Program in Media Arts & Sciences

Diffusion of Ideas, Practice, and Artifacts: Agent-based modeling of Social Network Effects on Collective Outcomes

by

Juan Carlos Barahona

Thesis Proposal 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

December 2006

Thesis Advisor

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

Reader 1

Mitchel Resnick
Head of the Academic Program
Media, Arts and Sciences
MIT Media Laboratory

Reader2

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

Table of Contents

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.

Agent-based modeling 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.

The adoption of an idea, practice or artifact is heavily influenced by social context, through both conscious and unconscious mechanisms. Thus better understanding of a community's web of interactions can improve our understanding and management of diffusion processes. Unfortunately data about social interactions is always sparse, and the parameters of analytic models are difficult to accurately establish. The system developed in this thesis combines sparse observation data with theoretical models to generate quantitative models of the latent social structure for specific communities. Network, context, and attribute features learned by the system are used to produce predictive stochastic models of the diffusion process.

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 failure than success stories. This document is a proposal for a doctoral thesis that 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 technical 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 will be complemented by literature review, theoretical analysis of competing methods and an agent-based model of diffusion of innovations that follow network based rules and algorithms.

The Importance of Diffusion of Ideas, Practices and Artifacts

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, frequently yet are dumped into communities without a clear strategy to manage their diffusion amongst its members[1][2].

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 thesis will provide meaningful stories or cases that help clarify powerful ideas, but will also provide hard data and statistical analysis to establish their empirical significance.

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 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].

Media Lab's life long commitment with ICT4D

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. We expect to produce a working prototype of a software tool that contributes with the analysis of the social structure and the evaluation of possible diffusion scenarios based on estimated sociometric parameters out of incomplete data.

Social Network Analysis

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 in the community should be the "entry" or starting points of the diffusion process as well[9]. 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 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 preliminary findings.

Empirical Evidence

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 influential and problem solvers by community members. We also came out with a procedure based on quantitative models to identify those key players in an efficient way. While 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 will use the unique opportunity of exploring, with real community data, the analysis of multiple networks. Methods that currently exist to analyze multiple networks include: multiplexity analysis; block models, network comparison methods; network algebras, comparison of observed networks with uniform random graph distributions of various types and most recently, and still under development, exponential random graph models[10].

This thesis addresses the challenges and opportunities of using sociometric data and analysis to bring about change in a community and also contributes with ideas and tools that potentially can improve our understanding of diffusion processes. The system developed in this thesis combines sparse observation data with theoretical models to generate quantitative models of the latent social structure for specific communities. Network, context, and attribute features learned by the system are used to produce predictive stochastic models of the diffusion process.

Expected Results and Evaluation

Our main goal is to improve the diffusion process of technological innovations as a way to foster wealth creation and quality of life in underserved communities. The way we propose to do it have a real potential to effectively contribute to the solution of the well documented problem of economic and social sustainability of ICT4D projects. Research work performed or underway is briefly mentioned in this introduction and a full description is presented in the second part of the background chapter. Our main hypotheses are:

1. Targeting the influential members of a community improves the diffusion of innovations.
2. Networks of advice and quantitative analysis of sociometric data provide a novel and effective way of finding the influential members of a community effectively.
3. The social structure of a community encapsulates otherwise unmanageable quantities of information about a community's dynamics. Targeting networks, not social classes, age, gender groups or institutions creates the basis for an emergence of local organizations, businesses and collaborations that organically provide the necessary support to achieve an effective diffusion process of technological innovations.

Evaluation of First Hypothesis

Evaluation of hypothesis 1 can be accomplished through:

a) Field observation of the diffusion of home computers in three neighboring communities of coffee growers in Costa Rica: *Santa Maria* where we have an on going project called Marketplace of Ideas using our proposed methodology, *San Marcos* where Media Lab's Lincos project was conducted a few years ago and *San Pablo* where none of these projects has being targeted. All of these communities have similar populations, are exposed to the same mass media and are subject of the same public policies providing a reasonable ground for comparison. Through a cross-sectional analysis, we expect to see higher rates of adoption of home computers in *Santa Maria* when compared with *San Marcos* where the effort and investment in the Lincos project was several orders of magnitude higher and also a higher rate compared with *San Pablo*. This town is important since confounding effects from the Lincos project in *Santa Maria* should be present in this community as well (which is even closer to *San Marcos* than *Santa Maria*). Through longitudinal analysis, we will analyze and document the adoption process of desktop computers among Santa Maria's producers and we will use sociometric measures to perform statistical evaluation of the network effect on the diffusion process.

b) Theoretical analysis of the diffusion process.

c) Simulation of the diffusion process comparing sociometric based alternatives against random contagion models. As a complement to the analyses I will use simulation techniques to generate a distribution of outcomes controlling for the usual socioeconomic measures to test the fitness of the observed network effect.

Evaluation of Second Hypothesis

We already conducted a study with a sample of 122 producers using an ethnographic study as control. We showed that our proposed methodology doubles the precision in finding the influential members in the community. We plan to substantially increase the amount of producers in our next and final evaluation and compare this with previous results. We recognize that the fact that there will not be a new traditional evaluation and that using the same control against newly collected information is a limitation, but it is not a serious one, as social structure in rural communities is not expected to change dramatically in two years. In the specific case of Santa Maria, we know that all main actors are still present in the community. On the other hand, it will help to evaluate whether the main structural characteristics of the advice networks prevail in time and our estimation of key parameters is robust to the use of a larger amount of relational data.

Evaluation of Third Hypothesis

To evaluate our third hypothesis, we will write a series of small case studies of the new businesses that arise in the community, how the technology has influenced different elements of the personal and collective life within the community and we will use descriptive sociometric measures and ethnographic techniques to trace and document the role of the "influentials in these entrepreneurial outcomes.

Expected Contributions

- Extends the capacity to gain insights into the long-term implications for diffusion processes of alternative network properties
- Extends the network analysis tool capabilities to gain insights into long-term implications of different policy measures in network attributes.
- Provides a computational tool to augment and assist theory building.
- Provides insight about how massive interaction data, mined from different sources can be used to improve the simulation and prediction capacity of diffusion processes using network structure information.
- Allows the analysis of potentially massive datasets (while most empirical research on social networks is limited to small datasets).
- Provides a new method for community intervention in ICT4D related projects that fully operationalizes the concept of social structure in diffusion of innovations theory.

Background

The Research Space of Diffusion of Innovations

The study of diffusion of innovation (DOI) leads to important questions that illustrate the research space of DOI [11, 12]: What is the process of invention and adaptation of technologies or diffusion of ideas subject to? What is the process that people go through as they adopt? What are the stages they go through? What influences them at each stage (sources) to move forward or reverse innovations? Why do some people (or collectivities) adopt before others? What are the characteristics associated with early adopters? What are the consequences with regard to economic growth and wealth distribution or public policy design and implementation? How many innovators are needed to achieve critical mass[2]? How do they need to be dispersed?

Personal relations are important in the diffusion of innovations because adoption of a new idea or product involves some level of risk and a learning curve and friends and acquaintances provide a valuable source of information and advice. Starting with Rogers, diffusion of innovations theory and research clearly recognizes that human relations are instrumental in the diffusion process of innovations because they act as channels of social contagion and persuasion [13]. For example, in Lazarfeld's two step communication model awareness results in the first step from exposure of the opinion leaders to mass media, and then these individuals become the source of contagion or persuasion to the members of their social networks [14]. Nevertheless, the amount of research that specifically attempts to deal with Social Network's effects is relatively small and has been developed within the community of social network analysts, mainly by people coming from health related fields. Valente distinguishes two types of network approach: structural network diffusion and relational network diffusion[15]. One can think of several reasons why something recognized as key as social structure effects are usually left out of most empirical work on DOI. The network approach is not only different to the traditional approach, it also introduces important challenges.

Diffusion of Innovations

Diffusion of innovations is treated usually as social contagion. Similarly 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 passed to their own networks and so on. First slow, then faster and faster, and then slower and slower as the amount of potential adopters decrease. 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, probably that is not the right network, or there is another more powerful effect other than network diffusion[13].

Network Diffusion Models and Statistical Methods

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

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[16]. It has one parameter and has limited applicability. It could be used to compare growth rates for various innovations

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 every one else. Thus, do not measure whether people who are connected to one another engage in the same behaviors. . 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.

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. We have used a quadratic statistic proposed by {Nyblom et. Al} which is a Special case of the Mantel test used by epidemiologists [17] and known by sociologists as QAP (for Quadratic Assignment Procedure) [18]

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.

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

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

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[19]. 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 time 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

$$\log \frac{\Pr(y_t = 1)}{\left(1 - \Pr(y_t = 1)\right)} = \alpha + \sum B_j X_j + \sum B_{kt} X_{k2} + \sum B_{(k+1)} \omega y_t + B_{kt} \omega y_t \qquad \textbf{(2)}$$

Where j are the socio demographic characteristics (xj); Bkt are parameter estimates for the matrix of time-varying socio demographic characteristics (Xkt); w 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 reanalyze of the physician data studied by Coleman et al and by Burt, and show the crucial improvement in the methodology to better asses that particular empirical study.

Empirical Studies on Diffusion of Innovations

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[20]:

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

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

Central: Degree, closeness, betweens, 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[13]:

- High network density[3] will contribute to easily and fast spread of an innovation. Low density (sparse networks) will act in the opposite direction [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. [Relational]
- 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[13]
- 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 association between adoption and exposure[4]. [personal characteristic]
- Lower thresholds indicate personal innovativeness. [personal characteristic]

The last two statements 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 be most likely 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[9, 13, 14].

As mentioned in the introduction, the Annex shows a table with 28 different diffusion studies that we have been able to identify. They include some form of Social Network Analysis, which are being considered as part of the literature review section of the proposed thesis. For the background chapter of this proposal it is worthy to comment on a few but relevant ones.

A very well known study on social networks and diffusion of innovations, is that of Medical Innovations by Coleman, Katz and Menzel, 1966[21]. It has being revisited several times and each time the review provides insights in different aspects of the practice of this emerging field. There is even a paper that does an ethnostatistical analysis of the several reviews[22]. This last paper brings up interesting issues that provide useful insights, considerations and guidance to carefully present the qualitative foundations for quantitative research. By providing context, Kilduff shows that at the time of the study Physicians had being experiencing what was perceived as an external threat. It may have induced a higher level social integration. Coleman et al.'s paper does not report the individual results for each network. They do report that friendship behaves different to advice and professional discussion, and that the latter behave very similar, but fails to explain why they resolve to average all in degree measures. In terms of validity, social integration is a rather general concept and one can recognize validity in the way they measure it. But it may also be the case that being in more meetings also meant not only network exposure but media exposure. In terms of reliability is a different story. These network data is based on self report, and research has shown that self report of network data induces considerable error. Burt, imputes competition in his reanalysis but as pointed out by Kilduff, fails to explain why in his analysis, where he is trying to build an argument for competition as opposed to collaboration, he drops the friendship network from the analysis, which is expected to bring cohesiveness to the analysis. In this case the validity of the measure is to be questioned because deliberatively is leaving out important pieces of information that are relevant for the argument the author is building.

Marsden and Podolny reanalyzed both previous analyses of the physician data set. Two important lessons from their review are the change in the regression technique, to account for the time-base nature of the process that is being modeled, as mentioned before, and showed how sensitive are statistical results to data – imputation choices like those made by Burt. We expect to have access to these well know data and test our ideas and models and explore how our models behave with it and possible ways to contribute in this on going discussion of network effects on DOI.

Literature on the Social Network Approach to DOI

Despite the literature review is still underway; the material that I have covered up to now allows me to address at least some of the salient issues in the study of social networks and diffusion of innovations relevant to my research questions and research design.

Since the early work on Diffusion of Innovations (DOI) [23], the social structure has being acknowledged. Computational models of diffusion of innovations have evolved mainly in the space of system dynamics [24]. This branch of modeling has produced lots of research from which new theoretical propositions have emerged, but in essence 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[25]. 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 [26] and promising platforms for social sensing devices are being designed [27], 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 on the observations ignoring the fundamental dependency that is inherent to social structure[28]. 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 [29].

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. In the other side of the spectrum are 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 [25].

Our 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 use to explore their implications in the flow of ideas and, eventually, other human dynamics. Thus, allowing improved understanding and management of diffusion processes.

Influential people and DOI

Most studies on innovation have been retrospective; they lack information on interpersonal communication networks, and more important, few have attempted to use the lessons from diffusion research to accelerate the diffusion of innovations [15]. Valente and Davis' work [30] 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. The thesis evaluate that the target is better defined as an influential member within multiple advice networks as one of its key contributions. 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.

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[31, 32] and some have been modeled mathematically[33, 34]. The analysis of the interplay of this 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 [35-37]. 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 diffusion of ideas, practices or artifacts[12, 38]. 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. The thesis will extend this models focusing on their ability to supplement missing relational data.

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. Recent developments in this area and MCMC algorithms have being developed which are able to produce Maximum Likelihood estimators. This is important because it allows 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. I am designing an agent –based software that captures multimodal information on the individual, his relations and their context to produce a model of the latent social network, evaluates the fitness of the model and through simulation estimates the expected outcome of a diffusion process.

These recent methodological advances come from scientific work on the study of HIV contagion. As opposed to other contagion process, HIV studies need to focus in the dyadic relationship. The change on the focus of research, from the subject to a pair of subjects and the existence or not of a link, produces different and often counterintuitive results with important impact in health policies. Diffusion of computers in rural areas of the world, among other environments might have something in common with this.

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 [39] or tetracycline [40], as opposed to the ever evolving modern technologies and their myriad of versions and the potential difficulties and complexities intrinsic to them.

Relevant Research Projects

Marketplace of Ideas as a general framework for ICT4D

In rural settings, and especially developing countries, it is often difficult to argue for the benefits of digital technologies and its relation with higher levels of productivity through better coordination, communication and knowledge sharing. Many times there are skeptical reactions from community members, who are usually focused on more traditional methods of economic development, and when it comes to arguing for provision of Internet services the skepticism becomes even greater. The common promises about Internet for the people are:

- the opportunity to do better business through the access to international market prices or even next town prices for their goods;
- the possibility to deal with the local government from home;
- medical assistance provided by a faraway physician on line;

...and so forth. However most villagers do not care about the international prices --- they must take the offered price in local commodities markets because of their small production size. Dealing with local government through the Internet is justifiably seen as just sending an email to an incompetent and unresponsive official in a distant municipality, and not much better than trying to `fix' the problem through the local babu. And as for medical and agricultural advice, that is only useful if the local health and agricultural systems have the correct drugs and trained personnel[41].

People are the content

In the seemingly endless number of studies and position papers concerning the introduction of digital technology into underserved communities, the first recommendation is to obtain more local content. Perhaps the key component of the Marketplace of Ideas proposal is to recognize human contact among the villagers as the main content, at least in the early stages. The goal should be finding easy and efficient ways for them to meet and share with their friends and peers within the community. For instance, having means to distribute simple but crucial information on time, such as "I just found a "x" fungi in my field, check yours to prevent infection" or "I am going to town with my truck half full in about an hour, who wants to use the remaining half and share the cost?" could make a world of difference.

Traditional ICT programs…the village phone or the central telecenters…are useless for communication within the community. Even if every villager has a phone it is not easy to call hundreds with such `opportunistic' messages, and traditional broadcast means such as radio are inefficient because they depend on every member listening at the same time. In contrast, the asynchronous digital media (where the receiver can be view the message whenever convenient) are ideally suited for such communications. Whether by two-way pagers, or SMS messages, or Internet email, simple asynchronous messages have the potential to raise the productivity of all the producers through better coordination of their activities. However, bringing connectivity to a rural community by providing the appropriate infrastructure for low cost communications is not enough. People need support and motivation to understand why it is worth their time and money to acquire these new tools and learn how to use them. Since resources are limited there is a need to seek the most effective and efficient methodologies to introduce these technologies into the community, and to spread their benefits to largest number of individuals within the community.

We argue that in rural areas, the existing social networks are the critical change agents in the community. Introduction of digital communications technologies must be done recognizing these key groups of people and connecting them to each other, allowing them to immediately find the value of the technology in the wealth of new `community communication content'. Our approach therefore begins by exploring the structure of the community to find the networks of influence, motivation and support among the community members.

Ideas as a source of wealth

Marketplace of Ideas is a metaphor to help think of ideas as a good that is produced, imported and shared by community members. Ideas can have both value and cost. They can be registered and protected in the form of patents, trademarks and copyrights or can be given away. Ideas also have opportunity, distribution and production costs. The following diagram describes our socio economic rationale linking social structure, diffusion of ideas and sustainable development.

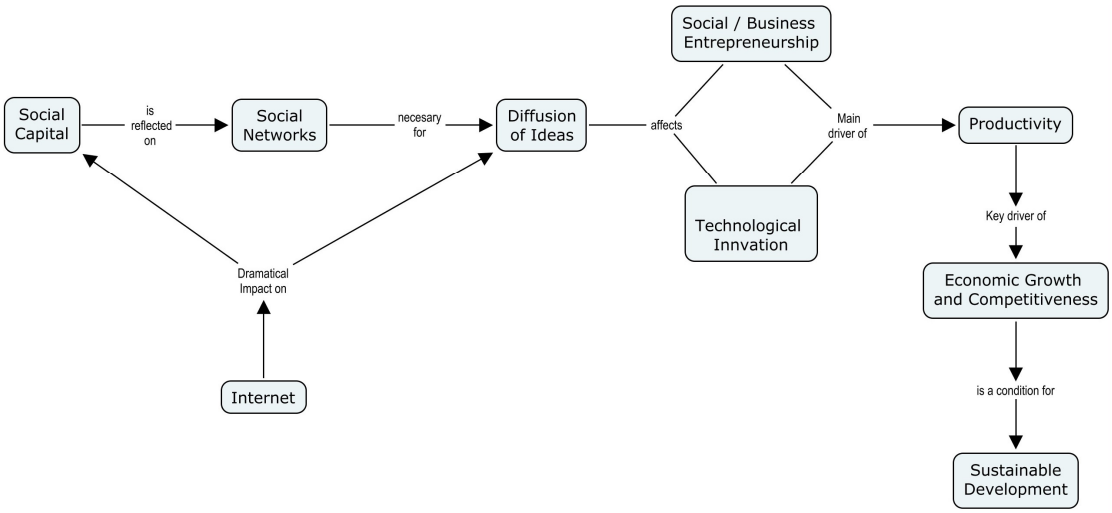


Fig. 1 Concept Map of the socio-economic rationale that links Social structure, diffusion of ideas and sustainable development.

Role of social networks

Identifying who to start with is key for success and finding ways to later reach the highest possible number of community members is an ethical obligation. Social network analysis could play a key role in both. Communities are collections of superimposed social networks build of different community agents or actors and the relations or links among them. This social networks act as the distribution channel for the flow of ideas. The process of an idea flowing through a social network is critical to its usefulness. The flow of ideas from one individual to the community and vice-versa will depend on the relations or links among them. The different channels and patterns will depend on the different levels of influence of different individuals.

Since the 1930´s cognitive and social psychologists have worked on the problems of sociometry and group dynamics, and have develop 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 understand better how ideas flow, how technology can help them spread more efficiently. We can use these methods to explore ways to foster growth within the Marketplace of Ideas as a method to reach the underserved members of the community. Social Network analysis should provide a useful insight on the regularities in the relationships and the key structural properties of the community. This approach should guide an intervention to enhance the abilities of key agents and to provide better or new links. These richer networks should be a vehicle for richer and more significant content for the different categories of actors within the community.

Under this view we developed a set of projects to explore further and evaluate this novel approach to the role of Information and Communications Technologies to foster development in a sustainable way. Those will be described later in this section.

From little intelligent communities to the marketplace of ideas

The MIT Media Laboratory has been experimenting with several different models in different settings with the goal of improving community access to digital technologies. The LINCOS telecenters

are one of these projects and they have provided important and interesting lessons about the dynamics of development in the several dozen different communities where LINCOS units have been tested.

LINCOS stands for "little intelligent communities". It was developed as a joint project with the Costa Rican Foundation for Sustainable Development, with the goal of being a prototype for the 21st century community center. Build on used commercial containers this high-end telecenters provided a mobile solution that can bring first-class educational material, medical advice, business communications, and the arts to every family by use of wireless communications technologies

The LINCOS approach relies mainly on the communities ability to organize it self around the telecenters and the services it provides. Different units in different communities have shown the LINCOS' flexibility and capacity to respond to different needs and situations, and in most of the LINCOS settings services related to education and communications have been of great demand. In terms of hardware, software and pedagogical approach, the LINCOS project has been quite successful. However in most cases sustainability has become a serious problem, as after a few months the organizational weakness of the community it serves starts to reflect on the unit's economic sustainability. To overcome these sustainability issues and to have a higher impact in the community we must also introduce a change in the relationship between the unit and the local business community as well as in the way we introduce these digital technologies in the community.

Fig. 2 LINCOS in San Marcos

How to introduce technologies that improve people's productivity and quality of life in a sustainable way remains as a very important problem to solve and it is in essence the motivational keystone of Marketplace of Ideas as a research theme.

For my thesis I will use data on the diffusion process of home computers in a rural community a few miles away from San Marcos de Tarrazu in Costa Rica, where the Lincos model was developed. Even though Santa Maria de Dota (SMD) is a different community, previous work from Media Lab faculty and former Media Lab and Sloan students within Los Santos Region, paved the way for my work and I am thankful for their efforts as a researcher and as a Costa Rican as well.

Fig. 3 Los Santos Region, Costa Rica

Since 2003, and with support from a local university, I have been gathering and analyzing information on people and organizations in SMD. I have explored different ideas through a series of projects conducted in that locality. This is a community of coffee growers in the southern mountains of Costa Rica called Santa Maria de Dota (Figure 6). The community has roughly 4300 inhabitants; coffee production and exports represent about 80% of their income. SMD 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. Aproximately 700 producers are registered at a local cooperative. These are some statistics that describe this group of producers: A sixth of the producers never completed primary school, 45% completed 6 years of primary education, 12% percent completed high school and 15% attended tertiary education. This meant that literacy is not a barrier in this community and there is a good chance to promote written short communications.

The average age is 48 years, and in the interviews most of them feel that their generation is too old to use computers, but some of them consider it as an opportunity for the youth. As an average, they have two kids living in their household and the average age is 16 years old.

One fifth of their households already have a computer at home that are in occasional use by their kids but no one has access to the Internet. It is important to mention that computers have been in ubiquitous in Costa Rican schools for at least fifteen years and dial up service to the Internet have a moderate cost, so it is surprising to find zero connections in place.

G-Lab, LINCOS and System Dynamics

My first contact with this community as a Media Lab student was a ML/LMF 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"[42]. 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 shot 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 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 being actually 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, 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. 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 will understate the strong intuition I 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. I started to think that "building" this links should be the core product of ICT4D efforts and not a byproduct.

Finding the "influentials": using a qualitative approach as baseline

The objective of initial intervention was to explore the effectiveness of social network information in establishing who the "key people" of the community were. The implicit hypothesis was that sociometric information could be effectively and efficiently be used to find who the influential members of a community are. To test this hypothesis it was necessary to have a "conventional" social sciences approach that will operate as a baseline or "ground truth" for further exploration.

The MIT Media Lab had established a consortium called Digital Nations with sponsors in different parts of the developing world. In the Central American region INCAE, which is the leading business school in Latin America, was the sponsor and an active participant in the San Marcos Project and they agreed to actively participate in this project in the nearby town called Santa Maria. Two local NGO's were also invited: Entebbe who had experience working with the Media Lab and CEMEDCO who had experience working with INCAE and its digital nations' projects. The latter was run at the time by a social psychologist and a social worker, both of them emeritus professors from University of Costa Rica.

Once all the stakeholders and collaborators had agreed that Santa Maria would be a good place to conduct this research, a set of interviews were conducted by the CEMEDCO team which had agreed on conducting an Ethnographic Diagnostic in the community[9, 43]. 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[44]. 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 ad-

dressing those concerns, the different organizations they were involved with, and what studies about the community they knew of. After 6 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 next step was aimed to validation of the original 53 names as influential members. The list was discussed for validation with a group of "community experts", who were identified simply as the most commonly mentioned or by their visibility as leaders. The expert's validation reduced the original list to 30 members.

They were expected to be the most influential members. Among those 30 influential community members, 19 were registered at the local cooperative as producers.

The third and last step to find the most influential members used a basic sociometric instrument. The group of 30 was invited and attended to a workshop sponsored by INCAE, where they completed a 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[45]. 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. In different words, these 19 were the influential amongst the influentials. Table 1 summarizes the three different exercises that lead to the baseline estimation described in this section.

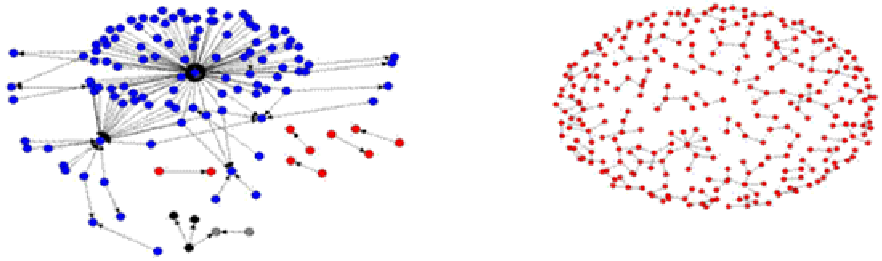
Table 1. The three exercises used to construct the baseline for this study.

Step		Community Members	Subset of Producers
1	Ethnographic Diagnostic	53	32`
2	Experts Validation	30	19
3	Sociometric Survey	19	16

A preliminary exploration of their social network showed a very different structure when they were asked two different questions. When they were ask about people with whom they talk to seek for advise in 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 inexistent 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 by 20%. Again, there was 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 is a very 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', should not be perceived as a threat to their position. This initial results suggested that it would have been a good idea to take social network measurements such as these before designing any intervention in a community.



(a) Production Advice (b) Personal Advice

Fig.4. **Networks of Advice**

Finding the influentials: Quantitative approach

The goal of this study was to explore theoretically and empirically different sociometric measures using the attribute and relational data we had previously collected. The exploration was aimed to find a quantitative model that would predict who the influential people in SMD are and within that model test the effect of the sociometric measures as compared with socioeconomic and demographic variables. The propositions and tests in this study were:

1) Are sociometric measures a good supplement to conventional social and economic status attainment measures in predicting who is influential.
2) Are patterns of advice as captured by structural properties of the advice networks a good predictor of who are the influential members of a community.
3) If 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 use 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

In [9] we explored the use of different centrality measures and proposed a new method to find the influentials based on the eigenvector centrality of the advice networks.

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i p_k) \tag{3}$$

We used the ethnographic study previously described as our "ground truth". After evaluation of a set of variables we developed the model described in Equation 4 to predict who is influential:

$$\log \frac{\Pr(y=1)}{(1-\Pr(y=1))} = \beta_0 + \beta_1 X + \beta_2 C + \beta_3 M + \varepsilon \tag{4}$$

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 and M their use of Media and ε the expected error.

While the conventional way of classifying the influential was extremely efficient with zero type II errors, it produced a false positive rate (type I) equal to 17.92%. But accuracy is not the convenient measure for the results we are interested in classifying. 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.

We also found that 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 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 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 certain threshold. It can also be effectively used as the first step to develop cognitive social structure studies [46].

Table 2. 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

Marketplace of Ideas: seeding the influential members

Data showed that this community's social network is characterized by 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 diffusion of information and communications technology in the town of Santa Maria. The claim here is that by selectively working with the influentials, we should be able to see the evolution of a more rapid diffusion process. Since we originally collected information on who had home computers and Internet access. We seeded the influentials and we expect to measure in the near future what is the current rate of adoption and how it compares with two nearby communities of coffee growers: San Marcos, which 6 years ago experimented with the LINCOS approach and San Pablo where no direct effort to promote ICT4D has been made.

1) **Technology workshop with the influentials**. Those considered influential were invited to a 1 day long seminar at INCAE during the morning they were exposed to some basic ideas on technology and development and during the afternoon they participated in a workshop to learn the basics to setup an email account and to start using it among themselves. No follow up or extra training was provided, expecting that being part of a well interconnected core, we expect that they took it from there and supported each other, as effectively seems to be happening and we hope to measure soon. One of the most influential have kept an advise relationship with Media Lab, and as part of the

Fig. 4 INCAE researcher at ICT workshop with influential community members. Here they are experiencing for the first time the use of e-mail.

methodology we are working on, we keep our contacts to a minimum and only through the "influentials".

2) **Tooling entrepreneurs**. Ten Media Lab used computers were collected and they were shipped to the community. Five were used by president of the women's association (one of the influential) and five were deployed with support of another two of the influential members in five different small convenience stores to be used by them as pleased and we collected information on their ego networks. The first gave rise to a local computer school that is still in place and have long replaced the original machines. The other five had mixed but very interesting results. Five data points do not allow for a correlation analysis but it was clear that the three that succeeded had reported considerably bigger personal networks. This project captured national attention and in 2004 was awarded a public recognition among the most innovative projects of that year.

Figure 5. Women's association computer training school started off with five used computers and today is a self sustained business.

Some evidence of the emergence of sustainability through entrepreneurship

There are several new business, practices and projects that can be listed as a sustainable outcome of Marketplace of Ideas. Some are private small business; others are extensions or change of practices in places where the influentials have direct impact including the local government and NGOs.

SANTOS PC. Without any direct intervention from Marketplace of Ideas, another member of the influential people that we worked with, realized the potential of importing used Pentium III that he were able to sell in $100. It is now an on going enterprise and a little kid that at some point was an emblematic image of the LINCOS project, is today a talented teenager working part time in charge of "IT support". Even though this is not a core hypothesis of the thesis, since it needs further exploration, these kind of small business are of worthy of pointing out as they seem to have emerged organically from the key social networks of the community and its new entrepreneurial members. The thesis should provide further definition of the new concept of developmental entrepreneurship, but at this point some salient characteristics seem to be present:

Figure 6. Santos PC

- They improve 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 "new comers" bridged their original social networks

The thesis will use mini-case studies to document some of these initiatives and possible case leads for others to explore.

Next Steps

Data Collection

With the help of INCAE, the local cooperative and people from the community a data base is in place with attribute information on most producers. There are also disparate amounts of data on over 1500 people from the community, mostly from public sources gather by INCAE and other local organizations. For this thesis the design of a sociometric survey is underway and should be ready by the end of 2006 as January represents a unique window of opportunity to collect data and interview people since most of the producers will be visiting the Cooperative since it will be the time of the year when they bring their coffee cherries to be processed. The four goals of the up coming field work are to find from different available sources (which already have being contated):

- Information to reduce as much as possible missing data on SMD producers.
- Information on home computer ownership in SMD, San Marcos and Leon Cortes producers.
- Geographic references to allow the exploration of geographical information in the context of this thesis.
- Material to support the mini-case studies mentioned above.

System prototype and simulation

A working prototype of the system is currently underway. Training data to test the algorithms that detect network effects on diffusion were collected and scripts have being tested. At the moment I am working with one UROP redesigning the GUI and working on the algorithms to model agent exchanges under network constrains and eventually introducing a small set of communication rules based in the mathematical modeling of some relevant social theories that have been used for studying communication networks[31, 32].

Timeline

	2007						
	Dec	Jan	Feb	Mar	Apr	May	Jun
Survey and field work preparation							
Model Development (network dependent diffusion simulation module)							
Data Collection							
Analysis							
Defendable Draft							
Defense							
Submit Final Copy of Dissertation							
Graduation							

Resources Required

Required resources are: UROP Assistance, Digital Video Recorder, travel expenses for field work

Bios

Juan Carlos Barahona

Originally from Costa Rica, he is a doctoral is a PhD student in the Human Dynamics Group at MIT Media Laboratory. His research is focused on the study of social networks and their role in diffusion of technologies, as relevant for entrepreneurial activity and socio-economic growth. He has served public positions as an advisor to the Minister of Education in Costa Rica. He has been consultant and advisor to the governments and companies in several Latin American countries on strategy and information systems design and development. He is author of case studies, papers, books and book chapters in topics of Competitiveness and Sustainable Development such as Education, Customs Reform and Information Technologies for Development.

Prior to Media Lab he was Deputy Director of the Latin American Centre for Competitiveness and Sustainable Development and has also worked on Sustainable Development with World Bank. He was the project leader for customs reform in Central America as part of a joint research-action initiative of the Centre for International Development at Harvard University, Centre of Strategy and Competition at the Harvard Business School, and INCAE. Prior to his academic career at INCAE he was founding partner of a software start up and a consulting firm.

PhD Candidate from Massachusetts Institute of Technology, MBA from INCAE and Bachelors in Business Administration and a major in Computer Science from UIA.

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. Barahona, J.C. and A.S. Pentland, *Advice and Influence*, in *Media Lab Technical Notes*. 2005, MIT Media Lab.
10. Robins, G. and P.E. Pattison, *Multiple Networks in Organizations*. 2006, Deparment of Psychology, School of Behavioral Science, University of Melbourne: Melbourne, Australia.
11. Hornik, R., *Some Reflections on Diffusion Theory and the Role of Everett Rogers.* Journal of Health Communication, 2004. **9**: p. 143-148.
12. Young, H.P., *The Diffusion of Innovations in Social Networks*. 2002, Center on Social and Economic Dynamics, The Brooking Institution & Santa Fe Institute.
13. 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.
14. Weimann, G., *The influentials: people who influence people*. 1994, Albany: State University of New York Press. xiv, 370 p.
15. Valente, T.W., *Network models of the diffusion of innovations*. Quantitative methods in communication. 1995, Cresskill, N.J.: Hampton Press. xiii, 171 p.
16. 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.
17. Mantel, N., *The detection of disease clustering and a generalized regression approach.* Cancer Research, 1967. **2**: p. 209-20.
18. 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.
19. Allison, P.D., *Event history analysis: regression for longitudinal event data*. 1984, Beverly Hills, Calif.: Sage Publications. 87 p.
20. 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.
21. Coleman, J.S., et al., *Medical innovation; a diffusion study*. 1966, Indianapolis: Bobbs-Merrill Co. xxi, 246 p.
22. Kilduff, M. and H. Oh, *Deconstructing Diffusion: An Ethnostatistical Examination.* Organizational Research Methods.
23. Rogers, E.M., *Diffusion of innovations*. 3rd ed. 1983, New York London: Free Press;
Collier Macmillan. xix, 453 p.
24. Sterman, J., *Business dynamics: systems thinking and modeling for a complex world*. 2000, Boston: Irwin/McGraw-Hill. xxvi, 982 p.
25. Morris, M., *Overview of Network Survey Designs*, in *Network Epidemiology*, M. Morris, Editor. 2004, Oxford University Press: Oxford.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.

32. Monge, P.R. and N.S. Contractor, *Theories of Communication Networks*. 2003, New York: Oxford University Press.
33. Gottman, J.M., et al., *The Mathematics of Marriage: Dynamic Nonlinear Models*. Bradford Books. 2002, Cambridge, Massachusetts: The MIT Press.
34. Hunt, E., *The Mathematics of Behavior*. First Edition ed. 2007, Cambridge: Cambridge University Press. 346.
35. 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.
36. 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.
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. Ryan, R. and N. Gross, *The Diffusion of Hybrid Seed Corn in Two Iowa Communities.* Rural Sociology, 1943. **8**(1): p. 15-24.
40. Coleman, J.S., E. Katz, and H. Menzel, *Medical Innovation: A Diffusion Study. New York: Bobbs Merrill.* 1966.
41. Barahona, J.C. and A.S. Pentland, *Marketplace of Ideas: The Role of Digital Technologies in Rural Settings.* Information for Development, 2003.
42. Pepper, S. *LFM G-Lab: Students Spend Spring Break Around the World*. [Internet Site] 2003 [cited 2006; News Article].
43. 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.
44. Goodman, L.A., *Snowball Sampling*. The Annals of Mathematical Statistics, 1961. **32**(1): p. 148-170.
45. Freeman, L.C., *Centrality in Social Networks Conceptual Clarification.* Social Networks, 1979. **1**: p. 215-239.
46. 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.

Investigators, Research Design	Description and obseved population	Type of Data (egocentric, whole network, sample)	Type of Network or networks	Question asked?	Dependend Variable	Latent Variable	Observed Explanatory variables	Source of Data (subject reports on self or on self and peers?)	Types of ties	Diffusion Model	Research Methods used	Sociometric Measurement	Description of Measurement	Validity of Measurement	Reliability of Measurement	Strong Assumptions	Results	Source of information on research either primary or secundary
Ryan and Gross, 1949	Hybrid seed corn adoption by																	
Coleman, Katz and Menzel, 1966	doctor's willingness to prescribe tetracycline in 4 Illinois towns. This work is a mile stone because was the first to especifically ask who talks to whom about an innovation	sample. 125 self-employed general practitioners, internists and pediatricians were interviewed. They represent 85% of the doctors practicing in relevant fields to the innovation in the four Midwestern cities surveyed.	advice, discussion frequency of contact among friends within the proffesion.	To whom did he most often turn for advice and information? With whom did he most often discussed his cases in the course of an ordinary week? Who were the friends, among his collegues, whom he saw most often socially?	First month the drug was used by a physician. Time of adoption, monitored at the local pharmacies.	Integration into the local medical community	staff priviledges at city hospitals, amount of hospital meetings attended, share office, many or few friends, advisors, discussion partener among fellow physicians. Many individual attributes of each doctor obtained by interview.	Prescriptions of the general practitioners who ran private practices were tracked over a 16 month period.			Indegrees in the friendship network. Other two networks are reported to yield similar findings. Average of the three indegree measurements					Attributes were a predictor of adoption. Especially the classification between proffesion oriented and patient oriented. Less integraded doctors were laggards in the process of adoption.	Original Paper
Rogers, PhD	Weed Spray in Iowa																	
	Diffusion of Modern Math Interviews of education superintendents in Pensylvania in the 50's. 38 out of sixty eight superintendents interviewed		Friendship	Interview question: who are your three best friends among education superintends?					Frienship ties. Symetrized. Subjects with 0 indegree are treated as isolates.					Education Chiefs are the gatekeepers or decision makers.				
Beal and Bohlen,																		
Katz, Levine and																		
Rogers and Kincaid,	Korean Family Planning																	
Entwisle and others,	Contraceptive adoption in																	
Dozier, 1977	Korean Family Planning																	
Kohler, 1997	Korean Family Planning																	
Montgomery and	Korean Family Planning																	
Burt, 1987	Reanalyses Coleman et al.'s Medical Innovations Study	sample. 125 self-employed general practitioners, internists and pediatricians were interviewed. They represent 85% of the doctors practicing in relevant fields to the innovation in the four Midwestern cities surveyed.	advice, discussion frequency of contact among friends within the proffesion.	To whom did he most often turn for advice and information? With whom did he most often discussed his cases in the course of an ordinary week? Who were the friends, among his collegues, whom he saw most often socially?		Competition between similarly situated physicians	Structural Equivalence					Focus on structural equivalence to analyse the role of competition between alter and ego.					Rejects original conclusion that claims that the more integrated physicians will innovate sooner.	
Marsden and Podolny, 1990	Reanalyses Coleman et al.'s Medical Innovations Study							Coleman et al. 1966 dataset on medical			Event History Analysis						Network exposure was not related with adoption	
Blume,1993	Stochastic Approacho																	
Ellison, 1993	Stochastic Approacho																	
Strang and Tuma, 1993	Revisited Coleman et al. 1966 and Marsden and Podolny							Coleman et al. 1966 dataset on medical			Postulated time variance in						They found evidence of contagion	
Rogers, 1995																		
Valente, 1995	Korean Family Planning																	
Valente, 1996	Korean Family Planning																	
Valente, 1997	Cameron women in voluntary organizations		Friends					Women were ask to name their friends to determine association with contraceptive use	Frienship ties. Symetrized. Subjects with 0 indegree are treated as isolates.									
Valente and Saba,	Contraceptive									Threshold Model	Controlled for							
Van den Bulte and Lilien, 2001	Revisited Coleman et al. 1966 and Marsden and Podolny and Strang and Tuma, 1993										Controlled for media exposition						Network Contagion Dissapeared	
Casterline, 2001	Reproductive Health																	
Neaigus et al., 2001	Substance Abuse																	
Bearman, Jones and Udry, 2000	Adolecents health data related with smoking behavior																	
Montgomery, 2001	Contraceptives in Ghana. Social influence on social behavior.	egocentric data									Controlled for tie characteristics						Strongly Associated with peers adoption in network	
	Adolecents health data related with smoking behavior										The influence of peers in smoking, while conditioning on the smoking						Event History Analysis	
Young, 2002	Stochastic Approacho							Artifitial Data	Dyadic Relations. Weighted Directed Edges Symetric representing influence		Mathematical demonstration	r-Close-knittedness	Ratio of internal degree to total degree. Measures the vulnerability of a set to unravelling			no isolates		
Bearman, Kholer and Watkins 2003	Kenya	egocentric data / longitudinal																
Boulay and Valente,	Contraceptive adoption in Nepal																	

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 Community. It uses over 74,000 words to describe the state of the art and new directions for making ICT work for the poor" and the word diffusion is not even used once in the whole document.

2 The critical mass is the minimum number of adopters needed to sustain a diffusion process. Criticality is achieved in the first second-order inflection point of the S-curve.

3 Proportion of existing to maximum possible links.

4 The amount of exposure than an individual needs to adopt is call the individuals threshold.