

UNIVERSITY OF TRENTO



SCHOOL OF SOCIAL SCIENCES

PHD PROGRAM IN ECONOMICS AND MANAGEMENT

DOCTORAL THESIS

PEER INFLUENCE AND  
KNOWLEDGE MANAGEMENT IN  
COMMUNITIES  
OF USERS AND PRACTICES

DISSERTATION SUBMITTED TO THE DOCTORAL SCHOOL  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY  
IN ECONOMICS AND MANAGEMENT

*Author:*

Gianna Giudicati

*Supervisor:*

Prof. Massimo Riccaboni

*Co-Supervisor:*

Prof. Enrico Zaninotto

April, 2012



## SUPERVISOR:

PROF. MASSIMO RICCABONI

IMT Lucca

## CO-SUPERVISOR:

PROF. ENRICO ZANINOTTO

University of Trento

## EVALUATION COMMITTEE

PROF. KELD LAURSEN

Copenhagen Business School

PROF. STEFANO SCHIAVO

University of Trento

PROF. GIANMARIO VERONA

Bocconi University

## FINAL EXAMINATION COMMITTEE

PROF.SSA LAURA MAGAZZINI

University of Verona

PROF. FEDERICO MUNARI

University of Bologna

PROF. MARCO ZAMARIAN

University of Trento



# Acknowledgements

My first thanks go to my supervisor, Prof. Massimo Riccaboni, for his continual and unconditional help and patience, and for having transmitted to me his love for research. I also express my gratitude to my co-supervisor, Prof. Enrico Zaninotto, for his support, advice and encouragement during these three years. Special thanks go to Profs. Lee Fleming, Francesca Gino and John Padgett for their valuable suggestions and comments, which improved the quality of my work. I would also like to thank the examiners Prof.s Keld Laursen, Stefano Schiavo and Gianmario Verona for their very useful feedback. I also have a debt of gratitude to the “Rockers” for their useful suggestions and advice.

Warm thanks to all my friends in Cifrem, especially Ivana Catturani, Simone Cerroni, Kiryl Haiduk and Lorenza Lorenzetti. Special thanks also to my second family in Boston, especially Italia and Elaine, and my friends Sam, Andrea, Aurora and especially Roberto for their smiles and help overcoming difficult moments.

I am forever grateful to my mom, my grandparents, my godmother Antonella, my best friend Silvia, my sweet flatmates Jessica, Chiara and Nadia, and all my friends for their support and especially their patience.

Lastly, a very special thank you goes to Nicola, for his unconditional support and love.



To SABINA,  
who always believed in me.





# Contents

<b>Abstract</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
<b>2 Co-experience Network Dynamics and Customer Retention: Lessons from the Dance Floor</b>	<b>9</b>
2.1 Introduction . . . . .	10
2.2 Conceptual framework . . . . .	12
2.3 Data and methodology . . . . .	16
2.4 Descriptive statistics . . . . .	20
2.5 Results . . . . .	24
2.6 Concluding discussion . . . . .	30
<b>3 Homophily and Peer Influence in Customer Co-presence Net- works</b>	<b>33</b>
3.1 Introduction . . . . .	33
3.2 Literature review and research propositions . . . . .	36
3.3 Contextual framework and methodological approach . . . . .	40
3.3.1 Data sources . . . . .	40
3.3.2 The co-presence network . . . . .	41
3.3.3 Influence in co-presence networks . . . . .	42
3.3.4 Influence, homophily and environmental effects . . . . .	44

---

3.4	Results . . . . .	45
3.4.1	Preliminary findings . . . . .	47
3.4.2	The effect of peer-influence on renewal decisions . . . . .	50
3.4.3	Confounding effects: customer profile and product features . . . . .	55
3.4.4	Timely peer-influence . . . . .	55
3.4.5	Observational learning attenuates overestimation of attendance . . . . .	57
3.5	Concluding Remarks . . . . .	61
<b>4</b>	<b>Drilling Wells and Killing Wells through Knowledge Networks</b>	<b>65</b>
4.1	Introduction . . . . .	65
4.2	Literature review . . . . .	69
4.3	Research questions and contextual framework . . . . .	74
4.4	Empirical framework . . . . .	79
4.5	Results . . . . .	83
4.5.1	Representativeness in the domain of experts . . . . .	83
4.5.2	Proactivity in the domain of information exchange . . . . .	87
4.5.3	Presence and individual overlap . . . . .	90
4.5.4	Activation of experts . . . . .	92
4.5.5	Retrieval . . . . .	95
4.5.6	Structure: small worlds and structural holes . . . . .	98
4.6	Discussion and conclusions . . . . .	104
<b>5</b>	<b>Conclusions</b>	<b>107</b>
<b>A</b>	<b>Propensity Score and Coarsened Exact Matching</b>	<b>111</b>
	<b>Bibliography</b>	<b>115</b>

# Abstract

Knowledge is highly dispersed in companies, customer networks and, more generally, in markets. Various forms of knowledge influence collective results: experiential knowledge derived from shared socialization, practical knowledge, and scientific knowledge. This work aims at better understanding of how these types of knowledge affect collective results in communities of customers and of practice. In particular, collective results are interpreted on one hand as individual commitments in communities of customers, and, on the other, companies' success in rapidly solving problematic urgencies through communities of practice. First, the effect of both experience and socialization is examined in the service sector, in order to investigate how the social structures of communities of customers influence commitment and retention. Second, communities of practice operating in high-risk sectors are analyzed to look for improvements in their structure by aiding the recombination of dispersed knowledge and more rapid access to those skills and competences required to solve urgent problem-solving processes.

Various methods and tools are used to deal with these issues. In particular, cross network analyses and new methodological approaches, originally applied in the field of social ethology, are applied to study reality-mined data collected by radio-frequency identification (RFID) tags. In addition, a set of recently developed non-parametric methods is devised to deal with new methodological problems arising in studying socialization in networks. Recent debates in

## ABSTRACT

---

the marketing literature have emphasized methodological criticalities in discriminating between peer influence, homophily, and other confounding factors in real-world social settings. A whole chapter is devoted to a methodological contribution to these debates.

# Chapter 1

## Introduction

Knowledge is highly dispersed in companies, customer networks and, more generally, in markets. The fast recombination of dispersed knowledge is crucial in decision-making and problem-solving processes. The idea that all things (not just novelties) are the result of recombinations can be traced back to the pre-Socratic philosopher Empedocles (ca. 490-430 BC) (Kirk et al. 1983), who proposed that the components of recombination were air, fire, earth, and water; similar ideas have been identified in Aristotelian thought and in other ancient societies (Russell and Marić 1945, Burnyeat 1981, Taylor 1990). The concept has continued to influence Western thought and can be recognized in the works of John Locke, Adam Smith, and Henri Poincaré: new ideas are formed by a combination of old ideas, which may be either empirical (that is, experienced) or theoretical. In particular, the “growth of knowledge” principle proposed by Smith is based on the connections of existing elements: the generative process of recombinant search exploits the connections between elements rather than the elements *per se* (Smith 1982, Loasby 2002). The importance of recombination has often resurfaced often in the last century of management literature. Schumpeter (1939) proposed three steps for the recombinant search process: identifying the simple components of an existing complex system separating

them to form single independent entities and substantially recombining them in a new manner (Schumpeter 1939, pages 84-86). Innovation, at both broader economic and technical levels and firm level, may be seen as the output of a recombinant search process of existing physical and non-physical materials (Nelson and Winter 1982).

There are various forms of knowledge which need to be recombined and which affect collective results: experiential knowledge derived from shared socialization, practical knowledge, and scientific knowledge. This work aims at better understanding of how the various types of knowledge affect collective results in community settings. Two types of communities are analyzed to cover the two main sides of innovation management dealing with knowledge creation and its spread: communities of consumers and users, and communities of practice and experts. In particular, collective results are interpreted on one hand as individual commitment in communities of customers, and, on the other, as companies' success in rapidly solving problematic urgencies through communities of practice. First, the combined effect of experience and socialization is investigated in the service sector, in order to examine how the social structures of communities of customers influence individual commitments and customer retention. Second, communities of practice operating in a high-risk sector are analyzed to identify improvements in their structure, aiding the recombination of dispersed knowledge and more rapid access to those skills and competences required to solve urgent problem-solving processes.

In addition, not all recombinable knowledge is coded, and difficulties arise in merging universal and reciprocal (open science, i.e. scientific publishing), or private and trust-based (private science, i.e. patenting) knowledge. In addition, individual decision processes are strongly influenced by information and knowledge acquired from past experience (Fazio and Zanna 1981, Smith and Swinyard 1982, Branigan et al. 1997) and from previous others' behav-

ior (Goffman and Newill 1964, Coleman et al. 1966, Valente 1996b, Geroski 2000, Stoneman 2002). On one hand, individuals sharing experiences are subject to “chameleon effects” (Ramanathan and McGill 2007). On the other, individual decisions are influenced by other people’s choices. Psychologists and sociologists have introduced models of “social proof” to analyze how people’s behavior influences others’ choices, in both experimental laboratories and real-world settings (Hornstein et al. 1968, Cialdini 2001, Bandura et al. 2006). Two main epidemic models have also been used in economics, sociology and managerial studies to represent the transmission of new ideas, products and information: *contagion* and *cascades* (Chan and Misra 1990, Bikhchandani et al. 1992, Banerjee 1992, Valente 1996b, Bikhchandani et al. 1998, Valente 2005, Oh and Jeon 2007). Since the 1960s, the literature on marketing has amply recognized that customers are influenced by other customers in a variety of decision-making situations. In this framework, the model of Bass (1969) predicts an S-shaped diffusion curve of new products triggered by a small number of subjects, called *innovators* or *influentials*. Starting from the mid-1990s, Valentes’s (1995) “influential hypothesis” has become central in literature on the diffusion of innovation, communication research, and viral marketing (Tucker 2008) Recent research in marketing also reveals the role of network positioning and influence dynamics in customers’ adoption decisions (Van den Bulte and Lilien 2001, Hill et al. 2006, Bell and Song 2007, Manchanda et al. 2008, Grinblatt et al. 2008, Iyengar et al. 2011) and definition of marketing policies (Aral and Walker 2010). In addition, the increased interest in on-line social networks has reinvigorated the idea of “leveraging” customer networks to accelerate the diffusion of new products.

Various methods and tools are used to deal with these issues. Social network analysis has been used to study both the analytical and descriptive aspects of social influence. Traditionally, static social networks have been studied

according to questionnaire data and web-based social networking tools. More recently, thanks to the widespread diffusion of ICT, data on co-presence networks and dynamic weak social relationships have been collected and analyzed both on-line and in real-world settings (Ebel et al. 2002, Eagle and Pentland 2006, Guimerā et al. 2006, Eagle et al. 2009). For example, urban planning and transportation engineering are increasingly taking into consideration data on real-world social dynamics collected through reality-mining solutions to interactive design products and solutions (Graham 1997, Townsend 2000, Ratti et al. 2006, Rojas et al. 2008). The analysis of reality-mined network data, both within and between organizational boundaries, can radically improve our knowledge about the effects of social ties on human decision-making. In particular, the diffusion of radio frequency identification (RFID) devices and wireless networks has recently been improved, thanks to their usefulness in allowing researchers to collect real time data on social networks (Barrat et al. 2008, Lee et al. 2008).

In this thesis, multiple-network analyses and new methodological approaches, originally applied in the field of social ethology, are applied to study reality-mined data collected by RFID tags. In addition, a set of recently developed non-parametric methods is devised to deal with new methodological problems arising in studying socialization in networks. Recent debates in the marketing literature have emphasized methodological criticalities in discriminating between peer influence, homophily, and other confounding factors in real-world social settings. Indeed, the role of peer influence may be overestimated when other effects, such as homophily (Aral et al. 2009) and other contextual and correlated effects (Manski 1993), are wrongly considered as real social effects, such as peer influences. People may behave similarly because of similar individual characteristics and common exposure to the same institutional environment, and not because of contagion or peer influence: the higher the similarity among



people, the higher the probability of attraction between them and alignment in their behavior (McPherson et al. 2003, Christakis and Fowler 2007, Christakis and Fowler 2008, Christakis and Fowler 2011). Social identity and group identification also affect the degree to which people are influenced by other people's behavior, especially through weak ties, even in moral and unethical decisions (Tajfel 1982, Tajfel and Turner 2004, Gino, Ayal and Ariely 2009, Gino and Galinsky 2010) and in group settings (Cialdini and Goldstein 2004, Cialdini et al. 2004, Gino, Gu and Zhong 2009, Gino, Ayal and Ariely 2009). Therefore, despite the amount of literature in social psychology and diffusion dynamics dealing with social influence, confusion still emerges in isolating contagion mechanisms and methodological issues arise in analyzing decision-making processes because of correlated effects impacting social proof (Manski 1993). A whole chapter is devoted to a methodological contribution to these debates.

This thesis is composed of three core papers, the main contributions and results of which are summarized as follows:

1. *Co-experience Network Dynamics and Customer Retention: Lessons from the Dance Floor.*

Experience and socialization are key factors in determining customer commitment and renewal decisions in the service sector. To analyze the combined effect of experience and socialization, in this paper we introduce the concept of co-experience networks, and a new methodological approach is devised to study reality-mined co-experience networks. Analyzing a network of health club members over a period of four years, we find that clients with long experience have lower probability of renewing their membership. Conversely, members who are central in the co-experience network are stable and tend to renew their membership. In addition, since the members of the same reference group align their

levels of commitment, renewal decisions are clustered in a small-world network. These findings contribute to our understanding of social dynamics and localized conformity in customer decision-making which can be used to plan marketing strategies to improve customer retention.

2. *Homophily and Peer Influence in Customer Co-presence Networks.*

The role of peer influence and observational learning in customer decision making is highly debated. On empirical grounds, methodological problems arise in discriminating confounding effects from peer influence. In this paper, the impact of customer co-presence networks upon contractual renewal decisions is analyzed by controlling for homophily and environmental effects. By applying a set of non-parametric methods to a longitudinal dataset of health club members, we find that observational learning attenuates overestimation of attendance and improves customer retention.

3. *Drilling Wells and Killing Wells through Knowledge Networks.*

The rapid recombination of dispersed knowledge in multinational companies operating in risk sectors is crucial in responding to unexpected and urgent crises. The communities of practice (CoPs) introduced by a company operating in the energy sector are analyzed, to examine how their structure and the roles and positions of their members improve a rapid access to dispersed knowledge. By overlapping multiple networks of co-authorships, e-mail exchange and communities of practice, we find structural holes among the phases of research and practice in the innovation process. There are still inefficiencies in the introduction of on-line communities of practice as tools to facilitate access to critical knowledge and to speed problem-solving procedures, especially for multinational companies operating in high-risk sectors.

# Chapter 2

## Co-experience Network

## Dynamics and Customer

## Retention: Lessons from the

## Dance Floor

### **Abstract**

Experience and socialization are key factors in determining customer commitment and renewal decisions in the service sector. To analyse the combined effect of experience and socialization, in this paper we introduce the concept of co-experience networks. A new methodological approach, originally applied in the field of social ethology, is devised to study reality-mined co-experience networks. By analysing a network of health club members over a period of four years, we find that clients with long experience have a lower chance of renewing their membership. Conversely, central members in the co-experience network are stable and tend to renew their membership. In addition, since the members of the same reference group align their levels of commitment, renewal decisions are clustered in a small-world network. These findings contribute to our understanding of social dynamics and localized conformity in customer decision-making that can be used to plan marketing strategies to improve customer retention.

**Keywords:** co-experience network, co-presence data, socialization dynamics, customer retention

## 2.1 Introduction

Understanding the role of experience in individual and collective decision-making is crucial in a wide range of disciplines. Research on customer behavior reveals that individual choices are strongly influenced by information acquired from past experience (Smith and Swinyard 1982, Fredrickson and Kahneman 1993, Branigan et al. 1997, Bolton et al. 2006) and through social ties (Valente 1995, Valente 1996b, Valente 1996a, Valente 2005, Watts and Dodds 2007b, Aral et al. 2008, Iyengar et al. 2010), although the relationship between the satisfaction of prior experience and customer retention is difficult to measure, since both experience and information are socially mediated. Subjective experience is created in a dynamic social environment and is shared with other people of the same community. Individuals who share experiences are subject to a “chameleon effect”, according to which the longer subjects share the same environment, the higher the probability that they will make aligned decisions (Ramanathan and McGill 2007). Therefore, co-experience leads consumers to take similar decisions, since they tend to align their opinions.

This paper contributes to the literature by investigating co-experience dynamics (Zajonc 1965, Ramanathan and McGill 2007, Tokarchuk et al. 2009, Cattuto et al. 2010). We define a co-experience network as a social network in which individuals repeatedly share the experience of specific situations in a given context. Co-worker networks of shared job activities, commuters using the same urban transport system, and on-line communities of software developers are examples of co-experience networks. In such a network, two individuals are linked if they share some experiences. In addition, the greater the number of experiences and the longer they are shared, the stronger the link between the two individuals (Battarbee 2003a, Battarbee 2003b, Battarbee

and Koskinen 2005, Bolton and Saxena-Iyer 2009). We develop an innovative methodology to study the relationship between co-experience and retention based on co-presence<sup>1</sup> data. We take advantage of a methodology originally developed by ethologists (Hinde 1976, Krause and Ruxton 2002, Whitehead 2008) to study the dynamics of social networks based on weak patterns of social interaction. Co-presence data were collected by radio-frequency identification (RFID) tags. Thanks to the spread of information and communication technologies (ICT), data on co-presence networks and weak social relationships have been collected and analysed both on-line and in real world settings (Ebel et al. 2002, Eagle and Pentland 2006, Eagle et al. 2009).

Our approach has important implications for customer relationship management (CRM) and strategy formation in the service sector. By characterizing the real-time dynamics of customer networks, our methodological approach may be further exploited to re-design co-experience networks for customer retention and improve the existing literature applying data mining and RFID solutions to CRM (Schloter and Aghajan 2005, Lee et al. 2008, Ngai et al. 2008, Heim et al. 2009).

The paper is organized as follows. Section 2.2 describes the conceptual framework. Section 2.3 illustrates the methodology of co-presence network analysis and the empirical settings of our research. Sections 2.4 and 2.5 discuss the descriptive statistics and our main empirical findings. A concluding discussion summarizes the results and explores future developments and the potential applications of our methodology.

---

<sup>1</sup>In this study, we consider co-presence as a mode which refers to all the circumstances, both spatial and temporal, during which people can reciprocally and instantaneously interact with each other (Zhao 2003).

## 2.2 Conceptual framework

The members of any community  $M = \{1, 2, \dots, i, \dots, m\}$  are connected by a network of shared experiences which we call co-experience network,  $N$ . In a given time frame, each community member can be active or inactive within the network. Members of a health club are active when they exercise at the gym. Similarly, members of an on-line forum are active when they chat on-line. The more the active spells of two community members overlap, the more intense is their relationship in the co-experience network. Since previous research indicates that decisions are based on previous experiences, we were interested in finding out whether and how individuals decide to continue being members of the community and how much time they want to spend as active members. We focus on the decision of individual  $i$  to belong to the community  $M$ . The decision may be either positive (renewal,  $R$ ) or negative (exit,  $E$ ). We define  $G_i \subset M$  as the reference group of subject  $i$ .  $G_i$  is the group of community members with whom subject  $i$  has co-experience ties.

Decision-making models in networks typically assume that decisions are functions of the status of customer neighborhood (Watts 2002). As argued by Grabisch and Rusinowska (2008), each subject has an original inclination either to put into effect an action, such as to renew a contract. This individual inclination may differ from the final decision partly because of other subjects' influences. We define  $s_i$  as a measure of the status or inclination of customer  $i$ . In its simplest form, the status of an individual may be either positive ( $s_i = 1$ ) or negative ( $s_i = 0$ ). In our case, each individual decides whether to be or not to be an active member of the community. By being active, an individual establishes new relationships in the co-experience network and influences other members' choices. When subject  $i$  enters the community at a given time by signing a contract, we define  $s_i^+$  as positive and equal to the

duration of the contract. After the expiry date,  $s_i$  becomes negative: in this case,  $s_i^-$  specifies the time elapsing from the expiry of  $i$ 's contract. In general, a high value of  $s_i^+$  is a sign of strong commitment, and a negative value of  $s_i^-$  corresponds to the decision to leave the network (negative commitment). We can define  $s(G_i)$  as the state or inclination of the reference group of  $i$  in the co-experience network, that is of all the nodes linked to  $i$  in the co-experience network. Thus, the inclination of the neighborhood is defined as the sum of the states of the reference group of customer  $i$  in the co-experience network. If the status of the reference group is above a given threshold, there is a higher probability that the individual decision is positive. This variable measures the level of commitment of each subject's neighborhood. To ensure that the status is between 0 and 1, we normalize the positive status by dividing the duration of the contract by the maximum possible duration (a year).

By assuming that individuals have leisure-time constraints and preference for variety, we expect that long experience in the community reduces the chances of membership renewal. Thus, there is a negative drift in the renewal probability. Workers should have lower renewal probability since their leisure-time constraints are more stringent (Becker 1965).

**Proposition 2.1** *The longer the presence in a co-experience network and the more stringent the leisure-time constraint, the higher the probability of leaving the network:*

$$Pr(R_i^t) > Pr(R_i^{t+\Delta}) \tag{2.1}$$

or equivalently

$$Pr(E_i^t) \leq Pr(E_i^{t+\Delta}) \forall i \in N \tag{2.2}$$

where  $Pr(R_i^t)$  is  $i$ 's probability to renew at time  $t$ .

The second aspect of interest is the relationship between duration and experience. By controlling for the duration of membership, human interactions lead to a more positive evaluation of both the experience and the service (Moore et al. 2005, Ramanathan and McGill 2007). A service environment which improves arousal and pleasure is also likely to be more interesting for customers, who may stay longer in a store and renew service contracts (Baker 1987, Baker et al. 1992, Oliver 1999). In the literature on network dynamics, preferential attachments and the “popularity is attractive” effect imply that the stability and centrality of nodes are positively related (Barabasi et al. 2002, Jeong et al. 2003, Wagner and Leydesdorff 2005). Thus, we expect that central nodes in the co-experience network have a higher probability of contract renewal.

**Proposition 2.2** *Central agents in the network have a higher persistence:*

$$\text{if } c_i > c_j \text{ then } Pr(R_i) > Pr(R_j), \forall i, j \in M \text{ with } i \neq j \quad (2.3)$$

where  $c_i$  is the centrality of  $i$  and  $c_j$  is the centrality of  $j$ , respectively.

Lastly, we were interested in verifying whether the strength of ties among individuals influence the alignment of renewal decisions. Since individual commitment and individual experiences are socially mediated, co-experience enhances alignment in decision-making processes. A “chameleon effect” takes place when subjects share experiences.



**Proposition 2.3** *The level of commitment of subject  $i$  is positively related to the status of the reference group. More specifically, the stronger the ties between two subjects, the higher the probability that their decision to renew contracts or to leave the club will be aligned:*

$$Pr(R_i|s(G_i)^+) \geq Pr(R_i|s(G_i)^-), \forall i \in M. \quad (2.4)$$

In cases in which  $i$  and  $j$  are neighbors, we expect that the link between subjects enhances their propensity of taking common action. In other words, we claim that co-experiences tend to align the level of commitment of individuals to that of their reference group.

Different factors can be used as an explanation for alignment in individual choices in co-experience network: ambient cues, such as common shared environments (Mehrabian and Russell 1974, Donovan and Rossiter 1982, Baker 1987, Battarbee 2003b), and also social cues, such as observational learning, homophily (Manski 1993, McPherson et al. 2003, Christakis and Fowler 2007, Christakis and Fowler 2008), and peer-effects (Bikhchandani et al. 1992, Banerjee 1992, Manski 1993, Sacerdote 2001, Glaeser and Scheinkman 2003, Soetevent 2006, Oh and Jeon 2007, Watts and Dodds 2007b). Therefore, within a social context, co-experience leads to alignment in decisions made by various subjects because of three main factors: contagion and peer-effects, homophily, and environmental effects.

In sum, by including the duration of membership in the analyses (Proposition 2.1), we study the impact of both sociality and experience on future commitment decisions (Propositions 2.2 - 2.3) (Aral et al. 2009, La Fond and Neville 2010).

## 2.3 Data and methodology

We test our model in the context of the health club service sector, a rapidly growing industry in high-income countries. In 2008, there were 122,473 clubs worldwide, with more than 117 million members and a total revenue of USD 68.2 billion (IHRSA 2009). The European market ranks first in the world, with 46,736 clubs and 33.2 USD billion of revenues, followed by North America and Asia. Italy, with 5.5 million members, is the fourth largest national market after the US, the UK and Spain, in terms of industry revenues (USD 4.4 billion). In this study, we analyse data provided by a health club in Tuscany, Italy. Our database includes comprehensive customer data and contract information for each member of the club from December 2004 to July 2008. Overall, the data consist of 4,649 coded individual members, 133,945 entry registrations, 103 types of contract<sup>2</sup>, and 4,892 subscriptions<sup>3</sup>.

Data on entries were collected through RFID tags on the personal badges used to access the club and they show that among the 4,649 coded individual members, only 4,378 were observed in the health club at least once during the analysed period of 1,145 days (see Table 2.1).

	2004	2005	2006	2007	2008	2004-2008
<b>Number of days</b>	12	315	312	316	190	1,145
<b>Number of active individuals</b>	370	1,157	866	1,090	895	4,378
<b>Number of entries</b>	915	33,295	26,503	38,199	35,033	133,945

Table 2.1: Summary statistics of the complete dataset

Figure 2.1 highlights the evident seasonality of club attendance: entries are concentrated at the beginning of the week, remain stable from Monday

---

<sup>2</sup>After grouping contracts according to type of activity, we carried out analyses on sixteen contract categories. The rationale for grouping the contracts and the courses was that all participants in the same room had similar relationships.

<sup>3</sup>The number of subscriptions was greater than the number of individual members, because each member sometimes had more than one contract. Specifically, the number of contracts for each individual ranged from 1 to 57.

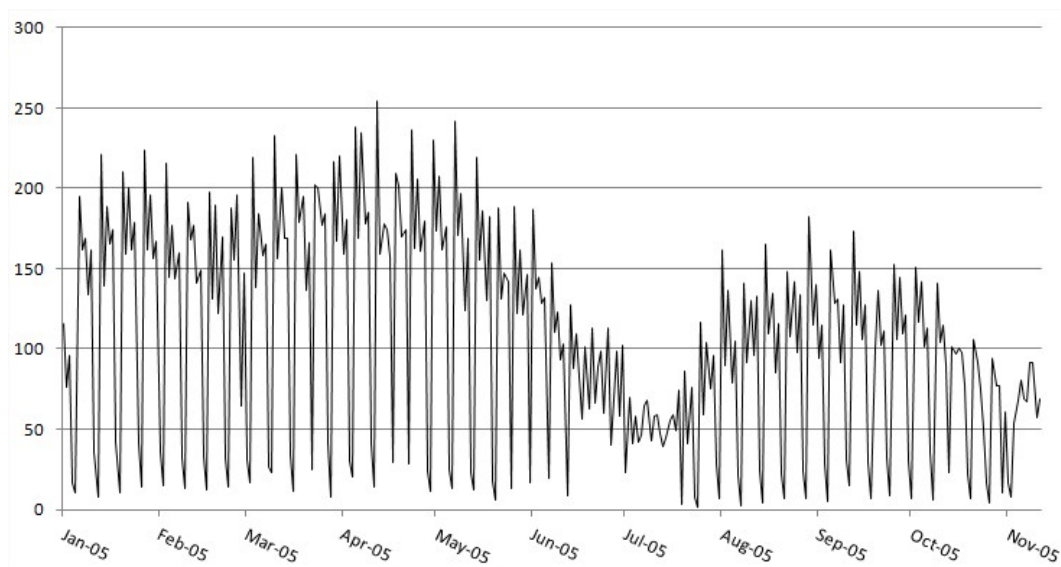


Figure 2.1: Daily number of entries in 2005

to Friday, and fall during weekends and summer holidays. A similar trend is observed in all years. Based upon an interview with the club owner, the maximum numbers of members are always well below the total capacity of the club. In addition, the RFID system to monitor entries was put in place when the club opened in 2004 and constantly used until 2008. In spite of this, in order to avoid any potential problem of initialization and finite period truncation, we decided to limit our analysis to the three central years (2005-2007). The number of club members varies over time. Table 2.1 shows that the number of identified individuals was about 1,100 in 2005 and 2007, whereas there was a sharp decrease in both total entries and active members in 2006. Thus, the total turnover of club members was negative in 2006 and positive in 2005 and 2007. Again according to the owner, the downturn of active membership in 2006 was due to the opening of a new gym in the same area and the exit of one of the club shareholders.

The club offers both traditional fitness programs and dance courses. In more detail, sixteen types of activities are available, including classical dance, belly dance, hip-hop, jiu jitsu, kick-boxing, spinning, tai chun, yoga, and Latin

American dances. We classified contracts in terms of type, duration and cost. In addition to standard contracts, the club uses other promotional tools which may influence both subscription renewal and recruitment of new customers (e.g., holiday-free offers, social dinners, Facebook groups). Unlike Della Vigna and Malmendier (2006), our dataset does not contain renewal default - that is, there are no clauses for automatic renewal of contracts.

The majority of club members (60%) choose monthly subscriptions, and about 25% subscribe to yearly contracts. Subscription fees are paid in advance, and data show that, for all contract types, the nominal price equals the effective price paid. The health club applies discounts, which are linearly related to contract length. On average, a 3.6% discount is applied for each month of contract duration (11% for 3 months). Thus, the same discount policy was consistently applied to all members throughout the study period. Information on course attendance and renewals showed that most customers' renewal decisions are taken immediately after contract expiry, and that 90% of new contracts are subscribed within 40 days of the expiry of the old ones. Therefore, a decision to leave the club is considered to be effective 40 days after the expiry of the contract in question.

To test our propositions, we mapped a network of *spontaneous social ties* (co-experience network) among club members, in which two subjects have a co-experience tie if they are co-present in the club. We define two individuals as entering together if the time-gap between their entries, recorded through RFID tags, is below a given threshold. Two subjects who frequently go to the club together are strongly associated in the co-experience network: the less they go together and the longer the time elapsing between two consecutive entries, the weaker the link between two individuals in the co-experience network. Since the time cut-off is discretionary, we apply different thresholds (from 5 to 30 minutes) as a robustness check.

The co-experience network is computed by applying a methodology currently used in social ethology (Whitehead 2008, Whitehead 2009, Sih et al. 2009). Two individuals are associated in a sampling period if the difference between the time of observation of co-presence is within a given time range. The association between two subjects is summarized by an association index, which is an estimate of the proportion of time which two individuals spend together. Among several measures of co-experience, we selected the half-weight association index defined as:

$$a_{ij} = \frac{x}{x + y_{ij} + 1/2(y_i + y_j)} \quad (2.5)$$

where  $x$  represents the number of sampling periods in which individuals  $i$  and  $j$  are jointly observed;  $y_i$  is the number of sampling periods when only  $i$  is identified;  $y_j$  the number of sampling periods in which only  $j$  has been observed, and  $y_{ij}$  the number of times both  $i$  and  $j$  are identified but not together. This ratio is unbiased, because each group has the same probability of being identified. In addition, since each individual is recorded as a member of the health club for only a small fraction of the sampling period, the half-weight index must be applied in order to avoid biases in the computation of  $y_i$ ,  $y_j$  and  $x$  (Whitehead 2008)<sup>4</sup>. Co-presence is not used to infer stronger social relationships (such as friendship) but rather to investigate the structure of weak ties, to measure co-experience, and to verify how stability and experience influence renewal decision.

---

<sup>4</sup>We replicated our analysis with various association indices and found that the results are robust to the selection of alternative measures of association.

## 2.4 Descriptive statistics

Although all the following analyses and results focus on the co-experience network, the descriptive statistics are enriched by matching other two types of social relationships. First, *exogenous social ties* have been analyzed: two individuals are linked by an exogenous social tie if they cohabituate or go to the club the first time together. The data show that 1,649 subjects live together, while 2,493 subjects have no other club's member reporting the same address. Secondly, two subjects are linked by an *induced social tie* if they attend the same course, that is to say the club management encourages socialization by scheduling group activities. In these two networks a link is identified whenever two individuals live together or attend the same course. By comparing the networks of cohabitants and shared activities in the club, we found that only 10.86% of the cohabitants attend the same courses. This result can be easily explained by the different tastes of cohabitants: mothers and sons, fathers and daughters, wives and husbands presumably attend different courses. Moreover, the networks of cohabitants and common entries in 5 minutes are largely independent: only 9.34%<sup>5</sup> of the cohabitants go to the club together. Therefore, almost all the cohabitants attending the same courses enter together, but they are only a minority of total co-entrants.

We now focus the attention on the co-experience network of *spontaneous social ties*. Figure 2.2 shows the co-experience network in 2008. To improve the legibility of the network, we applied a .10 cut-off to the association index and set the size of each node proportional to its centrality (number of links). Nodes represent the members for whom data about contractual renewals were available in 2008: green nodes stand for members who renewed contracts in the same year, and red nodes represent those who did not renew.

---

<sup>5</sup>This proportion remains stable and increases to 10.61% if we consider, respectively, 10 or 15 minutes as the time distances on which associations are defined.

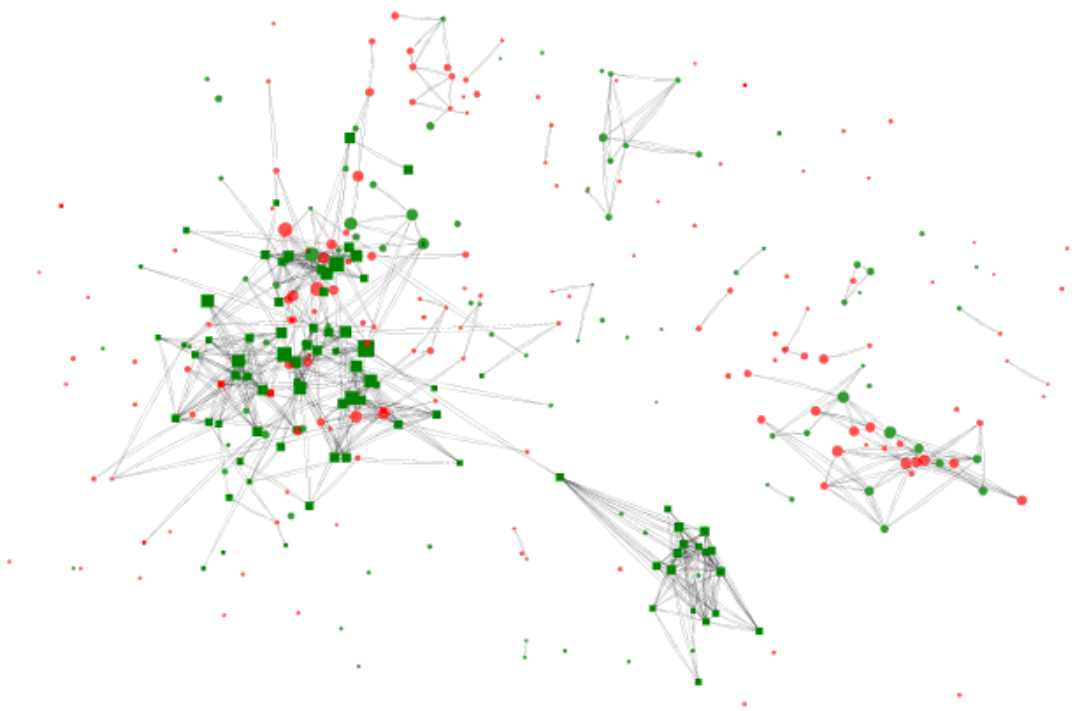


Figure 2.2: Co-presence network in 2008: each node is a club member, renewers are in green and people attending dance courses are square-shaped.

The network displays an heterogeneous small-world structure: clusters of highly connected nodes linked by few ties (Watts and Strogatz 1998a). The majority of the large central nodes are green. At first glance, this seems to confirm Proposition 2.2, according to which the central position of subjects in the network is positively related to the decision to renew contracts. The alignment in decision-making described in Proposition 2.3 is also visually confirmed by the fact that green and red nodes tend to fall into the same cluster. When we applied higher cut-offs, we found that the resulting communities were even more homogeneous and formed of either green or red nodes. Lastly, squares represent the nodes of people attending the dance courses (at the bottom of the figure). They are the more central nodes and most of them form a separate, dense group. We therefore conclude that renewal decisions tend to be shared at the level of local co-experience communities.

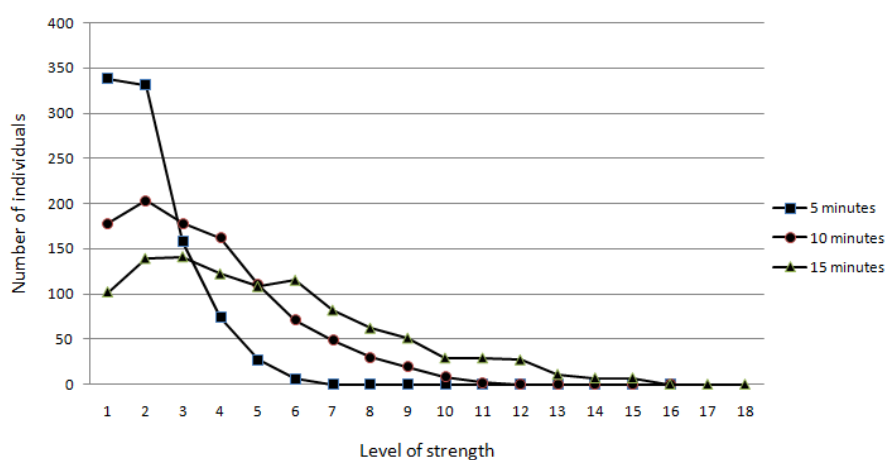


Figure 2.3: Connectivity distribution in 2007

Figure 2.3 reports the strength connectivity distribution at the three different frequencies of entry in 2007. To measure centrality we use the strength index also in the following analyses<sup>6</sup>: in our network, the strength centrality is a measure of the number of times an individual goes to the club (experience)

<sup>6</sup>We checked other measures of centrality (Wasserman and Faust 1994), which are strongly correlated with the strength index.



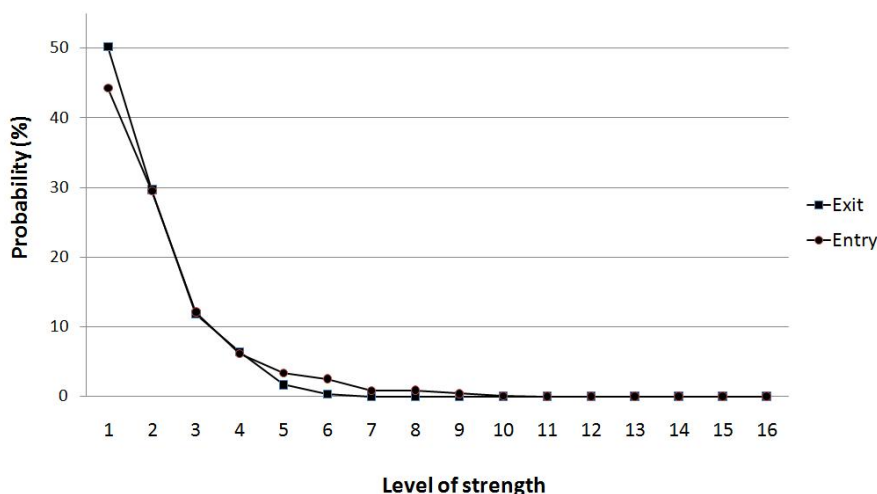


Figure 2.4: Entry and exit probability distribution by node strength (5 min)

weighted by the number of co-present members (co-experience). The connectivity distribution is relatively flat, and its skewness is negatively related to the length of the co-entrance time window (10-15 minutes): the longer the time window in question, the flatter the distribution of the number of subjects according to their strength. The same trend characterizes all the other analysed years. Since the variance of the strength is limited, we expect a more prominent role of central subjects (Watts 2002).

Figure 2.4 shows the probability distribution of entry and exit by node strength (co-entrance in a 5-minute time frame) and reveals that the majority of subjects entering or leaving the network have a lower level of strength centrality, although there are cases in which subjects enter the network at the maximum level of strength, and others in which central members have left the network (in 2007, for instance, ten members with the highest levels of strength left the network). These preliminary findings confirm the existence of a positive relationship between centrality and renewal decisions (Proposition 2.2).

## 2.5 Results

In this section, we test how renewal decisions are influenced by the position of club members in the co-experience network. We run a set of regressions to identify the variables affecting renewal decisions ( $R$ ) and, conditional upon renewal, the duration of the renewed contract ( $D$ ).  $R$  is a dummy variable (1 renewal, otherwise 0), and  $D$  is the duration of the new subscription (in number of days). As discussed in the previous sections, customers usually need time to decide whether or not to renew their contracts. Since the decision to exit is implicit, members cannot be classified simply by subdividing them into those who renew and those who do not. Since our dataset does not contain renewal default, there are no clauses for automatic contract renewal, and we consider a non-renewal decision as effective only 40 days after contract expiry.

The regressions aim at analysing how the probability of renewal is influenced by four groups of variables: an individual's characteristics; learning and experience; the list of activities offered by the health club; and co-experience network variables (see Table 2.2).

	Mean	Std. Dev.	Min	Max
<b>R</b>	0.678	0.467	0	1
<b>D</b>	89.031	145.358	0	1,692
<b>age</b>	25.810	14.396	5	72
<b>gender</b>	0.379	0.485	0	1
<b>worker</b>	0.360	0.480	0	1
<b>immigrant</b>	0.027	0.174	0	1
<b>log-distance</b>	1.196	1.221	-3.611	6.812
<b>duration</b>	0.099	0.147	0	1.291
<b>attendance</b>	0.072	0.182	0	1
<b>price</b>	50.049	115.254	0	900
<b>time</b>	2.147	0.848	0.016	4.816
<b>delay</b>	-0.043	0.251	-3.819	0.997
<b>dance</b>	0.086	0.280	0	1
<b>gym</b>	0.187	0.390	0	1
<b>status</b>	0.053	0.464	-3.488	0.997
<b>centrality</b>	3.367	1.622	1	8.722

Table 2.2: Summary statistics of the variables

The following variables are used to characterize individuals: *age* (in years), a dummy for gender (1, male; 0, female), and a dummy for employment status, *worker* (1, employed; 0, otherwise). The dummy variable *immigrant* indicates whether subjects are Italian citizens (1, immigrant; 0, otherwise), and *log-distance* is the logarithm of the distance between the home address of each subject and the health club. Two variables are used to describe members' participation in the club: *duration* is the time in years of the club membership, and *attendance* is the frequency of attendance, computed as the ratio between the number of effective entries of each subject in the club and the number of total possible entries. In addition, *price* represents the price of the last contract subscription by each member before the renewal decision, and it also serves as a proxy for the willingness to pay. The variable *time* is the moment at which the renewal decision was taken, and *delay* indicates the time elapsing from contract expiry. The two variables regarding the list of activities proposed by the club are the dummies *dance* (participation in dance courses) and *gym* (individual use of exercise machines and gym equipment). We use only dance and gym dummies because these are the two most frequently attended activities, with 55% and 13% of total subscriptions, respectively. Lastly, the network variables include both network indexes and neighbors' inclinations: *centrality* indicates the strength index of each individual, a measure of centrality in the co-experience network. The *status*  $s(G_i)$  specifies the inclination of the neighbors of each subject and their level of commitment. As noted above, status indicates if, at the moment of contract renewal, the reference group of the decision-maker in the co-experience network is mainly composed of people who have a long-term commitment to go to the gym (positive values) or includes a majority of people who decided to leave (negative values).

The results of the regressions are summarized in Table 2.3. The second

	Probit	Ordered Probit	
	R	D	D
<b>age</b>	0.007*** (0.002)	0.003** (0.001)	.004** (0.001)
<b>gender</b>	0.141** (0.071)	0.040 (0.044)	0.056 (0.044)
<b>worker</b>	-0.209*** (0.074)	-0.048 (0.045)	-0.037 (0.045)
<b>immigrant</b>	-0.172 (0.210)	-0.102 (0.128)	-.129 (0.129)
<b>logdistance</b>	0.015 (0.027)	0.011 (0.017)	0.021 (0.017)
<b>duration</b>	-2.233*** (0.318)	- -	-0.001 (0.150)
<b>attendance</b>	9.466*** (1.050)	- -	0.160* (0.093)
<b>price</b>	0.007*** (0.001)	0.003*** (0.000)	- -
<b>time</b>	-0.820*** (0.045)	-0.652*** (0.029)	-0.655*** (0.029)
<b>delay</b>	-0.416*** (0.125)	-1.339*** (0.080)	-1.194*** (0.076)
<b>dance</b>	2.597*** (0.351)	1.224*** (0.086)	1.309*** (0.087)
<b>gym</b>	1.480*** (0.248)	0.976*** (0.055)	1.426*** (0.054)
<b>status</b>	0.607*** (0.200)	0.260 ** (0.125)	0.335 *** (0.124)
<b>centrality</b>	0.243*** (0.027)	0.138*** (0.015)	0.163*** (0.016)
<b>N.observations</b>	3,269	3,297	3,278
<b>Pseudo <math>R^2</math></b>	63.74%	25.76%	21.65%

Notes: Standard errors in parentheses. \*Significant at the 10-per cent level. \*\*Significant at the 5-per cent level. \*\*\*Significant at the 1-percent level.

Table 2.3: Regression results

column lists probit regression results with dummy  $R$  as dependent variable<sup>7</sup>. Individual characteristics do not play a prominent role in renewal decisions, but the probability of renewal is influenced by socialization, experience, attendance, and kind of activities offered by the health club. More specifically, members' age has a positive and significant influence on the probability of renewal, whereas workers and women have a lower probability of renewing. Indeed, the constraint of being a worker and, especially, a young one reduces the flexibility, stability and propensity of renewing a contract. In contrast, distance from the health club and the nationality of members are not significant<sup>8</sup>.

The duration of membership (*duration*) has a negative effect on renewal decisions: the longer the membership in the club, the lower the probability of renewing. Therefore, long affiliation to the club decreases the positive influence of experience on renewal decisions. This result, together with the negative sign of the *worker* dummy, confirms Proposition 2.1. Instead, *attendance* exerts a positive effect on  $R$ : the higher the attendance frequency of each subject, the higher the probability of renewing the contract. This positive influence of individual experience on renewal decision confirms the first hypothesis of Bolton et al. (2006), according to which favorable (unfavorable) outcomes experienced over prior time periods positively (or negatively) influence renewal decisions for service contracts. The price of the previous contract also has a positive influence on the probability of renewing: subjects having higher willingness to pay are more willing to renew contracts. Conversely, the *time* dummy captures a decreasing probability of renewal, partially due to greater competition. The time elapsing from contract expiry (*delay*) also has a negative effect: the

---

<sup>7</sup>As a robustness check, we also ran a set of logit regressions, without noting any significant changes in the results.

<sup>8</sup>The result regarding the significance of distance may be biased by the fact that we do not consider the distance between club and workplace.

later a subject decides to renew a contract, the lower the propensity to renew it. The core activities organized by the club have a positive and significant influence on the probability of renewal: both *dance* and *gym* are effective in inducing participants to renew their contracts.

Lastly, all the co-experience network variables significantly influence the probability of renewal: both *centrality* and *status* have a positive and significant effect on the decision to renew. Thus, the more central a customer in the co-experience network, the lower the probability of leaving the network. In addition, since we control for the frequency of attendance, this variable controls for socialization at the gym: the larger the group of people with whom a client co-experience a service, the higher the probability of renewal. The state of the reference group in the co-experience network is also very important. A positive commitment of the reference group consistently has a positive and significant effect on the decision to renew the contract. This finding provides support for the statement in Proposition 2.2 and 2.3.

When the renewal decision is taken, members must decide on the duration of the new contract. The last column in Table 2.3 shows the results of ordered probit regressions<sup>9</sup>. For the ordered probit, we identify four thresholds, in order to examine decisions on contractual duration: one month, three months, a year, or longer. This choice is due to the fact that the demand is higher for these types of contracts. Also in this case, the decision regarding the duration of the renewed contract is not significantly influenced by individual characteristics: only the variable *age* has a positive and significant effect on the contractual duration.

In addition to the fact that *status* and *centrality* positively influence renewal decisions, these two variables also have a positive effect on the duration of

---

<sup>9</sup>The different number of observations between the two cases in Table 2.3 depends on the different variables considered and, in particular, on the different number of missing values.

the new contract. Members who are central in the co-experience network tend to subscribe long-term contracts, and tend to align the duration of the contract with the reference group in the co-experience network. The length of membership has a negative effect on contract duration (although it is not significant at the 10% level): therefore, members who have been in the club longer are more likely either not to renew the contract or, if they decide to renew it, to subscribe short-term. *Attendance* has a positive effect, although this effect on the duration of the new contract is less significant: therefore, the higher the attendance rate, the higher the propensity both to renew the contract and, in particular, to renew longer contracts.

The price of the previous contract, when included, has a positive influence on the probability of choosing long contracts, whereas *time* and *delay* negatively influence this probability: the later a subject decides to renew a contract, the lower the propensity to renew a long-term contract. The longer a member waits after contract expiry, the less likely that person will renew a long-term contract. The variables of activities *dance* and *gym* have a positive and significant influence on the decision to subscribe to longer contracts. Therefore, the decision to renew contracts and, more specifically, to renew long-term ones, does not depend simply on the socializing features of the activities, but rather on the type of activity and on how these activities are structured and organized. The successful organization of activities by the club is crucial toward creating, developing and improving socialization, experience and retention. All in all our results confirm that experience and socialization are both important for customer retention.

## 2.6 Concluding discussion

In this paper we introduce the concept of co-experience network and develop a method to analyse data on customer interactions collected by means of RFID tags or other reality-mining devices. On-line communities have been increasingly studied, but so far little attention has been devoted to the analysis of the dynamics of customer networks in real-world settings. To our knowledge, this is the first attempt to study the influence of co-experience networks on customer retention in the service sector. More in general, this paper contributes to analysis of customer behavior in co-experience networks. By analysing a network of health club members over a period of almost four years, we find that retention decreases with the duration of membership. Leisure-time constraints and preference for variety induce long-experienced subjects to leave the community. However, active club members and especially the central clients in the co-experience network are more likely to renew their membership and to subscribe long-term contracts. The longer a customer waits after contract expiry, the less likely that person will be to renew it. In addition, by controlling for individual characteristics and contractual options, we find that the commitment to go to the club is socially mediated. Health club members tend to align their commitment to the socially related group of co-members.

The innovative method used in this study may be further applied to customer and human resource management, to design co-experience networks. Our preliminary results may be used to improve the allocation of marketing efforts, for example, or to better allocate instructors in the co-experience network of the health club. Secondly, our results may be used to develop more efficient discount policies, which take into account not only contract characteristics, but also the interpersonal influence regarding renewal decisions. Customer relationship managers in the service sector do already implement customer re-



lationship strategies such as fidelity cards, “bring a friend” promotions, and special discounts on the subscription price. However, these strategies are not based on real time analyses of customer experience: the proposed results and methodology described in the paper may be applied toward improving marketing policies by developing and implementing personalized marketing actions based on customer roles and positions in the co-experience network.

Although we show that individual commitment is highly related to the decisions of their neighbors, this is not sufficient to infer that individual choices are influenced by their reference group. The next paper aims at differentiating peer influence from other confounding effects in aligned decision-making processes.



# Chapter 3

## Homophily and Peer Influence in Customer Co-presence Networks

### Abstract

The role of peer influence and observational learning in customer decision making is highly debated. On the empirical ground, methodological problems arise in discriminating confounding effects from peer influence. In this paper the impact of customer co-presence networks upon contractual renewal decision is analysed controlling for homophily and environmental effects. By applying a set of non-parametric methods to a longitudinal dataset of health club members we find that observational learning attenuates overestimation of attendance and improves customer retention.

**Keywords:** peer influence, homophily, social networks, observational learning, service marketing

### 3.1 Introduction

*“Who dances which dance with whom and when? [...] As the combinations of partners and dances unfold, collective dynamics emerge. Individual choices may cumulate into a cascade [...] or trends may cluster and find coherence only in small densely connected groups.”* (Powell et al. 2005)

In everyday life, social networks emerge from interactive experiences in our roles as consumers, workers, friends, or members of online and real-world social groups. Peer influence (Katz and Lazarsfeld 1955, Bass 1969, Valente 1995) and observational learning (Bikhchandani et al. 1992, Banerjee 1992, Cai et al. 2007, Zhang 2010, Moretti 2011) in social networks are relevant in many purchasing and non-purchasing decisions. While peer influence is a more generic concept based on various mechanisms through which subjects learn from others (Granovetter 1978, Valente 1996a, Stoneman 2002, Watts and Dodds 2007a), observational learning is a weak form of peer influence. Observational learning allows to gain informations from others by observing their behavior, without a direct contact or speech (Bikhchandani et al. 1998).

Various approaches have been implemented so far to study the co-evolution of influence dynamics and social networks, both theoretically and empirically (Manski 1993, Valente 2005, Watts and Dodds 2007a, Iyengar et al. 2011). The relation between individual behavior and position in social networks has already been demonstrated in customers' adoption decision studies (Van den Bulte and Lilien 2001, Hill et al. 2006, Bell and Song 2007, Manchanda et al. 2008, Grinblatt et al. 2008, Iyengar et al. 2011, Narayan et al. 2011) and marketing policies has been designed accordingly (Aral and Walker 2010).

Despite the crucial role of observational learning, this literature is still hampered by several difficulties in measuring the role of influence in network diffusion dynamics and customer behavior. Indeed, the role of peer influence may be overestimated when other effects, such as homophily (Aral et al. 2009) and other contextual and correlated effects (Manski 1993, Manski 2000, Sacerdote 2001, Soetevent 2006) are not considered. On the empirical ground, recent debates in the literature on diffusion and peer influence focus on new methodological approaches to disentangle among confounding factors (Snijders et al. 2007, Tucker 2008, Aral et al. 2009, Aral 2011, Christakis and

Fowler 2011, Iyengar et al. 2011). By analysing a global instant messaging network, Aral et al. (2009) show that more than half of behavioral contagion is due to homophily rather than to peer influence. Subjects may behave similarly, not because of social stimuli but due to common individual inclinations and tastes, or because they are exposed to the same environment as well. The quality of the available data and the robustness of the applied econometric methods are the two main concerns in the literature (Aral 2011, Iyengar and Van den Bulte 2011).

The goal of this paper is to empirically model observational learning on long-term commitment, its impact on customer decision making, and over-estimation of attendance. We contribute to the recent debates on diffusion dynamics by studying a group-level setting (Christakis and Fowler 2011), and exploiting the advantages of a unique dynamic dataset (Iyengar and Van den Bulte 2011, Aral 2011). As in Della Vigna and Malmendier (2006), we selected a long-time perspective but our data allow to disentangle confounding effects from observational learning by excluding automatic renewals. We improve the previous literature on customer behavior and observational learning by adding a dynamic perspective in the long run and using a rich observational dataset over a four year-period. Radio frequency identification (RFID) technology, customer relationship management, and questionnaire data has been exploited to discriminate between influence and confounding effects and to demonstrate the importance of peer influence on the decisions of members to renew or not renew their contracts. Innovative methodologies frequently used in social ethology have also been implemented to measure co-presence in a dynamic setting. In addition, we improve the efficiency in the discrimination analysis among confounding effects by implementing a set of matching non-parametric methods. In particular, we improve the previous studies based on the use of Propensity Score Matching (Aral et al. 2009) by adding a promising

approach called Coarsened Exact Matching (Iacus et al. 2011) allowing us to control for confounding effects in observational data.

The rest of the paper is organized as follows. Section 3.2 overviews the relevant literature and identifies the four research questions. Section 3.3 describes our research setting and methodological approach. Next, Section 3.4 specifies the variables created for the analysis and present the results. Section 3.5 concludes the paper with a discussion of implications for marketing practice and directions for future research.

## **3.2 Literature review and research propositions**

Social influence is a key factor in human decision-making processes, and helps to explain the diffusion of behavior and beliefs. Literature in innovation diffusion and individual behavior specifies theoretical explanations of how peer influence affects both individual choices and the diffusion of knowledge, ideas and products. However numerous difficulties have emerged in measuring and testing the presence and role of influence in individual decision-making.

We are interested in analysing the adoption and purchasing decisions of customers in real work settings. In the theoretical literature, behavioral models, at both social (Bass 1969, Chan and Misra 1990, Valente 1996a, Valente 1996b, Watts and Dodds 2007a, Valente 2005) and individual levels (Cialdini 2001, Goldstein et al. 2008), have been developed to describe peer influence in social contexts by focusing on the relational aspects of network dynamics. Increased interest in on-line social networks has also invigorated the idea of leveraging customer networks to accelerate the diffusion of innovation and new products (Watts and Dodds 2007b, Watts and Dodds 2007a, Iyengar et al. 2011, Narayan et al. 2011). The power of social proof has also been widely confirmed in experimental psychology (Hornstein et al. 1968) and in real world settings (Bandura

et al. 1963, Bandura 1965, Bandura et al. 2006).

Despite the amount of literature in social psychology and diffusion dynamics dealing with social influence, confusion about the evaluation of consumer network dynamics emerge in the marketing field when scholars try to isolate contagion and influence (Aral et al. 2009, Aral 2011). Behavioral alignment caused by social proof may indeed be evidence of various elements: peer influence, homophily and other confounding factors. Alignment in individual choices may be due to the similarity of tastes and attributes among subjects (homophily), rather than to endogenous peer effects (Manski 1993, McPherson et al. 2003, Christakis and Fowler 2007, Christakis and Fowler 2008, Fowler and Christakis 2008). In addition, one stream of literature focuses on environmental effects - that is, on the effects of the physical environment of a store on individual attitudes, emotions and choices as one of the most relevant confounding effects (Donovan and Rossiter 1982, Baker 1987). As we also demonstrated in Chapter 2, homogeneity in the behavior of group members may in fact be determined not only by social cues, but also by atmospheric factors or ambient cues, such as a common shared environment and co-experience (Mehrabian and Russell 1974).

This paper aims to prove that, despite the presence of both homophily and environmental effects, customer decision-making is conditioned by observational learning<sup>1</sup>.

**Proposition 3.1** *After controlling for homophily and environmental effects, peer influence in co-presence networks has a positive effect on individual decisions.*

---

<sup>1</sup>In this paper we measure observational learning as the most likely form of emerging peer-influence in co-presence networks. Co-presence allows to observe others' attendance and verify renewal decisions but in this context we cannot exclude other direct forms of social learning due to socialization inside the club. We also examined ex ante socialization outside the club by analysing exogenous social ties (cohabitation and co-entrace for the first time), but we find that the network of co-presence is only weakly dependent upon exogenous social ties.

By focusing the attention on homophily characteristics, we also specify the types of homophily that can affect individual decision-making processes. In particular, psychological closeness and physical proximity (Goldstein et al. 2008) are main causes of alignment in decision-making processes. Moreover, individual characteristics, product features and price have been defined by the recent literature to be important determinants of individual decisions and influential mechanisms (Aral 2011, Christakis and Fowler 2011). Differentiations based on individual characteristics, such as socio-demographic characteristics and willingness to pay, and product features, such as quantity, have been highlighted by the economic literature on price discrimination as well. By supporting Aral (2011)'s comments about the elements causing influence and recovering price discrimination sources, we characterize homophily according to four elements. More specifically, we expect that:

**Proposition 3.2** *Among homophily effects (socio-demographic characteristics, physical proximity, product usage, and product type), the type of products or activities chosen and the individual characteristics of customers play a predominant role in determining individual choices.*

In addition, the marketing literature contains an increasing amount of customer network data but, to our knowledge, there is a lack of literature describing influential factors in dynamic marketing settings (Iyengar et al. 2011, Iyengar and Van den Bulte 2011). In the last decade, data on co-presence networks and dynamic weak social relationships have been collected and analysed both on-line and in real world settings<sup>2</sup> (Ebel et al. 2002, Eagle and Pentland 2006, Guimerà et al. 2006, Eagle et al. 2009) but previous studies on

---

<sup>2</sup>An application of dynamic co-presence data in a real world setting has been proposed in Chapter 2.



contagion did not either consider social network data and the connectivity of nodes or control for time-varying shocks (Iyengar et al. 2011). Following Aral et al. (2009), who introduced an intertemporal factor in analysing the discrimination between homophily and peer influence, we use our dynamic dataset to examine the relation between influence and temporal distance in individual choices. In particular, we expect that:

**Proposition 3.3** *Influential effects on renewal decisions decay with the time between customers' renewal decisions.*

Lastly, we prove that level of attendance is influenced by the reference group of peers with whom customers frequently overlap. Since Della Vigna and Malmendier (2006) found that club members are overconfident about their time capacity to go to the club, we test how attendance is amplified or reduced when influence occurs. We find that individual attendance is influenced by the reference group of peers in the co-presence network. Our results are in line with those of Mas and Moretti (2009) and suggest that customer overconfidence is socially mediated (Della Vigna and Malmendier 2006). Following Della Vigna and Malmendier (2006)'s approach, focused on detailed analysis of contractual prices, we test how attendance is amplified or reduced when influence occurs.

**Proposition 3.4** *Among renewers, observational learning reduces over-estimation of future attendance.*

However, methodological difficulties arise when we want to discriminate between peer influence and confounding factors, such as homophily and environmental cues. As discussed in Aral (2011), in order to limit the evaluation bias, various methodological instruments have been used in the past to deal with this issue, e.g., actor-oriented models (Snijders et al. 2007), *ad hoc* methods (Christakis and Fowler 2007) and natural experiment methods

(Sacerdote 2001, Tucker 2008). More recently, random field experiments (Aral and Walker 2010, Aral 2011), peer effect models (Bramoullé et al. 2009), dynamic matching models (Aral et al. 2009) and structural models (Ghose and Han 2009) have also been introduced to analyse the topic more rigorously.

This paper proposes two methods to discriminate between these three factors determining alignment in individual decisions. The first is based on the Propensity Score Matching approach developed by Aral et al. (2009), to distinguish influence-based contagion from homophily-driven diffusion in dynamic networks through matched sample estimations. They compared influence estimates in matched sample estimation and random matching by applying the package developed by Becker and Ichino (2002). We improve on this by adding a Monotonic Imbalance Bounding (MIB) matching method, called Coarsened Exact Matching (CEM), recently developed by Iacus et al. (2009). The proposed analyses test whether an individual’s position in a social network of dance club members influences attendance and renewal decisions.

### **3.3 Contextual framework and methodological approach**

This section describes our setup. It includes our data sources, notation, definitions of the co-presence network and our target quantities of interest, as well as a brief summary of matching methods to control for confounding effects (see also the appendix).

#### **3.3.1 Data sources**

In this paper we study influence and contagion effects in a health club and focus on the decision of the members to renew or not renew their contracts.

We test our model in the health club service sector.

In this study we use analyse data provided by a health club in Tuscany, Italy. The dataset contains 4,645 contracts subscribed by the members of a dance club in the period December 2004-July 2008. Each member is reported as many times as the number of contracts subscribed. The data consist of 4,649 coded individual members, 133,945 entries recorded with RFID technology, 103 types of contracts, and 4,892 subscriptions<sup>3</sup>. In addition to fitness programs, the contracts offer 16 types of activities, including classical dance, belly dance, hip-hop, jiu jitsu, kick boxing, spinning, tai chun, yoga, and Latin American dancing. This dataset was further reduced by dropping contracts not renewed in their last 90 days. The final number of observations used for these analyses is 4,390.

#### 3.3.2 The co-presence network

The members of the community are connected by a spontaneous network of co-presences<sup>4</sup> in the community, and influence and contagion are determined by the presence of renewer neighbors in the co-presence network. The network was defined by applying a method usually used in social ethology to study associations and spontaneous social ties in animal societies (Whitehead 2008). By monitoring entries through the RFID system, we associate two subjects based on the time elapsed between their entries. Two subjects are associated if the period elapsing between their entries is less than a certain threshold. The association index is an estimate of the proportion of time that two individuals spend together. We use a half-weight association index<sup>5</sup>, defined as:

---

<sup>3</sup>In our data, there are no clauses for automatic renewal of contracts (renewal defaults).

<sup>4</sup>In this analysis, co-presence is used to analyse weak ties, and it is not aimed to inferring stronger relationships.

<sup>5</sup>The index is not biased because each group has the same probability of being identified. Because of the small fraction of time during which each individual was registered in the club, the half-weight index had to be applied, to avoid bias in the computations (Whitehead 2008).

$$a_{ij} = \frac{x}{x + y_{ij} + 1/2(y_i + y_j)} \quad (3.1)$$

where  $x$  is the number of sampling periods in which individuals  $i$  and  $j$  are jointly observed;  $y_i$  is the number of sampling periods when only  $i$  is identified;  $y_j$  is the number of sampling periods in which only  $j$  is identified, and  $y_{ij}$  is the number of sampling periods when both  $i$  and  $j$  are identified but not associated. We define  $G_i$  as the reference group of subject  $i$ : in other words,  $G_i$  is the set of  $i$ 's direct neighborhood in co-presence network - that is, of those subjects co-entering with  $i$ .

Despite the evidence that the level of commitment of subjects  $i \in M$  is positively related to the inclination and commitment of their reference group  $G_i$ , this is not sufficient to infer that subjects' decisions are influenced by the inclination of their reference groups.

### 3.3.3 Influence in co-presence networks

Let assume that member  $i$  must decide whether to renew or not renew her/his contract, and  $i$  co-enters with  $j$  according to co-presence network  $N$ . We then analyse influence and contagion by defining treatment as the presence of renewer ( $j$ ) in the reference group on a customer ( $i$ ). The probability to be treated is a function of three parameters:

$$Pr(T_i) = f(p_1, p_2, p_3) \quad (3.2)$$

where  $p_1$  is the number of days remaining until the expiry of  $i$ 's contract, and i.e., until time of  $i$ 's decision,  $p_2$  is the number of days which have already passed since subject  $j$  renewed<sup>6</sup>, and  $p_3$  is the time distance between

---

The robustness of the results was confirmed when other association indexes were applied.

<sup>6</sup>We include both the case in which  $j$  first subscribes, that in which  $j$  renews.

### 3.3. CONTEXTUAL FRAMEWORK AND METHODOLOGICAL APPROACH

the entries of  $i$  and  $j$ . A customer  $i$  is treated if he co-enters with at least one renewer, based on the previous three time constraints. Since the most frequently subscribed contracts are for one month, it make sense to consider values of  $p_1$  and  $p_2$  less than 30 days, while  $p_3$  may have different time frames (5 to 30 minutes).

Members are therefore treated if they enter the gym within the contract period, if their contract expires in less than  $p_1$  days, they co-enter at least once with a client who renewed no more than  $p_2$  days before their common entry in a  $p_3$  time frame. Figure 3.1 shows the time-line explaining this framework. Thus by writing  $T(5, 5, 5)$  we identify treated nodes by setting  $p_1 = 5$  days,  $p_2 = 5$  days and  $p_3 = 5$  minutes.

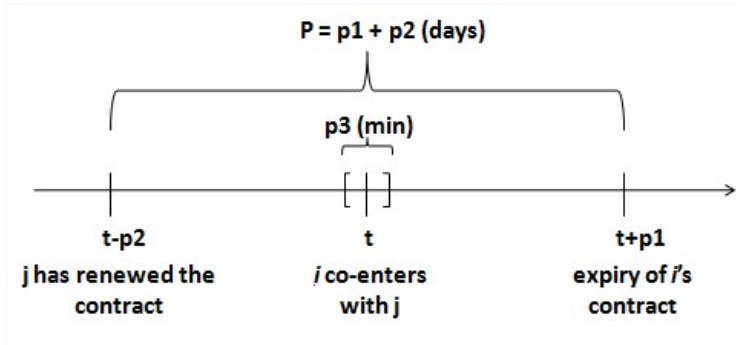


Figure 3.1: Time-line of the treatment framework

Figure 3.2 reports a graphical representation of the co-presence network in 2008<sup>7</sup>. Nodes represent members for which data on renewals are available: black nodes are subjects who renewed in 2008, and gray nodes those who did not. The two shapes of the nodes are to show treated subjects in the co-presence network: squares are treated and circles are untreated members ( $T(5, 5, 5)$ ). Clearly, the network has a small-world structure of tightly knit subgroups of co-entrants (Watts and Strogatz 2006). Since squares (circles)

<sup>7</sup>A .10 cutoff was applied to the association index to improve the legibility of the graph. The size of the nodes is proportional to their network centrality, that is, to the number of links each node has within the network.

are mostly black (gray), treated members tend to renew.

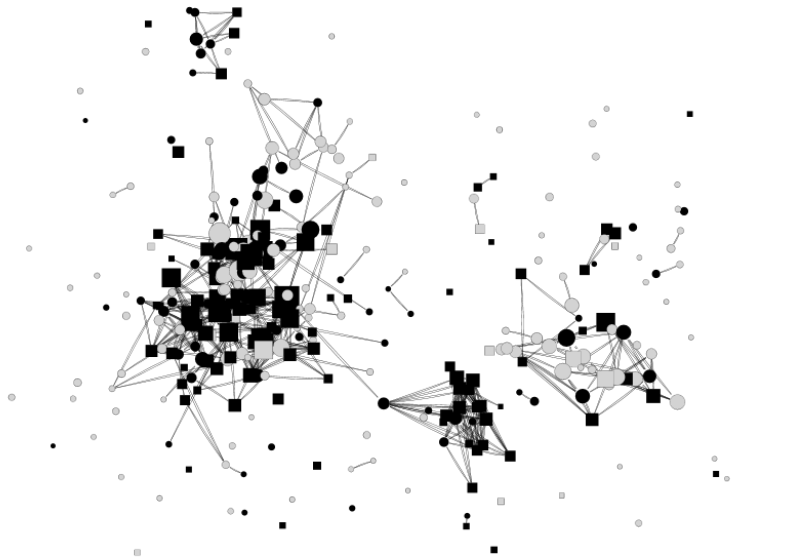


Figure 3.2: Co-presence network in 2008: each node is a dance club member, renewers are in black and treated members ( $T(5, 5, 5)$ ) are squared-shaped.

### 3.3.4 Influence, homophily and environmental effects

However, if influence is specified by the presence of renewer neighbors in the co-presence network (*treatment*), we cannot observe whether members having neighbors who have renewed their contracts (*treated*) would have done so if they had not had renewer neighbors (Aral et al. 2009). The selection bias introduced by non-randomly assigned treatments requires simple regression analysis to be improved by moving to matched sampling (Becker and Ichino 2002). Following Aral et al. (2009)’s procedure and methodological framework, we apply matched estimations and compare similar subjects having the same probability of being treated: we control for similarity among members - that is, for cases in which renewers are more likely to renew, not because of influential effects but because of similarities with their neighbors, and for other confounding environmental factors affecting individual decisions. We therefore condition the matches on the characteristics and attributes of both members and contracts,

in order to control for situations in which aligned behavior is verified even if no peer influence or contagion effects exist. In addition, because of the variation of influence over time, we consider temporal aspects in our definition of influence and treatment and propose dynamic matching over time.

We discriminate between homophily and other confounding factors (pre-treatment effects) which may be wrongly considered as influence. In order to measure homophily, we consider individual characteristics and the type of chosen activity to take into account both socio-demographic similarities and common tastes. As environmental cues, we consider two proxies: subjects as exposed to the same common environment if they enter the gym on average during the same hours of the day and if they are exposed to crowding of the noisy hours, when there are many people around. A more detailed definition of the statistic models used to measure homophily and environmental factors is reported in Section 3.4.2.

To discriminate between homophily, peer influence, and environmental factors, we apply two matching methods: Propensity Score Matching, also proposed by Aral et al. (2009), and the Coarsened Exact Matching (Iacus et al. 2011) (see appendix). In the next section we describe the variables and report a first logit regression to test whether influence affects renewal decisions. We then implement the two matching algorithms to test whether, after controlling for homophily, environmental factors, and both, peer influence is still a determinant of individual decisions to renew contracts.

## 3.4 Results

In the following analyses we controlled for various groups of variables: Table 3.1 lists the characteristics of individuals, entries, contracts, and environment.

	Mean	Std. Dev.	Min	Max
<b>R</b>	0.634	0.481	0	1
<b>influence</b>	0.095	0.294	0	1
<b>price</b>	149.048	181.681	0	900
<b>dance</b>	0.133	0.340	0	1
<b>gym</b>	0.550	0.497	0	1
<b>age</b>	24.707	15.634	5	72
<b>male</b>	0.337	0.472	0	1
<b>married</b>	0.219	0.414	0	1
<b>worker</b>	0.494	0.500	0	1
<b>citizenship</b>	0.974	0.156	0	1
<b>education</b>	2.44	1.040	0	4
<b>downtown</b>	0.260	0.439	0	1
<b>entries</b>	16.026	27.692	0	240
<b>noise</b>	1.040	3.104	0	46
<b>time of entry</b>	15.753	2.864	8	22

Table 3.1: Summary statistics of variables.

$R$  is a dummy indicating renewals (1 renewal, 0 no renewal) and *influence* is the variable specifying the applied treatment  $T(5, 5, 5)$ , which is 1 if the subject co-entered at least once with renewer neighbors in the co-presence network, 0 otherwise. In order to characterize contracts we considered *price* as representing the price of the expiring contract, and it can also be seen as a proxy for willingness to pay. The two variables regarding the set of activities proposed by the club are the dummies *dance*, participation in dance courses, and *gym*, individual use of exercise machines and gym equipment without the subscription to any type of course. The following variables were used to characterize individuals: *age*, in years; *male* is a dummy (1, male; 0, female); and *married* represents marital status (1 married; 0 otherwise). *Worker* indicates employment status dummy as proxy for economic independence (1, employed; 0, otherwise), and *education* the education level, based on certificated schooling. The dummy variable *citizenship* indicates whether subjects are Italian (1, Italian; 0, otherwise); *nationality* reports the citizenship of each member. *Downtown* is a dummy specifying subjects' addresses (1 city center; 0 otherwise). Also controlled were members' entries in the club, and *entries* was used



as the number of times each subjects entered the gym within each expiring contract period.

Lastly, we controlled for two proxies describing environmental factors. We used the results of a questionnaire from a consulting company to determine environmental characteristics. In particular, the questionnaire, which covered a sample of 249 gym clients in 2009, highlighted the level of noise and crowding of the gym as determinants for customer satisfaction. We used the questionnaire information about the time at which people complained of more noise, which overlapped with the time of children’s courses (Monday, Wednesday and Friday, 5PM to 6PM), to measure one of the confounding environmental effects. Accordingly, we defined the variable *noise* as the number of times each subject entered the gym during the crowded and noisy hours within each contract’s validity. As a second proxy, we used the *time of entry*, indicating the mean time (in hours after midnight) when each subject entered the gym during the validity of each expiring contract. This variable was a proxy for gym cleaning, because the cleaners were scheduled to arrive on a daily basis during the late evening.

### 3.4.1 Preliminary findings

The first proposed regression in Table 3.2 lists the logit regressions analysing how renewal decisions ( $R$ ) are influenced by the treatment, that is, how influential dynamics conditioned individual choices to renew membership. Four models are proposed, by adding new control factors to the regressions: more specifically, we increasingly controlled for influence, the characteristics of both activities and individuals, and common environmental factors.

Neither environmental factors nor socio-demographic characteristics played a significant role in determining the renewal decision, but renewal decision

Logit model	(1)	(2)	(3)	(4)
<b>Influence</b>	1.321*** (0.142)	1.192*** (0.155)	1.252*** (0.242)	1.225*** (0.246)
<b>price</b>		-0.003*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)
<b>gym</b>		-1.642*** (0.094)	-1.723*** (0.168)	-1.651*** (0.234)
<b>dance</b>		-0.001 (0.152)	0.389 (0.240)	0.571* (0.334)
<b>entries</b>			0.007** (0.002)	0.006** (0.003)
<b>education</b>			-0.170** (0.085)	-0.167* (0.096)
<b>downtown</b>			0.208 (0.134)	0.315** (0.152)
<b>age</b>			0.023*** (0.007)	0.012 (0.009)
<b>male</b>			-0.228* (0.124)	-0.254 (0.139)
<b>married</b>			0.058 (0.165)	-0.019 (0.187)
<b>worker</b>			0.051 (0.192)	0.262 (0.209)
<b>citizenship</b>			0.181 (0.309)	0.116 (0.329)
<b>Environmental effects</b>	NO	NO	NO	YES
<b>N.observations</b>	4,390	4,375	1,962	1,455
<b>Pseudo <math>R^2</math></b>	1.93%	21.06%	24.52%	24.09%

Notes: Standard errors in brackets. \*Significant at 10% level. \*\*Significant at 5% level. \*\*\*Significant at 1% level.

Table 3.2: Effects of influence on renewal decisions. Four logit models increasingly control for influence, characteristics of both activities and individuals, and common environmental factors.

was influenced by treatment and by the characteristics of both contracts and entries.

First, all regressions confirm the positive effect of influence on renewals: the likelihood of renewal increases when subjects have renewer neighbors. When other values of  $p1$  and  $p2$  are applied to define influence, the effect of treatment on the decision to renew does not change significantly. The results are also robust when longer time distances of co-entrance ( $p3$ ) are applied. Second, the *price* of the expiring contract has a negative effect on the probability of renewal: members were less willing to pay for expensive contracts. The activities organized by the club also had a significant impact on the probability of renewal: in particular, those activities which did not require attendance at organized courses (*gym*) significantly reduced renewal probability, unlike from organized activities like dance courses, which significantly increased  $R$  when all controls were included in the logit model. The way in which the club organizes courses and activities is therefore determinant in customer-satisfaction decisions: this confirms Aral (2011)'s discussion on the importance of product characteristics in determining individual behavior, and contagion and cascade mechanisms.

Socio-demographic characteristics do not play a significant role in determining renewal decisions. More specifically, neither employment nor marital status significantly affect renewal decisions, nor is members' citizenship significant. Nevertheless, when environmental factors are not considered, older women without secondary education have a higher probability of renewing. Indeed, being highly educated requires more effort, and less free time to go to gyms, than during secondary school. Living downtown, where the gym is located, also increases the probability of renewal: the effect is significant when common environmental characteristics are included. Lastly, whenever included, the number of entries significantly influences renewal decisions. The

more times members enter the gym during each contract duration, and therefore the more efficient the use of the contract, the higher the probability that they will renew their contracts.

We also control for environmental factors, but neither level of crowding nor members' time of entry significantly affect renewal decisions. The results of this first regression do not change significantly after also controlling for the mean price per entry as an additional independent variable.

Table 3.2 therefore shows the positive role played by peer influence in determining individual decisions. However, in this regression we overestimated peer influence effects because we were not controlling in detail for specific homophily effects and other confounding factors. We discriminate among these factors in the next section.

### **3.4.2 The effect of peer-influence on renewal decisions**

The effects for which we controlled in order not to overestimate peer influence in renewal decisions were homophily and confounding environmental effects. In the following analysis we applied Propensity Score Matching, as proposed by Aral et al. (2009), and Coarsened Exact Matching (Iacus et al. 2011) to test for confounding factors. Three tests were computed in order to test the size of the influence on members' renewal decisions: the first controlled only for homophily, the second only for environmental effects, and the third controlled for both.

The two control variables used as proxies for confounding environmental effects were *time of entry* and *noise*: we considered subjects as exposed to the same common environment if they were on average exposed to the same cleaning level and if they both entered during the crowded and noisy hours. Instead, In order to test for homophily, we used four groups of variables: customers'

socio-demographic characteristics, customers' location, type of activity, and use of contract. Variables characterizing the similarity of subjects were all the individual socio-demographic characteristics listed in Table 3.1, and price as a proxy for willingness to pay. The variable *proximity* summarized nine dummies representing the various districts of the city where customers live, in order to study location. We also used *type* as a homophily factor to specify the effects of 16 dummies, representing the 16 activities offered by the club. Similarity among subjects may also be evaluated on the basis of the same activities. Lastly, the number of entries was used to analyse the contract use by subjects. All variables considered in the two models were simultaneously included as control variables when both effects were tested.

	Mean Pscore	Std.Dev. Pscore	Balance property	ATT	treated	controls
<b>Environmental effects</b>	0.128	0.016	YES***	0.306*** (0.026)	421	2,801
<b>Homophily</b>	0.101	0.085	NO	0.181*** (0.029)	165	1,310
<b>Both effects</b>	0.138	0.117	YES***	0.178*** (0.032)	164	871

Notes: Standard errors in brackets. \*Significant at 10% level. \*\*Significant at 5% level. \*\*\*Significant at 1% level.

Table 3.3: Stratification Propensity Score Matching.

The results of Stratification Propensity Score Matching, are listed in Table 3.3, which lists the data on the size of the propensity score and of the treatment and control groups in the three tests. The propensity score algorithm determined the number of blocks within which the average propensity score of treated and untreated subjects did not differ. The balancing property was satisfied at the 1% significance level when both effects were treated simultaneously, and when only the environmental one was tested, but not for some of the control variables in the peer influence test. After running the algorithm, ATTs were computed under the common support condition, according to which com-

putation is restricted to the region of common support. The standard errors of treatment effects were bootstrapped and 100 bootstrap replications were performed. In all three tests, the ATT was significantly positive, confirming the positive effect of treatment on the probability of renewal after matching observations. First, in the peer influence analysis, we found that, even after controlling for homophily - that is, for the similarity of individual characteristics both socio-demographic and related to common tastes in gym activities, subjects having neighbors who renewed their contracts (treated) have a higher probability of renewing. Second, in the environmental effect analysis, we found that, of those who are exposed to the same environmental stimuli, people having renewer neighbors were more likely to renew. Lastly, after simultaneously controlling for homophily and confounding environmental effects, we found that the previous results remained robust (Proposition 3.1).

Once the propensity score had been computed, CEM was applied to improve the efficiency of the matching procedure. As Table 3.4 shows, the level of imbalance, measured by the multivariate  $L_1$  distance, decreased after applying CEM and therefore after deleting observations with no close matching on pre-treatment variables in either treated or control groups. For example, when we tested for both effects, CEM increased the percentages of the overlapping density of the two distributions of treated and control data from 9.5% to 22.6%.

	Environmental effect	Homophily	Both effects
$L_1$ Imbalance pre-matching	0.273	0.798	0.905
$L_1$ Imbalance post-matching	0.173	0.623	0.774
Number of strata	52	2,111	2,825
Number of treatment observations	421	421	421
Number of treatment obs matched	420	313	229
Number of control observations	3,969	3,969	3,969
Number of control obs matched	2,777	1,069	391

Table 3.4: Level of imbalance before and after CEM.

Automated coarsening was applied to establish the maximum level of imbalance *ex ante* and its output was used to estimate the SATT, the Sample Average Treatment effect on the Treated. CEM was applied with the same control variables used in propensity score analysis. Under the ignorability assumption, according to which treatment is independent of outcome, SATT was computed as a weighted difference in means between treated and control groups or, equivalently, as a weighted linear regression of  $R$  on treatment, that is, on the variable *influence*. In our case, we measured SATT as coefficient  $b$  of regression  $R = a + b \cdot influence + e$ . Table 3.5 lists the four models resulting from applying the CEM algorithm to each of the three analyses. Also in this case, we increasingly added groups of independent variables, to test for incremental cumulative effects. As shown in the (1) models for the three analyses of Table 3.5, the SATT estimate (coefficient of the variable *influence*) was always positive and significant at the 1% level. In order to check for remaining imbalance in the matched data, Table 3.5 lists the logit regressions obtained by adding other groups of control variables to the previous basic model.

CEM model	Environmental effects				Homophily				Both effects			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
<b>Influence</b>	1.275*** (0.269)	1.351*** (0.276)	1.507*** (0.432)	1.479*** (0.440)	1.477*** (0.144)	1.170*** (0.159)	1.239*** (0.251)	1.233*** (0.251)	1.498*** (0.388)	1.534*** (0.393)	1.794*** (0.640)	1.886*** (0.655)
<b>price</b>		-0.003*** (0.001)	-0.008** (0.041)	-0.009** (0.004)		-0.003*** (0.001)	-0.005*** (0.001)	-0.052*** (0.001)		-0.005** (0.002)	0.027 (0.019)	0.009 (0.021)
<b>gym</b>		-0.895*** (0.198)	-2.497*** (0.620)	-2.064*** (0.681)		-1.638*** (0.119)	-1.713*** (0.243)	-1.713*** (0.248)		-0.315 (0.002)	-2.888** (1.192)	-1.832 (1.420)
<b>dance</b>		0.010 (0.228)	0.380 (0.362)	0.741 (0.476)		0.248 (0.205)	0.500 (0.328)	0.494 (0.329)		0.175 (0.335)	0.489 (0.565)	0.929 (0.597)
<b>education</b>			0.190 (0.250)	0.144 (0.298)			-0.186* (0.102)	-0.195* (0.104)			0.381 (0.441)	0.431 (0.562)
<b>entries</b>			0.032 (0.024)	0.026 (0.028)			0.087*** (0.032)	0.007** (0.003)			0.075 (0.068)	0.140* (0.810)
<b>age</b>			0.006 (0.029)	0.012 (0.304)			0.017* (0.009)	0.018* (0.009)			0.009 (0.069)	-0.044 (0.077)
<b>male</b>			0.475 (0.487)	0.304 (0.498)			-0.286** (0.145)	-0.301** (0.147)			0.248 (1.421)	2.926 (1.187)
<b>married</b>			0.654 (0.499)	0.573 (0.549)			0.002 (0.196)	-0.009 (0.201)			-0.612 (1.040)	0.659 (1.226)
<b>worker</b>			0.498 (0.649)	0.604 (0.675)			0.201 (0.224)	0.208 (0.225)			0.957 (1.741)	1.747 (1.996)
<b>citizenship</b>			-0.950 (1.238)	-1.099 (1.272)			0.373 (0.350)	0.363 (0.350)			-	-
<b>neighborhood</b>	NO	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES
<b>environmental</b>	NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	YES
<b>N.observations</b>	1,382	1,382	638	480	3,197	3,185	1,424	1,424	620	620	260	260
<b>Pseudo R<sup>2</sup></b>	2.73%	8.83%	16.28%	20.99%	3.19%	22.81%	26.24%	26.31%	4.96%	7.25%	14.56%	25.01%

Notes: Standard errors in brackets. \*Significant at 10% level. \*\*Significant at 5% level. \*\*\*Significant at 1% level.

Table 3.5: Control regressions of effect of influence on renewal decisions with CEM. For each of the three tests four models increasingly control for influence, characteristics of both activities and individuals, and common environmental factors.



### 3.4.3 Confounding effects: customer profile and product features

Table 3.6 details the characteristics of homophily: customers' socio-demographic characteristics, location, contract use, and type of activity. The balance of the propensity score matching was satisfied when the four aspects were differentiated, and the positive value of ATT is confirmed. After controlling for individual characteristics and activities, the effect of peer influence on renewal decisions was lower than in the other two controls, demonstrating the importance of individual characteristics and type of activities in explaining customers' choices (Proposition 3.2).

	Mean Pscore	Std.Dev. Pscore	Balance property	ATT	treated	controls
<b>Product type</b>	0.106	0.043	YES***	0.181*** (0.017)	421	3,528
<b>Customer profile</b>	0.104	0.044	YES***	0.182*** (0.024)	234	1,961
<b>Customer location</b>	0.091	0.025	YES***	0.252*** (0.022)	344	3,400
<b>Contract usage</b>	0.095	0.015	YES***	0.305*** (0.020)	421	2,853

Notes: Standard errors in brackets. \*Significant at 10% level. \*\*Significant at 5% level. \*\*\*Significant at 1% level.

Table 3.6: Stratification Propensity Score Matching.

### 3.4.4 Timely peer-influence

Figure 3.3 shows the trend in the impact coefficient of treatment  $T$  on the probability of renewal  $R$  when various values of  $p_1$ ,  $p_2$  and  $p_3$ . Three cases are considered: first, both  $p_1$  and  $p_2$  may change, and are both equal to the same number of days  $\alpha$ ; second,  $p_1 = 15$  and  $p_2$  varies ( $p_2 = \alpha$ ), and vice versa. The reported impact coefficient is the coefficient of a logit regression with  $R$  as dependent variable and  $T(p_1, p_2, p_3)$  as the only independent variable. The impact coefficient is decreasing in all three cases, although the decline of the

treatment effect is less marked when  $p_1$  is taken as fixed. The impact of being treated on the renewal probability decreases when either the number of days remaining until the expiry of the treated contract or, although less significantly, the number of days already passed since subject the treater renewed, or both increase.

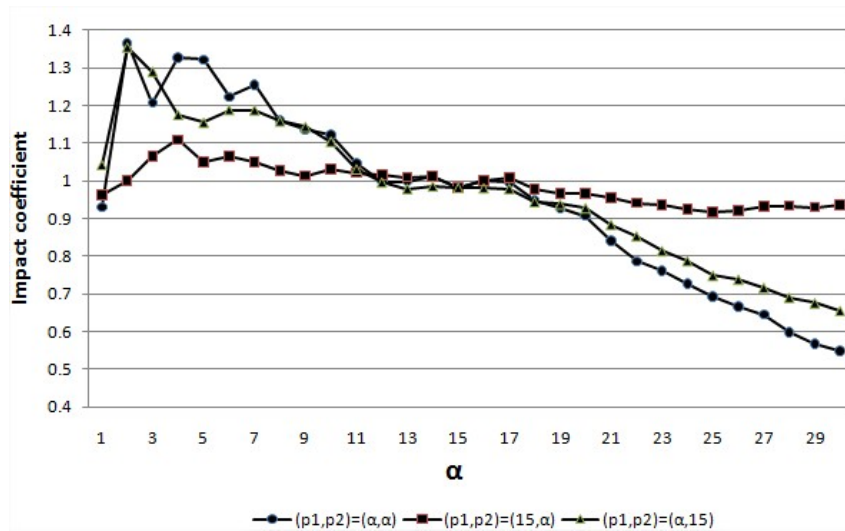


Figure 3.3: Impact of treatment on probability of renewal for different values of  $p_1$  (number of days until the expiry of  $i$ 's contract) and  $p_2$  (number of days already passed since subject  $j$  renewed);  $p_3 = 5$  minutes.

The robustness of the results in Sections 3.4.2 and 3.4.3 are confirmed for both propensity score matching and CEM analyses when various values of  $p_1$  and  $p_2$  are applied. More specifically, in order to test Proposition 3.3, we report the results obtained by applying the former when different values of  $p_1$  and  $p_2$  are applied (Table 3.7). We consider three ranges in the values of  $p_1$  and  $p_2$  and therefore derive three new intertemporal treatments: we take  $p_3$  as fixed and equal to 5 minutes, defined as  $1 \leq (p_1, p_2) \leq 5$ ,  $5 < (p_1, p_2) \leq 15$ , and  $15 < (p_1, p_2) \leq 30$  to define the three new treatments. In all three tests for environmental effects, homophily, and both effects there is a decreasing effect of influence on the probability of renewal. In addition, ATT becomes negative in the case in which  $i$ 's contract expires in more than two weeks and

$j$ 's contract expired more than two weeks before their common entry. The treatment must thus be applied within two weeks in order to be effective.

Proposition 3.3 is therefore confirmed, because we prove that the effect of peer influence in determining the alignment in individual decisions to renew decreases with the distance between customer renewal decisions. Also in this case, we test the effect of treatment on renewal probability by separately controlling for homophily characteristics and differentiating among homophily effects (lower part of Table 3.7). For all four groups of homophily variables, we see a decrease in the effect of influence on renewal decisions and the ATTs become negative when more than two weeks are chosen for defining the treatment.

### **3.4.5 Observational learning attenuates overestimation of attendance**

Della Vigna and Malmendier (2006) found that club members are overconfident about their effective capability to go to the club. Overconfidence is defined on the price per entry (Della Vigna and Malmendier 2006). By analysing the data we find that treated subjects pay less than half what non-treated members pay (20 vs 9.55 euro). Treated members are not overconfident because, after signing their contracts, for each entry they pay less than what they would have to pay for a one-entry ticket (10 euro).

To check more rigorously if the treatment has an impact on the price per entry we define a dummy variable equal to 1 if the price per entry for the new contract is higher than the price per entry of the previous expired contract, 0 otherwise. The dummy can be seen as a proxy for individual overestimation of attendance and it is used as dependent variable in the further analysis. The first column in Table 3.8 reports the logit regressions aimed to analyse

CHAPTER 3. PEER INFLUENCE IN CO-PRESENCE NETWORKS

	Time window	Mean Pscore	Std.Dev. Pscore	Balance property	ATT	Treated	Controls
<b>Environm. effects</b>	within 5 days	0.128	0.016	YES***	0.306*** (0.018)	421	2,801
	5-15 days	0.317	0.019	YES***	0.226*** (0.019)	1,308	2,229
	15-30 days	0.288	0.019	NO	-0.044*** (0.029)	943	2,327
<b>Homophily</b>	within 5 days	0.101	0.085	NO	0.181*** (0.029)	165	1,310
	5-15 days	0.271	0.106	NO	0.113*** (0.023)	487	1,274
	15-30 days	0.243	0.098	NO	-0.106*** (0.026)	434	1,347
<b>Both effects</b>	within 5 days	0.138	0.109	YES***	0.187*** (0.029)	164	871
	5-15 days	0.334	0.124	NO	0.094*** (0.026)	488	945
	15-30 days	0.299	0.109	NO	-0.100*** (0.026)	435	995
<b>Product type</b>	within 5 days	0.106	0.043	YES***	0.181*** (0.017)	421	3,528
	5-15 days	0.259	0.032	YES***	0.155*** (0.016)	1,308	2,957
	15-30 days	0.236	0.040	YES***	-0.026* (0.017)	942	3,099
<b>Customer profile</b>	within 5 days	0.104	0.044	YES***	0.182*** (0.024)	234	1,961
	5-15 days	0.250	0.050	YES***	0.128*** (0.021)	559	1,662
	15-30 days	0.222	0.040	YES***	-0.136*** (0.023)	496	1,730
<b>Customer location</b>	within 5 days	0.091	0.025	YES***	0.252*** (0.021)	344	3,400
	5-15 days	0.238	0.024	YES***	0.134*** (0.026)	892	2,852
	15-30 days	0.215	0.030	YES***	-0.081*** (0.020)	805	2,939
<b>Contract usage</b>	within 5 days	0.095	0.015	YES***	0.305*** (0.020)	421	2,853
	5-15 days	0.236	0.050	YES***	0.223*** (0.023)	1,038	2,231
	15-30 days	0.214	0.010	YES***	-0.037** (0.017)	943	2,325

Notes: Standard errors in brackets. \*Significant at 10% level. \*\*Significant at 5% level. \*\*\*Significant at 1% level.

Table 3.7: Discrimination on time with the Propensity Score Matching.

how individual overestimation is conditioned by influential dynamics. The variable price and number of entries are reported in the regression because not significantly correlated with the dependent variables. If dropped, the results remain robust.

<b>Logit model</b>	<b>Full model</b>	<b>Homophily model</b>
<b>Influence</b>	-0.782*** (0.220)	-0.702*** (0.259)
<b>price</b>	0.008*** (0.001)	0.033*** (0.005)
<b>dance</b>	-0.223 (0.224)	-0.165 (0.268)
<b>gym</b>	0.121 (0.296)	-1.631*** (0.602)
<b>age</b>	-0.006 (0.014)	0.049 (0.030)
<b>gender</b>	0.097 (0.215)	2.758*** (0.599)
<b>married</b>	0.037 (0.284)	1.887*** (0.537)
<b>worker</b>	0.348 (0.333)	1.173* (0.706)
<b>Italian citizen</b>	0.315 (0.480)	0.528 (1.268)
<b>certificate of studies</b>	0.082 (0.129)	0.320 (0.237)
<b>Pistoia center</b>	-0.165 (0.196)	-0.570* (0.329)
<b>number of entries</b>	-0.066*** (0.010)	-0.218*** (0.032)
<b>time of entry</b>	-0.078** (0.034)	-0.025 (0.065)
<b>noise</b>	-0.002 (0.036)	0.065 (0.079)
<b>N.observations</b>	799	438
<b>Pseudo <math>R^2</math></b>	12.63%	18.50%

Notes: Standard errors in parentheses. \*Significant at the 10-percent level. \*\*Significant at the 5-percent level. \*\*\*Significant at the 1-percent level.

Table 3.8: The impact of influence on price per entry

The results confirm Proposition 3.4 and show that observational learning reduces overestimation of attendance by helping subjects to take more efficient decisions about contract usage. Influence has indeed a significant and negative impact on price per entry and on the overestimation of attendance. Therefore, because of the strong tie linking  $i$  and  $j$ , originating influence, the probability

that  $i$ 's price for each entry during the new contract is lower than that in the old contract is higher when  $i$ 's and  $j$ 's decisions are aligned, rather than otherwise. Indeed, the motivation to attend the gym increases when other subjects having influence power attend themselves. As expected the likelihood of taking inefficient decisions decreases when subjects have renewer neighbors. This means, on the one hand, that treated subjects not only are more likely to renew, but that, when they do, the purchasing decisions are more aligned with their previous individual expectations. On the other hand, untreated renewers have a higher risk to take unefficient decisions about future attendance. Moreover, the price of the expired contract and the number of entries has a significant effect on attendance. In particular, the higher the price of the expired contract, and therefore the longer the contract, the higher the level of overconfidence. Indeed, subjects subscribing long term and expensive contracts have a higher probability to not be able to satisfy their early expected commitment. Moreover, those subjects who entered more times during the previous expired contract period are less overconfident because they had more chances to test their commitment over time. Finally, the results show that overconfidence is also significantly affected by some individual characteristics, such as gender, marriage and employment status, and the address.

Now we apply the same methodology to test in particular how similarities among subjects impact individual overconfidence. In this analysis we therefore consider only the variables characterizing the previous homophily test. We compute the propensity score matching and also in this case the standard errors of the treatment effects are bootstrapped and 100 replications of the bootstrap are performed. Despite the fact that the balancing property in propensity score matching is not satisfied at 10% level, the ATT is significantly negative at the 1% level and it is equal to -0.241. Therefore, we confirm the result according to which influence has a negative impact on individual

overestimation of attendance. Even after controlling for homophily, subjects having neighbors who renewed their contracts (treated) are more confident on their future commitment. The results are confirmed by computing CEM. After reducing the imbalance of 13%, the basic model  $dummy = a + b \cdot influence + e$  shows a negative impact of the treatment on overconfidence: at the 1% significance level we find  $b = -1.165$ . The second column in Table 3.8 reports the complete regression computed to check for remaining imbalance. The results are confirmed.

This preliminary findings show that influential dynamics mitigate overconfidence. Further work is needed to better understand the relationship between observational learning and overestimation of attendance. For example, more effort is needed to determine whether the decrease of evaluation errors for treated subjects is due either to a improvement in the available information set through social ties, or to socialization per se as a key element for customer satisfaction.

### 3.5 Concluding Remarks

This work addresses some of the question arised in recent debates in the literature of network diffusion about the discrimination among peer influence, homophily, and other confounding effects (Iyengar and Van den Bulte 2011, Aral 2011, Christakis and Fowler 2011). Our results corroborate the findings of Iyengar et al. (2011) for marketing practice by finding evidence of contagion through social ties. We demonstrate that a person's decision to belong to a club and to renew membership is largely influenced by the level of commitment of that person's reference group. We prove that members tend to align their commitment to the socially related group of co-enterers, independently of the similarities among them and their exposure to the same external stimuli.

Moreover, we answer to the emerging doubts on the lack of accuracy in the econometric procedures by applying rigorous matching models to study influence dynamics in social contexts. By defining *influence* as the presence of renewer neighbors in the co-presence network, we use these tools to rigorously discriminate between peer influence, homophily and environmental effects, which may be wrongly considered as peer influence. In addition, we contribute to the literature by disentangling various types of homophily and environmental effects, and confirming the importance of individual characteristics and product type as sources of homophily.

To our knowledge, this is one of the first attempts to examine the emergence of influence in a dynamic real-world co-presence networks. The analysis of reality-mined network data can indeed radically improve the knowledge about the effects of social ties and contagion dynamics on human decision-making. We contribute to the debate by exploiting a dynamic dataset and computing co-presence networks by the mean of instruments used in social biology aimed at measuring individual associations. By analysing the data over time, we find that treatment, used for subjects having neighbors who renewed their contracts, must be applied within two weeks from the co-entry in order to be effective and that the effect of peer influence in determining the alignment in individual decisions to renew decreases with the distance between renewal decisions.

Lastly, our results show that influence does have a significant and positive effect on attendance: renewers without renewer neighbors pay a higher price per entry for their new contracts and therefore use them inefficiently.

This work has limitations as well. Despite the richness of our dataset, we are not able to deal with some emerging questions about the discrimination between the roles played by influentials and influentiabiles in the contagion process. In addition, as underlined by Aral (2011) further work is also needed



to differentiate among the various processes of peer influences, such as persuasion, imitation, and social learning. The framework proposed in this paper allows to measure observational learning but we cannot exclude other forms of direct social learning (Zhang 2010). The use of active RFID can be helpful in analysing the magnitude of the role played by direct and indirect influence. Moreover questionnaires can be implemented to better define emerging social ties among subjects during the time spent in the club.



# Chapter 4

## Drilling Wells and Killing Wells through Knowledge Networks

### Abstract

The rapid recombination of dispersed knowledge in multinational companies operating in risk sectors is crucial in responding to unexpected and urgent crises. The communities of practice (CoPs) introduced by a company operating in the energy sectors are analyzed to investigate how their structure and the roles and positions of their members improve a quickly access to dispersed knowledge. By overlapping multiple networks of co-authorships, e-mail exchange and communities of practice, we find the emergence of structural holes among the phases of research and practice in the innovation process. There are still inefficiencies in the introduction of on-line communities of practice as tools to facilitate access to critical knowledge and to speed problem-solving procedures, especially for multinational companies operating in risk sectors.

**Keywords:** communities of practice, recombination of knowledge, multiple network analysis, structural holes

### 4.1 Introduction

On April 20 2010, failure of the blowout preventers (BOP) <sup>1</sup> caused an explosion and the release of 4.9 million barrels of oil in the transocean drilling rig licensed by British Petroleum (BP), the Deep Water Horizon, in the Gulf

---

<sup>1</sup>In order to contrast underground pressure, underwater valves called Blowout Preventers (BOP) are placed on the top of wellheads to prevent oil and gas escaping into the sea.

of Mexico. Due to a leak in the wellhead, the BP oil spill flowed for more than four months, until the resolute decision to cap the wellhead and the end of the spill process on September 19 2010<sup>2</sup>. The BP disaster has been recognized as the largest accidental oil spill in the history of the petroleum industry and, by July 2011, several countries were still contaminated by BP oil. Many reports have been written on the multiple causes of the damage, at both technical and managerial levels. In addition to BP's lack of testing BOPs, repeated failures in the evaluation of risks, costs, environmental impact, and the financial uncertainties of drilling offshore wells contributed toward worsening and extending the effects of the explosion and lowering managerial reactions. The slow responses and lack of urgency in BP's reactions have been widely debated and criticizes especially because of the declared predictability of some of the problems encountered. In addition, according to its CEO, Tony Hayward, BP made public relations mistakes in responding to the crisis, by communicating unfounded optimism during the many failed attempts to kill the well.

This paper examines the specific problem of rapid mobilization of dispersed knowledge in complex and global organizations operating in risk sectors and facing emergencies by focusing on network ties. Recent debates have proposed an overview of BP's changes in knowledge management (KM) programs and discussed the inefficient and late reactions of both managers and experts in problem-solving, because of the inability to recombine previously acquired knowledge rapidly (Mazzola and Distefano n.d., Schwartzman et al. 2011). Because of BPs' failure to manage the information system properly, the study of the knowledge spillover in this sector is especially interesting. More specifically, the aim of this work is to verify the efficiency of the knowledge management

---

<sup>2</sup>For a detailed description regarding the time-line of the BP disaster, see <http://www.guardian.co.uk/environment/2010/jun/29/bp-oil-spill-timeline-deepwater-horizon> accessed on November 8 2011.

system (KMS) and to check whether the introduction of specific KM tools can facilitate the recombination of coded and tacit knowledge by improving diffusion systems and reducing mobilization problems.

Not all knowledge is in fact coded, and many tools have been developed to create organizational designs based on the exchange of knowledge, which can be either universal and reciprocal (open science, i.e., scientific publishing) or private and trust-based (private science, i.e., patenting). Recently, the concept of knowledge as a private good, either owned by a single subject or interpreted as a commodity, has been replaced by that of knowledge as a public good, embedded in a community and shared among its members (Brown and Duguid 1991, Lave and Wenger 1991). In particular, in the last decade KM literature has increasingly focused on Communities of Practice (CoPs)<sup>3</sup> as innovative tools to improve knowledge exchange in organizational settings, and a whole theory has been developed to better understand organizational learning through CoPs (Fox 2000). The concept of CoPs as scientific collaboration networks among experts helps to identify and activate required knowledge by facilitating the recombination of experts' knowledge (Breschi and Lissoni 2005, Newman 2004, Newman 2001), increasing the extent to which experts are likely to share it (Newman 2001, Newman 2004, Breschi and Lissoni 2005), and balancing the exploitation of existing knowledge and the exploration of new (Lave and Wenger 1991).

---

<sup>3</sup>In the early 1990s Lave and Wenger (1991) introduced the term *community of practice* (CoP) to indicate an active system of relationships among subjects sharing the same activities in the same organizational unit, across units, or across company boundaries. According to the needs of multinational companies, Markus (2001) rearranged the definition proposed by Lave and Wenger (1991) by adding an across-boundary dimension, and defined a CoP as a group formed of subjects occupying the same position in various geographical locations for various companies and organizational units. To emphasize the importance of ICT in managing information systems in multinational companies, Wasko and Faraj (2005) defined an *electronic network of practice* as a self-organizing system, an open activity system focusing on a shared practice which exists primarily through computer-mediated communication. See Handley et al. (2006) and Amin and Roberts (2008) for an overview of the various definitions of *communities of practice*.

In 1994, BP implemented a “virtual teamwork” program based on the creation of drilling CoPs and the coordination of employees dispersed worldwide on oilrigs (Leavitt 2002). After the spill, despite an initial denial by the BP CEO, knowledge spillover collapsed and prompt coordination for problem-solving also failed, because of the inability to recombine previously acquired knowledge and collect new knowledge in a short period of time. The defect in KMS highlighted the need to study the information system in the energy sector, to avoid other failures and to improve KMS by facilitating rapid identification and activation of experts. BP’s failure to plug the leak could not be due only to incompetence, lack of concern or missing tools, but also to the inability or impossibility of identifying and reaching those technical experts able to recognize and solve emerging problems. Problems in recruiting key competences and in sharing knowledge even through CoP formation must be carefully examined. The need to identify and activate experts rapidly is an additional crucial aspect for problem-solving in increasingly dynamic environments, and problems arise whenever the two domains of information and knowledge exchange rates, and the identification and competence of experts do not overlap.

The focus of this paper is to better understand the efficacy of on-line CoPs introduced by a multinational company operating in the energy sector, to facilitate access to critical knowledge and easy rapid problem-solving procedures. In particular, five main issues are examined: (1) Who are the experts in specific topics? (2) Which role should they play in the activity of CoPs? (3) How active or proactive are they in sharing their knowledge? (4) Is their knowledge easily accessible to other colleagues? (5) How do communities improve the overlap between the experts’ domain and the e-mail exchange domain? This paper contributes to the existing literature by adding multiple levels in examining the overlap among the domains of research and practice and the problems

related to this topic. By exploring a biannual dataset of e-mail exchanges, the KM network is studied to examine the domain of the knowledge spillover both within and among CoPs, and the degree of knowledge diffusion within the whole company. A second domain is also investigated by taking into account the experts' domain and considering employees' productivity in patenting and publishing. To our knowledge, this is one of the first attempts to investigate simultaneously multiple networks of co-authorships, e-mail networks and CoPs, to check for improvements in KM tools.

The paper is organized as follows. Section 4.2 overviews the relative literature on recombination of knowledge through CoPs. Sections 4.3 and 4.4 describe the contextual and empirical frameworks, research questions and data sources. Section 4.5 provides the qualitative and quantitative results. Section 4.6 concludes.

## 4.2 Literature review

Dispersed knowledge is not easily reproducible because of spatial and functional dispersion in competitive environments characterized by rapid changes. The need to identify critical knowledge quickly is a key factor in facing unexpected contingencies and fast decision-making processes. Organizational efficiency, innovation and creativity are improved by reusing, merging and recombining existing knowledge, and the acquisition of new expertise given by interactions among members makes CoPs efficient platforms for improving organizational learning and recombining of organizational knowledge.

The mobilization of both external and internal knowledge requires discovering and recombining available knowledge. The components of the recombination process are usually called "factors" (Schumpeter 1939) or "components" (Fleming and Sorenson 2001). The elements entering the recombination pro-

cess are “any bits of knowledge or matter that inventors might use to build inventions” (Fleming and Sorenson 2004) or resources, skills, knowledge and technical systems which are effectively integrated in new and flexible ways to develop new component competencies (Henderson and Cockburn 1994). The components involved in the recombination process may be existing components, previously untried components, or new components created by the inventor. The choice of components is usually based on their availability, proximity and saliency according to the inventor’s aims, and the availability of the components is constrained by science, available skills, and culture (Fleming and Sorenson 2001). In particular, the availability of components at the time of recombination is based on two mechanisms: exploitation, or local search, and exploration (March 1991). Exploration vs exploitation analysis helps us to understand how to deal with various types of emerging problems within firms, such as completely new problems not previously defined (exploration) and specific problems which may be grouped into general issues and solutions in the neighborhood of the company’s current expertise (exploitation). Therefore, learning in organizational contexts may be improved by two factors: exploring new possibilities, still unknown, and exploiting old information, resources and routines. This may be translated into a decision by the firm to balance resources and efforts between the improvement, investigation and recombination of old processes and technologies, and the invention of new ones. For each type of problem, various types of knowledge and organizational dynamics are needed. Several network organizations and experts can be activated by each kind of problem. New knowledge, or creative knowledge, can indeed be created and shared by recombining peers’ knowledge and expertise (Fleming et al. 2007).

Strategy research has developed the implications of these ideas for firms and examined the main ways of defining their ability to go beyond local searches to



look for new components and implement new recombinations and reconfigurations<sup>4</sup>. In his works on firm growth, Penrose (1959) defined resource recombination as the main source of innovation in firms: available resources may indeed be recombined among themselves or with other resources in various ways, for rapid solutions in decision-making processes. Creative solutions to emerging contingencies do depend on combining existing and new knowledge (Pisano et al. 1997). The reconfiguration of assets, both local and across boundaries (Rosenkopf and Nerkar 2001) and the resulting innovation are therefore also key factors in increasing competitive advantages.

Despite efforts in examining the dynamics of individual creativity and upcoming knowledge in social contexts by analysing how networks influence the flow of information and ideas in recombinant search processes (Burt 2004, Uzzi and Spiro 2005, Fleming et al. 2007), only recently have collective dynamics of creativity been taken into account (Kurtzberg and Amabile 2001, Padgett and Powell 2009, Frigotto and Riccaboni 2010). In an organizational setting, the recombination process within firms is improved by introducing internal communities and boosting multiple network ties among their members. Members' participation in CoPs and frequent contacts in sharing information are necessary requirements for knowledge generation and the identification of creative solutions to emerging problems (Wenger 1998). However, CoPs are usually formed of small groups of subjects working on the same problems, a small number of whom actively participate in community life; the majority is mostly passive in relating with colleagues to deal with emerging issues<sup>5</sup> (Wenger et al. 2002).

---

<sup>4</sup>The ability to synthesize, practically use and implement existing and acquired knowledge has been variously defined in the strategic literature as *combinative capabilities* (Kogut and Zander 1992), *dynamic capabilities* (Pisano et al. 1997) and *architectural competences* (Henderson and Cockburn 1994).

<sup>5</sup>Literature on CoP theory categorizes communities' members as peripheral, full, or marginal, according to their level of participation in CoP activities (Wenger 1998, Handley et al. 2006).

The effectivity of CoPs in improving learning in practice and creativity increases whenever managerial interventions are put in place to motivate individual sharing through continuous recombination processes within the organization and by favouring social influence and intra-community ties (Wellman and Wortley 1990). The managerial need for intervention in knowledge-sharing is even more important for multinational companies operating worldwide to control knowledge flow. In such a context, professional virtual communities are widely implemented to collect knowledge from a specific area of expertise for problem-solving at work. The increased level of globalization and internationalization in which multinational companies operate enhances the crucial role of knowledge-sharing, because of the higher percentage of competitive advantages deriving from tacit knowledge shared through communications, feedback and networking. Once employees' willingness to contribute to the on-line system has been guaranteed (Sharratt and Usoro 2003, Jian and Jeffres 2006), on-line communication technologies can overcome difficulties in international sharing and improve interactivity among subjects. In the last decade, KM literature has focused on virtual CoPs and other web-based knowledge-sharing systems as the main tool in improving international exchange of knowledge among far-away subjects and circulation of best practices (Jarvenpaa and Staples 2000, Ardichvili et al. 2003, Bansler and Havn 2003, Ardichvili et al. 2006). Empirical studies have examined the application of virtual CoPs in KM systems in multinational companies, both in the IT sector - such as Hewlett-Packard (Sieloff 1999), Xerox and IBM - but also in the energy sector, such as Chevron, British Petroleum (Cohen and Prusak 1996) and Shell (Haimila 2001). In some empirical studies, positive effects emerged after the introduction of CoPs, because of the recombination of members' knowledge confirmed the positive impact of CoPs on organizational performance in stable environments (Schenkel and Teigland 2008).

However, many failures have been recorded in implementing of both virtual and conventional CoPs. Despite many empirical findings, a few theoretical studies focus on the reasons for the success or failure of CoPs. Ardichvili et al. (2003) examined knowledge-sharing in a multinational corporation and focused on motivations and barriers for knowledge-sharing within CoPs. The literature has little evidence as regard how and why workers decide to exchange and share knowledge (McDermott 1999). The major failures in CoPs implementation derive from inhospitable organizational contexts, socio-cultural factors, and the egoistic behavior of subjects, who exchange knowledge through private relationships (Roberts 2006). Despite companies' effort and investments aimed at improving internal communications (Szulanski 1996), subjects may not be willing to share knowledge within the whole organization because of both individual and organizational barriers<sup>6</sup>. Some of these issues are solved through governance mechanisms and incentive systems aimed at improving both the intrinsic and extrinsic motivation of the subjects involved and at encouraging participation (Wenger 1998, Osterloh and Frey 2000). Lastly, in addition to subjects' willingness to share knowledge (Brown and Duguid 1991, Wenger et al. 2002), benefits from the communities arise only when subjects overcome their preference for private knowledge and recognize CoPs as important knowledge-sharing instruments to spread ideas and solve problems (Ardichvili et al. 2003).

---

<sup>6</sup>Cultural influences (Ardichvili et al. 2006) and emotional and professional insecurity (Brown and Duguid 1991, Brown and Duguid 2001, Ardichvili et al. 2003) are some of the main examples of individual barriers to knowledge-sharing. Obsolete technologies and systems (Ardichvili et al. 2003), inappropriate culture and climate (De Long and Fahey 2000) and security issues are the main organizational factors reducing knowledge exchange.

### 4.3 Research questions and contextual framework

Communities are usually formed as *small-worlds*, structured by strong ties among the members of each community and weak ties connecting the communities with each other, in both scientific and organizational contexts (Granovetter 1973, Watts and Strogatz 1998b, Newman 2001). In this framework, recombination of knowledge is improved by the presence of subjects bridging communities and filling two types of structural holes. First, at a horizontal level, structural holes appear within one single network whenever difficulties emerge in managing the one single kind of relation linking members within it. However, the interpretation of CoPs as unique, specialized groups of affiliation is a myopic view of the various types of relations and ties needed to improve internal cohesion and homogenization of knowledge within a company. Although the usual organizational procedures are based on increasing specialization of individual roles, each subject involved in the innovation process should be involved in multiple roles and relations, to improve communications in the stages forming the innovation process. Whenever multiple networks are introduced, multiple relations among subjects may give rise to the appearance of structural holes among multiple networks (Padgett and Ansell 1993). In a multinational company operating in a complex, risk sector, both problems emerge, and the importance of recombining knowledge across all the stages of the innovation process is crucial to facilitate the generation of innovative solutions to emerging problems. The recent concept of *democratizing innovation* (or *user-centered innovation*<sup>7</sup>) and the increasing topicality of the Chain-Linked Model introduced by Kline and Rosenberg (1986) empha-

---

<sup>7</sup>This concept is based on a bottom-up perspective, in which the role of final consumers is crucial in innovation and KM processes. Due to improvements in ICT, users are increasingly able to combine their efforts and knowledge to develop their own products (Von Hippel 2009).

### 4.3. RESEARCH QUESTIONS AND CONTEXTUAL FRAMEWORK

size the increasing importance of improving feedback across all the stages of processes. The model describes innovation as a continuous process in which the required science, supporting applied research, and various types of necessary scientific knowledge are linked, through either direct communications or feedback, to guarantee a successful flow of information along the entire process (Figure 4.1).

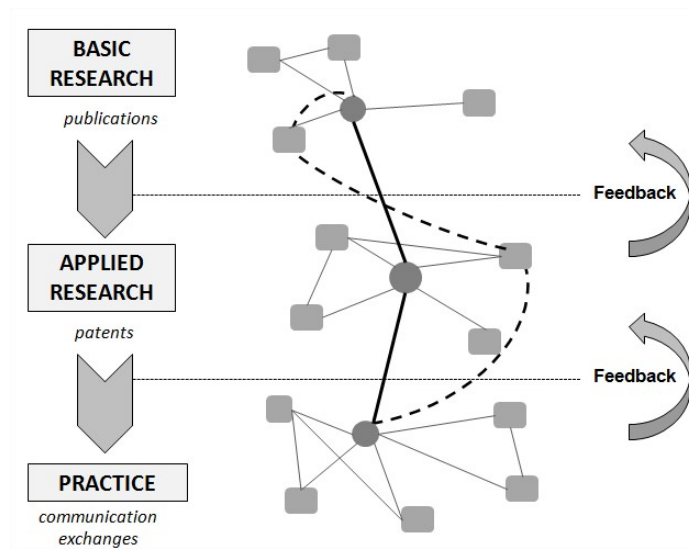


Figure 4.1: Top-down and bottom-up information flows in the Chain-Linked Model

This bottom-up perspective is not easy to implement in a multinational company, and the required variety in individual roles is difficult to relate to fixed internal communities. Indeed, vertical organization usually faces limitations and difficulties in combining basic (i.e., scientific publishing) and applied (i.e., patenting) research with more practical stages of the innovation process, because of the variety of communities and languages. This study examines both horizontal and vertical spillovers by focusing on five main issues regarding both CoPs and multiple networks (patents, publications, and e-mail exchange).

First, information and knowledge flow within a multinational company may be limited by communicational problems at a vertical level among basic re-

search, applied research and production in the various stages of the innovation process. However, the speed of communications can be improved whenever subjects are simultaneously central in multiple networks across the phases of the innovation process<sup>8</sup>. Their presence improves the speed of information and feedback, which would otherwise be lengthened, reduced or stopped. In addition, the efficiency of the communities of practice in linking the innovation stages is highly improved if central experts are involved: whenever a problem emerges, the experts in that field can be rapidly identified and reached through CoPs information exchange and the problem quickly solved.

**1 Representativeness.** *Are the top experts central in the networks to which they belong? Do the top scientists and inventors have official roles either as company experts or in CoP activity? Is there a core group of subjects supporting CoPs activity?*

Second, the results of Padgett and Ansell (1993) improved previous research in social network studies by emphasizing not only problems related to structural holes within single networks (Burt 1992) but also by adding overlapping dimensions and generalizing structural holes across networks. In this context, a simultaneous analysis of the information exchange network and co-authorship networks is appropriate, to check for the simultaneous presence of subjects in the networks, or *network overlap*.

**2 Presence.** *Is there any overlap between the domains of experts and of knowledge exchange? How many scientists and inventors are involved in e-mail exchange? How many top experts are simultaneously embedded in the scientific and sharing domains?*

---

<sup>8</sup>See black continuous lines connecting round node versus dotted lines linking no central subjects in Figure 4.1.

Third, once the presence and representativeness of top experts is tested, further analyses are needed to check for effective activation of these knowledge owners in sharing their skills and competences. Indeed, in spontaneous communities, knowledge exchange may be lower because of insufficient participation of the crucial subjects involved (Brown and Duguid 1991, Lave and Wenger 1991). Indeed, in addition to problems due to the low quality of the system itself, which needs time-consuming effort to be learned and applied by the employees, Probst and Borzillo (2008) identified the main determinants of CoPs failure in members' networks and roles. By empirically analysing 57 CoPs in major worldwide companies, Probst and Borzillo (2008) proved that successful CoPs should be guided by one or a few leaders, who incentivate transfers and control community activity. Therefore, the presence of leaders guiding community activities and knowledge transfer is another important successful element in CoPs implementation.

**3 *Activation.*** *Are the top experts effectively proactive in sharing their technical knowledge? Are they likely to use CoPs as a KM tool to improve and speed information-sharing?*

Fourth, the check for structural holes in both vertical communications among the stages of the innovation process and difficulties in horizontal communications among CoPs allow us to verify the effectiveness of the structure of the KM system in facilitating individual access to knowledge. Despite the effective activation of top experts in sharing their knowledge, the flow of information may be guaranteed by experts' neighbors in the e-mail network who are highly proactive in information exchange.

**4 Retrieval.** *How difficult is it to access required knowledge?*

*And, even without the likelihood that experts will share their competences, is their knowledge easily reachable by colleagues using the e-mail exchange system?*

Lastly, we move from an individual to a community perspective. Communities of practice are important tools in improving negotiation and identity in organizational settings (Wenger 2007). The idea of CoPs was introduced as a useful instrument in improving learning-in-practice by facilitating the recombination of experts' knowledge and balancing the exploitation of existing knowledge and exploration of the new knowledge (Lave and Wenger 1991). Previous KM tools have been improved by the introduction of communities because of the reduction in informational overloading and increased member participation (Jarvenpaa and Staples 2000, Jian and Jeffres 2006). Indeed, in the latter case, knowledge is exchanged through discussions and networking, because of community's rather than individuals' interest, and knowledge spillover is improved within rather than among communities (Breschi and Lissoni 2005). Collaboration networks of scientists in various fields, such as biology, medicine, physics and computer science, have already been widely analysed in the literature and proved to have a *small world* structure (Milgram 1967, Watts and Strogatz 1998b). Because of the structural properties of co-authorship networks, data on co-publishing and co-patenting have also been used in network analysis to test *small-world* approaches (Watts and Strogatz 1998a, Burt 2004, Uzzi and Spiro 2005) and social network theories (Wasserman and Faust 1994). This paper contributes by verifying the "small world hypothesis" also in communities of practice and structurally analysing CoPs as bridges between research and practice.



**5 Structure.** *How do communities improve the overlap between the experts' and the e-mail exchange domains? Do communities behave as small worlds? Do they facilitate both bottom-up feedback and top-down information flows?*

## 4.4 Empirical framework

In this paper we identify networks of co-publishing, co-patenting, and an e-mail exchange network respectively, to deal with basic research, applied research, and the intra-organizational practices of a multinational company operating worldwide in the energy sector. The networks of co-authorships and communication exchange resulting from these databases are increasingly closed to external subjects: publications can be read, used, and applied by all subjects, even those external to universities or research centers, patents can only be consulted by subjects external to the assignee but not applied; and internal information exchange deals with more applied intra-organizational issues, to which external subjects have no access for either strategic or privacy reasons.

Despite emerging doubts (Breschi and Lissoni 2005), a wide area in the innovation literature exploits patent citation data and co-authorship networks as proxies for knowledge flows and creativity (Agrawal et al. 2003, Yoon and Park 2004, Breschi and Lissoni 2005, Fleming et al. 2007). Knowledge flows depend not only on individual positions in the organizational chart and in CoP structure, but also on the labor market of inventors (Zucker et al. 1998, Almeida and Kogut 1999, Breschi and Lissoni 2005, Breschi and Lissoni 2009). Also at community level, the mobilization of knowledge and the search for novelties and innovative solutions based on weak ties derive from the evolutionary learning of less experienced members from more experienced ones,

who are considered the experts in specific areas (Lave and Wenger 1991). The literature on CoPs has also been improved by patent data, proving the importance of social ties and inventors' locations in knowledge transfer processes and knowledge spillover across CoPs (Agrawal et al. 2003, Agrawal et al. 2008). Social links, trust, reciprocity, identification, shared vision and shared language in communities of experts have a very positive influence on individual decisions to share knowledge in on-line communities (Chiu et al. 2006).

However, clarification is required here. Basic and applied research knowledge is shared not only through patents and publications, but also through learning-by-doing processes (Arrow 1962, Young 1993). In addition, because of the potential conflict between public and private contributions in industrial firms, companies may be over-secretive about their research and implement different incentive policies to boost patenting, scientific publishing, or both (Lim 2000, Kinney et al. 2004, Van Looy et al. 2006). Companies' interests, individual carriers, and the need to share knowledge must be carefully balanced. The extent to which companies rely on patents and implement various incentive mechanisms to support research usually depends on the industry in which they compete, on the products and services offered, and on the degree of innovation implemented (Lim 2000). In fact, in some specific sectors or industrial processes, innovations are not always reported in patents but are held as trade secrets, and few direct incentives are provided to publishers within companies (Kinney et al. 2004).

The analysed company operates in the field of oil and gas, petrochemicals and oilfield services construction, and is committed to improve the exploration, production, transformation and transportation of oil and gas. It operates in more than 50 countries and has more than 70,000 employees. Despite the ongoing economic crisis and the instability of the costs of energy sources, the energy sector is rapidly evolving, due to improvements in drilling technologies

and to governmental and environmental interventions. This sector is becoming even more knowledge-intensive because of its high risk and complexity. The role played by KM systems and knowledge spillover is therefore even more crucial in problem-solving and particularly in immediate and short-term decision-making processes.

Three datasets are examined to deal with these five questions: one dataset of e-mail exchange and two datasets of co-authorships in patenting and scientific publishing. The first dataset contains e-mail exchanges from 09/01/2009 to 04/05/2010, provided by one of the main divisions of the company. The exchange only covers the internal e-mail accounts of the company, and no private accounts were included. The dataset contains information on 15 communities of practice. The communities covered both practical fields, i.e., production, materials, and constructions, and other fields based on research, i.e., geology, geophysics, etc. For each CoP, at least two supervisors, an official facilitator<sup>9</sup> and at least one alternate, were identified to guide and coordinate CoP activity. The official roles were assigned to 15 facilitators, 14 alternates, and 3 second alternates. The effectively active senders and receivers of e-mails in the division totalled 1,567 and exchanged 103,474 messages in the time window in question. In this framework, e-mails were considered as one-to-one communications: whenever the receiver was the general account of the CoP, the e-mail was sent to all subjects belonging to the receiving CoP. Several pieces of information are available for each message: date, time, subject, and attachments, but also information about the senders and receivers, i.e., provenience, whether employees or external subjects and the relative CoPs.

In addition to communication problems, explored through the previous

---

<sup>9</sup>Several studies have been developed to identify gatekeepers and to differentiate between those who do not actively participate in the community, and those who are pro-active and improve the informational exchange by contributing to select, identify and transmit information. These latter are also called *boundary spanners* (Fleming and Waguespack, 2005) or *facilitators*.

dataset, inefficient and late reactions to critical urgencies were also due to the inability to recombine previously acquired knowledge rapidly, because the right people were not identified or reached when they are needed. Technical experts must be easy to recognize and consult whenever a problem emerges in their area of expertise. The co-authorship networks aim at verifying the existence of experts by analysing the patents and publications produced by employees since 1996<sup>10</sup>. A publication database (SciVerse-Scopus<sup>11</sup>) and a patent dataset (Delphion<sup>12</sup>) were used to obtain information on individual patent and publication rates, respectively. The former was an abstract and citation database of peer-reviewed literature and quality web sources, providing information about authors, co-authors, affiliation networks, citations, and other publication indexes. The latter provided collections of patents and analytical and productivity tools for patent analysis. These datasets gave the identity of those subjects who, according to their frequency of patenting and publishing, were considered here as experts. However, the experts defined according to these two criteria do not always coincide with the 49 skilled technicians appointed by the management of the company as official experienced specialists.

In the time window in question, Delphion reports 250 single patents produced by members of the analysed company, with a maximum number of patents per author of 29 and 305 inventors. The dataset on scientific publications lists 1,703 publishers and 1,638 publications recorded as documents having the analysed company as main affiliation. The overlap of the three networks was computed after a manual matching of each employee across the networks involved.

---

<sup>10</sup>The choice of the time window was due to the availability of data. In addition, knowledge needs a long time to ripen, and examination of individual productivity must therefore also focus on longer period previous the introduction of CoPs.

<sup>11</sup>SciVerse-Scopus registered by Elsevier Properties S.A., is accessible at: <http://www.scopus.com/home.url> (last access on August 10 2010).

<sup>12</sup>The link to Delphion by Thomson Reuters is <http://www.delphion.com> (last access on July 11 2010)

## 4.5 Results

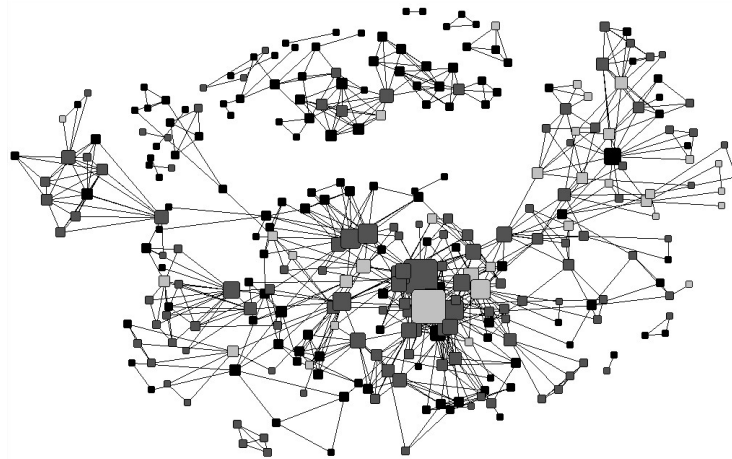
In our framework, social network instruments were used to study the datasets and answer the research questions. Three networks were examined to check for vertical integration in the innovation process: one patent network (P), one scientific network of publications (S), and one e-mail exchange network (E).

### 4.5.1 Representativeness in the domain of experts

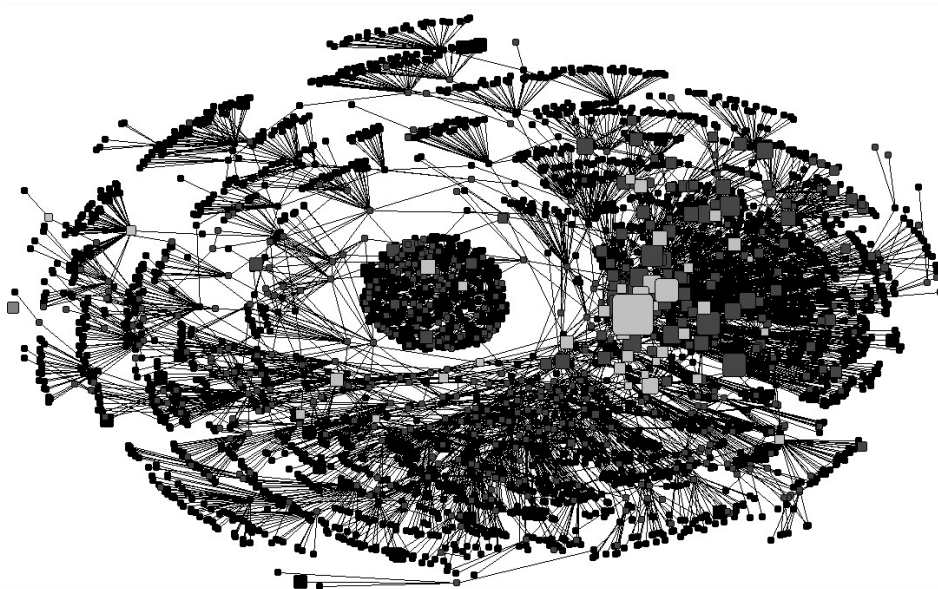
The presence of experts in the analysed company was tested by analysing patents and publications since 1996. All these subjects were considered as experts in their publication fields. A summary of the fields in which company employees published is provided by Scopus: more than half the publications deal with chemical engineering, chemistry, earth and planetary science, material science, and energy.

The network of patents (P) and the scientific network of co-authorships (S) are shown in Figures 4.2a and 4.2b and respectively represent co-authorships in patenting and publishing scientific papers in the period 1996-2010. Networks cover co-authorships and publication rates of users of the KMS system and therefore of the nodes in network E. Network S also covers the publications rates and co-authorships of those employees (not necessarily KMS users) with the highest number of publications. Based on frequency of publication, we include 101 subjects having at least one publication in the last three years and at least five in the time window 1996-2010. In both the networks in Figures 4.2a and 4.2b, node dimensions are proportional to degree centrality and are shown in black, dark gray and light gray if the nodes respectively belong to only one network, two of the three networks, or all three.

In order to check for the questions emerging in Issue 1, focusing on the



(a) *Network of patents.*



(b) *Network of publications.*

Figure 4.2: Networks of experts. Nodes are colored in black, dark grey, and light grey if the corresponding expert respectively belongs to only one network, to two of the three analyzed networks, and to all the three networks of experts and e-mail exchange.

productivity of CoP members is helpful in analysing both the representativeness of experts in CoPs and to check for experts' centrality in the networks. Despite the fluctuating trend in the number of co-authorships, the decreasing trend in number of patents produced by all the employees since 2003 reveals both a decrease in employees' production and the clear application of an internal policy aimed at reducing time spent on patenting and boosting either publishing or other practical and short-term lucrative activities (Figure 4.3). This is confirmed by the continuous activity of CoP members in publishing scientific papers. Figure 4.4 shows an increasing trade-off between the number of patents and last publication produced by CoP members: despite the remarkable number of CoP members who wrote their last publication before 1996, the majority of recent publications were produced in the last three years but the activity of patenting is decreasing. The activity of the few experts in CoPs therefore mainly deals with scientific publishing and not patenting.

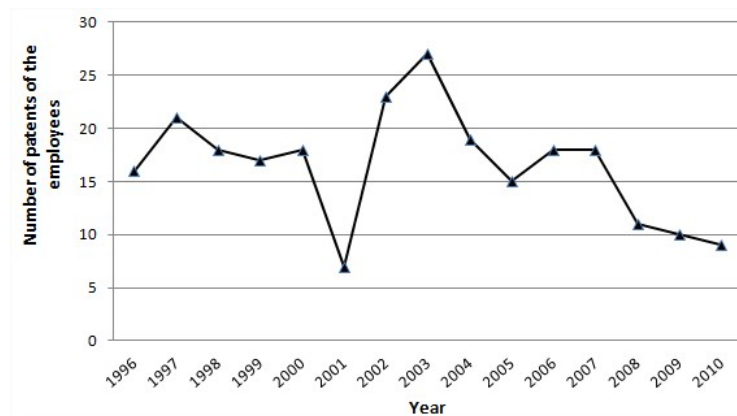


Figure 4.3: Trend of the number of patents

Despite the few members of CoPs involved in publishing, most of those members produced at least one publication in the last three years. Out of 667 CoP members, only 28% (4%) have had at least one publication (patent) since 1996. The distribution of members' publications and patents is listed in Table 4.1, showing the highest percentage of CoP members who had published and

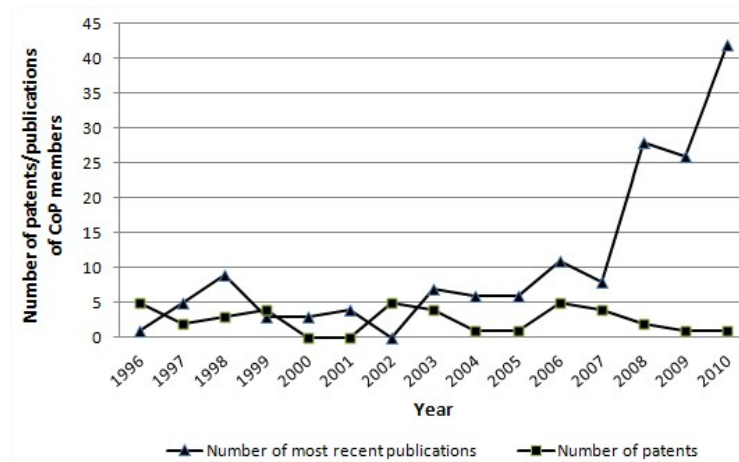


Figure 4.4: Trend of the number of patents and recent publications by CoP members

invented from 1 to 5 documents since 1996. Table 4.1 proves that, when they publish, most publishers and inventors focused on not more than 5 documents.

		N. of CoP members	
		(1)	(2)
Number of publications (1) and patents (2)	from 1 to 5	125	23
	from 6 to 10	30	2
	from 11 to 15	12	0
	from 16 to 20	8	0
	from 21 to 25	4	0
	from 26 to 30	2	0
	after 31	3	0
Total n. of subjects		184	25
% of the 667 CoP members		27.58	3.75

Table 4.1: Productivity of CoPs’ members since 1996.

The dispersion of the networks is also verified to check for subjects’ roles. Low density (1.31%) and low centrality, both when either degree centrality (2.72%) or betweenness centrality (5.72%) were examined, revealed dispersion of the network of patents: the various areas of patenting were not homogeneously connected through co-patenting. When publications are considered, the network dispersion is even more pronounced, and the centrality measure does not even reach 1%.

The results answer Issue 1 and verify the low representiveness of experts



in CoP activity, in both domains of patenting and scientific publishing. There are no central experts who can reduce the dispersion of the two networks in these domains. In addition, neither top scientists nor inventors play an official role in CoP activity: only two inventors out of 305 were officially assigned as an alternate and a skilled technician and, out of all the scientific publishers, only 33 were appointed as skilled technicians. However, of the 15 facilitators and 15 alternates involved in CoP activity, half were also publishers.

### 4.5.2 Proactivity in the domain of information exchange

Despite the crucial role played by private communication tools in small environments such as laboratories, small offices, and family businesses, professional e-mail exchange is a key success factor for multinational companies needing to connect and share information among worldwide subsidiaries, laboratories, and platforms. Therefore, although employees can communicate by phone, intranet chatting, or in person, the communication network is represented by focusing on e-mail exchanges.

A first representation of network E in the analysed division of the company is shown in Figure 4.5, computed by identifying as nodes those subjects using the KMS, that is, subjects who either send or receive e-mail.

The size of the nodes represents their proactivity in knowledge sharing, measured through an outdegree index based on number of messages sent. The nodes are colored according to the CoP to which they belong, and the operating line has blue nodes. Network E was drawn by the spring embeddedness technique, and a link between two subjects is identified whenever they exchange at least one e-mail in the time window (09/01/2009-04/05/2010).

To better understand the dynamics within the network, individual positioning is analysed. A description of the network was examined by computing

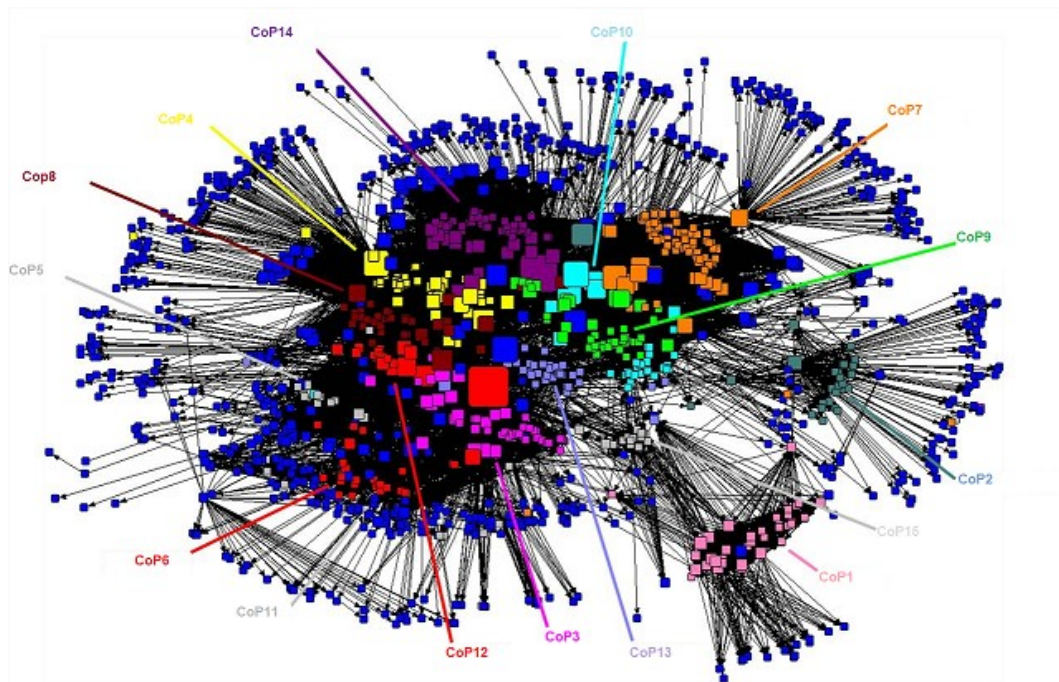


Figure 4.5: The e-mail exchange network. Colors based on the differentiation among CoPs.

the  $K$ -core index. The resulting network, after removal of pendent subjects, highlights the importance of a few key subjects in knowledge spreading by ensuring the persistence of the information flow. These preliminary qualitative results answer the last question in Issue 1 and support the hypotheses proposed by Wenger et al. (2002), according to which only a few subjects support CoP activity. This was confirmed by the absence of a few key players, which creates structural holes in the flow of information. In addition, the crucial role played by a few KMS users in facilitating the information flow among CoPs and more peripheral groups of CoP members is also emphasized when cut-offs on the number of messages sent per subject are introduced. In fact, a few dense star networks are clearly identifiable whenever only those subjects who send more than 12 messages were included in the network definition (Figure 4.6). The emerging stars are centered on either facilitators or alternates and, in particular, on those facilitators and alternates, who were also identified as

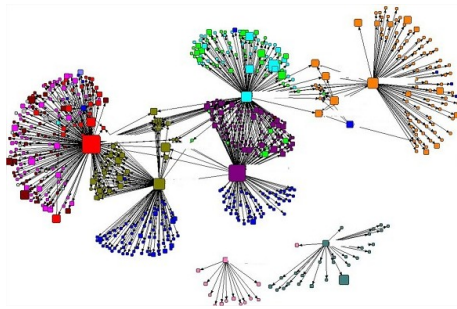


Figure 4.6: Network of e-mail exchange by applying a 12 e-mail cutoff

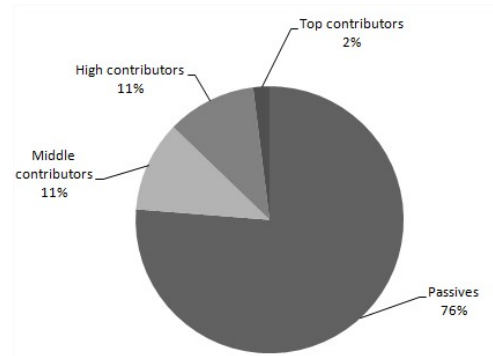


Figure 4.7: Contribution to knowledge-sharing

experts in the publication network.

At a broader level, these results show that the communication flow is not adequately supported by the company's employees. Interestingly enough, more than 76% of the company's members do not communicate with the colleagues and does not contribute to the information flow (Figure 4.7). The proactivity level of the network members is therefore insufficient to support efficient knowledge spillover. In addition, 90% of subjects receive communications from only one subject, that is, more or less the whole division is composed of pendent subjects. The high pendency of members is emphasized by the distribution of betweenness values: the great majority of subjects do not contribute toward linking other subjects and do not have any control on the relations and interactions among them. By identifying subjects contributing to knowledge-sharing through a combination of proactivity (outdegree) and brokerage (betweenness) measures, the data show that only 24% of KMS users represented in network E concur in sharing knowledge, and 64% of the contributors do not belong to any CoP. In addition, 63% of the skilled technicians are passive and 71% of them are passive, although they receive e-mails from other KMS users. This lack of proactivity is also due to the fact that 12 of the 49 skilled technicians are not in any CoP. In order to improve collaboration among CoPs in

knowledge-sharing, the commitment of CoP facilitators must be emphasized, and more clarity in role definition of both facilitators and alternates is recommended. The low proactivity of facilitators and the passivity of 44% of alternates do not contribute to introducing CoPs as an effective KM tool to improve knowledge-sharing.

In summary, by computing detailed analyses of individual roles in knowledge sharing through study of individual proactivity (outdegree, measured by the number of e-mails sent during the time window), peer-recognition (indegree, recorded by the number of received e-mails), and brokerage behavior and power (betweenness) we confirm the structure proposed by Wenger et al. (2002) and the influential hypothesis by Valente (2005) in CoP contexts: there are many subjects who rarely contribute to knowledge-sharing and a limited number of influential subjects playing a proactive role in encouraging discussion.

### 4.5.3 Presence and individual overlap

In this section the match among the three networks are examined to check for overlap between the domains of experts and of knowledge exchange. The answer to Issue 2 may be sought in recent multiple network theories measuring the overlap and simultaneous presence and role of subjects in the various networks. Individual roles are highly determined by structural holes, which create network disjunctures, and by the possibility of being simultaneously involved in various networks (Burt 1992, Ahuja 2000, Burt 2004). As proposed by Padgett and Ansell (1993), *multivocality identity* and the possibility of acting in various cross-cutting networks increase individual power. By in-depth analysis of the rise of Cosimo de' Medici (1389-1464) in Renaissance Florence in the 15th century, Padgett and Ansell (1993) proved the power of network

rules in explaining the beginning of Cosimo’s political power and generalized theories about structural holes in multiple networks. Cosimo’s style of control, derived from multivocality, is defined as “robust action”. His power was indeed determined by differentiated network strategies and by the multiple networks in which he was embedded at various levels of involvement: intermarriages, political and economic affiliation, and personal friendships.

In our framework, social network instruments are used to study the overlap of the three networks and the emergence of Cosimo-like subjects, able to connect basic and applied research with practice by simultaneously acting in network P (patents), S (scientific network of publications), and E (e-mail exchange network). Complete overlap between the information exchange through CoPs and co-authorship networks is found whenever all the experts, publishers of publications or patents, are CoP members and whenever an improvement in individual scientific publications and patenting corresponds to a higher degree of centrality in the information exchange network. Between the two extremes of total and no overlap, partial overlap exists when there are some subjects whose roles enable them to exploit the advantages of being embedded in multiple networks.

Subjects belonging to	Combination	Total overlap (%)	
3 networks	P S E	0.80	0.80
2 networks	- S E	8.54	11.21
	P S -	2.60	
	P - E	0.07	
1 network	P - -	2.82	87.99
	- S -	63.12	
	- - E	22.05	

Table 4.2: The overlap of the three networks: network of patent co authorships (P), network of scientific publishing (S), and network of e-mail exchange (E).

Table 4.2 lists the percentages of overlap for each combination of individual simultaneous presence in the networks, and the cumulative percentage of

subjects belonging to one, two or three networks. Most subjects are publishers and co-publishers who are not active in the KMS and not involved in e-mail exchange (- S -, 63%). In addition, almost 25% of the employees are KMS users who had neither published nor patented (- - E, 22%). Despite a high number of subjects belonging to either one network (87.99%) or two (11,21%), only 40 subjects in the three datasets were simultaneously active in all three networks.

Lastly, it was also interesting to verify the presence of Cosimo-like subjects and check the consequences of structural holes in information creation and sharing. Early results do not prove the presence of Cosimo-like subjects able to exploit central roles in all the three networks simultaneously.

This analysis answers Issue 2 and proves that less than 1% of the total number of subjects simultaneously belong to the domains of experts and of knowledge exchange. In addition, less than 9% of total subjects are either publishers or inventors using the e-mail exchange system. The results show no multivocality identity in individual roles, and no subject acts in the various domains to improve communication exchange by recombining theoretical and practical knowledge. Very low overlap between the domains of experts and of knowledge exchange was indeed verified.

#### **4.5.4 Activation of experts**

Examining the type of messages sent is helpful in differentiating among types of shared information and identifying the extent to which experts contribute to technical knowledge-sharing (Issue 3). Experts' contributions are especially crucial for technical aspects and problems directly related to production processes, not necessarily for other daily routines or organizational communications. Messages were therefore classified according to subject mat-

ter, and divided as follows:

1. *organizational issues*: either multi-messages aimed to inform employees about events (conferences, meetings, etc.) or organizational messages sharing formal documents (reports on meetings, journal subscriptions, etc.);
2. *technical issues*: messages dealing with specific technical matters;
3. *others*: messages automatically sent by the KM area or dealing with the CoP system (holiday greetings, welcome to new CoPs members, changes in CoPs structure, vacation schedules, automatic invitations to CoPs members to participate in CoPs, etc.);
4. *residuals*: messages without subject matter, error messages, “out of office” messages.

Despite the low percentages of organizational (16%) and other (17%) messages, most e-mails (61%) deal with technical messages<sup>13</sup>. In addition, a regular trend is recorded for organizational subjects, and the fluctuating trend of total e-mails is generated by the high number of technical messages in the first and last months of the year.

Technical e-mails were examined in more detail to check the proactivity and participation of experts on specific and technical issues. By focusing on the subject matter of each message, forwards and answers were identified by the prefixes Fw:, R:, Re:, etc.<sup>14</sup>. The survival length of each message shows that the first answer usually appears within 2 hours of the original message and, in case of more than one level of answer or forward, the mean time lag of each level of answer is no more than 12 hours. The time of persistence of

---

<sup>13</sup>An additional 6% were residual messages.

<sup>14</sup>Messages having the same subject, without any prefix but recorded at various times and sent by different people, are included in the analyses as both answers and forwards.

each message is on average 3 days. Messages may also be on stand-by for a certain period of time and then reconsidered as crucial issues: for these outsider cases, answers may be recorded even more than one month after the original message.<sup>15</sup>

Subjects' position	Forwards (%)	Answers (%)
P S E	1.36	2.13
- S E	55.98	50.10
P - E	0.00	0.00
- - E	42.66	47.78

Table 4.3: Percentages of forwards and answers divided by networks' overlap

Subjects' role	Forwards (%)	Answers (%)	Single senders(%)
Facilitator	32.61	22.24	4.39
Alternate	8.70	9.09	3.41
Second alternate	0.27	0.19	0.48
Skilled technician	3.80	3.09	4.87
No specific role	54.62	65.38	86.82

Table 4.4: Percentages of forwards and answers divided by subjects' role

How do core-people behave in e-mail exchanges? Do the authors of publications and patents answer the technical e-mails they receive? Tables 4.3 and 4.4 show the number of messages based on the characteristics of subjects sending either answers or forwards. First, despite the low number of subjects belonging to all three networks, the results show that more than half both answers and forwards are sent by subjects involved in scientific publishing (-SE). Therefore, technical problem-solving is partially enriched by the participation of experts publishing in those technical areas, whereas the authors of patents do not seem to be directly enrolled in knowledge-sharing. Also in this case, the company's interests must be considered, because incentive policies to boost either patenting or scientific publishing, or both, may be implemented. Second,

---

<sup>15</sup>This descriptive survival analysis has been computed by examining only those cases in which more than one level of answer or forward was reported.



Table 4.4 focuses on subjects' roles and lists the percentage of answers and forwards sent by CoPs facilitators, alternates and by skilled technicians. The latter column reports the numbers of single senders for both answers and forwards. Although only 4% of senders are facilitators, 32% and 22% of technical messages are respectively forwarded and re-discussed by CoP facilitators. This result shows that very few subjects play official roles in CoP, sending many messages to various colleagues. In addition, knowledge owners defined by the company do not seem to be technical communicators for problem-solving, and more than half the technical issues are discussed by people, who are not in the co-authorship network and play no specific role in the KMS. These results confirm the need to revise the official role and position of currently defined skilled technicians and to involve other people having publications on specific issues who are more likely to enrich discussions.

In summary, experts are only partially proactive in sharing their technical knowledge. In particular, only scientific publishers contribute to the spread of technical knowledge through CoP activity. Instead, inventors do not seem to contribute to problem-solving through a frequent involvement in knowledge spillover and are not likely to use CoPs as a KM tool to improve and speed information-sharing. As a conclusion, the number of discussions on technical issues, their survival time and experts' involvement must still be improved.

#### **4.5.5 Retrieval**

In order to measure the overlap among domains quantitatively and to check for difficulties in reaching the required knowledge (Issue 4), an indicator was constructed to measure the cost of accessing that knowledge. The concept of *geodesic distance* has been widely used in the literature on logistics and transports as a proxy for measuring costs in logistic networks (Combes and

Lafourcade 2003, Combes and Lafourcade 2005, Cavalletti 2011). By analogy, this analysis focuses on the measure of *farness* as a proxy to measure the costs in which a company incurs in guaranteeing quick and efficient knowledge access and recombination of knowledge whenever networks do not overlap sufficiently.

Comparisons among the normalized indexes of the three networks aimed at checking whether subjects belonging to one network are closer to or farther from their colleagues in another network, and therefore at determining the efforts and costs needed to reach the required information. Table 4.5 shows the formula of one of the five indexes and their basic explanation.

Index	Explanation
$C_{EP} = \frac{Farness_i^E \cdot Patents_i}{Tot.patents}$	It measures how close/far the authors of patents are from the other subjects within the network of e-mail exchange.

Table 4.5: Formula and description of one of the cost index to access knowledge.

$C_{EP}$  measures the cost to access patenting knowledge by the mean of the extent to which those subjects in the network P are far from the other neighbors in network E.  $Farness_i^E$  reports the normalized value of farness for subject  $i$  in network E and then it is weighted by number of patents to compute  $C_{EP}$ <sup>16</sup>.  $C_{EP}$  varies between 0 and 1: the closer the index is to 0, the closer the authors of patents are to all the subjects in network E, and the lower the costs to access knowledge. All the other indexes combining networks ( $C_{ES}$ ,  $C_{EPS}$ ,  $C_{PS}$ ,  $C_{SP}$ ) are constructed following the same basic formula and have the same boundaries.

The first row of Table 4.6 lists the values of indexes for total e-mail exchanges, and the second row the exchange of technical messages only. Viewing

---

<sup>16</sup>The same analyses were computed by changing the index and using farness values as weights: for example, the normalized farness of  $i$  in network P was used instead of the number of patents  $C_{EP} = \frac{Farness_i^E \cdot Farness_i^P}{\sum_i Farness_i^P}$ . For all indexes, all reported results were confirmed.

	$C_{ES}$	$C_{EP}$	$C_{PS}$	$C_{EPS}$	$C_{SP}$
Total e-mail exchange	0.380	0.144	0.064	0.042	0.023
Technical e-mail exchange	0.286	0.040	0.064	0.031	0.023

Table 4.6: Cost indexes in the e-mail network.

the whole e-mail dataset, the results emphasize the fact that experts are not close enough to those colleagues who can benefit from those messages. Therefore, the cost of employees to acquire expert knowledge is high. Indeed, the higher values of indexes  $C_{EP}$  and  $C_{ES}$  prove that those who generate knowledge through patenting or publishing respectively are far from all subjects within the e-mail network. The low level of closeness among these subjects confirms the lack of overlap between network E and the other two networks, causing higher costs. Instead, the considerable reduction in the values of  $C_{PS}$  and  $C_{SP}$  shows increasing overlap between networks P and S: subjects in network S are quite close to subjects in network P ( $C_{PS}$ ) and the very low value of  $C_{SP}$  proves the closeness of the authors of patents from other colleagues who publish. Experts incur lower costs whenever they need to reach other experts' knowledge, because experts, either publishing or patenting, are much closer to each other than to those who have a better chance of benefiting, exploiting, and sharing knowledge.

These results are also confirmed when analysis focuses on technical e-mails only. In this case, the general small decrease in the indexes indicates a small increase in the level of closeness among the networks for technical issues. In addition, those who patent are much closer to subjects belonging to technical e-mail exchange (lower value of  $C_{EP}$ ). This result demonstrates the greater overlap between the patent co authorship network and the network of technical e-mail exchange.

In summary, access to required technical knowledge through e-mail exchange is difficult, due to the farness of subjects belonging to the two domains.

Although publishers have been proved to be more proactive by answering and forwarding technical messages, their contributions are hard to reach by their colleagues through the e-mail exchange system.

#### **4.5.6 Structure: small worlds and structural holes**

The role played by CoPs in multinational companies is even more crucial in dynamic contexts, in which innovation and competitive advantage are closely dependent on tacit knowledge acquired by subjects and shared through communications, feedback, and networking within communities. However, many difficulties arise in directly connecting people in multinational companies, and critical situations in moving from self-regulated team working to structured ad hoc communities may prevent improvements in knowledge recombination and the extent of innovation (Nirenberg 1994). This section checks two points: do small-world structures limit the emergence of structural holes both within and among CoPs? Do CoPs increase the overlap between research and practice domains in the innovation process? (Issue 5).

Starting from a micro-perspective dealing with intra-community structure, results demonstrate that collaboration among subjects belonging to different communities can certainly improve: the average density in the whole dataset (3.8%) is in fact more than half the average density per CoP (7.8%). Inefficiencies are confirmed by checking for the small-world hypothesis (Watts and Strogatz 2006, Watts and Strogatz 1998b) and by studying the diffusion speed of information. Results on geodesic distances (the number of messages needed to link two subjects) and on maximum flow (the number of alternative ways to link two subjects) confirm the need for better connections among CoPs, because of the low probability of reaching colleagues and exchanging information among subjects belonging to differing CoPs. In addition, analysis of the small-

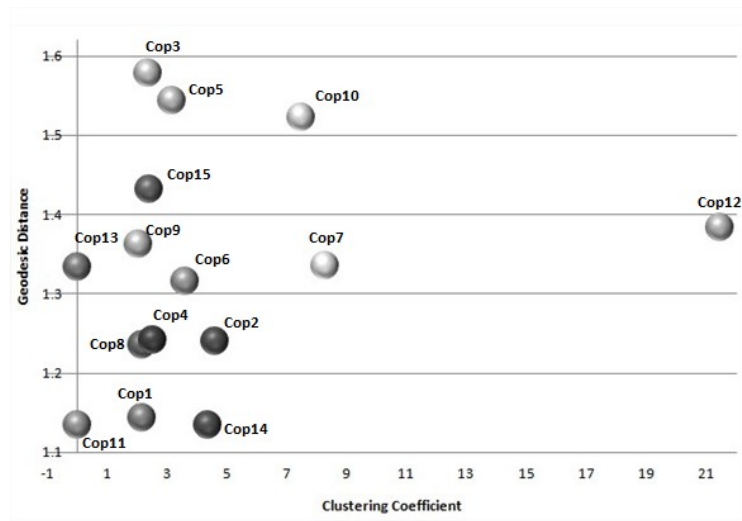


Figure 4.8: Positioning of CoPs based on the small-world hypothesis.

world phenomenon is checked by simultaneously combining the values of the clustering coefficient (CC) and the geodesic distance within each community: a small-world model is verified whenever a network reports a high CC and a low mean geodesic distance. Figure 4.8 (bottom right) shows only one community behaving as a small world. All the other communities, in both research and practical fields, are divided into two groups: most practical communities are positioned bottom-left in the graph, meaning that their members are close to each other because of their high exchange of information within communities, but they have tend to cluster together less to form a small world. In the remaining communities (top-left of graph), members are not connected either by short path length and do not tend to cluster together. As a result, CoPs cannot exploit efficiency in the diffusion of information and the improvement in learning guaranteed by small-world structures.

Moving to a macro-perspective, we focus on relations among CoPs. Figure 4.9 applies the Chain-Linked Model to the structure of CoPs in this specific framework. This analysis explores relations among CoPs to check for feedbacks flows across the various stages of the innovation process and for links between

research stages and more practical ones. In particular, CoPs are marked by

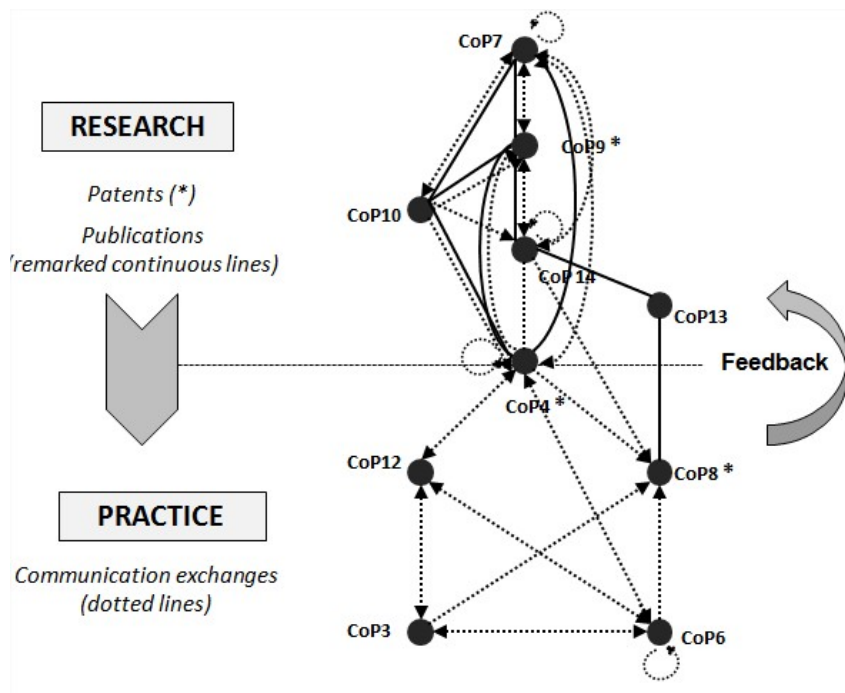


Figure 4.9: Interactions among CoPs in the Chain Link Model. CoPs are marked by stars (\*) whenever the members have registered at least one patent since 1996. Dotted and continuous lines represent ties based respectively on e-mail exchange and publication co-authorships between the members of the linked CoPs.

stars (\*) whenever their members have registered at least one patent since 1996. Dotted and continuous lines represent ties among CoPs based respectively on e-mail exchange and publication co-authorships between the members of the linked communities. Lastly, reflexive links are shown whenever CoPs have high numbers of information exchanges within their communities. To favor easy interpretation of the graph, only ties having strengths above the mean total strength are shown.

The two stages of research and practice in the analysed company are mainly linked through CoP 4, which is also the most central community with respect both to e-mail exchange and co-publishing. Despite the bridging role of CoP 4, research and production areas are completely separate, and publication

co-authorships between CoPs are grouped within those communities which mainly deal with more scientific knowledge, such as geology, geophysics and explorative techniques. Generally, no scientific knowledge results from the collaboration of communities belonging to different stages of the innovation process. Communities dealing with more practical fields, i.e., material management, production, and construction techniques, do not communicate with the research area, except for CoP 4, which deals with well drilling. This is confirmed by an overview of the answers and forwards to check CoP participation in technical discussions. Despite the 10% and 19% of forwards and answers respectively, sent by non-CoP members, the three CoPs giving the highest contribution to problem-solving through their members are those operating in the practice stage of the innovation process (CoP 6, CoP 12 and CoP 3) and CoP 4. Their members contribute to both forwards and answers by 20%, 13%, 8%, and 9 % respectively. Because of the more practical and technical fields in which these CoPs operate, the communities involved in the final stage of the innovation process have a higher need or interest in being involved in technical issues, and contribute to problem-solving by more frequent involvement in knowledge spillover.

A detailed analysis of the directions in e-mail exchange also allows us to check for the role of CoP4 in both bottom-up and top-down information flows. Results show the crucial role of CoP 4 as both sender and receiver of messages within the two areas of research and practice, and it is often embedded in sending the same messages to both areas. However, no direct flow, either top-down or bottom-up, is recorded through CoP 4.

Because of the demonstrated importance of the latter community in recombining knowledge from the two stages of the innovation process, the presence of Cosimo-like subjects in CoP 4 was checked to look for transversal subjects matching theoretical and practical languages and procedures, to facilitate the

coding and transfer of knowledge across the domains of research and practice. This check simultaneously focuses not only on individual centrality in the three networks (see Section 4.5.3) rather, individual centrality is considered together with the centrality of the CoP to which the subjects belong. For example, subject  $i$  is central only in the expert domain and not in that of e-mail exchange because s/he is embedded in generating knowledge and not in sharing it within the company. This is not a Cosimo-like subject *per se*. However, s/he is considered a Cosimo whenever her/his CoP is central in the e-mail exchange domain and co-members guarantee knowledge flow. The results reveal two Cosimo-like subjects within CoP 4: one was an expert in cement systems and a technical leader in cementing, the other won a company award for introducing a new technology allowing pressure to be controlled in drilling systems. Despite these two subjects are not Cosimos *per se* because they are central in only two of the three networks, the involvement and the centrality of their CoPs in both patenting activities and information flows compensate this lack and guarantee the link among the various stages of the innovation process.

The previous analyses answer to the first point and prove the absence of small world structures and the presence of structural holes among CoPs. Moving now to the second point, farness indexes were also computed to check whether CoPs increase the overlap between research and practice domains. The analysis focuses on each single CoP and considers total e-mail exchanges among each CoPs member as a network *per se* in order to check for increased overlap among the domains within each community. Despite the few cases in which we can compute  $C_{SP}$  and  $C_{PS}$  because of the lack of CoP members in co-authorship networks, some CoPs were found, especially ones dealing with more practical issues (CoPs 3, 4, 6, and 12), having lower values of farness indexes and therefore greater closeness between those members who are experts on certain topics and others able to reach and actively participate in knowledge-



sharing. In addition, the overlap among the three network for all the CoPs has been computed and reported in Table 4.7.

Network overlap (%)				
	P S E	P - E	- S E	- - E
CoP 1	6.67	0	15.56	77.78
CoP 2	0	0	24.24	75.76
CoP 3	1.49	1.49	14.93	82.09
CoP 4	6.25	0	37.5	56.25
CoP 5	3.45	0	10.34	86.21
CoP 6	0	0	25.00	75.00
CoP 7	0	0	29.79	70.21
CoP 8	2.08	0	16.67	81.25
CoP 9	3.70	0	64.81	31.48
CoP 10	0	0	78.05	21.95
CoP 11	0	0	0	100.00
CoP 12	0	0	45.45	54.55
CoP 13	3.23	0	22.58	74.19
CoP 14	0	0	54.05	45.95
CoP 15	0	0	16.67	83.33

Table 4.7: Network overlaps within CoPs

The data show an improvement in overlap whenever the KMS in CoPs is considered. Half the CoPs do have at least one member belonging to all three networks, and most of them benefit from publications produced and shared by their members (- S E): experienced members are therefore better positioned to relate with proactive KMS users. Lastly, improvements and interventions are needed to reduce the percentages of CoPs members using the KMS but not generating knowledge<sup>17</sup>.

In summary, despite these multivocal subjects in the bridging CoP4, whenever their community is deleted from the graph, a structural hole appears, clearly dividing research and practice. In addition, feedback between the stages of the process are guaranteed only by weak ties entering CoP4: the generation of innovative solutions to emerging problems is prevented by the absence of direct connections between researchers or inventors and the professionals who

<sup>17</sup>No information is available on subjects' positioning in the organizational chart to check for differentiations based on subjects's organizational role. This percentage is reduced whenever only technical e-mails are examined.

have to implement their research and inventions practically. The bottom-up perspective is only partially verified, and results confirm the previous hypothesis stating the limitations and difficulties faced by a multinational company in combining basic and applied research with more practical stages of the innovation process. Lastly, despite the higher overlap among the domains of research and practice when CoPs are considered as networks *per se*, no small-world structure can be exploited to improve robustness to perturbations and to speed learning (Simard et al. 2005).

## 4.6 Discussion and conclusions

Multivocality is a critical success factor in knowledge mobilization. However, difficulties arise in matching the required variety of individual roles with the need to group employees into single groups of affiliation, i.e., communities of practice, based on knowledge specialization. The success of CoPs as transversal tools with respect to exploration and exploitation derives from CoPs contributions to the recombination of dispersed knowledge. The case reported in this study emphasizes the risk which multinational companies face in coping with critical contingencies, which require urgent recombination of crucial knowledge and a short-term decision-making. This study examined the efficacy of on-line CoPs introduced by a multinational company operating in the energy sector to facilitate access to critical knowledge and accelerate problem-solving procedures.

Our focus is on better understanding of communication exchanges in multiple networks, rather than within single networks independently. This work contributes to existing literature by adding multiple levels in studying the overlap among the domains of information and knowledge exchange, and the problems related to this topic. In particular, two more levels are added to

the state-of-the-art in the literature on CoPs. By matching communication networks with experiential networks, we improve the knowledge of learning in practice (Fox 2000) and the current research in this field, mainly based on questionnaires and interviews.

Results give a myopic view of management in focusing only on e-mail exchanges as KM tools and reveal the need to go deeper into company strategy to improve incentive systems and the definition of facilitators' role within CoPs activities. Many appointed skilled technicians tend to be excluded from both CoPs and communication exchange, and they do not often have experience in publishing or patenting. Despite the difficulties in including professionals in on-line social networks, the system lacks subjects favoring communication exchanges between practice and research. CoPs turn out to be not adequately connected, and the resulting structural holes prevent the development of feedback along the chain of the innovation process. In fact, only a few people, mainly team leaders or technical award winners, are involved in the exchange of technical communications, both within and among communities, along the whole innovation process. Transversal communications need to be improved by the presence of subjects trying to match different languages and procedures, and codifying and translating knowledge among the various domains. The partial failure of this tool in the company in question is also influenced by the environmental context in which it works and by the relatively short period of time since its introduction. In addition, the strong institutionalization of competences characterizing the company inhibits resource recombination (Galunic and Rodan 1998). A cognitive dimension could also be added to this framework. As proposed by Ardichvili et al. (2003), further analyses can be implemented to verify whether the attribution of recognized status to the subjects involved enhances their contribution to CoPs.

The inefficiency of CoPs prevents the advantages of small and dense com-

munities from being exploited to improve members' commitment (See Chapter 2 and 3). The expectation that the spread of knowledge is more likely to come from CoP members is not confirmed, because of members' lack of participation, identification with the CoP, and recognition of CoPs as tools for information diffusion. According to these results, the effectiveness of CoPs could be greatly improved by modifying their structure and adding new members who could contribute to spillover but who are not included in the CoPs. The CoPs have received 143 new members according to analyses of proactivity, brokerage behavior and peer recognition. Future work is needed to check effective improvements in knowledge-sharing and exchange density after these changes in CoPs structure.

Further research on private communications is required to complete the description of the information and KM systems, and other communication tools, apart from e-mails (intranet chatting, private e-mail accounts, social network accounts, etc.) should be included to enhance the robustness of the results. In addition, other tools and methods can be used to improve the understanding of knowledge diffusion, such as network simulation and cluster analysis. In particular, network simulations can be useful to verify the impact of various CoPs' structures on knowledge spreading in order to find the most efficient configuration to enhance the use of CoPs and members' proactivity.

# Chapter 5

## Conclusions

Individual decisions are influenced by other people's choices. In everyday life, we have socially interactive experiences and continually acquire knowledge in our roles as consumers, workers, friends, or in general as members of groups. In addition, social influence is a key in human decision-making, explaining how beliefs and knowledge spread and sustain the evolution of organizational forms. In this thesis, we exploit various methods and data to study how several types of dispersed knowledge spread in networks of customers and co-workers, and how this knowledge affects collective results in communities of customers and communities of practice.

In Chapter 2, new technologies and methodologies originally developed by ethologists (Hinde 1976, Whitehead 2008) are applied to examine reality-mined network data in marketing and the influence of consumer co-experience on customer retention for health clubs. With instruments used in social biology, a model is developed to analyse the effects of co-experience for customer defection and choice of subscription plans. Considering various types and strengths of relationships, the analysis focuses on how an individual's position in a consumer network influences attendance and renewal decisions. We find that relationships in the club are crucial in membership renewal decisions. In

particular, the duration of the members' contract depends on the level of commitment of the other members with whom they usually go to the club. More specifically, we predict and empirically prove lower retention with longer experience, higher retention for more central subjects, and higher probability of decision alignment for subjects having stronger ties. Our work sheds new light on the influence of social networks on commitment decisions and contributes to understanding the effect of social dynamics and localized conformity on customer decision-making.

Social effects enter a field of research about which the amount of recent work has recently increased. Although these effects are interesting and may be important for marketing researchers and managers, empirically measuring causal peer effects is challenging. Other factors differing from peer influence and social dynamics are crucial elements in determining diffusion dynamics. More than half behavioral contagion is indeed due to homophily and other confounding factors, rather than to peer influence (Aral et al. 2009). This area of literature is still limited, and confusion about the evaluation of consumer network dynamics emerges in the marketing field when researchers try to isolate contagion and influence. This work recognizes these challenges and proposes a method to address at least homophily and environmental effects. Although in our case Chapter 2 provides evidence that family ties and other socio-demographic characteristics - such as gender, age and provenance - are not causes of homophily, improvements in this research are needed to distinguish clearly between neighbors influence and homophily. By applying a set of non-parametric methods to a longitudinal and dynamic dataset, Chapter 3 measures peer effects in the context of contractual renewals for a subscription service. This chapter addresses some of the questions which have arisen in recent debates in the literature on network diffusion about discrimination among peer influence, homophily, and other confounding effects (Aral 2011, Iyengar

and Van den Bulte 2011). In addition, we contribute to recent debates regarding the lack of accuracy in econometric procedures by applying to our context rigorous matching models to study influence dynamics in social contexts and to disentangle the various types of homophily and environmental effects. In particular, two matching estimators are used to control for homophily and other effects. We find evidence that peer effects play an important role in both contract renewal and attendance decisions by consumers and demonstrate that members in the consumer network tend to align their commitment to the socially related group of co-enterers, independently of their similarities and the common exposure to external stimuli. The econometric models proposed in Chapter 3 may be applied to all communities in which discrimination between homophily and influence need to be addressed.

Lastly, Chapter 4 moves from communities of consumers to communities of practice and checks for efficiency in knowledge spreading within companies by implementing on-line communities of practice as tools to facilitate access to critical knowledge and to speed problem-solving. By matching multiple networks of co-authorships, e-mail exchange and communities of practice, we find structural holes among the phases of research and practice in the innovation process, and several problems emerge in the efficient introduction of CoPs. By matching communication networks with experiential networks, we improve current research in this field and contribute to better understanding of communication exchanges in multiple networks. We find that communications and knowledge recombination need to be improved by introducing transversal subjects, trying to match different languages and procedures along the innovation process, and coding and translating knowledge among the various domains.

Further research is needed for a to better contribution to the open debates focusing on methodological issues: no definite solutions have been found to discriminate homophily clearly from other confounding factors, and multiple

network analysis must be improved by adding more levels to the analysis, especially when dealing with private communications. In addition, further work is needed to deal with the reflection problem (Manski 1993), endogenous tie formation (Nair et al. 2010), and unobserved heterogeneity in measuring peer effects. The definition of asymmetric ties or temporal ordering of decisions are two feasible solutions to overcome some of these problems. Lack of attention to these aspects may create a bias in the evaluation, measurement and effect of social influence. Lastly, other methods can be applied to increase the robustness of the results (i.e. hazard models).



# Appendix A

## Propensity Score and Coarsened Exact Matching

Propensity Score Matching is used to minimize the bias in measuring treatment effects (Rosenbaum and Rubin 1983, Becker and Ichino 2002). By controlling for confounding elements, this approach reduces bias by comparing treated and untreated subjects who are similar. Propensity score ( $\rho$ ) is defined for each subject as a summary of  $k$  pre-treatment characteristics  $C = c_1, c_2, \dots, c_k$ , on which the degree of similarity is computed and is the conditional probability of being treated, given pre-treatment characteristics (Rosenbaum and Rubin 1983).

This means that:

$$\rho = Pr(T = 1|C) = E(T|C) \tag{A.1}$$

where  $T = \{0,1\}$  is the dummy reporting exposure to treatment ( $T = 1$  if the subject belongs to the treated group, 0 otherwise) and  $C$  the set of pre-treatment characteristics. Subjects with the same propensity scores have the same distribution of pre-treatment characteristics, independently of treatment.

Treated and untreated subjects with the same propensity score are on average alike. Once  $\rho$  is computed, the Average effect of Treatment on the Treated (ATT) is estimated as the difference between available outcomes in the cases of treatment and non-treatment when subjects are treated, i.e.:

$$ATT = E(R_{1i} - R_{0i} | T_i = 1) \quad (\text{A.2})$$

where  $R_1$  and  $R_0$  are the two outcomes (renewal/not renewal) in case of treatment and non treatment, respectively. For each subject we have:

$$R_i = T_i \cdot R_{1i} + (1 - T_i) \cdot R_{0i} \quad (\text{A.3})$$

where  $R_{0i}$  ( $R_{1i}$ ) is unobserved if  $i$  is treated (or untreated). Subjects are then subdivided into groups based on the various identified intervals of propensity scores, which are invariant between treated and untreated subjects within each interval. The balance property is verified if, within each interval, the mean of the pre-treatment characteristics does not vary between treated and untreated subjects.

Because of the low number of cases in which observations have exactly the same propensity score, various approximate matching methods have been developed to determine the intervals within which propensity scores are considered identical, i.e. Stratification Matching, Nearest-Neighbor Matching, Radius Matching and Kernel Matching (Becker and Ichino 2002). Our analyses are based on the Stratification matching to estimate ATT based on propensity scores. This method consists of dividing the range of variation of the propensity score into intervals, so that, within each interval, treated and untreated subjects have on average the same propensity score. Also pre-treatment characteristics in each block are balanced and assignment of treatment is considered random. Thus, the difference between the average outcomes of treated and un-

treated subjects is computed and the ATT is the average of the ATT of each block, weighted by the number of treated subjects in each block.

In order to improve the efficiency of Propensity Score Matching, Coarsened Exact Matching (CEM) was then applied (Blackwell et al. 2009, Iacus et al. 2009). CEM was used not only to increase the number of matches in comparison to exact matching but also to reduce the imbalance in pre-treatment characteristics between treated and untreated groups. According to this method, the user chooses the maximum imbalance (coarsening) ex ante rather than ex post - for example, as a result of a propensity score process which is repeated until an acceptable balance is verified. The imbalance is given by  $L_1$  statistics (Iacus et al. 2009), which derive from the relative frequency of treated ( $f$ ) and untreated groups ( $g$ ) in each stratum. That is:

$$L_1(f, g) = \frac{1}{2} \sum_{l_1 \dots l_k} |f_{l_1 \dots l_k} - g_{l_1 \dots l_k}| \quad (\text{A.4})$$

The matched dataset ( $f_m, g_m$ ) yields a lower imbalance than that resulting from the complete dataset ( $f, g$ ), that is:

$$L_1(f_m, g_m) \leq L_1(f, g) \quad (\text{A.5})$$

where  $L_1 \in (0, 1)$ .  $L_1 = 0$  if the balance is complete and the distributions of the two groups overlap, and  $L_1=1$  in case of maximum imbalance and complete separation between distributions. Propensity score matching is indeed defined only on the basis of an achieved balance and not on measures of fit (Iacus et al. 2009). In addition, unlike propensity scores, which run in order to improve the balance of all variables, CEM anticipates that the reduction of the imbalance on one variable does not have any effect on the imbalance of the other variables. This method drops observations in such a way that the pre-treatment characteristics between treated and untreated groups are more

similar, and therefore the balance between the two groups is improved. CEM implies coarsening of pre-treatment variables into strata, exact matching of the coarsened data, retaining the original values of the matched data, and dropping any observation the stratum of which does not contain at least one treated and one untreated unit. More specifically, CEM eliminates observations which do not have close matching in both treated and control groups on the pre-treatment characteristics. The finer the coarsening (and the more the strata), the lower the maximum imbalance. Coarsening is an ex ante choice and can be automated or personalized by the user.

# Bibliography

- Agrawal, A., Cockburn, I. and McHale, J.: 2003, Gone but not forgotten: Labor flows, knowledge spillovers, and enduring social capital.
- Agrawal, A., Kapur, D. and McHale, J.: 2008, How do spatial and social proximity influence knowledge flows? evidence from patent data, *Journal of Urban Economics* **64**(2), 258–269.
- Ahuja, G.: 2000, Collaboration networks, structural holes, and innovation: A longitudinal study, *Administrative science quarterly* pp. 425–455.
- Almeida, P. and Kogut, B.: 1999, Localization of knowledge and the mobility of engineers in regional networks, *Management science* pp. 905–917.
- Amin, A. and Roberts, J.: 2008, Knowing in action: Beyond communities of practice, *Research Policy* **37**(2), 353–369.
- Aral, S.: 2011, Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion, *Marketing Science* **30**(2), 217–223.
- Aral, S., Brynjolfsson, E. and Van Alstyne, M.: 2008, Productivity effects of information diffusion in networks, <http://ssrn.com/abstract=987499>. SSRN Working Paper.

## BIBLIOGRAPHY

---

- Aral, S., Muchnik, L. and Sundararajan, A.: 2009, Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks, *Proceedings of the National Academy of Sciences* **106**(51), 21544.
- Aral, S. and Walker, D.: 2010, Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks, *Available at SSRN: <http://ssrn.com/abstract=1564856>* .
- Ardichvili, A., Maurer, M., Li, W., Wentling, T. and Stuedemann, R.: 2006, Cultural influences on knowledge sharing through online communities of practice, *Journal of Knowledge Management* **10**(1), 94–107.
- Ardichvili, A., Page, V. and Wentling, T.: 2003, Motivation and barriers to participation in virtual knowledge-sharing communities of practice, *Journal of knowledge management* **7**(1), 64–77.
- Arrow, K.: 1962, The economic implications of learning by doing, *The review of economic studies* pp. 155–173.
- Baker, J.: 1987, The role of the environment in marketing services: the consumer perspective, *The Services Challenge: Integrating for Competitive Advantage*, American Marketing Association, Chicago, IL pp. 79–84.
- Baker, J., Grewal, D. and Levy, M.: 1992, An experimental approach to making retail store environmental decisions, *Journal of retailing* **68**, 445–445.
- Bandura, A.: 1965, Influence of models' reinforcement contingencies on the acquisition of imitative responses, *Journal of Personality and Social Psychology* **1**(6), 589–595.
- Bandura, A., Ross, D. and Ross, S.: 1963, Imitation of film-mediated aggressive models, *Journal of abnormal and social psychology* **66**(1), 3–11.

- Bandura, A., Ross, D. and Ross, S.: 2006, Transmission of aggression through imitation of aggressive models, *Educational psychology in context: readings for future teachers* p. 57.
- Banerjee, A.: 1992, A simple model of herd behavior, *The Quarterly Journal of Economics* **107**(3), 797–817.
- Bansler, J. and Havn, E.: 2003, Building community knowledge systems: an empirical study of it-support for sharing best practices among managers, *Knowledge and Process Management* **10**(3), 156–163.
- Barabasi, A., Jeong, H., Neda, Z., Ravasz, E., Schubert, A. and Vicsek, T.: 2002, Evolution of the social network of scientific collaborations, *Physica A: Statistical Mechanics and its Applications* **311**(3-4), 590–614.
- Barrat, A., Cattuto, C., Colizza, V., Pinton, J., Van den Broeck, W. and Vespignani, A.: 2008, High resolution dynamical mapping of social interactions with active RFID, *Arxiv preprint arXiv:0811.4170* .
- Bass, F.: 1969, A new product growth for model customer durables, *Management Science* **15**, 215–227.
- Battarbee, K.: 2003a, Co-experience: the social user experience, *CHI'03 extended abstracts on Human factors in computing systems*, ACM, pp. 730–731.
- Battarbee, K.: 2003b, Defining co-experience, *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, ACM, pp. 109–113.
- Battarbee, K. and Koskinen, I.: 2005, Co-experience: user experience as interaction, *CoDesign* **1**(1), 5–18.

## BIBLIOGRAPHY

---

- Becker, G.: 1965, A theory of the allocation of time, *The economic journal* **75**(299), 493–517.
- Becker, S. and Ichino, A.: 2002, Estimation of average treatment effects based on propensity scores, *The Stata Journal* **2**(4), 358–377.
- Bell, D. and Song, S.: 2007, Neighborhood effects and trial on the Internet: Evidence from online grocery retailing, *Quantitative Marketing and Economics* **5**(4), 361–400.
- Bikhchandani, S., Hirshleifer, D. and Welch, I.: 1992, A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades, *Journal of Political Economy* **100**(5), 992.
- Bikhchandani, S., Hirshleifer, D. and Welch, I.: 1998, Learning from the behavior of others: Conformity, fads, and informational cascades, *The Journal of Economic Perspectives* **12**(3), 151–170.
- Blackwell, M., Iacus, S., King, G. and Porro, G.: 2009, cem: Coarsened exact matching in Stata, *Stata Journal* **9**(4), 524–546.
- Bolton, R., Lemon, K. and Bramlett, M.: 2006, The effect of service experiences over time on a supplier’s retention of business customers, *Management Science* **52**(12), 1811–1823.
- Bolton, R. and Saxena-Iyer, S.: 2009, Interactive services: A framework, synthesis and research directions, *Journal of Interactive Marketing* **23**(1), 91–104.
- Bramoullé, Y., Djebbari, H. and Fortin, B.: 2009, Identification of peer effects through social networks, *Journal of Econometrics* **150**(1), 41–55.



- Branigan, C., Moise, J., Fredrickson, B. and Kahneman, D.: 1997, Peak (but not end) ANS reactivity to aversive episodes predicts bracing for anticipated re-experience, *Society for Psychophysiological Research, Cape Cod, MA* .
- Breschi, S. and Lissoni, F.: 2005, Knowledge networks from patent data, *Handbook of quantitative science and technology research* pp. 613–643.
- Breschi, S. and Lissoni, F.: 2009, Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows, *Journal of Economic Geography* **9**(4), 439.
- Brown, J. and Duguid, P.: 1991, Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation, *Organization science* **2**(1), 40–57.
- Brown, J. and Duguid, P.: 2001, Knowledge and organization: A social-practice perspective, *Organization science* pp. 198–213.
- Burnyeat, M.: 1981, Aristotle on understanding knowledge, *Aristotle on Science* pp. 97–139.
- Burt, R.: 1992, *Structural holes*, Harvard Univ. Press.
- Burt, R.: 2004, Structural holes and good ideas, *American journal of sociology* pp. 349–399.
- Cai, H., Chen, Y. and Fang, H.: 2007, Observational learning: Evidence from a randomized natural field experiment.
- Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J. and Vespignani, A.: 2010, Dynamics of person-to-person interactions from distributed RFID sensor networks, *PloS one* **5**(7), e11596.

## BIBLIOGRAPHY

---

- Cavalletti, F.: 2011, A strategy for non-strictly convex distance transport cost and the obstacle problem, *Arxiv preprint arXiv:1103.2797* .
- Chan, K. and Misra, S.: 1990, Characteristics of the opinion leader: a new dimension, *Journal of Advertising* pp. 53–60.
- Chiu, C., Hsu, M. and Wang, E.: 2006, Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories, *Decision Support Systems* **42**(3), 1872–1888.
- Christakis, N. and Fowler, J.: 2007, The spread of obesity in a large social network over 32 years, *New England Journal of Medicine* **357**(4), 370.
- Christakis, N. and Fowler, J.: 2008, The collective dynamics of smoking in a large social network, *New England journal of medicine* **358**(21), 2249.
- Christakis, N. and Fowler, J.: 2011, Contagion in prescribing behavior among networks of doctors, *Marketing Science* **30**(2), 213–216.
- Cialdini, R.: 2001, *Influence: Science and practice*, Allyn and Bacon Boston.
- Cialdini, R. and Goldstein, N.: 2004, Social influence: Compliance and conformity, *Psychology* **55**.
- Cialdini, R., Reno, R. and Kallgren, C.: 2004, A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places, *Readings About the Social Animal* pp. 56–79.
- Cohen, D. and Prusak, L.: 1996, British petroleum's virtual teamwork program, *Case study, Ernst & Young Center for Business Innovation* .
- Coleman, J., Katz, E. and Menzel, H.: 1966, *Medical innovation: A diffusion study*, Bobbs-Merrill Co.

- Combes, P. and Lafourcade, M.: 2003, Core-periphery patterns of generalized transport costs: France, 1978-98, *Discussion Paper Series - Centre for Economics policy Research London* .
- Combes, P. and Lafourcade, M.: 2005, Transport costs: measures, determinants, and regional policy implications for France, *Journal of Economic Geography* **5**(3), 319.
- De Long, D. and Fahey, L.: 2000, Diagnosing cultural barriers to knowledge management, *Executive - Ada then Briercliff* - **14**(4), 113–127.
- Della Vigna, S. and Malmendier, U.: 2006, Paying not to go to the gym, *American Economic Review* **96**(3), 694–719.
- Donovan, R. and Rossiter, J.: 1982, Store atmosphere: an environmental psychology approach, *Journal of Retailing* **58**(1), 34–57.
- Eagle, N. and Pentland, A.: 2006, Reality mining: sensing complex social systems, *Personal and Ubiquitous Computing* **10**(4), 255–268.
- Eagle, N., Pentland, A. and Lazer, D.: 2009, Inferring friendship network structure by using mobile phone data, *Proceedings of the National Academy of Sciences* **106**(36), 15274.
- Ebel, H., Mielsch, L. and Bornholdt, S.: 2002, Scale-free topology of e-mail networks, *Physical Review E* **66**(3), 35103.
- Fazio, R. and Zanna, M.: 1981, Direct experience and attitude-behavior consistency, *Advances in experimental social psychology* **14**, 161–202.
- Fleming, L., Mingo, S. and Chen, D.: 2007, Collaborative brokerage, generative creativity, and creative success, *Administrative Science Quarterly* **52**(3), 443–475.

## BIBLIOGRAPHY

---

- Fleming, L. and Sorenson, O.: 2001, Technology as a complex adaptive system: evidence from patent data, *Research Policy* **30**(7), 1019–1039.
- Fleming, L. and Sorenson, O.: 2004, Science as a map in technological search, *Strategic Management Journal* **25**(8-9), 909–928.
- Fowler, J. and Christakis, N.: 2008, Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study, *British Medical Journal* **337**(dec04 2), a2338.
- Fox, S.: 2000, Communities of practice, foucault and actor-network theory, *Journal of Management Studies* **37**(6), 853–868.
- Fredrickson, B. and Kahneman, D.: 1993, Duration neglect in retrospective evaluations of affective episodes., *Journal of Personality and Social Psychology* **65**(1), 45.
- Frigotto, M. and Riccaboni, M.: 2010, A few special cases: scientific creativity and network dynamics in the field of rare diseases, *Scientometrics* pp. 1–24.
- Geroski, P.: 2000, Models of technology diffusion, *Research Policy* **29**(4-5), 603–625.
- Ghose, A. and Han, S.: 2009, An Empirical Analysis of User Content Generation and Usage Behavior in Mobile Digital Media, *ICIS 2009 Proceedings* p. 190.
- Gino, F., Ayal, S. and Ariely, D.: 2009, Contagion and Differentiation in Unethical Behavior, *Psychological Science* **20**(3), 393.
- Gino, F. and Galinsky, A.: 2010, When Psychological Closeness Creates Distance from Ones Moral Compass, <http://ssrn.com/abstract=1612528>.

- Gino, F., Gu, J. and Zhong, C.: 2009, Contagion or restitution? When bad apples can motivate ethical behavior, *Journal of Experimental Social Psychology* **45**(6), 1299–1302.
- Glaeser, E. and Scheinkman, J.: 2003, Non-market interactions, *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. 1, pp. 339–369.
- Goffman, W. and Newill, V.: 1964, Generalization of epidemic theory, *Nature* **204**, 225–228.
- Goldstein, N., Cialdini, R. and Griskevicius, V.: 2008, A room with a viewpoint: Using social norms to motivate environmental conservation in hotels, *Journal of Consumer Research* **35**(3), 472–482.
- Grabisch, M. and Rusinowska, A.: 2008, A model of influence in a social network, *Theory and Decision* pp. 1–28.
- Graham, S.: 1997, Cities in the real-time age: the paradigm challenge of telecommunications to the conception and planning of urban space, *Environment and Planning A* **29**, 105–128.
- Granovetter, M.: 1973, The strength of weak ties, *American journal of sociology* pp. 1360–1380.
- Granovetter, M.: 1978, Threshold models of collective behavior, *ajs* **83**(6), 1420.
- Grinblatt, M., Keloharju, M. and Ikäheimo, S.: 2008, Social influence and consumption: Evidence from the automobile purchases of neighbors, *The Review of Economics and Statistics* **90**(4), 735–753.

## BIBLIOGRAPHY

---

- Guimerā, R., Danon, L., Diaz-Guilera, A., Giralt, F. and Arenas, A.: 2006, The real communication network behind the formal chart: Community structure in organizations, *Journal of Economic Behavior & Organization* **61**(4), 653–667.
- Haimila, S.: 2001, Shell creates communities of practice, *KM World* **19**, 1–2.
- Handley, K., Sturdy, A., Fincham, R. and Clark, T.: 2006, Within and beyond communities of practice: Making sense of learning through participation, identity and practice\*, *Journal of Management Studies* **43**(3), 641–653.
- Heim, G., Wentworth Jr, W. and Peng, X.: 2009, The Value to the Customer of RFID in Service Applications\*, *Decision Sciences* **40**(3), 477–512.
- Henderson, R. and Cockburn, I.: 1994, Measuring competence? exploring firm effects in pharmaceutical research, *Strategic management journal* **15**(S1), 63–84.
- Hill, S., Provost, F. and Volinsky, C.: 2006, Network-based marketing: Identifying likely adopters via consumer networks, *Statistical Science* **21**(2), 256–276.
- Hinde, R.: 1976, Interactions, Relationships and Social Structure, *Man* **11**, 1–17.
- Hornstein, H., Fisch, E. and Holmes, M.: 1968, Influence of a model's feeling about his behavior and his relevance as a comparison other on observers' helping behavior, *Journal of Personality and Social Psychology* **10**(3), 222–226.
- Iacus, S., King, G. and Porro, G.: 2009, Causal inference without balance checking: Coarsened exact matching, *Cambridge, MA: Institute for Quantitative Social Science, Harvard University* .

- Iacus, S., King, G. and Porro, G.: 2011, Multivariate matching methods that are monotonic imbalance bounding, *Journal of the American Statistical Association* **106**(493), 345–361.
- IHRSA: 2009, IHRA’s 2009 Global Report: the state of the health club industry. Boston, Ma.
- Iyengar, R., Valente, T. and Van den Bulte, C.: 2010, Opinion leadership and social contagion in new product diffusion.
- Iyengar, R. and Van den Bulte, C.: 2011, Further reflections on studying social influence in new product diffusion, *Marketing Science* **30**(2), 230–232.
- Iyengar, R., Van den Bulte, C. and Valente, T.: 2011, Opinion leadership and social contagion in new product diffusion, *Marketing Science* **30**(2), 195–212.
- Jarvenpaa, S. and Staples, D.: 2000, The use of collaborative electronic media for information sharing: an exploratory study of determinants, *The Journal of Strategic Information Systems* **9**(2-3), 129–154.
- Jeong, H., Neda, Z. and Barabási, A.: 2003, Measuring preferential attachment in evolving networks, *Europhysics Letters* **61**(4), 567–572.
- Jian, G. and Jeffres, L.: 2006, Understanding employees’ willingness to contribute to shared electronic databases, *Communication Research* **33**(4), 242.
- Katz, E. and Lazarsfeld, P.: 1955, Personal influence: the part played by people in the flow of mass communications.
- Kinney, A., Krebbers, E. and Vollmer, S.: 2004, Publications from industry. personal and corporate incentives, *Plant physiology* **134**(1), 11.

## BIBLIOGRAPHY

---

- Kirk, G., Raven, J. and Schofield, M.: 1983, *The presocratic philosophers: a critical history with a selection of texts*, Cambridge Univ Pr.
- Kline, S. and Rosenberg, N.: 1986, An overview of innovation, *The positive sum strategy: Harnessing technology for economic growth* **275**, 305.
- Kogut, B. and Zander, U.: 1992, Knowledge of the firm, combinative capabilities, and the replication of technology, *Organization science* pp. 383–397.
- Krause, J. and Ruxton, G.: 2002, *Living in Groups*, Oxford University Press.
- Kurtzberg, T. and Amabile, T.: 2001, From guilford to creative synergy: Opening the black box of team-level creativity, *Creativity Research Journal* **13**(3-4), 285–294.
- La Fond, T. and Neville, J.: 2010, Randomization tests for distinguishing social influence and homophily effects, *Proceedings of the 19th international conference on World wide web*, ACM, pp. 601–610.
- Lave, J. and Wenger, E.: 1991, *Situated learning: Legitimate peripheral participation*, Cambridge university press.
- Leavitt, P.: 2002, Applying knowledge management to oil and gas industry challenges, *American Productivity and Quality Center*.
- Lee, L., Fiedler, K. and Smith, J.: 2008, Radio frequency identification (RFID) implementation in the service sector: A customer-facing diffusion model, *International Journal of Production Economics* **112**(2), 587–600.
- Lim, K.: 2000, The relationship between publications and patents by researchers at five companies, *Technical report*, Sloan Working Paper.
- Loasby, B.: 2002, The evolution of knowledge: beyond the biological model, *Research Policy* **31**(8-9), 1227–1239.



- Manchanda, P., Xie, Y. and Youn, N.: 2008, The role of targeted communication and contagion in product adoption, *Marketing Science* **27**(6), 961.
- Manski, C.: 1993, Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* pp. 531–542.
- Manski, C.: 2000, Economic analysis of social interactions, *The Journal of Economic Perspectives* pp. 115–136.
- March, J.: 1991, Exploration and exploitation in organizational learning, *Organization science* pp. 71–87.
- Markus, M.: 2001, Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success, *Journal of management information systems* **18**(1), 57–93.
- Mas, A. and Moretti, E.: 2009, Peers at work, *American Economic Review* **99**(1), 112–145.
- Mazzola, D. and Distefano, A.: n.d., Crowdsourcing and the participation process for problem solving: the case of bp.
- McDermott, R.: 1999, Learning across teams, *Knowledge Management Review* **8**(3), 32–36.
- McPherson, M., Smith-Lovin, L. and Cook, J.: 2003, Birds of a feather: Homophily in social networks.
- Mehrabian, A. and Russell, J.: 1974, An approach to environmental psychology.
- Milgram, S.: 1967, The small world problem, *Psychology today* **2**(1), 60–67.

## BIBLIOGRAPHY

---

- Moore, R., Moore, M. and Capella, M.: 2005, The impact of customer-to-customer interactions in a high personal contact service setting, *Journal of Services Marketing* **19**(7), 482–491.
- Moretti, E.: 2011, Social learning and peer effects in consumption: Evidence from movie sales, *The Review of Economic Studies* **78**(1), 356.
- Nair, H., Manchanda, P. and Bhatia, T.: 2010, Asymmetric social interactions in physician prescription behavior: The role of opinion leaders, *Journal of Marketing Research* **47**(5), 883–895.
- Narayan, V., Rao, V. and Saunders, C.: 2011, How peer influence affects attribute preferences: a bayesian updating mechanism, *Marketing Science* **30**(2), 368–384.
- Nelson, R. and Winter, S.: 1982, *An evolutionary theory of economic change*, Belknap press.
- Newman, M.: 2001, The structure of scientific collaboration networks, *Proceedings of the National Academy of Sciences of the United States of America* **98**(2), 404.
- Newman, M.: 2004, Who is the best connected scientist? a study of scientific coauthorship networks, *Complex networks* pp. 337–370.
- Ngai, E., Moon, K., Riggins, F. and Yi, C.: 2008, RFID research: An academic literature review (1995-2005) and future research directions, *International Journal of Production Economics* **112**(2), 510–520.
- Nirenberg, J.: 1994, From team building to community building, *National Productivity Review* **14**(1), 51–62.

- Oh, W. and Jeon, S.: 2007, Membership herding and network stability in the open source community: The Ising perspective, *Management science* **53**(7), 1086.
- Oliver, R.: 1999, Whence consumer loyalty?, *The Journal of Marketing* **63**, 33–44.
- Osterloh, M. and Frey, B.: 2000, Motivation, knowledge transfer, and organizational forms, *Organization science* pp. 538–550.
- Padgett, J. and Ansell, C.: 1993, Robust action and the rise of the medici, 1400-1434, *American Journal of Sociology* pp. 1259–1319.
- Padgett, J. and Powell, W.: 2009, The emergence of organizations and markets.
- Penrose, E.: 1959, The growth of the firm, *White Plains, New York: ME Sharpe* .
- Pisano, G., Shuen, A. and Teece, D.: 1997, Dynamic capabilities and strategic management, *Strategic Management Journal* **18**(7), 509–533.
- Powell, W., White, D., Koput, K. and Owen-Smith, J.: 2005, Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences<sup>1</sup>, *ajs* **110**(4), 1132–205.
- Probst, G. and Borzillo, S.: 2008, Why communities of practice succeed and why they fail, *European Management Journal* **26**(5), 335–347.
- Ramanathan, S. and McGill, A.: 2007, Consuming with others: Social influences on moment-to-moment and retrospective evaluations of an experience, *Journal of Consumer Research* **34**(4), 506–524.

## BIBLIOGRAPHY

---

- Ratti, C., Williams, S., Frenchman, D. and Pulselli, R.: 2006, Mobile Landscapes: using location data from cell phones for urban analysis, *Environment and Planning B: Planning and Design* **33**(5), 727.
- Roberts, J.: 2006, Limits to communities of practice, *Journal of Management Studies* **43**(3), 623–639.
- Rojas, F., Calabrese, F., Krishnan, S. and Ratti, C.: 2008, Real Time Rome.
- Rosenbaum, P. and Rubin, D.: 1983, The central role of the propensity score in observational studies for causal effects, *Biometrika* **70**(1), 41.
- Rosenkopf, L. and Nerkar, A.: 2001, Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry, *Strategic Management Journal* **22**(4), 287–306.
- Russell, B. and Marić, S.: 1945, *History of western philosophy and its connection with political and social circumstances from the earliest times to the present day*, Simon and Schuster.
- Sacerdote, B.: 2001, Peer Effects with Random Assignment: Results for Dartmouth Roommates\*, *Quarterly Journal of Economics* **116**(2), 681–704.
- Schenkel, A. and Teigland, R.: 2008, Improved organizational performance through communities of practice, *Journal of Knowledge Management* **12**(1), 106–118.
- Schloter, P. and Aghajan, H.: 2005, Wireless rfid networks for real-time customer relationship management, *Embedded and Ubiquitous Computing* pp. 1069–1077.
- Schumpeter, J.: 1939, *Business cycles*, Vol. 100, Cambridge Univ Press.

- Schwartzman, R., Ross, D. and Berube, D.: 2011, Rhetoric and risk, *Poroi* **7**(1), 9.
- Sharratt, M. and Usoro, A.: 2003, Understanding knowledge-sharing in online communities of practice, *Electronic Journal on Knowledge Management* **1**(2), 187–196.
- Sieloff, C.: 1999, if only hp knew what hp knows: the roots of knowledge management at hewlett-packard, *Journal of Knowledge management* **3**(1), 47–53.
- Sih, A., Hanser, S. and McHugh, K.: 2009, Social network theory: new insights and issues for behavioral ecologists, *Behavioral Ecology and Sociobiology* **63**(7), 975–988.
- Simard, D., Nadeau, L. and Kroger, H.: 2005, Fastest learning in small-world neural networks, *Physics Letters A* **336**(1), 8–15.
- Smith, A.: 1982, Lectures on jurisprudence (1766), *Indianapolis: Liberty Fund* .
- Smith, R. and Swinyard, W.: 1982, Information response models: An integrated approach, *The Journal of Marketing* **46**(1), 81–93.
- Snijders, T., Steglich, C. and Schweinberger, M.: 2007, Modeling the co-evolution of networks and behavior, *Longitudinal models in the behavioral and related sciences* pp. 41–71.
- Soetevent, A.: 2006, Empirics of the Identification of Social Interactions; An Evaluation of the Approaches and Their Results\*, *Journal of Economic Surveys* **20**(2), 193–228.
- Stoneman, P.: 2002, *The economics of technological diffusion*, Wiley-Blackwell.

## BIBLIOGRAPHY

---

- Szulanski, G.: 1996, Exploring internal stickiness: Impediments to the transfer of best practice within the firm, *Strategic management journal* **17**, 27–43.
- Tajfel, H.: 1982, *Social identity and intergroup relations*, Cambridge University Press Cambridge.
- Tajfel, H. and Turner, J.: 2004, The social identity theory of intergroup behavior.
- Taylor, C.: 1990, Aristotles epistemology, *Epistemology* **1**, 116.
- Tokarchuk, L., Shoop, K. and Ma, A.: 2009, Using co-presence communities to enhance social recommendation, *Wireless On-Demand Network Systems and Services, 2009. WONS 2009. Sixth International Conference on*, IEEE, pp. 169–172.
- Townsend, A.: 2000, Life in the real-time city: Mobile telephones and urban metabolism, *Journal of urban technology* **7**(2), 85–104.
- Tucker, C.: 2008, Identifying formal and informal influence in technology adoption with network externalities, *Management science* **54**(12), 287–304.
- Uzzi, B. and Spiro, J.: 2005, Collaboration and creativity: The small world problem, *ajs* **111**(2), 447–504.
- Valente, T.: 1995, *Network Models of the Diffusion of Innovations*, Hampton Press, Cresskill, NJ.
- Valente, T.: 1996a, Network models of the diffusion of innovations, *Computational & Mathematical Organization Theory* **2**(2), 163–164.
- Valente, T.: 1996b, Social network thresholds in the diffusion of innovations, *Social Networks* **18**(1), 69–89.

- Valente, T.: 2005, Network models and methods for studying the diffusion of innovations, *Models and methods in social network analysis* pp. 98–116.
- Van den Bulte, C. and Lilien, G.: 2001, Medical innovation revisited: Social contagion versus marketing effort, *American Journal of Sociology* **106**(5), 1409–1435.
- Van Looy, B., Callaert, J. and Debackere, K.: 2006, Publication and patent behavior of academic researchers: Conflicting, reinforcing or merely co-existing?, *Research Policy* **35**(4), 596–608.
- Von Hippel, E.: 2009, Democratizing innovation: the evolving phenomenon of user innovation, *International Journal of Innovation Science* **1**(1), 29–40.
- Wagner, C. and Leydesdorff, L.: 2005, Network structure, self-organization, and the growth of international collaboration in science, *Research Policy* **34**(10), 1608–1618.
- Wasko, M. and Faraj, S.: 2005, Why should i share? examining social capital and knowledge contribution in electronic networks of practice, *Mis Quarterly* **29**(1), 35–57.
- Wasserman, S. and Faust, K.: 1994, *Social network analysis: Methods and applications*, Cambridge Univ Pr.
- Watts, D.: 2002, A simple model of global cascades on random networks, *Proceedings of the National Academy of Sciences of the United States of America* **99**(9), 5766.
- Watts, D. and Dodds, P.: 2007a, Influentials, networks, and public opinion formation, *Journal of Consumer Research* **34**(4), 441–458.
- Watts, D. and Dodds, P.: 2007b, The accidental influentials, *Harvard Business Review* **85**(2), 22–23.

## BIBLIOGRAPHY

---

- Watts, D. and Strogatz, S.: 1998a, Collective dynamics of small-world networks, *Nature* **393**(6684), 440–442.
- Watts, D. and Strogatz, S.: 1998b, Small world, *Nature* **393**, 440–442.
- Watts, D. and Strogatz, S.: 2006, Collective dynamics of small-world networks, *The Structure and Dynamics of Networks* **393**, 440–442.
- Wellman, B. and Wortley, S.: 1990, Different strokes from different folks: Community ties and social support, *The American Journal of Sociology* **96**(3), 558–588.
- Wenger, E.: 1998, Communities of practice: Learning as a social system, *Systems thinker* **9**(5), 1–5.
- Wenger, E.: 2007, *Communities of practice: Learning, meanings, and identity*, Cambridge university press.
- Wenger, E., McDermott, R. and Snyder, W.: 2002, *Cultivating communities of practice: A guide to managing knowledge*, Harvard Business Press.
- Whitehead, H.: 2008, *Analyzing animal societies: quantitative methods for vertebrate social analysis*, University Of Chicago Press.
- Whitehead, H.: 2009, SOCPROG programs: analysing animal social structures, *Behavioral Ecology and Sociobiology* **63**(5), 765–778.
- Yoon, B. and Park, Y.: 2004, A text-mining-based patent network: Analytical tool for high-technology trend, *The Journal of High Technology Management Research* **15**(1), 37–50.
- Young, A.: 1993, Invention and bounded learning by doing, *Journal of Political Economy* pp. 443–472.



- Zajonc, R.: 1965, Social facilitation., *Science* .
- Zhang, J.: 2010, The sound of silence: Observational learning in the us kidney market, *Marketing Science* **29**(2), 315–335.
- Zhao, S.: 2003, Toward a taxonomy of copresence, *Presence: Teleoperators & Virtual Environments* **12**(5), 445–455.
- Zucker, L., Darby, M. and Armstrong, J.: 1998, Geographically localized knowledge: spillovers or markets?, *Economic Inquiry* **36**(1), 65–86.