Identifying informal knowledge networks through SNA, revealing the stickiness of communities.

Erwin Jonkeren* and Marleen Huysman**

*Knowledge Management Group

Wormerhoek 14k

2905 TX Capelle a/d Ijssel

The Netherlands

EmailL e.jonkeren@k-m-g.nl

**Vrije Universiteit Amsterdam

De Boelelaan 1105

1081 HV Amsterdam

The Netherlands

Email: mhuysman@feweb.vu.nl

'Submitted to OLKC 2006 Conference at the University of Warwick, Coventry on 20th - 22nd March 2006'. UK.

Abstract

The main purpose of the paper is to present a methodology that can be helpful to identify knowledge networks or knowledge communities and its relation to the formal organization. The methodology is used in a case study that is based on a multi-method approach applying both qualitative techniques, such as interviews observations and document analysis as well as quantitative techniques, namely statistical analysis of survey responses as well as social network analysis. The research was conducted at a Dutch research institute specialized in documentation of historic documents. All employees have their offices in the same building and at the same floor. Work is divided formally over seven research programmes related to a certain time period in the Dutch History (e.g. Middle Ages, 17th century, WO II) each consisting of yet smaller research projects. We found that the informal knowledge networks only to a small extent differ from the formal work units. In the discussion of this paper, we will elaborate on the method and its implications.

1 Introduction

Communities are informal self-organizing groups of people sharing a particular practice and / or an interest in a particular knowledge area[i]. Communities have 'full participants', who teach the new or less active and less knowledgeable members through legitimate peripheral participation (Lave and Wenger 1991). As a result of social learning, ongoing interaction and story telling, communities accumulates wisdom. Communities thus also serve as corporate memories (Brown & Duguid 1991). Brown & Duguid (1991) argue for interchanges between communities. They state that "out of the friction between competing ideas can come the sort of organizational sparks necessary for igniting organizational innovating". Innovation thus not only originates inside communities in the form of non-canonical practices, but also in the interaction between communities.

Communities can thus deliver several advantages for organizations. They foster innovation through the creation of non canonical practices or best-practices, they serve as learning mechanisms through which these best practices are spread and they accumulate wisdom thereby acting as corporate memories. In order to leverage on these communities, firms require the ability to make strong perspectives within a community as well as the ability to take the perspective of another into account (Boland & Tenkasi 1995).

Considering the advantages communities can deliver, they are a promising tool for knowledge management (Wenger & Snyder 2000). However, to achieve these advantages, the identification of the communities and its members is required. Yet, the identification of communities is one of aspects that makes the relation between knowledge management and communities uneasy. Communities are well-suited for social learning but because of their informal and flexible nature, it is problematic for organization and management to be aware of their nature and existence, let alone what is learned in these units (Huysman 2004). If management decides to cultivate communities like Wenger et al. propose it would be helpful to know which people already are forming communities so management knows who they must help to get more in contact with each other. There are various ways to study informal groups, such as through interviews and observations and through email tracking. Lately, a renewed interest from social network analysis has emerged, discussing SNA as a method to identify communities In this paper we will briefly review the existing methodologies. We will than discuss how we have used one of these methods, namely SNA, to visualize the informal flow of knowledge and identifying cliques around knowledge areas (See Cross et al). We will use the outcomes of the method by comparing it with the formal project structure of the organization under study. In the discussion part of this paper, we elaborate on the findings that the knowledge communities seem to stick to the actual work practices. We will first start the paper with a review of the various methods to analyse communities.

2. Review of methods to identify communities

Before going on to describing the method we have used, a short review of alternative methods to identify communities is given. Then we will discuss the negative aspects of these methods.

2.1 Ethnography

Communities originally have been observed and studied by means of ethnography as one of the most dominant methods in anthropology. The use of ethnography is still the most important methodology and has been used to study communities in their day to day activities. More specifically, ethnography is used to study situated learning in communities of practice from a cultural perspective. Based on cultural-interpretive research methods, learning is studied within e.g communities of system analysts (Ciborra and Lanzara, 1994), maintenance engineers (Orr, 1996), midwives (Jordan, 1989), flight crews and ground staff (Weick and Roberts 1993), claim processors (Wenger 1998), IT consultants (Teigland and Wasko 2000), flutemakers (Cook and Yanow 1993), and technicians (Barly 1996). The interpretive ethnographic methods try to reveal how the 'social world is constituted by the local production of meaningful action' (Suchman, p 58, 1987). The focus is not so much on the outcome or the achievement of learning but on the process of learning as it is actually taken shape as part of the day to day activities of communities. During their day to day interactions, people learn to become a practitioner, such as a photocopier repairman or a experienced midwife. Learning within these communities takes place through the communication of tacit knowledge. Or as Yanow (2000) puts it '..in interaction with and through the actifacts, leaving their embodied meanings unspoken' (p.255). This learning is very much tacit. In terms of Polanyi, practitioners learn to become a community member while focussing on something else.

The major drawback of this method is that it is extremely time consuming. It means that the research need to be participate in the periphery of the communities in order to be able to make sense of ongoing learning processes. Also, researchers can only observe small communities of practices.

2.2 Interviews

When interviews are used to identify communities the so called 'snowball sampling' method is used. Prospective members of a community are interviewed and the interviewers look for challenges and problems that people across units and teams have in common which would serve as the bases for communities (Wenger & Snyder, 2000). The interviewers also ask who else they should talk with about the challenges and problems an interviewee identifies. When, after interviewing several people, the same names keep getting mentioned a picture of the potential core group would come into view.

The advantage of this method is that it is relatively accurate. Interviewers can ask a large amount of questions and address uncertainties immediately. However, this method is time consuming and labour intensive, especially when communities need to be identified in large organizations (Tyler, Wilkinson & Huberman, 2003).. Another drawback is that when there are two communities that are disconnected, there is no guarantee of identifying both. If the snowball starts rolling in the wrong place, whole subsets of actors who are connected, but not attached to the starting point may be missed (Hanneman, 2001). As a consequence, it becomes uncertain if the picture of the network is complete. This would make it hard to assess to what extent communities are connected and to help them get connected. Therefore this method can only be applied when screening relatively small groups for communities

2.3 Ontology based Community identifier (ONTOCOPI)

This method, abbreviated with 'ONTOCOPI' was developed at the University of Southampton by O'Hara, Shadbolt and Alani (2002). The developers describe their method as follows:

"The insight behind ONTOCOPI is that if an ontology of the working domain of an organisation is created, then the links between the instances can be measured to indicate which are closely related. If certain canonical members of a COP can be isolated in advance, then ONA can be used to identify other instances related to them. The hypothesis underlying ONTOCOPI is that (some) informal relations can be inferred from the presence of formal relations. For instance, if A and B have no formal relation, but they have both authored papers (formal relation) with C, then that indicates that they might share interests (informal relation); clearly this is not necessarily true, but is a reasonable enough assumption to support COP identification."

This method is computer based; the data concerning the created ontologies, and links between instances are fed into the computer. A big advantage of this is that when this is done, community identification is easy and fast. However, this data about ontologies and links between instances is not always available to any given organization and collecting it would take a considerable amount of time. Also the method relies on the hypothesis that informal relations can be inferred from formal relations. Although this is true for some informal relations, like the authors already indicated, this is of course not true for all formal relations. This might imply that one also need to consider if the communities identified are really communities of practice. ONTOCOPI identifies communities by selecting an instance and then see who are closely related to this instance. However, the starting point would then determine who appears in the network. So, just as the interview method, this method is not well suited to get a picture of the complete 'community of communities' network and to help get communities connected

2.4 E-mail tracking

Tyler, Wilkinson and Huberman (2003) have developed an automated method to identify communities in organizations. This method makes use of Social Network Analysis (SNA) techniques. The aim of this research method is to uncover structures, patterns and regularities in relations among people. As a research method SNA offers a set of techniques to analyze social structures. This method consists of two basic steps (Tyler 2003). The first step is to use the headers of email logs to construct a graph. In this graph vertices are the senders or recipients of email and the links between vertices denote a direct mail between them. In the second step Tyler et al. (2003) use an algorithm that can identify communities embedded in a graph. The advantages of this method are that it is automated and scale free. It is a very efficient way of identifying communities and it organization size is no limitation. A major drawback of this method is that the 'domain' of the identified communities is not known. However, people may be part of different communities with different domains. Therefore this method may obscure communities and depict several communities as being one large component. Only additional interviews could establish the domain of communities and overcome these problems but that would be resource heavy and impair the big advantage of this method, its applicability on large networks.

3 Social networks analysis based on surveys

The method we present in this paper in detail is based on a combination of interviews and SNA techniques. It should be noted that SNA techniques to identify knowledge flow in communities are not new, see for well known contribution to this field (Cross and Borgatti 2004; Cross et al 2001). However, while their methodology is mainly intented to reveal the advise network within the organization, while the method we present in this paper is to reveal the existence of informal knowledge communities within the formal organizational structure.

The presented method consists of three steps. First, relevant knowledge areas need to be indicated to ascertain the domain of the communities. In the second step data on who is connected to whom is collected by means of a survey and fed into a computer program to construct a graph. This allows us to get a picture of the complete 'community of communities' network in the organization. The last step uses an algorithm to identify communities in the obtained graph. The advantage of this technique relative to the techniques discussed above is that it can map the organization as a community of knowledge communities in a relatively short period while not being too time consuming. Given the emphasis on knowledge communities as communities of practice, this method is mainly applicable in highly knowledge intensive organizations, such as research institutes and consultancy firms.

By identifying the domain and obtaining a picture of the complete 'community of communities' network concerning that domain, this method will not succumb to the problems the other methods have encountered that were outlined in the last section. Below we shortly present the 3 steps. Section four will flesh out the steps by discussing the methodology used in practice.

3.1 Establishing the relevant knowledge areas

This step is to identify the domain of the community or communities to be identified.

The way that the domain or knowledge areas are established may differ depending on the reason to identify communities. For example, when management wants to build a certain competence, the central topic(s) making up this competence may serve as the domain(s) of the community/communities to be identified. This means that management should establish these knowledge areas themselves. However, when management wants to cultivate communities, as in Wenger's et al. intention, interviews with employees will be necessary to determine which topics are much discussed. In case of the later, this part of the research might be the most time consuming, certainly in case it is combined with more ethnographic methods such as observations.

3.2 Collecting and processing the data

Because the aim is to identify communities in an organization the boundary of the actor set is easily established. All members of the organization should be included in the actor set. The questionnaire consists of two questions to map the knowledge flow in an organization so that a knowledge sharing network can be constructed. Two basic questions are used to collect the data. For each knowledge area two questions are asked: how often person i asks person j to share his or her knowledge and how often person i gives person j information without person j having asked for this knowledge. To indicate how often people asked questions or gave information respondents could choose from six categories ranging from ‘less then one a quarter’ (1) to more then once a day’ (6). As such the data is directional and is collected via a roster question format (Wasserman & Faust, 1994). Each respondent is given a list of all the other actors in the actor set. Using a roster format improves the recall ability (Storck & Richards, 1992). People also had a free choice in the amount of people they could name. Since there is no distinction made between email or face to face interaction every possible way of interacting is recorded. After receiving the filled-out questionnaires, data is fed into UCINET (Borgatti, Everett & Freeman, 2002) and is visualized with Netdraw. When entering the data in the software people are only connected when it was clear that both persons were engaging. After all ‘mutual engagement’ (Wenger, 1998) is one of the characteristics of a communities. So an edge between vertices was only created when person i indicated asking questions to person j. Assuming the question is answered, both people are then engaging. When person i indicated giving advice to person j, and person i and j did not ask each other questions, the edge was only created when person j also gave advice to person i to ensure that both persons were engaging.

In a series of studies about the accuracy of self report, as is the case in this research, Bernard, Killworth & Sailer (Bernard & Killworth 1977; Killworth & Bernard 1976;1979; Bernard, Killworth & Sailer 1980, 1981,1982) found that about half of what people report about their own interactions is incorrect. They concluded that research based on questions such as ‘who do you talk to?’ is not credible. In reaction to this conclusion Freeman, Romney and Freeman (1987) performed a follow up study. This study confirmed the findings of Bernard and his colleagues. However, what it also revealed was that the errors introduced by false recall and forgetting were systematically biased. Freeman et al. found that informants forget to mention people that they deal with infrequently and create false recalls around people who they deal frequently with. Therefore they concluded that the bias then works in the direction of consistency with long term patterns of interaction. In this research that is exactly what we want to investigate since interaction in communities is considered regular and frequent. The informant bias then works towards rather then against us.

3.3 The Girvan-Newman algorithm

When the data are fed into the computer the Girvan-Newman algorithm provided by Netdraw can identify whether there is a community structure in the network. A community structure is present in a network when there are subsets of vertices within which vertex–vertex connections are dense, but between which connections are less dense (Girvan & Newman, 2002). In this research, a vertex represents a person in the network. The figure below depicts a network containing a community structure.

- Figure 1 Sketch of a network with a community structure about here -

In this case there are three communities with dense internal links but between the three communities links are much less dense.

The algorithm is based on the principle of edge betweenness (Freeman, 1977). The betweenness of an edge is defined as the number of geodesics (shortest paths), connecting vertex pairs, that go through that edge summed over all vertex pairs in the network. What the Girvan-Newman algorithm does is calculate the betweenness of all edges in a network and then remove the edge that has the highest betweenness score. This step is then repeated until no edges remain. During this process a network will first be split into two. Then these two networks will be further split until they consist of only one individual. At the end of this process the algorithm has produced many ways of splitting up the network into separate communities. To determine which community structure is best the modularity (Newman & Girvan 2004) of all community structures identified by the algorithm has to be compared. The modularity is a measure that quantifies the strength of a community structure. The community structure with the highest modularity is considered to be the structure that reflects reality best. According to Newman and Girvan modularity scores above 0,3 indicate a strong community structure.

Respondents also indicate the frequency with which they share knowledge in order to have valued data because communities are similar to networks consisting of strong ties (Wenger, 1998). These strong ties are associated with a relatively high frequency

of interaction (Hansen, 1999). Homans (1958), cited previous work that demonstrated that "the more cohesive a group is, that is, the more valuable the sentiment or activity the members exchange with one another, the greater the frequency of interaction of its members" (Younger, 2004). When the algorithm can not find communities with relations that resemble a relatively low frequency of knowledge sharing included in the network, then these relations will be removed starting with the relationships with the lowest frequency (this is called dichotomization). The algorithm will then subsequently be applied on a network consisting of ties with higher frequencies then the one removed. This should allow the algorithm to identify cohesive groups such as communities. When the modularity of a community structure at a certain level of dichotomization is 0.3 or greater we can assume we have identified communities.
It may be possible that the algorithm identifies one person as being a community. This person could then for example be assigned to a particular community based on reciprocity, the amount of ties or strength of these ties.

4. Results

The research was performed at a scientific research institute in The Netherlands. Access to the institute was negotiated through the director who was the main sponsor of the research. All 56 employees of the institute have their offices in the same building and at the same floor. Work is divided over seven research programme's consisting of smaller research projects. There are a total of 32 researchers of which three are leading the research programme's and take part in the management team together with the director of the institute. There are 13 research assistants who function in a 'pool' to assist any of the researchers in tasks such as processing the research data. Five employees work for the ICT department who, for example, help researchers in the online publishing of their research. The remaining six employees have functions ranging from PR to the secretary. Because beforehand it is not known who participate in communities all 56 employees of the institute were part of the actor set.

4.1 Establishing the knowledge areas

As mentioned earlier the first step was to establish knowledge areas communities can form around. In this research it was decided to define topics around the knowledge needed at the institute to produce its output. To get a preliminary idea of these topics, a few publications were studied and observations where made and a project meeting was attended. Then to verify whether the identified knowledge areas indeed referred to the most salient organizational knowledge, a 40 minute interview with two researchers (of which one was also a member of the management team) and a research assistant was held. Before the interview we had four knowledge areas in mind. During the interviews three of these were confirmed by the interview subjects, one was dismissed and one was added. The definitive knowledge areas and a short description are given in table 1.

- Table 1 about here –

4.2 Collecting and processing the data

Of the 56 questionnaires that were handed out, 43 were returned resulting in a 77% response rate. Other network researchers have analyzed data sets with response rates between 90% en 65% (Storck & Richards, 1992). Thus the response rate is good enough to perform analysis. Storck & Richards also recommend asking questions that capture both sides of the relationship. This means not only asking questions such as 'who do you give advice"' but also, 'who do you receive advice from? When this is done, this reduces the risk of including non-existing relationships due to false memories of respondents that the studies performed by Bernard, Killworth & Sailer have found often to be present. However, because the collected data was part of a larger research that included more questions and because people already indicated that they were tired of surveys (they had completed several surveys in the period before this survey was performed) it was feared that asking additional questions would be too much for potential respondents and jeopardize the response rate. Therefore it was chosen not to capture both sides of the relationships. It was thought that the 'long term' bias that could be introduced by not doing this would not be a problem since it is exactly those 'long term' relationships that exist in communities that are the subject of our investigation.

Additionally, because the bias introduced by not confirming relationships tends to create a long term pattern in the data this does not really pose a problem because it is exactly such a long term pattern (relations making up a communities) that is the subject under investigation.

4.3 The Girvan-Newman algorithm

The Girvan-Newman algorithm was applied to the networks of all the five knowledge areas. Four networks showed a satisfactory community structure when a level of dichotomization was applied that included only relations with a frequency higher then once a quarter. The network of knowledge area four showed a satisfactory community structure when only relations were included with a frequency of once a quarter and higher. Below the networks and communities in them are showed and a table with the corresponding modularity scores will be given.

- Figure 2,3,4,5 and 6 about here –

- Table 2 about here

The nodes that are grouped with a circle around them are identified by the algorithm as communities. The modularity of the community structures are all but one well above 0.3 indicating that the algorithm has found a satisfactory community structure in all the networks. In only one network we can see a small disconnected component of three people. People that were not connected at the level of dichotomization are not displayed. In a few networks we can see components of two people. We have chosen to not mark these as communities since we believe that communities should be more than a dyad.

Verification

The issue to consider now is to verify if the identified communities are indeed communities or if they simply reflect the formal structure of the organization. To do this, the amount of relations between researchers that cross formal boundaries in communities identified by GN will be expressed as a percentage of the total amount of relations between researchers in those communities. Communities are often thought to cross these formal boundaries and by doing this we will get an indication to what extent relations in communities identified by GN cross formal boundaries. It was decided that only the relations of researchers would be investigated because it was only of these employees that we could assure they had no formal contacts outside their research programme. A relationship between person i from community i and person j from community i where person j is either a research assistant, ICT employee, pr or secretary employee and where person i is from a different formal department, or programme group, then person j could still cross formal boundaries. After all, the ICT employees and research assistants assist any of the researchers regardless of the programme group they are in. So only of the researchers we can say that they have no formal contacts outside their programme group and that any such contact would be informal. As further means to verify if we have successfully identified communities, structured interviews were held with four people working at the institute.

Relationships between and within departments

The table below gives insight into the rate of relationships between researchers in the same community but from different formal departments.

- Table 3 about here

As can be seen in the table, on average 9.4% of all relationships between researcher i and researcher j in community i cross formal organizational boundaries. Although marginally, at least some of the relations in communities are crossing formal organizational boundaries. So, if knowledge sharing between departments is not common, this might be an indication that we have not just mapped the formal relationships at the institute. The small percentage of relationships between researchers in the same community but from different departments could then be logical.

Both quantitative and qualitative research data from a research performed earlier at the institute give more insight into this matter. The aim of the mentioned research was to establish to what extent the organization is managing its knowledge. As part of this research, questions (see table three below) were asked about sharing knowledge within and between departments. Respondents could answer on a likert scale ranging from (1) ‘totally disagree’ to (5) ‘totally agree’.

- Table 4 about here –

Of the 56 questionnaires handed out for this research 37 were returned resulting in a 66% response rate. First we will compare the answers of the 23 researchers who returned the questionnaire. Then we will look at results of all the respondents.
Table 4 gives insight into the distribution of the answers over the possible answers.

- Table 5 about here -

When looking at the table it becomes clear that, for researchers, knowledge sharing within departments is much more common than between departments. Only 4.2 % agreed and 12.5 % partly agreed to share knowledge with colleagues from other departments against respectively 58.3% and 29.2% for knowledge sharing within the same department. The fact that contacts between researchers in the same community but from different departments are rare then becomes much more logical.

This is further reinforced by qualitative data from 9 semi-structured interviews held during the same research. These interviews were not tailored to see if there was knowledge sharing between departments. However, four out of nine interviewees gave indications that this was sparse. For example when asked "how did you learn your profession and how important has the organization been in this process?" three people indicated that they mostly learned from their colleagues within the same research programme. For example, one of these persons said *"when it happens that people get into discussion with each other, these are almost always people from the same department"*. What also was striking was that two people referred to the culture of the institute as being a 'culture of islands'. With this they both referred to a culture of islands between hierarchies and between departments.

On the basis of these quantitative and qualitative results it is not strange to see that only 9.4% of all relations between researchers in the same communities cross formal organizational boundaries and can be considered logical. This means that we do not have to conclude that we have simply mapped the formal structure of the organization. The following section will report on the interviews that were held to investigate what the nature of the communities then is.

Establishing the nature of the identified communities

Now that we know that the identified communities do not necessarily have to represent the formal structure of the organization we have to establish what the nature of the communities is. To address this, structured interviews were held with four people with different functions and from different communities. The interviews lasted between 10 and 20 minutes and took place in the offices of the interviewees. Eight questions were asked about the communities these people were part of according to the GN algorithm. Four questions were related to group characteristics, three questions were related to 'achievements' of the community and one question was about relational characteristics. The interviews were recorded with handwritten notes by the interviewer. The questions are given in the table below

- Table 6 about here -

Three of the four interviewees recognized the communities on all the knowledge areas the GN algorithm put them in. More specifically they described them as being their project group or multiple project groups making up a programme group. After being presented the communities one person was in he responded with *"that's funny, that's my project group"*. Another said *"these communities pretty much reflect the project groups, or multiple projects groups actually"*.

Questions 2a, 2b and 2c were framed to see if learning and problem solving takes place in the identified communities. The three subjects that recognized the communities all responded positively to the questions although one noted that although discussions and learning took place, it did not occur very often. So although we have not seen cross departmental communities, at least the communities the interviewees are part of seem to have these important community characteristics.

The responses to question three were quite similar. Three interviewees reported not having more difficulty at admitting not to know something to some outside his or her communities as opposed to someone from within his or her communities. One person said that he would have no difficulty doing this to anyone from within his communities and to all but a few outside his communities.

During the interviews we have also asked the subjects to name the people they considered to be the most knowledgeable in each of the knowledge areas. This was done to see if we can identify the full participants or ‘thought leaders’ of communities by using SNA techniques. The appropriate network measure for this is called ‘indegree’. The indegree of a node is a measure for the amount nodes adjacent to it or put simply, the amount of nodes it is connected to. According to social network theory the indegree of a node is an indication of the ‘popularity’ of a node (Wasserman & Faust, 1994). In the tables below we will rank the actors by their indegree. Only the actors with the highest indegree, the actors with the lowest indegree mentioned to be knowledgeable by the interview subjects and all the actors in between are shown. The total rank can be seen in the column ‘rank’. We compared the results of the interviews and the indegrees and see if the people with the highest indegree are also mentioned by the interview subjects as being the most knowledgeable. When the indegree was analyzed, relationships between people based on giving advice were excluded. When these would be included this would mean that the indegree of a person would also be affected when they receive advice, which is of course not an indication for expertise.

- Table 7,8,9,10,11 and 12 about here –

When we look at tables representing knowledge area two and three we see that the people named by the interview subjects are ranked highest out of thirteen. In the knowledge area one, they are ranked second and third out of thirteen. In knowledge are 4 they are ranked first, third and fifth out of ten and in knowledge area five they are ranked second, third, fourth and fifth out of twelve.

This illustrates what is already known in SNA literature. People with the highest indegree are generally considered to be popular or in this case ‘knowledgeable’. The method to identify communities presented in this paper thus also allows the identification of full participants in communities. This can be of importance as when the people identified as thought leaders are the ones that could act as knowledge broker between communities and the organization.

5 Conclusion and discussion

In this paper we reported on a study that was meant to identify knowledge communities and that uses a mixture of various methods. We have been quite detailed in reporting the research methods in order for other researchers to re-use the methodology and refine it.

In order to identify organizational knowledge used by all employees and that transcend the formal research programmes, we identified five knowledge areas. This was done by means of observing project meetings, document analysis and conducting interviews. The five identified knowledge areas relate to heuristic methods, communication methods, processing research findings, use of IT and more general historic knowledge. Most of these knowledge areas concern procedural knowledge or know how, which usually is of a tacit nature (Ryle 1945). We then handed out a survey to all employees posing questions such as 'to who do you give advise" and 'to whom do you go to, to ask questions" in relation to these five wide-ranging knowledge areas. The Girvan-Newman algorithm used to analyse the structure of communities (Newman & Girvan 2004) was applied to the networks of all the five knowledge areas. With the use of this social network analyses we spotted various knowledge communities. When comparing these result with the formal structure of the organization, we observed that the identified knowledge communities formed around collective organizational knowledge are for more than 90% situated within the formal boundaries of the sub-units (the research programmes). Moreover, we observed that there is only a slight difference in composition (membership) among the five different knowledge areas.

These findings suggest that people flock together in communities of practices related to their daily work practices. Wide-ranging knowledge used and shared by all members of the organization that crosses structural boundaries, sticks to these communities of practice.

The use of the methodology to identify communities has proven to be useful, at least if we use the assessment of the respondents as an indicator for the validity. Applying it in this particular organization, shows that communities do represent the informal knowledge flow and the informal 'thought leader', but also that it represents the formal work unit. This brings us to a discussion about the fluidity of communities, as has often been expressed in the literature on communities of practice. Our case shows that knowledge communities tend to stick within the locality of the formal work units. While this observation supports the literature on situated cognition (Lave 1988) sticky and leaky knowledge and boundary spanning (Duguid 2005, Carlile 2002, Beckhy 2003) locality of knowledge and situated learning (Suchman 1987, Allen 1977, Sole and Huysman 2001), it provides counter-evidence for the assumption that knowledge communities cross inter and intra-organizational divisions of work. Although the findings are based on one case study and thus cannot be generalized, it does bring into question the often expressed argument that communities can exist independent of work practices.

The method presented in this paper can be applied for a variety of knowledge management purposes. First, it shows which people make up communities with respect to a particular knowledge area. This can help in cultivating these communities as Wenger et al. (2002) have proposed. Second, it shows to what extent these communities are connected to each other. Based on this information one might want to decide to connect communities with the aim of innovation in mind (Boland and Tenkasi 1995). Last, this method allows for the identification of the thought leaders with respect to certain knowledge intensive communities. The methodology thus also allows for effective knowledge extraction from communities by collaboration with these individuals.

As with all methodologies, this particular methodology has its downsides as well. First, as is known with social network analysis, the proposed method is difficult to apply to very large organizations. The answer sheet has to contain all the names of the individuals in the actor set. Also, the method requires checking all the names per knowledge area as entry points into the computer. With an actor set of a few thousand one can imagine the amount of time it would take to enter the data.

Furthermore, it is important to stress the ethical concerns the use of this methodology creates. For one, it provides insight into the degree one is active in asking questions and giving advice. This could tempt management to take actions against people they consider underperformers in this area. Also, when management intervention is a consequence of the mapping, this could trigger socially biased answers and thus endanger the accuracy of mapping relationships and the success of the method.

Future studies will have to capture both side relationships as recommended by Stork & Richards (1992) in order to map relationships as accurately as possible. Also additional questions to measure the intensity of relationships should be asked. In this study only the frequency of asking questions and giving advice was measured. However, relationships in communities also have an affective component. Asking one or more questions to measure this affective component and then average it with the frequency component just like Hansen (1999) might result in a more accurate mapping of communities.

References

Allen, Thomas J. (1977), Managing the Flow of Technology, Cambridge (MA): MIT Press, reprinted 1993, chapters 4-8.

Andriessen, J.H.E. (2005) Archetypes of Knowledge Communities, In Communities and Technologies, Conference Proceedings, Dordrecht Kluwer Academics

Barley, S. (1986) Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments, Administrative Science Quarterly 31, pp 78-108.

Bechky, B. (2003) Sharing Meaning across occupational communities: the transformation of understanding on a production floor, Organization Science , 14/3

Bernard, H.R., & Killworth, P.D. (1977). Informant accuracy in social network data II. Human Communication Research, 4, 3-18

Bernard, H., Killworth, P., & Sailor, L. (1980). Informant accuracy in social network data IV. Social Networks, 2, 191-218.

Bernard, H.R., Killworth, P.D., & Sailor, L. (1982). Informant accuracy in social network data V: An experimental attempt to predict actual communication from recall data. Social Science Research, 11, 30-66.

Boland, R. and R. Tenkasi (1995), Perspective Making and perspective taking in communities of knowing, Organization Science

Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Botkin, (1999) Smart Business: How Knowledge Communities Can Revolutionize Your Company, New York: Free Press

Brown, J. S. en Duguid, P. (1991). Organizational Learning and Communities of Practice: Toward a Unified View of Working, Learning and Innovation. Organizational Science, 2(1), 40–57.

Ciborra, C.U. and G.F. Lanzara (1994), Formative contexts and information technology, understanding the dynamics of innovation in organizations. Accounting, Management and Information Technology, vol 4/2.

Creplet, F. O. Dupoiet, E. Vaast (2003) Episteme or practice? Differentiated communitarian structures in a biology laboratory, In Huysman, Wenger and Wulf, Communities and Technologies, Conference Proceedings, Dordrecht: Kluwer Academics

Cross, R.A. and S.P Borgatti (2004) The ties that Share: Relational Characteristics That Facilitate Information Sharing, In: Huysman, M.H. and Wulf Social Capital and Information Technology, Cambridge: MIT Press.

Cross, R. A. Parker, Prusak, L. and S.R. Borgatti (2001) Knowing what we know: supporting knowledge creation and sharing in social *networks Organizational Dynamic*s, Vol. 30, 2, pp.100-120

Carlile, P.R. (2002) A pragmatic view of knowledge and boundaries, boundary objects in new product development. Organization Science,

Cook, S.D.N. and D.Yanow (1993) Culture and Organizational Learning, Journal of Management Inquiry, 2/4.

Duguid, P. (2005) The Art of Knowing: Social and Tacit Dimensions of Knowledge and the Limits of the Community of Practice, The Information Society, vol 21, pp. 109-118

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. Social Networks, 1, p215-239.

Freeman, L.C., Romney, A.K., and Freeman, S. (1987). Cognitive structure and informant accuracy. American Anthropologist 89, p310-325.

Girvan,M. and Newman,M.E.J. (2002) Community structure in social and biological networks. Proc. Natl Acad. Sci. USA, (99),p7821–7826.

Hanneman, R.A. (2001) "Introduction to Social Network Methods," Mimeo. University of California, Riverside. http://faculty.ucr.edu/~hanneman/SOC157/NETTEXT.PDF

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly, 44, p82-111.

Huysman, M.H. (2000) "Rethinking Organizational Learning", Accountancy Management and Information Technology, vol 10.

Huysman, M. (2004) 'Communities of practice: Facilitating social learning while frustrating organizational learning'. In: Organizations as Knowledge Systems: Knowledge, Learning and Dynamic Capabilities. Tsoukas, H. and N. Mylonopoulos (eds). Hampshire: Palgrave Macmillan.

Jordan, B. (1989) Cosmopolitical obstetrics: some insights from the training of traditional midwives, Social Science and Medicine, 28/9.

Killworth, P.D., & Bernard, H.R. (1976). Informant accuracy in social network data. Human Organization, 35, 269-286.

Lave, J. (1988) Cognition in Practice: Mind, Mathematics and Culture in Everyday Life. Cambridge University Press

Lave, J. en Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge: Cambridge University Press.

Newman, M.E.J. and M. Girvan (2004) Finding and evaluating community structure in networks, Phys. Rev. E 69, 026113

O'Hara K., Alani H., and Shadbolt N. Identifying Communities of Practice: Analysing Ontologies as Networks to Support Community Recognition. In Proceedings of the 2002 IFIP World Computer Congress, Montreal, Canada, August 2002.

Ryle, Gibert. 1945. ”Knowing How and Knowing That.” Proceedings of the Aristotelian Society 46:1-16

Sole, D. and M.H. Huysman (2002), Communities of Practice and the role of Location: revealing limits of virtuality and knowledge, International conference on Information Systems (ISIS), Barcelona.

Stork, D. and Richards, W. D. (1992). Nonrespondents in communication network studies: Problems and possibilities. Group and Organization Management, 17 (2),p 193-209.

Suchman, L. A. (1987). Plans and Situated Actions: The Problem of Human Machine Communication, Cambridge: Cambridge University Press

Tyler J.R., Wilkinson D.M., Huberman B.A. (2003). Email as spectroscopy: Automated discovery of community structure within organizations. In: Communities and Technologies. M. Huysman, E. Wenger & V.Wulf (Ed.). Dordrecht: Kluwer Academic.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. New York: Cambridge University Press.

Weick, K.E. and K.H. Roberts (1993) Collective mind in organizations: heedful inerrelating on flight desks, Administrative Science Quarterly vol 38/3, pp. 357-381

Wenger, E. (1998) Communities of Practice. Learning, meaning and identity. Cambridge University Press, Cambridge.

Wenger, E., McDermott, R. & Snyder, W. (2002). Cultivating Communities of Practice, A Guide to Managing Knowledge. Boston Mass: Harvard Business School Press.

Younger, S. (2004) Reciprocity, normative reputation, and the development of mutual obligation in gift-giving societies. Journal of Artificial ocieties and Social Simulation 7 (1).

Appendix: figures and tables

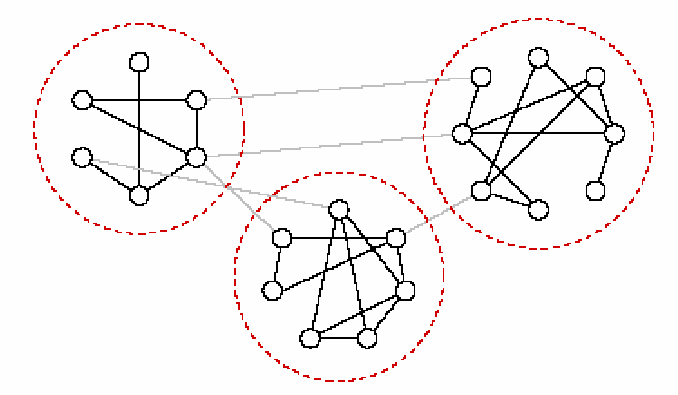

Figure 1

Knowledge area	Description
1: Historic knowledge	Knowledge of people, places etc. in history
2: Heuristic methods	Knowledge of methods about how to collect research data
3: Methods for opening up historic sources	Knowledge of methods about how to make research accessible for the public
4: Processing research results	Knowledge of processing research results generated by the researches
5: Use of electronic aid	Use of electronic aids at the institute such as the intranet and a database.

Table 1: identified knowledge areas

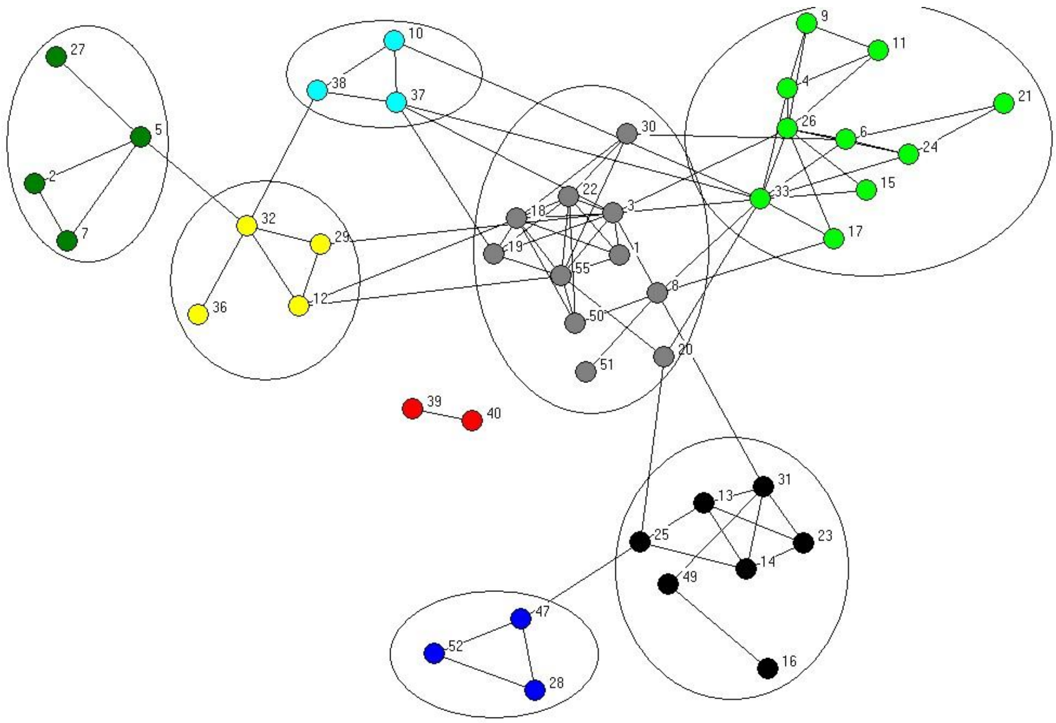

Figure 2: Community structure 'historic knowledge'

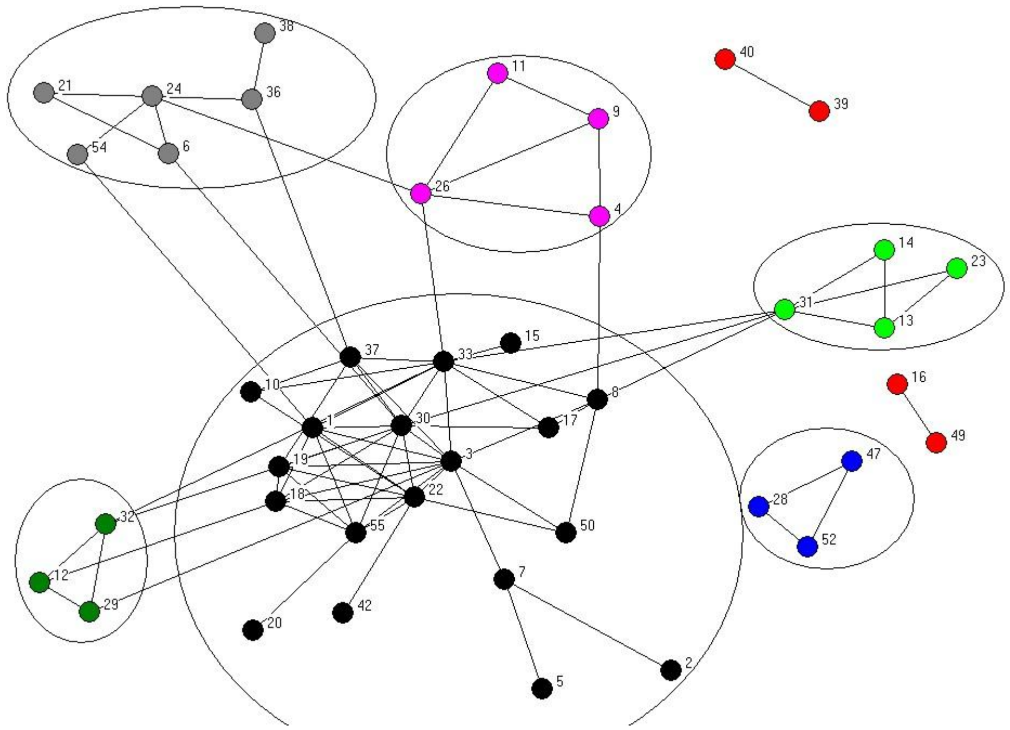

Figure 3: Community structure 'knowledge of heuristic methods'

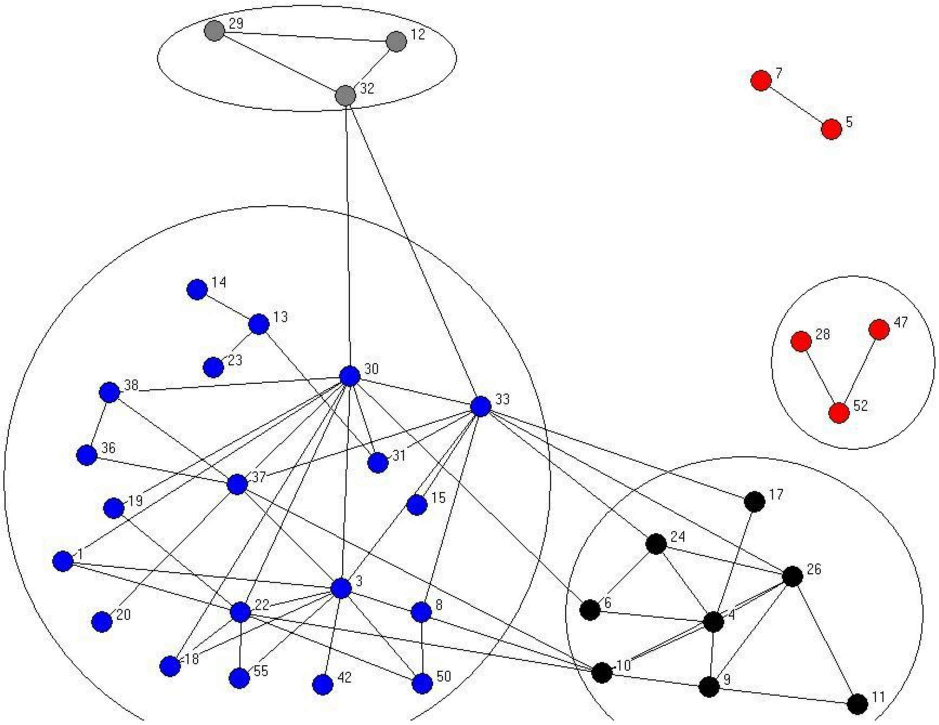

Figure 4: Community structure ‘knowledge providing access to historic resources’

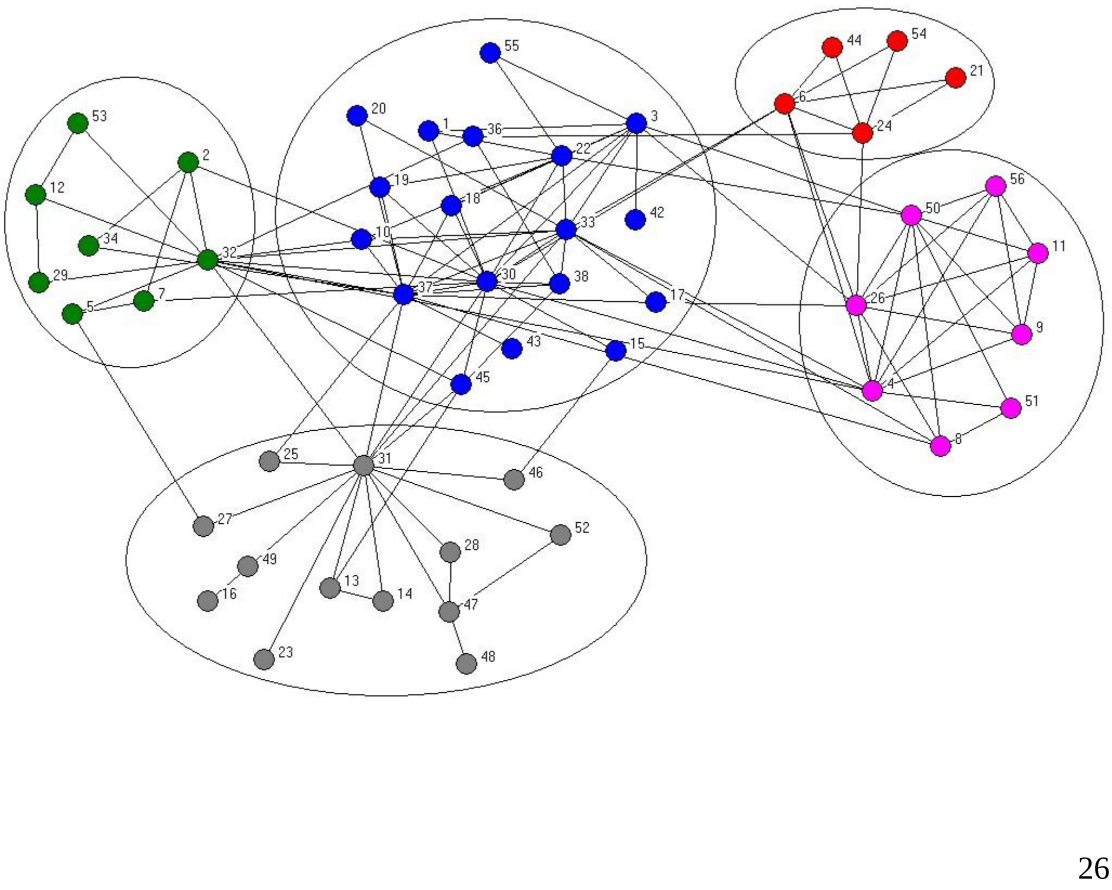

Figure 5: Community structure ‘knowledge of processing research results’

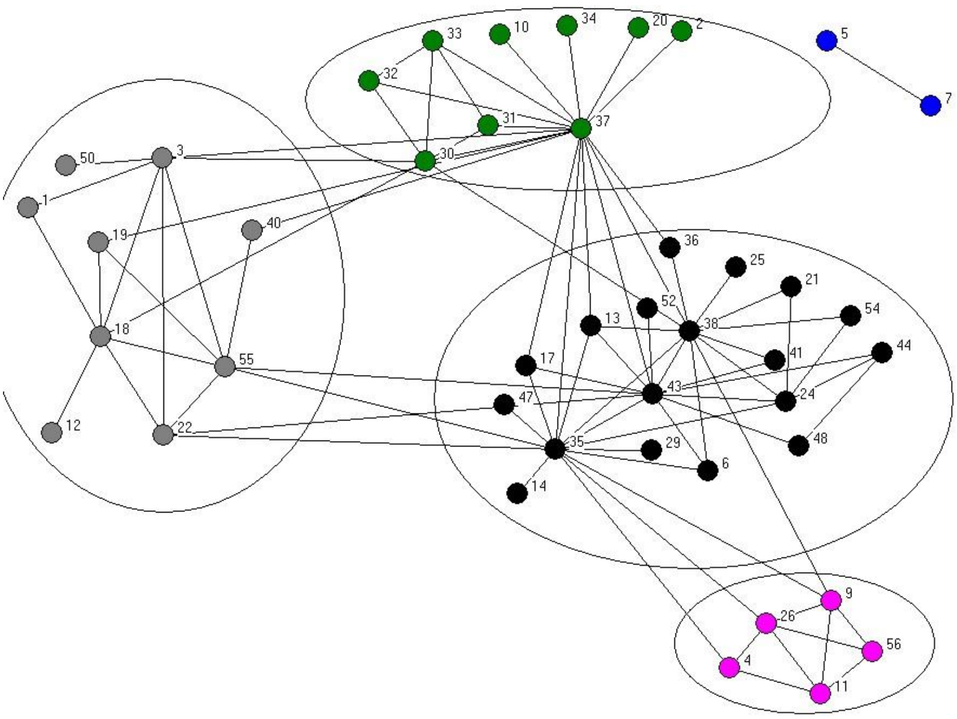

Figure 6: Community structure ‘knowledge about use of electronic aids’

Table 2: Modularity (Q) of community structures

Knowledge Area	Modularity (Q)
Historic knowledge	0,525
Heuristic Methods	0,414
Methods for providing access to historic resources	0,352
Processing research results	0,429
Use of electronics aids	0,406

Table 3: Investigating researchers' relationships

Community	# relations between researchers in communities	# number of relations between researchers in same community but from different formal departments	% relations between researchers in same community that cross formal boundaries
Historic Knowledge	30	2	6,6
Knowledge of Heuristic Methods	27	3	11,1
Methods for providing acces to historic resources	24	4	16,6
Processing research results	20	1	5,0
Use of electronic aids	13	1	7,7
Average			9,4

Table 4: propositions concerning knowledge sharing within and between departments

Propositions
A :I share knowledge with colleagues from my project/programme group
B: I share knowledge with colleagues outside my project/programme group

Table 5: distribution of answers on propositions concerning knowledge sharing within and between departments

Issue	Proposition	1	2	3	4	5
Researchers	A	0	0	12,5	29,2	58,3
	B	12,5	29,2	41,7	12,5	4,2

Table 6: interview questions

Issue	Questions
1: Group characteristics	1a: Would you say the by GN identified groups represents a group in real life?
	1b: How would you typify these groups?
	1c: Are there any people missing from this groups?
	1d: Are there people in the groups that should not be there?
2: Achievements/ Outcomes	2a: When you have gained experience in a particular knowledge area, would you share this with others in your communities?
	2b: Do you have discussions with people in these groups from which you learn?
	2c: Have you solved problems concerning a particular knowledge are with the people in your communities?
3: Relational characteristics	3: Would you feel more comfortable admitting not to know something to a member of one of your communities then to someone outside one of your communities?

Table 7: indegrees knowledge area 1

Knowledge area 1		
Rank (out of 13)	Person	Indegree
1	32	17
2	33	13
3	3	10
3	6	10
3	31	10

Table 8: indegrees knowledge area 2

Knowledge area 2		
Rank (out of 13)	Person	Indegree
1	32	13
2	3	12
3	33	10
3	30	10
3	1	10

Table 9: indegrees knowledge area 3

Knowledge area 3		
Rank (out of 13)	Person	Indegree
1	32	13
2	3	12
3	30	10
3	8	10

Table 10: indegrees knowledge area 4

Knowledge area 4		
Rank (out of 10)	Person	Indegree
1	32	11
2	4	8
2	31	8
3	33	7
3	30	7
4	8	6
4	10	6
4	17	6
4	37	6
4	13	6
4	50	6
5	6	5
5	3	5
5	20	5
5	1	5
5	11	5
5	2	5
5	38	5
5	24	5
5	56	5
5	34	5

Knowledge area 5		
Rank (out of 12)	Person	Indegree
1	43	32
2	38	30
3	35	26
4	37	23
5	33	8
5	3	8

Table 11: Indegrees knowledge area 5

Knowledge area	Persons
1	6, 32
2	3, 32, 33
3	30, 32, 33
4	20, 30, 32, 33
5	3, 25, 37, 38

Table 12: People mentioned to be knowledgeable categorized by knowledge area.

[i] Whereas the former is often referred to as communities of practice, the latter is labelled as communities of interests, knowledge communities, or epistemic communities. It can be argued argued that the characteristics of COP's are similar to such informal knowledge networks in knowledge intensive work practices. This is why we have decided to refer to 'communities' in general in the rest of this paper, while in fact referring to knowledge intensive communities of practice.