Combining Social Network Analysis with Semantic Relations to Support the Evolution of a Scientific Community
Andreas Harrer, Nils Malzahn, Sam Zeini, H. Ulrich Hoppe, University of Duisburg-Essen
harrer@collide.info, malzahn@collide.info, zeini@collide.info, hoppe@collide.info

Abstract: This paper presents an analytical approach to support organisational learning in terms of the evolution of a scientific community based on a combination of social network analysis and semantic relations. The primary and direct target of the method is to infer hidden or desirable links between subgroups in a networked community. The data source for these inferences comprises memberships in teams and thematic subgroups. The approach has been applied in a case study to a large scientific network on technology enhanced learning.

Introduction
One perspective of research on technology enhanced learning and scientific knowledge production deals with communities of practice (Wenger, 1999), and especially with the role of information and communication technologies in defining, maintaining and evolving these communities. In this view, networking infrastructures and communication mechanisms are enabling factors for community development. In a more recent article, Wenger, McDermott and Snyder (2002) make a transition from observing and analysing communities of practice towards "cultivating" such communities. The first of seven principles for cultivating communities of practice is the "design for evolution" (as a postulate). In the sequel, we will explore how a combination of analytic and semantic modelling techniques can be used to support the evolution of networked communities. This support relies, in first place, on enabling more informed decisions based on the identification of social patterns and particularly on the combination of information about topical (or thematic) and personal dynamics in the network.

Social network analysis is an established method to derive person-person relations in the form of sociograms from "traces" of communication in a networked community (cf. Wassermann & Faust 1994). Given the typical types of communication in community portals or forums, it is possible to relate the communication events also to specific topics which, in turn, can be associated with each other through an ontology (i.e. an explicit description of relevant concepts and their relations within a specific domain). Here, the ontology is an intellectual construct given a priori, whereas the sociogram results from an empirical analysis. Semantic background knowledge can be used to suggest additional interactions (e.g., of type "person A should talk to B") based on an ontology link (cf. Malzahn et al. 2005b). On the other hand, personal proximities which are not accounted for by the ontology may indicate an update of the semantic structure.

Our approach is exemplified with a community of practice in the field of technology enhanced learning (TEL), namely the Kaleidoscope Network of Excellence. The methodology has been originally developed and implemented in an interdisciplinary research project on "Virtual Work and Learning in Project Based Networks" (VIP-NET project no. 01HU0128, cf. Malzahn et al., 2005a). The FreeStyler application, a mind tool focusing on collaborative modelling and simulation, which has a direct interface to the CoNaVi application (Community Navigation Visualizer; see Malzahn et al., 2005b) provides an easy-to-use interface to these methods. It can be applied to various networked communities, such as learning communities and/or communities of practice, and makes use of different types of input data.

Use Cases and Methods of SNA for Community Support
In the following subsections we will define the basic issues of classic Social Network Analysis and then introduce our ideas for enhancing the community support by means of enriched SNA techniques. We will present the stimulation of new partnerships between network actors by the augmentation of social networks with semantic networks. The evolvement of a community goes hand in hand with the development of the conceptual framework of it. Therefore we propose an SNA based approach for the detection of conceptual changes (Slotta, Chi, & Joram, 1995) on community level. The liveliness of a community is influenced to a large extent by the trends that define the major strands of interest. We present several techniques how to spot trends to support communities in their evolution and self-reflection.
Bi-partite social networks of actors and topics

The data that networks can be built upon and the analyses that can be conducted are manifold. Social Network Analysis (or SNA, see Wassermann & Faust, 1994) relies usually on homogenous networks with one type of actors, namely persons, that are also called one-mode-networks. In these networks specific properties, such as centrality or prestige of a person or the overall centralization and density of a network can be computed by well-defined formulae.

Yet computer-supported interaction and cooperation typically involves mediating artefacts, such as written documents like emails or postings as well as diagrams or models being co-constructed by a group of learners. These “tangible” products can be seen as another type of entity to be included in the network. Exchange on the level of the original artefacts can also be provided by similarity-based search in a shared collection of objects (cf. Hoppe et al., 2005). If we restrict our analysis, in this sense, to “communication through artefacts” (Dix et al., 2004) this will result in networks only with relations spanning between elements of the two different categories (i.e. person-artefact). This network structure corresponds to the one of a bi-partite graph. Typical examples of these kinds of relations are, e.g., a person creating, deleting or modifying an artefact (cf. Ogata & Yano, 1998) such as a forum posting, or a person expressing interest in an artefact, such as subscribing to a thread or discussion board. In a generalization step, an artefact can be classified by its semantic content or theme. E.g., all forum postings classified under “SNA in CSCL” would be seen as related to one topic or theme. This would potentially reduce the number of non-person nodes from artefacts (postings) to topics and increase the number of persons gathered around one topic (as compared to the original postings). Similar bi-partite networks have been discussed in SNA under the notion of “affiliation networks” (Wassermann & Faust, 1994).

These networks provide the potential for a variety of analytical approaches and applications: Persons might be interested in other persons related to an object of their interest, to contact them for discussion and exchange; a researcher might be interested in relations between topics that have been created indirectly by people related to these topics indicating possible connectedness of the topics. These kind of mathematically inferred topic-to-topic-networks are called event overlap networks (Breiger, 1974) in the SNA literature.

In the following subsections we will discuss different use cases for SNA-based community support and explain the methods that facilitate this kind of support.

Team recommendation - Enriching bi-partite networks with expert knowledge

The example mentioned above of people being interested in other persons related to an object of their interest can be considered the computation of the path length 2 in the simplest case, i.e. the multiplication of the network matrix with itself. In bi-partite networks this will result in either person-person or topic-topic networks, i.e. one-mode networks that can be analysed further in the usual way. These strict networks structures might be interesting for analytical purposes, but have the danger of producing either already known or rather trivial insights: Imagine a member of a University's discussion network who participates in one specific discussion board; the computation of the person-person network will show him the people that he already saw in the discussion thread, thus not giving him information about further discussion partners.

A more interesting information could be the persons participating in a discussion thread that is related in some way to the thread he is interested in. This could be inferred by the creation of a topic-topic network, as discussed above, or by the explicit additional information about topics represented as a knowledge map, which tends to be available often through an expert (e.g. the moderator of the forum) or a shared conceptualisation, also known as an ontology (Gruber, 1993). Using this explicit additional information together with the bi-partite network results in a network structure that is no longer a bi-partite network, since nodes of the same type might have relations with each other. A schema for this combination of bi-partite networks with a knowledge map into a multi-mode network can be seen in figure 1.

In the following subsections we will discuss different use cases for SNA-based community support and explain the methods that facilitate this kind of support.
These multi-mode networks afford a different kind of computational method than the standard ones for affiliation networks that we call *ontology-facilitated navigation* and that is described in detail in (Malzahn et al., 2005b). Using weighted and typed relations in the additional ontology, the multi-mode network is transformed into a one-mode network (see figure 1 top right), which contains relations that was mediated only by the information expressed in the ontology. In our example of the discussion board, the student would get information about the persons involved in a separate thread she might not have been aware of, but that is relevant to her area of interest, instead of getting information about the persons she already knows. Since experiences with expert and diagnosis systems showed that users want to have explanations (Clancey, 1983) of computed results we decided to enable the user to color the relations between the topics in the ontology. The colors used in the ontology are then transferred to the resulting person-person network so that the user can identify the relations and therefore the artefacts that have led to the computed relation. Along the way the user gets the information how many different ways facilitate the relations because every color represents another path from one person to the other. This kind of information produced by our method of combining bi-partite networks with networks based on expert knowledge is mainly interesting for the future behavior of the user e.g. by contacting the newly identified persons for discussion about their joint interest. Thus the inferred network information might be used to recommend options and consequently evolve communities towards new kinds of relations.

Reconciling, evolving and validating community terms using SNA

In the previous section we have shown that ontologies are valuable tools to foster relations that might not be obvious to the majority of participants in a community. In this section we want to show how social network analysis can be used to validate presumed relations between artefacts in the observed community. When communities come into existence they usually develop their own vocabulary and standards (cf. Wenger, 1999; Zeini, 2005). This is done either implicitly by using the same terms for the same concept or explicitly by developing an external representation of their common understanding. Sometimes even the vocabulary itself is the object of common interest. This is e.g. the case with the notion of *collaboration script* in CSCL. A lively community will adapt itself to new developments in their field of interest. So should the ontology enable all members of the community to profit from new insights. The adaption will most likely manifest itself by the integration of new members into the community or by the re-orientation of - in the beginning some members of the community towards new topics (or artefacts) of interest. Sometimes this process is made very explicit by those people by announcing the next big issues and grand challenges (such as in Hoare & Milner, 2004) for a community, but most of the time the change is made silently and unnoticed by the community: there are persons who work on topics of two seemingly not connected topics - at least in terms of the agreed on knowledge map. If more and more people work on two topics not connected in the map the community should investigate these two topics concerning the nature of the link. This provides deeper insight in the topic and might strengthen a new research field.

Social network analysis can support the community by highlighting missing links, confirm existing links or even questioning presumably existing links in the community's ontology. We use a weighted and standardized co-
occurrence algorithm based on actors' relationships results in a network of artefacts that can be compared to the existing knowledge map. The resulting network distinguishes three types of relations in the reconciled map:

- **green relations** indicating that these relations were confirmed by the affiliation network, i.e. they are part of the given map and various persons are working on both topics.
- **red relations** indicating that these relations could not be confirmed by the affiliation network, i.e. they are part of the map but no one is working on both topics.
- **grey relations** indicating that these relations are emerging from the network, but they are currently not included in the map.

All three types of relations are valuable for the community. Although the green ones may seem to be trivial because they are already known, they confirm the validity of the given ontology. The red ones are important because the observer might want to investigate if either the ontology has to be corrected, because the currently existing relation is not valid (anymore) or obsolete so that no further work is spent (thus not observable). The most promising consequences might be drawn from the grey relations. These relations may indicate missing links in the ontology. If there are strong ties between two topics because of the amount or the reputation of the persons working on both topics the observer should carefully consider the inclusion of this link into the common knowledge of the community. So the analysis of relations between actors and topics can be used to evolve the common grounding of a community.

**Trendspotting - identifying topics of interest in a network**

Innovation and also research in and for the information society are to some extent driven by the emergence of new topics and trends. To identify these trends early is an important success factor. In science and research this corresponds to the identification of big issues and grand challenges (Hoare & Milner, 2004). In the commercial world, the dynamics of life-long-learning is driven by such trends. Personal and public communication can be a source of knowledge about trends. Similarly it is valuable to identify persons that are known to be trendsetters or early adopters of technology (Rogers, 1995). We propose to apply SNA-based techniques to support the identification of trends (“trend-spotting”) and suggest two related approaches:

**Trend analysis via temporal dimension**

A trend is usually associated with a new term or label coming up in the communication within a community. Thus, investigating changes in terminology over time can give an indication of the trends in a given field. For our approach of conceptualizing scientific and learning communities as multi-mode networks interacting around specific topics and artefacts, the changes in the network over time are the indicator for trends: on the one hand the differences within the network at specific points in time are important, on the other hand the relations that span across a time period give additional insights. This means that for our network analyses we consider the comparison between different "snapshots" of the network, preferably at well-defined moments, such as the beginning of a new period (e.g. a new year at university, a new period of funding for projects). Basically, entities and relations in the network could have persevered unchanged, emerged newly, or vanished. This information can be combined with relations over time that express the change or influence of a network element. This information enriches the plain information of differences, since it reflects also the potential change and evolution of persons, artefacts, or relations in a network. These phenomena in the network can be indicators for trends. They can be further operationalized with SNA methods by considering network properties, such as density of the network or centrality of a topic, in the perspective of their temporal change. An example for our approach of identifying trends in a network according to the temporal perspective will be given in the example section of this paper.

**Trend analysis via trusted authorities**

Another approach to identifying trends is to look at suspected trend setters or early adopters (cf. Rogers, 1995). These persons are usually considered as trusted authorities because of their expertise or influence on the community. It is not important that all members agree on a fixed set of authorities because this type of trend spotting relies on trust and beliefs. It is therefore bound to the personal judgement rather than a general agreement, although agreement may help to build up trust.

Accordingly, trends can be detected by examining the surrounding of such a trusted authority, i.e. the topics or persons the authority has recently established links with. This can be supported using visual navigation through the community networks by providing means to find authorities and focus on their neighborhood in a flexible way, e.g. by varying the degree of shown details.
Example - the Scientific Network Kaleidoscope in Technology-Enhanced Learning

As a proof of concept of our methods for fostering communities we chose the Kaleidoscope Network of Excellence (IST 507838). Kaleidoscope is a scientific network with institutions from academia and industry in the area of Technology Enhanced Learning (TEL). It brings together persons and teams from multiple fields of expertise and aims at the integration of concepts, institutions, and technology to strengthen the ties in the European research area. All in all Kaleidoscope consists currently of approximately 80 partner institutions and 1000 personal members of all old member states and several new member states of the European Union. Because of its size the Kaleidoscope network has a more complex structure than the usual two-mode networks discussed in the previous section. Kaleidoscope has a strong sub-community (Special Interest Group = SIG) interested in Computer-Supported Collaborative Learning, that has approximately 380 participants. For our analyses we will focus mainly on this community to identify major strands of work and recent trends of this European CSCL interest group. As formally captured data for the bi-partite networks we use the authorship of reports / deliverables within Kaleidoscope: a partner institution is connected to a Kaleidoscope activity if it is one of the contributors of a report. The resulting network of figure 2 is thus similar to a co-citation network. The labels “Dxx” in the boxes represent the work package numbers of the different activities. The small ellipsoids represent some of the teams that are part of Kaleidoscope.

![Figure 2: Bi-partite Team Deliverable Network](image)

Team recommendation and network mapping

When applying the standard operations of transforming the bi-partite network to a one-mode network (Wassermann & Faust, 1994), the collaboration between teams can be identified easily. While this is interesting for analytical purposes, it is not enough for creating substantial recommendations for the future, i.e. for community building activities, because these connections should be obvious for all the actors involved directly.

For the recommendation of non-trivial links in order to promote community building we extend conventional social network analysis with the following approach: Using either a personally created conceptualisation (personal view) of the field or by applying the shared conceptualisation of the community
(ontology, cf. Gruber, 1993) in combination with the bi-partite network, a weighted network can be created for recommendation of links. The algorithmic details of this approach have been presented in Malzahn et al. (2005b). In our example case we applied the knowledge map of our team’s personal view on Kaleidoscope's activities to the bi-partite authoring network to search for partners with similar interest. Because of the nature of the selected data set not all of the links that our team established during the past are part of the data. This is because some teams did not formally submit a deliverable (viz. not being a main author) for every activity they were participating in. So some of the links are missing and we can evaluate our link recommendation method by evaluating if these missing links are found. We created a knowledge map representing our team's view (cf. figure 3) of Kaleidoscope's activity interrelations with the help of the FreeStyler application in combination with CoNaVi.

The team’s perspective (knowledge map) on the Kaleidoscope network was then applied to the bi-partite network shown in figure 2. Our extended SNA approach indicates that there are links of interest between our team and other partners in Kaleidoscope as shown in the resulting one-mode network in figure 4 that have not been present in figure 2.

![Figure 3: Manually created knowledge map of our team’s perspective on the activities](image)

Looking at the set of proposed new partners we clearly realized that the algorithm was working plausibly because with some of the partners we were already in touch. Since the given data was not representing all of our activities in Kaleidoscope the algorithm could not take into account those links from the beginning. This kind of validation of the learned links is quite similar to the commonly used “leave one out” cross-validation approach in machine learning. The really new links motivated us to have a closer look at these Kaleidoscope partners’ work, the type of usage we designed this approach for. Admittedly we discarded some of the links again, but in the end we were stimulated by the algorithm's results to communicate through the artefact by reading their papers - which is in turn valuable for a community like Kaleidoscope because it generates a deeper understanding of the whole field of TEL.
Figure 4: One-mode network with teams linked to our team after ontology facilitation

Trend-spotting via trusted authorities

To demonstrate the method of trend-spotting via trusted authorities described generally in section Trend-spotting we assume that the user regards our team as a trusted authority for promising research topics. If this person analyses Kaleidoscope network with CoNaVi he or she would notice that UDuisburg is currently involved in five activities (see figure 5). These activities are marked with numbers indicating the amount of other teams involved in the corresponding activity. Considering his or her own interests and links towards certain topics the user is now able to investigate one of the topics further. Given that the user is currently not involved in mobile learning activities and discovers that UDuisburg is involved in such activities he might take participation in mobile learning activities into consideration, if he thinks it is a promising direction based on his trust in the supposed authority.

Figure 5: Looking at the scope of interest of a trusted authority

Trend-spotting in Kaleidoscope - a temporal perspective

Since the Kaleidoscope network renews its activity plan each year, a temporal shift of activities which might reflect trends and evolution within the network is an interesting issue for analysis and also for the strategic orientation towards the future. Kaleidoscope is currently in its third year period, thus we can refer to the data from 2 finished work programs and its extrapolation to the third year.
To highlight the evolution of the research network according to new research lines and changes of focus, we show the changes in the activities from one period to its succeeding period. Figure 6 shows the changes from Year 2 to Year 3. The different activity types are represented in different layers. Each arrow represents one transition between the points in time. For the sake of clarity we focussed on just exactly one year-transition to avoid too many arrows. Currently we are working on different filters to enable the user to select appropriate views on specific time ranges and activities of interest.

Figure 6: Evolution of Kaleidoscope from Year 2 to Year 3

Activities that have been continued in the same format from one year to the other are represented as middle-sized cubes; the dynamic evolution of the network can be seen in activities, that have been discontinued at the transition from one period to the next (represented as small cones), and activities that started in the new period, which are represented as large spheres. Especially interesting are the transformations from short-term activities, such as Joint Projects with duration of one year, into activities with a longer-term perspective, mainly Research Teams that are planned to be sustainable for a longer period. These direct transformations are shown as links between activities from one year to another, i.e. from a cone to a sphere. The figure shows that one of Kaleidoscope's major concerns is the evolution into a reliable and sustainable structure reflecting major research areas as well as current trends of Technology Enhanced Learning. This is especially visible in the number of Research Teams that grows from four in year 1 to eight in year 3.

In a complimentary analysis we extracted current trends in the Kaleidoscope CSCL community. First we identified the Kaleidoscope teams with a strong participation ratio with respect to CSCL: We only considered teams with at least 50% of their team members being also members in the CSCL Special Interest Group. This threshold was validated with the computation of the median of the ratios of teams, which resulted also exactly at a ratio of 50%. Then we identified the new Kaleidoscope activities that the “CSCL” teams have participated in since the year 2006 as an indicator for recent projects that might have emerged with CSCL aspects in mind. The activities that resulted from the combination of statistical analysis and the new bi-partite network are:

- **Learning Patterns for the design and deployment of Mathematical Games**, which supports the collaborative design of educational games by collecting re-usable patterns for mathematical games.
• **Integrating Collaborative, Inquiry and Experiential Learning** which brings together the strands of experiential and inquiry learning with collaborative learning, thus creating for the students rich learning experiences with scientific methods in a social context.

• **Computer-based Analysis and Visualization of Collaborative Learning Activities** which aims at the integration of computer-based methods into the analysis process of collaborative activities by means of capturing, processing, and visualization the collaboration.

Indeed all these activities have a CSCL tint, which can be interpreted twofold: on the one hand the CSCL community can be considered an important driving force for the Kaleidoscope network, on the other hand these topics may be potential trends for CSCL research also on a broader scope. Some of these analysis results might have been obvious to or at least assumed by people involved in the Kaleidoscope network, yet the social network approach can be used to assure these assumptions by using the SNA inspired measures that can be easily derived from the given data by our tools. In addition non-evident links from activities can be identified and be used for new collaboration opportunities between the projects.

**Discussion - Privacy Issues and Implications on Visualization**

Experiences from the research project 'VIP-NET - Virtual Work and Learning in Project Based Networks' led us to the conclusion that personal networks can sometimes be critical for analysis purposes. Some people disagree with the idea of opening their personal networks for scientific analysis, since these networks represent their social capital. On the other hand these people are often interested to explore their personal networks to exploit them more thoroughly. Another example where information may be affected by privacy issues is in the case of teaching. In (Harrer et al., 2005) for example we had to anonymize the results for the publication. For teachers it could also be problematic when external people (e.g. social network researchers) see the structural data of classroom activities, since there may be some students who are not as well integrated in the class as others.

These privacy aspects related to personal networks led us to the question how to support users to explore their own networks using social network analysis techniques. Our approach addressing the operationalization of social network analysis for non scientific users is to embed advanced measurements into the visualization. According to (Krempel, 2005) techniques exist to integrate structural properties of networks in the display. For this purpose we created the Weaver application, a 3D visualizer for social networks, which arranges and draws the nodes according to properties such as degree, centrality, or externally defined properties within a simple solution space. From the user view this means, that he or she is able to perceive the properties of a node immediately (e.g. what is the most central topic in the network). The figures 5 and 6 show nodes arranged by their type and shaped by additional information on their life-span.

**Conclusions**

This paper presented three approaches (team recommendation, knowledge map evolution, trend spotting), on how to gain information with SNA-related techniques in multi-mode networks. We showed how our tools can support different use cases of community support: CoNaVi in combination with FreeStyler enables its users to visualize and navigate through (enriched) multi-mode networks as well as evolve given knowledge maps with the help of empirical data. Weaver’s hierarchical views and filters allow for emphasizing temporal developments for trend spotting.

The proposed methods for fostering new personal links rely on the mathematical properties characterizing the structure of the network data. The approach dealing with trend-spotting uses a sociometrically inspired technique to enable the user to reflect on the current state and temporal evolution of the examined network providing a foundation to develop a strategy for self-development and/or re-positioning in the community. A typical use-case of trend-analysis is the support of freelancers. They have a constant need to discover potential trends in technology to be up-to-date for the next contract. Albeit we focused on a scientific community for our analyses because its rich structure provides a good demonstrator and testbed for our approaches, we think that the support is not limited to this kind of community, but can be applied as well for learning communities, especially the ones with computer-based learning support. Almost all universities provide forums accompanying the courses to exchange ideas and provide help for the students having either a problem within the current topics or with organizational issues. New users tend to cross-post their questions in several forums to be sure to find someone who answers to the problem. The proposed method about new personal links can be applied here to direct the novices directly to competent helpers. In Harrer et al. (2005) we combined SNA methods on a student community in a blended learning scenario.
with qualitative and statistical methods. In this earlier work we restricted our approach to the analytical perspective without giving feedback to the students on their network position. In upcoming courses we plan to combine the use of analysis feedback as sketched in Daradoumis et al. (2004) with our techniques of community support by directly giving feedback and recommendation.

References

Acknowledgements
This research project was inspired by our participation in the Kaleidoscope Network of Excellence. (IST No. 507838).
Thanks to Sabrina Ziebarth for her great job in implementing the Weaver application, which enabled us to visualize figure 6.