# Network modeling

V. Amati, J. Lerner, B. Nick

Dept. Computer & Information Science
University of Konstanz

Winter 2011/2012
(version 13 February 2012)

# Contents

# Chapter 1

# Introduction

To fully understand the models presented in the following pages some statistical concepts are necessary. These are briefly defined and recalled in this chapter.

## 1 Continuous random variable

**Definition 1.1.** Let $(\Omega, P)$ be a probability space. A (real-valued) random variable (r.v.) is a function $X : \Omega \to \mathbb{R}$. A random variable $X$ is called *(absolutely) continuous* if there exists a function $f_X(x) : \mathbb{R} \to \mathbb{R}^+$ such that

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du \quad \forall x \in \mathbb{R} \tag{1.1}$$

If $X$ is a continuous random variable, the function $f_X(x)$ in equation (1.1) is called the *probability density function* (p.d.f.). It is assumed that

- $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
- $P(X \in \mathbb{R}) = \int_{-\infty}^{+\infty} f_X(x)dx = 1$

The function $F_X(x)$ in equation (1.1) is called the *cumulative distribution function* (c.d.f.). Its domain is the real line and its codomain is the interval [0,1]. This function describes the probability that $X$ takes value less than or equal to a real number $x \in \mathbb{R}$ and is uniquely defined for each random variable. The cumulative distribution function has the following properties:

- $\lim_{x \to -\infty} = 0$
- $\lim_{x \to +\infty} = 1$
- $F_X(x)$ is a monotone nondecreasing function of $x$, i.e. $F_X(a) \leq F_X(b)$ for $a < b$
- $F_X(x)$ is continuous from the right, i.e. $\lim_{h \to 0} F_X(x+h) = F_X(x)$

*A note on notation:* a random variable will always be denoted with an uppercase letter and its realized value by the corresponding lower case letter. Thus, the random variable $X$ can take value $x$.

From the definition above, it follows that if $X$ is a continuous random variable with density $f_X(x)$, then its probability distribution function is given by

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du \quad \forall x \in \mathbb{R}$$

Conversely, if $X$ is an absolutely continuous random variable with distribution function $F_X(x)$, then its probability density function is given by

$$f_X(x) = \frac{\partial}{\partial x} F_X(x)$$

at those points $x$ where $F_X(x)$ is differentiable and $f_X(x) = 0$ at those points $x$ where $F_X(x)$ is not differentiable.
The density function $f_X(x)$ allows to compute all the probability statements about $X$. For instance, the probability that $X$ takes values between $a$ and $b$, with $a < b$ is computed by

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Recalling the geometric interpretation of a definite integral, the probability that a continuous random variable $X$ takes value in any interval $[a,b]$ is equivalent to the area under the density function $f_X(x)$ on the interval $[a,b]$ (Figure 1.1).

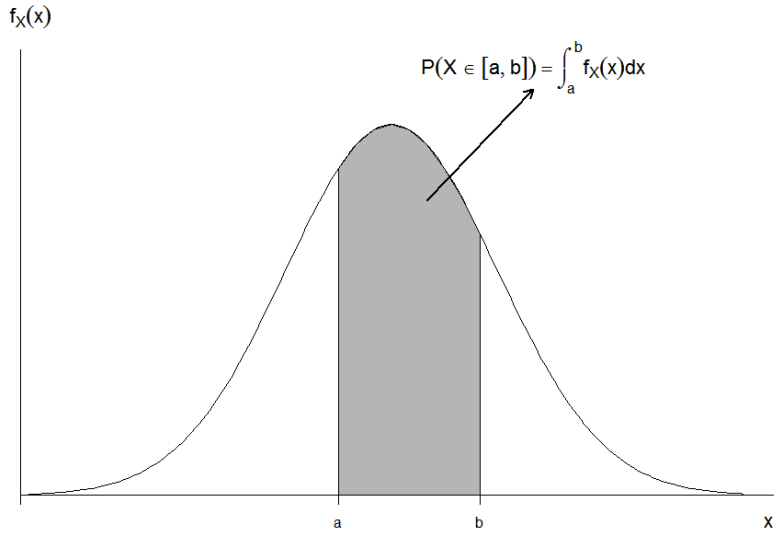



**Figure 1.1**

Thus, intuition suggests that if we let $b = a$ then

$$P(X = a) = \int_a^a f_X(x)dx = 0$$

In words, the probability that a continuous random variable will assume any particular value $x \in \mathbb{R}$ is zero. This explains why we cannot define the probability mass function in the continuous case but we have to generalize this notion using the probability density function.
The probability density function is also useful to define the expectation ($E$) of a continuous random variable $X$:

$$E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx$$

At the end of this section, we should point out that the word "continuous" in "continuous random variable" is not used in its usual sense. Although a random variable is a function and the notion of continuous function is fairly well established in mathematics, "continuous" here is not used in that usual mathematical sense. The justification for the adjective "continuous" comes from the absolute continuity of the cumulative distribution function $F_X(x)$. Roughly speaking, a function

is called absolutely continuous if it can be written as the integral of its derivative. Thus, the absolute continuity of the cumulative distribution function $F_X(x)$ is the regular definition of an absolute continuous function and the "continuous" could be considered just as an abbreviation of "absolutely continuous".
A second reason that leads to the use of the adjective continuous, relies on the range of values that a random variable can take. In contrasting discrete random variable with continuous random variable it turns out that a discrete random variable takes values on a countable set (finite or infinite) of values, whereas a continuous random variable takes values on an uncountable set. The connection between uncountable and the continuum justifies the use of the word continuous.
Random variables are used to describe the outcomes and the related probabilities of some random experiments. Different experiments have different characteristics, so that they are described by different random variables. Let us consider a random variable that will play a key role in the following.

## 1.1 The Exponential random variable

**Definition 1.2.** A continuous random variable $X$ whose probability density function (Figure 1.2) is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if} \quad x \geq 0 \\ \\ 0 & \text{if} \quad x < 0 \end{cases}$$

is said to be an *Exponential* random variable with parameter $\lambda > 0$.

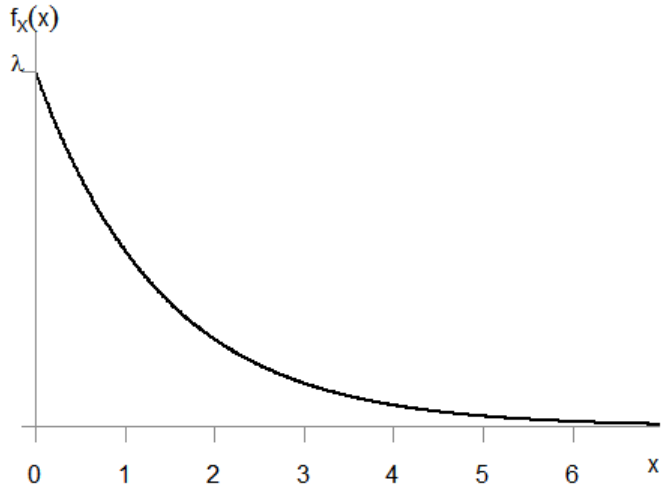



**Figure 1.2**

The cumulative distribution function of $X$ is

$$P(X \leq x) = \int_{-\infty}^{+\infty} f_X(x)dx = \int_0^x \lambda e^{-\lambda x}dx = 1 - e^{-\lambda x}$$

The Exponential distribution is widely used in applications since it models the lifetime of a component. For instance, it describes the amount of time that a light bulb works before burning itself out, or of a processor before breaking itself. In this context the parameter $\lambda$ can be interpreted as the speed at which an outcome occurs (e.g. the light bulb burns out, the engine breaks, the processor stops working,...) and it is called the *rate* of the random variable $X$. Formally an event is said to happen at rate $\lambda$, if the probability that it happens in a very short interval $(t + \Delta t)$ is approximately equal to $\lambda \Delta t$:

*Proof.*

$$\lim_{\Delta t\to 0}\frac{P(t<X<t+\Delta t)}{\Delta t}=\lim_{\Delta t\to 0}\frac{P(X<t+\Delta t)-P(X<t)}{\Delta t}=\lim_{\Delta t\to 0}\frac{1-e^{-(t+\Delta t)\lambda}-1+e^{-t\lambda}}{\Delta t}=$$

$$=\lim_{\Delta t\to 0}\frac{e^{-t\lambda}-e^{-(t+\Delta t)\lambda}}{\Delta t}=-\frac{\partial}{\partial t}e^{-t\lambda}\Big|_{t=0}$$

Then, we can write

$$P(t<X<t+\Delta t)=\lambda\Delta t+o(\Delta t)$$

where $o(\Delta t)$ is an infinitesimal of higher order than $\Delta t$ (i.e. a function of $\Delta t$ that goes to zero more quickly than $\Delta t$ does). □

The inverse $\frac{1}{\lambda}$ is the mean of the Exponential distribution and it is interpreted as the expected waiting time until an outcome occurs.
The Exponential random variable is also the only continuous random variable that has the memoryless property.

**Definition 1.3.** A random variable $X$ is *memoryless* if

$$P(X>s+t|X>t)=P(X>s)\quad\forall s,t>0$$

It is easy to prove the memoryless property for the Exponential random variable

*Proof.*

$$P(X>s+t|X>t)=\frac{P(X>t+s\cap X>t)}{P(X>t)}=\frac{P(X>t+s)}{P(X>t)}=\frac{1-P(X\leq t+s)}{1-P(X\leq t)}=$$

$$=\frac{1-1+e^{-\lambda(t+s)}}{1-1+e^{-\lambda t}}=e^{-\lambda s}=P(X>s)$$

□

Intuitively, if we think of $X$ as being the lifetime of a processor, the memoryless property states that the probability that the processor "lives" for at least $s+t$ hours given that it "has survived" $t$ hours is the same as the initial probability that it lives for at least $s$ hours. In other words, if the instrument is "alive" at time $t$ , then the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution; that is, the processor does not remember that it has already been in use for a time $t$.

# 2 Stochastic processes

**Definition 2.1.** A *stochastic process* $\{X(t),t\in T\}$ is a collection of random variables. Formally, a random process can be defined as a mapping that associates to each index $t\in T$ a random variable $X(t)$ defined on a probability space $(\Omega,P)$ and taking value in the set $\mathbb{R}$:

$$\forall t\in T\mapsto X(t):\Omega\to\mathbb{R}$$

Let $\mathcal{S}$ be the space in which the possible values of each $X(t)$ lie. $\mathcal{S}$ is called the *(*space state) of the process. The index $t$ is often interpreted as time and, as a consequence, $X(t)$ is the *state* of the process at time $t$. For example $X(t)$ can be the number of telephone calls that have arrived at a switchboard of a company at time $t$, the amount of money that a Gambler has after each cards game, and so on.
The realization of a stochastic process $\{X(t),t\in T\}$ is an assignment to each $t\in T$ of a possible value of X(t) (Figure 1.3).

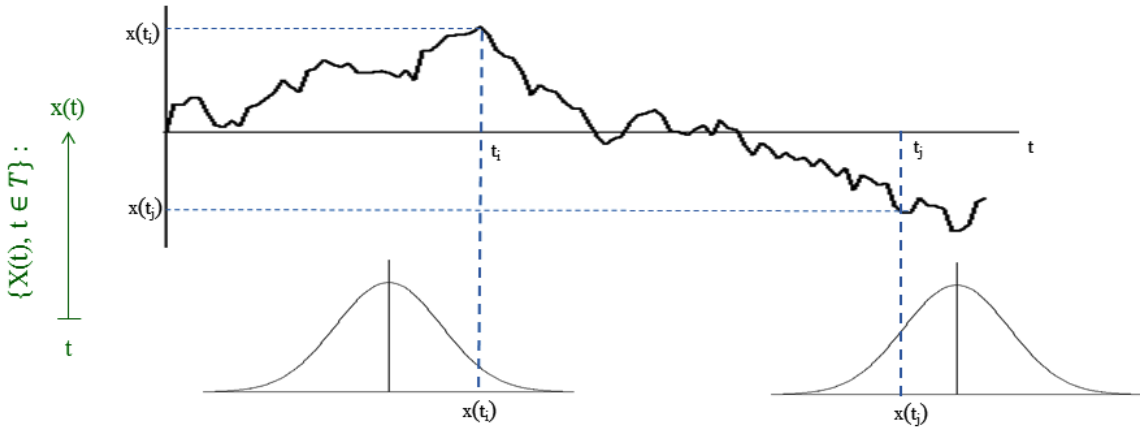



**Figure 1.3**

The main elements distinguishing stochastic process are related to the nature of the state space $\mathcal{S}$, the index set[1] $T$ and the dependence relations among the random variables $X(t)$ . In the following pages we will consider continuous-time Markov chains, i.e. finite state processes evolving in continuous time and which have the Markov property.

**Definition 2.2.** A stochastic process $\{X(t), t \in T\}$ has the *Markov property* if for any $t_i \in T$ the conditional distribution of the future, $\{X(t)|t > t_i\}$, given the present and the past $\{X(t)|t \leq t_i\}$, is a function only of the present $X(t_i)$. This implies that for any state $x \in \mathcal{S}$, and for any pair of time points $t_i < t_j$

$$P(X(t_j) = x(t_j)|X(t) = x(t) \text{ for all } t \leq t_i) = P(X(t_j) = x(t_j)|X(t_i) = x(t_i))$$

Roughly speaking, the Markov property states that the probability of any particular future behavior of the process, when its present state is known exactly, is not altered by additional knowledge concerning his past behavior. In other words the process retains no memory of where it has been in the past. Thus, only the current state of the process can influence where it goes next.
As we have seen a stochastic process is just a collection of random variables. Thus, we have to ask: what quantities characterize a random variable? The answer is obviously its distribution. However, here we are working with a lot of variables. Depending on the number of elements in the index set $T$, the stochastic process may have a finite or infinite number of components. In either case we will be concerned with the joint distribution of a finite sample taken from the process since Kolmogorov in the 1930s proved that if we can describe these finite dimensional joint distributions we completely characterized the stochastic process.
There is also a second way of describing the process. It consists in "building" the process specifying the "rules" that allow to go from a state to another. Since this way is more intuitive, we will refer to it in the following.

## 2.1 Continuous-time Markov chains

**Definition 2.3.** A *Continuous-time Markov chain* $\{X(t), t \geq 0\}$ is a finite state, continuous-time stochastic process having the Markovian property.

[1] If $\mathcal{S}$ is a countable set (e.g. $\mathcal{S} = \{0, 1, \ldots\}$), we refer to the process as "finite state" process. If $\mathcal{S}$ is an uncountable set (e.g. $\mathcal{S} = \mathbb{R}$), we refer to the process as "continuous state" process.
When T is a countable set, the stochastic process is said to be a *discrete-time* process. If T is an interval of the real line then the stochastic process is said to be a *continuous-time* process. For instance, the process $\{X(t), t = 0, 1, \ldots\}$ indexed by the nonnegative integers is a discrete-time process, while the process $\{X(t), t \geq 0\}$ indexed by the nonnegative real numbers is a continuous-time process.

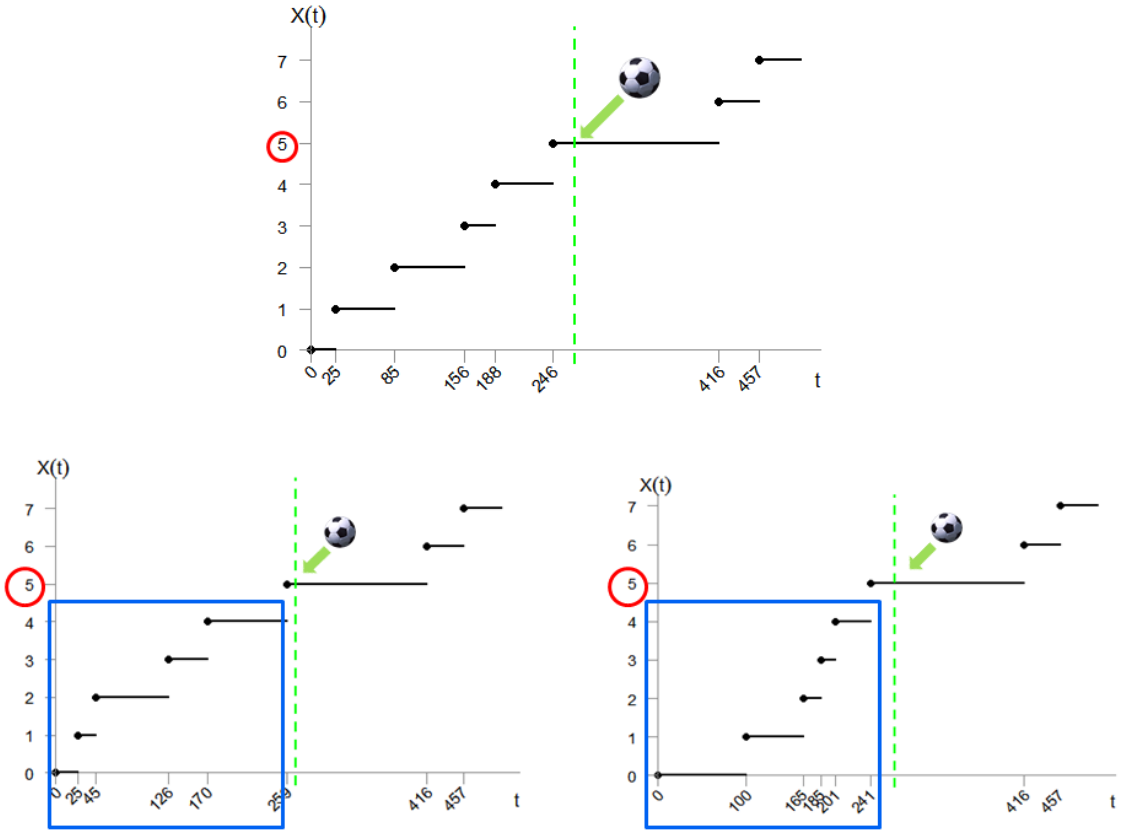



**Figure 1.4**

**Example 2.4.** Let $X(t)$ equals the number of goals that a given soccer player scores by time $t \in T$. Let assume that $T = [0,\infty)$ interpret the time that the soccer player played in official matches. Then $\{X(t), t \geq 0\}$ is a continuous- time Markov chain.

It is not difficult to understand why.

1. The state space of the process $\{X(t), t \geq 0\}$ is a finite set since at the end of his career the soccer player will have scored a certain number of goals $B$. The state space of the process is then described by the set
$$\mathcal{S} = \{0, 1, 2, \ldots, B\}$$

2. The time is continuous since we are taking into account the time played in official games and the player can score at any time.

3. The process $\{X(t), t \geq 0\}$ has the Markov property. To see this let us consider one of the possible realization of the process (Figure 1.4). If we want to determine the probability that the player will score the fifth goal at 269' (*future*), the only information we need is the number of goals that the player has scored till the 268' 59" (*present*, red circle) and it is not relevant to know the time at which each of the previous four goals was scored (*past*, blue box).

There are two basic ways we can describe continuous-time Markov chains. The first approach uses holding times and the jump chain. The second approach directly models the Markov chain in terms of the rate matrix (or infinitesimal generator).

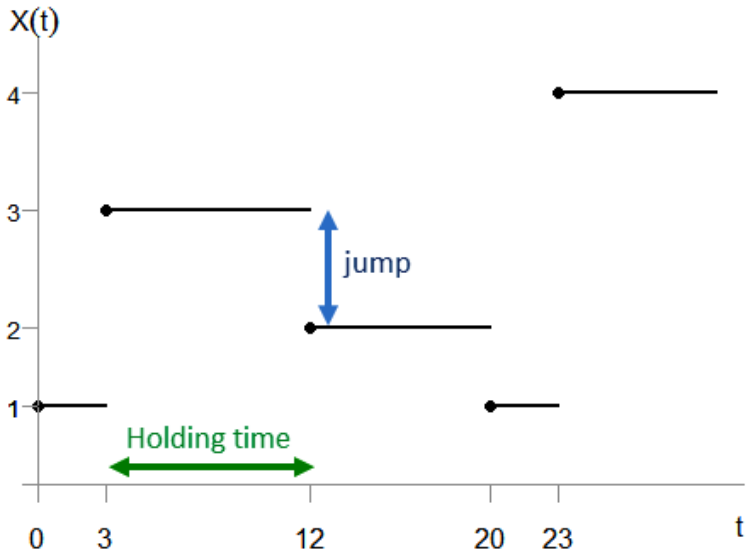



**Figure 1.5**

### Holding times and the jump chain

The idea behind the definition of holding times and the jump chain is that if we want to describe the process, we should characterize its realizations. In Figure 1.5 one realization of a continuous-time Markov chain is depicted. To describe it, we can list the times at which a change in the state happens and the corresponding state taken by the chain: the process starts in state 1 at time $t = 0$, then it jumps from 1 to 3 at time $t = 3$, from 3 to 2 at time $t = 12$, from 2 to 1 at time 20, and so on. In order to probabilistically describe this series of changes we should model the probability of staying in a state for a certain amount of time (e.g. the probability of remaining in state *1*) and the probability of jumping from a state to another (e.g. the probability of jumping from state *1* to state *3*).
Regarding the former, intuition suggests that for each state $i$, the amount of time we spend in that state (i.e. the *holding time*) is an exponentially[2] distributed random variable, with parameter $\lambda_i$. This can be motivated by the fact that the Markov property is a "forgetting" property, suggesting memorylessness in the distribution of the time a continuous-time Markov chain spends in any state. This amount of time can depend on the state, thus the parameter $\lambda$ is indexed by the state. We can observe that if $\lambda_i = 0$ we never leave the state $i$ once we enter.
The probability of jumping from a state to another is described by a *jump matrix* $P = (p_{ij} : i, j \in \mathcal{S})$ which satisfies the following properties:

1. $p_{ij} \geq 0 \quad \forall i, j \in S$
2. $\sum_{j \in \mathcal{S}} p_{ij} = 1 \quad \forall i \in S$

$P$ is a stochastic matrix (i.e., a matrix where each row sums to one) on the state space $\mathcal{S}$, which describes the next state we will go to when we leave a state $i$. In particular, $p_{ij}$ gives the probability of going to state $j$ when we make a jump out a state $i$. Formally:

$$p_{ij} = P(X(t') = j | X(t) = i, \text{given the opportunity to leave state } i),\ t' > t$$

[2]The exponential random variable models the time before something happen. Here the outcome is leaving the state "i".

The following matrix is an example of transition matrix for the process whose trajectory is represented by Figure 1.5

$$P = \begin{bmatrix} 0.1 & 0 & 0.6 & 0.3 \\ 0.8 & 0.1 & 0.1 & 0 \\ 0.05 & 0.5 & 0.05 & 0.4 \\ 0.6 & 0.1 & 0.15 & 0.15 \end{bmatrix}$$

The rows of $P$ sum to 1 and each cell gives the probability of going from a state to another. For instance, the probability of going from state 1 to state 3 is equal to 0.6, while it is not possible to go directly from state 1 to state 2 ($p_{12} = 0$), but we can reach 2 from 1 through 3. When each state of a chain is reachable (directly or indirectly) from any other state of the chain, the chain is said to be *irreducible*.

### The rate matrix

The other way to model a continuous-time Markov chain consists in specifying the rate matrix $Q$ of the process. A rate matrix $Q = (q_{ij} : i, j \in \mathcal{S})$ satisfies the following condition:

1. $0 < -q_{ii} < \infty \quad \forall i \in \mathcal{S}$
2. $q_{ij} > 0 \quad \forall i \neq j,\ i, j \in \mathcal{S}$
3. $\sum_{j \in \mathcal{S}} q_{ij} = 0 \quad \forall i \in \mathcal{S}$

The generic entry $q_{ij}$ of this matrix gives the rate of transition[3] from state $i$ to state $j$ and is strictly related to the waiting times and the jump matrix. In particular:

$$q_{ij} = \begin{cases} \lambda_i p_{ij} & \text{if } j \neq i \\ \\ -\lambda_i & \text{if } j = i \end{cases}$$

Thus, given the holding times and the jump matrix we can calculate the rate matrix. It can be proven that, given the rate matrix, we can compute the distribution of the holding times and the jump matrix. Thus, the rate matrix contains the same modeling information as the holding time/jump chain specification and the two methods of specifying a continuous-time Markov chain are equivalent.

[3] $Q$ is called the rate of the process since it can be proved that:

$$q_{ii} = \lim_{\Delta t \to 0} \frac{P(X(t+\Delta t) = i | X(t) = i)}{\Delta t} = -\lambda_i$$

$$q_{ij} = \lim_{\Delta t \to 0} \frac{P(X(t+\Delta t) = j | X(t) = i)}{\Delta t} = \lambda_i p_{ij}$$

# Chapter 2

# The Stochastic actor-oriented model

Networks are dynamic by nature. Ties may change over time: they can be established or they can be broken. These changes can be explained by the position of the actors within the network, as well as by their individual or dyadic characteristics. The stochastic actor-oriented model (SAOM) has the purpose to describe network evolution over time according to these potential explanatory sources.

## 1 Notations and data

Let us focus the attention on a set $\mathcal{N} = \{1, \cdots, n\}$ of actors over which a relation $\mathcal{R}$ is defined. We assume that the relations are non reflexive (i.e. $i \nrightarrow i \ \forall i \in \mathcal{N}$ ) and directed (i.e. $i \rightarrow j$ does not imply that $j \rightarrow i$ ) The two sets $\mathcal{N}$ and $\mathcal{R}$ define a network which is described by its adjacency matrix $x = (x_{ij}, i, j \in \mathcal{N})$, where

$$x_{ij} = \begin{cases} 1 & \text{if } i \rightarrow j \\ 0 & \text{otherwise} \end{cases}$$

We observed the network at $M \geq 2$ time points $t_1, \cdots, t_M$. The corresponding networks will be denoted by $x(t_1), \cdots, x(t_M)$. Furthermore, some actors'characteristics can be collected, such as gender, seniority rank, and so forth. These characteristics are called actors'attributes or covariates and will be denoted by $Z_1, \cdots, Z_H$ where $H$ is the number of collected attributes. The network observations $x(t_1), \cdots, x(t_M)$ and the covariates $Z_1, \cdots, Z_H$ constitutes the longitudinal panel data that we want to analyze to determine the leading forces which govern the network evolution. In more detail, the network evolution is the dependent variable and we want to describe it as a function of structural effects, explanatory random variables and explanatory dyadic variables.

## 2 Assumptions

To statistically define the model some assumptions concerning the evolving process should be formulated.

1. *Ties are state*. Network ties represent a state with a tendency to endure over time, rather than a brief event. For instance, if we consider ties determined by friendship, trust or co-operation, they can change but they endure over time. On the contrary, telephone calls or e-mail exchanges among a group of actors at any given time point are brief events, which start and finish in a short time.

2. *Distribution of the process*. The changing network is the outcome of a continuous - time Markov chain. The state space $\mathcal{X}$ of the model is the set of all possible adjacency matrices (digraphs) defined on the set of actors $\mathcal{N}$. Since each dyad can take value 0 or 1 and the number of possible ties in a network is $n(n-1)$ (we are considering directed relations!), the cardinality of $\mathcal{X}$ is $2^{n(n-1)}$ and $\mathcal{X}$ is a finite set.
   The continuous-time assumption means that the network evolves in continuous time, even if we observed it only at discrete time points $t_1, \cdots, t_M$. Formally we will say that there is a "latent process" (i.e. not observable) going on between network observation.
   The Markovian property suggests that for any point in time, the current state of the network determines probabilistically its further evolution, and there are no additional effects of the earlier past (Markov property). This hypothesis allows us to consider the dependencies between network ties as the result of processes where one tie is formed as a reaction to the existence of other ties.

3. *Opportunity to change*. At a given moment one probabilistically selected actor has the opportunity to change one of his outgoing tie or not to change.

4. *Absence of co-occurrence*. It is assumed that no more than one tie can change at any given moment $t$, i.e. only one actor has the opportunity to change one of his outgoing ties at $t$. This implies that tie changes are not coordinated, and depend on each other only sequentially, via the changing configuration of the whole network.

5. *Actor-oriented perspective*. We assume that the actors control their outgoing ties, i.e. they decide to change one of their outgoing ties according to their position in the network, their attributes and the characteristics of the other actors. This means that the actors do not change their ties at will but they want to maximize *a utility function* under the structural configuration of the network (i.e. actors are looking for the best rewarding in changing their outgoing ties given the current configuration of the network). This explains the "actor-based" nature of the model. Moreover it is assumed that actors have complete knowledge about the network and all the other actors, so that they can maximize the utility function. This maximization is based on immediate returns and not on long-run rewarding.

# 3 The formulation of the model

According to the previous assumptions, the evolution process can be decomposed into its smallest possible components, which are called *micro-steps*. At each micro-step one probabilistically selected actor might have the opportunity to change one of his outgoing ties, so that the utility function is maximized. Formally we can describe each micro-step as a pair of elements: the time at which one actor has the opportunity to change and the precise change which he made. The sequence of all micro-steps represents the *complete data* of the process.
The network evolution process can be then decomposed into two sub-processes: the *change opportunity process*, which describes how fast are the opportunities for change and the *change determination process*, which describes the precise tie that is changed when an actor has the opportunity to make a change. This two sub-processes should remind the definition of a continuous-time Markov chain according to the holding times and the jump chain (i.e. the waiting time until the next opportunity for a change made by an actor $i$ corresponds to holding time, while the probability of changing the link $x_{ij}$ given the opportunity for changing can be described by the transition matrix). The distribution of the waiting time and the transition matrix of the jump chain are modeled by the *rate function* and the *objective function* respectively.

## 3.1 The rate function

The waiting time between one opportunity of change for actor $i$ and another is modeled by an exponential distribution with parameter given by the *rate function*. This function describes the average frequency at which each actor has the opportunity to change and its simplest specification is obtained assuming that all actors have the same rate of change $\lambda$ between two consecutive observation moments. Then, the probability that an actor $i$ has the opportunity to make a change is equal to $1/n$.[1]

$$P(i \text{ has the opportunity of change}) = \frac{1}{n} \quad \forall i \in \mathcal{N}$$

The rate parameter $\lambda$ can assume nonnegative values and we can interpret it as the speed at which the opportunity to change occurs. In particular the higher its value is, the greater the number of changes between two observation moments is.
A more complex (and realistic) specification of the rate function takes into account that actors may change their ties at different frequencies, according to the position they have in the network and their covariates. For instance, we can imagine that "younger individuals might change their ties more frequently than older individuals, or that more central actors might change their ties more frequently than peripheral actors". In order to take into account this distinct aptitude we denote by $\lambda_i(\alpha, x)$ the rate of changes of the actor $i$, where $x$ is the current state of the network, and $\alpha$ is a vector of parameter. This notation expresses the dependence of the rate function from structural effects and actor covariates. The waiting times until the next opportunity for change by any actors follows an exponential distribution with parameter:

$$\lambda(\alpha, x) = \sum_{i=1}^{n} \lambda_i(\alpha, x)$$

Thus, given that an opportunity for change occurs, the probability $\pi(\alpha, x)$ that it is the actor $i$ who has the opportunity to change is given by:

$$P(i \text{ has the opportunity of change}) = \frac{\lambda_i(\alpha, x)}{\lambda(\alpha, x)}$$

## 3.2 The objective function

To define the transition matrix and the jump chain we should spend some words on the choice that an actor $i$ must face once he has the opportunity to change. Given the current state of the network $x$, $i$ can decide not to change anything or to change one of his outgoing ties, for instance the tie $x_{ij}$ directed to an actor $j$, into its opposite. In the following we will denote the network in which the tie from $i$ to $j$ is turned into its opposite by $x(i \rightsquigarrow j)$. Since we are considering simple digraphs, and a tie can assume values 1 or 0, according to the fact that it is present or it is absent, changing a tie into its opposite means that the tie variable changes from 1 to 0 or from 0 to 1. In the first case the tie is terminated, while in the second the tie is created. Thus, if a relation between $i$ and $j$ exists in the current state of the network ($x_{ij} = 1$) and $i$ decides to change it, the considered tie is deleted ($x_{ij} = 0$). Vice versa the tie is created.
This suggests that the set of admissible choices has cardinality equal to $n$: $n-1$ changes and 1 non-change. Consequently, the set of possible reachable states for the considered network, given the current state, contains $n$ elements: $n-1$ networks which are equal to that of the current state except for the value assumed by the changed tie $x_{ij}$ and 1 equal to the current state. It turns out that each actor can choose between a discrete finite set of alternatives, which are mutually exclusive

[1] Intuitively, since the actor has the same rate of changing, it is reasonable to assume that at each time point $t$ each actor has the same probability to be selected to change one of his outgoing tie.

(the selected actor can make only one change) and exhaustive (the actor can decide among all the other actors).
Just to clarify, we can consider Figure 2.1. Let us assume that actor 1 has the opportunity of changing. He can change his outgoing tie towards actor 2 (a)), or actor 3 (b)) or actor 4 (c)) or do nothing (d)). Then the corresponding Markov chain can jump from $x$ to $x(1 \rightsquigarrow 2)$, $x(1 \rightsquigarrow 3)$, $x(1 \rightsquigarrow 4)$ or stay in $x$, respectively. This are the only four possible states that the chain can reach. In fact, if we consider for example the two adjacency matrices depicted in situations e) it immediately appears that they are not admissible. On one hand the first matrix does not respect the "opportunity to change" assumption (actor 1 changed two of his outgoing ties) the second is against the "absence of co-occurrence assumption" (even actor 3 changed one of his outgoing ties). It follows that the transition matrix must assign probability 0 to the states represented by matrices differing in more than one element and positive probabilities to the states represented by situations from a) to d). At this point we need a rule on how to determine this nonzero probabilities.

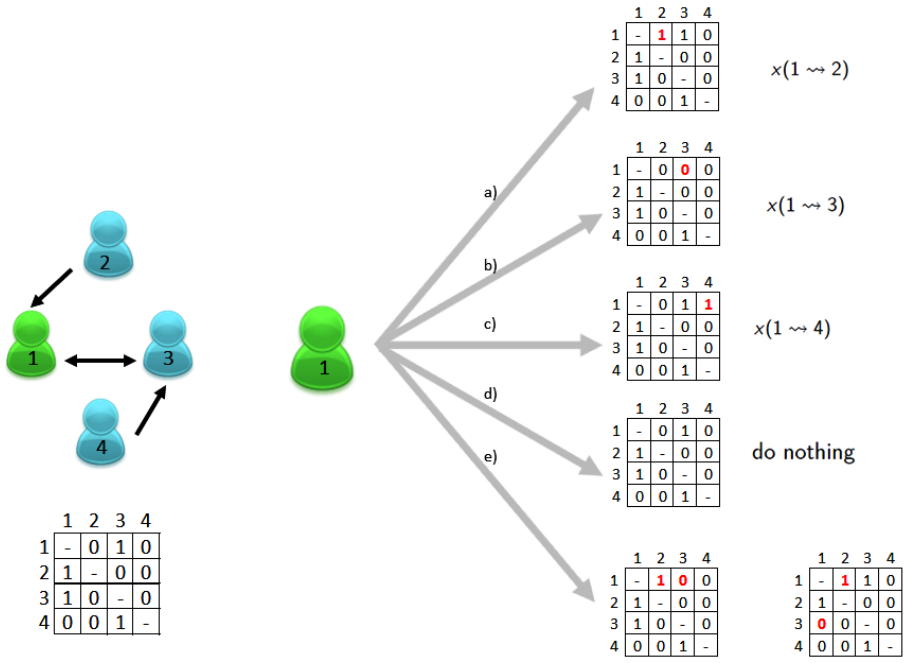



**Figure 2.1**

The idea is to use a random utility model, usually applied in situations where there are decision makers who face a choice between $n$ alternatives. Each decision maker choose the alternative that assure him the highest pay-off. The reward of an actor $i$ facing the choice $j$ is quantified by the utility function

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{2.1}$$

where $V_{ij}$ is the part of utility that a researcher can capture while $\varepsilon_{ij}$ is a random term. For a suitable choice of the distribution[2] of $\varepsilon_{ij}$ the probability that an actor $i$ faces the choice $j$ can be expressed by

$$p_{ij} = \frac{e^{V_{ij}}}{\sum_{j=1}^{n} e^{V_{ij}}} \tag{2.2}$$

[2]It is assumed that $\varepsilon_{ij}$ is distributed as a Type I Extreme (or Gumbel) distribution:

$$f_{\varepsilon}(x) = exp(-x - exp-x), \quad x \in \mathbb{R}$$

The choice of this distribution is mainly related to the mathematical tractability of the model.

In network context, the utility function corresponds to *objective function*. Informally speaking, the objective function expresses the degree of satisfaction of an actor towards the current state of the network and how likely it is for the actor to change the current state in a particular way.
Formally, we will denote the objective function by $f_i(\beta, x(i \rightsquigarrow j))$, where the subscript $i$ underlines that this is the utility function for the focal actor $i$, while the two quantities between parentheses express the idea that the objective function is a function of the state of the network obtained when $i$ changes his outgoing tie towards $j$, $x(i \rightsquigarrow j)$, and of statistical parameters $\beta$. The objective function is defined as the linear combination

$$f_i(\beta, x(i \rightsquigarrow j)) = \sum_{k=1}^{K} \beta_k s_{ik}(x(i \rightsquigarrow j)) + U_i(t, x, j) \tag{2.3}$$

where $\beta_k$ are statistical parameters, $s_{ik}(x(i \rightsquigarrow j))$ are the effects and $U_i(t, x, j)$ is a random utility term. The effects $s_{ik}(x(i \rightsquigarrow j))$ are relevant functions of the digraph which are supposed to play a key role in the network evolution. In other words they represent the leading forces of the underlying process that governs network changes from an observation moment to another. It is fundamental to specify that network effects are aspects of the network as perceived by the focal actor $i$.
The strength of each effect is represented by the corresponding parameter $\beta_k$, which should be estimated on the basis of the longitudinal network data observed. $\beta_k$ can assume any real value and can be interpreted as follows. If $\beta_k$ is equal to 0, it means that the corresponding effect plays no role in the network dynamics. If it assumes a positive value, then there is higher probability of moving into networks where the corresponding effect is higher. Vice versa if the parameter takes a negative value there is higher probability of moving into networks where the corresponding effect is lower.
The last term $U_i(t, x, j)$ of the objective function is the random term, distributed as a Gumbel distribution. Thus, according to equation (2.2), the probability that an actor $i$ changes his outgoing ties towards $j$ or leaves his outgoing tie variables unchanged is:

$$p_{ij} = \frac{exp\left(\sum_{k=1}^{K} \beta_k s_{ik}(x(i \rightsquigarrow j))\right)}{\sum_{h=1}^{n} exp\left(\sum_{k=1}^{K} \beta_k s_{ik}(x(i \rightsquigarrow h))\right)} \tag{2.4}$$

If $i = j$, then equation (2.4) represents the probability of not changing anything.

### The objective function effects

Several effects are proposed in order to specify the objective function. They can be determined endogenously or exogenously, according to the fact that they regard the structure of the network or actor attributes, i.e. covariates. A complete list of effects can be found in [1], but here only the basic ones are considered.
Each effect is characterized by a subscript $i$ to remind that these network aspects are determined from the point of view of the actor $i$, who has the opportunity to change. Furthermore, we use a more compact notation, denoting by $x'$ the resulting network instead of $x(i \rightsquigarrow j)$.
Let us start considering endogenous effects which depend on network structure.

#### *Endogenous effects*

In this paragraph a short description of the main endogenous effect is given.

- The *outdegree effect* corresponds to the number of outgoing ties of actor $i$ and is defined as:

$$s_{i\_out}(x') = \sum_{j} x'_{ij}$$

The corresponding parameter $\beta_{out}$ reflects the preference of activity for actor $i$. It takes negative values if the network is sparse (very low density).

- The *reciprocity effect* represents the number of mutual dyads in which the actor $i$ is involved:

$$s_{i\_rec}(x') = \sum_j x'_{ij} x'_{ji}$$

Denoting by $\beta_{\_rec}$ the associated parameter, $\beta_{rec}$ reflects the preference of reciprocal relations. It often assumes positive values, since a lot of social relations show a tendency towards reciprocation.

- The *transitive effect* can be modeled according to different effects. Each of them interprets a different situation. The more common structure is represented by the *transitive triads*, whose related effect counts the number of transitive patterns in which an actor $i$ is involved:

$$s_{i\_trans}(x') = \sum_{j,h} x'_{ij} x'_{ih} x'_{jh}$$

The corresponding parameter $\beta_{trans}$ reflects the preference of having relations to others who are related among themselves.

- The *three cycle-effect* is a sort of a generalized reciprocity which involves three actors $i$, $j$ and $h$. It expresses the idea that it is not important if $j$ reciprocates the tie from $i$, but it is important that he sends a tie to an actor $h$, who has an outgoing tie towards $i$.

$$s_{i\_cyc}(x') = \sum_{j,h} x'_{ij} x'_{jh} x'_{hi}$$

The corresponding parameter $\beta_{cyc}$ reflects the preference of having reciprocated in an indirect way through actor $h$.

- The *in-degree related popularity effect* is defined as the sum of the in-degrees of whom $i$ is related to:

$$s_{i\_pop}(x') = \sum_j x'_{ij} \sqrt{\sum_h x'_{hj}}$$

and the related parameter $\beta_{pop}$ expresses the idea that a very popular actor is more chosen. This means that popularity reinforce itself.

- The *out-degree related activity effect* represents the sum of the out-degrees of the others to whom $i$ is related:

$$s_{i\_act}(x') = \sum_j x'_{ij} \sqrt{\sum_h x'_{jh}}$$

The interpretation of these effects is similar to the previous one but it regards the out-degree. In particular, the related parameter $\beta_{\_act}$ expresses the idea that an actor is more active since the more popular. In this case popularity reinforces expansivity.

- The *indirect ties effect* is the number of actors $j$ to whom $i$ is indirectly tied to, through at least one intermediary:

$$s_{i\_ind}(x') = \#\{j : x'_{ij} = 0, \max_h(x'_{ih}, x'_{hj}) > 0\}$$

- The *balance effect* may also be called *structural equivalence with respect to outgoing ties*. It expresses a preference of actors to have ties to those other actors who have a similar set of outgoing ties as they have. It is defined by the similarity between the outgoing ties of actor $i$ and the outgoing ties of the other actors $j$ to whom $i$ is tied, to

$$s_{i\_bel}(x') = \sum_{j=1}^{n} x'_{ij} \sum_{h \neq i,j} \left( b_0 - \mid x_{ih} - x_{jh} \mid \right)$$

  where $b_0$ is a constant included to reduce the correlation between this effect and the density effect.

There are many other effects that are not considered here which regard very specific structural network properties. One can wonder how to choose among them. From a practical point of view, the choice of effects that should be included in the model is guided by theory. In fact, according to the kind of network in analysis, different hypotheses about the leading forces of network dynamics can be formulated. In order to test them one should include the related effects in the objective function.
For instance, if we consider "friendship", we know that transitivity ("the friend of my friend is also my friend") is a well-established network structure, and so we should include it in the model. Instead, if we are studying an advice network, we can imagine that people to whom a lot of other people ask advice are chosen the most. Maybe because if many people ask this actor for advice, it means that he is reliable and one can trust him. Thus, other people decide to ask to him. In this case, the in-degree related popularity effect plays a key role and should be tested.
Like all the other statistical models each effect is controlled for all the other effects included in the model. For this reason, one practical suggestion is to always include the density effect, since all the structural effects are related to the presence or absence of ties. Reciprocity is also fundamental in social relationships, and for this reason it should not be forgotten.

***Exogenous effects***
Let us now consider some effects related to actor attributes. Let $Z$ be a covariate, such as gender, ethnicity, seniority rank, etc. We denote by $z_i$ and $z_j$ the values or the categories assumed by the covariate on the actor $i$ and on the actor $j$, respectively. Some interesting effects are represented by:

- the *covariate-related popularity* that is defined by the sum of the covariate over all actors to whom $i$ has a relation:

$$s_{i\_cpop}(x) = \sum_j x_{ij} z_j$$

  The corresponding parameter $\beta_{cpop}$ reflects the aspiration of the actor to have relations with others who score high on $Z$.

- the *covariate-related activity* that is defined by $i$'s outdegree weighted by his covariate values

$$s_{i\_cact}(x) = \sum_j x_{ij} z_i$$

  The corresponding parameter $\beta_{cact}$ reflects the aspiration of actors with attribute $z_i$ to have relations.

- the *covariate-related similarity* that is the sum of measure of covariate similarity between $i$ and $j$:

$$s_{i\_csim}(x) = \sum_j x_{ij} \left( 1 - \frac{\left| z_i - z_j \right|}{R_Z} \right)$$

  where $R_Z$ is the range of $Z$.

Again, the choice of exogenous covariates to be included in the objective function should be determined according to hypotheses derived from theory. If we consider friendship network data gathered in adolescent groups, sociological theory suggests that girls trust girls and boys trust boys, showing evidence towards homophily with respect to this attribute. Thus, covariate-related similarity effect plays a key role and should be included in the objective function.
A positive parameter $\beta_k$, associated to popularity or activity effects will lead to association between the covariate and the receiver and sender tendency of an actor, respectively. In the same way a positive parameter related to the covariate- related similarity will lead to relations being formed particularly between actors who have similar values on the covariate.

### 3.3 An alternative formulation of the model

At this point we can specify the rate matrix $Q$ for the network evolution process, whose generic element is the rate of change from one network configuration to another.
The continuous time stochastic process is defined on the set $\mathcal{X}$ of all digraphs, or adjacency matrices. We denoted by $x$ the initial state of the network. At a certain time point an actor $i$ has the opportunity to change his outgoing ties towards an actor $j$. If he decides to change, the next state of the network is $x(i \rightsquigarrow j)$, where the link $x_{ij}$ between $i$ and $j$ is changed into its opposite. Thus, we can denote the rate of change from a state $x$ to $x' = x(i \rightsquigarrow j)$ using the following notation $q(x; x(i \rightsquigarrow j))$. Since we assume that at each time point only one actor may change, all transition rates for matrices belonging to $\mathcal{X}$ and differing in more than one element are equal to 0. Consequently, the rate of change between two states, denoted by the networks $x$ and $x'$, is given by:

$$q(x,x') = \begin{cases} \lambda_i(x) p_{ij}(x) & \text{if } x \text{ and } x' \text{ differ in the tie } (i,j) \\ -\lambda_i & \text{if } x = x' \\ 0 & \text{if } x \text{ and } x' \text{ differ in more than one tie} \end{cases}$$

If all the actors have the same opportunity to change then $\lambda_i(x) = \lambda$, and a simplest notation can be used.

### 3.4 Remark

1. It is assumed that the frequencies at which actors have the opportunity to make a change depends on time. To model the different speed of the change opportunities we should specify as many rate functions as the number of periods between two consecutive observation moments. Thus, if we observed a network at $M$ time points, we must specify $M-1$ rate functions. In other words, the parameter $\lambda$ is not constant over time, so there are $M-1$ parameters $\lambda_1, \cdots, \lambda_{M-1}$.

2. It is assumed that the preferences that drive actors action have the same intensities over time, i.e. the parameters $\beta_1, \cdots, \beta_k$ that are involved in the objective function are constant over time.

## 4 Simulating network evolution

This section presents a probabilistic algorithm to simulate network evolution for fixed parameter values. Simulating network evolution is meaningful, e.g., for the estimation of the parameter, for theoretical exploration of the model, for goodness of fit assessment, and for studying the sensitivity of the model to parameters.

The simulation is done exploiting the continuous-time Markov chain that defines the evolution process. Roughly speaking, it consists in reproducing a possible series of micro-steps between two observation moments $t_0$ and $t_1$ according to fixed parameter value and the network $x(t_0)$. Thus, it is convenient to construct the continuous-time Markov chain as the combination of its holding times and its jump process. We will assume that all actors have the same rate of change $\lambda$.
The simulation algorithm is as follows; $t$ indexes the time and $dt$ is the holding time between consecutive changes.

1. Set the time $t = 0$ and $x = x(t_0)$

2. Generate $dt$ according to an exponential distribution with parameter $n\lambda$

   The $n$ actors are acting independently given the current state of the network[3]. Assuming that each actor has the same individual change rate $\lambda$, the time until the next change by any actor has the exponential distribution with parameter $n\lambda$.

3. Select randomly the actor $i \in \mathcal{N}$, who makes the changes, with probability

$$P(i \text{ has the opportunity of change}) = \frac{1}{n} \quad \forall i \in \mathcal{N}$$

4. Select randomly the actor $j \in \mathcal{N}$, to whom $i$ changes his outgoing tie, with probability given by:

$$p_{ij} = \frac{exp\left(\sum_{k=1}^{K} \beta_k s_{ik}(x(i \rightsquigarrow j))\right)}{\sum_{h=1}^{n} exp\left(\sum_{k=1}^{K} \beta_k s_{ik}(x(i \rightsquigarrow h))\right)}$$

5. Set the time $t = t + dt$ and if $i \neq j$ then $x = x(i \rightsquigarrow j)$. If $i \neq j$ then $x = x$ .

6. Repeat step 2. to 5. until the stopping criterion is satisfied.

There are two different stopping rules for the simulations of the network evolution.
In the *unconditional* simulation, the simulations of the network evolution in each time period carry on until the predetermined time length (chosen as 1 for each time period between consecutive observation moments) has elapsed.
In the *conditional* simulation, in each period the simulations run on until a stopping criterion is reached that is calculated from the observed data. Conditioning is possible for each of the dependent variables, where *conditional* means conditional on the observed number of changes on this dependent variable. Conditioning on the network variable means running simulations until the number of different entries between the initially observed network of this period and the simulated network is equal to the number of entries in the adjacency matrix that differ between the initially and the finally observed networks of this period.
The algorithm for the unconditional estimation is summed up in Algorithm 1. The symbol $\sim$ should be interpreted as "is randomly generated from". For instance $dt \sim Exp(\lambda)$ means that the value taken by $dt$ is randomly drawn from an Exponential distribution of parameter $\lambda$.

[3]This is another way of interpreting the Markov property.

**Algorithm 1:** Network evolution

**Input**: $x(t_0)$, $\lambda$,$\beta$, $n$
**Output**: $x^{sim}(t_1)$
$t \leftarrow 0$
$x \leftarrow x(t_0)$
**while** $t < 1$ **do**
  $dt \sim Exp(n\lambda)$
  $i \sim Uniform(1,\dots,n)$
  $j \sim Multinomial(p_{i1},\dots,p_{in})$
  **if** $i \neq j$ **then**
    $x = x(i \rightsquigarrow j)$
  **else**
    $x = x$
  $t \leftarrow t + dt$
$x^{sim}(t_1) \leftarrow x$
**return** $x^{sim}(t_1)$

# 5 The parameter estimation

The formulation of the SAOM is based on the rate and on the objective functions and depends on $M-1+K$ statistical parameters. The aim of this section is to provide a description of the available estimation methods for completely specifying the model that interprets the network evolution process. Roughly speaking, estimating a parameter means determining its plausible value on the basis of the information deriving by the data collected by observation. In some situations finding an estimate is an easy and intuitive task, but this is not always the case. For this reason we need methods to properly assign a value to a parameter. In the following we will consider the Method of Moments and the Maximum Likelihood Estimation.

## 5.1 The Method of Moments

The logic of the Method of Moments (MoM) is quite straightforward. Let $X$ be a random variable with distribution $f_X(x;\theta)$, which depends on a p-dimensional vector of parameters $\theta = (\theta_1,\dots,\theta_p)$.

**Definition 5.1.** Let $X_1, X_2, \dots, X_n$ be a random sample from the probability distribution $f_X(x;\theta)$, where $f_X(x;\theta)$ can be a discrete probability mass function or a continuous probability density function. The $k$-th *population moment* (k = 1,2,...) is given by

$$E[X^k] = \sum_x x^k f_X(x;\theta) \qquad \text{(for the discrete case)}$$

$$E[X^k] = \int_{-\infty}^{+\infty} x^k f_X(x;\theta) \qquad \text{(for the continuous case)}$$

The corresponding $k$-th *sample moment* is

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

To estimate $\theta$, one can observe that the theoretical moments of a certain distribution usually depend on the statistical parameters $\theta = (\theta_1,\dots,\theta_p)$ which fully specify the distribution. Thus, the idea of the MoM is to estimate the parameter $\theta$ with the values that assure that the theoretical expected values are equal to their sample counterparts.

**Definition 5.2.** Let $X_1, \ldots, X_n$ a random sample from either a probability mass function or a probability density function $f_X(x;\theta)$ with $p$ unknown parameters. The method of moment estimators for $\theta = (\theta_1, \ldots, \theta_p)$ are found by equating the first $p$ population moments to the first $p$ sample moments:

$$E[X^1] = \mu_1$$

$$E[X^2] = \mu_2$$

$$\ldots$$

$$E[X^p] = \mu_p$$

and solving the resulting equations for the unknown parameters.

**Example 5.3.** The time to failure of an electronic module used in an automobile engine controller is tested at an elevated temperature to accelerate the failure mechanism. The time to failure is exponentially distributed with parameter $\lambda$. To estimate the rate parameter $\lambda$, eight units are randomly selected and tested, resulting in the following failure time (in hours):

$x_1 = 12.1 \quad x_2 = 5.7 \quad x_3 = 17.8 \quad x_4 = 16.5 \quad x_5 = 31.6 \quad x_6 = 7.7 \quad x_7 = 11.9 \quad x_8 = 22.7$

What is the estimate for $\lambda$ according to the observed data and the the MoM?

The first population moment of the Exponential random variable is

$$E[X] = \frac{1}{\lambda}$$

and the corresponding sample moment is

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

According to the MoM, the estimator for the parameter $\lambda$ is:

$$\frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} X_i}$$

and the corresponding estimate is

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{8}{126} = 0.063$$

The principle of the MoM can be easily generalized. Instead of using a set of powers of the random variable $X$ to estimate the unknown vector of parameters $\theta$, one can select a set of $p$ functions $s_k(X)$, $k = 1, \ldots, p$ ,and proceed in the same way. The population moment is computed as

$$E[s_k(X)] = \sum_x s_k(x) f_X(x) \qquad \text{(for the discrete case)}$$

$$E[s_k(X)] = \int_{-\infty}^{+\infty} s_k(x) f_X(x) dx \qquad \text{(for the continuous case)}$$

and the corresponding sample moment is

$$\gamma_k = \frac{1}{n}\sum_{i=1}^{n} s_k(X_i)$$

The functions $s_k(X)$ are usually called statistics[4]. Of course, they cannot be chosen at random, but they must be sensitive to the parameter $\theta$ in the sense that higher values of $\theta$ lead to higher values of $s(X)$.

There is one distinction that must be made clear, the difference between an estimate and an estimator. An *estimator* is a function of the sample, while an *estimate* is the realized value of an estimator (i.e. a number) that is obtained when a sample is actually taken. Notationally, when a sample is taken, an estimator is a function of the random variables $X_1, X_2, \ldots, X_n$, while an estimate is a function of the realized values $x_1, x_2, \ldots, x_n$. For instance, in Example 5.3 $\frac{n}{\sum_{i=1}^{n} X_i}$ is the estimator, while $\frac{n}{\sum_{i=1}^{n} x_i}$ is the estimate.
It is easy to imagine that the estimate of the parameter can vary according to the selected sample. Looking again at Example 5.3, we can assume to randomly select and test other eight units, resulting in the following failure time (in hours):

$x_1 = 9.5 \quad x_2 = 7.2 \quad x_3 = 13.4 \quad x_4 = 10.2 \quad x_5 = 15.0 \quad x_6 = 16.3 \quad x_7 = 13.9 \quad x_8 = 34.5$

The new estimate for the parameter $\lambda$ is now

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{8}{120} = 0.067$$

This value is close to that obtained in Example 5.3 but it is not the same. For this reason, we usually associate to an estimator its standard error, which is a measure of the variation of the estimates with respect to the different selected samples. We will consider the role and the interpretation of the standard error later on, when we will provide the parameter interpretation.

### The MoM for estimating the parameter of the SAOM

Let

$$\theta = (\lambda_1, \cdots, \lambda_{M-1}, \beta_1, \cdots, \beta_K)$$

be the $M-1+K$-dimensional parameter vector of the SAO model. From the previous paragraph it follows that, if we want to estimate $\theta$, we must find $M-1+K$ statistics, set the theoretical expected value of each statistic equal to its sample counterpart, and solve the resulting system of equations with respect to $\theta$.
Since each parameter is related to a specific effect, the logic is to determine the statistics as functions of the corresponding effects. For simplicity, let us assume to have observed a network at two time points $t_0$ and $t_1$. Estimation is done conditional on the first observation $x(t_0)$. This has the advantage that no model assumptions need to be invoked concerning the probability distribution that may have led to the first observed network $x(t_0)$, and the estimated parameters refer exclusively to the dynamics of the network.

[4]Formally a statistic is a real-value function of the random sample $X_1, X_2, \ldots, X_n$ .

- The rate parameter $\lambda$ describes the frequencies at which changes happen. Thus, a suitable statistic for the MoM is the total amount of changes between $t_0$ and $t_1$ and can be computed as follows:

$$s_\lambda(X(t_1), X(t_0)|X(t_0) = x(t_0)) = \sum_{\substack{i,j=1 \\ 1 \neq j}}^{n} \left| X_{ij}(t_1) - X_{ij}(t_0) \right| \tag{2.5}$$

- The rate parameter $\beta_k$ quantifies the role played by each effect in the network evolution. As already mentioned, if $\beta_k$ is positive, there is higher probability of moving into networks where the corresponding effect is higher (and vice versa if $\beta_k$ is negative). Thus, a suitable statistic for the parameter $\beta_k$ of the objective function is the number of the corresponding effect $s_{ik}(X(t_1))$:

$$s_k(X(t_1)|X(t_0) = x(t_0)) = \sum_{i=1}^{n} s_{ik}(X(t_1)) \tag{2.6}$$

  This statistic will be sensitive to $\beta_k$ because a high value of $\beta_k$ will lead to a high value of $S_k$

Having determined the statistics for two time points, it is easy to derive the corresponding statistics for more than two network observations. In order to do this, one should keep in mind that the rate parameter $\lambda$ is assumed to be constant within each time period, so that one should specify $M-1$ statistics $s_{\lambda_m}$ for the $M-1$ rate parameters $\lambda_m$. Then, we can generalize Equation (2.5) in the following way:

$$s_{\lambda_m}(X(t_m), X(t_{m-1})|X(t_{m-1}) = x(t_{m-1})) = \sum_{\substack{i,j=1 \\ 1 \neq j}}^{n} \left| X_{ij}(t_m) - X_{ij}(t_{m-1}) \right|, \quad m = 1, \ldots, M-1 \tag{2.7}$$

Regarding the parameters $\beta_k$ of the objective function, it is assumed that they are constant over the whole observation period. Thus, a general statistic for $\beta_k$ is the sum over all the observation time points of the corresponding configurations[5] $s_k(X(t_m))$:

$$\sum_{m=1}^{M-1} s_{mk}\left(X(t_m)|X(t_{m-1}) = x(t_{m-1})\right) = \sum_{m=1}^{M-1} s_{mk}(X(t_m), X(t_{m-1})) \tag{2.8}$$

Consequently, the MoM estimator for $\theta$ is defined as the solution of the system of equations:

$$\begin{cases} E_\theta\left[s_{\lambda_m}(X(t_m), X(t_{m+1})|X(t_m) = x(t_m))\right] = s_{\lambda_m}(x(t_1), x(t_0)) & m = 1, \cdots, M-1 \\ \\ E_\theta\left[\sum_{m=1}^{M-1} s_{mk}\left(X(t_{m+1})|X(t_m) = x(t_m)\right)\right] = \sum_{m=1}^{M-1} s_{mk}(x(t_{m+1}), x(t_m)) & k = 1, \cdots, K \end{cases} \tag{2.9}$$

where a small $x$ denotes the observed networks, and thus the second term of in (2.9) are the observed values of the statistics.
In the following, we will use a more compact notation. Let S be the $(M-1+K)$-dimensional vector of statistics involved in the estimation process and $s$ be its sample counterpart (i.e. the $(M-1+K)$-dimensional vector of the observed values of the statistics). Then, the system of equations (2.9) can be written in the following way:

$$E_\theta[S] = s$$

or equivalently as

$$E_\theta[S - s] = 0 \tag{2.10}$$

[5] A note on notation: the two indexes in (2.8) refer respectively to the observation period ($m$) and the effects ($k$), where $s_k(X(t_m))$ is defined by (2.6). The definition in (2.6) allows the use of a straightforward notation.

At this point, the problem is how to solve the system (2.10) with respect to the parameter $\theta$. Analytical and usual numerical procedures cannot be applied, since the theoretical expected values of the statistics cannot be calculated explicitly. However, it is rather straightforward to simulate random digraphs with the desired distributions. Therefore, stochastic approximation methods can be used to approximate the estimates.

**Definition 5.4.** *Stochastic approximation methods* are a family of iterative simulations-based (i.e. stochastic) algorithms that attempt to find zeros or extrema of functions which cannot be computed in an analytical way.

In the context of the parameter estimation of the SAOM, the stochastic approximation method consists in iteratively approximate the moment equations for fixed values of the parameter $\theta$ and updating the value of $\theta$ step by step, according to the "distance" between the approximated expected values and the corresponding observed values.

***The Monte Carlo Method***

The expected values of the statistics are approximated via Monte Carlo methods.

**Definition 5.5.** Let $X$ be a random variable with distribution function $f_x(x)$ and let us assume to be interested in estimating the expected value:

$$E[s(X)] = \int_{-\infty}^{+\infty} s(x) f_X(x) dx$$

where $s$ is a real valued function. The Monte Carlo methods requires generating a sample $(X_1, \cdots, X_q)$ from the distribution function $f_X(x)$ and to approximating the expected value with the empirical average, i.e.:

$$\overline{S} = \frac{1}{q} \sum_{l=1}^{q} s(x_l)$$

It can be proved that $\overline{S}$ converge to the true value of $E[s(X)]$ as $q \to \infty$.

Therefore, we can approximate the expected value of the vector of statistics $S$ simulating the network evolution $q$ time,s given a fixed value for $\theta$, and computing the mean of the values assumed by the statistics during each simulation.
For instance, the Monte Carlo approximation of the expected value for the rate parameter $\beta_{out}$ can be obtained in the following way:

1. Given $x(t_0)$ and $\theta$, simulate the sequence of the observed networks at time $t_1, \ldots, t_M$ $q$ times. Denote these sequences by

$$x^{(1)}(t_1),\ x^{(1)}(t_2),\ \ldots,\ x^{(1)}(t_M)$$

$$\ldots$$

$$x^{(q)}(t_1),\ x^{(q)}(t_2),\ \ldots,\ x^{(q)}(t_M)$$

2. For each sequence compute the value assumed by the statistics $S_{out}^{(i)} = \sum_{m=1}^{M-1} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}^{(l)}(t_m)$:

3. Approximate the expected value by

$$\overline{S}_{out} = \frac{1}{q} \sum_{l=1}^{q} S_{out}^{(l)}$$

***The Robbins-Monro step***

The iterative step for updating the value of $\theta$ is based on the Robbins-Monro step.

**Definition 5.6.** Let $\theta \in \Theta^p$ be a $p$-dimensional vector of parameters and $g(\theta;x) \in \mathbb{R}^p$ some vector-valued function which depends on $\theta$. The focus is to find the root $\widehat{\theta} \in \Theta^p$ of the equation

$$g(\theta;x) = 0 \tag{2.11}$$

Let $\theta_1$ be an initial guess of $\theta$. Then the Robbins-Monro algorithm to solve equation (2.11) is based on a sequence $\theta_i$ generated by the recursion:

$$\theta_{i+1} = \theta_i - a_i g(\theta;x)$$

where the gain $a_i$ is a sequence of positive numbers such that:

$$\lim_{(i\to\infty)} a_i = 0 \qquad \sum_{i=1}^{\infty} a_i = \infty \qquad \sum_{i=1}^{\infty} a_i^2 < \infty$$

**Remarks**

1. The conditions related to the sequence of gain provide a careful balance in having the gain $a_i$ decay neither too fast nor too slow. In particular, the gain should approach zero sufficiently fast $\left(\lim_{(i\to\infty)} a_i = 0,\ \sum_{i=1}^{\infty} a_i^2 < \infty\right)$, but should also approach it at a sufficiently slow rate $\left(\sum_{i=1}^{\infty} a_i = \infty\right)$ to avoid premature (false) convergence of the algorithm.

2. The sequence $\{\theta_i\}$ converges (almost sure) to the true value of $\theta$ as $i \to \infty$. It can be proved that optimal convergence of the Robbins-Monro algorithm can be achieved including in the iterative step a positive diagonal matrix $D$:

   $$\theta_{i+1} = \theta_i - a_i D^{-1} g(\theta_i;x)$$

   and estimating the value of the parameter $\theta$ not by the last value $\theta_i$, but by the average of the consecutively generated $\theta_i$ values[6].
   Usually $D$ is the diagonal matrix of the first order derivative of $g(\theta;x)$. Different methods can be used to estimate the diagonal elements of $D$. One possibility is the finite difference method, which estimates the $j-th$ diagonal element by

   $$\frac{\partial g(\theta;x)}{\partial \theta_j} = \lim_{\varepsilon_j \to 0} \frac{g(\theta + e_j\varepsilon_j;x) - g(\theta;x)}{\varepsilon_j}$$

   where $e_j$ is the $j$-th unit vector in $p$-dimension and $\varepsilon_j$ is a suitable constant.

We can use the Robbins-Monro step to find the root of the system of equations represented by (2.10). In the estimation algorithm $D$ denotes the diagonal matrix of the first order derivative matrix of $S$ with respect to $\theta$:

$$D = \frac{\partial}{\partial \theta} E_\theta[S|X(t_0) = x(t_0)]$$

[6]The choice of averaging over the generated $\theta_i$ values relies on a work by Ruppert [3]. In his technical report, Ruppert proved that the optimal convergence rate of the algorithm can be achieved when $D$ is a diagonal matrix and the sequence $a_i$ is a sequence of positive numbers converging to zero at the rate $i^{-c}$, $0.5 < c < 1$. To obtain this optimal convergence rate, the solution to (2.10) must be estimated not by the last value $\theta_i$, but by the average of the consecutively generated $\theta_i$ values

and the general step of the Robbins-Monro algorithm by:

$$\widehat{\theta}_{i+1} = \widehat{\theta}_i - a_i D^{-1}(\overline{S}_i - s)$$

At the end of the estimation procedure, the standard error associated to the estimates is computed. The standard error will play a key role in the interpretation of the results provided by the model. The algorithm is divided into three phases.

1. ***Phase 1***. A small number $q_1 = 7 + 3 * (M - 1 + K)$ of steps are made to estimate $D$, using the Monte Carlo method, and to provide a first update to the parameter $\theta$.
   Let $e_j$ be the $j$-th unit vector in $(M - 1 + K)$ dimensions (thus, $j = 1, \ldots, M - 1 + K$) and $\varepsilon$ a constant vector in $(M - 1 + K)$ dimensions whose components take values between 0.1 and 1 (but often are equal to 1).
   In step $i$ $(i = 1, 2, \ldots, q_1)$, given the network at time $t_0$, the network evolution is simulated from the initial value $\theta_0$ and from the values $\theta_j = \theta_0 + \varepsilon_j e_j$ and the $(M-1+K)$-dimensional vectors of statistics are computed according to the simulated networks. We denote by $S_{i0}$ and $S_{ij}$ such vectors, where the first index ($i$) refers to the step and the second index (0 or $j$) to the value of $\theta$ used for the simulation. Then, the difference quotients $d_{ij} = \varepsilon_j^{-1}(S_{ij} - S_{i0})$ are computed. $d_{ij}$ is the approximation of the derivative of the statistics $S_j$ at step $i$ (i.e. the approximation of the $j$-th element on the diagonal of $D$) using the finite difference method. At the end of this phase the expected value $\overline{S}$ and the $j - th$ diagonal element $\widehat{d}_j$ of $D$ are estimated using the following formulas:

$$\overline{S} = \frac{1}{q_1} \sum_{i=1}^{q_1} S_{i0} \qquad \widehat{d}_j = \frac{1}{q_1} \sum_{i=1}^{q_1} d_{ij}$$

   Finally a new value of $\theta$ is estimated via the Robbins-Monro step with $a_i = 1$:

$$\widehat{\theta}_{q_1} = \theta_0 - \widehat{D}^{-1}(\overline{S} - s)$$

   Phase 1 is summed up by Algorithm 2. The symbol $\sim$ means "generated from".

**Algorithm 2:** Robbins-Monro algorithm - Phase 1

**Input**: $\theta_0$, $s$, $q_1$,$\varepsilon_j$
**Output**: $\widehat{\theta}_{q_1}$, $\widehat{D}$

$i \leftarrow 0$
$d \leftarrow 0$
$S_0 \leftarrow 0$
**while** $i < n_1$ **do**
  $i \leftarrow i + 1$
  $S_{i0} \sim \theta_0$
  $S_0 \leftarrow S_0 + S_{i0}$
  **for** *j=1, ... (M+K-1)* **do**
    $S_{ij} \sim \theta_1 + \varepsilon_j * e_j$
    $d_{ij} \leftarrow \varepsilon_j^{-1}(S_{ij} - S_{i0})$
    $d \leftarrow d + d_{ij} e_j$
$\overline{S} \leftarrow \frac{1}{q_1} S_0$
$\widehat{d} \leftarrow \frac{1}{q_1} d$
$\widehat{D} \leftarrow diag(\widehat{d})$
$\widehat{\theta}_{q_1} \leftarrow \theta_0 - \widehat{D}^{-1}(\overline{S} - s)$
**return** $\widehat{\theta}_{q_1}$, $\widehat{D}$

2. ***Phase 2***. This is the main phase, consisting of $c$ sub-phases. The number of iteration per sub-phase is determined by a stopping rule but bounded for sub-phase $h$ between $q_h^- = 2^{4((M-1+K)-1)/3}(7+(M-1+K))$ and $q_h^+ = q_h^- + 200$. In each sub-phase $a_i$ is constant and equal to $a_h$. The only difference between the sub-phases is $a_i$.
Each sub-phase is ended after less than $q_h^+$ steps, as soon as the number of steps in this sub-phase exceeds $q_h^-$ while, for each coordinate $S_{ih}$, the sum within this sub-phase of successive products $(S_{ih}-s)(S_{(i-1)h}-s)$ is negative. It can be proved that as long as $(S_{ih}-s)(S_{(i-1)h}-s)$ is positive, it must be assumed that the sequence $\theta_i$ is still drifting toward the limit point rather then wandering around the limiting point, so that is not desirable decrease the gain $a_i$. If the upper bound $q_h^+$ is reached, then the sub-phase is terminated anyway. In each iteration step within each sub-phase $h$, $S_{ih}$ is generated according to the current parameter value $\widehat{\theta}_i$ . After each step, this value is updated according to the formula

$$\widehat{\theta}_{i+1} = \widehat{\theta}_i - a_h \widehat{D}^{-1}(\overline{S}_i - s)$$

At the end of each sub-phase, the average of $\widehat{\theta}_i$ over this sub-phase is used as the initial value $\widehat{\theta}_h$ of the parameter for the next sub-phase. The value of $a_h$ is divided by 2 when a new sub-phase is entered.
The average of $\widehat{\theta}_i$ over the last sub-phase is the eventual estimate for $\theta$.
Phase 2 is summed up by Algorithm 3.

**Algorithm 3:** Robbins-Monro algorithm - Phase2

**Input**: $\widehat{\theta}_1 = \widehat{\theta}_{q_1}$, $\widehat{D}$, $s$
**Output**: $\widehat{\theta}$

$\theta \leftarrow 0$
**for** $h = 1$ *to* $c$ **do**
  $i \leftarrow 0$
  **if** $i >= q_h^+$ *OR* $(i > q_h^-$ *AND* $(S_{ih}-s)(S_{(i-1)h}-s) < 0)$ **then**
    $i \leftarrow i+1$
    $S_i \sim \widehat{\theta}_i$
    $\widehat{\theta}_{i+1} \leftarrow \widehat{\theta}_i - a_h \widehat{D}^{-1}(\overline{S}_i - s)$
    $\theta \leftarrow \widehat{\theta}_{i+1} + \theta$
  $\widehat{\theta}_h \leftarrow \frac{1}{i-1}\theta$
  $a_{h+1} \leftarrow a_h/2$;
$\widehat{\theta} \leftarrow \theta_c$
**return** $\widehat{\theta}$

3. ***Phase 3***. A number $q_3 = 500$ is used to estimate the variance-covariance matrix $\Sigma_\theta$ of the estimator of $\theta$.The standard error is the square root of the diagonal elements of the variance-covariance matrix $\Sigma_\theta$. We will denote by $\widehat{\Sigma}_\theta$ the estimate of $\Sigma_\theta$. It can be proved that the estimate of $\Sigma_\theta$ is given by

$$\widehat{\Sigma}_\theta = \widehat{D}^{-1}\left[\frac{1}{q_3}(S_i - \overline{S})'(S_i - \overline{S})\right]\widehat{D}^{-1}$$

and the squared root of the diagonal matrix The square root of the diagonal value of $\Sigma_\theta$ are the standard errors associate to the estimates obtained through the Phase 2. The value $\widehat{\theta}$ is left unchanged during Phase 3 and it is equal to the value obtained after the last subphase of Phase 2. It is used to provide the new estimate $\widehat{D}$ of $D$ and of the expected value of the

statistics $\overline{S}$. Therefore, Phase 3 is similar to Phase 1.
Phase 3 is summed up by Algorithm 4.

**Algorithm 4:** Robbins-Monro algorithm - Phase 3

**Input**: $\widehat{\theta}$, $s$, $n_3$,$\varepsilon_j$
**Output**: $\widehat{\Sigma}_\theta$

$i \leftarrow 0$
$d \leftarrow 0$
$S_0 \leftarrow 0$
**while** $i < n_3$ **do**
  $i \leftarrow i+1$
  $S_{i0} \sim \widehat{\theta}$
  $S_0 \leftarrow S_0 + S_{i0}$
  **for** *j=1, ... M+K-1* **do**
    $S_{ij} \sim \widehat{\theta} + \varepsilon_j * e_j$
    $d_{ij} \leftarrow \varepsilon_j^{-1}(S_{ij} - S_{i0})$
    $d \leftarrow d + d_{ij} e_j$
$\overline{S} = \frac{1}{q_3} S_0$
$\widehat{d} = \frac{1}{q_3} d$
$\widehat{D} = diag(D)$
$\widehat{\Sigma}_\theta = \widehat{D}^{-1} \left[ \frac{1}{q_3} (S_{i0} - \overline{S})(S_{i0} - \overline{S}) \right] \widehat{D}^{-1}$
**return** $\widehat{\Sigma}_\theta$

## 5.2 The Maximum likelihood estimation

The Maximum likelihood (ML) estimation for the parameter of the SAOM is quite complex and requires advanced statistical concepts. For this reason, the aim of this paragraph is to provide only an intuitive idea of the algorithm.
The model assumptions allow us to decompose the process in a series of micro-steps. Each micro-step can be described by the sequence $\{(T_r, i_r, j_r), r = 1, \ldots, R\}$, where $T_r$ is the time point for an opportunity for change, $i_r$ denotes the actor who has the opportunity to change and $j_r$ is the actor towards whom the tie is changed. We denote by $R$ the total number of micro-steps between $t_0$ and $t_1$ and we assume that the time point $T_r$ are ordered increasingly, i.e. $t_0 = T_0 < T_1 < \ldots < T_R < t_1$.
If we knew the complete data of the network evolution process, then we could write down the likelihood of the network evolution process and compute the maximum likelihood estimation for $\theta$.

**Definition 5.7.** Given the sequence of micro-steps $\{(T_r, i_r, j_r), r = 0, \ldots, R\}$, the likelihood function of the network evolution process is defined by:

$$L(\theta) = \prod_{r=1}^{R} P((T_r, i_r, j_r))$$

Then, the estimate for $\theta$ is the vector of values $\widehat{\theta}$ such that:

$$\widehat{\theta} = arg \max_{\theta \in \Theta} L(\theta)$$

or equivalently, the vector of values $\widehat{\theta}$ such that:

$$U(\theta) = \frac{\partial}{\partial \theta} log(L(\theta)) = 0 \tag{2.12}$$

where $\frac{\partial}{\partial \theta} log(L(\theta))$ is the score function.

Since we cannot observe the complete data, the likelihood of the observed data $(x(t_1), \ldots, x(t_M))$ conditional on $x(t_0)$ cannot generally be computed. Again, the parameter estimation requires finding the root of a system of equations represented by (2.12) in which the functions that are set equal to 0 cannot be computed analytically. Thus, a stochastic approximation method must be applied. In order to approximate the score-function, the general data augmentation principle is employed. The idea is to augment the observed data so that an easily computable likelihood is obtained. Since the data augmentation can be done separately for each time period $(t_{m-1}, t_m)$ it is not restrictive to describe it only for two observations $x(t_0)$ and $x(t_1)$.

**Definition 5.8.** The *augmented data* (or *sample path*) consists of $R$ and the sequence $(i_1, j_1), \ldots, (i_R, j_R)$, without the time points $T_r$, that specifies the sequence of tie changes that brings the network from $x(t_0)$ to $x(t_1)$ . Formally we will denote the augmented data by

$$\underline{v} = \{(i_1, j_1), \ldots, (i_R, j_R)\}$$

and the set of all sample paths connecting $x(t_0)$ and $x(t_1)$ is denoted by $\mathcal{V}$.

The probability of the augmented data $\underline{v}$ conditional on $X(t_0) = x(t_0)$ and $X(t_1) = x(t_1)$ is proportional to

$$P(\underline{v}|x(t_0), x(t_1)) \propto \frac{(n\lambda)^R}{R!} e^{-n\lambda} \prod_{r=1}^{R} \frac{1}{\lambda} p_{i_r j_r}(\beta, x(T_{r-1})) \tag{2.13}$$

and can be used to approximate the likelihood of the network evolution process (and then the score function). Thus, we need a method to properly sample the augmented data from the distribution in (2.13). This is done through a Markov chain simulation defined on the state space $\mathcal{V}$. Roughly speaking the Markov chain defines the transition probabilities from a sample path $\underline{v} \in \mathcal{V}$ to another, so that the long-run distribution of the Markov-chain is equal to the distribution in (2.13). The transition probability from a sample path to another is defined by the Metropolis-Hastings algorithm.

**Definition 5.9.** The *Metropolis-Hastings algorithm* is defined by the following transition probabilities:

1. Given $\underline{v}_i = \underline{v}$, generate $\widetilde{\underline{v}}$ form the proposal distribution $u(\widetilde{\underline{v}}|\underline{v}_i)$
2. Take

$$\underline{v}_{i+1} = \begin{cases} \widetilde{\underline{v}} & \text{with probability} \quad \rho(\widetilde{\underline{v}}, \underline{v}) \\ \underline{v} & \text{with probability} \quad 1 - \rho(\widetilde{\underline{v}}, \underline{v}) \end{cases}$$

where

$$\rho(\widetilde{\underline{v}}, \underline{v}) = \min\left\{ \frac{P(\widetilde{\underline{v}}) u(\underline{v}|\widetilde{\underline{v}})}{P(\underline{v}) u(\widetilde{\underline{v}}|\underline{v})}, 1 \right\}$$

Then, the transition probabilities of the chain generate by the Metropolis-Hastings algorithm are given by $\rho(\widetilde{\underline{v}}, \underline{v}) u(\widetilde{\underline{v}}|\underline{v})$.

The proposal distribution $u(\widetilde{\underline{v}}|\underline{v})$ is defined so that the admitted changes from a state to another are the following:

- *Pairwise deletions*: one pair of indices $r_1$ and $r_2$ such that $(i_{r_1}, j_{r_1}) = (i_{r_2}, j_{r_2})$ is selected and the corresponding pairs $(i_{r_1}, j_{r_1})$ and $(i_{r_2}, j_{r_2})$ are deleted from $\underline{v}$

- *Pairwise insertions*: one pair of $(i, j) \in \mathcal{N}^2$ and two indices $r_1$ and $r_2$ are randomly chosen. The element $(i, j)$ is inserted in $\underline{v}$ immediately before $r_1$ and $r_2$

- *Single deletion*: one pair $(i_r, j_r)$ satisfying $i_r = j_r$ is randomly selected and deleted from $\underline{v}$

- *Single insertion*: one actor $i \in \mathcal{N}$ and an index $r$ are selected. The element $(i, i)$ is inserted immediately before $r$

- *Permutations*: for randomly chosen indices $r_1 < r_2$, the sequence $(i_{r_1}, j_{r_1}), \ldots, ((i_{r_2}, j_{r_2}))$ is randomly permuted

**Theorem 5.10.** The Metropolis-Hastings algorithm leads to an irreducible, aperiodic and reversible Markov-chain whose stationary distribution is given by (2.13).

*Proof.*

- *The Markov chain is irreducible*.
  Pairwise deletions and insertions and single deletion and insertion are sufficient for all $\underline{v} \in$ to communicate.

- *The Markov chain is aperiodic*.
  The graph associated to the resulting Markov-chain contains all the loops and thus the greatest common divisor of all cycles is one.

- *The Markov chain is reversible*. The detailed balance condition:

$$\rho(\widetilde{\underline{v}}, \underline{v}) u(\widetilde{\underline{v}} | \underline{v}) P(\underline{v}) = \rho(\underline{v}, \widetilde{\underline{v}}) u(\underline{v} | \widetilde{\underline{v}}) P(\widetilde{\underline{v}})$$

  is satisfied.

$$\rho(\widetilde{\underline{v}}, \underline{v}) u(\widetilde{\underline{v}} | \underline{v}) P(\underline{v}) = \min \left\{ \tfrac{P(\widetilde{\underline{v}}) u(\underline{v} | \widetilde{\underline{v}})}{P(\underline{v}) u(\widetilde{\underline{v}} | \underline{v})}, 1 \right\} u(\widetilde{\underline{v}} | \underline{v}) P(\underline{v}) =$$

$$= \min \left\{ \tfrac{P(\widetilde{\underline{v}}) u(\underline{v} | \widetilde{\underline{v}})}{u(\widetilde{\underline{v}} | \underline{v})}, P(\underline{v}) \right\} u(\widetilde{\underline{v}} | \underline{v}) =$$

$$= \min \left\{ \tfrac{u(\underline{v} | \widetilde{\underline{v}})}{u(\widetilde{\underline{v}} | \underline{v})}, \tfrac{P(\underline{v})}{P(\widetilde{\underline{v}})} \right\} u(\widetilde{\underline{v}} | \underline{v}) P(\widetilde{\underline{v}}) =$$

$$= \min \left\{ 1, \tfrac{P(\underline{v}) u(\widetilde{\underline{v}} | \underline{v})}{P(\widetilde{\underline{v}}) u(\underline{v} | \widetilde{\underline{v}})} \right\} u(\underline{v} | \widetilde{\underline{v}}) P(\widetilde{\underline{v}}) =$$

$$= \rho(\underline{v}, \widetilde{\underline{v}}) u(\underline{v} | \widetilde{\underline{v}}) P(\widetilde{\underline{v}})$$

□

The ML estimation algorithm then can be sketched in the following way:

1. For each $m = 1, \ldots, M-1$ makes a large number of Metropolis-Hastings steps yielding $v^{(i)} = (v_1^{(i)}, \ldots, v_{M-1}^{(i)})$

2. Compute the score function:

$$\frac{\partial}{\partial \theta} log(L(\widehat{\theta}_i; x; v_m^{(i)}))$$

3. Update the parameter estimate using the Robbins-Monro step

$$\theta_{i+1} = \theta_i + a_i D^{-1} U(L(\widehat{\theta}_i; x; v_m^{(i)}))$$

The estimate $\widehat{\theta}$ is calculated as the average of the $\theta_{i+1}$ values generated by this algorithm.

The Robbins-Monro algorithm and the ML estimation are implemented in the R library *RSiena* (Simulation Investigation for Empirical Network Analysis) program, now also available as an R library called *RSiena* (see the Rscript "estimation.R" for the practical implementation of the estimation procedure).

# 6 The parameter interpretation

To explain how to interpret the parameters of the SAOM model, we consider the results deriving from the Rscript "estimation.R". The analyzed data are an excerpt from the "Teenage Friends and Lifestyle Study" data set. The network is defined by friendship relations among pupils in a school in the West of Scotland. Only the 129 actors present at all three measurement points are considered in the analysis.
To find out whether it makes sense to analyze the data with the SAOM we need to check if the data are sufficiently informative to allow for the identification of the effects. In general, the total number of changes between consecutive observations should be large enough to provide the information for estimating parameters. To express quantitatively whether the data contain enough information, we can use the Jaccard index. The Jaccard index $J$ is defined by the following ratio:

$$J = \frac{N_{11}}{N_{11} + N_{01} + N_{10}}$$

where $N_{11}$ is the number of ties present at both wave, $N_{01}$ is the number of newly created ties and $N_{10}$ is the number of ties terminated. Based on experience with the method, the Jaccard values between consecutive waves should preferably be higher than 0.3. In our case the Jaccard indexes take value 0.304 for the first period and values 0.351 for the second period, thus data provide enough information for estimating parameters.
For illustrative purposes, we start considering a very simple model. Since the network was observed at three time points, two rate parameters are specified. Then, outdegree and reciprocity effects are included in the objective function. The out-degree effect always must be included in the model, while the reciprocity effect practically almost always must be included since most relations show a tendency towards reciprocity.

| | Estimates | s.e. | t-score |
|---|---|---|---|
| *Rate parameters:* | | | |
| Rate parameter period 1 | 8.5948 | ( 0.7091 ) | |
| Rate parameter period 2 | 7.2115 | ( 0.5751 ) | |
| | | | |
| *Other parameters:* | | | |
| outdegree (density) | -2.4147 | ( 0.0387 ) | -62.3875 |
| reciprocity | 2.7106 | ( 0.0811 ) | 33.4061 |

**Table 2.1:** A simple model

Table 2.1 shows first the estimated rate parameters. The rate parameter is the expected frequency,

between two consecutive network observations, with which actors get the opportunity to change a network tie. Rate parameters suggest that pupils had about 9 opportunities for change in the first period and about 7 in the second period (perhaps due to a higher friendship turnover between the first and the second school years). We should observe that the estimated rate parameters will be higher than the observed number of changes per actor for two main reasons. The former is that an actor can decide not to make any change when he has the opportunity to change a tie. The latter is that an actor may add and then withdraw the same tie during the period between two observation moments.
The parameters for the objective function, which the table shows next, are more interesting for the interpretation. In fact, they provide information about the leading forces of network evolution. In the previous pages, we pointed out that the strength of each effect $s_k$ is represented by the value assumed by the corresponding parameter $\beta_k$. The parameter $\beta_k$ can assume any real value and can be interpreted as follows. If $\beta_k$ is equal to 0, it means that the corresponding effect plays no role in the network dynamics. If it assumes a positive value, then there is higher probability of moving into networks where the corresponding effect is higher. Vice versa if the parameter takes a negative value there is higher probability of moving into networks where the corresponding effect is lower. Therefore, the first thing to do when interpreting the parameters of the SAOM is to establish which parameters are "significantly" different from 0, i.e.which effects are really important to explain network evolution. To do this we use hypotheses test. Several steps are necessary to perform a hypothesis test.

1. State the hypotheses.

   **Definition 6.1.** A *statistical hypothesis* is a statement about the parameters of a model. There are two different kind of hypothesis, which in the context of the SAOM model can be described in the following way:

   - The *null hypothesis*, denoted by $H_0$, is usually the hypothesis that the observed increase or decrease in the number of network configurations related to a certain effect results purely from chance. This means that the considered effect plays no role in the network dynamics and the null hypothesis can be expressed by:

     $$H_0 : \beta_k = 0$$

   - The *alternative hypothesis*, denoted by $H_1$, is the hypothesis that the observed increase or decrease in the number of network configurations related to a certain effect is influenced by some non-random cause. This means that there is a tendency towards or against the considered effect and the alternative hypothesis can be expressed by:

     $$H_1 : \beta_k \neq 0$$

   The hypotheses are stated in such a way that they are mutually exclusive. The observed data are used in ordered to decide between the two hypotheses, i.e. to decide if the data give evidence to the null or the alternative hypothesis.

2. Define a decision rule that allows discriminating between the null and the alternative hypothesis.

   **Definition 6.2.** The decision rule is defined by the test statistic

   $$\begin{cases} \left|\frac{\beta_k}{s.e.(\beta_k)}\right| \geq 2 & \text{reject } H_0 \\ \left|\frac{\beta_k}{s.e.(\beta_k)}\right| < 2 & \text{fail to reject } H_0 \end{cases}$$

The logic behind this decision rule is based on the standard error concept. The standard error $s.e.(\beta_k)$ is a measure of the accuracy associate to an estimate $\beta_k$ that takes into account the randomness introduced in the estimation process using simulations and only one set of data: different set of simulations and slightly different set of data can lead to slightly different estimates. For instance, let us assume that $\beta_k = 0.13$. The question is: is this value far enough from 0, so that we can conclude that the effect is important to explain network evolution? If we consider its $s.e.(\beta_k) = 0.9$, we can assume that a more or less plausible set of values that the parameter can assume is approximately $[0.04, 0.22]$. Therefore, in order to decide if a parameter $\beta_k$ is significantly different from 0, we should take into account its standard error.
The decision rule suggests that if the estimate of a parameter in its standard error unit is "small" enough (i.e. close enough to 0), then we can conclude that the data support $H_0$ and we conclude that the effect is not important (or not significant) for explaining the observed network evolution. On the contrary if the estimate of a parameter in its standard error unit is "big" enough (i.e. far enough from 0) then we can conclude that the data support $H_1$ and we conclude that the effect is important (or significant) for explaining the observed network evolution. In our context the "enough" is expressed by 2, and it is established using determined statistical tools.
The ratio between the estimate and their standard error is reported in the "t-score" column of Table 2.1.

At this point we can continue in interpreting Table 2.1. The t-scores related to outdegree and reciprocity effects are in absolute value greater than two, thus these two effects are significant in explaining network evolution. Their interpretation takes into account that the objective function expresses the attractiveness of the network for a given actor. Thus, theobjective functions are used to compare how attractive various different tie changes are.
The parameter of the outdegree effect is negative and suggests that the observed networks have low density. This means that on average the cost of establishing a friendship tie is higher than its benefit. The positive parameters related to reciprocity effect, suggests that there is strong tendency towards reciprocated ties. In more detail, let us consider the contribution to the objective function:

$$\beta_{out} \sum_{j=1}^{n} x_{ij} + \beta_{rec} \sum_{j=1}^{n} x_{ij} x_{ji} = -2.4147 \sum_{j=1}^{n} x_{ij} + 2.7106 \sum_{j=1}^{n} x_{ij} x_{ji}$$

where $\beta_{out}$ and $\beta_{rec}$ are the parameters associated to the considered effects. Adding a reciprocated tie (i.e., for which $x_{ji} = 1$) gives a positive contribution to the objective function

$$-2.4147 + 2.7106 = 0.2959$$

while adding a non-reciprocated tie (i.e., for which $x_{ji} = 0$) gives a negative contribution to the objective function (-2.4147). This means that reciprocated ties are valued positively and non-reciprocated ties are valued negatively by actors, thus actors will be reluctant to form reciprocated ties. Such ties will be formed nevertheless by chance and these are the basis on which reciprocation by others can start. We consider now a more detailed specification of the objective function. The effects related to the objective function must be determined according to hypotheses derived from theory.
In friendship context, sociological theory suggests that friendship relations tend to be reciprocated and that the statement "the friend of my friend is also my friend" is almost always true. Thus, the reciprocity transitive triplets effect are included in the model. Furthermore, it is well-known from psychological theory that pupils prefer to establish friendship relations with others that are similar to themselves. Thus, we include two covariate-related similarity effects with respect to gender

and smoking behavior, respectively (i.e. sex similarity and smoke similarity). These effects are controlled for the sender (i.e., sex ego, smoke ego) and receiver (i.e. sex alter, smoke alter) effects of the two attributes. The graphical interpretation and the verbal description of these effects are reported in Table 2.5.

| | Estimates | s.e. | t-score |
|---|---|---|---|
| *Rate parameters:* | | | |
| Rate parameter period 1 | 10.6809 | ( 1.0425 ) | |
| Rate parameter period 2 | 9.0116 | ( 0.8386 ) | |
| | | | |
| *Other parameters:* | | | |
| outdegree (density) | -2.8597 | ( 0.0608 ) | -47.0288 |
| reciprocity | 1.9855 | ( 0.0876 ) | 22.6765 |
| transitive triplets | 0.4480 | ( 0.0257 ) | 17.4558 |
| sex alter | -0.1513 | ( 0.0980 ) | -1.5445 |
| sex ego | 0.1571 | ( 0.1072 ) | 1.4659 |
| sex similarity | 0.9191 | ( 0.1076 ) | 8.5440 |
| smoke alter | 0.1055 | ( 0.0577 ) | 1.8272 |
| smoke ego | 0.0714 | ( 0.0623 ) | 1.1469 |
| smoke similarity | 0.3724 | ( 0.1177 ) | 3.1647 |

**Table 2.2:** A more complex model

The interpretation of the rate, outdegree and reciprocity parameters was provided before. Therefore, we focus our attention on the effects added to the previous model.
The parameter related to transitivity, measured by the transitive triplets effects, is significant and positive. It indicates a preference for being friends with friends'friends.
To interpret the three effects of the actor covariates related to gender, it is more useful to consider them simultaneously. The sender (ego) and receiver (alter) effect of gender are not significant, meaning that the activity or the popularity of the node does not depend on gender. On the contrary, the same gender effect is significant, suggesting that actors'similarity plays a key role on network evolution. To establish if gender similarity support or prevent tie formation we should look to the sign of the parameter. The parameter is positive, thus there is a tendency to choose friends with the same gender. To be more precise, we can consider again the contribution of gender to the objective function.
Gender was coded originally with 1 for male and with 2 for female. All actor covariates are centered, i.e. their mean value is subtracted from the original values. For instance, the variable for gender has mean $\bar{z} = 1.434$ (73 boys and 56 girls), which leads to the centered values $z_i - \bar{z} = -0.434$ for boys and $z_i - \bar{z} = 0.566$ for girls. For this variable, the model includes the *ego*, *alter* and *same* covariate effects. Let us denote the corresponding parameters by $\beta_{ego}$, $\beta_{alter}$ and $\beta_{same}$. The contribution of the single tie $x_{ij}$ to the objective function is provided by[7]:

$$\beta_{ego}(z_i - \bar{z}) + \beta_{alter}(z_j - \bar{z}) + \beta_{same}\mathbb{I}\{z_i = z_j\} = 0.1571(z_i - \bar{z}) - 0.1513(z_j - \bar{z}) + 0.9191\mathbb{I}\{z_i = z_j\}$$

where $\mathbb{I}\{z_i = z_j\}$ is the indicator function which takes the two values:

$$\mathbb{I}\{z_i = z_j\} \begin{cases} 1 & z_i = z_j \\ 0 & \text{otherwise} \end{cases}$$

Substituting the values $-0.434$ for boys and $0.566$ for girls yelds the following table:

[7]The alter and ego effects are not significant so they can be excluded by the following expression. For illustrative purposes, we does not concern about their significance.

| | Male | Female |
|---|---|---|
| Male | 0.9166 | 0.1546 |
| Female | -0.1538 | 0.9224 |

**Table 2.3:** Gender-related contributions to the objective function

Table 2.3 shows the preference for similar alters: in all rows, the highest value is at the diagonal ($z_j = z_i$). This result can be explained by the the fact that the ego and alter parameters are close to 0 (not significant), therefore the similarity effect is dominant. The interpretation is that pupils prefer to nominate friends of the same gender denoting segregation (or homophily) with respect to this actor covariate. Furthermore, the negative value associated to the the single tie form a girl to a boy, suggests that girls seem not to like male friends.
We can proceed in the similar way for the smoking behavior. Only the similarity in smoking behavior is significant and the positive value of the parameter suggests the pupils'preference in choosing friends with the same smoking behavior. The smoking behavior covariate is centered using the overall mean value from the original values of the covariate. In more detail, the smoking covariate has mean $\bar{z}_1 = 1.233$ (109 no-smokers, 10 occasional smokers and 10 regular smokers) in the first period and $\bar{z}_2 = 1.388$ (100 no-smokers, 8 occasional smokers and 21 regular smokers) in the second period. The overall mean of the two periods is given by the mean of these values, i.e. $\bar{z} = 1.310$. Thus, the centered values for the smoking covariate are: $z_i - \bar{z} = -0.310$ for no-smokers, $z_i - \bar{z} = 0.690$ for occasional smokers and $z_i - \bar{z} = 1.690$ for regular smokers. The contribution of the single tie $x_{ij}$ to the objective function is provided by the more complex expression:

$$\beta_{ego}(z_i - \bar{z}) + \beta_{alter}(z_j - \bar{z}) + \beta_{same}\left(1 - \frac{|z_i - z_j|}{R_z} - sim_z\right) =$$

$$= 0.0714(z_i - \bar{z}) + 0.1055(z_j - \bar{z}) + 0.3724\left(1 - \frac{|z_i - z_j|}{2} - 0.7415\right)$$

where $R_z$ is the range of the values assumed by the actor covariate (i.e. the difference between the highest and the smallest values, for the smoking behavior $R_z = 2$) and $sim_z$ is the mean similarity value. It is defined by the mean of all similarity scores $sim_z(ij)$

$$sim_z(ij) = 1 - \frac{|z_i - z_j|}{R_z}$$

The value of $sim_z$ is computed directly by the RSiena program through the *sienaDataCreate* function and here $sim_z = 0.7415$. Substituting the centered values for the smoking covariate yelds the following table:

| | no | occasional | regular |
|---|---|---|---|
| no | 0.0414 | -0.0734 | -0.1882 |
| occasional | -0.0393 | 0.2183 | 0.1035 |
| regular | -0.1200 | 0.1376 | 0.3952 |

**Table 2.4:** Smoking-related contributions to the objective function

We observed preference for similar alters with respect to smoking behavior. Table 2.4 shows that the net resulting preference for similar others is strongest for actors (egos) high on smoking behavior, and weakest for actors in the low range of smoking behavior.

| Effect | Formula | Effective transitions | Verbal description |
|---|---|---|---|
| Outdegree | $\sum_j x_{ij}$ |  | Overall tendency to send ties |
| Reciprocity | $\sum_j x_{ij}x_{ji}$ |  | Tendency to have reciprocal ties |
| Transitive triplets | $\sum_j x_{ij}\sum_h x_{ih}x_{jh}$ |  | Tendency toward triadic closure |
| Covariate alter | $\sum_j x_{ij}(z_j-\bar{z})$ |  | Main effect of alters covariate |
| Covariate ego | $\sum_j x_{ij}(z_i-\bar{z})$ |  | Main effect of egos covariate |
| Covariate similarity | $\sum_j x_{ij}(z_i-\bar{z})$ |  | Tendency to have ties to similar others (homophily) |

**Table 2.5:** Effects included in the model

# Chapter 3

# Modeling the co-evolution of networks and behaviors

## 1 The interdependence of network and behaviors

One of the natural extensions of the SAOM is related to the kind of changes that can happen in a social system represented by a network. Together with ties, also actors'characteristics can change over time. In the following, the term *"behavior"* will denote any changeable actor attribute (including opinions, attitudes, intentions, performances etc.).
Changes in ties and in actor's attributes are two interdependent processes. On the one hand, social network dynamics can depend on actor's characteristics, i.e. network evolution can be affected by actors' attributes. This is the result of a *selection process* in which relationship *partners* are selected according to their characteristics. The best known example of such process is the *homophily* mechanism, i.e. the formation of relations based on the similarity of two actors. On the other hand, changeable actors'characteristics can depend on the social network, i.e. behaviors can be affected by the configuration of the network. This is the outcome of an *influence process*, in which actors adjust their characteristics according to the characteristics of other actors to whom they are tied. A classical example is the assimilation process, by which connected actors become increasingly similar over time. Since hoophily and assimilation give rise to the same outcome (similarity of connected individuals), researchers have known for long time that the study of influence requires the consideration of selection and vice versa. A fundamental question is then whether this similarity is caused mainly by influence or mainly by selection. The answer to this question is then provided by the extension of the SAOM to a structure where the dependent variables consist of both tie variables and actor's behavior variables.

## 2 Data

In this section we follow the notation used in Chapter 2. Let $x$ be a binary adjacency matrix representing a network. The network is defined by a relation $\mathcal{R}$ collected over a set $\mathcal{N}$ of actors. For each actor we collected also a series of actor attributes. Let $V_1, \cdots, V_H$ denote the $H$ constant covariates (such as gender, ethnicity, and so on) and $Z_1(t), \cdots, Z_L(t)$ the $L$ behavior covariates (such as, smoking or drinking behaviors). It is assumed that the behavior variables are ordinal categorical variables , with values 1,2, etc.[1]

[1] A categorical variable is a variable with modalities in the form of categories, such as, for example, man and woman of the variable sex; when it is possible to order the modalities of a categorical variable, the variable is said to be an ordinal categorical variable. Examples are the professional variable with categories "qualified", "semiqualified", and "nonqualified" or the smoking behavior with categories "no", "occasional" and"regular".

Networks and behaviors are observed at $M \geq 2$ time points $t_1, \cdots, t_M$. The corresponding time series of observed networks and behaviors will be denoted by:

$$(x,z)(t_1),\ (x,z)(t_2), \cdots,\ (x,z)(t_M)$$

This time series together with the constant actor's attribute $V_1, \cdots, V_H$ constitutes the the *longitudinal network-behavior panel data* that we want to analyze in order to determine the leading forces which govern network and behavioral changes and to distinguish between selection and influence mechanism. In more detail the time series of observed networks and behaviors is the dependent variable of the model and our aim is to explain it as a function of structural effects and individuals'characteristics (constant or time-varying covariates). For simplivity in the next sections we will considered $L = 1$.

# 3 Assumptions

The assumptions for the actor-based model for the dynamics of networks and behaviors are extensions of the assumptions of network dynamics for the SAOM.

1. *Distribution of the process.* Changes between observational time points are modeled according to a continuous-time Markov chain. The state space $\mathcal{C}$ of the chain is comprised of all possible configurations arising from the combination of network and behaviors. The number of possible networks based on a set of $n$ actors is $2^{n(n-1)}$, while the possible distribution over the $n$ actors of a behavior variable with B categories is $B^n$. Therefore the cardinality of $\mathcal{C}$ is $|C| = 2^{n(n-1)} \times B^n$.
   The continuous-time assumption expresses the idea that the changes reflected in the longitudinal data typically accrue over the unobserved periods between the observation moments in continuous time. The importance of this consideration in explaining the co-evolution of networks and behaviors becomes clear when looking at Figure 3.1a. Let us assume, at $t_0$, actor $i$ considers actor $j$ a friend but they are not behaviorally similar, whereas at $t_1$, $i$ again considers $j$ a friend but now they are behaviorally similar. Then, one is tempted to conclude that this is the result of an influence process. However, the unobserved process that generated these data may have looked very different, as illustrated in the brackets. The relation between $i$ and $j$ may have ended, and $i$ may have changed his behavior in the absence of a relation with $j$. After the behavior of $i$ changed, $i$ may have renewed is the friendship with $j$. According to this scenario, the change in behavior happened in the absence of a tie form $i$ to $j$ and the configuration observed at time $t_1$ is actually the result of a selection process. Figure 3.1b illustrates how social influence can be misdiagnosed as homophilous selection. The Markov-chain assumption means that the changes actors make are assumed to depend only on the current state of the network and the behaviors, not on previous configurations.

2. *Opportunity to change.* At any given moment one probabilistically selected actor has the opportunity to change one of his outgoing tie or his behavior. The moments at which an actor has the opportunity for a tie change or a behavior change are modeled by two distinct rate functions.

3. *Absence of co-occurrence.* No more than one tie or one behavior can change at any given moment, i.e. only one actor has the opportunity to change (an outgoing time or one behavior) at each instant $t$.

4. *Actor-oriented perspective.* Actors control their outgoing ties as well as their own behavior, i.e. the changes in outgoing ties and behavior are determined by the position of the actor

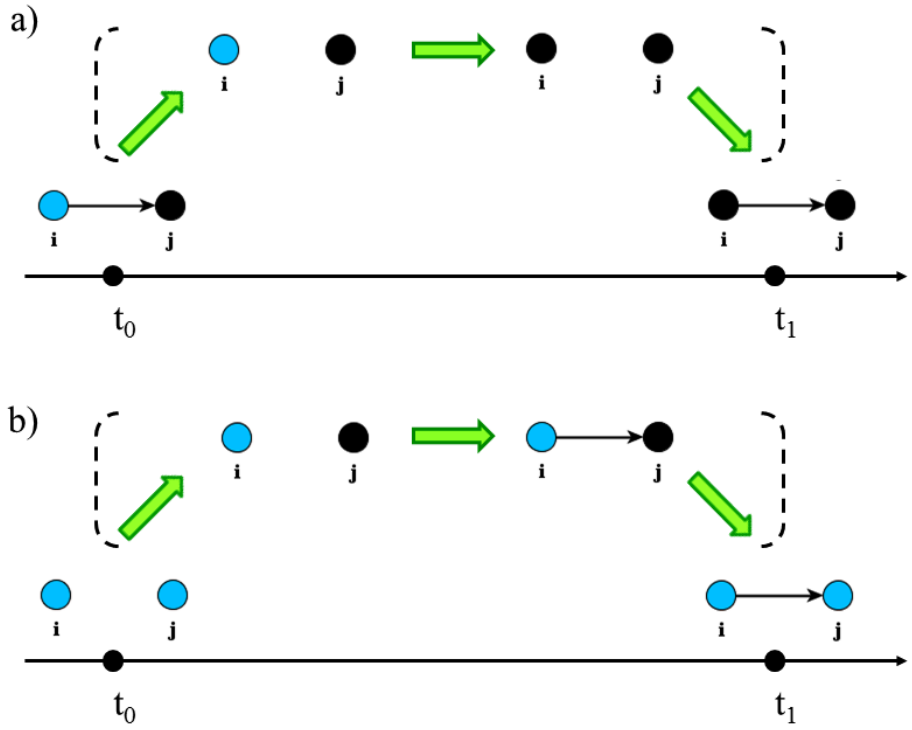



**Figure 3.1:** Selection and influence as alternative mechanisms

in the network, his individual characteristics, as well as the characteristics and the behavior of the other actors in the network. This means that actors do not change ties and behavior at will but they act in order to improve their preferences towards the resulting network and behavior configuration. These preferences are modeled by two distinct objective functions: one for the network and one for the behavior change.
As already mentioned the behavior variables are ordinal discrete variables. It is assumed that once an actor $i$ has the opportunity to change his behavior, he can increase or decrease the level of its behavior only by one unit. For instance, let us suppose that the current behavior level of actor $i$ is $z_i = l$. Once $i$ gets the opportunity to change his behavior, he can decide to increase his level by one unit ($z_i = l + 1$), to decrease his level by one unit ($z_i = l - 1$) or not to change $z_i = l$

Assumptions 2 and 4 suggest that network evolution and behavioral evolution are modeled according to two distinct processes: a network process and a behavior process. This does not mean that network and behavior interdependence is neglected. As it will be shown in the following section, the interdependence of networks and behaviors is modeled through the effects that specify the objective function for the network change and the behavioral change.

# 4 Model formulation

According to the previous assumptions, the network-behavior co-evolution process is decomposed into a series of micro-steps that consist of either the opportunity of changing one network tie and the corresponding tie changed or by the opportunity of changing a behavior and the corresponding unit changed in behavior. Consequently, every micro-step requires the identification of a focal actor who gets the opportunity to make a change and the identification of the change outcome.
To take into account the double nature of the micro-steps, the basic co-evolution model is de-

| | Occurrence | *Preference* |
|---|---|---|
| Network changes | Network rate function | Network objective function |
| Behavioral changes | Behavioral rate function | Behavioral objective function |

**Table 3.1:** Model components of the SAOM for network change

scribed by four components, expressing the amount of opportunities for a change (modeled by the rate functions) and the direction of a change (modeled by the objective functions). This structure is represented in Table 3.1.
The first observations of network $x(t_0)$ and behavior $z(t_0)$ serve as starting values of the evolution process, so that we do not concern about the process that may have generated the first network and behavior observation and only the subsequent changes of network ties and behavior are modeled.

## 4.1 The rate functions

The frequency by which actors have the opportunity to make a change is modeled by the *rate functions*, one for each type of change. The main reason for having two different rate functions for the behavioral and the network changes is that practically always, one type of decision will be made more frequently than the other. For instance, in the joint study of friendship and smoking behavior at high school, we would expect more frequent changes in the network than in behavior, because of the addictive nature of substance use and the high friendship turnover in adolescence.
The Markov chain assumption requires that the waiting times between micro steps must have exponential distributions. For each actor $i$, we denote by $T_i^{net}$ and by $T_i^{beh}$ the waiting time until $i$ gets the opportunity to make a network or a behavioral change, respectively. Therefore, $T_i^{net}$ and $T_i^{beh}$ are exponentially distributed with parameters $\lambda_i^{net}$ and $\lambda_i^{beh}$, respectively.These parameters indicate the speed at which the respective change is likely to occur. The parameters $\lambda$ that govern the rate functions are indexed by a $i$, because we assume that actors may change their network ties, or their behavior, at different frequencies. We can incorporate such activity differences between actors by letting the parameters $\lambda$ depend on actor attributes and network positions.
The Markov-property suggests that all waiting times are independent, given the current state of network and behavior. Thus, properties of the exponential distribution imply that, starting from the moment of the preceding micro step, the waiting time until the occurrence of the next micro step of either kind by any actor is exponentially distributed with parameter:

$$\lambda_{tot} = \sum_i (\lambda_i^{net} + \lambda_i^{beh})$$

It follows that the probability of a network micro-step is taken by actor $i$ is given by

$$P(i \text{ has the opportunity to change one of his tie}) = \frac{\lambda_i^{net}}{\lambda_{tot}}$$

and the probability of a behavioral micro-step is taken by actor $i$ is given by

$$P(i \text{ has the opportunity to change his behavior}) = \frac{\lambda_i^{beh}}{\lambda_{tot}}$$

If we consider the simplest specification of the rate functions, that assumes that all actors have the same rate of change between two observational moments, $T_i^{net}$ and $T_i^{beh}$ are exponentially distributed with parameters $\lambda^{net}$ and $\lambda^{beh}$, respectively. Therefore, the waiting time until occurrence of the next micro step of either kind by any actor is exponentially distributed with parameter:

$$\lambda_{tot} = \sum_i (\lambda_i^{net} + \lambda_i^{beh}) = n(\lambda^{net} + \lambda^{beh})$$

It follows that the probability of a network micro-step is given by

$$P(\text{network micro-steps}) = \frac{n\lambda^{net}}{\lambda_{tot}}$$

and the probability of a behavioral micro-step is given by

$$P(\text{behavioral micro-step}) = \frac{n\lambda^{beh}}{\lambda_{tot}}$$

and as a consequence

$$P(i \text{ has the opportunity to change one of his tie}) = \frac{\lambda^{net}}{\lambda_{tot}}$$

$$P(i \text{ has the opportunity to change his behavior}) = \frac{\lambda^{beh}}{\lambda_{tot}}$$

In the following we assume that both types of rate functions are constant across actors, and depend only on the observational period.

## 4.2 The objective functions

The probability of going from one state to another state of the co-evolution Markov chain is defined by the objective functions. Since we are considering the co-evolution of networks and behaviors, we use a slightly different notation. We denote the objective function for the network changes by $f_i^{net}(\beta, x(i \rightsquigarrow j), z)$. The definition and the interpretation of $f_i^{net}(\beta, x(i \rightsquigarrow j), z)$ were described deeply in Chapter 2. In this section we will focus on the objective function for behavioral change.

We denote the behavioral objective function by $f_i^{beh}(\gamma, z(l \rightsquigarrow l'), x)$. The three quantities between brackets express the dependence of the objective function on the statistical parameters $\gamma$, the behavior of actor $i$ when he changes his behavior form l to $l' \in \{l-1, l, l+1\}$ and the current network configuration $x$.

The mathematical formulation of $f_i^{beh}(\gamma, z(l \rightsquigarrow l'), x)$ is again a linear combination of a vector $\gamma$ of statistical parameters and a vector of effects denoted by $s^{beh}(x, z)$ plus a random utility term $\varepsilon_i(t, z, l, l')$

$$f_i^{beh}(\gamma, z(l \rightsquigarrow l'), x) = \sum_{w=1}^{W} \gamma_k s_{iw}^{beh}(x, z) + \varepsilon_i(t, z, l, l')$$

This function will be different from the objective function for network change since it needs to represent primarily how likely it is for the actor to change the current level of its behavior. Therefore, effects which are supposed to play a key role in the behavioral evolution must be different from those important for describing network evolution.

The strength of each effect is represented by the corresponding parameter $\gamma_k$, which should be estimated on the basis of the longitudinal network data. $\gamma_k$ can assume any real value and its interpretation is equivalent to that of the parameters of the objective function for network change.

The term $\varepsilon_i(t, z, l, l')$ is assumed to be distributed as a Gumbel distribution. Thus, according to equation (2.2), the probability that an actor $i$ changes his own behavior by one unit is:

$$p_{ll'}(\gamma; z(l \rightsquigarrow l'); x) = \frac{exp\left(\sum_{w=1}^{W} \gamma_w s_{iw}^{beh}(x, z(l \rightsquigarrow l'))\right)}{\sum_{l'' \in \{l+1, l-1, l\}} exp\left(\sum_{w=1}^{W} \gamma_w s_{iw}^{beh}(x, z(l \rightsquigarrow l''))\right)} \tag{3.1}$$

If $l = l'$, then equation (3.1) represents the probability that actor $i$ does not change his behavior.

## 4.3 The effects of objective function for behavioral change

Several effects are proposed to specify the objective function for behavioral change.

- *Basic shape effects*
  The linear shape effect $s^{beh}_{i\_linear}(x,z)$ and the quadratic shape effect $s^{beh}_{i\_quadratic}(x,z)$ are defined by the level of the behavior for the actor $i$ and its quadratic value, respectively:

$$s^{beh}_{i\_linear}(x,z) = z_i \qquad\qquad s^{beh}_{i\_quadratic}(x,z) = z_i^2$$

  The basic shape effects must be always included in the model specification, since they represent the relative preference for the specific value $z_i$ of the behavior. In the special case in which the behavior variable is dichotomous, the linear shape effect suffices, since each function of two values can be represented by a linear function.

- *Classical nfluence effects*
  The influence effects describe the influence of alter's behavior on ego's behavior. There are different ways to measure the influences from different alter:

  1. The *average similarity effect* $s^{beh}_{i\_avsim}(x,z)$ expressing the preference of actors to be similar in behavior to their alters, in such a way that the total influence of the alters is the same regardless of the number of alters (i.e., ego's outdegree $x_{i+}$):

$$s^{beh}_{i\_avsim}(x,z) = \frac{1}{x_{i+}} \sum_{j=1}^{n} x_{ij}(sim_z(ij) - sim_z)$$

     where

$$sim_z(ij) = 1 - \frac{|z_i - z_j|}{R_z}$$

     $R_z$ is the range of the behavior $z$ and $sim_z$ is the mean similarity value.
  2. The *total similarity effect* $s^{beh}_{i\_totsim}(x,z)$, expressing the preference of actors to be similar in behavior to their alters, in such a way that the total influence of the alters is proportional to the number of alters:

$$s^{beh}_{i\_totsim}(x,z) = \sum_{j=1}^{n} x_{ij}(sim_z(ij) - sim_z)$$

  The choice among these different representations of social influence has to be made on theoretical grounds.

- *Position-dependent influence effects*
  Network position could also have an effect on the behavior of dynamics. Therefore the outdegree effect $s^{beh}_{i\_out}(x,z)$ or $s^{beh}_{i\_ind}(x,z)$ indegree effect may be included in the objective function for network behavior.

$$s^{beh}_{i\_out}(x,z) = z_i \sum_{j=1}^{n} x_{ij}$$

$$s^{beh}_{i\_ind}(x,z) = z_i \sum_{j=1}^{n} x_{ji}$$

- *Effects of other actor variables.*
  For each actor's attribute a main effect on the behavior can be included in the model. This reflects the influence of the considered actor characteristics on changes in the behavior.

# 5 Simulating the co-evolution of networks and behaviors

This section presents a probabilistic algorithm to simulate the co-evolution of networks and behaviors for fixed parameter values. This algorithm is the extensions of the "network evolution" algorithm presented in Chapter 2. Therefore, the algorithm consists in reproducing a possible series of micro-steps between two observation moments $t_0$ and $t_1$ according to fixed parameter value, the initial network $x(t_0)$ and the initial behavior $z(t_0)$. The simulation algorithm is as follows.

1. Set the time $t = 0$, $x = x(t_0)$ and $z = z(t_0)$
2. Generate $T^{net}$ according to an exponential distribution with parameter $\lambda^{net}$
3. Generate $T^{beh}$ according to an exponential distribution with parameter $\lambda^{beh}$
4. If $\min\left\{T^{net}, T^{beh}\right\} = T^{net}$ a *network micro-step* is implemented:

   - Select the actor $i \in \mathcal{N}$, who makes the changes, with probability

   $$P(i \text{ has the opportunity to change one tie}| \text{ network micro-steps}) = \frac{\lambda^{net}}{\lambda_{tot}}$$

   - Select the actor $j \in \mathcal{N}$, to whom $i$ changes his outgoing tie, with probability:

   $$p_{ij}(\beta; x(i \rightsquigarrow j); z) = \frac{exp\left(\sum_{k=1}^{K} \beta_k s_{ik}(x(i \rightsquigarrow j), z)\right)}{\sum_{h=1}^{n} exp\left(\sum_{k=1}^{K} \beta_k s_{ik}(x(i \rightsquigarrow h), z)\right)}$$

   - If $i \neq j$ then $x = x(i \rightsquigarrow j)$. If $i = j$ then $x = x$
   - Set $t = t + T^{net}$

   Otherwise if $\min\left\{T^{net}, T^{beh}\right\} = T^{beh}$ a *behavioral micro-step* is implemented:

   - Select the actor $i \in \mathcal{N}$, who makes the changes, with probability

   $$P(i \text{ has the opportunity to change his behavior}|\text{behavioral micro-steps}) = \frac{\lambda^{beh}}{\lambda_{tot}}$$

   - Determine the behavioral change $l' \in \{l+1, l-1, l\}$ with probability:

   $$p_{ll'}(\gamma; z(l \rightsquigarrow l'); x) = \frac{exp\left(\sum_{w=1}^{W} \gamma_w s_{iw}^{beh}(x, z(l \rightsquigarrow l'))\right)}{\sum_{l'' \in \{l+1, l-1, l\}} exp\left(\sum_{w=1}^{W} \gamma_w s_{iw}^{beh}(x, z(l \rightsquigarrow l''))\right)}$$

   - If $l \neq l'$ then $z = z(l \rightsquigarrow l')$. If $l = l'$ then $z = z$
   - Set $t = t + T^{beh}$

5. Repeat step 2. to 4. until the stopping criterion is satisfied.

The stopping rules for the algorithm of the co-evolution of network and behaviors are the same of those for the algorithm of network evolution, but for the conditional simulation we can now define a stopping rule based also on the behavior variable. Conditioning on the behavior variable means running simulations until the number of different entries between the initially observed behavior vector and the simulated behavior vector is equal to the number of entries in the vector of the behavior variable that differ between the initially and the finally observed behavior variable of this period.
The algorithm for the unconditional estimation is summed up in Algorithm 5.

**Algorithm 5:** Co-evolution of network and behaviors

**Input**: $x(t_0)$, $z(t_0)$, $\lambda^{net}$, $\lambda^{beh}$, $\beta$, $\gamma$, $n$
**Output**: $x^{sim}(t_1)$,$z^{sim}(t_1)$
$t \leftarrow 0$
$x \leftarrow x(t_0)$
$z \leftarrow z(t_0)$
**while** $t < 1$ **do**
  $T^{net} \sim Exp(lambda^{net})$
  $T^{beh} \sim Exp(lambda^{beh})$
  **if** $\min\{T^{net}, T^{beh}\} = T^{net}$ **then**
    $i \sim Uniform(1,\ldots,n)$
    select $j$ with probability $p_{ij} = \frac{e^{f_i^{net}(\beta, x(i\rightsquigarrow j), z)}}{\sum_{h=1}^{n} e^{f_i^{net}(\beta, x(i\rightsquigarrow j), z)}}$
    **if** $i \neq j$ **then**
      $x = x(i \rightsquigarrow j)$
    **else**
      $x = x$
    $t \leftarrow= t + T^{net}$
  **else**
    $i \sim Uniform(1,\ldots,n)$
    select $l'$ with probability $p_{ll'} = \frac{e^{f_i^{beh}(\gamma, z(l\rightsquigarrow l'), x)}}{\sum_{l'' \in \{l+1, l-1, l\}} e^{f_i^{beh}(\gamma, z(l\rightsquigarrow l''), x)}}$
    **if** $l \neq l'$ **then**
      $z = z(l \rightsquigarrow l')$
    **else**
      $z = z$
  $t \leftarrow= t + T^{net}$
$x^{sim}(t_1) \leftarrow x$
$z^{sim}(t_1) \leftarrow z$
**return** $x^{sim}(t_1), z^{sim}(t_1)$

# 6 The parameter estimation

The estimation of the parameters of the SAOM for the co-evolution of network and behaviors are obtained mainly using the Method of Moments. The Maximum likelihood estimation is currently too time-consuming and is still under construction.
We will denote by $\theta$ the 2(M-1)+K+W dimensional vector of parameters ($M-1$ parameters for the network rate function, $M-1$ parameters for the behavior rate function, K parameters for the network objective function, W parameters for the behavior objective function). For network-behavior co-evolution the following four types of statistics for the model parameters were proposed:

- Network rate parameters for the period $m$

$$s^{net}_{\lambda_m}(X(t_m), X(t_{m-1})|X(t_{m-1}) = x(t_{m-1})) = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \left| X_{ij}(t_m) - X_{ij}(t_{m-1}) \right|$$

- Behavior rate parameters for the period $m$

$$s^{beh}_{\lambda_m}(Z(t_m), Z(t_{m-1})|Z(t_{m-1}) = z(t_{m-1})) = \sum_{i=1}^{n} |Z_i(t_m) - Z_i(t_{m-1})|$$

- Network objective function effects

$$\sum_{m=1}^{M-1} s^{net}_{mk}(X(t_m)|X(t_{m-1}) = x(t_{m-1})) = \sum_{m=1}^{M-1} s^{net}_{mk}(X(t_m), X(t_{m-1}))$$

- Behavior objective function effects

$$\sum_{m=1}^{M-1} s^{beh}_{mw}(X(t_m)|X(t_{m-1}) = x(t_{m-1})) = \sum_{m=1}^{M-1} s^{beh}_{mw}(X(t_m), X(t_{m-1}))$$

Consequently the MoM estimator for $\theta$ is provided by the solution of the system of equations:

$$\begin{cases} E_\theta \left[ s_{\lambda_m}(X(t_m), X(t_{m-1})|X(t_{m-1}) = x(t_{m-1})) \right] = s_{\lambda_m}(x(t_m), x(t_{m-1})) & m = 1, \cdots, M-1 \\ E_\theta \left[ s_{\lambda_m}(Z(t_m), Z(t_{m-1})|Z(t_{m-1}) = z(t_{m-1})) \right] = s_{\lambda_m}(z(t_m), z(t_{m-1})) & m = 1, \cdots, M-1 \\ E_\theta \left[ \sum_{m=1}^{M-1} s^{net}_{mk}(X(t_m)|X(t_{m-1}) = x(t_{m-1})) \right] = \sum_{m=1}^{M-1} s^{net}_{mk}(x(t_m), x(t_{m-1})) & k = 1, \cdots, K \\ E_\theta \left[ \sum_{m=1}^{M-1} s^{beh}_{mw}(X(t_m)|X(t_{m-1}) = x(t_{m-1})) \right] = \sum_{m=1}^{M-1} s^{beh}_{mw}(x(t_m), x(t_{m-1})) & w = 1, \cdots, W \end{cases}$$

This system of equation is solved using the Robbins-Monro algorithm, described in Chapter 2.

# 7 The parameter interpretation

In this section we consider the parameter interpretation of the SAOM model for the co-evolution of networks and behaviors. The analyzed data are the excerpt from the "Teenage Friends and Lifestyle Study" data set already used in Chapter 2.
In the literature there are several studies suggesting that the similarity in smoking behavior could be caused by selection of similar others as friends as well as by influence processes where friends adjust their smoking behavior to each other, or by a combination of these. Therefore, in the following we will use the SAOM for the co-evolution of networks and behaviors (see the Rscript "coevolution.R") to disentangle influence and selection processes. In particular we can try to answer these two questions:

1. Do pupils select friends based on similar smoking behavior?

2. Are pupils influenced by friends to adjust to their smoking behavior?

In this context, the dependent variables are represented by the friendship networks and the smoking attitudes gathered at the three time points. To find out whether it makes sense to analyze the data with a co-evolution model, we first need to check whether there is interdependence between networks and behavioral variables. Second we should check whether the data are sufficiently informative to allow for identification of effects.
One measure of interdependence is the Moran index.

$$I = -\frac{\overline{d}(\delta)}{(n-1)\overline{d}}$$

where

$$\overline{d}(\delta) = \frac{\sum_{ij} x_{ij}(z_i - \overline{z})(z_j - \overline{z})}{\sum_{ij} x_{ij}}$$

is the mean of the cross products of the centered behavioral variable for connected actors and

$$\overline{d} = \frac{\sum_{ij}(z_i - \overline{z})(z_j - \overline{z})}{n(n-1)}$$

is the overall mean of the cross products of the centered behavioral variable for all the possible pairs of actors in the network (i.e. without considering the presence of ties between the actors).
The Moran index can in theory range from -1 to +1, with values close to zero indicating independence between networks and behaviors, and the values +1 or -1 indicating identity or perfect complementarity of the behaviors of two friends (i.e. very strong interdependence). The computation of the Moran index for the friendship networks and smoking behaviors at the three measurement points leads to the values 0.244, 0.258 and 0.341. Since these values are greater than zero, there is considerable dependence between networks and behaviors.
The Jaccard indexes for the two periods allow for checking whether the data provide enough information for estimating the parameters. The values taken by the Jaccard indexes are higher than 0.3 (0.304 and 0.351), revealing that the data are sufficiently informative to allow for identification of effects. Since the two "preconditions" are satisfied, we can estimate the co-evolution model.
For illustrative purposes, we start considering a baseline model. Networks and the Behaviors were observed at three time points, thus four rate parameters are specified: two for the network rate functions and two for the behavioral rate functions. Then, outdegree and reciprocity effects are included in the network objective function, and the linear and the quadratic shape effects are included in the behavioral objective function[2].
Table 3.2 shows the estimates of the parameters. The table is divided up into two parts.
The former concern to the network dynamics and we have already learned how to interpret it. The rate parameter is the expected frequency, between two consecutive network observations, with which actors get the opportunity to change a network tie. The rate parameters suggest that pupils had about 9 opportunities for change in the first period and about 7 in the second period. The outdegree and the reciprocity effects are both significant (absolute t-scores $> 2$). The parameter of the outdegree effect is negative and suggests that the cost of establishing a friendship tie is higher than its benefit. The positive parameters related to reciprocity effect, suggests that there is strong tendency towards reciprocated ties.
The latter refers to the behavioral dynamics. The behavioral rate parameter is the expected frequency, between two consecutive network observations, with which actors get the opportunity to change their behavior by one unit. The rate parameters suggest that pupils had about 4 opportunities for change in the two considered periods. We should observe that the behavioral rate

[2]The quadratic shape effect is included because the number of the smoking behavioral categories is higher than 2.

| | Estimates | s.e. | t-score |
|---|---|---|---|
| *Network Dynamics* | | | |
| constant friendship rate (period 1) | 8.6287 | ( 0.6666 ) | |
| constant friendship rate (period 2) | 7.2489 | ( 0.5466 ) | |
| outdegree (density) | -2.4084 | ( 0.0407 ) | -59.12676 |
| reciprocity | 2.7024 | ( 0.0823 ) | 32.8337 |
| *Behavior Dynamics* | | | |
| rate smokebeh (period 1) | 3.8922 | ( 1.9689 ) | |
| rate smokebeh (period 2) | 4.4813 | ( 2.3679 ) | |
| behavior smokebeh linear shap | -3.5464 | ( 0.4394 ) | -8.0712 |
| behavior smokebeh quadratic shape | 2.8464 | ( 0.3628 ) | 7.8447 |

**Table 3.2:** A baseline model

parameters will be higher than the observed number of behavioral changes per actor for two main reasons: an actor can decide not to make any change when he has the opportunity to change his behavior; an actor may first increase and then decrease (or vice versa) by one unit his behavior during the period between two observation moments. Comparing the rate parameters of the network rate functions and the behavioral rate functions we note that the speed at which tie change occur is higher than the speed at which behavioral change occur.

The interpretation of the parameters of the behavioral objective function is more complex. We should keep in mind that the behavioral objective function expresses the attractiveness of different behavioral levels taking into account the current structure of the network and the behavior of the other actors. Both the linear and the quadratic shape effects are highly significant (absolute t-scores $> 2$). To interpret the parameters it should be noted that all actors'covariates are internally centered by RSiena, and that the mean values used for the centering are directly provided by the program. The overall mean value of the smoking behavior is 1.377. In more detail, the smoking covariate has mean $\bar{z}_1 = 1.233$ (109 no-smokers, 10 occasional smokers and 10 regular smokers) at the first measurement point, $\bar{z}_2 = 1.388$ (100 no-smokers, 8 occasional smokers and 21 regular smokers) at the second measurement point and $\bar{z}_3 = 1.512$ (93 no-smokers, 6 occasional smokers and 30 regular smokers) at the third measurement point. The overall mean of the three observation points is given by the mean of these values, i.e. $\bar{z} = 1.377$. Thus, the centered values for the smoking covariate are: $z_i - \bar{z} = -0.377$ for no-smokers, $z_i - \bar{z} = 0.623$ for occasional smokers and $z_i - \bar{z} = 1.623$ for regular smokers.

The contribution to the behavioral objective function by actor $i$ is:

$$f_i^{beh}(\gamma, x, z) = \gamma_{linear}(z_i - \bar{z}) + \gamma_{quadratic}(z_i - \bar{z})^2 =$$

$$= -3.5464_{linear}(z_i - \bar{z}) + 2.8464_{quadratic}(z_i - \bar{z})^2$$

and it is depicted in Figure 3.2

The behavioral objective function is a unimodal preference function, with the maximum/minimum attained for $z_i = \bar{z} - \frac{\gamma_{linear}}{2\gamma_{quadratic}} = 1.778$. The U-shaped suggests that changes in the behavior are drawn to the extreme of the range, with actors already low on the behavior to lower values and actors already high on the behavior to higher values. This result reflects the U-shaped distribution of the smoking behavior in the pupils'population.

The baseline model does not provide any information about selection and influence processes because the network dynamics are explained by the preference towards creating and reciprocating ties and the behavior dynamic are described only by the distribution of the behavior in the popula-

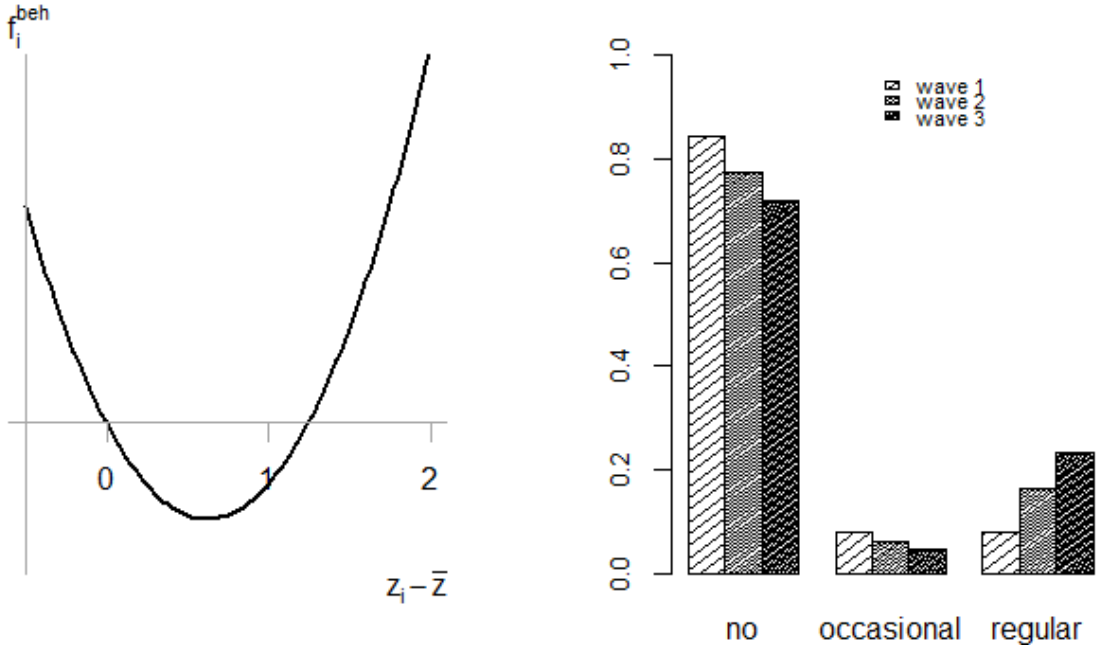



**Figure 3.2:** Shape of the behavioral objective function and distribution of the smoking behavior in the pupils'population

tion. If we want to disentangle the selection and influence effects a more complex model must be specified. This model should be able to simultaneously represent changes in friendship network structure and changes in smoking behavior among pupils. Therefore, we consider the following model specification.

The network objective function is specified according to the final specification of the SAOM presented in Chapter 2 (Table 2.5). The behavioral objective function includes, besides the baseline effects, the indegree, outdegree and average similarity effect related to smoking behavior (Table 3.5). The choice of the average similarity effect instead of the total similarity effect can be justified by the (sociological) theory that each pupil is influenced by its own group of friends in about the same way, so that it is not necessary to take into account the total amount of friends.

Table 3.3 shows the estimates of the parameters. The rate parameters suggest that actors'opportunity to change their outgoing ties is higher in the first period than in the second period (11 vs. 9 opportunities), while actors'opportunity to change their behavior is nearly the same during the two periods (nearly 4 opportunities).

The estimates for the network objective function parameters leads to the same conclusions of the model estimated at the end of Chapter 2. The main result is related to the similarity based on smoking behavior. Controlling for the influence process (i.e. including the behavior dynamics in the model), the parameter related to the smoking similarity is still positive and significant. Thus, there is evidence for smoking-based friendship selection.

The parameter estimates of the rate behavioral objective function reveal that the indegree and the outdegree effects turned out not to be significant contributors to the behavior dynamics (absolute t-scores $\leq 2$). Therefore, the contribution to the behavioral objective function is given by:

$$\gamma_{linear}(z_i - \bar{z}) + \gamma_{quadratic}(z_i - \bar{z})^2 + \gamma_{avsim} \frac{1}{x_{i+}} \sum_{j=1}^{n} x_{ij}(sim_z(ij) - sim_z) =$$

$$= -3.3573_{linear}(z_i - \bar{z}) + 2.8406_{quadratic}(z_i - \bar{z})^2 + 3.4361 \frac{1}{x_{i+}} \sum_{j=1}^{n} x_{ij}(sim_z(ij) - 0.7415)$$

| | Estimates | s.e. | t-score |
|---|---|---|---|
| *Network Dynamics* | | | |
| constant friendship rate (period 1) | 10.7166 | ( 1.4036 ) | |
| constant friendship rate (period 2) | 9.0005 | ( 0.7709 ) | |
| outdegree (density) | -2.8435 | ( 0.0572 ) | -49.6776 |
| reciprocity | 1.9683 | ( 0.0933 ) | 21.1077 |
| transitive triplets | 0.4447 | ( 0.0322 ) | 13.7964 |
| sex ego | 0.1612 | ( 0.1206 ) | 1.3368 |
| sex alter | -0.1476 | ( 0.1064 ) | -1.3871 |
| sex similarity | 0.9104 | ( 0.0882 ) | 10.3244 |
| smoke ego | 0.0665 | ( 0.0846 ) | 0.7857 |
| smoke alter | 0.1121 | ( 0.0761 ) | 1.4719 |
| smokebeh similarity | 0.5114 | ( 0.1735 ) | 2.9479 |
| *Behavior Dynamics* | | | |
| rate smokebeh (period 1) | 3.9041 | (1.7402 ) | |
| rate smokebeh (period 2) | 3.8059 | ( 1.4323 ) | |
| behavior smokebeh linear shap | -3.3573 | ( 0.5678 ) | -5.9129 |
| behavior smokebeh quadratic shape | 2.8406 | ( 0.4125 ) | 6.8864 |
| behavior smokebeh indegree | 0.1711 | ( 0.1812 ) | 0.9444 |
| behavior smokebeh outdegree | 0.0128 | ( 0.1926 ) | 0.0662 |
| behavior smokebeh average similarity | 3.4361 | ( 1.4170 ) | 2.4250 |

**Table 3.3:** A co-evolution model for the analysis of selection and influence mechanisms

Since the behavioral objective function depends not generally on the average behavior of the actor's friends, here we present a table only for the special case of actors all whose friend have the same behavior $z_j$. The result is given in Table 3.4.

| $z_j$ / $z_i$ | no | occasional | regular |
|---|---|---|---|
| 1 | 2.56 | -1.82 | -0.51 |
| 2 | 0.84 | -0.10 | 1.20 |
| 3 | -0.88 | -1.82 | 2.92 |

**Table 3.4:** Smoking-related contributions to the behavioral objective function

The interpretation of Table 3.4 is that each row corresponds to a given common behavior of the focal actor's friends; comparing the different values in the row shows the relative attractiveness of the different potential values of ego's own behavior. The row maximum is assumed at the diagonal for the non-smokers and for the regular smokers. This means that for low and high values of the common friends' behavior $z_j$, the focal actor prefers to have the same behavior as all these friends. The differences in the bottom row is larger than in the top row, indicating that in the case where the friends do not smoke at all, the preference (or social pressure) toward imitating their behavior is less strong than in the case where all the friends smoke a lot. The second row of the table suggests that if all the actor's friend are occasional smokers, then it is more rewarding for the focal actor to adjust his behavior towards a higher value of the smoking attitude.

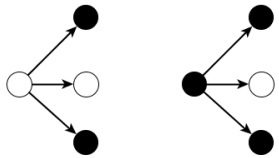
| Effect | Formula | Effective transitions | Verbal description |
|---|---|---|---|
| Shape: linear and quadratic | $z_i$ and $z_i^2$ |  | Basic shape of the observed distribution of the behavioral variable |
| Average similarity | $\sum_j x_{ij}x_{ji}$ |  | Assimilation to neighbors' average behavior |
| Indegree | $z_i \sum_j x_{ji}$ |  | Effect of own popularity in the network on behavior |
| Outdegree | $z_i \sum_j x_{ij}$ |  | Effect of own activity in the network on behavior |

**Table 3.5:** Effects for the behavioral dynamics

# Bibliography

R.M. Ripley and T.A.B. Snijders. Manual for siena version 4.0. *University of Oxford, Department of Statistics & Nuffield College*, 2010.

H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.

D. Ruppert. Efficient estimations from a slowly convergent robbins-monro process. 1988.

T. Snijders and M. Van Duijn. Simulation for statistical inference in dynamic network models. *lecture notes in Economics and Mathematical systems*, 1997.

T.A.B. Snijders. The statistical evaluation of social network dynamics. *Sociological methodology*, pages 361–395, 2001.

T.A.B. Snijders. Models for Longitudinal Network Data. *Models and methods in social network analysis*, page 215, 2005.

T.A.B. Snijders, G.G. Van de Bunt, and C.E.G. Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60, 2010.

C. Steglich, T.A.B. Snijders, and M. Pearson. Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1):329–393, 2010.