Diffusion of Ideas, Practices, and Artifacts: Network Effects on Collective Outcomes

by

Juan Carlos Barahona

Master in Business Administration
INCAE Business School, 1993

Submitted to the Program of Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MEDIA ARTS AND SCIENCES
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 2007

Signature of Author: _____
Program in Media Arts and Sciences
August 3rd, 2007

Certified by _____
Alex (Sandy) Pentland
Toshiba Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by _____
Deb Roy
Chairperson, Departamental Committee on Graduate Studies
Program in Media Arts and Sciences

Diffusion of Ideas, Practices, and Artifacts: Network Effects on Collective Outcomes

by

Juan Carlos Barahona

Submitted to the Program of Media Arts and Sciences,
School of Architecture and Planning, on July, 2007
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

ABSTRACT

Important ideas, practices and artifacts often fail to reach their target population efficiently or fail to reach altogether. Surprisingly, most projects aimed to bring technology to underserved communities of the world lack an explicit diffusion strategy and/or lack an implementation strategy that acknowledges the social structure that binds together the members of the targeted community. Without the knowledge of social structures efficient diffusion of technological innovations becomes an unreachable goal.

Socioeconomic and behavioral information can be combined with sparse social structure data to derive quantitative estimates of a community's social dynamics, allowing improved understanding and management of diffusion processes. We found that patterns of advice and use of media provide and effective way to identify the influential members of a community. We set up a large scale experiment in a rural community using our model and tested our proposed method of intervention and found strong evidence of an improved diffusion process which is significantly related to the communities' network of advice.

The adoption of an idea, practice or artifact is heavily influenced by social context, through both conscious and unconscious mechanisms. By targeting social networks, not social classes, age, gender groups or institutions, we can create the basis for the emergence of local organizations and businesses that organically provide the necessary support to achieve effective diffusion of technological innovations. In our experiment, the introduction of a few powerful ideas at the core of the communities' social networks helped to create a social context where the new innovations created economic and social value. In addition, these new business create a richer social context from which further new innovations are expected to emerge.

Thesis Supervisor: Alex (Sandy) Pentland
Title: Toshiba Professor of Media Arts and Sciences

Diffusion of Ideas, Practices, and Artifacts:
Network Effects on Collective Outcomes

by

Juan Carlos Barahona

Thesis Committee:

Thesis Advisor: _____

Alex (Sandy) Pentland
Toshiba Professor of
Media Arts and Sciences
MIT Media Laboratory

Thesis Reader: _____

Mitchel Resnick
LEGO Papert Professor of
Learning Research
Media, Arts and Sciences
MIT Media Laboratory

Thesis Reader: _____

Randy Zadra
Managing Director,
Institute for Connectivity
in the Americas and
Visiting Scientist
MIT Media Laboratory

To Maritza

For your inspiring courage to dream and live

Acknowledgments

This dissertation could not have been written without the intellectual guidance and judicious support of my advisor Professor Alex(Sandy) Pentland. During the past five years he has been a constant source of encouragement, ideas and challenges. I am also very grateful for having an exceptional doctoral committee and wish to thank Mitch Resnick for his continuous support during the doctoral program and Randy for his enthusiastic belief in the positive impact that this research may have. I also want to thank Professor Henry Lieberman's valuable feedback in preparation for the critique and defense of this thesis.

I want to thank past and present graduate students in my research group (the Pentlandians). It is a true privilege to be surrounded by so talented people. I learned so much from all of them. Daniel, Tammie and Ben, you are a great team and I wish you the best. Among all, I want to give special thanks to Tanzeem Choudhury for her warmth and welcoming attitude and guidance in my early days at MIT. I am also grateful to my office mates, Akshay Mohan's generosity and friendship and Weng Don's intelligent critique of research ideas and sense of humor. I also want to thank a very important team member, Mary Heckbert who really makes a difference as the group manager.

A special note of gratitude goes to Professor Bakhtiar Mikhak and his loving family. I was encouraged by him to enter the doctoral program at the Media Lab and since then he has been a source of personal and professional advice and I am fortunate to count him among my dearest friends. Also, I should mention Wesley Williams and the many hours of discussions that eventually led to chapter three and the rest of my thesis. Abby and Claudia who's support and companionship made the work of writing this document a lot easier. Manuel Esquivel, my UROP did an extraordinary work collecting and organizing the data I used in chapter four.

Next, I want to acknowledge the deep influence upon my research and teaching skills that Professor Peter Marsden has made. I first met Professor Marsden in his Social Network Analysis Seminar at Harvard. After the first few classes, my research work had a new direction. I owe him and Professor Chris Winship most of the analytical tools I used in this thesis. Both set for me the golden standard in teaching excellence.

My work at MIT since day one, has been related with a collection of projects in Costa Rica. I owe so much to so many people that it makes it impossible to list all of their names. But I should at least mention the leadership and unstintingly provided support of former INCAE president Dr. Roberto Artavia, current INCAE president Dr. Arturo Condo and former dean Dr. Alberto Trejos. I was fortunate enough to have the intellectual and material support of two Costa Rican NGOs. The first one, Fundación Costa Rica para el Desarrollo Sostenible was a source of constant support. I am grateful to all of them but especially, I am grateful to President José María Figueres and Dr. Juan Barrios. CEMEDCO under the leadership of emeritus professor Leda Beirute was instrumental in the early stages of the research project. She and her colleagues provided a testing ground for new ideas, and many hours of hard work with members of the Los Santos Community members, helping to set up the experiments, registering observations and collecting data.

This research would not have been possible without the generosity of the people of Los Santos. I have no words to thank Rodrigo Jiménez for his leadership, his constant support and his friendship. Roberto Mata and Elías Calderón and many people that work with them providing material support and data that were key to complete the experiments.
Among the many new friends that I met during my visits to Costa Rica, I want to thank the support and friendship of Ivannia and Natalia Ureña, their hospitality and generosity know no limits.
Finally, I want to thank my family. Moving to Boston was I huge challenge for my wife Vivian. She reinvented herself in different and positive ways. Her growth can only be parallel to the admiration I feel for her. María and Alberto, thank you for your patience, your love and your decision to share with me this wonderful ride. Last but not least, my father Dr. Rodrigo, thank you for always keeping the bar high!

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Classical writers such as John Stuart Mill and Karl Marx speculated that the standard of living could not rise indefinitely unless advances in technology increased the yield of the means of production. Neoclassical growth theory, based on capital accumulation, supports this intuition [1]. The last decade has witnessed increasing efforts to develop and deploy Information and Communication Technologies for Development (ICT4D) in underserved communities. Unfortunately, there are many more failures than success stories. This thesis provides a novel approach to social and economic development at the village level. It extends a new perspective on the role and dynamics of diffusion of technological innovations. The thesis will develop arguments for the need to understand the role of social networks and why they should be a unit of analysis for understanding, describing and dealing with community dynamics. It will present evidence of the importance of influential members of the community in optimizing diffusion processes. It will propose and test an original method to quantitatively and efficiently identify the influential members of a community. It will present an empirical study that demonstrates how the new method doubles the precision and reduces, by orders of magnitude, the time, effort and financial resources of current methods of finding these "influentials". These claims are planned, developed and evaluated through empirical work done within an experimental setting in the southern mountains of Costa Rica. Empirical

evaluation is complemented by literature review and theoretical analysis of competing methods.

1.1 Motivation

Wealth, innovation and economic growth are the result of ideas, collaboration and the ability to transform those ideas into valuable tools, products or services. This is true at personal, business, community and national level. Ideas come in a variety of forms, including knowledge, values, attitudes, information and coordination. A better flow of ideas generates wealth. ICT4D should provide to the right people a means to enhance substantially the way and frequency of their contacts and the quantity and quality of their relationships. The last decade has seen an increased effort to diffuse the use of ICT to improve the life of billions living in underserved communities. Unfortunately, most of them simply fail and early success stories become ephemeral as the critical point for massive adoption is never reached and efforts to push technological innovations become extinguished.

For a decade an impressive collection of world class universities and corporations have being trying to solve their sustainability problem and a solution that actually works has not yet been achieved. We think it is because the problem is not being framed adequately. The core of the problem is that modern digital technologies embody a diverse conglomerate of complex ideas, practices and artifacts, yet they are frequently dumped into communities without a clear strategy to manage their diffusion amongst its members[1][2].

---

1 UNDP's 2005 book titled "Community-based Networks and Innovative Technologies: New Models to Serve and Empower the Poor is a recent and powerful example of how social structure and diffusion of technological innovations concepts are totally absent in the International Development

Around the world thousands of stories and photo opportunities for politicians and corporate marketers emerged [3, 4]; however, it is hard to find statistics of the many ICT4D projects that failed to fulfill their promises. Research on ICT4D such as telecenters, in general, is at best sparse and anecdotal[5]. Telecenters are a good example of this problem. The financial agencies no longer know what to do with them as affirmed by Klaus Stoll, president of the largest and earliest regional telecenters network in the world[6].

The proposed ideas in this thesis come from the search for an alternative solution to the paramount challenge of fostering economic and social development out of the urban settings. This thesis will provide hard data and statistical analysis to establish the empirical significance of the key elements of the proposal, but will also provide stories or cases that help clarify powerful ideas that are substantial for our closing discussion linking community development as a problem of diffusion of ideas, practices and artifacts.

1.1.1 This is an old and unsolved problem

The spread of new ideas has a dramatic impact on our lives, but we still need to improve our understanding and ability to manage the diffusion of ideas and innovations. Diffusion of Innovations pioneer Everett Rogers documented the case of scurvy in the British Navy, a good example of how inefficient the flow of ideas in an organization or community can be: Basically, the effectiveness of lemon juice in preventing this fatal illness was discovered by a British Navy Captain in 1601 and it took almost 200 years for this innovation to be adopted by the Navy, where upon scurvy was immediately wiped out [7]. New media is not enough to diffuse ideas. Today, after the impressive media development of the past 100 years, we can

---

Community. It uses over 74,000 words to describe the state of the art and new directions for making

observe with dismay how similarly simple solutions to similarly serious health problems suffer comparable rates of diffusion. One modern example is iodine deficiency, which for more than 80 years has been recognized as the most common cause of preventable brain damage and mental retardation. It is solved by simply adding iodine to salt, yet one out of every three human beings has some degree of iodine deficiency [8].

1.1.2 Media Lab's life long commitment with technology and development

Since its early days, the Media Laboratory has been an important contributor to developing technologies and methods that improve people lives around the world. Our work is aimed at learning from our recent past and current projects in this endeavor. We hope to contribute with a new framework of analysis, an original methodology of intervention based on quantitative analysis of the community's web of interactions.

1.1.3 G-Lab, LINCOS and System Dynamics

My first contact with this community as a Media Lab student was a Media Lab/Leaders for Manufacturing Program class project. It was set to explore "how digital technology could be applied to improve the economics of a rural community in a developing country, where infrastructure is minimal and marketing techniques must differ from those in the developed world"[9]. The context was two of the three main towns of Los Santos Region in Costa Rica. In one of them, San Marcos, LINCOS had recently shut forever its telecenter. Conventional System Dynamics ideas and tools were used to understand the social dynamics that first helped and later condemned the telecenters initiative in this region. An unexpected result was

---

ICT work for the poor" and the word diffusion is not even used once in the whole document.

found. By most commonly used standards within the international development community their local LINCOS may be thought by some as failed, but we arrived to the opposite conclusion when we realized the dramatic and mostly ignored impact the telecenter had on some of the people we were interviewing. Their new technological abilities and the project itself worked as a "weaver", linking people to social and business networks to which they do not belong. The network was a real source of financial and social capital that was actually being used by a few emerging local entrepreneurs. What LINCOS had unknowingly provided was a few business ideas using digital tools and communication technologies and a few people managed to use their new skills to connect and become a member of richer networks within the community. They were now able to provide a service and in exchange access human and financial support to develop their small business, a few of them initially in direct competition to LINCOS' services. Initially we tried to model our ideas using the System Dynamics concepts and software, but the experience proved to me that it was an inadequate tool to capture these phenomena. Chapter 2 will discuss the limitation of this tool to model the phenomena we are interested in. Business people frequently think in terms of clusters, or "geographic concentrations of interconnected companies". Using this mainly descriptive framework, one may say that with the new digital tools and skills a few San Marcos community members became part of the "local coffee cluster", but it understates the strong intuition we got from revisiting LINCOS with the G-Lab team: ICT was unintentionally used to build new *links* among community members that will change their socioeconomic status within the community in a sustainable way. We started to think that "building" these links should be the core product of ICT4D efforts and not a byproduct.

1.2 Thesis Organization

In Chapter 3, we claim that the relevant unit of analysis for the diffusion of advanced technologies is the community's social networks of advice, and that flow of ideas within these networks can be used to identify the "influentials" in order to better promote rapid diffusion of ideas. Diffusion efforts should have as an objective to increase the collective knowledge base of the new technologies and surpass the critical threshold of adoption. Valente and others have shown that sociometric concepts such as centrality are key to accelerating diffusion. We have found in our own research that the most influential people in the community should be the "entry" or starting points of the diffusion process as well[10]. We have also found that they are not always visible members of the community, and we have also found a quantitative method based on sociometric measures of advice networks that is an efficient way to find the influentials. We have also tested our intuitions and ideas by seeding a few powerful ideas in a real community with promising findings.

We have accumulated data on the social structure of a Costa Rican village and have conducted interventions aimed at people selected by their positional attributes within the social network structure. We have found out the key role played in diffusion by people considered as influential and as problem solvers by community members. Because the thesis' main focus and contribution is planned to be around a theoretically sound and empirically tested method of finding the influentials to optimize the diffusion of technological innovations, we conducted an experiment on diffusion of technological ideas. Its results are presented in Chapter 4. They represent strong evidence of the positive impact of our intervention in the rate of adoption of technology. The broader implications are discussed around business cases presented in Chapter 5. These cases show the impact at personal level

on five entrepreneurs, and how their new businesses are creating a richer "ecosystem" that should benefit the community as a whole.

Sociometric data is expected to become more available and easier to record and process, as mobile phones, computers of all sizes and the Internet become ubiquitous and better algorithms for data mining from those devices evolve. This work is part of a larger research agenda aimed at designing methods and applications informed by the structural properties of human dynamics to improve the flow of ideas and innovations.

Chapter 2: Social Networks and Diffusion of Innovations

2.1 Social Network Analysis

Social Network Analysis provides an alternative to the traditional individualism of most mainstream social research[11]. Social Network Analysis has been used in many different disciplines such as anthropology, biology, business (organizational behavior and marketing), computers science, ecology, epidemiology, ethology, mathematics, political science, psychology, communication science, statistics and lately physics and Artificial Intelligence.

With the renewed interest of physicists on networks, the explosion of networking sites over the Internet and the success of popular science books like Linked, The Tipping Point and The Wisdom of Crowds; the pervasiveness of networks has become evident and probably one of the defining paradigms of modern culture[12].

Literature claims at least three sources of concepts and tools in the development of Social Networks as a research field:

1) Linton C. Freeman, in his book "The Development of Social Network Analysis" defines the field as an approach to social research that displays four features: structural intuition; systematic relational data; graphic images and mathematical and computational models. For each feature he tracked back to the earliest works. So, according to Freeman, August Comte, the founder of Sociology, was the first scholar

who proposed a way of looking at society in terms of its actor's interconnections. For the systematic relational data he points at Swiss entomologist Pierre Huber and his work on bumblebees in 1802. As for the graphic component, he talks about the use of trees to describe kinship in the ninth century and the diagrams of Louis Henry Morgan on kinship data in 1858. As for mathematical modeling, he cites the algebraic models of Alexander MacFarlane. Finally, all four features came together in the 1930's when sociometry was introduced in New York by the Viennese Jacob Levy Moreno and Helen Hall Jennings.

2) Graph theory. In the sciences, the study of networks goes back to 1736, with the works of Leonard Euler (the Königsberg Bridge Problem). To solve a popular riddle he used a graph, i.e. "a mathematical object consisting of points also called vertices or nodes, and lines, also called edges or links, which abstracts away all the details of the original problem except for its connectivity"[13]. Graph theory spread from mathematics to other disciplines where abstracting details to study the topology of a latent network was useful, like applications in engineering, operations research, computer science and among all sociology and anthropology. Social Network Analysis borrowed from graph theory important nomenclature that allows describing abstract models and also became a practical tool for the analysis of empirical data[14].
3) Anthropologically-inclined organizational field workers based in the Harvard Business School during the 1920s (The Hawthorne Studies)[12].

2.2 Diffusion of Innovations Research

Business literature also provides a source of diffusion of innovation research. System Dynamics has being the leading modeling tool of diffusion of innovations in the business

research community. The modeler using System Dynamics will represent real-world processes in terms of stocks (inventories, materials, people, money, etc), flows between these stocks, and information that determines the values of the flows. It is an aggregated view, thus concentrating in policies. The modeler has to describe the system behavior as a number of interacting feedback loops, balancing and reinforcing. It is basically a system of differential equations. As long as the model works only with aggregates, the items in the same stock are indistinguishable, the modeler has to think in terms of global structural dependencies and has to provide accurate quantitative data for them[15].

In this field, Sterman's work on Business Dynamics[16] has been highly influential and provides a useful text book and tool to model a dynamic system. This particular tool helps to identify the interactions of multiple variables affecting complex systems. The limitation to its use comes from the difficulty of estimating the right parameters for the system being modeled. Many applications of business dynamics tools are weakened with assumptions hard to sustain out of the classroom. In the specific case of diffusion of innovations, it is traditionally modeled using the logistic or the bass model, and the network effect is recognized with vague concepts like "word of mouth".

2.3 Social Networks Approach to Diffusion of Innovations

The barriers to see more frequent use of sociometric data lay in the challenges of data measurement, model definition and the selection of research methods and different network weighting mechanisms to fit the selected model and test the results. The computing power to handle relatively complex analysis and potentially huge data sets is relatively new.

2.4 Diffusion of Innovations and Social Influence

Diffusion of innovations is treated usually as social contagion. Similar to the spread of disease, there is a chain reaction phenomenon. First there are a few adopters, then members of their networks adopt, and then they pass to their own networks and so on. First slow, then faster and faster, and then slows down again as the amount of potential adopters decreases. Mathematically, this phenomenon is best described by a logistic S-shape curve and it can be easily constructed from a simple frequency tabulation of the adoption time. If one measures a network to model a process of diffusion and the resulting graph has a poor or inexistent fit to a logistic curve, that is probably not the right network, or there is another more powerful effect other than network diffusion[17].

Social influence is an important process in diffusion of innovations analysis. It can be modeled, at least, using three different classes of network weight matrices that can be produced by the same network data[18]:

Relational: Direct ties, indirect ties, two-mode incidence (subjects and events).

Positional: Percent positive matches, Euclidean distance, regular equivalence.

Central: Degree, closeness, betweeness, eigenvector, integration/radiality, flow centrality and information centrality, etc.

Assuming constant the perceived risk or advantage of an innovation and personal characteristics, if the fit to a logistic model is good and the network is a plausible explanation of the diffusion process then one can expect that[17]:

- High network density[2] will contribute to the easy and fast spread of an innovation. Low density (sparse networks) will act in the opposite direction [relational]
- Diffusion will be slower and less comprehensive than in a more connected network [relational]
- If there are pockets of interconnectivity, diffusion spreads fast within the densely connected subgroups and slower between groups.
- In a bi-component network diffusion will be faster than in components with cut-points or bridges [structural]
- The larger the ego network of a vertex within the network, the earlier it will adopt an innovation [relational]
- The greater the amount of innovators in a vertex's ego network, the faster it will adopt. [Relational]
- The greater the amount of members in a vertex's ego network that have adopted the higher the likelihood that it will adopt in which case exposure is a function of time and distance.
- A central position is likely to lead to early adoption [central]
- Diffusion from a central vertex (core) is faster than from a vertex in the margins of the network (periphery). [Structural Property]
- Core members tend to be more connected among themselves than periphery members, and periphery members tend to be connected to the core members, therefore when most of the core has adopted, then most likely criticality has being achieved and an avalanche of adoptions should occur[17]
- Since some people are more influential or persuasive than others, and some people are more difficult to persuade than others, one should not expect a strong statistical

---

[2] Proportion of existing to maximum possible links.

association between adoption and exposure in their local network in the early stages of diffusion[3]. [personal characteristic]

- Lower thresholds should be associated to a propensity to innovativeness. [personal characteristic]

The last two hypotheses are not directly related with the relational or structural approach, and involve time. We add them because once time is incorporated, the early adopters will most likely be those with a low threshold. And in general, innovativeness and low thresholds are supposed to be related to broad media use, high education level, high socioeconomic status, cosmopolitan contacts and cosmopolitan (outside local community) news preferences[10, 17, 19].

2.4.1 Network Diffusion Models and Statistical Methods

Network diffusion models used by sociologists and the business communities among others are logistic growth (basic model), the Bass Model, Spatial autocorrelation and Network auto correlation.

2.4.2 Logistic Growth

It is based in the cumulative patterns of diffusion. This model follows a growth pattern, which has proven to be a consistent pattern through decades of empirical research on DOI. It is approximated by a logistic function[20]. It has one parameter and has limited applicability. It could be used to compare growth rates for various innovations. It is described by

---

3 The amount of exposure than an individual needs to adopt is called the individual's threshold.

$$y_t = b_0 + \frac{1}{1+e^{-b_1 t}} \tag{2.1}$$

Where t equals time, $-b_1 t$ is the rate parameter to be estimated and b0 is the intercept.

2.4.3 The Bass Model

This model has two parameters: the rate for innovation (b0) and the rate for imitation (b1). It assumes perfect social mixing, i.e. everyone interacting with everyone else. Thus, it does not measure whether people who are connected to one another engage in the same behaviors. It is described by

$$y_t = b_0 + (b_1 - b_0) y_{t-1} - b_1 \left(y_{t-1}\right)^2 \tag{2.2}$$

The parameter b0 can be interpreted as external influence or innovativeness. The parameter b1 can be interpreted as internal influence or interpersonal persuasion. This model is used to forecast expected levels of diffusion. Interpretation of parameters is highly dependent on the time scale used to measure diffusion. It incorporates the % of adopters at each time point. Thus, it makes a better estimate of the growth attributable to personal network persuasion.

2.4.4 Spatial autocorrelation

This model measures the spread between contiguous areas. It uses proximity data to produce a network of connections based on distance. The model tests for spatial association (geographic clustering) of adoption. It is described by

$$I = \frac{N \sum_i^N \sum_j^N D_{ij} (y_i - \bar{y})(y_i - \bar{y})}{S \sum_i^N (y_i - \bar{y})^2} \tag{2.3}$$

Where N is the sample size, y is adoption, S is the sum of distances in D and D is the distance matrix (proximities). I measures the degree to which nodes that are connected to one another deviate from the average behavior in the network similarly or differently. A high "I" indicates positive or negative difference from the average score.

2.4.5 Network autocorrelation models

Network models measure personal or network exposure as opposed to random mixing. In network exposure models it is very important to control for clustering[18]. It is described by

$$E_i = \frac{\sum \omega_{ij} y_i}{\sum \omega_i} \tag{2.4}$$

Where E_i is the estimated proportion of contacts who are adopters, it is a measure of autocorrelation, y_i is a vector of adoptions, ω_i is the social network weight matrix of direct contacts (relational) or centrality to reflect influence by opinion leaders (central) or it can be transformed to represent the degree of structural equivalence (positional).

Testing social influence requires at least two time points to model a simple dynamic process expressed by

$$\log \frac{\Pr(y_t = 1)}{\left(1 - \Pr(y_t = 1)\right)} = \alpha + \sum B_k X_k + B_{(k+1)} \omega_t y_t + B_{(k+2)} \omega_{(t-1)} y_{(t-1)} \tag{2.5}$$

Panel data collected at two time periods are adequate for most research needs and can provide evidence of network influence on behavior. Nonetheless, there are confounding and non-observed variable problems. This model extends to a more powerful and robust model that is increasingly being use by the social network analysis community.

Event history analysis, also called survival analysis or Cox proportional hazard model is a regression model for longitudinal event data[21]. It allows managing a substantive number of time points. There are two types. Discrete type deals with binary outcome. The continuous is used in the case in which the dependent variable is timed to an event. Data needs to be reshaped to a case-time format, and then coefficient estimates are produced using Maximum Likelihood estimators. The independent variable is if adoption occurs or not. It is described by

$$\log \frac{\Pr(y_t = 1)}{(1 - \Pr(y_t = 1))} = \alpha + \sum B_j X_j + \sum B_{kt} X_{k2} + \sum B_{(k+1)} \omega y_t + B_{kt} \omega y_t \tag{2.6}$$

Where j are the socio demographic characteristics (x_j); B_{kt} are parameter estimates for the matrix of time-varying socio demographic characteristics (X_{kt}); and ω is the social network weight matrix. It assumes a static network.

Event history analysis is an important methodology for the analysis of diffusion of technologies since it takes into account the time-sensitive nature of this kind of data. Marsden and Podolni, for example, use this technique in their reanalysis of the physician data studied by Coleman et al and by Burt, which shows the crucial improvement in the methodology to better asses that particular empirical study.

2.4.6 Empirical Studies

Since the early work on Diffusion of Innovations (DOI) [22], the social structure has been acknowledged. Computational models of diffusion of innovations have evolved mainly in the space of system dynamics [16]. This branch of modeling has produced lots of research from

which new theoretical propositions have emerged, but in essence the System Dynamics tradition is independent of social network information.

Conducting empirical studies of the effect of social structure on DOI, however, remains quite a challenge. It requires important changes in research design and there is a notorious lack of published literature to systematically guide researchers interested in using the network approach[23]. To start, collection of data poses important limitations and trade offs. Network Data Questionnaires are usually long, and sometimes complex, and represent an enormous burden on the interviewer and the respondent. Even with recent advances with sensor networks to collect relational data [24] and while promising platforms for social sensing devices are being designed [25], important technical challenges are involved. Another important challenge is related with issues of confidentiality and human subject data protection. And even when those hurdles are overcome the greatest obstacle is ahead: most statistical tools assume independence of the observations ignoring the fundamental dependency that is inherent to social structures[26]. Recent statistical developments and algorithms are now available to estimate a limited but promising set of network parameters and they deal specifically with the dependence problem, but they are limited and still being developed [27].

Network Survey Design is also a key decision with important empirical and methodological consequences. Most of the work on the past decades on social network analysis is based on complete network data. This is equivalent to a census or saturation sample. On the other side of the spectrum is the analysis based on ego-networks where subjects are sampled and their "alters" studied assuming that there is no connection between the different sets of "egos" and their "alters", which is hardly the case in most communities,

and is of little or no interest in the study of diffusion process. In traditional research methods, a sample is the midway solution and it is expected to be a less expensive approach. Unfortunately, there is very little sampling theory when it comes to network structure [23].

One approach to this problem is to use partial information to learn the parameters that best describe the network of interest and use those parameters to generate through simulation a model of the complete network. This network then can be analyzed to learn the properties that best describes it and the consequences of such structural properties can be used to explore their implications in the flow of ideas and, eventually, other human dynamics. Thus, allowing improved understanding and management of diffusion processes. Unfortunately, the mathematics to produce specific type of networks according to a set of parameters are at best, in their early childhood stages [28].

2.5 Social Emergence

Relatively new tools like multi-agent models, may prove to be better suited for this task. Agents could be programmed to have beliefs (what it knows about the state of the world) and desires (goals). When the agent reasons about the state of the world and its goals and it makes a decision, its plan is called intentions. This particular kind of agents is called BDI (belief-desire-intention) by the artificial intelligence community[29]. Work developed in the last decade, with rather simple agents, has produced "interesting results about the emergence of social networks, trade, markets, cultural differentiation and evolution"[30, 31].

2.5.1 Agent-based generation of network data

The adoption of an idea, practice or artifact is heavily influenced by social context, through both conscious and unconscious mechanisms, but it becomes, at the end, an individual

decision. Nevertheless, how heterogeneous individuals behave with each other generate collective results that are not explained by the sum of the parts. There are different theories of how individuals engage in relations[32, 33] and some have been modeled mathematically[34, 35]. The analysis of the interplay of these different theories and how past experience and adaptation to the past experience causes mathematical analysis to be very limited in its ability to derive the dynamic consequences of the aggregate behavior. This is what makes agent-based modeling a practical method of analysis [31, 36, 37]. Surprisingly, the use of this tool to study the effect of social networks in diffusion is practically inexistent. We think that this method can be used to generate parameters useful to deal with sparse relational data problems that inevitably will emerge from the limitations on data collection and measurement we mentioned. We expect to extend concepts from the diffusion literature on economics, epidemics and herd behavior to study and model the diffusion of ideas, practices or artifacts[38, 39]. Agent-base models of human behavior that include network, context, and attribute features to produce predictive stochastic models of the diffusion process are rare and most lack validation against real data.

2.5.2 Exponential Random Graph Models

Exponential Random Graph Models are the most promising statistical models to represent social networks since they can represent structural tendencies, such as transitivity, that define complicated dependence patterns not easily modeled by more basic probability models. This model has also been referred to as a p^* model [40]. Recent developments in this area and MCMC algorithms have being developed which are able to produce Maximum Likelihood estimators. This is important because it would allow the use of simulation to evaluate the fitness of empirical observations.

Structures such as transivity and heterogeneity of degrees can be represented using ERGM using new specifications such as: geometrically weighted degree distributions, alternating k-triangles, and alternating independent two-paths.

These recent methodological advances come from scientific work on the study of HIV contagion. As opposed to other contagion processes, HIV studies need to focus on the dyadic relationship. The change on the focus of research, from the subject to a pair of subjects and the existence or non existence of a link, produces different and often counterintuitive results with important impacts on health policies. Diffusion of computers in rural areas of the world, among other environments, might have something in common with this.

Chapter 3: The Flow of Advice and the Diffusion of Innovation

3.1 Structural Perspectives on Diffusion of Innovations

Most empirical research on diffusion of innovations confirms the premise that new ideas and practices spread through interpersonal communications. However, most foundational studies have focused on the spread of relatively simple and "static" technologies, such as weed spray in Iowa [7], hybrid seed corn [41] or tetracycline [42], as opposed to ever evolving modern technologies and their myriad of versions and the potential difficulties and complexities intrinsic to them.

The key to transfer those simple technologies is awareness and imitation. In other words, P gets the idea through personal communication with O (awareness) and P decides to imitate O (adoption), later P passes the information to Q and so on. This approach leads to the interest in parameters such as the rate of diffusion and how it correlates with proximity, communication or influence. Valente et al. [43] studied and confirmed the association between friendship ties and the adoption of contraceptive choices in Cameroonian women. Their model defined network exposure as

$$E_i = \frac{\sum \omega_{ij} y_j}{\sum \omega_i} \tag{3.1}$$

Where, ω is the social network weight matrix and y is the vector of adoptions. The network exposure is measured on direct contacts. ω can be transformed to reflect other social influence processes through a family of relational, positional and centrality measures.

Their approach implies at least four different levels of decision to design a study of the network effect on diffusion:

1. The election of the type of network to observe and register. It could be a network of friendship, advice or any other convenient type.
2. If influence or other behavior determines P's probability of adoption, what set of structural features of networks capture such behaviors (relational, positional or centrality)?
3. Within each set, which measures should be used? (There are probably more than a dozen different types of centrality measures). And once the above decisions are made, still there is an issue of fine tuning to decide the weight attached to each factor, generally based on social distance. For example, if O influence P and P influences Q. Should the influence of P and Q reflect the fact that O may or may not be connected to a highly central or an isolated N?

ICT for development projects usually come in the form of computers for schools, community centers or other public or quasi-public spaces. In rural areas, probably more often, they come in the form of telecenters that embody a variety of different media that offers a wide range of potential solutions for community problems, all the way from telemedicine to e-commerce, as discussed in Chapter 2. In terms of ICT for development public policies, most discussions revolve around Internet access issues.

Those types of innovations are substantially different from the technologies mentioned above. They are knowledge intensive and for their adoption to be sustained over time there needs to be a continuous flow of information and support to keep up with the pace of new versions or even just to keep it functional. Voice over IP and wireless Internet solutions are frequently praised for their promising potential to serve isolated communities. But, updating to a newer version of hardware or software may cause operative systems to crash. In that

moment, what may seem a simple operation (update a driver for instance) can become a real problem. It may come from previous experience (knowledge), advice (another villager has the knowledge and the villager has direct or indirect access to him or her) or from specialized technical assistance, which depending on the type of source and the relative isolation, could be scarce and expensive to acquire. In this particular setting exposure to the friendship network is probably not enough. We will describe below an exercise where we gathered 30 members of a community because they were considered "influential" people that were important to "solve" the problems of the community. Among them we asked them who they considered influential and collected information about friendship and about advice. We found advice to be highly correlated with being influential.

We claim that the relevant unit of analysis for the diffusion of advanced or complex technologies is the community's social networks of advice, and that flow of ideas within these networks can be used to identify the influentials in order to better promote rapid diffusion of ideas. Rapid diffusion should be an objective to increase the collective knowledge base of the new technologies and surpass the critical threshold of adoption. In the next Chapter we will discuss an experiment in which work with the influentials was used to improve the adoption of technology.

3.2 Influential people and DOI

Most studies on innovation have been retrospective; they lack information on interpersonal communication networks, and more importantly, few have attempted to use the lessons from diffusion research to accelerate the diffusion of innovations [44]. Valente and Davis' work [45] suggests, through simulation, the possibility of achieving a critical mass in a much shorter time by carefully selecting the opinion leaders of a social network. This chapter

evaluates the hypothesis that a target is better defined as an influential member within multiple advice networks as one of the key contributions of this work. In general, identifying who are the influential members improves the design of diffusion strategies, regardless of what is being diffused through the network. In practice, the selection of influentials is usually accomplished by using conventional wisdom and traditional sociological theory, e.g. by looking for those with higher social and economic status and leaders of formal and informal organizations within the community. Selection is usually done after the definition of general criteria to select participants or "beneficiaries", ignoring the underlying network structure. In other words, many projects by design define a profile that usually tends to make the population of interest very homogeneous (e.g. programs designed to reach the poorest of the poor, or a specific gender within an income bracket) without consideration of the social network into which they are embedded and their links to the sources of economic and social capital.

We claim that the most influential problem solvers in the community should be the "entry" or starting points of the diffusion process as well and that networks of advice contain enough information to recognize them. For the community as a whole patterns of centrality should emerge even when network information does not include the totality of possible nodes and relationships. To decide what centrality measure to use and how to adjust any time of weight, we shall review what the mean in terms of the flow of advice.

3.3 Centrality and Advice

Since research on the idea of centrality applied to human communications was introduced in the late 40's by Bavelas at the Group Networks Laboratory at M.I.T., centrality has been related to reputations of power and influence over a community [46].

The most frequent form of organization of a social structure is the center-periphery pattern. It consists of a) a subgroup of relatively central prestigious actors who are connected by direct or short indirect ties and b) a subgroup of peripheral actors who are directly connected to the central actors rather than to other peripheral actors. In this form of organization, central actors tend to be resourceful and cohesively joined to other actors [47].

Within the family of centrality measures, there are four prominent ones due to their strong and distinct qualities[48]. They are also foundational in the field of social network analysis: degree, betweenness, closeness and eigenvector centralities.

3.3.1 Degree Centrality

The most simple and natural way of describing the concept of centrality is the star configuration. The center in this structure possesses 3 unique properties: it has the maximum degree [49-51]; it falls on the geodesics (shortest path[4] linking a given pair of points) between the largest possible number of other points and, since it is located at the minimum distance from all other points, it is maximally close to them (Freeman, 1978/79). Mathematically it is defined by equation (2).

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i p_k) \tag{3.2}$$

Where $a(p_i p_k)=1$ if and only if p_i and p_k are connected by a line, otherwise it is 0.

Its strength is that is intuitively simple, it's just a count. It is widely used in social network analysis. It assumes that all nodes are "equal" in the sense that if one node is

[4] A path is defined as a sequence of adjacent nodes in which no node is visited more than once

connected to two isolates and another is connected to two stars, both nodes will have a centrality of two. Depending on the process of diffusion that we are interested in, it can make a big difference. In aids contagion for example, its risk of contagion is obviously lower for someone having sex with the first node that is only connected with two isolates, than having sex with the second that because of the centrality of its partners has a higher risk of having the disease.

3.3.2 Betweeness Centrality

Betweeness [46] usually indicates a node that can control the flow of information bridging disparate regions of the network. Betweeness centrality is a more precise way of measuring a vertex centrality, in the sense that it takes into account that some low degree centrality vertexes may still be in a position indispensable for certain transactions[52]. According to Nooy et al. "targeting" the actors with the highest betweeness-centrality is a good strategy for launching an innovation[17]. They claim that if the first adopters are central and directly linked, their neighbors have higher exposure rates, so they are more likely to adopt. This is a quantitative approach, a qualitative approach may consider that by getting all the central nodes to adopt in a short period of time the system has been changed as the individual thresholds (perceptions of risk) of core and periphery actors has been changed, as a consequence of the social signaling and monitoring processes. For the diffusion of certain technologies the threshold may be lower also because of an increased value as a result of an increase in the number of adopters, especially communication technologies like mobiles, sms, email and the like[5].

[5] Some call these interactive technologies.

Because of its reliance on non-directed paths and geodesics, betweeness cannot be easily estimated for directed data [40], which is an important loss of information in cases where persuasion or advise play an important role in the diffusion of innovations. Its assumptions are that the traffic will choose the shortest path, and if confronted with equally short paths, it will randomly choose only one. Traffic moves one to one instead of copying itself or being broadcast from a node. A second assumption is that it is not diffusing randomly. Since it is taking only the shortest path, then it "knows" its target from the origin [53]. These last assumptions make Freeman's betweeness centrality measure unsuitable to be used in contexts where these assumptions do not hold, like the spread of computer viruses, diseases and other infections, or information movement in most cases. An innovation where imitation has an important effect may be less vulnerable to these strong assumptions than, for example, an innovation where persuasion is more important that imitation.

3.3.3 Closeness centrality

Closeness is the theoretic distance of a given node to all other nodes and it is commonly used in the study of diffusion. As opposed to degree centrality, this measure takes into account indirect connections. In a directed graph the outgoing arcs will be related to the amount of steps one actor needed to reach the other actors. In terms of flow it is ordinarily interpreted as an index of the expected time until arrival of something flowing within the network[54].

The critical assumption of this measure is that information is following the shortest path or parallel duplication –where all paths are followed simultaneously, including the shortest path as well. It only works on connected or strongly connected graphs.

This is a disadvantage because it means that you may not be able to use this measure to compare subgroups if there are regions less connected within the network and one or more

components belong to that area. In our study, the networks of advice found and registered are not well connected. This limitation impedes the use of this measure in the current analysis.

3.3.4 Eigenvector centrality

It is the property of a node that has a high eigenvector score and that is connected to others who are also high scorers. This is measured by the principal eigenvector of the adjacency matrix of a network. It was designed to work with valued data but works on binary information as well. The use of eigenvector centrality is convenient when the status of an actor is a function of the status of those with who he is in contact (Bonacich, 1972). Given an adjacency matrix A, the eigenvector centrality of node i is:

$$c_i = \alpha \sum A_{ij} c_j \tag{3.3}$$

Where α is a required parameter to give the equations a non-trivial solution ($\alpha = 1/\lambda$, i.e. the reciprocal of the eigenvalue) and has no substantive interpretation.

It is usually interpreted as a measure of influence. It assumes that traffic moves via unrestricted walks and does not assume that things flowing will be transferred or copied to one neighbor at a time, so this measure is ideal for influence type processes [54]. Diffusion of innovations where social influence is an important variable should obtain better results using this centrality measure.

3.3.5 Eigen decomposition of Advice Networks

Eigenvector centrality generalizes to what is known as Bonacich Power Centrality or Alpha Centrality [55]. Among the measures of centrality the eigenvector method seems appropriate in cases where persuasion depends on the perceived knowledge or authority of the vertex that

is the source of influence or advice. Because one should expect that receiving advice/information from someone who is more central should add more to one's centrality than being advised by an isolated member of the community, this is an attractive measurement of centrality. It is represented by the following equation:

$$c_i(\alpha,\beta) = \sum_{j=1}^{n} A_{ij}(\alpha + \beta c_j) \tag{3.4}$$

The value of α is used to normalize the measure and has no substantive interpretation. We use UCINET [56] to estimate Bonacich Power Centrality and in their solution the normalization parameter is automatically selected so that the sum of squares of the node centralities is the size of the network [57]. The parameter β is an attenuation factor which gives the amount of dependence of each node's centrality on the centralities of the nodes it is adjacent to. It can be interpreted as the degree to which an individual's status is a function of the statuses of those to whom he is connected. Beta is an adjustable weight that can take positive and negative values, depending on the specific phenomena under analysis. There are cases like bargain where the advantage comes from being connected to less powerful individuals. In communication networks Beta should be positive, as one benefit from the information available to one's alters. Bonacich [58] suggests that in a communication network, a low positive value of Beta would be appropriate if most communications were local and not transmitted beyond the dyad. Since the nature of personal advice implies information on specific personal concerns, it seems reasonable to expect that most interaction happens at the dyad level. But who advises the advisor should also be relevant. Thus, in the case of technical/business as well as personal advice it seems reasonable to choose the maximum value for β (note that if α and $\beta = 0$ then equation 4 is equal to equation 2). It was

done using β= very close to the absolute value of the reciprocal of the largest eigenvalue of both the adjacency matrices.

An important property of this measure is that it allows for negative values of β . In the case of advice it may have valuable applications since it is possible to think of negative values, for example, some actors may have conflict with a source of advice, potentially affecting the flow of information and ideas in the system. Therefore, even in the presence of a small β, conceptually eigenvector centrality is the adequate point centrality measure for advice networks. This parallels the "graph-theoretic concept of "vulnerability". This is not to be developed in this thesis. Cook et al. [59] among others have developed relevant work on the effect of negative edges in communication networks. Degenne and Forsé [52] offer a useful discussion about technical and sociological criteria to choose the value of β .

3.4 Empirical Analysis

We want to fit a model to predict which people are influential based on: conventional economic and demographic attributes, graphic-theoretical characteristics of the individuals and a measure that captures their use of advanced media.

We want to estimate the following logistic model to predict who is influential:

$$\log \frac{\Pr(y=1)}{(1-\Pr(y=1))} = \beta_0 + \beta_1 \mathrm{X} + \beta_2 C + \beta_3 M + \varepsilon \tag{3.5}$$

Where y equals 1 if the respondent is influential, X is a set of socio-economic and/or demographic characteristics, C is a set of sociometric measures based on the eigenvector centralities, M their use of Media and ε the expected error.

The propositions and tests in the form of a hypothesis are:

1. Sociometric measures are an important supplement to conventional social and economic status attainment measures in predicting who is influential.
2. Patterns of Advice received and given is a good predictor of who are the influential members of a community.
3. If Hypothesis 2 is true, there must be an important correlation with the early adoption of tools that are used to support and enhance communication, which leads to Hypothesis number 4.
4. If the use of media technology can be used as a predictor of influence, then a propensity to be an early adopter is correlated with patterns of advice and the use of media technology.

3.5 Data

3.5.1 Sampling Region

We explored these ideas using data collected in 2003 from a community of coffee growers in the southern mountains of Costa Rica called Santa Maria de Dota. The community has roughly 4300 inhabitants; coffee production and exports represent about 80% of their income.

Santa Maria de Dota is a well-established and integrated rural community. An interesting characteristic of the region is the structure of land ownership, mostly very small producers with 1 or 2 acres, with not much land available to grow their crops. In being so small, coordination and diffusion of information is key to production, processing, and commercialization of their coffee beans. Entrepreneurship is in high demand, as they cannot divide their land among their offspring, thus forcing them to generate their own jobs or to migrate.

The homogeneous social and economic characteristics of this population are expected to produce a relatively small effect from the social and demographic characteristics. Most producers are organized in a local cooperative called COOPEDOTA. It is collectively owned by the coffee growers registered as members of the cooperative. COOPEDOTA receives their coffee cherries and processes them into coffee beans. Whole beans and ground coffee is commercialized by the cooperative on behalf of the producers.

3.5.2 Baseline

In order to establish a "ground truth" or baseline it is necessary to establish who are the influential members that the model is expected to capture in a more effective and efficient way.

During the summer of 2003, a team of two senior Costa Rican researchers, trained in social sciences, former professors at the university of Costa Rica and currently members of an NGO ("CEMEDCO"), volunteered to conduct an Ethnographic Diagnostic in Dota [10, 60]. They were familiar with the general ideas of the social networks approach but not with its methods. They used the `snowball' methodology to establish who to interview next[61]. Their goal was to identify key members in the community. Key members were understood to be people that influence the community's decisions and whose opinions and decisions have the potential to affect the socioeconomic development of the community as a whole. People were asked about who plays important roles within the community when they are trying to solve problems, and what decision makers were known and respected by the people. They were also asked about their main concerns and what organizations were active in addressing those concerns, the different organizations they were involved with, and what studies about the community they knew of. After six visits to the community and dozens of interviews they

reported 53 influential members, among them 32 were registered as members in the local cooperative.

The list with the 53 names was discussed for validation with a group of "community experts", identified by CEMEDCO's researchers based on the knowledge about the community they gained. The expert's validation reduced the list to 30 members. They were expected to be the most influential members. Among those 30 influential people, 19 were registered producers.

The group of 30 was invited and attended to a workshop sponsored by INCAE (a top business school and research facility in Latin America), where they completed a sociometric survey. A roster with their names was presented to them, and they were asked to provide information on friendship, advice and influence. This produced dyadic data. We used Freeman's in-degree centrality as a scale of influence. Only those that were considered influential by their peers were considered the "truly" influential people or baseline. Only 19 had an in-degree measure bigger than zero and among them 16 were registered producers. Table 3.1 summarizes the three different exercises that lead to the baseline estimation we described.

Table 3.1 The three exercises used to construct the baseline for this study.

	Community Members	Subset of producers
Ethnographic Diagnostic	53	32
Community Experts Validation	30	19
Sociometric Survey	19	16

3.5.3 Data Collection

All active producers have to personally approach the mill office to collect either a check or an equivalent form of payment for their processed crop. Usually, most of them arrive during the first three to five days. The producers were interviewed as they approached the mill during the peak four days. By the end of the fourth day 84.72% of all payments had being collected according to the administration files. Their arrival seems to follow an apparent log-normal distribution. One hundred and twenty three surveys were collected through a short interview (see Figure 3.1).

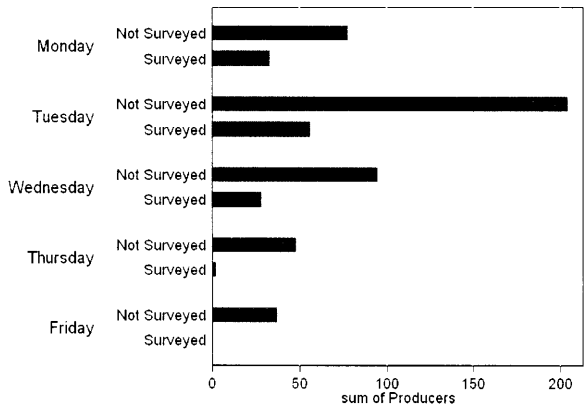


Figure 3.1 Distribution of Respondents as they approached the Coffee Mill to collect their payments during the week of data collection.

There were 123 respondents. (One was dropped because he was the son of a producer and his main occupation was not his family farm.) The respondents were asked to mention the names of the people who provide them with advice using the free recall method (no roster of names was shown) and they also were asked from the names they provided which were cooperative members. The names were validated later against with the cooperative membership records. There was no limit to the number of names that could be recorded. In

social network surveys it is especially problematic because each missing answer becomes a gap in the social network under study [62]. Advice is not a troublesome dimension of social relations, and in general, subjects where not uncomfortable answering questions about it. The interview was done in the premises of the cooperative, which probably legitimated the willingness of the cooperative management to support the study. Also, an interviewer filled out the questionnaire and the interview was kept as short as possible. All of these factors contributed to an unexpectedly high response rate for the overall questionnaire and a 100% response rate for the questions related to advice. We know that a sample is often not representative of a network because the structure of a random sample seldom matches the structure of the overall network. Therefore, we must be careful about generalizations about the social structure of the population, but accepting the limitations of our data set, we do believe that it is large enough to capture the main patterns of the flow of advice.

3.5.4 Graph-theoretic Data Sets

There are three generic social boundary specification strategies [63] : formal membership criteria based on the node's attributes; an event based approach and a relational approach based on social connectedness. In this paper we are using each of these methods to set the boundaries for three possible data sets. The overall criterion to select the interviewees was membership to the coffee cooperative. Those that actually had a chance to participate in the study were selected upon the event that they show up during the week of data collection. The open question (with no roster) on who you look to for advice... generated names of producers as well as names of other members in the community. The total list of names presented the possibility to define two different data sets based on a relational criterion mixed with an attribute criterion:

a) Those mentioned (connected) by the interviewees, who are registered producers and had been interviewed (n=122).

b) Those mentioned (connected) by the interviewees who are registered producers (n=169)

c) Those mentioned (connected) by the interviewees either registered producers or not (n=298).

In each case two n x n matrices were created. All of them are a one-mode matrix **A** where the (i, j) entry in the matrix is denoted by X_{ij} and represents the value of the tie from actor i to j. In this case this is a dichotomous relation where

$X_{ij} = 1$ if $i \rightarrow j$, $X_{ij} = 0$ otherwise.

We chose to use a) as the data set to work with. We can treat it like a whole network, since all the respondents sending nominations will have an equivalent likelihood of being nominated by his or her peers[6].

---

6 Dataset are available at www.media.mit.edu/~barahona/datasets

3.5.5 Attribute Data

Since recollection of sociometric data using a paper survey places a burden on the respondent and the interviewer attribute data has been collected from different secondary public or semi-public sources, as much as possible. The master database has 1296 records corresponding to community members. In the case of the producers non-sensitive data was provided by the cooperative, other sources as phone books and qualified informants have being used as well. Some data collection on attributes was done specifically for this study other than the surveys as is the case of neighborhood status that we will describe later.

3.5.6 Explanatory Variables

Lipset, cited by Blau and Duncan ([49-51]) says that "position in the social structure is usually associated with a certain level of income, education, family structure, community reputation and so forth". This paper tried to follow as much as possible Lipset's intuition to construct an equation that predicts a person's influence using the socioeconomic and demographic variables used in most theories about influence.

Income. The INCAE survey did not ask for income, nor did the cooperative have this data available. However, we had access to the amount of coffee beans they sold in 2003 to the cooperative to be processed. Since coffee is the main source of income for the vast majority it should be a good proxy for income and the records of coffee processed were reliable since no other company nearby was offering a better price than them, not to mention a legal obligation of exclusivity and a natural restriction associated with costs of transportation from farm to mill. Since we had access to the exact home address of every producer, we created a supplementary "social status" variable based on the local perception of the social status of the producer's neighborhood. A list of all neighborhoods was produced and

provided to a young local health professional, a local taxi driver, and to a business man who is in the construction business. They were asked independently to assign a value from 1 to 5, according to their perception, of the socio-economic status of each neighborhood. When there was no consensus, two votes decided the assigned status. There was no case in which all three answers were different.

Age, education and gender were provided by the cooperative. A dichotomous variable call "Mature" was created to capture this age range, from 35 to 70 years old, reflecting what a producer described as "the age when you and society know who you really are". Gender has no significant correlation with being influential which is not a surprise in this community7. Education data is consistent with this observation. When comparing the level of education of all male and female producers they share the same average amount of years ($\bar{x}$ =8.7 years, $p = 0.0332$).

3.5.7 Graph-theoretic variables

The correlation of power-centrality with the response variable is higher for the personal advice network than the one corresponding to the technical/business advice network. This difference across domains may suggest that the influential's advice is most sought after in interpersonal issues. This is consistent with the results of a study conducted by the Allensbach Institute on a German national sample (n=3843) reported by Weimann [19]. They found that in the financial and political domain the influentials had clear dominance, but

---

[7] Costa Rica is known in Latin America as a pioneer in women rights and as an international advocate of women and children rights (e.g. the country adopted the Law of Responsible Fatherhood, which gave women the legal right to name and receive support from fathers who did not recognize their children when born out of wedlock, leaving the alleged father to bear with a legal process and the use of genetic evidence to proof her

compared with these and 16 other domains in their study, the influentials advice is most sought after in "dealing with others" and "recreation".

TABLE 3.2 Description of selected variables in their original dimensions, some usual transformations and interaction variables that were tested

Variable	Mean	Standard Deviation	Min	Max	Partial Correlation	p-value
Influential	0.13	0.34	0	1		
Age in years	48.80	13.50	23	87	-0.0408	0.6820
Squared Age	2560.01	1463.30	529	7569	0.0536	0.5910
Mature (range 35-70)	0.80	0.40	0	1	0.2037	0.0390
Education in Years	8.70	4.30	3	18	0.1553	0.1170
Respondent has Secondary Education	0.14	0.35	0	1	-0.0080	0.9360
Gender (Male)	0.80	0.40	0	1	-0.0912	0.3590
Perceived socio-economic status of Neighborhood	3.31	0.63	2	4	-0.0520	0.6020
Volume of Coffee Crop	3112.30	3020.60	83.5	12681.3	-0.0897	0.3680
Log of the Volume of Coffee Crop	7.50	1.10	4.4	9.4	0.0840	0.3990
Freeman Indegree for the personal advise network	0.16	0.39	0	2	0.0921	0.3550
Freeman Indegree for the economic advise network	0.74	7.70	0	85	-0.0377	0.7050
Freeman Outdegree for the personal Advice Network	0.16	0.39	0	2	-0.0528	0.5960
Freeman Outdegree for the economic Advice Network	0.74	0.49	0	2	-0.1054	0.2890
Advice Centrality Index	0.23	0.54	0	2	0.2911	0.0030
Innovation (is early adopter of e-mail, fax and mobile)	0.36	0.63	0	3	0.2213	0.0250
Respondent has 3 or more channels of Communication	0.08	0.28	0	1	-0.0669	0.5020
Interaction of Mature x Years of Education	6.93	5.19	0	18	-0.1676	0.0910
Interaction of Mature x secondary education	0.10	0.30	0	1	0.0789	0.4280
Interaction of Mature x ACI	0.20	0.51	0	2	0.0642	0.5200
Interaction of ACI x Innov/comm	0.28	1.04	0	6	0.0118	0.9060

A paired correlation of the power centrality measures also shows this relationship (pair wise correlation=.43, $p<.001$). This is strong evidence that there is a correlation between both matrices of advice. When structural autocorrelation is present, Krackhart [64] recommends the use of Quadratic Assignment Procedure (QAP) to test the independence of the coefficients, since OLS can become severely biased under this condition. This is because the assumption of zero covariance between any two errors [65] is not met. Each person in a dyad will contribute to (N-1) dyads, and hence there is a high likelihood that the error that characterizes one dyad involving ego is similar to the error characterizing another dyad involving ego, or that the errors are "auto correlated". QAP attempts to solve this problem [66]. In this procedure the relation matrices are permuted to examine whether the results are

---

wrong in court). People in Los Santos are familiar with women as members or president of the board of

artifacts of the structure of the network rather than genuine relations among the actors. A hypothesis test using QAP effectively suggests the existence of a correlation between both advice matrices (Pearson Correlation= 0.062, p=0.005).

To avoid the problem of confounded variables we constructed a new variable ACI (Advice Centrality Index) to reflect the combined effect of both the personal and business/technical advice domains. We first dichotomized each power centrality variable using a 2.5 cut-off after inspecting the data (see Figures 3.2 and 3.3). Then the new variable resulted from summing up the "power advisors" of each network. Therefore the values for the new variable are 0 for non advisors, 1 for those who are power advisors in one of the networks and 2 for those powerful advisors in both networks.

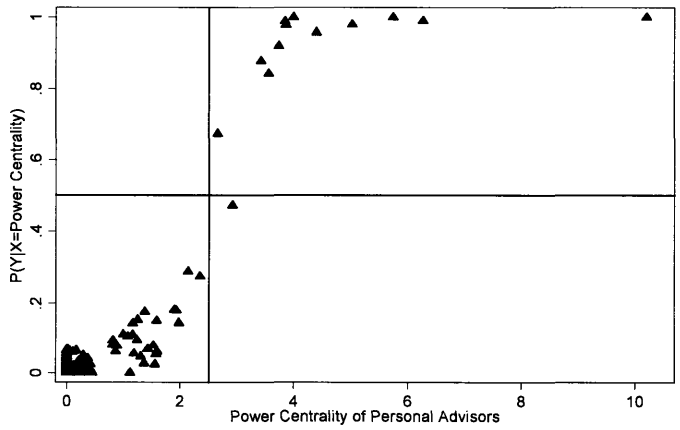


Figure 3.2 Probability of being an influential producer conditioned on Power Centrality of Personal Advice

businesses and civic organizations, driving a 4x4 taxi or running a mechanical workshop.

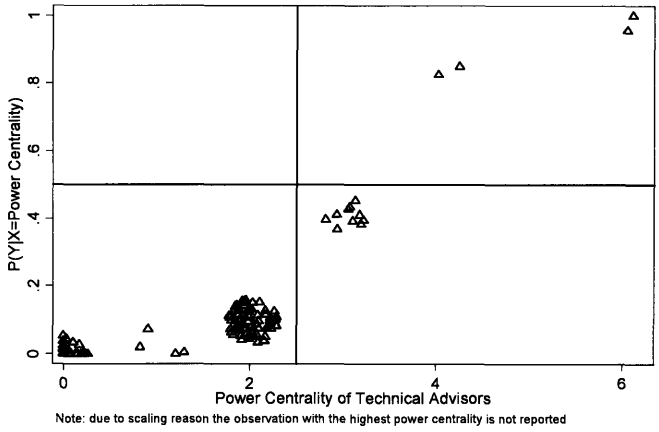


Figure 3.3 Probability of being an influential producer conditioned on Power Centrality of business/economic Advice

3.5.8 Media Technology and Innovation

Since media technology plays an important role in the flow of ideas, availability and use of communication tools should also play a role in the community member's capacity to influence. Research shows that the strength of ties between nodes is associated with multiple relationships and the use of more media to communicate [67]. In the process, communicators will reach a common understanding of the media and work together to a joint communications solution [68]. So, we explored the use of communication technologies in the community.

Most of the producers have access to a land phone, fax, mobile phone and email. As one might expect, having access to the latter is more difficult do to infrastructure limitations. So we created a simple ordinal variable called channel that adds up the number of channels a subject employs. By observing the data (mode 1, average 1.54, s. d. 0.85) we chose having 3

or more as the cutoff value (91.8% had 2 or less) to create a new binary variable to distinguish those having an exceptional number of communication channels.

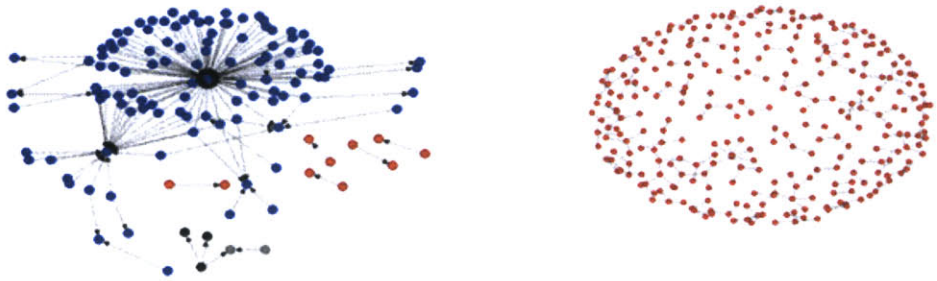
One fourth of the respondents had a computer at home but only 5.7% of all respondents used e-mail, and the correlation between having a computer and using e-mail was rather weak ($\chi^2 = 3.95$, $\rho = 0.047$). Thus, independently of having a computer at home or not, it seems fair to expect that the few using e-mail are early adopters. The second and third least popular channels were faxes and mobile phones (8% and 22%). To capture the propensity to be early adopters and the use of multiple channels for communication we used the presence of e-mail, fax and mobile phone as a proxy for the pattern of adoption of new communication channels. We called the variable Innovativeness.

3.6 Results

A preliminary exploration of their social network showed a very different structure when they were asked two different questions. When they were asked about people with whom they talk to in order to seek advise on different matters that are not related with production or commerce, their answers were limited to very few people, most of them close family and with almost non existent links to other producers.

When asked about who they seek for information or advice related with their business it was surprising to find that only two names were mentioned frequently. One name is mentioned by 69% of the producers the other is mentioned by 20%. Again, there were almost no links to the other producers. These hubs are sociometric stars. If communications are verbal and more than two-thirds report the same source of advice, how frequent could such interactions be?

These survey results suggest the hypothesis that there may be a poor flow of ideas through direct contact among the producers: almost everyone lives in semi-isolation, communicating almost exclusively among those of closest kinship. At the same time they show a few important hubs that hold the various social networks in place. Consequently, any attempt to use technology to strengthen this community must include these `hubs', and should not be perceived as a threat to their position. These initial results suggested the convenience of this kind of information when designing an intervention in a community.



(a) Production (b) Personal

Figure 3.4 Networks of Advice

All variables, transformations, and interactions presented in Table 3.2 were divided into three subsets: socio-economic conventional, sociometric, and Media/Innovation. Stepwise regression was used only for the subset of socioeconomic and demographic variables. Stepwise regression [69] was used first to discriminate among the subset of socio economic and demographic variables and then to compare our variables for centrality and for innovativeness, and to screen for possible interaction effects among the variables. No significant interaction effects were found. We run the hierarchical stepwise regression using ρ=0.25 in the forward steps and ρ=0.10 in the backward steps.

For this particular data set (n=122), we found only the variable "Mature" (being within the age range 35 to 70 years) being significant among the socio economic and demographic variables. This should be no surprise, as this is a particularly homogeneous group of people. The Alpha Centrality Index was used as our sociometric measure, as discussed above.

We then tested "Mature", Innovativeness and the Alpha Centrality Index against the null hypothesis of being simultaneously zero. We conducted a Wald test after running a logistic regression against the binary response variable (isInfluential). We obtained strong evidence to reject the hypothesis that the effects of these variables are simultaneously equal to zero ($\chi^2 = 11.30$, df =3, ρ=0.0102). Table 3.3 describes the equation of the logit regression model. The second, third and fourth columns present the results of running a logistic regression independently for each variable against the response variable. Column 4 is the full model.

TABLE 3.3 Logistic Regression Results for the components and the final model.

Variable	X (Age)	M (Innov.)	C (Alpha)	$X+M+C$ Full Model
LR chi2 (a)	2.61	29.65	48.07	58.40
D. of Freedom	1	1	1	3
Prob > chi2	0.1061	0.0000	0.0000	0.0000
Pseudo-R^2 (b)	0.0275	0.3127	0.5070	0.6160
Log likelihood	-46.0992	-32.5814	-23.3687	-18.2056

(a) The likelihood-ratio chi-square is defined as 2(L1 - L0), where L0 represents the log likelihood for the "constant-only" model and L1 is the log likelihood for the full model with constant and predictors.

(b) Technically, R2 cannot be computed the same way in logistic regression as it is in OLS regression. The pseudo-R2, in logistic regression, is defined as (1 - L1)/L0, where L0 represents the log likelihood for the "constant-only" model and L1 is the log likelihood for the full model with a constant and predictors. This statistic will equal zero if all coefficients are zero. It will come close to 1 if the model is very good.

Table 3.4 presents two nested models and the full model. Model 1 stands for the sociometric and demographic variables, in this case age, which was not significant by itself. Model 2 combines Innovativeness with Mature and was significant at 1%. The full model adds the centrality measure. For the combined model the strongest association is for the sociometric variable, and the weakest is age.

TABLE 3.4 Odds Ratios and p-values of the Main Effects Model

	Model (1)	Model (2)	Model (3)
Mature	4.157	5.008	14.916
	-0.179	-0.146	(0.072)*
Innovativeness		13.534	10.101
		(0.000) ***	(0.014) **
Alpha Centrality Index			35.586
			(0.000) ***
Observations	122	122	122
Pseudo R-squared	0.028	0.343	0.616
Log Lik Intercept Only	-47.405		
Log Lik Full Mod	-46.099	-31.121	-18.206
Likelihood Ratio LR		32.567	58.398

p values in parentheses

* Significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.5 describes the estimated unstandardized coefficients for the full model.

TABLE 3.5 Estimated Coefficients, Standard Errors, z-Scores, TwoTailed p-Values and 95% Confidence Intervals for the Final Logistic Regression Model (n=122)

	Coeff.	*Std.Err.*	*Z*	*P>|z|*	*95% Conf. Interval*	
Mature	2.7024	1.503	1.8	0.072	-0.2434	5.6484
Innovativeness	2.3127	0.9435	2.45	0.014	0.4634	4.1619
Advice Centrality Index	3.572	0.8932	4	0	1.8213	5.3227
cons	-7.157	1.8903	-3.79	0	-10.8619	-3.4522

The following histograms show how many "influential producers" are predicted by each component of the model. The histograms in Figures 3.5 and 3.6 show that even though a

majority of early adopters are not influential, amongst the influential the majority are innovators.

Among influentials and non-influentials, producers of mature age are the dominant ones, but the ratio of mature subjects that respects those outside of the 40-70 range is much higher for the influentials.

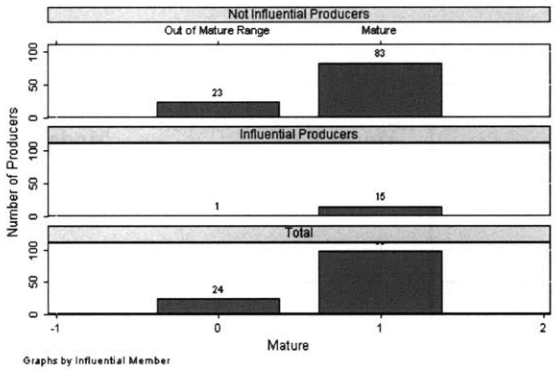


Figure 3.5 Distribution of Influential and non influential Producers According to Age (n=122).

Figure 3.7 shows the power of the advice centrality index to predict influentials. 100% of producers with ACI of 2 are influentials, 45% of those with an ACI of 1 are influentials, and only 4% of those with an ACI of 0 are influential.

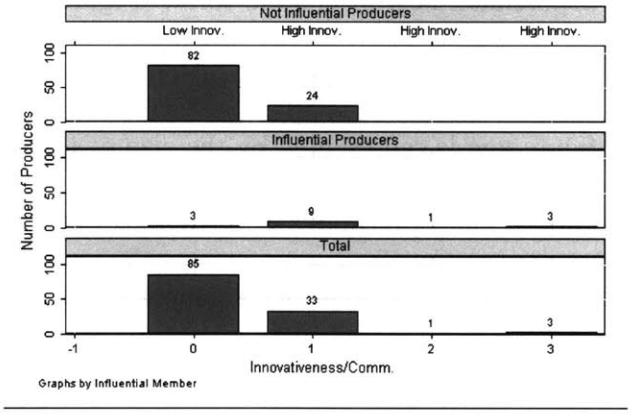


Figure 3.6. Distritution of producers according to innovativeness between influentials and non-influentials (n=122)

If all variables are held equal to zero, the probability of being influential is close to zero ($Pr(y|x) = 0.0015$) and someone meeting the three criteria has a probability of 0.9216 of being influential.

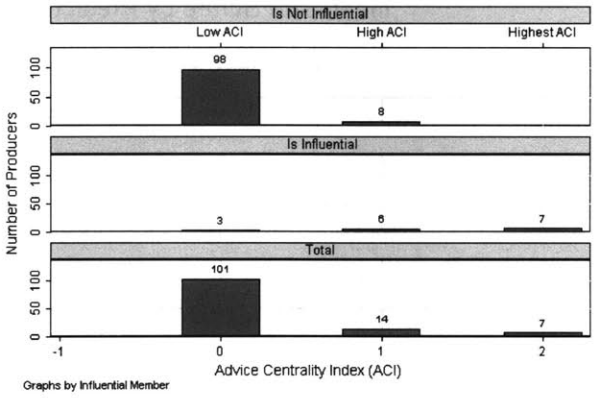


Figure 3.7 Distribution of producers according to Advice Centrality Index clearly discriminating influentials from non-influentials (n=122).

This can be appreciated best in graphic form. Figure 3.8 combines the three effects and presents the predicted values for the sample. Note how the expected "S-shape" curve is shaped. In the next figure age is represented by the size of the marker and a diamond shape means that the subject is an innovator in the terms defined in this paper. Although on the left hand side of the Figure there are mature aged people and innovators, they as a group are dominant within those with high alpha centrality.

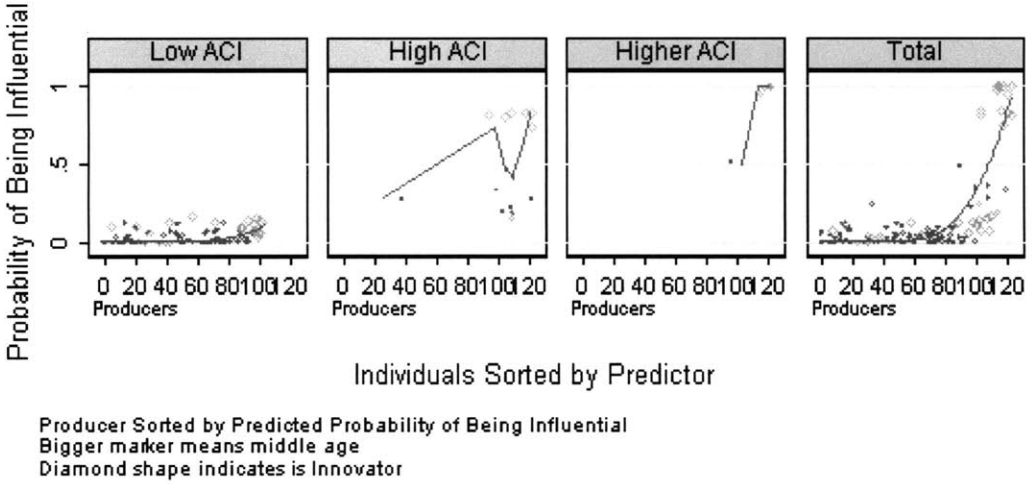


Figure 3.8 Associated probability of being an influential producers according to the combination of Advice Centrality Index, Age and pattern of adoption of new communication channels.

3.7 Discussion

This experiment suggests that patterns of advice captured by sociometric measures are a powerful predictor of influence. The model is effective for classification of who the influential producers are, according to the success and failures as a result of the model. In terms of accuracy (total correctly classified divided by total population), our classifier was 95.08% accurate and the ethnographic study has 85.42% accuracy. But accuracy is not the right metric, since it implies that all errors are equal [70]. We argue that in this context there are much higher costs associated with type I errors (false positive) than with type II (false negative).

In the context were it is desirable to tell apart who belongs to the group of influential members and who does not, with the purpose of working with them to foster an optimized diffusion process, both errors have very different consequences. For example, imagine someone gathers 11 influential members of the same community and none of the non influential members is present. They will recognize each other as influential and they will easily recognize what other influential people should be there, in case they are missing. It is so because core people tend to have a dense collection of relationships among themselves [71]. This structure has being recognized and documented in community influence systems [72]. Thus, missing a few will tend to be autocorrected by the knowledge and well established relationships of the core group. Now imagine the scenario were they are together, but share the room with other people that are not influential. It may be confusing to recognize what the group is about for them. The rules of engagement will somehow be different for the members of the two different groups and the effectiveness will suffer, in turn raising the organizational cost. To correct this, they or someone else would have to ask the "false

influential members" to leave, which would imply a social and emotional cost. To use a measure that is adequate to compare the conventional methodology and ours in these terms, let us introduce the corresponding confusion matrices in Table 3.6.

TABLE 3.6 Confusion Matrices

		Predicted by Model	
		Negative	Positive
Actual	Negative	105	1
	Positive	5	11

		Predicted by Conventional Methods	
		Negative	Positive
Actual	Negative	87	0
	Positive	19	16

While the conventional way of classifying the influential is extremely efficient with zero type II errors, it produces a false positive rate (type I) equal to 17.92%. These values for our model are 31.25% and 0.94% respectively. It is an important difference that is blurred by the accuracy measure. Instead, we should use the proportion of the predicted positive cases that were correct. This ratio is called in the machine learning literature the precision of the classifier, also known as the positive predicting value. In these terms our results suggest that we can get a 91.66% precision as opposed to 45% estimated for the ethnographic study.

Being an "established" member of the community and being an innovator plays a significant but much less important role. The findings are consistent with our intuition: influence follows the flow of advice and information. The ability to capture the dynamics of diffusion of ideas has the potential to have a very positive impact in the way ideas are

promoted and especially in the way that technology is deployed in underserved communities, by making interventions more effective and efficient by nurturing the flow of advice.

There are several different reasons to consider these results as useful and worthy of more testing. From an empirical point of view, it shows that sociometric information could have a significant role in helping identify influential members of a community, especially under conditions where the population of interest is highly homogeneous. Many settlements, housing projects, or communities are very homogeneous in their attribute values, giving more importance to relational sociometric measures.

The "advice centrality index" also has advantages in terms of efficiency. It is well known that traditional socioeconomic surveys have serious problems. Many people don't like to answer income or social status related questions. As a result data quality is poor and large survey samples are required. However, this research suggests that a light and neutral question like "Who do you look for when you need technical or business information" or "who do you look to for personal advice", can provide enough information to recognize the influential members of the group, those who are key for the diffusion of ideas and innovations. It is important to note that satisfactory results were obtained working with a partial data network.

Improved precision through the use of our proposed sociometric method can have a major effect, particularly with costly interventions. For example, the diffusion of technological innovations with a high learning curve, where almost personal support and follow up is needed for long periods of time, is difficult and expensive, but crucial to pass a certain threshold. It can also be effectively used as the first step to develop cognitive social structure studies [73].

Sociologists and more recently economists have devoted considerable attention to the impact of social structure and networks on the economy [74]. However these have been few attempts to translate this work into practical field methods. This work is one of the first of its kind.

There are still theoretical and empirical problems to solve before practical use of the abundant information about social networks can be used by communities. We foresee a role for machine-learning tools that can be used to develop stochastic models and methods to reconstruct whole networks out of partial and incomplete information. A future direction for this research is to test the model under conditions where the boundaries of the network are more diffuse and replication of the study with a different sample or different population will be sought. In the next chapter we present an application of these ideas to promote and study the diffusion of a technological innovation through the homes of the coffee producers described in this chapter.

Chapter 4: Network Effects in the Adoption of PCs

We have previously claimed that the social structure of advice provides enough data to derive quantitative estimates of a community's social dynamics. In Chapter 3 we use the eigenvector centrality of a community's network of advice to propose a method that allows the classification of members of a community as most likely influentials or not in an efficient way.

Since the 1930´s cognitive and social psychologists have worked on the problems of sociometry and group dynamics, and have developed methods to look at group structure and at the flow of information and ideas through groups. More recently Social Network analysis has become an important multidisciplinary field of knowledge and research. It has been used to better understand how ideas flow and how technology can help them spread more efficiently. But, what about the diffusion of ideas about technology itself?

We discussed earlier the role of technology in economic growth in general, and in particular, the critical role of information technologies in fostering the development of rural and underserved communities. It makes economic, social and ethical sense to improve the process of diffusion of technological innovations. In other words, it makes sense to accelerate the rate of adoption of ICT in such communities. In this chapter we describe an experiment that took place between 2003 and 2007, in the same community of coffee growers described

in section x.x. aimed to increase their rate of adoption of technology by selectively working with the influentials.

4.1 Marketplace of Ideas

The influential producers from the previously identified group were assumed to be the key agents to seed an alternative view of the use of IT within their own lives, families and business and the potential impact on the social and economic devolpment of SMD. This was done in hope that it would trigger a vast diffusion process throughout the whole community.

The intervention had an explicit yet somewhat controversial strategy. We chose to work exclusively with the influentials, regardless of social class, formal position, income level, gender or any of the conventional categories usually used by the international development community. We focused on the flow of ideas and named the intervention the "Marketplace of Ideas".

We used this method to explore ways to foster growth within the marketplace of ideas and reach most community members. Social network analysis provided a useful insight on the regularities in the relationships and the key structural properties of the community. This approach guided an intervention to enhance the abilities of key agents and to provide better or new links. These richer networks should become a vehicle for new and more significant content for the different actors within the community.

4.2 Project Activities

Data showed that this community's social network is characterized by a core-periphery structure and by design we chose to work with the most influential members of the core. Three relatively small activities were used to seed the diffusion of information and

communications technology in the town of Santa Maria. The claim here is that by selectively working with a few agents, we should be able to see the evolution of a more rapid diffusion process. As a baseline, we had collected information on who had home computers and Internet access in 2003. As a control group, we had the periodic statistics on home computer adoption at the regional and national level.

Figure 4.1 INCAE researcher at ICT workshop with influential community members. Here they are experiencing for the first time the use of e-mail.

4.2.1 Technology workshop with the influentials

Those considered influential were invited to a one day long seminar at INCAE Business School. During the morning they were exposed to some basic ideas on technology and development and during the afternoon they participated in a hands-on workshop to learn how to set up an email account and to start using it among them. No follow-up or extra training was provided as a group. Because they are members of a well interconnected core, we expected them to support each other, as it effectively happened. Some of the most influential have kept an advisory/collaborative relationship with Media Lab, specially in 2003 and 2004

but most of 2004 and 2005 we kept our contacts to a minimum and only through few "influentials" and mainly over the long distance conversations.

4.2.2 Context improvement

Ten used desktop computers from the Media Lab were collected and were shipped to Costa Rica. They were Pentium II and Pentium III and were set up to run on Windows. One of the most influential members used resources of his organization to bring them from the Atlantic port to the community. Five were kept for a project described below. The other five were used for computer training in a setup organized by one of the most influential members and run by the local women's association (whose president was also one of the influentials). This gave rise to a local computer school that is still in place and has extended its services to an Internet Café. This case as those that will be discussed in the next chapter, fit the idea that in "the last mile" diffusion is more effective in the hands of people with enough social capital and influence. In this particular case, they manage to start training themselves, which trigger a local business opportunity and a local demand for home computers. As we will discuss later, this demand created the opportunity for another business to emerge, importing and selling used computers, which later evolved into a computer shop that also provides maintenance and repairing services. In an organic way, new organizations appeared and enriched the context.

Another avenue of enrichment was through "weak ties"[74]. Some key members developed a collaborative relationship with new actors who in different ways provided initial technical support. There were three main organizations that provided some support. INCAE Business School provided the time of an engineering student who visited SMD two or three

times a month during a three months period to meet with some of the influentials. He also answered their phone calls and, later, their emails. He was a sort of a "digital weaver", providing information and pointers to sources of resources and creating new links that influenced and supported the influentials. CEMEDCO, the NGO we described in the previous chapter. And the Costa Rican Foundation for Sustainable Development, the local promoters of Lincos provided technical assistance to the new training school. Their relationship evolved through time and the SMD computer school has been described by some as the "Lincos without the container".

Figure 4.2 Women's association computer training school started off with five used computers and today is a self sustained business.

4.2.3 Community Help Desk and Communications Hub

This project was intended to promote a communications platform for the community. It had a short life. It is briefly reported because during a short period of time it became a source of ideas and impacted "the influentials" as they gained national notoriety when the project was awarded among the most innovative projects of 2004 by the local financial press[75].

We promoted the idea of a LINCOS container that would work as a helpdesk, trying to scale up the "digital weaver" concept. It was supposed to work also as a communications hub for a town wide wireless network and an asynchronous e-mail and phone service. Both the container and the asynchronous e-mail and phone service collapsed shortly. On the other hand the wireless internet around the cooperative is still working.

4.3 Measuring the network outcome in the diffusion process

We did not visit the community during 2005 or 2006. We went back during January and February of 2007 to evaluate the results of our proposed intervention strategy. We looked at the adoption of home computers as a proxy for the mindset change that helped the decision of fully integrating ICT in their families' lives.

We wanted to know if the strategy was effective and if that was the case, if it can be attributed to a network effect. In question format:

1. Is there a significant increase on home computer adoption?
2. Is there a network effect in the process of adoption?
3. Can we relate this to the influence within the network of advice?

4.4 Data

Data was collected during January and February of 2007 in Costa Rica. Most producers of Coopedota live in Dota. Table 4.1 describes the distribution of the cooperative members in the different neighborhoods and the number of people interviewed in each of them.

Table 4.1 Distribution of Registered Coffee Producers and sample used by Neighborhood

Neighborhood	Sample	Total	Neighborhood	sample	Total
Barrio Las Flores	1	1	El INVU	5	7
Barrio Las Nubes	10	8	El Llano	16	34
Calle Higueronal	58	107	El Rodeo	1	6
Calle Hogar de Ancianos	1	1	El Vapor	1	1
Calle IMAS	1	1	Guadalupe	2	17
Calle San Rafael	46	108	Las Arias	1	1
Calle a Copey	38	102	Providencia	29	37
Calle a San Marcos	7	7	Rincon de Los Solis	4	4
Calle al Cidral	3	8	San Lorenzo	4	14
Calle del Albergue	6	6	San Marcos	3	3
Carrizal	1	1	San Pedro	5	20
Copey Centro	3	3	San Rafael	1	1
Dota Centro	47	101	Tibas	2	2
El Guayabal	7	8	Urb. J.Ma.Ureña	1	1
El IMAS	9	29	Urb. Las Tres Marias	2	2
Total				**315**	**641**

4.4.1 Data Collection

Coopedota's role in data collection was crucial. As a cooperative, they are used to collect and manage information from the producers and the producers are used to provide information to the Coopedota's management. They sent a letter to all producers informing them about the data collection process and asking them to support it. In regard to materials and visual identification, they printed all forms with their logo and provided the team of interviewers

with the company's t-shirts. The producers' trust on the cooperative mostly has had a positive influence on the high response rate.

4.4.2 Survey

Structured interviews were conducted using a paper survey. This instrument was filled by the interviewer in the presence of the interviewee as they talked. Data capturing was made manually using a local clerk and Microsoft Access Forms to condition different data entries to the expected value formats. This provision helped to maintain a relatively low entry error rate. When human data entry is used, 100% rekeying and verification is a common feature. We did not use a similar test. In the US, studies have showed that the error rate in this form of data entry tends to be less than 1% of the total fields[76]. However, when obvious mistakes were found all corrections have being documented and the original digital files were kept intact.

Paper based survey and human data entry tends to be regarded as costly and avoided in the US, but that was not the case in Dota. The characteristics of the population, the favorable conditions to hire local help and the relatively small size of the target population allowed a cost-effective manual process and made seeking other types of digitally supported data impractical.

To answer the questions in this chapter the survey could have being significantly shorter. We chose to collect more data that may be used in further exploration of the project's impact or be useful for some of the possible research directions that will be discussed in the future work section of the concluding remarks. Table 4.2 describes all the data gathered in this survey and some secondary sources of additional information.

4.4.3 Study Design

A time framework was set to collect the data. The goal was to interview all Coopedota's registered producers if possible. Otherwise, we should try to collect as much information as possible.

A random sample would have been smaller and more effective for socio-demographic data, but in terms of sociometric or network data, the more we can get, the higher the probability of capturing the dominant features of the social structure.

Members of a small team of Costa Rican Interviewers were assigned to different routes that covered all the relevant neighborhoods. The collection strategy was to start in each route with those that were relatively harder to reach, which were very few, given the region's good infrastructure and the small size of farms. The time to visit the farmers was scheduled according to the dynamics of field labor in order to maximize the likelihood of finding the producers in their farm and with available time for the interview. When they were found busy but willing to participate, an appointment was set to visit them at a different time either at their farm on a following date or at night in their homes.

Table 4.1. Personal, Household and Network Data Captured by Paper Survey and Secondary sources

Personal Information	
FETY	Fertilizer type
FEBR	Fertilizer brand
hasPC	Personal Computer at Home
PCAd	Place of acquisition
FirstPC	Is her first PC?
PCUses	What is the computer used for
Birth	Date of Birth
Cedula	National Unique ID
Phone	Telephone Number
Educ	Education Level
HHH	Name of HouseHold Head
reltoHHH	Relationship to HHH
Household Information	
HHH	Household Code
node	Coded name of Household Member
for household members:	
AGE	Age
Edyrs	Education
isAsoc	If is a registered member of Coopedota
reltoHHH	Relation with HHH
usesPC	member uses PC?
Pcuses	How is PC used?
Neigh	Neighborhood
Address	Household Adress
Internet	has Internet?
IntCafe	uses Café Internet?
CompGame	has Computer Games
yrCGames	First Year with Comp. Games
hasEMail	member has email
useofEMail	what email is used for
SurveyInfo	
Start	Initial time
End	Ending time
Iname	Interviewer Name
Producer to Producer Conversational Network	
alter	Name of Alters
freq	Frequency of Contact
hasPC	Alters' ownership of computer according to subject
PC Advice Network	
alter	Name of Alters
freq	Frequency of Contact
hasPC	Alters' ownership of computer according to subject
Personal Advice Network	
alter	Name of Alters
freq	Frequency of Contact
hasPC	Alters' ownership of computer according to subject
Economic Advice Network	
alter	Name of Alters
freq	Frequency of Contact
hasPC	Alters' ownership of computer according to subject
Reported relationships with Core Community Members	
alter	Aquaintance
alter	Frienship
alter	Economic Advice
alter	Personal Advice
alter	Frequency of Conversation
Secondary Sources	
Nstatus	Socioeconomic Status of Neighborhood
KW	KW consumption
Crop	Crop Size
HHHLoc	Household Location

Given that this is a very homogeneous and stable community and the big size of our sample (around half of the population), we are comfortable assuming that it is representative and has a normal functional form, for all estimation purposes of the socioeconomic or demographic data. It is common to find more than one producer living in the same household, since both husband and wife could be producers and even some of the older

children. One way of looking at the dataset without mistaking producers that live in the same household is by using the subset of household heads. Table 4.3 describes this sub-sample.

Table 4.3. Household Heads: Sample Composition

Demographic Characteristics of study sample		
Demographic Characteristics	Number of cases	Percentage of sample
Gender		
Male	186	78.48
Female	51	21.52
Total	237	100
Age group (years)		
18-39	25	10.5
40-64	148	62.18
65+	65	27.31
Total	238	100
Education		
<High School	184	81.42
=High School	16	7.08
>High School	26	11.5
Total	226	100

Network data considerations are to be referred to the previous discussion in Chapter 3 about the 2003 data. Nevertheless, we observed this time a much bigger proportion of the total network and collected advice information on 1328 dyads formed by 679 subjects. These data was used as a complete network to extract the eigenvector centrality of each subject.

4.4.4 Missing Data and Imputations

Most surveys are complete. Once in a while a question was missed. More often, data is missing because of omissions when the paper surveys were processed into the database. Whenever possible, missing data was imputed by going back to the paper instruments or by follow up phone calls. Very few outliers were found and they were tracked back to data processing mistakes, which were corrected. All corrections and imputations are recorded as part of the STATA scripts written to manage and analyze these data. For all estimation purposes we followed a procedure based on completely recorded units[77].

We built a rich data set with at least some information on over 2300 people.

4.5 Methods

Torsten Hägerstrand, in his dissertation *Innovationsförloppet ur korologisk* synpunk, cited by Skillnäs [78], presented a model of diffusion for simulation that took into account the spatial neighborhood. He concluded that the frequency of contact decreases with the distance from the source of information. Likewise, a frequent contact would be associated with nearness to the source of information. Nyblon et. Al [79] extend the definition of neighbors to a social system, space or some combination. They explain the general approach as follows:

> "If adoption is a function of diffusion among neighbors, ...then pairs of adopters should be found more frequently among neighbors than we would expect by chance. In contrast, if adoption is principally a function of independent decisions based on, say, economic factors (e.g. government subsidies) or individual dispositions, then we would expect adopters to be uniformly randomly distributed with respect to being neighbors".

We created a connection matrix with the answer to the questions about "producers you look for advice" and the variable "is an adopter" is define by the answer to the question about having or not having a computer at home. So,

N is the number of producers 1, 2, 3, ..., N

Each producer is classified into one of two classes. This denoted by the categorical variable x_i, where $x_i = 0$ if i is a non adopter, $x_i = 1$ otherwise. Each pair is either connected or disconnected. The connection is directional and denoted by $\delta_{ij} = 1$ if $i \rightarrow j$,

$\delta_{ij} = 0$ otherwise. The connection matrix is then $\Delta = (\delta_{ij})$. As a consequence, Δ determines the "advice" neighborhood of all producers in the sample.

The neighborhood of producer i consists of all individuals j with $\delta_{ij} = 1$. Since it is a directed graph, we count the incoming and outgoing edges separately. The in-degree of j is $q_j = \sum_i \delta_{ij}$ and the out-degree is $k_i = \sum_j \delta_{ij}$. In this experiment, k_i is the number of people producer i reports as an advisor. Likewise, q_j is the number of times j was reported as being someone's advisor. If the relationship that defines the social space or "social neighborhood" is advice and i cites j as its advisor then k_i is the number of advisors the farmer has. We assume that frequency of contact is a measure of influence, then k_i is the measure of how much i is influenced by its neighbors, while q_j measures how influential j is.

In matrix terms, the column vector $\mathbf{k}$ with components k_i (out-degrees) results from multiplying Δ by a vector of ones $\mathbf{k} = \Delta\mathbf{1}$. Likewise, $\mathbf{q} = \Delta'\mathbf{1}$. Since $\sum k_i = \sum q_j = \sum \sum \delta_{ij}$ is the total number of edges, that can be denoted as S, the $S = \mathbf{k'1} = \mathbf{q'1} = \mathbf{1'\Delta 1}$.

We want to know whether the adopters are clustered in a way that they are more likely to be connected with one another than what might happen randomly. Following Niblom et al.[79] the statistic for measuring network association is

$$Q_{11} = \sum i = 1^N \sum j = 1^N \delta_{ij} x_i x_j = \mathbf{x'\Delta x} \qquad (4.1)$$

Q_{11} gives the number of pairs $i \to j$ in which both are adopters. If $i \to j$ and $j \to i$, then this pair is counted twice. The association between non adopters is given by

$$Q_{00} = \sum i = 1^{N} \sum j = 1^{N} \delta_{ij}(1 - x_i)(1 - x_j) = \mathbf{(1 - x)'\Delta(1 - x)} \tag{4.2}$$

The remain two associations are $Q_{01} = \mathbf{(1 - x)'\Delta x}$ and $Q_{10} = \mathbf{x'\Delta(1 - x)}$, in terms of Q_{11}

$$Q_{10} = k'x - Q_{11}$$

$$Q_{01} = q'x - Q_{11}$$

$$Q_{00} = \mathbf{1'\Delta 1} - k' - q'x + Q_{11}$$

This allows us to produce an association table with the following values

Table 4.4 Association Table

			Total
	Q_{00}	Q_{01}	**k'(1 - x)**
	Q_{10}	Q_{11}	**k'x**
Total	**q'(1 - x)**	**q'x**	S

Note that a high Q_{11} indicates high association between adopters. Note that the increase in Q_{11} can be compensated by an increase in **q'x** or **k'x**. Therefore, a high Q_{11} does not imply a high Q_{00}. Also, note that **k'x** and **q'x** are large when the adopters are more likely to be found among those having large numbers of neighbors (large **k'x**) or those belonging to many neighborhoods (large **q'x**).

To measure the influence on ego from alters decision on adopting we used the normalized eigencentrality[80] of the advice connection matrix Δ, in order to produce a weighted measure of influence. This weight was applied to the adoption vector $\mathbf{x}$ to calculate the a weighted measure of indegrees for each subject. This construct represents the amount of influence to adopt each individual gets from his advisors taken into account the influence the advisors get from their own advisors. In matrix terms, $\mathbf{I} = \Delta' * \mathbf{x} * \mathbf{c}'$.

4.6 Results

As expected the advice network has a core-periphery structure which is evident in Figure 4.3. There are two components in this graph, the small one is most likely a result of sampling and not necessarily means that these people are disconnected from the rest of the network of advice. It is important to observe that most of the people are linked to the main component. The size of the node is proportional to the eigencentrality of each actor in this network.

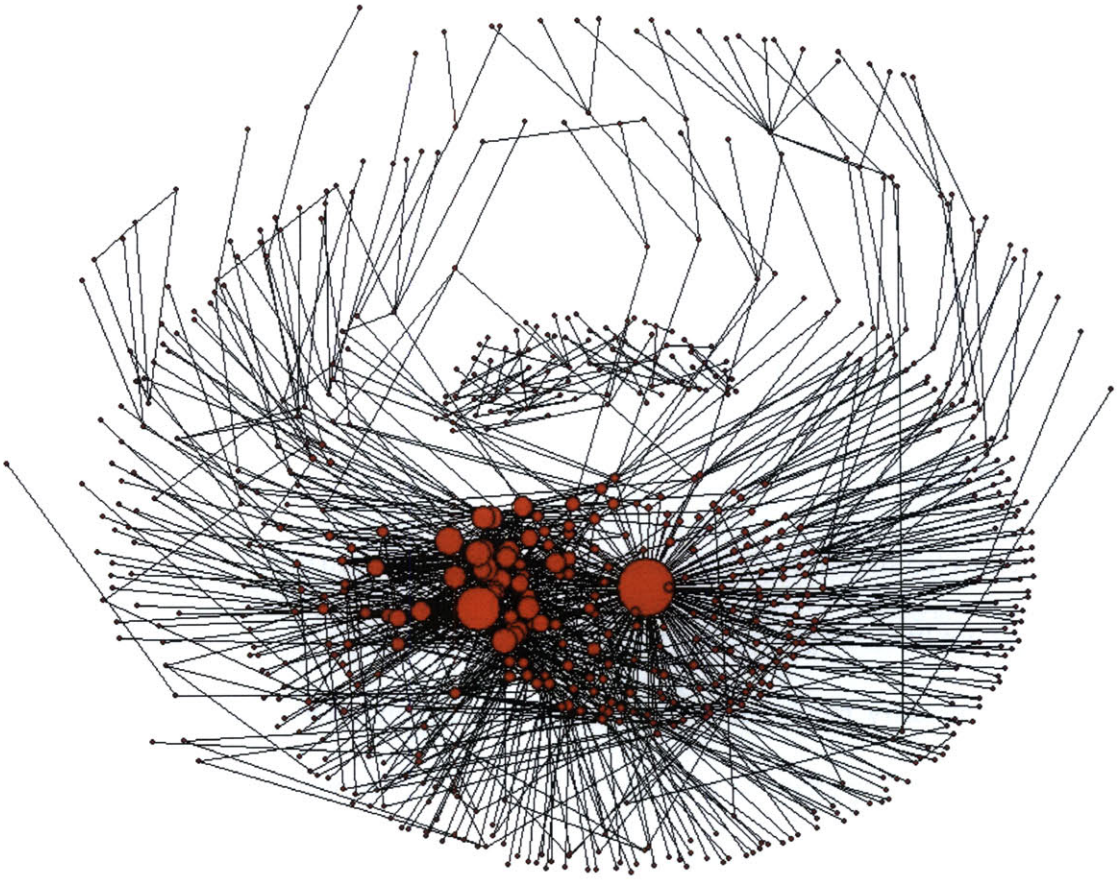


Figure 4.3 **Network of Advice. The size of the node is proportional to the eigencentrality of each actor.**

Costa Rica's political division is organized into provinces. Each province is divided into "Cantones" which, in turn, are divided into districts. For planning purposes the Costa Rican government has arranged the country's cantones into seven planning regions [81]. So, one direct way to control the experiment is to compare Dota's PC adoption rate with those of the country, Dota's region and its neighboring Central Pacific region.

According to this division, Dota belongs formally to the Central Region and represents its most southern frontier (See Figure 4.4), even though a mountain range clearly separates this rural region from the rest and more urban "cantons" of the Central Region. In many

senses, they are geographically, economically and socially closer to the Central Pacific Region than to the rest of the Central Region.

The National Institute of Statistics prepares a yearly home survey called "Encuesta de Hogares". Since 2002, they have collected information about the ownership of different appliances and electronic equipments in Costa Ricans' homes. This provides a baseline to compare the adoption rate of home computers among Coopedota's registered producers with the national adoption rate, its own region and the neighboring Central Pacific Region. We can compare these data with the two data points corresponding to the cross-section analysis of 2003 and 2007. Results are shown in Figure 4.5.

Figura 4.4 **Socioeconomic Regions of Costa Rica. The yellow star shows Dota's geographic position within Central, and bordering with Central Pacific Region.**

In 2003, Coopedota's registered coffee producers' rate of adoption of home computers was 16.2%, well bellow the national and Central Region levels of 22.2% and 28.7%, respectively. Their adoption rate was closer to its neighbors' rate in the Pacific Central Region (11.1%). Providing that all government initiatives to promote the use of technology are not substantially different in the Los Santos Region compared with the rest of the country, and given that Marketplace of Ideas as a project was the only new source of influence, the data shows an outstanding level of adoption for this group of producers at 43.5%, almost doubling the 2003 rate and surpassing the national and regional levels. (See Figure 4.4).

This provides a clear answer to the first question of this chapter, now we need to study if there is a network effect in this process.

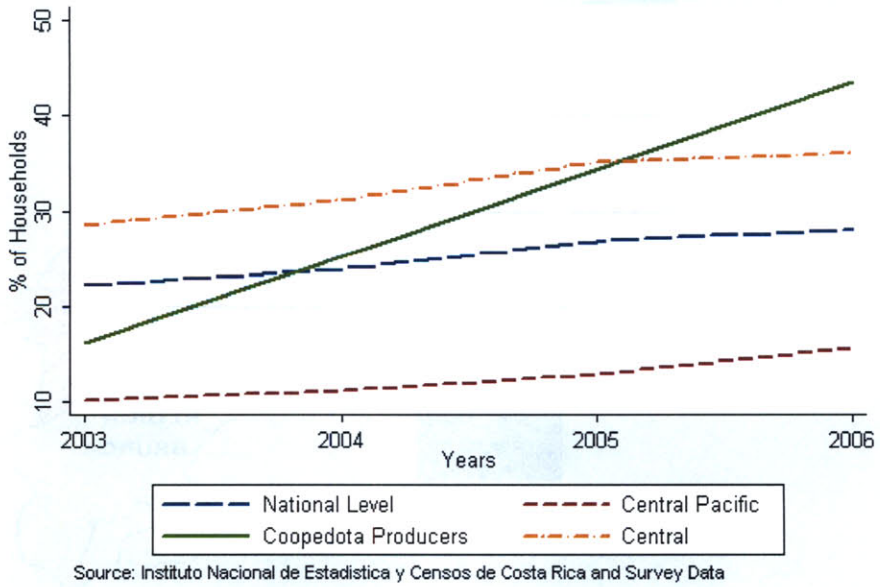


Figura 4.5 Home Computer Adoption at the country, regional and Coopedota producers level.

We should check if the distribution of those adopters is random or if there is a network bonding the adopters. We had 1328 dyads recorded in the Advice Network. Of those, 549 (Q_{11}) pairs of adopters of home PCs were observed (see Association Table bellow). The expected number of adopter pairs is 330.53 with standard deviation 62.77. Given a large expected value, we can safely use the normal approximation, leading to a $z = 3.48$. Therefore, the Network Effect is statistically significant at all reasonable levels.

Table 4.5 Association Table of the Advice Network

	0	*1*	Total
0	238	97	335
1	444	549	993
Total	682	646	1328

We have found a network effect, but it could disappear when adjusted for possible confounding variables like gender, age or education level. (See Table 4.6)

Table 4.6 Household Heads: with and without home computer

	stats	age	edYrs	Gender
No Home Computer	*n*	135	126	134
	mean	59.2444	6.12698	0.791045
	p50	60	6	1
Home Computer	*N*	99	96	99
	mean	51.8687	8.34375	0.767677
	p50	50	6	1
Total	*N*	234	222	233
	mean	56.1239	7.08559	0.781116
	p50	55	6	1

There are no gender differences between the group of household heads with and without computers ($t = -0.42$, $\Pr(|T| > |t|) = 0.6714$, cl=.99). Differences were found in age and years of education. In average, those with a home computer were over seven years

younger ($t = 4.2722$, $\Pr(|T| > |t|) = 0.0000$, cl=.99). On average, those with home computers had two more years of education ($t = -5.2568$, $\Pr(|T| > |t|) = 0.0000$, cl=.99).

To study the advice network effect we looked at each node's indegree weighted by the normalized eigenvector centrality of each of its "neighbors" (qYnw). Figure 4.6 shows the neighborhood influence as a predictor of who has adopted a home computer. The construct works as expected. To adjust for education and age we used the following model:

$$\log \frac{\Pr(y=1)}{(1-\Pr(y=1))} = \beta_0 + \beta_1 wInfluence + \beta_2 edYrs + \beta_3 age + \varepsilon \tag{4.5}$$

Where y equals 1 if the respondent has a home PC, qYnw is the neighborhood influence, edYrs is a continuous variable representing education, age is age treated as continuous and ε is the expected error.

Table 4.7 Independently Estimated Coefficients, Standard Errors, z-Scores, Two-Tailed p-Values and 95% Confidence Intervals with Dependent Variable = hasPC

Variable	O.R.	Std. Err.	z	P>|z|	95% CI	
wInfluence	1.05212	0.0114	4.68	0.000	1.0300	1.0747
edYrs	1.20214	0.0366	6.05	0.000	1.1325	1.2761
age	0.97089	0.0059	-4.87	0.000	0.9594	0.9825

Independently all variables significantly related with the adoption of computers as shown in Table 4.7.

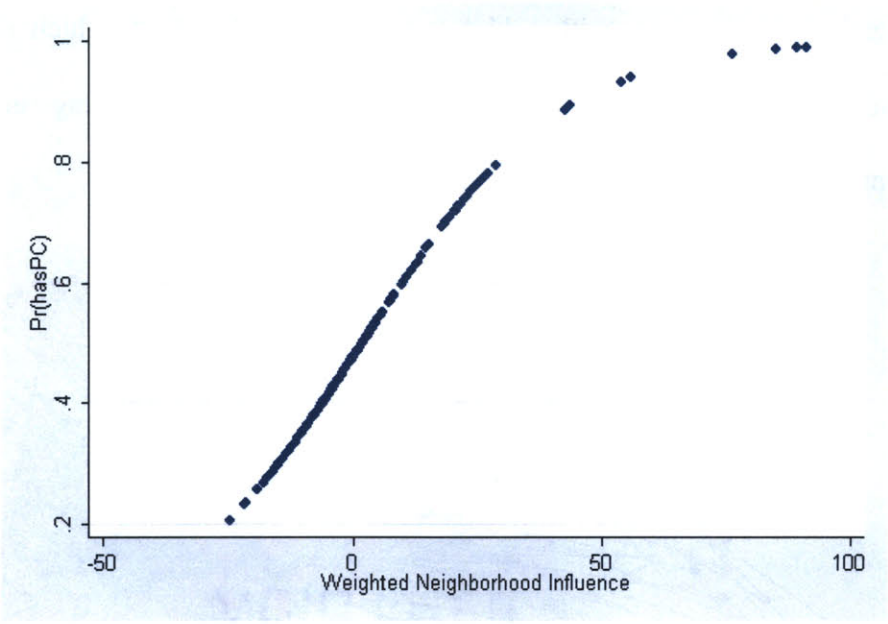


Figura 4.6 Predicted value of hasPC conditioned on the Weighted Neighborhood Influence (wInfluence)

Adjusting for age, education and interactions, the advice network effect remains significant. Controlling for age and education, as the neighborhood influence level increases the odds of adopting increase by 19%. Note that conditioned by the interactions, age and education are no longer significant. (See Table 4.8).

Table 4.8 Estimated Coefficients, Standard Errors, z-Scores, Two-Tailed p-Values and 95% Confidence Intervals with Dependent Variable = hasPC

Variable	O.R.	Std. Err.	z	P>|z|	95% CI	
qYnw1	1.1937	0.0661	3.20	0.001	1.0710	1.3305
edYrs	1.1036	0.1082	1.01	0.315	0.9106	1.3374
age	0.9728	0.0148	-1.81	0.070	0.9442	1.0022
lqxe	0.9935	0.0029	-2.23	0.026	0.9878	0.9992
lqxa	0.9988	0.0009	-1.40	0.160	0.9971	1.0005
lexa	1.0010	0.0020	0.49	0.626	0.9971	1.0048

Figure 4.7 shows in blue those nodes that adopted a home computer. The size of the node is proportional to its eigencentrality. The image reflects how adoption seems to be

propagating among those with a higher advice centrality first, which is consistent with a previously observed pattern in research about influentials being influenced by other influentials [19].

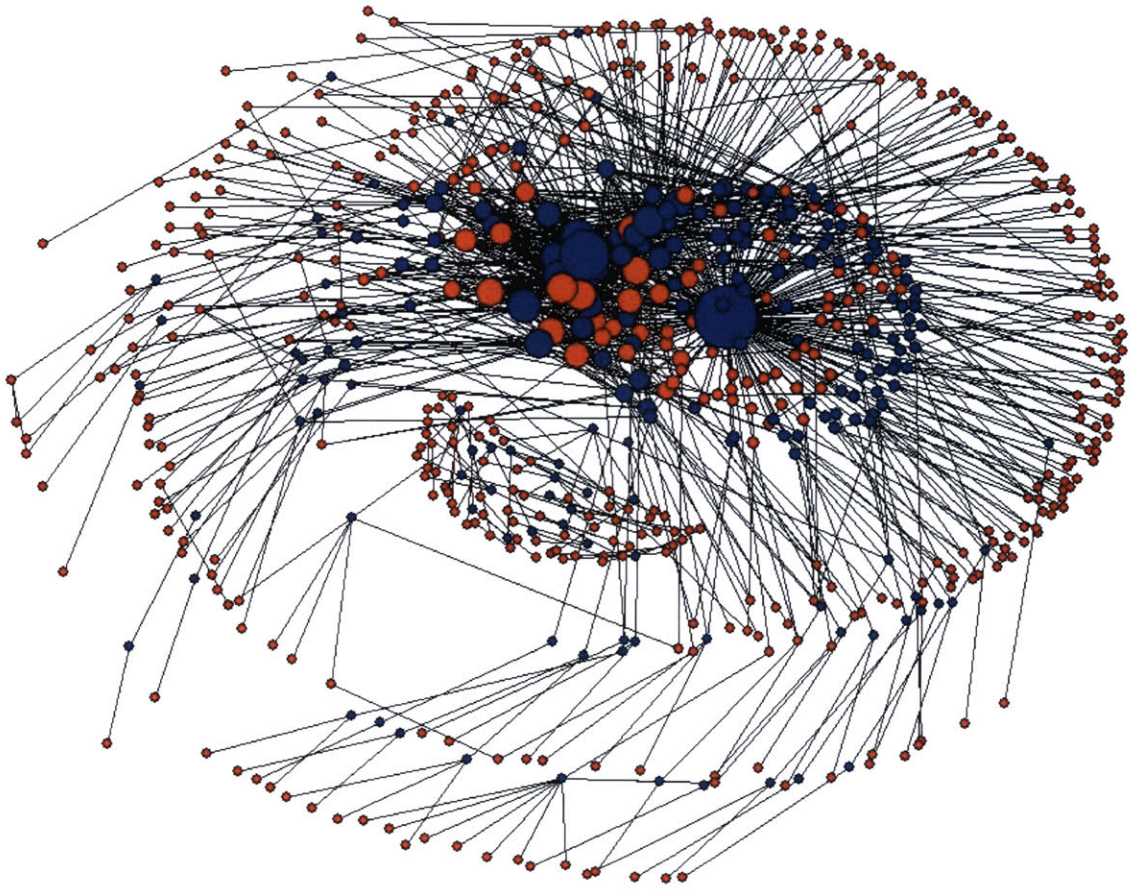


Figure 4.7 **Adoption of home computers within a network of advice. The size of the nodes is proportional to the eigencentrality.**

4.7 Conclusion

We were able to show a successful impact on the rate of adoption, which is overwhelmingly high if compared with the neighboring region, its own region and the national level. We were able to test quantitatively the existence of a significant network effect. We also tested the

network effect for possible confounding effects, and found that when controlled by these effects and their interactions, the network effect remains and appears to be stronger.

Besides the objective of testing a hypothesis, one should explore the ulterior consequences of the accelerated adoption of new ideas and technologies within the community. Surveys are rather blunt instruments of data collection. They are useful to produce rigorous estimates such as the ones just described. But they will be inadequate to capture the intricate and complex behaviors of human societies as they become better tooled and act in a richer environment. The next and final chapter takes a different approach to address the question of the overall importance of addressing the question of community development from this new perspective.

Chapter 5: Network Approach to Development

> "But in capitalist reality as distinguished from its textbook picture, it is not that kind of competition [price competition] that counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organization (...) -- competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives." (Schumpeter, 1943, p.84)

We maintain that the social structure of a community encapsulates otherwise unmanageable quantities of information about a community's dynamics. We claim that targeting networks, not social classes, age, gender groups or institutions, creates the basis for the emergence of local organizations, businesses and collaborations that organically provide the necessary support to achieve an effective diffusion process of technological innovations. This claim translates into the three hypotheses presented in Chapter 1:

1. Networks of advice and quantitative analysis of sociometric data provide a novel and efficient way to effectively find the influential members of a community.
2. Targeting the influential members of a community through this procedure improves the rate of adoption of innovations.
3. Targeting networks, not social classes, age, gender groups or institutions, creates the basis for the emergence of local organizations and entrepreneurs providing long-term sustainability.

Hypotheses 1 and 2 were quantitatively addressed in Chapter 3 and Chapter 4, respectively. Hypothesis 3 will be treated in this chapter with a different approach. The first half of this chapter presents an economic argument that links the social network effects on the

diffusion of technological innovations with the leading theories on competitiveness and sustainable development. The purpose of the argument is also to frame the contributions of this thesis within the most general discussion on international development. The second half of this chapter is a collection of five small business cases that show from the community members' perspective, how their businesses emerged and are using technology in ways that are shaping a richer "ecosystem" or "business climate" that potentially benefits the community at large. They do it by providing services in a sustainable way, independently of external and usually uncertain donors' funding and the burden they create when their funding is over.

5.1 The microeconomics of sustainable development

While there are many definitions of sustainable development, the general concept is mostly about increasing the well-being of current generations, while preserving productive capacity for future generations. In very broad terms, it means sustainable economic development that supports long term sustainable improvement in a nation's standard of living, which in turn is determined by the productivity of a nation's economy[82]. The macroeconomics and political economy of competitiveness and economic development are becoming better understood and, while there is no consensus, there is mainstream agreement that the best conditions for development are those that provide a stable political environment, sound political and legal institutions, prudent fiscal and monetary policies, responsible debt management, cost effective size of government, a market driven economy with a dose of well defined regulations, openness to international markets, high and sustained rates of investment in human capital (mainly health and education), physical capital and technology.

Many agree that those conditions are necessary but not sufficient. Porter and Christensen argue:

> "As important – or even more so – are the microeconomic foundations of economic development, rooted in firm operating practices and strategies as well as in the business inputs, infrastructure, institutions and policies that constitute the environment in which a nation's firms operate (…) Some economists think that if the proper macroeconomic conditions can be put in place, the rest will take care of itself. If governments operate efficiently, aggregate savings are ample and inflation is controlled, the lower interest rates will lead firms to make the investments necessary to enhance competitiveness. If government resources are allocated to education, the resulting rise in human capital will translate into jobs with higher wages. If the government removes distortion in prices and exchange rates, firms will become more innovative and sophisticated. There is some truth in that, because lowering the cost of capital, raising the rate of investment and removing distortions certainly matter. However, the gap between macroeconomic policies and company competitiveness is a wide one. A myriad of intervening circumstances at the microeconomic level must be understood and addressed by the private sector and through government policies if a nations' prosperity is to improve."

We think that their view of the dynamics of business and cluster formation at the micro-level can be complemented by an "operational" or quantitative approach to the notion of social capital, which can provide a valuable layer of information. We also think that social network analysis can provide the heuristics to actually move beyond the general concept of

social capital to well defined metrics that are useful to assist development initiatives, especially in rural settings in third world countries, where we think their argument needs to be taken one step further.

For most rural communities in underdeveloped nations the divide between the public and private sectors is blurred, many government policies are hardly enforced and institutions are weak or simply do not exist. In such cases, the norms, behaviors and resource allocation are highly influenced, complemented or replaced by the local culture and the social capacity of the community members. In such cases, their arguments may not work best to describe the dynamics at the micro-level. In this context, the discussion may benefit stepping down to an even finer granularity, by looking at the individual, not the firm, or the institutions, but at the community's web of interactions. From those interactions, the organizational processes that enable the transformation of new ideas (knowledge) into economic knowledge may or may not emerge.

The critique of mainstream economics about its lack in addressing the individual is, at least, about a hundred years old. According to Fagerber[83], the term "methodological individualism" was coined by Schumpeter in 1908 and through the last century his ideas have evolve into different strands of evolutionary economics. Following Faberger, the different strands within evolutionary economics are tied together at their core by three interrelated arguments that define evolutionary dynamics. "The first of these specifies the evolutionary processes and the third is concerned with the relationship between evolution, cognition and action".

The basic argument is that innovation is the main driver of economic development in the long run. The second argument is about evolutionary processes being characterized by

strong regularities: a sequence of innovation and imitation. Innovators are amply rewarded at first, but these advantages decrease as imitators enter the scene. The role of innovation as a pointer to further change is another important regularity: an (important) innovation facilitates the development of certain types of applications, leading to links between innovations or technologies sharing the same context. The role of incremental innovations is related to this. Learning, based on accumulated experience with the important innovation, sets out a path for other innovations. The last regularity described by Faberger has to do with the influence of users in inducing, improving and selecting innovations. The third argument is about the role actors (and cognition) play in the evolutionary process. It looks at economic knowledge as a set of routines (for action) that are reproduced (remembered) through practice. Firms are forced to abandon the ideal of the "rational man" and go for a more economic (realistic) strategy by the combined effect of randomness (open-ended character of evolution) and the impossibility of the decision maker to consider all the effects and counter effects.

This framework is useful to discuss the potential impact of our contributions to the process of diffusion of technological innovations in the context of development. The cases we present in this chapter as positive outcomes of our experiments in SMD also have a good fit with this particular way of understanding economic growth. we present them from the perspective of the individual and her connections with her social system, in particular with her connections with the influential members.

5.2 Social Capital and the Influentials

Social Capital is the social structure which facilitates the exchange of ideas, coordination and cooperation. Since 1916, when the concept was introduced, many definitions have been

proposed but all of them place exchange through connections among individuals as the central component of social capital[84].

Underneath the concept of social capital lies a network of trust. This network is the social infrastructure that allows the exchange of ideas, practices and artifacts. In all societies there is an interaction between social capital and institutional capital. The latter, supplements the first. Banking, for example, started within a brotherhood, a tight social system of mutual trust, which made acceptable paper notes representing gold. Today the financial system supplements the lack of trust between someone who has the money (investor) and someone that needs it (lender) since they do not know each other. The bank that "knows" both provides the necessary link for the resources to flow from one end to the other and vice versa. This is true in many daily human interactions. When institutional capital is low, as is frequently the case in underserved communities, the efficiency and effectiveness of resource allocation depends heavily on social capital, limiting access to resources and ideas to those that are disconnected from the main social networks. The role of the influentials is huge in promoting innovations when the institutions that typically provide guidance, support and incentives for individuals are not in place [85] (see Figure 5.1).

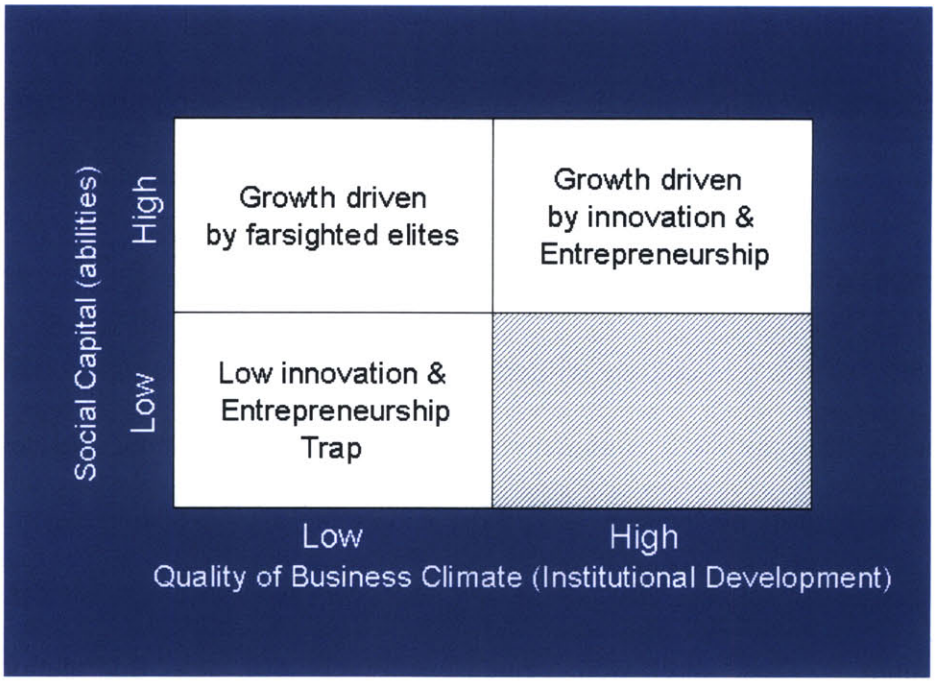


Figure 5.1 Social Capital and Business Climate

The influentials are already connected in many ways. They are aware of each other via collaboration or competition. Their social capital is as well part of the social capital of the community as a whole as far as they are part of the same social networks. In our experiment, helping them to become the first in collectively adopt asynchronous digital communications had a high probability of adoption because it simply makes sense for them. If their peers become accessible at the same time they do, they will be able to tap into the new media immediately. If competitors within their networks are also becoming accessible through the new media, neither would they want to be left behind. Given that core-periphery network structures are common and influentials are associated with central positions within social structures, the diffusion of technological innovations was expected to occur more rapidly.

The following business cases are a qualitative approach to illustrate ways in which the adoption of technology in Santa Maria seems to be having an impact that is consistent with the rationale we described in this chapter. They are about entrepreneurs and their businesses.

5.3 Case Studies

The following five cases were written following the case method tradition of western business schools. They follow a common structure: a description of the business and the use of technology from the perspective of the entrepreneur or decision maker, as well as their connections to the rest of the community. All five cases close with their plans and hopes for the future. These cases were written exclusively with data, information and perceptions provided by the community member during a face to face interview. These cases are to be published by INCAE Business School as teaching materials and are coauthored by Rene Zuleta, an INCAE student who conducted the interviews.

Case 1: Los Santos Coffee Shop, is the case of a well established business man who creates a new firm based on new ideas. He is an influential, exploiting new ideas and creating new channels of commercialization.

Case 2: Simprima, S.A. Inspired by his father's previous work and new tools, the son of an influential starts a new business and introduces innovations by using the new tools to change his supply chain.

Case 3: NutriSalud, S.A. It is about a young professional, daughter of an influential, who migrates back to her town and is providing innovative health services.

Case 4: Santos PC, S.A. A young woman, of humble origin, becomes a business owner selling computers and providing technical support. She gets support to launch and develop her business from one of the influentials.

Case 5: Academia San Marcos, S.A. is a former employee, who gets support from another influential, to start up a computer school in a nearby town, after he had helped another influential to start up a computer school and Internet Café in SMD.

5.4 Case 1: Los Santos Café Coffee Shop

R.A. is a well established business man. He is founding partner of a local hardware store in the neighboring San Marcos. He is also an active member of the civil society and business community. As an agronomist, he was aware of the quality of different coffee farms throughout the region and it bothered him that none of the local cafeterias at the time were reflecting or promoting the special characteristics of Los Santos coffee production.

In April 2002 he inaugurated the first coffee shop in San Marcos de Tarrazú; it was a very small locale and not too different from other coffee places in town. In 2003, after being exposed to different ideas related to the world's renaissance of neighborhood coffee shops, as a meeting and working place, he started to gather information and test ideas through conversations with friends and acquaintances, especially with those that had been abroad about a different business model for his cafeteria. He was among the first to have Internet in his business, and had quickly become a daily user. It became the first source of ideas, equipment specifications and even coffee recipes. After being in business for a year and a half, he moved to a larger space, this time with a different strategy. These ideas basically followed models of existing coffee shops that offer high quality coffee, diversity of beans and preparations, and an inviting environment for people to spend time in the cafeteria, which allowed him seven to fifteen times the price of a regular coffee in a traditional coffee shop.

As an active member of the community, and having been influenced by the ideas about social networks and the flow of ideas, it was easy for him to realize the social importance and economic value of the physical space he had created. It was also a learning space. He introduced a key innovation: he was the first in the Los Santos Region to train a certified a "barista"[8]. Los Santos Café was starting to become a meeting place for planned or serendipitous conversations.

Also in 2004, he opened a second coffee shop, outside of town, next to a heavily transited road. He planned to benefit from the road as a traffic generator. An interesting segment was that of tourism in their way to other destinations that had the chance to learn about Los Santos while enjoying a cup of coffee during their short stop. In January 2007 he moved his first cafeteria to a larger space, this time designed to provide space to people who may want to use the wireless internet service that he is planning to provide soon. This new place was also better equipped to fully exploit the talent of his well trained employees.

5.4.1 Imitation and Competition

In 2006, COOPEDOTA, R.L. opened its own cafeteria targeting the same segments R.A. was targeting, offering a similar experience and price range. Coopedota's cafeteria is also exploiting its closeness to the coffee mill and the knowledge of its "baristas" to offer a coffee tour, which is beginning to attract the attention of national tour operators. These tours are sometimes complemented with folklore presentations of a local youth association. R.A.'s response to his new competitor is geographic growth. In February 2007, Los Santos Café

---

[8] Also known as a coffee "sommelier", someone who is highly competent in the preparation of espresso-based coffee drinks and has a comprehensive knowledge about different varieties, roast degree, latte art, etc.

opened its first coffee shop in San Jose, Costa Rica's capital. He is also expanding his product lines to include catering services. Now his research is oriented to the establishment and operation of franchises. In the near future, RA hopes that his coffee shops will be all around the country as a beacon of Los Santos quality and hospitality.

5.4.2 Discussion

Among the influential members of the network we chose to work with, R.A. is probably among the most sophisticated ones. He had learned to use email before any other of the influential members and was an enthusiastic supporter of our project and ideas. His energy and connections within the community were of great help. Most likely, a conventional approach to developmental projects will leave someone as R.A. out because he is certainly not the poorest of the poor; nor does he belong to any of the usual "beneficiary" categories. But, he is an innovator, and he is someone with enough resources and has the drive to take the risk of converting new ideas into economical ideas. Other influential members have respect for his opinions and ideas. To be connected with him and his network is a source of social capital.

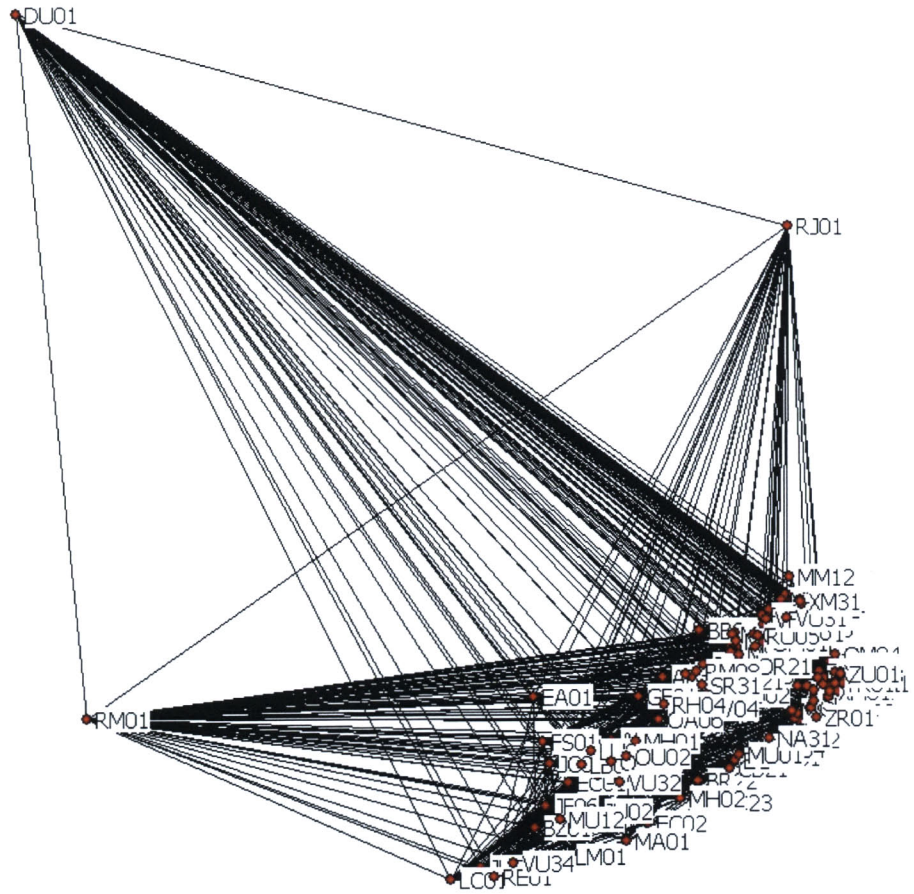


Figure 5.2 **Principal Components Layout of the Advice Network, it shows the position of RJ01 described in case 1 and RM01 described in case 2.**

5.5 Case 2: Ganadería Sinprima R y M.

In 2003, M.M. started the cattle-raising business as an alternative to coffee production motivated by the crop's crisis. He went into business with his uncle, but soon became independent. He started using artificial insemination, but later M.M. decided to move to an embryo transfer process, a technique his father had pioneered in SMD 10 years earlier. His father R.M. had abandoned the cattle business when he moved to coffee production and later became Coopedota's general manager. We will come back to R.M. in the discussion of this case.

Sinprima had a herd of 25 breeding cows and until a few months the company was buying semen from a local provider. This genetic material came from the same three bulls that his provider was offering the local competitors.

5.5.1 Technology and the supply chain

In his own words, M.M. was motivated to explore the use of the Internet out of "shame" because his father had become an intensive user. Once he tried his father's computer, with a little help from him, M.M. changed his ideas about the use of computers, and he also became a frequent user and bought one for his home. Now he frequently searches the Internet for information about cattle handling and cross-breeding, particularly from Austria, Germany, Canada, and South Africa. Local competition takes place particularly in livestock fairs. Cattle breeders attend these fairs to gain prestige, compete, get publicity, and acquire information on better crossings. However, for him the Internet has become the main source of information to improve his calving practices and trait selection.

His productivity and the quality of his products changed dramatically with the new practices he has incorporated. But the most dramatic change comes from a business innovation. Instead of using the same local provider of genetic material, he is using the internet to locate and use a much broader international offer of providers, from whom he is buying frozen embryos and semen.

Product improvement has reflected on Sinprima's sales and prestige and M.M. has started to share his knowledge and techniques with fellow Santa Maria breeders, as he understands that he will benefit if the region starts to position its name as a source of superior breeding.

5.5.2 Plans for the Future

Future plans to improve the business include the development of a webpage to promote his products and services. He plans to design a webpage showing pictures of cattle for sale, including their genetic line, so customers could be assured of product quality. Attending fairs is costly and he hopes to avoid some of that cost by attracting potential buyers through his webpage. He is considering the idea of becoming an embryo exporter himself. Beyond his business, the adoption of technology has also impact his family dynamics. He is married to a local school teacher and their children are still too little to attend school, but that did not stop them from playing with the computer that M.M. bought for home from Santos PC, a new local PC provider. M.M. and his wife have become first hand witness the potential of the home computer to support a learning environment for the children. As a teacher, she is now concerned about the need for a similar environment for the students at her school.

5.5.3 Discussion

Sinprima is an interesting case because its emergence is not directly related with the coffee "cluster". M.M. started using the Internet because he was motivated by his father, who is one of the influential members in the advice network discussed in previous chapters. He used the "new technology" to produce a different innovation: a change in his supply chain. As other cattle producers in his network benefit from his knowledge or competitors start to imitate him, a different but related network will be enriched by the new ideas and will also transform those innovations into economical and social value. This may be a good illustration of the Schumpetarian observation of how one important innovation tends to facilitate (induce) other innovations[83].

The same way he bridged the flow of ideas to the cattle breeders, his wife will probably play a role in the diffusion of ideas and the propensity of adoption of other teachers and families. M.M. was not the only family member impacted by his father. His mother is an influential woman within the local Catholic Church and a leader in the catechism movement and the leader of a female network of about 70 volunteers that helps the Church to prepare the community's boys and girls for first communion. She was among the first to learn how to use powerpoint. She saw an opportunity to improve the way they were teaching cathecism. Thus, she was instrumental in registering several dozen of women in computer classes. Those women were among the first groups of clients that helped jump start the Women's Association new computer training school.

5.6 Case 3: NutriSalud, S.A.

In May 2003, V. started practicing her profession at her private outpatient nutritional counseling clinic, NutriSalud, S.A. located in San Marcos de Tarrazú. On average, NutriSalud is serving per month 40 patients (25% men, 63% women, 12% children). Nutritional Services as a private practice is new to Los Santos, and NutriSalud became the first and until to now the only private practice available in the region.

It opened offering services 2 days per week. V. was born and raised in Dota, and after finishing college she got a job offer in San Jose. By reducing her work load, she managed to open her new business and at the same time she kept her job. This combination allowed her to get acquainted with patient management and gain professional experience.

One of the main challenges she had to face was creating a local demand for her services. The Los Santos population was unaware of the importance of having a nutrition care

treatment or asking for advice on this matter. This was also true for other health professionals in the area. Physicians, nurses nor psychologists were fully aware of her services and they were not a source of reference.

5.6.1 Exporting services

The idea of providing nutrition counseling to patients at a distance came from V.'s relatives, friends, and acquaintances living in the United States who had requested her services by email. They were part of a group of Los Santos people who had migrated to the US as a consequence of the coffee crisis.

The NutriSalud webpage http://www.nutrisalud.co.cr/ was among the first business pages set up by local entrepreneurs, and even though they are not sure if the website is actually attracting new customers, V. sees it as a tool that eventually may attract clients, and in the meanwhile it can play a substantial role supporting her practice and her patients as she advances its site development in the near future.

Today a sixth of her patients live abroad, and taking care of them is an activity that is usually done from home, where she has Internet access. She sees growth potential in this niche were she can differentiate her service offering it at local prices, in Spanish, and with all the familiarity involved in sharing common roots with her clientele.

5.6.2 Looking ahead

In order to fully use the Internet, she wants to take the NutriSalud webpage to the next level. She foresees the need for developing a program that enables a more personalized interaction supporting their nutrition care plan, case analysis, follow-up, and assessment tools.

NutriSalud is also exploring the corporative market. So far, it has been counseling some cooperatives, such as CoopeTarrazú R.L, Coopedota R.L, and CoopeSantos.

At a personal level, V. knows that she will continue using The Internet to keep up to date with her profession and to search for ideas to improve her practice.

5.7 Case 4: Santos PC

In 2002, L.V. started working with the Lincos Project in a Hewlett-Packard project as promoter of community services rendered by Lincos. This promotion was an attempt to improve Lincos economic sustainability. Her job consisted of door to door service sales. While working at Lincos she learned about basic computer technical maintenance, customer service, and graphic design.

When the Lincos project was shut down by the end of 2000, L.V. couldn't find a job in Los Santos, so she migrated to San Jose were she found one as a part-time teller at a financial institution. During that time, she also started a little business on the side as a freelancer in Los Santos. Using the tools she had gotten acquainted with she was designing coffee package labels, professional visit cards, fliers, etc.

5.7.1 Entrepreneurial attempts

In 2004 she partnered with a former Lincos female co-worker to import printer toner from China. The contact was made through e-mail and soon they had their first shipment to be sold. The results, however, were not as expected. They were new to the business, the brand was also new in the country, and companies had regular suppliers of this kind of product.

Nevertheless, the experience gave them the idea of selling another product: used personal computers. There was a computer school in SMD run by a women's organization

and recently another school from a former worker of COOPEDOTA was quickly growing in San Marcos. It must create a demand for home computers, they thought, and they were soon convinced of offering low-cost used computers that rural people could afford.

L.V. and her friend invited R.A. to join them as an equity partner. This alliance provided them with working capital, a small office space in the back of his hardware store and a close source of business advice. The first computers were some Pentium I bought in San Jose from a second-hand computer dealer.

Around 2005, business growth allowed them to hire someone for a part time job to help them in dealing with their hardware purchases from suppliers. At the same time, they refurbished the site to build shelves and began carrying some second-hand accessories, such as floppies, CD-ROMs, monitors, etc. That same year they tapped into a different market segment. There was an opportunity with businesses that required a low-cost computer to run their basic administrative and accounting tasks.

5.7.2 The Sales Strategy

The strategy they used in launching sales was to start selling on the outside and then move in, that is to say, from remote districts toward the main towns. This resulted from their vision of targeting people who did not have access to expensive hardware, for example, a coffee grower with 5 kids who needed a computer. With this in mind, they would go to schools offering their computers to schools, children, and parents at PTA meetings. As expected, the business continued to grow and they hired an additional person. She had some basic knowledge of computer maintenance; they provide her training on how to explain technology to people and how to deliver basic computer training. This is a differentiating characteristic of Santos PC. They were sensitive and in many ways shared with their customers the difficulty of not only

grasping new concepts like "ram memory" but also pronouncing their names, as the computers and software came bundled with a lot of Anglicism (driver, usb, etc.).

With time they saw that customers who had a computer for some time now wanted to buy a better one, so they expanded their business line to include new computers, computer parts and accessories. These implied new suppliers, but this time they found them over the Internet, avoiding at least the local intermediary.

As the region's computer base grows, the demand for repair and maintenance services is also growing. To this end, they conditioned a technical computer workshop. It was the first one in Los Santos. Before this workshop was opened, all repairs were done by technicians visiting customers or by sending the computers to workshops in San Jose.

In late 2006, the ADSL service was launched in San Marcos, thus leading Santos PC to start offering new services, such as online purchasing, where people come and order products (particularly PlayStation 3 games), and they purchase the items mostly through Amazon using courier services, i.e., JetBox.

5.7.3 Plans for the Future

Currently, Santos PC has hired two additional workers to help with maintenance. They are now expecting to open a new store in San Marcos and to structure a sales department, while keeping the workshop space in El Carpintero.

L.V. continues designing logos, coffee labels, visit cards, and similar jobs, and she expects to have a website soon.

5.7.4 Discussion

We chose this case because it illustrates a young woman of humble origins was able to mathc her abilities with the opportunities provided by an enriched rural context. Her personality traits, her new tools and the change in her social network and context allowed her to flourish as an entrepreneur, recombining innovations to produce economic and social value. It is important because it shows how her new abilities allowed her to become a member of a richer network (coffee cluster) where she formed an alliance with on of the "influentials". It also shows the importance of the influentials as people with the ability to mobilize resources for change. In this case R.A. is mobilizing financial capital (investment), business ideas and infrastructure (ideas) and his own capacity to assume the associated risk (social support) within his network. It is the increased social capability that creates the opportunities they both seized.

We know that our work with the advice network is one instantiation of different ways in which the community's "social capability" can be improved in order to benefit from the flow of ideas, practices and artifacts. Nevertheless, this case is consistent with the claims we sustained in the previous chapters.

5.8 Case 5: Academia de Cómputo San Marcos S.A.

M.A. started working for Coopedota in 1990, a few months after he graduated from high school in Santa María de Dota. In 1992 he became a certified private accountant. He worked as an accounting assistant for Coopedota R.L in charge of transactions, deposits, and check issuance, among other things. Some of these tasks were done on the computer, although most of the information was handled manually.

In 2003 a major information system was introduced that also brought about changes in the organization and he ended up in charged of the IT department.

His experience and training at Coopedota allowed him to witness technological changes in a business, get constant training, and keep current in his profession. It also expanded his vision concerning the use of technologies and their application in everyday life. Coopedota was also involved in supporting the creation of a computer training school managed by the women's association, which started with a donation of 5 used computers. This gave him the idea of establishing his own academy in the neighboring town of San Marcos.

In January 2004, he launched Academia de Cómputo San Marcos, legally established as a corporation. Initially, the academy was only open for evening classes as he continued working at Coopedota during the day. In june of 2005 he quit his job to work full time at his own business.

In the Los Santos area there were no professionals teaching information technology, and that is why M.A. had to design, update, and revise the manuals (basic course teaching materials), according to the new programs that would come out to the market.

He started with eight computers, which were subsequently increased to eleven. One year later, in 2005, they began offering Internet services, opening to the public during the day as an Internet Café and offering computer classes in the evening.

Demand grew, as well as customer needs, and so he decided to diversify the service by offering not only Internet, but also printing, document typing, bookbinding, Internet information search, and photocopying services.

As a reference, IT courses in 2004 started with an average of 30 students and they are currently (2007) working with 77 students.

Courses offered by Academia de Cómputo included Accounting and English (basic, intermediate, and advanced). The academy grants a Computer Operator certificate after completion of Excel, Word, Power Point, and Access courses.

All of the courses have the value added of being accredited by the national learning institute, or Instituto Nacional de Aprendizaje (INA), which has become a major differentiating factor, since it is the only educational body accredited by this institution, thus providing it with support and prestige.

In addition to the above mentioned services, since September 2004 the academy offers new and used computer hardware sales and repair services. In September 2004 they were handling roughly a $1500 inventory, while in April 2007 that figure went up to about $10,000.

In 2004, M.A. had two workers. Currently, he works with eight more people. As for teachers, at the beginning he was the only one; now there are four teaching IT courses. Most teachers and workers come from the area.

Direct competition faced by the academy happens mostly in computer hardware sales and repair. Stores and businesses strongly competing against the academy are: Santos PC, Ferretería Nidia, and some bookstores, most of them located in Tarrazú.

Having his own business has been a very enriching work and personal experience to M.A., since it has allowed him to overcome challenges and to innovate in the area, concerning the use of technology.

The academy has trained parents interested in managing technologies in a suitable manner, in order for them to help their children do their homework. The oldest student was a 52-year-old man.

For the future, M.A. is planning on expanding academy programs and courses, as well as its infrastructure. For instance, he plans to deliver courses in programming, webpage design, and Autocad. He will also start courses aimed only at children.

In expanding his business, however, some limitations are the lack of infrastructure and not having his own land and/or facilities.

5.9 Cross-case Comparison

They selected cases show how:

- They improved the business environment
- They enriched the "social ecology", which creates market incentives for further investment
- They improved the communities' capacity to respond to market incentives and social needs
- They improved the flow of ideas as the "newcomers" bridged their original social networks
- The role of the influentials in creating or supporting new businesses
- Entrepreneurs taking an important innovation and remixing it to generate other innovations that create economic and social value

5.10 Innovation, Entrepreneurship and Social Networks

People, families, firms and communities always have an "established way of operation". When new ideas appear that challenge "the way things are done" threatening with an increased level of uncertainty, the natural reaction is that of rejection. Only a few members of society have the psychological drive or motivation to adopt and recombine innovations to bring them from ideas to successful economic implementation. Schumpeter assumed they were normally distributed[83], which may or may not be the case (we do not know on what basis or what are alternative functional forms). We do think that as the network grows in quality and quantity of nodes and links in a way that the flow of ideas, practices and artifacts is improved, the likelihood of new entrepreneurs emerging within the social system will increase, and their personal success will translate into the generation of economic and social value for the community to prosper as a whole.

By targeting the influentials we are tapping at the heart of the community, they are not central by choice, they are central because the complex social dynamics of the society allow them to be in that position of power. Necessarily, to be central in a network of advice, their opinions are actions must be regarded valuable by most of the network members. We have seeing how a few influentials in one or two degrees are connected to them. These cases show how people of a different socioeconomic background became entrepreneurs and innovators and benefited from the social capital of a few influentials.

The cases we chose are a good support to conclude this chapter stressing the importance of social capital for decision making. Scholars have address the importance of shared information to entrepreneurship and economic growth [86], but this should not get

confused with information access through the Internet. As important as it is, in decision making social capital provides much more than the facts to make a decision. Potential risks and benefits are weighted through conversations with those whose advice matters to the decision maker. The cases discussed in this chapter about firms using technology and innovating, but all of them were clearly developed with financial and cognitive capital of community members that belong to the core of the influentials or are connected with them. Before businesses emerging, there were a social dynamic that does not operate in a random vacuum. These dynamic is strongly determined by the structure of the community's web of interactions.

Chapter 6: Conclusion and Future Work

"There's still much to do... still so much to learn."
Jean-Luc Picard,
"The Neutral Zone",
Stardate 41986.0[87]

6.1 Contributions

Technological change is the main contributor to economic growth. In 1956 and 1957 MIT professor and Nobel laureate Robert Solow surprised many with his conclusions: investment (in machinery) cannot be a source of growth in the long run. He argued that technological change, not investment, drives long-run growth. He estimated that technological change accounted for seven eighths of U.S. growth per worker over the first half of the 20th century[88].

The main contribution of this thesis is to have developed a method to quantitatively and efficiently identify influential members in a community. We have used the ability to predict who is influential to guide the diffusion of technological innovations. We set up a large scale experiment in a community of coffee growers in Costa Rica and studied the adoption of home computers as a proxy for the diffusion of ideas, practices and artifacts. We found strong network effects in that were rigorously analyzed to establish its statistical significance. As Solow showed, accelerating the rate of adoption of technologies is of paramount importance in pursuing economic growth.

We also documented five business cases that we believe are the outcome of our intervention. In a Shumpeterian approach to organic economic and social growth, we argue that diffusing powerful ideas and tools through out the community, from the core members of the advice structure to the periphery, generates a context that fosters entrepreneurship and innovation, which is actually the mechanism that translates ideas, practices and artifacts into economic value and social prosperity, which are the ultimate goals of our research.

Specific contributions of this dissertation include:

- Developing methods for computationally modeling the social structure component of a diffusion process. By identifying networks of advice and estimating its eigenvector centrality we were able to combine socio-economic, behavioral and sociometric data to predict who is influential.
- The first large scale experiment that uses sociometric and behavioral data to succesfully diffuse technology in a rural community. The results are very strong and promising, and the methodology simple enough to facilitate further testing in other communities.
- This work generated a unique data set that allows further exploration of other networks, other possible effects that haven't been modeled in this thesis or other research methods to establish the size of the network effect in this and other social outcomes.
- Costa Rica has host different research efforts from faculty and students of the Media Laboratory since it was founded. We benefit from the good standing of that relationship which has being brokered mainly by government. Our careful work with the Los Santos Comunity, the understanding of its social structure, the amount of data we collected, and in general the goodwill created by this project, sets the stage for the continuation of this research and the support to other research agendas from the Human Dynamics Group and the Media Laboratory at large.

- The potential impact of this approach as a solution to the global need of diffusion strategies in IT4D projects is empirically tested using both quantitative and qualitative evidence from a community of coffee growers in southern Costa Rica. This method is well documented and ready for further trials in similar communities around the world.

6.2 Future Directions

There are different possible directions for future research that extends this work. The next final sections describe some of them.

6.2.1 Influence and Advice

Network Analysis offers a systematic tool to study the flow of ideas and the structure that accommodates it. Overtime, human attitudes and abilities will form patterns of communications and those patterns will help to predict behavior. Our research strongly suggests that patterns of advice are good predictors of influence. Future research can explore better or complementary ways of defining who is influential and who is not, and test the capacity of advice networks to predict it. Likewise, advice as a latent variable could be subject of exploration. What kind of advice works better as a predictor of influence? What is the best way to aggregate different advice networks? Should it be an open question, or a question bounded by a timeframe will provide more precise results? Is any network of advice useful for this kind of prediction, or is it sensible to certain network properties, transivity for example?

6.2.2 Eigenvector Centrality and other measures of centrality.

We discussed Bonacich centrality and it is the basis of the advice centrality index used in chapter three. The β parameter was a very important addition to the previous centrality indexes. By assigning a zero, positive or negative value to this parameter the researcher has the capacity of defining three different types of centrality. However, this assignment is done as a function of context and working hypothesis. Simulation and sensitivity analysis could be use to explore possible ways to assign this value conditioned to other criteria such as network properties. There are other measures of centrality found in literature that could be explore, beyond the ones considered for this research.

6.2.3 Network Modeling

Sociologists, anthropologists, social psychologist, and others are sources of different communication theories. These theories provide the basis for different rules of engagement such as: self interest, collective action, balance theory, resource dependence theory, exchange theory, homophily, proximity, evolutionary or coevolutionary. There are current efforts in the network science community to develop the algorithms that model the network structures that will result from the prevalence of such rules of engagement. This research is a rich source of future research that will explore the effect of these rules and their interaction in the emergence of characteristic network patterns.

Once we have this ability, new avenues of research and professional practice will emerge. As we move rapidly into a civilization phase characterized by the omni presence of sensors and digital information on one's behavior, the ability to automatically and dynamically capture the latent communications network structure will be available.

Therefore, the observed network properties may be deconstructed to infer the underlying rules of engagement of a team, group or community.

6.2.4 Network Sampling

We discuss the existing technical challenges of network sampling. On the other hand, in terms of diffusion and the role of influentials, may be these challenges can be overcome by the data collection strategy. For example, social scientists are familiar with the problem of rare events. The statistical analysis base on samples of rare event data with dozens to thousands of times fewer zeros can lead to sharply underestimated probabilities of rare events. This is overcome without losing consistency by focusing in the collection of "1s" and sampling the zeros. Our dataset and others can be use to explore the impact of random sampling using known techniques such as Monte Carlo simulations, and then contrast them with alternative sampling strategies, to measure the impact on this specific application (optimizing a diffusion process). The current research to advance the parameterization and estimation of exponential Random Graph Models [89, 90] and other generative models such as the influence model[91] may also provide in the future the technical means to evaluate the fitness of a network sample and its estimates.

6.2.5 Heterogeneous communities

It is clear that the community we chose for our experiments is very homogeneous. It is convenient in many ways, as discuss previously. But future research can extend this work by testing it in more heterogeneous communities. Will the core periphery structure prevail? If not, what network models and properties should be considered to manage and improve the diffusion of ideas, practices and artifacts?

6.2.6 Network Effects Estimation

There are other approaches worth of further exploration. It may result useful to compare the techniques used in this thesis with the results that can be produced by using a multi-level analytic approach. It could be used to study and compare individual and network effects, an approach that can be also useful in dealing with heterogeneous individuals or multiple networks [92].

References

1. Grossman, G.M., *Innovation and growth in the global economy*. 2000, Cambridge, MA: MIT Press.
2. Siochrú, S.Ó. and B. Girard, *Community-based Networks and Innovative Technologies: New models to serve and empower the poor*. Making ICT Work for the Poor, ed. U.N.D. Program. 2005: UNDP.
3. Stoll, K. *Somos@Telecentros: The Story So Far and Lessons Learned*. ICT for Development 2003 [cited; Available from: http://topics.developmentgateway.org/ict/sdm/previewDocument.do~activeDocumentId=442785.
4. Mardle, E. *Telecenters: How Did We Lose the Plot?* ICT for Development 2005 [cited 2005; Available from: http://topics.developmentgateway.org/ict/sdm/previewDocument.doc.
5. Harris, R. and M. Bhatarai, *Fact-Finding Review of ICT in Development in a Rural-Urban Setting*. 2003, United Nations Human Settlements Programme: Hong Kong.
6. Barahona, J.C. and A.S. Pentland, *Entrepreneurship, Networks and Technology*. 2005.
7. Rogers, E., *Diffusion of Innovations*. 2003, New York.
8. Pretell, E.A., et al., *Iodine Nutrition Improves in Latin America.* Thyroid, 2004. **14**(8): p. 590-599.
9. Pepper, S. *LFM G-Lab: Students Spend Spring Break Around the World.* [Internet Site] 2003 [cited 2006; News Article].
10. Barahona, J.C. and A.S. Pentland, *Advice and Influence*. 2005, MIT Media Lab.
11. Freeman, L.C., *The Development of Social Network Analysis*. 2004, Vancouver, BC, Canada: Empirical Press.
12. Kilduff, M. and W. Tsai, *Social Networks and Organizations*. 2005, London: Sage Publications.
13. National Research Council (U.S.) Committee on Network Science for Future Army Applications, *Network Science*, ed. N.A.o.S. (U.S.). 2005: National Academies Press.
14. Newman, M., A.-L. Barabási, and D.J. Watts, eds. *The Structure and Dynamics of Networks*. Princeton Studies in Complexity, ed. P.W. Anderson, et al. 2006, Princeton University Press: New Jersey.
15. Borshchev, A. and A. Filippov, *From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools*. 2005.
16. Sterman, J., *Business dynamics: systems thinking and modeling for a complex world.* 2000, Boston: Irwin/McGraw-Hill. xxvi, 982 p.
17. Nooy, W.d., A. Mrvar, and V. Batagelj, *Exploratory social network analysis with Pajek.* Structural analysis in the social sciences. 2005, New York: Cambridge University Press. xxvii, 334 p.
18. Valente, T.W., *Network Models and Methods for Studying the Diffusion of Innovations*, in *Models and Methods in Social Network Analysis*, P.J. Carrington, J. Scott, and S. Wasserman, Editors. 2005, Cambridge University Press: Cambridge.
19. Weimann, G., *The influentials: people who influence people*. 1994, Albany: State University of New York Press. xiv, 370 p.

20. Hosmer, D.W. and S. Lemeshow, *Applied logistic regression*. 2nd ed. Wiley series in probability and statistics. Texts and references section. 2000, New York: Wiley. xii, 373 p.
21. Allison, P.D., *Event history analysis: regression for longitudinal event data*. 1984, Beverly Hills, Calif.: Sage Publications. 87 p.
22. Rogers, E.M., *Diffusion of innovations*. 3rd ed. 1983, New York London: Free Press; Collier Macmillan. xix, 453 p.
23. Morris, M., *Overview of Network Survey Designs*, in *Network Epidemiology*, M. Morris, Editor. 2004, Oxford University Press: Oxford.
24. Choudhury, T.K. and Massachusetts Institute of Technology. Dept. of Architecture. Program in Media Arts and Sciences., *Sensing and modeling human networks*. 2004. p. 105 p.
25. Olguin, D., J.A. Paradiso, and A.S. Pentland, *Wearable Communicator Badge: Designing a New Platform for Revealing Organizational Dynamics*. 2006, Media Laboratory, Massachusetts Institute of Technology. p. 3.
26. Wellman, B., *Structural Analysis: from Metaphor to Substance*, in *Social Structures: A Network Approach*, B. Wellman and S.D. Berkowitz, Editors. 1988, Cambridge University Press.
27. Snijders, T.A.B., P.E. Pattison, and M.S. Handcock, *New specifications for exponential random graph models*. 2004, Center for Statistics and the Social Sciences. p. 46.
28. Borgati, S.P., *Personal Conversation*, J.C. Barahona, Editor. circa May, 2006: Boston College, Chestnut Hill Campus.
29. Bigus, J.P. and J. Bigus, *Constructing intelligent agents using JAVA*. 2nd ed. Professional developer's guide series. 2001, New York: Wiley. xxii, 408 p.
30. Gilbert, N., *Simulation For The Social Scientist*. 2ND ED. ed. 2005, Buckingham: OPEN UNIVERSITY.
31. Epstein, J.M., R. Axtell, and 2050 Project., *Growing artificial societies: social science from the bottom up*. Complex adaptive systems. 1996, Washington, D.C.: Brookings Institution Press. xv, 208 p.
32. Monge, P.R. and J.N. Cappella, *Multivariate techniques in human communication research*. Human communication research series. 1980, New York: Academic Press. xxiv, 552 p.
33. Monge, P.R. and N.S. Contractor, *Theories of Communication Networks*. 2003, New York: Oxford University Press.
34. Gottman, J.M., et al., *The Mathematics of Marriage: Dynamic Nonlinear Models*. Bradford Books. 2002, Cambridge, Massachusetts: The MIT Press.
35. Hunt, E., *The Mathematics of Behavior*. First Edition ed. 2007, Cambridge: Cambridge University Press. 346.
36. Axelrod, R. and L. Tesfatsion, *Agent-Based Modeling in the Social Sciences*, in *Handbooks in Economics Series*, L. Tesfatsion and K.L. Judd, Editors. 2005: North-Holland.
37. Perez, P., et al. *Diffusion and Social Networks: Revisiting Medical Innovation with Agents*. in *MODSIM 2005 International Congress on Modelling and Simulation*. 2005. Melbourne, Australia: Modelling and Simulation Society of Australia and New Zealand.

38. Brock, W.A. and S.N. Durlauf, *Discrete Choice with Social Interactions.* The Review of Economic Studies, 2001. **68**(2): p. 235.
39. Young, H.P., *The Diffusion of Innovations in Social Networks*. 2002, Center on Social and Economic Dynamics, The Brooking Institution & Santa Fe Institute.
40. Wasserman, S. and K. Faust, *Social network analysis: methods and applications*. Structural analysis in the social sciences; 8. 1994, Cambridge; New York: Cambridge University Press. xxxi, 825 p.
41. Ryan, R. and N. Gross, *The Diffusion of Hybrid Seed Corn in Two Iowa Communities.* Rural Sociology, 1943. **8**(1): p. 15-24.
42. Coleman, J.S., E. Katz, and H. Menzel, *Medical Innovation: A Diffusion Study. New York: Bobbs Merrill.* 1966.
43. Valente, T.W., et al., *Social Network Associations with Contraceptive Use Among Cameroonian Women in Voluntary Associations.* Social Science and Medicine, 1997(45): p. 677-687.
44. Valente, T.W., *Network models of the diffusion of innovations*. Quantitative methods in communication. 1995, Cresskill, N.J.: Hampton Press. xiii, 171 p.
45. Valente, T.W. and R.L. Davis, *Accelerating the Diffusion of Innovations Using Opinion Leaders.* The ANNALS of the American Academy of Political and Social Science, 1999. **566**(1): p. 55-67.
46. Freeman, L.C., *Centrality in Social Networks Conceptual Clarification.* Social Networks, 1979. **1**: p. 215-239.
47. Friedkin, N.E., *Structural Bases of Interpersonal Influence in Groups: A Longitudinal Case Study.* Americal Sociological Review, 1993. **58**(December): p. 861-872.
48. Everett, M.G. and S.P. Borgati, *Extending Centrality*, in *Models and Methods in Social Network Analysis*, P.J. Carrington, J. Scott, and S. Wasserman, Editors. 2005, Cambridge University Press: Cambridge.
49. Blau, P.M., et al., *Structures of power and constraint: papers in honor of Peter M. Blau.* 1990, Cambridge [England]; New York, NY, USA: Cambridge University Press. x, 495 p.
50. Blau, P.M. and O.D. Duncan, *The American occupational structure*. 1967, New York: Wiley. xvii, 520 p.
51. Blau, P.M., O.D. Duncan, and A. Tyree, *The American occupational structure*. 1978, New York: Free Press. xvii, 520 p.
52. Degenne, A. and M. Forse, *Introducing Social Networks*. 1999 ed. Introducing Statistical Methods, ed. U.o.B. Daniel B. Wright. 1999, Trowbridge, Wiltshire, Great Britain: SAGE.
53. Borgatti, S.P., *Centrality and Nework Flow.* Social Networks. **Accepted**.
54. Borgatti, S.P., *Centrality and Aids.* Connections, 1995. **18**(1): p. 112-114.
55. Bonacich, P. and P. Lloyd, *Eigenvector-like measures of centrality for asymmetric relations.* Social Networks, 2001(23): p. 191-201.
56. Borgatti, S.P., M.G. Everett, and L.C. Freeman, *Ucinet 6 for Windows: Software for Social Network Analysis*, H.A. Technologies, Editor. 2002.
57. Borgatti, Everett, and Freeman, *UCINET 6 for Windows*. Reference Guide, ed. I. Analytic Technologies. 2002.
58. Bonacich, P., *Power and Centrality: A Family of Measures.* The American Journal of Sociology, March, 1987. **92**(5): p. 1170-1182.

59. Cook, K.S., R.M. Emerson, and M.R. Gillmore, *The Distribution of Power in Exchange Networks: Theory and Experimental Results.* The American Journal of Sociology, 1983. **89**(2): p. 275-305.
60. Beirute, L.M. and L.F. Mayorga, *Diagnostico de Actores Clave para el Desarrollo Socioeconomico de Dota: Un acercamiento a las redes sociales.* 2003, Centro de Mediacion de Conflictos, CEMEDCO. p. 20.
61. Goodman, L.A., *Snowball Sampling.* The Annals of Mathematical Statistics, 1961. **32**(1): p. 148-170.
62. De-Lange, D., F. Agneessens, and H. Waege, *Asking Social Network Questions: A Quality Assessment of Different Measures.* Metodoloski zvezki, 2004. **1**(2): p. 351-378.
63. Marsden, P.V., *Network Data and Measurement.* Annual Review of Sociology, 1990. **16**: p. 435-463.
64. Krackhardt, D., *Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data.* Social Networks, 1988(10): p. 359-381.
65. Wooldridge, J.M., *Introductory Econometrics: A Modern Approach.* 3RD ED. ed. 2006, Mason: SOUTH-WESTERN.
66. Martin, J.L., *A General Permutation-Based QAP Analysis Approach for Dyadic Data.* Connections 22, 1999. **2**: p. 50-60.
67. Haythornwaite, C. *Tie Strength and the Impact of New Media.* in *Hawaii International Conference on System Sciences.* 2001. Maui, Hawaii: IEEE.
68. DeSanctis, G. and M.S. Poole, *Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory.* Organization Science, 1994. **5**(2): p. 121-147.
69. Berk, R.A., *Regression analysis: a constructive critique.* 2004, Thousand Oaks, Calif.: Sage Publications. xix, 259 p.
70. Provost, F., T. Fawcett, and R. Kohavi. *The Case Against Accuracy Estimation for Comparing Induction Algorithms.* in *15th international conference on machine learning.* 1998.
71. Borgatti, S.P. and M.G. Everett, *Models of core/periphery structures.* Social Networks, 2000. **21**(4): p. 375.
72. Laumann, E.O. and F.U. Pappi, *Networks of collective action: a perspective on community influence systems.* Quantitative studies in social relations. 1976, New York: Academic Press. xx, 329 p.
73. Krackhardt, *Cognitive social structures.* Social networks, 1987. **9**.
74. Granovetter, M., *The Impact of Social Structure on Economic Outcomes.* Journal of Economic Perspectives, 2005. **19**: p. 33-50.
75. *Internet en la pulperia*, in *El Financiero.* 2005: San Jose, Costa Rica. p. 40.
76. Groves, R.M., *Survey methodology.* Wiley series in survey methodology. 2004, Hoboken, NJ: J. Wiley. xix, 424 p.
77. Little, R.J.A. and D.B. Rubin, *Statistical analysis with missing data.* Wiley series in probability and mathematical statistics. Applied probability and statistics. 1987, New York: Wiley. xiv, 278 p.
78. Skillnäs, N., *Modified Innovation Diffusion - A Way to Explain the Diffusion of Cholera in Linkoping in 1866? A Study in Methods.* Human Geography, 1999. **81**(4): p. 243-260.

79. Nyblom, J., et al., *Statistical Analysis of Network Data - an Application to Diffusion of Innovation.* Social Networks, 2003. **25**(2): p. 175-195.
80. Barahona, J.C. and A. Pentland. *Advice Networks and Local Diffusion of Technological Innovations*. in *Communities and Technologies*. 2007. Michigan State University: Springer.
81. Alvarado, R., *Regiones y Cantones de Costa Rica*, I.y. Desarrollo, Editor. 2003, Instituto de Fomento y Asesoria Municipal.
82. Porter, M.E. and C.R. Christensen, *Measuring the Microeconomic Foundations of Economic Development*, in *Geneva Report on Economic Development*. 1999.
83. Fagerberg, J., *Schumpeter and the revival of evolutionary economics: an appraisal of the literature.* Journal of Evolutionary Economics, 2003. **13**(2): p. 125-159.
84. Putnam, R., *Making democracy work: Civic traditions in modern Italy.* 1993, Princeton, NJ: Princeton University Press.
85. Wilson, E.J., *The information revolution and developing countries*. 2004, Cambridge, Mass.: MIT Press. xiv, 431 p.
86. Fountain, J.E., *Social Capital: A Key Enabler of Innovation*, in *Investing in Innovation*, L.M. Branscomb and J.H. Keller, Editors. 1999, MIT Press: Cambridge, MA. p. 85-111.
87. Conway, J., *The Neutral Zone*, in *Star Trek: The Next Generation*. 1987, Paramount Pictures: U.S.A.
88. Easterly, W.R., *The elusive quest for growth: economists' adventures and misadventures in the tropics*. 2001, Cambridge, Mass.: MIT Press. xiii, 342 p.
89. Snijders, T.A.B., *Markov Chain Monte Carlo Estimation of Exponential Random Graph Models.* Journal of Social Structure, 2002. **3**(2).
90. Hunter, D.R., S.M. Goodreau, and M.S. Handcock, *Goodness of Fit of Social Network Models*. 2005, Center for Statistics and the Social Sciences, University of Washington.
91. Dong, W. and A. Pentland. *Influence Modeling and Network Discovery*. in *NetSci'07*. 2007. New York: ISCE Publishing, Mansfield, MA.
92. Snijders, T.A.B. and R.J. Bosker, *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. 1999, London; Thousand Oaks, Calif.: Sage Publications. viii, 266 p.