

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

Tiago Henrique Lopes Ferreira

Instituto Superior Técnico de Lisboa
Lisbon, Portugal
tiagohenrique@ist.utl.pt

Abstract. Communities have explored the virtual world as a tool to improve their communication. However, when the number of interactions was manageable in its face-to-face manner, the same was not true when the Internet became the main communicator. The number of interactions grows at a pace that is very hard for communities to control. As a consequence connections get lost or forgotten and communities lose the chance to perceive individuals' relations. It is in this gap that the "Newsletter Tracking System" (NTS) comes as an automatic tool that allows communities to capture connections between individuals through their interactions with newsletters. By storing the data on individuals' interactions, NTS discovers implicit connections between individuals and fosters an implicit community. In addition, it uses clustering algorithms to allow communities to better understand how individuals relate to each other and it proposes a "Connection Degree" model (CD) to measure the connections' strength among individuals. NTS and CD were developed and evaluated within Nano-Tera.ch scientific community. At the end, the results showed that implicit communities can be an advantage for real communities to better organize individuals, share knowledge, and promote teamwork.

Keywords: Community, Newsletter Tracking System, Connection Degree, Implicit Connection

1 Introduction

New technologies have changed the way people interact by providing new approaches to communicate, share, and stay connected to each other. The Internet has revolutionized the computer and communication worlds like nothing before [13], 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 [15]. Likewise, communities started to change from groups to networks and to take advantage of the Internet as communication tool [23]. In 1993, the concept of "Virtual Communities" [25] came to live and also the research on the relation between communities and the Internet. Virtually, communities have no restriction on the number of connections and the way they can explore them. When tracking all individuals' connections was extremely difficult in the physical world it becomes an easier target in the virtual sphere. Technology brought people together as well as their own interests, curiosities, hobbies, professions, and so on. Each time people go online they become exposed and their interactions can be stored as connections with something or someone.

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

Both explicit and implicit connections are important to understand individuals and communities' network. Explicit connections ensure the knowledge-base on individuals' relations and implicit relations improve that knowledge. An explicit behavior "is controllable, intended, made with awareness, and requires cognitive resources" [6]. Individuals have clear sense of explicit activities and this makes explicit relations extremely important when defining individuals and communities' networks. However, the study on explicit connections can be very limited if individuals do not share behavior and interests. To fill the gap, implicit connections can be used and explored as long as individuals continue to interact with content and people. Implicit connections are

based on individuals' unconscious interactions and can be tracked if they occur virtually. In the presented study an interaction between an individual and the newsletter is stored as an explicit connection between the individual and the subject. Then the explicit connection "individual-subject" is translated into an implicit connection "individual-individual". Implicit connections are important to help a network of connections to be evolved beyond individuals' explicit behavior. The more interactions an individual does, the stronger is the implicit knowledge in the network [21].

The problem arises when capturing and analyzing individuals' implicit connections at the rate they grow. Data management becomes difficult to handle and implicit patterns harder to find. As consequence, "the connections between individuals, groups, and information becomes lost, or forgotten, and individuals and groups become more isolated" [11]. In the presented case-study of Nano-Tera.ch [16], a scientific community at Switzerland, the governing bodies were aware of the complexity on capturing researchers' connections as the community grew. The goal of exploring implicit connections was difficult to achieve due to the increase number of interactions and the high complexity on capturing researchers' interactions. Nano-Tera.ch was interested in understanding how its community was implicitly organized but had no way to capture and promote interactions between researchers. In fact, Nano-Tera.ch needed an automatic tool able to track the connections among researchers and organize them into results. In addition, there was the goal of finding a way to classify implicit connections in order to measure the connection's strength and thus understand the connections' influence in the discovered universe.

The presented research developed a tool with Nano-Tera.ch to enable communities to capture implicit connections between individuals and also to design a way to measure the connection degree among individuals. The "Newsletter Tracking System" (NTS) is that developed web tool to discover implicit connections between individuals based on their interactions with newsletters. Each interaction is stored and translated into a relation between the individual and the content. At the end, individuals that have interacted will be related through newsletters' content and therefore implicitly related to each other. The "Connection Degree" (CD) model is based on explicit and implicit behavior and it proposes a way to measure the connections' strength among individuals. Individuals are organized into a network of connections and each connection is related to a connection degree value expressing its strength in the network. The higher the connection degree is, the higher the importance of the relationship is for the community.

This paper is structured in 6 sections. The presented introduction as the section 1 and the section 2 as the description of NTS as a tool for communities to explore implicit connections. Section 3 explains the CD model and ends with section 4 where it is described the results according to the practical case study of Nano-Tera.ch community. By the end, section 5 explores some of the related work and section 6 presents several research conclusions.

2 Newsletter Tracking System

Communities have taken their step into the virtual world and have included the web tools in their habits. Emails, forums and blogs became part of communities' ways to interact. However, when the number of connections between individuals increases, the task of controlling the implicitly growing is difficult to monitor.

Through interactions with newsletter the "Newsletter Tracking System" (NTS) proposes a way to capture individuals' interactions, discover implicit connections, and so foster an implicit community. NTS uses electronic mail technology to reach individuals, and newsletters to promote interactions and discover implicit connections among individuals. According to Bellotti Ducheneaut "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" [5]. 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. The use of newsletters comes as the way to capture individuals' interactions. It gives communities the freedom to design their content and define the newsletter according to individuals' interests. In addition, communities can use their newsletters to extract individuals' interactions while keeping them updated.

In the context of the NTS, a newsletter is defined as a set of news collected into an HTML file. Its content can be defined by several authors and thus result on a collaborative common information space. The periodicity (e.g. monthly) defines the pace at which information reaches individuals and it is defined by the community itself.

2.1 System Overview

The NTS is an online tool that allows communities to foster an implicit community based on individuals' interactions with newsletters. In fact, the NTS supports the relationship between the community and the individuals by providing the tracking tool for the community (**Fig. 1**). The community itself is managed by a "Community Manager" that is responsible for triggering the tasks at the NTS. Based on the goal of "Fostering 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 "dependor-dependee"¹, where the "Community Manager" (the "dependor") depends on the "Newsletter Tracking System" (the "dependee") to achieve the goal.

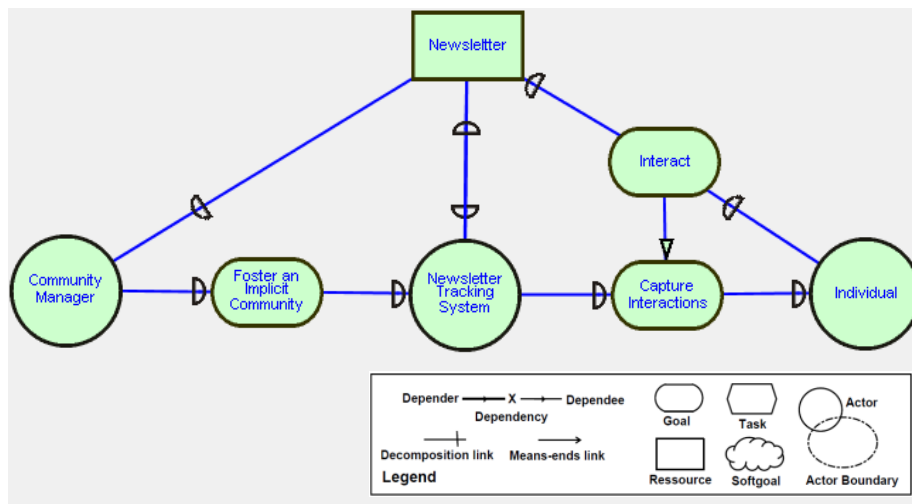


Fig. 1. System Overview

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 with the community manager. The NTS depends on the community manager to provide the newsletter so it can perform the tasks, satisfy the goals, and also provide the resources. Without all the input elements no further links can be followed and the model stops.

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 a 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 interact. 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", "Newsletter Tracking System", 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 "dependor" 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" (**Fig. 2**). 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.

¹ The used terminology as well as the presented schemas are based on i* framework defined by Yu Eric [24]. It "conceives of software-based information systems as being situated in environments in which social actors relate to each other in terms of goals to be achieved, tasks to be performed, and resources to be furnished" [9].

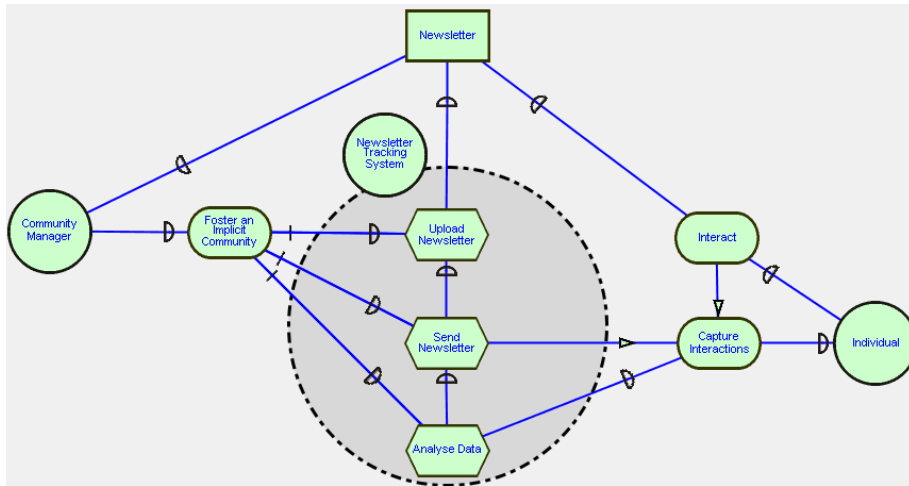


Fig. 2. Newsletter Tacking System Overview

The dependency links show the order in which 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 “Analyse Data” task. Thus, 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 “Analyse Data”. The task “Send Newsletter” has the end goal of “Capture Interactions” that is needed to perform the last task of “Analyse 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.

2.2 Upload Process

The process of capture and detect implicit connections among individuals starts with a community uploading a newsletter into the NTS. The upload process is described as the first interaction between a community and the NTS. At this stage, a community reveals its interest on capturing individuals’ interaction and on fosters an implicit community. The process can be represented by Fig. 3 where the “Community Manager” and the “Newsletter Tracking System” 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.

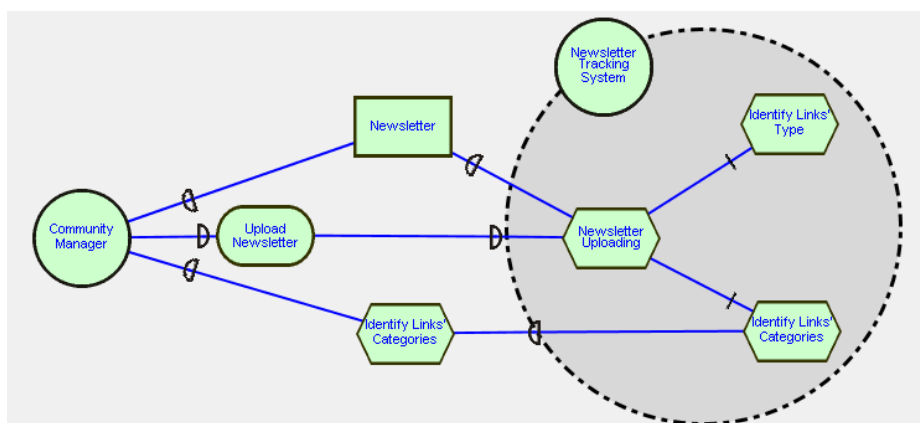


Fig. 3. Upload Process

The newsletter is described as the main resource for the system since it is the element that is shared with individuals. 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 individuals’ needs, the higher the number of interactions. This responsibility is given to communities once they know better what individuals’ desire and expect.

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 a list of categories provided by the community manager. 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 the individuals will be related with categories and implicit connected to each other based on these categories.

2.3 Send Process

To meet the goal of discovering implicit connections among individuals, the community needs to reach individuals’ emails and let the NTS capture the interactions with the newsletters. The process is described by **Fig. 4** 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”.

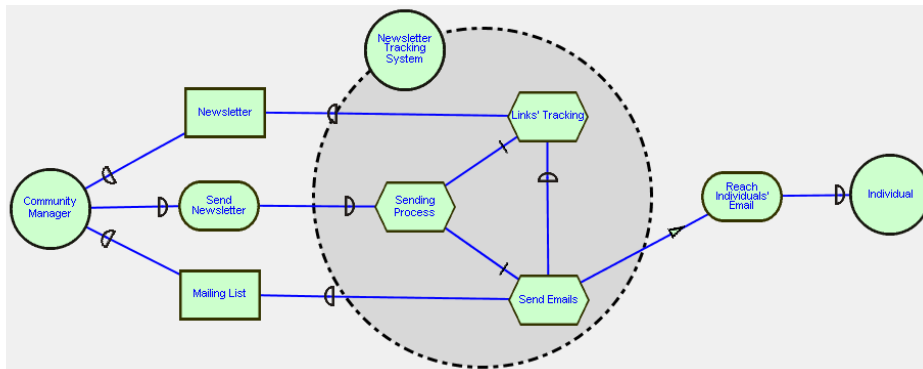


Fig. 4. Send Process

Once the dependency links are respected the NTS starts the task of tracking the newsletter’s link by replacing them by 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 working. The “code” defines one of the possible actions: opening the email, clicking a link, sharing the newsletter, or seeing it online. The “user-id” identifies the individual who triggered the action, the “link-id” which link was clicked, and the “newsletter-id” the newsletter in what the action was performed. The elements are all automatically generated by the NTS and put together in order to build a tracked newsletter for each individual. The newsletters are then the key elements to perform the task of “Send Emails”, where the community manager performs its last interaction with the NTS by providing the “Mailing List” resource, which contains the list of emails to which the newsletter is going to be sent.

The 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 the links.

2.4 Tracking Process

The core component in the NTS boils down to the tracking process, where all individuals’ interactions are capture and stored into the 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 **Fig. 5** 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”.

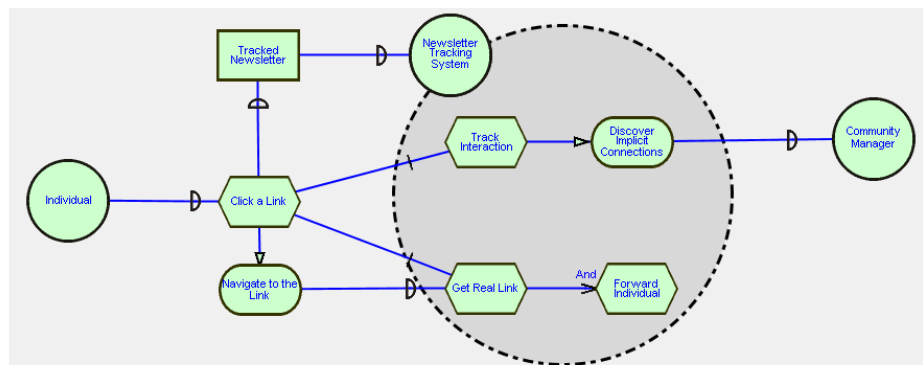


Fig. 5. Tracking Process

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. The 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 forward the individual. Once the individual navigates to the 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 foster 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.

2.5 Data Analysis

The captured data on individuals’ interactions is the most important asset and the one that allows communities to foster implicit communities. 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 and allowing communities to export it.

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, such as time passed – total number of days that have passed since the newsletter was sent –, the total number of individuals who interacted with the newsletter, and the total number of clicks. This information is useful to have an overview of the newsletter’s impact and to monitor the results. On a further analysis the NTS divides the presentation of the data into a set of sections:

1. **Links' Type.** Organizes the links by their type – “Text” or “Image” – and presents the percentage of clicks on both types. The analysis allows communities to understand the best approach to design newsletters. If individuals interact more with image-based or text-based links.
2. **Links' Categories.** Presents the categories of the newsletter followed by percentage of clicks gather on each. At the end, communities will be able to understand to which categories individuals showed more interest, and who was the category with higher impact on individuals. This categorization is also used to relate individuals with categories and thereby discover implicit connections among them.
3. **Individuals' Category Clustering.** On clustering, the NTS uses individuals' interactions to cluster them by categories. A relation “individual-link” is translated into a relation “individual-category” and the individual is added to the category cluster followed by his total number of clicks on the category. The clustering allows communities to discover all the “category-individual” relations and thus followers.
4. **Implicit Connections.** The NTS translates the relations “individual-category” into implicit connections “individual-individual” and assigns them a “Connection Degree” in order to communities have a way of measure the connections. The value is based on individuals' interactions with categories and on their explicit preferences on the categories. However, in order to visualize the implicit community-based the NTS uses external tools such as Vizster [12] and NodeXL [17].
5. **Data Export.** The process of exportation is what allows communities to export the data in order to use it on external tools. The NTS enables the data to be exported in the file formats of XLS and XML. The goal is to allow communities to exploit the data the way they want and do not limit its exploitation to what NTS offers.

At this stage communities are able to have an overview of individuals' implicit connections. Once the newsletters reach individuals it all comes to individuals' interactions. The NTS will automate the process of sending the newsletters to individuals and track each newsletter in order to capture interactions. The system is also responsible for capture every click and translate it from a relation “individual-link” to a relation “individual-category”. At the end, the relations will be used to foster an implicit community based on individuals' implicit connections.

3 Connection Degree

With communities having the ability to discover implicit connections among individuals, it becomes important to understand the value of each connection in the discovered universe. In a scenario where all individuals click in all the categories, they will be all implicit connected and the community will find it harder to take conclusions. On the other hand, with a way to measure the value of each connection in the network, the community will have the chance to create their own thresholds and filter the implicit connections.

The connection degree model proposes a way to measure the connection strength between every two individuals by calling it “Connection Degree” (CD). The higher the CD, the stronger the relationship between two individuals. The CD model uses both explicit and implicit connections to calculate the explicit and implicit degrees in the CD. While the implicit degree is based individuals' implicit connections, the explicit degree comes from individuals' preferences on the newsletters' categories.

An individual can express his categories interest by explicitly checking a category as preferred. This process is done through the NTS, where individuals can navigate to a “Preferences 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 when no explicit action is performed. The explicit degree will affect the final CD in a “Category Importance” (ci) value, defined by the community and that represents the importance of the individuals' preferences in the equation.

$$explicit\ degree\ (Individual):\ ed(I) = \begin{cases} ci, & \text{positive preference} \\ 0, & \text{no preference shown} \\ -ci, & \text{negative preference} \end{cases} \quad 0 \leq ci \leq 1 \quad (1)$$

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

$$n \times m = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix} \quad (2)$$

Regarding the relations “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 implicit value is then calculated based on the following equation.

$$(I_i, I_j) \times c_l = \min(a_{il}, a_{jl}) + (ed_i + ed_j) \quad (3)$$

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 (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 plus the sum of both explicit degrees. The first part of the equation represents the minimum value on the individuals’ categories-based relation and the second part expresses the individuals’ explicit interests 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 relations “individual-category” to the implicit relations “individual-individual”. Individuals are organized into a matrix $n \times m$ where rows (n) and columns (m) represent individuals (I) and the values the sum of all (3) equations for all l categories ($\sum_{i=0}^l (3)$).

$$n \times m = \begin{bmatrix} \sum_{i=0}^l (I_1, I_1) \times c_i & \cdots & \sum_{i=0}^l (I_1, I_m) \times c_i \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^l (I_n, I_1) \times c_i & \cdots & \sum_{i=0}^l (I_n, I_m) \times c_i \end{bmatrix} \quad (4)$$

The final value for the CD between every two individuals is calculated by performing a syntax sum of both resulted matrix from (2) and (4) but with the multiplication operator. At the end only one of the sides of the matrix is taken into account, i.e. a lower or upper triangular matrix, and excluded the diagonal. This will ignore duplicated pair and take only one CD into account.

$$\text{connection degree} : n \times m = \begin{bmatrix} (2)_{11} \times (4)_{11} & \cdots & (2)_{1m} \times (4)_{1m} \\ \vdots & \ddots & \vdots \\ (2)_{n1} \times (4)_{n1} & \cdots & (2)_{nm} \times (4)_{nm} \end{bmatrix} \quad (5)$$

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 and highlight the most important implicit connections.

The CD model brings to the NTS value on measuring implicit connections and allows communities to have a better overview of the implicit community. The NTS is responsible for capturing and discovering implicit connection and the CD model for assigning every connection a CD value.

4 Evaluation: The Nano-Tera.ch Case Study

The evaluation of this research was done inside the Nano-Tera.ch community. A scientific Swiss federal program with more than 40 projects on the subjects of “Security”, “Health”, and “Environment” [16]. Nano-Tera.ch diversity goes from different projects to the hundreds of researchers around the world. Thus, in order to capture knowledge on the Nano-Tera.ch community, the management structure at Nano-Tera.ch decided to run the research project “Community Knowledge Development” (CKD) [6]. Within other goals, the CKD was trying to understand how researchers were related and how connections could be explored in order to improve researchers’ work and promote collaboration.

The research project CKD supported the development and the evaluation of both NTS and CD model at Nano-Tera.ch community. This project took place during 6 months and the community management designed 6 newsletters to send to the community and to evaluate the system itself. However, the presented results are based on the top 3 of the newsletters regarding interest and reliability. The overall information on the newsletters was more than 3000 of emails sent (1000 per newsletter) and about 1750 of captured clicks. In addition, Nano-Tera.ch tried to maintain the same thousand of individuals (i.e. emails sent) for every newsletter in order to collect data from different sources but for the same individuals.

4.1 Links' Type

The results focus on trying to understand to what kind of information-exposure individuals interact more - "Text" or "Image". Each newsletter's topic was introduced with an "image" followed by the "text" to which individuals had the chance to interact and reach the same information. This analysis is important for communities to better design newsletters. Individuals at Nano-Tera.ch community have showed their strong interest on information as "Text" with more than 96% of interactions on links linked to text and 4% on links assigned to an image.

4.2 Categories

Nano-Tera.ch decided to define a category for each of the research topics – Health, Environment, and Security – plus a topic related with the community itself – Nano-Tera.ch. The categories were all organized so that all were present in the newsletters. At the end, the results showed individuals preferences on Health (38%), followed by Environment (27%) and Security (22%), and finally Nano-Tera.ch (8%). In fact, the results on Health can be possibly explained due to the higher number of projects that were born on the Health category and to the Nano-Tera.ch strong research on Health.

4.3 Connection Degree

To illustrate the results on the CD model it was used the external tool NodeXL [17]. Based on the Microsoft Excel, the NodeXL enables the exploration of a community based on every input value. In this 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 D higher than 7, in order to highlight the most important connections and have a clearer view of the relevant nodes on the network (Fig. 6).

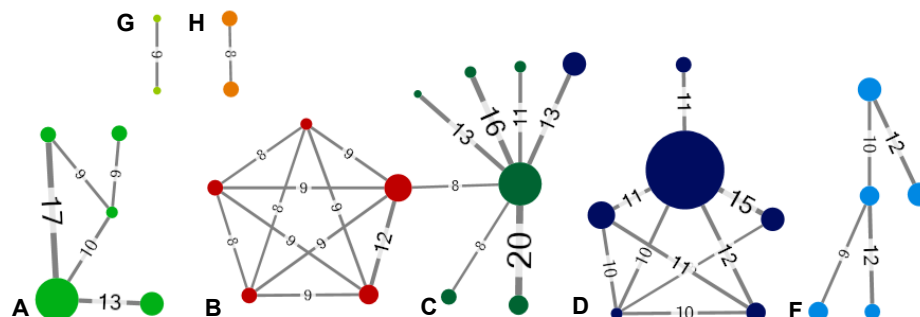


Fig. 6. Connection Degrees

The size of the nodes represents the total number of clicks that the individual have 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") have also a high CD between them, followed by the cluster B ("Health" and "Security"), the cluster F ("Environment"), and the two isolated clusters G and H with almost all categories. The CDs at the cluster C shows that the "Security" category is the strongest between individuals with a high CD.

4.4 Clusters Detection

The results on clustering detection were performed using the external tool Vizster [12]. Vizster is a social network visualization system that uses Girvan-Newman algorithm to discover clusters in a network of connections.

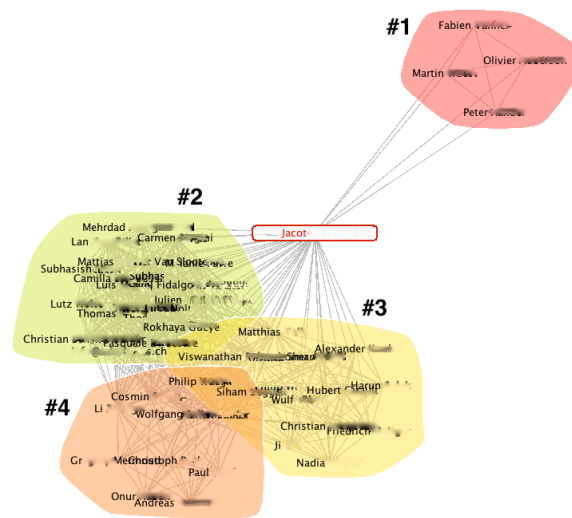


Fig. 7. Clusters Detection

The NTS organizes implicit connections into a network by defining individuals as nodes and connections as edges. Through data exportation, communities are able to use Vizster to visualize NTS discovered network and organize individuals into clusters. The results on the Nano-Tera.ch community showed that individuals are organized into 4 different clusters (Fig. 6). The details show that the Cluster #1 is mainly based on individuals with strong interest on “Nano-Tera” category; the Cluster #2 on individuals with high interest on “Health”; and the Cluster #3 and #4 on individuals highly interested on “Security” and “Environment” respectively.

4.5 Discussion

Once the data is gathered and the results achieved, communities move efforts to better understand the results and extract their value. The presented discussion is based on the Nano-Tera.ch case study and describes some of the conclusions reached at Nano-Tera.ch offices. On links’ type, individuals revealed their text-oriented focus by interacting with a higher number of text-based links than image-based. The results showed that individuals have a strong scientific focus and care about the newsletters’ content. A brief description of a topic will better promote individuals to read it and click it. On the other hand, images had a strong power on balancing newsletter design but a low level of interaction. In fact, “Images” help creating newsletter’s context and “Text” allows individuals to go deeper in the subjects.

On categories, “Health” has proved its high number of interactions and individuals’ preferences on the subject. In fact, the results reflect the higher number of projects on Health and the Nano-Tera.ch main focus on Health issues. The balance between both categories of “Environment” and “Security” can be explained with the approximate number of projects and with the proper division of the topics per newsletters. The fact that Nano-Tera.ch designed all newsletters with fresh news leads individuals to interact with almost all categories in order to stay updated. On the other hand, “Nano-Tera.ch” category showed that individuals are interested in their own community.

By applying the Girvan-Newman algorithm through Vizster, Nano-Tera.ch can divide individuals into 4 main clusters where relations represent implicit connections. Clusters’ size showed 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 showed that Nano-Tera.ch community is strongly divided into its main categories and that individuals within a category have a strong connection between them. The overlaps highlight individuals that connect categories and are valuable nodes. In fact, those individuals turn out to be some of the most influential people in the community and with a high number of interactions. In the middle, Nano-Tera.ch has the most valuable individual, implicit connected to all clusters and to a high number of individuals.

The final value on the CD allows communities to place implicit connection at the same level and compare them. The CD model assigns to each connection a CD value that describes how strong two individuals are relat-

ed. CD is established between any two of the main researchers at Nano-Tera.ch. Although they were connected on their research area, the CDs on the cluster C show that the category has the strongest CD. This fact can be translated to individuals being high related as a team. Moreover, the result can be explained by the low interest on the other categories rather than “Security”. The CD on the Cluster C is also high because individuals at security projects have clicked on security links but ignored the other category links. It is also interesting to note that CDs support the Girvan-Newman clustering algorithm by showing that individuals in a group have a strong CD between them. Even with a sample of all implicit connections, the clustering algorithm continuous to cluster individuals into the four main categories: “Nano-Tera”, “Security”, “Health”, and “Environment”.

Nano-Tera.ch used these results to promote collaboration between researchers and to present to each research the people that showed interest on their work. Nano-Tera.ch also identified the main followers and asked them to promote the newsletters and to help reaching the maximum of researchers. In addition, the results were also used to promote conferences on a topic that researchers showed interest and to answer simple questions such as the number of interested people that may attend.

5 Related Work

The main work on using email as a source to extract information started with Schwatz and Wood using email headers to extract shared interests between people through graph theory [21]. However, that approach is very limited to the emails’ headers and does not take into account the message’s body and the subject, which can be the richest source on individuals’ interests. PeCo (Ogata and Yano, 1998) [18] collected users’ relationship through email headers (“From”, “To”, and “Subject”) but as well as the previous solution does not focus on discovering and explore implicit connections in a network. On the other hand, McArthur and Bruza discovered implicit connections by mining semantic associations from people’s communications [11]. 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 implicitly connected. However, that solution does not have any measure connections’ value.

Together with the referred works there are several track engines used for marketing purposes, which are also able to send newsletters and track users’ interactions. An example is the system developed by Foulger, Chipperfield, Cooper and Storms [7]. The system generates an email template and uses it to track all the receivers’ interactions. However, the detection of key user through a connection degree can be harder to achieve or difficult to understand once the model works in a black box.

Barão and Silva proposed an holistic and complex model to define the Relational Capital Value (RCV) of organizations as well as online communities [2, 3]. Explicit and 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.

6 Conclusion

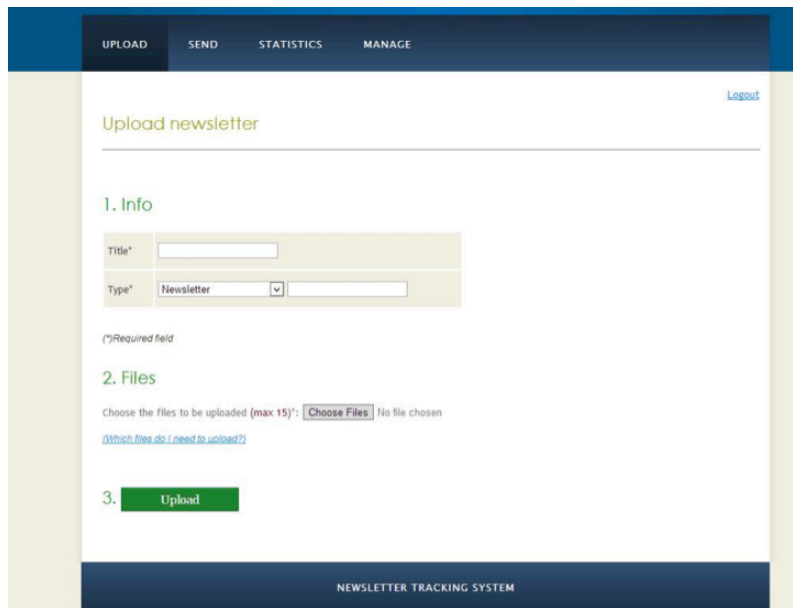
The NTS allows communities to improve the quality of their knowledge on individuals’ relationships by introducing a web system able to send newsletters and gather individuals’ interactions. Based on explicit and implicit connections NTS is able to bring to light hidden relationships and to measure their CD trough the CD model.

At the end, communities are able to foster an implicit community and to explore individuals’ connections based on analysis tools like NTS or exporting the data to external tools for further analysis. The NTS brings value on its ability to capture and expose individuals’ implicit relationships through a network. By designing newsletters, communities are able to better understand individuals and to improve the way they explore individuals’ implicit connections. We believe that the NTS and the CD model can help communities to have a more valuable overview of their network.

References

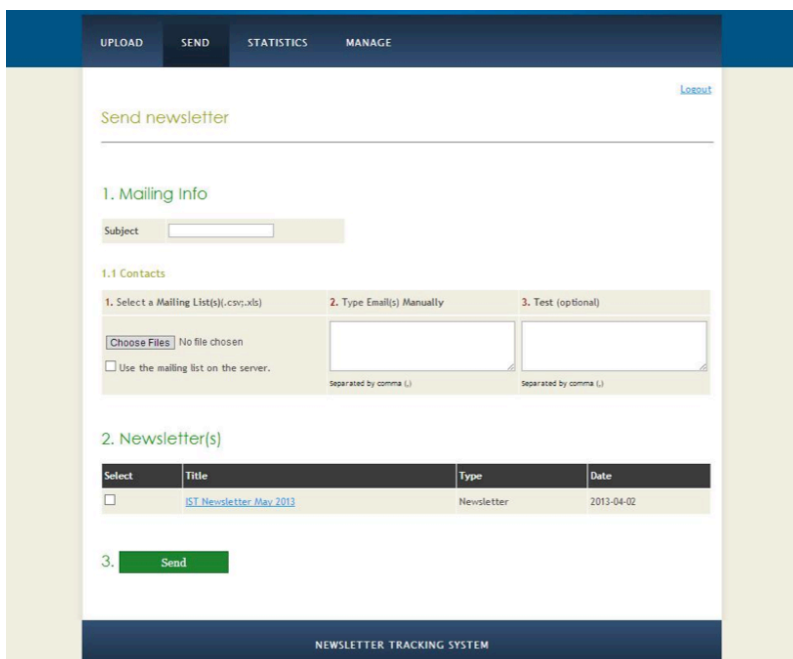
1. Aggarwal, C.: An introduction to social network data analytics. Springer Science And Business Media, LLC 2011 (2011)
2. Barão, A., Silva, A.: A Model to Evaluate the Relational Capital of Organizations (SNARE-RCO), Conference of Enterprise Information Systems (Centeris'2011), Springer (2011)
3. Barão, A., Silva, A.: How to value and monitor the relational capital of knowledge-intensive organizations?, Research on Enterprise 2.0: Technological, Social, and Organizational Dimensions, IGI Global (2012)
4. Bell, G.: Building Social Web Applications. O'Reilly Media, Inc. (2009)
5. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., E. Grinter, R.: Quality Versus Quantity: E-Mail-Centric Task Management and Its Relation With Overload. *Human-Computer Interaction*, vol. 20, pp. 89-138. , Lawrence Erlbaum Associates, Inc. (2005)
6. CKD research project: <http://www.nano-tera.ch/members/263.php>
7. Foulger, M., Chipperfield, T., Cooper, J., Storms, A.: System and method related to generating and tracking an email campaign. IC Planet (2006)
8. Harley, J., Blismas, N.: An Anatomy of Collaboration Within the Online Environment. Springer-Verlag Berlin Heidelberg 2010, 14-34 (2010)
9. i* Wiki: <http://istar.rwth-aachen.de>
10. m. boy, d., B. Ellison, N.: Social Network Sites: Definition, History, and Scholarship. In: *Journal of Computer-Mediated Communication*, vol. 13, pp. 210-230 (2008)
11. McArthur, R., Bruza, P.: Discovery of implicit and explicit connections between people using email utterance. In: Kluwer Academic Publishers, pp. 21-40 (2003)
12. Heer, J., Boyd, D.: Vizster: Visualizing Online Social Networks. In: 2005 IEEE Symposium on Information Visualization (2005)
13. M. Leiner, B., G. Cerf, V., D. Clark, D., E. Kahn, R., Kleinrock, L., C. Lynch, D., Postel, J., G. Roberts, L., S. Wolff, S.: The past and future history of the Internet. In: *Communications of the ACM*, vol. 40, pp. 102-108, New York (1997)
14. M Ridings, C., Gefen, D., A0072inze, B.: Some antecedents and effects of trust in virtual communities. In: *The Journal of Strategic Information Systems*, vol. 11, pp. 271-295 (2002)
15. Musser, J., O'Reilly, T.: *Web 2.0 Principles and Best Practices*. O'Reilly Media, Inc. (2006)
16. Nano-Tera.ch: <http://www.nano-tera.ch/topdownbottomup/index.html>
17. NodeXL: <http://nodexl.codeplex.com/>
18. Ogata, H., Yano, Y.: Collecting organizational memory based on social networks in collaborative learning. In: *WebNet*, pp. 822-827 (1998)
19. O'Reilly, T.: *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. In: O'Reilly Media, Sebastopol (CA) USA, pp. 17-37 (2007).
20. Papacharissi, Z.: *A Networked Self-Identity, Community and Culture on Social Network Sites*. Routledge (2010)
21. Schwartz, M., Wood, D.: Discovering shared interests among people using graph analysis of global electronic mail traffic. In: *Communication of the ACM* (1993)
22. Swan, K.: Building Learning Communities in Online Courses: the importance of interaction. In: *Education, Communication & Information*, vol. 2 (2002)
23. Wellman, B., Boase, J. Chen, W.: The Networked Nature of Community: Online and Offline. In: *It&Society*, vol. 1, pp. 151-165 (2002)
24. Yu, E.: *Modelling Strategic Relationships for Process Reengineering*. Doctoral Dissertation, University of Toronto (1996)
25. Zaphiris, P., Ang, C.: *Social Computing and Virtual Communities*. Chapman and Hall/CRC, 1^a Edition (2009)

Appendix: Newsletter Tracking System Screenshots



The screenshot shows the 'Upload newsletter' page in the Newsletter Tracking System. The navigation bar at the top includes 'UPLOAD', 'SEND', 'STATISTICS', and 'MANAGE'. A 'Logout' link is in the top right. The page title is 'Upload newsletter'. The main content is divided into three sections: 1. Info, 2. Files, and 3. Upload. In the 'Info' section, there is a 'Title*' text input field and a 'Type*' dropdown menu set to 'Newsletter'. A note indicates that the asterisk denotes a required field. The 'Files' section prompts the user to 'Choose the files to be uploaded (max 15):' with a 'Choose Files' button and 'No file chosen' text. A link '(Which files do I need to upload?)' is provided. The 'Upload' section contains a green 'Upload' button. The footer of the page reads 'NEWSLETTER TRACKING SYSTEM'.

Fig. 8. NTS Upload Page



The screenshot shows the 'Send newsletter' page in the Newsletter Tracking System. The navigation bar at the top includes 'UPLOAD', 'SEND', 'STATISTICS', and 'MANAGE'. A 'Logout' link is in the top right. The page title is 'Send newsletter'. The main content is divided into three sections: 1. Mailing Info, 1.1 Contacts, and 2. Newsletter(s). In the 'Mailing Info' section, there is a 'Subject' text input field. The '1.1 Contacts' section has three sub-sections: '1. Select a Mailing List(s).csv.xls' with a 'Choose Files' button and 'No file chosen' text; '2. Type Email(s) Manually' with a text input field and 'Separated by comma (,)' label; and '3. Test (optional)' with another text input field and 'Separated by comma (,)' label. There is a checkbox 'Use the mailing list on the server.' below the file selection. The '2. Newsletter(s)' section contains a table with the following data:

Select	Title	Type	Date
<input type="checkbox"/>	ST Newsletter May 2013	Newsletter	2013-04-02

The 'Send' section contains a green 'Send' button. The footer of the page reads 'NEWSLETTER TRACKING SYSTEM'.

Fig. 9. NTS Send Page