathbf{x}(t-1) = \mathbf{W}^T$, $\forall t > 0$, thus the right derivative can be reduced to

$$\frac{1}{(1-\gamma)^t}\partial_+\mathbf{s}_t^T = \partial_+\mathbf{f}_t^T\mathbf{W}^T \prod_{r=1}^{t-1} \mathbf{Diag}(\partial_+\mathbf{f}_r)\mathbf{W}^T. \tag{3.19}$$

From Eq. (3.18), $\partial_+ f_{j,r}(\cdot) \in \{0, 1\},\ \forall v_j \in V, r > 0$, hence the overall right derivative in (3.19) is a restricted version of the (bidirectional) derivative in the case of the corresponding linear dynamics where $\partial f_{j,r}(\cdot) = 1,\ \forall v_j \in V, r > 0$, and

$$\frac{1}{(1-\gamma)^t}\partial\mathbf{s}_t^T = \mathbf{1}^T\left(\mathbf{W}^T\right)^t. \tag{3.20}$$

Therefore, the right derivative can be useful if the propagation is dominated by the linear part, but will almost always be 0 if the lower bounds are close to the upper bounds. Suppose at a particular time step t', $l_{j,t'} = h_{j,t'} = l_{t'},\ \forall v_j \in V$, then $f_{j,t'}(x) = l_{t'}H(x - l_{t'})$, where $H(\cdot)$ is the Heaviside step function, and

$$\frac{df_{j,t'}(x)}{dx} = l_{t'}\frac{dH(x-l_{t'})}{dx} = l_{t'}\delta(x - l_{t'}), \tag{3.21}$$

where $\delta(x)$ is the *Dirac delta function* with $\delta(x) = +\infty$ if $x = 0$ and 0 otherwise. Accordingly, the derivative $\partial s_t(x_j(0))$ for each node v_j at any time steps $t \geq t'$ can only be 0 or $+\infty$, which is not very informative. We conclude here that the derivative information generally has limited use in understanding the change of the overall influence, and accordingly in the IM problem.

3.4 Influence maximisation

Now, we proceed to a key algorithmic problem associated with information propagation, the *influence maximisation* (IM) problem, i.e. to maximise the overall influence on the whole network, and here we are interested in the constraint of a limited

number of initially activated nodes, determined by the *budget size*, corresponding to the limited resources, e.g. time and energy to influence more people as discussed in Sec. 3.2. In this section, we will first introduce a new formulation of the IM problem in Sec. 3.4.1, and give general solution methods to this task in Sec. 3.4.2. In these two sections, we focus on the general features that the IM problem could have in practice. Then we turn to the special cases when the dynamics are governed by the GIP model in Sec. 3.4.3, and further propose a customised algorithm in Sec. 3.4.4.

3.4.1 Problem formulation

With a given information propagation process, and a given function $s_j(\cdot)$ for the overall influence on each node v_j, the overall influence on the whole network is naturally,

$$s(\mathbf{x}(0)) = \sum_j s_j(\mathbf{x}(0)), \tag{3.22}$$

where $\mathbf{x}(0)$ is the initial state vector. For example, the function for individual influence $s_j(\cdot)$ can encode the state of the nodes at the end of the propagation process, and then the IM problem recovers the one in the literature associated with the classic models, while $s_j(\cdot)$ can also be Eq. (3.3) associated with the GIP model we proposed in Sec. 3.3. The IM problem is then to maximise $s(\mathbf{x}(0))$ with respect to $\mathbf{x}(0)$, subject to the constraint of limited budget size,

$$|\{v_j : x_j(0) > 0\}| \leq k, \tag{3.23}$$

where $k \in \mathbb{Z}^+$ is the budget size.

With objective (3.22) and constraint (3.23), we then formulate the IM problem as a *mixed-integer (nonlinear) programming* (MINLP),

$$\begin{aligned} \max_{\mathbf{x},\mathbf{z}} \quad & s(\mathbf{x}) \\ s.t. \quad & x_j \leq h_{j,0} z_j, \\ & x_j \geq l_{j,0} z_j, \\ & \sum_j z_j \leq k, \\ & x_j \in \mathbb{R},\ z_j \in \{0,1\},\ \forall j, \end{aligned} \tag{3.24}$$

where $0 < l_{j,0} \leq h_{j,0}$ restrict the initial level of influence of node v_j, $k \in \mathbb{Z}^+$ is the budget size, and the objective function $s(\cdot)$ is the overall influence on the whole

network as in Eq. (3.22). The continuous variables in vector $\mathbf{x}$ correspond to the initial state values, while the extra binary variables in vector $\mathbf{z}$, of the same dimension, correspond to whether or not to set positive initial state values, and they are added to appropriately impose the constraint (3.23).

The difficulty of the optimisation problem lies in the objective function $s(\mathbf{x})$. Take the GIP model together with the function (3.3) for individual influence as an example. (i) $s(\mathbf{x})$ is not always smooth and even discontinuous, since each $f_{j,t}(x)$ in (3.2) can be nonsmooth at $h_{j,t}$ and discontinuous at $l_{j,t}$. (ii) A closed-form of $s(\mathbf{x})$ cannot be obtained generally, except when $f_{j,t}(x) = x,\ \forall t > 0,\ v_j \in V$, in (3.2). (iii) The derivative information is rarely very useful in finding a maximal point, as discussed in Sec. 3.3.5. However, even in this case, we can show that the evaluation of the objective function can be solved efficiently, as in Theorem 4. This is a bonus in the deterministic setting [105]. Hence, it is necessary to treat the objective function as an input-output (black-box) system and resort to derivative-free methods (DFMs) for general solutions, which we will discuss in the following section.

Theorem 4. *Given a network $G(V, E)$ with the weight matrix $\mathbf{W}$ and an initial state vector $\mathbf{x}(0)$, and with the GIP model governing the information propagation process and Eq. (3.3) as the function for individual influence, the problem of computing the objective function $s(\mathbf{x}(0))$ in the MINLP (3.24) (i.e. Eq. (3.22)) can be solved in $O(|E|t_\epsilon)$ time, where t_ϵ is the number of time steps required for the convergence of the underlying process with tolerance $\epsilon > 0$,where*

$$\left\|(1-\gamma)^t \mathbf{x}(t)\right\|_2 < \epsilon, \quad \forall t \geq t_\epsilon.$$

Proof. The time complexity follows from Alg. 1. In each iteration t, each nonzero element of the weight matrix $\mathbf{W}$ has only one chance to be used to potentially adjust the state value $\mathbf{x}^{(t)}$, and there are overall $O(|E|)$ such elements. Therefore, the time complexity of each iteration is $O(|E|)$, and the overall evaluation has time complexity $O(|E|t_\epsilon)$, dependent on the number of steps towards convergence, t_ϵ. □

Algorithm 1 Influence evaluation.

Input: A network $G(V, E)$ with its weight matrix $\mathbf{W}$ where $W_{ij} > 0$ if $(v_i, v_j) \in E$, parameters $\{l_{j,t}\}$, $\{h_{j,t}\}$ in the GIP model, time-discounting factor γ, the initial state $\mathbf{x}(0)$ where $x_j(0) \in [l_{j,0}, h_{j,0}]$ if and only if $v_j \in \mathcal{A}_0$ (0 otherwise), and the tolerance ϵ.
Output: The value of the objective function in the MINLP (3.24), s.
Set $t \leftarrow 0$, $\mathbf{x}^{(0)} \leftarrow \mathbf{x}(0)$, and $s \leftarrow 0$.
Mark all the out-neighbours of $\mathcal{A}_0$ as potentially activated nodes, $\mathcal{N}_0 \leftarrow \bigcup_{v_j \in \mathcal{A}_0} \mathcal{N}^{out}(v_j)$.
while $\left\|(1-\gamma)^t \mathbf{x}^{(t)}\right\|_2 > \epsilon$ **do**
    $\mathcal{A}_{t+1}, \mathcal{N}_{t+1} \leftarrow \emptyset$, and $\mathbf{x}^{(t+1)} \leftarrow \mathbf{0}$;
    for each potentially activated node $v_j \in \mathcal{N}_t$ **do**
        $x_j^{(t+1)} = f_{j,t}(\sum_{i \in \mathcal{A}_t} W_{ij} x_i^{(t)})$;
        if $x_j^{(t+1)} > 0$ **then**
            $\mathcal{A}_{t+1} \leftarrow \mathcal{A}_{t+1} \cup \{v_j\}$;
            $\mathcal{N}_{t+1} \leftarrow \mathcal{N}_{t+1} \cup \mathcal{N}^{out}(v_j)$;
            $s \leftarrow s + (1-\gamma)^{t+1} x_j^{(t+1)}$;
        end if
    end for
    $t \leftarrow t + 1$;
end while

3.4.2 General solution methods

Since we cannot assume the objective to fall in a simple family, e.g. polynomials, *model-based* methods are not appropriate in the current problem. This effectively limits the choice among algorithm classes for its solution to that of *direct-search* methods. Among such algorithms, the mesh adaptive direct search (MADS) method is one of the few that has local convergence analysis when the objective function is not necessarily Lipschitz continuous [152]. Therefore, we consider the MADS for mixed variables (MV) [1] as a general solution to the IM problem, which can be implemented by the software NOMAD [10, 93]. In the following, we provide a brief overview of the MADS, which paves the way for the customised method we will propose in Sec. 3.4.4.

In the MADS for mixed variables (MV), each vector $\mathbf{y}$ is partitioned into its continuous and discrete components, $\mathbf{y}^c$ and $\mathbf{y}^d$, respectively. For the MINLP (3.24), $\mathbf{y}^c = \mathbf{x}$ and $\mathbf{y}^d = \mathbf{z}$. We denote the maximum dimensions of the continuous and discrete variables by n^c and n^d, respectively, thus $\mathbf{y}^c \in \Omega^c \subseteq \mathbb{R}^{n^c}$ and $\mathbf{y}^d \in \Omega^d \subseteq \mathbb{Z}^{n^d}$.

The general problem under consideration is a minimisation problem,

$$\min_{\mathbf{y}\in\Omega} f(\mathbf{y}),$$

where $f : \Omega \to \mathbb{R} \cup \{\infty\}$, and the domain, or the feasible region, is the union of continuous domain across possible discrete variable values, i.e.

$$\Omega = \bigcup_{\mathbf{y}^d\in\Omega^d} (\Omega^c(\mathbf{y}^d) \times \{\mathbf{y}^d\}),$$

where $\Omega^c(\mathbf{y}^d)$ indicates that the continuous domain can change with different discrete variable values, and $\Omega = \Omega^c$ if $n^d = 0$. Hence, we need to modify the objective of the MINLP (3.24) as $\min_{\mathbf{x},\mathbf{z}} -s(\mathbf{x})$ when applying the algorithm, i.e. to set $f = -s$. The constraints are incorporated in the domain, and are treated by the extreme barrier approach f_Ω, where $f_\Omega(\mathbf{y}) = f(\mathbf{y})$ if $\mathbf{y} \in \Omega$ and ∞ otherwise.

The MADS is a local search method, aiming to find a *local minimiser*, therefore it is important to define the *local neighbourhood* and then *local optimality*. For the continuous variables, the neighbourhood is well-defined as the open ball, $B_\epsilon(\mathbf{y}^c) = \{\mathbf{y}_1^c \in \mathbb{R}^{n^c} : \|\mathbf{y}_1^c - \mathbf{y}^c\|_2 < \epsilon\}$ with $\epsilon > 0$. However, different notions of the discrete neighbourhood exist. One common choice for integer variables is $\mathcal{N}(\mathbf{y}) = \{\mathbf{y}_1 \in \Omega : \mathbf{y}_1^c = \mathbf{y}^c, \|\mathbf{y}_1^d - \mathbf{y}^d\|_1 \leq 1\}$. With a user-defined discrete neighbourhood, the classical definition of local optimality can be extended to the domain of mixed variables as follows.

Definition 1. *A point $\mathbf{y} = (\mathbf{y}^c; \mathbf{y}^d) \in \Omega$ is said to be a local minimiser of a function f on Ω with respect to the set of neighbours $\mathcal{N}(\mathbf{y}) \subset \Omega$ if there exists an $\epsilon > 0$ such that $f(\mathbf{y}) \leq f(\mathbf{y}_2)$ for all $\mathbf{y}_2$ in the set*

$$\Omega \cap \left(\bigcup_{\mathbf{y}_1\in\mathcal{N}(\mathbf{y})} B_\epsilon(\mathbf{y}_1^c) \times \mathbf{y}_1^d \right),$$

where $B_\epsilon(\mathbf{y}^c) = \{\mathbf{y}_1^c \in \mathbb{R}^{n^c} : \|\mathbf{y}_1^c - \mathbf{y}^c\|_2 < \epsilon\}$ is an open ball, and $\mathcal{N}(\mathbf{y})$ is a user-defined discrete neighbourhood.

The MADS algorithm is characterised by an optional `search` step, a local `poll` step, and an `extended poll` step, where the objective f_Ω is evaluated at specific points defined on an underlying mesh M_r at each iteration r. The goal of each iteration is to find a feasible improved mesh point from the current iterate, $\mathbf{y} \in M_r$

s.t. $f_\Omega(\mathbf{y}) < f_\Omega(\mathbf{y}^{(r)})$, and the algorithm will output the point it converges to. The mesh M_r at each iteration r is a central concept in this method, which is formed as the direct product of Ω^d with the union of a finite number of lattices in Ω^c,

$$M_r = \bigcup_{q=1}^{q_{max}} M_r^q \times \Omega^d, \tag{3.25}$$

where $q = 1, ...q_{max}$ indicates each combination of discrete variable values, and the lattice M_r^q is defined through previously evaluated points, the positive spanning directions and the mesh size parameter Δ_r^m which dictates its coarseness. In the `poll` step, the method evaluates the discrete neighbourhood $\mathcal{N}(\mathbf{y}^{(r)})$, and the points whose continuous parts are close to the current iterate along certain directions, $P_r(\mathbf{y}^{(r)})$, controlled by Δ_r^m and the poll size parameter Δ_r^p. The `extended poll` step is triggered when the `poll` step fails to find an improved point. It consists of a finite sequence of `poll` steps performed around the points in $\mathcal{N}(\mathbf{y}^{(r)})$ whose objective values are sufficiently close to the incumbent value, i.e. $f_\Omega(\mathbf{y}^{(r)}) \leq f_\Omega(\mathbf{y}) \leq f_\Omega(\mathbf{y}^{(r)}) + \xi_r$, for some user-defined tolerance $\xi_r \geq \xi$ (e.g. $\xi_r = \max\{\xi, 0.05\left|f(\mathbf{y}^{(r)})\right|\}$), and we denote such set of nodes by $\mathcal{N}_r^{\xi_r}(\mathbf{y}^{(r)})$. We summarise the main ideas of the MADS in Alg. 2.

Algorithm 2 Mesh adaptive direct search for mixed variables (MADS-MV).

1: Initialisation: Set $\xi > 0$ and $\xi_0 \geq \xi$. Let $\mathbf{y}^{(0)} \in \Omega$ such that $f_\Omega(\mathbf{y}^{(0)}) < \infty$, set $\Delta_0^p \geq \Delta_0^m > 0$. Set iteration $r = 0$.
2: `SEARCH` step (optional): Evaluate f_Ω on a finite set of trial points on the mesh M_r (3.25). If an improved mesh point is found, the `SEARCH` step may terminate, skip the next `POLL` step and go directly to step 5.
3: `POLL` step: Evaluate f_Ω on the set $P_r(\mathbf{y}^{(r)}) \cup \mathcal{N}(\mathbf{y}^{(r)}) \subset M_r$ (i.e. close to the current iterate), until an improved mesh point is found, or until all points have been exhausted. If an improved mesh point is found, go to step 5.
4: `EXTENDED POLL` step: Perform a finite sequence of `poll` step starting from each point $\mathbf{y} \in \mathcal{N}_r^{\xi_r}(\mathbf{y}^{(r)}) \subseteq \mathcal{N}(\mathbf{y}^{(r)})$ with $f_\Omega(\mathbf{y}^{(r)}) \leq f_\Omega(\mathbf{y}) \leq f_\Omega(\mathbf{y}^{(r)}) + \xi_r$, until an improved mesh point is found or until all points have been exhausted.
5: Parameter update: Coarsen Δ_{r+1}^m and Δ_{r+1}^p when an improved mesh point is found and refine them otherwise. Update $\xi_r \geq \xi$, increment $r \leftarrow r + 1$, and go to step 2.

As mentioned before, the MADS is among the few algorithms that can relax the assumptions for convergence analysis to include discontinuous functions. To conclude the analysis, we mention that there are many heuristic algorithms in derivative-free methods [90], but without theoretical performance guarantees. We refer the reader

to the work of Boukouvala et al. [22] for a thorough review of DFMs in conjunction with MINLP problems.

3.4.3 Special cases

In the previous sections, we have analysed the highly general features of the IM problem, and given general solution methods accordingly. Hereafter, we turn our attention to the IM problem with the GIP model governing the information propagation process and Eq. (3.3) as the function for individual influence. In this section, we consider two special cases of the GIP model, in order to shed light on other more general scenarios.

The first special case is when the lower bounds, $\{l_{j,t}\}$, are sufficiently small in the GIP model, where we can show that the objective function is continuous and concave with respect to the continuous variables $\mathbf{x}$ as in Theorem 5. In this case, any local maximum is a global maximum, therefore the direct search algorithms can have global convergence, though it is only w.r.t. the continuous variables since the optimality of the integer part is still local, from Definition 1.

Theorem 5. *If $\{l_{j,t}\}, \{h_{j,t}\}$ are as in Lemma 1, then the objective function $s(\cdot)$ in the MINLP (3.24) is continuous and concave w.r.t. the continuous variables $\mathbf{x}$.*

Proof. By Lemma 1, the lower bounds are effectively 0. Therefore, in the GIP model (3.2), $f_{j,t}(x) = 0$ if and only if $x = 0$, $\forall v_j \in V,\ t > 0$. Hence, $f_{j,t}(x)$ is equivalent to another bound function $\tilde{f}_{j,t}(x)$ associated with the same upper bound $\tilde{h}_{j,t} = h_{j,t}$ but a different lower bound $\tilde{l}_{j,t} = 0$.

We first note that if $f_{j,t}(x)$ is continuous for all $v_j \in V$ and $t > 0$, then by the properties of composite continuous functions, the objective function $s(\cdot)$ is also continuous. Now, we focus on $f_{j,t}(x)$, and show that it is continuous by proving the continuity of $\tilde{f}_{j,t}(x)$. (1) There are two parts of the function that are always continuous, i.e. when $\tilde{l}_{j,t} < x < \tilde{h}_{j,t}$ and $x > \tilde{h}_{j,t}$. (2) We can show that the function is also continuous at the boundary points, where

$$\lim_{x \to \tilde{h}_{j,t}^-} \tilde{f}_{j,t}(x) = \tilde{h}_{j,t} = \tilde{f}_{i,t}(\tilde{h}_{j,t}) = \lim_{x \to \tilde{h}_{j,t}^+} \tilde{f}_{j,t}(x),$$

and

$$\lim_{x \to \tilde{l}_{j,t}^+} \tilde{f}_{j,t}(x) = 0 = \tilde{f}_{j,t}(\tilde{l}_{j,t}).$$

Hence, $\tilde{f}_{j,t}(x)$ is continuous for all $x \geq 0$.

Then for the concavity, we note that if $f_{j,t}(x)$ is concave for all $v_j \in V$ and $t > 0$, then since it is also nondecreasing (and the linear function $h(\mathbf{x}) = \mathbf{W}^T\mathbf{x}$ is also concave and nondecreasing), the objective function $s(\cdot)$ is concave by the properties of composite concave functions. Now, we consider specifically $f_{j,t}(x)$, and show that it is concave by the concavity of $\tilde{f}_{j,t}(x)$, i.e. $\forall x, y \geq 0$ and $\beta \in [0, 1]$,

$$\tilde{f}_{j,t}((1-\beta)x + \beta y) \geq (1-\beta)\tilde{f}_{j,t}(x) + \beta\tilde{f}_{j,t}(y). \tag{3.26}$$

(1) When $0 = \tilde{l}_{j,t} \leq x, y < \tilde{h}_{j,t}$ or $x, y \geq \tilde{h}_{j,t}$, (3.26) is true by the concavity of linear functions and constant functions, respectively.
(2) When $\tilde{l}_{j,t} \leq x < \tilde{h}_{j,t} \leq y$, $\tilde{f}_{j,t}(x) = x < \tilde{h}_{j,t}$ and $\tilde{f}_{j,t}(y) = \tilde{h}_{j,t} \leq y$. Then, if $(1-\beta)x + \beta y \geq \tilde{h}_{j,t}$,

$$\tilde{f}_{j,t}((1-\beta)x + \beta y) = \tilde{h}_{j,t} = (1-\beta)\tilde{h}_{j,t} + \beta\tilde{f}_{j,t}(y) \geq (1-\beta)\tilde{f}_{j,t}(x) + \alpha\tilde{f}_{j,t}(y);$$

otherwise $(1-\beta)x + \beta y < \tilde{h}_{j,t}$,

$$\tilde{f}_t((1-\beta)x + \beta y) = (1-\beta)x + \beta y = (1-\beta)\tilde{f}_{j,t}(x) + \beta y \geq (1-\beta)\tilde{f}_{j,t}(x) + \beta\tilde{f}_{j,t}(y).$$

(3) When $\tilde{l}_{j,t} \leq y < \tilde{h}_{j,t} \leq x$, (3.26) is true by exchanging x, y in case (2). Hence, $\tilde{f}_{j,t}(x)$ is concave for all $x \geq 0$. □

Furthermore, we can also show that the objective function is Lipschitz continuous by noting that $0 \leq f_{j,t}(x) \leq x$, $\forall t > 0, v_j \in V$ and potential value x. In this case, there are methods with global convergence, for example the new derivative-free line-search type algorithms [66]. Meanwhile, the condition of the lower bounds could be looser in practice, since the propagation does not necessarily go through the edge(s) of the smallest weight in every step. Therefore, we can have a larger region of the parameters where the objective function is (Lipschitz) continuous and concave.

The other special case is when not only $\{l_{j,t}\}$ are sufficiently small but $\{h_{j,t}\}$ are sufficiently large in the GIP model, i.e. the linear-dynamics extreme as analysed in Sec. 3.3.2. Hence,

$$\mathbf{x}(t) = \mathbf{W}^T\mathbf{x}(t-1) = \left(\mathbf{W}^T\right)^t\mathbf{x}(0),$$

and the objective function is then,

$$\begin{aligned} s(\mathbf{x}(0)) &= \sum_j \sum_{t=1}^{\infty} (1-\gamma)^t x_j(t) = \sum_{t=1}^{\infty} \mathbf{1}^T \left((1-\gamma)\mathbf{W}^T\right)^t \mathbf{x}(0) \\ &= \mathbf{1}^T \left(\left(\mathbf{I} - (1-\gamma)\mathbf{W}^T\right)^{-1} - \mathbf{I} \right) \mathbf{x}(0) = \mathbf{c}^T \mathbf{x}(0), \end{aligned} \tag{3.27}$$

where $\mathbf{c} = ((\mathbf{I} - (1-\gamma)\mathbf{W})^{-1} - \mathbf{I})\mathbf{1}$ is the Katz centrality with factor $(1-\gamma)$, $\mathbf{I}$ is the identity matrix, and the penultimate equation is obtained given that condition (3.5) is true in the linear-dynamics extreme. Hence, the objective function is linear, thus (Lipschitz) continuous, concave and smooth. The exact solution(s) in this case is achievable as in Theorem 6.

Theorem 6. *When $\{l_{j,t}\}$, $\{h_{j,t}\}$ are as in Theorem 2, then the exact solution(s) to the MINLP (3.24) is*

$$x_j^* = \begin{cases} h_{j,0}, & \text{if } j \in \mathcal{A}, \\ 0, & \text{otherwise}, \end{cases} \quad z_j^* = \begin{cases} 1, & \text{if } j \in \mathcal{A}, \\ 0, & \text{otherwise}, \end{cases} \tag{3.28}$$

where $\mathcal{A} = \{j_1, ..., j_k\}$ s.t. $h_{i,0}c_i \leq h_{j,0}c_j$, $\forall i \notin \mathcal{A}, j \in \mathcal{A}$, $\mathbf{c} = ((\mathbf{I} - (1-\gamma)\mathbf{W})^{-1} - \mathbf{I})\mathbf{1}$ is the Katz centrality with factor $(1-\gamma)$, and the uniqueness of the solution depends on the uniqueness of set $\mathcal{A}$.

Proof. By Theorem 2, the GIP model reaches the linear-dynamics extreme. Then as in (3.27), the objective function $s(\cdot)$ is linear in $\mathbf{x}$,

$$s(\mathbf{x}) = \mathbf{c}^T \mathbf{x},$$

where $\mathbf{c} = ((\mathbf{I} - (1-\gamma)\mathbf{W})^{-1} - \mathbf{I})\mathbf{1}$. For illustrative purposes, we split the proof into two parts. (1) We first analyse the MINLP (3.24) solely w.r.t. $\mathbf{x}$ while fixing the integer variables $\mathbf{z}$. The problem can then be decomposed into n sub-problems, where for each $v_j \in V$,

$$\begin{aligned} \max_{x_j} \quad & \tilde{s}_j(x_j) \coloneqq c_j x_j \\ s.t. \quad & x_j \leq h_{j,0} z_j, \\ & x_j \geq l_{j,0} z_j, \\ & x_j \in \mathbb{R}. \end{aligned} \tag{3.29}$$

Because $c_j, z_j \geq 0,\ \forall v_j \in V$, we can show that the optimal solution to each sub-problem (3.29) is $x_j^* = h_{j,0} z_i$, and the optimal value is $\tilde{s}_j^* = c_j h_{j,0} z_j$. (2) Then we consider the MINLP (3.24) w.r.t. $\mathbf{z}$ when $\mathbf{x}$ is at its optimal value, where

$$\begin{aligned} \max_{\mathbf{z}} \quad & \sum_j \tilde{s}_j^* = \sum_j c_j h_{j,0} z_j \\ s.t. \quad & \sum_j z_j \leq k, \\ & z_j \in \{0,1\}, \forall j. \end{aligned}$$

We can show that the optimal solution is to set $z_j = 1$ if node j is ranked among the top k according to its coefficient in the objective, $c_i h_{j,0}$. This gives the solution (3.28), and the uniqueness of the solution depends on the uniqueness of the top k nodes. □

Hence, the exact solution(s) when the GIP model is in the linear-dynamics extreme is to activate the k nodes of the highest product of their Katz centrality and their maximum initial values. This relates the IM problem to a well-studied centrality measure in networks, the Katz centrality. Furthermore, this solution can also serve as a warm start in the search algorithms for the MINLP (3.24), as what we will do in the following section, where the number of steps required to find a good solution is potentially proportional to how far the underlying propagation is from being the linear dynamics.

3.4.4 Customised direct search method

Here, we exploit one feature of the objective that $s(\mathbf{x}(0))$ is non-decreasing in $\mathbf{x}(0)$, which is inherited in the proof of Theorem 6. Accordingly, we propose a customised method for the MINLP (3.24).

To see why $s(\mathbf{x}(0))$ is non-decreasing in $\mathbf{x}(0)$, we first show that $\mathbf{x}(t)$ is non-decreasing in $\mathbf{x}(0), \forall t > 0$. This can be proven by induction: (i) $\mathbf{x}(1)$ is non-decreasing in $\mathbf{x}(0)$, since each $f_{j,1}(\cdot)$ and each linear function $h_j(\mathbf{x}) = \sum_i W_{ij} x_i$ in the GIP model are non-decreasing; (ii) suppose $\mathbf{x}(t)$ is non-decreasing in $\mathbf{x}(0), \forall t \leq t'$, then we can show that $\mathbf{x}(t'+1)$ is non-decreasing in $\mathbf{x}(0)$, since $\mathbf{x}(t'+1)$ is non-decreasing in $\mathbf{x}(t')$ by the same logic as (i) and $\mathbf{x}(t')$ is non-decreasing in $\mathbf{x}(0)$ by the induction hypothesis. Hence, $s(\mathbf{x}(0)) = \sum_j \sum_{t=0}^{\infty} (1-\gamma)^t x_j(t)$ is non-decreasing in $\mathbf{x}(0)$.

Therefore, maximising the objective $s(\mathbf{x})$ with respect to $\mathbf{x}$ and $\mathbf{z}$ in the MINLP (3.24) is equivalent to the maximisation with $\mathbf{x}$ and $\mathbf{z}$ reaching their highest possible

values, particularly $x_j = h_{j,0} z_j$ and $\sum_j z_j = k$. Hence, the problem is effectively reduced to the following problem only w.r.t. the binary vector $\mathbf{z}$,

$$\begin{aligned} \max_{\mathbf{z}} \quad & s(\mathbf{h}_0 \odot \mathbf{z}) \\ s.t. \quad & \sum_j z_j = k, \\ & z_j \in \{0,1\}, \forall j, \end{aligned} \tag{3.30}$$

where $\mathbf{h}_0 = (h_{j,0})$, and $\odot$ denotes the element-wise (Hadamard) product. Then, the domain Ω^d is a natural mesh to search at each iteration r,

$$M_r = \Omega^d = \{\mathbf{z} \in \{0,1\}^n : \sum_j z_j = k\}. \tag{3.31}$$

The constraints are incorporated in the domain, and are treated by the extreme barrier approach s_{Ω^d}, where $s_{\Omega^d}(\mathbf{z}) = s(\mathbf{h}_0 \odot \mathbf{z})$ if $\mathbf{z} \in \Omega^d$ and $-\infty$ otherwise. We define the neighbourhood function of binary variables $\mathbf{z}$ to be,

$$\mathcal{N}(\mathbf{z}) = \{\mathbf{y} \in \{0,1\}^n : \|\mathbf{y} - \mathbf{z}\|_1 \leq d\}, \tag{3.32}$$

where $d \in Z^+ \backslash \{1\}$, since $\|\mathbf{y} - \mathbf{z}\|_1 \geq 2$ if $\mathbf{y} \neq \mathbf{z}$ and $\mathbf{y}, \mathbf{z} \in \Omega^d$, where the shortest distance of 2 occurs when exchanging only one element of value 1 with another of value 0.

We then propose the following *customised direct search* (CDS) algorithm for the revised problem (3.30). In this algorithm, we start from the exact solution (in Theorem 6) when the GIP model is at the linear-dynamics extreme. Then at each iteration r, in the `poll` step, we search the local neighbourhood of the current candidate $\mathbf{z}^{(r)}$, until a point with sufficient improvement in the objective value has been found or all points have been exhausted. In the termination check, if an improved point has been found, the algorithm will go back to the optional `search` step, but will decrease the required improvement if a sufficiently improved point has not been found; if no improvement has been found, the algorithm outputs the current iterate and terminates; see Alg. 3 for more details. The default parameter values are set to be $\zeta = 0.1$, $\delta = 0.5$ and $d = 2$.

Algorithm 3 Customised direct search (CDS).

1: Initialisation: Set $0 < \zeta, \delta < 1$. Let $\mathbf{z}^{(0)} \in \Omega^d$ such that $z_j^{(0)} = 1$ if node $j \in \mathcal{A} = \{j_1, ..., j_k\}$ where $h_{i,0}c_i \leq h_{j,0}c_j$, $\forall i \notin \mathcal{A}, j \in \mathcal{A}$, and $\mathbf{c} = (\mathbf{I} - (1-\gamma)\mathbf{W})^{-1}\mathbf{1}$ is the Katz centrality. Set iteration $r = 0$.

2: `SEARCH` step (optional): Evaluate s_{Ω^d} on a finite subset of trial points on the mesh M_r (3.31), until a sufficiently improved mesh point $\mathbf{z}$ is found, where $s_{\Omega^d}(\mathbf{z}) > (1+\zeta)s_{\Omega^d}(\mathbf{z}^{(r)})$, or all points have been exhausted. If an improved point is found, then the `SEARCH` step may terminate, skip the next `POLL` step and go directly to step 4.

3: `POLL` step: Evaluate s_{Ω^d} on the set $\Omega^d \cap \mathcal{N}(\mathbf{z}^{(r)}) \subset M_r$ as in (3.32), until a sufficiently improved mesh point $\mathbf{z}$ is found, where $s_{\Omega^d}(\mathbf{z}) > (1+\zeta)s_{\Omega^d}(\mathbf{z}^{(r)})$, or all points have been exhausted.

4: Termination check: If an improvement is found, set $\mathbf{z}^{(r+1)}$ as the improved solution, while decreasing $\zeta \leftarrow \delta\zeta$ if a sufficient improvement has not been found, increment $r \leftarrow r+1$, and go to step 2. Otherwise, output the solution $\mathbf{z}^{(r)}$.

Therefore, local convergence is guaranteed in the termination step, by Definition 1. Global convergence could possibly be obtained with a sophisticatedly developed `search` step and a better understanding of the landscape of the objective function, in order not to be trapped in bad local optima. However, the downside of a global method is its time complexity, thus we leave the `search` step optional. Instead, the CDS method incorporates the problem's features and circumvents the worst-case complexity by initialising the search process with the exact solution when the GIP model reaches the linear-dynamics extreme, and we postulate that the local optima near this special solution are sufficiently good. We leave the detailed discussion of the time complexity to Appendix B.1.

From the current CDS method, there are two dimensions to further improve the quality of the output. We note that the current problem is equivalent to selecting a set of nodes to give value 1 (and others 0). Accordingly, there are two known methods of global convergence: (i) *brute-force*, where all node sets of size k are evaluated in order to choose an optimal one; (ii) *random sampling*, where randomly chosen node sets are evaluated, and this method has global convergence asymptotically if it samples densely enough. The two dimensions of improvement are motivated by these two global methods. On the one hand, we can enlarge the distance in defining the neighbourhood, which necessarily searches more points in the domain. Further, if the neighbourhood is as large as the whole domain, it reduces to the brute-force method. On the other hand, we can restart the searching process, i.e. steps 2,3,4 in Alg. 3, from other unexplored points randomly, which works in the same logic as the

`search` step. This strategy will give global convergence asymptotically, similar to the random sampling method.

3.5 Numerical experiments

In this section, we experimentally illustrate the rich behaviour of the GIP model, and evaluate the performance of the CDS method for the IM problem in both small and large, both synthetic and real networks. Throughout the section, $l_{j,0} = h_{j,0} = 1$, $\forall v_j \in V$, $\gamma = 0$, and we apply exclusively the threshold-type bounds (3.13) with condition (3.14), thus the lower bounds vary according to the lower bound threshold θ_l and the upper bounds also change with the upper bound threshold θ_h.

3.5.1 The proposed model for information propagation

We start from the general features of the GIP model. In accordance with Sec. 3.3, we show that the GIP model can have both the threshold effect and the locally linear effect via tuning the lower and upper bounds, respectively. Such effects cannot happen simultaneously in either the linear dynamics or the (extended) LT model, but may co-exist in real systems.

Stochastic block models. We start from simple networks generated from the two-block planted $SBM(0.9, 0.1)$, where an edge is placed between the nodes in the same community with probability $p_{in} = 0.9$ and in the different communities with $p_{out} = 0.1$. We choose these values to construct networks of assortative communities, and also to have a large difference between node sets in the same community and those not, for visualising purposes[6]. The networks have size $n = 50$ and $n_c = 2$ communities, where we label the nodes in communities one and two as 0 to 24 and 25 to 49, respectively; see Fig. 3.5 for one realisation. We assign a uniform weight $\alpha = 0.1$, to account for moderate level of trust among agents. Therefore, $\theta_l = 1$ corresponds to the critical lower bounds for the linear-dynamics extreme, where any $\theta_l > 1$ cannot always result in linear dynamics. The difference in the propagation behaviour will be quantified by the time-dependent influence of all nodes,

$$s(t) = \sum_j \sum_{t'=0}^{t} (1-\gamma)^{t'} x_j(t'), \tag{3.33}$$

[6]Note that as long as the probabilities satisfy $p_{in} > p_{out}$, similar phenomena will occur.

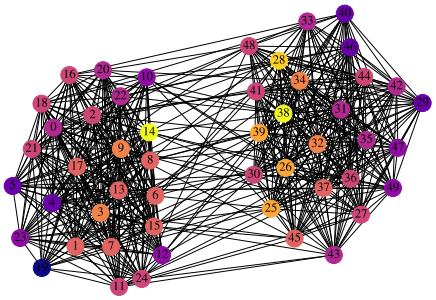


Figure 3.5: One realisation of the two-block $SBM(0.9, 0.1)$, where the lighter the colour of a node is, the higher the degree of the node is.

where $\mathbf{x}(t') = (x_j(t'))$ is the state vector at time step t' with the updating function (3.2) and a given initial state vector $\mathbf{x}(0) = (x_j(0))$. Note that $\lim_{t\to\infty} s(t) - \sum_j x_j(0)$ is the objective in the IM problem.

We first show the threshold effect imposed on top of the linear dynamics, by tuning the lower bounds of the GIP model to gradually deviate it from the linear-dynamics extreme, as in Sec. 3.3.2.2. Specifically, we consider the following two initially activated node sets as in Claim 1: (i) $\{0, 1\}$ from the same community; (ii) $\{0, 25\}$ from the different communities. The results numerically verify such effect, since the time-dependent influence from the two node sets are similar when $\theta_l = 1$, while set (i) triggers a propagation with higher influence as θ_l slightly increases; see Fig. 3.6. We note that the performance of set (ii) has larger variance, and this is because it largely depends on the inter-community edges whose existence has much smaller probability. Here, the results from set (ii) are concentrated at the values slightly above the mean (not as high as set (i)), while they also contain values substantially lower than the mean.

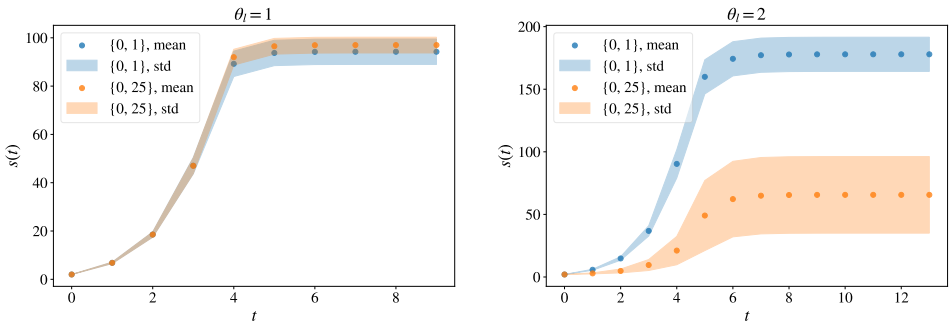


Figure 3.6: The time-dependent influence, from two different initially activated node sets, with $\theta_l = 1$ (left, the critical value for the linear-dynamics extreme) and $\theta_h = 2$ (right, a larger value) while θ_h being large (here 9000), on 1000 samples of $SBM(0.9, 0.1)$.

We then illustrate the locally linear effect added on top of the extended LT model, by tuning the upper bounds of the GIP model to differentiate it from the LT extreme. Here, we fix $\theta_l = 2$, the smallest integer for the threshold effect to take place, while increasing θ_h from 2 to a higher value, as in Sec. 3.3.3.2. We then consider the following two larger initially activated node sets: (i) $\{0, 1, 2, 3\}$ from the same community; (ii) $\{0, 1, 25, 26\}$ evenly distributed in the two communities. The numerical results justify such effect, because the time-dependent influence from set (i) has larger increase than the other when the upper bound threshold θ_h increases; see Fig. 3.7. We also observe that the propagation triggered by set (ii) has consistently higher influence than the other.

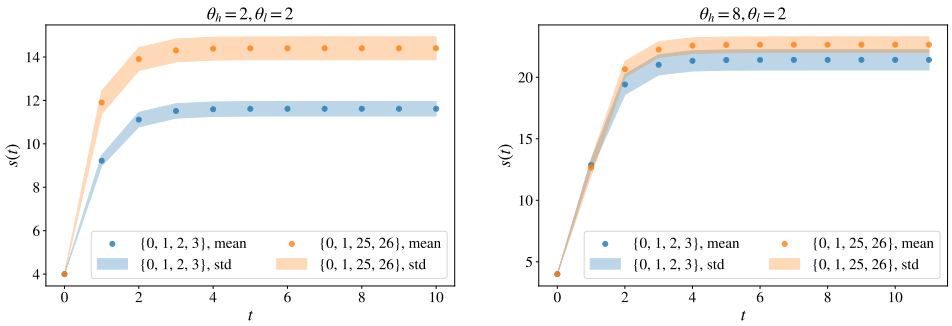


Figure 3.7: The time-dependent influence, from two different initially activated node sets, with $\theta_l = 2$ while $\theta_h = 2$ (left, the linear-threshold extreme) and $\theta_h = 8$ (right, a larger value), on 1000 samples of $SBM(0.9, 0.1)$.

Finally, we integrate the two aspects and provide a whole picture of these general features of the GIP model, by changing the upper and lower bounds simultaneously. Here, we consider both node-set pairs, $\{0, 1\}$ versus $\{0, 25\}$ and $\{0, 1, 2, 3\}$ versus $\{0, 1, 25, 26\}$, and quantify the their relative behaviour by the *ratio* δ of the overall influence (3.3) following the GIP model. We observe consistent patterns of the previously analysed features: (i) the node sets in the same community have consistently higher influence as θ_l exceeds the critical value for the other set, 1 for $\{0, 25\}$ and 2 for $\{0, 1, 25, 26\}$ (which are equal to the numbers of nodes distributed in each community); (ii) the overall influences from the node sets in each pair are increasingly similar as θ_h increases in general; see Fig. 3.8. We also notice that there is a regime where increasing the upper bounds will enlarge the (relative) difference between the node sets in each pair, which emphasises the nonlinearity in the GIP model.

Composite networks. Furthermore, we investigate the feature discussed in Sec. 3.3.4 numerically. Specifically, we construct a network by connecting two very different

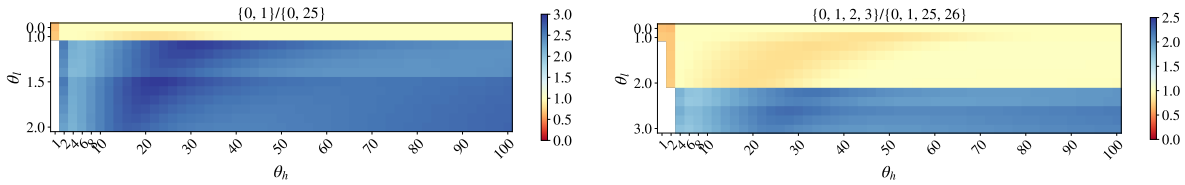


Figure 3.8: The ratios δ of the overall influence from two pairs of initially activated node sets, $\{0, 1\}$ to $\{0, 25\}$ (left) and $\{0, 1, 2, 3\}$ to $\{0, 1, 25, 26\}$ (right), with changing upper (x-axis) and lower bound (y-axis) thresholds, on 1000 samples of $SBM(0.9, 0.1)$.

structures, an Erdős Rényi (ER) random graph and a regular lattice, which we refer to as *composite network*, and explore the difference between the propagation processes from the GIP model at the LT extreme (i.e. the extended LT model) and that with other parameters, or in a general case.

The composite network is composed of the following: (i) a regular lattice of size n_o with mean degree d_o, (ii) an ER random graph of the same size n_o and with connecting probability $p_{er} = d_o/n_o$, and (iii) edges randomly placed between the two parts with a small probability p_o. Here, we choose $d_o = 4$, $n_o = 25$, $p_o = 0.01$, thus the network is of the same size as the previous SBM with $n = 50$. We label the nodes in the regular lattice as 0 to 24 and in the random part as 25 to 49; see Fig. 3.9 for one realisation. We assign a uniform weight $\alpha = 0.1$, and set the following parameters for the two

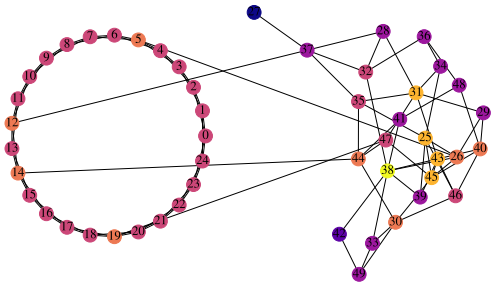


Figure 3.9: One realisation of the composite graph, where nodes 0 to 24 are from a regular lattice and nodes 25 to 49 are from a ER random graph, and the lighter the colour of a node is, the higher the degree of the node is.

cases of the GIP model: (i) $\theta_l = \theta_h = 2$ for the LT extreme, and (ii) $\theta_l = 2, \theta_h = 16$ for the general case for comparison. To emphasise the differences in the propagation spread, we consider the *number* of nodes that have positive influence up to time step

t,

$$n_a(t) = \sum_j I_{\{\sum_{t'=0}^{t}(1-\gamma)^{t'}x_j(t')>0\}},$$

where $I_{\{\cdot\}}$ is the indicator function, and $\lim_{t\to\infty} s_j(t) - x_j(0)$ is the overall influence on each node as in Eq. (3.3) with $s_j(t) \coloneqq \sum_{t'=0}^{t}(1-\gamma)^{t'}x_j(t')$.

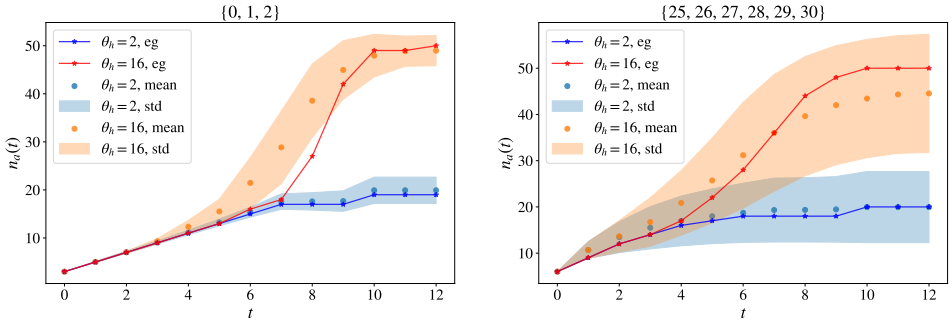


Figure 3.10: The number of nodes of positive influence up to increasing time steps (x-axis), from the GIP model at the LT extreme ($\theta_l = \theta_h = 2$) and in the general case ($\theta_l = 2, \theta_h = 16$), with the initially activated nodes in the regular lattice (left) and the ER random part (right), where "eg" corresponds to the results on the composite network in Fig. 3.9 and others are from the results on 1000 samples of the random composite network.

We observe that, particularly clearly in the composite network in Fig. 3.9, the set from either part of the network can eventually influence the whole network when the GIP model is in the general case, while only certain parts of the network can be reached when the GIP model is at the LT extreme; see Fig. 3.10. The two propagation processes initially proceed similarly in terms of the number of nodes with positive influence, while after this "initial preparation" phase, the general case gradually reaches more nodes and finally the whole network. The results are generally consistent with those obtained from the 1000 samples of the random composite network. Since the randomness is mostly from the ER model part, the results have more variance when the initially activated nodes are from this part.

3.5.2 The proposed method for influence maximisation

We now proceed to the proposed CDS method for the IM problem (with the GIP model governing the information propagation process and Eq. (3.3) as the function for individual influence). In the following, we first explore the constraint of limited

budget size for the IM problem in Sec. 3.5.2.1. We then compare the performance of the proposed CDS method with the global optima in Sec. 3.5.2.2, through networks of a relatively small scale. We finally examine its performance on relatively large networks in Sec. 3.5.2.3, both real and synthetic, by comparison with other state-of-the-art approaches.

3.5.2.1 Budget size

In this section, we consider the IM problem formulation, and particularly the constraint of limited budget size by exploring the dependence of the optimal objective value on the budget size k. We consider exclusively small networks here, because algorithms with global convergence, such as the brute-force method, require $O(n^k)$ evaluations of the objective, which prohibits their application to large networks.

Specifically, we consider a network of $n = 20$ nodes, generated from a two-block SBM with connecting probability $p_1 = 0.5$ in one community, $p_2 = 0.25$ in the other, and $p_{12} = 0.05$ between the two communities. The probabilities in the two communities are set to be different in order to separate the nodes in each community. We again assign uniform weight $\alpha = 0.1$. We compare optimal objective values s^* obtained with different budget sizes through the *relative optimal objective value*, s^*/s^*_{max}, where s^*_{max} is the optimal value from $k = n = 20$, i.e. the maximum objective value with respect to all possible k.

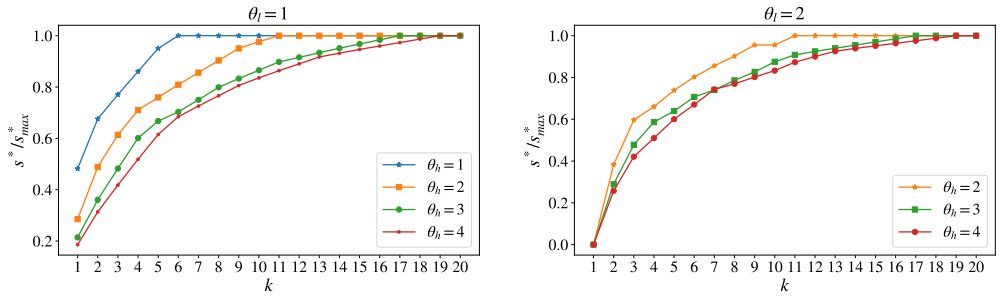


Figure 3.11: The change of relative optimal objective value s^*/s^*_{max} with respect to the budget size k (x-axis) on the SBM, when the GIP model has varying upper bound threshold θ_h while the lower bound threshold $\theta_l = 1$ (left) and $\theta_l = 2$ (right).

We observe that the optimal objective value reaches its maximum level at a smaller budget size k, as the upper bounds decrease; see Fig. 3.11. When $\theta_l = \theta_h = 2$, only activating $k = 11$ nodes initially can achieve the maximum level of influence on the network, and for $\theta_l = \theta_h = 1$, only $k = 6$ nodes are needed. This property of saturation also illustrates the rationality for the IM problem, where one aims to

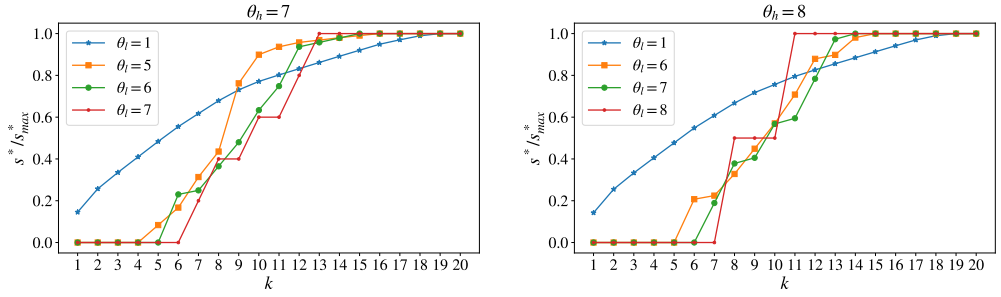


Figure 3.12: The change of relative optimal objective value s^*/s^*_{max} with respect to the budget size k (x-axis) from the SBM, when the GIP model has varying lower bound threshold θ_l while the upper bound threshold $\theta_h = 7$ (left) and $\theta_h = 8$ (right).

influence a large portion of the network from a small set of initially activated nodes: it is not only because of limited resources, but also because activating more nodes does not necessarily benefit the optimisation substantially.

Interestingly, when increasing the lower bounds, the early-saturation characteristic is not significant, but the step effect is increasingly explicit; see Fig. 3.12. When $\theta_l = \theta_h$, the optimal objective function changes as a step function w.r.t. the budget size k, and when $\theta_l = 1$, it is closer to being linear. When θ_l lies in between these two extreme cases, the optimal objective function interpolates these two shapes, with both linear-like increase and step effect, consistent with the GIP model capturing features both from the linear dynamics and the extended LT model.

3.5.2.2 Comparison with global optima

We now proceed to the proposed CDS method. From Sec. 3.4.4, we know that the CDS method is a local algorithm, thus it is important to explore how good the output is with respect to the global optima. However, as discussed in the previous section, it is practically hard for global algorithms like brute-force to obtain an optimal solution in large networks. Hence, we exclusively consider networks of a relatively small scale in this section.

In order to measure the *goodness* of the output from the CDS method, we consider the following two measures: accuracy and rank. The *accuracy* is defined as the relative value to a global optimum,

$$\tau(s; s^*) = s/s^*,$$

where s is an output from the proposed algorithm, and s^* is a global optimum. $\tau(\cdot) \in [0, 1]$ and a higher accuracy implies a better solution. Since the current problem

is purely combinatorial in terms of the initially activated node set, we also consider the output's *rank*,

$$\phi(\mathcal{A}_0; V, k) = \frac{|\{\mathcal{A} \subset V : |\mathcal{A}| = k, s(\mathbf{h}_0 \odot \mathbf{z}_{\mathcal{A}}) > s(\mathbf{h}_0 \odot \mathbf{z}_{\mathcal{A}_0})\}| + 1}{|\{\mathcal{A} \subset V : |\mathcal{A}| = k\}|},$$

where $G(V, E)$ is the underlying network, k is the budget size, $\mathcal{A}_0$ is the set of initially activated nodes corresponding to the output, $\mathbf{z}_{\mathcal{A}}$ is the binary vector whose jth element is 1 if and only if $v_j \in \mathcal{A}$, and $s(\mathbf{h}_0 \odot \mathbf{z})$ is the objective function of the revised problem (3.30). $\phi(\cdot) \in (0, 1]$, and a lower rank implies a better solution. Hence, in order to obtain the parameters in the measures, we select the *brute-force* as the reference global algorithm.

Specifically, we consider the two cases: (i) differentiating θ_l and θ_h while maintaining k, and (ii) varying k while fixing θ_l and θ_h. In case (ii), we also evaluate the performance of the initial point of the CDS method, i.e. selecting nodes by their Katz centrality (since $h_{j,0} = 1,\ \forall v_j \in V$). As representative examples, we show results from the following networks: the classic karate club network, two-block SBMs, and the composite network as in Fig. 3.9.

Karate club network. The karate club network is a social network of a university karate club [159]. It captures $n = 34$ members of the club, and has $|E| = 78$ edges indicating pairs of members who interact outside the club. This real network is extensively used in network analysis for various purposes, and is relatively small, thus is suitable here.

When $k = 3$, the CDS method can successfully find an optimal solution in all different choices of the upper and lower bounds; see Fig. 3.13. When $\theta_l \leq 1$, the algorithm returns the top 3 nodes of the highest Katz centrality (i.e., nodes $0, 32, 33$) when θ_h is large, while return those that are relatively far from each other when θ_h is small (i.e., nodes $0, 31, 33$). As θ_l increases to 2, the algorithm generally provides node sets of smaller Katz centrality but with more overlap of their neighbourhoods (e.g., $8, 32, 33$). Since different combinations of upper and lower bounds correspond to various properties of the underlying propagation process as discussed in Sec. 3.3, these demonstrate that the proposed method can capture the general properties of the IM problem. Furthermore, the time consumed by the CDS method is always less than 5% of the brute-force's which is between 1s and 4s, thus the CDS method is also much more efficient.

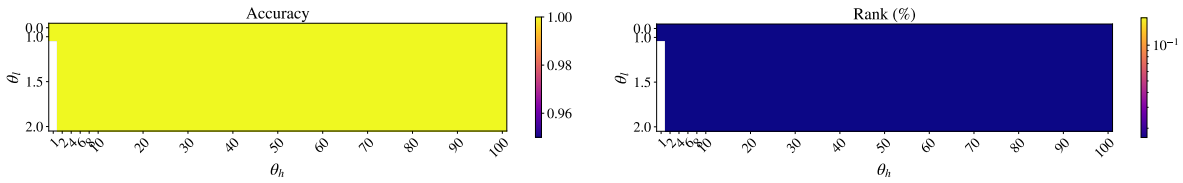


Figure 3.13: Performance of the CDS method on the karate club network in terms of the output's accuracy (left) and the rank (%, right), subject to changing upper (x-axis) and lower (y-axis) bound thresholds of the GIP model, θ_h and θ_l, respectively, when $k = 3$.

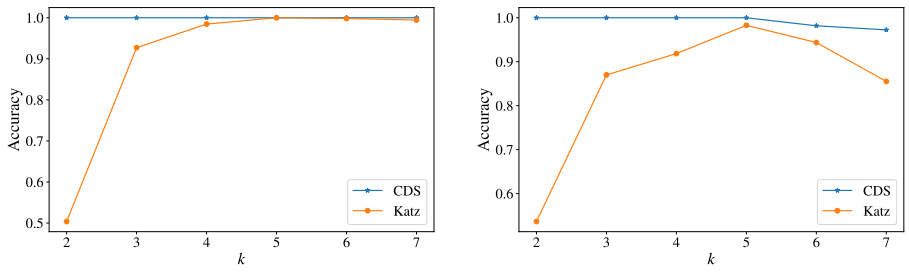


Figure 3.14: Performance of the CDS method on the Karate club network in terms of the output's accuracy, subject to changing budget size k (x-axis) when $\theta_l = 2, \theta_h = 16$ (left) and $\theta_l = \theta_h = 2$ (right), where the results from selecting nodes by the Katz centrality ("Katz") are shown to understand the change of performance of the initial points of the CDS method.

We then change the budget size k in two different cases of bounds: (i) $\theta_l = 2$, $\theta_h = 16$, and (ii) $\theta_l = 2 = \theta_h$, corresponding to the extended LT model. We observe that the CDS method can always find a global optimal solution in case (i), while the performance drops slightly as k increases but is still close-to-optimal in case (ii); see Fig. 3.14. Moreover, the time consumption of the CDS method increases approximately linearly as k rises, while for the brute-force, it changes exponentially, which makes it practically hard to obtain a global optimum under larger budget sizes.

Stochastic block models. The SBM considered here has the same size and weights as in Sec. 3.5.1, but different probabilities in the two communities, $p_1 = 0.3$ in one and $p_2 = 0.12$ in the other, in order to distinguish nodes in different communities. The connecting probability between the communities is set to be a smaller value $p_{12} = 0.01$ here.

When $k = 4$, the CDS method can find a solution either globally optimal or fairly close to optimal with all different combinations of the upper and lower bound

thresholds; see Fig. 3.15. There are only 5 cases when the CDS method cannot find a globally optimal solution. However, in such worst-case scenarios, the solutions still have accuracy over 0.95 and rank less than 0.001% in overall 230, 300 possibly initially activated sets, i.e. the CDS method can still output a top 2 set.

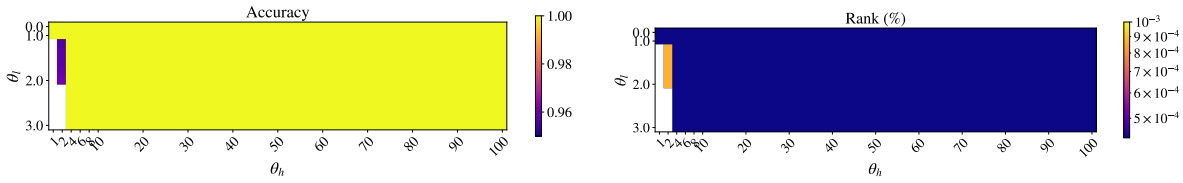


Figure 3.15: Performance of the CDS method on the SBM in terms of the output's accuracy (left) and rank (right), subject to changing upper (x-axis) and lower (y-axis) bound thresholds of the GIP model, θ_h and θ_l, respectively, when $k = 4$.

We further explore the performance of the CDS method when increasing the budget size k and $\theta_l = 2, \theta_h = 2$, one of the pairs with worst-case performance when $k = 4$. We observe that the outputs from the CDS method generally have high accuracy (greater than 0.9), and consistently low rank; see Fig. 3.16. There is a drop in accuracy when k becomes larger. Apart from the drop in performance of the initial point, it is also partially because the fixed neighbourhood size, i.e. 2 here, becomes increasingly restrictive. The former is because the current dynamics are far from the linear dynamics. The latter is inherited in the CDS method being local, since one needs to define the radius of local neighbourhood.

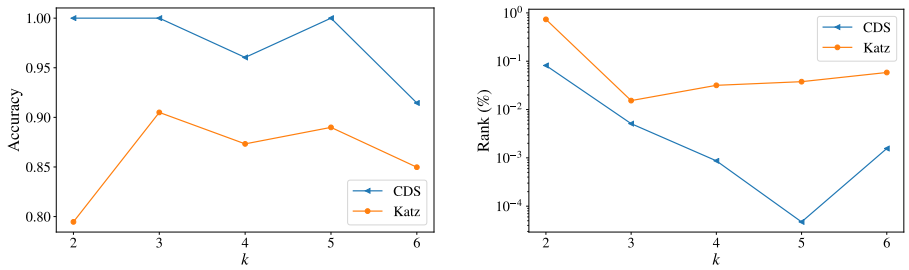


Figure 3.16: Performance of the CDS method on the SBM in terms of the output's accuracy (left) and rank (right), subject to changing budget size k (x-axis) when $\theta_l = 2, \theta_h = 2$.

Composite network. The composite network here particularly refers to the one in Fig. 3.9, constructed by connecting a regular lattice of size $n_o = 25$ and mean degree $d_o = 4$, and a ER random graph of the same size n_o and probability $p_{er} = d_o/n$, with

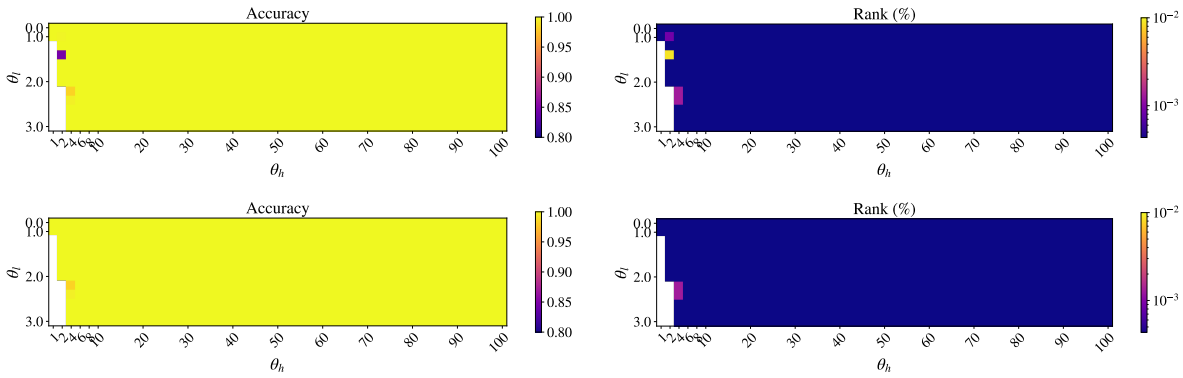


Figure 3.17: Performance of the CDS method (top) and that with community restart (bottom) on the composite network in terms of the output's accuracy (left) and the rank (%, right), subject to changing upper (x-axis) and lower (y-axis) bound thresholds of the GIP model, θ_h and θ_l, respectively, when $k = 4$.

edges randomly placed between the two parts with a small probability $p_o = 0.01$. As discussed in Sec. 3.5.1, there is noticeable difference in performance when the GIP model is at the LT extreme versus in the general case in this network. It is then interesting to explore the performance of the CDS method for the corresponding IM task, on this particular structure.

When $k = 4$, the CDS method can again find an optimal or close-to-optimal solution with different choices of the upper and lower bound thresholds: see Fig. 3.17. There are only 4 cases where the CDS method cannot output a globally optimal solution, and the worst-case scenario occurs when $\theta_l = 1.4$ and $\theta_h = 2$, where the accuracy is about 0.85 and the rank is slightly below 0.009% in overall $230,300$ candidate node sets, i.e. the CDS method can still output a top 20 set in this case.

Since there are certain cases where a global optimum cannot be reached by the (plain) CDS method, we explore one improvement strategy here: to restart the search process, i.e. steps 2,3,4 in Alg. 3, from other unexplored points. We achieve this by its noticeable community structure. Specifically, we propose the following *community restart strategy*: (i) construct a set containing different splits of k into the two communities, $\mathcal{S} := \{(k_1, k_2) : k_1 + k_2 = k, k_1, k_2 \in \mathbb{N}\}$; (ii) construct the set of initial points corresponding to activating k_1 and k_2 nodes of the highest values of $h_{j,0}c_j$ where c_j is the Katz centrality of node v_j, in communities 1 and 2, respectively, $\forall (k_1, k_2) \in \mathcal{S}$; (iii) restart the search process from each point in (ii) if it has not been explored yet. We observe that the community restart strategy can assist the CDS method to find a global optimum when the lower bound threshold $\theta_l < 2$; see Fig. 3.15. Now, the worse-case scenario has accuracy 0.98 and rank 0.001%, i.e. it can now output a top 2

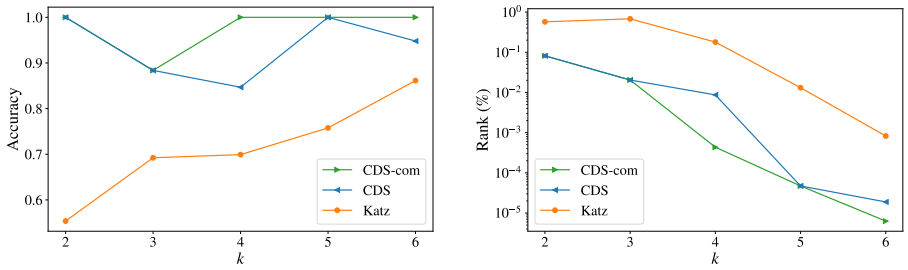


Figure 3.18: Performance of the CDS method ("CDS") and that with community restart ("CDS-com") on the composite network in terms of the output's accuracy (left) and rank (right), subject to changing budget size k (x-axis) when $\theta_l = 1.4, \theta_h = 2$.

set. We refer the reader to [135] for more theoretical results on the interplay between the community structure and complex contagions.

We then explore the performance of the CDS method, together with the community restart strategy, when varying the budget size k and $\theta_l = 1.4, \theta_h = 2$, the pair with worst-case performance when $k = 4$. The performance of the CDS method can be further improved by the community restart strategy, particularly when k is large, which also verifies the valuable information in the community structure; see Fig. 3.18. We also observe that the performance of the CDS method is generally good, with accuracy greater than 0.8 and consistently low rank, thus we will only examine the CDS method in the following.

3.5.2.3 Comparison with state-of-the-art

In the following, we explore the performance of the CDS method on both synthetic and real networks on a relatively large scale, by comparing them with the state-of-the-art methods. Hereafter, we denote the set of initially activated nodes, $\mathcal{A}_0$, as the *seed set*. We note that the vast majority of greedy algorithms following the work of Kempe et al. [85] are not applicable here, because every node sets that are not large enough[7] will return an objective value 0, and then these algorithms lack an appropriate approach to select the first few nodes, which has significant consequences in their following steps. We instead compare the CDS method with the following ones.

(i) **Random sampling** ("Random"). Randomly selecting k nodes of the network, and return them as the seed set. We repeat the process n_s times, and output

[7]It corresponds to the node sets of sizes less than θ_l in networks with uniform weights, while in weighted networks, we also need to incorporate the exact weights around each node compared to the mean weight α.

the one with the highest objective value.

(ii) **Degree-centrality method** ("Degree"). Centrality is an important measure which quantifies the significance of nodes in networks [12]. The degree-centrality method is to select the k nodes of the highest degree (centrality) in the network as the seed set.

(iii) **Katz-centrality method** ("Katz"). Finding the k nodes of the highest Katz centrality in the network as the seed set.

Since method (i) has random components, we will repeat the methods n_r times and analyse their averaged performance. In this section, we compare the performance of different methods directly through the overall influence s, i.e. the objective value in the IM problem (3.24) (or equivalently (3.30)).

The networks under consideration are composed of a large two-block SBM, and a real collaboration network. On the one hand, SBMs are considered here because community structure is a common feature in many real networks, and also to extend the previous analysis in Sec. 3.5.2.2 to a larger scale. Collaboration networks, on the other hand, are extensively used in IM experiments, because researchers believe that such networks capture key features of social networks [117]. The specific one we select also has ground-truth communities as metadata. In the following experiments, we choose $n_s = 100$ and $n_r = 10$.

Stochastic block model (SBM). Consider a relatively large network of size $n = 1000$, generated by a two-block SBM with different probabilities in the two communities, $p_1 = 0.015$ and $p_2 = 0.006$, and the connecting probability between the two being $p_{12} = 0.0005$. These values are chosen to maintain roughly the same mean degree as the SBM in Sec. 3.5.2.2, and we assign the same uniform weight $\alpha = 0.1$. We observe that the CDS method outperforms all the reference algorithms in all possible budget sizes; see Fig. 3.19. Further from the results, the performance of the degree centrality is similar to the Katz centrality, and we then prove that it is theoretically expected in Appendix B.2.

Collaboration network. The collaboration network is constructed by a comprehensive list of research papers in computer science provided by the DBLP computer science bibliography, where two authors are connected if they publish at least one paper together [158]. There are intrinsic communities defined by the publication venue, e.g. journals or conferences. Here, we randomly select two such venues (no. 6035 and

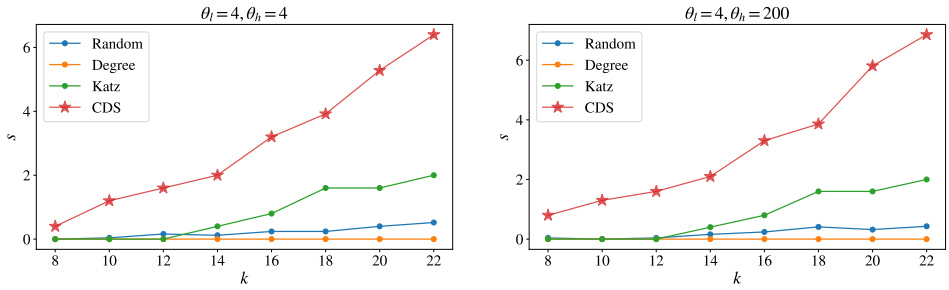


Figure 3.19: Overall influence s from different node selection algorithms applied to the SBM, subject to changing budget size (x-axis), when the GIP model has $\theta_l = \theta_h = 4$ (left) and $\theta_l = 4$, $\theta_h = 200$ (right).

6335) whose sizes are over 500 and the related authors are relatively densely connected so that the reinforcement within groups is more likely to happen. The resulting network is unweighted and connected, containing $n = 1016$ nodes and $|E| = 3469$ edges. Here, we maintain a uniform weight $\alpha = 0.1$. Since the collaboration network contains several nodes of much higher degrees than others, the performance of the linear dynamics is largely dominated by these nodes. Accordingly, we observe that both degree-centrality and Katz-centrality methods perform competitively to the CDS method when θ_h is relatively far from θ_l. However, the CDS method still outperforms others, and the distance is relatively larger as θ_h becomes smaller; see Fig. 3.20.

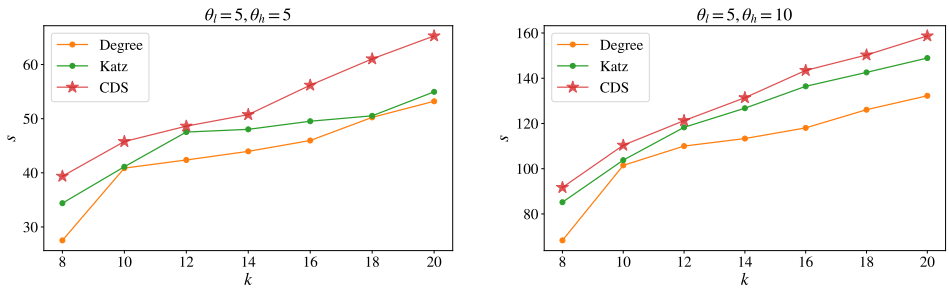


Figure 3.20: Overall influence from different node selection algorithms applied to the collaboration network, subject to changing budget size (x-axis), when the GIP model has $\theta_l = \theta_h = 5$ (left) and $\theta_l = 5$, $\theta_h = 10$ (right), where the results from random sampling is ignored because they are close to 0.

3.6 Conclusions

To understand how information propagates through social networks has many practical implications. Among the vast amount of work in this field, the IC model and the LT model are among the most popular choices of models. However, their characteristics, such as the binary state variables and no feedback mechanism, although simplifying their analysis, neglect important features of information propagation in real systems, e.g. higher influence from people of higher activity and reinforcement within close-contact groups. Therefore, we propose a novel class of information propagation model, the GIP model, which unifies the mechanisms underlying the classic models on the one hand, while extending them to address the aforementioned issues on the other hand. More importantly, the GIP model has features that each single model does not possess but may occur in real systems, as discussed in Sec. 3.3. We leave the investigation on real social networks to future work.

The general features of the GIP model necessarily lead to the IM problem with more general properties, such as the objective not being submodular. However, breaking the boundary of such restrictive properties is necessary for the IM problem to embrace a wider family of information propagation models in practice [72, 101]. Therefore, we introduce MINLP to the IM problem, and provide derivative-free methods as general solution methods. Furthermore, we propose the CDS method particularly suited for the IM problem when the information propagation is governed by the GIP model, and numerically demonstrate its close-to-optimal performance in various scenarios through experiments. One can also consider fine tuning each step of the CDS method as possible extensions to further improve the performance.

In summary, we have unified the mechanisms underlying the IC model and the LT model into a novel class of information propagation model, and propose a general framework for the IM problem that is applicable to a broad range of functions describing the overall influence in this chapter. As the two classic models are widely accepted for information propagation, we believe that the GIP model has the potential to explain more propagation phenomena on, but not restricted to, social networks. Meanwhile, the proposed IM framework provides a systematic approach to handle the case when the objective does not have restrictive properties, which provides insights into solving the IM problem in more realistic scenarios.

Chapter 4

Dynamics on signed networks and optimisation

4.1 Introduction

Dynamics on simple networks where there is only one single type of connection have been studied extensively. However, two competing types of node interactions often play an important part in shaping the system behaviour, e.g., activatory or inhibitory functions in biological systems, trustful and mistrustful connections in social or political networks, and cooperative or antagonistic relationships in the economic world [4, 58, 109]. Therefore, *signed networks*, where each connection can be either positive or negative, and the notion of balance, have become important ingredients of models and tools for many research fields over recent years. The concepts were motivated and suggested by problems in social psychology [32, 74] and have stimulated new methods for analysing social networks [88, 145, 156], biological networks [144] and so on. Signed networks also play an important role in various branches of mathematics, such as group theory, topology and physics [18, 20, 29–31]. However, most existing results are based on unweighted signed networks. Hence, the first contribution of this chapter is to extend the results in characterising the structure of signed networks to their weighted version, and also to show consistent patterns of two classic but distinct dynamics on signed networks.

Our interest in signed networks originates from the work in extracting two competing roles in Chapter 2. We should note that both the complementary and substitute relationships between products could affect how the demand changes of some products propagate to the whole system, and it is then important to consider them simultane-

ously. Therefore, signed networks are appropriate models to understand the dynamic of demand changes of the interrelated products, and then the following decision making. As in Chapter 3, we still contextualise the problem in social networks, and consider it from the perspective of information propagation and influence maximisation, which is a central topic in social, behavioural, and economic sciences and has been extensively studied over the last two decades [11, 23, 33, 39, 95, 113, 114, 125, 126].

However, most of the existing studies in information propagation and influence maximisation consider simple networks exclusively with positive connections, and the problem in signed networks is relatively unknown. Among such work, the mainstream is to extend the two classic information propagation models in simple networks, the independent cascade (IC) model and linear threshold (LT) model, to signed networks while maintaining the desired properties of the associated influence maximisation (IM) problem [38, 103]. Nevertheless, the expected influence spread for a given initially activated node set becomes increasingly difficult to estimate in signed networks, and the classic models also suffer from certain drawbacks as discussed in Chapter 3. Hence, as a second contribution, we extend the general class of information propagation model proposed in Chapter 3 to signed networks, where the extended model maintains the desired properties discussed before, such as possessing the mechanisms underlying the classic models. We also consider the associated IM problem, where we can still formulate it as a mixed integer nonlinear programming (MINLP), and refer to derivative-free methods as general solutions. Then we consider a particular case where each node only has one choice of initial state value, which could correspond to product range optimisation with the aim of finding the optimal range of products to stock in a store, and propose a customised method accordingly, with local convergence. As the last part in the main body of the thesis, we also show how the work in different chapters can collectively contribute to a better understanding of the business through an illustrative example. Last but not least, even though we refer to the retail industry as a practical example, we believe that the proposed methods are generally applicable to various scenarios when two competing types of relationships are important in shaping the dynamics.

4.2 Literature review

4.2.1 Structure of signed networks

Let $G = (V, E, \mathbf{W})$ be an undirected signed network, where $V = \{v_1, v_2, \ldots, v_n\}$ is the node set, an edge $(v_i, v_j) \in E$ is an unordered pair of two distinct nodes in the set V, and the signed (weighted) adjacency matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ describes the nonzero edge weights. Each edge in E is associated with a sign, positive or negative, characterising G as a signed network. Specifically, if there is no edge between nodes v_i, v_j, we set $W_{ij} = 0$; otherwise, $W_{ij} > 0$ denotes a positive edge, while $W_{ij} < 0$ denotes a negative edge. The degree of a node v_i is defined as

$$d_i = \sum_j |W_{ij}|, \tag{4.1}$$

and motivated by graph drawing [89], the signed Laplacian matrix in the literature is normally defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \tag{4.2}$$

where the signed degree matrix $\mathbf{D}$ is the diagonal matrix with $\mathbf{d} = (d_i)$ on its diagonal. Accordingly the signed random walk Laplacian is

$$\mathbf{L}_{rw} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W}, \tag{4.3}$$

where $\mathbf{I}$ is the identity matrix. We say the network is unweighted if $|W_{ij}| \equiv 1$, $\forall (v_i, v_j) \in E$, and has uniform weights if $|W_{ij}| \equiv c$, $\forall (v_i, v_j) \in E$, for some constant c. Most work in the literature is based on unweighted signed networks, hence in this section, we assume G to be unweighted unless otherwise mentioned, i.e. $\mathbf{W} = \mathbf{A}$ where $\mathbf{A} = (A_{ij})$ is the (unweighted) adjacency matrix with

$$A_{ij} = \begin{cases} sign(W_{ij}), & \text{if } W_{ij} \neq 0; \\ 0, & otherwise, \end{cases} \tag{4.4}$$

and function $sign(\cdot)$ indicates the sign of a value.

Introduced in 1940s [78] and primarily motivated by social and economic networks, a fundamental notion in the study of signed networks is the so-called *structural balance* [32]. A signed graph is structurally balanced if there is no cycle with an odd number

of negative edges. The following structure theorem was proved in [74].

Theorem 7 (Structure Theorem for Balance [74]). *A signed graph G is structurally balanced if and only if there is a bipartition of the node set into $V = V_1 \cup V_2$ with V_1 and V_2 being mutually disjoint and one of them being nonempty, s.t. any edge between the two node subsets is negative while any edge within each node subset is positive.*

From another aspect, Harary [75] defined a signed graph G to be antibalanced if the graph negating the edge sign is balanced. Thus, G is antibalanced if and only if there is no cycle with an odd number of positive edges. By reversing the edge sign, Harary gave the following antithetical dual result for antibalance [75].

Theorem 8 (Structure Theorem for Antibalance [75]). *A signed graph G is structurally antibalanced if and only if there is a bipartition of the node set into $V = V_1 \cup V_2$ with V_1 and V_2 being mutually disjoint and one of them being nonempty, s.t. any edge between the two node subsets is positive while any edge within each node subset is negative.*

There is a further line of research in the weakened version of structural balance, where a graph is weakly balanced if and only if no cycle has exactly one negative edge in G [46,54]. But due to its lack of dynamical interpretation, we focus on the original version of structural balance in this chapter.

The properties of being balanced or antibalanced can be characterised by the eigenvalues of both the signed Laplacian matrix (4.2) and the signed random walk Laplacian (4.3). Specifically, Kunegis et al. showed that, as in the case of the unsigned Laplacian, the signed Laplacian matrix is still positive semi-definite, and it is positive definite if and only if the underlying signed network does not have a balanced connected component [89]. Similarly, the smallest eigenvalue of the signed random walk Laplacian vanishes if and only if the network has a balanced connected component. Meanwhile, it is also known that a signed network has an antibalanced connected component if and only if the largest eigenvalue of the signed random walk Laplacian equals 2 [98]. The counterpart for the signed Laplacian has also been explored, and we refer the reader to [82,160] for more details in the results of the spectral properties of signed networks.

However, the literature in analysing the properties of the signed networks that are neither balanced nor antibalanced is limited. Among them, Atay and Liu characterised such signed networks through the idea of Cheeger inequality [8]. They defined the signed Cheeger constant through how far the network is from having a balanced

connected component, and managed to estimate the smallest eigenvalue of the Laplacian matrices from below and above with it. They obtained similar results concerning antibalance and the spectral gap between 2 and the largest eigenvalue via an antithetical dual signed Cheeger constant. We refer the reader to [15, 81] for more work on this aspect.

4.2.2 Dynamics and optimisation

There are various research directions in analysing the dynamics over signed networks, following the dynamical relationship specified along the positive and negative connections. For example, different consensus algorithms with positive and negative edges have been proposed and investigated [5, 79, 102, 111, 112, 139, 140, 157]. There exist two basic types of interactions along the negative edges: the "opposing negative dynamics" [5] where nodes are attracted by the opposite values of the neighbours, and the "repelling negative dynamics" [139] where nodes tend to be repulsive of the relative value of the states with respect to the neighbours instead of being attractive. Shi et al. [138] provide a recent review on dynamics on signed networks, where they integrate the work in this field through extending the classic DeGroot model with the two aforementioned rules on negative edges and provide a systematic approach to analyse the steady-state behaviour of the dynamics on signed networks. There is another line of research to explore the evolution of edge weights in signed networks and their interactions with the balanced structure, but since it is out of scope of our current analysis, we refer the reader to [109] and references therein.

The problem of information propagation and influence maximisation on networks only with positive connections has been studied extensively in the literature, but relatively less has been explored in the context of signed networks. Among such work, Li et al. [100] extended the voter model to signed networks with the "opposing negative dynamics". They provided systematic characterisation of both short-term and long-term behaviour of the model, through the balanced structure of the underlying signed networks. Then they considered the IM problem, where they defined the objective function based on subtracting the number of negatively influenced nodes from that of positively influenced nodes on either time scale, and proposed methods making use of its special property as a linear set function. There is also a stream of research in extending the classic information propagation models to signed networks while maintaining the desirable properties in the IM problem, e.g., the objective function being monotonic and submodular, so that the related greedy algorithms can maintain

the theoretical approximation guarantee for the IM problem [38, 103]. However, the objective function, commonly defined as the number of positively influenced nodes at the end of the process, becomes increasingly difficult to estimate in signed networks, where many simulations are necessary. Meanwhile, as discussed in Chapter 3, the binary state space could be limited in practice and the feedback between nodes is also an important feature to incorporate in the model. Hence, we will tackle the problem in terms of the GIP model we have developed in Chapter 3. Specifically, we will extend it to signed networks and analyse the associated IM problem.

4.3 Signed networks

In this section, we illustrate the interesting properties of signed networks that are significant to further understand the problem of information propagation and influence maximisation. Specifically, we will show how a signed network connects and differentiates from its unsigned counterpart from both the structural and the dynamical perspectives, and how the separate behaviour interacts with the structure balance. Throughout the section, we consider connected, undirected, weighted signed networks[1] $G = (V, E, \mathbf{W})$ where $\mathbf{W}$ is the signed (weighted) adjacency matrix, and the corresponding networks ignoring the edge sign $\bar{G} = (V, E, \bar{\mathbf{W}})$ where $\bar{\mathbf{W}}$ is the unsigned (weighted) adjacency matrix with $\bar{W}_{ij} = |W_{ij}|,\ \forall i, j$.

4.3.1 Structural properties

We first specify our definitions of structural balance. We define a signed graph to be *balanced* as in Theorem 7, and *antibalanced* as in Theorem 8. Finally, we define all the remaining signed graphs to be *strictly unbalanced* in Definition 2. Since we focus on the structural properties here, we use the terms "network" and "graph" interchangeably.

Definition 2 (Strict Unbalance)**.** *A signed graph G is strictly unbalanced if G is neither balanced nor antibalanced.*

With the definitions, we should note that balanced graphs and antibalanced graphs are not always mutually exclusive. For example, a four-node path with edge sign $(-, +, -)$ is both balanced and antibalanced, and so is a four-node cycle with edge sign $(-, +, -, +)$. We then show that the intersection between balanced

[1]For disconnected networks, we can consider each of its connected components. We allow self-loops in the networks, but multiple edges are not present.

graphs and antibalanced graphs only contains signed trees in Proposition 9, and (balanced/antibalanced) bipartite graphs as in Proposition 10. As in the literature, we define a path, walk, or cycle to be *positive* if it contains an even number of negative edges, and *negative* otherwise.

Proposition 9. *Every signed tree is both balanced and antibalanced.*

Proof. We consider an arbitrary signed tree graph, $T = (V, E, \mathbf{W})$. We first show that T is balanced. For a node $v_i \in V$, there is only one path from v_i to other nodes in V. Hence, we can partition each node $v_j \in V \backslash \{v_i\}$ according to the sign of the path from v_i to v_j, where V_1 contains v_i and the nodes of positive paths, while V_2 contains the nodes of negative paths.

We can show that V_1, V_2 is the partition corresponding to the balanced structure. Suppose there is an edge (v_h, v_l) in V_1, then there are two paths of positive sign from v_i to v_h and v_l, and one of them does not go through edge (v_h, v_l). WLOG, the path to v_h, denoted P_h, does not go through (v_h, v_l). Then $P_h + (v_h, v_l)$ is a path from v_i to v_l, then it is positive. Hence, edge (v_h, v_l) is positive. Similarly, we can show that each edge in V_2 is positive, while each edge between S and $\bar{S}$ is negative. Hence, each signed tree is balanced.

We now show that T is also antibalanced. We first construct another tree by negating the edge sign, $T' = (V, E, -\mathbf{W})$, and then following the above procedure, we can show that T' is balanced. Hence, T is antibalanced. □

Proposition 10. *A non-tree signed graph G which is both balanced and antibalanced has to be bipartite.*

Proof. The balanced graphs can be equivalently defined as that all cycles have even number of negative edges, and the antibalanced graphs can be equivalently defined as that all cycles have even number of positive edges. Hence, a non-tree signed graph G is both balanced and antibalanced if and only if every cycle has both an even number of positive edges and an even number of negative edges. This can only happen when there is no odd cycle, thus G is bipartite, given that it is not a tree. □

Now, with a better understanding of the structural balance, we characterise signed networks through the spectral properties. Specifically, we show that the eigenvalues and eigenvectors of a signed graph G are closely related to those of its unsigned counterpart $\bar{G}$ in Theorem 11, and further characterise the leading eigenvalue and eigenvector in Proposition 12. Similar results have been obtained in the literature but on unweighted graphs or other matrices, e.g. the signed Laplacian [89].

Theorem 11 (Spectral Theorem of Balance and Antibalance)**.** *Let* $\mathbf{W} = \mathbf{U}\Lambda\mathbf{U}^T$ *and* $\bar{\mathbf{W}} = \bar{\mathbf{U}}\bar{\Lambda}\bar{\mathbf{U}}^T$ *be the unitary eigendecomposition of* $\mathbf{W}$ *and* $\bar{\mathbf{W}}$*, respectively, where* $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ *and* $\bar{\mathbf{U}}\bar{\mathbf{U}}^T = \mathbf{I}$*. Let* V_1, V_2 *denote the corresponding bipartition for either balanced or antibalanced graphs, and* $\mathbf{I}_1$ *denote the diagonal matrix whose* (i, i) *element is* 1 *if* $i \in V_1$ *and* -1 *otherwise.*

1. *If* G *is balanced,*

$$\Lambda = \bar{\Lambda},\ \mathbf{U} = \mathbf{I}_1\bar{\mathbf{U}}.$$

2. *If* G *is antibalanced,*

$$\Lambda = -\bar{\Lambda},\ \mathbf{U} = \mathbf{I}_1\bar{\mathbf{U}}.$$

Proof. If G is balanced, $\mathbf{W} = \mathbf{I}_1\bar{\mathbf{W}}\mathbf{I}_1$ by definition. Then,

$$\mathbf{W} = \mathbf{I}_1\bar{\mathbf{W}}\mathbf{I}_1 = \mathbf{I}_1\bar{\mathbf{U}}\bar{\Lambda}\bar{\mathbf{U}}^{-1}\mathbf{I}_1 = (\mathbf{I}_1\bar{\mathbf{U}})\bar{\Lambda}(\mathbf{I}_1\bar{\mathbf{U}})^{-1},$$

where the second equality is by $\mathbf{I}_1\mathbf{I}_1 = \mathbf{I}$. It is the eignedecomposition of $\mathbf{W}$ by the uniqueness. Hence, $\Lambda = \bar{\Lambda},\ \mathbf{U} = \mathbf{I}_1\bar{\mathbf{U}}$.

While, if G is antibalanced, $\mathbf{W} = -\mathbf{I}_1\bar{\mathbf{W}}\mathbf{I}_1$ by definition. Then,

$$\mathbf{W} = -\mathbf{I}_1\bar{\mathbf{W}}\mathbf{I}_1 = -\mathbf{I}_1\bar{\mathbf{U}}\bar{\Lambda}\bar{\mathbf{U}}^{-1}\mathbf{I}_1 = (\mathbf{I}_1\bar{\mathbf{U}})(-\bar{\Lambda})(\mathbf{I}_1\bar{\mathbf{U}})^{-1},$$

where the second equality is by $\mathbf{I}_1\mathbf{I}_1 = \mathbf{I}$. It is the eignedecomposition of $\mathbf{W}$ by the uniqueness. Hence, $\Lambda = -\bar{\Lambda},\ \mathbf{U} = \mathbf{I}_1\bar{\mathbf{U}}$. □

Remark. *For directed signed graphs, we can show that (i) the relationships between the eigenvalues still hold, and (ii) the general eigenvectors of the two matrices have the same correspondence as the eigenvectors in Theorem 11, where the proof follows similarly but in terms of their Jordan canonical forms.*

Proposition 12. *Suppose* $\bar{G}$ *is not bipartite, thus aperiodic. Let* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ *denote the eigenvalues of* $\mathbf{W}$ *with the associated eigenvectors* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$*,* $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \cdots \geq \bar{\lambda}_n$ *denote the eigenvalues of* $\bar{\mathbf{W}}$ *with the associated eigenvectors* $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \ldots, \bar{\mathbf{u}}_n$*, and* $\rho(\cdot)$ *denotes the spectral radius. Let* V_1, V_2 *denote the corresponding bipartition for either balanced or antibalanced graphs.*

1. *If G is balanced, $\lambda_1 = \rho(\mathbf{W}) > 0$, and this eigenvalue is simple and the only one of the largest magnitude, where $|\lambda_i| < \lambda_1,\ \forall i \neq 1$.*

2. *If G is antibalanced, $\lambda_n = -\rho(\mathbf{W}) < 0$, and this eigenvalue is simple and the only one of the largest magnitude, where $|\lambda_i| < -\lambda_n,\ \forall i \neq n$.*

Meanwhile, the associated eigenvector, $\mathbf{u}_1$ for balanced graphs and $\mathbf{u}_n$ for antibalanced graphs, is the only one of the following pattern: it has positive values in one node subset in the bipartition (e.g. V_1) and negative values in the other (e.g. V_2).

Proof. Since $\bar{\mathbf{W}}$ is an non-negative matrix, and $\bar{G}$ is irreducible and aperiodic, then by Perron-Frobenius theorem, (i) $\rho(\bar{\mathbf{W}})$ is real positive and an eigenvalue of $\bar{\mathbf{W}}$, i.e. $\bar{\lambda}_1 = \rho(\bar{\mathbf{W}})$, (ii) this eigenvalue is simple s.t. the associated eigenspace is one-dimensional, (iii) the associated eigenvector, i.e. $\bar{\mathbf{u}}_1$, has all positive entries and is the only one of this pattern, and (iv) $\bar{\mathbf{W}}$ has only 1 eigenvalue of the magnitude $\rho(\bar{\mathbf{W}})$.

Then, if G is balanced, from Theorem 11, (i) $\mathbf{W}$ and $\bar{\mathbf{W}}$ share the same spectrum, and (ii) $\mathbf{U} = \mathbf{I}_1\bar{\mathbf{U}}$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ and $\bar{\mathbf{U}} = [\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \ldots, \bar{\mathbf{u}}_n]$ containing all the eigenvectors, and $\mathbf{I}_1$ is the diagonal matrix whose (i,i) element is 1 if $i \in V_1$ and -1 otherwise. Hence, $\lambda_1 = \bar{\lambda}_1 = \rho(\bar{\mathbf{W}}) = \rho(\mathbf{W})$, and this eigenvalue is simple and the only one of the largest magnitude. Meanwhile, $\mathbf{u}_1 = \mathbf{I}_1\bar{\mathbf{u}}_1$, thus it has the pattern as described and is the only one of this pattern. The results of antibalanced graphs follow similarly. □

Remark. *For a bipartite undirected graph G, if it is balanced, it will also be antibalanced, by Proposition 10. Then we can show that (i) the relationship of the spectral radius in Proposition 12 holds and the corresponding eigenvalues are still simple, but each is not the only one of the largest magnitude, since both $\lambda_1 = \rho(\mathbf{W})$ and $\lambda_n = -\rho(\mathbf{W})$ if G is balanced or antibalanced, and (ii) the patterns of $\mathbf{u}_1, \mathbf{u}_n$ in Proposition 12 still hold, and correspond to the bipartitions for the balanced structure and the antibalanced structure, respectively.*

Finally, we consider strictly unbalanced graphs. It is a relatively unexplored area, and the existing results are mostly with regard to the signed Laplacian matrices [8,89]. Here, we show a general property in terms of the (weighted) adjacency matrix that its spectral radius is smaller than the unsigned counterpart given that the signed network is neither balanced nor antibalanced. Together with Theorem 11, this is the only case when the contraction of the spectral radius occurs. Hence, if we consider dynamics

given by the signed (weighted) adjacency matrix, the corresponding state values obtained from strictly unbalanced graphs will have smaller magnitude compared with those obtained from either balanced or antibalanced graphs.

Lemma 13. *If G is strictly unbalanced, then $\exists v_i, v_j \in V$ and $l \in \mathbb{Z}^+$ s.t. there are two walks of length l between nodes v_i, v_j of different signs.*

Proof. We construct a directed signed graph G' by making each edge in G bidirectional in G' while maintaining the same sign in both directions. We note that G is strictly unbalanced if and only if G' is strictly unbalanced, and also that the statement, i.e., $\exists v_i, v_j \in V$ and $l \in \mathbb{Z}^+$ s.t. there are two walks of length l between nodes v_i, v_j of different signs, is true in G if and only if it is true in G'. Hence, we prove the lemma through G'.

By construction, G' contains cycles of length 2. (i) If G' is periodic, then G' is bipartite, because of the presence of length-2 cycle(s). Suppose the statement is not true, i.e. all walks of the same length between each pair of nodes $v_h, v_l \in V$ have the same sign, then all cycles are positive, noting that the edges connecting the same pair of nodes have the same sign, thus G' is balanced, which leads to contradiction. (ii) Otherwise, G' is aperiodic, then the statement is true by Proposition 3.5 in [100]. □

Theorem 14. *G is strictly unbalanced if and only if $\rho(\mathbf{W}) < \rho(\bar{\mathbf{W}})$.*

Proof. We first note that if $\rho(\mathbf{W}) < \rho(\bar{\mathbf{W}})$, then G is strictly unbalanced, since the spectral radius will be the same if G is balanced or antibalanced by Theorem 11.

For the other direction, if G is strictly unbalanced, by Lemma 13, $\exists v_i, v_j \in V$ and $l_1 \in \mathbb{Z}^+$ s.t. there are two walks of length l_1 between nodes v_i, v_j of different signs. Then

$$\left|(\mathbf{W}^{l_1})_{ij}\right| < (\bar{\mathbf{W}}^{l_1})_{ij},$$

where $(\mathbf{W})_{ij}$ indicates the (i, j) element of a matrix $\mathbf{W}$. Hence, for sufficiently large l_2, the walks between each pair of nodes will be able to go through the two walks of different signs between nodes v_i, v_j, thus $\forall v_h, v_k \in V$,

$$\left|(\mathbf{W}^{l_2})_{hk}\right| < (\bar{\mathbf{W}}^{l_2})_{hk}.$$

Then for each vector $\mathbf{x} = (x_h) \in \mathbb{R}^n$ and $\|\mathbf{x}\|_2 = 1$, we can find $\bar{\mathbf{x}} = (|x_h|)$ s.t. $\|\bar{\mathbf{x}}\|_2 = 1$ and

$$\left|(\mathbf{W}^{l_2}\mathbf{x})_h\right| = \left|\sum_k (\mathbf{W}^{l_2})_{hk} x_k\right| \leq \sum_k \left|(\mathbf{W}^{l_2})_{hk} x_k\right| < \sum_k (\bar{\mathbf{W}}^{l_2})_{hk} |x_k| = (\bar{\mathbf{W}}^{l_2}\bar{\mathbf{x}})_h,$$

where $(\mathbf{x})_h$ indicates the h-th element of an vector $\mathbf{x}$. Therefore, $\left\|\mathbf{W}^{l_2}\mathbf{x}\right\|_2 < \left\|\bar{\mathbf{W}}^{l_2}\bar{\mathbf{x}}\right\|_2$. Hence, by definition,

$$\left\|\mathbf{W}^{l_2}\right\|_2 = \max_{\|\mathbf{x}\|_2=1} \left\|\mathbf{W}^{l_2}\mathbf{x}\right\|_2 < \max_{\|\mathbf{y}\|_2=1} \left\|\bar{\mathbf{W}}^{l_2}\mathbf{y}\right\|_2 = \left\|\bar{\mathbf{W}}^{l_2}\right\|_2,$$

then $\rho(\mathbf{W})^{l_2} = \rho(\mathbf{W}^{l_2}) < \rho(\bar{\mathbf{W}}^{l_2}) = \rho(\bar{\mathbf{W}})^{l_2}$, and finally $\rho(\mathbf{W}) < \rho(\bar{\mathbf{W}})$. □

4.3.2 Dynamical properties

We now characterise the dynamics on signed networks. Specifically, we consider the following dynamics which are closely related to the information propagation: (i) the linear dynamics and (ii) the extended LT model, which we will specify in Secs. 4.3.2.1 and 4.3.2.2, respectively. Throughout this section, we use $x_i(t) \in \mathbb{R}$ to represent the state value of node v_i, and $\mathbf{x}(t) = (x_i(t))$ denotes the vector consisting of $x_i(t)$. In analysing the dynamics, we are interested in the following properties: (i) the asymptotic behaviour,

$$x_j(\infty) = \lim_{t\to\infty} x_j(t),$$

and (ii) the weighted sum of state values,

$$s_j = \sum_{t=1}^{\infty} (1-\gamma)^t x_j(t),$$

where γ is a time-discounting factor that guarantees convergence. This quantity will be interpreted as the overall influence on each node v_j later in this chapter.

4.3.2.1 The linear dynamics

Starting from an initial vector $\mathbf{x}(0)$, the linear dynamics evolves as follows,

$$\mathbf{x}(t) = \mathbf{W}^T\mathbf{x}(t-1), \quad \forall t > 0, \tag{4.5}$$

where $\mathbf{W}$ is the signed (weighted) adjacency matrix. In unsigned networks, every edge represents the potential to spread the information or influence others, thus at each time step t, the linear dynamics sum all possible walks from a node to another of length t via $\bar{\mathbf{W}}^t$ in Eq. (3.4). However, in signed networks, there are also edges of the negative sign, which means there is potential to reinforce the information flow in the opposite direction. The way the linear dynamics incorporate this feature is to add a negative sign if the corresponding walk is negative via $\mathbf{W}^t$ in Eq. (4.5).

Hence, the state vector at each time step $t > 0$ is

$$\mathbf{x}(t)^T = \mathbf{x}(0)^T \mathbf{W}^t,$$

and the vector containing the overall influence on each node is

$$\mathbf{s}^T = \sum_{t=1}^{\infty} (1-\gamma)^t \mathbf{x}(t)^T = \mathbf{x}(0)^T \sum_{t=1}^{\infty} (1-\gamma)^t \mathbf{W}^t = \mathbf{x}(0)^T \mathbf{S},$$

where $\mathbf{S} \coloneqq \sum_{t=1}^{\infty} (1-\gamma)^t \mathbf{W}^t$, and each entry S_{ij} indicates the overall influence from node v_i onto node v_j if $x_i(0) = 1$. The following condition on the factor γ and the spectral radius $\rho(\mathbf{W})$ is required to guarantee convergence,

$$\gamma > 1 - 1/\rho(\mathbf{W}). \tag{4.6}$$

Hence, for the asymptotic state values, we analyse the evolution of $\mathbf{W}^t$ over time, and for the overall influence, we explore the properties of $\mathbf{S}$.

Starting from the unitary decomposition, $\mathbf{W} = \mathbf{U}\Lambda\mathbf{U}^T = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^T$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of $\mathbf{W}$ and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are the associated eigenvectors, we have

$$\mathbf{W}^t = \sum_{i=1}^{n} \lambda_i^t \mathbf{u}_i \mathbf{u}_i^T, \tag{4.7}$$

and

$$\mathbf{S} = \sum_{t=1}^{\infty} (1-\gamma)^t \mathbf{W}^t = \sum_{i=1}^{n} \left(\sum_{t=1}^{\infty} (1-\gamma)^t \lambda_i^t \right) \mathbf{u}_i \mathbf{u}_i^T = \sum_{i=1}^{n} \frac{(1-\gamma)\lambda_i}{1-(1-\gamma)\lambda_i} \mathbf{u}_i \mathbf{u}_i^T. \tag{4.8}$$

Hence, the behaviour of $\mathbf{W}^t$ is dominated by the eigenvectors associated with the eigenvalues significantly different from 0, and for sufficiently large t, the behaviour can be approximated by the leading eigenvector and leading eigenvalue. Similarly,

the behaviour of $\mathbf{S}$ is dominated by the eigenvectors associated with the eigenvalues whose infinite sum $(1-\gamma)\lambda_i/(1-(1-\gamma)\lambda_i)$ is sufficiently far from 0. Furthermore, we note that when the leading eigenvalue is positive, its infinite sum will definitely have the largest magnitude, but it is not necessarily the case when the leading eigenvalue is negative. Consider the function $f(x) \coloneqq x/(1-x)$ with $x \in (-1, 1)$. $f(x)$ is increasing in x, has positive value when $x \in (0, 1)$ and negative value when $x \in (-1, 0)$, and $|f(x)| > |f(-x)|$ where $x > 0$. Then the largest magnitude a negative value could approach is $\lim_{x\to-1} |f(x)| = 1/2$, while $f(x) \geq 1/2$ when $x \geq 1/3$ and $f(x)$ goes to infinity as x approaches 1. Therefore, positive eigenvalues are more important, and specifically, if $(1-\gamma)\lambda_1 \geq 1/3$, the corresponding infinite sum has the largest magnitude and could potentially dominate the behaviour of $\mathbf{S}$.

In the following, we consider specific balanced structures in signed networks, specifically balance, antibalance and strict unbalance. We denote the eigenvalues of the (weighted) adjacency matrix ignoring the edge sign $\bar{\mathbf{W}}$ by $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \cdots \geq \bar{\lambda}_n$, with the associated eigenvectors $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \ldots, \bar{\mathbf{u}}_n$, thus $\bar{\mathbf{W}} = \sum_{i=1}^{n} \bar{\lambda}_i \bar{\mathbf{u}}_i \bar{\mathbf{u}}_i^T$. We denote the partition corresponding to the balanced or antibalanced structure by V_1, V_2, and the diagonal matrix corresponding to the partition by $\mathbf{I}_1$ which is a diagonal matrix with its (i, i) element being 1 if $v_i \in V_1$ and -1 otherwise.

Balanced networks. If the network is balanced with partition V_1, V_2, then from Theorem 11, we have

$$\mathbf{W}^t = \mathbf{I}_1 \left(\sum_{i=1}^{n} \bar{\lambda}_i^t \bar{\mathbf{u}}_i \bar{\mathbf{u}}_i^T \right) \mathbf{I}_1 = \mathbf{I}_1 \bar{\mathbf{W}}^t \mathbf{I}_1, \tag{4.9}$$

thus the magnitude of the elements in $\mathbf{W}^t$ evolves as in the simple network ignoring the edge sign (i.e. $\bar{\mathbf{W}}^t$), and the difference is in the sign pattern: the (i, j) element is positive if nodes $v_i, v_j \in V_1$ or $v_i, v_j \in V_2$ and negative otherwise. This means that at each time step t, the influence on a node from the nodes in the same partition is positive, while the influence from the nodes in the other partition is negative. Further, if the network is irreducible and aperiodic, then from Proposition 12, when t is sufficiently large,

$$\mathbf{W}^t \approx \bar{\lambda}_1^t \mathbf{I}_1 \bar{\mathbf{u}}_1 \bar{\mathbf{u}}_1^T \mathbf{I}_1,$$

the asymptotic behaviour can be approximated by the term associated with the leading eigenvalue and the associated eigenvector, which has the same sign pattern.

For the overall influence, we have

$$\mathbf{S} = \sum_{t=1}^{\infty}(1-\gamma)^t\mathbf{W}^t = \mathbf{I}_1\left(\sum_{t=1}^{\infty}(1-\gamma)^t\bar{\mathbf{W}}^t\right)\mathbf{I}_1, \tag{4.10}$$

thus similarly, the elements in $\mathbf{S}$ have the same magnitude as in the simple network ignoring the edge sign, but only the elements corresponding to the node pairs in the same partition are positive (negative otherwise). Both are expected since $\mathbf{W}^t$ maintains the same magnitude as $\bar{\mathbf{W}}^t$ and the same sign pattern for each time step t. For this reason, if one aims to maximise the overall influence on a signed network, a reasonable start is to activate nodes in the partition that are more closely connected with each other (than the other partition). Further, if the network is irreducible and aperiodic, then from Proposition 12 and the property of $f(x) = x/(1-x)$ $(x \in (1,1))$, we can approximate $\mathbf{S}$ by

$$\frac{(1-\gamma)\bar{\lambda}_1}{1-(1-\gamma)\bar{\lambda}_1}\mathbf{I}_1\bar{\mathbf{u}}_1\bar{\mathbf{u}}_1^T\mathbf{I}_1,$$

and the accuracy depends on how far the leading eigenvalue is from the remaining ones in terms of the magnitude (of the function value $f(\cdot)$).

Antibalanced networks. If the network is antibalanced with partition V_1, V_2, then from Theorem 11, we have

$$\mathbf{W}^t = \mathbf{I}_1\left(\sum_{i=1}^{n}(-\bar{\lambda}_i)^t\bar{\mathbf{u}}_i\bar{\mathbf{u}}_i^T\right)\mathbf{I}_1 = (-1)^t\mathbf{I}_1\bar{\mathbf{W}}^t\mathbf{I}_1, \tag{4.11}$$

thus again, the magnitude of the elements in $\mathbf{W}^t$ evolves as in the simple network ignoring the edge sign, but here the sign pattern alternates over time: when t is odd, the (i,j) element is *negative* if nodes $v_i, v_j \in V_1$ or $v_i, v_j \in V_2$ and positive otherwise; while t is even in the following step, the (i,j) element becomes *positive* if nodes $v_i, v_j \in V_1$ or $v_i, v_j \in V_2$ and negative otherwise. Hence, the antibalanced structure is highly unstable. Further, if the network is irreducible and aperiodic, then from Proposition 12, when t is sufficiently large,

$$\mathbf{W}^t \approx (-\bar{\lambda}_1)^t\mathbf{I}_1\bar{\mathbf{u}}_1\bar{\mathbf{u}}_1^T\mathbf{I}_1,$$

the asymptotic behaviour can be approximated by the term associated with the leading eigenvalue and the associated eigenvector, where the sign pattern is the same as

Eq. (4.11).

Then for the overall influence, we have

$$\mathbf{S} = \sum_{t=1}^{\infty}(1-\gamma)^t\mathbf{W}^t = \mathbf{I}_1\left(\sum_{t=1}^{\infty}(-1)^t(1-\gamma)^t\bar{\mathbf{W}}^t\right)\mathbf{I}_1. \tag{4.12}$$

Hence here, the elements in $\mathbf{S}$ have smaller magnitude than those in the simple network ignoring the edge sign, which is expected from the alternating sign pattern of $\mathbf{W}^t$ over time. Also, the sign pattern of $\mathbf{S}$ depends on the specific sign and magnitude of edge weights: for each element (i, j), if nodes $v_i, v_j \in V_1$ or $v_i, v_j \in V_2$, the value will be *negative* if there are more weighted walks of odd length than those of even length connecting them, and positive otherwise, while if nodes are in different parts, the value will be *positive* if there are more weighted walks of odd length than those of even length connecting them, and negative otherwise.

Remark. *As an extreme example, we can consider a bipartite network where there are only edges between V_1 and V_2 and all edges are positive. Then the network is antibalanced with the partition V_1, V_2. Also, there are only walks of even length connecting the node pairs in the same parts, while walks of odd lengths connecting node pairs in different parts. Hence by the above, all elements in $\mathbf{S}$ is positive. As validation, this is true because all edges are positive, and the initial state values are also positive.*

Strictly unbalanced networks. In all the remaining networks, neither are they like balanced networks where all walks of length $t + 1$ have the same sign as the walks of length t connecting each pair of nodes v_i, v_j for each $t > 0$, nor are they like antibalanced graphs where all walks of length $t+1$ have the opposite sign as the walks of length t connecting each pair of nodes v_i, v_j for each $t > 0$. Hence to characterise its performance, we propose the following measures to quantify how far a network is from being balanced or antibalanced. Motivated by the signed Cheeger inequality [8], we measure the distance from being balanced by

$$d_b(G) = \lambda_{min}(\mathbf{L}_{rw}(G)), \tag{4.13}$$

and the distance from being antibalanced by

$$d_a(G) = 2 - \lambda_{max}(\mathbf{L}_{rw}(G)), \tag{4.14}$$

where $\mathbf{L}_{rw}(G)$ is the random walk Laplacian of the signed network G as in Eq. (4.3), and $\lambda_{min}(\cdot), \lambda_{max}(\cdot)$ return the smallest and the largest eigenvalues, respectively. We note that in simple networks, the smallest eigenvalue of the random walk Laplacian is trivially 0 (corresponding to the transition matrix has a trivial largest eigenvalue 1) and the smallest nontrivial one is important from many aspects, including the relaxation time of random walks [110], while in signed networks, the smallest eigenvalue is nontrivial.

Therefore, depending on how far the signed network is from being balanced or antibalanced, it can have performance closer to that of balanced or antibalanced networks.

(i) If $d_b(G) < d_a(G)$, we expect G to be closer to being balanced. Then $\forall v_i, v_j \in V, t > 0$, we expect that most walks of length $t+1$ connecting nodes v_i, v_j have the same sign as most walks of length t, thus $\mathbf{W}^t$ tends to maintain the same sign pattern over time.

(ii) If $d_b(G) > d_a(G)$, we expect G to be closer to being antibalanced. Then $\forall v_i, v_j \in V, t > 0$, we expect that most walks of length $t+1$ connecting nodes v_i, v_j have the opposite sign as most walks of length t, thus $\mathbf{W}^t$ tends to alternate the sign pattern over time.

From Theorem 14, $\rho(\mathbf{W}) < \rho(\bar{\mathbf{W}})$ where $\rho(\cdot)$ is the spectral radius or the eigenvalue of the largest magnitude, thus when t is sufficiently large, elements in $\mathbf{W}^t$ will have smaller magnitude than those in $\bar{\mathbf{W}}^t$. Meanwhile, $(1-\gamma)^t\mathbf{W}^t$ converges to $\mathbf{0}$ in a shorter time steps than the cases when the signed network is balanced or antibalanced, thus fewer summands are significant for the overall influence in $\mathbf{S}$. However, this does not necessarily mean that the elements in $\mathbf{S}$ will be smaller than the other two cases.

4.3.2.2 The linear threshold model

Starting from an initial vector $\mathbf{x}(0)$, the extended LT model evolves as follows on signed networks, where $\forall v_j \in V, t > 0$,

$$x_j(t) = \begin{cases} \theta_{j,t}, & \sum_i W_{ij} x_i(t) \geq \theta_{j,t}, \\ -\theta_{j,t}, & \sum_i W_{ij} x_i(t) \leq -\theta_{j,t}, \\ 0, & \text{otherwise}, \end{cases} \tag{4.15}$$

where $\theta_{j,t}$ is the threshold to trigger the propagation, either positively or negatively, and $\mathbf{W}$ is the signed (weighted) adjacency matrix. The model is an extension of (3.11) to signed networks, where a node cannot only take a *positive* state value when the sum of its neighbours' state values is sufficiently positive, but also a *negative* state value if the sum is sufficiently negative, corresponding to the "opposing rule". Hence, the dynamics only have threshold effect on the linear product, and we still refer to it as the extended LT model in this section[2].

A theoretical understanding of threshold models on simple networks is still an active area of research, and here we consider the even more challenging case of signed networks. Hence, we start from a specific network structure, regular (ring) lattices with uniform magnitude of the edge weight, α, and analyse the behaviour of the extended LT model when the whole neighbourhood of a node (including itself), referred to as the *central node*, is activated. Specifically, we denote the degree of nodes in the regular lattices as $\bar{d}$, and apply the following geometric series (cf. the threshold-type bounds (3.13) in Chapter 3) for the threshold values

$$\theta_{j,t} = (\theta_l \alpha)^t l_0,$$

where $\theta_l > 0$ is the uniform threshold (for all nodes), and $l_0 > 0$ is the magnitude of the initially activated state value with $x_j(0) = \pm l_0$ if node v_j is activated initially and 0 otherwise. Then the updating function (4.15) is now

$$x_j(t) = \begin{cases} (\theta_l \alpha)^t l_0, & \sum_{v_i \in \mathcal{A}_{t-1}} A_{ij} \geq \theta_l, \\ -(\theta_l \alpha)^t l_0, & \sum_{v_i \in \mathcal{A}_{t-1}} A_{ij} \leq -\theta_l, \\ 0, & \text{otherwise}, \end{cases} \tag{4.16}$$

where $\mathbf{A}$ is the signed adjacency matrix as in Eq. (4.4), and $\mathcal{A}_t = \{v_i : x_i(t) \neq 0\}$. Hence, θ_l here is the threshold on the number of neighbours that are positively activated over those that are negatively activated (in the previous time step). For the overall influence on each node, we set $\gamma = 0$ for simplicity.

We start from analysing the extended LT model on simple regular lattices (ignoring the edge sign), and then proceed to signed regular lattices through their balanced structures. In simple regular lattices, $x_i(t) \geq 0$, $\forall v_i \in V, t \geq 0$, hence the whole neighbourhood of the central node that are activated initially have positive state

[2]Note that we modify the name to be the extended signed LT model later in this chapter to emphasise the incorporation of negative edges

values. We find that when

$$\theta_l \leq \theta_l^* = \bar{d}/2, \tag{4.17}$$

$\exists v_i \in V, t > 0,\ s.t.\ x_i(t) > 0$, i.e. some node has positive state value, or is successfully influenced, at certain time step other than the initial start[3], and the condition is the same as the one in [34]. Specifically, at each $t > 0$:

(i) $x_j(t) = (\theta_l \alpha)^t l_0$ if $v_j \in \mathcal{A}_{t-1}$;

(ii) $\bar{d} - 2(\lceil \theta_l \rceil - 1)$ more nodes that are closest to $\mathcal{A}_{t-1}$ will be activated with the same state value $(\theta_l \alpha)^t l_0$, if there is any.

Therefore, $x_i(t) > 0,\ \forall v_i \in V$, for sufficiently large t. For the sum of individual influence or the overall influence on the whole network, suppose the regular lattice is infinitely large, then

$$\begin{aligned} s = \sum_j s_j &= \sum_{t=1}^{\infty} \left((1 + \bar{d}) + (\bar{d} - 2(\lceil \theta_l \rceil - 1))t \right) (\theta_l \alpha)^t l_0 \\ &= \frac{\theta_l \alpha l_0}{1 - \theta_l \alpha} \left((1 + \bar{d}) + (\bar{d} - 2(\lceil \theta_l \rceil - 1)) \frac{1}{1 - \theta_l \alpha} \right). \end{aligned} \tag{4.18}$$

In the following, we will specify the behaviour of the extended LT model on signed regular lattices that are balanced, antibalanced or strictly unbalanced. Here, $x_i(t) < 0$ is possible for each node v_i at each time step $t \geq 0$, and particularly we will specify whether to positively or negatively activate a node initially.

Balanced regular lattices. We consider the activations that are consistent with the balanced partition, where we positively activate the central node v_i, and for each of its neighbours v_j, we positively activate it with $x_j(0) = l_0$ if $W_{ij} > 0$ and negatively activate it with $x_j(0) = -l_0$ otherwise; see Fig. 4.1 for example.

Similarly, we find that when condition (4.17) is true, $\exists v_i \in V, t > 0,\ s.t.\ x_i(t) \neq 0$, i.e. some node has nonzero state value, or is successfully influenced, at certain time step other than the initial start, and we refer to this phenomenon as "certain propagation" on the signed networks. However, with the edge sign, there are more interesting patterns, where at each $t > 0$:

[3] In this specific case, under the same condition, (i) every nodes will have positive state values at some time step, and (ii) only $d/2 + 1$ consecutive nodes are needed to be activated initially in order to trigger the propagation with feature (i).

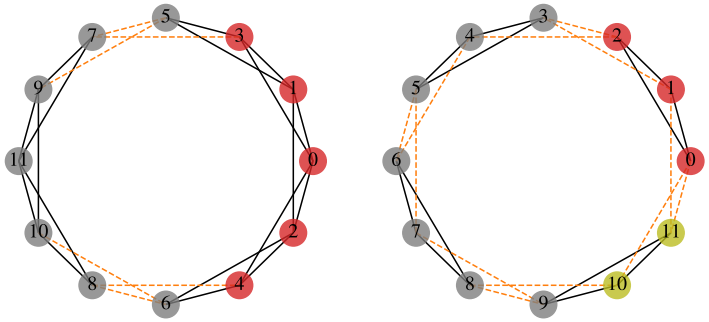


Figure 4.1: Example of signed regular lattices of degree 4 with different structurally balanced configurations, where positive edges are in black, negative edges are dashed in orange, and the whole neighbourhood of node v_0 is in different colour(s) from the others (in grey), with the ones that are positively activated in red and the others that are negatively activated in green.

(i)

$$x_j(t) = \begin{cases} (\theta_l \alpha)^t l_0, & \text{if } v_j \in \mathcal{A}_{t-1}^+, \\ -(\theta_l \alpha)^t l_0, & \text{if } v_j \in \mathcal{A}_{t-1}^-, \end{cases}$$

where $\mathcal{A}_t = \mathcal{A}_t^+ \cup \mathcal{A}_t^-$ with $\mathcal{A}_t^+ = \{v_j : x_j(t) > 0\}$ and $\mathcal{A}_t^- = \{v_j : x_j(t) < 0\}$;

(ii) $\bar{d} - 2(\lceil \theta_l \rceil - 1)$ more nodes that are closest to $\mathcal{A}_{t-1}$ will be activated if there is any, where

$$x_j(t) = \begin{cases} (\theta_l \alpha)^t l_0, & \text{if } \exists v_{i_1} \in \mathcal{A}_{t-1}^+ \text{ or } v_{i_2} \in \mathcal{A}_{t-1}^- \text{ s.t. } A_{i_1 j} > 0 \text{ or } A_{i_2 j} < 0, \\ -(\theta_l \alpha)^t l_0, & \text{if } \exists v_{i_1} \in \mathcal{A}_{t-1}^+ \text{ or } v_{i_2} \in \mathcal{A}_{t-1}^- \text{ s.t. } A_{i_1 j} < 0 \text{ or } A_{i_2 j} > 0. \end{cases}$$

Hence, each $x_j(t)$ has the same magnitude as the state value on the corresponding simple regular lattice. For the sign pattern, $\mathcal{A}_{t-1}^+ \subset \mathcal{A}_t^+$ and $\mathcal{A}_{t-1}^- \subset \mathcal{A}_t^-$, $\forall t > 0$, i.e. the nodes, once activated, remain active and maintain the sign of their state values over time, which is similar to the evolution of $\mathbf{W}^t$ in linear dynamics when the underlying signed network is balanced.

For the overall influence, we consider the following two distinct cases of the specific distribution of the edge sign, where we assume the network is infinitely large.

(a) There are only a few negative edges: if we label the nodes symmetrically with regard to node v_0 as in the left of Fig. 4.1, only edges between the nodes of label less than or equal to $\bar{d} + t_0(\bar{d} - 2(\lceil \theta_l \rceil - 1))$, for some $t_0 \in \mathbb{N}^+$, and those of

higher values are negative. Hence, we allow the balanced partition which does not include node v_0 to have infinitely large size.

(b) There are many negative edges if we label the nodes anticlockwise (on the ring), then nodes of labels $(\bar{d}/2+1)i$ up to $(\bar{d}/2+1)i+\bar{d}/2$ have positive edges within them but negative edges outside, for each $i \in \mathbb{N}$, as in the right of Fig. 4.1. The network has size $(\bar{d}+2)j$ with $j \in \mathbb{N}^+$ sufficiently large. Hence, we allow both balanced partitions to have infinitely large sizes.

In both cases, we treat v_0 as the central node, and activate its whole neighbourhood as before; see Fig. 4.1 for examples.

In case (a), $\mathcal{A}_t^- = \emptyset,\ \forall t \le t_0$, and $\mathcal{A}_t^+ = \mathcal{A}_{t_0}^+,\ \forall t > t_0$. Hence, the overall influence on the whole network is

$$\begin{aligned} s = & \sum_{t=1}^{t_0} \left((1+\bar{d}) + (\bar{d} - 2(\lceil \theta_l \rceil - 1))t \right) (\theta_l \alpha)^t l_0 \\ & + \sum_{t=t_0+1}^{\infty} \left((1+\bar{d}) + (\bar{d} - 2(\lceil \theta_l \rceil - 1))(t_0 - (t-t_0)) \right) (\theta_l \alpha)^t l_0 \\ = & \frac{\theta_l \alpha l_0}{1-\theta_l \alpha} \left((1+\bar{d}) + (\bar{d} - 2(\lceil \theta_l \rceil - 1)) \frac{1 - 2(\theta_l \alpha)^{t_0}}{1-\theta_l \alpha} \right). \end{aligned}$$

It is very close to (and smaller than) the result in the corresponding simple regular lattice in (4.18), and is positive for most choices of parameters, $\bar{d}, \theta_l, \alpha, t_0$, although it can also be negative in some special cases. For example, when $t_0 = 0$,

$$s = \frac{\theta_l \alpha l_0}{(1-\theta_l \alpha)^2} \left(2\lceil \theta_l \rceil - 1 - \theta_l \alpha (1+\bar{d}) \right),$$

and it can be negative if α is sufficiently large (note $2\lceil \theta_l \rceil \le \bar{d}$), e.g., when the level of trust/distrust among agents is fairly high.

While in case (b), there are certain periodic patterns in every $(\bar{d}+2)$ time steps, and one can obtain the overall influence by specifying the behaviour in the first $(\bar{d}+2)$ time steps. Take $\lceil \theta_l \rceil = \bar{d}/2$ for example, and let the *net value* denote the difference from the number of nodes that start to have positive state values minus those that start to have negative ones, then $\forall i \in \mathbb{N}$,

(i) at $t = (\bar{d}+2)i + 1$, the net value is -2;

(ii) when $(\bar{d}+2)i + 1 < t \le (\bar{d}+2)i + 1 + \bar{d}/2$, the net value is 0;

(iii) at $t = (\bar{d}+2)i + 2 + \bar{d}/2$, the net value is 2;

(iv) when $(\bar{d}+2)i + 2 + \bar{d}/2 < t \leq (\bar{d}+2)(i+1)$, the net value is 0.

Recall that the nodes, once activated, remain activated and the same sign over time. Hence, the overall influence from the nodes in $\mathcal{A}_{\bar{d}+2} \backslash \mathcal{A}_0$ is

$$\begin{aligned} s^{(p)} &= -2\sum_{t=1}^{\infty}(\theta_l\alpha)^t l_0 + 2\sum_{t=\bar{d}/2+2}^{\infty}(\theta_l\alpha)^t l_0 \\ &= -2\frac{\theta_l\alpha l_0}{1-\theta_l\alpha}\left(1-(\theta_l\alpha)^{\bar{d}/2+1}\right). \end{aligned}$$

Hence, the overall influence from the nodes in $\mathcal{A}_{(\bar{d}+2)(i+1)} \backslash \mathcal{A}_{(\bar{d}+2)i}$ is $(\theta_l\alpha)^{(\bar{d}+2)i} s^{(p)}$, $\forall i \in \mathbb{N}$. Note that the net value is 1 at $t = 0$. Hence the overall influence on the whole network is

$$\begin{aligned} s &= \sum_{t=1}^{\infty}(\theta_l\alpha)^t l_0 + \sum_{i=0}^{\infty}(\theta_l\alpha)^{(\bar{d}+2)i} s^{(p)} = \frac{\theta_l\alpha l_0}{1-\theta_l\alpha} + \frac{s^{(p)}}{1-(\theta_l\alpha)^{\bar{d}+2}} \\ &= \frac{\theta_l\alpha l_0}{1-\theta_l\alpha}\left(1 - \frac{2}{1+(\theta_l\alpha)^{\bar{d}/2+1}}\right) < 0. \end{aligned}$$

Furthermore, we find that the alternating occurrence of the net values -2 and 2 in the periodic pattern is a general feature for all possible θ_l. Hence, $s^{(p)} < 0$, $\forall \theta_l$, and then the overall influence s is more likely to be negative in this case. For finite lattices, we need to specify whether the number of nodes is odd or even, and whether it is a multiple of the period $(\bar{d}+2)$, which may affect the results, although we do not expect the effect to be substantial if the network is sufficiently large.

The results in cases (a) and (b) imply that not only the balanced structure but also the actual portion of positive/negative edges have impact on the overall influence. Specifically, we observe that when there are more positive edges, the overall influence tends to be positive, while it tends to be negative when there are more negative edges. We should also notice that the observed phenomenology is due to the specific choice of model parameters, and a possible variation of the model in which the bounds depend on the proportion of negative links and positive links of nodes might display another phenomenology.

Antibalanced regular lattices. Corresponding to the analysis in balanced regular lattices, we consider the activations that are consistent with the antibalanced

partition, where we positively activate the central node v_i, and for each of its neighbour v_j, we positively activate it with $x_j(0) = l_0$ if $W_{ij} < 0$ and negatively activate it otherwise; see Fig. 4.2 for example.

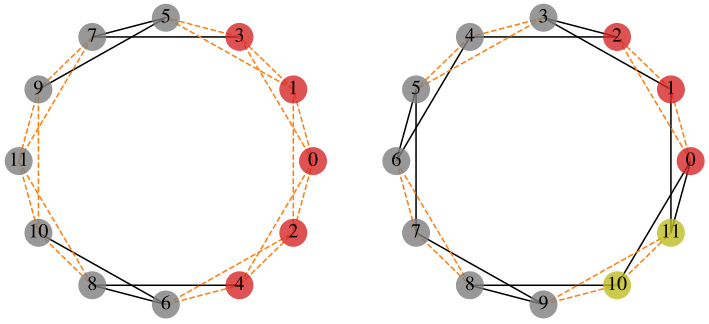


Figure 4.2: Example of signed regular lattices of degree 4 with different structurally antibalanced configurations, where positive edges are in black, negative edges are dashed in orange, and the whole neighbourhood of node v_0 are in different colour(s) from the others (in grey), with the ones that are positively activated in red and the others that are negatively activated in green.

Again, we find that when condition (4.17) is true, there is certain propagation on the structurally antibalanced regular lattice, but an alternating sign pattern, where at each $t > 0$:

(i)

$$x_j(t) = \begin{cases} -(\theta_l \alpha)^t l_0, & \text{if } v_j \in \mathcal{A}_{t-1}^+, \\ (\theta_l \alpha)^t l_0, & \text{if } v_j \in \mathcal{A}_{t-1}^-, \end{cases}$$

where $\mathcal{A}_t = \mathcal{A}_t^+ \cup \mathcal{A}_t^-$ with $\mathcal{A}_t^+ = \{v_j : x_j(t) > 0\}$ and $\mathcal{A}_t^- = \{v_j : x_j(t) < 0\}$;

(ii) $\bar{d} - 2(\lceil \theta_l \rceil - 1)$ more nodes that are closest to $\mathcal{A}_{t-1}$ will be activated if there is any, where

$$x_j(t) = \begin{cases} -(\theta_l \alpha)^t l_0, & \text{if } \exists v_{i_1} \in \mathcal{A}_{t-1}^+ \text{ or } v_{i_2} \in \mathcal{A}_{t-1}^- \text{ s.t. } A_{i_1 j} > 0 \text{ or } A_{i_2 j} < 0, \\ (\theta_l \alpha)^t l_0, & \text{if } \exists v_{i_1} \in \mathcal{A}_{t-1}^+ \text{ or } v_{i_2} \in \mathcal{A}_{t-1}^- \text{ s.t. } A_{i_1 j} < 0 \text{ or } A_{i_2 j} > 0. \end{cases}$$

Hence again, each $x_j(t)$ has the same magnitude as the state value on the corresponding simple regular lattice. However, for the sign pattern, $\mathcal{A}_{t-1}^+ \subset \mathcal{A}_t^-$ and $\mathcal{A}_{t-1}^- \subset \mathcal{A}_t^+$, i.e. the nodes, once activated, remain active but alternate the sign of their state

values in every time step, which is similar to the evolution of $\mathbf{W}^t$ in linear dynamics when the underlying signed network is antibalanced.

For the overall influence and also in response to the analysis in balanced networks, we consider the following two distinct cases of the distribution of edge sign, where we assume the network is infinitely large.

(a) There are only a few positive edges: if we label nodes symmetrically with regard to node v_0 as in the left of Fig. 4.2, only edges between nodes of label less than or equal to $\bar{d} + t_0(\bar{d} - 2(\lceil \theta_l \rceil - 1))$, for some $t_0 \in \mathbb{N}^+$, and those of higher values are positive. Hence, we allow the antibalanced partition which does not include node v_0 to have infinitely large size.

(b) There are many positive edges: if we label the nodes anticlockwise (on the ring), then nodes of labels $(\bar{d}/2+1)i$ up to $(\bar{d}/2+1)i + d/2$ have negative edges within them but positive edges outside, for each $i \in \mathbb{N}$, as in the right of Fig. 4.2. The network has size $(\bar{d}+2)j$ with $j \in \mathbb{N}^+$ sufficiently large. Hence, we allow both antibalanced partitions to have infinitely large sizes.

Again in both cases, we activate the whole neighbourhood of node v_0; see Fig. 4.2 for example.

From the evolution of state values in balanced and antibalanced regular lattices, we find that the results of an antibalanced regular lattice can be obtained by replacing α by $-\alpha$ in the results of the balanced one constructed by negating the edge sign. We can then obtain the overall influence through the same technique. Hence, in case (a), the overall influence on the whole network is

$$\begin{aligned} s &= -\frac{\theta_l \alpha l_0}{1+\theta_l \alpha} \left((1+\bar{d}) + (\bar{d} - 2(\lceil \theta_l \rceil - 1)) \frac{1 - 2(-\theta_l \alpha)^{t_0}}{1+\theta_l \alpha} \right) \\ &\leq -\frac{\theta_l \alpha l_0}{1+\theta_l \alpha} \left((1+\bar{d}) - (\bar{d} - 2(\lceil \theta_l \rceil - 1)) \frac{1}{1+\theta_l \alpha} \right) < 0. \end{aligned}$$

While in case (b), the overall influence on the whole network when $\lceil \theta_l \rceil = \bar{d}/2$ is

$$s = \frac{\theta_l \alpha l_0}{1+\theta_l \alpha} \left(-1 + \frac{2}{1 + (-\theta_l \alpha)^{\bar{d}/2+1}} \right) > 0.$$

Now, the "net value" is defined as the number of nodes that start to have state value $(-\theta_l \alpha)^t l_0$ over those that start to have state value $-(-\theta_l \alpha)^t l_0$, and its alternating occurrence of -2 and 2 in the periodic pattern is still a general feature for all possible θ_l. But to determine the sign of the overall influence here, we should also take into

account whether the occurrence is in odd or even time steps. However, since we find that the "net value" is always -2 at the odd time steps $t = (\bar{d}+2)i+1,\ \forall i \in \mathbb{N}$, causing the sum of influence on the corresponding nodes to be positive, and these nodes also have larger magnitude of the overall influence than all nodes that start to have nonzero state values in later time steps $(\bar{d}+2)i+1 < t \leq (\bar{d}+2)(i+1)$ (within a period), the overall influence s is very likely to be positive in this case.

Therefore, we observe consistent patterns in the overall influence between balanced and antibalanced regular lattices, where it will very likely be negative if there are more negative edges, while it will potentially be positive if there are more positive edges. Still, the portion of positive edges versus that of negative edges matters.

Strictly unbalanced regular lattices. In all the remaining configurations, neither are they balanced where the nodes, once activated, remain active and maintain the sign of their state values over time, nor are they antibalanced where the nodes, once activated, remain active but alternate the sign of their state values in every time step. There could be conflicts in the sign of a node's neighbours' state values multiplying the edge weights, hence it is more likely for the sum to be less than the threshold, and for these nodes to have state value 0 accordingly. Hence, the propagation in strictly unbalanced lattices generally terminates within less number of time steps.

We can still find the same condition (4.17) on θ_l to trigger certain propagation on the strictly unbalanced regular lattice, but generally not all nodes will have nonzero state values in the propagation process. For the overall influence, depending on how far it is from being balanced by (4.13) or antibalanced by (4.14) plus the distribution of negative edges as analysed before, we can estimate it by the corresponding cases in balanced or antibalanced regular lattices.

General signed networks. In this section, we have analysed the behaviour of the extended LT model on signed regular lattices, from the perspective of the balanced structure. For general signed networks, we can consider the performance of extended LT model as follows. (i) We interpolate the signed network locally by signed regular lattices of different degrees, or signed trees where complex contagions (e.g., $\theta_l > 1$ on the signed networks with uniform magnitude of the edge weight) can hardly proceed. (ii) Then we can estimate the behaviour of the extended LT model on the whole network by interpolating that on the corresponding signed regular lattices.

4.4 General class of information propagation model

In this section, we extend the GIP model proposed in Sec. 3.3, with continuous state variables while allowing the feedback between nodes, to signed networks. The full description of the model is in Sec. 4.4.1. We then show that it maintains the properties that it can be reduced to the IC model for one end, and the LT model for another, both of which have been further extended for signed networks in Secs. 4.4.2 and 4.4.3, respectively, as well as its general features via the corresponding differences from the two classic models.

4.4.1 Model description

We consider a signed social network $G(V, E, \mathbf{W})$ that is connected, weighted and undirected[4]. Each edge (v_i, v_j) can be considered as a channel connecting nodes v_i and v_j from which the information flows, and the magnitude of the associated weight W_{ij} can indicate, for example, the tendency to agree (when $W_{ij} > 0$) or disagree (when $W_{ij} < 0$) with each other. Correspondingly, we suppose that there are also two opposite directions of the influence on each node, and we refer to the one in the positive direction as *positive influence*, e.g. agreeing on an argument or buying more products, while the other in the negative direction as *negative influence*, e.g. disagreeing on the argument or buying fewer products. Hence, the influence on each node can be positive or negative. We maintain $x_i(t)$ to represent the state value of node v_j at each discrete time step $t \geq 0$, but now $x_i(t) \in \mathbb{R}$ which can be negative. $\mathbf{x}(t) = (x_i(t))$ again denotes the vector consisting of $x_i(t)$.

To extend the GIP model for signed networks, the main problems are (i) how to incorporate negative edges in the mode and (ii) when the nodes should take negative state values. For the former, we apply the "opposing rule" (cf. Sec. 4.2.2) where a negative connection will tune the state value towards the opposite direction, while for the latter, we allow a node to take negative state value if the the accumulated attempts from all its neighbours satisfies the same requirements as the positive counterpart but in the opposite direction. Hence, the general class of signed information propagation

[4]The analysis can be extended to directed networks, but corresponding to the analysis of signed networks in Sec. 4.3, we consider undirected networks here.

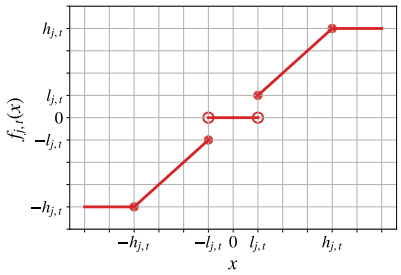


Figure 4.3: Example bound function $f_{j,t}$ of node v_j at time step t where $h_{j,t} = 4l_{j,t} > 0$.

(GSIP) model we propose here is still a bounded-linear dynamics,

$$x_j(t) = f_{j,t}(\sum_i W_{ij}x_i(t-1)), \quad \forall t > 0,\ v_j \in V, \tag{4.19}$$

$$\text{where } f_{j,t}(x) = \begin{cases} 0, & |x| < l_{j,t}, \\ x, & l_{j,t} \le |x| < h_{j,t}, \\ sign(x)h_{j,t}, & |x| \ge h_{j,t}, \end{cases}$$

is the time-dependent bound function, which is odd or rotational symmetric where $f(-x) = -f(x)$ for all possible x, of each node v_j (see Fig. 3.2 for an example), $sign(\cdot) : \mathbb{R} \to \{-1, 0, 1\}$ is the function indicating the sign, $\mathbf{W} = (W_{ij})$ with $W_{ij} \in \mathbb{R}$ is the signed (weighted) adjacency matrix of the underlying signed network, $\{l_{j,t}\}$ and $\{h_{j,t}\}$ are the time-dependent lower and upper bounds of each node v_j, respectively, with $0 \le l_{j,t} \le h_{j,t}$. The initial states $\mathbf{x}(0)$ are given, with $x_j(0) \in [-h_{j,0}, -l_{j,0}] \cup \{0\} \cup [l_{j,0}, h_{j,0}]$ and $l_{j,0} > 0$.

In the GSIP model, Eq. (4.19) determines the time evolution of influence at each time step, and a node v_j is *influenced*, or *active*, at time step t if $x_j(t) \neq 0$. In signed networks, we specify the following two cases, where a node v_j is *positively influenced* at t if $x_j(t) > 0$ while *negatively influenced* at t if $x_j(t) < 0$. We represent the *overall influence* on each node v_j as

$$s_j = \sum_{t=1}^{\infty}(1-\gamma)^t x_j(t), \tag{4.20}$$

where $\gamma \in [0, 1)$ is a time-discounting factor which guarantees convergence.

In order to interpret the GSIP model (4.19) and the underlying process more intuitively, we construct a small undirected signed social network that is structurally balanced, with seven agents and a uniform weight $\alpha = 0.4$; see Fig. 4.4. For illustrative

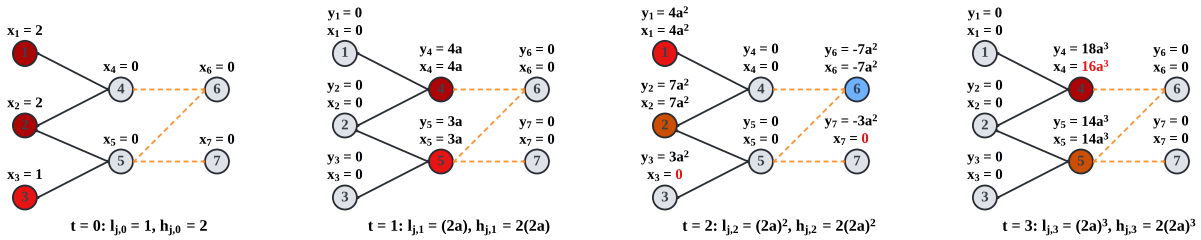


Figure 4.4: Illustration of the information propagation in the first few steps following the GSIP model on a signed network, where positive edges are solid in black and negative edges are dashed in orange, $y_j(t) = \sum_i W_{ij} x_i(t-1)$ is the linear product, $x_j(t) = f_{j,t}(y_j(t))$ is the state value, the network has weights of uniform magnitude α, and the bounds are set to be $l_{j,t} = (2\alpha)^t, h_{j,t} = 2(2\alpha)^t,\ \forall v_j \in V$, with the node colour indicating the level of influence at each step.

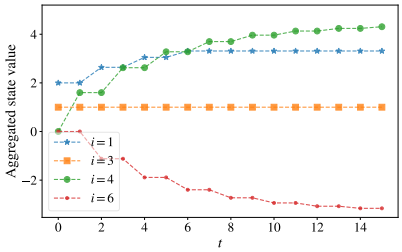


Figure 4.5: The change of the sum of influence on selected nodes along time t, from the network in Fig. 4.4 with uniform weight $\alpha = 0.4$.

purposes, we still apply the bounds $l_{j,t} = 0.8^t$ and $h_{j,t} = 2 \times 0.8^t$, $\forall t \geq 0,\ v_j \in V$, set $\gamma = 0$, and activate nodes v_1, v_2 with value 2 and node v_3 with value 1 at $t = 0$, the same as Sec. 3.4.1. Then at $t = 2$, v_4 can still independently influence v_1, consistent with the IC model, while collective effort is needed to influence v_7, consistent with the LT model; see Fig. 4.4. The coexistence of the mechanisms underlying both classic models is an outstanding feature of the GIP model which we developed for simple networks, and the current extension for signed networks is able to maintain it. Further at $t = 2$, nodes v_4 and v_5 can collectively influence node v_6 but negatively since they are both negatively connected with v_6. This indicates the collective effort required by the LT model but in the opposite direction. Moreover, there is still reinforcement of states among nodes v_2, v_4, v_5, v_6, since the state values of v_2, v_4, v_5 become more and more positive over time while that of v_6 becomes more and more negative; see Fig. 4.5. The corresponds to the phenomena where groups of close friends keep receiving positive feedback from each other, while their common enemies keep receiving negative feedback thus going in the opposite direction.

4.4.2 Limiting case I: the linear dynamics

When the upper bounds are effectively infinity and the lower bounds are effectively 0, the GSIP model reduces to the linear dynamics of unbounded state variables, as analysed in Sec. 4.3.2.1. Hence, condition (4.6) on the time discounting factor γ and the spectral radius $\rho(\mathbf{W})$ is required to guarantee the convergence of the overall influence.

The linear dynamics can be considered as an extension to the classic IC model for continuous state variables and deterministic processes in signed networks. In the classic *probabilistic* IC model, each edge weight $W_{ij} > 0$ indicates the probability that node v_i can influence node v_j in $Bernoulli(W_{ij})$, thus the expected amount of contribution to the state value of v_j from v_i in one time step. Then for signed networks, we assume that a negative edge weight $W_{ij} < 0$ corresponds to the probability that node v_i can negatively influence node v_j in $-Bernoulli(-W_{ij})$, thus still the expected amount of contribution to the state value of v_j from v_i in one time step. Then for the expected or *deterministic* case, if we assume that (i) the expected value is the actualinfluence on each node at each time step, and (ii) the state values have the *no-memory* property where the ability either to be influenced or influence others at the current time step t is independent of its previous states, then $x_j(t) = \sum_i W_{ij} x_i(t-1)$, $\forall v_j \in V$, i.e. it has the same updating function as the linear dynamics in Eq. (4.5). We refer to this model as the extended signed independent cascade (SIC) model hereafter.

Relation between the models. We now give the detailed correspondence between the GSIP model and the linear dynamics by specifying "effectively infinity" for the upper bounds and "effectively 0" for the lower bounds.

Theorem 15. *If* $l_{j,t} = 0 \leq \mathbf{h}_0 |\mathbf{W}^t|_{:,j} \leq h_{j,t}$, $\forall t > 0, v_j \in V$, *where* $\mathbf{h}_0 = (h_{j,0})$, *and we let* $|\mathbf{W}^t|_{:,j}$ *denote the* j*-th column of the absolute values in* $\mathbf{W}^t$, *the GSIP model is equivalent to the linear dynamics* (4.5).

Proof. We show that given such bounds in the GSIP model,

$$\mathbf{x}(t)^T = \mathbf{x}(0)^T \mathbf{W}^t, \quad \forall t > 0, \tag{4.21}$$

by induction.

When $l_{j,t} = 0,\ \forall t > 0, v_j \in V$, there is no threshold effect from the current lower bounds, i.e. $\forall t > 0,\ v_j \in V$ s.t. $|\sum_i W_{ij} x_i(t-1)| > 0$,

$$\left| \sum_i W_{ij} x_i(t-1) \right| \geq l_{j,t}.$$

We now check the saturation effect from the upper bounds. (i) When $t = 1$, for each node v_j,

$$\left| \mathbf{x}(0)^T \mathbf{W}_{:,j} \right| \leq \sum_i |W_{ij}||x_i(0)| \leq \sum_i |W_{ij}| h_{i,0} \leq h_{j,1},$$

since $|x_i(0)| \leq h_{i,0},\ \forall i$. Hence,

$$x_j(1) = f_{j,1}(\mathbf{x}(0)^T \mathbf{W}_{:,j}) = \mathbf{x}(0)^T \mathbf{W}_{:,j},$$

and $\mathbf{x}(1)^T = \mathbf{x}(0)^T \mathbf{W}$.

(ii) Suppose $\mathbf{x}(t)^T = \mathbf{x}(0)^T \mathbf{W}^t,\ \forall t \leq t'$. Then, for each node v_j

$$\left| \mathbf{x}(t')^T \mathbf{W}_{:,j} \right| = \left| \mathbf{x}(0)^T \mathbf{W}_{:,j}^{t'+1} \right| \leq \sum_i \left| W_{ij}^{t'+1} \right| |x_i(0)| \leq \sum_i \left| W_{ij}^{t'+1} \right| h_{i,0} \leq h_{j,t'+1},$$

where the equality is by the induction hypothesis, and the second inequality is again by $|x_i(0)| \leq h_{i,0},\ \forall i$. Hence,

$$x_j(t'+1) = f_{j,t'+1}(\mathbf{x}(0)^T \mathbf{W}_{:,j}^{t'+1}) = \mathbf{x}(0)^T \mathbf{W}_{:,j}^{t'+1},$$

and $\mathbf{x}(t'+1)^T = \mathbf{x}(0)^T \mathbf{W}^{t'+1}$. As in Sec. 4.3.2.1, (4.21) characterises the linear dynamics. □

We notice that the requirement on the lower bounds is much stricter in signed networks (cf. Theorem 2 in simple networks), since the existence of negative edges could cause the state values to be arbitrarily small at some time steps. Meanwhile, we note that when the network has weights of uniform magnitude α and the initial state values are also uniform with $l_{j,0} = h_{j,0} = l_0,\ \forall v_j \in V$, then the requirement of the lower bounds can be relaxed to $l_{j,t} \leq l_0 \alpha^t,\ \forall t > 0,\ v_j \in V$.

Differences between the models. We then consider how the general properties of the GSIP model differ from those of the linear dynamics. Specifically, we analyse the imposed *signed threshold effect* when increasing the lower bounds, where nodes

with more overlap in their *positive neighbourhoods* of various distances will be able to achieve a higher influence on all nodes, through the signed stochastic block model (SSBM) which is slightly modified from the version in [45]. The SSBM considered here is constructed by the following components: (i) a planted SBM, $SBM(p_{in}, p_{out})$, where the probabilities of an edge to occur inside each community and between the two communities being p_{in} and p_{out}, respectively; (ii) an initial balanced configuration, where edges inside each community being positive while those between the two communities are negative; (iii) a probability $\eta \in [0, 1]$[5] to flip the edge sign. We denote this planted SSBM by $SSBM(p_{in}, p_{out}, \eta)$ and also allow self-loops that characterise the likelihood to maintain ones' states.

Claim 3. *With $l_{j,0} = h_{j,0} = l_0,\ \forall v_j \in V$, and $\{h_{j,t}\}$ as in Theorem 2 in the GSIP model, $SSBM(p_{in}, p_{out}, \eta)$ with two equally sized communities, $\mathcal{B}_1, \mathcal{B}_2$, and weights of uniform magnitude α, while allowing self-loops, has the following properties at $t = 1$:*

1. *when $l_{j,1} \leq l_1^* = l_0\alpha,\ \forall v_j \in V$ (the linear-dynamics case), the expected influence $\mathbb{E}[\sum_j (1-\gamma)x_j(1)]$ from the initially positively activated node set (i) $\mathcal{A}_0 = \{v_{i_1}, v_{i_2}\} \subset \mathcal{B}_1$ is the same as that from (ii) $\mathcal{A}_0 = \{v_{j_1}, v_{j_2}\}$ with $v_{j_1} \in \mathcal{B}_1,\ v_{j_2} \in \mathcal{B}_2$, and is positive if*

$$(p_{in} - p_{out})(1 - 2\eta) > 0; \tag{4.22}$$

2. *when $l_1^* < l_{j,1} \leq 2l_1^*,\ \forall v_j \in V$[6], the expected influence $\mathbb{E}[\sum_j (1-\gamma)x_j(1)]$ from set (ii) is 0, and that from set (i) is positive under the same condition (4.22).*

Proof. For the SSBM, $\mathbf{W} = \alpha\mathbf{A}$, where $\mathbf{A}$ is the signed (unweighted) adjacency matrix, and for each pair of nodes v_i, v_j, $A_{ij} \sim \text{Cat}(p_{ij}r_{ij}, p_{ij}(1 - r_{ij}), 1 - p_{ij})$, where Cat denotes the categorical distribution, the three probabilities in the distribution correspond to values $1, -1, 0$, and

$$p_{ij} = p_{in}\delta(\sigma_i, \sigma_j) + p_{out}(1 - \delta(\sigma_i, \sigma_j)),$$
$$r_{ij} = (1 - \eta)\delta(\sigma_i, \sigma_j) + \eta(1 - \delta(\sigma_i, \sigma_j)),$$

[5]The original range is $[0, 1/2)$ in [45] to obtain perturbations from (at least) weakly balanced networks. Here we extend it to be the whole range because networks close to being antibalanced (in the two-block case) could also be interesting.

[6]In the specific case here, the upper bound $2l_1^*$ is equivalent to requiring that at most two initially activated neighbours are needed to influence a node at $t = 1$.

with $\sigma_i \in \{1, 2\}$ indicating the block membership of each node v_i, and $\delta(i, j)$ being the delta function where $\delta(i, j) = 1$ if and only if $i = j$ and 0 otherwise. Further, we denote the linear part of the state vector by $\mathbf{y}(t+1) = \mathbf{W}^T\mathbf{x}(t)$, then for each node v_j at each time step $t > 0$,

$$\begin{cases} x_j(t) &= f_{j,t}(y_j(t)), \\ y_j(t) &= \sum_i W_{ij}x_i(t-1) = \alpha \sum_i A_{ij}x_i(t-1). \end{cases}$$

Then at $t = 1$[7], $y_j(1)$ is then

$$\begin{aligned} &\alpha \sum_i A_{ij}x_i(0) = \alpha l_0 \sum_{v_i \in \mathcal{A}_0} A_{ij} \\ &= \alpha l_0 \left(\sum_{v_i \in \mathcal{A}_0 \cap \mathcal{B}_{\sigma_j}} \xi_{in} + \sum_{v_i \in \mathcal{A}_0 \setminus \mathcal{B}_{\sigma_j}} \xi_{out} \right) \\ &= \alpha l_0 \left(\zeta_{k_{\sigma_j}, p_{in}(1-\eta), p_{in}\eta, 1-p_{in}} + \zeta_{k-k_{\sigma_j}, p_{out}\eta, p_{out}(1-\eta), 1-p_{out}} \right), \end{aligned}$$

where $\mathcal{A}_0 = \{v_i : x_i(0) \neq 0\}$ is the (given) set of initially (here positively) activated nodes, $k = |\mathcal{A}_0|$, $k_i = |\mathcal{A}_0 \cap \mathcal{B}_i|$, $i = 1, 2$, $\xi_{in} \sim \text{Cat}(p_{in}(1-\eta), p_{in}\eta, 1-p_{in})$, $\xi_{out} \sim \text{Cat}(p_{out}\eta, p_{out}(1-\eta), 1-p_{out})$, $\zeta_{n_m, p_1, p_2, p_3} \sim \text{Multin}(n_m, p_1, p_2, p_3)$, and $\text{Multin}(n_m, p_1, p_2, p_3)$ denotes the multinomial distribution with the three probabilities corresponding to values $1, -1, 0$.

Hence, for set (i),

$$y_j(1) = \alpha l_0 \left(\zeta_{2, p_{in}(1-\eta), p_{in}\eta, 1-p_{in}} \delta(\sigma_j, 1) + \zeta_{2, p_{out}\eta, p_{out}(1-\eta), 1-p_{out}} \delta(\sigma_j, 2) \right),$$

while for set (ii),

$$y_j(1) = \alpha l_0 \left(\zeta_{1, p_{in}(1-\eta), p_{in}\eta, 1-p_{in}} + \zeta_{1, p_{out}\eta, p_{out}(1-\eta), 1-p_{out}} \right).$$

When $l_{j,1} \leq l_1^* = l_0\alpha$, $x_j(1) = y_j(1)$, $\forall v_j \in V$. Then for set (i),

$$\begin{aligned} \mathbb{E}\left[\sum_j (1-\gamma)x_j(1) \right] &= (1-\gamma)n_b \times \alpha l_0 \times (2p_{in}(1-\eta) - 2p_{in}\eta + 2p_{out}\eta - 2p_{out}(1-\eta)) \\ &= 2(1-\gamma)n_b \alpha l_0 (p_{in} - p_{out})(1-2\eta), \end{aligned}$$

[7]Note that in the classical SBM, the expressions of initially activated nodes can be different from others in the same community, due to the common assumption of no self-edges. However, we allow self-loops in our analysis, thus also include $v_j \in \mathcal{A}_0$ in the sum for $y_j(1)$.

where n_b is the size of each community, and for set (ii),

$$\mathbb{E}\left[\sum_j (1-\gamma)x_j(1)\right] = 2(1-\gamma)n_b \times \alpha l_0 \times (p_{in}(1-\eta) - p_{in}\eta + p_{out}\eta - p_{out}(1-\eta))$$
$$= 2(1-\gamma)n_b\alpha l_0(p_{in} - p_{out})(1-2\eta).$$

Hence, the expected influence values at $t = 1$ are the same in the two cases, and is positive if condition (4.22) holds true.

However, when $\alpha l_0 = l_1^* < l_{j,1} \leq 2l_1^* = 2\alpha l_0$, for set (ii),

$$P(y_j(1) = 2\alpha l_0) = p_{in}(1-\eta)p_{out}\eta = P(y_j(1) = -2\alpha l_0).$$

Hence,

$$\mathbb{E}[x_j(1)] = 2\alpha l_0 P(y_j(1) = 2\alpha l_0) - 2\alpha l_0 P(y_j(1) = -2\alpha l_0) = 0,$$

and the overall influence at $t = 1$ is 0. While for set (i),

$$P(y_j(1) = 2\alpha l_0) = p_{in}^2(1-\eta)^2\delta(\sigma_j, 1) + p_{out}^2\eta^2\delta(\sigma_j, 2),$$

and

$$P(y_j(1) = -2\alpha l_0) = p_{in}^2\eta^2\delta(\sigma_j, 1) + p_{out}^2(1-\eta)^2\delta(\sigma_j, 2).$$

Hence,

$$\mathbb{E}\left[\sum_j (1-\gamma)x_j(1)\right] = (1-\gamma)n_b \times 2\alpha l_0 \left(p_{in}^2(1-\eta)^2 + p_{out}^2\eta^2 - p_{in}^2\eta^2 - p_{out}^2(1-\eta)^2\right)$$
$$= 2(1-\gamma)n_b\alpha l_0(p_{in}^2 - p_{out}^2)(1-2\eta),$$

which is positive if condition (4.22) holds true. □

If condition (4.22) holds true, then either p_{in} is large while η is small, or p_{out} is large while η is large. Both of the two cases contribute to the fact that (a) node sets in the same community have more shared positive neighbours (of various distances) than the ones distributed in the two, and (b) there are more positive edges in the network (cf. Sec. 4.3.2). Further from Claim 3, we find that even though the nodes are positively activated initially, the overall influence could be negative if there are

more negative edges in the network, and particularly for set (ii) in case 2, the overall influence could also be zero even when the probability to activate a node is nonzero. Both of the features cannot exist in simple networks.

4.4.3 Limiting case II: the linear threshold model

With all upper bounds and the corresponding lower bounds equal to some thresholds that could change for different nodes or time steps, all the bound functions in the GSIP model only have *threshold* effect on the *linear* product, as explored in Sec. 4.3.2.2.

The GSIP model in this case can be treated as an extension to the classic LT model for continuous state variables and deterministic parameters in signed networks. Firstly, we maintain the linear activation strategy as in (3.1), but both when the weighted sum of active neighbours' state values is sufficiently positive (to give value 1) and sufficiently negative (to give value -1). Secondly, with continuous variables, we set the activated state value to be the threshold value, thus control the source of nonlinearity to be only the activation. Note that the state values can change magnitude over time, thus we allow time-dependent thresholds $\{\theta_{j,t}\}$. Thirdly, we also assume the no-memory property as in Sec. 4.4.2. These jointly give the updating function (4.15), and we refer to this extension as the extended signed linear threshold (SLT) model hereafter.

From another perspective in extending the classic LT model, we can maintain the magnitude of the state value at each time step, but instead of a single value of 1 or -1 for being positively and negatively influenced, respectively, each node v_j can take values in a range $[-m_j, -1 \cup [1, m_j]$ depending on how strong the positive or negative influence attempts from its neighbours are, i.e. $\sum_i W_{ij} x_i(t-1)$. Specifically, as in (4.19), a node v_j starts to take nonzero state value if the magnitude of the sum is at least a threshold, denoted l'_j here, but further, the magnitude of the state value increases from 1 to the highest level m_j, as the magnitude of the sum increases from l'_j to a higher value h'_j. This gives a more direct extension to the LT model (on signed networks), and we name it the multi-valued signed linear threshold (MSLT) model.

Explicitly, it has the updating function,

$$x_j(t) = f_j(\sum_i W_{ij}x_i(t-1)), \quad \forall t > 0,\ v_j \in V, \tag{4.23}$$
$$\text{where, } f_j(x) = \begin{cases} 0, & |x| < l'_j, \\ \frac{m_j - 1}{h'_j - l'_j}(x - sign(x)l'_j) + sign(x), & l'_j \le |x| < h'_j, \\ sign(x)m_j, & |x| \ge h'_j, \end{cases}$$

is the time-independent bound function that is odd or rotational symmetric, $sign(\cdot) : \mathbb{R} \to \{-1, 0, 1\}$ is the function indicating the sign, and $x_j(0) \in [-h'_{j,0}, -1] \cup \{0\} \cup [1, h'_{j,0}]$ with $h'_{j,0}$ being the highest magnitude that node v_j's initial state value can take. We will show later that the two extensions to the LT model are equivalent through their equivalences with the GSIP model.

Relation between the models. To further disentangle the relationship between the GSIP model and the threshold models, we consider the following *modified threshold-type* bounds, where for each node v_j at each time step $t > 0$,

$$\begin{aligned} l_{j,t} &= (\theta_{l,j}\alpha)^t l_{j,0}, \\ h_{j,t} &= \theta_{hh,j}\theta_{hl,j}^{t-1}\alpha^t h_{j,0}, \end{aligned} \tag{4.24}$$

$\alpha = \sum_{(v_i,v_j)\in E} |W_{ij}|/|E|$ is the mean magnitude of weights[8], $\theta_{l,j}$ is the threshold for the lower bounds, and $\theta_{hh,j}, \theta_{hl,j}$ are the higher and lower thresholds for the upper bounds, respectively, with $0 \le \theta_{l,j} \le \theta_{hl,j} \le \theta_{hh,j}$. It is called the modified version to separate the one we proposed on simple networks (3.13), where the difference is that the common ratio in the upper bounds (i.e. $\theta_{hl,j}\alpha$) can be different from that in the lower bounds (i.e. $\theta_{l,j}\alpha$). The modification is necessary in general weighted signed networks, since for the GSIP model to reach the linear-dynamics extreme, the latter is 0 while sufficiently large upper bounds are required in Theorem 15. With such bounds, the GSIP model reduces to the extended SLT model if $l_{j,0} = h_{j,0}$ and $\theta_{l,j} = \theta_{hl,j} = \theta_{hh,j},\ \forall v_j \in V$. Furthermore, we will also show that the GSIP model can also be equivalent to the MSLT model in Theorem 16.

[8]With the threshold-type bounds, the condition $|\sum_i W_{ij}x_i(0)| \ge l_{j,1}$ at $t = 1$ is equivalent to $|\sum_i (W_{ij}/\alpha)(x_i(0)/l_{j,0})| \ge \theta_{l,j}$. Hence, α will not affect the activation so long as the relative weight W_{ij}/α does not change (e.g., $W_{ij}/\alpha = 1$ if the network has uniform edge weight).

Theorem 16. *If the modified threshold-type bounds have uniform thresholds s.t.*

$$\theta_{l,j} = \theta_l, \theta_{hl,j} = \theta_{hl}, \theta_{hh,j} = \theta_{hh}, \quad \forall v_j \in V \tag{4.25}$$

the GSIP model with such bounds plus $\theta_{hl} = \theta_l$ *and* $l_{j,0} = 1,\ \forall v_j \in V$*, is equivalent to the MSLT model with* $l'_j = \theta_l\alpha$*,* $h'_j = \theta_{hh}\alpha h_{j,0}$*,* $m_j = (\theta_{hh}h_{j,0})/\theta_l$*, and* $h'_{j,0} = h_{j,0},\ \forall v_j \in V$*, in terms of the overall influence where if we denote the time-discounting factors for the GSIP model and the MSLT model as* γ, γ'*, respectively, we set* $\gamma' = 1-(1-\gamma)\theta_l\alpha$*. Specifically, if we denote the state values from the GSIP model as* $x_j(t)$ *and those from the MSLT model as* $x'_j(t)$*,*

$$x_j(t) = (\theta_l\alpha)^t x'_j(t), \quad \forall t \geq 0,\ v_j \in V. \tag{4.26}$$

Proof. We first note that if (4.26) is true, then for each node $v_j \in V$,

$$\sum_{t=0}^{\infty}(1-\gamma)^t x_j(t) = \sum_{t=0}^{\infty}(1-\gamma)^t(\theta_l\alpha)^t x'_j(t) = \sum_{t=0}^{\infty}(1-\gamma')^t x'_j(t),$$

thus the network has the same overall influence from the two models.

We then show that (4.26) is true by induction on the time step t. (i) At $t = 0$, $x_j(0) = (\theta_l\alpha)^0 x'_j(0)$, since $x_j(0) = x'_j(0),\ \forall v_j \in V$. (ii) Suppose $x_j(t) = (\theta_l\alpha)^t x'_j(t),\ \forall v_j \in V$, is true for all $t \leq t'$, then for each node v_j at $t = t' + 1$, if we denote $y_j(t'+1) = \sum_i W_{ij}x_i(t')$,

$$x_j(t'+1) = \begin{cases} 0, & |y_j(t'+1)| < l_{j,t'+1}, \\ y_j(t'+1), & l_{j,t'+1} \leq |y_j(t'+1)| < h_{j,t'+1}, \\ sign(y_j(t'+1))h_{j,t'+1}, & |y_j(t'+1)| \geq h_{j,t'+1}, \end{cases} \tag{4.27}$$

where $l_{j,t'+1} = (\theta_l\alpha)^{t'+1}$ and $h_{j,t'+1} = \theta_{hh}\theta_l^{t'}\alpha^{t'+1}h_{j,0}$, by the GSIP model. We now consider the state value $x'_j(t'+1)$ from the MSLT model in the three different cases, and compare it with the state value $x_j(t'+1)$ in (4.27). (1) When

$$\left|\sum_i W_{ij}x_j(t')\right| < l_{j,t'+1} = (\theta_l\alpha)^{t'+1} \Leftrightarrow \left|\sum_i W_{ij}(\theta_l\alpha)^{t'}x'_i(t')\right| < (\theta_l\alpha)^{t'+1}$$
$$\Leftrightarrow \left|\sum_i W_{ij}x'_j(t')\right| < \theta_l\alpha = l'_j,$$

we have $(\theta_l\alpha)^{t'+1}x'_j(t'+1) = 0 = x_j(t'+1)$.
(2) When

$$\left|\sum_i W_{ij}x_j(t')\right| \geq h_{j,t'+1} = \theta_{hh}\theta_l^{t'}\alpha^{t'+1}h_{j,0} \Leftrightarrow \left|\sum_i W_{ij}(\theta_l\alpha)^{t'}x'_i(t')\right| \geq \theta_{hh}\theta_l^{t'}\alpha^{t'+1}h_{j,0}$$
$$\Leftrightarrow \left|\sum_i W_{ij}x'_j(t')\right| \geq \theta_{hh}\alpha h_{j,0} = h'_j,$$

we then have

$$\begin{aligned}(\theta_l\alpha)^{t'+1}x'_j(t'+1) &= (\theta_l\alpha)^{t'+1}sign(\sum_i W_{ij}x'_i(t'))m_j \\ &= sign(\sum_i W_{ij}x_i(t')/(\theta_l\alpha)^t)(\theta_l\alpha)^{t'+1}(\theta_{hh}h_{j,0})/\theta_l \\ &= sign(\sum_i W_{ij}x_i(t'))h_{j,t'+1} = x_j(t'+1).\end{aligned}$$

(3) Finally, in the remaining case when

$$l_{j,t'+1} \leq \left|\sum_i W_{ij}x_j(t')\right| < h_{j,t'+1} \Leftrightarrow\ l'_j \leq \left|\sum_i W_{ij}x'_j(t')\right| < h'_j,$$

the state value is in the linear regime where

$$\begin{aligned}x'_j(t'+1) &= \frac{m_j-1}{h'_j-l'_j}(\sum_i W_{ij}x'_j(t') - sign(\sum_i W_{ij}x'_i(t'))l'_j) + sign(\sum_i W_{ij}x'_i(t')) \\ &= \frac{(\theta_{hh}h_{j,0})/\theta_l - 1}{\theta_{hh}\alpha h_{j,0} - \theta_l\alpha}(\sum_i W_{ij}x'_j(t') - sign(\sum_i W_{ij}x'_i(t'))\theta_l\alpha) + sign(\sum_i W_{ij}x'_i(t')) \\ &= \frac{1}{\theta_l\alpha}\sum_i W_{ij}x'_j(t') = \frac{1}{\theta_l\alpha}\sum_i W_{ij}\frac{1}{(\theta_l\alpha)^{t'}}x_j(t') \\ &= \frac{1}{(\theta_l\alpha)^{t'+1}}\sum_i W_{ij}x_j(t') = \frac{1}{(\theta_l\alpha)^{t'+1}}x_j(t'+1).\end{aligned}$$

Hence, we have shown that $x_j(t'+1) = (\theta_l\alpha)^{t'+1}x'_j(t'+1), \forall v_j \in V$. □

Further, if the network has weights of uniform magnitude α, then setting $h'_j = l'_j = \theta_l\alpha$ in the MSLT model is equivalent to requiring that for each node v_j, it can be positively (negatively) influenced at $t > 0$ if there are at least θ_l more neighbours of positive (negative) linear products between their state values and the edge weights in the previous step, i.e. $W_{ij}x_i(t-1)$ than those of negative (positive) ones. This is

reminiscent of the constant threshold model [12], but extended to signed networks. The equivalence can pass to the GSIP model and the extended SLT model. Hence, $\theta_l = \theta_{hl} = \theta_{hh} = 1$ corresponds to simple contagions, and $\theta_l = \theta_{hl} = \theta_{hh} > 1$ corresponds to complex contagions, but both have been extended to signed networks where the difference between the positive and negative influence attempts is considered, rather than the number of active neighbours as in simple networks. For general weighted networks, it is also necessary to take into account the actual edge weights surrounding a node. Due to the desired correspondence demonstrated here, we will consider exclusively the modified threshold-type bounds hereafter.

Differences between the models. We then how the general properties of the GSIP model deviate from those of the extended SLT model. Specifically, we consider the locally *signed linear effect* when increasing the upper bounds through θ_{hh} in (4.25) for the modified threshold-type bounds (4.24), where *one more positively/negatively* active node (than those that are negatively/positively active) can influence its neighbours, which cannot happen when $\theta_{hh} = \theta_{hl} = \theta_l > 1$.

Claim 4. *With $l_{j,0} = h_{j,0} = l_0,\ \forall v_j \in V$, and the modified threshold-type bounds (4.24) satisfying (4.25) applied in the GSIP model, suppose at a particular time $t' \geq 0$, there is a currently inactive node v_{j_*} s.t.*

$$\sum_i A_{ij_*} sign(x_i(t')) = \pm 1,$$

where $\mathbf{A}$ is the signed (unweighted) adjacency matrix and the set $\mathcal{A}_{t'} = \{v_i : x_i(t') \neq 0\}$ contains the active nodes at t'. Then if the network has weights of uniform magnitude α,

1. *when $\theta_{hh} = \theta_{hl} = \theta_l > 1$, node v_{j_*} can never have nonzero state value at $t = t' + 1$;*
2. *when $\theta_{hh} > \theta_{hl} = \theta_l > 1$, node v_{j_*} can have nonzero state value at $t = t' + 1$ given a sufficiently large θ_{hh}.*

Proof. Since the underlying network has weight of uniform magnitude α, $\mathbf{W} = \alpha\mathbf{A}$. (1) When $\theta_{hh} = \theta_{hl} = \theta_l > 1$, $x_i(t') = sign(x_i(t'))(\theta_l\alpha)^{t'} l_0,\ \forall v_i \in \mathcal{A}_{t'}$. Then we have

$$\left| \sum_i W_{ij_*} x_i(t') \right| = \alpha(\theta_l\alpha)^{t'} l_0 < \theta_l\alpha(\theta_l\alpha)^{t'} l_0 = l_{j_*,t'+1},$$

thus node v_{j_*} cannot have positive state value at $t'+1$.

(2) When $\theta_{hh} > \theta_{hl} = \theta_l > 1$, then nodes can have different state values at each time step, and the highest possible magnitude the state of one node can have at time t' is $\theta_{hh}\theta_l^{t'-1}\alpha^{t'}l_0$. Suppose there is only one node in $\mathcal{A}_{t'}$ that is node v_j's neighbour, which we denote by v_{j_0}, then node v_{j_*} can have positive state value at $t'+1$ if

$$\begin{aligned}\left|W_{j_0 j_*}\theta_{hh}\theta_l^{t'-1}\alpha^{t'}l_0\right| \geq l_{j_*,t'+1} &\Leftrightarrow \alpha\theta_{hh}\theta_l^{t'-1}\alpha^{t'}l_0 \geq (\theta_l\alpha)^{t'+1}l_0 \\ &\Leftrightarrow \theta_{hh} \geq \theta_l^2,\end{aligned}$$

which can be achieved given that θ_{hh} is sufficiently large. Note that in other cases when there are more nodes in $\mathcal{A}_{t'}$ that are also v_j's neighbours, the weighted sum, $\sum_i W_{ij}x_i(t')$, could have higher magnitude. □

We can take the two-block planted SSBM for example again as in Sec. 4.4.2, and consider the case when increasing θ_{hh} from 2 to 4 while maintaining $\theta_l = \theta_{hl} = 2$. Accordingly, we analyse the performance of the following two sets of four initially positively activated nodes: (i) $\mathcal{A}_0 = \{v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}\} \subset \mathcal{B}_1$; (ii) $\mathcal{A}_0 = \{v_{j_1}, v_{j_2}, v_{j_3}, v_{j_4}\}$ with $v_{j_1}, v_{j_2} \in \mathcal{B}_1$ and $v_{j_3}, v_{j_4} \in \mathcal{B}_2$. We numerically verify that the increase in the expected influence (in one time step) from (i) is more than that from (ii) if condition (4.22) holds true, and the decrease is more otherwise in Sec. 4.6.1. This is consistent with Claim 3 because nodes in $\mathcal{B}_1$ have a higher probability to reach a higher or the highest magnitude of the state value in the propagation from (i), thus they have a higher chance to influence their neighbours, either positively or negatively. Furthermore, in signed networks, we can also see that it is the difference of the influence attempt in one direction (e.g., positive) over the other (e.g., negative) that matters, hence not only connections are important, but also the sign of the connections and the sign of the state values.

4.5 Influence maximisation

Now, we proceed to the key algorithmic problem of influence maximisation (IM), where the objective is to maximise the overall influence on all nodes at the end of the propagation process. Here, we maintain the constraint corresponding to the case of limited resources to activate a node, thus the number of nodes people can influence at the beginning of the process is determined by the budget size. In this section, we first introduce the formulations of the IM problem, corresponding to different cases,

and also discuss the complexity of the problem in Sec. 4.5.1. Then we analyse the special cases when the dynamics is governed by the GSIP model (4.19) in Sec. 4.5.2, and then extend the customised direct search method for the IM problem on signed networks in Sec. 4.5.3.

4.5.1 Problem formulation

With a given information propagation process, and a given function $s_j(\cdot)$ for the overall influence on each node v_j, the overall influence on the whole network is naturally,

$$s(\mathbf{x}(0)) = \sum_j s_j(\mathbf{x}(0)), \tag{4.28}$$

where $\mathbf{x}(0)$ is the initial state vector. For example, the function $s_j(\cdot)$ can count the number of nodes of positive state values at the end of the propagation process, and then the IM problem recovers one of the most common formulation in the literature, associated with the classic models extended for signed networks. While, we can also consider Eq.(4.20) associated with the GSIP model we proposed in Sec. 4.4 for $s_j(\cdot)$. The IM problem is then to maximise $s(\mathbf{x}(0))$ with respect to the $\mathbf{x}(0)$, subject to the constraint of limited budget size,

$$|\{v_j : x_j(0) \neq 0\}| \leq k, \tag{4.29}$$

where $k \in \mathbb{Z}^+$ is the budget size.

With objective (4.28) and constraint (4.29), we then formulate the IM problem as a *mixed-integer (nonlinear) programming* (MINLP),

$$\begin{aligned} \max_{\mathbf{x},\mathbf{y},\mathbf{z}} \quad & s(\mathbf{x}) \\ s.t. \quad & x_j \leq h_{j,0} z_j, \\ & x_j \geq -h_{j,0} z_j, \\ & x_j + C(1 - y_j) \geq l_{j,0} z_j \\ & -x_j + C y_j \geq l_{j,0} z_j \\ & \sum_j z_j \leq k, \\ & x_j \in \mathbb{R},\ y_j, z_j \in \{0, 1\},\ \forall j, \end{aligned} \tag{4.30}$$

where C is a sufficiently large constant, $0 < l_{j,0} \leq h_{j,0}$ restrict the initial magnitude

of influence of each node v_j, $k \in \mathbb{Z}^+$ is the budget size, and the objective function $s(\cdot)$ is the overall influence on the whole network as in Eq. (4.28). The variables in vector $\mathbf{x}$ correspond to the initial states, while the extra binary variables in vector $\mathbf{y}, \mathbf{z}$, of the same dimension, are added to appropriately impose the constraint (4.29), where each variable z_j corresponds to whether or not to activate node v_j and each variable y_j corresponds to whether or not to positively activate node v_j.

From Sec. 3.4.1, the IM problem in simple networks where the only activation option for each node is to positively activate it, is already difficult. Hence, we further assume that the optimal solution restricted to the case when only positive initial activation is considered is sufficiently good, and then consider a reduced version of the MINLP (4.30), where

$$\begin{aligned} \max_{\mathbf{x},\mathbf{z}} \quad & s(\mathbf{x}) \\ s.t. \quad & x_j \leq h_{j,0} z_j, \\ & x_j \geq l_{j,0} z_j \\ & \sum_j z_j \leq k, \\ & x_j \in \mathbb{R},\ z_j \in \{0,1\},\ \forall j, \end{aligned} \tag{4.31}$$

where we remove the binary vector $\mathbf{y}$, since here once a node is activated initially, it will have positive state value, where $x_j > 0$.

The difficulty of the current optimisation problem (4.31) (and also (4.30)) still lies in the objective function $s(\mathbf{x})$. Take the GSIP model together with the function (4.20) for individual influence as an example. (i) $s(\mathbf{x})$ is not always smooth and even discontinuous, since each $f_{j,t}(x)$ in (4.19) can be nonsmooth at $\pm h_{j,t}$ and discontinuous at $\pm l_{j,t}$. (ii) A closed-form of $s(\mathbf{x})$ cannot be obtained generally, except when $f_{j,t}(x) = x,\ \forall t > 0,\ v_j \in V$, in (4.19). (iii) Now both left and right derivatives do not generally exist. However, we can still show that the evaluation of the objective function can be solved efficiently, as in Proposition 17. Hence, it is then necessary to treat the objective function as an input-output (black-box) system, and resort to derivative-free methods (DFMs) for general solutions.

Proposition 17. *Given a signed network $G(V, E, \mathbf{W})$ with the signed weight matrix $\mathbf{W}$ and an initial state $\mathbf{x}(0)$, with the GSIP model governing the information propagation process and Eq. (4.20) as the function for individual influence, the problem of computing the objective function $s(\mathbf{x}(0))$ in the MINLP (4.31) (i.e. Eq. 4.28)*

can be solved in $O(|E|t_\epsilon)$ time, where t_ϵ is the number of time steps required for the convergence of the underlying process with tolerance $\epsilon > 0$, where

$$\left\|(1-\gamma)^t \mathbf{x}(t)\right\|_2 < \epsilon, \quad \forall t \geq t_\epsilon.$$

Proof. The time complexity follows from Alg. 4. In each iteration t, each nonzero element of the weight matrix $\mathbf{W}$ has only one chance to be used to potentially adjust the state value $\mathbf{x}^{(t)}$, and there are overall $O(|E|)$ such elements. Therefore, the time complexity of each iteration is $O(|E|)$, and the overall evaluation has time complexity $O(|E|t_\epsilon)$, depending on the number of steps towards convergence, t_ϵ. □

Algorithm 4 Influence evaluation on signed networks.

Input: A signed network $G(V, E, \mathbf{W})$ with its weight matrix $\mathbf{W}$ where $W_{ij} \neq 0$ if $(v_i, v_j) \in E$, parameters $\{l_{j,t}\}$, $\{h_{j,t}\}$ in the GSIP model, time-discounting factor γ, the initial state $\mathbf{x}(0)$ where $x_j(0) \in [-h_{j,0}, -l_{j,0}] \cup [l_{j,0}, h_{j,0}]$ if and only if $v_j \in \mathcal{A}_0$ (0 otherwise), and the tolerance ϵ.
Output: The value of the objective function in the MINLP (4.31), s.
Set $t \leftarrow 0$, $\mathbf{x}^{(0)} \leftarrow \mathbf{x}(0)$, and $s \leftarrow 0$.
Mark all the neighbours, both positive and negative, of $\mathcal{A}_0$ as potentially activated nodes, $\mathcal{N}_0 \leftarrow \bigcup_{v_j \in \mathcal{A}_0} \mathcal{N}(v_j)$.
while $\left\|(1-\gamma)^t \mathbf{x}^{(t)}\right\|_2 > \epsilon$ **do**
    $\mathcal{A}_{t+1}, \mathcal{N}_{t+1} \leftarrow \emptyset$, and $\mathbf{x}^{(t+1)} \leftarrow \mathbf{0}$;
    for each potentially activated node $v_j \in \mathcal{N}_t$ **do**
        $x_j^{(t+1)} = f_{j,t}(\sum_{i \in \mathcal{A}_t} W_{ij} x_i^{(t)})$;
        if $x_j^{(t+1)} \neq 0$ **then**
            $\mathcal{A}_{t+1} \leftarrow \mathcal{A}_{t+1} \cup \{v_j\}$;
            $\mathcal{N}_{t+1} \leftarrow \mathcal{N}_{t+1} \cup \mathcal{N}(v_j)$;
            $s \leftarrow s + (1-\gamma)^{t+1} x_j^{(t+1)}$;
        end if
    end for
    $t \leftarrow t + 1$;
end while

Since we cannot assume the objective to fall in a simple family, e.g. polynomials, model-based methods may not fit the current problem, either. This again limits the choice among algorithm classes for its solution to that of direct-search methods. Among them, the mesh adaptive direct search (MADS) method is still the one of the few that has local convergence analysis when the objective function is not necessarily

Lipschitz continuous [152], and accordingly, we refer to the MADS for mixed variables (MV) [1] as a general solution to the IM problem (4.31) (and also (4.30)), with implementation in NOMAD [10,93]. We refer the reader to Sec. 3.4.2 for more details.

4.5.2 Special cases

In the previous section, we have analysed the highly general features of the IM problem, and given general solution methods accordingly. Hereafter, we turn our attention to the IM problem with the GSIP model governing the information propagation process and Eq. 4.20 as the function for individual influence. In this section, we consider two special cases of the GSIP model, with the aim of understanding the general scenarios better.

The first special case is when the lower bounds are sufficiently low in the GSIP model, specifically $l_{j,t} = 0,\ \forall v_j \in V$. We can show that in this case, the objective function is continuous with respect to the continuous variables $\mathbf{x}$ as in Proposition 18.

Proposition 18. *If $l_{j,t} = 0,\ \forall v_j \in V$, then the objective function $s(\cdot)$ in the MINLP (4.31) (and (4.30)) is continuous w.r.t. the continuous variables $\mathbf{x}$.*

Proof. We first note that if $f_{j,t}(x)$ is continuous for all $v_j \in V$ and $t > 0$, then by the properties of composite continuous functions, the objective function $s(\cdot)$ is also continuous. Now, we show the continuity of each $f_{j,t}(x)$. (1) We first note that when $l_{j,t} = 0$, the regime $|x| < l_{j,t}$ vanishes, and $f_{j,t}(x) = x$ when $|x| < h_{j,t}$. (2) There are then three parts of the function that are always continuous, i.e. when $-h_{j,t} < x < h_{j,t}$, $x > h_{j,t}$ and $x < -h_{j,t}$. (3) We can show that the function is also continuous at the boundary points, where

$$\lim_{x \to h_{j,t}^-} f_{j,t}(x) = h_{j,t} = f_{i,t}(h_{j,t}) = \lim_{x \to h_{j,t}^+} f_{j,t}(x),$$

and

$$\lim_{x \to -h_{j,t}^+} f_{j,t}(x) = -h_{j,t} = f_{j,t}(-h_{j,t}) = \lim_{x \to -h_{j,t}^-} f_{j,t}(x).$$

Hence, $f_{j,t}(x)$ is continuous for all possible $x \in \mathbb{R}$. □

Furthermore, we can also show that the objective function is Lipschitz continuous by noting that $0 \leq |f_{j,t}(x)| \leq |x|,\ \forall t > 0, v_j \in V$ and potential value x. In this

case, the problem is open to methods with global convergence, for example the new derivative-free line-search type algorithms [66]. Meanwhile, the condition of the lower bounds could be looser in practice, where for example, if the signed network has weights of uniform magnitude α and the initial state values are uniform with $l_{j,0} = h_{j,0} = l_0,\ \forall v_j \in V$, then the requirement of the lower bounds can be relaxed to $l_{j,t} \leq l_0\alpha^t,\ \forall t > 0,\ v_j \in V$, as discussed in Sec. 4.4.2.

The other special case is when not only $\{l_{j,t}\}$ are sufficiently small but $\{h_{j,t}\}$ are sufficiently large in the GSIP model, i.e. the linear-dynamics extreme as analysed in Sec. 4.4.2. Hence,

$$\mathbf{x}(t) = \mathbf{W}^T\mathbf{x}(t-1) = \left(\mathbf{W}^T\right)^t\mathbf{x}(0),$$

and the objective function is then,

$$\begin{aligned} s(\mathbf{x}(0)) &= \sum_j \sum_{t=1}^{\infty} (1-\gamma)^t x_j(t) = \sum_{t=1}^{\infty} \mathbf{1}^T \left((1-\gamma)\mathbf{W}^T\right)^t \mathbf{x}(0) \\ &= \mathbf{1}^T \left(\left(\mathbf{I} - (1-\gamma)\mathbf{W}^T\right)^{-1} - \mathbf{I}\right)\mathbf{x}(0) = \mathbf{c}^T\mathbf{x}(0), \end{aligned} \tag{4.32}$$

as analysed in Sec. 4.3.2.1, where $\mathbf{c} = ((\mathbf{I} - (1-\gamma)\mathbf{W})^{-1} - \mathbf{I})\mathbf{1}$ is the *signed Katz centrality* with factor $(1-\gamma)$, $\mathbf{I}$ is the identity matrix, and the penultimate equation is obtained given that condition (4.6) is true in the linear-dynamics extreme. Hence, the objective function is linear, thus (Lipschitz) continuous, concave and smooth. The exact solution(s) in this case is achievable as in Theorem 19.

Theorem 19. *When $\{l_{j,t}\}$, $\{h_{j,t}\}$ are as in Theorem 15, then the exact solution(s) to the MINLP (4.31) is*

$$x_j^* = \begin{cases} h_{j,0}, & \text{if } j \in \mathcal{A}, \\ 0, & \text{otherwise}, \end{cases} \quad z_j^* = \begin{cases} 1, & \text{if } j \in \mathcal{A}, \\ 0, & \text{otherwise}, \end{cases} \tag{4.33}$$

where $\mathcal{A} = \{j_1, ..., j_{n_k}\}$ s.t. $c_j > 0$ and $h_{i,0}c_i \leq h_{j,0}c_j,\ \forall i \notin \mathcal{A}, j \in \mathcal{A}$, $n_k = \min\{k, |\{c_j : c_j > 0\}|\}$, $\mathbf{c} = ((\mathbf{I} - (1-\gamma)\mathbf{W})^{-1} - \mathbf{I})\mathbf{1}$ is the signed Katz centrality with factor $(1-\gamma)$, and the uniqueness of the solution depends on the uniqueness of set $\mathcal{A}$.

Proof. By Theorem 15, the GSIP model reaches the linear-dynamics extreme. Then

as in (4.32), the objective function $s(\cdot)$ is linear in $\mathbf{x}$,

$$s(\mathbf{x}) = \mathbf{c}^T\mathbf{x},$$

where $\mathbf{c} = ((\mathbf{I} - (1-\gamma)\mathbf{W})^{-1} - \mathbf{I})\mathbf{1}$. For illustrative purposes, we split the proof into two parts. (1) We first analyse the MINLP (4.31) solely w.r.t. $\mathbf{x}$ while fixing the integer variables $\mathbf{z}$. The problem can then be decomposed into n sub-problems, where for each $v_j \in V$,

$$\begin{aligned} \max_{x_j} \quad & \tilde{s}_j(x_j) \coloneqq c_j x_j \\ s.t. \quad & x_j \leq h_{j,0} z_j, \\ & x_j \geq l_{j,0} z_j, \\ & x_j \in \mathbb{R}. \end{aligned} \tag{4.34}$$

Hence, for v_j with $c_j > 0$, the optimal solution is $x_j^* = h_{j,0} z_i$, while for those with $c_j < 0$, the optimal solution is $x_j^* = l_{j,0} z_i$. (2) Then we consider the MINLP (4.31) w.r.t. $\mathbf{z}$ when $\mathbf{x}$ is at the optimal value, where

$$\begin{aligned} \max_{\mathbf{z}} \quad & \sum_j \tilde{s}_j^* = \sum_{j:c_j>0} c_j h_{j,0} z_i + \sum_{j:c_j<0} c_j l_{j,0} z_i \\ s.t. \quad & \sum_j z_j \leq k, \\ & z_j \in \{0,1\}, \forall j. \end{aligned}$$

We can show that the optimal solution is to set $z_j = 1$ if node v_j has $c_j > 0$ and is also ranked among the top k according to its coefficient in the objective, $c_i h_{j,0}$. This gives the solution in (4.33), and the uniqueness of the solution depends on the uniqueness of the top k nodes. □

Hence the exact solution(s) when the GSIP model is in the linear-dynamics extreme is to first restrict to nodes of positive signed Katz centrality, and then choose the top k nodes, if there are any, of the highest product of their signed Katz centrality and their maximum initial values. This relates the IM problem to a well-studied centrality measure, the Katz centrality, although now extended to signed networks. Furthermore, this solution can serve as a warm start in the search algorithms for the MINLP (4.31) in general, where the number of steps required to find a good solution is potentially proportional to how far the underlying propagation is from being the linear dynamics.

4.5.3 Customised methods

Here, the objective function $s(\mathbf{x}(0))$ is not necessarily non-decreasing in the initial state value $\mathbf{x}(0)$, as a result of negative connections in opposing the states or the influence of the corresponding neighbours. Hence, the problem (4.31) is not reducible, and both the continuous variables in $\mathbf{x}$ and the binary variables in $\mathbf{z}$ need to be determined[9]. To incorporate the features explored in Sec. 4.5.2, we can utilise the exact solution in the linear-dynamics extreme as a warm start in the MADS method, with the assumption that the local optima near this solution are sufficiently good.

In real cases, it is highly likely that people only have several choices of initial state of a node, or only one choice. For example, in product range optimisation, each product can only be available or not, which could correspond to a single choice of the initially activated value. In this case, $l_{j,0} = h_{j,0},\ \forall v_j \in V$, then $x_j = l_{j,0}z_j = h_{j,0}z_j$, thus the MINLP (4.31) is equivalent to the following problem,

$$\begin{aligned} \max_{\mathbf{z}} \quad & s(\mathbf{h}_0 \odot \mathbf{z}) \\ s.t. \quad & \sum_j z_j \leq k, \\ & z_j \in \{0,1\}, \forall j, \end{aligned} \tag{4.35}$$

where $\mathbf{h}_0 = (h_{j,0})$, and $\odot$ denotes the element-wise (Hadamard) product. Note the difference with the revised problem (3.30) on simple networks is that the number of initially activated nodes is not necessarily the budget size k in the optimal solution(s). Now, the domain Ω^d is a natural mesh to search at each iteration r,

$$M_r = \Omega^d = \{\mathbf{z} \in \{0,1\}^n : \sum_j z_j \leq k\}. \tag{4.36}$$

The constraints are incorporated in the domain, and are treated by the extreme barrier approach s_{Ω^d}, where $s_{\Omega^d}(\mathbf{z}) = s(\mathbf{h}_0 \odot \mathbf{z})$ if $\mathbf{z} \in \Omega^d$ and $-\infty$ otherwise. We define the neighbourhood function of binary variables $\mathbf{z}$ to be,

$$\mathcal{N}(\mathbf{z}) = \{\mathbf{y} \in \{0,1\}^n : \|\mathbf{y} - \mathbf{z}\|_1 \leq d\}, \tag{4.37}$$

where $d \in Z^+ \backslash \{1\}$, and 2 is the shortest distance to allow exchanging one element of value 1 with another of value 0.

[9]In the implementation of the MADS method, we need to enlarge the discrete neighbourhood size to be at least 2 for the current problem.

We then propose the following customised direct search (CDS) method by modifying the one proposed in Sec. 3.4.4 for the problem (4.35). Specifically, we update (i) the domain and the mesh to be (4.36), (ii) the initial start to be the solution when the GSIP model reaches the linear-dynamics extreme as in Theorem 15, and (iii) the rule for sufficient improvement to incorporate negative objective values,

$$s_{\Omega^d}(\mathbf{z}) > (1+\zeta)\left|s_{\Omega^d}(\mathbf{z}^{(r)})\right|,$$

where $\mathbf{z}^{(r)}$ is the maximal point in iteration r. Since the termination criteria remain unchanged, the local convergence of the algorithm is maintained. We refer the reader to Sec. 3.4.4 for more details.

4.6 Experiments

In this section, we experimentally illustrate the rich behaviour of the GSIP model on signed networks, and also evaluate the performance of the CDS method for the IM problem (4.35) in both small and large, both synthetic and real signed networks. More importantly, we will demonstrate how the proposed framework of information propagation and influence maximisation is connected with extracting two competing roles in Chapter 2, and also how this work can be useful in practice, through the example in Sec. 2.5.1. Throughout this section, $l_{j,0} = h_{j,0} = 1$, $\forall v_j \in V$, and we apply exclusively the modified threshold-type bounds in (4.24) with condition (4.25) and $\theta_l = \theta_{hl}$ (unless mentioned otherwise), thus the lower bounds vary according to the lower bound threshold θ_l while the upper bounds are also affected by the higher upper bound threshold θ_{hh}.

4.6.1 The proposed model for information propagation

We start from the general features of the GSIP model. In accordance with Sec. 4.4, we show that the GSIP model can have both the threshold effect and the locally linear effect by tuning the lower and upper bounds, respectively. Specifically in signed networks, both effects can have two opposite directions, corresponding to propagating the positive and negative information, which will cause extra complexity to the associated IM problem. Similar to the case of simple networks, these two effects cannot happen simultaneously in either the linear dynamics or the (extended) LT model, but may co-exist in real systems.

Corresponding to the dynamical analysis in Sec. 4.3.2, we consider the following two types of signed networks: (i) one close to being balanced, generated from the two-block planted $SSBM(0.9, 0.1, 0.05)$, and (ii) the other close to being antibalanced, generated from the two-block planted $SSBM(0.9, 0.1, 0.95)$. For the former, an edge is placed between the nodes in the same community with probability $p_{in} = 0.9$ and in the different communities with $p_{out} = 0.1$, where for each edge inside the communities, it has probability $\eta = 0.05$ to have negative sign (positive otherwise), while for each edge between the communities, it has probability $\eta = 0.05$ to have positive sign (negative otherwise). The latter is constructed in the same way except that $\eta = 0.95$. The signed networks have size $n = 50$ and $n_c = 2$ communities, where we label the nodes in communities one and two as 0 to 24 and 25 to 49, respectively; see Fig. 4.6 for one realisation. We assign a uniform magnitude $\alpha = 0.1$ to the edge weights and set the time-discounting factor as $\gamma = 0.6$, to account for moderate level of trust/distrust among agents. Therefore, $\theta_l = 1$ corresponds to the critical lower bounds for the linear-dynamics extreme, where any $\theta_l > 1$ cannot always result in linear dynamics. The difference in the propagation behaviour will be quantified by the time-dependent

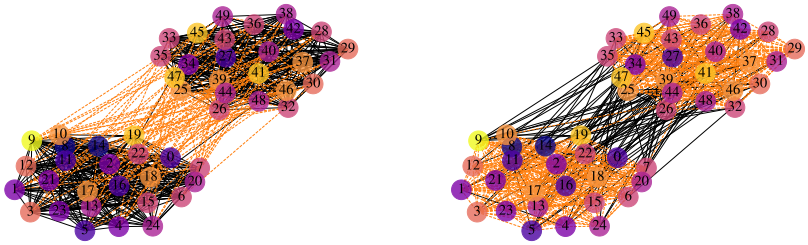


Figure 4.6: One realisation of the two-block $SSBM(0.9, 0.1, 0.05)$ (left) and $SSBM(0.9, 0.1, 0.95)$ (right) where positive edges are in black while negative edges are in orange (dashed), and the lighter the colour of a node is, the higher the degree (4.1) of the node is.

influence of all nodes,

$$s(t) = \sum_j \sum_{t'=0}^{t} (1-\gamma)^{t'} x_j(t'), \tag{4.38}$$

where $\mathbf{x}(t') = (x_j(t'))$ is the state vector at time step t' with the updating function (4.19) and a given initial state vector $\mathbf{x}(0) = (x_j(0))$. Note that $\lim_{t\to\infty} s(t) - \sum_j x_j(0)$ is the objective in the IM problem.

We first show the imposed threshold effect on top of the linear dynamics, by tuning the lower bounds of the GSIP model to gradually deviate it from the linear-dynamics

extreme, as in Sec. 4.4.2. Specifically, we consider the following two initially activated node sets as in Claim 3: (i) $\{0, 1\}$ from the same community; (ii) $\{0, 25\}$ from the different communities. Firstly, when $\theta_l = 1$, which approaches the linear-dynamics extreme, we observe that the time-dependent influence from the two node sets are similar; see Fig. 4.7. Specifically, in the signed networks close to being balanced, the overall influence keeps increasing until convergence, while in the networks close to being antibalanced, the overall influence alternates the direction of change, which indicates that the influence at each time step keeps changing signs. These observations verify the analysis of linear dynamics on balanced and antibalanced networks in Sec. 4.3.2.1. Secondly, as θ_l increases, set (i) is expected to trigger a propagation with higher overall influence when the network is close to being balanced, while it triggers one with lower overall influence when the network is close to being antibalanced. These imply the two opposite directions in the threshold effect: one corresponds to thresholding the not-positive-enough influence attempts to be 0, which decreases the overall influence, while the other corresponds to suppressing the not-negative-enough influence attempts to be 0, which increases the overall influence. Hence, it is important to understand the balanced structure of signed networks. Finally, when θ_l exceeds the critical value of 1, the overall influence from set (ii) is always 0. This is due to the coexistence of both positive and negative influence on the nodes, and here we have a special case where they are of the same level. As analysed in Sec. 4.4.2, both the coexistence of positive and negative influence, and the possibility to have zero overall influence while nonzero individual influence, are unique in signed networks.

We then illustrate the locally linear effect added on top of the extended SLT model, by changing the upper bounds of the GSIP model to differentiate it from the LT extreme. Here, we fix $\theta_l = 2$, the smallest integer for the threshold effect to take place, while increasing θ_{hh} from 2 to a higher value, as mentioned in Sec. 4.4.3. We then consider the following two larger initially activated node sets: (i) $\{0, 1, 2, 3\}$ from the same community; (ii) $\{0, 1, 25, 26\}$ evenly distributed in the two communities. Firstly, when $\theta_l = \theta_{hh} = 2$, i.e. in the LT extreme, set (ii) is expected to trigger a propagation with relatively higher overall influence when the underlying signed networks are close to being balanced, while it tends to trigger one with relatively lower overall influence when the signed networks are close to being antibalanced; see Fig. 4.8. This again indicates the two directions of the threshold effect. Furthermore, we observe that while increasing until convergence in the close-to-balanced networks, the overall influence alternates the direction of change in the close-to-antibalanced networks, which leads the difference between the two node sets to be smaller. These

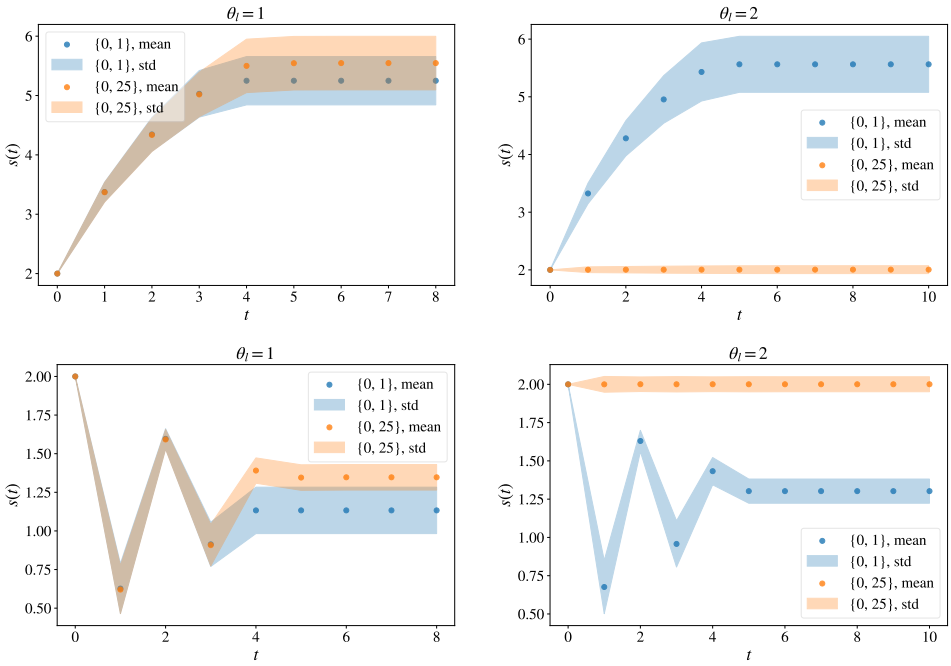


Figure 4.7: The time-dependent influence, from two different initially activated node sets, with $\theta_l = 1$ (left, the critical value for the linear-dynamics extreme) and $\theta_l = 2$ (right, a larger value) while θ_{hh} being large (here 9000), on 1000 samples of $SSBM(0.9, 0.1, 0.05)$ (top) and $SSBM(0.9, 0.1, 0.95)$ (bottom).

observations supplement the analysis of the extended SLT model on signed regular lattices in Sec. 4.3.2.2. Secondly, as θ_{hh} becomes larger, the increase in the overall influence of set (i) is larger than that of set (ii) in the signed networks close to being balanced, while the decrease of the overall influence of set (i) is larger than that of set (ii) in the signed networks close to being antibalanced. These then imply the two opposite directions in the locally linear effect: one corresponds to nodes of higher state values to positively influence others, which increases the overall influence, and the other corresponds to nodes of higher state values to negatively influence others, which decreases the overall influence. Hence, the balanced structure of signed networks again plays a key role in characterising the behaviour of the dynamics. Thirdly, we observe that the overall influence (ignoring the initial state values) is positive in the signed networks close to being balanced where they have more positive edges, while it is negative in the ones close to being antibalanced with more negative edges, which also supplements to the analysis of the overall influence in Sec. 4.3.2.2.

Finally, we integrate the two aspects and provide a whole picture of the general features of the GSIP model, by changing the upper and lower bounds simultaneously. Here, we consider both pairs of node sets, $\{0, 1\}$ versus $\{0, 25\}$ and $\{0, 1, 2, 3\}$ ver-

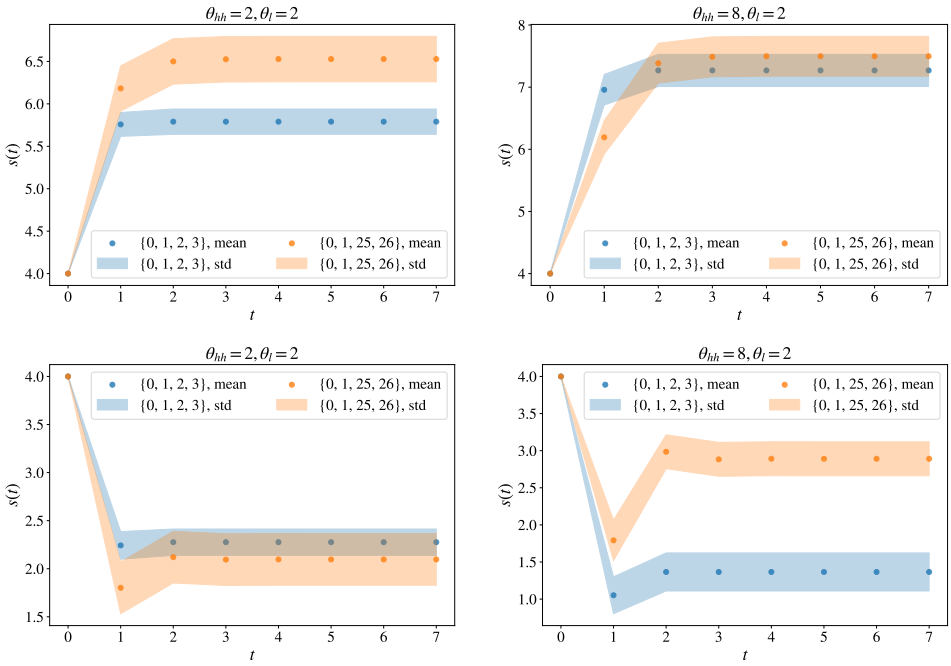


Figure 4.8: The time-dependent influence, from two different initially activated node sets, with $\theta_l = 2$ while $\theta_{hh} = 2$ (left, the LT extreme) and $\theta_{hh} = 8$ (right, a larger value), on 1000 samples of $SSBM(0.9, 0.1, 0.05)$ (top) and $SSBM(0.9, 0.1, 0.95)$ (bottom).

sus $\{0, 1, 25, 26\}$, and quantify the change of behaviour through the *ratio* δ of the time-dependent influence of all nodes as t goes to infinity, i.e. $\lim_{t\to\infty} s(t)$.[10] We observe consistent patterns of the previously analysed features: (i) in the signed networks close to being balanced, the node sets in the same community have consistently higher influence as θ_l exceeds the critical value for the other set, 1 for $\{0, 25\}$ and 2 for $\{0, 1, 25, 26\}$ (which are equal to the numbers of nodes distributed in each community), while in the signed networks close to being antibalanced, the node sets evenly distributed in the two communities have consistently better performance under the same condition; (ii) the node sets in each pair have generally more similar performance as the upper bounds become larger, although there is more complexity in between; see Fig. 4.9. However, the differences in the two types of signed networks are also clear. The ratios δ in the signed networks close to being balanced are similar to the results obtained in simple networks as in Fig. 3.8. This is expected, to some extent, since balanced networks share the same spectrum as the corresponding unsigned ones ignoring the edge sign, by Theorem 11. While in the signed networks close to being antibalanced, the node sets distributed in the two communities generally have better

[10]This quantity is considered to avoid negative values.

performance than the ones in the same community, except in limited cases, e.g., when $\theta_l \leq 1$ and θ_{hh} no greater than the critical value mentioned before. This is because here most edges are negative and within communities, and then nodes in the same communities have a higher probability to negatively activate others. These separating characteristics imply the importance of understanding their balanced structure.

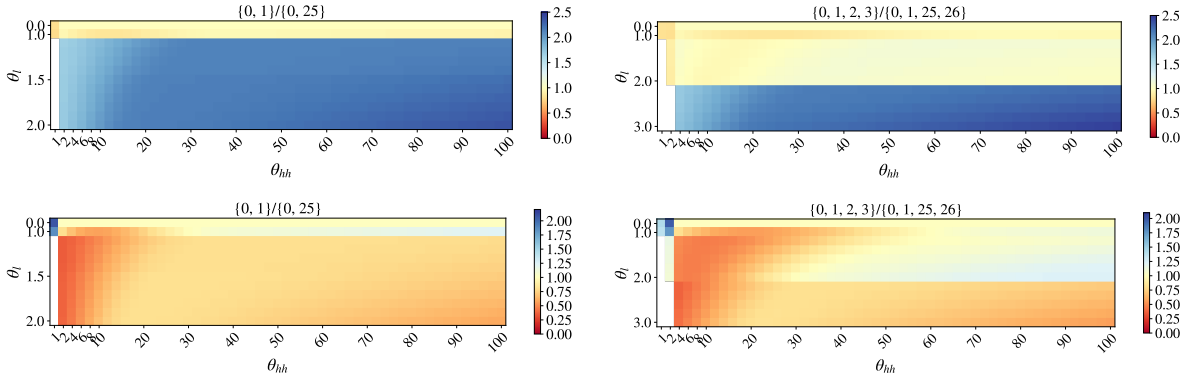


Figure 4.9: The ratios of $\lim_{t\to\infty} s(t)$ in Eq. (4.38) from two pairs of initially activated node sets, $\{0,1\}$ to $\{0,25\}$ (left) and $\{0,1,2,3\}$ to $\{0,1,25,26\}$ (right), with changing upper bound (x-axis) and lower bound (y-axis) thresholds, on 1000 samples of $SSBM(0.9, 0.1, 0.05)$ (top) and $SSBM(0.9, 0.1, 0.95)$ (bottom).

4.6.2 The proposed method for influence maximisation

We now proceed to the proposed CDS method for the IM problem (4.35). In the following, we first test the accuracy of the CDS method on relatively small networks. We then examine its performance on relatively large networks through comparison with other state-of-the-art approaches adapted for signed networks. Similar to Sec. 4.6.1, we also incorporate the balanced structure of signed networks through considering both the ones that are close to being balanced and those that are close to being antibalanced.

4.6.2.1 Comparison with global optima

From Sec. 4.5.3, we know that the CDS method is a local algorithm, thus it is important to explore how good the output is with respect to the global optima. However, algorithms with global convergence, such as the brute-force method, now require $O(n^{k^2})$ evaluations of the objective function, which prohibits their application to large networks. Hence in this section, we consider exclusively signed networks of a relatively small scale.

In order to measure the *goodness* of the output, we consider the following two dimensions again, accuracy and rank, but both are extended to signed networks. The *accuracy* is now defined as one minus the relative error with respect to a global optimum,

$$\tau(s; s^*) = 1 - |(s^* - s)/s^*|,$$

where s is an output from the proposed algorithm, and s^* is a global optimum, with the convention that $\tau = 0$ if $s^* - s > |s^*|$. It is different from the one defined in Sec. 3.5.2.2 because $s < 0$ is possible in signed networks. Still, $\tau(\cdot) \in [0, 1]$ and a higher accuracy implies a better solution. For the *rank* of the output, we now need to compare it with all sets of sizes no greater than k,

$$\phi(\mathcal{A}_0; V, k) = \frac{|\{\mathcal{A} \subset V : |\mathcal{A}| = k, s(\mathbf{h}_0 \odot \mathbf{z}_{\mathcal{A}}) > s(\mathbf{h}_0 \odot \mathbf{z}_{\mathcal{A}_0})\}| + 1}{|\{\mathcal{A} \subset V : |\mathcal{A}| \leq k\}|}, \tag{4.39}$$

where V is the node set, k is the budget size, $\mathcal{A}_0$ is the set of initially activated nodes corresponding to the output, $\mathbf{z}_{\mathcal{A}}$ is the binary vector whose jth element is 1 if and only if $v_j \in \mathcal{A}$, and $s(\mathbf{h}_0 \odot \mathbf{z})$ is the objective function of the problem (4.35). Still, $\phi(\cdot) \in (0, 1]$ and a lower rank implies a better solution. Hence, in order to obtain the parameters in both measures, we select the *brute-force* as the reference global algorithm.

Specifically, we separate the two cases: (i) differentiating θ_l and θ_{hh} while maintaining the same k, and (ii) varying k while fixing θ_l and θ_{hh}. For the underlying signed networks, we consider both synthetic ones from signed stochastic block models, and a classic example of real signed networks, the Highland tribes network.

Signed stochastic block model. The SSBM considered here has the same size and weights as in Sec. 4.6.1, but different probabilities in the two communities, $p_1 = 0.3$ in one and $p_2 = 0.15$ in the other, in order to distinguish nodes in the two communities. The connecting probability between the communities is set to be a smaller value $p_{12} = 0.05$[11]. We also consider the following cases of the flipping probability, $\eta = 0.05$ and $\eta = 0.95$, corresponding to the cases when the underlying signed network is close to being balanced and antibalanced, respectively. In both cases, we assign a uniform magnitude $\alpha = 0.1$ to the edge weights, and set the time-discounting factor as $\gamma = 0$.

[11]Here we set the probability between the two communities to be larger than the one analysed in Sec. 3.5.2.2 to allow more edges of negative sign.

When $k = 4$, the CDS method can find a solution either globally optimal or fairly close to optimal when $\theta_l \leq 2$ in the signed network close to being balanced, while for the one close to being antibalanced, it happens when $\theta_l > 1.4$ or $\theta_l \leq 1$; see Fig. 4.10. For the close-to-balanced network, the performance of the CDS method drops when θ_l becomes larger, or closer to the budget size k, and in the worse-case scenario, the output has accuracy around 0.73 with rank 0.005% in overall 251, 175 sets (of sizes no greater than k as in (4.39)), thus the CDS method can still output a top 13 set. While for the close-to-antibalanced network, the CDS method cannot find a global optimal solution in some cases when θ_l is around 1.3, and in the worst-case scenario, the accuracy of the output is 0.75 with rank 0.002%, i.e. the CDS method can still output a top 7 set.

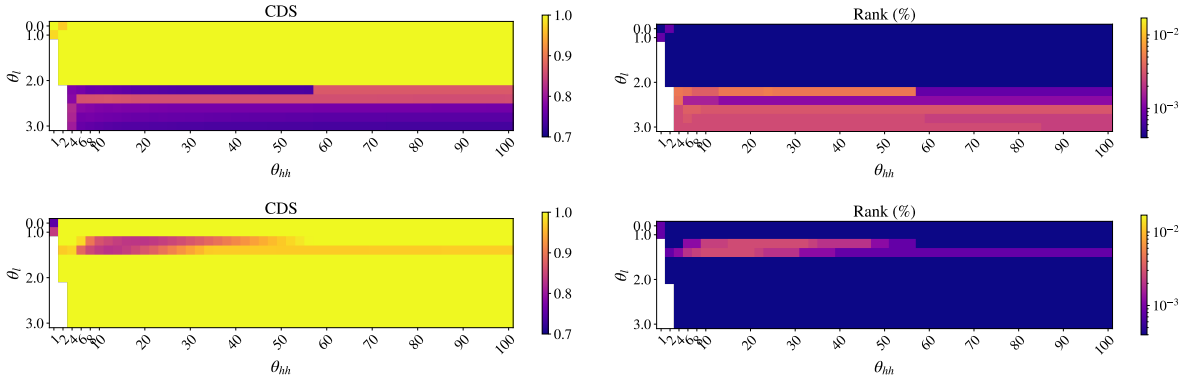


Figure 4.10: Performance of the CDS method on the SSBM with $\eta = 0.05$ (top) and $\eta = 0.95$ (bottom) in terms of the output's accuracy (left) and rank (right), subject to changing upper (x-axis) and lower (y-axis) bound thresholds of the GSIP model, θ_{hh} and θ_l, respectively, when $k = 4$.

We further explore the performance of the CDS method when increasing the budget size k. Specifically, for the signed network close to being balanced, we consider $\theta_l = 3, \theta_{hh} = 6$, while for the one close to being antibalanced, we choose $\theta_l = 1.4, \theta_{hh} = 10$. These combinations correspond to the cases where the CDS method cannot find a globally optimal solution when $k = 4$. We observe better performance of the CDS method as the budget size k becomes larger, in both networks of very different balanced structures; see Fig. 4.11. Note that we do not explore the cases when k is greater than 6 because there are already over 10 million sets to evaluate in the brute force when $k = 6$.

Highland tribes network. We now consider a classic example of real signed networks, the Highland tribes network. The network contains $n = 16$ tribes as nodes,

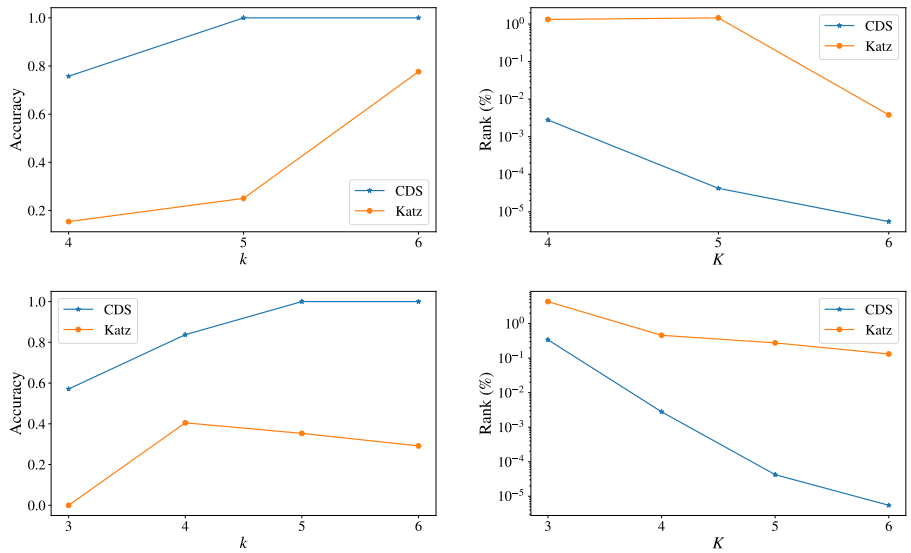


Figure 4.11: Performance of the CDS method in terms of the output's accuracy (left) and rank (right), subject to changing budget size k (x-axis) when $\theta_l = 3, \theta_{hh} = 6$ on the SSBM with $\eta = 0.05$ (top) and $\theta_l = 1.4, \theta_{hh} = 10$ on the SSBM with $\eta = 0.95$ (bottom).

and a positive edge indicates friendship while a negative edge indicates enmity. Hage and Harary [73] used the Gahuku-Gama system of the Eastern Central Highlands of New Guinea, described by Read [130], to illustrate a clusterable signed network. This network is weakly balanced, where the enemy of an enemy can be either a friend or an enemy, thus is neither balanced nor antibalanced; see Fig. 4.12. This is consistent with the distance $d_b(G) = 0.155$ from being balanced and $d_a(G) = 0.529$ from being antibalanced. We assign a uniform magnitude $\alpha = 0.1$ to the edge weights, and set the time-discounting factor as $\gamma = 0$.

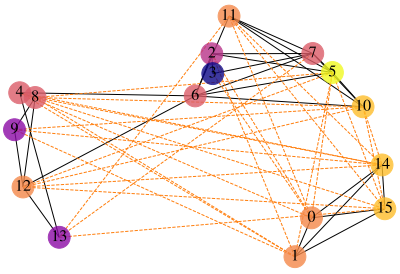


Figure 4.12: Highland tribes network, where positive edges are in black while negative edges are in orange (dashed), and the lighter the colour of a node is, the higher the degree (4.1) of the node is.

When $k = 5$, the CDS method can successfully find either a globally optimal or close-to-optimal solution with different choices of upper and lower bounds in this

signed network; see Fig. 4.13. When $\theta_h < 4$, an optimal set could contain less than k nodes. When $\theta_l \leq 1$, the nodes returns by the algorithm are ranked among the top $k+1$ by the signed Katz centrality (i.e., in $2, 3, 5, 6, 7, 11$) when θ_{hh} is large, while are relatively far from each other when θ_{hh} is small (e.g., $2, 3, 6, 7, 12$). As θ_l increases to 3, the algorithm generally provides node sets of smaller signed Katz centrality but with more overlap of their positive neighbourhoods (e.g., $2, 5, 6, 7, 10$). There are only 5 cases when the CDS method cannot find a global optimum, where the upper bounds are close to the lower bounds. However, even in these worst-case scenarios, the CDS method can still output the sets with accuracy greater than 0.93 and rank in the top 0.03% in overall $7,000$ node sets, i.e. it can still output a top 2 set.

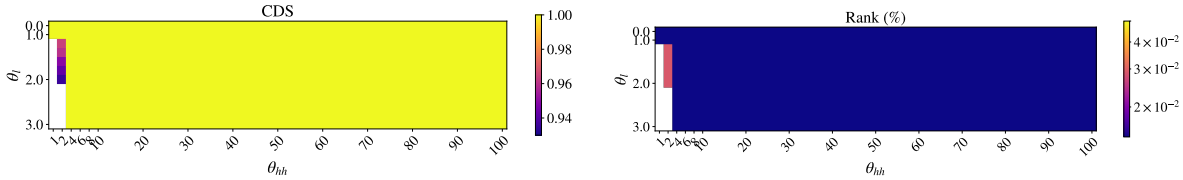


Figure 4.13: Performance of the CDS method on the Highland tribes network in terms of the output's accuracy (left) and the rank (%, right), subject to changing upper (x-axis) and lower (y-axis) bound thresholds of the GSIP model, θ_{hh} and θ_l, respectively, when $k = 5$.

We then change the budget size k in one of the worst-case scenarios when $\theta_l = 2$ and $\theta_{hh} = 2$. We observe that the accuracy of the output from the CDS method maintains at a relatively high level (greater than 0.9) while increasing the budget size k, with consistently low rank; see Fig. 4.14. We note that the signed network has a significant portion of negative edges, and here exactly half of all edges have negative sign, hence it is very possible that the optimal set which maximises the overall influence does not take the whole budget, i.e. $\left|\{v_j : x_j^*(0)! = 0\}\right| < k$ where $\mathbf{x}^*(0)$ is a true optimal solution (obtained by the brute force). We verify that it is the case in the highland tribes networks, where a set of size 4 can maintain the optimality when further increasing the budget size to be greater than 4 in the experiments.

4.6.2.2 Comparison with the state-of-the-art

In the following, we examine the performance of the CDS method on both synthetic and real networks on a relatively large scale, where we compare them with the state-of-the-art methods extended for signed networks. Hereafter, we denote the set of initially activated nodes, $\mathcal{A}_0$, as the *seed set*. We note that the vast majority of

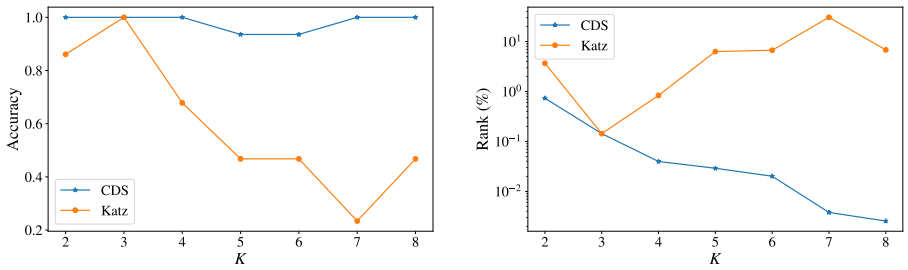


Figure 4.14: Performance of the CDS method on the Highland tribes network in terms of the output's accuracy in terms of the output's accuracy (left) and the rank (%, right), subject to changing budget size k (x-axis) when $\theta_l = 2, \theta_{hh} = 2$.

greedy algorithms following the work of Kempe et al. [85] are not applicable here, because node sets that are not large enough[12] will return an overall influence 0 and then these algorithms lack an appropriate approach to select the first few nodes, which has significant consequences in their following steps. We instead compare the CDS method with the following ones.

(i) **Random sampling** ("Random"). Randomly selecting an integer k_r between 1 and k inclusively, then randomly selecting k_r nodes of the signed network, and returning them as the seed set. We repeat the process n_s times, and output the one with the highest objective value.

(ii) **Degree-centrality method** ("Degree"). We define the signed degree here as

$$d_i^{(s)} = \sum_j W_{ij}, \quad \forall v_i \in V,$$

where $\mathbf{W}$ is the signed (weighted) adjacency matrix. Hence, the signed degree of a node can take negative values. The degree-centrality method is then to select the k nodes of the highest positive degree in the network as the seed set if possible; otherwise return all nodes of positive degree.

(iii) **Katz-centrality method** ("Katz"). Finding the k nodes of the highest positive signed Katz centrality (as in Sec. 4.5.2) in the signed network as the seed set if possible; otherwise return all nodes of positive Katz centrality.

Since method (i) has random components, we will repeat the method n_r times and analyse their averaged performance. In this section, we compare the performance of

[12]It corresponds to the node sets of sizes less than θ_l in networks with weights of uniform magnitude, while in weighted networks, we also need to incorporate the exact weights around each node compared to the mean absolute weight.

different methods directly through the overall influence s, i.e. the objective value in the IM problem (4.35) (or equivalently (4.28)).

The signed networks under consideration are composed of two large SSBM, one close to being balanced and the other closed to being antibalanced, and a real signed network among philosophers. On the one hand, SSBMs are considered here because community structure is a common feature in real networks, and also to extend the previous analysis in Sec. 4.6.2.1 to a larger scale. On the other hand, the Philosophers' network is a classic example in signed networks of a relatively large scale, where each edge indicates relatively strong interactions between philosophers.

Signed stochastic block model. Consider a relatively large signed network of size $n = 500$, generated by two-block SSBMs with different probabilities in the two communities, $p_1 = 0.03$ and $p_2 = 0.015$, and the connecting probability between the two being $p_{12} = 0.005$. These values are chosen to maintain roughly the same mean degree as the SSBMs in Sec. 4.6.2.1. We also consider the following cases of the flipping probability: (i) $\eta = 0.05$ for the signed network close to being balanced, and (ii) $\eta = 0.95$ for the signed network close to being antibalanced. We assign the same uniform magnitude $\alpha = 0.1$ for the edge weights and the time-discounting factor $\gamma = 0$. We then observe that, in both cases, the CDS method outperforms all the reference algorithms in all possible budget sizes; see Fig. 4.15. Furthermore, there is a drop in performance of the CDS method when increasing the budget size k from 18 to 20 in the close-to-antibalanced network and when $\theta_{hh} > \theta_l$, which is possible since the current search processes are independent of each other when the budget size is different. A potential extension when the solution can be obtained in a sequential manner is to (re-)start the search process from the optimal set corresponding to a smaller budget size.

Philosophers' network The Philosophers' network describes two types of relationships, master-pupil and acquaintance, between philosophers from different schools of thoughts, recorded by Collins [43]. Collins studies various types of philosophical thoughts in ancient Greece, China, Japan, India, the medieval Islamic and Jewish world, medieval Christendom, and modern Europe, and his book represents division and conflicts among philosophers based on the relationships. Here, we apply the flattened version of this network in [6], and select the largest connected component. The resulting network has size $n = 218$ and $|E| = 259$. In the experiments, we maintain

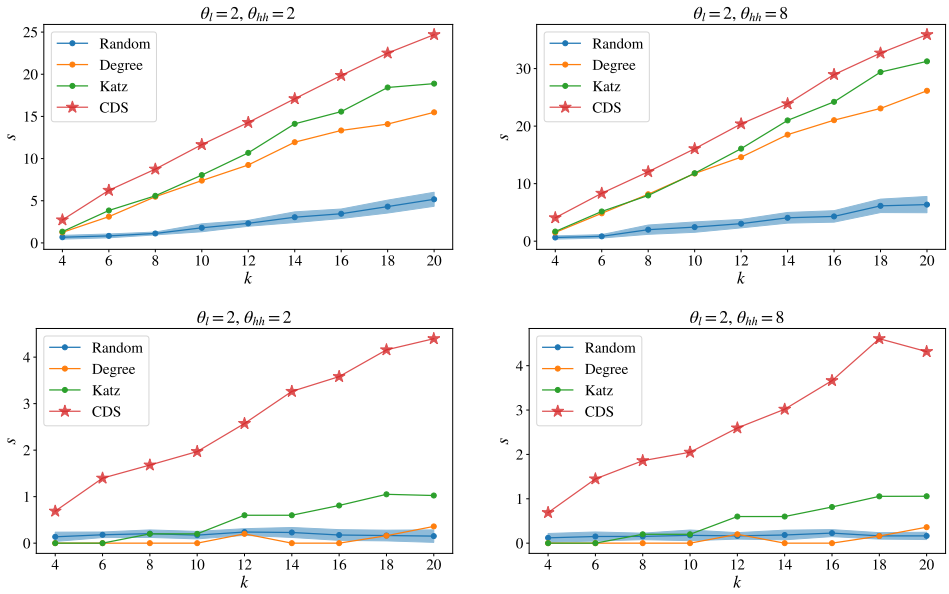


Figure 4.15: Overall influence s from different node selection algorithms applied to the SSBM with $\eta = 0.05$ (top) and $\eta = 0.95$ (bottom), subject to changing budget size (x-axis), when the GSIP model has $\theta_l = \theta_{hl} = \theta_{hh} = 2$ (left) and $\theta_l = \theta_{hl} = 2$, $\theta_{hh} = 8$ (right).

a uniform magnitude $\alpha = 0.1$ for the edge weight and the time-discounting factor $\gamma = 0$.

We observe that the CDS method outperforms all the other reference methods, and its distance with others becomes larger as the higher upper bound threshold θ_{hh} becomes closer to the lower bound threshold θ_l, i.e. the GSIP model approaches the LT extreme; see Fig. 4.16. We also find that the performance of both centrality methods is close to the random sampling in the LT extreme. Meanwhile, the Katz-centrality method performs better when the upper bounds become larger, which is expected from the fact that it is closely related to the exact solution when the GSIP model reaches the linear-dynamics extreme, i.e. when both bounds are sufficiently loose.

4.6.3 An illustrative example for real applications

In this section, we illustrate how the information propagation model and the influence maximisation framework we have developed are connected with extracting two competing roles in networks discussed in Chapter 2, and also how these studies together can be useful in retail, through the example in Sec. 2.5.1.

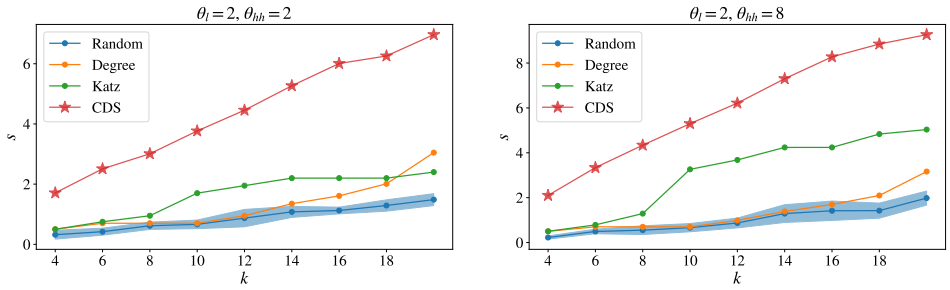


Figure 4.16: Overall influence s from different node selection algorithms applied to the signed Philosophers' network, subject to changing budget size (x-axis), when the proposed model has $\theta_l = \theta_{hl} = \theta_{hh} = 2$ (left) and $\theta_l = \theta_{hl} = 2$, $\theta_{hh} = 8$ (right).

From Chapter 2, we can estimate the degree of complementarity and substitutability between products through the complementarity scores ($W_{ij}^{(c)}$) and substitutability scores ($W_{ij}^{(s)}$), respectively. Then we can build a signed network of product nodes out of them: (i) since the complementarity scores are directly obtained from the purchase incidence, we can define the *unit* of a product as its typical sales quantity (in one basket or shopping trip) for the following analysis in sales changes, and then use ($W_{ij}^{(c)}$) directly; (ii) while the substitutability scores are obtained from the characteristics related to complements, hence to connect the scores with sales changes, we further assume that the substitution is one-to-one[13], and use ($W_{ij}^{(s)} P_{ij}^{(s)}$) instead, where $P_{ij}^{(s)} = W_{ij}^{(s)} / (\sum_l W_{il}^{(s)})$. We give some examples showing the rationale behind applying the transformed scores to construct the signed network in Fig. 4.17.

Hence, the signed (weighted) adjacency matrix we construct for further analysing the dynamic of the sales changes is

$$\mathbf{W}_{cs} = \mathbf{W}^{(c)} - \mathbf{W}^{(s)} \odot \mathbf{P}^{(s)}, \tag{4.40}$$

where $\mathbf{W}^{(c)}, \mathbf{W}^{(s)}$ are the complementarity and substitutability scores, respectively, $\odot$ represents element-wise (Hadamard) matrix product, and $\mathbf{P}^{(s)} = \mathbf{D}^{(s)-1}\mathbf{W}^{(s)}$ with $\mathbf{D}^{(s)} = \mathbf{Diag}(\mathbf{d}^{(s)})$ and each element being $d_i^{(s)} = \sum_l W_{il}^{(s)}$. Note that (1) the signed (weighted) adjacency matrix $\mathbf{W}_{cs}$ in (4.40) is generally asymmetric, although it is still symmetric in the illustrative example here, and (2) the signed network characterised by $\mathbf{W}_{cs}$ is antibalanced. Furthermore, to estimate the co-purchase from random noise, we also add a weight matrix generated from ER random graphs to the constructed

[13]That is to say, decreasing the sales of a product by one unit will increase the sales of its substitutes by at most one unit overall.

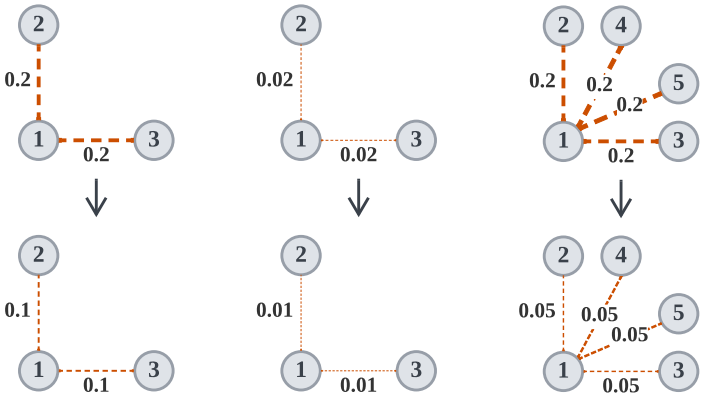


Figure 4.17: Example networks showing the substitutability scores ($(W_{ij}^{(s)})$, top), and the transformed scores to construct the signed network ($(W_{ij}^{(s)} P_{ij}^{(s)})$, bottom), where we show how the transformed score can separate the left from the middle where the structures are the same, and also from the right where the score values are the same.

signed network,

$$\mathbf{W} = \mathbf{W}_{cs} + \alpha_{er} \mathbf{A}_{er}, \tag{4.41}$$

where factor α_{er} controls the level of influence from random noise and $\mathbf{A}_{er}$ is the adjacency matrix from the ER random graph with the connecting probability between nodes being p_{er}. Here, we choose $\alpha_{er} = 0.05$ and $p_{er} = 0.05$. Then the mean absolute weight is approximately $\alpha = 0.25$. We set the time discounting factor in computing the influence (4.20) as $\gamma = 0.6$.

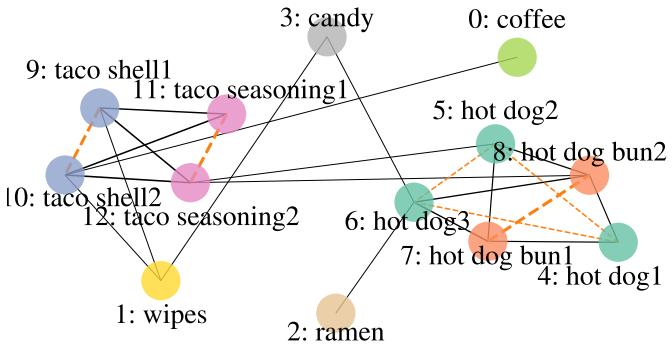


Figure 4.18: One realisation of the signed product network with weight matrix $\mathbf{W}$ as defined in (4.41).

We start from understanding the influence from increasing the sales of specific set of products on the whole system. We understand that there could be certain level of

critical mass to trigger the sales change of a product, and also limits on how much the sales can change, e.g. due to limited stock, hence we assume that the dynamic of the *sales changes* follow the GSIP model (4.19), and then the sales change of each product can be considered as an increasing function of the influence (4.20) on each node. Specifically, we are interested in increasing the sales of the following two pairs of complements, (i) hot dog2 and hot dog bun2, and (ii) taco shell2 and taco seasoning2, i.e. (i) $\{5, 8\}$ and (ii) $\{10, 12\}$ as in Fig. 4.18, since normally retailers do not reduce the price of substitutes to boost their sales simultaneously. We consider the following two scenarios: (a) when the GSIP model reaches the linear-dynamics extreme and another when it slightly deviates from the extreme by increasing the lower bound threshold θ_l, and (b) when the GSIP model reaches the LT extreme and another when it slightly deviates from the extreme by increasing the higher upper bound threshold θ_{hh}.

When the GSIP model is at the linear-dynamics extreme, the two node sets have similar performance, although set (ii) is expected to trigger a propagation of a slightly higher overall influence than set (i); see Fig. 4.19. However, when the lower bounds increase slightly, set (i) can then trigger a propagation of a higher overall influence. Hence, if the sales of each product increase as long as its complements are bought more or its substitutes are bought less, and there is no limit on how much the sales can change, then increasing the sales of the two pairs of complements will have similar results in terms of the sales changes of all products. However, if there is an extra requirement on the sales changes of a product's complements and substitutes to trigger the propagation, then increasing the sales of hot dog2 and hot dog bun2 is recommended.

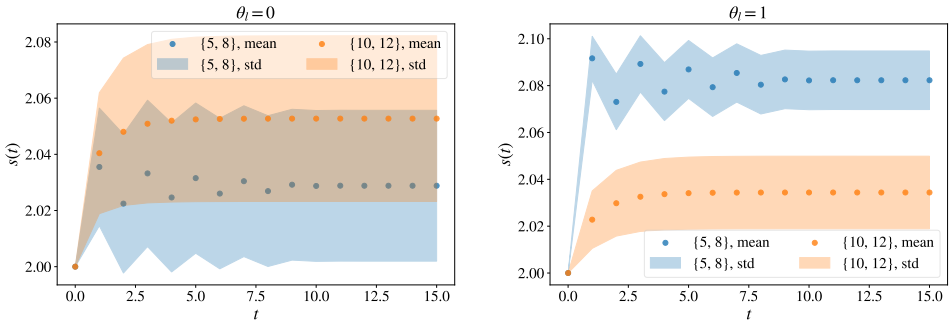


Figure 4.19: The time-dependent influence, from two different initially activated node sets, with $\theta_l = 0$ (left, the critical value for the linear-dynamics extreme) and $\theta_l = 1$ (right, a larger value) while $\theta_{hl} = 2$ and θ_{hh} being large (here 9000), on 1000 samples of ER random graphs in the signed product network.

When the GSIP model is at the LT extreme, set (i) can trigger a propagation of nonzero overall influence, while set (i) cannot achieve it; see Fig. 4.20. However, when the upper bounds increase slightly, the two node sets have similar performance in terms of the overall influence. In other words, if the sales of each product only change a certain amount given that the difference between the sales changes of its complements and those of its substitutes is sufficiently large, then increasing the sales of hot dog2 and hot dog bun2 is expected to trigger a higher increase in the overall sales. However, if the sales can change more in each propagation step, then increasing the sales of both pairs will have similar results in terms of the sales changes of all products.

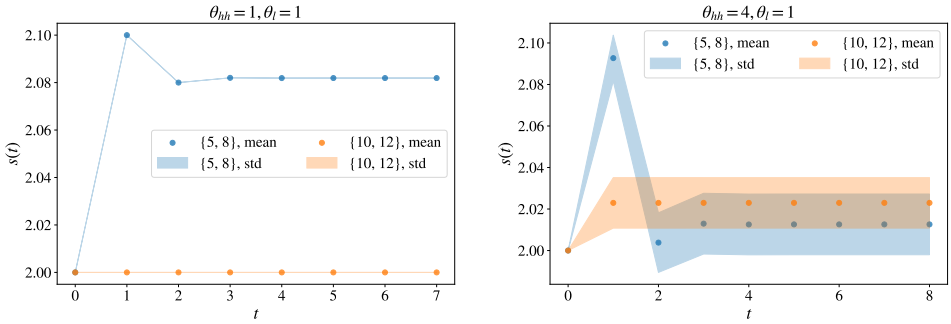


Figure 4.20: The time-dependent influence, from two different initially activated node sets, with $\theta_l = \theta_{hl} = 1$ while $\theta_{hh} = 1$ (left, the LT extreme) and $\theta_{hh} = 4$ (right, a larger value), on 1000 samples of ER random graphs in the signed product network.

Finally, we tune the lower bound threshold θ_l and the higher upper bound threshold θ_{hh} simultaneously, to further explore the performance of the two node sets. Specifically, we consider the two cases of θ_{hl}: (a) $\theta_{hl} = 2$, fixed at a relatively high value, which results in higher upper bounds by the definition of the modifed threshold-type bounds (4.24); (b) $\theta_{hl} = \theta_l$, which pushes the GSIP model to the LT extreme when $\theta_{hh} = \theta_l$. We find consistent patterns in the two cases: if $\theta_l \leq 1$, the two node sets have similar performance, although set (ii) is expected to trigger a propagation of a slightly higher overall influence; if θ_l is around 1, set (i) is expected to have better performance in terms of the overall influence, while the difference becomes smaller as θ_{hh} increases; if θ_l is around 2, set (i) is again expected to have higher overall influence; see Fig. 4.21.

We now move on to the scenario when retailers can freely choose the set of products so as to maximise the overall sales increase, which corresponds to the IM problem on the signed product network. Following the previous analysis, we set $k = 2$, and corresponding to different characteristics of the underlying population, we consider

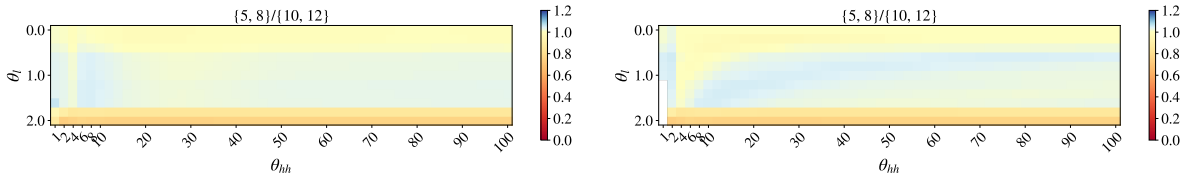


Figure 4.21: The ratios of $\lim_{t\to\infty} s(t)$ in (4.38) from the initially activated node sets, $\{5, 8\}$ to $\{10, 12\}$, with changing θ_{hh} (x-axis) and θ_l (y-axis) while $\theta_{hl} = 2$ (left) and $\theta_{hl} = \theta_l$ (right), on 1000 samples of ER random graphs in the signed product network.

different combinations of upper and lower bounds. We first analyse the true optimal set, in order to understand the nature of the problem more. We observe that when θ_{hh} is low, the node sets that span the two complementary pairs tend to have larger overall influence, and when θ_l is high, the node sets inside each complementary pair are preferred in terms of the overall influence; see Fig. 4.22. Hence, if the underlying population has a relatively strict bound on how much more one can buy, then it is recommended that retailers do not choose the product in the same complementary pair to boost their sales; while if the underlying population has a relatively higher threshold to buy more products, then boosting the sales of the products in the same complementary pair(s) is preferred.

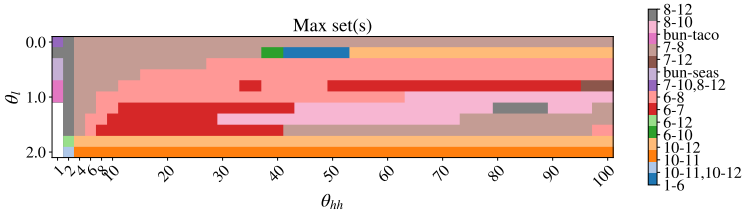


Figure 4.22: The optimal set(s) of products that maximise the overall influence on the signed product network in Fig. 4.18, subject to changing θ_{hh} (x-axis) and θ_l (y-axis) while $\theta_{hl} = \theta_l$ of the GSIP model and $k = 2$, where "bun-seas" corresponds to node pairs "7-11,7-12,8-11,8-12" and "bun-taco" corresponds to node pairs "7-9,7-10,7-11,7-12,8-9,8-10,8-11,8-12".

We then examine the performance of the CDS method. We find that the CDS method can still find a globally optimal or close-to-optimal solution in most cases; see Fig. 4.23. The worse-case scenarios occur when the upper bounds are relatively low while the lower bounds are relatively high. There is one output of accuracy 0.25 with rank 9% in overall 91 node sets, and there are overall 23 cases when the output has accuracy less than 0.5, although they are still at least top 8 in all possible sets. These demonstrate the complexity in the problem, and some heuristics, such as the

community restart strategy we propose in Sec. 3.5.2.2, can be considered to further improve the performance.

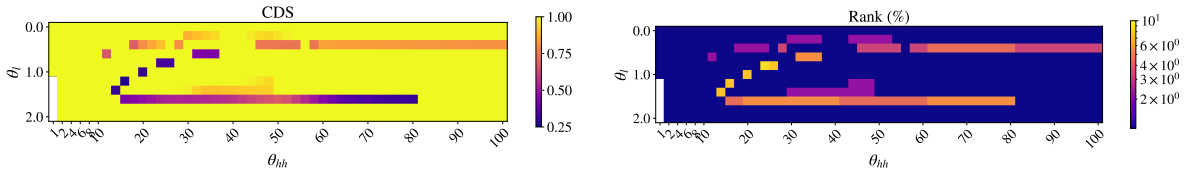


Figure 4.23: Performance of the CDS method on the signed product network in Fig. 4.18 in terms of the output's accuracy (left) and the rank (%, right), subject to changing θ_{hh} (x-axis) and θ_l (y-axis) while $\theta_{hl} = \theta_l$ of the GSIP model and $k = 2$.

Up to now, we have considered the optimisation problem in terms of the overall sales changes in the system. However, it can also be extended for more sophisticated optimisation problems. Take the problem to maximise the overall revenue or profit as an example. To achieve this end, we could include the price or the profit of products as coefficient in computing the overall influence (4.20), where

$$s_j = p_j \sum_{t=1}^{\infty} (1-\gamma)^t x_j(t), \tag{4.42}$$

and p_j indicates the relative importance of node v_j. Then the problem follows similarly, and the proposed framework still works.

4.7 Conclusions

Signed networks have gained more and more popularity over recent years, due to their extensive applications in various domains. However, most results are obtained for unweighted signed networks. Therefore, in this chapter, we first analyse general weighted signed networks in terms of the structural and dynamical properties. We classify signed networks into balanced, antibalanced and strictly unbalanced networks by the distribution of their negative edges, and then characterise the spectral properties of the (weighted) adjacency matrix in each type. We then focus on two classic dynamics that are closely related to information propagation and influence maximisation, the linear dynamics and the extended LT model on signed networks. Considering the problem from the angle of the balanced structure, we are able to show that both dynamics have separating behaviour when the underlying signed networks have different balanced structures, and more importantly, there are consistent

patterns in the two dynamics when they have the same balanced structure, such as the evolution of the sign of state values.

Equipped with the fundamental understanding of signed networks, we then move on to the important problem of information propagation and influence maximisation. As discussed in Chapter 3, the GIP model we propose is able to capture extra features that may occur in real systems, such as the reinforcement among close friends, hence we still consider this model here in this chapter. Specifically, we propose the GSIP model by extending the GIP model from simple networks to signed networks with the "opposing rule", and further show that the GSIP model can still recover the classic dynamics, the extended SIC model on the one hand and the extended SLT model on the other, both of which have been extended to incorporate negative connections. For the associated IM problem, we can still formulate it as a MINLP, and refer to DFMs as general solution methods. We have shown that the exact solution in the linear-dynamics extreme is still closely related to the Katz centrality but extended to signed networks. We have then considered a specific case when there is only one choice of initially activated state value for each node, and provided another CDS method through appropriately modifying the one that we have developed in Sec. 3.4.4 for signed networks, with local convergence. Finally, we numerically illustrate the interesting features of the GSIP model, where both the threshold effect and the locally linear effect can have two opposite directions, and also the close-to-optimal performance of the CDS method.

Even though contextualised in social networks, our work can be generalised to other domains including the retail industry. Specifically, we have shown how our work in the three chapters are closely connected with each other through the illustrative example in Chapter 2. We first construct a signed network from the complementarity and substitutability scores developed in Chapter 2. We then show how the GSIP model that we have developed in this chapter can help retailers compare the overall sales changes of all products from imposing some changes on different pairs of complementary products, subject to different characteristics of the underlying population. Finally, we apply the CDS method to automatically find a pair of products with the maximal change of sales of all products, where the performance is close-to-optimal in most cases. Hence, the whole thesis forms a data-driven approach for sales related decisions, e.g. promotions, on a range of products, from the easily accessible sales transaction data. We believe the whole framework can be useful not only to retailers, or companies with broad product catalogue, but also in other contexts, such as

trading networks, ecological systems and social networks, where the combination of cooperative and competitive relations shapes the dynamical behaviour of the systems.

Chapter 5

Conclusions and future work

As I have discovered during my thesis, industrial problems may inspire fascinating mathematical research questions. With our industrial partner, Tesco, in retail, we have considered the two closely related questions:

(i) to understand the intrinsic relationships between products, including complements and substitutes;

(ii) to analyse the interrelated demand changes between products together with the following decision making.

To answer both questions, we resort to a topological structure emphasising the connection between entities, networks. Specifically, we formulate question **(i)** in terms of bipartite product-purchase networks, where we characterise the two competing product relationships as two different connectivity patterns between product nodes and then formulate the problem as to extract the two competing roles from bipartite networks. We then formulate question **(ii)** in the projected networks of products connected by the two competing types of relationships, where we connect the problem with information propagation and influence maximisation on signed networks.

Inspired by a very specific problem, the methods developed in this thesis are actually part of a broader picture. For instance, in the answer to question **(i)**, the notion of role is an important mesoscale organisation principle, which finds applications in multiple networks. In the literature, most role extraction methods contribute to the problem in terms of groups of nodes sharing generally similar connectivity patterns, directly or indirectly. However, there could be roles of specified characteristics and subtle differences from each other that are particularly interesting in certain contexts, such as the roles corresponding to the complementary and substitute relationships between products. Therefore, it is necessary and important to extend the current role

extraction methods for the roles with predetermined connectivity patterns. A similar statement could be made for backbone extraction methods, aiming at finding significant connections in the projection of bipartite networks. Our method clearly relates to this field of research, but it is tailored to the problem considered in our industrial context.

In Chapter 2, we have developed a customised role extraction method for the two competing roles in bipartite networks. Starting from the characteristics of the specific roles corresponding to the complementary and substitute relationships between products, we first propose several null models to determine the significant relations between the nodes. We then propose measures to quantify the degrees of the two competing types of relationships, motivated by random walks on networks. Finally, we apply community detection algorithms to obtain groups of nodes that are mostly complementary or substitute to each other, i.e. the two competing roles at the mesoscale level. These methods as a whole provide an indirect solution to the specific problem of extracting two competing roles in bipartite networks on the one hand, and also the general problem with respect to roles with specific characteristics on the other hand. We believe that the overall framework we propose, consisting of significance tests with null models, measures based on random walks and community detection algorithms, could also provide insights into the role extraction problem in general.

There are also several directions to explore in the future. Firstly, with the hope to capture more valuable information in networks, it could be interesting to develop a method that directly uncover the degrees of complementarity and substitutability from bipartite networks, without any intermediate steps as it is done here. Secondly, for the practical problem of understanding the product relationships, one could also incorporate the temporal perspective and further explore the connection between the network structure and the cross-elasticity. Finally, the assumption regarding the effect of price could be restrictive for some products, thus it is important to consider price in a meaningful way for future research.

For question **(ii)**, we first consider the answer on networks solely composed of positive connections. To understand how information propagates through social networks is a central theme in social, behavioural, and economic sciences, with important theoretical and practical implications, such as the influence maximisation (IM) problem for viral marketing. Among the vast amount of work in this field, the independent cascade (IC) model and the linear threshold (LT) model are two widely adopted choices of models. However, they have drawbacks in explaining certain features of the information propagation in real systems, e.g. people of influence are more likely

to influence others and there could also be reinforcement within close-contact groups. These limits call for more general models for explaining information propagation, and also more general framework for the associated IM problem.

In Chapter 3, we have proposed a novel class of information propagation model, the GIP model, which unifies the mechanisms underlying the classic models on the one hand, while extending them to address their drawbacks on the other hand. More importantly, we have shown that the GIP model can have features that each single model does not possess but are necessary for real systems. Furthermore, we have introduced mixed integer nonlinear programming (MINLP) to the IM problem, and provided derivative-free methods (DFMs) as general solutions. We have also proposed the customised direct search (CDS) method particularly suited to the IM problem with the GIP model governing the underlying information propagation process, and shown its close-to-optimal performance in various scenarios through experiments. Since the two classic models contain widely accepted mechanisms for information propagation from different perspectives, we believe that the GIP model has the potential to explain more propagation phenomena on, but not restricted to, social networks. Meanwhile, the proposed IM framework provides a systematic approach to handle the case when the objective does not have desired properties, which provides insights into solving the IM problem in more realistic scenarios. For future directions, one could consider investigating the GIP model on real systems, and also fine tuning each step in the CDS method to further improve its performance.

In Chapter 4, we proceed to signed networks with both positive and negative connections, corresponding to the two competing relationships between products. Signed networks have become popular models and tools for many research fields over recent years, since two competing types of node interactions generally co-exist in real systems and they play an equally important role in characterising the behaviour. However, the current development of signed networks is limited from either a structural or a dynamical perspective. Specifically for the information propagation and influence maximisation, most of the work focuses on extending the two classic models to signed networks while maintaining the desired properties in the associated IM problem. However, the expected influence spread is fairly difficult to estimate, and the classic models also suffer from certain drawbacks as discussed before.

Hence, we first characterise general weighted signed networks in terms of both the structural and the dynamical properties. Specifically, we consider the signed networks from the angle of the balanced structure, where we classify them into being balanced, antibalanced or strictly unbalanced. We then distinguish the spectral properties of the

signed (weighted) adjacency matrix in each type, and also show consistent patterns of two interesting dynamics, the linear dynamics and the extended LT model on signed networks. Furthermore, for the information propagation, we propose the GSIP model by extending the GIP model from simple networks to signed networks with the "opposing rule", while maintaining its desired properties as discussed before. For the associated IM problem, we still formulate it as a MINLP and refer to DFMs as general solution methods. We also modify CDS method for the current IM problem when the GSIP model governs the information propagation process, in a specific case when there is only one choice of initially activated value for each node, which could correspond to the practical problem of product range optimisation. We finally demonstrate the interesting feature of the GSIP model and the close-to-optimal performance of the CDS method.

The whole thesis ends with an illustrative example which connects our work in different chapters of the thesis. In this example, we first utilise the method in Chapter 2 to estimate the two competing relationships between products from the simulated sales transaction data. Then a signed network of products connected by both the complementary and the substitute relationships is constructed to further perform the analysis regarding the sales changes, from the work that we have developed in Chapters 3 and 4. Specifically, the GSIP model can be applied to understand the potential consequences in activating, or imposing sales changes on, given node sets, while the IM method can be used to choose an optimal set which will lead to the maximal change of sales or revenue.

As one can notice throughout the thesis, the work is inspired by the retail context, and has several constraints in terms of the type of data we use and the specific problems we consider. However, such constraints have fuelled up our creativity, and led to new insights into more general problems that had been considered in the literature, such as the bipartite backbone extraction, role extraction, information propagation and influence maximisation. Hence, we emphasise at the end that the proposed set of methods contribute to not only general businesses, but also the aforementioned theoretical problems.

Appendix A

Extracting two competing roles in bipartite networks

A.1 Bipartite Configuration Models (BiCMs)

The configuration model creates a network with a given degree sequence $\{d_i\}$, by assigning d_i half-edges (or stubs) to each node $v_i \in V$ and joining two chosen stubs uniformly at random until no more stubs are left [120, 123]. The BiCM takes the bipartite features into account, where two degree sequences are given, dividing the nodes into two subsets, $V_1 = \{u_1, u_2, \ldots, u_{n_1}\}, V_2 = \{v_1, v_2, \ldots, v_{n_2}\}$, s.t. $V_1 \cup V_2 = V$, $V_1 \cap V_2 = \emptyset$, and edges are only allowed between V_1 and V_2. We denote the two given degree sequences $\{d_l^{(1)}\}$ for V_1 and $\{d_i^{(2)}\}$ for V_2. Hence, $\sum_{l=1}^{n_1} d_l^{(1)} = \sum_{i=1}^{n_2} d_i^{(2)} = m$, where m is the number of edges, and n_1, n_2 are the numbers of nodes in V_1 and V_2, respectively.

A.1.1 Probability of edge (u_l, v_i)

Consider one stub of node $u_l \in V_1$, the probability to connect to one of the $d_i^{(2)}$ stubs of node v_i among all m stubs in V_2 is $d_i^{(2)}/m$. Since there are $d_l^{(1)}$ stubs of u_l, if we approximate the connecting process as $d_l^{(1)}$ independent Bernoulli trials with probability $d_i^{(2)}/m$[1], then the probability of nodes u_l and v_i to be connected, i.e. edge

[1]Note the actual probability depends on how many stubs in V_2 are left, and also how many stubs from node v_i remain.

(u_l, v_i), is,

$$p_{li} = 1 - \left(1 - \frac{d_i^{(2)}}{m}\right)^{d_l^{(1)}} \approx 1 - \left(1 - d_l^{(1)}\frac{d_i^{(2)}}{m}\right) = \frac{d_l^{(1)} d_i^{(2)}}{m},$$

where $m \gg 1$, and we assume a cutoff of the node degree, thus the maximum degree is bounded. We will use the approximation in the following analysis since it is appropriate in large networks. Then we note that the probability of an edge between two nodes only depends on their degrees if they are in different subsets (0 otherwise).

Since BiCMs do not exclude multi-edges, it is then important to know how probable it is to obtain a multi-edge. Suppose node $u_l \in V_1$ and $v_i \in V_2$ are already connected, the probability of getting another edge between them is then the probability of an edge between a node of degree $d_l^{(1)} - 1$ and another of degree $d_i^{(2)} - 1$. Hence, the probability of obtaining at least two edges between u_l and v_i is approximately $d_l^{(1)} d_i^{(2)} (d_l^{(1)} - 1)(d_i^{(2)} - 1)/m^2$. Suppose the processes to form multi-edges between every pairs of nodes are independent Bernoulli trials with possibly different probabilities, then the expected number of pairs with multi-edges is approximately

$$\sum_{l=1}^{n_1}\sum_{i=1}^{n_2} \frac{d_l^{(1)} d_i^{(2)} (d_l^{(1)} - 1)(d_i^{(2)} - 1)}{m^2} = \left(\frac{\langle d^{(1)2}\rangle - \langle d^{(1)}\rangle}{\langle d^{(1)}\rangle}\right)\left(\frac{\langle d^{(2)2}\rangle - \langle d^{(2)}\rangle}{\langle d^{(2)}\rangle}\right),$$

where $\langle d^{(z)q}\rangle = \sum_{i=1}^{n_z} d_i^{(z)q}/n_z, z \in \{1, 2\}$, is the q-th moment of the degree sequence $\{d_i^{(z)}\}$. Hence the number of pairs having multi-edges approximately stays constant as long as the moments are constant and finite, and will be negligible if the network is sufficiently large.

A.1.2 Common neighbours

The common-neighbour pattern is important in characterising nodes in bipartite networks, hence we now consider the number of common neighbours between nodes v_i and v_j in V_2, $cn_{ij}^{(2)}$, and that between nodes u_h and u_l in V_1, $cn_{hl}^{(1)}$, follows naturally.

For a node $u_l \in V_1$, we know the probabilities of the edges (u_l, v_j) and (u_l, v_i), but if u_l is already connected with v_j, the probability to also connect with v_i will be approximately $(d_l^{(1)} - 1)d_i^{(2)}/m$. Hence, the probability of v_i, v_j sharing a common neighbour u_l is then approximately

$$p_{ilj}^{(2)} = \frac{d_l^{(1)} d_j^{(2)} (d_l^{(1)} - 1) d_i^{(2)}}{m^2},$$

Since v_i, v_j can have any node in V_1 as their common neighbour, if we consider the whole process as n_2 independent Bernoulli trials with possibly different probabilities, $cn_{ij}^{(2)}$ is then a Poisson binomial random variable, with the expected value approximately

$$\mu_{ij}^{(2)} = \sum_{l=1}^{n_1} \frac{d_l^{(1)} d_j^{(2)} (d_l^{(1)} - 1) d_i^{(2)}}{m^2} = \frac{d_i^{(2)} d_j^{(2)}}{m} \frac{\langle d^{(1)2} \rangle - \langle d^{(1)} \rangle}{\langle d^{(1)} \rangle}.$$

Similarly for u_h, u_l in V_1, $cn_{hl}^{(1)}$ is a Poisson binomial random variable with the mean value approximately

$$\mu_{hl}^{(1)} = \sum_{i=1}^{n_2} \frac{d_h^{(1)} d_i^{(2)} d_l^{(1)} (d_i^{(2)} - 1)}{m^2} = \frac{d_h^{(1)} d_l^{(1)}}{m} \frac{\langle d^{(2)2} \rangle - \langle d^{(2)} \rangle}{\langle d^{(2)} \rangle}.$$

Hence the expected number of common neighbours is dependent on the degrees of both nodes, and the moments of the degree sequence of the other subset, approximately.

A.2 Poisson approximation

The Poisson binomial distribution is the discrete probability distribution of a sum of independent Bernoulli random variables that are not necessarily identically distributed, i.e. with parameters $p_1, p_2, ..., p_n$ that are possibly different.

Since a Poisson binomial random variable is the sum of n independent Bernoulli distributed variables, its mean and variance are simply the sums of the means and the variances of the corresponding Bernoulli distributions, respectively, i.e. $\mu = \sum_{i=1}^{n} p_i$, $\sigma^2 = \sum_{i=1}^{n} p_i(1 - p_i)$.

The distribution of a Poisson binomial random variable X can be approximated by that of a Poisson random variable, $Pois(\lambda)$, with the same mean value $\lambda = \mu$, where Barbour *et al.* [13] showed that

$$d_{TV}(\mathcal{L}(X), Pois(\mu)) \leq \min(1, \mu^{-1}) \sum_{i=1}^{n} p_i^2 \tag{A.1}$$

where $d_{TV}(\cdot, \cdot)$ is the total variation distance between two distributions, $X = \sum_i^n X_i$ is the Poisson binomial random variable with $X_i \sim Bernoulli(p_i)$, and $\mathcal{L}(X)$ is the distribution of X.

In our case, the composing probability is p_{ilj}, the likelihood for each pair of product nodes v_i and v_j to both connect with transaction node u_l in the BiCM, and is

approximately

$$p_{ilj} = \frac{d_i^{(2)} d_l^{(1)} d_j^{(2)} (d_l^{(1)} - 1)}{m^2},$$

where $d_i^{(2)}$ is the degree of product node v_i, $d_l^{(1)}$ is the degree of transaction node u_l, and m is the number of edges in the bipartite network. We can then evaluate the error bounds to determine whether the Poisson approximation is appropriate. For the sales data, we consider $\omega_{ij} = \min(1, \mu^{-1}) \sum_l p_{ilj}^2$, i.e., the error bound in (A.1), and find that the maximum value of ω_{ij}s is around 0.09, with more than 99.5% pairs of $\omega_{ij} \leq 0.01$, 98.6% pairs of $\omega_{ij} \leq 0.005$ and more than 89.3% pairs of $\omega_{ij} \leq 0.001$, thus the Poisson approximation is guaranteed to perform well. Hence, we apply the Poisson approximation in the corresponding experiments.

A.3 Measures

The original measure (2.5) is derived from the weighted cosine similarity between random walkers starting from each pair of nodes after one step. Specifically, for each node $v_i \in V_2$, suppose that an impulse $\mathbf{y}_i(0) = \mathbf{e}_i \in \{0, 1\}^{n_p}$, with value 1 only in its i-th element, is injected on the V_2 side at time $t = 0$. We record the response of the system after a one-step random walk $\mathbf{y}_i(1) = \mathbf{P}^T \mathbf{y}_i(0)$, where $\mathbf{P} = \mathbf{D}^{(2)-1} \mathbf{A}^{(b)T}$, $\mathbf{A}^{(b)} = (A_{li}^{(b)})$ is the biadjacency matrix, and $\mathbf{D}^{(2)} = \mathbf{Diag}(d_i^{(2)})$ is the diagonal matrix with the degrees of nodes in V_2 on its diagonal [134]. We set the relative importance of each node $v_l \in V_1$ as the inverse of its degree $d_l^{(1)}$, hence the weighted cosine similarity between the responses $\mathbf{y}_i(1)$ and $\mathbf{y}_j(1)$ corresponding to nodes $v_i, v_j \in V_2$,

respectively, is

$$
\begin{aligned}
sim_o(i,j) &= \frac{\mathbf{y}_i(1)^T \mathbf{W}_{cos} \mathbf{y}_j(1)}{||\mathbf{y}_i(1)||_{\mathbf{W}_{cos}} ||\mathbf{y}_j(1)||_{\mathbf{W}_{cos}}} \\
&= \sum_{l=1}^{n_t} \frac{\frac{A_{li}^{(b)}}{\sum_{k=1}^{n_t} A_{ki}^{(b)}} \frac{1}{d_l^{(1)}} \frac{A_{lj}^{(b)}}{\sum_{k=1}^{n_t} A_{kj}^{(b)}}}{\sqrt{(\sum_{h=1}^{n_t} \frac{A_{hi}^{(b)}}{\sum_{k=1}^{n_t} A_{ki}^{(b)}} \frac{1}{d_h^{(1)}} \frac{A_{hi}^{(b)}}{\sum_{k=1}^{n_t} A_{ki}^{(b)}})(\sum_{h=1}^{n_t} \frac{A_{hj}^{(b)}}{\sum_{k=1}^{n_t} A_{kj}^{(b)}} \frac{1}{d_h^{(1)}} \frac{A_{hj}^{(b)}}{\sum_{k=1}^{n_t} A_{kj}^{(b)}})}} \\
&= \sum_{l=1}^{n_t} \frac{A_{li}^{(b)} A_{lj}^{(b)}}{d_i^{(2)} d_l^{(1)} d_j^{(2)} \sqrt{(\sum_{h=1}^{n_t} \frac{A_{hi}^{(b)}}{d_i^{(2)2} d_h^{(1)}})(\sum_{h=1}^{n_t} \frac{A_{hj}^{(b)}}{d_j^{(2)2} d_h^{(1)}})}} \\
&= \sum_{l=1}^{n_t} \frac{A_{li}^{(b)} A_{lj}^{(b)}}{d_l^{(1)} \sqrt{(\sum_{h=1}^{n_t} \frac{A_{hi}^{(b)}}{d_h^{(1)}})(\sum_{h=1}^{n_t} \frac{A_{hj}^{(b)}}{d_h^{(1)}})}},
\end{aligned}
$$

where $\mathbf{W}_{cos} = \mathbf{Diag}(1/d_l^{(1)})$ is the weight matrix for the cosine similarity, and $||\mathbf{y}||_{\mathbf{W}} = \sqrt{\mathbf{y}^T \mathbf{W} \mathbf{y}} = ||\mathbf{W}^{1/2}\mathbf{y}||_2$ with $\mathbf{W}$ (symmetric) positive-definite.

A.4 Sales data

We now give details of how to calibrate the parameters, including the significance levels, α_m and α_l, and the thresholds, θ_c for $\mathbf{W}^{(c)}$ and θ_s for $\mathbf{W}^{(s)}$. We interpret the criterion of maintaining the same community structure as the NMI between the two underlying partitions being greater than 0.8, and use the map equation to detect the communities [132]. This value is chosen for sufficient consistency but limited freedom of variance between partitions [92]. Note the variant of bipartite ER models is used as the underlying null model, and we choose 0.05 as the baseline significance level in both cases. We select the thresholds θ_c, θ_s by the quantiles of the nonzero values, q_c, q_s, and use $0, 0.7$ as the baseline threshold quantiles for $\mathbf{W}^{(c)}, \mathbf{W}^{(s)}$, respectively. Thus by applying the extra rules of significance-level and threshold selection, we obtain $\alpha_m = 0.01$, $\alpha_l = 0.2$, $q_c = 0.35$ with $\theta_c = 0.011$ and $q_s = 0$, for the scores induced by the original measure; see Fig. A.1. The results for the scores induced by the randomised configuration measure are exactly the same as the above ones, (which can be inferred from the score values being very close to each other, as in Table A.1) thus is omitted here. We clarify here that the complementarity scores $\mathbf{W}^{(c)}$, and the substitutability scores $\mathbf{W}^{(s)}$, only refer to the scores after thresholding.

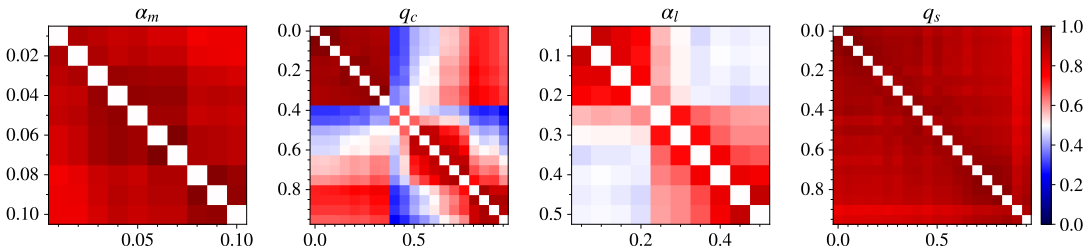


Figure A.1: Pairwise NMI between the partitions of $\mathbf{W}^{(c)}$ varying α_m (**leftmost**) and varying the threshold quantile q_c with $\alpha_m = 0.01$ (**middle-left**), and of $\mathbf{W}^{(s)}$ varying α_l with $\alpha_m = 0.01, q_c = 0.35$ (**middle-right**) and varying the threshold quantile q_s with $\alpha_m = 0.01, q_c = 0.35, \alpha_l = 0.2$ (**right-most**), where the original measure is used, and both x- and y-axis labels are shown in the corresponding titles.

We then show the pairwise rankings from the scores induced by the randomised configuration measure. Comparing with Table 2.1, we find that the results from the randomised configuration measure agree with those from the original measure. Plus, the score values induced by both measures are very close to each other, which is why they have exactly the same community structure in the parameter calibration phase, as mentioned before.

Table A.1: Products of the three highest complementarity scores and substitutability scores with the query products, where the scores are induced by the randomised configuration measure.

Query product	Complement		Substitute	
Organic Blueberries	0.14	Organic Raspberries	0.50	Blueberries
	0.058	Organic Strawberries	0.13	Green Seedless Grapes
	0.047	Organic Cherry Tomatoes	0.070	Tomatoes on the Vine
Loose Cucumbers	0.095	Salad Tomatoes	0.54	Organic Loose Cucumbers
	0.085	Baby Plum Tomatoes	0.21	Courgette Spaghetti
	0.077	Tomatoes on the Vine	0.18	Sliced Runner Beans
Salad Tomatoes	0.095	Loose Cucumbers	0.83	Tomatoes on the Vine
	0.062	Iceberg Lettuce	0.79	Baby Plum Tomatoes
	0.045	Mixed Peppers	0.73	Cherry Tomatoes

A.5 Robustness

The set of methods we have developed in Chapter 2 for extracting the product relationships from sales transaction data is robust to temporal shifts, on the condition

that the underlying customers maintain their purchase habits during different time periods, i.e. they keep treating certain products as complements or substitutes. It is not necessarily true for every time period, and from our industrial collaborators, customers do change behaviours over time. Hence, it is a promising direction to further incorporate the temporal features of the proposed complementarity/substitutability scores, e.g. how the scores change over time, as mentioned in Sec. 2.6.

For our current methods, it is also interesting to explore how the proposed scores change within the three-month period we have chosen. Hence, (i) we split the three-month sales transaction dataset into two one-month-and-a-half sales transaction datasets; (ii) we then compare the results from the two datasets, in terms of the score values and the extracted role structures. Note that we now have a shorter time period, thus noise will play a relatively larger role in the analysis. Hence, we further filter out the products that are sold less frequently than once a week, which brings our analysis down to 169 products, and 19,498 and 19,670 transactions in splits 1 and 2, respectively. As in Sec. 2.5.2, we use the variant of ER model as the underlying null model, and compare the results from the original measure, with the same choice of reference significance level (0.05 in both cases) and reference threshold quantiles ($0, 0.7$ for the degrees of complementarity and substituteability, respectively).

Split	α_m	q_c (θ_c)	α_l	q_s
1	0.01	0.35 (0.016)	0.25	0
2	0.01	0.30 (0.014)	0.25	0

Table A.2: Parameters chosen for the customised role extraction method on splits 1 and 2.

In general, it is true in both complementarity and substitutability scores that if the values are larger than 0 in both splits of the data, they are close to each other, see Fig. A.2 where points $\{(x, y) : x > 0, y > 0\}$ approach the identity line. However, there are also many points have $x = 0, y > 0$, or $y = 0, x > 0$, i.e. the relationship is significant in one split but not in the other. Since the same significance levels are chosen for both splits (see Table A.2), the insignificance from the other split stems from the change in customer behaviour. Note that the substitutability scores are computed through the complementarity scores, thus there is also part of the points lying in either $x = 0$ or $y = 0$ obtained from sharing no complements[2], which further caused by the changing behaviour of customers.

[2]There are 59.2% of points $\{(x, y) : x = 0, y > 0\}$ and 49.1% of points $\{(x, y) : x > 0, y = 0\}$ fall in this category.

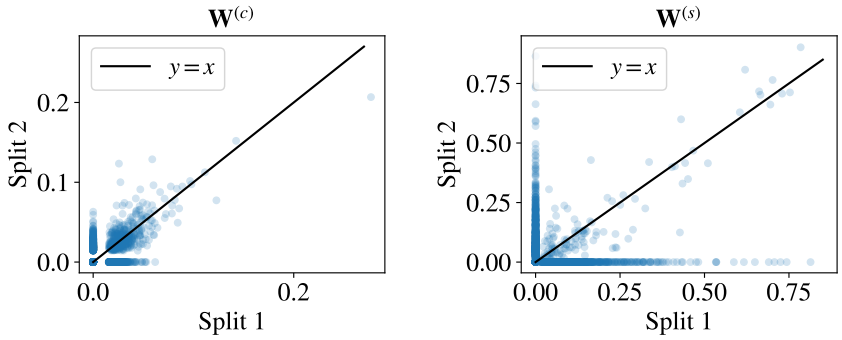


Figure A.2: Scatter plots of the complementarity scores (left) and the substitutability scores (right) from split 1 (x-axis) and 2 (y-axis), where the identity line $y = x$ is shown in black for reference.

Further to quantify how far the score values in one split are from the other, we define the *relative distance* D between the scores $\{X_{ij}^{(1)}\}$ in split 1 and $\{X_{ij}^{(2)}\}$ in split 2 to be,

$$D = \frac{1}{n_p(n_p-1)} \sum_{i \neq j} D_{ij}, \quad \text{where } D_{ij} = \frac{\left| X_{ij}^{(1)} - X_{ij}^{(2)} \right|}{mean(X_{ij}^{(1)}, X_{ij}^{(2)})},$$

and n_p is the number of product nodes. Here, we use the arithmetic mean, $mean(x, y) = (x + y)/2$. The resulting relative distance of the complementarity and the substitutability scores are 0.13 and 0.17, respectively, thus both are relatively small.

Finally, we compare the roles extracted from both splits. The complement roles from the two splits have NMI 0.74 and AMI 0.62, which indicates that the complement role structures from the two datasets largely agree with each other. While for the substitute roles obtained from the two sources, they have NMI 0.58 but AMI 0.16, where the difference between the two values is from the relatively large number of roles. Hence, there is slightly more change in the substitute role structures, which is expected since it also carries the changes from the complementary relationship.

To conclude, in our chosen period, the temporal shifts will cause a certain amount of change in the complementary and substitutability score values, but will not cause dramatic change in the mesoscale role structure. Hence, as we mentioned in the very beginning, analysing the temporal change of the scores an interesting topic to explore in the future.

A.6 Formulation as unsupervised learning

The sales data can be seen as n_t observations $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_t})$ of a random n_p-dimensional vector $\mathbf{X} = (X_1, X_2, \ldots, X_{n_p})$ having joint density $\mathbb{P}(\mathbf{X})$. The goal of unsupervised learning is to directly infer the properties of this probability density without the help of labels or degree-of-error for each observation [76]. Our goal here is specifically to find (i) (complements) the item set $\mathcal{J}_i^{(c)} \subset \{1, 2, ..., n_p\}$ for each product i, s.t. $\forall j \in \mathcal{J}_i^{(c)}$,

$$\mathbb{P}(X_j = 1 | X_i = 1), \tag{A.2}$$

is significantly large; (ii) (substitutes) the item set $\mathcal{J}_i^{(s)} \subset \{1, 2, ..., n_p\}$ for each product i, s.t. $\forall j \in \mathcal{J}_i^{(s)},\ \exists j' \in \mathcal{J}_i^{(c)}$,

$$\mathbb{P}(X_j = 1 | X_i = 0, X_{j'} = 1), \tag{A.3}$$

is significantly large.

Instead of approximating the probabilities (A.2) and (A.3) by the fractions of the corresponding observations in the data, which is not straightforward for the latter, we take a network perspective. In the product-purchase network, for an upcoming transaction node $u_{l'}$,

$$\mathbb{P}(X_j = 1 | X_i = 1) = \mathbb{P}(A_{l'j}^{(b)} = 1 | A_{l'i}^{(b)} = 1) = \frac{\mathbb{P}(A_{l'j}^{(b)} = 1, A_{l'i}^{(b)} = 1)}{\mathbb{P}(A_{l'i}^{(b)} = 1)},$$

and similarly,

$$\begin{aligned}\mathbb{P}(X_j = 1 | X_i = 0, X_{j'} = 1) &= \mathbb{P}(A_{l'j}^{(b)} = 1 | A_{l'i}^{(b)} = 0, A_{l'j'}^{(b)} = 1) \\ &= \frac{\mathbb{P}(A_{l'j}^{(b)} = 1, A_{l'j'}^{(b)} = 1) - \mathbb{P}(A_{l'j}^{(b)} = 1, A_{l'i}^{(b)} = 1, A_{l'j'}^{(b)} = 1)}{\mathbb{P}(A_{l'i}^{(b)} = 0, A_{l'j'}^{(b)} = 1)}.\end{aligned}$$

Then, to find the corresponding sets, we restrict set $\mathcal{J}_i^{(c)}$ to search values in

$$\mathcal{I}_i^{(c)} = \{j : \mathbb{P}(A_{l'j}^{(b)} = 1, A_{l'i}^{(b)} = 1) > \alpha_m\},$$

i.e. products i and j are bought together significantly more frequently with significance

level α_m (feature **(1)** in Sec. 2.3.2); we also restrict set $\mathcal{J}_i^{(s)}$ to take values in

$$\mathcal{I}_i^{(s)} = \{j : \mathbb{P}(A_{l'j}^{(b)} = 1, A_{l'i}^{(b)} = 1) < \alpha_l,\ \mathbb{P}(A_{l'j}^{(b)} = 1, A_{l'j'}^{(b)} = 1) > \alpha_m,\ \exists j' \in \mathcal{I}_i^{(c)}\},$$

i.e. products i and j are bought together significantly less frequently with significance level α_l, but more frequently with some products in $\mathcal{I}_i^{(c)}$ (feature **(3)** in Sec. 2.3.2).

Next, we propose the measures to estimate the conditional probabilities. Based on random walks on networks, the estimated value of (A.2) is taken to be the original directed measure (2.6), where $\forall j \in \mathcal{I}_i^{(c)}$,

$$\hat{\mathbb{P}}(A_{l'j}^{(b)} = 1 | A_{l'i}^{(b)} = 1) = sim_{od}(j, i) = \frac{\sum_{l=1}^{n_t} \frac{A_{lj}^{(b)} A_{li}^{(b)}}{d_l^{(1)}}}{\sum_{h=1}^{n_t} \frac{A_{hi}^{(b)}}{d_h^{(1)}}}. \tag{A.4}$$

We also modify the conditional probability to provide a symmetric version

$$\mathbb{P}_m(j, i) = \frac{\mathbb{P}(A_{l'j}^{(b)} = 1, A_{l'i}^{(b)} = 1)}{\sqrt{\mathbb{P}(A_{l'i}^{(b)} = 1)\mathbb{P}(A_{l'j}^{(b)} = 1)}},$$

and approximate it by the original measure (2.5), where

$$\hat{\mathbb{P}}_m(j, i) = sim_o(j, i) = \frac{\sum_{l=1}^{n_t} \frac{A_{lj}^{(b)} A_{li}^{(b)}}{d_l^{(1)}}}{\sqrt{(\sum_{h=1}^{n_t} \frac{A_{hi}^{(b)}}{d_h^{(1)}})(\sum_{h=1}^{n_t} \frac{A_{hj}^{(b)}}{d_h^{(1)}})}}. \tag{A.5}$$

We also propose their randomised versions in order to remove specific noise effects from our estimates; see Eqs. (2.8) and (2.7) for examples.

The step to restrict the set of possible items is particularly important for estimating conditional probability (A.3), since we will remove the condition $X_i = 0$, and instead approximate $\mathbb{P}(A_{lj}^{(b)} = 1 | A_{lj'}^{(b)} = 1)$, $\forall j' \in \mathcal{I}_i^{(c)}$, in an integrated way. The measure we propose directly for this purpose is the directed substitutability measure, where $\forall j \in \mathcal{I}_i^{(s)}$,

$$sim_{sd}(j, i) = \tag{A.6}$$
$$\frac{\sum_{j' \in \mathcal{I}_j^{(c)} \cap \mathcal{I}_i^{(c)}} \min(\hat{\mathbb{P}}(A_{l'j}^{(b)} = 1 | A_{l'j'}^{(b)} = 1), \hat{\mathbb{P}}(A_{l'i}^{(b)} = 1 | A_{l'j'}^{(b)} = 1)) \hat{\mathbb{P}}(A_{l'i}^{(b)} = 1 | A_{l'j'}^{(b)} = 1)}{\sum_{p \in \mathcal{I}_i^{(c)}} \hat{\mathbb{P}}(A_{l'i}^{(b)} = 1 | A_{l'p}^{(b)} = 1)^2}.$$

Note the $\hat{\mathbb{P}}(\cdot|\cdot)$ in Eq. (A.6) is an estimate of the probability in (A.2), and we can consider, for example, estimates (A.4) and (A.5). We also modify the normalisation method to propose a symmetric version – the substitutability measure, where

$$sim_s(j,i) = \frac{\sum_{j' \in \mathcal{I}_j^{(c)} \cap \mathcal{I}_i^{(c)}} \hat{\mathbb{P}}(A_{l'j}^{(b)} = 1 | A_{l'j'}^{(b)} = 1) \hat{\mathbb{P}}(A_{l'i}^{(b)} = 1 | A_{l'j'}^{(b)} = 1)}{\sqrt{(\sum_{p \in \mathcal{I}_i^{(c)}} \hat{\mathbb{P}}(A_{l'i}^{(b)} = 1 | A_{l'p}^{(b)} = 1)^2)(\sum_{p \in \mathcal{I}_j^{(c)}} \hat{\mathbb{P}}(A_{l'j}^{(b)} = 1 | A_{l'p}^{(b)} = 1)^2)}}. \quad \text{(A.7)}$$

Now with the item sets and the corresponding estimates, we can output $\mathcal{J}_i^{(c)}$ and $\mathcal{J}_i^{(s)}$ for each product i, with any threshold/lower bound of the estimated probabilities. From our analysis of the sales data, 0.01, 0.2 are reasonable values of α_m and α_l, respectively, and 0.01, 0 are sensible thresholds for the estimations (A.5) and (A.7), respectively. Following the path of network analysis, we can treat this process as a bipartite projection, and further analyse the community structure of the relationships determined by these probabilities, i.e. complements and substitutes, as in Sec. 2.5.2.

Appendix B

Unifying dynamics on networks and optimisation

B.1 Time complexity of the proposed method

In this section, we discuss the time complexity of the CDS method that we proposed in Sec. 3.4.4, through decomposing it into two parts: (I) the evaluation of the objective function, and (II) the number of evaluations required until convergence. From Theorem 4, for a given network, $O(|E|)$ is fixed, and then the worst-case scenario of (I) corresponds to the highest possible value t_ϵ could take. We know that the state values from the GIP model cannot exceed those obtained from the linear dynamics, thus the linear dynamics correspond to an achievable upper bound for t_ϵ. Hence, the CDS method circumvents the worst-case complexity by starting from the solution associated with the linear dynamics. For (II), we further decompose it into the product of (i) the size of discrete neighbourhood that is feasible, and roughly, (iia) the number of steps towards convergence, or more precisely, (iib) the number of evaluations divided by the neighbourhood size. (i) is $k(n-k)$, by Eqs. (3.31) and (3.32) with the choice of $d = 2$, and is equivalently $O(n)$ since k is normally assumed to be $O(1)$. The only remaining part for a full description of the time complexity is (iia) or (iib), both are related to the rate of convergence which is difficult to provide a theoretical guarantee given the general properties of the objective function of the MINLP (3.24). Instead, we conjecture that the complexity of (iib) does not increase significantly when varying the network size n, i.e. approximately $O(1)$, given the same mean degree of nodes, and verify it empirically through experiments on ER random graphs.

Specifically, we construct ER random graphs of the same expected node degree, $< d >= (n-1)p$ where n is the network size and p is the connecting probability in ER random graphs, but of increasing size n. We then apply the CDS method to the IM problem on this series of networks to explore the dependence of its time complexity on the network size n. The experiments are performed with different combinations of parameters in order to take into account their effects, including the upper bounds, the lower bounds, and the budget size k. Here, we assign a uniform edge weight $\alpha = 0.1$ to all the networks, set $l_{j,0} = h_{j,0} = 1$, $\forall v_j \in V$, $\gamma = 0$, and apply exclusively the threshold-type bounds in (3.13) with condition (3.14), as in Sec. 3.5.

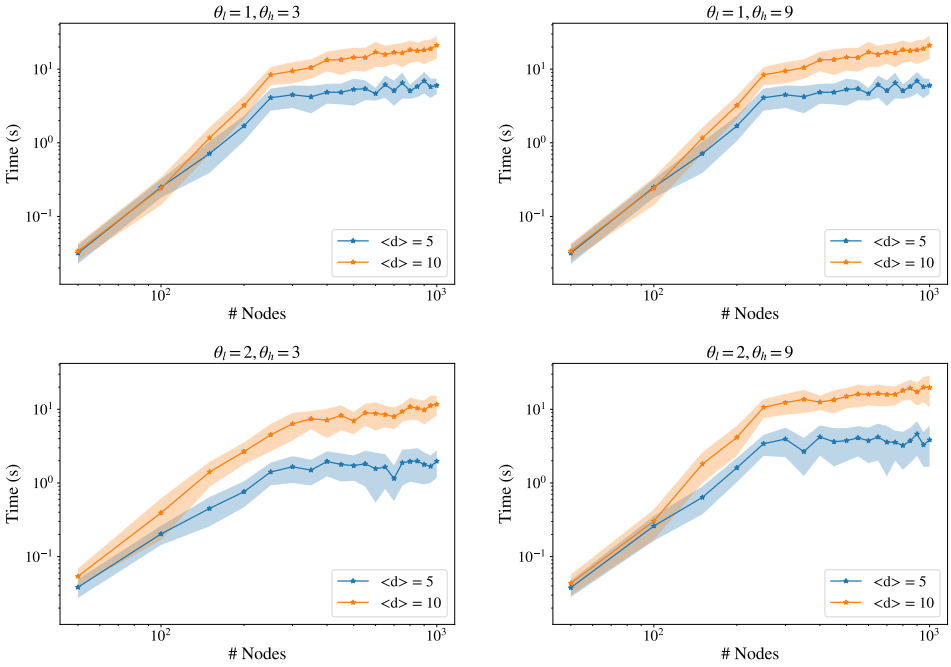


Figure B.1: Time consumption of the CDS method on ER random graphs of increasing sizes (x-axis) when $k = 3$, where we consider different combinations of parameters (θ_l, θ_h) while maintaining the same mean degree 5 (blue) and 10 (orange), and the results are obtained from 50 samples of the ER graphs of each size, with dots showing the mean value and shades indicating the standard deviation.

Overall, the time increases linearly when the network is small, and once the network size reaches a critical value, the increase slows down; see Fig. B.1 for $k = 3$ and Fig. B.2 for $k = 5$, with different combinations of parameters (θ_l, θ_h). Having different mean degree, upper bounds, lower bounds, or budget sizes normally causes shifts in the trends but not the overall shape.

We now integrate the experiment results with the theoretical understanding. For (I), we know from Alg. 1 that only neighbours of currently active nodes are considered

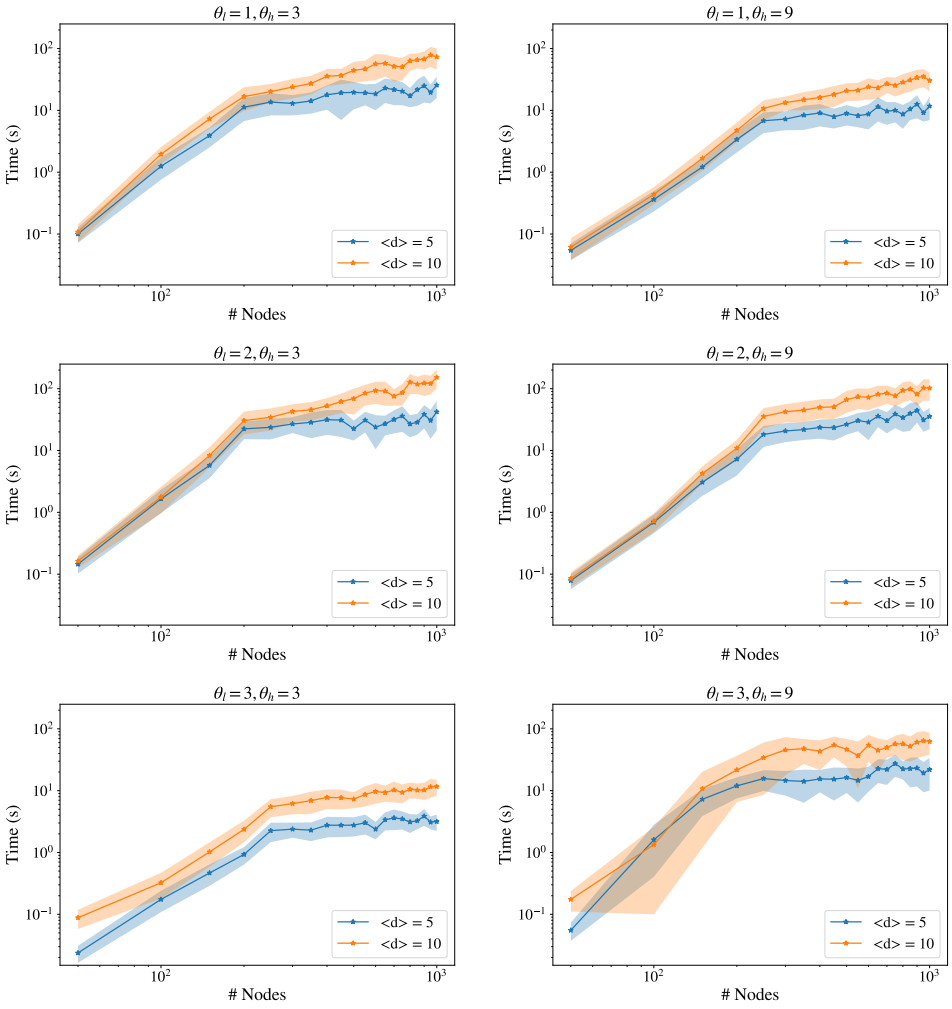


Figure B.2: Time consumption of the CDS method on ER random graphs of increasing sizes (x-axis) when $k = 5$, where we consider different combinations of parameters (θ_l, θ_h) while maintaining the same mean degree 5 (blue) and 10 (orange), and the results are obtained from on 50 samples of the ER graphs of each size, with dots showing the mean value and shades indicating the standard deviation.

when calculating the state value in the next step. Since we maintain the same mean degree in our setting, the difference in time lies in t_ϵ which is at least $O(1)$. For (iia), the size of the feasible discrete neighbourhood $k(n-k)$ increases linearly in n, and we also note that when $\theta_l > 1$, to include nodes that share common neighbours with other initially activated nodes is the only way to improve the overall influence (on networks with uniform edge weights), which is also applied in the algorithm (in refining the mesh at each time step) to reduce the complexity to be lower than $O(n)$. From the experiments, we observe linear increase, and also that the increase slows down significantly after certain critical value of n. Therefore, it is reasonable to conjecture that (iib), the number of evaluations required by the CDS method divided by the feasible neighbourhood size, is approximately $O(1)$ in the IM problem in real networks.

These all together result in the value of (II) to be approximately $O(n)$. To measure the goodness of this complexity value, we compare it with a global algorithm, the brute-force method, where each node set will be evaluated. Then, overall $\binom{n}{k}$ evaluations are needed, which is proportional to $O(n^k)$ since again $k = O(1)$. Therefore, $O(n^{k-1})$ more multiples of evaluations are needed in the global algorithm than the CDS method.

The time complexity of the CDS method can be further improved by parallel programming, although it is described in Alg. 3 and currently performed serially. For instance, the tasks to evaluate the objective function of different candidates in a discrete neighbourhood can run in parallel. There are also other optimisation techniques to further reduce the complexity, e.g. coarse evaluation of the objective function with a higher tolerance at early stage. All these can be included in the future work.

B.2 Stochastic block models (SBMs)

In this section, we discuss the properties of SBMs mentioned in the Chapter 3 in more detail. We first show that when increasing the upper bounds in the GIP model, the increase in the expected influence (in one time step) from the node set in one community is higher than that from the other evenly distributed in the two communities, as in Sec. 3.3.3.2, and then prove that a node is expected to have higher Katz centrality if it has higher degree centrality than others, as in Sec. 3.5.2.3.

Claim 5. *With $l_{j,0} = h_{j,0} = l_0$, and $l_{j,1} = 2\alpha l_0,\ \forall v_j \in V$ in the GIP model, if $SBM(p_{in}, p_{out})$ of two equally-sized* [1] *communities, $\mathcal{B}_1, \mathcal{B}_2$, and uniform edge weight α, while allowing self-loops, satisfies*

$$p_{in} > p_{out},\ (1 - p_{in}) > p_{out}, \tag{B.1}$$

or

$$p_{out} > p_{in},\ (1 - p_{out}) > p_{in}, \tag{B.2}$$

then the increase in the expected influence, $\mathbb{E}[\sum_j (1-\gamma)x_j(1)]$, from the initially activated node set (i) $\mathcal{A}_0 = \{v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}\} \subset \mathcal{B}_1$, is larger than that from (ii) $\mathcal{A}_0 = \{v_{j_1}, v_{j_2}, v_{j_3}, v_{j_4}\}$ with $v_{j_1}, v_{j_2} \in \mathcal{B}_1, v_{j_3}$ and $v_{j_4} \in \mathcal{B}_2$, when $h_{j,1}$ rises from $h_{j,1} = l_{j,1} = 2\alpha l_0$ to $h_{j,1} = 2l_{j,1} = 4\alpha l_0,\ \forall v_j \in V$.

Proof. For the SBM, $\mathbf{W} = \alpha\mathbf{A}$, where $\mathbf{A}$ is the (unweighted) adjacency matrix. We maintain the same notations for the block membership of each node v_i, $\sigma_i \in \{1, 2\}$, the linear part of the state vector, $\mathbf{y}(t+1) = \mathbf{W}^T\mathbf{x}(t)$, the size of each community n_b, and the binomial random variable $\zeta_{n_m,p} \sim Bin(n_m, p)$, as in the proof of Claim 1.

In case (i),

$$y_j(1) = \alpha l_0(\zeta_{4,p_{in}}\delta(\sigma_j, 1) + \zeta_{4,p_{out}}\delta(\sigma_j, 2)),$$

while in case (ii),

$$y_j(1) = \alpha l_0(\zeta_{2,p_{in}} + \zeta_{2,p_{out}}).$$

When $h_{j,1} = l_{j,1} = 2\alpha l_0,\ \forall v_j \in V$,

$$\mathbb{E}\left[\sum_j x_j(1)\right] = \sum_j 0P(y_j(1) < l_{j,1}) + 2\alpha l_0 P(y_j(1) \geq 2\alpha l_0). \tag{B.3}$$

[1] Note the same results can be obtained with arbitrary community sizes, but extra conditions on both the community sizes and the probabilities are required.

When $h_{j,1} = 4\alpha l_0,\ \forall v_j \in V$,

$$\begin{aligned}\mathbb{E}\left[\sum_j x_j(1)\right] = \sum_j \big(&0P(y_j(1) < l_{j,1}) + 2\alpha l_0 P(y_j(1) = 2\alpha l_0) \\ &+ 3\alpha l_0 P(y_j(1) = 3\alpha l_0) + 4\alpha l_0 P(y_j(1) = 4\alpha l_0)\big)\,.\end{aligned} \tag{B.4}$$

Hence, the increase in the expected influence, ignoring the factor $1-\gamma$, is the difference between the two equations (B.4) and (B.3), i.e.

$$\Delta = \sum_j \alpha l_0 P(y_j(1) = 3\alpha l_0) + 2\alpha l_0 P(y_j(1) = 4\alpha l_0).$$

In case (i),

$$\begin{aligned}\Delta_{(i)} &= \sum_j \alpha l_0 \left(\binom{4}{3} p_{in}^3 (1-p_{in})\delta(\sigma_j, 1) + \binom{4}{3} p_{out}^3 (1-p_{out})\delta(\sigma_j, 2)\right) \\ &\quad + 2\alpha l_0 \left(p_{in}^4 \delta(\sigma_j, 1) + p_{out}^4 \delta(\sigma_j, 2)\right) \\ &= n_b \times \alpha l_0 \times 2\left(2p_{in}^3(1-p_{in}) + 2p_{out}^3(1-p_{out}) + p_{in}^4 + p_{out}^4\right).\end{aligned}$$

While in case (ii),

$$\begin{aligned}\Delta_{(ii)} &= \sum_j \alpha l_0 \left(\binom{2}{1} p_{in}(1-p_{in}) p_{out}^2 + p_{in}^2 \binom{2}{1} p_{out}(1-p_{out})\right) + 2\alpha l_0 \left(p_{in}^2 p_{out}^2\right) \\ &= 2n_b \times \alpha l_0 \times \left(2p_{in}^2 p_{out}(1-p_{out}) + 2p_{out}^2 p_{in}(1-p_{in}) + 2p_{in}^2 p_{out}^2\right).\end{aligned}$$

Hence,

$$\Delta_{(i)} - \Delta_{(ii)} = 2n_b \alpha l_0 \left(2\left(p_{in}^2 - p_{out}^2\right)\left(p_{in}(1-p_{in}) - p_{out}(1-p_{out})\right) + \left(p_{in}^2 - p_{out}^2\right)^2\right),$$

which is positive given conditions (B.1) or (B.2). □

Theorem 20. *In a two-block SBM with the connecting probabilities in the two communities, $\mathcal{B}_1, \mathcal{B}_2$, being p_1, p_2, respectively, and between the two being p_{12}, while allowing self-loops, if a node v_i has higher expected degree centrality than another node v_j, where*

$$\mathbb{E}\left[\sum_r A_{ir}\right] > \mathbb{E}\left[\sum_r A_{jr}\right], \tag{B.5}$$

and $\mathbf{A} = (A_{ij})$ *is the (unweighted) adjacency matrix, then node* v_i *has higher expected Katz centrality than node* v_j*, where*

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \alpha_{katz}^t \sum_r A_{ri}^t\right] > \mathbb{E}\left[\sum_{t=1}^{\infty} \alpha_{katz}^t \sum_r A_{rj}^t\right], \tag{B.6}$$

and α_{katz} *is the discounting factor.*

Proof. In such SBM, for each pair of nodes v_i, v_j, A_{ij} is an independently distributed Bernoulli random variable, with success probability p_1 if $v_i, v_j \in \mathcal{B}_1$, p_2 if $v_i, v_j \in \mathcal{B}_2$, and p_{12} otherwise. Hence, nodes in the same communities are equivalent, and there are only two distinct expected values in both centralities, one for each community.

For the degree centrality[2],

$$\mathbb{E}\left[\sum_r A_{ir}\right] = \begin{cases} n_1 p_1 + n_2 p_{12}, & v_i \in \mathcal{B}_1, \\ n_1 p_{12} + n_2 p_2, & v_i \in \mathcal{B}_2, \end{cases}$$

where n_1, n_2 are the sizes of communities $\mathcal{B}_1, \mathcal{B}_2$, respectively. Hence, condition (B.5) can only happen when nodes v_i and v_j are in different communities.

Without loss of generality, we assume $v_i \in \mathcal{B}_1$, and then $v_j \in \mathcal{B}_2$. We show that (B.6) holds true by proving the following stronger relationship where for each $t > 0$,

$$\mathbb{E}\left[\sum_r A_{ri}^t\right] > \mathbb{E}\left[\sum_r A_{rj}^t\right]. \tag{B.7}$$

We show it by induction on t. (i) When $t = 1$,

$$\mathbb{E}\left[\sum_r A_{ri}\right] = n_1 p_1 + n_2 p_{12} > n_1 p_{12} + n_2 p_2 = \mathbb{E}\left[\sum_r A_{rj}\right],$$

where the inequality is true by condition (B.5). (ii) Suppose (B.7) is true for all $t \leq t'$.

[2]Note that in the classical SBM, the expressions can be slightly different, due to the common assumption of no self-edges. However, we allow self-loops in our analysis, thus also include v_i in computing the expected degree of node v_i

Then when $t = t' + 1$,

$$\mathbb{E}\left[\sum_r A_{ri}^{t'+1}\right] = \mathbb{E}\left[\sum_r \sum_q A_{rq}^{t'} A_{qi}\right] = \sum_{v_q \in \mathcal{B}_1} \mathbb{E}\left[\sum_r A_{rq}^{t'}\right] p_1 + \sum_{v_q \in \mathcal{B}_2} \mathbb{E}\left[\sum_r A_{rq}^{t'}\right] p_{12}$$
$$= \mathbb{E}\left[\sum_r A_{ri}^{t'}\right] n_1 p_1 + \mathbb{E}\left[\sum_r A_{rj}^{t'}\right] n_2 p_{12}, \tag{B.8}$$

where the second equality is by independence, and the last equality is by equivalence among nodes in the same communities. Similarly,

$$\mathbb{E}\left[\sum_r A_{rj}^{t'+1}\right] = \mathbb{E}\left[\sum_r A_{ri}^{t'}\right] n_1 p_{12} + \mathbb{E}\left[\sum_r A_{rj}^{t'}\right] n_2 p_2. \tag{B.9}$$

Hence, the difference (B.8) - (B.9) is

$$\mathbb{E}\left[\sum_r A_{ri}^{t'}\right] n_1(p_1 - p_{12}) + \mathbb{E}\left[\sum_r A_{rj}^{t'}\right] n_2(p_{12} - p_2)$$
$$> \mathbb{E}\left[\sum_r A_{rj}^{t'}\right] (n_1(p_1 - p_{12}) + n_2(p_{12} - p_2)) > 0,$$

where the first inequality is by induction hypothesis, and the last inequality is by condition (B.5). □

Bibliography

[1] M. Abramson, C. Audet, J. Chrissis, and J. Walston. Mesh adaptive direct search algorithms for mixed variable optimization. *Optim. Lett.*, 3:35–47, 2009.

[2] Y. Ahn, S. Ahnert, J. Bagrow, and A. Barabási. Flavor network and the principles of food pairing. *Sci. Rep.*, 1:196, 2011.

[3] K. Ailawadi, B. Harlam, J. César, and D. Trounce. Quantifying and improving promotion effectiveness at CVS. *Market. Sci.*, 26(4):566–575, 2007.

[4] C. Altafini. Dynamics of opinion forming in structurally balanced social networks. *PLoS ONE*, 7(6):1–9, 2012.

[5] C. Altafini. Consensus problems on networks with antagonistic interactions. *IEEE Trans. Automat. Control*, 58(4):935–946, 2013.

[6] S. Aref, L. Dinh, R. Rezapour, and J. Diesner. Multilevel structural evaluation of signed directed social networks based on balance theory. *Sci. Rep.*, 10(1):15228, 2020.

[7] M. Asllani, T. Carletti, F. Di Patti, D. Fanelli, and F. Piazza. Hopping in the crowd to unveil network topology. *Phys. Rev. Lett.*, 120:158301, 2018.

[8] F. Atay and S. Liu. Cheeger constants, structural balance, and spectral clustering analysis for signed graphs. *Discrete Math.*, 343(1):111616, 2020.

[9] S. Athey and S. Stern. An empirical framework for testing theories about complementarity in orgaziational design. Technical report, Nat. Bur. Econ. Res., 1998.

[10] C. Audet, S. Le Digabel, V. Montplaisir, and C. Tribes. NOMAD version 4: Nonlinear optimization with the MADS algorithm. *preprint*, arXiv:2104.11627v2, 2021.

[11] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, pages 519–528. ACM, 2012.

[12] S. Banerjee, M. Jenamani, and D. Pratihar. A survey on influence maximization in a social network. *Knowledge Inf. Sist.*, 62:3417–3455, 2020.

[13] A. Barbour, L. Holst, and S. Jason. *Poisson approximation*. Oxford University Press, 1992.

[14] A. Barten. Consumer demand functions under conditions of almost additive preferences. *Econometrica*, 32(1/2):1–38, 1964.

[15] F. Belardo. Balancedness and the least eigenvalue of laplacian of signed graphs. *Linear Algebra Appl.*, 446:133–147, 2014.

[16] P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan. Mixed-integer nonlinear optimization. *Acta Numer.*, 22:1–131, 2013.

[17] S. Berry, A. Khwaja, V. Kumar, A. Musalem, K. Wilbur, G. Allenby, B. Anand, P. Chintagunta, W. Hanemann, P. Jeziorski, and A. Mele. Structural models of complementary choices. *Market. Lett.*, 25:245–256, 2014.

[18] L. Bieche, J. Uhry, R. Maynard, and R. Rammal. On the ground states of the frustration model of a spin glass by a matching method of graph theory. *J. Phys. A Math. Gen.*, 13(8):2553–2576, 1980.

[19] T. Bijmolt, H. Van Heerde, and R. Pieters. New empirical generalizations on the determinants of price elasticity. *J. Marketing Res.*, 42(2):141–156, 2005.

[20] Y. Bilu and N. Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(1):495–519, 2006.

[21] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier. Maximizing social influence in nearly optimal time. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 946–957. SIAM, 2014.

[22] F. Boukouvala, R. Misener, and C. Floudas. Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO. *European J. Oper. Res.*, 252:701–727, 2016.

[23] A. Bovet and H. Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nat. Commun.*, 10(7), 2019.

[24] E. Breugelmans, K. Campo, and E. Gijsbrechts. Shelf sequence and proximity effects on online grocery choices. *Market. Lett.*, 18:117–133, 2007.

[25] R. Briesch, P. Chintagunta, and E. Fox. How does assortment affect grocery store choice? *J. Marketing Res.*, 46:176–189, 2009.

[26] A. Browet and P. Dooren. Low-rank similarity measure for role model extraction. In *Proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems*, pages 1412–1418, 2014.

[27] G. Burdock. *Fenaroli's handbook of flavor ingredients.* CRC Press, Boca Raton, 5 edition, 2004.

[28] S. Burer and A. Letchford. Non-convex mixed-integer nonlinear programming: A survey. *Surv. Oper. Res. Manag. Sci.*, 17:97–106, 2012.

[29] P. Cameron. Cohomological aspects of two-graphs. *Math. Z.*, 157(1):101–119, 1977.

[30] P. Cameron, J. Seidel, and S. Tsaranov. Signed graphs, root lattices, and coxeter groups. *J. Algebra*, 164(1):173–209, 1994.

[31] P. Cameron and A. Wells. Signatures and signed switching classes. *J. Combin. Theory Ser. B*, 40(3):344–361, 1986.

[32] D. Cartwright and F. Harary. Structural balance: A generalization of heider's theory. *Psychol. Rev.*, 63(5):277–293, 1956.

[33] D. Centola. The spread of behavior in an online social network experiment. *Science*, 329:1194–1197, 2010.

[34] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *Amer. J. Sociol.*, 113(3):702–734, 2007.

[35] F. Chen, X. Liu, D. Proserpio, I. Troncoso, and F. Xiong. Studying product competition using representation learning. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1261–1268. ACM, 2020.

[36] N. Chen. On the approximability of influence in social networks. *SIAM J. Discrete Math.*, 23:1400–1415, 2009.

[37] W. Chen, C. Castillo, and L. Lakshmanan. *Information and Influence Propagation in Social Networks*. Morgan & Claypool, 2013.

[38] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *Proceedings of the 11th SIAM International Conference on Data Mining*, pages 379–390. SIAM, 2011.

[39] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1029–1038. ACM, 2010.

[40] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *2010 IEEE International Conference on Data Mining*, pages 88–97. IEEE, 2010.

[41] L. Christensen, D. Jorgenson, and L. Lau. Transcendental logarithmic utility functions. *Am. Econ. Rev.*, 65(3):367–383, 1975.

[42] M. Cohen, J. Kalas, and G. Perakis. Promotion optimization for multiple items in supermarkets. *Manag. Sci.*, 67(4):2340–2364, 2021.

[43] R. Collins. *The Sociology of Philosophies*. Harvard University Press, 2009.

[44] N. Crossley, A. Mechelli, P. Vértes, T. Winton-Brown, A. Patel, C. Ginestet, P. McGuire, and E. Bullmore. Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci.*, 110(28):11583–11588, 2013.

[45] M. Cucuringu, P. Davies, A. Glielmo, and H. Tyagi. SPONGE: A generalized eigenproblem for clustering signed networks. In K. Chaudhuri and M. Sugiyama, editors, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, volume 89, pages 1088–1098. PMLR, 2019.

[46] J. Davis. Clustering and structural balance in graphs. *Hum. Relat.*, 20(2):181–187, 1967.

[47] Y. De Montjoye, L. Radaelli, V. Singh, and A. Pentland. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347:536–539, 2015.

[48] A. Deaton and J. Muellbauer. An almost ideal demand system. *Am. Econ. Rev.*, 70(3):312–326, 1980.

[49] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Adv. Complex Syst.*, 03(01n04):87–98, 2000.

[50] M. Degroot. Reaching a consensus. *J. Amer. Statist. Assoc.*, 69(345):118–121, 1974.

[51] E. Demaine, M. Hajiaghayi, H. Mahini, D. Malec, S. Raghavan, A. Sawant, and M. Zadimoghadam. How to influence people with partial incentives. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 937–948. ACM, 2014.

[52] A. den Boer. Dynamic pricing and learning: Historical origins, current research, and new directions. *Surv. Oper. Res. Manag. Sci.*, 20(1):1–18, 2015.

[53] C. Donnat and S. Holmes. Tracking network distances: An overview. *Ann. Appl. Stat.*, 12(2):971–1012, 2018.

[54] D. Easley and J. Kleinberg. *Positive and Negative Relationships*, pages 107–136. Cambridge University Press, 2010.

[55] T. Elrod, G. Russell, A. Shocker, R. Andrews, L. Bacon, B. Bayus, J. Carroll, R. Johnson, W. Kamakura, P. Lenk, J. Mazanec, V. Rao, and V. Shankar. Inferring market structure from customer response to competing and complementary products. *Market. Lett.*, 13:221–232, 2002.

[56] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[57] E. Even-Dar and A. Shapira. A note on maximizing the spread of influence in social networks. In X. Deng and F. Graham, editors, *Internet and Network Economics*, pages 281–286, Berlin, Heidelberg, 2007. Springer.

[58] G. Facchetti, G. Iacono, and C. Altafini. Computing global structural balance in large-scale signed social networks. *Proc. Natl. Acad. Sci.*, 108(52):20953–20958, 2011.

[59] D. Fanelli and A. McKane. Diffusion in a crowded environment. *Phys. Rev. E*, 82:021113, 2010.

[60] J. Farrell. Network structure and influence of the climate change countermovement. *Nature Clim. Change*, 6(4):370–374, 2016.

[61] M. Fay and M. Proschan. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Stat. Surv.*, 4:1–39, 2010.

[62] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486:75–174, 2010.

[63] S. Fortunato and D. Hric. Community detection in networks: A user guide. *Phys. Rep.*, 659:1–44, 2016.

[64] S. Gabel, D. Guhl, and D. Klapper. P2V-MAP: Mapping market structures for large retail assortments. *J. Marketing Res.*, 56:557–580, 2019.

[65] Z. Ghalmane, C. Cherifi, and M. Cherifi, H. El Hassouni. Extracting backbones in weighted modular complex networks. *Sci. Rep.*, 10(1):15539, 2020.

[66] T. Giovannelli, G. Liuzzi, S. Lucidi, and F. Rinaldi. Derivative-free methods for mixed-integer nonsmooth constrained optimization. *Comput. Optim. Appl.*, 82(1):293–327, 2022.

[67] A. Goyal, W. Lu, and L. Lakshmanan. CELF++: Optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th International Conference Companion on World Wide Web*, pages 47–48. ACM, 2011.

[68] G. Grimmett and D. Stirzaker. Two limit theorems. In *Probability and Random Processes*, pages 216–225. Oxford University Press, New York, 3 edition, 2001.

[69] D. Guilbeault, J. Becker, and D. Centola. Complex contagions: A decade in review. In *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks*, pages 3–25. Springer, Cham, 2018.

[70] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci.*, 102(22):7794–7799, 2005.

[71] Ö. Gür Ali, S. Sayin, T. Woensel van, and J. Fransoo. SKU demand forecasting in the presence of promotions. *Expert Syst. Appl.*, 36(10):12340–12348, 2009.

[72] F. Gursoy and D. Gunnec. Influence maximization in social networks under deterministic linear threshold model. *Knowl. Based Syst.*, 161:111–123, 2018.

[73] P. Hage and F. Harary. *Structural models in anthropology.* Cambridge University Press, 1983.

[74] F. Harary. On the notion of balance of a signed graph. *Michigan Math. J.*, 2(2):143–146, 1953.

[75] F. Harary. Structural duality. *Behav. Sci.*, 2(4):255–265, 1957.

[76] R. Hastie, T. Tibshirani and J. Friedman. Unsupervised learning. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, pages 485–585. Springer, New York, 2009.

[77] X. He, H. Du, M. Feldman, and G. Li. Information diffusion in signed networks. *PLoS ONE*, 14(10):e0224177, 2019.

[78] F. Heider. Attitudes and cognitive organization. *J. Psychol.*, 21(1):107–112, 1946.

[79] J. Hendrickx. A lifting approach to models of opinion dynamics with antagonisms. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 2118–2123, 2014.

[80] P. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.*, 76(373):33–50, 1981.

[81] Y. Hou. Bounds for the least laplacian eigenvalue of a signed graph. *Acta Math. Sin. (English Series)*, 21(4):955–960, 2005.

[82] Y. Hou, J. Li, and Y. Pan. On the laplacian eigenvalues of signed graphs. *Linear Multilinear Algebra*, 51(1):21–30, 2003.

[83] J. Huber, A. Gossmann, and H. Stuckenschmidt. Cluster-based hierarchical demand forecasting for perishable goods. *Expert Syst. Appl.*, 76:140–151, 2017.

[84] T. Kawamoto and M. Rosvall. Estimating the resolution limit of the map equation in community detection. *Phys. Rev. E*, 91:012809, 2015.

[85] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146. ACM, 2003.

[86] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. *Theory Comput.*, 11:105–147, 2015.

[87] A. Kök, M. Fisher, and R. Vaidyanathan. Assortment planning: Review of literature and industry practice. In *Retail Supply Chain Management: Quantitative Models and Empirical Studies*, pages 175–236. Springer, Boston, 2015.

[88] J. Kunegis, A. Lommatzsch, and C. Bauckhage. The slashdot zoo: Mining a social network with negative edges. In *Proceedings of the 18th International Conference on World Wide Web*, pages 741–750, New York, NY, USA, 2009. ACM.

[89] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. De Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 559–570. SIAM, 2010.

[90] M. Laguna, F. Gortázar, M. Gallego, A. Duarte, and R. Martí. A black-box scatter search for optimization problems with integer variables. *J. Global Optim.*, 58:497–516, 2014.

[91] R. Lambiotte, J. Delvenne, and M. Barahona. Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE Trans. Network Sci. Eng.*, 1:76–90, 2014.

[92] A. Lancichinetti and S. Fortunato. Community detection algorithms: A comparative analysis. *Phys. Rev. E*, 80:056117, 2009.

[93] S. Le Digabel. Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm. *ACM Trans. Math. Software*, 37, 2011.

[94] E. Leicht, P. Holme, and M. Newman. Vertex similarity in networks. *Phys. Rev. E*, 73:026120, 2006.

[95] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1:5–43, 2007.

[96] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 420–429. ACM, 2007.

[97] C. Lewis. *Demand forecasting and inventory control*. Routledge, 2012.

[98] H. Li and J. Li. Note on the normalized laplacian eigenvalues of signed graphs. *Australas. J. Combin.*, 44(1):153–62, 2009.

[99] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, and J. Wu. Weighted networks of scientific communication: the measurement and topological role of weight. *Phys. A*, 350:643–656, 2005.

[100] Y. Li, W. Chen, Y. Wang, and Z. Zhang. Voter model on signed social networks. *Internet Math.*, 11(2):93–133, 2015.

[101] Y. Li, J. Fan, Y. Wang, and K. Tan. Influence maximization on social graphs: A survey. *IEEE Trans. Knowledge Data Engrg.*, 30:1852–1872, 2018.

[102] J. Liu, X. Chen, T. Başar, and M. Belabbas. Exponential convergence of the discrete- and continuous-time altafini models. *IEEE Trans. Automat. Control*, 62(12):6168–6182, 2017.

[103] W. Liu, X. Chen, B. Jeon, L. Chen, and B. Chen. Influence maximization on signed networks under independent cascade model. *Appl. Intell.*, 49(1):912–928, 2019.

[104] F. Lorrain and H. White. Structural equivalence of individuals in social networks. *J. Math Sociol.*, 1(1):49–80, 1971.

[105] Z. Lu, W. Zhang, W. Wu, J. Kim, and B. Fu. The complexity of influence maximization problem in the deterministic linear threshold model. *J. Comb. Optim.*, 24:374–378, 2012.

[106] H. Ma, H. Yang, M. Lyu, and I. King. Mining social networks using heat diffusion processes for marketing candidates selection. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 233–242. ACM, 2008.

[107] H. Mann and D. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, 18(1):50–60, 1947.

[108] M. Mantrala, M. Levy, B. Kahn, E. Fox, P. Gaidarev, B. Dankworth, and D. Shah. Why is assortment planning so difficult for retailers? A framework and research agenda. *J. Retailing*, 85:71–83, 2009.

[109] S. Marvel, J. Kleinberg, R. Kleinberg, and S. Strogatz. Continuous-time model of structural balance. *Proc. Natl. Acad. Sci.*, 108(5):1771–1776, 2011.

[110] N. Masuda, M. Porter, and R. Lambiotte. Random walks and diffusion on networks. *Phys. Rep.*, 716–717:1–58, 2017.

[111] Z. Meng, G. Shi, and K. Johansson. Multiagent systems with compasses. *SIAM J. Control Optim.*, 53(5):3057–3080, 2015.

[112] Z. Meng, G. Shi, K. Johansson, M. Cao, and Y. Hong. Behaviors of networks with antagonistic interactions and switching topologies. *Automatica*, 73:110–116, 2016.

[113] E. Mossel and S. Roch. Submodularity of influence in social networks: From local to global. *SIAM J. Comput.*, 39(6):2176–2188, 2010.

[114] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. *Phys. A*, 374:457–470, 2007.

[115] M. Newman. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E*, 64:016131, 2001.

[116] M. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 64:016132, 2001.

[117] M. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci.*, 98(2):404–409, 2001.

[118] M. Newman. Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci.*, 101(suppl 1):5200–5205, 2004.

[119] M. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74:036104, 2006.

[120] M. Newman. The configuration model. In *Networks*, pages 434–445. Oxford University Press, New York, 2 edition, 2018.

[121] M. Newman. The empirical study of networks. In *Networks*, pages 15–104. Oxford University Press, New York, 2 edition, 2018.

[122] M. Newman. Fundamentals of network theory. In *Networks*, pages 107–270. Oxford University Press, New York, 2 edition, 2018.

[123] M. Newman, S. Strogatz, and D. Watts. Random graph with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, 2001.

[124] W. Nicholson and C. Snyder. Demand relationships among goods. In *Microeconmic Theory: Basic principles and extensions*, pages 187–206. Cengage Learning, Mason, 11 edition, 2012.

[125] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, 2015.

[126] R. Pastor-Satorras and A. Vespignani. Epidemics in the internet. In *Evolution and Structure of the Internet: A Statistical Physics Approach*, pages 180–210. Cambridge University Press, Cambridge, 2004.

[127] T. Peixoto. Bayesian stochastic blockmodeling. In *Advances in Network Clustering and Blockmodeling*, pages 289–332. John Wiley & Sons, Ltd, West Sussex, 2019.

[128] M. Porter, J. Onnela, and P. Mucha. Communities in networks. *Notices Amer. Math. Soc.*, 56(9):1082–1097, 2016.

[129] J. Preusse, J. Kunegis, M. Thimm, S. Staab, and T. Gottron. Structural dynamics of knowledge networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 506–515, 2013.

[130] K. Read. Cultures of the Central Highlands, New Guinea. *Southwest. J. Anthropol.*, 10(1):1–43, 1954.

[131] L. Ribeiro, P. Saverese, and D. Figueiredo. Struc2vec: Learning node representations from structural identity. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 385–394. ACM, 2017.

[132] M. Rosvall, D. Axelsson, and C. Bergstrom. The map equation. *Eur. Phys. J. Spec. Top.*, 178:13–23, 2009.

[133] F. Ruiz, S. Athey, and D. Blei. SHOPPER: A probabilistic model of consumer choice with substitutes and complements. *Ann. Appl. Stat.*, 14(1):1–27, 2020.

[134] M. Schaub, J. Delvenne, R. Lambiotte, and M. Barahona. Multiscale dynamical embeddings of complex networks. *Phys. Rev. E*, 99:062308, 2019.

[135] G. Schoenebeck, B. Tao, and F. Yu. Think globally, act locally: On the optimal seeding for nonsubmodular influence maximization. *Inf. Comput.*, 2022.

[136] M. Serrano, M. Boguñá, and A. Vespignani. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci.*, 106(16):6483–6488, 2009.

[137] P. Shakarian, A. Bhatnagar, A. Aleali, E. Shaabani, and R. Guo. The independent cascade and linear threshold models. In *Diffusion in Social Networks*, pages 35–48. Springer, Cham, 2015.

[138] G. Shi, C. Altafini, and J. Baras. Dynamics over signed networks. *SIAM Rev.*, 61(2):229–257, 2019.

[139] G. Shi, M. Johansson, and K. Johansson. How agreement and disagreement evolve over random dynamic networks. *IEEE J. Sel. Areas Commun.*, 31:1061–1071, 2013.

[140] G. Shi, A. Proutiere, M. Johansson, J. Baras, and K. Johansson. The evolution of beliefs over signed social networks. *Oper. Res.*, 64(3):585–604, 2016.

[141] I. Song and P. Chintagunta. A discrete–continuous model for multicategory purchase behavior of households. *J. Marketing Res.*, 44(4):595–612, 2007.

[142] X. Song, B. Tseng, C. Lin, and M. Sun. Personalized recommendation driven by information flow. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 509–516. ACM, 2006.

[143] V. Srivastava, J. Moehlis, and F. Bullo. On bifurcations in nonlinear consensus networks. In *Proceedings of the 2010 American Control Conference*, pages 1647–1652. IEEE, 2010.

[144] P. Symeonidis, N. Iakovidou, N. Mantas, and Y. Manolopoulos. From biological to social networks: Link prediction based on multi-way spectral clustering. *Data Knowl. Eng.*, 87:226–242, 2013.

[145] P. Symeonidis and N. Mantas. Spectral clustering for link prediction in social networks with positive and negative links. *Soc. Netw. Anal. Min.*, 3:1433–1447, 2013.

[146] H. Theil. The information approach to demand analysis. *Econometrica*, 33(1):67–87, 1965.

[147] Y. Tian and R. Lambiotte. Unifying information propagation models on networks and influence maximisation. *preprint*, arXiv:2112.01465, 2021.

[148] Y. Tian, S. Lautz, A. Wallis, and R. Lambiotte. Extracting complements and substitutes from sales data: a network perspective. *EPJ Data Sci.*, 10(1):45, 2021.

[149] V. Traag, L. Waltman, and N. Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.*, 9:5233, 2019.

[150] H. van Heerde and S. Neslin. *Sales Promotion Models*, pages 13–77. Springer, Cham, 2017.

[151] E. van Nierop, D. Fok, and P. Franses. Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements. *Market. Sci.*, 27(6):1065–1082, 2008.

[152] L. Vicente and A. Custódio. Analysis of direct searches for discontinuous functions. *Math. Program.*, 133:299–325, 2012.

[153] N. Vinh, J. Epps, and J. Bailey. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.*, 11:2837–2854, 2010.

[154] C. Wang, W. Chen, and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Min. Knowl. Discov.*, 25:545–576, 2012.

[155] D. White and K. Reitz. Graph and semigroup homomorphisms on networks of relations. *Soc. Netw.*, 5:193–234, 1983.

[156] L. Wu, X. Ying, X. Wu, A. Lu, and Z. Zhou. Examining spectral space of complex networks with positive and negative links. *Int. J. Social Network Mining*, 1(1):91–111, 2012.

[157] W. Xia, M. Cao, and K. Johansson. Structural balance and opinion separation in trust–mistrust social networks. *IEEE Trans. Control Netw. Syst.*, 3(1):46–56, 2016.

[158] J. Yang and J. Leskovec. Defining and evaluating network communities based on ground-truth. *Knowl. Inf. Syst.*, 42:181–213, 2015.

[159] W. Zachary. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.*, 33:452–473, 1977.

[160] T. Zaslavsky. Signed graphs. *Discrete Appl. Math.*, 4(1):47–74, 1982.

[161] T. Zhou, J. Ren, M. Medo, and Y. Zhang. Bipartite network projection and personal recommendation. *Phys. Rev. E*, 76:046115, 2007.