Università
Ca'Foscari
Venezia

Department
of Management

Working Paper Series

C. Colapinto, E. Sartori, and M. Tolotti

A two-stage model for diffusion of innovations

Working Paper n. 16/2012
October 2012

ISSN: 2239-2734

A two-stage model for diffusion of innovations

Cinzia Colapinto*, Elena Sartori† and Marco Tolotti‡

October 4, 2012

Abstract. The objective of this paper is to provide an analytical framework to study the whole process of diffusion of innovations, new products or ideas: we take into account knowledge transfer in a complex society, decisional process for adoption and key features in the spread of new technologies. For this purpose, we propose a probabilistic model based on an interacting population connected through new communication channels (such as social media) where potential adopters are linked with each other at different connection degrees. Our diffusion curve is the result of an emotion driven decision process following the awareness phase. Finally, we are able to recover stylized facts highlighted by the extant literature in the field.

Keywords: Awareness and adoption, diffusion of innovations, emotion driven decision making, hubs, random utility model, social media marketing.

JEL Classification Numbers: C63 - M31 - D83

*Department of Management, Ca' Foscari University of Venice, S.Giobbe - Cannaregio 873, I - 30121 - Venice, Italy; cinzia.colapinto@unive.it

†Department of Management, Ca' Foscari University of Venice, S.Giobbe - Cannaregio 873, I - 30121 - Venice, Italy; esartori@unive.it

‡**Corresponding author**. Department of Management, Ca' Foscari University of Venice, S.Giobbe - Cannaregio 873, I - 30121 - Venice, Italy; tolotti@unive.it

1 Introduction

Diffusion of innovations is a process related to the spread of a new idea, product or technology through the members of a social system. It is one of the more discussed topics in the fields of behavioral science, including anthropology, sociology and marketing. Two of the milestones widely recognized in this field are the models developed by Rogers (1962) and Bass (1969). The former identifies different steps through which the innovation decision process develops (knowledge, persuasion, decision, implementation, confirmation) and different adopters categories as ideal types, i.e., innovators (venturesome), early adopters (respectable), early majority (deliberate), late majority (skeptical) and laggards (traditional). The latter proposes an analytical model for the timing of initial purchase of new products; he assumes that the population consists of "innovators", who typically early adopt independently of the others, and "imitators", who are influenced in their choice by the media and by the number of previous buyers.

In the last decade audience fragmentation, consumers' empowerment, media convergence and technology development have affected both demand and supply sides changing consumption patterns and brand/product communication. The prominent role of new media and communication channels (internet, social media and social networks) calls for a revision of the classical diffusion models. We believe, in particular, that some minimal features must be included: (*i*) heterogeneous *connection degrees* influencing awareness; (*ii*) a *multilevel diffusion mechanism* recalling the original Rogers' diffusion stages; (*iii*) *behavioral motives* such as imitation and peers' pressure, driving the emotional decision process; (*iv*) *fragmentation* of the population of potential adopters.

The aim of this paper is to provide a general (and tractable) framework that encompasses all these ingredients. Our main result is a closed form equation driving the *adoption curve*. It is derived relying on large deviation techniques, as the limit of aggregate statistics describing a population of interacting agents. With this respect, we can say that our approach is inspired by Schelling's *Micromotives and Macrobehavior* (1978). Indeed, we firstly implement the single agent decision process, where characteristics such as connectivity degree, private willingness to adopt and imitation propensity are considered. Secondly, we choose appropriate state variables, identifying, at the micro level, *awareness* (knowledge) and *adoption*, and we study their time evolution, depending on agents' characteristics. Finally, a convergence result over the population size, i.e., when the number of actors grows to infinity, provides, at the macro level, the awareness and the adoption dynamics.

There are recent contributions that focus separately on some of the aspects highlighted above (connectivity, delayed adoptions, behavioral motives and heterogeneous agents). Among the others, Goldenberg *et al.* (2009) analyze the role of *hubs* (highly interconnected individuals in a social network) in the adoption process. Van Den Bulte and Joshi (2007) allow for *heterogeneity* in the population's characteristics, taking into account influential and non influential agents. New trends and directions within the marketing literature on the field are reviewed in Peres *et al.* (2010). A different strand of literature develops *agent-based models* (see Kiesling *et al.* (2011) for a survey); the aim of these papers is to model the micro-motives behind the decision process, thus introducing specific characteristics of single actors that are difficult to capture by means of a macro/aggregate equation. The price they pay is the loss of tractability and the need to use numerical methods in order to provide intuition about dynamics and equilibria. In Fanelli and Maddalena (2012), a delayed equation is used in order to model the gap between awareness and adoption. The task is pursued directly at the aggregate level, without addressing the decision process itself.

Our methodology allows us to take advantage of the micro-structure of the decision process (as in the agent-based approach) while maintaining tractability. On the other hand, relying on the closed form equation for the adoption curve, we recover some of the stylized facts suggested by previous empirical studies (see Goldenberg *et al.* (2009) and Van Den Bulte and Joshi (2007)) such as delayed adoptions compared to awareness trends and atypical (non S-shaped) adoption curves. Finally, thanks to the general setup for population's characteristics we are able to disentangle *hubs* and *influentials* (characters that are often overlapped); empirical evidence suggests in fact that their role in the diffusion process is different (see Goldenberg *et al.* (2009)). Our model provides an analytical support for this hypothesis.

The paper is organized as follows. Connectivity, the behavioral rationale of possible adopters and the two-stage diffusion mechanism are introduced in Section 2. In Section 3 we find the asymptotic aggregate awareness' and adoption's dynamics. Section 4 describes a baseline example, where we divide agents in subgroups with respect to their connectivity degree and their taste for innovation. Section 5 concludes the paper. Appendix A contains all technical proofs.

2 Awareness and adoption

We assume that an innovative product/service is launched through a new communication channel such as a social network. Nowadays, social media are more and more common in the marketing mix chosen to sustain the launch of

a new product/service. We consider a population of N actors linked within the network. They form the population of possible adopters.
To summarize information concerning agents' characteristics, we define *identity vectors* as follows.

Definition 2.1 (Identity vector)
Let $\theta_i = (\beta_i, p_i, q_i)$, for $i = 1, \dots, N$, be the identity vector of actor i, where

- $\beta_i \geq 0$, *measures the connection degree (connectivity);*
- $p_i \geq 0$, *represents the propensity to adopt (innovation coefficient);*
- $q_i \geq 0$, *represents the propensity to conform (imitation coefficient).*

θ_i, $i = 1, \dots, N$, *are i.i.d. random vectors with distribution η^θ.*

The identity vector collects the basic parameters useful to represent in a simple way the features highlighted in the introduction (network structure and emotional motives behind the adoption). We will see in the sequel how these parameters are used in order to characterize the awareness process and the decision problem behind adoption.

We aim at monitoring a *two-stage diffusion process*: awareness and adoption. Awareness shows whether an actor has been reached so far by the information. Being aware represents a necessary step for the agent to consider the possibility of adopting the innovation. We now describe the two levels diffusion process on a time window $[0, T]$ (possibly $T = +\infty$).

Definition 2.2 (Awareness and adoption) *We define the awareness and the adoption processes, respectively $\boldsymbol{x} = \{(x_1(t), \dots, x_N(t)), t \in [0, T]\}$ and $\boldsymbol{y} = \{(y_1(t), \dots, y_N(t)), t \in [0, T]\}$ where*

$$\begin{cases} x_i(t) = 1 & if\ agent\ \ i\ has\ been\ reached\ by\ the\ information\ by\ t; \\ x_i(t) = 0 & otherwise. \end{cases}$$

$$\begin{cases} y_i(t) = 1 & if\ agent\ \ i\ has\ adopted\ by\ t; \\ y_i(t) = 0 & otherwise. \end{cases}$$

We are interested in studying the joint evolution of $(\boldsymbol{x}, \boldsymbol{y})$, based on characteristics of the agents and of the network. In particular, we now propose suitable *transition intensities*[1] for the awareness and the adoption processes

[1]Transition intensities are the rates of probability of having a jump in one component of a jump process. For instance, the transition intensity $\lambda_i^x(t)$ for process $\boldsymbol{x}$ at time t is defined as

$$\lambda_i^x(t) = \lim_{h \to 0} \frac{1}{h} \mathbb{P}(x_i(t+h) \neq x_i(t) | \, \boldsymbol{x}(t), \boldsymbol{y}(t), \theta_i).$$

.

taking into account the whole network structure and the emotional motives driving the adoption process of the single agents.

The spread of information (awareness)

In the first stage we study the information exposure and its diffusion process. The awareness process spreads by contagion through the social network as follows

$$x_i: \; 0 \mapsto 1 \quad \text{with intensity} \quad \lambda_i^x = \exp(\beta_i \cdot m_N^x), \tag{1}$$

where

$$m_N^x = \frac{1}{N}\sum_{j=1}^{N} x_j \tag{2}$$

denotes the (cumulative) proportion of actors reached by the information[2] and where $\beta_i \geq 0$ is the parameter measuring the level of popularity of agent i in the network; we call it *connectivity*[3].

The rationale is the following: the more an actor is interconnected within the network, the higher is the probability of becoming aware of the innovation. Intensities as in (1) are suitable to model a viral effect as in classical contagion models (see Barucci and Tolotti (2012) or Dai Pra and Tolotti (2009)), taking into account heterogeneity in the identity vectors of the agents. In Section 4 we will see how to reproduce significant examples such as the *hub vs. non-hub world* analyzed in Goldenberg *et al.* (2009).

The emotion driven decision process (adoption)

Once reached by the information, the actor has to decide whether to adopt or not, optimizing his/her own utility function: any agent compares the utility of adoption and non adoption and chooses for the highest. To describe the decision process we take into account the emotional motives that guide the agent in his/her choice. Considering new and social media it is evident that engagement and emotional drivers play a crucial role. Thus we consider utility functions $U_{[i,t]} : \{0; 1\} \to \mathbb{R}$ for $i : 1, \ldots, N$ and $t \in [0, T]$ defined as follows.

$$U_{[i,t]}(y) = y \cdot \big[x_i(t) \cdot (p_i + q_i \, m_N^y(t) + \varepsilon_i)\big], \tag{3}$$

[2]Note that in principle we could use the statistics $\tilde{m}_N^x = \frac{1}{N-1}\sum_{j\neq i} x^j$, excluding x_i by the sum. Eventually, the marginal contribution of the single x_i will disappear at the limit, so the two formulations are basically equivalent.

[3]We can think of β_i as a popularity measure increasing with the *network degree* $d_i > 0$, i.e., the number of links departing from node i. We prefer to avoid a proper graph structure to maintain tractability.

where

$$m_N^y(t) = \frac{1}{N} \sum_{j=1}^{N} y_j(t) \tag{4}$$

is the proportion of adoptions occurred up to time t. This utility is time dependent and it is formed by three components: a private component, a social one and a random term[4]. In particular, the parameter $p_i \geq 0$, expresses the propensity to adopt (taste for innovation), $q_i \geq 0$ describes the measures the propensity to conform (imitation coefficient). Note that parameters p_i and q_i recall Rogers' classifications and Bass' modeling ideas. ε_i is a random term which introduces heterogeneity at the level of the utilities of the actors in the network. We assume $\varepsilon_i, i = 1, \ldots, N$ are i.i.d. with distribution function η^ε, where $\eta^\varepsilon(z) = \mathbb{P}(\varepsilon \leq z)$ attributes positive mass on the negative numbers. Note, moreover, that the identity vector $\theta_i = (\beta_i, p_i, q_i)$, as in Definition 2.1, and ε_i play two different roles. Although both θ_i and ε_i are agent's characteristics, θ_i is known at time zero and is related to the "actors' personality", whilst ε_i is revealed only at the time where the decision takes place and is "decision related".
Some remarks about utilities $U_{[i,t]}$ are needed.

- If $x_i(t) = 0$, $U_{[i,t]}(\cdot) = 0$. In other words, if the agent is not aware, his/her utility is zero.

- If $x_i(t) = 1$, $U_{[i,t]}(1) = p_i + q_i\, m_N^y(t) + \varepsilon_i$, which is increasing in both the innovation and imitation coefficient and in the proportion of adoptions already occurred. Note, moreover, that $U_{[i,t]}(1)$ could be negative, because of ε_i.

- $U_{[i,t]}(0) = 0$. Utility of non-adoption is zero. This means that, in deciding whether to adopt, any agent compares $U_{[i,t]}(1)$ with zero[5].

- In facing an instantaneous utility, actors are somewhat myopic. This assumption, although rather simplistic, can be justified in the context of new media: agents acting in such a context experience very fast and emotional decision-making processes (see Berger and Milkman (2012)). With this respect, the contribution of future utility can be neglected.

[4]This approach is referred to as *random utility model*. See, for instance, Brock and Durlauf (2001) for more details.

[5]Adoption based on utilities $U_{[i,t]}$ can be seen as obeying a threshold model. Given p_i, q_i and ε_i, as soon as $m_N^y(t)$ is large enough, it becomes worth for agent i to adopt. With this respect, ε_i can be seen as a term introducing a random (idiosyncratic) threshold level for adoption.

Following Blume and Durlauf (2003) and Barucci and Tolotti (2012) we assume that agents may decide to adopt at random times according to utilities as in (3). This is in line with Rogers' seminal idea of *implementation* and *confirmation* (see Rogers (1962)). The actor can adopt the new product or reject it and decide to postpone the decision to a future time. Under these assumptions, transition rates for the adoption process are described in the following proposition.

Proposition 2.3 *Consider agents whose utilities are given by* (3)*. Assume, moreover, that agents are randomly asked to adopt at exponentially distributed random times with parameter* 1*. Then, the transition intensities of process* $\boldsymbol{y}$*, defined in definition 2.2, are*

$$\lambda_i^y(t) = x_i(t) \cdot (1 - \eta^\varepsilon(-p_i - q_i m_N^y(t))). \tag{5}$$

Proof. See Appendix A.

3 Awareness' and adoption's dynamics

We have formalized the decision process behind the adoption and the dynamics of awareness through the social network. In particular, we have provided transition intensities for the processes $\boldsymbol{x}$ and $\boldsymbol{y}$. We are ready to state the main result of this paper. It provides the proper dynamics of the two-stage system (awareness and adoption) when the number of actors goes to infinity.

Theorem 3.1 *Consider the identity vectors $\theta_i = (\beta_i, p_i, q_i)$, $i = 1, \ldots, N$, as in Definition 2.1 and the Markov process $(\boldsymbol{x}, \boldsymbol{y})$ characterized by the following transition rates:*

$$\begin{array}{lll} x_i = 0 \mapsto x_i = 1 & \text{with intensity} & \lambda_i^x = \exp\{\beta_i m_N^x\}, \\ y_i = 0 \mapsto y_i = 1 & \text{with intensity} & \lambda_i^y = x_i \cdot \big(1 - \eta^\varepsilon(-p_i - q_i m_N^y)\big). \end{array} \tag{6}$$

Assume, moreover, that at time $t = 0$, $x_i(0) = y_i(0) = 0$, for $i = 1, \ldots, N$. Then, for $N \to \infty$, $(m_N^x(t), m_N^y(t))$ weakly converges to the pair (m_t^x, m_t^y), defined as

$$m_t^x = \int m_t^x(\theta') d\eta^\theta(\theta'); \quad m_t^y = \int m_t^y(\theta') d\eta^\theta(\theta'), \tag{7}$$

where $\theta' = (\beta', p', q')$, $(m_0^x, m_0^y) = (0, 0)$ and

$$\begin{cases} \dot{m}_t^x(\theta) = (1 - m_t^x(\theta)) \exp\{\beta\, m_t^x\} \\ \\ \dot{m}_t^y(\theta) = (m_t^x(\theta) - m_t^y(\theta))(1 - \eta^\varepsilon(-p - q\, m_t^y)). \end{cases} \tag{8}$$

Proof. See Appendix A.

Some remarks are needed. First of all notice that this result is asymptotic. It describes the behavior of a population of infinite (heterogeneous) agents. Finite volume approximations could be computed, but this is out of the scope of this paper. Indeed, equations (7)-(8) provide a generalization of the standard Bass differential equation driving the adoption curve (see Example 3.2 below for a comparison with the seminal Bass model).

Example 3.2 (Back to Bass) *The original Bass adoption curve (see Bass (1969)) can be recovered as a special case of equation* (8). *Just consider the following specifications:*

- $\boldsymbol{x} \equiv \mathbf{1}$, *meaning that only adoption is considered, without treating any network effect;*
- $p_i = p$, $q_i = q$ *for all* $i = 1, \dots, N$, *dealing with no heterogeneity;*
- *concerning the noise terms, put* $\varepsilon_i \sim Unif[-1, 1]$, *i.e., assume that the additional term added to the evaluation of private and social utility is uniformly distributed across actors in the community.*

Then, it is easy to see that (8) *reduces to*

$$\dot{m}_t^y = (1 - m_t^y) \cdot (p + q\, m_t^y)\,, \tag{9}$$

with initial condition $m_0^y = 0$. *This equation gives rise to the Bass adoption curve, where* m_t^y *represents the cumulative proportion of potential buyers who have already adopted.*

Equations (8) are very general: they apply in principle to a continuum of heterogeneous agents. To help the intuition and to recover some features analyzed by recent empirical literature, in Section 4 we apply this result to the case of a population of actors belonging to (few) different groups.

4 A significant case: hubs and innovators

We consider the case where the population can be split into four subgroups, taking into account the degree of connectivity of each individual and his/her propensity for innovation. In particular, we consider two connectivity levels (*hubs* and *non-hubs*) and two innovativeness levels (*innovators* and *followers*). Therefore, we have innovative hubs, follower hubs, innovative non-hubs and follower non-hubs, each group characterized by a different identity vector (see Table 1). In Table 1 we denote by $\alpha_h \in (0,1)$ and $\alpha_I \in (0,1)$, respectively, the proportion of hubs and of innovators in the whole population. Clearly, $\alpha_n = 1 - \alpha_h$ and $\alpha_F = 1 - \alpha_I$.

	Innovator (I)	Follower (F)	
hub (h)	$\theta_{hI} = (\beta_h, p_I, q_I)$	$\theta_{hF} = (\beta_h, p_F, q_F)$	α_h
non-hub (n)	$\theta_{nI} = (\beta_n, p_I, q_I)$	$\theta_{nF} = (\beta_n, p_F, q_F)$	$\alpha_n = 1 - \alpha_h$
	α_I	$\alpha_F = 1 - \alpha_I$	

Table 1: Subgroups, relative identity vectors and their proportions.

We assume, furthermore, that ε_i are uniformly distributed over $[-1,1]$ as in Example 3.2. Under these specifications, we obtain the following result.

Corollary 4.1 *Let consider the system of four sub-populations θ_k, with $k \in \{hI, hF, nI, nF\}$ as defined in Table 1. Assume, moreover, that ε_i are uniformly distributed over $[-1,1]$.*
Then awareness and adoption (m_t^x, m_t^y) evolve according to the following differential system:

$$\begin{cases} \dot{m}_t^x = \alpha_h\alpha_I \cdot m_t^x(\theta_{hI}) + \alpha_h\alpha_F \cdot m_t^x(\theta_{hF}) + \alpha_n\alpha_I \cdot m_t^x(\theta_{nI}) + \alpha_n\alpha_F \cdot m_t^x(\theta_{nF}) \\ \dot{m}_t^y = \alpha_h\alpha_I \cdot m_t^y(\theta_{hI}) + \alpha_h\alpha_F \cdot m_t^y(\theta_{hF}) + \alpha_n\alpha_I \cdot m_t^y(\theta_{nI}) + \alpha_n\alpha_F \cdot m_t^y(\theta_{nF}) \end{cases}$$

where, for $k \in \{hI, hF, nI, nF\}$,

$$\begin{cases} \dot{m}_t^x(\theta_k) = (1 - m_t^x(\theta_k)) \exp\{\beta_k \, m_t^x\} \\ \\ \dot{m}_t^y(\theta_k) = (m_t^x(\theta_k) - m_t^y(\theta_k))(p_k + q_k \, m_t^y). \end{cases} \tag{10}$$

Note that equations (10) are the exact counterpart of (8) in the case of four sub-groups of homogeneous actors (see Table 1); in particular, they describe how the awareness and the adoption evolve in such a population. Because of the introduction of coupled dynamics and the presence of heterogeneous actors, they cannot be analytically solved, thus, in order to discuss some interesting stylized facts, we provide some numerical computations.

In this case example, parameters are the identity vectors $\theta_k = (\beta_k, p_k, q_k)$, with $k \in \{hI, hF, nI, nF\}$, and the population distribution $\alpha = (\alpha_h, \alpha_I)$. For the numerical simulation we choose the following parameters based on the extant literature:

- $\beta_h = 8 \times \beta_n$ (assumption taken from Goldenberg *et al.* (2009));

- $q_I = 0$ (extreme case: no imitative component in innovator actors);
- $p_F = 0$ (extreme case: no innovative component in follower actors);
- $\alpha_h = 0.1$ (small proportion of hubs);
- $\alpha_I = 0.25$ (assumption taken from Van Den Bulte and Joshi (2007)).

	Innovator (I)	Follower (F)	
hub (h)	$\theta_{hI} = (4.0, 5, 0)$	$\theta_{hF} = (4.0, 0, 1)$	$\alpha_h = 0.1$
non-hub (n)	$\theta_{nI} = (0.5, 5, 0)$	$\theta_{nF} = (0.5, 0, 1)$	$\alpha_n = 0.9$
	$\alpha_I = 0.25$	$\alpha_F = 0.75$	

Table 2: Values of the parameters used in Figures 1 and 2.

The particular values chosen are synthesized in Table 2.

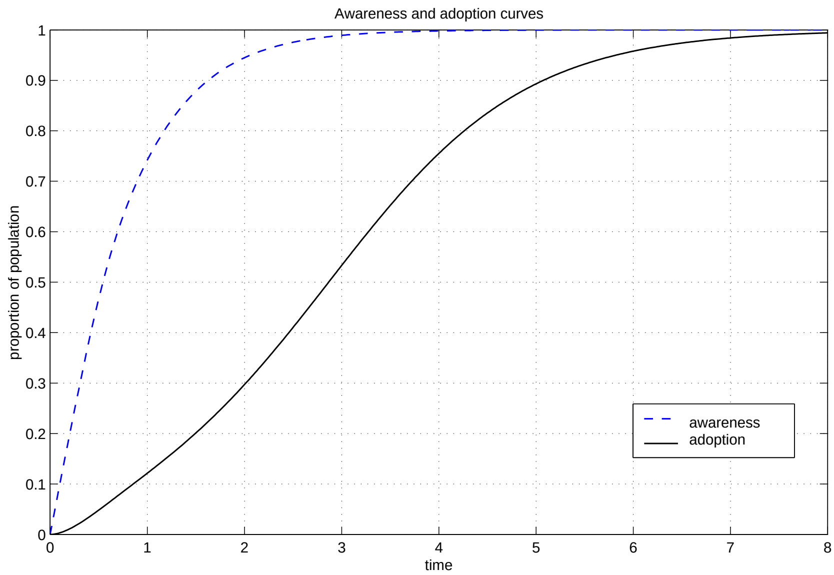

Figure 1: Different shapes for awareness' and adoption's curves

Figure 1 shows that the adoption (as predictable) is delayed and can have a different shape compared to the awareness, which seems to be faster in this case. Social media are useful for awareness campaigns as they are able to find and target a specific group of people, implying furthermore, a social aspect (see Harris and Rae (2009)). They mainly affect the first stage of the process. Observing diffusion data of such messages on Facebook or Youtube, one can end up with steep (increasing and concave) curves for the awareness similar to what represented in Figure 1. Actually, the speed of knowledge on the social network does not necessarily imply that the adoption should grow exponentially fast.

Figure 2 shows the different roles of innovator hubs and follower hubs in the adoption process. We find that, as conjectured and found in Goldenberg *et al.* (2009), the formers push the speed of adoption at the very beginning of the process, giving an initial positive signal to the network, the follower hubs make the adoption rates more persistent in time, having a strong impact on the total number of adoptions. Moreover, we show that the adoption process is not necessarily monotone in time, as conjectured and empirically found in Van Den Bulte and Joshi (2007).

5 Conclusions

We proposed a framework for modeling diffusion of innovations through the *emotional web*, taking into account, from one hand, a two-stage process (awareness and adoption) and, from the other, the behavioral motives driving the adoption process.

We studied the process of aggregate awareness and adoption finding macroscopic equations that describe how these processes evolve over time.

Our results are twofold. From one side we enriched models *à la* Bass with Roger's paradigm of a *complex* adoption process (synthesized here by the two-stage and emotion driven adoption process). From the other hand, our micro founded approach enabled us to explain some stylized facts observed in previous empirical works, such as Goldenberg *et al.* (2009) and Van Den Bulte and Joshi (2007). Indeed, we discussed the different shapes of the awareness' and adoption's curves, confirming the need to monitor both of them to better describe the whole diffusion process. We have, also, proposed a significant example, where the disentangled effects of *innovator hubs* and *follower hubs* can be monitored; eventually, we showed in certain cases the emergence of a non S-shaped adoption curve.

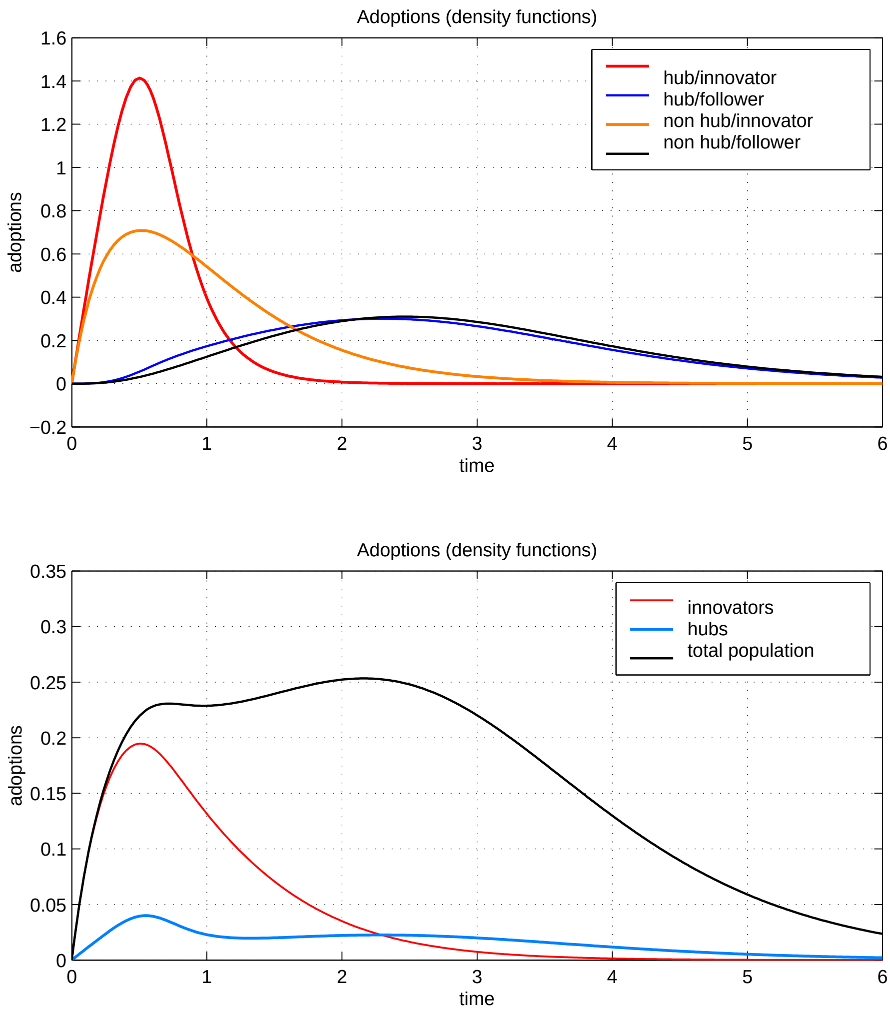

Figure 2: The effects of *innovator hubs* and *follower hubs* and the non monotonicity in the adoption curve.

A Proofs

Proof of Proposition 2.3

Consider an actor (say i), who, at time t, has been reached by the information about the innovation ($x_i(t) = 1$), but has not yet adopted ($y_i(t) = 0$). We call τ the next random time at which agent i is asked to decide whether to adopt or not. We are interested in computing the (rate of) probability of adopting on the interval $(t, t+h]$, conditioned above current information. Note that the event $\{$*to adopt on* $(t, t+h]\}$ is the intersection of two events, respectively, $\{\tau \in (t, t+h]\}$ and $\{U_{[i,\tau]}(1) \geq U_{[i,\tau]}(0)\}$. Concerning the former, the arrival time is independent from the history of the process. Hence

$$\mathbb{P}\big(\tau \in (t, t+h] | \, \boldsymbol{x}(t), \boldsymbol{y}(t), \theta_i\big) = \mathbb{P}\big(\tau \in (t, t+h]\big) = 1 - e^{-h}.$$

The probability of the latter, conditioned on the event $\{\tau \in (t, t+h]\}$, can be computed as follows.

$$\mathbb{P}\big(U_{[i,\tau]}(1) \geq U_{[i,\tau]}(0) | \, \boldsymbol{x}(t), \boldsymbol{y}(t), \theta_i, \tau \in (t, t+h]\big) =$$

$$x_i(\tau) \cdot \mathbb{P}\big(p_i + q_i m_N^y(\tau) + \varepsilon_i \geq 0 \big| \tau \in (t, t+h]),$$

where the term $x_i(\tau)$ makes 0 the probability of adoption for a non-informed agent. Therefore we have

$$\lim_{h \to 0} \frac{1}{h} \mathbb{P}\big(y(t+h) = 1 | \, \boldsymbol{x}(t), \boldsymbol{y}(t), \theta_i\big) =$$

$$= \lim_{h \to 0} \frac{1}{h} \, \big[(1 - e^{-h}) \cdot x_i(\tau) \cdot \mathbb{P}\big(p_i + q_i m_N^y(\tau) + \varepsilon_i \geq 0 | \tau \in (t, t+h]\big)\big] =$$

$$= x_i(t) \cdot \big(1 - \eta^\varepsilon(-p_i - q_i m_N^y(t))\big),$$

where η^ε is the distribution of ε_i. The last equality follows from the fact that, the probability of having two jumps of the processes $(\boldsymbol{x}, \boldsymbol{y})$ shrinks when the time interval tends to zero. Hence, $x_i(\tau) = x_i(t)$ and $m_N^y(\tau) = m_N^y(t)$. ∎

Proof of Theorem 3.1

We split the main body of the proof into two technical lemmas. The first one states a law of large numbers for a sequence $(\rho_N)_N$ of suitable empirical measures (see equation (12)). The second one characterizes the unique limiting (deterministic) measure Π_* such that

$$\lim_N \rho_N \to \Pi_*.$$

In particular, it provides a Fokker-Plank (forward) equation useful to describe the time evolution of Π_*.

Before stating and proving the two lemmas, we refresh a couple of facts about the process $(\boldsymbol{x}, \boldsymbol{y})$ and we introduce some useful notations. Conditioned on a realization of the identity vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ (see Definition 2.1), $(\boldsymbol{x}, \boldsymbol{y})$ evolves as a continuous time Markov chain on the state space $\{0,1\}^{2N}$ with infinitesimal generator acting on functions $h : \{0,1\}^{2N} \to \mathbb{R}$ as follows

$$\begin{aligned} G^N_{[\boldsymbol{\theta}]}\, h(\boldsymbol{x}, \boldsymbol{y}) = & \sum_{i=1}^{N} (1 - x_i)\lambda_i^x \left[h\left(\boldsymbol{x}^i, \boldsymbol{y}\right) - h\left(\boldsymbol{x}, \boldsymbol{y}\right) \right] + \\ & \sum_{j=1}^{N} (1 - y_i)\lambda_j^y \left[h\left(\boldsymbol{x}, \boldsymbol{y}^j\right) - h\left(\boldsymbol{x}, \boldsymbol{y}\right) \right] , \end{aligned} \tag{11}$$

where $\boldsymbol{x}^i := (x_1, \dots, 1 - x_i, \dots, x_N)$, analogously $\boldsymbol{y}^j$ and where λ_i^x and λ_j^y are respectively defined in (1) and (5). Note that the terms $(1 - x_i)$ and $(1 - y_i)$ inhibit transitions of the kind $1 \mapsto 0$.

We denote by $x_i[0,T]$ (resp. $y_i[0,T]$) the trajectory of x_i (resp. y_i) on $[0,T]$ and with $\mathcal{D}([0,T])$ the Skorohod space of right continuous, piecewise constant functions defined on the interval $[0,T]$. Finally, define the (random) empirical measure.

$$\rho_N(\boldsymbol{x}([0,T]), \boldsymbol{y}([0,T]), \boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\{x_i[0,T], y_i[0,T], \theta_i\}} \,. \tag{12}$$

$\rho_N \in \mathcal{M}_1$, where $\mathcal{M}_1$ denotes the space of probability measures on $\mathcal{D}^2([0,T]) \times \mathbb{R}^3$ endowed with the weak convergence topology. Note, moreover, that $m_N^x(t)$ and $m_N^y(t)$ are empirical averages of the form

$$\frac{1}{N} \sum_{i=1}^{N} f(x_i[0,T], y_i[0,T], \theta_i) =: \int f \, d\rho_N.$$

Therefore, we now provide a law of large numbers for the sequence $(\rho_N)_N$ (Lemma A.1), characterizing the limiting measure Π_* (Lemma A.2).

Lemma A.1 *There exists a unique* $\Pi_* \in \mathcal{M}_1$ *such that*

$$\rho_N \to \Pi_* \qquad \textit{almost surely}$$

in the weak topology.

Lemma A.2 *It holds* $\Pi_* = \Pi_*^\theta \otimes \eta^\theta$*, where* η^θ *has been defined in Definition 2.1 and where* Π_*^θ *can be characterized as follows. Let* $\pi_t^x(\theta) := \Pi_*^\theta(x(t) = 1)$ *and* $\pi_t^y(\theta) := \Pi_*^\theta(y(t) = 1)$*, then* $(\pi_t^x(\theta), \pi_t^y(\theta))$ *is the unique solution of*

$$\begin{cases} \frac{\partial}{\partial t}\pi_t^x(\theta) = (1 - \pi_t^x(\theta)) \exp\left\{\beta \int \pi_t^x(\theta') d\eta^\theta(\theta')\right\} \\ \\ \frac{\partial}{\partial t}\pi_t^y(\theta) = (\pi_t^x(\theta) - \pi_t^y(\theta)) \left(1 - \eta^\varepsilon\left(-p - q\int \pi_t^y(\theta') d\eta^\theta(\theta')\right)\right) \\ \\ \pi_0^x(\theta) = \pi_0^y(\theta) = \delta_0 \end{cases} \tag{13}$$

for $t \in [0, T]$ *and for any fixed* $\theta = (\beta, p, q)$*.*

The proofs of these results are postponed to the end of the section. Assuming the validity of the two lemmas, we can prove Theorem 3.1.
Note that $m_N^x(t) = \int x\Pi_t(\rho_N)(dx, dy, d\theta)$ and $m_N^y(t) = \int y\Pi_t(\rho_N)(dx, dy, d\theta)$. Therefore, as a corollary of Lemma A.1, we have that $(m_N^x(t), m_N^y(t))$ weakly converges to the couple (m_t^x, m_t^y), where

$$m_t^x = \int x\, \Pi_t(\Pi_*)(dx, dy, d\theta)\,; \quad m_t^y = \int y\, \Pi_t(\Pi_*)(dx, dy, d\theta)$$

and $\Pi_t(\Pi_*)$ is the time-t projection of Π_*.
It remains to show that (m_t^x, m_t^y) solve (7) and (8). These equations are derived relying on Lemma A.2. Indeed, note that $m_t^x = \int m_t^x(\theta') d\eta^\theta(\theta')$, where $m_t^x(\theta) = \int x\, \Pi_t(\Pi_*^\theta)(dx, dy) = \pi_t^x(\theta)$ (the same for m_t^y). Therefore, equations (7) and (8) immediately follow from (13). ■

Proof of Lemma A.1

To prove the law of large numbers, we need to define a new function F, $F : \mathcal{M}_1 \to \mathbb{R}$, where $\mathcal{M}_1$ is the space of probabilities on $\mathcal{D}^2([0, T]) \times \mathbb{R}^3$. This function is crucial to specify the good rate function of the large deviation principle used to derive the law of large numbers. To this aim, we need to introduce an auxiliary process (x, y) on the state space $\{0; 1\}^2$ evolving according to the following rates of transition. For $Q \in \mathcal{M}_1$,

$$x : 0 \mapsto 1 \quad \textit{with intensity} \quad (1 - x(t)) \cdot \lambda_Q^x(t),$$
$$\lambda_Q^x(t) = \exp\left\{\beta \int v(t)\, Q(dv[0, T], dw[0, T], d\theta)\right\};$$
$$y : 0 \mapsto 1 \quad \textit{with intensity} \quad (1 - y(t)) \cdot \lambda_Q^y(t),$$
$$\lambda_Q^y(t) = x(t) \left(1 - \eta^\varepsilon\big(-p - q\int w(t)\, Q(dv[0, T], dw[0, T], d\theta)\big)\right).$$

We denote by τ_x (resp. τ_y) the jump time of the process x (resp. y). Define $F(Q)$ as

$$F(Q) = \mathbb{E}^Q \left[f(x[0,T], y[0,T], \theta) \right], \tag{14}$$

where

$$\begin{aligned} f(x[0,T], y[0,T], \theta) = & \int_0^T (1 - x(t)\big(1 - \lambda_Q^x(t)\big) dt + x(T) \ln \big(\lambda_Q^x(\tau_x^-)\big) \\ & + \int_0^T (1 - y(t)\big(1 - \lambda_Q^y(t)\big) dt + y(T) \ln \big(\lambda_Q^y(\tau_x^-)\big). \end{aligned}$$

Let P_N be the law of the process induced by the generator (11) and $\mathcal{P}_N$ the probability distribution of $\rho_N(\boldsymbol{x}[0,T], \boldsymbol{y}[0,t], \boldsymbol{\theta})$ under P_N. Having defined a suitable function F, we can apply Theorem 1 of Dai Pra and Tolotti (2009). That result allows to characterize a large deviation principle for the sequence of measures $(\mathcal{P}_N)_N$, in the following sense. Define

$$I(Q) := H(Q|\,W) - F(Q), \tag{15}$$

where W is the law of the process induced by the generator (11) assuming independence (i.e., $\lambda_i^x = \lambda_i^y = 1$ for all $i = 1, \dots, N$) and where $H(Q|\,W)$ denotes the relative entropy of Q with respect to W. Then $(\mathcal{P}_N)_N$ satisfies a large deviation principle with good rate function I. Therefore,

$$\mathbb{P}\big(d(\rho_N, \Pi_*) \geq \varepsilon\big)$$

converges to zero with exponential rate in N, where $d(\cdot, \cdot)$ is any metric that induces the weak topology on $\mathcal{M}_1$. The law of large numbers follows from this result, applying the Borel-Cantelli lemma. ■

Proof of Lemma A.2

The measure Π_* as defined in Lemma A.1 is the unique element of $\mathcal{M}_1$ such that $\Pi_* = \Pi^\theta \otimes \eta^\theta$, where Π^θ is the law of the Markov process on $\{0;1\}^2$ with initial distribution $\delta_0 \otimes \delta_0$ and time-dependent generator $L_t^{[\Pi_*;\,\theta]}$ defined as

$$\begin{aligned} L_t^{[\Pi_*;\,\theta]} f(x,y) = (1-x) \cdot e^{\beta \int x(t)\, \Pi_*(dx[0,T], dy[0,T], d\theta)} \cdot \big(f(1-x, y) - f(x,y)\big) + \\ x \cdot (1-y) \cdot e^{p + q \int y(t)\, \Pi_*(dx[0,T], dy[0,T], d\theta)} \cdot \big(f(x, 1-y) - f(x,y)\big) \end{aligned} \tag{16}$$

(see Theorem 2 of Dai Pra and Tolotti (2009)). Since the marginals of a Markov process are solutions of the corresponding *forward equation*, defining

$\pi_t := \Pi_t\left(\Pi^\theta\right)$, the $t-$projection of Π^θ, we have that $\dot{\pi}_t = \mathcal{L}_t^{[\Pi_*;\,\theta]}$ where $\mathcal{L}_t^{[\Pi_*;\,\theta]}$ is the adjoint of $L_t^{[\Pi_*;\,\theta]}$:

$$\begin{aligned}(\mathcal{L}_t^{[\Pi_*;\,\theta]}q)(x,y) = x \cdot e^{\beta \int x(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)} \cdot q(1-x,y) - \\ (1-x) \cdot e^{\beta \int x(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)} \cdot q(x,y) + \\ x \cdot y \cdot e^{p+q \int y(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)} \cdot q(x,1-y) - \\ x \cdot (1-y) \cdot e^{p+q \int y(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)} \cdot q(x,y).\end{aligned} \tag{17}$$

More specifically,

$$\begin{aligned}
\dot{\pi}_t(1,1) &= e^{\beta \int x(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}\pi_t(0,1) + e^{p+q \int y(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}\pi_t(1,0);\\
\dot{\pi}_t(1,0) &= e^{\beta \int x(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}\pi_t(0,0) - e^{p+q \int y(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}\pi_t(1,0);\\
\dot{\pi}_t(0,1) &= -e^{\beta \int x(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}\pi_t(0,1);\\
\dot{\pi}_t(0,0) &= -e^{\beta \int x(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}\pi_t(0,0).
\end{aligned}$$

Note that $\pi_t(0,1) \equiv 0$, since $\pi_0(0,1) = 0$ and $\dot{\pi}_t(0,1)/\pi_t(0,1) \leq 0$. Therefore, the unique non-negative solution is the null one. Put now $\pi_t^x = \pi_t(1,1) + \pi_t(1,0)$ and $\pi_t^y = \pi_t(1,1)$, then it is easy to see that

$$\begin{cases} \dot{\pi}_t^x = e^{\beta \int x(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}(1-\pi_t^x) \\ \dot{\pi}_t^y = e^{p+q \int y(t)\, \Pi_*(dx[0,T],dy[0,T],d\theta)}(\pi_t^x - \pi_t^y) \\ \pi_0^x = \pi_0^y = 0 \end{cases}$$

which are exactly equations (13). ∎

References

Barucci, E. and Tolotti, M. (2012) Social interaction and conformism in a random utility model, *Journal of Economic Dynamics & Control*, 36(12): 1855-1866.

Bass, F. M. (1969) A new product growth for model consumer durables, *Management Science*, 15(5): 215-227.

Berger, J. and Milkman, K. L. (2012) What makes online content viral?, *Journal of Marketing Research*, 49(2): 192-205.

Blume, L. and Durlauf, S. (2003) Equilibrium concepts for social interaction models, *International Game Theory Review*, 5(3): 193-209.

Brock, W. and Durlauf, S. (2001) Discrete choice with social interactions, *Review of Economic studies*, 68(2): 235-260.

Dai Pra, P. and Tolotti, M. (2009) Heterogeneous credit portfolios and the dynamics of the aggregate losses, *Stochastic Processes and their Applications*, 119(9): 2913-2944

Fanelli, V. and Maddalena, L. (2012) A time delay model for the diffusion of a new technology, *Nonlinear Analysis: Real World Applications*, 13: 643-649.

Goldenberg, J., Han, S., Lehmann, D. R. and Hong, J. W.(2009) The role of hubs in the adoption process, *Journal of Marketing*, 73(2): 1-13.

Harris, L. and Rae, A. (2009) Social networks: the future of marketing for small business. *Journal of Business Strategy*, 30(5): 24-31.

Katz, E. and Lazarsfeld, P. F. (1955) *Personal Influence*, New York: The Free Press.

Kiesling, E., Günther, M., Stummer, C. and Wakolbinger, L. M. (2011) Agent-based simulation of innovation diffusion: a review, *Central European Journal of Operations Research*, Physica Verlag: 1-48.

Norton, J. A. and Bass, F. M. (1987) A diffusion theory model of adoption and substitution for successive generations of high-technology products, *Management Science*, 33(9): 1069-1086.

Parker, P. and Gatignon, H. (1994) Specifying competitive effects in diffusion models: An empirical analysis, *International Journal of Research in Marketing*, 11(1): 17-39.

Peres R., Muller, E. and Mahajan V. (2010) Innovation diffusion and new product growth models: A critical review and research directions, *International Journal of Research in Marketing*, 27: 91-106.

Rogers, E. M. (1962) *Diffusion of Innovations*, New York: The Free Press.

Schelling, T. C. (1978) *Micromotives and Macrobehavior*, New York: Norton.

Van den Bulte, C. and Joshi, Y. V. (2007) New product diffusion with influentials and imitators, *Marketing science*, 26(3): 400-421.