# Mining and Visualizing Research Networks using the Artefact-Actor-Network Approach



**DOI:**
10.1007/978-1-4471-4054-2_10

**Document status and date:**
Published: 14/06/2012

**Document Version:**
Version created as part of publication process; publisher's layout; not normally made publicly available

# Mining and visualizing Research Networks using the Artefact-Actor-Network approach

Wolfgang Reinhardt[1], Adrian Wilke[1], Matthias Moi[1], Hendrik Drachsler[2], and Peter Sloep[2]

**Abstract** Virtual communities are increasingly relying on technologies and tools of the so-called Web 2.0. In the context of scientific events and topical Research Networks, researchers use Social Media as one main communication channel. This raises the question, how to monitor and analyze such Research Networks. In this chapter we argue that Artefact-Actor-Networks (AANs) serve well for modeling, storing and mining the social interactions around digital learning resources originating from various learning services. In order to deepen the model of AANs and its application to Research Networks, a relevant theoretical background as well as clues for a prototypical reference implementation are provided. This is followed by the analysis of six Research Networks and a detailed inspection of the results. Moreover, selected networks are visualized. Research Networks of the same type show similar descriptive measures while different types are not directly comparable to each other. Further, our analysis shows that narrowness of a Research Network's subject area can be predicted using the connectedness of semantic similarity networks. Finally conclusions are drawn and implications for future research are discussed.

**Key words:** research networks, learning networks, research 2.0, social media, social network analysis, visualization, semantic similarity, community mining

---

University of Paderborn
Department of Computer Science, Computer Science Education Group
Fuerstenallee 11, 33012 Paderborn, Germany
{wolle,wilke,moisun}@mail.upb.de · Open University of the Netherlands
Centre for Learning Sciences and Technologies
6401 DL Heerlen, The Netherlands
{firstname.lastname}@ou.nl.de

## 1 Introduction

With the recent rise of Social Media tools like Twitter and Facebook the Web-based interaction in virtual communities of like-minded people keeps up growing. Lately, Learning Networks and research communities make use of the communication and collaboration features of Social Media platforms. This increases the productivity of the involved participants, enhances mutual awareness, and increases a community's nexus. In the last years we have witnessed the wide application of Social Media to higher education courses and scientific conferences, the discussion about political and environmental phenomena as well as the usage in research communities and enterprises. The analysis of such online activities enables researchers to reveal patterns in communication, detect and visualize cliques of people, or trace the trails of discussions in a community. Most of these analyzes however only reflect the social part of the interactions and thus are able to make claims about the structure of a virtual community but not about the respective digital objects.

In this chapter we present the derivation of the concept of Research Networks (Section 2) and put the concept in the context of Research 2.0 in Section 3. Following we present the model of Artefact-Actor-Networks and its reference implementation in Section 4 that was used for mining different types of Research Networks: conferences, university courses and hashtag communities (Section 5). In addition, we give an insight into data storage with Semantic Web technologies. We explore the artefacts and their relations to online actors in three learning services (Delicious, SlideShare, and Twitter). Besides the descriptive analysis of the communities we apply metrics from Social Network Analysis (SNA) and visualize the networks based on different factors such as semantic similarity (Section 6). From the analysis of the networks we aim at bridging the gap between the use of Social Media tools, as a mean for communication and exchange, and the missing awareness for one's own and activities of others in such settings. Furthermore, we will show the strengths of the Artefact-Actor-Network approach for identifying interesting relations between activities of users, the artefacts they generate and the larger image those activities produce towards pattern recognition in Learning Networks' activities. The chapter closes with the discussion of the results of the Research Network mining in Section 7 and gives an outlook how they could be used in future research towards awareness-support for participants in Research Networks.

## 2 Research Networks and levels of member participation

An online community, e-community, virtual community or online social network is to be understood as a group of people that interact using electronic means of cooperation. Examples of such cooperation media are email, telephone, instant messaging services and more recently Social Software. Lately, online communities have become a valuable and widespread used supplement for groups that work together in face-to-face contexts but they are also existing exclusively in the online world. Online communities may be centered around professional, educational, recreational, political topics; they may be organizational, topical or regional and most often assemble people around specific objects (also see [6; 10]).

Rheingold coined the term virtual communities and claims that they form "*when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace*" [39]. Kim adds that web communities are "*a group of people who share a common interest or purpose; who have the ability to get to know each other better over time. There are two pieces to that definition. That second piece getting to know each other better over time means that there needs to be some mechanism of identity and communication*" [22]. The mere existence of an online community does not mean that there are any strong personal relations between its participants; uncovering the very liberal use of the term community and the term of virtual communities as such [20; 31]. Wellman on the other hand defines community as "*networks of interpersonal ties that provide sociability, support, information, a sense of belonging, and social identity. I do not limit my thinking about community to neighbourhoods and villages. This is good advice for any epoch and especially pertinent for the twenty-first century*" and further elaborates that "*we find community in networks, not groups. Although people often view the world in terms of groups (Freeman 1992), they function in networks. In networked societies: boundaries are permeable, interactions are with diverse others, connections switch between multiple networks, and hierarchies can be flatter and recursive*" [54].

In blended learning, classroom learning is combined with web-based learning that may uses organizational learning management systems (LMS) or more open approaches win which the learners may decide on the tools the want to use. The learner's Personal Learning Environment (PLE, [55]) provides access to all learning resources, useful people and learning services he might need for pursuing his learning goals. Recently, the term *Learning Networks* has been coined for such online communities of learners. According to Koper et al. [23], Learning Networks (LNs) are online communities in which users share existing information and cooperatively create new knowledge. This way, Learning Networks help participants to develop their skills and competences in rather non-formal, unplanned and ad-hoc learning situations and educational contexts. Different from formal education there are little learning goals for the whole Learning Network as well as diffuse, hard-to-phrase individual ones. As Koper points out [24], the participants of a Learning Network could:

- exchange experience and knowledge with each other,
- collaborate on common research questions and tasks,

- offer and get support to/from other participants in the Learning Network (e.g. questions, answers, remarks),
- set up focussed working groups,
- support each other when encountering learning problems, and
- use tools and services to create, share, find and access learning resources.

Each Learning Network – being a social network – is composed of people that share a similar interest or follow a similar goal. The commitment to the common interest or goal, the timeframe of the Learning Network's existence, the size of the networks and other properties vary between Learning Networks but for all that Learning Networks are providing their participants with resources, services and agents to support their learning purposes. The *participants* in Learning Networks have clearly defined or rather blurred learning goals; they could be help seekers as well as mentors, coaches, teachers or lurking bystanders. The *resources* in a Learning Network are all digital artefacts that might help the participants to accomplish their learning goal or that make them aware of a lack of personal competence that they want to eliminate. Learning resources may include any audio or video file, blog post, wiki page or learning resources as well as entire courses in LMS. Part of those resources were already existing before the nascency of the Learning Network, others are created by the participants and all of those resources can be used by several LNs at a time. Sloep elaborates that *learning services* are software tools that increase a Learning Network's viability [44]. Koper adds that such web services are designed to facilitate the participants to exchange experience and knowledge, to stimulate active participation in the Learning Network, to assess and develop the participants' competences, to find relevant peers and experts that could offer support in solving a certain problem, and to facilitate ubiquitous learning [24]. According to Koper examples of Learning Networks are [23]:

- Communities of teachers who exchange experience on how to handle certain pedagogical issues in the classroom.
- Employees of a company that need to update themselves about the functions of a new product their company released recently.
- Students who cooperatively write a composition on a given topic.
- Lawyers who exchange experience and knowledge when a new law is introduced within their field.
- Researchers that exchange information to find solutions for a specific problem. They update each other with new findings and cooperatively solve problems, co-author documents, attend face-to-face events and carry out joint projects in a geographically and timely separated manner.

As a matter of course, there exist a range of other Learning Networks with different participants, resources and services. If the participants in a Learning Network are scholars, the resources used and services in place are related to their research activities or the execution of research projects we call such Learning Networks *Research Learning Networks* or briefly Research Networks (RNs). It is common to all of those Learning Networks that we find differing levels of member participation.

## 2.1 Levels of interaction

As Kim [22] elaborates, we find differences in the interactions in Research Networks that make use of structured means of communication (such as bulletin boards, mailing lists or chat rooms) and such Research Networks where interactions are mediated through bottom-up, individual-centered tools (e.g. blogs, microblogs or social networking sites). In almost all Research Networks, there are patterns of social interaction and user contributions. It does not matter if the participation in the Research Network includes posing questions and answering some, tagging resources or creating own learning resources, creating discussion threads or linking online learning repositories; it is a rule-of-thumb that only 1% of the participants create new content, 10% interact with this content and 89% will just consume the content that is there [2]. This inequality pattern is even worse within Wikipedia, the most well-known Research Network where participants jointly create a high-quality online encyclopedia. In September 2010, the English version of Wikipedia had 35,222 active users[1] which is only 0.027% of the 130 million unique visitors it has worldwide (it is 0.07% of the 45 million unique visitors it has in the United States alone) [14; 56].

This unbalanced participation patterns can be found in most Research Networks and Social Networking Services[**?** ] (SNS) and can be explained by technical and motivational reasons. If there are technical hurdles that hinder the learner to participate in the Research Network's activities or if the participants sense a lack of compensation for their work, the participation in the Research Network will probably not set up. As the reasons for a learner's participation is both varying and not singular, Research Networks should incentivize participants with multiple types of motivation in order to engage them and keep them engaged. Lately Wikipedia undermined the sovereignty of its users and demotivated some of them with the ongoing controversy around Deletionism versus Inclusionism [7] and the force of producing higher quality articles with a range of external references. This, together with increasing administrative processes needed to edit articles, resulted in a decline of active users in the Learning Network by 12,3% (11,170 users) between January and September 2010. Many Wikipedians lost their feeling of belonging to a community of equivalents, thus trashing their identity in the Learning Network.

[1] A Wikipedian is counted as being active, if he contributed to Wikipedia articles at least 5 times in a month.

Table 1: Levels of participation in Research Networks (based on [18])

| Participation | Status | Lifecycle |
|---|---|---|
| Peripheral | Visitor | The participant is an outsider and has little or no structured participation in the Research Network (he is lurking). |
| Inbound | Novice | The participant is a introduced as newcomer to the Research Network and heading towards full participation in the Network's activities. |
| Insider | Regular | The participant is a fully committed inhabitant of the Learning Network. |
| Boundary | Leader | The participant is a leader in the Research Network sustaining his membership with active participation and the brokering of information and interactions. |
| Peripheral | Elder | The participant is about to leave the Research Network because of new goals, extended relationships to new Research Networks and new positions. |

Another explanation approach for those participation patterns comes from a more sociological point of view. Kim suggested that there is a membership life cycle [22] and Lave and Wenger presented the model of Legitimate Peripheral Participation (LPP) [26], both claiming that there is a participation life cycle for participants in communities such as Research Networks. Table 1 synthesizes the ideas whereupon participants start their life in a Research Network as a visitor or lurker that are only watching interactions and consuming existing content but are not directly adding new content. At some point learners start participating in the Learning Network's activities and become novices. After having contributed to the RN with both active social and content interaction, the learner becomes a regular participant. If a learner further engages in the RN's activities he might become a leader that sustains his membership through multifaceted activities. After being in a Research Network for some time, a participants might become an elder that is about to leave the network because of new learning goals or matured knowledge in the domain. It needs to be pointed out, that a learner can always be part of many Research Networks at a time; so while he is a leader in one, he might be visitor in another one and regular participant in a third Research Network. At each time and in any Research Network, participants on a lower level of participation must feel engaged and motivated by the fellow participants and be technically empowered to 'graduate' to a higher level.

## 3 Research Networks and Research 2.0

Lately, Research Networks are increasingly dependent on Web 2.0 tools, technologies and techniques to their daily practices. In Technology Enhanced Learning (TEL), the adoption of Web 2.0 is already actively researched under such notations as Learning 2.0 [34], Personal Learning Environments [55], Open Learning Environments [16; 28] or Learning Networks [25]. The application of Web 2.0 to Research Networks is often squired with the terms Research 2.0 or Science 2.0 and aims at leveraging the same opportunities for research. Research 2.0 is a rather young concept but there are already numerous controversial positions, oscillating between new tools and technologies, methods and practices (cf. [47]). Waldrop [52] for example, relates Science 2.0 to "*new practices of scientists who post raw experimental results, nascent theories, claims of discovery and draft papers on the Web for other to see and comment on*" and Shneiderman [43] comprehends the term as "*new technologies [that] continue to reorder whole disciplines [... as ...] increased collaboration [is stimulated] through these socio-technical systems*".

Focusing on the change of practices mentioned in Waldrop's definition, Kieslinger and Lindstaedt [21] are underlining the Science 2.0 focus on "*improving, enhancing [and] speeding up feedback cycles*". Underwood et al. [48] postulate even further that Research 2.0 offers more potential that the mere optimization of science efficiency: participation in research can be broadened beyond existing scientific communities. Research 2.0 as "*technology enhanced participatory science*" [48] could then unbolt science allowing 'everyday scientists' [40] to participate globally and pervasively in research and collaboration. Butler [5] sees a key feature of Research 2.0 in "*dynamic interactions between [scholars] in real time*" at the same time criticizing the slow adoption of these new technologies and practices in the scholarly daily routines. Waldrop also claims that Science 2.0 allows for a richer dialogue in Research Networks such as collaborative brainstorming, meta conversations, or an open discourse of "*critiquing, suggesting, sharing of ideas and data*" among previously unknown parties [52]. Ullmann et al. point out that this way, "*Science 2.0 is supposed to enable efficient crowd-sourcing of ideas and refinement of knowledge through open debate.*" [47]. As Nielsen remarks, the scholarly system has hardly changed since the creation of the first scientific journal in the 17th century. With the Internet, WWW and Research 2.0 becoming mainstream, science will "*change more over the next 20 years than in the past 300 years*" [27]. He goes on and elaborates that Research 2.0 is the "*first major opportunity to improve this collective long-term memory [the scientific journal system], and to create a collective short-term working memory, a conversational commons for the rapid collaborative development of ideas*".

There is some controversy about whether Research Networks are driven by new technologies or new practices and the reciprocal relationship between those two aspects. Where understanding new practices allows for their implementation into tools, the existence of new tools reshape existing practices and often allow for the appropriation of new practices not foreseen in their design. Finally, Shneiderman controversially asserts Science 2.0 a change in research methodology that would be complementary to the Science 1.0 focus on predictive models and laboratory controlled testing (see also Gillet's elaborations on the transition from Science 1.0 to Science 2.0 in [13]). Research 2.0 would therefore take place embedded in the real world through large-scale, rigorous observations and their validity would be empirically investigated using qualitative and quantitative analysis. Objectors of this understanding point out that many scientific fields including social sciences or natural sciences already rely on this scientific methodology. In spite of that, there seems to be an agreement amongst scholars that Research 2.0 and its new socio-technical systems are more cooperative, more efficient, productive and open, are fostering engagement and focussing on the sharing of new ideas.

Despite the many undoubted advantages of Research 2.0, many authors mention a reluctant adoption of the new learning services by researchers. In some disciplines the revolution of Research 2.0 is even passing by without researchers noticing the changes [5]. A recent study conducted by Procter et al. [33] with 1,477 UK researchers reveals that the adoption of Research 2.0 in scholarly communications "*has reached only modest levels so far*" whereas there are certain learning services that have been rapidly adopted. Especially in the context of scientific events and higher education courses, services like Twitter and SlideShare have proven to be heavily used to share messages and learning resources with a wider public [17; 35]. Duval even says that "*In fact, Twitter is more relevant to me now than any [other] research2.0 application*" [8]. Another category of learning service that is widely adopted within the scholarly practice are social bookmarking systems such as Delicious or Diigo [53]. What is common to those learning services is the fact that they are built around clearly defined digital social objects [10] and not intended for the usage in the scholarly system in the first place. Instead, researchers adopted and reshaped the learning services in order to make them better suited to the scholarly routines of work.

Summing up, it can be stated that participants in Research Networks use learning services in varying intensity with the goal to open up the previously closed world or research. They share ideas and learning resources with each other and cooperatively create new knowledge that becomes part of the collective memory. Not all learning services are used equally and not all researchers use all existing learning services. In fact, we even observe different usage of different services with one person, meaning that they differentially behave in different learning services. In order to mine Research Networks and the respective learning services we therefore should differentiate between the different handles of a person in the learning services, allowing for the separated inspection of a user's behavior. Also, we should be able to distinguish between the single learning services like Twitter or Delicious within a Research Network in order to recognize pattern that might exist in one service but not within the other. This will also allow us to compare the usage of learning services in different Research Networks.

The overall goals of Research Network mining are thus: expert finding and recommendation, learning resource clustering and recommendation, pattern recognition within and across Research Networks, community detection within Research Networks, awareness raising about a network's behavior and structure, and the analysis of a participant's research network trajectory. In the following section we introduce the approach of Artefact-Actor-Networks (AANs) to support these Research Network mining goals.

# 4 The AAN approach for Research Network mining

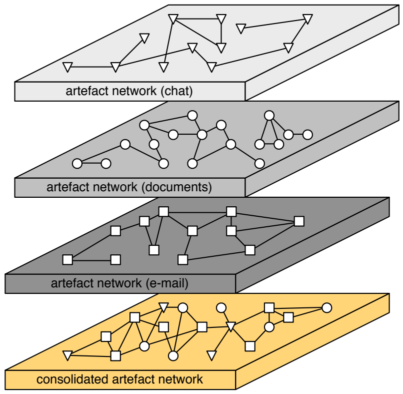
Artefact-Actor-Networks are an approach for mining resources of various kinds of source networks. It comprises two main parts: The theoretical foundation and a reference implementation. In the theoretical part, a concept for a consolidation of social networks and artifact networks of documents is explained. Resources of mined networks are stored by a distinction between artefacts, actors, and keywords. The practical implementation of this concept was put into practice using Semantic Web technologies. Section 4.1 introduces the fundamentals for a system for finding experts and communities, retrieving information, analyzing and visualizing Research Networks. The reference implementation of this system is introduced in Section 4.2.

## *4.1 Theoretical approach*

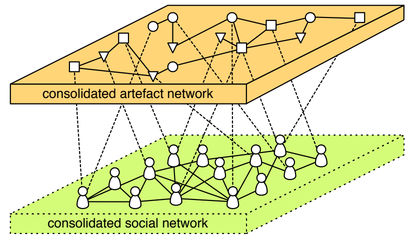
Artefact-Actor-Networks (AANs) were first introduced by Reinhardt, Moi and Varlemann in 2009 [36] and serve as an approach to semantically intertwine social networks with so-called artefact networks. We distinguish two general types of layers - the artefact and actor layers. Both types can have arbitrary sub-layers to specialize the type of an artefact or actor. This can be understood like the hierarchy concept of higher level programming languages. Furthermore artefacts and actors can be connected trough typed relations, so called semantic relations to manifest the semantic context. Examples for semantic relations are *isCoWorker* to connect actors, *references* to connect artefacts and *isAuthor* to connect artefacts with actors.

### 4.1.1 Layer in Artefact-Actor-Networks

Using Artefact-Actor-Networks an actor's participation in the life cycle of artefacts as well as significant connections to other actors will be outlined. Artefact-Actor-Networks consolidate multilayered social networks and artefact networks in an integrated network. Therefore, we consider the communication and collaboration with each learning service or artefact supply (e.g. Twitter, chats, email or scientific documents) as a single layer of the respective network. We unite these single layers in both social and artefact networks to consolidated networks that contain all actors and artefacts respectively (cf. Figure 1a). While in the consolidated social network we can only make statements concerning the relations between actors and in the consolidated artefact network we can only analyze the relations between artefacts, Artefact-Actor-Networks (cf. Figure 1b) also contain semantic relations between actors and artefacts. The recently discussed semantic relations can be found in each layer respectively between each layers.




(a) Consolidated artefact network resulting from three layers




(b) Artefact-Actor-Network with semantic relations between artefacts and actors

Fig. 1: Schematic assembling of an AAN

#### 4.1.2 Use of ontologies

As introduced, we distinguish different types of layers in AANs. To model an Artefact-Actor-Network with it's layers we use a ontologies to specify semantical and hierarchical relations. Using current techniques like OWL [50] and RDF(S) [51] the inheritance of classes and relations can be accomplished. Every class represent an special type of artefact and actor, which are the base classes. By following this approach, querying specialized information becomes possible and allows to change between different abstraction levels. On the base level there are only artefacts and actors without further specialization. If we were interested in an aggregated analysis of all artefacts or actors, we would simply query the base class whereas querying specific classes allows for more focused analyses. Figure 2 depicts the ontologies used in the AAN reference implementation.

AANBase and Co.

All our ontologies inherit from the *AANBase* ontology. It holds the base classes *artefact*, *actor* and *keyword*, which are the most general classes in any Artefact-Actor-Network. An artefact can have arbitrary many keywords. Each keyword can be specialized as a category or tag class.

Figure 2 also shows the *AANOnline* ontology, which describes artefacts and actors of the WWW. 'Between' the *AANTwitter* and *AANOnline* ontology there is the *AANMicroblog* ontology which abstracts from the various microblog services and allows to extent the whole ontology in the future. The

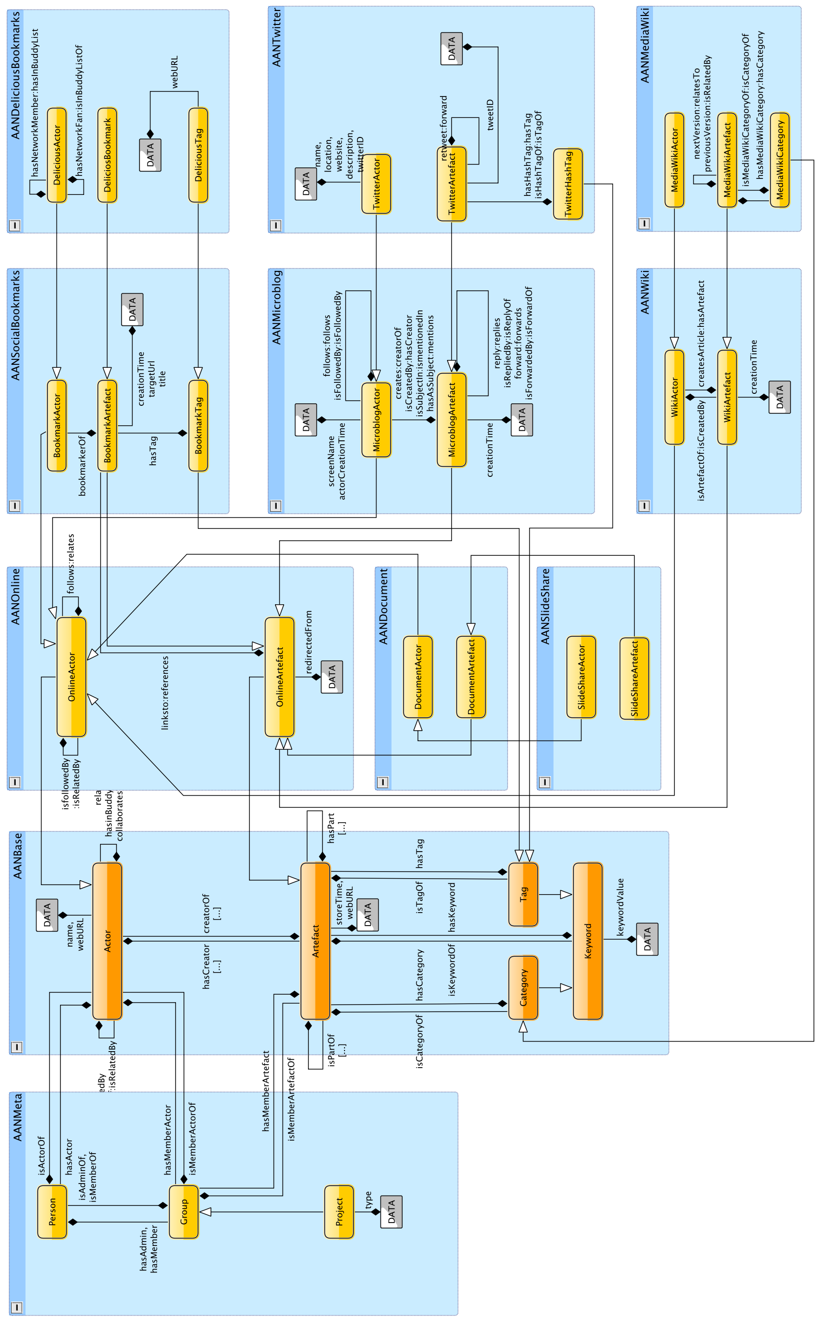


Fig. 2: Simplified overview of the ontologies available in the AAN reference implementation

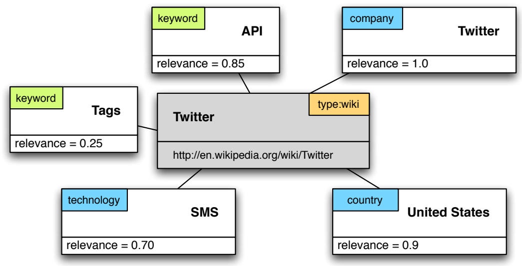



Fig. 3: Relevant keywords and named entities for a wiki artefact about Twitter

same holds true for the other most specialized ontologies like *AANMediaWiki* or *AANDeliciousBookmarks*.

Using all ontologies in place, the layers of the Artefact-Actor-Network can be described and distinguished. The *AANBase* ontology represents the consolidated artefact and actor layer. More special layers are *AANMicroblog* or *AANWiki*. The most specialized layers *AANSlideShare*, *AANMediaWiki*, *AANTwitter* and *AANDeliciousBookmarks* can be inferenced to get a more aggregated view of the network.

#### 4.1.3 Analysis of artefacts

The semantic relations between artefacts and between actors can most often be extracted automatically, like references or citations without considering the content of the artefacts. In order to extract differentiated information about domain experts or the like, not explicitly existing relations between actors in the same domain need to be extracted from the content of their artefacts. If two artefacts are semantically similar, then there is also a more or less strong relation between two concerned actors.

To determine the semantic similarity of two artefacts, we need metadata of the objects. There are numerous ways of obtaining metadata for artefacts. We will not cover all these possibilities. Amongst others, the metadata contains semantically relevant information such as keywords or named entities. Semantic metadata can be extracted through external libraries and services like OpenCalais [38] or AlchemyAPI [29]. Picture 3 shows exemplary keywords and named entities (*technology*, *country*, *company*) for a wiki artefact about the Twitter micro-blogging service.

We have to calculate the relevance for every extracted keyword and named entity, which describes the semantical relevance of the metadata for describing the artefact. Several techniques of information retrieval and natural language processing can be used for the calculation of this relevance. One of these techniques is the inverse document frequency (tf-idf) [41; 42], to determine how good a keyword separates an artefact from all other artefacts. Tf-idf uses the fact that if the keyword has a large frequency in the whole set of keywords, it has only small relevance to describe an artefact. Processing of the relevance has to be done in continuous intervals, caused by the fact that tf-idf is based on the existing keyword corpus from the set of artefacts and thus has to be re-calculated as soon as new artefacts are stored.

Two artefacts are semantically similar, if the semantic metadata of the artefacts are similar. To determine the semantic similarity, we compare the relevance of the metadata of two artefacts. We distinguish metadata of artefacts in different concepts like *keywords* or *named entities*. Examples for *named entities* are *companies*, *technologies* or *persons*. Every artefact may have several concepts. An artefact interprets its referenced concepts as attributes. By using RDF to represent artefacts, we have no redundantly stored concepts. A concept may be referenced by many artefacts in the network. To compute the similarity between two artefacts, there must exist at least one equal concept between them. Otherwise the semantic similarity is zero. For a better understanding of our concept we divide the process to calculate the semantic similarity into short steps.

### Relevance of concepts for an artefact

As discussed previously an artefact may have arbitrary many concepts with specified relevances. Services like OpenCalais [38] and AlchemyAPI [29] deliver information about keywords and named entities with their respective relevance for the artefact. Directly extracted keywords can be weighted through information retrieval methods like tf-idf.

### Normalizing of relevances

The relevance of the attributes are absolute values with no respect to other attributes. But to compute the semantic similarity between two artefacts it is necessary to normalize the values to get the weight of one relevance in respect to all others. In our approach we normalize the attributes to the value 1. Denote that all relevance factors are mapped into the continuous interval $(0, 1]$. The sum of all relevances is at most one.

### Computation of semantic similarity

To compute the similarity between two artefacts, we take into account all common attributes of the artefacts. Pairwise, the difference between the normalized values is calculated and weighted by the minimum of the normalized values of both attributes. Then all pairs will be summed up. The resulting value is the similarity of both artefacts in respect to the weight of their attributes. Hereafter, we present some definitions which are necessary to calculate the semantic similarity in AANs. Let $A$ be an artefact, then $C_A$ denotes the set over all concepts of the artefact $A$ which relevance $r_A(c)$ is greater 0. Let $A$ and $B$ be artefacts, then $C_{A,B} = C_A \cap C_B$ denotes the set of the common concepts.

Let $A$ be an artefact and $c \in C_A$ a concept of this artefact, then $r_A(c)$ denotes the relevance of the concept $c$ referred to artefact $A$. The normalized relevance of the concept $c$ referred to artefact $A$ will be calculated as follows

$$n_A(c) = \frac{r_A(c)}{\sum_{i \in C_A} r_A(i)} \tag{1}$$

To calculate the semantic similarity between two artefacts $A$ and $B$ we iterate over all common concepts $C_{A,B}$. At every iteration step the semantic similarity in respect to the current concept is calculated and summed up. The semantic similarity between two artefacts is then given by

$$SemSim_{A,B} = \sum_{c \in C_{A,B}} \Big( \min \big( n_A(c), n_B(c) \big) \cdot ConSim_{A,B}(c)^2 \Big) \tag{2}$$

with

$$ConSim_{A,B}(c) = 1 - |n_A(c) - n_B(c)| \tag{3}$$

The function $ConSim_{A,B}(c)$ (ConceptSimilarity) calculates the semantic similarity between two artefacts $A$ and $B$ in respect to concept $c$ by subtracting the absolute value of the difference of the relevances from 1. The greater the difference of the relevances of a common concept the lesser the semantic similarity by the concept $c$.

$SemSim_{A,B}$ is a linear function (cf. Figure 4). For a common concept between two artefacts $A$ and $B$ the relevancies are on the x- and y-axis. The value of the semantic similarity is represented by the z-axis. If the relevance $x$ equals $y$, the semantic similarity is maximal for a given concept.

For example a common concept *SMS* which is a technology, the relevance of this concept must not necessarily be equal to both artefacts. If the relevancies are same, then $ConSim$ returns 1, which means that the semantic similarity value will not be weakened, because the current concept is identically important to both artefacts. The minimum of the normalized relevancies in the first part of the formula guarantees, that the semantic similarity value in every iteration is not greater then the smallest relevance. If two artefacts have the same concepts and for every concept equal relevance, then it must be that the semantic similarity is exactly 1. Differences on relevancies for common concepts affect alleviative to the semantic similarity between two artefacts. In an evaluation process we decided to square $ConSim$ which means that a small difference of the relevance will affect less alleviative.

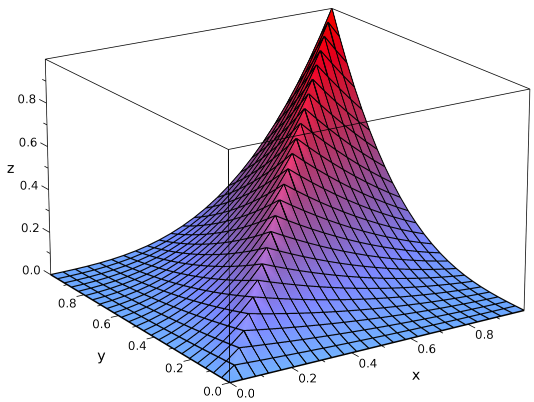



Fig. 4: Plot of the *SemSim* formula

### *4.2 AAN reference implementation*

The different requirements of the AAN concept makes various demands on an implementation. With regards to a pool of possible data sources, the storage of semantic relations, and various goals of analysis by different components, there is a need for a dynamic system. Such a dynamic system can be designed with the OSGi Service Platform [30], which is a specification to develop modularized architectures with the Java programming language. With this basis, modularized components, called bundles, can be defined. A bundle consists of executable code and additional resources. Its functionality is offered by services, which are defined by interfaces. In this way a service can be provided by different bundles. This means, that a specific task can be executed by several bundles, e.g. the analysis of a resource can be done by different specialized components. Another advantage of the OSGi Service Platform is the dynamic treatment at runtime. Bundles can be in different states, they can be installed, started, and stopped at runtime. Thus it is possible to add a more recent version of a bundle without restarting the system.

The architecture of the reference implementation is divided into three main blocks of bundles, in which tasks of the fields data acquisition, data storage, and analysis are performed. Figure 5 shows the main parts of bundles and interfaces for the data flow. A more detailed insight is given in [37].

Bundles in the crawling block are responsible for the data acquisition. This block comprises three main types of bundles: CrawlerManager, Crawler, and Parser. The purpose of CrawlerManagers is to define tasks, by which resources of given URIs are processed. The first URI and additional parameters of an overall job can be given by users. This is why CrawlerManagers are accessible by web services. Beside an URI, a user can define when a job is started, if a job has to be repeated after some time, and how deep a network has

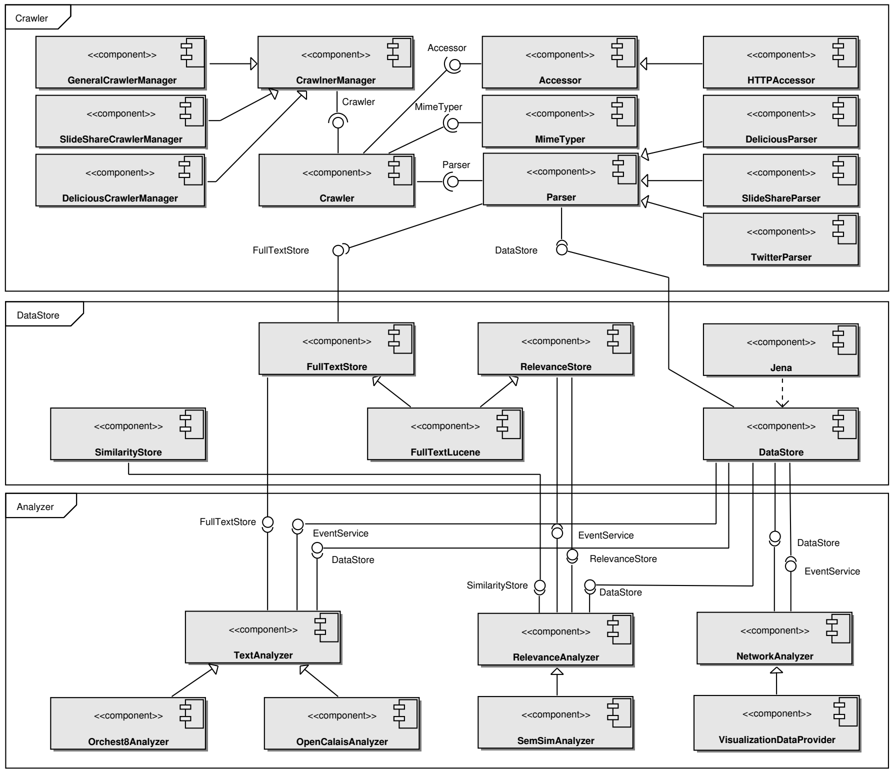



Fig. 5: Architecture of the AAN reference implementation

to be accessed. General jobs, like of simple websites, can be executed by the GeneralCrawlerManager. If a network requires special handling, there is the possibility to define adapted CrawlerManagers, e.g. the DeliciousCrawlerManager, which was implemented to fit the requirements of Delicious feeds. With this component, it is possible to define, if resources of an actor or a keyword is of interest. After an overall crawling job is started, the Crawler component works through the job by a working chain, consisting of Accessor, MimeTyper, and Parser. The Accessor component accesses the resource given by an URI and temporarily stores the data of the URI locally. In the next step, the MimeTyper component determines the MIME type of the resource. Finally, a fitting Parser component is extracting relevant data. A suitable parser is chosen by the MIME type and the URI of a resource. There are two types of parsers, general and special parsers. General parsers are built up to handle resources like conventional websites. A general parser extracts hyperlinks and sub-documents like images and adds additional tasks for the extracted links. A specific parser is specialized to extract characteristic metadata of different interfaces, e.g. structured XML contents. Extracted data is hand over to bundles of the DataStore block. After a parser finishes work on a resource, a task is completed.

There are three types of data, which are stored by bundles of the DataStore block: full texts, semantic data, and data describing the similarity of resources. The first two types are hand over by parsers. Full texts can be the main content of a website, a presentation, a microblog entry or other artifacts. The FullTextStore is realized using Apache Lucene, a Java-based indexing and search implementation. With this component, URIs and their related full texts can be stored and loaded again. Stored full text can be used for further analysis, e.g. clustering or keyword extraction to classify artifacts. Another exciting matter is the storage of semantic data. Objects, their semantic relations, and additional metadata can be stored by the DataStore bundle. The Semantic Web framework Jena [19] serves as the basis of this component. It provides a model, whose schema is specified by the AAN ontology. By the use of this model, statements in the form of triples can be stored (e.g. an actor is the creatorOf an artifact). Further, it is possible to make queries, written in the RDF query language SPARQL [49]. This offers various opportunities for defining specialized requests. For example, some metadata properties for all artifacts of a special class, reduced to a set of related authors and keywords can be queried. Such requests can be used for selecting data of interest and to analyze the data or visualize relations between requested objects. The third type of stored data is used indirectly by the analyzing component SemSim to store data about semantic similarity between artifacts.

Based on the harvested data, there are various opportunities for analysis. The first two developed components in the Analyzer block are bundles of the type TextAnalyzer, with which relevant entities (like describing keywords) for representing a full text can be extracted. The developed bundles are listening on upcoming events, which are fired in the DataStore block. These events occur every time a new resource is stored. The bundles OpenCalaisAnalyzer and Orchestr8Analyzer are based on the web services of OpenCalais [38] and the Orchestr8 AlchemiAPI [29]. Their functionality, efficiency and accuracy are described in detail in [4]. The returned metadata is stored within the semantic model and provides describing data for further analysis. Another bundle of the type RelevanceAnalyzer is the SemSimAnalyzer, which translates an approach of computing semantic similarity into practice. Like presented in [36], the SemSimAnalyzer computes the similarity of artifacts in pairs. Therefore all common attributes of two artifacts are taken into account pairwise. By the minimum of the normalized values of two attributes, the difference between the normalized values is calculated and weighted. The resulting values form the basis of the semantic similarity presented in this chapter. Finally, the harvested and calculated data has to be extracted and transformed for further analysis. This is done by the VisualizationDataProvider. With this bundle it is possible to extract subnetworks in form of data describing graphs. Subgraphs consist of edges and nodes. Nodes represent resources, like artifacts or actors, of the different network sources. Edges either represent semantic similarity or relations describing the relationships in the networks themselves. By web services, artifact and actor classes of interest, which have to be defined in the ontology, can be requested. Further, context keywords (e.g. a keyword or tag) can be given. Resources, described by these context keywords, are extracted and exported in the Graph Exchange XML Format (GEXF). Such files can be opened, analyzed, and visualized by software like Gephi [12].

## 5 Research Networks analyzed in this article

In this chapter we analyzed six different Research Networks using the Artefact-Actor-Networks approach as described in the previous sections. The Research Networks were chosen because of their widespread adoption of Research 2.0 services and established practices within the particular communities. In our exploration we focused on the analysis of three types of learning services: 1) Twitter[2], 2) SlideShare[3] and 3) Delicious[4]. As described in [33; 46; 53] those services are especially good adopted by researchers for supporting scholarly communications. In our analysis, we incorporated three types of Research Networks: a) such networks that are formed around a scientific event like conferences or workshops, b) such networks that arise in the context of higher education courses and c) networks that accrue from the usage of a common tag[5]. In detail we analyzed four scientific conferences, one university seminar and one hashtag community. Those Research Networks differ in context, size, structure, voluntariness of participation and their age. Table 2 presents an aggregated overview about the data we used for our analysis (labeled with 'analyzed'). The table shows that there are differences in the number of analyzable and analyzed data. For the case of Delicious we compared the number of bookmarks in the system ('Web') with those that are accessible via publicly available interfaces. The reasons for the partially significant differences are to be explained with restrictions in the Delicious API limiting the number of bookmarks you can access. For the learning service Twitter we compared the number of tweets that were accessible directly via Twitter's search interface ('Web'), a third-party Twitter archive called TwapperKeeper[6] ('TwK') and those crawled using the AAN reference imple-

[2] `http://www.twitter.com`

[3] `http://www.slideshare.net/`

[4] `http://www.delicious.com/`

[5] Such Learning Networks are also known as *hashtag communities* as they spring up around the accidental or planned usage of a tag, meaning a keyword or term associated with a piece of digital information. Some authors use the term *Communities of Interest* for describing such virtual communities with shared problems and goals [11].

[6] In TwapperKeeper (`http://www.twapperkeeper.com/`) someone has to manually create an archive for a hashtag. The software then stores all the tweets associated with that hash-

Table 2: Research Networks investigated in this article (data as of 22.09.2010)

| | Type | Tag(s) | Delicious | | Twitter | | SlideShare | |
|---|---|---|---|---|---|---|---|---|
| PLE 2010 | conference | ple_bcn | Web | 196 | Web | 1 | Web | 20 |
| | | plebcn | | | TwK | 6,772 | | |
| | | | analyzed | 181 | analyzed | 6,542 | analyzed | 0 |
| ALT-C 2010 | conference | altc2010 | Web | 345 | Web | 1 | Web | 5 |
| | | | | | TwK | 6,723 | | |
| | | | analyzed | 245 | analyzed | 6,679 | analyzed | 5 |
| FSLN 2010 | seminar | fsln10 | Web | 384 | Web | 0 | Web | 17 |
| | | | | | TwK | 768 | | |
| | | | analyzed | 383 | analyzed | 689 | analyzed | 17 |
| SOLO 2010 | conference | solo2010 | Web | 124 | Web | 28 | Web | 4 |
| | | solo10 | | | TwK | 4,925 | | |
| | | | analyzed | 118 | analyzed | 4,635 | analyzed | 4 |
| PLE | hashtag community | ple | Web | 22,599 | Web | 76 | Web | 595 |
| | | | | | TwK | 71,761 | | |
| | | | analyzed | 2,314 | analyzed | 2,908 | analyzed | 595 |
| ED-MEDIA | conference | edmedia | Web | 190 | Web | 0 | Web | 14 |
| | | | | | TwK | 2,120 | | |
| | | | analyzed | 128 | analyzed | 1,993 | analyzed | 14 |

Due to a problem accessing SlideShare contents, the artefacts for PLE 2010 were not analyzed in this chapter.
Web = artefacts available on the websites, TwK = TwapperKeeper

mentation. Finally we show the number of artefacts available in the learning service SlideShare for each of the Research Networks.

Following, we briefly introduce the selected Research Networks and name the hashtags that were used by the participants of the network in order to identify their output as belonging to the Research Network.

The selected conferences were chosen because they dealt with topical themes in the context of Research 2.0 and personalized learning. The conferences attracted many well-known researchers and provided a broad range of social networking opportunities. Moreover, the participants of the conferences were affine with the usage of various learning services in scientific events.

In detail we analyzed the 1st PLE Conference 2010, the 17th international conference of the Association for Learning Technology (ALT-C), the 2010 Science Online London conference, and the ED-MEDIA conferences 2009 and 2010.

The 1st PLE Conference (used hashtags were #ple_bcn and #plebcn) took place July 7-9 2010 in Barcelona, Spain and was intended "*to produce a space for researchers and practitioners to exchange ideas, experience and research around the development and implementation of Personal Learning Environments (PLEs) including the design of environments, sociological and educational issues and their effectiveness and desirability as (informal) learn-*

tag and makes them accessible via an open interface. As of 20.03.2011 the API capabilities have been removed from TwapperKeeper. The same functionality can be achieved with an Open Source version of TwapperKeeper (yourTwapperKeeper) that can be installed on a local web server.

*ing spaces*" [32]. The conference provided opportunities for unconferencing events [15] and was squired by a rich range of Social Media offers such as a YouTube channel[7], a Twitter account for the conference and a dedicated Crowdvine[8] site with 116 registered participants.

The 17th international conference of the Association for Learning Technology (ALT-C) was held in Nottingham, UK from September 7-9 2010. The participants of this Research Network used the hashtag altc2010 and were also supported with a Crowdvine site to extend social interaction amongst the more than 400 registered participants. ALT-C 2010 was targeted towards "*practitioners, researchers and policy-makers from all sectors to explore, reflect, and learn*" [1].

The 2010 Science Online London (hashtags were solo2010 and solo10) conference took place September 3-4, 2010 in London, UK and was amongst others hosted by the popular reference management maker and scientific social network provider Mendeley[9]. The organizers of the conference were asking "*How is the web changing the way we conduct, communicate, share, and evaluate research? How can we employ these trends for the greater good?*" and answered "*This September, a brilliant group of scientists, bloggers, web entrepreneurs, and publishers will be meeting for two days to address these very questions.*" [45]. The event was promoted and transacted using the social event management software Eventbrite[10], accompanied with a dedicated Twitter account and pictures on Flickr[11].

The Research Network that uses the hashtag *edmedia* is made up of participants of the ED-MEDIA conference series, run by the Association for the Advancement of Computing in Education. "*This annual conference serves as a multi-disciplinary forum for the discussion and exchange of information on the research, development, and applications on all topics related to multimedia, hypermedia and telecommunications/distance education*" [3]. In particular we investigated learning resources that were published in the context of the 2009 and 2010 conferences. ED-MEDIA attracts participants from all over the world and encourages online interactions with the providence of a group blog, a dedicated Twitter account, a conference group on Ning[12], and a Flickr account.

Besides the four scientific conferences, we also analyzed an interdisciplinary seminar that took place in two geographically separated German Universities. The educational design of the seminar entitled *Future Social Learning Networks* demanded to cooperate in teams of two using mainly Social Media as mean for sharing and communication. The usage of Twitter and Delicious

[7] `http://www.youtube.com/`

[8] `http://www.crowdvine.com/`

[9] `http://www.mendeley.com/`

[10] `http://www.eventbrite.com`

[11] `http://www.flickr.com/`

[12] `http://www.ning.com/`

was mandatory for all participants in the seminar, whereat the students could additionally use any other Social Media services that would support them in achieving their learning goals [17].

Finally we analyzed the hashtag community that formed around the usage of the tag PLE. We chose this tag as it is the acronym for a term widely discussed in the domain of technology enhanced learning: Personal Learning Environments [9; 55].

## 6 Mining of selected Research Networks

In this section we describe the process of mining the selected Research Networks. This comprises the description of our hypotheses, the analysis procedure and the data-mining of artefact- and actor-level data. Finally, the most important results of the analysis of the mined data is presented. During the analysis of the Research Networks we use descriptive measures about the structure of Artefact-Actor-Networks and the networks that stem from semantic similarity between artefacts and actors. Those measures are defined as follows:

*Bookmark ratio* The bookmark ratio describes the quantitative relation how often a Web resource has been bookmarked in the learning service Delicious from different participants of a Research Network.

*Artefact/actor ratio* The artefact/actor ratio describes the quantitative relation how many artefacts an actor has a relation to.

*Density* The density of a network measures how close the network is to complete. A complete network has all possible edges and the density equals to 1.

Connectedness The connectedness denoted the average degree of an artefact or actor in the respective network and thus measures the number of relations to other artefacts or actors.

### *6.1 Hypotheses and procedure*

Our analysis of Research Networks using Artefact-Actor-Network theory was lead by the following hypotheses:

H1: The analysis of all Research Networks will show similar results based on the descriptive metrics.

H2: The analysis of all Research Networks of the same type (e.g. conferences) will show similar results based on the descriptive metrics.

H3: The hashtag community will have the lowest density of all Research Networks on the artefact level.

H4: The narrower the subject of a Research Network, the higher the semantic similarity of the associated artefacts will be.
H5: The similarity of artefacts and actors of a Research Network is independent of a vivid social interaction within the Research Network.

In order to test the above hypotheses, we obtained the data using the AAN reference implementation as described in Section 4.2, selected the relevant subsets of the data and exported them for visual analytics to the Graph Exchange XML Format[13]. We then used Gephi [12] to visually explore the resulting visualizations, calculated the descriptive measures and tested the hypotheses.

## 7 Discussion and Outlook

In this chapter we have introduced the notation of Research Networks as being special Learning Networks of scholars pursuing their individual learning goals (Section 2). The participants in Research Networks use different learning services to exchange experience, collaboratively elaborate common research questions, offer each other support in solving tasks and to create, share and find learning resources. We have given an overview about the different levels of interaction in Research Networks (Section 2.1) and discussed the possible learning trajectories of participants in Research Networks. We have further explored the application of Web 2.0 tools, technologies and techniques in Research Networks (under the umbrella of terms like Research 2.0 or Science 2.0). We discussed the fact that Research Networks now are driven by new technologies, practices and methods and presented overall goals for mining Research Networks.

In Section 4 we introduced the approach of Artefact-Actor-Networks (AANs) for mining Research Networks. AANs semantically intertwine social networks with artefact networks. Both network types have multiple layers for each learning service used in the Research Network what allows the layer-wise analysis as well as the consolidated one. Objects in the networks are artefacts and actors that are connected via semantic relations. Moreover, we presented a reference implementation for AANs that was developed as modular application in Java using the OSGi Service Platform. We used the reference application for analyzing six research networks. They were introduced in Section 5 and classified according to their type. We described the mining of the Research Networks in Section **??** and showed that Research Networks of the same type are comparable based on descriptive statistics.

[13] Altogether, we exported 144 subsets, containing data of the six Research Networks, the three learning services (Delicious, SlideShare, Twitter) and a consolidated set, the three levels (artefact-, actor-, and combined Artefact-Actor-Networks), and two different graph types (semantic similarity of objects and the networks themselves).

We found that hashtag communities are least dense on the artefact level but best connected when comparing the semantic similarity of artefacts. Further, our analysis showed that the narrower the subject of a Research Network, the higher the semantic similarity of the associated artefacts will be no matter if there are vivid social interactions between the participants in the Research Network.

The results of the research presented here will have to be further validated in prospective experiments. Future experiments should try to mine learning trajectories of participants and to identify boundary objects that connect various Research Networks. Moreover, we should extend the possible data sources to incorporate more learning services; for conferences for example we will consider publications as another type of artefact in the future. The analyses in this chapter took place subsequent to the happenings in the selected Research Networks. In order to better support the awareness of participants in Research Networks, we should prospectively provide them with real time analyses that help them better assessing their knowledge, recognizing competence deficits and being aware about the network structure and evolvement. For future research, we will explore how the presented results can be applied to the context of scientific events in order to raise awareness about the topical narrowness of an event and to predict discussed themes.

## References


[1] ALT-C 2010 Organizing Committee. ALT-C 2010: Into something rich and strange. Available online `http://www.alt.ac.uk/altc2010/index.html` (accessed 31 December 2010), 2010.

[2] Charles Arthur. Free Our Data debate goes public, hydrogen-powered London buses, US and ICANN, ISPs and file sharing, the 1% rule and much more... Available online `http://www.guardian.co.uk/technology/blog/2006/jul/20/freeourdatad` accessed 31.12.2010, July 2006.

[3] Association for the Advancement of Computing in Education. AACE - ED-MEDIA. Available online `http://www.aace.org/conf/edmedia/` (accessed 31 December 2010), 2010.

[4] Ivana Bosnic, Katrien Verbert, and Erik Duval. Automatic Keywords Extraction – a Basis for Content Recommendation. In *Proceedings of the Fourth International Workshop on Search and Exchange of e-le@rning Materials 2010 (SE@M10)*, pages 51–60, 2010.

[5] D. Butler. Science in the web age: Joint efforts. *Nature*, 438:548–549, 2005.

[6] Karin Knorr Cetina. Sociality with Objects: Social Relations in Postsocial Knowledge Societies. *Theory Culture Society*, 14(4):1–30, 1997.

[7] Crowdsourced. Deletionism and inclusionism in Wikipedia. Available online http://en.wikipedia.org/w/index.php?title=Deletionism_and_inclusionism_in_Wikipedia&oldid=392147959 (accessed 31 December 2010), December 2010.

[8] Erik Duval. Science 2.0 approach to research. Available online http://erikduval.wordpress.com/2010/06/10/science-2-0-approach-to-research/ (accessed 31 December 2010), June 2010.

[9] EDUCAUSE. 7 Things You Should Know About Personal Learning Environments. Report, available online http://net.educause.edu/ir/library/pdf/ELI7049.pdf (accessed 31 december 2010), EDUCAUSE Learning Initiative, 2009.

[10] Jyri Engeström. Why some social network services work and others don't — or: the case for object-centered sociality. Available online http://bit.ly/eJA7OQ (accessed 31 December 2010), April 2005.

[11] G. Fischer. Communities of Interest: Learning through the interaction of multiple knowledge systems. In *24th Annual Information Systems Research Seminar In Scandinavia (IRIS'24)*, pages 1–14, 2001.

[12] Gephi NGO. Gephi - the open graph viz platform. Available online http://gephi.org (accessed 31 December 2010), December 2010.

[13] Denis Gillet. eResearch: Community of Practice and Social Media for PhD Students. Available online http://www.slideshare.net/dgillet/eresearch-4693039 (accessed 31 December 2010), July 2010.

[14] Google Inc. Doubleclick ad planner by google (for en.wikipedia.org). Available online https://www.google.com/adplanner/planning/site_profile?identifier=en.wikipedia.org (accessed 31 December 2010), December 2010.

[15] K. Hamlin. Unconferencing. Available online http://www.kaliyasblogs.net/Files/unconferencing.pdf (accessed 31 December 2010), February 2008.

[16] M. Hannafin, S. Land, and K. Oliver. *Instructional design theories and models: A new paradigm of instructional theory*, chapter Open learning environments: Foundations, methods, and models, pages 115–140. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA, 1999.

[17] Nina Heinze and Wolfgang Reinhardt. *Educating Educators with Social Media*, volume 1 of *Cutting-edge Technologies in Higher Education*, chapter Future Social Learning Networks at Universities – an Exploratory Seminar Setting, pages 153–170. Emerald Publishing Group, 2011.

[18] Jenny Hughes, editor. *Teacher's Aids on Creating Content for Learning Environments - The E-learning Handbook for Classroom Teachers*, chapter Chapter 29: Communities of sharing. Go! Onderwijs Van De Vlaamse Gemeenschap, 2009.

[19] Jena team. Jena – A Semantic Web Framework for Java. Available online http://openjena.org (accessed 31 December 2010), 2010.

[20] S. Jones. *Cybersociety: Computer-mediated communication and community*, chapter Understanding community in the information age, pages 10–35. Sage Publications Ltd, Thousand Oaks, CA., USA, 1995.

[21] B. Kieslinger and S. Lindstaedt. Science 2.0 practices in the field of technology enhanced learning. In *Proceedings of the 1st International Workshop on Science 2.0 for TEL at the 4th European Conference on Technology Enhanced Learning (EC-TEL'09)*, 2009.

[22] Amy Jo Kim. *Community Building on the Web: Secret Strategies for Successful Online Communities*. Addison-Wesley, Boston, MA, USA, 1 edition, 2000.

[23] Rob Koper, editor. *Learning Network Services for Professional Development*. Springer, 2009.

[24] Rob Koper. *Learning Network Services for Professional Development*, chapter Introduction, pages 3–11. Springer, 2009.

[25] Rob Koper and Peter Sloep. Learning Networks - connecting people, organizations, autonomous agents and learning resources to establish the emergence of effective lifelong learning. Programme plan, Open University of the Netherlands, 2003.

[26] J. Lave and E. Wenger. *Situated learning: Legitimate peripheral participation*. Cambridge University Pres, 1991.

[27] Michael Nielsen. The future of science: Building a better collective memory. Available online `http://michaelnielsen.org/blog/the-future-of-science-2/` (accessed 31 December 2010), July 2008.

[28] K. Oliver and M. Hannafin. Developing and refining mental models in open-ended learning environments: A case study. *Educational Technology, Research and Development*, 49(4):5–32, 2001.

[29] Orchestr8 LLC. Alchemyapi. Available online `http://www.alchemyapi.com/` (accessed 31 December 2010), July 2009.

[30] OSGi Alliance. OSGi Alliance. Available online `http://www.osgi.org/` (accessed 31 December 2010), 2010.

[31] Luciano Paccagnella. Getting the seats of your pants dirty: Strategies for ethnographic research on virtual communities. *J. Computer-Mediated Communication*, 3(1), 1997.

[32] PLE Conference Organizing Committee. PLE Conference. Available online `http://pleconference.citilab.eu/` (accessed 31 December 2010), 2010.

[33] Rob Procter, Robin Williams, James Stewart, Meik Poschen, Helene Snee, Alex Voss, and Marzieh Asgari-Targhi. Adoption and use of web 2.0 in scholarly communications. *Phil. Trans. R. Soc. A*, 368:4039–4056, 2010.

[34] C. Redecker, K. Ala-Mutka, M. Bacigalupo, A. Ferrari, and Y. Punie. Learning 2.0: The Impact of Web 2.0 Innovations on Education and Training in Europe. Jcr scientific and technical report, European Commission, Joint Research Centre (JRC), 2009.

[35] Wolfgang Reinhardt, Martin Ebner, Guenter Beham, and Cristina Costa. How people are using Twitter during conferences. In V. Hornung-Prähauser and M. Luckmann, editors, *Creativity and Innovation Competencies on the Web. Proceedings of the 5th EduMedia 2009, Salzburg*, pages 145–156, 2009.

[36] Wolfgang Reinhardt, Matthias Moi, and Tobias Varlemann. Artefact-Actor-Networks as tie between social networks and artefact networks. In *Proceedings of CollaborateCom 2009*, 2009.

[37] Wolfgang Reinhardt, Tobias Varlemann, Matthias Moi, and Adrian Wilke. Modeling, obtaining and storing data from social media tools with Artefact-Actor-Networks. In Martin Atzmueller, Dominik Benz, Andreas Hotho, and Gerd Stumme, editors, *LWA 2010 Workshop Proceedings*, pages 323–330, Kassel, 2010.

[38] Thomson Reuters. Calais: Connect. Everything. Available online `http://www.opencalais.com/about` (accessed 31 December 2010), 2009.

[39] Howard Rheingold. *The virtual community: Homesteading on the electronic frontier*. Addison-Wesley, Reading, Mass., USA, 1993.

[40] D. De Roure. "the new e-science". Available online `http://www.slideshare.net/dder/new-escience-edinburgh-late-edition-presentation` (accessed 31 December 2010), November 2007.

[41] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523, 1988.

[42] G. Salton and M. J. McGill. *Introduction to modern information retrieval*. McGraw-Hill, 1983.

[43] Ben Shneiderman. Science 2.0. *Science*, 319(5868):1349–1350, 2008.

[44] Peter Sloep. *Learning Network Services for Professional Development*, chapter Social Interaction in Learning Networks, pages 13–16. Springer, 2009.

[45] Solo 2010 Organizing Committee. Science Online London 2010. Available online `http://scienceonlinelondon.org/` (accessed 31 December 2010), 2010.

[46] O. Tacke. Open Science 2.0: How Research and Education can benefit from Open Innovation and Web 2.0. In *Proceedings of the 1st Symposium on Collective Intelligence (COLLIN 2010)*, 2010.

[47] Thomas D. Ullmann, Fridolin Wild, Peter Scott, Erik Duval, Bram Vandeputte, Gonzalo Parra, Wolfgang Reinhardt, Nina Heinze, Peter Kraker, Angela Fessl, Stefanie Lindstaedt, Till Nagel, and Denis Gillet. Components of a research 2.0 infrastructure. In M. Wolpers, P. A. Kirschner, M. Scheffel, S. Lindstaedt, and V. Dimitrova, editors, *Sustaining TEL: From Innovation to Learning and Practice (EC-TEL 2010)*. Berlin: Springer/Physica, 2010.

[48] J. Underwood, R. Luckin, H. Smith, K. Walker, D. Rowland, G. Fitzpatrick, J. Good, and S. Benford. Reflections on participatory science for telsci2.0. In *Proceedings of the 1st International Workshop on Science*

*2.0 for TEL at the 4th European Conference on Technology Enhanced Learning (EC-TEL'09)*, 2009.

[49] W3C, World Wide Web Consortium. SPARQL Query Language for RDF. Available online `http://www.w3.org/TR/rdf-sparql-query/` (accessed 31 December 2010), January 2008.

[50] W3C, World Wide Web Consortium. Owl Web Ontology Language. Available online `http://www.w3.org/TR/2004/REC-owl-features-20040210/` (accessed 31 December 2010), 2010.

[51] W3C, World Wide Web Consortium. RDF Vocabulary Description Language 1.0: RDF Schema. Available online `http://www.w3.org/TR/2004/REC-rdf-schema-20040210/` (accessed 31 December 2010), 2010.

[52] M.M. Waldrop. Science 2.0. *Scientific American*, 298(5):68–73, 2008.

[53] Katrin Weller, Ramona Dornstädter, Raimonds Freimanis, Raphael N. Klein, and Maredith Perez. Social Software in Academia: Three Studies on Users' Acceptance of Web 2.0 Services. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, 2010.

[54] Barry Wellman. Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research*, 25(2):227–252, 2001.

[55] S. Wilson, O. Liber, M. Johnson, P. Beauvoir, P. Sharples, and C. Milligan. Personal Learning Environments: Challenging the dominant design of educational systems. *Journal of e-Learning and Knowledge Society*, 2, 2007.

[56] Erik Zachte. Wikipedia statistics. Available online `http://stats.wikimedia.org/EN/TablesWikipediansEditsGt5.htm` (accessed 31 December 2010), November 2010.