Università degli Studi di Pisa

Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Matematica

Discrete Models of Information Diffusion in Networks

Tesi di Laurea Magistrale

21 luglio 2014

Candidate
Alessio Muscillo

Advisors

Prof. Juan Tejada
(Universidad Complutense de Madrid)

Prof. Paolo Acquistapace
(Università degli Studi di Pisa)

Examiner

Prof. Dario Andrea Bini
(Università degli Studi di Pisa)

Anno Accademico 2013/2014

Contents

Chapter 1

Introduction

The last decades have seen the growth of **network science** mostly due to the birth and development of the *Internet*. The access to large databases, web networks and the possibility of heavy computer simulations have promoted new approaches in the research methods which are now more based on mathematical quantitative analysis. Obviously, the economical factors linked to the so-called *Digital/Web/Internet Economy* have played a huge role in stimulating the applications and the implementation of the new techniques to a wide variety of fields. In fact, this variety may be clearly seen by considering the **interdisciplinarity** of this field of study, whose it is one of the most important and interesting assets, and it reflects the *melting pot* that often characterizes this as well as many others branches of applied mathematics, especially the ones dedicated to the analysis of non-linear phenomena.

Network science deals with telecommunication networks, computer networks, biological networks, semantic networks, social networks to name but a few. In addition, the methods and techniques used are widely diverse and come from graph theory, statistics, statistical mechanics, data mining etc. (see, for example [Castellano *et al.*, 2009]). However, because of the abovementioned interdisciplinarity, a network scientist has always to handle (at least) two points of view: the first one concerning a specific model, with specific relationships and individuals, and the second one concerning general (or *topological*) properties. Think, for example, about the differences that exist between the ties in a social network, where the relationships indicate friendship, and a biological network simulating genes' interactions. Although these two worlds are totally different, once one is considering them as a network (or as a weighted graph, for instance) they may have some common properties, concepts, ideas and, so, sometimes the same techniques may be used.

Which are the characteristics of a network? Whenever one has a system with sub-units that are linked together, one may think of a network. One of the advantages of using this approach is that it has different levels, so, while the understanding of the whole system may be very complicated, it may be also possible that the knowledge of its *elemental* and *interacting* parts is quite deep. The characteristics or properties considered in a network may vary depending on the point of view that one adopts as well as the level of resolution of the tools that one uses. There are "global" properties such as the *size* of the network, its *diameter* and properties that combine "global" and "local" information, such as how much the network is *clustered* or which are the *central nodes* in the network.

Based on such properties one may define different typologies of networks, such as *complex networks*, which are characterized by having non-trivial topological features, in opposition with other types of networks such as *random graphs*, where patterns or structure are less predominant. Of course, on the one hand, one may be interested in trying to understand where (and why) these features come from: if a social structure has high *clustering coefficient* or a *hierarchical structure*, who are the bosses in the network? where do their power come from? On the other hand, one does not want to waste this information, whereas would like to take advantage of these properties to capture and predict some behaviors or dynamics more easily, effectively and efficiently.

Why and when to use a network as model? Why to model a population, for example, as a network instead of homogeneously? One of the possible answers to this question is, in our opinion, that using a network model may allows you to describe phenomena that may not be modelled well at the level of a homogeneous population. Think, for example, of the adoption of political views: one would tend to have political ideas that are somehow aligned to the ones of her friends, even if these ideas are in the minority if considered in a nation scale. Just to name but a few applications or examples, one can think about:

- a network consisting of a *social group*, like friends, sharing ideas or opinions, in which, for example, the intensity of the relationship depends on the intensity of the friendship, as in [Salganik *et al.*, 2006] and [Tang *et al.*, 2009];
- *diffusion of innovations* as well as *viral marketing*, i.e. the study of how a new product/idea can spread thanks to a word-of-mouth process, as in [Valente, 1995], [Richardson, Domingos, 2001], [Richardson, Domingos, 2002];
- a *public service grid*, in which servers and users are connected, like an electric power grid, as in [Asavathiratham, 2000];

- a *terrorist network*, in which the interest could be finding the influent people or the ones who play a key-role in the organization, as in [Lindelauf *et al.*, 2013];
- a *biological network* simulating genes' interactions, as in [Moretti *et al.*, 2010] or *food webs* as in [Estrada, 2007].

Organization of this work Generally speaking, the models we may think of are environments or structures consisting of a grid of individuals tied by relationships, where we want to use a network approach to study a group dynamic. In particular, we focus on models of networks in which relationships allow information transfer and where passing this information may be considered as a *diffusion mechanism*. The amount of passed information or the effectiveness of this passing is measured by the *influence* that an individual has over her neighbors in the network. In our models, we also try to take into account that the combined or *aggregated* influence of a group of neighbors may be different from the influence that a single individual is able to exert. Then, we study how this *diffusion* dynamic in the network may depend on the diffusion mechanism.

More precisely, this work is organized as follows.

- In chapter 2 and 3 we describe two different types of discrete models of diffusion in networks, respectively *threshold models* and *cascade models*. The diffusion mechanism is different as well as the ways through which neighbors' influence is measured and aggregated. These two model are shown to be stochastically equivalent. They are also considered in their *progressive* and *non-progressive* versions, meaning that the nodes remain active forever or may deactivate-activate again and again; we also describe a way to reduce non-progressive models to progressive ones.
- Chapter 4 is dedicated to the study of "local" properties: the main question to answer is: which are "reasonable" ways of aggregating the influence of the neighbors? First we try to define what *influence aggregation* is and then we study its properties by means of well-known aggregation functions such as *t-conorms* and *co-copulae*.
- Finally, in chapter 5 we focus on the *diffusion* in a network: we consider some "global" properties of the diffusion and show that some of them are inherited from "local" properties, which have the advantage of being much more tractable with respect to the mathematical uncontrollability and computational intractability of diffusion in huge network. In fact, we show that *maximizing the diffusion* in a network is a *NP-hard* problem, but however we describe two algorithms for approximating the solutions: a *greedy algorithm* and a *Shapley-value based algorithm*.

Chapter 2

Threshold Models

Among *diffusion* phenomena in networks, here we focus on those based on *threshold* mechanisms. In general, the word "threshold" is used when the probability for an event to occur changes quickly as some parameter varies. Although the notions of threshold and phase transition originated in physics, they now play an important role in many fields of mathematics, such as probability, statistics, computer science, and in other sciences, such as economics and social science (see [Zapata, Gauthier, 2003], [Kalai, Safra, 2006]).

In mathematics as well as in other sciences, it is quite easy to find examples of threshold phenomena. Think, for example, about the *logistic equation* used for modeling *population growth*:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = rP\left(1 - \frac{P}{K}\right),$$

where $P(t)$ is the population at time t and the constant r defines the growth rate, while K is the maximal load allowed in the environment. It is not hard to prove that K represents a threshold, because if the population starts above it, i.e. $P(0) > K$, then it decreases and tends asymptotically to K. Conversely, if $P(0) < K$, then the behavior is increasing to the same asymptote.

Another simple example is given by an election between two candidates, Alice and Bob, where the result is determined by majority and every voter, independently one from another, votes for Alice with probability p and for Bob with probability[1] $1 - p$. In this example, a simple application of the *Condorcet's Jury Theorem*, seen as a consequence of the *weak law of large numbers*, gives us that the threshold is the probability $p = 1/2$. When there are (infinitely) many voters, the probability of Alice winning the election changes from being close to 0, in case $p < 1/2$, to being close to 1, in

[1] No abstention is allowed.

case $p > 1/2$. Others examples of such behaviors, especially related to *technological transitions*, may be founded in [Zeppini *et al.*, 2013].

In this chapter we focus on models of *diffusion in networks* (see [Prakash *et al.*, 2012]), where each node has its own threshold, which often is a random variable that tries to represent the non-exact knowledge of the behavior of the node. To understand better what we are dealing with, we may use some examples from [Granovetter, 1978]. Think, for instance, of a riot or a strike: each person decides to join the strike if enough colleagues have already joined it. At the same time, each person has a specific threshold, that is the proportion of colleagues she would have to see joining the strike before she would do so. One advantage of this approach is that, since thresholds are "continuous percentages", this prevents from only classifying people in radical or conservative, allowing a person to have a "moderate" opinion. On the other hand, this gradualness does not compromise an individual to make a choice and so the model to be "decisional".

Another simple example is the application of such models to diffusion of technological innovations (see also [Valente, 1995]). Let us consider a situation in which there are two competitive softwares providing instant messaging services. A person would switch from the old technology to the new one depending on how many friends are already using the second; more exactly, only if a certain part of her friends have already chosen it and, in fact, they are acting as a *critical mass* to her. Furthermore, not any friend has to have the same "value" in terms of influence, meaning that possibly there are some nodes that are "close friends" to whom she sends texts very often, and so these tend to influence her more than other general acquaintances. Lastly, notice that not any node has to have the same "threshold", indeed each node may have its own threshold depending on the aptitude or inclination to adopt the new technology. In fact, this is a very well-studied phenomenon in sociological models for adoption of new technologies, in which people are categorized in adopter classes, according to their facility to adopt the new technology.[2]

These examples show two common ideas, which will be also core ideas for the models treated in this chapter:

- we have a *network* where each individual either has to make a (binary) decision or has to choose between two mutually exclusive alternatives or one of them;

- this decision is *influenced* by the neighbors, whose influence *accumulates* in some way; in addition, there is a *threshold* mechanism such that there is a change from an alternative to the other one only if

[2]See "The diffusion of hybrid seed corn in two Iowa communities (1943)" by Ryan and Gross, as well as "Technology adoption lifecycle" on Wikipedia.

there is a number or proportion of influencing neighbors that is great enough.

Perhaps the most intuitive model which gathers these ideas is the so-called **Linear Threshold Model**, which we will extensively study later (see also [Chen *et al.*, 2010], [Saito *et al.*, 2010]), also pointing out some of what we think are its bad aspects. In this model, the network is constituted by a weighted graph, neighbors' influence is simply summed up and if it exceeds a certain threshold, then the node is activated or "convinced"; then, this same node will become an influencing one and it will try to activate its neighbors in the following time-steps, and so on for every node throughout the whole network. Besides intuition, further important reasons to analyze and focus on this specific model are that it is easily generalizable to more complex models, its applicability to a wide variety of situations, and, finally, its mathematical treatability and algorithmic results (see also the last chapter).

2.1 General Threshold Model

The *General Threshold Model* was introduced by [Kempe *et al.*, 2003] as a generalization of a variety of models that had been already studied, including the very well-known *Linear Threshold Model* proposed by [Granovetter, 1978]. All these discrete diffusion models have in common the same core idea: for every inactive node in the network, the influence of its neighbors accumulates in a certain way and has to exceed a node-specific threshold in order for the node to become active.

Intuitively speaking, the threshold could represent how hard it is to convince a person to adopt a new technology that a part of her friends have already adopted. The mechanism by which the influence of a group on a node is represented is one of the key-features that makes the difference among different threshold models. This is especially true if one tries to make a model in which the influence of a group may not be reduced to the mere sum of the its components.

2.1.1 Network Structure

In the General Threshold Model the network structure is represented by a directed graph $G = (V, E)$, where $V = \{1, ..., n\}$ is the set of the nodes and $E \subseteq \{(i, j) : i, j \in V\}$ is the set of the arcs or edges. Associated to each node $j \in V$, there are some objects:

- a **threshold** θ^j, which is an uniformly distributed random variable in $[0, 1]$;
- a set of **neighbors** V^j, defined by $V^j = \{i \in V : (i, j) \in E\}$;

- a set function $f^j : 2^{V^j} \to [0,1]$, called **activation function**, that maps every subset S of j's neighbor set V^j to a number $f^j(S) \in [0,1]$, subject to the condition that $f^j(\emptyset) = 0$.

Before starting to describe how the diffusion process actually works, let us try to understand or interpret the network structure. The set S of active neighbors of j represents the subset of j's friends who actually try to convince her or, in the different context of a contagious disease which is spreading through the population, it may be read as the subset of ill friends who could infect j. Now, the way how the influences of the group S combine together is described by the function f^j. Eventually, as already sketched above and as it will be clearer in what follows, the threshold θ^j will represent a measure of j's propensity to accept the idea or technology or whatever is the object of interest which is spreading through the network.

Remark 2.1. In many versions of the *General Threshold Model*, as in the *Linear Threshold Model* in the section 2.3.1, the network structure is a *weighted directed graph* $G = (V, E, W)$, where $W = \{w_{ij}\}_{(i,j)\in E}$ is the set of the **influence weights** and each w_{ij} is in the interval $(0,1]$ and represents the amount of influence that the j's neighbor $i \in V^j$ exerts over j. The case $w_{ij} = 0$ is excluded because it would mean that i has no direct influence over j, meaning that either they are not friend, in which case $(i,j) \notin E$, or i's influence is irrelevant for j, so it is not taken into account in the model.

In this general approach, the only requirement on the function f^j made by [Kempe *et al.*, 2003] was the *monotonicity*, although in Conjecture 4.3 in the same paper they also considered functions that were *submodular*, i.e. functions $f : 2^V \to \mathbb{R}$ such that

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

for all $S \subseteq T \subseteq V$ and for all $i \notin T$, being V any set. In this context, it is clear that the monotonicity is an intuitive property to require, because it depicts a situation in which the more people try to influence a node, the more effective the result is. Concerning the submodularity, it might also be considered somehow intuitive: in fact, it means that the more people have tried to influence a node, the less an additional neighbor can make the difference, which is often referred to as a condition of *diminishing returns*. Moreover, this *local* condition will turn out to be crucial for proving the above-mentioned conjecture (see [Mossel, Roch, 2007]) and preserving the *global* submodularity of the whole process, as we will explain in the last chapter.

2.1.2 Progressive and Non-progressive Threshold Model

Given the network structure as described above, now we want to specify the dynamic of the model, that is the *diffusion mechanism*, which defines

how the process evolves over time. First of all, the evolution of the process happens in steps, which means that the time is discrete and always denoted with the letter $t \in \mathbb{N}$. Each step will include some of the basic mechanisms that compose the entire process.

Before starting to define the actual diffusion process, we need to distinguish between two substantially different cases: *progressive* and *non-progressive* models. The significant difference is that in the progressive case a node can only *progress* from inactivity to activity, while in the other case going in the opposite direction is allowed as well. This distinction is visible and clear if we are trying to model a contagious disease which is spreading through the population: the progressive case corresponds to a disease where a sick person acquires immunity after the infection, such as the *measles*, whereas the opposite case of a *flu* would be modeled by the non-progressive one.

Progressive Threshold Model

Let us consider a network structure as described above and let us define a time-step process associated with it. In every time step, each node has only two available states: *inactive* or *active*. Being a progressive model, the dynamic for every node may only change from being inactive to active, and of course this change in the state needs at least one time step to happen. Once a node becomes active, it will remain so forever.

Formally speaking, let $A_t \subseteq V$ be the set of nodes that are currently active at time t and, by construction, the set of inactive nodes at time t is $V \setminus A_t$. The diffusion process evolves as follows:

- the process starts at time $t = 0$ with a given set $A_0 \subseteq V$ of initially active nodes and with a vector of thresholds $\theta^1, ..., \theta^n$, which are chosen independently and uniformly distributed in $[0, 1]$;
- at each time step $t > 0$:
  - if a node $j \in A_{t-1}$ was active in the previous step $t - 1$, then it will remain active, so $j \in A_t$;
  - otherwise, given an inactive node $j \in V \setminus A_{t-1}$, it will be activated and then belong to A_t if

$$f^j\left((A_{t-1})^j\right) \geq \theta^j,$$

    where $(A_{t-1})^j$ denotes[3] the set of j's neighbors who are active at time $t - 1$. If the condition is not fulfilled, then j will remain inactive and belong to $V \setminus A_t$;

[3]Whenever the context is clear, throughout we will use indiscriminately the notations $(A_t)^j$ or $A_t \cap V^j$ to indicate the set of j's neighbors that are also active at time t. Clearly, when the second notation is used, $V^j \subseteq V$ indicates the set of j's neighbors.

- the process stops when no more activations are possible.

Remark 2.2. Notice that the diffusion process cannot last more than n steps. In fact, the set sequence $\{A_t\}_t$ of active nodes at time t is increasing in t, so there exists a step $T \leq n$ such that the process *stabilizes*, which is exactly when no more activations are possible:

$$A_0 \subseteq A_1 \subseteq ... \subseteq A_T = A_{T+1} = ...$$

Obviously, it does not mean that the *steady state* has to be the so-called *consensus* in the whole network, which means that at the end of the process there could still be some inactive nodes, whose neighbors were not strong enough to influence them.

Remark 2.3. Once the random thresholds $\theta^1, ..., \theta^n$ are established at the beginning of the process, it then evolves deterministically.

Non-progressive Threshold Model

At first glance, the *Non-progressive Threshold Model* could seem just a simple generalization of the progressive case, where the reversibility in the activity state of a node is allowed. Indeed, we will show later in section 2.2 that the non-progressive case may be reduced to the progressive one. However, the non-progressive case opens the door to many questions which simply might not make sense in the progressive context. This also means that non-progressive models are generally harder to deal with and this is well highlighted by comparing the vastness of the literature dedicated to the study of progressive diffusion models and the scarcity for non-progressive ones (see, for example [Yang *et al.*, 2014]).

In progressive models, once a node becomes active (or influenced or infected), it will remain so forever; conversely, in non-progressive models an active node may become inactive again. This might suggest that such models are better in situations such as *non-monopolistic markets* ([Fazli *et al.*, 2012]), where the customers can choose between two competing providers and switch company whenever certain conditions fail to occur: for example, in the case of instant messaging companies, a customer chooses the service that the majority of her friend (or her closest friends) has already adopted.

Another fundamental difference between progressive and non-progressive models is in the initially activated nodes at time $t = 0$, previously indicated with $A_0 \subseteq V$. From the point of view of progressive models, it is natural to think of A_0 as a set of seeds from which the process starts; nevertheless, in the non-progressive case it may happen that some initially activated node $j \in A_0$ immediately deactivates in the following step and this, of course, is not desirable at all if one is trying to maximize the diffusion in the network by adequately choosing the nodes in A_0. So, in non-progressive contexts,

another idea of seeding spontaneously comes out: instead of activating all nodes in A_0 at the beginning of the process, each initially selected node $j \in A_0$ has an activation time $\tau_j \geq 0$ at which it becomes really active and so "contagious". This means that in the non-progressive models is more common to talk of **interventions** instead of *initially activated nodes*, where, of course, an *intervention* is defined by selecting a node $j \in V$ and its activation time $\tau_j \geq 0$ at the beginning of the process. In particular, notice that the starting point of a non-progressive process is the set of interventions instead of the set of initially activated nodes.

Remark 2.4. In the progressive case, if every τ_j is 0, then the idea of interventions simply coincides with the original one of initially activated nodes A_0. However, even if the times are $\tau_j > 0$, if we compare two progressive processes with the same thresholds $\theta^1, ..., \theta^n$, the first (I) being a standard one starting from A_0 and the second (II) being the one with the correspondent interventions with $j \in A_0$ at times τ_j, then the result is that the sets $A_t^{(I)}$ and $A_t^{(II)}$ of active nodes at time t may be different, but the sets of active nodes at the end of the process $R^{(I)}$ and $R^{(II)}$, i.e. when it stabilizes, they do coincide.
This may be seen as a special case of the *Progressive Threshold Model with quiescent times*, see section 2.3.2, where only the nodes of the interventions have quiescent times $\tau_j > 0$, while the other nodes have set $\tau_j = 0$.

Using a slightly different notation from the previous section, we can now formally describe the non-progressive process. Let $\widetilde{A}_t \subseteq V$ be the set of currently active nodes at time t and $I = \{(j,t) \in V \times \mathbb{N} : t = \tau_j\}$ be the set of *interventions*, where the pair $(j,t) \in I$ corresponds to activating the node j at time $\tau_j = t$. The evolution of the diffusion process is, then:

- at time $t = 0$ the process starts with a set of interventions $I \subseteq V \times \mathbb{N}$. If $(j,0) \in I$, then $j \in \widetilde{A}_0$;
- at time $t > 0$, every node $j \in V$ chooses a threshold θ_t^j randomly from the interval $[0,1]$ and independently from the other nodes.
  - If $(j,t) \in I$, then the node j is active in the step t, i.e. $j \in \widetilde{A}_t$;
  - otherwise, the node j is active in the step t if

$$f^j\left((\widetilde{A}_{t-1})^j\right) \geq \theta_t^j.$$

    If this inequality is not fulfilled, $j \notin \widetilde{A}_t$. (As notation, $(\widetilde{A}_{t-1})^j$ is the subset of $\widetilde{A}_{t-1}$ constituted by j's neighbors.)

First of all, notice that in this process every node has to choose its threshold at every step t (possibly with the only exceptions of the nodes that are

intervening in that step, for which the threshold does not affect their activation), so there is no relation between the sets $\widetilde{A}_t$, unlike it happened in the previous case. Secondly, the process does not end nor stabilizes necessarily: this naturally originates non-trivial questions about the possible convergence of the process to a steady state (see [Fazli *et al.*, 2012] and [Grabisch, Rusinowska, 2011]). Furthermore, the same concept of *influence of the set* A_0 cannot be easily generalized and it is not even clear what should be its intuitive meaning; we will see in the next section what is the role of the set of *interventions* and that there will be one possible answer by considering a horizon time $T \in \mathbb{N}$ and then taking into account the sum of the number of steps where every node j is has been active.

2.2 Diffusion Function

Diffusion Function in the progressive case

Let us consider the progressive model: given the initial set $S \subseteq V$ and the thresholds $\theta^1, ..., \theta^n$, at the end of the diffusion process we will have the set of active nodes at termination, say $R_{\theta^1,...,\theta^n}(S) \subseteq V$. We have already observed that the process evolves deterministically, once the thresholds are given. On the other hand, if S is fixed, then $R_{\theta^1,...,\theta^n}(S)$ is a random variable, depending on the random thresholds, which indicates the set of active nodes. So, it is natural to take the average of this random variable, and in particular:

Definition 2.5. *The* ***diffusion function*** *of the Progressive General Threshold Model on G is the expected number of active nodes at the end of the process, i.e.*

$$\sigma : 2^V \longrightarrow [0, n], \qquad \sigma(S) := \mathbb{E}_{\theta^1,...,\theta^n} [|R(S)|],$$

where $S \subseteq V$ is the set of initially active nodes, $R(S)$ the (random) set of active nodes at the end of the process and $|\cdot|$ indicates the cardinality of a set.

Of course, assumptions on the elements constituting the model, such as what random variables are used as thresholds or what type of function f^j are chosen, have serious repercussions on the function $\sigma(\cdot)$. Hence, it is natural to wonder what kind of properties σ has and, possibly, what are the consequences that a choice on the model has on it.

In [Kempe *et al.*, 2003], motivated by algorithmic reasons related to an optimization problem called *Target Set Selection* problem, which we will consider in the last chapter, the requirements on the model were:

- the thresholds $\theta^1, ..., \theta^n$ are independent and uniformly distributed in $[0, 1]$;

- the functions $f^1, ..., f^n$ are monotone and submodular.

In particular, the authors managed to prove that in the special case of the Linear Threshold Model (LT) (see next section 2.3.1), in fact the correspondent diffusion function $\sigma_{LT}(\cdot)$ preserves these properties, i.e. it is monotone and submodular as well. This naturally leaded the authors to state a conjecture (Conjecture 4.3 in [Kempe *et al.*, 2003]), claiming that every diffusion model verifying the previous assumptions would have had a diffusion function σ monotone and submodular.

This conjecture was finally proved in [Mossel, Roch, 2007]. From our point of view, this is especially important because it proves that some "local" properties may preserve "globally" under diffusion processes and, since the global diffusion process is much harder to control than its local elements, this gives a useful tool to build monotone and submodular progressive diffusion models. This latter characteristic will be particularly useful because it may guarantee good properties in terms of optimization and approximation, as we will see in the last chapter of this work.

2.2.1 Reduction: from Non-progressive to Progressive

So far, we have seen that in case of progressive models, some properties on the diffusion process can be guaranteed. On the other hand, the situation of non-progressive models is not as clear as in the progressive one. In fact, it is not even obvious what may be considered as "influence of a set S" or the "spread of the initial set S". In other words, it is not easy to understand how to generalize the idea used to define σ for the progressive case , since in the non-progressive model there might not be a proper "end" for the process, while in the progressive case the process had to unfold and end in at most $n-1$ time steps.

Remark 2.6. As already mentioned, in the literature there are not many works about non-progressive models, nevertheless in some of them one can find studies concerning the convergence of the considered stochastic process: see, for example, [Grabisch, Rusinowska, 2011], [Forster *et al.* 2012], [Fazli *et al.*, 2012], [Prakash *et al.*, 2012] and [Ning, 2012].

However, if it is assumed that the non-progressive model can only run for $\tau < \infty$ steps, then it is possible to reduce this process to an equivalent progressive model. To do so, let $G = (V, E)$ be the graph of a given non-progressive model and let τ be fixed. Now, we define a new graph $G^{(\tau)} = (V^{(\tau)}, E^{(\tau)})$, specifically a *layered graph*, such that:

- there are $\tau \cdot (n+1)$ vertices $V^{(\tau)}$, i.e. the set V is considered to be copied τ times and we denote these copies by $\{V_t\}_{0 \leq t \leq \tau}$. Each node j of G is adequately copied and labeled such that $j_t \in V_t$, for all

$0 \leq t \leq \tau$ and, so the set of vertices of $G^{(\tau)}$ is

$$V^{(\tau)} = \bigcup_{0 \leq t \leq \tau} V_t;$$

- if there is an edge (i, j) in G (with weight $w_{ij} > 0$), then in $G^{(\tau)}$ there will be an edge (with the same weight) on any (i_{t-1}, j_t), for all $0 < t \leq \tau$. In other words, any node $j_t \in V_t \subset G(\tau)$ is connected precisely with all nodes $i_{t-1} \in V_{t-1}$ such that (i, j) was an edge in G;
- if the set of interventions is $I \subset V \times \{0, ..., \tau\}$, then the set of initially active nodes $A_0^{(\tau)} \subset V^{(\tau)}$ is defined by: $j_t \in A_0^{(\tau)}$, i.e. initially active, if and only if $(j, t) \in I$;
- the model on $G^{(\tau)}$ is progressive.

Remark 2.7. Notice that in the progressive model on $G^{(\tau)}$ the dynamics is quite restricted and the activations may only follow the layered structure of the graph: since there only are edges from the "level" $t - 1$, i.e. V_{t-1} to the level V_t, then the node j_t is activated in the whole process if and only if it is activated exactly at time t.

With the construction above, it may be proved that a limited/finite non-progressive process is equivalent to its progressive equivalent (Theorem 5.1 in [Kempe *et al.*, 2003]), more precisely

Proposition 2.8. *A node $j \in G$ is active at time t in the non-progressive general threshold process if and only if $j_t \in G^{(\tau)}$ is activated in the correspondent progressive process.*

Remark 2.9. If the graph is weighted $G = (V, E, W)$, then the reduction to $G^{(\tau)}$ has also to take into account the weights $w_{ij} \in W$: in order to do so, it suffices that the every edge (i_{t-1}, j_t) has the corresponding weight $w_{ij} > 0$.

Diffusion Function in the non-progressive case

We already observed that in the non-progressive case it is not possible to imitate what we have done for the progressive case in order to get a diffusion function, because the non-progressive process is not necessarily finite and does not necessarily stabilize.

Even so, if we take a finite time horizon τ then we may consider the diffusion function $\sigma^{(\tau)}$ associated to the associated progressive model $G^{(\tau)}$ as defined above and simply define it as the diffusion function corresponding to the non-progressive process, when limited to τ.

Definition 2.10. *The **diffusion function** for the Non-Progressive Threshold Model on G and a time horizon $\tau \in \mathbb{N}$ is, by definition, the diffusion function $\sigma^{(\tau)}$ associated to the reduced progressive model on $G^{(\tau)}$.*

It is worth trying to understand what $\sigma^{(\tau)}$ represents for the original non-progressive model:

- with the notations used in the previous sections, let $\widetilde{A}_t$ be the set of active nodes at time t, with $0 \leq t \leq \tau$, of the non-progressive model in G;
- analogously, let
$$A_0^{(\tau)} \subseteq A_1^{(\tau)} \subseteq ... \subseteq A_\tau^{(\tau)} \subseteq V,$$
be the sets of active nodes at time t of the progressive process $G^{(\tau)}$. Notice that the set $A_t^{(\tau)}$ contains, by definition, all nodes of $G^{(\tau)}$ that are active at time t, which also comprises nodes of the previous layers $V_0, ..., V_{t-1}$ of the layered graph $G^{(\tau)}$.

By definition, the set of reachable nodes at the end of the process in $G^{(\tau)}$ is the random variable $R\left(A_0^{(\tau)}\right) = A_\tau^{(\tau)}$ and the expectation of its cardinality is the diffusion function

$$\sigma^{(\tau)}(A_0) = \mathbb{E}\left[\left|R\left(A_\tau^{(\tau)}\right)\right|\right].$$

Proposition 2.8 tells us that the node $v \in \widetilde{A}_t$ is active at time t if and only if v_t is activated at time t, i.e. $v_t \in A_t^{(\tau)} \setminus A_{t-1}^{(\tau)}$, so we may write[4]

$$R\left(A_0^{(\tau)}\right) = A_\tau^{(\tau)} = A_0^{(\tau)} \sqcup \left(A_1^{(\tau)} \setminus A_0^{(\tau)}\right) \sqcup ... \sqcup \left(A_\tau^{(\tau)} \setminus A_{\tau-1}^{(\tau)}\right),$$

then, taking the cardinality

$$\left|R\left(A_0^{(\tau)}\right)\right| = \left|A_0^{(\tau)}\right| + \left|A_1^{(\tau)} \setminus A_0^{(\tau)}\right| + ... + \left|A_\tau^{(\tau)} \setminus A_{\tau-1}^{(\tau)}\right|.$$

This means that the diffusion function $\sigma^{(\tau)}$ considered in the non-progressive case counts the (expected number of) times that every single node has been active.

2.3 Special Cases

2.3.1 Linear Threshold Model

The *Linear Threshold Model* is perhaps the easiest and most intuitive example of threshold model: neighbors' influence weights are simply summed up and a node becomes active if this sum exceeds its threshold. Its extreme treatability and versatility explains why it is one of the most widely studied models, since its introduction by [Granovetter, 1978]. Indeed, in the seminal work by [Kempe *et al.*, 2003] the results about the submodularity of the

[4]Remind that $A_0^{(\tau)} = \widetilde{A}_0$, by construction.

diffusion process (and with that approximability) were only proved for the Linear Threshold Model and not for the General Threshold Model. Also, it is extensively used in practical experiments and simulations as well as in many variations on the basic models (see [Pathak *et al.*, 2010]).

Formally speaking, let the network structure be a directed weighted graph $G = (V, E, W)$ and let us assume that for every node j it holds that

$$\sum_{i \in V^j} w_{ij} \leq 1,$$

where w_{ij} is the influence of i over j and $V^j \subset V$ is the set of j's neighbors, as usual. Now, the Linear Threshold Model derives from the general one, when the functions f^j are defined by

$$f^j(S \cap V^j) = \sum_{i \in S \cap V^j} w_{ij},$$

where S is any subset of V, so as $S \cap V^j$ is the subset of S consisting of j's neighbors. Having said this, the diffusion mechanism and process remain the same as in the General Threshold Model, in the progressive case as well as in the non-progressive one.

Despite all the above mentioned advantages, the Linear Threshold Model has an obvious limitation: it can only be applied when the restriction $\sum w_{ij} \leq 1$ stands, which means that it cannot be used with any kind of network. Although it might seem just a normalization condition, it implies an unpleasant (and possibly unrealistic) consequence: if the node j has a lot of friends, then they necessarily have to to have little weights in order for the sum to stay below 1.

Another disadvantage is that it could be even too easy to activate/convince a node, since that by using the sum, it is quite easy to *saturate* and to reach the upper limit 1. For example, if the node j had only two friends and both were active and had an influence weight of 0.5, even if their individual influence is not too high, they would manage to activate j regardless of the threshold $\theta^j \leq 1$, since the sum is 1. These and other properties will be better analyzed in detail in the chapter dedicated to *Influence Aggregation Models*.

2.3.2 Progressive Threshold Model with quiescent times

This model is a generalization of the standard progressive threshold model with the addition of a *waiting time* τ_j for every node $j \in V$. More precisely, the process evolves in the same way, with the exception that if a node j is activated at time t, it will effectively become active at time $t + \tau_j$. Note that in the case $\tau_j = 0$ for all $j \in V$, then this is the original progressive threshold model.

Perhaps surprisingly, the two diffusion models are in fact equivalent, as proved in [Kempe *et al.*, 2005].

Proposition 2.11. *Let be $R^{(w)}(S)$ the distribution[5] of active nodes at the end of a general progressive threshold model with waiting times τ_j, with $j \in V$, when $S \subseteq V$ is the set of initially active nodes. Then it is the same of $R(S)$, which is the distribution of the general progressive threshold where no waiting time is allowed.*

From this proposition, it follows that the *diffusion function* $\sigma(\cdot)$ is the same and, in particular, it will verify the same properties, such as monotonicity and submodularity.

2.3.3 Competitive Influence Threshold Models

A natural extension for diffusion processes in network is a competitive setting, where there are two (or more) competitive technologies in the network that are trying to spread. Even though this might seem an intuitive generalization, in this environment is not hard to find examples and models where the diffusion function $\sigma(\cdot)$ is not submodular nor even monotone, as it may be seen in [Borodin *et al.*, 2010].

More precisely, the framework is the following: in a weighted network $G = (V, E, W)$, let $I_A \subseteq V$ and $I_B \subseteq V$ be the initial adopters of the two different (and competing) technologies A and B; each node j in the network may be A-active, B-active or inactive. The diffusion function $\sigma_{I_B}(I_A)$ tries to capture what is the diffusion of the technology A in presence of B.

Weight-Proportional Competitive Linear Threshold Model The first model is a generalization of the Linear Threshold Model and it is progressive and happens in discrete steps, as usual, and also it has the same limitation on the total weight over a node j, i.e. $\sum_{i \in V^j} w_{ij} \leq 1$. Using the notation A_t and B_t for indicating respectively the A-active and B-active nodes at time t, the diffusion process unfolds as follows:

- at time $t = 0$, let $A_0 = I_A$, $B_0 = I_B$ be the sets of initially active nodes and $\theta^1, ..., \theta^n$ the node-specific thresholds;
- at time $t > 0$:
  - if a node $j \in A_{t-1}$ was A-active, it will remain so in the current step: $j \in A_t$. Analogously, if $j \in B_{t-1}$ then $j \in B_t$;
  - if a node $j \notin A_{t-1} \cup B_{t-1}$ was inactive, then it will activate in the current step if the weight of its active neighbors (no matter

[5] For ease of notation, here we omit the dependence on the thresholds $\theta^1, ..., \theta^n$.

if A-active or B-active) exceeds its threshold θ^j, i.e.

$$\sum_{i \in (A_{t-1} \cup B_{t-1}) \cap V^j} w_{ij} \geq \theta^j.$$

Otherwise, j remains inactive, i.e. $j \notin A_t \cup B_t$. Now, once j has been activated, then it will become A-active with probability

$$\mathbb{P}\{j \in A_t | \, j \notin A_{t-1} \cup B_{t-1}\} = \frac{\sum_{i \in A_{t-1} \cap V^j} w_{ij}}{\sum_{i \in (A_{t-1} \cup B_{t-1}) \cap V^j} w_{ij}},$$

which corresponds to the proportion of j's active neighbors that are A-active. Otherwise, j will become B-active.

This model coincides with the usual Linear Threshold Model with single technology when $I_B = \emptyset$. So, the *Target Set Selection*[6] for this problem is at least hard as in the case of the *Linear Threshold Model*, which means that it is *NP-hard*, as we will see in the last chapter. Furthermore, it is possible to find instances of this process whose diffusion functions are not monotone or not submodular. Roughly speaking, it is worth noting that in this model, since both A-active and B-active nodes contribute in the activation of other nodes, it may indeed happen that the increase of B-active nodes will make possible more activations but then these newly activated nodes will be A-active.

Separated-Threshold Model for Competing Technologies In the previous model, a node became active regardless the actual type of activity (A-active or B-active) of its neighbors. So, a natural extension is: what if each node has different sensibilities with respect to the different technologies and the influence of a neighbor changes if the technology her is currently supporting is different?

More formally, in this case the network $G = (V, E, W^A, W^B)$ comprises two different weights for each edge w_{ij}^A and w_{ij}^B, and every node j has two (random) thresholds $\theta_A^j, \theta_B^j \in [0, 1]$. The diffusion mechanism is quite similar to the previous models:

- at time $t = 0$, it starts with the initially active nodes I_A, I_B respectively and with the assignment of the thresholds θ_A^j, θ_B^j for every node $j \in V$;
- at time $t > 0$, being a progressive model, any A-active (respectively B-active) node will remain so in the following time steps. An inactive node j will become A-active if

$$\sum_{i \in A_{t-1} \cap V^j} w_{ij}^A \geq \theta_A^j,$$

[6] Also called *Top-k Set Selection Problem.*

and analogously B-active if the same condition holds for B-active nodes. In case j is both A-activated and B-activated, then a random choice (either a coin-flip or any tie-breaking function) will decide if $j \in A_t$ or $j \in B_t$.

It can be proved (see [Borodin *et al.*, 2010]) that this process is monotone, in contrast with the previous model, and this follows from the fact that here only A-activations contribute to further A-activations and the same happens with B-activations. However, there exist instances of this diffusion process that are not submodular.

Chapter 3

Cascade models

In the previous chapter, we described models of diffusion where the influence of a group of neighbors over a node was measured and fixed, while the probabilistic part of the process was given by a random node-specific threshold, which represented the node's sensitivity with respect to the idea that was spreading or its propensity to become active. Of course, this was justified because we tried to model a non-exact knowledge of each node's sensitivity.

Perhaps a more intuitive approach is the one used to define Cascade Models: in such models, the influence of a neighbor (as well as a group of neighbors) over a node is directly given in terms of probability of activation. For example, let us consider the example of a *contagious disease* which is spreading through a population: it is intuitive to think that the more two people are friends and spend time together, the more the probability that the infection passes from one to the other increases.

In particular, in some models such as the *Independent Cascade Model*, the weights of the edges in the network graph are directly connected to the activation probability of each node. Although this might seem more intuitive and even more reasonable, managing different probabilities has to be done carefully: this will be especially clear later, when we will need the so-called *order-independence hypothesis* in order for the *General Cascade Model* to be well-defined.

For this reason, we think that the exposition of this chapter will be clearer if organized in the opposite way with respect to the previous one: first, we will define the *Independent Cascade Model*, which is conceptually the simplest one, but allows us to explain almost all the main ideas that will be used in more general models, which will be described after. The last section of the chapter will be dedicated to make an equivalence between the General Threshold Model and the *General Cascade Model*. The main references for this chapter are [Kempe *et al.*, 2003], [Kempe *et al.*, 2005]; others will be explicitly included in the text.

3.1 Independent Cascade Model

As already mentioned, the *Independent Cascade Model* is, with no doubt, the simplest one as well as one of the most studied and used in many applications (see, for example, [Kimura *et al.*, 2009], [Saito *et al.*, 2010]) among all cascade models and, despite its clear limitations, is very useful to capture the main ideas that lie at the core of most of them. Roughly speaking, in this model when a node i becomes active, it has a single chance to activate each currently inactive neighbor j and i succeeds with probability p_{ij}, independently from the other events in the process.

The network structure for the *Independent Cascade Model* is a weighted directed graph $G = (V, E, W)$, where $V = \{1, ..., n\}$ is the set of nodes, $E \subseteq V \times V$ is the set of edges and $W = \{p_{ij} \in [0, 1] : (i, j) \in E\}$ is the set of *activation weights*. Since there are no others objects or entities (such as thresholds) related to the structure, it is slightly simpler than the one used for threshold models, nevertheless some assumptions will be needed to specify exactly the dynamics of the process (also depending on if it is progressive or non-progressive).

3.1.1 Progressive Independent Cascade Model

First of all, let us precisely define some hypotheses about the activation mechanism of the model:

- when a node $i \in V$ becomes active, say at time t, then it has a single chance to activate each currently inactive neighbor j, i.e. **no multiple attempts** are allowed and, also, the nodes are only **temporarily contagious** for 1 time step. This means that once a node has made all its attempts, being contagious, it becomes *non-contagious* but still remains *active*;

- this activation successfully happens with probability p_{ij}, **independently** of all other events, in particular from other neighbors' attempts. vice versa, with probability $1 - p_{ij}$ the node i fails in its (unique) attempt;

- in case more than one node can try to activate the same node, the order in which they will make their attempts is **random**.

We use the same notations of the previous chapter, i.e. $A_t \subseteq V$ indicates the set of all active nodes at time $t \geq 0$ and $S^j = S \cap V^j$ is the subset of S constituted by j's neighbors. It is worth noticing that $A_t \setminus A_{t-1}$ is the set of **contagious nodes** (or **recently activated nodes**) at time t.[1] Having said

[1] As convention, the initially active nodes coincide with the (initially) contagious ones, i.e. $A_{-1} := \emptyset$.

this, it is possible now to define the evolution of the *Progressive Independent Cascade Model*:

- the process starts with a given set $A_0 \subseteq V$ of *initially contagious/active* nodes;
- at any time $t \geq 0$, let i be a contagious node, i.e. $i \in A_t \setminus A_{t-1}$. Given an inactive neighbor j, that is $j \in V^i$ and $j \notin A_t$, it will be successfully activated (and, so, become contagious in the next step) by i with a probability p_{ij}, in which case it will hold that $j \in A_{t+1}$;
- the process ends when there are no more contagious nodes.

It is worth noticing that the Progressive (General) Threshold Model ends in at most $n-1$ steps, while this model does it in at most n steps. This little difference is due to the fact that here in the step t it is said which are the nodes that will become active in the following step $t+1$, whereas in the threshold model the active nodes at time t were established in the same step t. In other words, it only depends on the definition of each single *step* in the two processes and, indeed, the step in the cascade model is from the point of view of an already active node; on the other hand, the step in the threshold model is defined from the point of view of an inactive node.

Apart from this difference, both processes share the same step-by-step-diffusion idea and, hence, both lead to the same definition of influence of a given initial active set S, i.e. the *diffusion function* $\sigma(S)$. In particular, as every progressive model, the sets of active nodes constitute a monotone sequence

$$A_0 \subseteq A_1 \subseteq ... \subseteq A_{n+1} \subseteq V$$

and all of them are subsets of the finite set V. Now, let $A_0 = S$ be the set of initially active nodes and, associated to it, let us consider the random variable $R(S)$, which is given by the set of active nodes at the end of the process. Taking the average of the cardinality of R gives us the definition of **diffusion function**

$$\sigma(S) = \mathbb{E}\left[|R(S)|\right],$$

which is the *expected number of active nodes at the end of the process*, when the initial active set is S. The *Independent Cascade Model* was among the first models where the submodularity of the diffusion function could be proved in [Kempe *et al.*, 2003].

Proposition 3.1. *Given an arbitrary instance of the Independent Cascade Model, the resulting diffusion function $\sigma(\cdot)$ is monotone and submodular.*

Remark 3.2. As it happened for the Progressive Threshold Model, there is a procedure to "convert" the Progressive Independent Cascade in a deterministic process. Precisely, we start with a set $S = A_0$ and, at the beginning

of the process, we flip a coin with probability p_{ij}, for each edge $(i,j) \in E$, to establish if the edge (i,j) is *live* or not. Somehow surprisingly, the set of nodes that are reachable via live-edge paths, starting from the nodes in S, coincides with the set of active nodes at the end of the process $R(S)$, when the activation probabilities have given the same results as the flipping coins (specifically, see Claim 2.3 in [Kempe *et al.*, 2003]). We will use this same reasoning in the last chapter.

3.1.2 Non-progressive Independent Cascade Model

As every non-progressive model, this also will allow the nodes to switch from *inactivity* to *activity* and vice versa. Another analogy with the non-progressive case seen in the first chapter is in the set of *interventions* $I \subset V \times \mathbb{N}$, which is used instead of the set of initially active nodes, where $(j,t) \in I$ if and only if the node j is activated at time $\tau_j = t$, by definition.

The *Non-progressive Independent Cascade* behaves exactly as its progressive correspondent, but here the key idea is that, at each time step, an *active* node j has a probability β_j of *becoming inactive*: in other words, every active node j has a probability of $1 - \beta_j =: p_{jj}$ of activating itself in the following step, and if it fails and its active neighbors fail as well, then it will deactivate in the following step. As above, j's neighbors have probabilities p_{ij} to activate it independently one from another and from the other events. Again, in case of multiple active neighbors, they will try to activate j in a random order. Besides, concerning multiple activation attempts from the same neighbor, as in the progressive model they are *not allowed* in a single time step; of course, being a non-progressive model, when a node j changes repeatedly from being inactive to active, each time it becomes active again, it has another (single) chance of activating its currently inactive neighbors and, so, multiple attempts in diverse steps are allowed, in this broader sense. Lastly, notice that here the concept of *contagious* node is the same as *active* node, since every active node is considered *recently activated* and so contagious, indeed.

Using the notation of $\widetilde{A}_t$ for the set of currently active nodes at time t, the non-progressive process of the *Independent Cascade Model* may be described according to the following steps:

- it starts with a set of given interventions $I \subset V \times \mathbb{N}$;
- at each step $t \geq 0$, given any node $j \in V$, then:
  - if $(j,t) \in I$, then the node j is forced to become active at time t, so $j \in A_t$;
  - otherwise:
    - if $j \notin \widetilde{A}_t$ is inactive, each j's currently active neighbor $i \in \widetilde{A}_t \cap V^j$ has a probability p_{ij} of succeeding in the attempt of

activating it, independently from the attempts of the other neighbors. If at least one attempt is successful, then j will become active in the following step, i.e. $j \in \widetilde{A}_{t+1}$;

* if $j \in \widetilde{A}_t$ is currently active, the set of attempting nodes is the union of the previous one and j itself, that is

$$\left(\widetilde{A}_t \cap V^j\right) \cup \{j\}.$$

Therefore, as before, $j \in \widetilde{A}_{t+1}$ will remain active in the next step $t+1$ if either a node $i \in \widetilde{A}_t \cap V^j$ successfully activate it, which occurs with probability p_{ij}, or j itself does not deactivate, and that happens with a probability of $p_{jj} = 1 - \beta_j$. If none of the previous events succeed, then j will become inactive, that is $j \notin \widetilde{A}_{t+1}$.

As already noticed in the previous chapter, the fundamental difference between progressive models and non-progressive ones, in terms of *diffusion process*, is that in the latter the sequence of subsets of active nodes $\{\widetilde{A}_t\}_t$ is not monotone and, in fact, is not necessarily finite (nor convergent whatsoever).

Remark 3.3. If the set of active nodes at time t is $\widetilde{A}_t$, the probability that a node $j \in V$ becomes active in one step is:

$$\mathbb{P}\{j \in \widetilde{A}_{t+1} \mid \widetilde{A}_t\} = 1 - \prod_{i \in \widetilde{A}_t \cap V^j} (1 - p_{ij}),$$

where j is not necessarily in $\widetilde{A}_t$ nor out of it. It is worth noticing that this probability only depends on the currently active nodes $i \in \widetilde{A}_t$ considered as a set and not as a sequence/vector, i.e. *the order does not matter*. This feature will be somehow kept even in more general models.
Besides, because of this, one can easily see that a non-progressive model can be written as a *(time-homogeneous) Markov chain*, while in a progressive model the "future" closely depends on the "past" and not only on the "present".

3.1.3 Reduction: from Non-progressive to Progressive

Similarly to what was done in the previous chapter for the reduction from a non-progressive (and finite) threshold process to a progressive one, here it is possible to follow the same approach.

Now, given the structure of a non-progressive independent cascade $G = (V, E, W)$ and given a finite time horizon $\tau \in \mathbb{N}$, let us define a *layered graph* associated to it $G^{(\tau)} = (V^{(\tau)}, E^{(\tau)}, W^{(\tau)})$, such that:

- each $j \in V$ has a copy $j_t \in V_t$, for all $0 \leq t \leq \tau$; so, $V^{(\tau)} = \bigcup_{t \leq \tau} V_t$;

- if (i, j) is an edge in G, then in $G^{(\tau)}$ there will be an edge (i_{t-1}, j_t), for all $0 < t \leq \tau$, with the same weight (probability) p_{ij};
- since any weight β_j in G is associated to a "virtual" edge from j to j itself with weight p_{jj} in G, then, in $G^{(\tau)}$, it will correspond to an edge (j_{t-1}, j_t) from j at time $t-1$ to j in the following time step;
- the interventions I in G correspond to setting as initial set $A_0^{(\tau)}$ the nodes $j_t \in V^{(\tau)}$ such that $(j, t) \in I$.

With this construction, a non-progressive independent cascade process on G is equivalent to a progressive independent cascade on G^{τ}.

Proposition 3.4. *A node $j \in G$ is active at time t in the non-progressive cascade threshold process if and only if $j_t \in G^{(\tau)}$ is activated in the correspondent progressive process.*

3.2 General Cascade Model

As the name suggests, the *General Cascade Model* is a generalization of the *Independent Cascade Model* in which, in particular, we want to take into account that the influence of a single person over another one may be different from the influence of an entire group of people, which means that we do not have necessarily to reduce a group to the mere sum of its components, as if they were acting independently. This model was firstly introduced in [Kempe *et al.*, 2003], but more extensively studied only later in [Kempe *et al.*, 2005]. The fundamental idea behind the *General Cascade Model* is that when a node i tries to activate a node j, its capacity of influence depends on how many people have already tried (and failed) in activating j. This means that we can model, for example, situations in which the more people try to activate j, the less a single additional node can succeed in doing it or vice versa, and this could be the case of a *market* that tends to become *saturated* and, so, nodes' propensity to be activated changes over time.

Progressive General Cascade Model As the *Progressive Independent Cascade Model*, here it is assumed that *active* nodes are *contagious* immediately after their activation, whereas later they become *non-contagious* but still remain active. However, the main change in this model is that the hypothesis concerning the independence of the attempts from different nodes is removed.

More formally, given an inactive node j and a contagious neighbor i, we define $p_j(i; S)$ as the probability that i succeeds in activating j, when S is the set of j's active neighbors that have already tried and failed their attempt so far. By using the same notations as for progressive models, such

as $A_t \subseteq V$ for the subset of active nodes at time t, then S is a subset of A_t and, in particular, it comprises both non-contagious nodes, i.e. active nodes from previous steps, as well as currently contagious ones. This is indeed a generalization of the *Independent Cascade Model*, which is obtained when $p_j(i;S) = p_{ij}$, for all $i, j \in V$ and $S \subseteq V$, are taken constant and independent from the history of the process.

Recall that in the independent model, when multiple contagious nodes tried to activate j, their attempts were sequenced *randomly*. Here, in order for the model to be well-defined, it is necessary to understand that this approach may not be followed directly. For example, in principle, it is not clear nor obvious how the order of the elements of the set S changes (or should change) the probability $p_j(i;S)$. To solve this problem, it suffices to assume the **order-independence hypothesis**, which states that this order does not affect the probability of j to become active in the end. Precisely:

- let $U \subseteq V^j$ be a set of contagious nodes that are trying to influence $j \in V$ and let $u_1, ..., u_r$ and $u_{\pi(1)}, ..., u_{\pi(r)}$ be two permutations of the elements of U;
- the hypothesis is that for any set $T \subseteq V$ disjoint from U, representing the non-contagious nodes that have already attempted, it holds that

$$\prod_{k=1}^{r}(1 - p_j(u_k; T \cup \{u_1, ..., u_{k-1}\})) = \\ = \prod_{k=1}^{r}\left(1 - p_j\left(u_{\pi(k)}; T \cup \{u_{\pi(1)}, ..., u_{\pi(k-1)}\}\right)\right), \quad (3.1)$$

  where $\{u_0\} = \{u_{\pi(0)}\} = \emptyset$ as convention.

Remark 3.5. It is easy to interpret this hypothesis and see that the above expression represents what we needed: since the probability that u_k fails in activating j when T and $u_1, ..., u_{k-1}$ have already tried (in this order) is $1 - p_j(u_k; T \cup \{u_1, ..., u_{k-1}\})$, then the probability that at least one of the contagious nodes $u_1, ..., u_r$ succeeds in its attempt to activate j is exactly given by

$$1 - \prod_{k=1}^{r}(1 - p_j(u_k; T \cup \{u_1, ..., u_{k-1}\}).$$

In addition, the importance of the *order-independence hypothesis* will be even clearer because it will allow an equivalence between threshold and cascade models (see next section 3.3).

Having said this, the process of the *Progressive General Cascade Model* evolves as usual:

- it starts with a set of initially active nodes A_0 at time $t = 0$;

- at time $t > 0$, let $j \notin A_t$ be an inactive node and
  - let $U = V^j \cap (A_t \setminus A_{t-1})$ be the set of its currently contagious neighbors and let us indicate its elements with $U = \{u_1, ..., u_r\}$;
  - let $T = V^j \cap A_{t-1}$ be the set of j's neighbors that have already tried to activate it, up to the step $t - 1$;
  - for $k = 1, ..., r$, the node u_k tries to activate j and succeeds with a probability $p_j(u_k; T \cup \{u_1, ..., u_{k-1}\})$, being $u_1, ..., u_{k-1}$ the currently contagious nodes that have already failed in their attempt and T the non-contagious ones that had tried and failed in previous time steps.
- the process ends when there are no more contagious nodes.

Remark 3.6. Considering only *contagious* nodes being the actual *influencing* nodes, instead of generally active ones, seems to be a clear limitation for these models, which is present in the *Independent Cascade* as well as the *General Cascade*. This might be especially true if one tries to focus more on a "cumulative effect" of a whole influencing group. The point of view of cascade models is, instead, to capture the behavior of single nodes as part of bigger or more complex structures. Despite this intuition, it will be possible to prove (see next section 3.3) that cascade models are equivalent to threshold models, where the focus on "cumulative effect" is evident.

Non-progressive General Cascade Model With this model, we aim to generalize the *Non-progressive Independent Cascade Model* in the same way we have just done for the progressive model, meaning that we want to include correlations among the attempts made by different active nodes in the framework of the non-progressive cascade. Therefore, being a non-progressive model, here every active node has a probability of becoming inactive and every inactive node may be activated by its neighbors at every time step. In practice, there is no difference between being *active* and *contagious*, i.e. all active nodes are considered contagious. In other words, in case a node remains active many times over consecutively, we will consider that it has been activated in each one of these time step. However, at every time step, each active node j has a probability of β_j to become inactive, which is considered to be a probability of j to auto-deactivate itself, that is to say that $p_{jj} = 1 - \beta_j$.

Concerning the correlation among the attempts of different neighbors, here the situation is slightly simpler with respect to the above progressive model, since in a non-progressive model each time step may be considered almost separately from the others. As above, the probability that an active node i succeeds in activating its currently inactive neighbor j is $p_j(i; S)$, where S indicates the set of nodes that have already tried to activate j.

Notice the difference between this model and the progressive one: here S may only contain contagious (= active) nodes, while in the progressive model S could include nodes that had tried in previous time-steps. With the usual notations for non-progressive models, being $\widetilde{A}_t \subseteq V$ the set of active nodes at time t, we have $S \subseteq \widetilde{A}_t$. Lastly, here the *order-independence hypothesis* may be weakened, since only currently contagious nodes are considered to affect the activation probability. Hence, we only need the **(weak) order-independence hypothesis**:

$$\prod_{k=1}^{r}(1 - p_j(u_k; \{u_1, ..., u_{k-1}\})) = \prod_{k=1}^{r}(1 - p_j(u_{\pi(k)}; \{u_{\pi(1)}, ..., u_{\pi(k-1)}\})$$

for all $j \in V$ and for all permutation $u_{\pi(1)}, ..., u_{\pi(r)}$ of the r elements $u_1, ..., u_r$ that are trying to activate j. Notice that the (strong) order-independence hypothesis 3.1 of the previous section implies the weak version by taking $T = \emptyset$.

The formal evolution of the process is straightforward:

- the process starts with a set $\widetilde{A}_0 \subseteq V$ of initially activated nodes;
- at time $t \geq 0$, given a generic node $j \in V$, it will be active at time $t+1$ if there exists an active neighbor $i \in A_t \cap V^j$, i.e. a neighbor that is active in the current time step t, that succeeds in activating j. This will successfully happen with a probability $p_j(i; S)$, where $S \subseteq A_t \cap V^j$ indicates the set of nodes that have already attempted.

Remark 3.7. In analogy to what already done for the *Independent Cascade Model* and the *General Threshold Model*, it is possible to *reduce* a finite non-progressive process to an equivalent progressive one and this operation would connect the diffusion functions σ and $\sigma^{(\tau)}$ of both processes and their properties, such as monotonicity and submodularity, in particular.

3.3 Cascade-Threshold Equivalence

In [Kempe *et al.*, 2003] and [Kempe *et al.*, 2005] the authors show an equivalence existing between the *Progressive General Cascade* and the *Progressive General Threshold Model*. More specifically, they showed that given a cascade model, it is possible to make a threshold model such that the probability that a node j is activated at time t in the cascade model is the same as that one of j being activated in the equivalent threshold model, and vice versa.

Hereafter in this section, let us consider a *Progressive General Cascade Model* on a network $G = (V, E)$ with *activation probabilities* $p_j(i; S)$, for $i, j \in V$ and $S \subseteq V$. Analogously, on the same network G let $f^j(S)$, with $j \in V$ and $S \subseteq V$, be the *activation function* associated to a *Progressive General Threshold Model*.

Definition 3.8. *Two models Progressive Cascade and Progressive Threshold on the network G are called* ***equivalent*** *if, for any set of nodes $T \subseteq V$ and time $t \geq 0$, the probability that T is the set of active nodes at time t is the same in both processes.*

We want to give now a method by which it is possible to construct an equivalent model, starting from a given one. First, let us start with the probabilities $p_j(i; S)$ of a given cascade model. In order to define a threshold model, given an inactive node $j \in V$ and a set of active neighbors $S = \{u_1, ..., u_r\}$, we need to define the correspondent activation function $f_j(S)$, and to do so we make the following choice

$$f^j(S) := 1 - \prod_{k=1}^{r} (1 - p_j(u_k; \{u_1, ..., u_{k-1}\})). \tag{3.2}$$

It is worth noticing that the function $f^j(\cdot)$ is well-defined because the *order-independence hypothesis* on the probabilities $p_j(\cdot; \cdot)$ guarantees that they depend only on the set $S = \{u_1, ..., u_r\}$ and not on the order of the elements $u_1, ..., u_r$.

Before verifying the equivalence, let us make the converse procedure: starting from the functions $f^j(S)$ of a progressive threshold, we define the probabilities

$$p_j(i; S) := \frac{f^j(S \cup \{i\}) - f^j(S)}{1 - f^j(S)} \tag{3.3}$$

for any $S \subseteq V^j$ and $i \notin S$.

Lemma 3.9. *If the function f^j is defined by the probability p_j via* (3.2)*, then p_j verifies* (3.3)*, and vice versa.*

The Lemma 1 of [Kempe *et al.*, 2005] may be written as follows:

Lemma 3.10. *Given a Progressive Cascade Model and a Progressive Threshold Model where, respectively, the probabilities p_j verify* (3.3) *and the activation functions f^j verify* (3.2)*, then: for any time $t \geq 0$ and any sets $T, T' \subseteq V$, the probability that exactly the nodes of T are active at time t and those of T' are active at time $t + 1$ is the same under both processes.*

Corollary 3.11 (**Cascade-Threshold Equivalence**)**.**
If the Cascade model and the Threshold model satisfy (3.3) *and* (3.2)*, respectively, then they are equivalent.*

Remark 3.12. Concerning non-progressive models, the equivalence is quite easy to prove when both have the same time horizon $\tau < \infty$. In this case, one may reduce both non-progressive processes to their progressive correspondent in the layered graph $G^{(\tau)}$ and then apply the previous corollary.

Chapter 4

Influence Aggregation Models

In the previous two chapters we have considered a network and focused on "global" models which are processes involving the whole structure. Here, we take a more "local" point of view: in particular, we focus on a single node in the network and try to understand how this node *aggregates* the influence of its neighbors.

In order to understand what we exactly mean by *aggregation influence* and to formalize our intuitive idea of it, we make a list of "reasonable" or "desirable" properties, that we expect an *influence aggregation function* to satisfy. Then we study the relations among these properties and this leads us to consider some aggregation functions called *t-conorms* and *co-copulas*. This *aggregators* were introduced in *Fuzzy Logic* and *Probability*: when Karl Menger in 1942 proposed a probabilistic generalization of the concept of metric space by substituting the distance $d(a, b)$ with a distribution function F_{ab}, whose value $F_{ab}(x)$ is the probability that the distance between a and b is less than x, one of the first difficulties was the generalization of the *triangular inequality*. This was the birth of the *triangular norms* (or *t-norms*). On the other hand, *(co)-copulas* are used to define parametric measures of dependence between random variables. In the very different context of *Fuzzy Logic*, which is a special type of *many-valued logic*, *t-norms* and *t-conorms* are used as generalization of the usual *and*-$\wedge$ and *or*-$\vee$ binary operators of the "classical" true-false logic.

Finally, in the last section we show how starting from a "reasonable" way of aggregating influences leads to a "local" aggregation in a network context, by means of the *activation functions* introduced in the context of threshold models. In particular, we show how two of the "desirable" properties, *monotonicity* and *submodularity*, may pass from an *influence aggregation function* to its correspondent *activation function*.

4.1 Preliminaries

In this section we point out some introductive definitions, properties and concepts which will be vastly used in the rest of the chapter.

4.1.1 The role of the Associativity

Since the situations we are dealing with include $n \geq 2$ friends/neighbors, then it will be natural to work with multivariate functions. However, we want an adequately elastic framework that allows us to change between having two friends or $n > 2$ friends, without meeting particular problems. The key property, in this context, is the **associativity**, thus defined (only) for a bivariate function:

Definition 4.1. *Given a function*[1] $F : [0,1]^2 \to [0,1]$*, it is called* ***associative*** *if*

$$F(F(a,b),c) = F(a,F(b,c)), \qquad \forall\, a,b,c \in [0,1].$$

Notice that this is a property that we want for our aggregation function, mainly for two reasons: first of all, intuitively speaking it describes the situation in which I have already accumulated the influence of two friends when another one shows up, but we do not want the order of appearance to matter. In other words, we want that accumulating opinions of three people is the same that first accumulating the opinions of two and, after, another one. Second, the associativity allows us to define easily a k-variate functions, simply by iteration as we see here:

Definition 4.2. *Given a bivariate function* $F : [0,1]^2 \to [0,1]$*, we define its multivariate iterations as*

$$F^{(k)}(x_1,...,x_k) := F(F(...F(F(x_1,x_2),x_3),...,x_{k-1}),x_k), \qquad \forall\, k \geq 2,$$

for all $x_1,...,x_k \in [0,1]$*. Equivalently, the definition may be written inductively as:*

$$F^{(k)}(x_1,...,x_k) = F^{(2)}\left(F^{(k-1)}(x_1,...,x_{k-1}),x_k\right), \qquad F^{(2)} = F.$$

Throughout, whenever the context is clear, we will use indiscriminately the notation F *or* $F^{(2)}$*.*

As we will see later, this definition will have the great advantage that all properties we define for a bivariate function will extend to its multivariate versions. As notation, throughout we use F to denote the bivariate function $F^{(2)}$ and $F^{(k)}$ for its k-variate iteration.

[1] Notice that here the use of functions with domain in $[0,1]$ is just for ease of writing. Obviously, the only condition that one needs is that $F : A \times B \to C$ is such that C is a subset of $A \cap B$, whichever sets are A, B, C.

4.1.2 Submodularity

Submodularity is an important property required in the context of spread in networks, as we will see in the last chapter of this work. Similarly, here, we will see that it is important when you are thinking more *locally*, focusing on a single node of the network. Let us recall now the definition of submodularity for a set function.

Definition 4.3. *Given a set Ω, a set function $f : 2^\Omega \to \mathbb{R}$ is called **submodular** if*

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

for all $S \subseteq T \subseteq \Omega$ and $\omega \in \Omega \setminus T$.

The idea is very intuitive: the bigger a set is, the less the *marginal contribution* that a new member can add is. It is known that this definition is equivalent to:

$$f(S) + f(T) \geq f(S \cap T) + f(S \cup T)$$

for all $S, T \subseteq \Omega$. The latter definition is less immediate, but it makes clear why and how it is possible to extend the idea of submodularity for vector functions.

Definition 4.4. *A function $F : \mathbb{R}^n \to \mathbb{R}$ is called **submodular** if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*

$$\text{(SubM)} \qquad F(\mathbf{x} \wedge \mathbf{y}) + F(\mathbf{x} \vee \mathbf{y}) \leq F(\mathbf{x}) + F(\mathbf{y}),$$

where $\wedge$ and $\vee$ are the min *and* max *taken by components.*

In case of *regular* functions, there is an equivalent notion for the submodularity, which is more useful in an operational context. For references, check [Milgrom, Roberts, 1990].

Lemma 4.5 (Topkis' Characterization Theorem).
Let $F : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Then F is submodular if and only if

$$\frac{\partial^2 F}{\partial x_i \partial x_j}(\mathbf{x}) \leq 0$$

for all $\mathbf{x} \in \mathbb{R}^n$ and $i \neq j$.[2]

[2]There are no restrictions on $\frac{\partial^2 F}{\partial x_i^2}$.

4.1.3 Aggregation Functions and T-conorms

Aggregation functions are the very main objects of this chapter; in particular, we will focus on particular examples of aggregation functions called t-conorms, by means of which we want to model a *cumulative* behavior of aggregating influences. For general references about aggregation functions, one may check [Beliakov *et al.*, 2008], while for more specific references about t-conorms, [Klement *et al.*, 2000] may be used.

Definition 4.6. *An **aggregation function** is a function $f : [0,1]^n \to [0,1]$ with the properties:*

(BCs) ***boundary conditions**: $f(0,...,0) = 0$ and $f(1,...,1) = 1$;*

(Mon) ***monotonicity**: if $\mathbf{x} \leq \mathbf{y}$ then $f(\mathbf{x}) \leq f(\mathbf{y})$, where the order $\leq$ in $[0,1]^n$ is defined by components, i.e. $\mathbf{x} \leq \mathbf{y}$ if and only if $x_i \leq y_i$, for all $i = 1,...,n$.*

Definition 4.7. *An aggregation function f is called **disjunctive** if for all $\mathbf{x} \in [0,1]^n$ it is bounded by:*

$$\text{(Disj)} \qquad f(\mathbf{x}) \geq \max(\mathbf{x}) = \max\{x_1, ..., x_n\}.$$

Definition 4.8. *A **triangular conorm** (or simply **t-conorm**[3]) is a binary function $S : [0,1] \times [0,1] \to [0,1]$ which verifies the following properties for all $a, b, c \in [0,1]$:*

(Sym) ***symmetry/commutativity:** $S(a,b) = S(b,a)$;*

(Assoc) ***associativity:** $S(S(a,b),c) = S(a,S(b,c))$;*

(Mon) ***monotonicity:** $S(a,b) \leq S(a,c)$, whenever $b \leq c$;*

(Neut) ***0 is neutral element:** $S(a,0) = a$.*[4]

Example 4.9. The *geometric mean* is an aggregation function

$$f(\mathbf{x}) = \sqrt[n]{x_1 ... x_n},$$

which is not disjunctive. Any *weighted arithmetic mean*, with weights $w_1, ..., w_n \in [0,1]$ such that $\sum w_i = 1$, is an aggregation function

$$f(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i$$

which is neither symmetric nor disjunctive. In fact, usually *means* are not disjunctive, since they are bounded between their minimum and their maximum.

[3] Some texts use the notation **s-norm** for these functions.

[4] Definition 4.4 may be further generalized by using in the same formula a t-norm and its dual t-conorm, respectively, instead of the min and the max.

Example 4.10. The function $f : [0,1]^2 \to [0,1]$ defined by

$$f(x_1, x_2) = 1 - (1 - x_1)(1 - x_2)^2$$

is a disjunctive aggregation function which is not a t-conorm, since it does not verify the symmetry (SYM).

Example 4.11. The following are often considered the basic t-conorms, for all $a, b \in [0,1]^2$ defined by:

- the **maximum**: $S_{max}(a,b) = \max\{a,b\}$;
- the **probabilistic sum**: $S_{prob}(a,b) = a + b - ab$;
- the **Lukasiewicz t-conorm**: $S_{Luk}(a,b) = \min\{a+b, 1\}$;
- the **drastic t-conorm**

$$S_{dr}(a,b) = \begin{cases} 1 & \text{if } (a,b) \in]0,1]^2, \\ \max\{a,b\} & \text{otherwise.} \end{cases}$$

Remark 4.12.

- A t-conorm is a binary operation $S = *$, therefore, it is possible to consider the pair $([0,1], *)$ as a semigroup.
- The monotonicity (MON) is given in its minimal form, since a t-conorm is monotone in both components due to the commutativity (SYM).
- The boundary condition (BC) is also given in its minimal form. In fact, all t-conorms satisfy these additional boundary conditions which follow directly from the properties in the definition:

$$\begin{aligned} &\text{(NEUT)} && S(0,t) = t, \\ &\text{(ANN)} && S(1,t) = S(t,1) = 1, \end{aligned}$$

  for all $t \in [0,1]$.
- As already observed above, the associativity of a t-conorm guarantees that it is possible to extend a bivariate t-conorm to a n-ary operation by induction. Whenever we do so, we will use the notation $S^{(k)}$ for the k-variate function iteratively defined by

$$S^{(k)}(x_1, ..., x_k) = S^{(2)}\left(S^{(k-1)}(x_1, ..., x_{k-1}), x_k\right), \quad \forall x_1, ..., x_k \in [0,1],$$

  based on the (bivariate) t-conorm $S^{(2)}(\cdot,\cdot)$.

Some of the most studied t-conorms are *Archimedean* and *continuous*, mostly because for these it is possible to make a useful and detailed representation in terms of simpler *generators*.

Definition 4.13. *A t-conorm S is called* ***Archimedean*** *if*

$$(\text{Arch}) \qquad \forall\, a, b \in (0,1), \quad \exists\, k \in \mathbb{N} : \quad S^{(k)}(a, ..., a) > b.$$

A t-conorm S is called ***continuous*** *if it is continuous in every point of $[0,1]^2$, with respect to the Euclidean topology of $\mathbb{R}^2$ and $\mathbb{R}$.*[5]

Lemma 4.14. *A t-conorm verifies* (Arch) *if and only if it has the* ***Limit Property****, which is defined by:*

$$(\text{Lim}) \qquad \lim_{k\to\infty} S^{(k)}(a, ..., a) = 1, \quad \forall\, a \in]0,1[;$$

Remark 4.15. Among the prototypical examples of t-conorm given above, only the *probabilistic sum* and the *Lukasiewicz t-conorm* are both continuous and Archimedean. The *maximum* is continuous but it is not Archimedean, since it cannot satisfy (Lim); conversely, the *drastic sum* is Archimedean but not continuous. In particular, this shows that continuity is independent of the Archimedean property (and Limit property, as well).

In order to write the representation theorem, first we need to introduce the concept of *pseudo-inverse* of a monotone function, which has a particularly easy form and can be divided further in two subcases for non-increasing and non-decreasing functions giving the following definition, respectively:

Definition 4.16 (**Pseudo-inverse of a monotone function**)**.**
Given a non-constant non-increasing function $\varphi : [0,1] \to [0,1]$, its pseudo-inverse is the function $\varphi^{(-1)} : [0,1] \to [0,1]$ defined by

$$\varphi^{(-1)}(z) = \sup\{x \in [0,1] :\, \varphi(x) > z\}, \qquad \forall\, z \in [0,1].$$

Given a non-constant non-decreasing function $\psi : [0,1] \to [0,\infty]$, its pseudo-inverse is the function $\psi^{(-1)} : [0,\infty] \to [0,1]$ defined by

$$\psi^{(-1)}(z) = \inf\{x \in [0,\infty] :\, \psi(x) > z\}, \qquad \forall\, z \in [0,\infty].$$

Lemma 4.17 (**Representation of continuous Archimedean t-conorms**)**.**
For a function $S : [0,1]^2 \to [0,1]$, the following are equivalent:

1. *S is a continuous Archimedean t-conorm;*
2. *S has a* ***multiplicative generator****, i.e. there exists a function $\varphi : [0,1] \to [0,1]$, which is uniquely determined up to a positive constant exponent, such that:*
   - *φ is continuous and strictly decreasing;*

[5] Obviously, being the t-conorms defined on the compact sets $[0,1]^2$ and $[0,1]$, the continuity is equivalent to the uniform continuity.

- $\varphi(0) = 1$;
- *it holds*

$$S(a,b) = \varphi^{(-1)}(\varphi(a) \cdot \varphi(b)), \qquad \forall\, (a,b) \in [0,1]^2;$$

3. *S has an **additive generator**, i.e. there exists a function $\psi : [0,1] \to [0,+\infty]$ which is uniquely determined up to a positive multiplicative constant such that:*

   - *ψ is continuous and strictly increasing;*
   - $\psi(0) = 0$;
   - *it holds*

$$S(a,b) = \psi^{(-1)}(\psi(a) + \psi(b)) \qquad \forall\, (a,b) \in [0,1]^2.$$

Notice that the lemma states that a function $S(\cdot,\cdot)$ defined as above is necessarily a t-conorm and, in addition, it is continuous and Archimedean. For precise references, check the section 5.2 of [Klement *et al.*, 2000].

Remark 4.18. Some basic properties of the pseudo-inverse function are summarized as follows in the case of a multiplicative generator:

- $\varphi^{(-1)} : [0,1] \to [0,1]$ is non-increasing, continuous;
- $\varphi^{(-1)}|_{[0,\varphi(1)]} \equiv 1$ and $\varphi^{(-1)}$ is strictly decreasing in $[\varphi(1),1]$;
- $\varphi^{(-1)}(1) = 0$;
- $\left(\varphi^{(-1)}\right)^{(-1)} = \varphi$ if and only if $\varphi(1) = 0$.

For an additive generator the properties are basically the same, with the only exception of the last one, which becomes: $\left(\psi^{(-1)}\right)^{(-1)} = \psi$ if and only if $\psi(1) = +\infty$.
Recall that in this case $\psi^{(-1)}$ is a so-called **proper function**, because it takes values in the extended real semi-line $[0,+\infty]$ and it is such that $\psi^{(-1)}(x) < +\infty$ for at least one $x \in [0,\infty[$.

Remark 4.19. The relation between multiplicative and additive generators is expressed by the formulas

$$\varphi(x) = \exp(-\psi(x)), \quad \psi(x) = -\log \varphi(x), \qquad \forall\, x \in [0,1]$$

Example 4.20. The function $\varphi_{prob} : [0,1] \to [0,1]$, defined by

$$\varphi_{prob}(x) = 1 - x,$$

is a multiplicative generator of the probabilistic t-conorm S_{prob}, which is continuous and Archimedean. In fact, since φ_{prob} is a bijection, then its

pseudo-inverse coincides with its inverse, which coincides with φ_{prob} itself, i.e. $\varphi_{prob}^{(-1)}(x) = \varphi_{prob}^{-1}(x) = \varphi_{prob}(x) = 1 - x$. So, it is easy to check that

$$\varphi_{prob}^{(-1)}\left[\varphi_{prob}(x) \cdot \varphi_{prob}(y)\right] = 1 - (1-x)(1-y) = x + y - xy = S_{prob}(x, y).$$

Example 4.21. The functions $\psi_{Luk}(x) = x$ and $\varphi_{Luk}(x) = \exp(-x)$ are respectively an additive and a multiplicative generator for the Lukasiewicz t-conorm S_{Luk}. Notice that, in particular:

$$\varphi_{Luk}^{(-1)}(y) = \begin{cases} 1 & \text{for } x \in [0, 1/e] \\ -\log(y) & \text{for } x \in [1/e, 1]. \end{cases}$$

The generators also help when one tries to extend a bivariate t-conorm to a k-ary function, by associativity. Indeed, from Lemma 4.4.1 of [Alsina *et al.*, 2006], we get:

Lemma 4.22. *If S is a continuous and Archimedean t-conorm additively generated by $\psi : [0,1] \to [0,\infty]$, then*

$$S^{(k)}(x_1, ..., x_k) = \psi^{(-1)}(\psi(x_1) + ... + \psi(x_k))$$

for all $k \geq 2$ and $x_1, ..., x_k \in [0,1]$.

Remark 4.23 (**Duality t-norms/t-conorms**)**.**
In the literature, it is more common to talk about t-norms, which are aggregation functions that are *dual* to t-conorms in the following sense: given a t-conorm $S(x, y)$, its **dual t-norm** is $T : [0,1]^2 \to [0,1]$, often indicated by $T = S^*$, defined by

$$T(x, y) := 1 - S(1 - x, 1 - y), \qquad \forall\, (x, y) \in [0,1]^2.$$

Almost all of the results for t-norms may become, by duality, results for their dual t-conorms and this technique will be extensively used in the whole chapter. For example, if S is a t-conorm which has an additive generator $s : [0,1] \to [0,\infty]$, then its dual t-norm T has an additive generator t, meaning that

- $t : [0,1] \to [0,\infty]$ is continuous, strictly decreasing and such that $t(1) = 0$;
- for all $(x, y) \in [0,1]^2$, it holds

$$T(x, y) = t^{-1}(t(x) + t(y)).$$

Moreover, the duality between S and T replicates on the respective generators by means of the following relation:

$$t(x) = s(1 - x), \qquad \forall\, x \in [0,1].$$

4.2 Influence Aggregation Functions

This long section is organized as follows: in order to understand what we mean by *influence aggregation*, first we make a list of "reasonable" or "desirable" properties. Then we study the relations that exist among them and this will lead us to the concept of *t-conorm*. We study some properties of t-conorms and define an *influence aggregation function* as a particular type of t-conorm, with a particular focus on *submodular t-conorms*. Finally, we describe a family of *influence aggregation functions* which allows us to somehow "gradually" move between the two most important t-conorms that we will consider, i.e. the *probabilistic* S_{prob} and the *Lukasiewicz* S_{Luk}.

4.2.1 Properties of Influence Aggregation or "What is Influence Aggregation?"

We all have an intuitive idea of *influence* that an agent has on another one, and an idea of the mechanism by which influences of many agents tend to *aggregate* together. In order to understand what we exactly mean by influence aggregation and to avoid the vagueness that the intuition may have, the first and natural approach for a mathematician is to try to figure out what properties this vague idea has. A fundamental step in this direction is the translation of these intuitive properties into mathematical language. For example, our ideal concept of influence aggregation verifies that the more influencing agents we have, the higher the total influence should be; clearly, this mathematically will be represented by a sort of monotonicity.

In what follows, we try to translate these intuitive properties and describe them in a formal way. Then, we study how they link together, trying to point out the relations and implications existing among those properties. The result of this process will be the definition of *influence aggregation function*, which will only gather the essential and minimal properties, avoiding redundances.

The first step, so, is to make a list of properties that we imagine to be satisfied by the intuitive mechanism of influence aggregation when we are in the following situation: we consider (up to) n agents all influencing a fixed one, where the influence of a single agent is represented by a number in $[0, 1]$. In particular, not only we will have to study how to aggregate $k \leq n$ given influences, but also we will want our framework to be enough elastic to deal with the possible addition of further agents to those k already considered. In other words, we will also study what happens when we pass from k to $k + 1$ influencing agents.

(Dom) Because we have $k \leq n$ numbers in $[0, 1]$ representing the influence of each agent, and the result of the aggregation of those numbers, i.e. the total influence, is another number in $[0, 1]$, it is natural to consider

some functions $F^{(k)} : [0,1]^k \to [0,1]$ as objective of our study, whose **domains** are the sets $[0,1]^k$ respectively.

(BC) In case we have no influencing agent, the total influence has to be 0, i.e. it holds the **boundary condition**

$$F^{(k)}(\underbrace{0,...,0}_{k \text{ times}}) = 0,$$

for every $k \leq n$.

(Mon) The more influence weight each agent has, the higher has to be the result of the aggregation of all influences, i.e. we want each $F^{(k)}$ to satisfy the **monotonicity**: for all $k \leq n$

$$F^{(k)}(x_1,...,x_k) \leq F^{(k)}(y_1,...,y_k),$$

if $x_i \leq y_i$ for all $i = 1,...,k$.

(Assoc) The order through which we aggregate the various influences does not matter and, in particular, we may either choose to aggregate k influences and then add another one or directly aggregate $k+1$ influences, obtaining the same outcome. Formally, this is what we call **associativity**:

$$F^{(k+1)}(x_1,...,x_{k+1}) = F^{(2)}\left(F^{(k)}(x_1,...,x_k), x_{k+1}\right)$$

for all $2 \leq k \leq n-1$ and $x_1,...,x_{k+1} \in [0,1]$.

(Neut) If we have only one influencing agent, there has to be no modification in the total influence, i.e. 0 is the **neutral element** for every $F^{(k)}$:

$$F^{(k)}(0,...,0,x_i,0,...,0) = x_i$$

for all $k \leq n$, $i = 1,...,k$ and $x_i \in [0,1]$.

(Ann) If we have (at least) a *guru* among the influencing agents, i.e. an agent with an influence equal to 1, then the total influence is 1 regardless the influences of other agents, i.e. 1 is the **annihilator** for every $F^{(k)}$:

$$F^{(k)}(x_1,...,x_{i-1},1,x_{i+1},...,x_k) = 1,$$

for all $k \leq n$ and $x_1,...,x_k \in [0,1]$, regardless of the position of the 1.

(Sym) The calculation of the total influence only has to take into account the influence weights of the agents, i.e. the strength of their tie, and not their name. In other words, each agent only matters for the amount

of its influence, which is often called anonymity. Mathematically, this is represented by the **symmetry** or **commutativity** of every $F^{(k)}$:

$$F^{(k)}\left(x_{\pi(1)},...,x_{\pi(k)}\right) = F^{(k)}(x_1,...,x_k),$$

for all $k \leq n$, any permutation π of k elements and $x_1,...,x_k \in [0,1]$.

(Cont) The aggregation of influences has to be gradual, in other words, little variations in the inputs correspond to a little variation in the output, i.e. the functions $F^{(k)}$ have to be **continuous** in all point $(x_1,...,x_k) \in [0,1]^k$. Formally, for all $k \leq n$, for all $(x_1,...,x_k) \in [0,1]^k$ and $\varepsilon > 0$, there exists an open neighborhood $I_\varepsilon(x_1,...,x_k) \subset [0,1]^k$ of $(x_1,...,x_k)$ such that

$$F^{(k)}(y_1,...,y_k) \in \left(F^{(k)}(x_1,...,x_k) - \varepsilon, F^{(k)}(x_1,...,x_k) + \varepsilon\right),$$

whenever $(y_1,...,y_k) \in I_\varepsilon(x_1,...,x_k)$.

(Disj) Since we are considering an accumulation of influences, we want that the result of aggregating various influences has to be greater than each single influence; hence, the aggregation has to have a **disjunctive** behavior:

$$F^{(k)}(x_1,...,x_k) \geq \max\{x_1,...x_k\},$$

for all $k \leq n$ and $x_1,...,x_k \in [0,1]$.

(Card) If we are aggregating k influences and then another agent shows up, the passage from k to $k+1$ influences has to positively contribute in increasing the total influence, no matter how little is the influence of the new agent. This means that we want the mechanism of aggregation to verify what we call **cardinal strict monotonicity**: for all $\mathbf{x} \in [0,1]^k$ such that $F^{(k)}(\mathbf{x}) < 1$, it holds

$$F^{(k)}(\mathbf{x}) < F^{(k+1)}(\mathbf{x},\varepsilon), \qquad \forall\, 2 \leq k \leq n-1,\ \forall\, \varepsilon \in (0,1].$$

(WSM) We have already considered the monotonicity (Mon), but we also want to take into account that even a small enhancement in the influence weight of a single agent matters, in case, of course, the total influence has not already reached its maximum value of 1. This means that every $F^{(k)}$ has to satisfy what we call **weak strict monotonicity**:[6]

$$F^{(k)}(\mathbf{x}) < F^{(k)}(\mathbf{y}), \qquad \text{whenever} \quad \mathbf{x} \lneqq \mathbf{y} \quad \text{and} \quad F^{(k)}(\mathbf{x}) < 1,$$

where $(x_1,...,x_k) \lneqq (y_1,...,y_k)$ indicates that $x_i \leq y_i$ for all components but there exists at least one index j such that $x_j < y_j$.

[6]The adjective "weak" is due to the fact that after we will consider *strict t-conorms* and a concept of "strong" monotonicity. Compare the definition 4.43.

(SubM) Another property emerges when we take a careful look at the contribution of a single agent to the total influence. It may be intuitive to consider that, in situations where influences are put together, the passage from two influencing agents to three has to be different from the passage from 1000 to 1001 influencing agents, if the added agent is the same. More precisely, we imagine that in the first situation the change is bigger than the second one. In other words, we focus on the so-called *marginal contributions* of each agent, and we want them to be decreasing. As already observed in the first section of this chapter 4.1.2, this property is mathematically represented by the **submodularity** of every $F^{(k)}$, defined by:

$$F^{(k)}(\mathbf{x} \wedge \mathbf{y}) + F^{(k)}(\mathbf{x} \vee \mathbf{y}) \leq F^{(k)}(\mathbf{x}) + F^{(k)}(\mathbf{y}), \qquad \forall\, \mathbf{x}, \mathbf{y} \in [0,1]^k$$

where $\wedge$ and $\vee$ indicates the min and max taken by components, i.e. $\mathbf{x} \wedge \mathbf{y} = (\min\{x_1, y_1\}, ..., \min\{x_k, y_k\}) \in [0,1]^k$ and analogously $\mathbf{x} \vee \mathbf{y} \in [0,1]^k$.

(NSat) The last property will be somehow optional, which means that it may be either considered or not depending on the aggregation model that one wants to consider. It concerns the possibility to reach the full influenceability only in presence of (at least) an agent that acts as a *guru*. This may be described by what we call **non-saturation property** of any $F^{(k)}$:

$$F^{(k)}(x_1, ..., x_k) = 1 \qquad \Longrightarrow \qquad \exists\, i \in \{1, ..., k\} :\ x_i = 1,$$

for all $k \leq n$ and $(x_1, ..., x_k) \in [0,1]^k$.

4.2.2 From binary to n-ary functions

Now that we have defined the properties that we will study and deal with, the situation is the following: our framework comprises some functions $F^{(2)}, ..., F^{(n)}$ and some properties, which may be categorized in two different classes, the first including characteristics of each single $F^{(k)}$ separately, the second linking together the functions as part of the same "family", such as (Assoc) or (Card). The advantage of using this kind of structure is that it is enough elastic and adaptable to deal "homogenously" and "uniformly" with different situations, where we have from two to n influencing agents. Nevertheless, the clear disadvantages of this approach are that we have to clarify how these functions are related, i.e. what we really mean with "homogenously" and "uniformly", and that every property should be studied for each one of the $F^{(k)}$.

However, these problems are solved by the following proposition, which tells us that if we define the functions $F^{(k)}$ by associativity, starting from a

given bivariate $F^{(2)}$ as shown in the section 4.1.1, and if this $F^{(2)}$ has the above properties, then all these extend automatically to any other $F^{(3)}, ..., F^{(n)}$. This, in particular, tells us that in order to study the properties in the list above, the only thing that we need to do is studying them just for $F^{(2)}$, as long as the family $\{F^{(2)}, ..., F^{(n)}\}$ is defined by associativity (Assoc).

Proposition 4.24. *Let $F^{(2)} : [0,1]^2 \to [0,1]$ be a bivariate function satisfying* (BC), (Mon), (Assoc), (Neut), (Ann), (Sym), (Cont), (Disj), (WSM), (SubM) *and* (NSat) *and let $F^{(3)}, ..., F^{(n)}$ be defined iteratively by associativity*

$$F^{(k)} : [0,1]^k \longrightarrow [0,1]$$
$$F^{(k)}(x_1, ..., x_k) = F^{(2)}\left(F^{(k-1)}(x_1, ..., x_{k-1}), x_k\right).$$

Then each function $F^{(k)}$ has the above-mentioned properties (BC), (Mon), (Neut), (Ann), (Sym), (Cont), (Disj), (WSM), (SubM), (NSat) *and, in addition, the whole family $\left\{F^{(2)}, ..., F^{(n)}\right\}$ verifies* (Assoc) *and* (Card).

Proof. Most parts of the proof are made by induction, where the hypothesis on $F^{(2)}$ is the base case.

(BC) The assumption that $F^{(2)}$ verifies (BC) is the first step of the inductive proof. Then, it follows from the definition and the inductive hypothesis on $F^{(k-1)}$ that:

$$F^{(k)}(0, ..., 0) = F^{(2)}\left(F^{(k-1)}(0, ..., 0), 0\right) = F^{(2)}(0, 0) = 0.$$

(Neut) Using the same technique, let us assume the property for $F^{(2)}$ and $F^{(k-1)}$. Now, if $1 \leq i \leq k-1$, then

$$F^{(k)}(0, ..., 0, x_i, 0, ..., 0) = F^{(2)}\left(F^{(k-1)}(0, ..., 0, x_i, 0, ..., 0), 0\right) = \\ = F^{(2)}(x_i, 0) = x_i.$$

Otherwise, if $i = k$, then, using (BC),

$$F^{(k)}(0, ..., 0, x_k) = F^{(2)}\left(F^{(k-1)}(0, ..., 0), x_k\right) = F^{(2)}(0, x_k) = x_k.$$

Clearly, an alternative proof may be simply made assuming the symmetry (Sym).

(Ann) Proceeding exactly as in the previous step one gets either

$$F^{(k)}(x_1, ..., x_{i-1}, 1, x_{i+1}, ..., x_k) = \\ = F^{(2)}\left(F^{(k-1)}(x_1, ..., x_{i-1}, 1, x_{i+1}, ..., x_{k-1}), x_k\right) = \\ = F^{(2)}(1, x_k) = 1,$$

or $F^{(k)}(x_1, ..., x_{k-1}, 1) = F^{(2)}\left(F^{(k)}(x_1, ..., x_{k-1}), 1\right) = 1$.

(Sym) For $k = 3$ it follows straightforward from the symmetry of $F^{(2)}$, indeed

$$F^{(3)}(a,b,c) = F^{(2)}(F^{(2)}(a,b),c) = F^{(2)}(F^{(2)}(b,a),c) = F^{(3)}(b,a,c)$$

and analogously it may be done in the other similar cases. For $k \geq 3$ in general, noting that any permutation of k elements may be written as product of 2-cycles, this implies that it is suffices to use the associativity (Assoc), in order to group together the elements two by two, and then to apply the symmetry (Sym) for pairs of elements.

(Mon) Given $(\mathbf{x}, x') \leq (\mathbf{y}, y') \in [0,1]^k$, where $\mathbf{x}, \mathbf{y} \in [0,1]^{k-1}$ and $x', y' \in [0,1]$, then using that $F^{(2)}$ and $F^{(k-1)}$ verify (Mon) and applying induction, one gets:

$$\begin{aligned} F^{(k)}(\mathbf{x}, x') = F^{(2)}\left(F^{(k-1)}(\mathbf{x}), x'\right) \leq F^{(2)}\left(F^{(k-1)}(\mathbf{y}), x'\right) \leq \\ \leq F^{(2)}\left(F^{(k-1)}(\mathbf{y}), y'\right) = F^{(k)}\left(\mathbf{y}, y'\right). \end{aligned}$$

(Disj) The disjunctivity, in general, is implied by (Mon) and (Neut). Indeed, given $\mathbf{x} \in [0,1]^k$, let us set $\overline{x} = \max\{x_1, ..., x_k\}$. Then

$$F^{(k)}(\mathbf{x}) \geq F^{(k)}(0, ..., 0, \overline{x}, 0, ..., 0) = \overline{x} = \max\{x_1, ..., x_k\}.$$

However, the property may be easily verified by induction as follows:

$$\begin{aligned} F^{(k+1)}(x_1, ..., x_{k+1}) = F^{(2)}\left(F^{(k)}(x_1, ..., x_k), x_{k+1}\right) \geq \\ \geq \max\left\{F^{(k)}(x_1, ..., x_k), x_{k+1}\right\} \geq \\ \max\{\max\{x_1, ..., x_k\}, x_{k+1}\} = \max\{x_1, ..., x_{k+1}\}. \end{aligned}$$

(Cont) It is known[7] that the continuity for monotone k-variate functions $[0,1]^k \to [0,1]$ is equivalent to the continuity in each component. Hence, in our case it is suffices to use the continuity of $F^{(2)}$ and $F^{(k-1)}$ to obtain the continuity of $F^{(k)}$.

(WSM) Let us consider $x, y \in [0,1]^k$ such that $x \lneqq y$ and $F^{(k)}(x) < 1$. Without loss of generality, it may be assumed that

$$\begin{cases} x_1 < y_1 \\ x_i \leq y_i, \ \forall\, i = 2, ..., k. \end{cases}$$

Now, first of all, this assumption that we have just made implies two things:

[7] See, for example, Lemma 2.1.2 of [Alsina *et al.*, 2006]: *Any function which is continuous and non-decreasing in each component is jointly continuous.*

- $(x_1, ..., x_{k-1}) \lneqq (y_1, ..., y_{k-1})$;
- $F^{(k-1)}(x_1, ..., x_{k-1}) < 1$; otherwise, if it were not true, then it would hold that $F^{(k)}(x_1, ..., x_k) = F^{(2)}\left(F^{(k-1)}(x_1, ..., x_{k-1}), x_k\right) = F^{(2)}(1, x_k) = 1$, because of (BC).

In addition, due to the inductive hypothesis it holds true that

$$F^{(k-1)}(x_1, ..., x_{k-1}) < F^{(k-1)}(y_1, ..., y_{k-1}).$$

Consequently, one has that

$$\left(F^{(k-1)}(x_1, ..., x_k), x_k\right) \lneqq \left(F^{(k-1)}(y_1, ..., y_k), y_k\right)$$

and then

$$F^{(2)}\left(F^{(k-1)}(x_1, ..., x_k), x_k\right) \equiv F^{(k)}(x) < 1,$$

which follows from the definition and from one hypothesis. Finally:

$$F^{(k)}(x) = F\left(F^{(k-1)}(x_1, ..., x_k), x_k\right) < \\ < F\left(F^{(k-1)}(y_1, ..., y_k), y_k\right) = F^{(k)}(y),$$

which is what we needed.

(SubM) Concerning the submodularity, see next section.

(NSat) As usual, the proof is inductive: $F^{(2)}$ satisfies the property and let us assume that so does $F^{(k-1)}$. Now, if $F^{(k)}(x_1, ..., x_k) = 1$, then using the definition one gets

$$1 = F^{(k)}(x_1, ..., x_k) = F^{(2)}\left(F^{(k-1)}(x_1, ..., x_{k-1}), x_k\right)$$

which implies either $F^{(k-1)}(x_1, ..., x_{k-1}) = 1$ or $x_k = 1$. In the first case, from the inductive hypothesis on $F^{(k-1)}$, we deduce that there exists $1 \leq i \leq k-1$ such that $x_i = 1$, which implies the thesis; while in the latter case, it is directly proved.

(Assoc) The associativity simply follows from the construction.

(Card) In this step, we do not use induction. It suffices to use the properties (WSM) and (Neut) of $F^{(2)}$: take $\mathbf{x} \in [0,1]^k$ such that $F^{(k)}(\mathbf{x}) < 1$ and $\varepsilon \in]0,1]$; then

$$F^{(k+1)}(\mathbf{x}, \varepsilon) = F^{(2)}\left(F^{(k)}(\mathbf{x}), \varepsilon\right) > F^{(2)}\left(F^{(k)}(\mathbf{x}), 0\right) = F^{(k)}(\mathbf{x}).$$

□

Submodularity: from binary to n-ary functions

A very useful concept related to the submodularity is given by the following definition and lemma.

Definition 4.25. *Given $F : [0,1]^n \to [0,1]$, $\boldsymbol{\xi} \in [0,1]^n$, $i_j \in \{1,...,n\}$, for all $j = 1,...,k$, such that $i_1 \leq ... \leq i_k$, then the k-**dimensional section** of F is the k-ary function $F_{(\boldsymbol{\xi},(i_1,...,i_k))} : [0,1]^k \to [0,1]$, defined by*

$$F_{(\boldsymbol{\xi},(i_1,...,i_k))}(\mathbf{z}) = F(\mathbf{x}), \qquad \forall\, \mathbf{z} \in [0,1]^k,$$

where $\mathbf{x} \in [0,1]^n$ is such that

$$x_i = \begin{cases} z_j, & \text{if } \exists\, j : \ i = i_j, \\ \xi_i, & \text{otherwise.} \end{cases}$$

Definition 4.26. *A (binary) function $g : \mathbb{R}^2 \to \mathbb{R}$ has **decreasing differences** if for all $b \leq b' \in \mathbb{R}$ the function*

$$\begin{aligned} &\mathbb{R} \longrightarrow \mathbb{R} \\ &t \longmapsto g(t,b') - g(t,b) \end{aligned}$$

is decreasing (in t).

Lemma 4.27. *An n-ary function $F : \mathbb{R}^n \to \mathbb{R}$ has decreasing differences if any 2-dimensional section of F has decreasing differences, i.e. for every $\mathbf{x} \in \mathbb{R}^n$ and different indexes $i, j = 1,...,n$, any function of two variables $F_{\mathbf{x},(i,j)} : \mathbb{R}^2 \to \mathbb{R}$ defined by*

$$F_{\mathbf{x},(i,j)}(a,b) = F(x_1,...,x_{i-1},a,x_{i+1},...,x_{j-1},b,x_{j+1},...,x_n),$$

has decreasing differences.

For references about the lemma below, see the Proposition 2.3 in [Manzi, 2010] for the first part and the section 2.3 of [Simchi-Levi *et al.*, 2005] for the second.

Lemma 4.28.

1. *An n-ary function $G : [0,1]^n \to [0,1]$ is submodular if and only if its 2-dimensional sections are submodular.*

2. *$G : \mathbb{R}^n \to \mathbb{R}$ is submodular if and only if it has decreasing differences.*

Proposition 4.29. *Let $F^{(2)} : [0,1]^2 \to [0,1]$ be a bivariate function which is symmetric* (Sym)*, associative* (Assoc) *and monotone* (Mon)*, and let its multivariate iterations $F^{(k)} : [0,1]^k \to [0,1]$ for $k = 3,...,n$ be defined by associativity. Then if $F^{(2)}$ verifies submodularity* (SubM)*, so does every $F^{(k)}$.*

Proof. By the symmetry of $F^{(k)}$ and the second part of the previous lemma, it suffices to verify that, fixed a vector $\boldsymbol{\xi} \in [0,1]^{k-2}$, the function

$$G_{\boldsymbol{\xi}}(a,b) := F^{(k)}(a,b,\xi_1,...,\xi_{k-2}) = F^{(2)}\left(F^{(2)}(a,b), F^{(k-2)}(\boldsymbol{\xi})\right)$$

has decreasing differences. So, it has to be proven that given $b \leq b'$ and $t \leq t'$, it holds

$$G_{\boldsymbol{\xi}}(t',b') - G_{\boldsymbol{\xi}}(t',b) \leq G_{\boldsymbol{\xi}}(t,b') - G_{\boldsymbol{\xi}}(t,b)$$

which corresponds to

$$F^{(2)}\left(F^{(2)}(t',b'),A\right) - F^{(2)}\left(F^{(2)}(t',b),A\right) \leq \\ \leq F^{(2)}\left(F^{(2)}(t,b'),A\right) - F^{(2)}\left(F^{(2)}(t,b),A\right),$$

where, for ease of notation, we have defined $A := F^{(k-2)}(\boldsymbol{\xi})$. Now, using the properties of associativity, we can replace the expression above with

$$F^{(3)}\left(t',b',A\right) - F^{(3)}\left(t',b,A\right) \leq F^{(3)}\left(t,b',A\right) - F^{(3)}\left(t,b,A\right)$$

and, eventually, by using the definition of $F^{(3)}$:

$$F^{(2)}\left(t',F^{(2)}(b',A)\right) - F^{(2)}\left(t',F^{(2)}(b,A)\right) \leq \\ \leq F^{(2)}\left(t,F^{(2)}(b',A)\right) - F^{(2)}\left(t,F^{(2)}(b,A)\right).$$

The last inequality follows from (Mon) and the hypothesis, which guarantees that $F^{(2)}$ has decreasing differences. □

Alternative proof. Defined $G_{\boldsymbol{\xi}}(a,b)$ as above, the inequality to be proven may be written as follows

$$F^{(2)}\left(F^{(2)}(t',b'),F^{(k-2)}(\boldsymbol{\xi})\right) - F^{(2)}\left(F^{(2)}(t',b),F^{(k-2)}(\boldsymbol{\xi})\right) \leq \\ \leq F^{(2)}\left(F^{(2)}(t,b'),F^{(k-2)}(\boldsymbol{\xi})\right) - F^{(2)}\left(F^{(2)}(t,b),F^{(k-2)}(\boldsymbol{\xi})\right)$$

which can be re-written, by associativity, in the following form

$$F^{(2)}\left(t',F^{(k-1)}(b',\boldsymbol{\xi})\right) - F^{(2)}\left(t',F^{(k-1)}(b,\boldsymbol{\xi})\right) \leq \\ \leq F^{(2)}\left(t,F^{(k-1)}(b',\boldsymbol{\xi})\right) - F^{(2)}\left(t,F^{(k-1)}(b,\boldsymbol{\xi})\right).$$

Now, in order to prove the last inequality, one only needs to observe that, since $b \leq b'$, we have $F^{(k-1)}(b',\boldsymbol{\xi}) \geq F^{(k-1)}(b,\boldsymbol{\xi})$ by the monotonicity of $F^{(k-1)}$, and finally that it has been assumed $t \leq t'$, so the inequality is the mere application of the decreasing differences of $F^{(2)}$. □

Remark 4.30. It is worth noticing that the situation of the above proposition is much more intricate if one does not assume the associativity and the symmetry of F: see [Manzi, 2009].

4.2.3 Relations among the properties of Influence Aggregation

Having made a list of properties, what we want to do now is trying to understand what relations exist among them. Once we realize which of them implies the others and which are independent of the others, then it will be possible to make a list of the *minimal* or *essential* properties, which will constitute the basis for the formal definition of Influence Aggregation Function (see Corollary 4.63). As already observed before, thanks to Proposition 4.24, in order to fully study the implications among the properties of the previous section, it suffices to study them just for the bivariate function $F^{(2)} = F$.[8]

Remark 4.31. Because of the symmetry (Sym), some properties of F may be rewritten only considering the first component, such as:

(Neut) $F(x, 0) = x$ for all $x \in [0, 1]$.

(Ann) $F(x, 1) = 1$ for all $x \in [0, 1]$.

(Mon) $F(x, y) \leq F(x, z)$ for all $x, y, z \in [0, 1]$ such that $y \leq z$.

Remark 4.32. As already mentioned in the preliminary section 4.1.3, a function $F : [0, 1]^2 \to [0, 1]$ exactly verifying (Mon), (Neut), (Sym) and (Assoc) is called **t-conorm**. In addition, any such a function necessarily satisfies:

- (Bc), which directly follows from (Neut): $F(0, 0) = 0$;
- (Disj), which follows from (Mon) and (Neut): $F(x, y) \geq F(\max\{x, y\}, 0) = \max\{x, y\}$;
- (Ann), which follows from (Disj) and, so, from (Mon) and (Neut) as well: $F(x, 1) \geq \max\{x, 1\} = 1$.

To summarize:

$$\text{t-conorm} \implies \text{(Mon), (Neut)} \implies \begin{cases} \text{(Bc)} \\ \\ \text{(Disj)} \implies \text{(Ann)}. \end{cases}$$

On the other hand, the opposite direction is somehow illustrated by the following result, which is the Lemma 2.1.13 of [Alsina *et al.*, 2006].

Lemma 4.33. *If $F : [0, 1]^2 \to [0, 1]$ verifies* (Bc), (Assoc), (Ann) *and* (Cont)*, then it also satisfies* (Mon) *and* (Neut)*.*

[8] Throughout, whenever the context is clear, we use both notations F and $F^{(2)}$ indiscriminately.

It is not hard to check that the properties that define a t-conorm are independent from each other:

Lemma 4.34. *Consider the properties* (Mon), (Neut), (Sym) *and* (Assoc). *There exist functions* $G : [0,1]^2 \to [0,1]$ *that satisfy only three of them and do not satisfy the remaining one.*

Proof. Let us define the functions $G_i : [0,1]^2 \to [0,1]$ for $i = 1,2,3,4$ such that:

$$G_1(x,y) = \begin{cases} 1/2 & \text{if } (x,y) \in]0,1[^2, \\ \max\{x,y\} & \text{otherwise;} \end{cases}$$

$$G_2(x,y) \equiv 1;$$

$$G_3(x,y) = \begin{cases} 1 & \text{if } (x,y) \in [0,1/2] \times [0,1[, \\ \max\{x,y\} & \text{otherwise;} \end{cases}$$

$$G_4(x,y) = 1 - (1-x)(1-y)[1 - \min\{x,y\}].$$

It holds that G_i does not verify (Mon), (Neut), (Sym) and (Assoc), respectively for $i = 1,2,3,4$. However, each G_i satisfies the remaining three properties. □

In the next proposition, we verify that the rest of "desirable" properties from the list of the section 4.2.1 is independent of the ones which define a t-conorm.

Lemma 4.35. *The properties defining a t-conorm* (Mon), (Neut), (Sym) *and* (Assoc) *are independent of* (Cont), (Card), (WSM), (SubM) *and* (NSat).

Proof. We base this proof on showing some examples of t-conorms (from the preliminary section) which do or do not verify the properties in the statement.

(Cont) The drastic t-conorm S_{dr} is not continuous, while so is the probabilistic S_{prob}.

(Card) The maximum S_{max} does not verify the property, while the Lukasiewicz t-conorm S_{Luk} does. To prove that, we first need to write the k-ary iteration of both, which are quite immediate to find:

$$S^{(k)}_{max}(x_1,...,x_k) = \max\{x_1,...,x_k\},$$

$$S^{(k)}_{Luk}(x_1,...,x_k) = \min\left\{1,\ \sum_{i=1}^{k} x_i\right\}.$$

Now, it is obvious that $S^{(k)}_{max}(1/2,...,1/2) = 1/2 = S^{(k+1)}_{max}(1/2,...,1/2)$; whereas, for any $\varepsilon \in]0,1]$ and any $\mathbf{x} \in [0,1]^k$ such that $S^{(k)}_{Luk}(\mathbf{x}) < 1$, it holds that

$$S^{(k)}_{Luk}(\mathbf{x},\varepsilon) = \min\{1,\ x_1 + ... + x_k + \varepsilon\} > x_1 + ... + x_k = S^{(k)}_{Luk}(\mathbf{x}).$$

(WSM) Obviously, the maximum S_{max} does not verify this property, since $S_{max}(1/2, 1/2) = 1/2 = S_{max}(1/2, 0)$, despite that $(1/2, 1/2) > (1/2, 0)$. On the other hand, S_{Luk} verifies it: thanks to (SYM), without loss of generality we can assume $(x_1, y_1) < (x_2, y_2)$ with $x_1 \leq x_2$ but $y_1 < y_2$ and both $S_{Luk}(x_1, y_1), S_{Luk}(x_2, y_2) < 1$. Then, by definition

$$S_{Luk}(x_1, y_1) = x_1 + y_1 < x_2 + y_2 = S_{Luk}(x_2, y_2).$$

(SUBM) The drastic S_{dr} is not submodular: for example, by taking two vectors $(x, 0), (0, y)$, with $x, y > 0$ such that $x + y < 1$, one gets that $(x, 0) \vee (0, y) = (x, y)$ and $(x, 0) \wedge (0, y) = (0, 0)$ and so:

$$S_{dr}((x, 0) \vee (0, y)) + S_{dr}((x, 0) \wedge (0, y)) = 1 + 0 > \\ > x + y = S_{dr}(x, 0) + S_{dr}(0, y).$$

On the other hand, for the smooth $S_{prob} \in \mathcal{C}^2$ one can easily differentiate twice, use the Topkis' Theorem 4.5 and see that S_{prob} is submodular:

- $\dfrac{\partial S_{prob}}{\partial x}(x, y) = \dfrac{\partial}{\partial x}[1 - (1 - x)(1 - y)] = 1 - y,$
- $\dfrac{\partial^2 S_{prob}}{\partial x \partial y}(x, y) = -1 \leq 0.$

(NSAT) The Lukasiewicz t-conorm S_{Luk} obviously does not verify the property, since $S_{Luk}(1/2, 1/2) = \min\{1, 1/2 + 1/2\} = 1$, whereas the probabilistic $S_{prob}(x, y) = 1 - (1 - x)(1 - y)$ is such that $S_{prob}(x, y) = 1$ if and only if $x = 1$ or $y = 1$.

□

By checking accurately the previous proof, one may see that we have proved something stronger than what is in the statement, in particular:

Corollary 4.36. *A continuous t-conorm does not necessarily satisfy neither* (CARD)*, nor* (WSM)*, nor* (NSAT)*.*

Proof. It suffices to notice that S_{max} is a continuous t-conorms which does not verify the required properties. □

Now that we are trying to find the relations that exist among the "desirable" properties, we want to show an example in order to prove that there exists (at least) one function which indeed satisfies all of them.

Example 4.37. A first (and fundamental) example of function verifying all the "desirable" properties of the section 4.2.1 is the *probabilistic t-conorm*

$$S_{prob} : [0, 1]^2 \to [0, 1], \qquad S_{prob}(x, y) := x + y - xy.$$

From the proof above, it follows that S_{prob} is continuous, submodular and does not saturate. There only remain to verify (CARD) and (WSM).

- (CARD) is easy to prove, because

$$S^{(k)}_{prob}(x_1, ..., x_k) = 1 - \prod_{i=1}^{k}(1 - x_i).$$

- Concerning (WSM), let us take $(x_1, y_1) < (x_2, y_2)$ such that $S_{prob}(x_1, y_1) < 1$ and, without loss of generality, assume that $x_1 < x_2$ and $y_1 \leq y_2$. Then

$$\begin{cases} 1 - x_1 > 1 - x_2 \\ 1 - y_1 \geq 1 - y_2 \end{cases} \implies (1 - x_1)(1 - y_1) > (1 - x_2)(1 - y_2),$$

  which implies $S_{prob}(x_1, y_1) < S_{prob}(x_2, y_2)$.

Example 4.38. Another fundamental example is the *Lukasiewicz t-conorm* S_{Luk}, defined by $S_{Luk}(x, y) := \min\{x + y, 1\}$. It is also a t-conorm verifying all the "desirable" properties but (NSAT), as already seen in the previous proof. As also already mentioned there, it satisfies (CARD) and (WSM), and, then, the only remaining properties to be verified are (CONT) and (SUBM):

- Since S_{Luk} is monotone, in order to verify the continuity it suffices to prove that it is continuous in each component. Fixed $y \in [0, 1]$, then

$$x \longmapsto S_{Luk}(x, y) = \min\{x + y, 1\} = \begin{cases} x + y & \text{for } x \in [0, 1 - y], \\ 1 & \text{for } x \in [1 - y, 1], \end{cases}$$

  which is a piecewise linear continuous function.

- To verify the submodularity, let us define two subsets of $[0, 1]^2$:

$$A := \left\{(a, b) \in [0, 1]^2 : \ a + b < 1\right\}, \qquad B := A^c,$$

  It is obvious that $S_{Luk}(a, b) < 1$ if and only if $(a, b) \in A$, while $S_{Luk}(a, b) = 1$ if and only if $(a, b) \in B$; in addition to that, note that the restriction $S_{Luk}|_A$ is linear (as well as $S_{Luk}|_B$.) Now, let us prove that

$$S_{Luk}(\mathbf{x} \vee \mathbf{y}) + S_{Luk}(\mathbf{x} \wedge \mathbf{y}) \leq S_{Luk}(\mathbf{x}) + S_{Luk}(\mathbf{y}),$$

  for all $\mathbf{x}, \mathbf{y} \in [0, 1]^2$. If both $\mathbf{x}, \mathbf{y} \in A$, then it suffices to use the linearity of $S_{Luk}|_A$, that implies the submodularity[9]. In the other case when at least one of $\mathbf{x}, \mathbf{y} \in B$, for example if $\mathbf{x} \in B$, then $S_{Luk}(\mathbf{x} \wedge \mathbf{y}) \leq S_{Luk}(\mathbf{y})$ by monotonicity, and

$$S_{Luk}(\mathbf{x} \vee \mathbf{y}) \leq 1 = S_{Luk}(\mathbf{x}).$$

  So, the inequality of the submodularity holds.

[9] In general, a linear function is obviously submodular (as well as supermodular, with the obvious definition that takes the opposite inequality). In fact, a linear function is modular, which means that the inequality is indeed an equality.

Continuous and Archimedean T-conorms

In the following, we will see that the object of our interest is a subclass of the continuous and Archimedean t-conorms. In particular, we will link the *Archimedean property* (ARCH) seen in the section 4.1.3 to the *weak strict monotonicity* (WSM); then we will focus on the *non-saturation* (NSAT) and, eventually, on the *submodularity* (SUBM).

Proposition 4.39. *Let $F : [0,1]^2 \to [0,1]$ be an associative function and $F^{(k)}$ its k-ary iteration. If F satisfies* (NEUT) *and* (WSM)*, then it also satisfies* (CARD)*.*

Proof. Given $\mathbf{x} \in [0,1]$ such that $F^{(k)}(\mathbf{x}) < 1$ and $\varepsilon \in]0,1]$. In particular, from (NEUT) it follows that $F\left(F^{(k)}(\mathbf{x}), 0\right) = F^{(k)}(\mathbf{x}) < 1$ and, so, (WSM) may be applied and one gets:

$$F^{(k+1)}(\mathbf{x}, \varepsilon) = F\left(F^{(k)}(\mathbf{x}), \varepsilon\right) > F\left(F^{(k)}(\mathbf{x}), 0\right) = F^{(k)}(\mathbf{x}).$$

□

Proposition 4.40. *Let S be a continuous t-conorm. Then:*

$$S \text{ is Archimedean} \iff S \text{ is weakly strict monotone.}$$

Proof.

($\Leftarrow$) The proof follows from the following lemma, which is taken from [Klement *et al.*, 2000].

Lemma 4.41. *A number $a \in (0,1)$ is called* ***idempotent element*** *for the t-conorm S if $S(a,a) = a$. It holds that:*

1. *If S is continuous, then: a idempotent if and only if $S(a,x) = \max(a,x)$ for all $x \in [0,1]$;*
2. *If S is continuous and it has no idempotent elements, then S is Archimedean.*

This lemma implies that it suffices to prove that S has no idempotent elements. Proceeding by *reduction ad absurdum*, suppose that there exists $a \in (0,1)$ idempotent for S. Then, one can choose x, y such that

$$0 < x < y < a < 1,$$

and, in particular, this implies (by the lemma):

$$(x,a) \lneqq (y,a) \quad \text{and} \quad S(x,a) = \max(x,a) = a < 1.$$

So, by using the weak strict monotonicity, one gets:

$$a = \max(x,a) = S(x,a) < S(y,a) = \max(y,a) = a,$$

which is a contradiction.

($\Rightarrow$) Using the representation theorem for continuous and Archimedean t-conorms from the section 4.1.3, once we assume that S is a continuous, Archimedean t-conorm, then it has to have a *multiplicative generator* ψ, with

$$S(x_1, x_2) = \psi^{(-1)}(\psi(x_1)\psi(x_2)).$$

Now, given $(x_1, x_2), (y_1, y_2) \in [0,1]^2$ with $\mathbf{x} \lneqq \mathbf{y}$ and $S(\mathbf{x}) < 1$, then we need to prove that $S(\mathbf{x}) < S(\mathbf{y})$. Without loss of generality, it may be assumed that $x_1 < y_1$ and $x_2 \leq y_2$, and for the strict decreasingness of ψ:

$$\begin{cases} \psi(x_1) > \psi(y_1) \\ \psi(x_2) \geq \psi(y_2) \end{cases} \text{, and then } \underbrace{\psi(x_1)\psi(x_2)}_{B:=} > \underbrace{\psi(y_1)\psi(y_2)}_{A:=},$$

which may simply be re-written as $A < B$. As a result, the thesis becomes $\psi^{(-1)}(B) < \psi^{(-1)}(A)$.
The strictly decreasingness of ψ implies that $\psi^{(-1)}$ is also strictly decreasing on $\psi([0,1]) = [\psi(1), 1]$, while it is constant elsewhere, i.e. $\psi^{(-1)}|_{[0,\psi(1)]} \equiv 1$. Now, in order to clarify the situation, it is worth splitting the cases, depending on the position of A and B with respect to $\psi(1)$ in the interval $[0,1]$:

- if $\psi(1) \leq A < B$, then $\psi^{(-1)}(A) < \psi^{(-1)}(B)$, for the strict decreasingness of $\psi^{(-1)}$;
- if $A \leq \psi(1) < B$, then $1 = \psi^{(-1)}(A) < \psi^{(-1)}(B) = S(x)$, which is true by hypothesis;
- if $A < B \leq \psi(1)$, then $\psi^{(-1)}(A) = \psi^{(-1)}(B) = 1$, but it cannot be possible, since we have assumed $\psi^{(-1)}(B) = S(x) < 1$.

So, the proof is completed.

□

Remark 4.42. From what we have seen so far, *continuous and Archimedean t-conorms* verify all "desirable" properties but (NSAT) and (SUBM).

From the previous remark, one may easily understand the reason why in the next sections we will concentrate on studying the relations that exist between **continuous Archimedean t-conorms** and the two properties **non-saturation** (NSAT) and **submodularity** (SUBM).

Non-Saturation and Strict T-conorms

First, we partition the *continuous and Archimedean t-conorms* right in the two subclasses **strict t-conorms** and **nilpotent t-conorms**, in a certain (and precise) sense "represented" respectively by S_{prob} and S_{Luk}. Then, we find that the *strict t-conorms* are exactly those verifying the *non-saturation* property, whereas the *nilpotent t-conorms* saturate, exactly as S_{Luk} does.

Definition 4.43.

- *A t-conorm S is called* ***strictly monotone*** *(on $[0,1[^2$) if*

$$\text{(SM)} \qquad S(x,y) < S(x,z) \quad \textit{whenever} \quad x < 1,\ y < z.$$

- *A t-conorm S satisfies the* ***cancellation law*** *if*

$$\text{(CL)} \qquad S(x,y) = S(x,z) \quad \Longrightarrow \quad x = 1 \text{ or } y = z.$$

The *cancellation law*, as the name suggests, highlights the behavior of (*strict*) *t-conorms* as semigroup/binary operations in $[0,1]^2$. However, the following lemma gives the equivalence between the two previous properties, and the proof may be found in [Klement *et al.*, 2000].

Lemma 4.44. *For any t-conorm S, it holds that S is strictly monotone if and only if it satisfies the cancellation law.*

The following proposition justifies the term "weak" used in the definition of the *weak strict monotonicity* (WSM).

Proposition 4.45. *Let S be a t-conorm. If S is strictly monotone, then it is weakly strict monotone, i.e.*

$$\text{(SM)} \quad \Longrightarrow \quad \text{(WSM)}$$

Proof. Let us take $(x_1,x_2),(y_1,y_2) \in [0,1]^2$. Without loss of generality, it can be assumed that

$$\begin{cases} x_1 \leq y_1, \\ x_2 < y_2, \end{cases} \quad \text{and} \quad S(x) < 1.$$

We need to prove that $S(x) < S(y)$, and in order to do that it suffices to apply the strict monotonicity, which may be automatically applied if we simply prove that $x_1 < 1$. But, if it were $x_1 = 1$, then we would have $S(x_1,x_2) = S(1,x_2) = 1$, which contradicts $S(x) < 1$. □

Definition 4.46. *Given a continuous t-conorm S, it is called:*

- ***strict*** *if it is strictly monotone* (SM)*;*
- ***nilpotent*** *if any $x \in (0,1)$ is a nilpotent element, i.e.*

$$\forall\, x \in (0,1), \quad \exists\, k \geq 2 : \ S^{(k)}(x,...,x) = 1.$$

Remark 4.47. Notice that if in Proposition 4.45 the t-conorm S were also continuous, then it would be strict, by definition. By using the equivalence between (WSM) and (ARCH) seen in the previous section, this would guarantee that a strict t-conorm S is necessarily Archimedean.

Example 4.48. As already sketched out, the *probabilistic t-conorm* S_{prob} is *strict*, while the *Lukasiewicz t-conorm* S_{Luk} is *nilpotent*. To verify these assertions:

- for S_{prob} it is easier to verify the equivalent cancellation law (CL). Indeed, $S_{prob}(x,y) = S_{prob}(x,z)$ means

$$1-(1-x)(1-y) = 1-(1-x)(1-z),$$

from which either $1-x=0$ or $1-y=1-z$, i.e. $x=1$ or $y=z$;

- for S_{Luk}, since

$$S^{(k)}_{Luk}(x_1,...,x_k) = \min\left\{1,\ \sum_{i=1}^{k} x_i\right\},$$

then, given $x \in (0,1)$, it suffices to take $k \in \mathbb{N}$ such that $k > 1/x$ to obtain $\sum_{i=1}^{k} x = kx > 1$ and, so

$$S^{(k)}_{Luk}(x,...,x) = \min\left\{1,\ \sum_{i=1}^{k} x\right\} = 1,$$

as we wanted.

Of course, S_{prob} cannot be nilpotent and S_{Luk} cannot be strict. Indeed:

- $S^{(k)}_{prob}(1/2,...,1/2) = 1 - \prod_{i=1}^{k} \frac{1}{2} = 1 - \frac{1}{2^k} < 1, \quad \forall\, k \in \mathbb{N}$;
- $S_{Luk}(0.5, 0.6) = \min\{1,\ 0.5+0.6\} = 1 = S_{Luk}(0.5, 0.7)$.

The example above is not a coincidence: indeed, it is known that *continuous, Archimedean t-conorms* can only be either *strict* or *nilpotent*. In addition to this, not only the class of continuous, Archimedean t-conorms may only contain strict or nilpotent t-conorms, but also the probabilistic t-conorm S_{prob} and the Lukasievicz t-conorm S_{Luk} are the two main prototypes of such t-conorms, as precisely stated in the lemma below.

Lemma 4.49. *Every continuous, Archimedean t-conorm S is necessarily either strict or nilpotent. Moreover, there exists a strictly increasing bijection $f : [0,1] \to [0,1]$, called* isomorphism*, such that:*

- *S is strict if and only if it is isomorphic to S_{prob}, i.e. for all $x, y \in [0,1]$:*

$$S(x,y) = f^{-1}\Big(S_{prob}\big(f(x), f(y)\big)\Big) \equiv f^{-1}\Big(1 - \big(1-f(x)\big)\big(1-f(y)\big)\Big);$$

- *S is nilpotent if and only if it is isomorphic to S_{Luk}, i.e.*

$$S(x,y) = f^{-1}\Big(S_{Luk}\big(f(x),f(y)\big)\Big) \equiv f^{-1}\Big(\min\big\{f(x)+f(y),1\big\}\Big).$$

The previous lemma is closely related to the representation by (*additive* and *multiplicative*) *generators* that continuous Archimedean t-conorms have, given by the *Representation Theorem* of section 4.1.3. That theorem may be summarized and further refined as follows.

Lemma 4.50.

$$\begin{array}{c} S:[0,1]^2 \longrightarrow [0,1] \\ (\textsc{Cont}),\ (\textsc{Arch})\ \textit{t-conorm} \end{array} \iff \begin{array}{c} \exists\ \psi:[0,1]\longrightarrow[0,+\infty] \\ \textit{continuous, strictly increasing,} \\ \psi(0)=0,\ \textit{such that} \\ S(x,y)=\psi^{-1}\big(\psi(x)+\psi(y)\big). \end{array}$$

Furthermore:

- *S is strict if and only if $\psi(1)=+\infty$;*
- *S is nilpotent if and only if $\psi(1)<+\infty$.*

Though continuous, Archimedean t-conorms are well separated in two different classes, there is always a way to uniformly approximate them as limit of strict t-conorms:

Lemma 4.51. *Every continuous, Archimedean t-conorms is the uniform limit of strict t-conorms.*

Proof. It suffices to adapt, by duality, Theorem 2.1.12 of [Alsina *et al.*, 2006]. □

Finally, we want to study better the relation between the non-saturation property and the strict monotonicity, particularly in the case of continuous, Archimedean t-conorms. Let us recall that a t-conorm S has the *non-saturation* property (NSat) if $S(x,y)=1$ happens only in the case $x=1$ or $y=1$.

Proposition 4.52. *Let S be a continuous, Archimedean t-conorm. Then:*

$$S \textit{ is strict} \iff S \textit{ verifies non-saturation.}$$

Proof.

($\Leftarrow$) Let us assume that S is not strict, which necessarily implies that S has to be nilpotent. There has to exist a nilpotent element $x\in(0,1)$ such that $S^{(k)}(x,...,x)=1$, for some $k\in\mathbb{N}$. This obviously contradicts the non-saturation property, since $x<1$.

($\Rightarrow$) Conversely, let us say that $S(x, y) = 1$ and we need to prove that $x = 1$ or $y = 1$. Suppose, by contradiction, that $x < 1$ and $y < 1$; then we may find $\varepsilon > 0$ such that $y < y + \varepsilon < 1$. Now, being S a t-conorm, in particular it is monotone, then:

$$1 = S(x, y) \leq S(x, y + \varepsilon).$$

On the other hand, since $x < 1$ and $y < y + \varepsilon$, then by using the strict monotonicity one could conclude that $S(x, y) < S(x, y + \varepsilon)$, which is a contradiction.

□

Submodular T-conorms and Co-copulae

Studying the *submodularity* (SUBM) of bivariate functions naturally leads us to the concept of *co-copula*. Such functions, which originally came out in *Probability*, also will allow us to exhibit further relations that insist among the "desirable" properties of *Influence Aggregation* of the section 4.2.1. In particular, we will concentrate on the connection between (*continuous* and *Archimedean*) *t-conorms* and *co-copulae* and, eventually, on how this affects their (additive and multiplicative) *generators*.

It is worth noticing that the focus here is on the bivariate co-copulae (i.e. in two dimensions) and this is due to the fact that it is not obvious how to generalize this definition to n-ary functions, called *n-co-copulae*. The standard extension of an associative 2-co-copula is not necessarily a n-co-copula and this has to do with the concept of n-decreasingness deeply. In our case of bivariate functions, the concepts of *sub/supermodularity, infra/ultramodularity* and 2-*de/increasingness* all collapse in the same definition and their connections with *concavity/convexity* are also simpler. For further references about this, see [Manzi, 2009], [Marinacci, Montrucchio, 2003] and [Marinacci, Montrucchio, 2008].

Definition 4.53. *A (2-dimensional)* ***co-copula*** *is a function $C : [0, 1]^2 \to [0, 1]$ that verifies:*

(ANN) *1 is the annihilator: $C(x, 1) = C(1, x) = 1$, for all $x \in [0, 1]$;*

(NEUT) *0 is the neutral element: $C(x, 0) = C(0, x) = x$, for all $x \in [0, 1]$;*

(SUBM) *C is submodular:*

$$C(\mathbf{x} \vee \mathbf{y}) + C(\mathbf{x} \wedge \mathbf{y}) \leq C(\mathbf{x}) + C(\mathbf{y}),$$

for all $\mathbf{x}, \mathbf{y} \in [0, 1]^2$.

Remark 4.54. The properties (SubM) and (Mon) are logically independent, since the function[10] $T_{min}(x, y) = \min\{x, y\}$ is monotone but not submodular (although it also verifies (Sym), (Assoc) and (Cont)); conversely, the function $F(x, y) = 1-(1-2x)(1-2y)$ is submodular but not monotone.

Remark 4.55 (**Duality copulae/co-copulae**)**.**
We have heavily used the duality between t-norms and t-conorms (see Remark 4.23), for adapting the results that are present in the literature (almost always for t-norms) to our case, which involves t-conorms. Something similar will happen with the co-copulae: in the literature is more common to talk about **copulae**, which are functions $C^* : [0, 1]^2 \to [0, 1]$ *dual* to the *co-copulae*, by means of the same duality relation of t-norms/conorms, i.e.

$$C^*(x, y) = 1 - C(1 - x, 1 - y), \qquad \forall\, (x, y) \in [0, 1]^2.$$

It is worth noticing that in the literature, given a *copula* H, the associated **co-copula** defined by duality is $H : [0, 1]^2 \to [0, 1]$, with

$$H^*(x, y) := 1 - H(1 - x, 1 - y), \qquad \forall\, (x, y) \in [0, 1].$$

However (see, for example, the Definition 9.16 of [Klement *et al.*, 2000]), by convention, the **dual copula** of H is the function $\widetilde{H} : [0, 1]^2 \to [0, 1]$, defined by

$$\widetilde{H}(x, y) := x + y - H(x, y), \qquad \forall\, (x, y) \in [0, 1].$$

Here, we will never use the concept of dual copula. Moreover, since we will deal with some cases when a function F is at the same time a t-conorm and a co-copula, by the (same) duality relation, its dual $F^*(\mathbf{x}) = 1 - F(\mathbf{1} - \mathbf{x})$ will be a t-norm and a copula.
Lastly, the lemma below explains how the *submodularity* of a co-copula (or t-conorm) shifts to a *dual* property of its associated copula (or t-norm), called *supermodularity*.

Proposition 4.56. *Given a function $F : [0, 1]^n \to [0, 1]$, let $F^* : [0, 1]^n \to [0, 1]$ be defined by duality as*

$$F^*(\mathbf{x}) = 1 - F(\mathbf{1} - \mathbf{x}) \qquad , \forall\, \mathbf{x} \in [0, 1]^n,$$

where $\mathbf{1} = (1, ..., 1) \in \mathbb{R}^n$. Then F is submodular if and only if F^* is ***supermodular****, i.e. for all $\mathbf{x}, \mathbf{y} \in [0, 1]^n$*

$$F^*(\mathbf{x} \vee \mathbf{y}) + F^*(\mathbf{x} \wedge \mathbf{y}) \geq F^*(\mathbf{x}) + F^*(\mathbf{y}).$$

[10] T_{min} is a **t-norm** which is the *dual t-norm* of S_{max}.

Proof. The proof is based on the fact that the maximum $\vee$ and the minimum $\wedge$ are taken by components and, so, given $\mathbf{x}, \mathbf{y} \in [0,1]^n$, then:

$$\begin{aligned}\mathbf{1} - \mathbf{x} \wedge \mathbf{y} &= (\mathbf{1} - \mathbf{x}) \vee (\mathbf{1} - \mathbf{y}),\\ \mathbf{1} - \mathbf{x} \vee \mathbf{y} &= (\mathbf{1} - \mathbf{x}) \wedge (\mathbf{1} - \mathbf{y}).\end{aligned}$$

And this implies that:

$$\begin{aligned}F^*(\mathbf{x} \wedge \mathbf{y}) &= 1 - F(\mathbf{1} - \mathbf{x} \wedge \mathbf{y}) = 1 - F((\mathbf{1} - \mathbf{x}) \vee (\mathbf{1} - \mathbf{y})),\\ F^*(\mathbf{x} \vee \mathbf{y}) &= 1 - F(\mathbf{1} - \mathbf{x} \vee \mathbf{y}) = 1 - F((\mathbf{1} - \mathbf{x}) \wedge (\mathbf{1} - \mathbf{y})).\end{aligned}$$

Now, assuming that F is submodular, then

$$\begin{aligned}&F^*(\mathbf{x} \wedge \mathbf{y}) + F^*(\mathbf{x} \vee \mathbf{y}) =\\ &\quad = 2 - [F((\mathbf{1} - \mathbf{x}) \vee (\mathbf{1} - \mathbf{y})) + F((\mathbf{1} - \mathbf{x}) \wedge (\mathbf{1} - \mathbf{y}))] \geq\\ &\qquad \geq 2 - [F(\mathbf{1} - \mathbf{x}) + F(\mathbf{1} - \mathbf{y})] = F^*(\mathbf{x}) + F^*(\mathbf{y}).\end{aligned}$$

The converse implication follows from the fact that $(F^*)^* = F$. □

Adapting Lemma 1.4.2 of [Alsina *et al.*, 2006] and Proposition 9.8 from [Klement *et al.*, 2000] for co-copulae gives us:

Lemma 4.57. *If C is a co-copula, then it is 1-Lipschitz, i.e.*

$$|C(x_1, y_1) - C(x_2, y_2)| \leq |x_1 - x_2| + |y_1 - y_2|,$$

for any $x_1, x_2, y_1, y_2 \in [0,1]$.
In particular, every co-copula C:

- *satisfies* (Mon)*;*
- *is continuous in each variable;*
- *is jointly continuous* (Cont)*.*

Remark 4.58. The classes of t-conorms and co-copulae intersect but do not coincide. Indeed, we have already checked that S_{max}, S_{prob} and S_{Luk} verify all the above properties and, so, they are (associative and commutative) co-copulae. Conversely, the continuous co-copula $C : [0,1]^2 \to [0,1]$ defined by

$$C(x, y) = 1 - (1 - x)(1 - y) + xy(1 - x)^2(1 - y)$$

is not commutative and, so, cannot be a t-conorm.

Proposition 4.59. *There exist (even strict) t-conorms that are not submodular. In particular, the submodularity* (SubM) *is independent of all the others "desirable" properties listed in the section 4.2.1 (even when they are considered jointly).*

Proof. First of all, of course we already know that S_{prob} verifies all those properties including the submodularity. On the other hand, we want to find a function which verifies all of them but (SUBM): in particular, for this purpose, we will define a *strict t-conorm* $S : [0,1]^2 \to [0,1]$ by using its *additive generator* $\psi : [0,1] \to [0,+\infty]$. Let us consider

$$\psi(x) = \begin{cases} \dfrac{\sqrt{2}}{2}\sqrt{x}, & \text{for } x \in [0,1/2], \\ \dfrac{1}{2-2x} - \dfrac{1}{2}, & \text{for } x \in [1/2,1]. \end{cases}$$

Now, it is clear that ψ is continuous, strictly increasing, such that $\psi(0) = 0$ and, in addition, $\psi(1^-) = \lim_{x\to 1^-} \psi(x) = +\infty$. Then, from Lemma 4.50, it follows that by defining

$$S(x,y) = \psi^{-1}\big(\psi(x) + \psi(y)\big), \qquad \forall\, (x,y) \in [0,1]^2,$$

we obtain a strict t-conorm. In particular, this means that S necessarily verifies all the "desirable" properties (including (NSAT)), with the only possible exception of (SUBM). Surely, we want to prove that S does not satisfy (SUBM). In order to do that, we need to find the inverse ψ^{-1} and, as we will see later, it will be enough to consider the restriction $\psi\big|_{[0,\frac{1}{2}]}$, whose inverse is

$$\left(\psi\big|_{[0,\frac{1}{2}]}\right)^{-1}(z) = 2\,z^2, \qquad \forall\, z \in [0,1/2].$$

Now, if we take $x, y \in]0,1/2]$ such that they are close enough to 0, for example $0 < x = y = 1/100$, then:

$$S(x,y) = \psi^{-1}\big(\psi(x) + \psi(y)\big) = \psi^{-1}\left(\frac{\sqrt{2}}{2}\sqrt{x} + \frac{\sqrt{2}}{2}\sqrt{y}\right) =$$

$$= 2\left[\frac{\sqrt{2}}{2}(\sqrt{x} + \sqrt{y})\right]^2 = \big(\sqrt{x} + \sqrt{y}\big)^2 = x + y + \sqrt{xy}.$$

And, now, it is easy to verify that S is not submodular, because if we choose $\mathbf{x} = (x,0)$ and $\mathbf{y} = (0,y)$ with the same $x, y \in]0,1/2]$ close to 0 as before, then:

$$S(\mathbf{x} \wedge \mathbf{y}) + S(\mathbf{x} \vee \mathbf{y}) = S(x,y) + S(0,0) = x + y + \sqrt{xy} >$$

$$> x + y = S(x,0) + S(0,y) = S(\mathbf{x}) + S(\mathbf{y}),$$

which contradicts the submodularity. □

From the above remark, it follows that, so far, *continuous t-conorm* and (*continuous*) *co-copulae* do not coincide. However, the situation changes drastically in presence of the associativity (ASSOC) as proved by the following lemma. Notice that the opposite implication cannot be true, as already checked in a previous remark with T_{min}.

Lemma 4.60. *Each associative co-copula is a continuous t-conorm.*

The next proposition finally allows us to conclude the characterization of all functions that satisfy all the "desirable" properties in terms of (additive) generators. But first we need a lemma:

Lemma 4.61. *A function $u : \mathbb{R}^n \to \mathbb{R}$ is convex (respectively concave) if and only if the function $u^* : \mathbb{R}^n \to \mathbb{R}$, defined by $u^*(\mathbf{x}) = u(\mathbf{1}-\mathbf{x})$, is convex (respect. concave), where $\mathbf{1} = (1, ..., 1) \in \mathbb{R}^n$.*
Consequently, a t-conorm S is concave if and only if its dual t-norm $T(\mathbf{x}) = 1 - S(\mathbf{1} - \mathbf{x})$ is convex.

Proof. For all $\mathbf{x}, \mathbf{y}, \lambda \in \mathbb{R}$, we may write

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y} = 1 - [\lambda(\mathbf{1}-\mathbf{x}) + (1-\lambda)(\mathbf{1}-\mathbf{y})].$$

In particular, if we assume that u is convex, then

$$\begin{aligned} u^*\big(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\big) &= u^*\Big(1 - \big[\lambda(\mathbf{1}-\mathbf{x}) + (1-\lambda)(\mathbf{1}-\mathbf{y})\big]\Big) = \\ &= u\big(\lambda(\mathbf{1}-\mathbf{x}) + (1-\lambda)(\mathbf{1}-\mathbf{y})\big) \leq \lambda u(\mathbf{1}-\mathbf{x}) + (1-\lambda)u(\mathbf{1}-\mathbf{y}) = \\ &= \lambda u^*(\mathbf{x}) + (1-\lambda)u^*(\mathbf{y}), \end{aligned}$$

so, u^* is convex as well. The opposite direction of the proof follows straightforward, because $(u^*)^* = u$. Finally, the last assertion comes from the fact that a function u is concave (resp. convex) if and only if its opposite $-u$ is convex (resp. concave). □

Proposition 4.62. *A continuous and Archimedean t-conorm S is submodular, i.e. an Archimedean co-copula, if and only if it is additively generated by a convex function.*

Proof. The proof follows from the duality with t-norms and Theorem 2.2.9 of [Alsina *et al.*, 2006]. In particular, S is submodular if and only if $T = S^*$ is supermodular; in addition to this, the additive generator $s(x)$ of S is convex if and only if $t(x) = s(1-x)$ is convex. □

Finally, by putting together all what we have seen so far, we may write the definition of what we call **Influence Aggregation Function**:

Corollary 4.63. *Given $n \in \mathbb{N}$, an **influence aggregation function** is a family of functions $\{F^{(2)}, ..., F^{(n)}\}$ defined by associativity, starting from a strict and submodular t-conorm $F^{(2)} = F : [0,1]^2 \to [0,1]$, as follows:*

$$F^{(k)}(x_1, ..., x_k) = F^{(2)}\left(F^{(k-1)}(x_1, ..., x_{k-1}), x_k\right), \quad \forall\, x_1, ..., x_k \in [0,1].$$

It satisfies all the "desirable" properties listed in 4.2.1. In addition, there exists an additive generator $f : [0,1] \to [0,\infty]$*, which is continuous, strictly increasing, convex, such that* $f(0) = 0$ *and* $f(1) = +\infty$*, and*

$$F^{(k)}(x_1, ..., x_k) = f^{-1}(f(x_1) + ... + f(x_k)),$$

for all $x_1, ..., x_k \in [0,1]$ *and* $k = 2, ..., n$*.*

Although generators provide a lot of useful information about a t-conorm, nevertheless, given a t-conorm, it is not easy, in general, to find or reconstruct its generator. Some techniques are described, for example, in [Navara, Petrık, 2008]. However, somehow avoiding to use generators, our concern, in the last part of this section, is how to easily understand whether a t-conorm S is submodular.

Proposition 4.64. *Let* $S : [0,1]^2 \to [0,1]$ *be a t-conorm and let us define the subset* $A := \{(x,y) \in [0,1]^2 : \ S(x,y) < 1\}$*. Then:* S *is submodular on the subset* A *if and only if* S *is submodular on the whole domain* $[0,1]^2$*.*

Proof. Obviously, if S is submodular on $[0,1]^2$, then, as special case, one gets the submodularity of S when restricted to the vectors of A. Let us now prove the converse implication. Take $\mathbf{u}, \mathbf{v} \in [0,1]^2$. There are three possible cases:

- if both $\mathbf{u}, \mathbf{v} \in A$, then, by construction, the inequality that defines the submodularity holds:

$$S(\mathbf{u} \wedge \mathbf{v}) + S(\mathbf{u} \vee \mathbf{v}) \leq S(\mathbf{u}) + S(\mathbf{v});$$

- if both $\mathbf{u}, \mathbf{v} \notin A$, then the right side of the inequality is equal to $1+1 = 2$, while the left side may only be less or equal, for the monotonicity of S;

- if $\mathbf{u} \in A$ and $\mathbf{v} \notin A$ (the symmetric case is analogous), then using the monotonicity of S gives:

$$\begin{cases} S(\mathbf{u} \wedge \mathbf{v}) \leq S(\mathbf{u}), \\ 1 \geq S(\mathbf{u} \vee \mathbf{v}) \geq S(\mathbf{v}) = 1 \ \Rightarrow \ S(\mathbf{u} \vee \mathbf{v}) = S(\mathbf{v}) = 1, \end{cases}$$

  and both together imply the thesis $S(\mathbf{u} \wedge \mathbf{v}) + S(\mathbf{u} \vee \mathbf{v}) \leq S(\mathbf{u}) + S(\mathbf{v})$.

□

In case S is a continuous t-conorm, then the previous result may be slightly strengthened.

Lemma 4.65. *Let* S *be a continuous t-conorm and* $A := \{(x,y) \in [0,1]^2 : S(x,y) < 1\}$*, as above. Then:* S *is submodular on* $[0,1]^2$ *if and only if it is submodular on the open subset* $\mathring{A} = \{(x,y) \in (0,1)^2 : \ S(x,y) < 1\}$*.*

Proof. If S is submodular, as a special case we get that it is submodular on the subset $\mathring{A}$. Let us now assume that S is submodular only on the subset $\mathring{A}$ and then, from the previous lemma, it will suffice to prove that S is submodular on the entire A, in order to obtain the thesis.

First of all, notice that, from the properties of continuous t-conorms, A may be partitioned in $A = \mathring{A} \sqcup [0,1) \times \{0\} \sqcup \{0\} \times [0,1)$, where the symbol $\sqcup$ denotes the union of disjoint sets. Now, let us take $\mathbf{u}, \mathbf{v} \in A$ arbitrarily and prove that the inequality

$$S(\mathbf{u} \wedge \mathbf{v}) + S(\mathbf{u} \vee \mathbf{v}) \leq S(\mathbf{u}) + S(\mathbf{v})$$

holds. We have a few cases:

- If both $\mathbf{u}, \mathbf{v} \in \mathring{A}$, then the thesis follows from the hypothesis.
- If both $\mathbf{u}, \mathbf{v} \in [0,1) \times \{0\} \sqcup \{0\} \times [0,1)$, then, from the monotonicity of S, we could trivially obtain the thesis in case both vectors were in the same axes, because in that case we would have either $\mathbf{u} \leq \mathbf{v}$ or $\mathbf{v} \leq \mathbf{u}$. So, let us assume that neither $\mathbf{u} \not\leq \mathbf{v}$ nor $\mathbf{v} \not\leq \mathbf{u}$, which means that:

  $$\mathbf{u} = (x, 0) \in (0,1) \times \{0\}, \qquad \mathbf{v} = (0, y) \in \{0\} \times (0,1).$$

  Since $x, y \in (0,1)$, then there exists a number $k \in \mathbb{N}$ such that $0 < 1/k < \min\{x, y\}$ and we may define the sequences

  $$\mathbf{u}_n = (x, 1/n), \qquad \mathbf{v}_n = (1/n, y), \quad \forall n \geq k.$$

  Studying the convergence[11] of these vectors, we have that:

  $$\mathbf{u}_n \xrightarrow[n\to\infty]{} \mathbf{u}, \qquad \mathbf{v}_n \xrightarrow[n\to\infty]{} \mathbf{v};$$

  in particular

  $$\mathbf{u}_n \wedge \mathbf{v}_n = (1/n, 1/n) \xrightarrow[n\to\infty]{} \mathbf{u} \wedge \mathbf{v} = (0,0),$$

  and obviously $\mathbf{u}_n \vee \mathbf{v}_n = (x, y) = \mathbf{u} \vee \mathbf{v}$. Now, using that $\mathbf{u}_n, \mathbf{v}_n \in \mathring{A}$ and S is submodular on $\mathring{A}$, we may write the inequality

  $$S(\mathbf{u}_n \wedge \mathbf{v}_n) + S(\mathbf{u}_n \vee \mathbf{v}_n) \leq S(\mathbf{u}_n) + S(\mathbf{v}_n).$$

  Passing to the limit and using the continuity of S on $[0,1]^2$, we get the thesis:

  $$S(\mathbf{u} \wedge \mathbf{v}) + S(\mathbf{u} \vee \mathbf{v}) \leq S(\mathbf{u}) + S(\mathbf{v}).$$

[11] The convergence may be taken, for example, in the **Euclidean norm** $\|\mathbf{x}\|_2 := \sqrt{x_1^2 + ... + x_n^2}$. It is not important, since in $[0,1]^n \subset \mathbb{R}^n$ all norm are equivalent and generate the same topology.

- The only case left is when $\mathbf{u} = (u_1, u_2) \in \mathring{A}$ and $\mathbf{v} \in A \setminus \mathring{A}$, which comprises two sub-cases: $\mathbf{v} = (0, v)$ or $\mathbf{v} = (v, 0)$, with $v \in (0, 1)$. Let us consider, for example, the first case in which $\mathbf{v} = (0, v)$ and then, in order to avoid the trivial case in which $\mathbf{u} \leq \mathbf{v}$ or $\mathbf{v} \leq \mathbf{u}$, let us assume that $v > u_2$. Using the same ideas of the previous step, we may define the vectors $\mathbf{v}_n = (1/n, v)$, with $n \geq k$ and $k \in \mathbb{N}$ such that $1/k < x_1$. Now, let us consider

$$\mathbf{u} \wedge \mathbf{v}_n = (1/n, x_2) \quad , \quad \mathbf{u} \vee \mathbf{v}_n = (x_1, v).$$

  First of all, notice that both $\mathbf{u} \wedge \mathbf{v}_n, \mathbf{u} \vee \mathbf{v}_n \in \mathring{A}$, and then the inequality defining the submodularity holds. Then, it suffices to pass to the limit in order to obtain the thesis.

□

Remark 4.66. In the previous proposition, continuity is a necessary condition. To prove this, let us consider the **drastic t-conorm**, which we have already defined previously and can be re-written as

$$S_{dr}(x, y) = \begin{cases} x, & \text{if } y = 0, \\ y, & \text{if } x = 0, \\ 1, & \text{otherwise.} \end{cases}$$

Notice that S_{dr} is not continuous and that $A = [0, 1) \times \{0\} \cup \{0\} \times [0, 1)$. Trivially, S_{dr} is submodular on $\mathring{A}$, since $\mathring{A} = \emptyset$. However, it is not submodular on $[0, 1]^2$, as one can show simply by taking $\mathbf{u} = (1/3, 0)$ and $\mathbf{v} = (0, 1/3)$ and calculating

$$1 = S_{dr}(0, 0) + S_{dr}(1/3, 1/3) = S_{dr}(\mathbf{u} \wedge \mathbf{v}) + S_{dr}(\mathbf{u} \vee \mathbf{v}) > S_{dr}(\mathbf{u}) + S_{dr}(\mathbf{v}) = 0.$$

4.2.4 Comparisons and Orderings

Pointwise Order

Definition 4.67. *We say that the t-conorm S_1 is **weaker** than the t-conorm S_2, and we write $S_1 \leq S_2$, if $S_1(x, y) \leq S_2(x, y)$ for all $(x, y) \in [0, 1]^2$.*

Lemma 4.68. *$S_1 \geq S_2$ if and only if $T_1 \leq T_2$, where $T_i = S_i^*$ is the dual t-norm.*

Remark 4.69. The four basic t-conorms of the section 4.1.3 are ordered as follows:

$$S_{max} < S_{prob} < S_{Luk} < S_{dr}.$$

However, the pointwise order $\leq$ is not a total order[12], meaning that there exist incomparable t-conorms. More precisely, there exist pairs of continuous and Archimedean incomparable t-conorms and even incomparable pairs in the class of nilpotent or strict t-conorms.

[12] Also called linear order in literature.

Proposition 4.70. *If S is a submodular t-conorm (as well as $S = C$ is a co-copula), then*

$$S_{max} \leq S \leq S_{Luk}.$$

Proof. The inequality on the left side always holds. For the right side, starting from the definition of submodularity

$$S(\mathbf{x} \vee \mathbf{y}) + S(\mathbf{x} \wedge \mathbf{y}) \leq S(\mathbf{x}) + S(\mathbf{y}),$$

and then taking $\mathbf{x} = (x, 0)$ and $\mathbf{y} = (0, y)$ gives

$$S(x, y) \leq x + y,$$

since 0 is the neutral element of S. Finally, $S(x, y) \leq \min\{x + y, 1\} = S_{Luk}(x, y)$. □

Proposition 4.71. *Let S_1, S_2 be continuous and Archimedean t-conorms with additive generators $s_1, s_2 : [0, 1] \to [0, +\infty]$ respectively. Then:*

- *$S_1 \geq S_2$ if and only if $s_1 \circ s_2^{-1}$ is subadditive;*
- *if $s_1 \circ s_2^{-1}$ is concave, then $S_1 \geq S_2$;*
- *if s_1/s_2 is non-increasing on $(0, 1)$, then $S_1 \geq S_2$;*
- *if $u \mapsto \frac{s_1 \circ s_2^{-1}(u)}{u}$ is non-increasing on $]0, s_2(1)]$, then $S_2 \geq S_1$;*
- *if s_1 and s_2 are continuously differentiable on $(0, 1)$ and s_1'/s_2' is non-decreasing on $(0, 1)$, then $S_1 \geq S_2$.*

Proof. The proof can be easily obtained using the correspondent and dual results for t-norms of the section 2.2 of [Alsina *et al.*, 2006], and modifying the relations involving the additive generators, considering that: given a t-conorm S with additive generator s, its dual t-norm $T = S^*$ has additive generator t such that

$$t(x) = s(1 - x).$$

Hence, in general, it holds that:

$$s(x) = t(1 - x), \quad t^{(-1)}(u) = 1 - s^{(-1)}(u),$$

and, in particular: $t'(x) = -s'(1 - x)$. Applying these formulae, one gets:

- $s_1 \circ s_2^{-1}(u) = t_1 \circ t_2^{-1}(u)$;
- $\frac{s_1}{s_2}(u) = \frac{t_1}{t_2}(1 - u)$;
- $\frac{s_1'}{s_2'}(u) = \frac{t_1'}{t_2'}(1 - u)$;

and the proof is concluded accordingly. □

Proposition 4.72.

- *Given two continuous and Archimedean t-conorms S_1, S_2, then*

$$S_{max} \leq \min\{S_1, S_2\}.$$

- *Given two strict (respectively nilpotent) t-conorms S_1, S_2, then there exists a strict (resp. nilpotent) t-conorm S such that*

$$S \leq \min\{S_1, S_2\}.$$

- *Consequently, given two continuous and Archimedean t-conorms S_1, S_2, then there exists a continuous and Archimedean t-conorm S such that $S \leq \min\{S_1, S_2\}$.*

Dominance (or Domination)

Another concept that emerges when considering comparisons among t-conorms is the *dominance* (or *domination*), which turns out to be stronger than the *pointwise order* seen before, but still a non-linear order. For references, check the sections 4.2 of [Alsina *et al.*, 2006] and 6.3 of [Klement *et al.*, 2000].

Definition 4.73. *The t-conorm S_1* ***dominates*** *the t-conorm S_2, in symbols $S_2 \ll S_1$ if*

$$S_1\big(S_2(a,b), S_2(c,d)\big) \geq S_2\big(S_1(a,b), S_1(c,d)\big), \qquad \forall\, a,b,c,d \in [0,1].$$

Lemma 4.74.

- *If $S_2 \ll S_1$ then $S_2 \leq S_1$; however, the converse implication is not true, even when restricted to the class of strict t-conorms, i.e.*

$$S_2 \leq S_1 \quad \not\Rightarrow \quad S_2 \ll S_1.$$

- *Each t-conorm S is $\ll$-bounded between:*

$$S_{max} \ll S \ll S_{dr}.$$

- *$\ll$ is an order among the t-conorms, but it is not a total order, meaning that there exist two t-conorms that are incomparable with respect to $\ll$.*[13]

[13] As far as we know, up to now it is not still totally clear what are the set of t-conorms where $\ll$ is a linear order, i.e. where also the transitivity holds.

Lemma 4.75. *Given a strict t-conorm S, there always exists another strict t-conorm $S^g \gg S$ which dominates S and, in addition, the additive generator ψ^g of S^g originates from ψ, which is the one of S, according to the formula:*

$$\psi^g(x) = \int_0^{\psi(x)} \exp\left(g(\log u)\right)\, \mathrm{d}u,$$

being $g : \mathbb{R} \to \mathbb{R}$ a non-decreasing, continuous and convex function such that $\int_0^\infty \exp\left(g(\log u)\right)\, \mathrm{d}u = \infty$.

4.2.5 Concavity, Submodularity and Inframodularity

We have already seen that the submodularity of a continuous and Archimedean t-conorm is equivalent to the convexity of its additive generators. In addition to this, the relations among the concavity/convexity and the submodularity do not end here. On the one hand, [Alsina, Tomas, 1988] stated that *smooth concave t-conorms* (i.e. $\mathcal{C}^2$) cannot exist; on the other hand, the *Lukasiewicz t-conorm* S_{Luk} is indeed concave, though not $\mathcal{C}^2$, of course. The properties *submodularity* and *concavity/convexity* are, in general independent; besides, as stated already in [Lovász, 1983], some times the *submodularity* behaves in a way similar to the concavity, while other times similar to the convexity.

In the end of this section, we will introduce another concept, the *inframodularity*, which generalizes both concavity and submodularity at once. Nevertheless, all these notions collapse in the same when reduced to one dimension. For references, check [Marinacci, Montrucchio, 2003], [Marinacci, Montrucchio, 2008], [Manzi, 2009], [Manzi, 2010], [Cardin, Manzi, 2009] and [Klement *et al.*, 2011].

Proposition 4.76.

- *If S is a concave*[14] *t-conorm, then $S_{Luk} \leq S$.*
- *In particular, if S is continuous then it is necessarily nilpotent. Equivalently, there are not concave*[15] *strict t-conorms.*
- *S_{Luk} is the only concave*[16] *t-conorm.*
- *If S is a nilpotent t-conorm and has a concave additive generator, then (S is Schur-concave) and*

$$S_{Luk} \leq S.$$

[14] This hypothesis can be weakened considering S quasi-concave or even Schur-concave.

[15] There are no quasi-concave or Schur-concave strict t-conorm as well.

[16] This result can be strengthened: S_{Luk} is the only Schur-concave or quasi-concave co-copula.

Proof. Use Lemma 4.1.12, Theorems 4.1.13 and Theorem 4.1.15 of [Alsina *et al.*, 2006] and that S (Schur-)concave if and only if $T = S^*$ (Schur-)convex, whereas, for the additive generators, s is concave if and only if the same holds for $t = s^*$. □

Inframodularity

The *submodularity*, as property, originated naturally and intuitively in the section 4.2.1 when we tried to take into account the property of *diminishing returns* (see the following remark). As we will see in the next section 4.3 and in the next chapter, it will also play a fundamental role to guarantee the *submodularity* of the *diffusion function* σ in a network. Here we study the *inframodularity* which is the same for set-functions but is a stronger property for vector-functions (like the aggregation functions and t-conorms). It also has deep relations with the concavity/convexity.

Remark 4.77. Recall that, if $f : 2^\Omega \to \mathbb{R}$ is a set-function, then the following properties defining the *submodularity* of f are equivalent:

- $f(A \cap B) + f(A \cup B) \leq f(A) + f(B)$, for all $A, B \subseteq \Omega$;
- $f(A \cup \{\omega\}) - f(A) \geq f(B \cup \{\omega\}) - f(B)$, for all $A \subseteq B \subseteq \Omega$ and $\omega \in \Omega \setminus B$.

Although those properties are equivalent for set-functions, we will see here that when we deal with vector-functions they are not, and correspond to the *submodularity* and *inframodularity*, respectively.

Motivated by the previous result, we have generalized the concept of submodularity for vector-functions[17] $F : [0,1]^n \to [0,1]$ as follows:

$$\text{(SubM)} \qquad F(\mathbf{x} \vee \mathbf{y}) + F(\mathbf{x} \wedge \mathbf{y}) \leq F(\mathbf{x}) + F(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in [0,1]^n$, stating that this is equivalent to asking a property of *diminishing returns*.

Definition 4.78. *A function $F : [0,1]^n \to [0,1]$ is called* ***inframodular*** *if it has* decreasing increments*, i.e.*

$$\text{(InfM)} \qquad F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) \geq F(\mathbf{y} + \mathbf{h}) - F(\mathbf{y}),$$

for all $\mathbf{x}, \mathbf{y}, \mathbf{h} \in [0,1]^n$ such that $\mathbf{x} \leq \mathbf{y}$, $\mathbf{h} > 0$ and $\mathbf{x} + \mathbf{h}, \mathbf{y} + \mathbf{h} \in [0,1]^n$.

The following proposition states that the inframodularity for t-conorms (as well as co-copulae) is equivalent to the ultramodularity of their dual t-norm (copula).

[17] In our context, we limit ourselves to the case of aggregation functions, in $[0,1]^n$.

Proposition 4.79. *A function $F : [0,1]^n \to [0,1]$ is inframodular if and only if the (dual) function $G(\mathbf{x}) := 1 - F(\mathbf{1} - \mathbf{x})$ is* ultramodular*, i.e. it verifies the opposite inequality in* (InfM).

Proof. First of all, notice that

$$G(\mathbf{x}+\mathbf{h}) - G(\mathbf{x}) = 1 - F(\underbrace{\mathbf{1}-\mathbf{x}-\mathbf{h}}_{=:\mathbf{z_x}}) - 1 + F(\underbrace{\mathbf{1}-\mathbf{x}}_{\mathbf{z_x}+\mathbf{h}}) = F(\mathbf{z_x}+\mathbf{h}) - F(\mathbf{z_x}).$$

Now, if $\mathbf{x} \leq \mathbf{y}$ then $\mathbf{z_x} = \mathbf{1}-\mathbf{x}-\mathbf{h} \geq \mathbf{1}-\mathbf{y}-\mathbf{h} = \mathbf{z_y}$ and, so, if we assume that F is inframodular, then

$$\begin{aligned} G(\mathbf{x}+\mathbf{h}) - G(\mathbf{x}) = F(\mathbf{z_x}+\mathbf{h}) - F(\mathbf{z_x}) \geq \\ \geq F(\mathbf{z_y}+\mathbf{h}) - F(\mathbf{z_y}) = G(\mathbf{y}+\mathbf{h}) - G(\mathbf{y}). \end{aligned}$$

The opposite direction is analogous. □

The *inframodularity* has deep relations both with the submodularity and with the (scalar) concavity.

Lemma 4.80. *The inframodularity is stronger than the submodularity*[18]*, i.e.*

$$(\text{InfM}) \quad \Longrightarrow \quad (\text{SubM}).$$

Lemma 4.81. *In the scalar case, with $n = 1$, a concave function $F : \mathbb{R} \to \mathbb{R}$ necessarily verifies* (InfM)*. The converse implication is true provided F is continuous.*

Lemma 4.82. *For an aggregation function $F : [0,1]^n \to [0,1]$, the following are equivalent:*

1. *F is inframodular;*
2. *every 2-dimensional section of F is inframodular;*
3. *every 2-dimensional section of F is submodular and every 1-dimensional section is concave.*

Lemma 4.83. *If $F : [0,1]^n \to [0,1]$ is a function twice-derivable, then:*

- *F is submodular if and only if $\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0$, for all $i \neq j$;*
- *F is inframodular if and only if $\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0$, for all i, j.*

Remark 4.84. Intuitively speaking, in the context of *influence aggregation*, the *inframodularity* is a property that says, in addition to the *submodularity*, that also in the same component the *marginal gain* has to be decreasing, which means that if the weight $w_i \in [0,1]$ of the agent i increases, then its *marginal contribution* to the total influence is concave.

[18] Whenever the domain is a lattice, as in the case of $[0,1]^n$.

The next result recalls those seen in the section 4.2.2: the *inframodularity* passes from a binary function to all its k-ary iterations. However, in this proposition the *associativity* will play a fundamental role.

Proposition 4.85. *Let $F : [0,1]^2 \to [0,1]$ be a binary associative, monotone and symmetric function. If F is inframodular, then each k-ary iteration $F^{(k)}$ defined by associativity is also inframodular.*

Proof. This proposition can be deduced by Corollary 3.3 in [Klement *et al.*, 2011], but here we give an alternative (and slightly weaker) proof. From the previous lemma, in order to prove that $F^{(k)}$ is inframodular, it suffices to prove that every 2-section is submodular and every 1-section is concave.

First of all, since F is inframodular, it is also submodular and we already proved in section 4.2.2 that every $F^{(k)}$ has to be submodular as well. From that section, we also know that $F^{(k)}$ is submodular if and only if every 2-section of $F^{(k)}$ is submodular. Hence, it remains to prove that every 1-section of $F^{(k)}$ is concave. By applying the previous lemma to the case of $k = 2$, i.e. to F, we get that the 1-sections of F are concave. Now, consider a 1-section of $F^{(k)}$: notice that by symmetry, we can only consider the last component of $F^{(k)}$ as variable, i.e. fix $\boldsymbol{\xi} \in [0,1]^{k-1}$ and take

$$z \longmapsto F^{(k)}(\boldsymbol{\xi}, z) = F\left(F^{(k-1)}(\boldsymbol{\xi}), z\right), \qquad \forall z \in [0,1]$$

and, since this univariate function coincides with a 1-section of F, it is concave. This proves that every 1-section of $F^{(k)}$ is concave and concludes the proof. □

Remark 4.86. Although, as stated in a previous lemma, there is (almost) an equivalence between *inframodularity* and *concavity* in the scalar case, when $n > 1$ they are independent properties: there are concave functions that are not inframodular and, viceversa, inframodular functions that are not concave. For example, consider the functions[19]:

- $F(\mathbf{x}) = 1 - \|\mathbf{1} - \mathbf{x}\|_2 = 1 - \left(\sum_{i=1}^{n}(1-x_i)^2\right)^{1/2}, \quad \forall\, \mathbf{x} \in [0,1]^n,$
- $S_{prob}(\mathbf{x}) = 1 - \prod_{i=1}^{n}(1-x_i), \qquad \forall\, \mathbf{x} \in [0,1]^n.$

Then:

- F is concave, because the Euclidean norm $\|\cdot\|_2$ is convex in $[0,1]^n$; then $-\|\cdot\|_2$ turns out to be concave and so is F. However, F is not

[19] $\|\cdot\|_2$ indicates the standard Euclidean norm $\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$ in $\mathbb{R}^n$.

inframodular, since

$$\frac{\partial F}{\partial x_i}(\mathbf{x}) = (1-x_i)\left(\sum_{i=1}^{k}(1-x_i)^2\right)^{-1/2} = \frac{1-x_i}{\|\mathbf{1}-\mathbf{x}\|_2},$$

and, consequently, for $j \neq i$

$$\frac{\partial^2 F}{\partial x_i \partial x_j}(\mathbf{x}) = (1-x_i)(1-x_j)\left(\sum_{i=1}^{k}(1-x_i)^2\right)^{-3/2} \geq 0, \quad \forall \mathbf{x} \in [0,1]^n,$$

and, obviously, the inequality is strict for an adequate choice of $\mathbf{x} \in [0,1]$. Then F is not inframodular (and not even submodular.)

- On the other hand, S_{prob} is inframodular, because

$$\frac{\partial S_{prob}}{\partial x_i}(\mathbf{x}) = \prod_{j=1,\ldots,n:\, j\neq i}(1-x_j),$$

and so, for any $\mathbf{x} \in [0,1]^n$:

$$0 \geq \frac{\partial^2 S_{prob}}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} -\prod_{k=1,\ldots,n:\, k\neq i,j}(1-x_k), & \text{for } i \neq j, \\ 0, & \text{for } i = j. \end{cases}$$

However, S_{prob} is not concave, because no smooth $\mathcal{C}^2$ t-conorm can be concave (see [Alsina, Tomas, 1988]).

Here are some interesting results in our context: the first involves the pointwise order and the inframodularity, while the second one concerns the additive generator of an inframodular t-conorm.

Lemma 4.87. *If S is an inframodular co-copula or an inframodular t-conorm, then*

$$S_{prob} \leq S \leq S_{Luk}.$$

Proof. We already know that the inequality on the right side is always true, in the case S is a co-copula as well as S is an ultramodular (and, so submodular) t-conorm. The left side follows from applying the duality among t-norms and t-conorms to Remark 2.10 (iii) of [Klement *et al.*, 2011], using that S is an inframodular co-copula if and only if $T = S^*$ is an ultramodular copula. Then, $T \leq T_{prob}$ becomes $S_{prob} \leq S$. □

Proposition 4.88. *Let S be a continuous, Archimedean and submodular t-conorm* [20] *with a twice differentiable additive generator $s : [0,1] \to [0,\infty]$. Then S is inframodular if and only if $x \mapsto 1/s'(x)$ is a concave function.*

[20] Or equivalently, an Archimedean co-copula.

Proof. The proof is based using the duality between additive generators of t-norms and t-conorms applied to Theorem 4.3.1 of [Manzi, 2009]. In particular, the theorem says that the dual t-norm $T = S^*$ of S, defined by $T(\mathbf{x}) = 1 - S(\mathbf{1} - \mathbf{x})$, is ultramodular if and only if its additive generator t is such that $1/t'$ is convex.

From Proposition 4.79, it follows that S is inframodular if and only if T is ultramodular. On the other hand, considering their respective additive generators, it holds that $t(x) = s(1 - x)$ for all $x \in [0, 1]$ and, obviously, both are twice differentiable by hypothesis. In particular, $t'(x) = -s'(1-x)$ and, so

$$\frac{1}{t'(x)} = -\frac{1}{s'(1-x)}, \qquad \forall\, x \in [0, 1].$$

This means that the following are equivalent, the last passage following from Lemma 4.61:

- S is inframodular;
- $S^* = T$ is ultramodular;
- $1/t'(x) = -1/s'(1 - x)$ is convex;
- $1/s'(1 - x)$ is concave;
- $1/s'(x)$ is concave.

So, the proof is concluded. □

4.2.6 A Family of Influence Aggregation Functions

Let us consider **Frank's t-conorms** $S_\alpha : [0, 1]^2 \to [0, 1]$, defined for any parameter $\alpha \in \mathbb{R} \setminus \{0\}$ by:

$$S_\alpha(x, y) = 1 - \frac{1}{\alpha} \log \left[1 + \frac{(e^{\alpha(1-x)} - 1)(e^{\alpha(1-y)} - 1)}{e^\alpha - 1} \right],$$

for all $(x, y) \in [0, 1]^2$.

This family of t-conorms has very important properties and it was first introduced and studied by Maurice J. Frank already in the '70s, when trying to find a pair (T, S) constituted by a t-norm and a t-conorm respectively, that satisfies the functional equation (see sections 2.6 and 3.1 of [Alsina *et al.*, 2006] and appendix A.3 of [Klement *et al.*, 2000])

$$T(x, y) + S(x, y) = x + y, \qquad \forall\, (x, y) \in [0, 1]^2.$$

In fact, a pair (S^*_α, S_α) is a solution of this equation[21]. Although we are not interested in these equation here, the study of this family indeed leads to prove some interesting properties for our context, in particular:

[21] Where S^*_α is the dual t-norm of S_α.

- taking the limits for the parameter α and (possibly) applying the *de l'Hopital rule*, one gets that

$$S_{-\infty} = S_{max}, \qquad S_0 = S_{prob}, \qquad S_{+\infty} = S_{Luk};$$

- S_α is continuous and Archimedean if and only if $\alpha \in]-\infty, +\infty]$; more precisely, the S_α's are strict t-conorms when $\alpha \neq \pm\infty$ while $S_{+\infty} = S_{Luk}$ is the only nilpotent t-conorm of the family;

- the family $\{S_\alpha\}_{\alpha\in\overline{\mathbb{R}}}$ is ordered with respect to $\leq$:

$$\alpha_1 \leq \alpha_2 \quad \Longrightarrow \quad S_{\alpha_1} \leq S_{\alpha_2};$$

- for fixed $(x, y) \in [0, 1]^2$, the map $\alpha \mapsto S_\alpha(x, y)$ is continuous.

- the additive generators $s_\alpha : [0, 1] \to [0, 1]$ are, respectively, given by

$$s_\alpha(x) = \begin{cases} -\log(1-x), & \text{if } \alpha = 0, \\ -\log\left[\frac{e^{\alpha(1-x)}-1}{e^\alpha-1}\right], & \text{if } \alpha \in \mathbb{R}\setminus\{0\}, \\ x & \text{if } \alpha = +\infty. \end{cases}$$

What we want to do now is applying what we have seen so far and, especially by using the generators s_α, prove that we can gradually (continuously) move upward from the *probabilistic* S_{prob} to the *Lukasiewicz t-conorm* S_{Luk}, via a family of strict and inframodular t-conorms, which is exactly $\{S_\alpha\}_{\alpha\geq 0}$.

Proposition 4.89. *S_α is inframodular if and only if $\alpha \geq 0$.*

Proof. In the literature, it is already known that S_α is submodular if and only $\alpha \geq 0$. Indeed, since s_α is twice derivable for any $\alpha \neq \pm\infty$, a straightforward calculation shows that

- $\dfrac{\mathrm{d}s_\alpha}{\mathrm{d}x}(x) = -\dfrac{\alpha}{e^{\alpha(x-1)} - 1}, \qquad \forall\, x \in [0, 1]$
- $\dfrac{\mathrm{d}^2 s_\alpha}{\mathrm{d}x^2}(x) = \dfrac{\alpha}{4\left[\sinh\left(\frac{\alpha(x-1)}{2}\right)\right]^2}, \qquad \forall\, x \in [0, 1]$

so that s_α is convex if and only if $\frac{\mathrm{d}^2}{\mathrm{d}x^2}s_\alpha \geq 0$, which occurs if and only if $\alpha \geq 0$. This proves that S_α is submodular if and only if $\alpha \geq 0$.

However, in order to prove that such S_α's are even inframodular, we have to show that $1/s'_\alpha$ is concave for all $\alpha \geq 0$. Now

$$g_\alpha(x) := \frac{1}{s'_\alpha(x)} = -\frac{e^{\alpha(x-1)} - 1}{\alpha}, \qquad \forall\, x \in [0, 1]$$

so, differentiating twice gives:

- $\dfrac{\mathrm{d}g_\alpha}{\mathrm{d}x}(x) = -e^{\alpha(x-1)}, \qquad \forall\, x \in [0,1]$
- $\dfrac{\mathrm{d}^2 g_\alpha}{\mathrm{d}x^2}(x) = -\alpha e^{\alpha(x-1)}, \qquad \forall\, x \in [0,1],$

and, of course, g_α is concave if and only if $\alpha \geq 0$. □

4.3 Activation Functions

The *threshold models* we dealt with in chapter 2 are based on node-specific *activation functions*. With this expression, we mean set functions $f^j : 2^{V^j} \to [0,1]$ that are monotone (and possibly submodular). In that context, they may be interpreted as ways of measuring the total influence of entire groups of j's neighbors.

Here, with a natural construction, we define an *influence vector* $\mathbf{x} \in [0,1]^n$ which represents at the same time the neighbors' activity levels (active-inactive) and their influence weights; by doing this, "aggregating influences" by means of activation functions in threshold models turns out to be the same as considering an aggregation function $F : [0,1]^n \to [0,1]$. So, the "reasonable ways" may be defined through properties of F and some of the techniques seen so far in this chapter may be used. In particular, we will see that if F is submodular and monotone, the same holds for its correspondent activation function f.

Let us now make the situation more formal and clear, by introducing some definitions.

Definition 4.90. *We consider the following framework that we call* ***structure of influence*** $(V, \mathbf{w})$*:*

- *a set $V = \{1, ..., n\}$ representing the friends/**neighbors** (of the node v of a given network/graph);*
- *a set of weights $\{b_i\}_{i \in V}$, where $w_{iv} = w_i \in [0,1]$ represents the influence of i over v; and associated to this, a vector of **influence weights** $\mathbf{w}_v = \mathbf{w} \in [0,1]^n$*
- *the **active neighbors** are represented by a subset $A \subseteq V$, to which we associate the indicator function $1_A : V \to \{0,1\}$, defined by $1_A(i) = 1$ if and only if $i \in A$, i.e. i is active.*

Notice that here we fix on a node v in a network and study what happens in its neighborhood. In particular, this means that, throughout, whenever the context is clear, we will avoid to write v for ease of notations. Moreover, as convention, when we are considering the whole network, if the node i is not a neighbor of v, then we will define the weight $w_{iv} := 0$.

We already observed that the widely studied *Linear Threshold Model* of the first chapter has the limitation that the weights are needed to sum less than 1. Such condition, in our opinion, is perhaps counter-intuitive: if we require that $\sum_{i \in V} w_{iv} \leq 1$, then each single node v of the network cannot have many heavy influencing friends, i.e. with weights $w_{iv} \approx 1$, because the sum of these influences has to be limited. Besides, it is worth noticing that a simple re-scaling or re-normalization of the weights is not possible, in general, because it is not clear how this "local" re-scales may affect the whole process and network.

Definition 4.91. *An **activation function** for a structure of influence* $(V, \mathbf{b})$ *is a function* $f : 2^V \to [0,1]$ *that is* monotone *and* submodular.

The name obviously originates from the General Threshold Model, in which a node v becomes active when its threshold function $f_v(S)$ exceeds its threshold θ_v, given the active neighbors $S \subseteq V$.

Now, we want to show that it is possible to define a "natural" structure, associated to every possible *structure of influence* which can be made fixing the node v and its neighbors V.

Definition 4.92. *Given an ordered set of neighbors* $V = \{1, ..., n\}$ *of a node* v *in a network*[22]*, we define the **vector of state** as the function*

$$\begin{aligned} \mathbf{X} : 2^V \times [0,1]^n &\longrightarrow [0,1]^n \\ (A, \mathbf{w}) &\longmapsto \mathbf{X}(A, \mathbf{w}) = \sum_{i=1}^{n} 1_A(i) w_i \cdot \mathbf{e}_i, \end{aligned}$$

where $\mathbf{e}_i \in \mathbb{R}^n$ *is the* i*th vector the standard basis and* $\cdot$ *is the product of a scalar with a vector.*

The next result puts together all the properties of the *vector of state* $\mathbf{X}$.

Proposition 4.93.

1. *Fixed* $\mathbf{w} \in [0,1]^n$*, the set-function* $\mathbf{X}(\cdot, \mathbf{w}) : 2^V \to [0,1]$
   - *is non-decreasing with respect to the inclusion of subsets of* V*, i.e.*
     $$A \subseteq B \subseteq V \quad \Longrightarrow \quad \mathbf{X}(A, \mathbf{w}) \leq \mathbf{X}(B, \mathbf{w}),$$
     where $\leq$ *is taken by components;*
   - *preserves the lattice structures of* $\left(2^V, \cap, \cup\right)$ *and* $([0,1]^n, \wedge, \vee)$*, i.e. for all* $A, B \subseteq V$
     - $\mathbf{X}(A \cap B, \mathbf{w}) = \mathbf{X}(A, \mathbf{w}) \wedge \mathbf{X}(B, \mathbf{w})$,

[22] For ease of comprehension, we omit the index v, and we use $\mathbf{w}$ instead of $\mathbf{w}_v$ as well as X instead of $X_v^{\mathbf{w}}$.

– $\mathbf{X}(A \cup B, \mathbf{w}) = \mathbf{X}(A, \mathbf{w}) \vee \mathbf{X}(B, \mathbf{w})$.

In particular $\mathbf{X}(\cdot, \mathbf{w})$ *is a homeomorphism of algebras if and only if* $\mathbf{w} = (1, ..., 1)$.

2. *Fixed a subset* $A \subseteq V$, *then* $\mathbf{X}(A, \cdot) : [0,1]^n \to [0,1]$ *is non-decreasing and linear, i.e.*

   - *if* $\mathbf{x} \leq \mathbf{y}$ *then* $\mathbf{X}(A, \mathbf{x}) \leq \mathbf{X}(A, \mathbf{y})$;
   - *for all* $\alpha, \beta \in [0,1]$ *and* $\mathbf{x}, \mathbf{y} \in [0,1]^n$:

   $$\mathbf{X}(A, \alpha\mathbf{x} + \beta\mathbf{y}) = \alpha \cdot \mathbf{X}(A, \mathbf{x}) + \beta \cdot \mathbf{X}(A, \mathbf{y}).$$

3. *In general, it holds that for all* $A, B \subseteq V$ *and* $\mathbf{x}, \mathbf{y} \in [0,1]^n$

   - $\mathbf{X}(A, \mathbf{x}) \wedge \mathbf{X}(B, \mathbf{y}) = \mathbf{X}(A \cap B, \mathbf{x} \wedge \mathbf{y})$,
   - $\mathbf{X}(A, \mathbf{x}) \vee \mathbf{X}(B, \mathbf{y}) = \mathbf{X}(A \setminus B, \mathbf{x}) + \mathbf{X}(B \setminus A, \mathbf{y}) + \mathbf{X}(A \cap B, \mathbf{x} \vee \mathbf{y})$.

Proof.

1. The fact that $\mathbf{X}(\cdot, \mathbf{w})$ is monotone follows straightforward from the definition. Now, let us prove the two algebraic properties. They are based on the following equalities, which are true for any $A, B \subseteq V$ and $i \in V$:

   $$1_{A \cap B}(i) = 1_A(i) 1_B(i) = \left\{ \begin{array}{ll} 1, & \text{if } i \in A \cap B, \\ 0, & \text{if } i \notin A \cap B, \end{array} \right\} = \min\{1_A(i), 1_B(i)\},$$

   and analogously

   $$1_{A \cup B}(i) = 1_A(i) + 1_B(i) - 1_{A \cap B}(i) = \left\{ \begin{array}{ll} 1, & \text{if } i \in A \cup B, \\ 0, & \text{if } i \notin A \cup B, \end{array} \right\} = \\ = \max\{1_A(i), 1_B(i)\}.$$

   By applying them to our case, we get what we needed:

   - $\mathbf{X}(A \cap B, \mathbf{w}) = \sum_i 1_{A \cap B}(i) w_i \mathbf{e}_i = \sum_i \min\{1_A(i), 1_B(i)\} w_i \mathbf{e}_i = \mathbf{X}(A, \mathbf{w}) \wedge \mathbf{X}(B, \mathbf{w})$;
   - $\mathbf{X}(A \cup B, \mathbf{w}) = \sum_i 1_{A \cup B}(i) w_i \mathbf{e}_i = \sum_i \max\{1_A(i), 1_B(i)\} w_i \mathbf{e}_i = \mathbf{X}(A, \mathbf{w}) \vee \mathbf{X}(B, \mathbf{w})$.

2. The properties of $\mathbf{X}(A, \cdot)$ both follow by a direct calculation from the definition. In particular:

   $$\mathbf{X}(A, \alpha\mathbf{x} + \beta\mathbf{y}) = \sum_{i=1}^{n} 1_A(i)(\alpha x_i + \beta y_i)\mathbf{e}_i = \\ = \alpha \sum_{i=1}^{n} 1_A(i) x_i \mathbf{e}_i + \beta \sum_{i=1}^{n} 1_A(i) y_i \mathbf{e}_i = \alpha \cdot \mathbf{X}(A, \mathbf{x}) + \beta \cdot \mathbf{X}(A, \mathbf{y}).$$

3. The last two properties are similar to the ones proved in the first step of this prove.

□

Remark 4.94. Notice that $\mathbf{X}(\cdot, \mathbf{w})$ is not an homomorphism of algebras because it does not verify the condition $\mathbf{X}(V, \mathbf{w}) = \mathbf{1}$, when $\mathbf{w} \neq \mathbf{1}$. In addition to this, if $A = \emptyset$ then $\mathbf{X}(\emptyset, \mathbf{w}) = \mathbf{0}$, but the viceversa is not necessarily true.

4.3.1 From Aggregation to Activation

This last subsection is, indeed, the link between the aggregation functions that we have studied in the previous sections before and the activation functions, which are the objects treated here. In particular, we want to show that if we aggregate the influences with a submodular aggregation function, then necessarily this generates an activation function which is also submodular (and monotone).

Definition 4.95. *Given a structure of influence $(V, \mathbf{w})$ for a node v and an aggregation function $F^v : [0,1]^n \to [0,1]$, we define the associated **activation function** by*

$$f^v(S) := F\big(X_v^{\mathbf{w}}(S)\big), \qquad \forall\, S \subseteq V.$$

Proposition 4.96. *Let $(V, \mathbf{w})$ be a structure of influence for the node v. If the aggregation function F^v is monotone and submodular, the same holds for its associated activation function f^v.*

Proof. For ease of notation, we omit v. For the monotonicity of f, given $A \subseteq B$ then $\mathbf{X}(A, \mathbf{w}) \leq \mathbf{X}(B, \mathbf{w})$ from the previous lemma. So, using the monotonicity of F, the assertion becomes clear.

Concerning the submodularity, let $A, B \subseteq V$ be two sets. From the previous lemma and by using the submodularity of F, we have

$$\begin{aligned} f(A \cap B) + f(A \cup B) &= F\left(\mathbf{X}(A \cap B, \mathbf{w})\right) + F\left(\mathbf{X}(A \cup B, \mathbf{w})\right) = \\ &= F\left(\mathbf{X}(A, \mathbf{w}) \wedge \mathbf{X}(B, \mathbf{w})\right) + F\left(\mathbf{X}(A, \mathbf{w}) \vee \mathbf{X}(B, \mathbf{w})\right) \leq \\ &\leq F(\mathbf{X}(A, \mathbf{w})) + F(\mathbf{X}(B, \mathbf{w})) = f(A) + f(B). \end{aligned}$$

□

Proposition 4.97. *Let $(V, \mathbf{w})$ be a structure of influence. If $F, G : [0,1]^2 \to [0,1]$ are two aggregation functions such that $F \leq G$, then their associated activation functions f, g also verify $f \leq g$.*

Remark 4.98. Given a bivariate associative and symmetric aggregation function (for example, a t-conorm) $S^{(2)} : [0,1]^2 \to [0,1]$, let $S^{(k)} : [0,1]^k \to [0,1]$ be its k-ary iterations, made by associativity. Now, consider a structure of influence $(V, \mathbf{w})$ and a set of active neighbors $A \subseteq V$. Let us denote with $A = \{i_1, ..., i_k\}$, where $|A| = k \leq n = |V|$. Then:

$$f(A) = S^{(n)}(\mathbf{X}(A, \mathbf{w})) = S^{(n)}(w_{i_1}, ..., w_{i_k}, 0, ..., 0) = S^{(k)}(w_1, ..., w_k).$$

This, in particular, means that our framework is enough flexible that it can take into account different network structures with different cardinalities. However, changing from a structure where the node v has k friends to another when it has $n \geq k$ friends happens somehow "homogeneously". In addition, in our approach only the active friends matter, which means that v's inactive neighbors exert *no dissuasive effect* on v.

Chapter 5

Diffusion Maximization

In this chapter we study the *diffusion function* σ, which was defined in the first chapter. In particular, firstly we focus on some properties of σ, seen as a "global" function which involves the whole process in the network, and we see that it inherits some of the "local" properties from a model. Then, in the second section we show that maximizing the diffusion function is a *NP-hard* problem and, so, in the third section we describe two algorithms for approximating the solution of this maximization problem.

5.1 Properties of the Diffusion Function

The study of (some of) the properties of the *diffusion function* σ, introduced in the first chapter, is the main purpose of this section. Since the diffusion function takes into account the whole diffusion process in the whole network, we tend to call its properties "global". As anyone could imagine, in general the behavior of σ is quite hard to predict and so are its "global" properties. Hence, our focus here is on those properties that may be somehow inherited from some more "local" components of the network, which, in some sense, are much more controllable.

In particular, we will see that if we consider a *threshold model* whose "local" *activation functions* are *monotone* and *submodular*, then the result diffusion function σ turns out to be monotone and submodular as well. These same properties are obtained if in a *cascade model*, the probabilities are taken *decreasing*.

In the next section, it will be finally clear why the monotonicity and submodularity of σ are so important. But first, let us recall the definition of *diffusion function* σ.

Definition 5.1. *Given a diffusion model (Progressive Cascade or Threshold) on a network $G = (V, E)$, with $|V| = n$, let $A_0 \subseteq V$ be the set of* initially active nodes *and let $R(A_0) \subseteq V$ the random variable that describes the* active nodes at the end *of the diffusion process, when it starts exactly*

*from A_0. The **diffusion function** $\sigma(A_0)$ is the average of the cardinality of $R(A_0)$, i.e.*

$$\sigma : 2^V \longrightarrow [0, n]$$
$$\sigma(A_0) = \mathbb{E}\left[|R(A_0)|\right].$$

Remark 5.2. Being σ an average which is based on the random variable $R(A_0)$, from a practical point of view its (at least approximated) computation requires an adequate number of repeated simulations.

The remark above stimulates the study of "local" properties of the models in the hope that it can help to find "global" properties of the *diffusion function* σ. In their seminal work [Kempe *et al.*, 2003], the authors were able to prove the two following results.

Proposition 5.3. *Given an arbitrary instance of the Linear Threshold Model (LT), the resulting diffusion function $\sigma_{LT}(\cdot)$ is monotone and submodular.*

Proposition 5.4. *Given an arbitrary instance of the Independent Cascade Model (IC), the resulting diffusion function $\sigma_{IC}(\cdot)$ is monotone and submodular.*

Sketch of the proof.
The monotonicity of σ_{IC} is obvious: the more initially activated nodes we have, the wider the diffusion will be. The proof of the submodularity of σ_{IC} is based on the same reasoning used in Remark 3.2: given an instance of the Independent Cascade Model on the graph $G = (V, E)$, we start by flipping a coin with probability p_{ij} for every edge (i, j) in the graph at the beginning of the process. The coin flip indicates whether the node i has succeeded in activating its neighbor j: in such a case, we call *live* the edge (i, j). Now:

- if we start from a set $A \subseteq V$, then a node $j \in V$ is activated (at the end of the process) if and only if there exists a path from some node in A to j consisting entirely of live edges;
- if we indicate with f the set of outcomes of all coin flips on edges, then we may define the deterministic function $\sigma^f : 2^V \to [0, 1]$ where $\sigma^f(A)$ is the total number of nodes that are activated by the process, when it starts in A and the live edges are established by f;
- if we denote with $R(i, f) \subseteq V$, for all $i \in V$, the set of nodes that are reachable from i via live-edges paths (and when the live-edges are established by f), then

$$\sigma^f(A) = \left| \bigcup_{i \in A} R(i, f) \right|.$$

Having said this, we want to prove that the diffusion function σ_{IC} is submodular: let $A, B \subseteq V$ be two sets such that $A \subseteq B$ and take $i \notin B$.

- consider $\sigma^f(A \cup \{i\}) - \sigma^f(A)$: this is the number of nodes that are in $R(i,f) \setminus \bigcup_{j \in A} R(j,f)$, i.e. the nodes that are reachable from i and that are not already reached from some node in A;

- analogously, consider $\sigma^f(B \cup \{i\}) - \sigma^f(B)$. Since $A \subseteq B$, the marginal contribution of i to the reached nodes from B, i.e. $\bigcup_{j \in B} R(j,f)$, has to be less than the one given to $\bigcup_{j \in A} R(j,f)$. So:

$$\sigma^f(A \cup \{i\}) - \sigma^f(A) \geq \sigma^f(B \cup \{i\}) - \sigma^f(B),$$

  i.e. σ^f is submodular.

Finally, if one thinks of f as a sample point in the probability space defined by all possible outcomes of the flip coins, then the diffusion function σ_{IC} may be seen as the average

$$\sigma_{IC}(A) = \sum_f \mathbb{P}[f] \cdot \sigma^f(A).$$

Then σ_{IC} is submodular because it is a non-negative linear combination of submodular functions. □

Only in Theorem 3 of [Kempe *et al.*, 2005], the authors were able to prove an analogous result in a more general context:

Proposition 5.5. *Let us consider an instance of the General Cascade Model (GC). If the probabilities $p_j(i;\cdot)$ are non-increasing for all $j \in V$ and $i \in V^j$, i.e. $p_j(i;A) \geq p_j(i;B)$ whenever $A \subseteq B \subseteq V$, then the diffusion function $\sigma_{GC}(\cdot)$ is monotone and submodular.*

By using the *Cascade-Threshold equivalence* seen in the section 3.3, we also get the following:

Proposition 5.6. *Let us consider an instance of the General Threshold Model (GT). If the activation functions $f^j : 2^{V^j} \to [0,1]$, for $j \in V$, are non-decreasing and satisfy the* ***normalized submodularity****, i.e.*

$$\text{(nSubM)} \qquad \frac{f^j(A \cup \{i\}) - f^j(A)}{1 - f^j(A)} \geq \frac{f^j(B \cup \{i\}) - f^j(B)}{1 - f^j(B)},$$

for all $A \subseteq B \subseteq V^j$ and all $i \notin B$, then the resulting diffusion function $\sigma_{GT}(\cdot)$ is monotone and submodular.

Proof. Such a threshold model is equivalent to a cascade model whose the probabilities are exactly defined by

$$p_j(i;A) = \frac{f^j(A\cup\{i\}) - f^j(A)}{1-f^j(A)}, \qquad \forall\, A \subseteq V^j,\ i \notin A.$$

Since the hypothesis implies that $p_j(i;\cdot)$ are decreasing, we can apply the previous proposition and get that $\sigma_{GT} = \sigma_{GC}$ is monotone and submodular. □

Remark 5.7.

- Notice that, for a non-decreasing set-function $f : 2^V \to [0,1]$, the *normalized submodularity* is stronger than the *submodularity*: in fact, since $f(A) \leq f(B)$, then $1 - f(A) \geq 1 - f(B)$ and, so, in (nSubM) we have
$$\frac{1}{1-f(A)} \leq \frac{1}{1-f(B)}, \qquad \forall\, A \subseteq B.$$

- We can try to define (nSubM) for associative and symmetric vector-functions $F : [0,1]^2 \to [0,1]$ as follows[1]:
  - **normalized submodularity**:
    $$\text{(nSubM)} \qquad \frac{F^{(k+1)}(\mathbf{x},\varepsilon) - F^{(k)}(\mathbf{x})}{1-F^{(k)}(\mathbf{x})} \geq \frac{F^{(k+1)}(\mathbf{y},\varepsilon) - F^{(k)}(\mathbf{y})}{1-F^{(k)}(\mathbf{y})},$$
    for all $\mathbf{x} \leq \mathbf{y} \in [0,1]^k$ and $\varepsilon \in (0,1]$.
  - **normalized inframodularity**:
    $$\text{(nInfM)} \qquad \frac{F^{(k)}(\mathbf{x}+\mathbf{h}) - F^{(k)}(\mathbf{x})}{1-F^{(k)}(\mathbf{x})} \geq \frac{F^{(k)}(\mathbf{y}+\mathbf{h}) - F^{(k)}(\mathbf{y})}{1-F^{(k)}(\mathbf{y})},$$
    for all $\mathbf{x},\mathbf{y},\mathbf{h} \in [0,1]^k$, with $\mathbf{h} > \mathbf{0}$, such that $\mathbf{x} \leq \mathbf{y}$ and $\mathbf{y}+\mathbf{h} \in [0,1]^k$.

  It is clear that (nInfM) implies both (nSubM), (InfM) (and, consequently, also (SubM)). However, the standard inframodularity (InfM) does not imply (nSubM), and one may use S_{Luk} as a counter-example. As a consequence, the standard submodularity does not imply its normalized version.

- If an associative symmetric aggregation function $F : [0,1]^2 \to [0,1]$ satisfies the normalized submodularity (nSubM), then the same holds for its associated activation function $f = F^{(n)} \circ \mathbf{X}$, which is defined as in the section 4.3.

[1] As in the previous chapter, $F^{(k)}$ indicates the k-iteration of $F \equiv F^{(2)}$ obtained by associativity.

Finally, in [Mossel, Roch, 2007] the authors prove that the *monotonicity* and *submodularity* of the activation functions of a threshold model are inherited from the diffusion function:

Proposition 5.8. *Let us consider a General Threshold Model on the network $G = (V, E)$. If the activation functions f^j are monotone and submodular for all $j \in V$, then the resulting diffusion function σ_{GT} is monotone and submodular.*

5.2 Maximization: Top-k Set Selection Problem

Definition 5.9. *Given a network $G = (V, E)$, with $|V| = n$, and a diffusion process (Cascade or Threshold Model), an instance of the* ***Top-k Set Selection Problem*** *is:*
"given $k \leq n$, find a set $A \subseteq V$ with cardinality k that maximizes the diffusion function $\sigma(\cdot)$."
In other words:

$$\begin{cases} \textit{maximize } \sigma(A) \\ \textit{subject to } A \subseteq V : |A| = k. \end{cases}$$

The *top-k set selection problem* is *NP-hard* for both Cascade and Threshold Model. To show that it suffices to see it in their respective special cases, as seen in [Kempe *et al.*, 2003]:

Proposition 5.10.

1. *The top-k set selection problem is NP-hard for the Linear Threshold Model.*

2. *The top-k set selection problem is NP-hard for the Independent Cascade Model.*

Sketch of the proof.

1. The proof is based on reducing an instance of the so-called *Vertex Cover problem*, which is *NP-complete*, to a top-k set selection problem. Let us consider an instance of the vertex cover problem: given an undirected graph $G = (V, E)$ and an integer k, the problem is to find a set $S \subset V$ of k nodes such that each edge of G has at least one endpoint in S.

   We now define a corresponding instance of the top-k set selection problem. First, we take the graph G and make it directed considering each edge in both directions. Now, if one can find a vertex cover S of size k in G, then it is possible to deterministically obtain $\sigma(A) = n$, simply by taking $A = S$. On the other hand, this is the only way to get a set $A \subseteq V$ such that $\sigma(A) = n$.

2. Analogously, the argument here uses a reduction from the *Set Cover problem*, which is also *NP-complete*.

□

Now, the general cases may be consequently deduced now and something more precise is said in the next corollary, whose proof may be based on a reduction from the *Set Cover problem* as well.

Corollary 5.11. *The top-k set selection problem is NP-hard also for the General Threshold Model and the General Cascade Model. More precisely: it is NP-hard to approximate the influence maximization problem to within a factor of* $n^{1-\varepsilon}$*, for any* $\varepsilon > 0$.

The *NP-hardness* of the *top-k set selection problem* stimulates the research of algorithms to approximate the solutions. Two of such algorithms are the objects of the study of the next section.

5.3 Approximation Algorithms

5.3.1 Greedy algorithm

The algorithm proposed in [Kempe *et al.*, 2003] and [Kempe *et al.*, 2005] is based on a greedy hill-climbing strategy: since we want to maximize the diffusion function σ, at each step we take the node that maximizes its *marginal contribution* to the diffusion. This simple idea is well clear in the following:

Algorithm 1 Greedy Algorithm

1: Start with $A = \emptyset$
2: **for** $i \leq k$ **do**
3: let v_i be a node that maximizes $\sigma(A \cup \{v_i\}) - \sigma(A)$;
4: set $A \leftarrow A \cup \{v_i\}$;
5: **end for**
6: **return** A

Remark 5.12. The *greedy algorithm* provides an "approximated solution" for the *top-k set selection problem*, however, since it is *NP-hard* to compute the/an optimal solution of the problem, especially with very big networks, then it is a problem how to evaluate the goodness of this approximation.

Motivated by the previous remark, one may understand the importance of the following result.

Proposition 5.13. *Let $A^* \subseteq V$ be an optimal solution for the* top-k set selection problem*, i.e.A^* maximizes σ among the sets of cardinality k, and let $A \subseteq V$ be the set given by the* Greedy Algorithm *1. Then*

$$\sigma(A) \geq \left(1 - \frac{1}{e}\right) \cdot \sigma(A^*).$$

In other words, A provides a solution which is at least a 63%-approximation of the optimal solution A^*.

Remark 5.14. Notice that the algorithm needs to calculate σ, which may be hard or even prohibitive, for example in the case of a very big network. To our knowledge, it is still an open question how to compute exactly σ by an efficient method. However, by simulating a random process many times, one may get arbitrarily good approximations of $\sigma(A)$: more precisely, for any $\varepsilon, \delta > 0$ and $n \in \mathbb{N}$, it is possible to compute an approximation of $\sigma(A)$ that is in $[\sigma(A) - \varepsilon, \sigma(A) + \varepsilon]$ with probability at least $1 - \delta$, polynomially in ε, δ and n. Moreover for any $\varepsilon' > 0$, this approximation allows us to find a node $v \in V$ whose marginal gain $\sigma(A \cup \{v\}) - \sigma(A)$ is a $(1 - \varepsilon')$-approximation of the maximal marginal gain.

In the light of the above remark, the previous proposition may be written more precisely:

Proposition 5.15. *Let A^* be an optimal solution and let A be the result of the* Greedy Algorithm*, where at each step the node v_i is the $(1 - \varepsilon')$-approximation of the best node. Then*

$$\sigma(A) \geq \left(1 - \frac{1}{e} - \varepsilon''\right) \cdot \sigma(A^*),$$

where ε'' depends polynomially on ε'.

Remark 5.16. Some practical experiments made in [Kempe *et al.*, 2003] have shown that the *Greedy Algorithm* provides a solution that is often better than other solutions provided by other *heuristics*. Among others, there are comparisons with:

- *High-degree Heuristic*: the selected nodes are the ones that have the highest **degree**, which is defined as the number of (outgoing) edges of a node;
- *Distance Centrality Heuristic*: the **distance** between two nodes is defined as the length of the minimal path that connects the nodes. This heuristic, then, chooses the k nodes that have the minimal average distance.

However, there exist other algorithms/heuristics, which possibly perform better than the *Greedy Algorithm*, for example the *Shapley-value based Algorithm*, which we will see in the next section, and a *Genetic Algorithm* (see [Wang *et al.*, 2013]).

5.3.2 Shapley-value based algorithm

The **Shapley value**, introduced in *(cooperative) game theory*, has been used in networks to define the *social power* and the *centrality*, see [Gómez *et al.*, 2003]. Using these ideas, in [Narayanam, Narahari, 2010] the authors developed an algorithm that is based on the Shapley-value, by which they find the most k influential/central/power nodes and select them as a solution of the *top-k set selection problem.*

Preliminaries: Cooperative Game Theory and Networks

Definition 5.17. *A **cooperative game** with* transferable utility *(*TU game*) is a pair* (V, v) *where*

- $V = \{1, ..., n\}$ *is the set of* players*;*
- $v : 2^V \to \mathbb{R}$*, with* $v(\emptyset) = 0$*, is called* characteristic function *of the game*

The subsets $S \subset V$ are called *coalitions* of players and $v(S)$ is the *value of the coalition*, representing how much the coalition S can gain (without any help from other players in $V \setminus S$).

In cooperative game theory, the concept of *solution* is not univocal. In general, a solution is a "proposal" of the allocation of the total value of the game among the individuals. Intuitively speaking, there are some proposals which are more likely to be accepted by the players, whereas others are almost surely rejected.

Example 5.18. Let us consider the following game. Three people inherit a house, but in the notary deed they find out that the house can go to them only if at least two of them agree to share it. This game may be modeled as follows: with $V = \{1, 2, 3\}$ and $v : 2^V \to 0, 1$, where

$$v(S) = \begin{cases} 1, & \text{if } |S| \geq 2, \\ 0, & \text{otherwise.} \end{cases}$$

The formalization means exactly that: if the players 1, 2 and 3 agree, they get the house; if two of them agree, they also get the house; in any other case, i.e. if they cannot find an agreement, they lose the house.
In this context, the problem of finding a "solution" for this game is: once the coalitions are established, say that $S = \{1, 2\}$ agree and they get the house, how to "fairly" divide the whole value $v(S) = 1$ between 1 and 2? Obviously, one reasonable allocation to 1, 2 and 3 could be written respectively as the vector $(1/2, 1/2, 0)$, but then, player 3 could offer an agreement to player 1 and propose a distribution that is more favorable for 1, i.e. something like

$$\left(\frac{1}{2} + \varepsilon,\ 0,\ \frac{1}{2} - \varepsilon\right),$$

with any $\varepsilon > 0$. Of course, player 1 would agree, since it gains more. Analogously, player 2 could do the same trick and propose an agreement between 2 and 3, which is more convenient for 3:

$$\left(0,\ \frac{1}{2}-\varepsilon,\ \frac{1}{2}+\varepsilon\right).$$

Clearly, this argument may be repeated infinitely many times and, so, it is not trivial how to find a somehow "stable" solution/allocation (if it exists)[2]. Notice that if the game were only between two players, i.e. $V = \{1, 2\}$, meaning that they get the house if and only if both agree on sharing it, then the solution $(1/2, 1/2)$ would be stable.

One of the main "solution concepts" used in cooperative game theory is the *Shapley value*, so called in honor of Lloyd S. Shapley. It has been used in a lot of applications in many different fields (see [Moretti, Patrone, 2008]).

Definition 5.19. *The* ***Shapley value*** *of the game* (V, v) *is the vector* $\boldsymbol{\phi} \in \mathbb{R}^n$ *defined by:*

$$\phi_i = \sum_{S \subseteq V \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!}\big[v(S \cup \{i\} - v(S)\big], \qquad \forall\, i \in V,$$

where ϕ_i *represents the payoff of the player* $i \in V$*.*

Remark 5.20. The idea behind the Shapley value is that it represents the importance or value of the player i in the game. Notice that it is based on the marginal contributions of i to every possible coalition S.

Remark 5.21. The Shapley value has the following desirable properties (for a solution concept):

- **efficiency**: the total value of the game $v(V)$, reached when all players act together as one coalition, is distributed, i.e. $\sum_{i \in V} \phi_i = v(V)$.
- **symmetry**: if the players i and j have the same importance/value in the game, i.e. $v(S \cup \{i\}) = v(S \cup \{j\})$ for any $S \subseteq V \setminus \{i, j\}$, then the Shapley value assigns to them the same amount $\phi_i(v) = \phi_j(v)$;
- **linearity**: if two games (V, v_1) and (V, v_2) are combined together, then each player gets what she could get independently in each game: $\phi_i(v_1 + v_2) = \phi_i(v_1) + \phi_i(v_2)$;
- **null player**: if the player i does not apport anything to any coalition, i.e. $v(S \cup \{i\}) = v(S)$ for all $S \subseteq V$, then its Shapley value is $\phi_i(v) = 0$.

[2]This may be formally proved by introducing other "solution concepts" such as the **core** of a game.

Example 5.22. In the example above, for the symmetry of the game, the Shapley value of each player has to be exactly $\phi_i(v) = v(V)/3 = 1/3$, for all $i = 1, 2, 3$. Indeed, taking for example $i = 1$, then the coalition $S \subseteq V \setminus \{1\}$ are exactly $\emptyset, \{2\}, \{3\}, \{2, 3\}$, so:

$$\begin{aligned}\phi_1(v) = \sum_{S \in V \setminus \{1\}} \frac{|S|!(3-|S|-1)!}{3!}\big[v(S \cup \{1\}) - v(S)\big] = \\ = \frac{1}{3}\big[v(\{1\}) - v(\emptyset)\big] + \frac{1}{6}\big[v(\{1,2\}) - v(\{2\})\big] + \\ + \frac{1}{6}\big[v(\{1,3\}) - v(\{3\})\big] + \frac{1}{3}\big[v(\{1,2,3\}) - v(\{2,3\})\big] = \\ = 0 + \frac{1}{6} + \frac{1}{6} + 0 = \frac{1}{3}.\end{aligned}$$

In graph theory, the Shapley value was firstly used by Roger B. Myerson in [Myerson, 1977] to define the so-called **Myerson value**. He took a game v and a graph $G = (V, E)$ and considered the Shapley value of the restricted game $v|_G$, in which the coalitions S are possible only if the individuals of S are connected in the associated graph G. The Shapley value of this restricted game $\boldsymbol{\phi}(v|_G)$ is the Myerson value.

Later, in [Gómez *et al.*, 2003], the authors combined the *Shapley value* and the *Myerson value* and found out that: if we have a game (V, v) and a graph $G = (V, E)$, then taking the difference between the Myerson value and the Shapley value generates a "centrality" measure for the node i in the network G:

$$\gamma_i(G) = \phi_i(v|_G) - \phi_i(v), \qquad \forall\, i \in V.$$

The interpretation of $\gamma_i(G)$ is the following: since the Shapley value is a measure of the value of the player i, by taking the difference between the game $v|_G$ and v, we get exactly how much i gains from occupying its own position in the network G, independently of the game v.

Example 5.23. Let us consider the game above where the player 1 and 3 cannot communicate, i.e. the is the following graph $G = (V, E)$ where $V = \{1, 2, 3\}$ and the only existing edges are $E = \{(1, 2), (2, 3)\}$, i.e.

$$(1) \longleftrightarrow (2) \longleftrightarrow (3).$$

The situation has changed and the characteristic function of this *graph restricted game* is

$$v|_G(S) = \begin{cases} 1, & \text{if } |S| \geq 2 \text{ and } 2 \in S, \\ 0, & \text{otherwise.} \end{cases}$$

The Myerson value is the Shapley value for the game $v|_G$, so proceeding as

before:

- $\phi_1(v|_G) = \phi_3(v|_G) = \frac{1}{3} \cdot 0 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 0 + \frac{1}{3} \cdot 0 = \frac{1}{6}$
- $\phi_2(v|_G) = \frac{1}{3} \cdot 0 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 + \frac{1}{3} \cdot 1 = \frac{2}{3}$,

which highlights how player 2 here plays a crucial role in order to get the house, whereas player 1 and 3 have less power than before. In fact, in the graph G the node (2) is more central than the others.[3]

Shapley-value based Heuristic Algorithm

In order to find the solution of the *top-k set selection problem*, the approach proposed in [Narayanam, Narahari, 2010] is the following: given a diffusion model in a network $G = (V, E)$, let us consider a cooperative game (V, v), where V is the set of nodes/players and the characteristic function $v : 2^V \to \mathbb{R}$ is defined by means of the diffusion function, i.e. $v(S) = \sigma(S)$, assuming that $v(\emptyset) = 0$. Intuitively speaking, the value of the coalition S (i.e. set of initially active nodes) is given by how much it can diffuse in average.

Having said this, the fundamental steps in the algorithm and, indeed, of this section are:

1. using the Shapley value to calculate a *ranking list* of the most central/influential nodes;
    - the Shapley value is hard to compute exactly, so it needs to be approximated;
2. select k nodes from the list;
    - the naive approach of selecting the first/highest k nodes may not be the optimal choice if, for example, the top-ranked nodes are clustered together.

Construction of the ranking list based on the Shapley value The exact calculation of the Shapley value is a hard computational problem, because in order to determine the marginal contributions of the player i one has to deal with $n!$ permutations. Given the importance of this value, a number of approximations have been proposed (see, for example [Aadithya *et al.*, 2010].)

The approach followed here is based on the following observation: the Shapley value of the player i is the average marginal contribution that player i gives to any coalition, when one assumes that all orderings have the same probability. To formalize this, let us consider the set $\Omega = \mathcal{P}(V)$ of all $n!$

[3] Whatever "centrality" measure one may adopt here (see [Garg, 2009]).

permutations of V. Now, given $\pi \in \Omega$, let us indicate with $P_i^{(\pi)} \subseteq V$ the set of all nodes that appear before the node i in the permutation π. Then, the Shapley value coincides with

$$\phi_i(v) = \frac{1}{n!} \sum_{\pi \in \Omega} \left[v\left(P_i^{(\pi)} \cup \{i\}\right) - v\left(P_i^{(\pi)}\right) \right].$$

To approximately calculate ϕ_i, we can use a randomly sampled set of permutations $\widetilde{\Omega}$, whose cardinality $|\widetilde{\Omega}| = \mathcal{O}(n)$ is polynomial in n. Then, we proceed as in the **Algorithm 2** of [Narayanam, Narahari, 2010] and get, as output a set of ordered nodes of V called $RankList = \{i_1, ..., i_n\}$.

How to choose the top-k nodes from *RankList*? It is quite intuitive to think that if the nodes are (adequately/badly) clustered and they are the first in *RankList*, this may not be a good choice for the *top-k set selection problem*: in fact, it may be possible that we choose the k nodes and they diffuse very well and effectively in their neighborhood/cluster but do not manage to "go out" of their clusters and infect other nodes of the network. Motivated by this observation, the choice of the k nodes from *RankList* would try to avoid *adjacent* nodes, as in the **Algorithm 3** of [Narayanam, Narahari, 2010]:

- let *Top-kList* be the output list and start with $Top\text{-}kList = \emptyset$.
- take the first node i_1 of *RankList* and add it to *Top-kList*;
- take the second node i_2 of *RankList*. If i_2 is not adjacent to i_1, add it to *Top-kList*;
- in general, at the step r, take the node i_r and add it to *Top-kList* if it is not adjacent to any of the nodes already in *Top-kList*;
- after a certain number of steps, it may be possible that every node j is either in *Top-kList* or adjacent to a node in *Top-kList*. If this happens and the cardinality of *Top-kList* is still $c < k$, then add to *Top-kList* the remaining $k - c$ highest ranked nodes in $RankList \setminus Top\text{-}kList$;
- at the end of this process, declare the nodes in *Top-kList* as the top-k nodes.

This procedure gives a heuristical solution to the *top-k set selection problem*.

Other remarks about the Shapley-value based Algorithm

The λ-coverage Problem Another very intuitive problem related to diffusion in networks is choosing a minimal set capable to activate/influence a given percentage λ of the nodes in the network. Formally:

Definition 5.24. *Given a diffusion process in a network $G = (V, E)$, the* ***λ-coverage Problem*** *is:*
"given (a percentage) $\lambda \in [0, 1]$, find a set $A \subseteq V$, whose cardinality is minimal, such that $|\sigma(A)|/|V| \geq \lambda$."

Indeed, a minor modification of the *Shaple-value based Algorithm* proposed above provides a solution for the λ-coverage problem:

1. make a *RankList* of the nodes as above;
2. use the **Algorithm 3** of [Narayanam, Narahari, 2010] with $k = n$: this gives us *Top-nList*;
3. determine the smallest value $k \leq n$ such that the first k nodes in *Top-nList*, chosen as initially activated nodes, give a diffusion at the end of the process which activates at least $\lambda \cdot |V|$ nodes.

Results with Non-Submodular Activation Functions in Threshold Models In [Narayanam, Narahari, 2010] the authors claim that their *Shapley-value based Algorithm* works well also in non-submodular situations. In particular, they propose two example of *threshold models* with link-specific weights w_{ij}, where the *activation functions* f^j are non-submodular[4]:

- *Multiplication Threshold Model*:

$$f^j_{mult}(A) = \prod_{i \in A} w_{ij},$$

  where the weights $w_{ij} \in [0, 1]$ are such that $\sum_{i \in V^j} w_{ij} \leq 1$ and w_{ij} represents the level of influence of i over j.

- *Minimum Threshold Model*:

$$f^j_{min}(A) = \min_{i \in A} \{\alpha_i \cdot w_{ij}\},$$

  where the (fixed) parameters α_i are non-negative for all $i \in V^j$.

However, in our opinion and interpretation, these two models do not correspond to an *aggregative* or *cumulative effect* and, also, they do not deal with *positive influences*, as we did in the previous chapter. More precisely, both models come out from the use of *t-norms* instead of *t-conorms*, respectively the *probabilistic t-norm* T_{prob} and the *minimum t-norm* T_{min}, which are dual to S_{prob} and S_{max}. Now, we think that the problem with f^j_{mult} and f^j_{min} is that when another active neighbor i comes or when its influence increases, these functions decrease, instead of increasing. This behavior, which is a *conjunctive behavior*, leads us to think that the influences

[4]Here we use the same notations of the first chapter.

taken into account here are not interpretable as *positive*, and, for example, the inactive neighbors here exert a "dissuasive effect" (see also [Chen *et al.*, 2011]).

Bibliography

[Aadithya *et al.*, 2010] Aadithya K.V., Ravindran B., Michalak T.P., Jennings N.R., "Efficient computation of the shapley value for centrality in networks", *Internet and Networks Economics*, pp. 1-13. *Springer Berlin-Heidelberg*, 2010.

[Alsina, Tomas, 1988] Alsina C., Tomas M.S., "Smooth Convex t-Norms do not exist", *Proceedings of the American Mathematical Society*, vol. 102, No. 2, pp. 317-320. 1988.

[Alsina *et al.*, 2006] Alsina C., Frank M.J., Schweizer B., "Associative Functions: Triangular Norms and Copulas", *World Scientific Publishing Co. Pte. Ltd.*, 2006.

[Asavathiratham, 2000] Asavathiratham C., "The Influence Model: a tractable representation for the Dynamics of Networked Markov Chains", *Ph.D. Thesis, MIT*, 2000.

[Beliakov *et al.*, 2008] Beliakov G., Pradera A., Calvo T., "Aggregation Functions: A Guide for Practitioners", *Studies in Fuzziness and Soft Computing*, *Springer*, 2008.

[Borodin *et al.*, 2010] Borodin A., Filmus Y., Oren J., "Threshold Models for Competitive Influence in Social Network", *Lecture Notes in Computer Science; Internet and Network Economics; 6th International Workshop, WINE 2010, Stanford, CA, December 13-17, 2010*, vol. 6484, pp. 539-550, Springer Berlin Heidelberg, 2010.

[Cardin, Manzi, 2009] Cardin M., Manzi M., 2009, "Supermodular and ultramodular aggregation evaluators", *Mathematical Methods in Economics and Finance*, **4** (1), 2009, 1-9, 2013.

[Castellano *et al.*, 2009] Castellano C., Fortunato S., Loreto V., "Statistical physics of social dynamics", *Reviews of modern physics*, 81 (2), 591. 2009.

[Chen *et al.*, 2010] Chen W., Yuan Y., Zhang L., "Scalable influence maximization in social networks under the linear threshold model", *Data*

Mining (ICDM), 2010 IEEE 10th International Conference on, IEEE, 2010.

[Chen *et al.*, 2011] Chen W., Collins A., Cummings R., Ke T., Liu Z., Rincon D., Sun X., Wang Y., Wei W., Yuan Y., "Influence Maximization in Social Networks when Negative Opinions may emerge and propagate", *SDM* (pp. 379-390), 2011.

[Estrada, 2007] Estrada E., "Characterization of topological keystone species: local, global and meso-scale centralities in food webs", *Ecological Complexity*, 4 (1) pp. 48-57. 2007.

[Fazli *et al.*, 2012] Fazli M., Ghodsi M., Habibi J., Khaliladabi P.J., Mirrokni V., Sadeghabad S.S., "On the Non-progressive Spread of Influence through Social Networks", *LATIN 2012: Theoretical Informatics*, *Lecture Notes in Computer Science*, vol. 7256, 2012, pp. 315-326.

[Forster *et al.* 2012] Förster M., Grabisch M., Rusinowska A., "Ordered Weighted Averaging in Social Network", *Documents de travail du Centre d'Economie de la Sorbonne, Universit de Panthon-Sorbonne (Paris 1)*, 2012.

[Garg, 2009] Garg M., "Axiomatic Foundations of Centrality in Networks", *Available at SSRN 1372441*, 2009.

[Grabisch, Rusinowska, 2011] Grabisch M., Rusinowska A., "A model of influence based on aggregation functions", *Documents de travail du Centre d'Economie de la Sorbonne, 11058, Universit de Panthon-Sorbonne (Paris 1)*, 2011.

[Granovetter, 1978] Granovetter M., "Threshold Models of Collective Behavior", *The American Journal of Sociology*, vol. 83, no. 6, pp. 1420-1443, 1978.

[Gómez *et al.*, 2003] Gómez D., González-Arangena E., Manuel C., Owen G., del Pozo M., Tejada J., "Centrality and power in social networks: a game theoretic approach", *Mathematical Social Sciences*, 46 (2003) 27-54.

[Kalai, Safra, 2006] Kalai G., Safra S., "Threshold phenomena and influence: perspective from mathematics, computer science and economics", *Computational complexity and statistical physics*, (2006): 25-60.

[Kempe *et al.*, 2003] Kempe D., Kleinberg J., Tardos E., "Maximizing the Spread of Influence through a Social Network", *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137-146, 2003.

[Kempe *et al.*, 2005] Kempe D., Kleinberg J., Tardos E., "Influential Nodes in a Diffusion Model for Social Networks", *Proceedings of the 32nd international conference on Automata, Languages and Programming*, p. 1127-1138, 2005.

[Kimura *et al.*, 2009] Kimura M., Saito K., Nakano R., Motoda H., "Finding influential nodes in a social network from information diffusion data", *Social Computing and Behavioral Modeling*, pp. 1-8, Springer US, 2009.

[Klement *et al.*, 2000] Klement E.P., Mesiar R., Pap E., "Triangular norms", *Springer*, 2000.

[Klement *et al.*, 2011] Klement E.P., Manzi M., Mesiar R., "Ultramodular aggregation functions", *Information Sciences*, **181** (19), 2011, 4101-4111.

[Lindelauf *et al.*, 2013] Lindelauf R.H.A., Hamers H.J.M., Husslage B.G.M., "Cooperative game theoretic centrality analysis of terrorist networks: The cases of Jemaah Islamiyah and Al Qaeda", *European Journal of Operational Research*, vol. 229, no. 1, p. 230-238, 2013. doi = "http://dx.doi.org/10.1016/j.ejor.2013.02.032", url = "http://www.sciencedirect.com/science/article/pii/S0377221713001653".

[Lovász, 1983] Lovász L., "Submodular Functions and Convexity", *Mathematical Programming: The State of the Art - Bonn 1982* - 11th International Symposium on Mathematical Programming, *Springer-Verlag Berlin Heidelberg*, 1983.

[Manzi, 2009] Manzi M., "New construction methods for copulas and the multivariate case", *PhD Thesis, Università degli Studi di Padova - Dipartimento di Matematica*, 2009.

[Manzi, 2010] Manzi M., "The concept of supermodularity in aggregation functions and copulas", *Seminario di Dottorato 2010/2011, Università di Padova - Dipartimento di Matematica*, 2010. "http://www.academia.edu/1136369/The_concept_of_supermodularity_in_aggregation_functions_and_copulas".

[Marinacci, Montrucchio, 2003] Marinacci M., Montrucchio L., "Ultramodular Functions", *Applied Mathematics Working Paper Series*, no. 13/2003; *International Centre for Economic Research*, 2003.

[Marinacci, Montrucchio, 2008] Marinacci M., Montrucchio L., "On Concavity and Supermodularity", *Journal of Mathematical Analysis and Applications*, vol. 344 (2008), pp. 642-654.

[Milgrom, Roberts, 1990] Milgrom P., Roberts J., "Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities", *Econometrica*, vol. 58, No. 6 (Nov., 1990), pp. 1255-1277.
"http://www.jstor.org/stable/2938316".

[Moretti *et al.*, 2010] Moretti S., Fragnelli V., Patrone F., Bonassi S., "Using coalitional games on biological networks to measure centrality and power of genes", *Bioinformatics*, vol. 26, no. 21, p. 2721-2730, 2010.

[Moretti, Patrone, 2008] Moretti S., Patrone F., "Transversality of the Shapley value", *Top* 16 (1), 1-14, 2008.

[Mossel, Roch, 2007] Mossel E., Roch S., "On the Submodularity of Influence in Social Network", *The American Journal of Sociology*, vol. 83, no. 6, p. 1420-1443, 2007.

[Myerson, 1977] Myerson R.B., "Graphs and Cooperation in Games", *Mathematics of Operation Research*, vol. 2, no. 3 (1977), 225-229.

[Narayanam, Narahari, 2010] Narayanam R., Narahari Y., "A Shapley-value based approach to discover influential nodes in social networks", *IEEE Transaction on Automation Science and Engineering*, 99 (2010), 1-18.

[Navara, Petrık, 2008] Navara M., Petrık M., "Two methods of reconstruction of generators of continuous t-norms", *IMPU*, 6, pp. 1016-1021, 2008.

[Ning, 2012] Ning Li, "Influence on Information Networks and Sparse Representation of Metric Spaces", *PhD thesis, University of Honk Kong*, 2012.
url = http://hdl.handle.net/10722/191190.

[Pathak *et al.*, 2010] Pathak N., Banerjee A., Srivastava J., "A Generalized Linear Threshold Model for Multiple Cascades", *ICDM '10 Proceedings of the 2010 IEEE International Conference on Data Mining*, pp. 965-970, *IEEE Computer Society*, 2010.

[Prakash *et al.*, 2012] Prakash Aditya B., Chakrabarti D., Valler N., Faloutsos C., "Threshold conditions for arbitrary cascade models on arbitrary networks", *Knowledge and information systems*, vol. 33, n. 3, pp. 549-575, 2012, *Springer*.

[Richardson, Domingos, 2001] Richardson M., Domingos P., "Mining the network value of customers", *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 56-66. ACM, 2001.

[Richardson, Domingos, 2002] Richardson M., Domingos P., "Mining knowledge-sharing sites for viral marketing", *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 61-70. ACM, 2002.

[Saito *et al.*, 2010] Saito K., Kimura M., Ohara K., Motoda H. "Selecting information diffusion models over social networks for behavioral analysis", *Machine Learning and Knowledge Discvery in Databases*, pp. 180-195, Springer Berlin-Heidelberg, 2010.

[Salganik *et al.*, 2006] Salganik M.J., Dodds P.S., Watts D.J., "Experimental study of inequality and unpredictability in an artificial cultural market", *science* 311 (5762), 854-856. 2006.

[Sheldon *et al.*, 2012] Sheldon D., Dilkina B.N., Elmachtoub A.N., Finseth R., Sabharwal A., Conrad J., Gomes C.P., Shmoys D.B., Allen W., Amundsen W., Vaughan W., "Maximizing the Spread of Cascades using Network Design", *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI2010)*, 2012.

[Simchi-Levi *et al.*, 2005] Simchi-Levi D., Chen X., Bramel J., "The Logic of Logistics: Theory, Algorithms, and Application for Logistics and Supply Chain Management", 2nd Edition, *Springer Series in Operation Research*, 2013.

[Tang *et al.*, 2009] Tang J., Sun J., Wang C., Yang Z., "Social influence analysis in large-scale networks", *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 807-816. ACM, 2009.

[Valente, 1995] Valente T.W., "Network models of the diffusion of innovations", Vol. 2, No. 2. Cresskill, NJ: *Hampton Press*, 1995.

[Wang *et al.*, 2013] Wang C., Deng L., Zhou G., Jiang M., "A Global Optimization Algorithm for Target Set Selection Problems", *Information Sciences* (2013),
doi: http://dx.doi.org/10.1016/j.ins.2013.09.033

[Yang *et al.*, 2014] Yang Z., Tang J., Xu B., Xing C., "Active Learning for Networked Data Based on Non-progressive Diffusion Model", *Proceedings of the 7th ACM international conference on Web search and data mining*, pp. 363-372, 2014.
"url = http://doi.acm.org/10.1145/2556195.2556223".

[Zapata, Gauthier, 2003] Zapata H., Gauthier W., "Threshold models in theory and practice", *Southern Agricultural Economics Association Annual Meeting*, pp. 1-5. 2003.

[Zeppini *et al.*, 2013] Zeppini P., Frenken K., Kupers R., 2013, "Threshold models of technological transitions", *Eindhoven Center for Innovation Studies (ECIS) working paper series*, No. 13-06, "http://EconPapers.repec.org/RePEc:dgr:tuecis:wpaper:1306".