Discover Implicit Relations Between Individuals in a Community based on a Newsletter Tracking System

Tiago Henrique Lopes Ferreira

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Examination Committee

Chairperson: Prof. Ernesto José Marques Morgado
Supervisor: Prof. Alberto Manuel Rodrigues da Silva
Members of the Committee: Prof. Mário Jorge Costa Gaspar da Silva

May 2013

Discover Implicit Relations Between Individuals in a Community based on a Newsletter Tracking System

Tiago Henrique Lopes Ferreira

TÉCNICO LISBOA

Abstract

In the last twenty years, Internet has changed the way people communicate and interact with each other. Thanks to the online technology, people are able to overcome geographical limitations and share common interests in a community base. The concept of "Online Communities" brought new ways to explore members' connections and interactions.

However, when the number of interactions was manageable in the community's face-to-face manner, the same was not true for online communities. The number grows at a pace that is very hard to control and, as a consequence, connections between members get lost or forgotten and communities lose their chance to perceive members' relationships.

It is on this gap that the "Newsletter Tracking System" (NTS) comes as a tool that allows communities to discover relationships between members. Due to the members' interactions with newsletters, NTS brings to light implicit connections between them and fosters an implicit community. In addition, it uses clustering algorithms to allow to better understand how members relate to each other and how connections can be explored.

This research also proposes a "Connection Degree" model (CD) to measure the connections' strength among members. The CD complements the discovery of implicit connections by offering a way to measure the importance of each connection in the network.

NTS and CD were developed and evaluated within the Nano-Tera.ch scientific community and, at the end, the results show that implicit communities can be discovered and blended with real communities to better organize their members, share knowledge, and promote teamwork.

Keywords

Community

Newsletter Tracking System

Connection Degree

Implicit Connection

Explicit Connection

Resumo

Nos últimos 20 anos, a Internet mudou a forma como as pessoas comunicam e interagem entre si. Com a tecnologia *on-line*, estas são agora capazes de superar as limitações geográficas e partilhar interesses comuns em comunidade. Por sua vez, o conceito de "*Online Communities*" (Comunidades *On-line*) trouxe novas formas de explorar as relações entre membros de uma comunidade.

No entanto, enquanto o número de relações entre membros era possível de gerir numa comunicação cara-a-cara, o mesmo não é verdade quando a *Internet* se transforma no principal meio de comunicação. O número de interações e relações cresce a um ritmo rápido e difícil de gerir manualmente. Como consequência, as relações entre membros perdem-se ou ficam esquecidas, e as comunidades perdem a oportunidade de perceber e explorar as mesmas.

É nesta lacuna que o "Newsletter Tracking System" (NTS) – Sistema de Rastreamento de *Newsletters* - surge como uma ferramenta capaz de permitir às comunidades capturarem as relações entre os seus membros. Através das interações com as *newsletters*, o NTS permite a descoberta de relações implícitas entre os diferentes membros e a conceção de uma comunidade implícita. Além disso, o NTS utiliza algoritmos de *clustering* para permitir que as comunidades percebem melhor as relações entre os seus membros

Esta dissertação propõe ainda o modelo "*Connection Degree*" (CD) – Grau de Ligação – para medir a força das relações entre os membros. O modelo complementa a descoberta das relações implícitas oferecendo uma forma de medir a importância das mesmas na rede.

O NTS e o CD foram desenvolvidos e avaliados dentro da comunidade científica Nano-Tera.ch. No final, os resultados mostram que a descoberta de relações implícitas é uma vantagem para as comunidades reais organizarem melhor seus membros, partilharem conhecimento e promoverem trabalho em equipa.

Palavras-chave

Comunidade

Newsletter Tracking System

Connection Degree

Relação Implícita

Relação Explícita

Table of Contents

List of Figures

List of Tables

List of Equations

1. Introduction

This dissertation presents a research that looks for discovering implicit relationships between members in a community. Introducing the “Newsletter Tracking System” that gathers members’ information and proposes a “Connection Degree” model that allows to define the grain of connection between members and improve the quality of community relations.

1.1. Context

This research was developed in cooperation with INESC-ID Lisbon, École Polytechnique Fédérale de Lausanne (EPFL), and Nano-Tera.ch scientific community, through a common interest on social networks and community exploitation. The research focuses on social interaction and its value as a way to gather members’ information inside a community.

The study on social networks in the Internet started in 2005 when Tim O’Reilly first brought the concept of “Web 2.0” (O’Reilly, 2007). For the first time in History the power of sharing was being moved from those who were technology experts to all who have access to the internet. Web 2.0 emerged as a solution for developers, companies, and users with or without computer expertise. Along with its new way of sharing content, Web 2.0 moved the power of interaction from its physical mode to the cyber space where people can virtually interact and establish new connections overcoming geographical constrains. Also communities had the chance to become virtually and use virtual technology to collaborate.

The Internet has become the main chain providing to people the infrastructure to cooperate, coordinate, and change the way they communicate and relate to each other (Harley & Blismas, 2010). Interactions become a vast web of virtual connections growing at an incredible pace. Whereas conventional interactions had physical constraints, non-conventional interactions were able to eradicate physical barriers and bring people together through the web. When the number of physical interactions was limited, the notion of being virtually restricted is far from being reached. Any web user can create his own content and build his network of connections with not geographical boundaries. In fact, the number of connections can rapidly rise from dozens to hundreds, and an individual that was firstly connected to a small set of individuals can now be connected to anyone in the World. The communities moved virtually and with them the concept of “Online Communities” came to life in 1993 (Zaphiris & Ang, 2009).

Research has shown how powerful online communities can be and how enriching they are regarding members’ interaction. Technology brought people together as well as their own interests, curiosities, hobbies, professions, and so on. Each time people go online they become incredibly exposed to the virtual engine which is able to capture individuals’ information and add it to their online profile. Moreover, along with their tracked path, the interactions are stored as connections between individuals and something or someone. Connections can then be expressed as explicit connections, if they are fully and clearly expressed by individuals (e.g. home address), or implicit connections, every time they are not expressed (e.g. based on individuals' common interests). Both connections are important to understand individuals' relationships and the community as a whole system. While, explicit connections ensure the initial set of relations between individuals, implicit connections are crucial to help the knowledge on connections to evolve and go beyond what individuals express.

The exploitation of virtual technology brought communities the chance to discover hidden relationships among members and to explore those connections in order to improve their knowledge on the way members related to each other.

1.2. Motivation

The ability to communicate thought the web powered individuals with a wide network of contacts. By moving individuals' connections to the virtual bubble, technology can be used to study individuals' connections and understand their behavior. Communities can study their members' interactions in a way that it will help them to better understand and respond to their needs and improve community's performance.

This research brought the chance to study the actual subject of online communities while helping a real community to understand their members' interests and relationships. Nano-Tera.ch is a scientific community at EPFL, University of Lausanne, Switzerland, and it was looking for a way to study their members' connections and have an overview of its scientific community organization. The notion that individuals inside the community were connected was clear but it was left the notion of how those relationships were made and how relevant they were for the community. Although Nano-Tera.ch had clear information on members' explicit connections it was missing a way to improve their knowledge on members' relationships and have a stronger perception of their network of interactions.

Unlike explicit connections, which are limited to what individuals express, implicit connections are based on individuals' behavior and can help communities to discover hidden relationships and better understand how members relate to each other and how those connections can be exploited. Particularly in scientific communities, the way members are connected is important to discover new ways to relate researches and improve scientific work.

Having the ability to help a scientific community to explore their researchers' connections while developing a real use web application helped to follow the idea and cooperate with Nano-Tera.ch. In addition, the project was planned to be implemented at the Computer Science School at EPFL with a multicultural and multidisciplinary team. A major motivation was linked to the fact that Nano-Tera.ch is a community spread around the world where connections cross continents.

1.3. Problem Statement

Research suggests that visualizing explicit and implicit connections to discover connections between individuals can enable the viewer "to find and understand connections between bits of information" (Dörk, Carpendale, & Williamson, 2011). However, the problem arises when it comes to capture and analyze individuals' implicit connections at the rate they grow: data management becomes difficult to handle and implicit connections harder to find. As a consequence, “the connection between individuals, groups, and information becomes lost, or forgotten, and individuals and groups become more isolated” (McArthur & Bruza, 2003a). Moreover, the more members a community has, the higher the number of interactions will be regarding implicit behavior.

While explicit connections, such as “friend requests” on Facebook, can be easily gathered due to individuals’ actions, implicit connections depend on individuals’ willingness to share interests and interact with common social objects, such as Blog (Barão & Rodrigues da Silva, 2012), Learning Objects (Dinis & Rodrigues da Silva, 2009; Rodrigues da Silva, 2011), or a Newsletter (T. Lopes Ferreira, Rodrigues da Silva, Bradley, Nitesh, & Agrawal, 2011; T. Lopes Ferreira & Rodrigues da Silva, 2012; T. Lopes Ferreira & Rodrigues da Silva, 2012). Individuals do not need to explicitly express their relations but actively interact with social objects and have the freedom to do it their way. This option gives space to interactions to grow and brings the need to make the gathering process as automatic as possible in order to avoid connections loss. The problem with Nano-Tera.ch was the creation of this automatic process in a way that it could be adjusted to their needs and used to understand their community’s interactions based on online newsletters. In addition, the need for an open source tool become part of the problem once the available tools such as, Atomic Email Tracker (Massmailsoftware, 2013) and GoTrackNow (Gotracknow, 2013), where commercial and do not fulfilled the requirements of Nano-Tera.ch.

1.4. Proposed Solution

This research proposes a solution to automatically capture members’ implicit connections and presents a model to measure the connection strength between them. The “Newsletter Tracking System” (NTS), the proposal solution, is based on newsletters to capture members’ interactions and discover implicit relationships among them. (The system uses electronic mail technology and PHP to send and collect information on members’ interactions, and JpGraph library to build visualizations of the gathered connections).

The “Connection Degree” model (CD) lies on the NTS’ discovered connections to measure the strength of members’ relationships. It describes a proposed model that takes into account both explicit and implicit connections to calculate a connection degree for each pair of members. CD is able to bring a value to every community’s connection by placing all members’ interactions at the same level and applying a common model of evaluation. The model enables communities to discover which connections are the most crucial (i.e. have the highest CD values).

Both NTS and CD were developed to help organizations to embrace social technologies as a medium to enrich their knowledge on members and better understand how its community is organized. With a better overview of members’ relationships, a community is able to improve their vision as a network of connections, both inside (how members are related to each other inside the community) and outside (how members relate to individuals outside the community).

1.5. Contributions

During this research several papers were published, namely:

- Lopes Ferreira, T., Rodrigues da Silva, A., Bradley, P., Nitesh, K., & Agrawal, M. (2011). Discover Implicit Relationships Between Researchers UsingEmail Tracking. *CAPSI'2011*.
- Lopes Ferreira, T., & Rodrigues da Silva, A. (2012). Improving the Quality of the Community Relations Knowledge Using Implicit Connections. *2012 Eighth International Conference on the Quality of Information and Communications Technology*, 327–332.
- Lopes Ferreira, T., & Rodrigues da Silva, A. (2012). Foster an Implicit Community Based on a Newsletter Tracking System. In R. Meersman, H. Panetto, T. Dillon, S. Rinderle-Ma, P. Dadam, X. Zhou, S. Pearson, et al. (Eds.), *On the Move to Meaningful Internet Systems: OTM 2012* (Vol. 7565, pp. 398–415). Berlin, Heidelberg: Springer Berlin Heidelberg.

The research also consisted on the development of an open source tool – the "Newsletter Tracking System"[1] – that enables communities to track members' interactions on newsletters without any associated costs along with the freedom to edit the source code. Moreover, the proposed "Connection Degree" model allows communities to compare implicit connections by assigning them a value that represents their strength on the network.

1.6. Results

This thesis produced the following results:

- **Newsletter Tracking System**: A web-based solution that allows the upload and send of HTML newsletters to a set of electronic mail contacts. It also gathers contacts' interactions with newsletters' content by tracking the links on them. At the end, the solution creates an implicit network, exposes the results based on visuals (e.g. charts), and allows to download the data for further analysis.
- **Connection Degree**: It is a model based on individuals' interactions with newsletters and explicit preferences on categories. It assigns to each edge a value representing its strength in the discovered set. For each implicit connection between every two individuals, the CD model calculates its connection degree, meaning that the higher the degree the stronger the connection.
- **Analysis on Nano-Tera.ch**: Both proposed solutions – NTS and CD – were evaluated at Nano-Tera.ch community and the results are further presented on this research. It is interesting to understand how the application of both solutions can improve a community's knowledge on their members and to which conclusions a community can reach through the ability of gathering individuals' interactions.

[1] The developed application can be tested and explored through a free online server in which it is deployed. The link to access the NTS is http://ist-tf-nts.net23.net/ with the username "nts-test" and the password "test-nts" to log in.

1.7. Dissertation Structure

This dissertation document is structured in 7 chapters. Chapter 2 focus on the background research and it starts by presenting the key concepts on Web 2.0 and how its development was crucial for individuals to interact online. It goes deeper in a community context and it studies interaction as the core of human activity and the emerge of online communities as triggers for individuals to interact through the web. By last if focus on scientific communities and in the two ways in which members can create connections, namely explicit or implicit activities. On the NTS and the CD backgrounds it focus on the previous approaches and how they relate to each other as well as which abilities they enable communities to perform.

In the next chapter, Chapter 3, both the problem and the proposal are presented. The chapter starts by introducing the goal and goes deeper into the problem, namely its data issues regarding collection, storage, process, and visualization. In a second part it introduces the proposed solutions – NTS and CD. The NTS as the solution to gather data on individuals' preferences and to foster a network based on implicit connections, and the CD as the model to bring to light the strength of each implicit connection regarding the discovered set.

The follow chapter, Chapter 4, describes in detail the "Newsletter Tracking System" (NTS) and how its process of upload, send, track, and analyze work. Which data communities need to have in order to start the NTS engine and which data is tracked in order to allow a deeper analysis on members' implicit connections. It also introduces the Girvan and Newman's divisive algorithm that was used to divide members into clusters according to their interactions with the newsletters.

Chapter 5 outlines the "Connection Degree" model (CD) and describes in detail how the model works based on members' explicit and implicit activity. It introduces how both explicit and implicit degrees are calculated in the final CD equation and how both solutions – NTS and CD – related to each other, mainly how solutions complement and add value to each other.

The evaluation chapter, Chapter 6, is based on the evaluation performed at Nano-Tera.ch community. It starts by introducing Nano-Tera.ch community and then presents the analysis based on the NTS and CD results. At the end, the results are discussed and the conclusions (reached at Nano-Tera.ch) are shared in order to show the value of both solutions.

The last chapter, Chapter 7, presents the thesis conclusions and main ideas along with its main contributions and future work.

2. Background

This chapter describes the background research for this dissertation to better prepare and understand the path followed during the system development. It starts by an introduction of Web 2.0 as a platform and how it changed the way people interact to each other. It explores the concept of online communities with the main focus on online interactions (e.g. how people interact online) and scientific communities (e.g. how scientific communities adopted online technologies as productive tools).

On a second part, this chapter introduces the background regarding the developed system (“Newsletter Tracking System”) and the use of explicit and implicit connections. It describes existing solutions and their problems as well as a brief analysis on explicit and implicit connections regarding peoples’ interactions. In the end, it focus on the “Connection Degree” model and its previous approaches.

2.1. Communities

Thanks to Web 2.0, which now allows anyone to be a content generator, a new generation of web-based communities was born. A generation represented by its innovation in terms of user interaction and information sharing (Isaías et al., 2009). Although the individuals publishing content in the early 1990s where those who were experts in computing, Web 2.0 empowers any individual to create and manage their own content reducing their dependency on computer professionals (Zaphiris & Ang, 2009). Adding to this, the progress of Web technology brought new technology enablers such as online social networks (e.g. LinkedIn), social computing developments tools (e.g. wikis), documents sharing systems (e.g. email), and other collaborative environments (e.g. teleconferencing) that changed the way people interact and share knowledge (Heer & Boyd, 2005).

2.1.1. Online Interaction

Web 2.0 places users at the core of its success and empowers them with the ability to create and manage their own content. The key aspect of interaction offers users the chance to interact with online content and with each other. The information becomes fluid rather than static (B. Gibson, 2007). People can virtually communicate exceeding physical barriers and motivating online cooperation and interaction. “It’s a story about community and collaboration on a scale never seen before” (Grossman, 2006). Internet is the biggest channel of knowledge with millions of Wikipedia articles and YouTube videos based on people’s collaboration and commitment.

As more and more social interactions take place online, people will develop new implicit ways for online social behavior (Farnham, Chesley, McGhee, Kawal, & Landau, 2000). Electronic mail, teleconferencing, wikis, and blogs are just some of the new ways that Internet brought, which changed the way people interact. However, traditional approaches to research tend to ignore the importance of interaction in shaping human behavior (Zaphiris & Ang, 2009). Using a symbolic interactionist framework, we see that interaction is defined as the reciprocal actions of two or more actors within a given context (Vrasidas & Glass, 2002). Interaction is an ongoing process that lies in a context and also creates context, which is important to understand human activity and behavior, and therefore study moment-to-moment events that may lead to further interaction among people (Zaphiris & Ang, 2009). The following framework, Figure 1, places interaction at the center of human activity (Vrasidas & Zembylas, 2004).

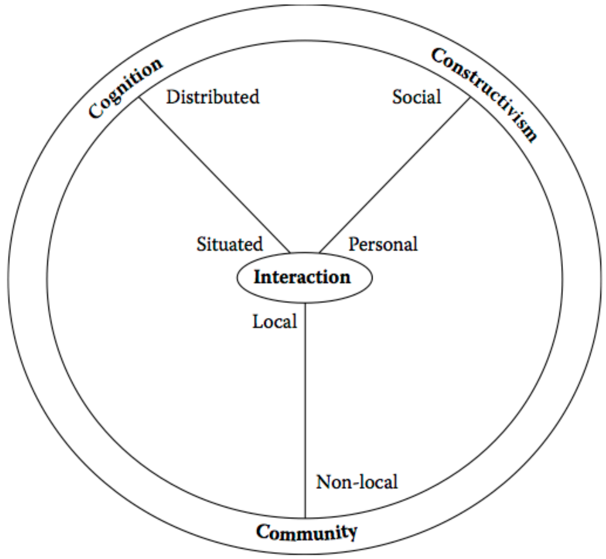


Figure 1. Interaction at the center of human activity

The framework places interaction in the middle of a three-dimensional space, where cognition, constructivism, and community, have a crucial role. Cognition, is described as "situated" when knowledge is constructed during a situation, i.e. is part of the learning processes (Brown, Collins, & Duguid, 1989), and "distributed" when knowledge is among online participants and their physical and socio-political worlds. Constructivism is related to the way the knowledge is constructed (Vrasidas, 1995), and it is classified as "personal" constructivism, when it is constructed in the head of the learner, and "social" or sociocultural constructivism, when it is built in "communities of practice" through social interaction (Brown et al., 1989). On the other hand, the community dimensional is described as "local", when interaction occurs in a local environment significant place of organizing and coordinating social life, and "non-local" when communities and individuals are physically apart from each other and use technologies of communication and cooperation to facilitated the participation (Zaphiris & Ang, 2009).

This analysis on interaction allows the understanding on how interaction occurs, and in what dimensions it should be addressed in order to achieve high levels of interactions from individuals. Although they are placed in the core of Web 2.0, the most effective way to gather individuals' connections and build a network of nodes comes by placing interaction as the core of their activities, i.e. by encouraging individuals to interact with online content and help the network of links to be as accurate as possible.

2.1.2. Online Communities

In the last years, the growth of the Internet has helped the rapid emergence of online communities (Li, Zeng, Mao, & Wang, 2008). By providing users with the ability to interact with each other, Internet allows users to build communities based on shared goals, interests, needs, or activities (Whittaker, Isaacs, & O'Day, 1997). A community can be defined as a group of people interacting in a virtual environment, with a common purpose, guided by a set of norms and policies, and supported by technology (Preece, 2001). In fact, individuals that belong to the same community are able to share resources and provide information, support, and services with each other in a known context of social connections.

In 1993, the concept of "Online Communities" came to live and with it the research about the relation between communities and the Internet (Zaphiris & Ang, 2009). Its understanding depends on one's perceptive and definition, but certainly that the question on "community", with all its affective and historical complications, will continue to frame popular understanding of Facebook, MySpace, and other social network sites (boyd, 2006). Facebook, for example, proclaims in its login page as the way that "helps you connect and share with the people in your life" (Facebook, 2012) by building your own network of connections and bringing to light the concept of community. Myspace, although it changed its core market to music fans, it describes itself as a place to "find old friends" and "make new friends" based on peoples music interests. It describes as a place to "connect", and to create your our music "community" (MySpace, 2012).

Although the concept of online communities is not new, the discussion around the topic continues to reverberate in academic discourse (Wellman & Gulia, 1999). Some authors reject the concept as a "confused oxymoron" (Lockard, 1997), while others see it in terms of network of personal relationships (Rheingold, 2000; Wellman & Gulia, 1999), and bring the notion of "social capital" that comes from people's interaction and participation on virtual communities (Ellison, Steinfield, & Lampe, 2007). However, for many authors (Chua, 2009; Fogel & Nehmad, 2009; Sohn, 2008) the community metaphor appears to be accepted with no criticism and used without further elaboration.

The concept of community has an unsettled intellectual history dating back nearly 200 years (Parks, 2010). It evokes feelings of friendliness, trust, and belonging that are often deemed lacking in ruthless, individualistic times (Bauman, 2001). The nostalgic about online communities has a deep root in its traditional concept. Rheingold, for example, speculated that its popularity is a "response to the hunger for community that has followed the disintegration of traditional communities" (Rheingold, 1993). In fact, the conceptualization of community may have "strong" and "weak" requirements. "Strong" conceptualization is usually applied to groups of people who share physical space, that are self-sufficient within that space, and which ties include their physical connections (Kinton, 1975; Weinreich, 1997). On the other hand, "weak" conceptualization emerges when the community is viewed as a culture, a set of ideas and interpersonal feelings rather than a physical space (Anderson, 2006; Bender, 1986; Calhoun, 1980). Within this framework, "online communities" are defined as "social groups that display the psychological and cultural qualities of strong community without physical proximity" (Willson, 2006).

On the same basis, Parks (Parks, 2010) summarizes the elements that define a community (Table 1). The first two requirements, sharing geographic space and self-sufficiency, represent the traditional "strong" community representation and are those which are less important when leading with communities that are mostly virtual in which elements do not share any physical space and base their interactions and communication through the web. On the other hand, the ability to engage on collective action is generally considered to be an essential test to the authenticity of any community, including virtual communities (Jones, 1998). Also, for a community to be sustained, it must engage in information-sharing rituals on a regular basis (Willson, 2006).

Table 1. Definition of community requirements

Defining requirements of a community
Less relevant for virtual communities • Sharing geographic space • Self-sufficiency
More relevant for virtual communities • Ability to engage in collective action • Shared rituals, social regulation • Patterned interaction among members • Identification, a sense of belonging and attachment • Self-awareness of being a community

A community is also made of patterns of interaction among members that grow from a regularized information sharing (Carey, 1992; Jones, 1998). Members must be able to share information among them and contribute to each other with content that is relevant and interesting for the community's purpose. From this information sharing and content distribution there are patterns of interaction that emerge and can be useful to understand community's structure and individuals' connections. Finally, definitions of community specify that members exhibit attachments to one another and to the community itself (Kanter, 1972; Willson, 2006). These emotional bonds do need to be experienced by all the community members but by the majority which create and have personal attachments to at least some other members. "Communities are defined as shared, close, and intimate" (Jensen, 1990).

The last requirement is about each member be self-awareness of being in a community. Although the concept of community implies the existence of more than one member, it only can be applied when the majority of the members are self-aware that they belong to a community and know what being a community means. A group might qualify as a online community if its members engaged in collection action, shared rituals, variety of relational linkages, and are emotionally bonded to each other, in a way that conferred a sense of belonging to group to which they identify themselves with (Parks, 2010).

Internet allowed many communities to move from the physical world to the virtual environment, from offline to online or, in some cases, to complement both processes and be able to interact physically and virtually. At best, communities can maintain their face-to-face interactions and enrich members' connections by breaking geographic barriers and promoting interactions on a web-based approach. The advantage lies on the ability to easily gather members' data capable of improving the user experience with the system (Mitzlaff, Atzmueller, Benz, Hotho, & Stumme, 2011). In particular, user-generated content on regularly distributed publication such as newsletters, provides primary data about topics of interest as well as shared patterns.

The study of online communities, namely in a scientific context, is important in order to understand how community members are willing to make contributions and be active. Moreover, a community is based on relationships that need to be understood due to its importance regarding contribution, motivation, and effectiveness. The goal lies on encouraging constructive participation and commitment. The higher the number of active and effective contributions, the better members work as a community and can take advantage from shared values, principles, and knowledge.

2.1.3. Scientific Communities

Although the concept of social network is not new, Web 2.0 has helped some authors define social network as a network of interactions or relationships, where nodes represent individuals and edges the relationships or interactions between them (Aggarwal, 2011; Freeman, 2004; Jackson & Watts, 2002). In a classical definition, we could define social network as a network based purely on human interactions (Aggarwal, 2011), where people interact face-to-face, or through personal friendships. This non-conventional definition, characterized by its geographical limitations, is given by sociology, which is responsible for the study of social networks in its convectional way where people need personal contact to promote information sharing.

However, regardless the conventional or non-conventional way of social networks, its representation is extremely powerful to understand the interactions and the connections among people. The ability to represent social networks depends on the way we define connections – direct links (e.g. a common activity) or undirected links (i.e. something unspecified), described as explicit and implicit relations respectively. In both ways, these connections establish new relationships and bring new ways of interaction.

Also in the scientific world, the interactions between researchers have been overcoming the geographical limitations of conventional networks and exploring the power of Web 2.0 (Papacharissi, 2010). Their exploitation is a valuable asset to expand contacts, share knowledge and promote successful relationships among potential research partners. However, the ability to manage the data that comes from online environments becomes more difficult to keep the pace and "the connections between individuals, groups and information become lost, or forgotten, and individuals and groups become ever more isolated" (McArthur & Bruza, 2003a). Still, this could lead to non-collaboration and poor knowledge sharing due to the non-exploitation of hidden relationships between researchers, and between researchers and information.

This non-collaboration, namely due to poor exploitation of Web technologies, is especially harmful to scientific communities who need a broad collaboration. Every time a connection between two or more researchers is missed, the opportunity to achieve partnerships and expand contacts is lost, meaning that none of the connections, visible or not, are used for the benefit of a community of researchers. Typically, connections named as visible are those who people clearly express or define. It is a synonym for directed or explicit links. On the other hand, when we define a connection between two or more individuals that is not explicit stated by them it is called non-visible connection, or indirect or implicit link. In both ways, it is possible to connect two or more individuals.

The analysis of collaboration between researchers within a community brings new ways of understanding the relationships, the structure of the different research fields, and the community itself. “The creation of scientific knowledge is a far richer, more complex, more intuitive and especially a more socially embedded process than the formal process defined in methodological textbooks” (Tuire & Erno, 2001). The introduction of a social network analysis explores the evolution of Web 2.0 and empowers scientific communities with a more convenient way of capture and maintain knowledge about the community’s relations.

Within a scientific perspective, the discovery of connections among researchers helps to identify common interests or topics of study. The relationships between researchers help to understand the structure of a scientific community and to formalize shared patterns. Unlike traditional networks, online networks are able to capture information beyond the one provided by individuals, as well as help defining interactions between them (Rangwala & Jamali, 2010).

2.1.4. Community Discovery

Data sets on individuals’ interactions are originally from different world domains and can be represented in the form of interaction networks with a meaningful way. The association of the data with the Web 2.0 technologies leads to a diverse range of data representation in the form of network of connections. Analysis of such networks can result in the discovery of important patterns and potentially shed lights on important properties on community structure (Parthasarathy, Ruan, & Satuluri, 2011). Moreover, the study of relationship networks, referred also as network of science, can provide insights into their structures, properties, and behaviors (Barnard & Simon, 1994; Girvan & Newman, 2002; M. E. J. Newman, 2006). Extracting such community structure and leveraging them to predict the emergent, critical, and causal nature of such networks in a dynamic setting has being growing on importance. Communities need to understand their structure and nature in order to better adapt to the dynamic characteristics of environment.

However, the extraction of such structure is a grand challenge and requires communities’ commitment on understanding the topological properties of its networks as well as the requirements imposed by directed and dynamic[2] networks. The more involvement a community has on understanding its structure the better it can take advantages from the discovery of patterns, relationships, and groups.

With the advance of the Internet and more recently Web 2.0 the number of applications of community discovery has arisen. In the E-commerce, for example, customers are grouped according to their profiles in order to empower personalized recommendation engines (Reddy, Kitsuregawa, Sreekanth, & Rao, 2002). Also in mobile ad-hoc net-

[2] By dynamic, we consider to any network that changes. This includes time-varying networks and networks that change due to external factors. (e.g. government approvals, or trust issues).

works, individuals are organized to enable efficient message routing and posting (Sun et al., 2009). In this context it is important to distinguish core members of the community from those on the borders (analogous to edge routers), and be able to bring to light the community structure. At the most fundamental level, community discovery "allow us to summarize the interactions within a network concisely, enabling richer understanding of the underlying social phenomenon" (Parthasarathy et al., 2011) and may lead to patterns discovery, identification of influential nodes, or sub-communities within a boarder community (Domingos & Richardson, 2001; Kempe, Kleinberg, & Tardos, 2003; Leskovec, Adamic, & Huberman, 2007).

Algorithms for community discovery can vary on a number of important dimensions such as, their approach to the problem and performance characteristics. One of the dimension on which algorithms vary is how they lead users to control the granularity of the division of the network into communities. Algorithms such as agglomerative and divisive clustering or MCL allow the user to indirectly control the granularity of the output communities through certain parameters. Agglomerative algorithms start by having each node in the social network in its own community, and at each step merges communities that are deemed to be sufficiently similar. On the other hand, divisive algorithms perform in the reverse way. They begin with the entire network as a unique community, and at each step, choose a certain community and divide it into two parts. Both algorithms stop when the desired number of communities is reached or the remaining communities are too dissimilar from the other to be merged or divided any further.

The output from the two hierarchical clustering algorithms is often based on a "dendogram", which is a binary tree, where leaves represent nodes and each internal node a community. In the case of divisive algorithms, a parent-child relationship indicates that the parent community, i.e. parent node, was divided to obtain the child communities. As for the agglomerative algorithms, the same parent-child relationship indicates that the child communities, represented by the child nodes, were agglomerated to obtain the parent community.

All these aspects and subjects were taken into account when developing the proposed system ("Newsletter Tracking System") and are discussed in the following chapter.

2.2. Tracking Systems

Online communities enable people to stay connected with no geographical limitations. People can communicate, collaborate, and share content with no need to interact face-to-face. The facilities brought by technology, such as messaging, web portals, and electronic mail are very powerful tools to allow individuals to communicate between them. As their use becomes part of people's social life, the number of interactions increases and thereby the interest to explore them.

2.2.1. Implicit and Explicit Relations

Every time people get to know each other they establish a new relationship. Logic and philosophy define it as a link between individuals or objects, mathematics as a generalization of arithmetic relations (Conville & Rogers, 1998), and computer science as a link between components or objects (Vrasidas, 1995). As for the presented research, a relationship is classified as a connection between two individuals, where both are connected due to a shared interest or common topic.

Connections can be defined as explicit connections, if they are clearly expressed by individuals (e.g. individuals' friends or followers), or implicit connections, every time they are implied (e.g. individuals' interests). Facebook is a good example on explicit and implicit connections. Although friend requests result from the explicit activity of sending a request and therefore creating an explicit friendship relation, the action of clicking a friends' post can result into the implicit connection with the post's topic and with another individuals that have clicked the same post.

Both explicit and implicit connections are important to understand members' network and communities' structure. An explicit behavior "is controllable, intended, made with awareness, and requires cognitive resources" (Dienes & Perner, 1999), and an implicit behavior comes from unacknowledged actions or actions outside of awareness (Schacter, 1992). The concepts can also be applied to a network of relationships, where nodes represent individuals and edges the relationships or interactions between them (Aggarwal, 2011; Freeman, 2004; Isaías et al., 2009).

Because individuals have clear sense of an explicit activity, this makes explicit behavior extremely important when defining individuals' relationships with each other and with content (Dörk et al., 2011). Whenever an individual explicitly express an action or behavior, this is translated into an explicit connection between the individual and something or someone. However, the study on explicit connections can be very limited if individuals do not share any of their interests. To fill this gap, the exploration of implicit connections can be helpful once it allows the discovery of new connections based on individuals' behavior and activity. Implicit connections can be used and explored as long as individuals interact with content and people, which in the virtual world can occur easily thanks to the Web technology.

2.2.2. Previous Approaches

Internet became the main chain providing infrastructure for people to collaborate, coordinate, and cooperate with each other (Harley & Blismas, 2010). It has revolutionized the computer and communication worlds like nothing before (Leiner et al., 1997), and today it reaches any field and affects the way society builds connections. People can create their own network of contacts and share information with anyone around the world. The more people interact, the more their network of interactions grows (Musser & O'Reilly, 2007).

The properly discovering and sharing of connection amount individuals is an important part of people's lives and knowledge construction in a community (Papacharissi, 2010). The use of email technology to discover individuals' connections (which is the methodology used on this research) is not a new idea and has been study in the last twenty years (N. B. Ducheneaut, 2002). In 2003, Ducheneaut and Bellotti commented that "even people having offices next to each other, still use email as a principal communication medium" (N. Ducheneaut & Bellotti, 2003). Additionally, previous works have also used email has a source to discover implicit relations. Schwatz and Wood, used email headers to extract shared interests between people using graph theory (Schwartz & Wood, 1992). However, this approach suffers from a lack of specificity since it ignored the message text and subject, hiding the context in which the interests arise. PeCo collected users relationship through email "From" and "To" headers by extracting keywords from the message text (Ogata & Yano, 1998). Still, is missing the analysis of the core message that can bring to light hidden relationships which are not represented in the "Subject", "From", and "To" fields of an email message. In 2003, McArthur and Bruza discovered implicit connections by mining semantic associations from people's communications. They proposed a model called HALe that automatically creates a dimensional representation of words based on the email corpus and uses it to discover a network of people implicit connected (McArthur & Bruza, 2003b). However, the solution does not focus on the interactions between people and content, especially with a scientific perspective.

More recently, Al Chakra, O'Doherty, Rice, and Yap (2008) developed a system that uses email technology to track information dispersion based on content substitution. The solution identifies entities to receive the email and "embeds a unique serial number by strategically changing words in the email content" (Chakra, O'doherty, Rice, & Yap, 2009). When a leak occurs the application is able to identify its source and thus information dispersion. The system has a business perspective and its designed to be used in scenarios where the protection of data is crucial. In addiction, its application domain is business focus and its features are limited to the identification of information dispersion. When it comes to discover hidden relationships between people, the system is not able to answer this need and to give a reliable result.

Foulger, Chipperfield, Cooper, and Storms invented a system and a method to generate and track an email campaign (Foulger, Chipperfield, Cooper, & Storms, 2010). The system generates an email campaign template from an email target database, which uses to create a custom email for each email target. The email campaign engine sends each custom email to the corresponding email target where it includes a custom uniform resource locator (URL) to identify the source of each interaction. Every time an email target selects a link the action is store by the server and then used for statistical purposes. However, even though the same track method is used, the system is not designed to explore the implicit connections between email targets and does not allow communities to change the system according to their needs and specifications. The system is well designed for marketing purposes but limited for data mining in a scientific perspective.

Barão and Silva proposed an holistic and complex model to define the Relational Capital Value (RCV) of organizations as well as online communities (Barão & Rodrigues da Silva, 2011; Barão & Rodrigues da Silva, 2012). Explicit but also implicit relational connections (such as these discovered by the NTS) are important for the RCV model application, hence for the determination of online communities relational value.

Table 2. Comparison between the analyzed approaches

Name	Authors	Year	Description	Problem	Discover Connections		Features		
					Explicit	**Implicit**	**Send Email**	**Track Interactions**	**Visualize Data**
-	Schwatz & Wood	1993	Used **email headers** to detect shared interests between people using graph theory.	Ignores the **context** of the message; Requires a **known starting** set of any specific interest area.	N	Y	N	N	N
PeCo	Ogata & Yano	1998	Used **From & To headers** to collect relationships between people, and **keywords** to extract "expertise" from message text by morphological analysis.		N	Y	N	N	N
HALe	McArthur & Bruza	2003	**Mining semantic associations** from people's email communications.	Does not focus on the **interactions** between people and emails' content.	Y	Y	N	N	N
-	Al Chakra, O'Doherty, Rice, Yap	2009	Tracks information dispersion based in the **content substitution**.	No connections **discovery**.	N	N	Y	N[3]	N
-	Foulger, Chipperfield, Cooper, Storms	2010	Generates and tracks an email campaign, and uses **URLs** to identify the source of each interaction.	No connection **discovery**.	N	N	Y	Y	N

[3] Although it tracks information dispersion it does not track interactions (i.e. clicks) that individuals do with on their emails' content.

2.3. Connection Degree Related Work

A graph *G* consists of a set of two elements, namely vertices *V* and edges *E*, and it is represented as *G=(V, E)* (Trudeau, 1994). Vertices, also called as nodes or points, are represented by dots, joined by lines or curves, which in turn are represented by edges. Edges can be direct or undirected. A direct graph or digraph is an ordered pair *D=(V, A)* where *V* represents a set of vertices and *A* a set of ordered pairs of vertices, called arcs, direct edges, or arrows (Diestel, 2006). On the other hand, an undirected graph is one where edges have no orientation, meaning that the edge represented by *(a,b)* is identical to the edge *(b,a)*. They are not ordered pairs but sets of two vertices.

Apart from its direction, edges can also have an associated label (weight), which is typically represented by a numeric value. In this case, a graph is called weighted graph (Fletcher, Hoyle, & Patty, 1993) and the weights may represent, for example, lengths, capacities, or costs, depending on the problem represented by the graph. Mainly, in applications, the weight may be a measure to the length of a route, the capacity of a line, or the energy needed to move between two locations along a route.

The concept behind connection degree is related with the association of a label (the degree) to every edge in the graph. This numeric value represents the weight that a particular edge has in the graph, meaning the "importance" of a relationship between two nodes regarding the universe of edges represented in the graph. The value assigned is based on both nodes' interactions with the newsletters. In this case, nodes represent individuals, i.e. community members. The proposed model tries to understand how strong a connection between two individuals is regarding their clicks on the newsletters. Moreover, when assigning the connection degree, a distinction is made between explicit and implicit behavior, which result from explicit and implicit data collection.

Table 3. Explicit and implicit data collection

Explicit and Implicit Data Collection
Explicit data collection include: • Asking individuals to designate their preferences on news' categories.
Implicit data collection include: • Collection individuals' interactions with the news; • Gathering individuals' interaction with news' categories.

The connection degree lies on both explicit and implicit data collection to better understand individuals' preferences and connect them. For every two individuals explicit and implicit interactions are taken into account to discover individuals' degree (weight) of connection. However, the proposed approach suffers from some problems:

1. **Cold Start:** Because the model requires a higher amount of information to better accurate its connections, the start is known as cold start due to the weak amount of data gather on individuals' interactions.
2. **Scalability:** The number of individuals and news produces a large amount of interactions to be processed. Once the calculation is performed to every pair of individuals, the vast amount of computation power may raises scalability problems when the number of interactions grows.

The main concept behind the connection degree (i.e. build a profile to each individual based on their interactions with the newsletters) is not a new. The same idea is described in recommender systems, which typically predict the "rating" or "preference" that a user would give to a new item (e.g. book, movie, music) or social element (e.g. people or groups) (Ricci, Rokach, & Shapira, 2011). This prediction is based on a user profile built by the recommender system itself and based on the characteristics of the previous items selected by the user or on the user's social environment.

Typically, recommendations systems apply knowledge discover techniques to face the problem of making personalized recommendations for information, products, or services (Sarwar, Karypis, Konstan, & Reidl, 2001). They produce a list of recommendations based on two approaches – content-based or collaborative filtering (Ricci et al., 2011). The content-based filtering approach uses the characteristics of the items to find and recommend to the user new items with similar properties. In particular, it gathers information on the items that user has selected in the past, and compares their characteristics with various candidate items and recommends the best matching (Mooney & Roy, 2000).

The approach uses an item profile (i.e. attributes and features) in order to characterize the item within the system and be able to relate users and items. It creates a content-based profile of users based on a weight vector of items' attributes, which represents the importance of each attribute to the user. The weights can be computed through a variety of techniques such the machine learning methods as Bayesian classifiers, cluster analysis, decision trees, and artificial neural networks, which determine the probability of a user to select/like an item (Pazzani & Billsus, 2007).

The issue regarding the content-based filtering lies on understanding whether a system is able to properly learn user preferences through user's actions and use them across other content types. The limited recommendation of items with common features and attributes of those selected by the user in the past, reduces the value that recommendation can bring to users when recommending items outside their know preferences that somehow could gain user's interests. The amount of recommended items are significantly less than when other approaches are taken into account (e.g. collaborative filtering). For example, when content-based filtering can be extremely useful when browsing for news, the same is not true in music, video, or products.

On the other hand, collaborative filtering builds a model based on users' past behavior and focus on the interactions between user and item, and uses that information to predict items that user may have interest in (Melville & Sindhwani, 2010). The approach is based on gathering and analyzing information on user's behavior, activities, and preferences, and predict what user will like based on other user's similarities. The advantage of the collaborative filtering is that it does not need to analyze each of the items and understand their features and attributes in order to be capable of performing an accurately recommendation.

The method is based on filtering information or patterns through the collaboration of multiple agents and data sources. Collaborative filtering does automatic predictions (filtering) about user's interests by collecting preferences from many users (collaborating). The assumption is that if a person A share the same view that a person B on an issue *z*, A is more likely to have a common opinion as B in another issue *x*. Typically, applications that use collaborative filtering lead with a large amount of data, gathered from user's interactions and behavior (whether explicit or implicit behavior).

Along with the content-based and collaborative filtering there are the hybrid recommender systems, which combine both approaches and bring what is the most effective into a unique method. The approach can be implemented by performing the content-based and collaborative filtering and then combine them; by adding content-based capabilities to the collaborative filtering or vice-versa; or by joining both approaches into a unique method and take what is the best about each of them (Adomavicius & Tuzhilin, 2005). This combination can also be used to overcome some of the problems in recommender systems such as cold start and scalability.

Although the connection degree defines a weight for every link (i.e. each connection between every two individuals) it does not go further on suggestion news based on individuals' preferences. It takes a passive approach by offering a way to measure connections' strength in a manner analogous to all but leaves the decision making to the community members – i.e. those who are responsible for analyzing the data and decide what to do. However, in the further chapters we will discuss the proposed "Connection Degree" model and how it helps communities to better understand their members' preferences and its structure.

3. Problem and Proposal

Although the concept of social network is not new, Web 2.0 has helped some authors define social network as a network of interactions or relationships, where nodes represent individuals and edges the relationships or interactions between them (Aggarwal, 2011; Freeman, 2004; Jackson & Watts, 2002). In a classical definition, we define social network as a network based on human interactions (Aggarwal, 2011), where people interact face-to-face or through personal friendships. This conventional definition, characterized by its geographical limitations, is given by sociology and describes an interaction where people need personal contact to promote information sharing.

Regardless the conventional or non-conventional ways of social networks, its representation is extremely important to understand connections among people. Even knowing that individuals relate to each other, the source of these relationships can come from several ways (e.g. face-to-face or online). Meanwhile, a connection can be defined as directed link (e.g. a common activity) or undirected links (e.g. a topic of interest), described as explicit and implicit relations respectively. However, both types of connections establish new relationships, bring new ways to relate individuals, and discover patterns among them.

Especially in the scientific world, the interactions between researchers have been overcoming the geographical limitations of conventional networks and exploring the power of Web 2.0 (Papacharissi, 2010). Their exploitation is a valuable asset to expand contacts, share knowledge, and promote successful relationships among potential research partners. However, if the ability to explore Web technologies is not properly exploited, information becomes lost and researchers more isolated.

Within the case study of Nano-Tera.ch, they needed to discover implicit connections among researchers in order to identify groups of researchers with common interests or topics of study. These relationships helped to understand the structure and behavior inside the community and to formalize shared patterns. Moreover, implicit connections were described as those who derived from members' actions and behavior with the content, namely HTML content. In particular, a set newsletters were designed and sent to researchers to capture their clicks on the newsletters' topics.

In order to identify implicit relationships within a community the proposed solution, named "Newsletter Tracking System" (NTS), uses electronic mail technology and HTML language to spread the information among individuals; PHP scripting to capture interactions and discover hidden relationships; and JpGraph graphic library to present the results. The procedure begins by creating an HTML file, called newsletter, with the information that needs to be spread across the community, goes through the process of sending to the individuals' e-mails, and ends with the evaluation of the data that is generated through their interactions with the newsletters. Every time an individual does a click in a newsletter link the NTS stores the action to further analysis.

In the end, not only the community analysts benefit from an implicit analysis but also individuals who can discover or be informed about people who showed interest in the same topics as them.

3.1. Goal

Online communities enable people to stay connected without being geographically restricted, meaning that they can communicate, collaborate, and share content with no need to interact face-to-face. Facilities brought by technology enable individuals to communicate to each other and overcome the constrains of time and space. As the use of online technologies keeps spreading at an incredible rate and interactions between individuals grow, the usefulness of these interactions becomes clear and an aspect to explore.

By taking into account that Web technologies can improve the way people interact as well as enable the study of their behavior, there is an opportunity to gather individuals' information and explore it in a community context. Meaning that, if a community is based on a set of members connected in a certain aspect, a further exploitation of these connections can help communities to better understand how members related to each other and how the community is organized. In particular, how members are connected between them regarding both explicit and implicit relations with content.

This way, the goal of this research lies on **explore individuals' interactions with social objects (digital content) and, based on that, discover implicit connections among them**. Through interaction with content (e.g. newsletter) individuals are related with it and therefore implicit related with each other. At the end, implicit connections bring to light hidden relationships and help communities to understand how their members relate and therefore how the community itself can be organized.

3.2. Problem

As Web technology becomes more powerful and its use more ubiquitous, it is increasingly important to understand what benefits may arise from its use and exploitation, namely in the context of communities. Individuals who belong to a community describe the connections among them and build their own network of contacts based their preferences and needs. As the number of individuals grows it becomes important to keep the pace in order to understand how a community is organized and how individuals relate to each other. However, as a community, there are a set of problems that may rise regarding the goal of having an updated overview of its structured and individuals' connections.

- **Data Collection:** When the number of individuals is high, the process of data collection becomes a challenge regarding the traditional methods such as face-to-face or surveys. Communities should be able to collect data on individuals' interactions without require too much effort from human resources as well as collect it with no time and geographical restrictions. Because individuals can interact with the content whenever they want and wherever they are, convectional ways to collect data do not allow communities to easily gather data and keep the pace on individuals' connections.

- **Data Storage:** As soon as communities are able to collect the data on individuals' interactions, it comes the challenge of storing it in a way that is secure and ready to process. Because data comes from individuals' behavior it is crucial for communities to ensure that it is safe and properly processed. Moreover, individuals should agree or sign for the process of data collection, otherwise the data cannot be collected.
- **Data Processing:** Once the data is collected and rightly stored, it only turns into information if communities are able to process it. Although the data by itself has no value it becomes extremely powerful when it is converted into information. Information refers to data organized in a way that has meaning and therefore allows communities to perform analysis, discussions, and take decisions. Communities should be able to organize data, process it, and extract their value through an efficient and practical approach.
- **Data Visualization:** The process of extracting value from the data can also be support and easily performed if communities have the ability to analyze it with visuals (e.g. graphs or tables). Visuals allow communities to better understand individuals' interactions and have an overview of its organization. However, most communities lead with the problem of getting a visual perception of their data in a way that helps them to perform analysis and take decisions.

In a scientific context, where the protection of intellectual property is taken very seriously, the power of the Web technology is also being explored as an opportunity to grow and expand a network of contacts. ResearchGate is an example of an online network where researchers can "expand their contacts, share knowledge and find potential research partners" (Researchgate, 2012) more easily and conveniently. The new set of Web technologies is changing the way scientific communities can explore their researchers' behavior. The classical face-to-face approach can be moved to the virtual world of online interactions, meaning that the information about researchers that was firstly gathered in a face-to-face process or through bureaucratic processes can now be collected using Web technology.

The problem relies on **finding the best strategy to collect the data as well as an effective way to treat it**. Along with the data collection on individuals' preferences comes the responsibility of handling it. The decision of exploring Web technologies in order to capture more information about members, namely their interactions with content, makes the data personal and powerful, and the processing sensitive to manage. Especially when the number of members in a community reaches thousands and interactions become countless, it is important that communities have efficient methods to perform their analysis and build visuals. Otherwise, data processing and representation becomes confusing and impractical, making difficult the study about implicit relationships and therefore a community analysis.

Along with the problems listed above communities also face the issue of high costs with the acquisition, maintenance, and evolution of known solutions. Namely in the context of scientific communities, these should be able to build their own solution or adequate an existing solution to their needs and specifications. Moreover, in order to not spend too much money, an open source solution is the most valuable requirement that communities set when building or acquiring a solution. For all these problems the thesis proposes the "Newsletter Tracking System" (NTS) that is briefly explained in the next section and fully presented in the Chapter 4.

On the other hand, there is also **the problem of having a high number of interactions between individuals and content**. In this case, the amount of implicit connections among individuals will be high and the implicit network will be a mess of links, which hamper data analysis and the chance of extract value from the implicit connections. Every time an individual interacts with the newsletter, it is created a connection between the individual and the newsletters' content, which is then used to implicitly discover relations among them. If the number of interactions is high, in the worst case scenario it could happen that individuals are all related to each other and it becomes very difficult to analyze the data and therefore understand individuals' relations inside a community.

To solve this problem the presented research proposes a model called "Connection Degree" (CD) that assigns to each connection a value that represents the importance of that connection in the network. The higher the value (i.e. the connection degree), the greater the importance. This model is also briefly presented in the next chapter and fully explained in the Chapter 5.

Both solutions, NTS and CD, were designed to solve the presented problems and to better fit the needs of a community, namely the need of discovering how their members relate to each other and how the community is organized.

3.3. Proposal

Through interactions with newsletters the proposed solution focuses on trying to gather connections and grow a network of implicit relationships. It provides a "Newsletter Tracking System" (NTS) able to capture and exploit implicit connections between individuals and presents a "Connection Degree" model (CD) to define the strength of those connections. Every time an individual clicks a news (in the newsletter) the action is gathered and stored as an implicit connection between the individual and the news. At the end individuals are all related with the news and then implicit related to each other with a CD.

The schema below (Figure 2) presents an overview on how both solutions, NTS and CD, related with each other and with the community itself. Note that, the schemas are designed using the i* Framework, which "offers the notions of *actor*, *goal*, and (actor) *dependency* , and uses these as a foundation to model early and late requirements, architectural and detailed design" (Castro, Kolp, & Mylopoulos, 2002). This framework tries to overcome the mismatch information system suffer between analysis and design. While "operational environment is understood in terms of actors, responsibilities, objectives, tasks and resources, information systems are conceived as a collection of modules, entities, data structures and interfaces." The framework lies on the goal of helping improving the quality of information systems by applying a common languages in order to avoid misunderstandings and communication problems.

Regarding Figure 2, the community, represented by the actor community manager, relates to both NTS and CD through a goal dependency.

- **NTS:** The community manager has the goal dependency of "Discover an Implicit Network" regarding the NTS, meaning that the community manager (the depender) depends on the NTS (the dependee) to achieve his goal. On the other hand, it is expected that the NTS do whatever is necessary in order to achieve the goal, which in turn depends on the resource "Newsletter" to be completed.
- **CD:** The goal dependency is based on the willingness of the community manager to "Discover Implicit Connections' Importance." However, in order to achieve that goal, the community manager depends on the actor CD, which therefore depends on the resource "Implicit Network" provided by the NTS. This dependency order shows us that in order to discover the implicit connections' importance the community manager has to first use the NTS to discover an implicit network and use the CD to achieve the goal.

Both solutions complement each other and satisfy different community needs. The proposed NTS helps communities to discover implicit connections among members and foster an implicit network based on those connections. The links come from the members' interactions with the newsletters, which is an imperative resource needed by the NTS. On the other side, the CD is able to satisfy the community goal of "Discover the implicit connections' importance" based on the implicit network discover by the NTS. This way, the CD complements NTS by assigning a value to each link between individuals and bringing to light connections' importance. As for the NTS, it adds value to the CD by proving an implicit network based on individuals' interactions.

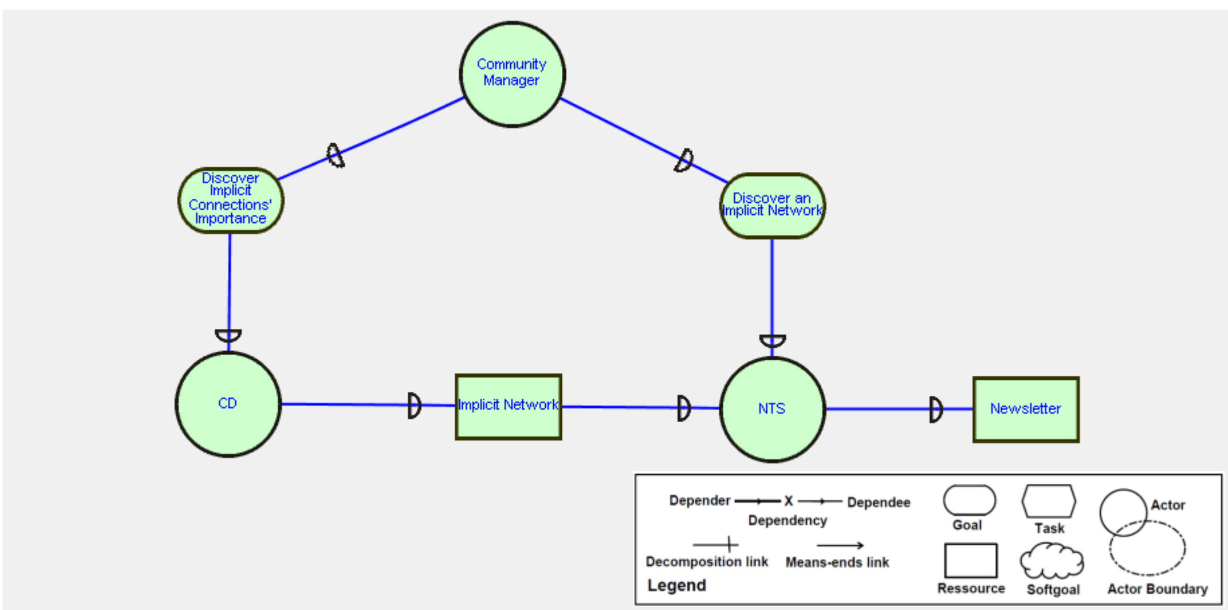


Figure 2. Proposed solution

3.3.1. Newsletter Tracking System

The NTS is an online tool that uses email technology and PHP scripting to send and collect individuals' information. The core engine is responsible for preparing the newsletters and gathering the information from individuals' interactions, i.e. it captures the data and exploits it into implicit relations and statistical results. A community interested on improving their knowledge on individuals' connections can build their own newsletters and use the NTS to send, capture, and process the gathered data. On the other side, individuals can receive the newsletters on their email addresses and share them with online social networks. NTS is also able to capture the interactions that come from the outside, i.e. from shared emails, social networks, and online newsletter.

On the NTS approach, the process of building a network of implicit connections between individuals is break down into several stages in which information is treated in different ways. Initially community managers are the main player once they are the best to know what kind of news their community would like to receive. Then, community managers are responsible for creating their own newsletters, with the only restriction of an HTML format, and to upload them into the system. After that the NTS will track the links in order to capture individuals' interaction and store the newsletters on the server. At the end, the community manager is responsible for sending the emails to the list of selected individuals with whom he wants to share the newsletter.

Data Analysis

One of the main features about the proposed solution is its statistical and data analysis purposes. Once the data on individuals' interactions is stored, the analysis feature allows communities to study members' connections and discover implicit relationships between them. NTS enables the data to be organized in several ways and uses JpGraph library to present it into online visuals.

The analysis is crucial for the NTS to discover implicit relationships among individuals and therefore allow communities to study their members' interactions and enrich their knowledge on connections. It also enables communities to download the data and explore it the way they want.

3.3.2. Connection Degree

While information on individuals' connections increases, it is important to understand how relevant discovered connections are and how they can be organized by its value in the network. The presented solution tries to answer this need by proposing the "Connection Degree" model that uses the two type's explicit and implicit connections as input values to calculate the connection degree for every two individuals. Both connections are gathered to calculate the explicit and implicit degree in the final equation, respectively.

The explicit degree is based on individuals' preferences on the categories that a community offers on its newsletters, and the implicit degree is calculated based on the individuals' interactions with news and categories. The action of clicking the news is stored as an implicit connection between the individual and the news. Once the news is linked to a category, the action also implicitly relates the individual with the news' category. This value for the explicit degree is included in the equation that measures the level of connection between an individual and the categories. The calculation ends by organizing individuals into a matrix where rows and columns represent individuals and values the connection degree between them.

The proposed model on the connection degree tries to assign to each connection a value that enables communities to measure their universe of discovered connections and understand which are the most important. In Chapter 0 the CD is explored in detail and discussed the added value to the NTS.

4. Newsletter Tracking System

Communities have taken a step into the virtual world and included web tools in their rituals of collaboration. Technologies such as email, forums, and blogs became part of communities' ways to interact and to allow members to stay connected. The appearance and evolution of Web 2.0 technologies allowed communities to move virtually and take advantage from what virtual technologies can offer as interaction enablers. Moreover, communities can maintain their physical structure and use web technology to overcome the obstacles of time and space.

Time – Since the newspapers, where information was delivery with one day of delay, radio and television decreased this delay to hours, minutes, or even to live streaming, with only few seconds separating reporters from viewers. With the web and mobile phones, people are constantly reporting to each other and receiving information updates each second (e.g. twitter, Facebook). Moreover, the web brought to individuals the power to select their own content and receive information based on their interests. Because the information is available 24 hours a day, readers can access it when it is more appropriate for them and have no time constraints regarding their schedules.

Space – Media such as newspapers, radio, and television deliver information to a relatively large geographical region: district, state, country, etc. However, although their efficiency on time delivery has been improving according to the technological evolution it is very difficult to find localized information, at a community level. With the web, any community, independent of its size, can share information relevant for its members and deliver information within a specific community focus wherever the web access is possible. The goal is to facilitate the collaboration across members by hiding the constraint of distance and allowing every member to receive and consume the information whatever is his physical location.

Due to online technology, people are able to communicate, collaborate, and share knowledge overcoming time and space restrictions. Virtual facilities brought people and groups close together along with their profiles and related information. Also communities took their step into the online world and have included web tools into their habits. Web technologies become part of communities' ways to interact, increasing the number of connections among individuals inside a community. However, the task of controlling the growth of those connections is difficult to accomplish in the old by hand way. As a consequence communities lose the track to connections as well as their chance to better understand and explore their members' relationships.

Through interactions with newsletter the "Newsletter Tracking System" proposes a way to capture individuals' interactions, discover implicit connections, and foster an implicit community. It uses electronic mail technology to reach individuals, and newsletters to promote interactions and discover implicit connections among them. According to Bellotti Ducheneaut (2002) "even colleagues having offices next to each other, or sitting in plain sight of each other, still use e-mail as a principal communication medium" (Bellotti, Ducheneaut, Howard, Smith, & Grinter, 2005). Also the fact that email is used worldwide and one of the most known tools makes it one of the tools to better reach individuals and capture interactions. As for the use of newsletters it comes as the way to capture individuals' interactions. It gives communities the freedom to design their own content and define the newsletter according to members' interests. Communities can also keep their newsletters viral and reuse them as a means to extract interactions while keeping their members updated. The next chapter explains why the use of newsletter was applied to this research.

4.1. Social Affordances of Newsletters

The theory on affordances was originated on Gibson work about visual perception (J. J. Gibson, 1986) but has since been applied to texts, social technologies, and social settings more generally (Fayad & Weeks, 2006; Graves, 2007; Hutchby, 2001). By "social affordances" it is mean to the possibilities for action that are called forth by a social technology or environment, in this particular case, by newsletters. A pencil, for example, "calls forth" writing; telephones "calls forth" talking; and books "calls forth" reading. The study on affordances provides a framework for identifying the characteristics of newsletters that "call forth" the constitutive elements of community.

Along with Nano-Tera.ch government bodies it was agreed that there are at least three types of social affordances required for the formation of virtual communities, namely on a newsletter basis: affordance of subscription, interaction, and connection.

4.1.1. Affordance of Subscription

The ritual of receiving the newsletters is depended on members to subscribe it and thus take part of the social environment that comes from newsletters distribution and topics of discussion. There are two aspects that can be directly observed. The commitment between members and the community that agree to take part on the same effort of information sharing. The community by gathering and spreading information, and members by consuming that information and using it to create knowledge and improve research work.

The other aspect is the access to the content that members are able to reach through newsletters and that community is able to offer as a way to promote collaboration and cooperation between them. Although information on the community's projects is available to all members, the problem lies on content gathering and filtering. Members do not have the time to be aware of all the projects' updates as well as to find ways to help each other. The subscription allows them to have a trusted distribution channel to spread information on their projects and to receive other projects' news in a most efficient and effective way. The reunion of members' efforts for a common purpose encourages community practices of sharing and help.

4.1.2. Affordance of Interaction

For each of the newsletters, members are able to interact with the content and navigate on projects' information and updates. Members can satisfy their willingness to know more on projects by clicking on the newsletter's content, namely text or images. Although the two ways of interaction differ from a design perceptive, they both are based on hyperlinks which are able to forward members to the web pages with more information about the content he/she has clicked.

The ability to interact, even in a simple way, encourages members to navigate beyond the information provided by the newsletters and learn more about the projects on the community. However, the level of interaction and engagement can grow thanks to the chance for members to express themselves through comments, suggestions, and thoughts on the newsletters' subjects. The opportunity to become active members and share visions and knowledge brings value to the community and enables individuals to help each other.

4.1.3. Affordance of Connection

A community can be defined by the proximity or similarity of members, where nodes are proximal if they are linked (boyd, 2006). This definition translates a community into a network of connections where ties reveal members' connections and proximity with other members. The ability to establish and discover new connections enables members to experience the involvement, identification, attachment, and sense of belonging to a community. However, in the case of Nano-Tera.ch, the exploitation of the newsletters as connection enablers was not possible due to its process of handling the newsletters. Its process lied on spreading the newsletter without having a way for members to connect to each other and bringing to light the sense of community. Nano-Tera.ch miss the fact that when members discover that many existing members have common interests, they become more active, due to the motivation of sharing knowledge and find new ways of cooperation. In fact, the ability to sense and respond interactively and to be able to leverage the wisdom of crowds (or communities) can be extremely fruitful and useful structure (Parthasarathy et al., 2011).

Although the general terms of "community" and "online community" have been study for several years and have enduring rhetorical and cultural appeal, the concept of community is notoriously slippery. By analyzing and extracting the most common ideas from the previous literature, it is possible to get a sense of what is necessary to designate a social group as a "community". Most of these requirements are enable and engage shared rituals, social regulation, and collective action through interactions and the ability to create relational linkages among members, which promote emotional bonds, sense of belonging, and self-identification within the community. The wide utilization of the affordances of subscription, interaction, and connection would help a community to provide a fertile ground for the development of an online community. However, the community metaphor merits a close examination for each particular case because it resonates deeply with individual and cultural aspirations (Parks, 2010).

4.2. System Overview

The NTS is an online tool that allows communities to foster an implicit community based on members' interactions with newsletters. It supports the relationship between the community and members by providing a tracking tool for the community (Figure 3). The community itself is managed by a "Community Manager" that is responsible for triggering the tasks at the NTS. Based on the goal of "Discovering an Implicit Community", the community manager depends on the NTS to achieve it. This relationship is based on a goal dependency and expressed as a relation "depender-dependee," where the "Community Manager" (the "depender") depends on the "NTS" (the "dependee") to achieve the goal.

The NTS is responsible for making the decisions that are necessary to achieve the goal and the community manager does not care how the NTS goes about achieving it. On the other hand the NTS has a resource dependency regarding the community manager. The NTS depends on the community manager to provide the newsletter so it can perform its tasks, satisfy the goals, and also provide the resources. Without all the input elements no further links can be followed and the model stops.

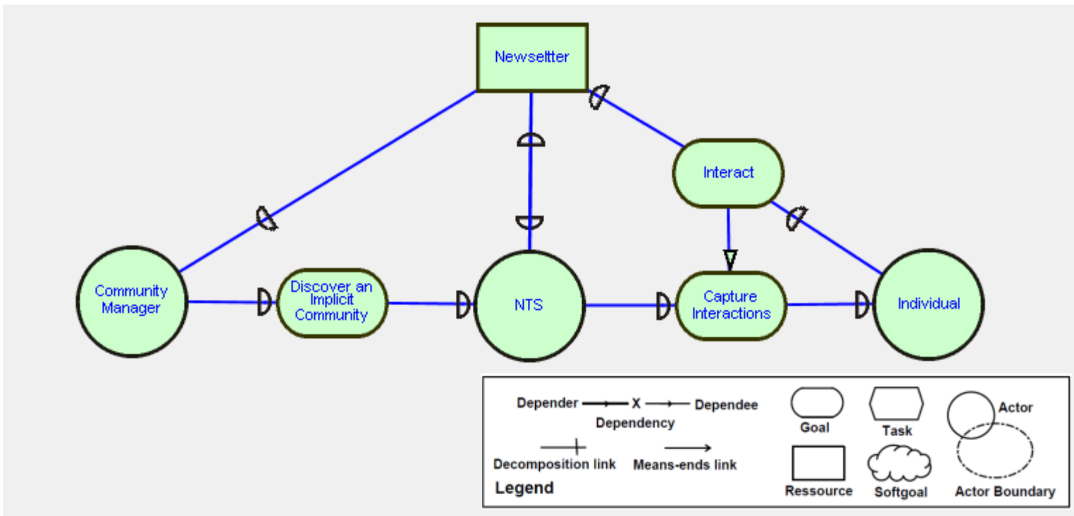


Figure 3. System overview

The relationship between the NTS and the individual is also based on goal and resource dependencies. In order to satisfy the goal of "Capturing Interactions", the NTS depends on the individuals to "Interact" with the "Newsletter" that is resource dependent on the NTS. The dependency happens in both directions. The NTS needs the individuals to interact with the newsletter, and the individuals need the NTS to provide the newsletter in order to achieve the goal of interaction. If some of the elements in the relationships does not do its role as a "dependee", both parts end up not achieving their goals. The link between the goals "Interact" and "Capture Interactions" represents a "means-ends" link. The mean of interact has an end of capture interactions, which are then used for the NTS.

The presented model is based on a dependency model of goals, where the actors "Community Manager", "NTS", and "Individual" are depended on each other based on goals. The direction of the dependency link defines the way the goal is achieved, i.e. which actor, task, or resource the goal depends on. Also, on the relationships where the "Newsletter" is a resource, the "Community Manager" represents the "depender" regarding the link with the NTS (the "dependee"), and the "Individual" the "dependee" in the relation with the NTS.

In a deep exploitation of the NTS, this is based on the three main tasks of "Upload Newsletter", "Send Newsletter", and "Analyze Data" (Figure 4). The community manager's goal of foster an implicit community is decomposed into three different tasks that are trigger by him and performed by the NTS. Although the relationship between the community manager and the NTS is a goal dependency, it can be described as a decomposition of three tasks that need to be performed in order to achieve the goal.

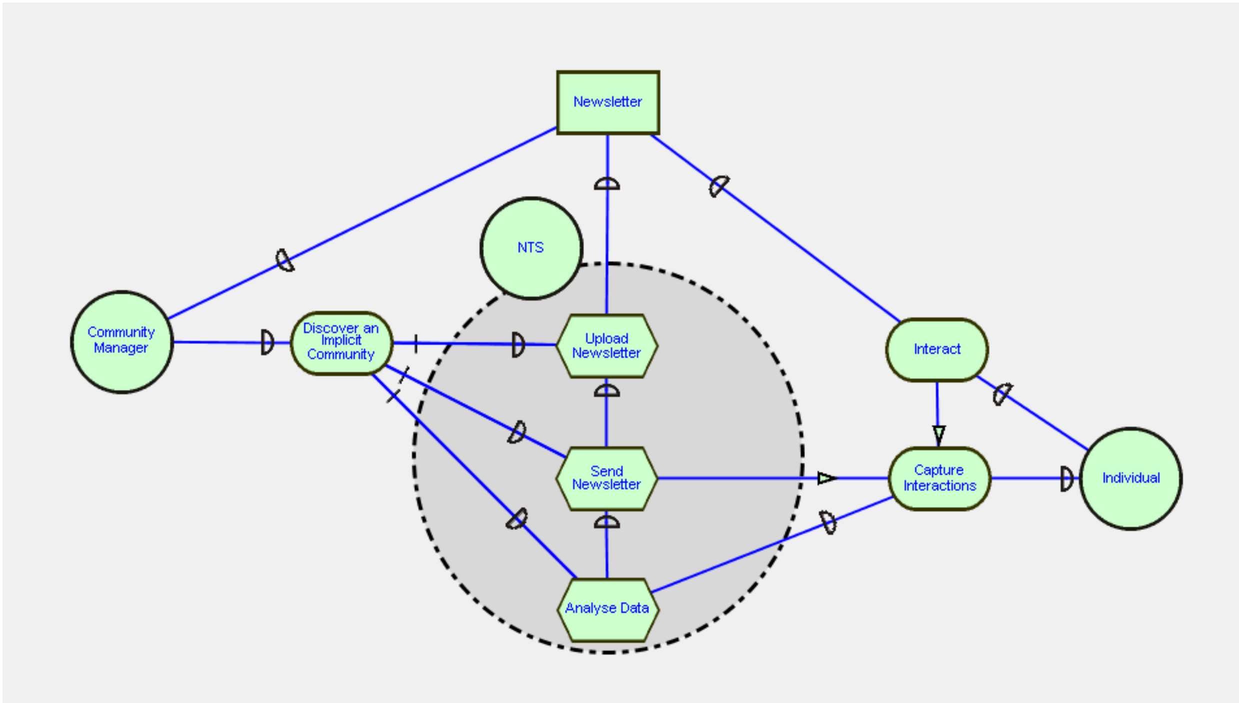


Figure 4. "Newsletter Tacking System" overview

The dependency links show the order in which the tasks must be performed. Only the tasks that do not have any dependency links going out can be performed right away the community manager wants to meet his goal. In this case all the tasks are "dependers" and need their "dependees" to run. By following the dependency links, the interpretation is that the "Upload Newsletter" task can be performed as soon as the "Community Manager" provides the "Newsletter." Then the "Send Newsletter" task, and by the end the "Analyze Data" task.

In order to the "Community Manager" achieve the goal of "Foster an Implicit Community" he needs to trigger on the "Newsletter Tracking System" the task "Upload Newsletter" by providing the "Newsletter" as the input, then the "Send Newsletter" task, and finally ask for the system to "Analyze Data". The task of "Send Newsletter" has the end goal of "Capture Interactions" that needs to performed for the last task of "Analyze Data". Again the "Individual" is responsible for meeting the goal of "Interact" and close the cycle of the dependency links. Once all the dependency links are respected the goals can be reached and the NTS is able to help communities to fostering an implicit community based on individuals' interactions with newsletters.

At NTS web page (Figure 5) communities can choose one of the three options of "Upload" (#1), "Send" (#2), and "Analyze" (#3). Each option forwards the community to a different web page where it is possible to perform the task asked in the front page. The presented page lies on the goal of helping communities to understand which order they should follow in order to properly gather individuals' data, namely, (1.) Upload, (2.) Send, and then (3.) Analyze the data.

Figure 5. NTS features web page

4.3. Upload Process

The process of capture and detect implicit connections among individuals starts with the community manager uploading a newsletter into the NTS. The upload process is described as the first interaction with the NTS. At this stage, the community manager reveals its interest on capturing members' interaction and fostering an implicit community. The process can be represented by Figure 6 where the "Community Manager" and the "NTS" are the only actors. The model starts with the dependency goal of "Upload Newsletter" between the community manager and the NTS. In order to the community manager satisfy his goal of uploading the newsletter he needs the NTS to perform the task "Newsletter Uploading". On the other hands, the NTS needs the "Newsletter" resource given by the community manager. Thus, he should first design the newsletter and then meet the goal of uploading it.

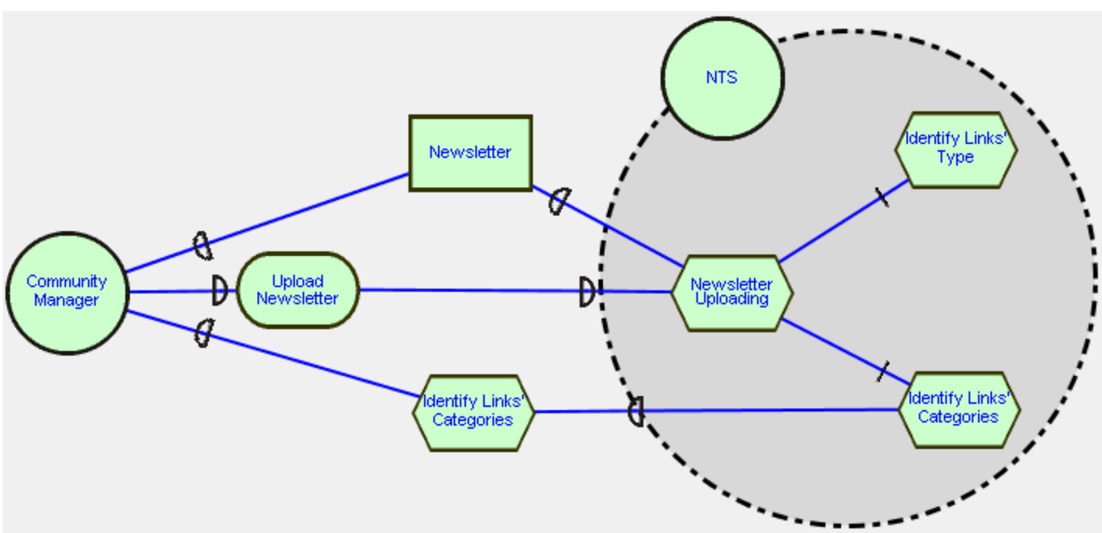


Figure 6. Upload newsletter task

The newsletter is described as the main resource for the system since it is the element that is shared with members. The community manager is responsible for choosing the content, designing the newsletter, and be aware of the newsletter's importance as a promoter of interactions. The better a newsletter meets the members' interests, the higher the number of interactions. This responsibility is given to communities once they know better what members' desire and expect. Figure 7 presents an example of a newsletter designed by Nano-Tera.ch community.

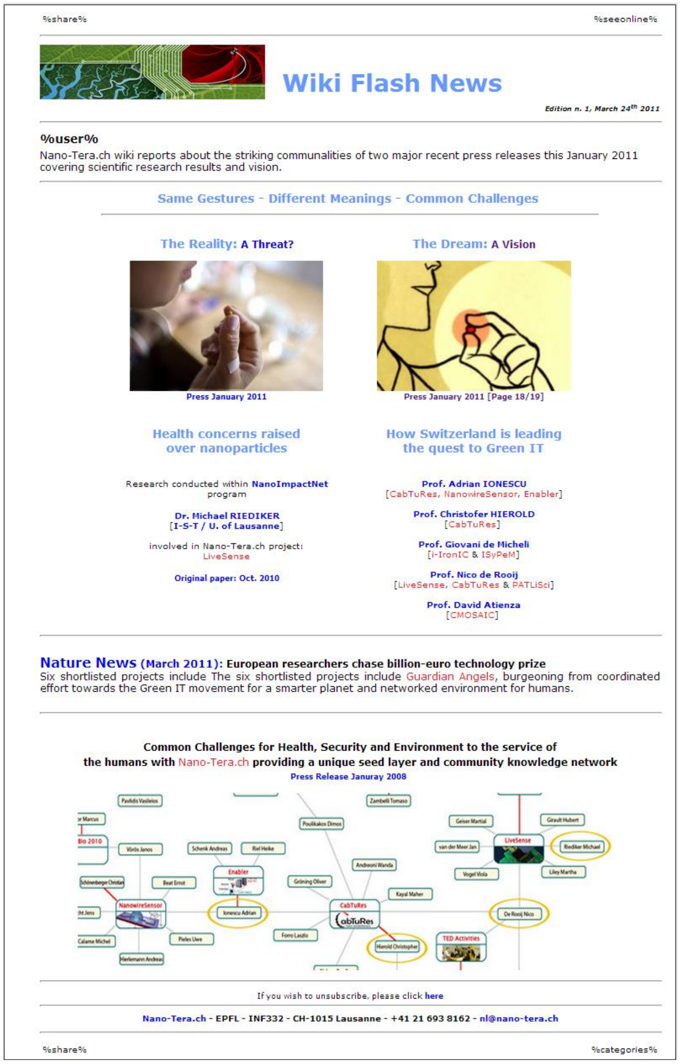


Figure 7. Newsletter example

As soon as the newsletter is ready to send, the community manager is able to initiate the upload process by triggering the task "Newsletter Uploading" on the NTS and providing the newsletter as a resource. At this stage the NTS is able to decompose the task into two different tasks - "Identify Links' Type" and "Identify Links' Categories." On the first task the NTS will process the newsletter and identify the type of all the links. The type is defined as the way links can be illustrated and can be identified as "Text" if the link is represented by text or as "Image" is an image extends for the link. This analysis allows communities to understand how individuals prefer the information to be exposed.

The next task of “Identify Links’ Categories” on the NTS depends on the “Community Manager” to perform it. For each link the community manager has to identify its category in order to the NTS complete the process of uploading. The task dependency is based on the community manager’s responsibility to define which categories fit best the links. The NTS presents him the uploaded newsletter with all the links followed by a combo box of categories, which is filled, based on the list of categories provided by the community manager (Figure 8). This categorization is what allows the NTS to discover implicit connections among individuals. Each time an individual clicks a link, the action is stored as a relation between the individual and the category of the link. At the end, individuals will be related with categories and implicit connected to each other based on those categories.

In a practical analysis the community has a web page designed to upload a newsletter (Figure 9), where it writes the newsletter “Title” (#1) and selects the “Type” (#2), which by default is set as “Newsletter”. This last option allows communities to upload different types of HTML files to gather information (e.g. invitations). On the last step the community manager can upload the newsletter’s files (e.g. HTML file and images) by selection the option “Choose Files” (#3) and then the “Upload” button (#4). At end of this step the newsletter will be uploaded into the system and the community manager is able to go forward on the tracking process, namely into the send step.

Figure 8. Process of selecting links' categories

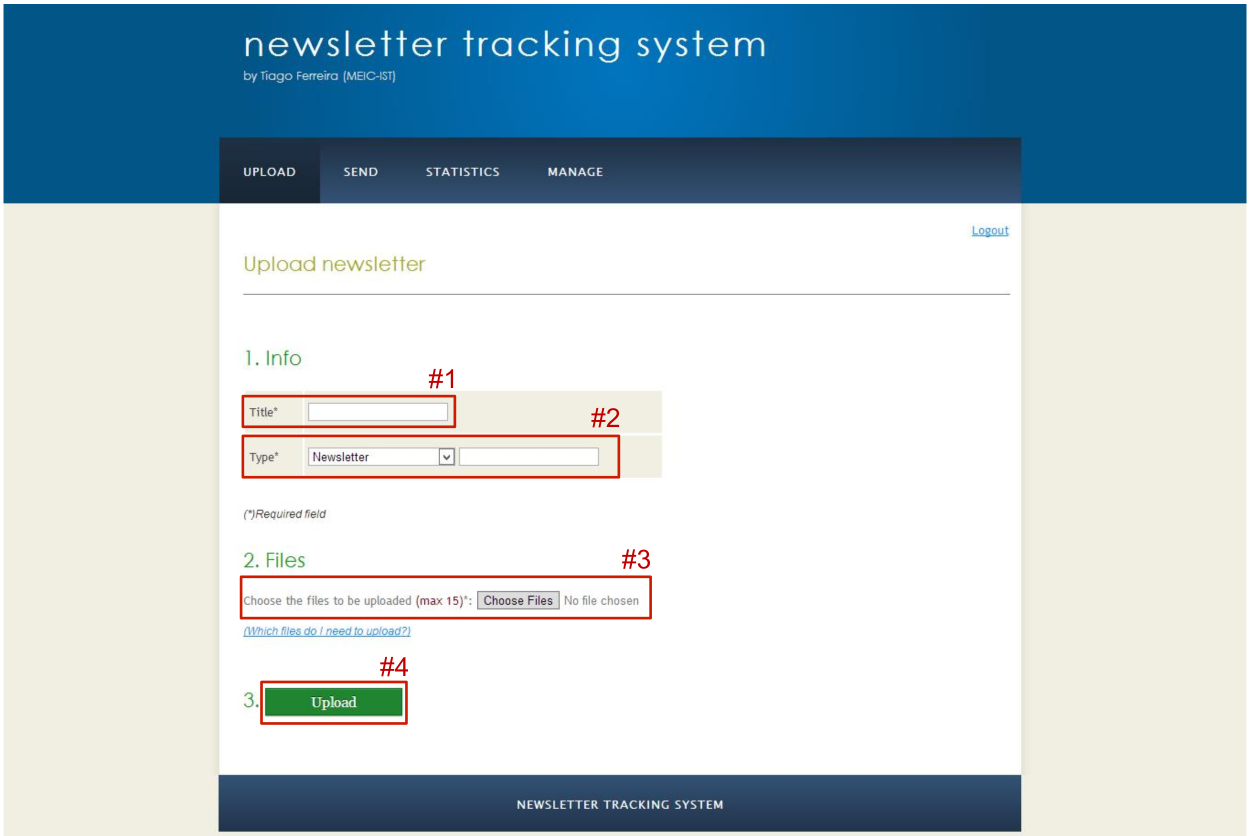


Figure 9. Upload web page

4.4. Send Process

To meet the goal of discovering implicit connections among individuals, the community needs to reach individuals' emails and let the NTS to capture their interactions with the newsletters. The process is described by Figure 10 and starts with the goal of "Send Newsletter", triggered by the "Community Manager", and task dependency on the NTS "Sending Process". Once the request reaches the NTS the task is decomposed into two different tasks, which according to the dependency links direction should start with the "Links' Tracking" task. However, the resource dependency on the "Newsletter" comes from the community manager through the "Upload Process".

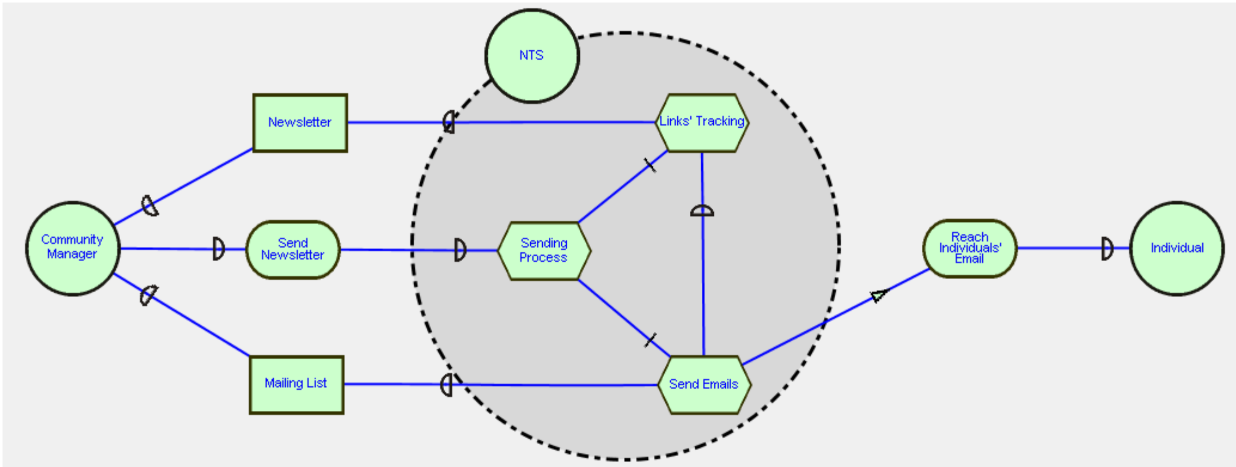


Figure 10. Send newsletter task

Once the dependency links are respected the NTS starts the task of tracking the newsletter's link by replacing them with malicious links. Each link is based on a standard URL defined by the NTS and shaped according to the following schema.

base-url/code/user-id/link-id/newsletter-id/

The "base-url" represents a common prefix to all the links, namely the path to the server where the NTS is deployed. The "code" defines one of the possible actions: opening the email, clicking a link, sharing the newsletter, or seeing the newsletter online. The "user-id" identifies the individual who trigged the action, the "link-id" which link he had clicked, and the "newsletter-id" the newsletter in which the action was performed. The elements are all automatically generated by the NTS and put together in order to build a tracked newsletter for every individual.

The newsletters are then the key elements to perform the task of "Send Emails", where the community manager executes his last interaction with the NTS by proving the "Mailing List" resource, which contains the list of all the emails to which the newsletter is going to be sent. This task of sending the emails with a tracked newsletter has the end of "Capture Individuals' Interaction". Once the newsletter reaches the "Individual" all the remaining process depends on him, namely on his interactions with the newsletter. The dependency process starts with individuals opening their emails with the tracked newsletter and ends with individuals' clicking on the links.

The practical analysis can be explained using Figure 11 that shows the web page designed for communities to send the newsletter to a list of contacts. The process of sending starts with the community filling the "Subject" (#1) of the email and then selecting the list of contacts to which the newsletter is going to be send. At this point, the community has the following options:

1. **Select a Mailing List (#2):** This option consists of uploading a comma-separated values (.csv) or a Microsoft excel (.xls) file (or a set of files) with the list of contacts. The goal lies on allowing to import a great number of contacts at once. The file should have per line the data for each of the individuals and follow the following format for the columns:
    a. **Column 1:** Last name
    b. **Column 2:** First name
    c. **Column 3:** Email address
2. **Type Email(s) Manually (#3):** At this option, communities are able to write a set of email addresses to which the newsletter is going to be send. Although this option does not allow individuals to receive the newsletter personalized with their first and last name, it offers the ability to send the newsletter to a set of contacts without the need to develop and prepare an .csv or .xls file. Moreover, this option can also be used in addiction to the uploaded file where both, file(s) and typed email(s), will be processed and will receive the newsletter.
3. **Test (#4):** The last option is designed for communities to receive feedback during the process of sending the newsletter. The emails inserted at this stage will be intercalated, starting by the end, with the list of contacts provided in options #2 and #3. For instance, if a file with 20 emails is upload in #1, more 10 email are added manually in #2, and 3 emails are inserted into the option #4, the process of sending will be the following: 10 emails from #1, 1 email from #4, the last 10 emails from #1, 1 email from #4, and finally 10 emails from #2, and 1 email from #4. Thus, communities are able to understand if the process of sending went well and if all the email were processed.

Once both subjects and contacts are defined, the community manager has to select the newsletter(s) to be spread across the list of emails by selecting the combo box(es) from the list of all uploaded newsletters (#5). As shown in #5 the information of the “Title”, “Type”, “Date” of upload are presented along with the ability to see the newsletter. The last step is performed by clicking the “Send” button (#6) where the selected newsletter(s) is(are) sent to the list of contacts via email with the defined subject.

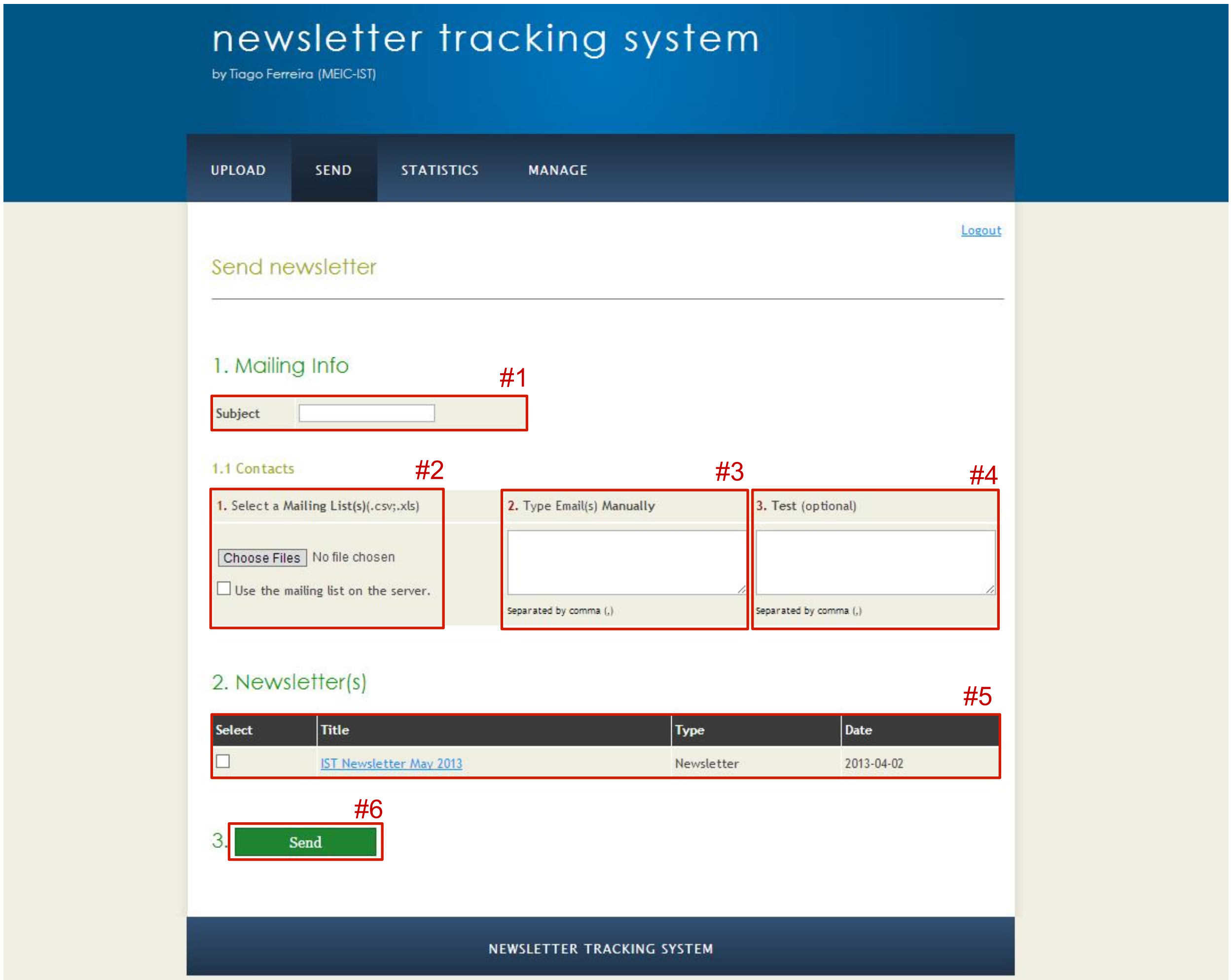


Figure 11. Send Web Page

4.4.1. Tags

When designing the newsletters, the community manager should include a set of tags that will be translated into features during the sending process. The position of the tags defines the place where the actions associated to the tags will be performed. Although the tags are optional they are crucial to enable individuals to better explore newsletters and to spread them across their own network of contacts. The actions associated to the tags are the following:

- **%share%** (Figure 12. #1): Designed to allow individuals to share the newsletter across several social networks – digg, stumbleupon, twitter, Technorati, Facebook, newsvine – or through email. The tag is translated into a text defined by the community in which it tries to convince the individuals to share the newsletter (Figure 14. #1).
- **%seeonline%** (Figure 12. #2): The tag is translated into a link in which individual can click to see the newsletter in the browser. This option is designed to allow receivers to read the newsletter especially when it appears baldy on their emails or if they prefer to see it online instead of their personal email. The text that appears is also defined by the community (Figure 14. #2).

- **%user%** (Figure 12. #3): This tag is designed to personalize each newsletter with the first and last name of the individual that is going to receive it. It creates empathy between individuals and their personalized newsletters and motivates them to interact and explore them. However, this ability is only available when the community manager uses a mailing list from a file. Otherwise, the tag will be removed from the newsletter (Figure 14. #3).
- **%categories%** (Figure 13): In order to allow individuals to define their preferences on the newsletters' categories, the community manager must include the categories' tag, which is then replaced by a link that individuals can click (Figure 15). By clicking it, individuals are then forward to a web page where all the categories are listed and they can choose which of those they would like to receive. A checked box means a positive preference and an unchecked box represents individual's lack of interest in that particular category (Figure 16).

Figure 12. Share, see online, and user tags

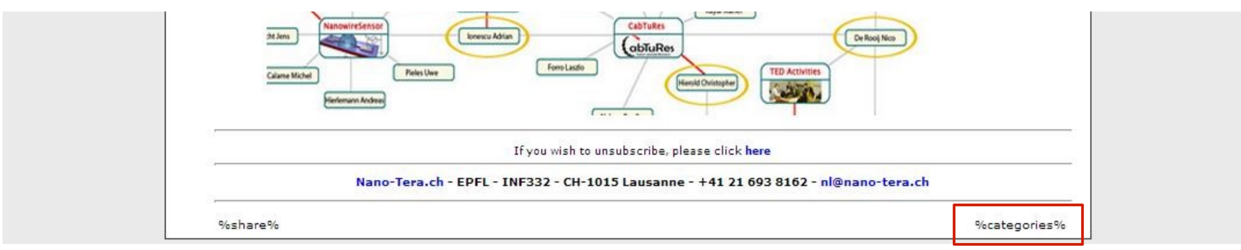


Figure 13. Categories tag

Figure 14. Text on share, see online, and user tags

Figure 15. Link on categories tag

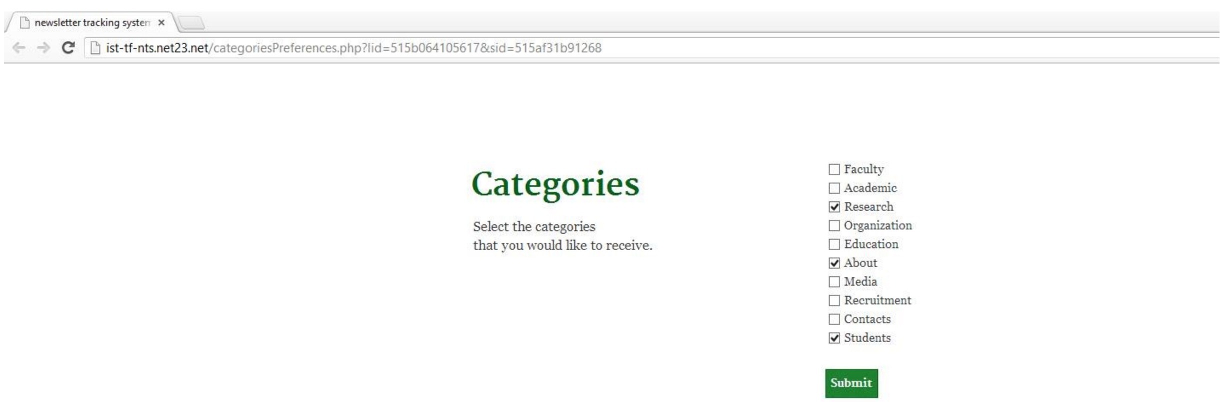


Figure 16. Categories example web page

4.5. Tracking Process

The core component in the NTS boils down to the tracking process, where all individuals' interactions are capture and stored into a database. Each time an individual clicks a link in the newsletter, the action is stored into the NTS as an interaction between the individual and the newsletter. The follow schema in Figure 17 explains how the tracking process is managed in the NTS and how actors play their roles. "Individual" plays the main role as the tracking process booster by performing the task of "Click a Link" on the "Tracked Newsletter". The task is then decomposed into the tasks "Track Interaction" and "Get Real Link".

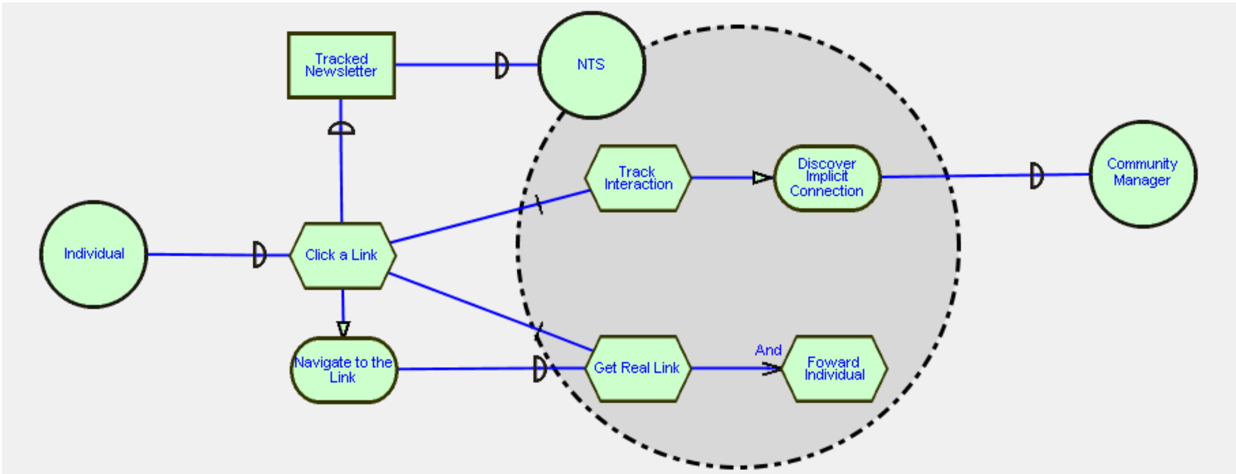


Figure 17. Tracking process

Before storing the interaction the NTS performs the task of getting the real link. It converts the tracked link into the original link to which the individual wants to navigate. This task is completed when all the tasks under it are also performed. In this case the contribution link "And" symbolizes that the parent "Get Real Link" is satisfied if the offspring "Forward Individual" is also satisfied. The NTS should satisfy the goal of "Navigate to the Link" by translating the tracked link into the real link and forwarding the individual. Once the individual navigates to the real link, the NTS loses the individual's track and no more interactions are stored.

The task "Track Interaction" is based on the data storage of the interaction. At this stage, the NTS stores all the information compiled on the tracked link. The system stores the information about the clicked newsletter, link, individual, and time. The information is stored into the database and used for data analysis. At the end, the task of tracking the interaction has the end of "Discover Implicit Connections" among individuals, where the "Community Manager" appears as the actor responsible for triggering the processes in order to meet the goal.

From the newsletter design to the fostering of an implicit community, the NTS is a solution based on the actors "Individual" and "Community Manager" to get the resources, perform the tasks, and achieve the goals. Through dependency links the schemas show how the solution works and how all the pieces fit together in order to achieve the goals.

Still in the tracking process, the NTS is responsible for storing the data on individuals' interactions in the newsletters. Although the data comes from the unique source of the newsletter, it represents different actions on the NTS. We then describe the different actions that can happen on the NTS and what they describe.

- **Open**: The action of "Open" comes from individuals' action of opening the newsletters' email and describes that the individual has opened his email and loaded the newsletter. This action is detected through an embedded image of 1x1 pixel at the bottom of the email message. This image is also known as "web beacon" or "tracker image." When the individual opens his email, the tracker image is downloaded from the server where the NTS is deployed and then counted as “Open.” However, if an individual has his email set to not show email images, even if he opens and reads the newsletter the NTS will not be able to capture the “Open” action. In this case, the only way to detect the action is to the individual to click a link. Then on the request for the link, the NTS will know that the individual read the email.

- **Share**: This action describes individuals’ willingness to share the newsletter. It results from his click on the share link described in the Chapter 4.4.1 under the tag %share%. In this action the NTS tracks that the individual have clicked to share the newsletter and that is available do it. However, it does not guarantee that the individual have really shared the newsletter. In fact, after selecting the option to share the newsletter, the individual is forwarded to a web page where he can select the way the wants to perform the sharing. However, this process is explained in the point bellow named “Social Network.” The information on individuals’ willingness to share is important for communities to understand the availability of their individuals to spread information.
- **See Online**: On this action individuals ask to see the newsletter in their Brower instead of reading it on their email addresses. It describes individuals’ interest on reading the newsletter but also their restrictions on the way they desire to do it. This feature is also designed to allow individuals to read the newsletter even if for some reason they are not able to properly read it on their emails (e.g. if they have the option to show the images off). However, despite individuals being forward to their Brower, the presented newsletter is still tracked in order to gather individuals’ interactions.
- **Click**: The action of “Click” results from individuals’ clicks on the newsletter’s links. It describes their interest on a particular link and subject. Each time an individual clicks a link it is stored as a relation between the individual and the link. In fact, the described action of “Click” is the most important due it is value as the core of fostering implicit connections among individuals. The most clicks the NTS is able to track the better will be the data on individuals’ preferences and therefore to discover implicit connections between them.
- **Social Network**: When chosen the option of sharing the newsletter, the individual is forward to a web page where he can select the way he desires to do the sharing – namely through digg, stumbleupon, twitter, Technorati, Facebook, newsvine, or email (Figure 18). On this step the gathered information notes the way the individual decided to share the newsletter. This information will allow communities to identify the most used ways to share, and keep the track over the shared newsletters once it creates a unique newsletter able to identify the source of the network where it was shared. For example, when a newsletter that was shared on Facebook is clicked, the NTS will know that the source of that click came from Facebook. This social network tracking is also available on the online newsletter through individuals’ Brower (Figure 19).

Once the request reaches the NTS in some of the ways listed above, the next step is to store the data on the action. The stored data is the minimum so that the NTS can gather it as quick as possible and forward the individual according to his request. The gathered data is described bellow.

- **Newsletter ID**: It represents the unique identifier for the newsletter and gives the information to which newsletter the individual has interact with.
- **Individual ID**: As for the newsletters, the NTS assigns to each individual a unique identifier in order to know who has performed the interaction.
- **Link ID**: It identifies the link to which the individual has interact with. It enables the NTS to discover what were the links with a higher percentage of interest from individuals.
- **Date**: The NTS registers the date (*yyyy/mm/dd*) on which the action was performed.
- **Hour**: It represents the detailed information on the date record. It is stored in the format *hh:mm:ss* and along with the date enables the NTS to understand the evolution of the interactions over time.
- **IP Address**: Represents the numerical label assigned to each device and it enables the NTS to get information on geographical location and information spreading. For instance, if two different IP addresses come from the same individual's ID, the probability that the individual had shared his newsletter is high.

After the data is stored the NTS is able to enable data analysis and allow communities to take value from the data. The following chapter (Chapter 4.6) describes the process of data analysis as well as the several ways in which the data can be treated by communities.

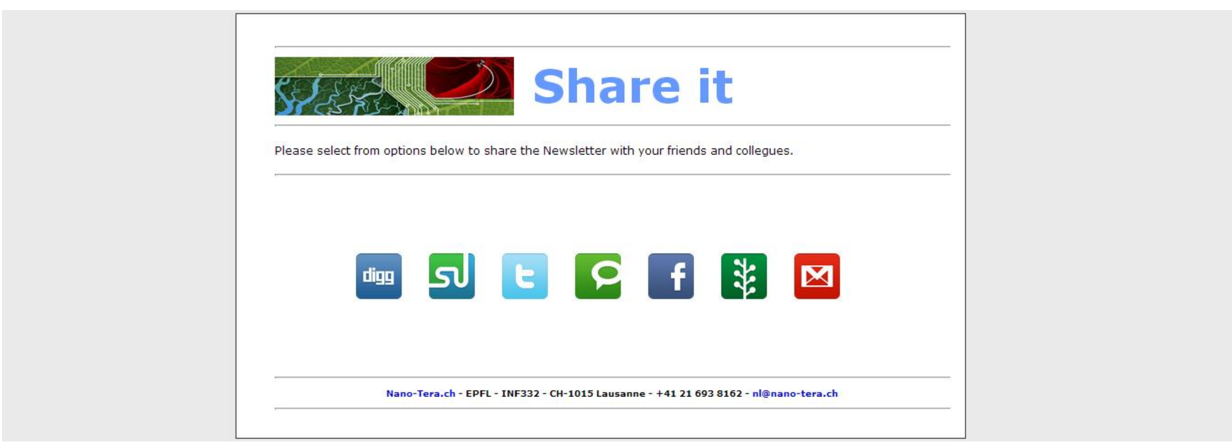


Figure 18. Web page to share the newsletter

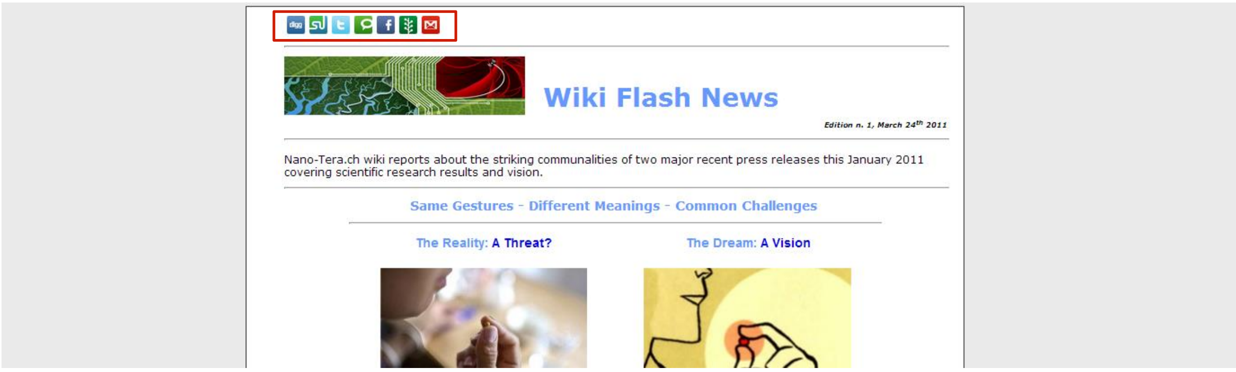


Figure 19. Share options at the online newsletter

4.6. Data Analysis

The captured data on individuals' interactions is the most valuable asset and the one that allows communities to discover implicit relations among members. The more data the NTS is able to capture, the stronger will be the results on individuals' implicit connections and the greater the value of the discovered networks. The NTS is able to expose the captured data by organizing it into visuals, through JpGraph, and allowing communities to export it.

Moreover, the NTS organizes the statistics on the newsletters by showing each newsletter at a time. The NTS presents a list of all the newsletters that were uploaded into the system and communities are able to select one newsletter at time and see the results until that moment. The results are divided into different sections that are presented below.

4.6.1. General Information

Although the analysis on a particular newsletter is available right after the newsletter is sent, it is up to the community manager decide when the analysis should be carried out. To help on this decision the NTS provides overall information on the newsletter statistical data, which may also help understanding the impact that the newsletter had on the community (Figure 20).

- **Total of Clicks (#1):** Represents the total number of clicks that the newsletter had receive. These clicks can come from the set of individuals to which the newsletter was initially send but also from any other sources that may have access the newsletter, for instance, a social network. At this point, any click is counted as one even if the individual had already clicked the link. This information allows communities to understand how interactive individuals were regarding the newsletter.
- **Total of Individuals that Clicked (#2):** This value shows the total number of individuals that have clicked the newsletter regardless the amount of clicks that they have made. The presented value enables communities to know the number of individuals that showed commitment and engage on a particular newsletter. However, at this point, the communities are not able to know which individuals are behind this number.

- **Total of Individuals that Open It (#3):** The presented value represents the total number of individuals that have open the newsletter at their email address. Once the process of spreading a newsletter across a community is done through email, it is important for communities to understand how many individuals have open their email and compare the data with the total number of individuals that have clicked the newsletter. This analysis will allow communities to understand the engagement that individuals had regarding the newsletter and the level of interest shown on a particular subject. However, this data only takes into account the individuals to which the newsletter was send through email.
- **Time (#4):** The value represents the total time that have passed since the newsletter was spread across the community. It is represented in the format of days and months, if applicable.

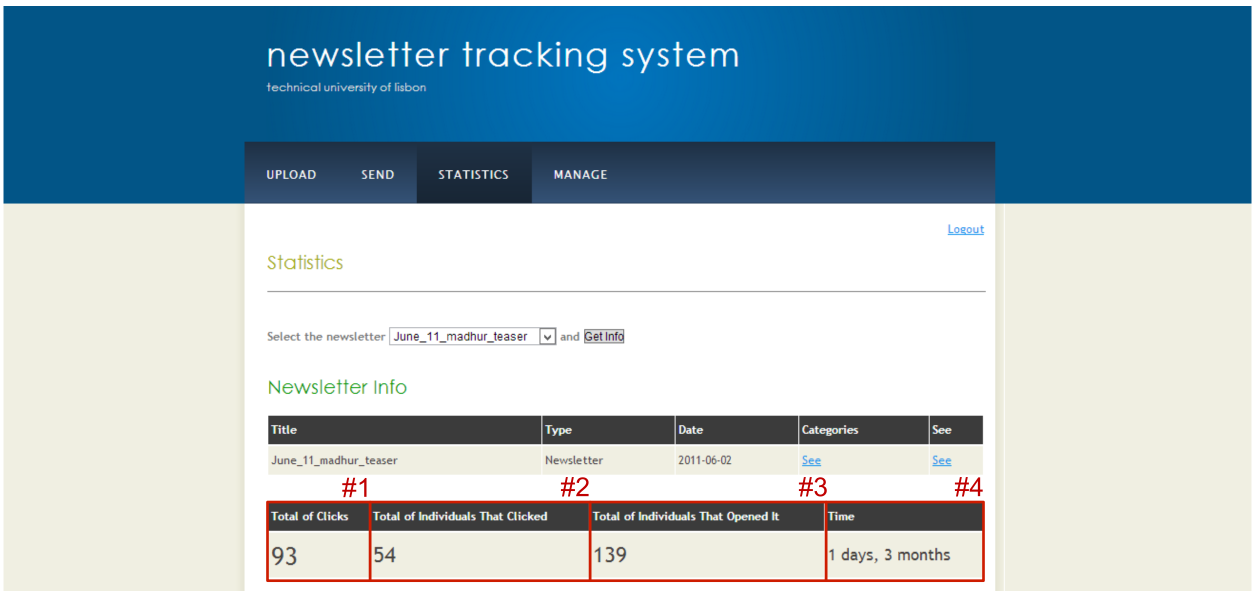


Figure 20. Statistics web page (general newsletter info)

This overall information is useful for communities to have an overview of the newsletter's impact and to monitor the results. At any time communities can access this information and decide whether or not they should go deeper in the data analysis and explore the data provided by the NTS. Based on the time passed along with the total number of clicks and the amount of individuals that have clicked, it is possible to understand if it makes sense to wait for more interactions or if the data analysis should be performed.

On a further analysis the NTS is prepared to better explore individuals' interactions, namely individuals' clicks on the newsletters. Every time a click happens the NTS stores it and then gathers all the data to display it through statistical results. The visuals on the interactions with the links are some of the ways used to present the results.

4.6.2. Links' Type

The design of the data in a newsletter can be identified as "Text" or "Image" (Figure 21), meaning that every link, to which individuals can interact through clicks, is assigned to a set of words, i.e. "Text" (#3), or to a visual, i.e. "Image" (#4). Each time an interaction occurs, the NTS stores the action of clicking and identifies the source as "Text" or "Image". This analysis aims to enable the understanding of what type of appearance individuals interact more, therefore if the members of a community are text- or image-focus. On a practical case, this information can be useful to help communities to better design newsletters in order to motivate individuals to interact. The higher the number of interactions, the more consistent will be the data on the implicit connections.

Figure 21. Share and see online options and links' type

The actions of "See Online" (#1) and "Share" (#2) are also presented on this analysis, with "See Online" being the action of asking to see the newsletter in the Brower and "Share" the willingness to share the newsletters with social networks or through email. This operation opens a new web page with the ability to share the newsletter across several social networks – digg, stumbleupon, twitter, Technorati, Facebook, newsvine – and through email. However, although the action of "Share" is performed there is no guarantee that the individual took the action until the end and shared the newsletter across the social network. This constraint lies on the fact that the last step for sharing at any social network is controlled by the social network itself.

Once the individual selects the social network and decides to share the newsletter, a new version is created out from the newsletter that he had receive. This step is designed to modify the links in order to capture the clicks that come from the social network with both information – from which social network and who share it. This allows communities to understand which social networks had shown more interest on the newsletter as well as who are the individuals with a strong impact on social communities.

4.6.3. Links' Categories

The ability to perceive to which subjects members reveal their interest is achieved by categorizing the links in the newsletters. Communities should categorize the links by firstly defining the categories and then assigning them to the links. Although each category can be assigned to one of more links, each link can only be assigned to one category. This process is done during the upload process and is explained in the Chapter 4.3.

At the end, the better links are categorized, the more accurate will be the results on individuals' preferences. The results are then presented as pie charts where each slice represents a category, and the value the percentage of clicks it received. At the chart's legend the percentages are translated into the total number of clicks. By going deeper into the links' categories analysis there is the option to see the details on each category and check the following information (Figure 22):

- **Category (#1)**: The category's name.
- **Total number of clicks (#2):** Represents the total number of clicks that a category received. However, it does not take into account the source of the clicks, meaning that all clicks can come from one or several individuals.
- **Elements (#3):** It presents the list of all individuals that have clicked any category's link, namely, individuals' first and last name, email address, and the total number of clicks. It also shows the other categories to which the individual have clicked and the total number of clicks.

Figure 22. Table of categories

4.6.4. Links' Ranking

At this analysis the links are all ordered by the total number of clicks received (Figure 23). The first link is the most clicked and the last one, the link with the lowest interaction. The links who did not received any click are not listed and do not appear in the raking. Associate to each link is also the category to which it belongs.

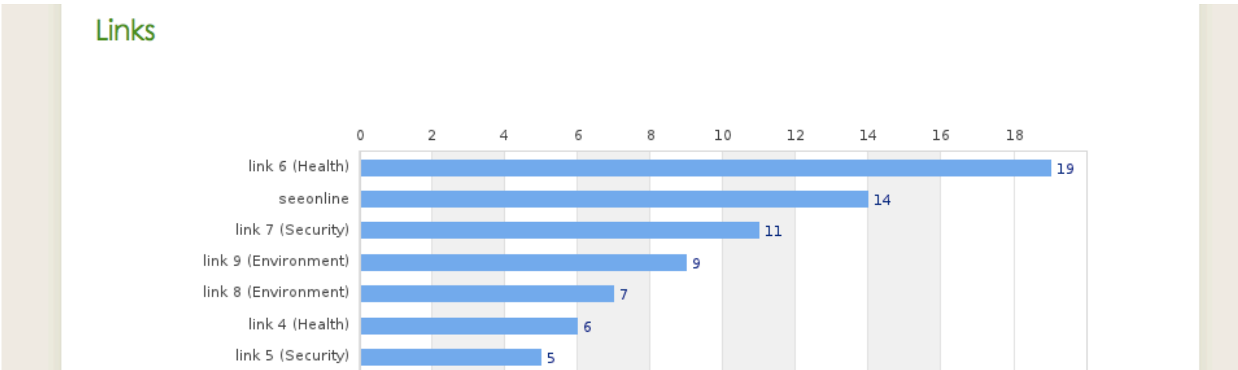


Figure 23. Links' ranking

Although the links' raking is useful to understand which links triggered the highest interest, this exposition is not practical and hard to associate with the newsletter due to the appearance of the links' ID and not the name, which will also be difficult to understand. However, in order to solve this problem, the NTS has the ability to present the newsletter together with the total number of clicks that every link received and the category to which the link belongs (Figure 24). This information is placed near to the link so the community manager can have a visual analysis on the newsletters interaction and understand which links had a higher impact. Moreover, it allows communities to know links' name and navigate through it.

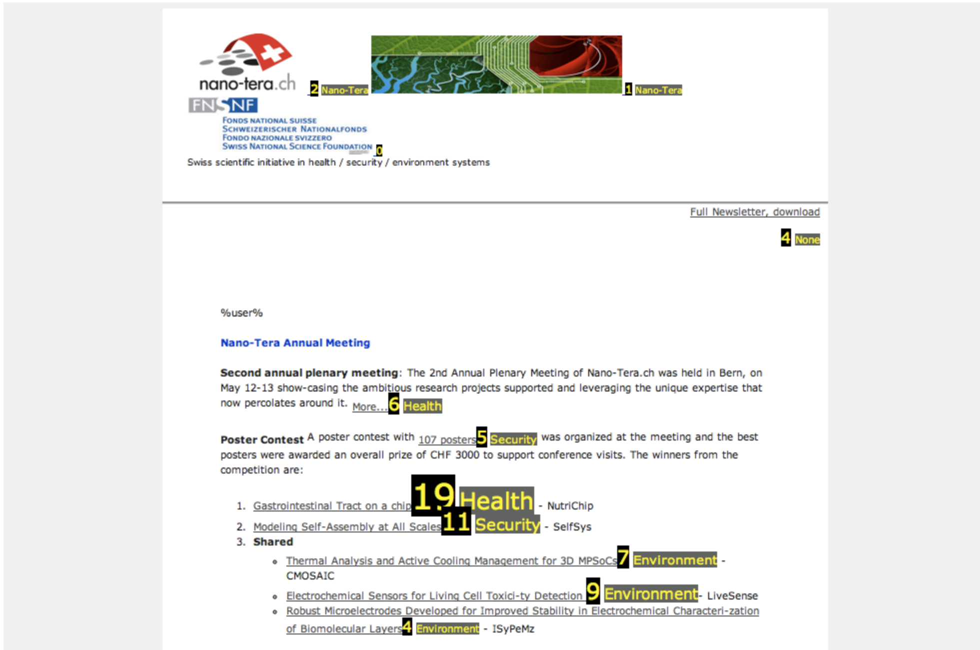


Figure 24. Total of clicks per link in the newsletter

4.6.5. Date

The total number of clicks is also organized in a date axis, where *y* represents the total number of clicks received in the date *x*. This analysis allows the understanding of interaction evolution where communities can observe the first newsletter's impact on individuals and how the total number of interactions evolves over time (Figure 25). The visual begins on the date (year-month-day) on which the newsletter was send and ends on the time past (days and months) since the last click. This analysis helps communities to realize which dates are more suitable to send the newsletters in order to achieve the best impact and therefore the higher interaction. Unfortunately the visuals do not go into time detail (hour, minute, and second) that could be useful to discover which time of the day is the best to send the newsletter. However, this information is recorded into the NTS logs and can be access through the server.

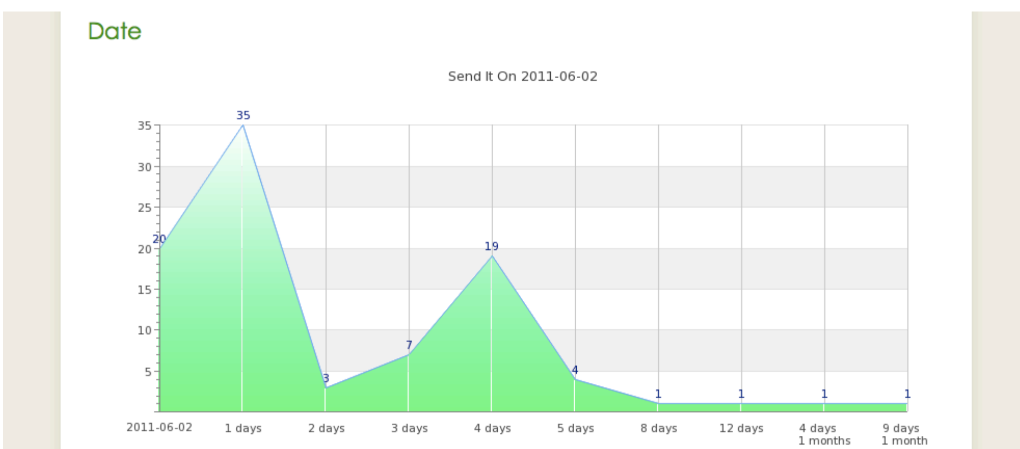


Figure 25. Number of clicks over time

4.6.6. Individuals

As well as links, individuals are also organized into a bar graph with the first individual being the one who has the higher number of clicks. The data is ordered in decreasing order regarding the total number of clicks per individual. This option allows communities to identify which individuals are more active and revealed more interest on the newsletters. On a further analysis individuals are organized into a table with the following information per column (Figure 26):

- **Info (#1):** Individuals' first and last name and email address.
- **Clicks (#2):** Total number of clicks performed on the newsletter.
- **Categories (#3):** List of categories and the total number of clicks on each. This data brings to light to which categories the individual showed more interest as well as the strength through the number of clicks.
- **Implicit Relations (#4):** List of all individuals to which the individual is implicit related. For each implicit relation it is presented the individual's first and last name, email address, and the categories on which individuals are implicit related. This information enables the exploitation of every implicit connection and lists the individual's network.
- **Results (#5):** Shows the results on individual's clicks per link on the newsletters as presented in Figure 24.

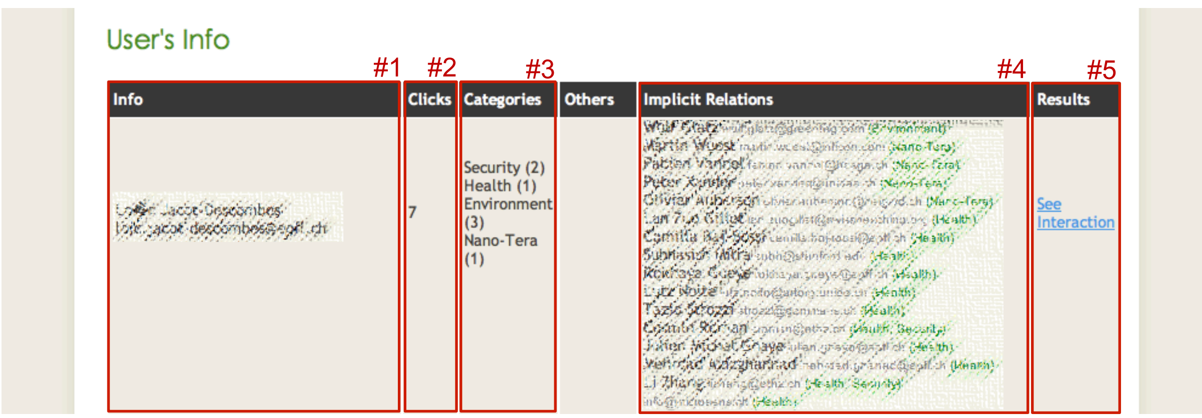


Figure 26. Individuals' implicit connections

4.6.7. Vizster

"Vizster" was implemented by Heer and Boyd (2005) with the goal of "building a visualization system that end-users of social networking services could use to facilitate discovery and increased awareness of their online community" (Heer & Boyd, 2005), in other words, an interactive system to visualize online social networks and allow the exploitation of community structure. A network is presented as an egocentric network once it places an individual as the center and then his immediate connections, creating an indirect graph of social activity. An animated layout is used to organize the network and to provide a clear visualization. It also allows a useful linkage view, by clicking two nodes and highlighting the common connections between them.

For the community identification and visualization Vizster is extremely powerful as it is able to identify and represent community clusters at the various stages of Girvan-Newman Algorithm allowing the user to search for alternative community configurations. In addition, the application has a search engine, which enables communities to search for keywords highlighting the nodes where the keywords exist. The network design is based in the Shneiderman's mantra of "overview first, zoom and filter, then details-on-demand", instead of the classical approach of "start with what you know, then grow" (Heer & Boyd, 2005) and has an X-ray mode in order to explore profile attributes. Since Vizster is also an open source tool, communities can use its functionalities to identify and visualize clusters as well as shaping it according to their needs.

On the analysis of links' categories (see Chapter 4.6.3) NTS offers the possibility to download the collected data into a XML file which can then be used in Vizster. The tool allows communities to explore implicit relations and discover how individuals are organized by grouping them according to their interactions on the newsletter. At the end, individuals are grouped and communities have an overview of how individuals relate to each other (Figure 27).

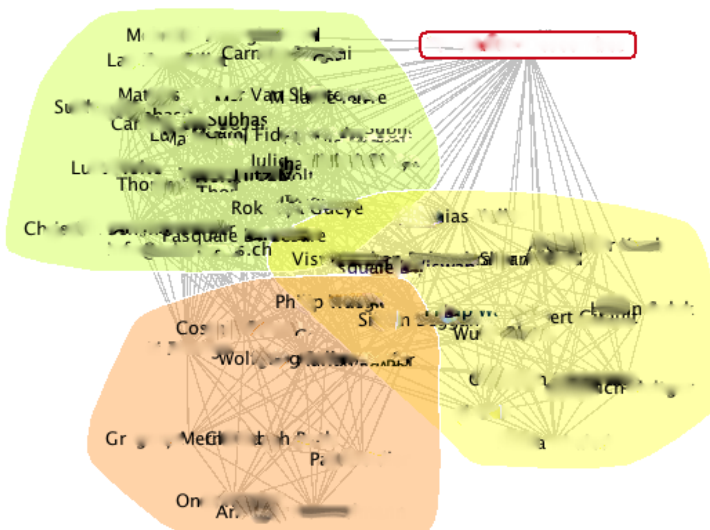


Figure 27. Example of clusters discovered with Vizster

5. Connection Degree

With communities having the ability to discover implicit connections among members, it becomes important to understand the value of each connection in the discovered universe. In a scenario where all members interact with all the categories, they will be all implicit connected and the community manager will find it hard to take any conclusion. On the other hand, if there is a way to measure the value of each connection in the network and assign to each connection a value from a common formula, the community manager will have the chance to create his own threshold and filter the implicit connections.

The "Connection Degree" (CD) model proposes a way to measure the connection strength between every two individuals. The higher the CD, the stronger the relationship between the two individuals.

The model estimates the connection degree based on both explicit and implicit connections. It takes the two types and uses them to calculate the corresponding explicit and implicit degrees on the CD formula. While the implicit degree is based on individuals' implicit interaction on the newsletter's links and categories, the explicit degree comes from individuals' explicit preferences on the categories.

Explicit Degree

An individual can express his preferences on categories by explicitly checking a category as preferred. This process is done through the NTS, where individuals can navigate to a preferences web page through their newsletters and check or uncheck their preferences on the categories. Thus, a checked category is understood as a positive preference, an unchecked category as a negative preference, and an unknown preference if no explicit action is performed. The explicit degree is defined for each category and will affect the final CD in a "Category Importance" (ci) value, defined by the community manager and that represents the importance of the individuals' preferences in the equation.

$$explicit\ degree_{for\ a\ category}\ (Individual): ed_c(I) = \begin{cases} ci, & positive\ preference \\ 0, & no\ preference\ shown \\ -ci, & negative\ preference \end{cases} \quad 0 \leq ci \leq 1$$

Equation 1. Explicit degree for an individual on a category

For instance, on a newsletter with a ci of *0.4* and the categories *A*, *B*, and *C*, if an individual *I* selects the categories *A* and *C* as preferred, his explicit degree for each category will be the following.

$$ed_A(I) = 0.4;\ ed_B(I) = -0.4;\ ed_C(I) = 0.4$$

The value of *0* is the value assigned by default to each of the ed_c, meaning that the individual has not yet shown any preference. When the individual decides to selects his preferences on the categories, the value *0* is replaced by the *ci* if he checks the category or *–ci* otherwise.

Implicit Degree

Implicit degree is calculated based on individuals' clicks on the newsletters. The action of clicking a link is stored as an implicit connection between the individual and the link ("individual-link"), and between the individual and the link's category ("individual-category"). The connections are then used to calculate the implicit degree on the CD equation. On the first relation "individual-link", the connections are organized into a matrix $n \times n$ where both rows (n) represent individuals and the values (v_{nn}) the total number of links that both individuals have clicked in common.

$$n \times n = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{bmatrix}$$

Equation 2. Calculation of the total number of clicks in common for every two individuals

On the relation "individual-category", individuals are also organized into a matrix where rows represent all pairs of every two individuals (I_i, I_j) and columns the categories (c_l). The value is represented on the formula:

$$(I_i, I_j) \times c_l = min(a_{il}, a_{jl}) + (ed_i + ed_j)$$

Equation 3. Implicit value for every two individuals (i and j) on the category (l)

Where a_{il} represents the total number of clicks that individual I_i gave in the category c_l and the ed_i the explicit degree (Equation 1) for the individual I_i. The implicit value for the two individuals (I_i, I_j) in the category (c_l) is then calculated based on the minimum value of both individuals' total number of clicks in the category $(min(a_{il}, a_{jl}))$ plus the sum of both explicit degrees $(ed_i + ed_j)$. The first part of the equation represents the minimum value on both individuals' categories-based relation and the second part expresses individuals' explicit interest on the categories. The explicit degrees are added to this equation once it contains the calculation on individuals' implicit connections per category.

The next step translates the implicit relations "individual-category" into "individual-individual", where individuals are organized into a matrix $n \times n$ with rows and columns (n) representing individuals (I) and the values the sum of all Equation 3 for all l categories ($\sum_{i=0}^{l}$ (Equation 3)).

$$n{\times}n = \begin{bmatrix} \sum_{i=0}^{l} (I_1, I_1){\times}c_i & \cdots & \sum_{i=0}^{l} (I_1, I_n){\times}c_i \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^{l} (I_n, I_1){\times}c_i & \cdots & \sum_{i=0}^{l} (I_n, I_n){\times}c_i \end{bmatrix}$$

Equation 4. Implicit degree for every two individuals based on all categories

The final value for the CD between every two individuals is calculated by performing a syntax sum of both resulted matrix from Equation 3 and Equation 4 but with the multiplication operator. At the end only one of the sides of the matrix is taken into account, i.e. the lower or upper triangular matrix, and excluded the diagonal. This will ignore duplicated pairs and take only one CD into account.

$$connection\ degree : n{\times}n = \begin{bmatrix} (\text{Equation } 2)_{11}{\times}(\text{Equation } 4)_{11} & \cdots & (\text{Equation } 2)_{1n}{\times}(\text{Equation } 4)_{1n} \\ \vdots & \ddots & \vdots \\ (\text{Equation } 2)_{n1}{\times}(\text{Equation } 4)_{n1} & \cdots & (\text{Equation } 2)_{nn}{\times}(\text{Equation } 4)_{nn} \end{bmatrix}$$

Equation 5. Final value for the "Connection Degree"

The multiplication of both values brings together the implicit connection on links and categories, and reveals the final connection degree for every two individuals. The multiplication as the final operator helps to highlight connections where a click can make a difference. Once the results are based on newsletter's interaction, each click should have a significant value so it can positively influence the CD. The CD will allow communities to compare implicit connections and discover which ones are more important.

The CD model brings to the NTS the value on measuring implicit connections and allowing communities to have a better overview of the implicit relations between individuals. While NTS is responsible for capturing and discovering implicit connections, the CD model is designed to assign to each connection a CD value.

Connection Degree and the Newsletter Tracking System

While the NTS is able to capture members' interactions (with newsletters) and foster an implicit community, the CD is able to add a connection degree value to each members' connections. This way, the CD is calculated when it has a defined set of nodes and links (i.e. a graph) on which it can measure links' connection degree. Moreover, it is designed in order to have the NTS as the source of its graphs, meaning that the CD will calculate links' CD based on the methodology adopted by the NTS, namely individuals' clicks on newsletters' links and preferences on the categories.

The strong relationship between the two solutions starts on the development of the NTS through the categories' tag (see Chapter 4.4.1). During the sending process the NTS replaces the categories' tag for a link through which individuals can set their preferences on the newsletter' categories. These categories are defined by the community manager and apply to all newsletters (see Chapter 4.3). Once individuals reach the web page to define their preferences, a list of all categories appears to check or uncheck the categories' check boxes (Figure 28). Therefore, a positive preference is described as a checked box and a negative preference as an unchecked box. For instance, in Figure 28 the individual has defined the categories “Academic”, “Education”, and “Recruitment” as his positive preferences. As for the others, because none of the boxes are checked his preferences are seen as negative. This action will impact individual’s explicit degree and therefore his connection degree with other individuals. In particular, the impact is defined by a “Category Importance” (ci) (see Equation 1) which is set by the community manager and represents the level of importance that the action of selecting the categories’ preferences should have in the final calculation of the CD.

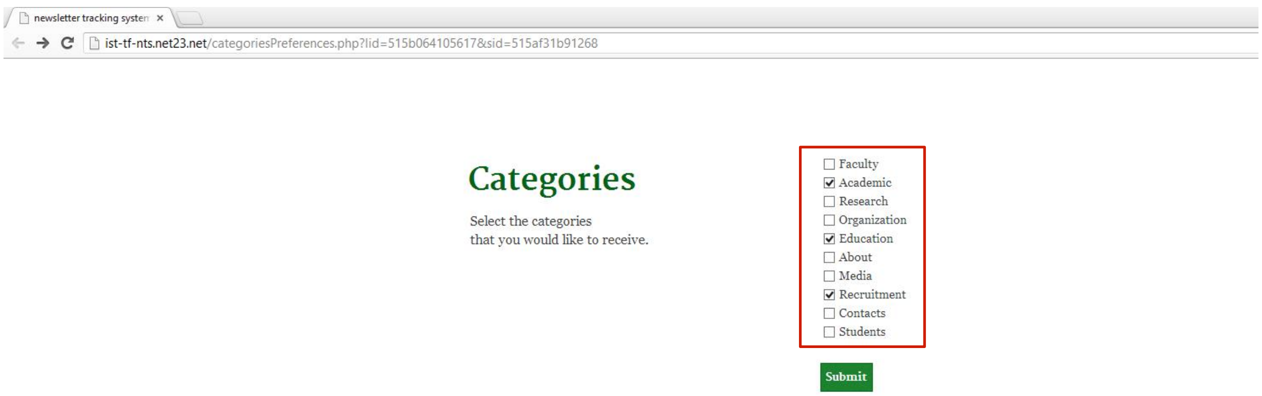


Figure 28. Example of a subscriber categories preferences

On the NTS statics panel the community manager has the ability to download the connection degree’s data (Figure 29 #1). This information is downloaded to all newsletters and takes into consideration the interactions gathered by the NTS on all the newsletters. When clicking the link, the NTS applies the CD model into the implicit network discovered until that moment. By gathering all individuals’ interactions with newsletters and their preferences on the categories, the NTS is able to run the CD model and to build a downloaded file for the community manager to access the CD values for every individuals’ implicit connection.

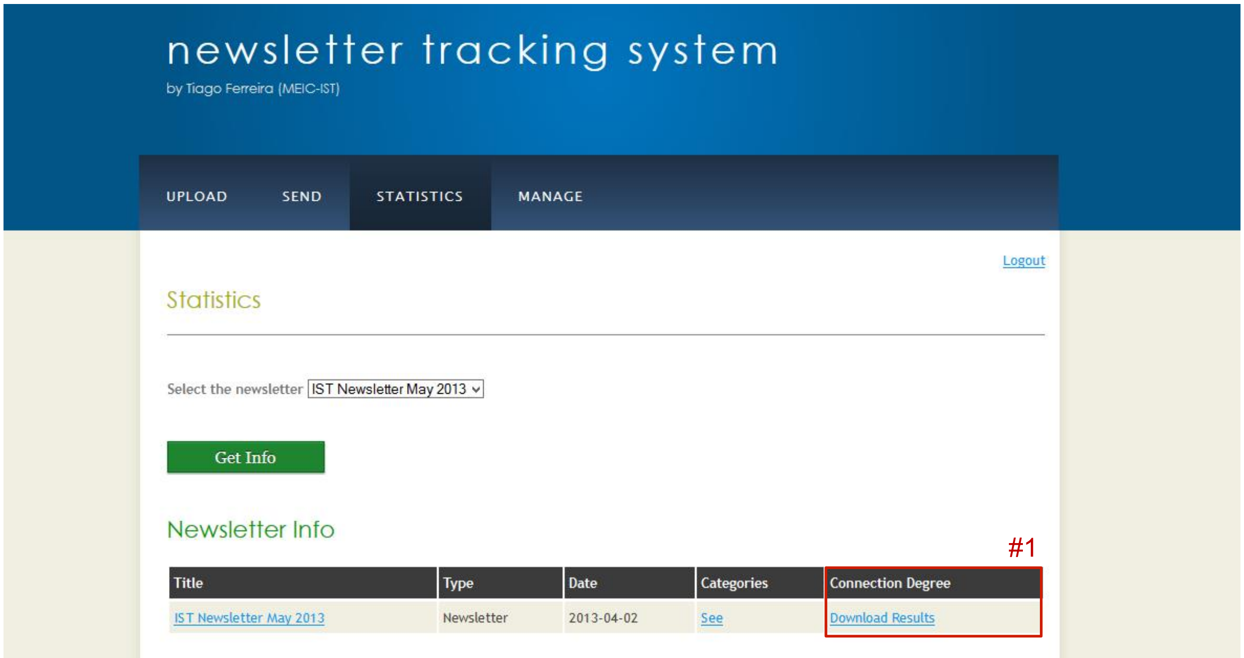


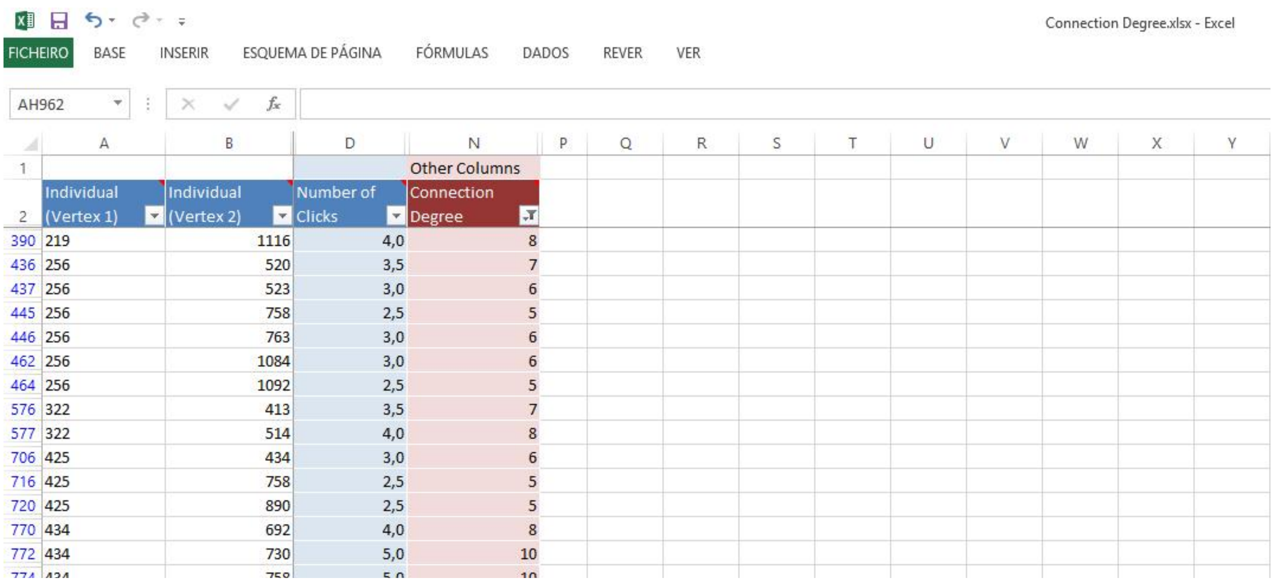
Figure 29. CD download option on the NTS

At the end, the downloaded data is organized into pairs of individuals where each line represents a pair of individuals ($n \times m$) and the connection degree assigned to the relationship $I_n \times I_m$ (see Equation 5). The results are downloaded as comma-separated values (CSV) file which can then be processed by the community manager. Through the example on Figure 30 we can see that the downloaded file has the following format:

- **Column 1**: List of individuals IDs with CDs. The total number of times that an individual ID appears represents the total number of implicit relations that that individual has with other individuals. Therefore, each line represents a unique implicit relationship.
- **Column 2**: Represents the list of individuals IDs that have a CD with the individual ID on the Column 1. Both columns 1 and 2 are filled with the individuals IDs defined by the NTS. This option of showing individuals' IDs instead of their personal information is to respect their privacy.
- **Column 3**: CD value assigned to the relation between the two individuals from Column 1 and 2.

Figure 30. Example of CD downloaded file

The presented file format has the purpose of allowing the community manager to take advantage of NodeXL template. “NodeXL is a free, open-source template for Microsoft Excel 2007 and 2010 that makes it easy to explore network graphs” (NodeXL, 2012). With NodeXL, the community manager can enter the NTS discovered network edge list in a worksheet, click a button and see the graph along with the CDs, all in the familiar environment of the Excel window. By copying the three columns into NodeXL template the community manager can use it to cluster individuals and do analysis. The following example (Figure 31) shows how the three columns were copied into the NodeXL template, namely Column 1 into the first column, named “Individual (Vertex 1)”, Column 2 into the column “Individual (Vertex 2)”, and the last column which represents the CD value (Column 3) into the fourth column labeled as “Connection Degree.” The presented example also contains a column with the total number of clicks that both individuals have in common, which could be useful to know its relation with the CD.



	A	B	D	N
1				Other Columns
2	Individual (Vertex 1)	Individual (Vertex 2)	Number of Clicks	Connection Degree
390	219	1116	4,0	8
436	256	520	3,5	7
437	256	523	3,0	6
445	256	758	2,5	5
446	256	763	3,0	6
462	256	1084	3,0	6
464	256	1092	2,5	5
576	322	413	3,5	7
577	322	514	4,0	8
706	425	434	3,0	6
716	425	758	2,5	5
720	425	890	2,5	5
770	434	692	4,0	8
772	434	730	5,0	10

Figure 31. NodeXL template data example

Each line in the CD Excel file represents an edge of the implicit network. Thus, the individuals’ ID in the Column 1 and 2 represent the vertex 1 and 2 of the edge respectively. As for the connection degree value, this represents the importance of that edge on the graph. The proposed CD model reveals how strong two individuals are connected in the network. The higher the value, the stronger is the relationship and therefore the importance in the network.

Once the CD values are copied into the NodeXL Template the community manager is able to use the template features to explore the NTS implicit network based on the CD values. The NodeXL allows the community manager to filter the data, do clustering and grouping, explore graph metrics (e.g. Betweenness, Closeness, Clustering Coefficient), and create visuals of the network. In fact, the overall features offered by NodeXL can be explored on their documentation by at NodeXL web site (NodeXL, 2012).

In short, both NTS and CD are related to each other and able to add value to what they propose. By discovering an implicit network of relations between individuals, the NTS is capable of explore individuals' interaction with newsletters and also their value in the network. This ability allows the community manager to understand how a community is organized and how individuals related to each other, along with their preferences on the categories. However, it is useful for a community to also understand in what way an implicit connection is important in the universe of discovered connections. To answer this question, the CD suggests a model in which the individuals' total number of interactions and preferences in categories are taken into account in order to calculate a connection degree value for each pair of individuals. This ability enables, for instance, the community manager to bring to light the most important connections.

The relation between the solutions NTS and CD comes with the value that both are able to add to each other. The NTS provides to the CD the implicit network on which it can calculate the degrees as well as the data on individuals, namely their total number of clicks and preferences. On the other side, the CD is capable of add value to the implicit network by assigning a value to every edge and revealing the importance that an edge has in the network. Therefore how important the edge is for the community.

However, the real value comes from the community manager being able to understand the data and use it to promote the community development. Once the data is available and the tools to explore the data are provided, it comes to the community manager to exploit the data and transform it into valuable information for the community.

6. Evaluation

This chapter discusses the evaluation made to the NTS and CD within a scientific community and it shows its features, benefits, and capabilities. Both proposed solutions were tested in order to understand how successfully the community requirements were handle and the value that a community was able to take from the gathered data. At the end, the evaluation was crucial to reach a final version with an higher quality regarding community needs.

6.1. Nano-Tera.ch case study

The evaluation of this research was done within Nano-Tera.ch community – a scientific Swiss federal program with more than 40 projects on the subjects of "Security", "Health", and "Environment". Nano-Tera.ch diversity goes from different projects to hundreds of researchers around the world. Thus, in order to capture knowledge on the Nano-Tera.ch community, the management structure decided to run the research project "Community Knowledge Development" (CKD). Within other goals, the CKD was trying to understand how researchers were related to each other and how connections could be explored in order to improve researchers' work and promote collaboration.

The research project CKD supported the development and evaluation of both NTS and CD model at Nano-Tera.ch between February – July 2011. The project took place during the 6 months and the community management office designed 6 newsletters to send to the Nano-Tera.ch community in order to evaluate the proposed solution. However, the presented results are based on the top 3 of the newsletters due to their interest and reliability on results. The overall information on the newsletters involved more than 3000 of emails sent (1000 per newsletter) and about 1750 of captured clicks. In the next chapters we will explore the results gathered at Nano-Tera.ch evaluation, organized in by the different ways offered through the NTS.

Background

The scientific community of Nano-Tera.ch is a community based in Switzerland and has more than 40 projects in health, security, and environment. Its vision challenges the convergence of expertise from different fields as a strategic approach to stimulate innovation. Towards this strategic goal, Nano-Tera.ch is one of the scientific communities trying to understand its researches' behavior and implementing a strong research on interactive communities by offering two different approaches. "Inside-out Perspective" where each project has its own Wiki Page and it is "strongly encourage to participate and share information such as abstracts, news, didactic videos, and interesting results published" (Nano-Tera.ch, 2011). And the "Outside-in Perspective" where Nano-Tera.ch has defined "several themes to expand the vision of the application potentials for each research field" (Nano-Tera.ch, 2011). The goal is to trigger the interest from the outside communities about Nano-Tera.ch projects, as well as gain larger exposure to obtain contributions from the outside world.

On the February 2011, Dr. Peter Edward Bradley, the Executive Director of the Nano-Tera.ch scientific program for Health-Security-Environment Systems Engineering supported by Prof. Chris Tucci, Professor of Management of

Technology, had initiated a knowledge based community development strategy and found interesting to join efforts in order to develop a system capable of gathering knowledge on a community based. Thus, the proposed solution of the NTS was able to answer Nano-Tera.ch needs and enable the progress of gathering knowledge on the community using an automatic and invisible approach.

On the next chapters will be presented the results from Nano-Tera.ch evaluation and how they were applied in order to take advantage from the gather knowledge. The results are organized into the several ways as the options offered by NTS for presenting the results. On Chapter 6.1.1 the results are organized according to their type (e.g. Text or Image), Chapter 6.1.3 takes into account links' categories and arranges them through categories, Chapter 6.1.4 is based on the CD model and presents the results using the CD values and grouping the individuals according to it, lastly Chapter 6.1.5 uses Vizster software to organize and present the results.

Although the data is presented in different ways, it comes from the same source, namely individuals' interactions with newsletters, and results from NTS automated engine to capture individuals clicks and store them into the database. We should note that all the subscribers involved in the study were notified about the goal of gathering information on their interactions and what was the purpose of Nano-Tera.ch in doing so.

6.1.1. Overall

The evaluation carried at Nano-Tera.ch was performed between the months of February and July 2011. During this 6 months were designed 6 different newsletters in which were include several news about each of the Nano-Tera.ch research fields. However, the results presented and discussed in this thesis are based no the top 3 of newsletters regarding their level of interest and reached conclusions.

The results are based on more than 3000 of emails sent – 1000 emails per newsletter, which is the same thousand for all the 3 newsletters in order to gather information on the same individuals. If we perfectly divide the total of clicks (1732 of clicks) per newsletter we get about 577 of clicks per newsletter, meaning that for the total of 1000 emails sent (i.e. 1000 different emails IDs and therefore different individuals) were collected 577 clicks. By looking at the result in detail we discover that only 237 individuals have clicked on the newsletters, which means that in average individuals have clicked in about 2.4 times per newsletter. However, this analysis is not perfectly correct once there are individuals who have clicked more than once. The following tables (Table 4 and Table 5) summarizes the overall results gather from these 3 newsletters.

Table 4. Overall results

Description	Data
Total of newsletters	3
Total of emails sent (an email per individual)	3000
Total of individuals who opened the newsletters	374
Total of clicks	1732
Total of individuals who clicked	237
Average of clicks per individual	2.4

Table 5. Total number of individuals on each set of clicks

Number of Clicks	Total Number of Individuals
1	128
2	43
3	26
4	9
5	13
6	5
7	6
8	5
12	1
20	1
Total	237

6.1.2. Links' Type

The results on links' type try to understand to what kind of information-exposure individuals interact more. Nano-Tera.ch designed their newsletters based on information as "text" and "image" (see Figure 21). Each newsletter's topic was introduced with an "image" followed by a "text" to which individuals could interact and reach the same content. Although the action of clicking any of these types has the same impact on discovering implicit connections, this distinction is important to better design newsletters. If individuals interact more with images, it may be a good idea to design a newsletter based mostly on images.

The purpose of this analysis lies on trying to understand how newsletter can be designed in order to get the maximum number of interactions. The better newsletters' content encourages individuals to interact, the higher will be the total number of interactions. Because individuals' interactions are the unique source for the NTS, communities should focus on trying to understand how to get a valuable level of interactions and therefore improve their results.

This way, the analysis offered through links' types enables communities to gather information on the first contact that individuals have with their content – i.e. newsletter design. Thus, if a community is able to better design its newsletters in order to catch individuals' curiosity and willingness to click, the better will be the gather knowledge on individuals' interactions.

It is also important to notice how communities should properly select the content to include in the newsletters. Although the design could be important, the newsletters' content is decisive on convincing individuals to interact and explore more on the subject. By raising individuals' interest, communities can more easily get their commitment on interaction and therefore gather more information on them.

Regarding Nano-Tera.ch case study, individuals have showed their strong interest on information exposed as "text" with more than 96% of interactions on links linked to text and 4% on links assigned to images (Figure 32). The results reveal that when facing with an image and text, Nano-Tera.ch members highly click on the text in order to found more about the subject.

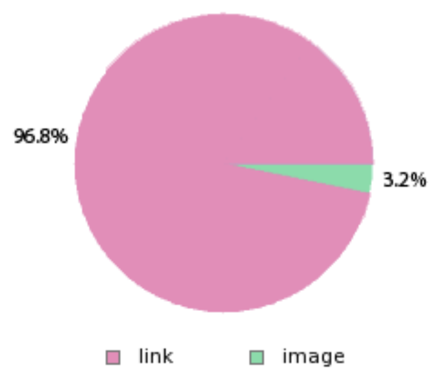


Figure 32. Links' type results

6.1.3. Categories

The results on categories are some of the most important in the context of this research. Categories are what define the groups and what help communities to understand which subjects have an higher interest from individuals. The properly definition of categories in the NTS along with their assigned to newsletters' links will defined the quality of these results. Meaning that, at this point communities should have their categories defined and each of the newsletter's links assigned to one of the categories (see Chapter 4.3).

With links assigned to categories, a click on each of these links is also translated into an interaction with its category. Although individuals interact with links by clicking them, they are implicitly interacting with categories and therefore proving the NTS with information about their preferences. Thus, the results on categories come from individuals interactions with clicks and not directly to the categories. This way, a community should defined newsletters' links in order to better meet the categories.

In the case of Nano-Tera.ch, it was defined a category for each of its research topics – Health, Environment, and Security – plus a topic related with the community itself – Nano-Tera.ch. The categories were all organized so they all were present in the newsletters, meaning that each newsletter was designed in order to have content on each of the categories. With this, Nano-Tera.ch wanted to understand, on each newsletter, what were the individuals' preferences regarding all the research topics.

At the end, the results showed individuals preferences on the category of Health (38%), followed by Environment (26.6%) and Security (21.5%), and finally Nano-Tera.ch (7.6%) (Figure 33). It is interesting to notice that although Nano-Tera.ch has several categories, the main stream of clicks was to Health category, which may be related to the fact of having an higher number of projects on Health regarding other categories. Another interesting aspect is the percentage get by Nano-Tera.ch category, meaning that almost 8% of the clicks were with the purpose of better know Nano-Tera.ch community. The small number of 8% has more value when compared with the total number of links assigned to each category. Although the links were equally divided into the different categories, the number of links assigned to Nano-Tera.ch community was relatively small.

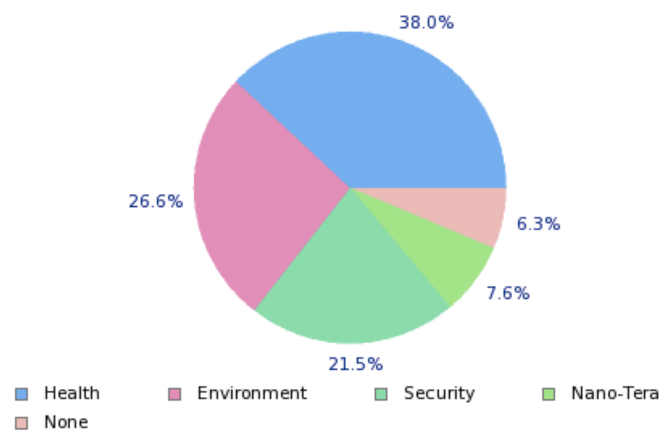


Figure 33. Links' categories results

6.1.4. Connection Degree

Once the data on individuals' interactions is gathered, the NTS is able to build an implicit network and present it to the community manager, which is then capable of having a better understanding of how individuals relate to each other. However, in a complex network where the total number of edges is impractical it is very hard to understand individuals' relations and take value from the implicit network. In this case, the CD is able to add to every edge a value that reflects how strong a connection between every two individuals is. The higher the CD value, the stronger the connection. Regarding complex networks, the value of the CD also comes with the ability to provide to communities a way to filter connections. If a network has a high number of edges, these can be filtered based on restrictions applied to the CD (e.g. edges with a CD higher than 7).

In order to illustrate the results on the CD model it was used the NodeXL template. Based on Microsoft Excel, the NodeXL enables the exploration of communities organized as a graphs. In Nano-Tera.ch case, individuals (nodes) and implicit connections (edges) were organized into a graph and the calculated CDs were used as input values to calculate edges' width and define edges' labels. The presented results are based on the implicit relations with a CD higher than 7, in order to highlight the most important connections and have a clearer view of the relevant nodes on the network (Figure 34).

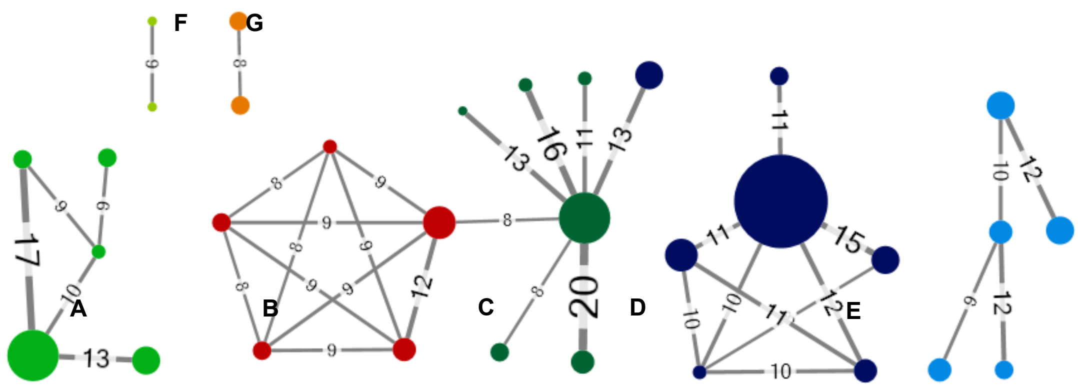


Figure 34. Connection degrees

The size of the nodes represents the total number of clicks that the individual has performed and the label on the edges the CD degree calculated to the implicit connection. The colors represent the Girvan-Newman algorithm only applied to the presented connections. By looking at the results, the higher CD has a value of 20, with both individuals having more than 6 clicks on the exactly same links on the newsletters. Adding to this, their strong relation is based on the "Security" category, represented by cluster C. The biggest node represents the individual with more than 20 clicks in the newsletters and with a strong interaction on the "Health" category (cluster D). On the other side, individuals at the cluster A ("Nano-Tera.ch") have also a high CD between them, followed by the cluster B ("Health" and "Security"), the cluster E ("Environment"), and the two isolated clusters F and G with almost all categories. The CDs at the cluster C show that the "Security" category is the strongest between individuals with a high CD.

6.1.5. Clusters Detection

With the NTS ability to implicit connect individuals in a network, it is interesting to explore those connections and understand how individuals relate to each other. Meaning that, instead of having a network with individuals implicit connected through edges and do not treat the data, it may be a valuable asset to exploit groups and bring to light hidden patterns as well as understand how individuals can be organized regarding their preferences.

On Nano-Tera.ch case study it was clear its interest on understanding how individuals could be grouped. By knowing that Nano-Tera.ch would have a better overview on individuals' relationships and how to organize the community according to its categories (i.e. subjects of research). We note that the names on the followed results are overshadowed in order to preserve individuals identity. Moreover, the results were achieved through the tool Vizster

(Heer & Boyd, 2005), which is a social network visualization system that uses Girvan-Newman algorithm to discover clusters in a network of connections. In fact, the use of Vizster to perform this final task brings value to the tool itself and helped to have as major focus the NTS discovered network instead of developing a way to build a network visual. Moreover, Vizster properly answers community needs on clustering detection as well as keeps the requirement of being an open-source tool achievable.

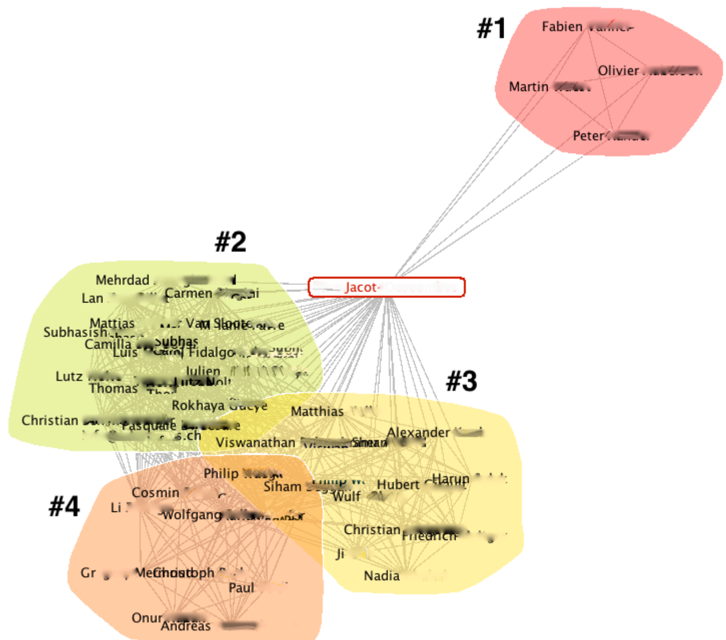


Figure 35. Clusters detection

By downloading the data through the NTS (see Chapter 5), communities are able to use Vizster to visualize the discovered implicit network and organize individuals into clusters. The results on the Nano-Tera.ch community showed that individuals are organized into 4 different clusters (Figure 35). The details show that the Cluster #1 is mainly based on individuals with strong interest on “Nano-Tera.ch” category; the Cluster #2 on individuals with high interest on “Health”; and Cluster #3 and #4 on individuals highly interested on “Security” and “Environment” respectively. From the image we can also see that the clusters’ size decreases by starting on the category “Health”, “Security”, “Environment”, and “Nano-Tera.ch”, which translates into the Cluster #2, namely on “Health”, having the highest number of individuals while “Nano-Tera.ch” cluster (#1) having the smallest number of individuals in it.

It is interesting to note that the results go along with the total number of clicks in the categories (see Chapter 6.1.3), meaning that not only “Health” category has the highest number of clicks as well the highest number of individuals in its clusters. This shows that the high percentage of clicks on “Health” category comes from different individuals and not from a small groups of individuals highly interested who have clicked several times on links related to “Health”. Therefore, “Health” is the category with the highest value in quantity (total number of clicks) and quality (total number of individuals on its cluster).

Along with the implicit network and because the NTS also has individuals' personal information (e.g. names, email, and institution of work) this analysis on clustering allows the community manager to not only understand how individuals can be organized as well as who the individuals are. Instead of having a unique ID numbers, which continues to bring value on the understanding of individuals clustering, the NTS offers the ability to identify the individuals through their personal information and therefore bring a face to each node. The value comes when the community manager is able to identify every individual and understand how they relate to each other. However, the data on individuals' personal information is only available if individuals themselves subscribe to the newsletter and are willing to give this information. Further on this research we will discuss the security aspects that come with this treatment on individuals' personal information as well as tracking through the newsletters.

6.2. Discussion

Once the data is gathered, the community manager can move forward and work on the results in order to extract their value. The presented discussion is based on the Nano-Tera.ch case study and describes some of the conclusions reached at its offices. For each newsletter the CDK team, leader by Dr. Peter Bradley, discussed the data on the NTS and compared it to the data gathered from the previous newsletters. On every discussion the team was able to take the challenge of exploring the data in the context of Nano-Tera.ch. Thanks to CDK team strong view of all Nano-Tera.ch network the discussions were interesting and satisfying regarding Nano-Tera.ch community needs.

The first discussion is related with the reduced number of clicks (1732 clicks) and the total number of individuals that have clicked (237 individuals). Regarding the total number of individuals to which the newsletters were sent (1000 individuals) only 24% of them have interacted with the newsletters' content while 37% have opened them (374 individuals). This way 137 individuals have ignored its content while 626 individuals have not performed any action. The answer to these results can be related to the poor popularity that newsletters have and to the lack of importance that individuals may give to newsletters. Yet, some email services apply span filters that mark newsletters as span and in these cases emails do not even reach individuals' main mail boxes. In addiction, there are the options that individuals can choose to receive their emails which may block the newsletter's content, for instance, if an individual receives his emails in text format, the action of opening will not be detected and he will see the tracking codes and therefore understand that his being tracked. Although individuals are aware of being tracked, seeing the tracking codes every time they receive a newsletter does not stimulates them to interact with the content.

On the links' type results, individuals have revealed a strong text-oriented focus by interacting with an high number of text-based links (97%). Although the number of images was significantly reduced regarding the text-based links, the percentages are highly different – 97% (text-based) versus 3% (image-based). The results show that individuals at Nano-Tera.ch have a strong scientific focus and strongly care about newsletters' text content. Thus, individuals were encouraged to click with the content not by the images but through the text description on the news. On these results, we can understand how important the provided text is regarding the process of gathering interactions. Once individuals are moved by the text, Nano-Tera.ch highly invested on its newsletters' design by adding catching phases to tease individuals to click and know more about the subjects. This strategy allowed Nano-Tera.ch to get individuals at its community interested in the community's work while getting their interactions with newsletters.

Regarding the results on categories, "Health" proved to be the one with the highest level of interest with 38% of the total number of clicks, followed by "Environment" (27%), "Security" (22%), and "Nano-Tera.ch" (8%). The results (see Chapter 6.1.3) meet the expectations since Nano-Tera.ch community has a larger number of heath projects and therefore a greater number of individuals interested in the subject. On the other hand, the balance between both categories of "Environment" and "Security" can be explained due to the approximate number of projects on the subjects, meaning that the amount of individuals on environment and security projects is similar. However, what is interesting to analyze is the total number of individuals that have shown interest in the categories different from those in which they work. This analysis can be accomplish through the grouping analysis, presented later.

On "Nano-Tera.ch" category it is exciting to note that individuals showed interest on their own community. Although some of the individuals (to which the newsletters were sent) were contacts outside Nano-Tera.ch when we look at the results in detail we see that the majority of clicks on "Nano-Tera.ch" category came from people inside the community. The category with the value "None" results from links with no assigned category such as the link to unsubscribe the newsletters or to access Nano-Tera.ch contact form.

By applying the Girvan-Newman algorithm through Vizster (Figure 35), we can see the clear division of Nano-Tera.ch individuals into 4 main clusters where edges represent implicit connections and nodes the individuals who belong to the community. Clusters' size show that Cluster #2 is the biggest, followed by the Clusters #3, #4, and #1, with their main categories on "Health", "Security", "Environment", and "Nano-Tera.ch" respectively. The results emphasize once more the higher number of individuals interested in "Health" category, meeting the expectations due to the higher number of projects on health. However, what is interesting to note is that even though individuals had the chance to interact with all categories, they tend to interact with the categories to which they are involved in.

The Girvan-Newman algorithm shows that Nano-Tera.ch community is strongly divided into its different topics of research with individuals within each of these topics having a strong connection between them. However, it also brings to light that although individuals inside their topics of research are strongly connected among them and highly interested on their topic of research, they also have showed a low interest in the remaining community's topics. On this analysis, Nano-Tera.ch should promote and encourage people from different projects to talk with each other and get more involved. Having individuals strictly focus on their research fields may be missing contributions that people from outside could provide to those working in the research field.

The overlaps highlight individuals that connect different categories and are valuable nodes, meaning that they have interacted with different categories and turn out to be highly connected to other individuals too. This data is crucial for Nano-Tera.ch to understand which individuals are more influential and have a widespread interest. It is exciting to note that these individuals are some of the most important people at Nano-Tera.ch that also work in several projects. In the middle, Nano-Tera.ch has the most valuable individual, which is implicit connected to all clusters and to a high number of individuals.

The final analysis on the "Connection Degree" enables communities to place implicit connections at the same level and compare them. The CD model assigns to each edge a CD value that describes how strong two individuals relate to each other. In the Nano-Tera.ch results only relations with a CD higher than 7 were include (Figure 34). The colors represent the Girvan-Newman algorithm and show that individuals can be grouped into 7 different clusters – Cluster A ("Nano-Tera.ch"), B ("Health" and "Security"), C ("Security"), D ("Health), E ("Environment"), and F and G (almost all categories).

Although individuals are all strongly connected on their clusters, the CD values show that people inside Cluster C are more highly connected and have the highest CD value of 20. If we look deeper into the results this edge results from the fact that both individuals have clicked more than 6 links in common, which reflects a very similar level of interest. As for the other connections (in C), these also have a strong CD, highlighting the fact that people who work in the "Security" topic are strongly connected with each other, followed by individuals inside "Health" category (Cluster D) and "Nano-Tera.ch" (Cluster A).

The results also show which individuals have interacted more with the newsletters. By assigning to each node a size that reflects the total number of clicks in the newsletters we can see that the individual with the highest level of clicks is at Cluster D ("Health") with 20 clicks, followed by the individual at Cluster A (12 clicks) and Cluster C (8 clicks). This higher number of clicks result form individuals high level of interest and go along with the highest CD values, meaning that inside Nano-Tera.ch these individuals are valuable nodes and important connections across the community.

Although individuals' names do not appear in the CD results we can note that individuals with the higher number of clicks and connected with the highest CD values are presented in the cluster detection (Figure 35) has the ones in the overlapped areas. These results prove once again that individuals in the overlaps correspond to those with high levels of connections and therefore more important regarding the goal of promoting individuals' collaboration.

In its application, Nano-Tera.ch used the results mainly to promote collaboration between individuals as well as to understand how individuals related to each other. It started by sharing with each individual the people that have showed interest on their work through their discovered network at NTS. With this information-sharing, Nano-Tera.ch allowed individuals to bring to light their own implicit network and to uncover missing connections between contacts that could result in future collaborations.

Nano-Tera.ch also identified its main nodes by looking at the overlapped areas on the clusters detection (Figure 35). It understood the value of these individuals as connection contacts between individuals from different areas. By talking with them Nano-Tera.ch tried to understand how it could bring people from different research fields to talk to each other and to help each other on their projects. In the same analysis, Nano-Tera.ch was able to discover their main followers and understand what were the main aspects that made them follow the newsletters strongly. In order to gather individuals' feedback and improve newsletters content, Nano-Tera.ch took an active position next to their followers and tried to bring to the CKD project some insights and fresh inputs. This work helped to design better newsletter and raise the total number of clicks over time.

In addition, the results were also used in some practical scenarios, such as to promote conferences and workshops on the topics with high level of interest. Regarding the presented results, Nano-Tera.ch have promoted an workshop on “Health” category in which the some of projects were presented followed by discussions. The workshop was communicated to those individuals that have showed an high level of interest in the topic. Based on the analysis offered by the NTS, Nano-Tera.ch was also able to predict the number of interested people in conferences and workshops and therefore the number of expected attenders. Along with the information on individuals’ preferences, there is individuals’ personal information, such as the local of work who helped to understand which places were the better to organize events. With Nano-Tera.ch community spread around the world the geographical location can be an highly influence aspect on attendance and community costs.

These are some of the actions taken by Nano-Tera.ch regarding the gathered results from the NTS and CD. In fact, the management office highlighted the challenge of working the data in order to help Nano-Tera.ch to work better and achieve better results. The CKD team have also showed their exciting on what it could be achieved through the results offered by the proposed solutions. At the end, both solutions were shared across Nano-Tera.ch community and future aspects were raised in order to improve the work done with the NTS and CD.

7. Conclusions

Communities are made of individuals who can easily communicate with each other, produce, and share information. While traditional connections among individuals were limited to their physical constraints, technology allowed to overcome these limitations and bring people close together. However, as the ability to people to communicate increases, the number of connections rises and with it the value of exploring these relationships becomes visible. Relationships that in the past were translated into a small set of connections (where nodes represent individuals and edges the known relationships) can now be translated into an extensive network of both explicit and implicit connections.

However, as the use of technology becomes part of people's life and the total number of connection increases, the goal of exploring individuals' interactions becomes more difficult to achieve. The poor technology on information-gathering through social objects (e.g. newsletters) creates a barrier for communities to discover implicit connections among individuals. Therefore the extensive goal of exploring individuals' interactions on social objects (digital content) and, based on that, discover implicit connections among them, becomes a challenge.

To answer this need, the presented research proposes a tracking system based on the digital content of newsletters – "Newsletter Tracking System". Through interactions with newsletters, the NTS solution focuses on trying to gather individuals' connections and grow a network of implicit relationships among them. Moreover, a connection can be seen as explicit or implicit if they come from different actions in the newsletter, such as clicking a link or selecting categories' preferences. Each time an individual clicks in a news it creates an implicit connection between himself and the news' category. At the end, individuals are all implicit connected to each other through these categories. On the other hand, explicit connections come from individuals explicit behavior regarding categories' preferences.

The NTS allows communities who want to know more about their individuals to do it through newsletters, but requires them to design their own newsletters and gather individuals electronic mail contacts in order to spread the newsletters. Once these requirements are meet, the NTS enables communities to upload the newsletters, send them, and then analyze the gathered data. The main process of gathering individuals' information boils down to individuals' desire to interact with newsletter content and thus perform clicks on the news' links. For each click, the NTS stores information on it such as individual's ID, newsletter's ID, link's ID, data and hour of the click, and the geographical location.

In the end, individuals are all organized into a network of connections where communities can apply their analysis and understand how individuals relate with each other and which connections are the most important. Based on explicit and implicit connections, NTS is able to bring to light hidden relationships and explore new connections that may rise new collaborations across the community.

On the other side, in a scenario where individuals are all highly implicit connected to each other it creates the challenge of filtering individuals' connections in order to understand which relationships are the most important in that universe. To answer this need, the presented research proposes a "Connection Degree" model which assigns to each implicit connection a value representing the importance of that connection in the network. The CD is calculated based on individuals' interactions with newsletters' news (which are then translated into implicit interactions with links and their categories) and their explicit behavior regarding the preferences on the categories.

For each implicit connection between every two individuals the CD takes into account the total number of clicks in common and the categories where they relate to each other. The final value also contains individuals' preferences on categories (positive, negative, or unknown) and communities' categories importance value, which represents the importance that individuals' explicit preferences on categories has in the CD calculation. All these aspects allow communities to have on each edge a value representing individuals' connection strength. The higher the value, the stronger the relationship. At the end, communities are able to apply filters to the CD and bring to light the most important connections in the network.

In short, communities are able to foster implicit relationships and explore individuals' connections based on the analysis offered by the NTS. By designing newsletters, communities are able to better understand how individuals relate to each other and to improve the way they explore its implicit connections. Along with the CD they can bring to light the most important connections and the individuals who have revealed high degrees on their relationships.

7.1. Main contributions

We believe that the NTS and the CD model can help communities to have a better overview on their individuals' relationships and thereby improve their collaborative work. The NTS is able to bring to light hidden relationships by building an implicit network based on individuals' interactions with newsletters. As for the CD, this is capable of assigning to each connection a value representing its importance in the network.

NTS value begins with the use of a well known social object (i.e. newsletter) as its main source to discover relationships. By allowing communities to keep an easy way of spreading information, the NTS uses the newsletters to create a tracking version in order to gather interactions and store information on individuals' connections. Along with this, communities are able to make their sending process fully automated by providing to the NTS the newsletter and the list of electronic mails of individuals. The followed process of capture the clicks is fully controlled by the NTS avoiding human error.

In order to make the analysis simple and allow communities to take value form the gathered data, the NTS exposes it through several visuals and organizes it by different aspects such as links' type and categories. NTS also takes advantage from existing tools (e.g. Vizster) thereby extending its feature and allowing a deeper data exploitation. In the extreme, communities can access all the data organized in a CSV file with no restrictions. Unlike existing tracking tools, the NTS is based on an open-source strategy allowing communities to evolve the solution the way it better fits their needs.

As for the CD, its main contribution lies on its ability to filter individuals' implicit connections and highlight the most important ones. By having individuals implicit connected to each other, communities are capable of understanding their relationships but unable to perceive how strong a relationship is regarding the discovered set. Applying the CD enables them to bring to each connection a strength value. The higher the value, the stronger is the relationship.

While keeping the use of newsletters, the CD is calculated based on individuals' clicks on links and explicit preferences in categories. The proposed model does not asks communities and individuals to further information or actions in order to calculate CD values. If no preference is shown by individuals, the model keeps working with unknown preferences and bases its CD values on individuals' clicks on links. Also in the CD, it is used existing tools (e.g. NodeXL) to expose the data and allow analysis on it. Moreover, the use of CSV files do not require communities to install additional software to access and work the data.

7.2. Future work

While the work on the thesis was being developed it was impressive the number of new ideas and developments that were emerging overtime. Gladly we manage to develop some of these suggestions but the time was too short to get everything done and some of the ideas were too ambitious regarding the timeline. The follow suggestions on the future work are some of the proposals that appear in collaboration with Nano-Tera.ch community and that would be interesting to have it implemented.

- **Business Intelligence**: The ability to transform raw data into meaningful and useful information for business purposes is incredible important for business to take value from the data. Include into the NTS some business intelligence methodologies and technology (e.g. analytics and data mining) would be interesting for communities to better identify and develop new opportunities. The properly identification of new opportunities and the effective implementation of strategies to create value can provide a competitive market advantage and a long-term stability. Even though from a community perspective these competitive advantage may be different from a business perspective, these questions have a greater impact when we rise the option of using the NTS in a business context.
- **Business perspective**: Although this research is hardly focus on a community-based analysis, it may be interesting to develop and analyze the NTS in a business perspective. In this case, a business will be interesting in use it to gather data on their customers' preferences and therefore identify new opportunities. In fact, this type of analysis is not new but the creation of value in this area is very difficult, however the development of the NTS as an open-source tool brings value on its initial investment and its ability to be developed according to the requirements of the business.
- **Newsletters' personalization**: Once the process of sending the newsletters is done for each individual and regarding the NTS ability to gather data on individuals' preferences, it would be interesting to have the capability to personalize each individuals' newsletter according to their preferences. In a future development the NTS should be able to pick up the news from the uploaded newsletter and create a personalized news-

letter to each individual according to the gather data on his preferences and clicks. A personalized newsletters is able to better get individuals interest and interactions.

- **Analysis over time**: The offered analysis on the NTS is focus on each newsletter. The community has to select the newsletter on which it wants to analyze the data and then the NTS exposes the data into several ways (see Chapter 4.6). An aspect to improve will be to offer the ability to analyze the data over time on all the newsletters. In addition to the deeper analysis on each newsletter, it would be exciting to have the chance to analyze the results over time by placing all newsletters' results in one big analysis. This will allow to easily answer questions such as the evolution of the number of clicks over time. However, we note that this analysis can be done manually by checking the data on each newsletter and place it in one common file (e.g. excel file).
- **Geographical analysis**: With the ability to work from anywhere and be connected at time brings a worldwide capacity. This aspect raises the interesting question of analyze individuals' interactions according to their geographical location, mainly in a context where the community is spread around the world. The NTS should include in its analysis a geographical analysis with the total number of clicks distributed by countries. We note that this feature was not implemented in the Nano-Tera.ch context due to individuals' privacy, which do not agree on the tracking of their geographical location when clicking on the newsletters.
- **Interaction with outside contacts**: Although Nano-Tera.ch was mainly focus on gathering data on individuals inside its own community projects, a further step will be to have the newsletters sent to contacts outside the community (e.g. individuals in other communities working in the same topics of research). This step will allow to understand how possible collaboration with outside contacts could be born from the implicit networks. Although the option was put on hold due to the desire of protecting Nano-Tera.ch individuals' private information and work, it would be interesting to also gather data on individuals outside the community.
- **Focus on other social objects**: Although the presented results are focus on the social object "newsletter", they are not strictly apply to newsletters. The analysis could be performed in other social objects (e.g. videos, learning objects, and images) as long as they would have the ability to individuals to interact with links and the NTS is able to replace the original links for malicious links. Meanwhile, the NTS open-source development allows communities to evolve it in the way that better meets its needs.
- **Flexibility on the CD**: The CD enables to assign to each implicit edge a value representing its strength in the discovered universe. However, this value its strictly calculated based on newsletters, namely individuals' clicks and preferences on categories. In a further development it will be interesting to apply the CD to any social object. A hypothesis will be to have the CD focus only on individuals' interactions and reach individuals' preferences based on those interactions.

- **Privacy**: An important aspect raised with this research is related with individuals' privacy. Since individuals are being tracked their privacy is placed in question. However, during this research all individuals were aware of their tracked and which where the purposes of this research. Individuals had also the chance to unsubscribed the newsletter and ask to their information to be deleted. In this case, the data on individuals' interaction is fully deleted. In future developments, the question regarding the privacy should always be taking into account in order to ensure individuals' privacy and awareness.

It is exciting to notice that there is a lot of room for improvement while the actual NTS and CD is capable of allowing communities to reach interesting results and improve their knowledge on individuals. In the future, we hope that both solutions could go further and better help communities improving their work and maybe go into in a business perspective.

8. References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734–749.

Aggarwal, C. C. (Ed.). (2011). *Social Network Data Analytics*. Springer.

Anderson, B. (2006). *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (New Editio.). Verso.

Barão, A., & Rodrigues da Silva, A. (2011). A Model to Evaluate the Relational Capital of Organizations (SNARE RCO). *Proceedings of Conference of Enterprise Information Systems (Centeris'2011)*, Springer

Barão, A., & Rodrigues da Silva, A. (2012). What is the Value of Your Network ? *Proceedings International Workshop on Social Networks and Data Processing (SNDP) in conjunction with The 2012 IEEE/WIC/ACM International Conference on Web Intelligence/Intelligent Agent Technology*, IEEE Computer Society.

Barão, A., & Rodrigues da Silva, A. (2012). How to value and monitor the relational capital of knowledge-intensive organizations? *Handbook of Research on Enterprise 2.0: Technological, Social, and Organizational Dimensions*, IGI Global.

Barnard, S. T., & Simon, H. D. (1994). A Fast MultiLevel Implementation of Recursive Spectral Bisection for Partitioning Unstructured Problems. *Concurrency Pratice and Experience*, *6*(2), 101–118.

Bauman, Z. (2001). *Community: Seeking Safety in an Insecure World* (1st ed.). Polity.

Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., & Grinter, R. E. (2005). Quality Versus Quantity: E-Mail-Centric Task Management and Its Relation With Overload. *Human-Computer Interaction*, *20*(1), 89–138.

Bender, T. (1986). *Community and Social Change in America*. Rutgers U.P.

boyd, danah. (2006). Friends, Friendsters, and Top 8: Writing community into being on social network sites. *Peer-Reviewed Journal of Internet*, *11*.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition the Culture of Learning, *18*(1), 32–42.

Calhoun, C. J. (1980). Community: Toward a variable conceptualization for comparative research. *Social History*, *5*(1), 105–129.

Carey, J. W. (1992). *Communication as Culture, Revised Edition: Essays on Media and Society*. Routledge.

Castro, J., Kolp, M., & Mylopoulos, J. (2002). Towards requirements-driven information systems engineering: the Tropos project. *Information Systems*, *27*(6), 365–389.

Chakra, A., O'doherty, S. P., Rice, J., & Yap, B. K. (2009). Embedding a unique serial number into the content of an email for tracking information dispersion. *US Patent Application Publication*.

Chua, C. E. H. (2009). Why Do Virtual Communities Regulate Speech? *Communication Monographs*, *76*(2), 234–261.

Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. *Journal of Direct, Data and Digital Marketing Practice*, *9*(3), 231–244.

Conville, R. L., & Rogers, L. E. (1998). *The Meaning of Relationship in Interpersonal Communication*. *The Meaning of Relationship in Interpersonal Communication*. Praeger.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *The Behavioral and brain sciences*, *22*(5), 735–808.

Diestel, R. (2006). *Graph Theory*. (3rd ed. Edition, Ed.). Springer.

Dinis, P., & Rodrigues da Silva, A. (2009). Application Scenarios for the Learning Objects Pool. *Journal of Universal Computer Science*, *15*(7), 1455–1471.

Domingos, P., & Richardson, M. (2001). Mining the network value of customers. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01* (pp. 57–66). New York, New York, USA: ACM Press.

Dörk, M., Carpendale, S., & Williamson, C. (2011). EdgeMaps: visualizing explicit and implicit relations, 78680G–78680G–12.

Ducheneaut, N. B. (2002). Information, Communication & Society The Social Impacts of Electronic Mail in Organizations : A case study of electronic power games using communication genres. *Information, Communication & Society*, *5*(2), 153–188.

Ducheneaut, N., & Bellotti, V. (2003). Ceci n'est pas un Objet? Talking About Objects in E-mail. *Human-Computer Interaction*, *18*(1), 85–110.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, *4*(12).

Ewing, T. (2008). Participation cycles and emergent cultures in an online community. *International Journal of Market Research*, *50*(5), 575–590.

Facebook. (2012). Facebook.com.

Farnham, S., Chesley, H. R., McGhee, D. E., Kawal, R., & Landau, J. (2000). Structured online interactions: improving the decision-making of small discussion groups. *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00*, 299–308.

Fayad, A.-L., & Weeks, J. (2006). Photocopiers and Water-coolers -The Affordances of Informal Interaction.

Fletcher, P., Hoyle, H., & Patty, C. W. (1993). *Foundations of Discrete Mathematics*. Brooks/Cole.

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, *25*(1), 153–160.

Foulger, M. G., Chipperfield, T. R., Cooper, J. S., & Storms, A. C. (2010). System and method related to generating and tracking an email campaign. *US Patent.*

Freeman, L. C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press.

Gibson, B. (2007). Enabling an accessible web 2.0. *W4A2007 - Keynote with the 16th International World Wide Web Conference*, 1–6.

Gibson, J. J. (1986). *The Ecological Approach To Visual Perception* (New Ed.). Psychology Press.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–6.

Gotracknow. (2013). www.gotracknow.com.

Graves, L. (2007). The Affordances of Blogging: A Case Study in Culture and Technological Effects. *Journal of Communication Inquiry*, *31*(4), 331–346.

Grossman, L. (2006). You - Yes, You - Are TIME's Person of the Year. *Time Magazine*.

Harley, J., & Blismas, N. (2010). An Anatomy of Collaboration Within the Online Environment. (M. Anandarajan, Ed.)*Springer-Verlag Berlin Heidelberg*, 15–34.

Heer, J., & Boyd, D. (2005). Vizster: Visualizing Online Social Networks. *Visualization, 2005. INFOVIS 2005. IEEE*, *3*, 2–9.

Hoegg, R., Martignoni, R., Meckel, Mi., & Stanoevska-Slabeva, K. (2006). Overview of business models for Web 2.0 communities. *Proceedings of GeNeMe 2006*, 1–17.

Hutchby, I. (2001). Technologies, Texts and Affordances. *Sociology*, *35*(2), 441–456.

Isaías, P., Miranda, P., & Pífano, S. (2009). Critical Success Factors for Web 2.0–A Reference Framework. *Online Communities and Social Computing*, *LNCS 5621*(Springer-Verlag Berlin Heidelberg 2009), 354–363.

Jackson, M. O., & Watts, A. (2002). The Evolution of Social and Economic Networks. *Journal of Economic Theory*, *106*(2), 265–295.

Jensen, J. (1990). *Redeeming Modernity: Contradictions in Media Criticism*. SAGE Publications.

Jones, S. G. (1998). Information, Internet, and Community: Notes Toward an Understanding of Community in the Information Age. *Cybersociety 2.0: Revisiting Computer-Mediated Community and Technology*. SAGE Publications, Inc.

Kanter, R. M. (1972). *Commitment and Community: Communes and Utopias in Sociological Perspective*. Harvard University Press.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03* (p. 137). New York, New York, USA: ACM Press.

Kinton, J. F. (1975). *The American community: Creation and revival : a multidisciplinary perspective*. Social Science and Sociological Resources (Aurora, Ill.).

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., et al. (1997). The Past and Future History of the Internet, *40*(2), 102–108.

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web*, *1*(1), 5–es.

Li, X., Zeng, D., Mao, W., & Wang, F. (2008). Online communities: a social computing perspective. *Intelligence and Security Informatics*, 355–365.

Lockard, J. (1997). Progressive Politics, Electronic Individualism and the Myth of Virtual Community. In D. Porter (Ed.), *Internet culture* (pp. 219–232). New York : Routledge.

Lopes Ferreira, T., Rodrigues da Silva, A., Bradley, P., Nitesh, K., & Agrawal, M. (2011). Discover Implicit Relationships Between Researchers UsingEmail Tracking. *CAPSI'2011*.

Lopes Ferreira, T., & Rodrigues da Silva, A. (2012). Improving the Quality of the Community Relations Knowledge Using Implicit Connections. *2012 Eighth International Conference on the Quality of Information and Communications Technology*, 327–332.

Lopes Ferreira, T., & Rodrigues da Silva, A. (2012). Foster an Implicit Community Based on a Newsletter Tracking System. In R. Meersman, H. Panetto, T. Dillon, S. Rinderle-Ma, P. Dadam, X. Zhou, S. Pearson, et al. (Eds.), *On the Move to Meaningful Internet Systems: CoopIS 2012* (Vol. 7565, pp. 398–415). Berlin, Heidelberg: Springer Berlin Heidelberg.

Massmailsoftware. (2013). www.massmailsoftware.com/tracker/.

McArthur, R., & Bruza, P. (2003a). Discovery of implicit and explicit connections between people using email utterance. *Proceedings of ECSCW*, (September), 21–40.

McArthur, R., & Bruza, P. (2003b). Dimensional representations of knowledge in online community. *Chance Discovery*.

Melville, P., & Sindhwani, V. (2010). Recommender Systems. *Encyclopedia of Machine Learning*.

Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., & Stumme, G. (2011). Community Assessment Using Evidence Networks. *M. Atzmueller et al. (Eds.): MUSE/MSM 2010*, *6904*, 79–98.

Mooney, R. J., & Roy, L. (2000). Content-based book recommending using learning for text categorization. *Proceedings of the fifth ACM conference on Digital libraries - DL '00* (pp. 195–204). New York, New York, USA: ACM Press.

Musser, J., & O'Reilly, T. (2007). Web 2.0: Principles and best practices. *O'Reilly Media, Inc.*

MySpace. (2012). new.myspace.com.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577–82.

Newman, M., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2).

NodeXL. (2012). http://nodexl.codeplex.com/.

O'Reilly, T. (2007). What is Web 2.0: Design Patterns and Business Models for the next Generation of Software. *Communications & strategies*, (65), 17–37.

Ogata, H., & Yano, Y. (1998). Collecting Organizational Memory Based on Social Networks in Collaborative Learning. *WebNet*, 822–827.

Papacharissi, Z. (Ed.). (2010). *A Networked Self: Identity, Community, and Culture on Social Network Sites*. Routledge.

Parks, M. R. (2010). Social Network Sites as Virtual Communities. *A Networked Self: Identity, Community and Culture on Social Network Sites* (pp. 105–120). Routledge.

Parthasarathy, S., Ruan, Y., & Satuluri, V. (2011). Community Discovery in Social Networks: Applications, Methods and Emerging Trends. In C. C. Aggarwal (Ed.), *Analytics, Social Network Data* (pp. 79–113). Boston, MA: Springer US.

Pazzani, M. J., & Billsus, D. (2007). Content-Based Recommendation Systems. *The Adaptive Web*, 325–341.

Pinney, J. W., & Westhead, D. R. (2005). Betweenness-based decomposition methods for social and biological networks . *Interdisciplinary Staticsa and Bioinformatics Leeds*, 87–90.

Preece, J. (2001). Sociability and usability in online communities: determining and measuring success. *Behaviour & Information Technology*, *20*(5), 347–356.

Rangwala, H., & Jamali, S. (2010). Co-Participation Networks Using Comment Information. *Association for the Advancement of Artificial Intelligence*, 315–318.

Reddy, P. K., Kitsuregawa, M., Sreekanth, P., & Rao, S. S. (2002). A Graph Based Approach to Extract a Neighborhood Customer Community for. In S. Bhalla (Ed.), *Databases in Networked Information Systems* (pp. 188–200). Berlin, Heidelberg: Springer Berlin Heidelberg.

Researchgate. (2012). Researchgate.net.

Rheingold, H. (1993). *The Virtual Community: Homesteading on the Electronic Frontier*. The MIT Press.

Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier* (Revised.). The MIT Press.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to Recommender Systems Handbook. *Recommender Systems Handbook* (1st ed.). Springer.

Rodrigues da Silva, A. (2011). Padrões de Interfaces Sociais no Desenho do BOA Social Interface Patterns for the Design of BOA System. *INFORMÁTICA NA EDUCAÇÃO: teoria & prática*, 167–181.

Sarwar, B., Karypis, G., Konstan, J., & Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the tenth international conference on World Wide Web - WWW '01* (pp. 285–295). New York, New York, USA: ACM Press.

Schacter, D. L. (1992). Implicit Knowledge: New Perspectives on Unconscious Processes. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(11113-11117).

Schwartz, M. F., & Wood, D. C. M. (1992). Discovering Shared Interests Among People Using Graph Analysis of Global Electronic Mail Traffic. *Communications of the Association for Computing Machinery*.

Sohn, D. (2008). *Social Network structures and the Internet: Collective Dynamics in Virtual Communities*. Cambria Press.

Sun, Y., Han, J., Zhao, P., Yin, Z., Cheng, H., & Wu, T. (2009). RankClus: integrating clustering with ranking for heterogeneous information network analysis. *Proceedings of EDBT'09.*

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* (p. 336). Doubleday; Anchor.

Ta, D. R. W., & Borgatti, S. P. (1994). Betweenness centrality measures for directed graphs, *16*, 335–346.

Thompson, J. (2008). Don't be afraid to explore Web 2.0. *Phi Delta Kappan*, *89*(10), 711, 778.

Trudeau, R. J. (1994). *Introduction to Graph Theory* (2nd editio.). Dover Publications.

Tuire, P., & Erno, L. (2001). Exploring invisible scientific communities : Studying networking relations within an educational research community. *Higher Education*, *42*(4), 493–513.

Vrasidas, C. (1995). Constructivism versus objectivism: implications for interaction, course design, and evaluation in distance education. *International Journal of Educational Telecommunications*, 1–17.

Vrasidas, C., & Glass, G. V. (2002). A Conceptual Framework for Studying Distance Education. *Distance Education and Distributed Learning* (pp. 31–55). Information Age Publishing.

Vrasidas, C., & Zembylas, M. (2004). Online professional development: lessons from the field. *Education + Training*, *46*(6/7), 326–334.

Weinreich, F. (1997). Establishing a point of view toward virtual communities. *Computer-Mediated Communication*, *3*(2).

Wellman, B., & Gulia, M. (1999). Net surfers don't ride alone: Virtual communities as communities. *Networks in the global village*.

Whittaker, S., Isaacs, E., & O'Day, V. (1997). Widening the net: workshop report on the theory and practice of physical and network communities. *ACM SIGCHI Bulletin*, *29*(3), 27–30.

Willson, M. A. (2006). *Technically Together: Rethinking Community within Techno-society*. Peter Lang Publishing Inc.

Zajicek, M. (2007). Web 2.0: Hype or Happiness ? *16th International World Wide Web Conference*.

Zaphiris, P., & Ang, C. S. (2009). *Social Computing and Virtual Communities* (1st ed., p. 303). Chapman and Hall/CRC.