Analysis of Dynamic Resource Access Patterns in Online Courses

Tobias Hecking, Sabrina Ziebarth, H. Ulrich Hoppe
University of Duisburg-Essen, Germany
hecking@collide.info

ABSTRACT: This paper presents an analysis of resource access patterns in two recently conducted online courses. One of these has been a master level university lecture taught as a blended learning course with a wide range of online learning activities and materials, including collaborative wikis, self-tests, and thematic videos. The other course has been offered in the form of a MOOC. As a specialty of this course, master level students from two different universities could participate as a regular university class and receive credits for successful completion. In both courses, online learning resources such as videos, scientific literature, and wikis played a central role. In this context, the motivation for our research was to investigate characteristic patterns of resource usage of the learners. In order to gain deeper insights into the usage of learning materials, we have adapted methods from social network analysis and applied them to dynamic bipartite student-resource networks built from event logs of the students' resource access. In particular, we describe the clustering of students and resources in such networks and propose a method to identify patterns of the cluster evolution over time.

KEYWORDS: Learning resources, MOOCs, SPOCs, social network analysis

1 INTRODUCTION

Learning materials play an important role in online courses. When there is no or very limited direct communication between learners and a knowledge conveying person (such as a teacher or tutor), learners have to be provided with a set of learning resources that enable self-directed acquisition of knowledge. To achieve this, traditional reading materials and lecture videos can be accompanied by quizzes for self-assessment and by wikis or forums to support collaborative writing, peer reviewing, and discussing the material (Ziebarth & Hoppe, 2014). The combination of the aforementioned resource types and activities provides the typical setting of what we call a "resource-intensive online course."

This paper presents a study on the student-resource interactions in two of such online courses, both using Moodle[1] as a learning management system (LMS). Network analysis methods were used to investigate the dynamics of relations between students and resources in order to identify characteristic patterns of the courses. The first case study focuses on a resource-intensive blended master-level university course. While in previous instances of these lecture the LMS was only used to distribute lecture slides and exercise assignments, this course followed a new approach inspired by the concept of small private online courses (Fox, 2013) that combine elements from massive open online courses (MOOCs) with traditional university classes. The course took up elements of both xMOOCs and connectivist MOOCs (cMOOCs) (cf. Rodriguez, 2013), including the supply of lecture videos, reading materials, weekly online exercises, quizzes, and discussion forums as well as collaborative activities such as the production of wiki articles based on group writing and peer reviewing activities. These

---

[1] https://moodle.org/

collaborative online activities completely replaced the previous presence-based and primarily teacher-centric exercise. The course was not opened to the outside, thus there was no "massification" beyond the expected audience of about 40 students from four M.Sc. programs, all related to computer science and interactive media. In a second case study, the same method was applied to data gathered from an online course designed as a relatively small MOOC with 173 participants on computer-mediated communication (CMC). The focus of resource usage in this course was more on videos and reading material, which led to different resource access behaviour in comparison to the first case study.

The contributions of this paper are twofold: first, a new method for learning analytics is introduced that allows identifying and tracing cohesive clusters of students and resources over time; second this method is applied to the two aforementioned case studies.

Based on the log protocols of resource access of the students, relations between students and learning resources are modelled as bipartite networks. In such networks, a link between a student and a learning resource can be interpreted as interest of the student in this resource. This interest either can be intrinsically motivated — for example, by curiosity — or externally stimulated by assignments and course design. Cohesive clusters of students and resources indicate a concentration of interests and activities around certain materials by specific subgroup of students in a given time segment. This does not necessarily imply a direct interaction between students, although this may be the case especially for content created by groups of learners (such as wiki articles). The tracking of the evolution of such mixed clusters of students and learning resources further describes how the affiliations between students and resources change over time. This information can be used to characterize different online courses according to the specific evolution of resource usage, e.g. courses with homogeneous or diverse resource usage.

The proposed approach was applied to two case studies to gain insights into the resource usage in different kinds of online courses. The main research questions are as follows:
- Do students use content produced by peers, or do they prefer the materials provided by teachers?
- Which resources are important for which groups of students?
- How do the relations between students and learning resources evolve over time?
- Are there differences regarding the resource access patterns of students of different study programmes participating in the same course?
- How can the evolution of resource access over time be characterized in different online courses?

Based on the answers to these questions, we hope to improve the design of future instances of the courses.

The rest of the paper is structured as follows: Section 2 describes the theoretical background concerning the role of learning resources and activities in general and more specific in the context of online learning. In section 3, we introduce our analysis methods before we explain the analysis of the mentioned case studies in section 4. Section 5 concludes the main results and gives an outline of possible future research directions.

2   BACKGROUND

2.1 The Role of Learning Resources in Online Courses

Online learning environments and their ability to facilitate distributed learning activities enable a wide variety of course designs, especially in academic education. More than a decade ago, Hill and Hannafin (2001) introduced the concept of resource-based learning environments (RBLEs), where learners are provided with various static and dynamic learning resources in conjunction with tools for manipulation and search in order to satisfy different learning needs. Today, massive open online courses (MOOCs) constitute a current trend in practical applications of technology-enhanced learning. Because of the large number of participants in such online courses, the personal supervision of learners by teachers is not possible. Therefore, lessons are given as video lectures. Exercises often comprise online quizzes/tests that can be automatically evaluated as well as other tasks reviewed by peers (e.g., see Belanger & Thornton, 2013). While in xMOOCs students mainly act as consumers of predefined learning contents, connectivist MOOCs (cMOOCs), as introduced by G. Siemens and S. Downs (c.f. Fini, 2009), focus on the production and sharing of knowledge artefacts like blogs or wiki articles by and between learners (Rodriguez, 2013). Apart from targeting a broad audience, the shift of learning activities into online environments can be of advantage even on smaller scale (Bruff et al., 2013; Harrer et al., 2007). Thus, "blended learning courses" provide hybrid approaches that combine the benefits of online learning and face-to-face sessions. Such course designs enable learners to consume learning resources and share and discuss results asynchronously — independent of time and place — while still having classes/events with physical attendance for community building and supervision (Garrison & Kanuka, 2004). In general, learning resources of various types play an important role in course design regarding the support of autonomous learning. It has also been argued that due to the variety of media and delivery channels online courses are particularly suited to address the needs of different learner types (Grünewald et al., 2013).

2.2 Analysis Methods for Learning Resource Usage

By investigating the relations between learners and learning resources, we expect to gain deeper insight into the function of the learning community. In general, analysis methods for learning resource usage can be characterized in two dimensions, namely structural and sequence analysis.

Modelling relations between learners and the resources they employ can be used to reflect structural characteristics of the resource usage of learners. Nachmias and Segev (2003) showed that a large proportion of the resources provided in a web-based learning environment is used by the students. However, by comparing the resource usage on the level of individuals, large differences regarding the quantity of different resources used could be identified. Hoppe et al. (2005) introduced the concept of social-thematic navigation through the sharing of learner-created "emerging learning objects." In this view, relations between groups of learners induced by thematically related learning objects indicate learners with a common interest. The learners could interact indirectly mediated by those objects without necessarily having person-to-person communication. Social network analysis methods have also been applied to networks of people and artefacts to support the evolution of a knowledge-creating community in terms of the identification of people with common interests as well as trend analysis (Harrer et al., 2007). The work presented in Romero, Ventura, and García (2008) evaluates data-mining methods for resource usage of students in LMS including statistics, clustering, and classification.

The dimension of sequential analysis is mostly concerned with the behaviour of learners over time. There are, especially, data mining methods used to identify typical sequences of resource access over time (Perera et al., 2009). This also enables the identification of different learning paths for different learners that can be compared to the intended learning path by the teacher (Pahl & Donnellan, 2002; Romero et al., 2008). More recently, Kizilcec, Piech, and Schneider (2013) discovered engagement patterns of learners in MOOCs based on their access to video lectures and assessments in order to classify the users into subpopulations.

The method described in this paper enables structural analysis of networks between students and learning resources. While apart from sequential analysis most methods rely on static snapshots of the state of a learning course, and thus are not able to capture the dynamics, this work explicitly considers the variation of the student-resource access patterns over time. Further, the thematic classification of learning resources used together by groups of students are taken into account, which will be described in section 3.5.

3 ANALYSIS METHODS

3.1 Building Student-Resource Networks from Students Action Logs

Actor-resource networks are bipartite networks. Bipartite networks (Wasserman & Faust, 1994), also called "two-mode" networks or "affiliation networks," contain two distinct types of vertices. Edges can only occur between vertices of different types.

In the case of bipartite actor-resource networks, one set of vertices represents actors who have connections to a second set of vertices that represent resources. In our study, we investigate bipartite networks of students and learning resources over time. The networks are based on the students' activities within the learning management system Moodle where the relations of students and resources can be observed by event logs of resource access. Whenever a student accesses a resource (opens a lecture video, for example), an event is generated and stored in the Moodle database. These event logs contain the identification of the student, the name, and identification of the resource, as well as the timestamp of the event. Based on these log files, a time series of bipartite student-resource networks can be extracted by sliding a specified time window over the stream of resource access events. Each network of the series (time slice) contains all relations between students and resources that occurred within the corresponding time window.

3.2 Clustering of Student-Resource Networks

Actors in evolving social networks usually develop a community structure. This means that certain subsets of actors tend to be more densely connected to each other than to outsiders (Backstrom et al., 2006; Watts & Strogatz, 1998). The detection of such cohesive subgroups, or subcommunities of actors in a social network, is a common task in social network analysis (Fortunato, 2010). Subcommunity detection methods can partition a graph into disjointed sets of nodes while maximizing the modularity between clusters (Girvan & Newman, 2002). Other methods do not cluster all nodes exhaustively and allow overlaps between clusters so that nodes cannot necessarily be assigned to a subgroup uniquely. A

prominent example for methods that detect overlapping subgroups is the Clique Percolation Method (CPM) (Palla et al., 2005).

In this study, however, we investigate bipartite graphs where students are connected to learning resources they accessed during a certain period. In such actor-resource networks, it is also possible to identify densely connected substructures. Thus, a bipartite cohesive subgroup, also called bipartite cluster, is a bipartite subgraph where a set of actors is more densely connected to a set of resources than to resources outside the cluster. This can help researchers to understand which students are interested in which resources, and how such affiliations change over time.
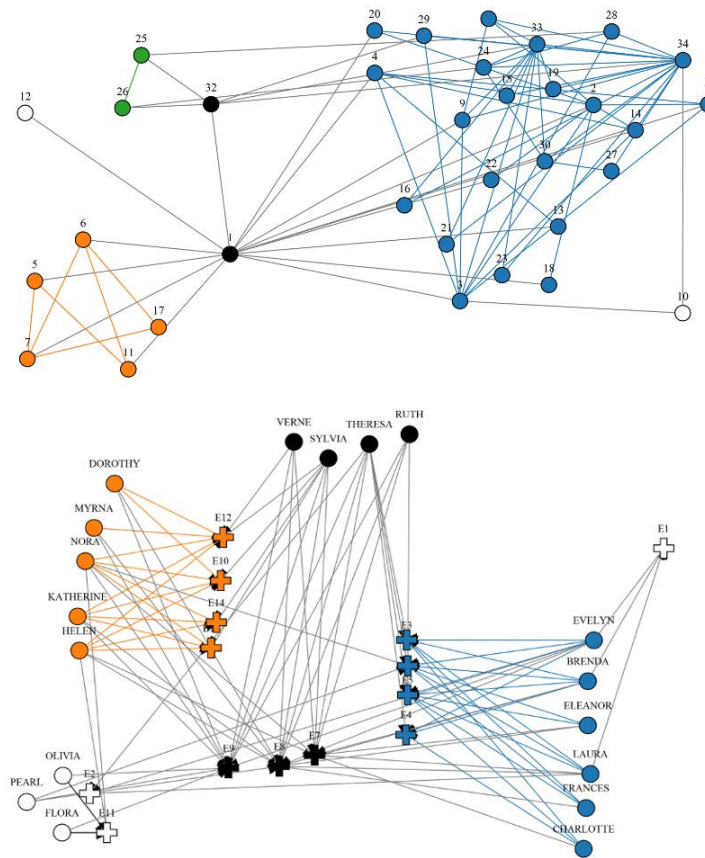

Figure 1: Comparison of clustering of a one-mode social network and a bipartite actor-resource network. The node shapes indicate different node types. Black nodes are in more than one cluster.

In comparison to clusters in one-mode networks, a bipartite cluster $BC = \{A, R\}$ consists of two parts, namely a set of actors (students: $A = \{a_1, a_2, \ldots, a_n\}$) and a set of resources ($R = \{r_1, r_2, \ldots, r_m\}$). Figure 1 shows the difference between clusters in one-mode and two-mode networks. For example, the orange bipartite cluster in the bottom part of the figure contains the actor group $A = \{Dorothy, Myrana, Nora, Cathrin, Helen\}$ and the resource group $R = \{E10, E11, E14, E15\}$. In contrast to subgroup detection in one-mode social networks, where social ties between actors are directly observable, the actors within a bipartite actor-resource cluster are not necessarily connected by social relations. Moreover, the actors share a common interest in or awareness of the resources of the bipartite cluster. In the following, the set of actors in a bipartite cluster is called "actor group of the

cluster" and the set of resources is called "resource group of the cluster."

The methods for subgroup detection in one-mode networks mentioned above cannot be applied directly to identify bipartite subgroups. One modification of a subgroup detection method that works for bipartite graphs is based on the Clique Percolation method (Palla et al., 2005). The Biclique Communities method (Lehmann, Schwartz, & Hansen, 2008) relies on the definition of a $K_{a,b}$ biclique. This is a maximal connected bipartite subgraph with a nodes of the first mode and b nodes of the second mode. Thus, if a set of a actors all are connected to each of the b resources, they form a $K_{a,b}$ biclique. A bipartite subgroup also called biclique community in Lehmann, Schwartz, and Hansen (2008) is defined as the union of a series of adjacent $K_{a,b}$ bicliques. Two $K_{a,b}$ bicliques are adjacent if they share at least $a-1$ nodes of the first mode and $b-1$ nodes of the second mode. Figure 2 depicts an example of two adjacent $K_{2,2}$ bicliques. Since only nodes that are part of a $K_{a,b}$ biclique are assigned to clusters, the method produces non-exhaustive clustering. Further, it is possible that nodes are assigned to more than one cluster, which results in overlapping clusters.
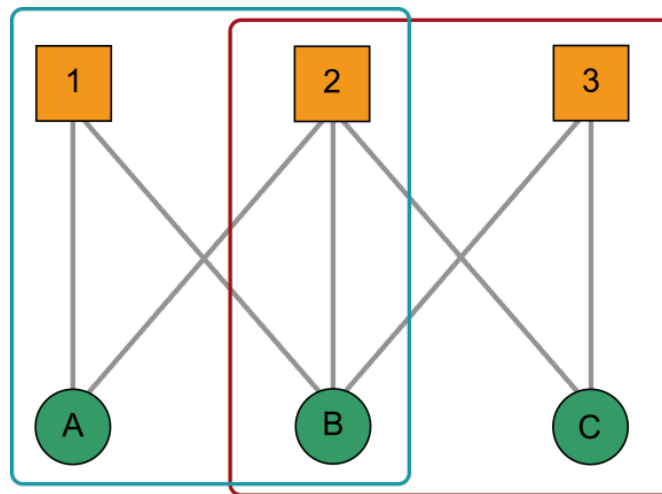

Figure 2: Example of two adjacent $K_{2,2}$ bicliques. First clique: {A, B, 1, 2}, second clique: {B, C, 2, 3}.

3.3 Detection of Evolution Patterns in Student-Resource Networks

The analysis of a static snapshot of a social network or actor-resource network can give various insights into their structures (Wasserman & Faust, 1994). However, since most networks evolve over time, the community structure also changes. In section 3.1 the transformation of event logs of resource access into subsequent snapshots (time slices) of an evolving bipartite network has been described. It is likely that a subgroup detection method will uncover different subgroups in each of the time slices. Since subgroups do not appear and disappear at random, there can be similar subgroups in consecutive time slices. This phenomenon has recently been studied in social network analysis (Leskovec, Kleinberg, & Faloutsos, 2005; Palla, Barabasi, & Vicsek, 2007). In social networks, a group of people may stay in contact for a certain period. New members may join the group and others may leave at certain points but a stable majority is considered as the same subgroup over several points in time. In such situations, one can speak of the same subcommunity with different instances in different time slices or snapshots of a dynamic network.

Palla, Barabasi, and Vicsek (2007) define six basic lifecycle events for subcommunities in one-mode social networks. These events are:

- Birth: A subcommunity is identified the first time.

- Growth: A subcommunity acquires new members but its core stays the same.

- Contraction: A subcommunity loses members over time.

- Merge: The members of distinct subgroups merge to one subgroup at a later point in time.

- Split: One subcommunity splits into two new sub-communities.

- Death: A subcommunity disappears over time.

In bipartite networks of students and learning resources, one may also find patterns of the evolution of student-resource clusters. The difference to one-mode networks is that in bipartite networks the previously described events can happen to the actor and resource groups of the bipartite clusters separately. By applying a bipartite subgroup detection method to different slices of an evolving student-resource network, one can identify traces of dynamically evolving actor groups and resource groups. Figure 3 depicts an example of such traces. The circle- and the square-shaped nodes represent actor and resource groups respectively. Each group of actors forms a bipartite cluster with a group of resources at a certain point in time, indicated by a dotted connection. According to the change of affiliation of students to resources, the detected bipartite clusters can change over time. However, similar actor groups (student groups) can be detected over time as part of different bipartite clusters. Thus, one can identify relations between similar actor or resource groups of subsequent time slices. These relations are depicted as arrows in Figure 3.
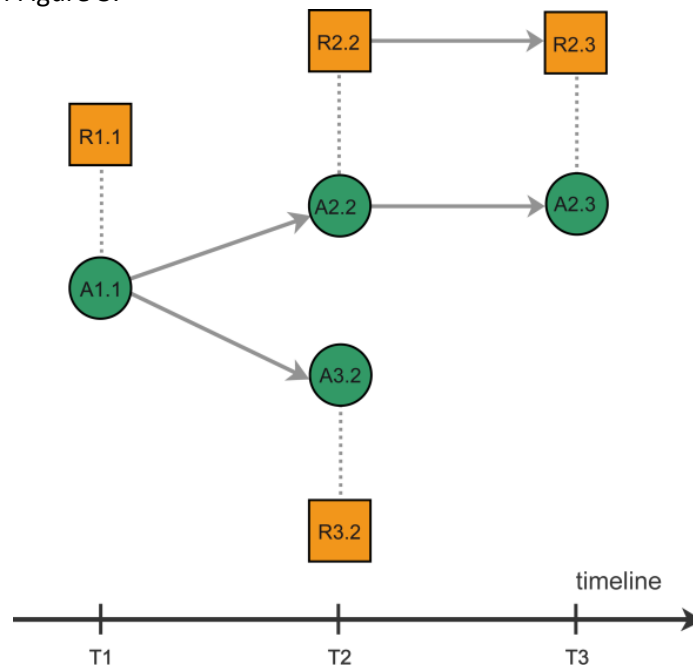

Figure 3: Traces of dynamic actor (green circles) and resource groups (orange squares) of bipartite clusters over time.

Table 1 shows those relations that allow the tracing of actor and resource groups over time, which results in a timeline as in Figure 3.

3.4 Tracing the Evolution of Dynamic Actor and Resource Groups

Given a list of subsequent time slices of k evolving actor-resource networks $(G_1, G_2, \ldots, G_k)$, the first step is to apply the Biclique Communities method, described in section 3.2, to each of these graphs. The result is an ordered list of k sets of bipartite clusters (or "biclique communities") for the k time slices $(\Theta_1 = \{BC_{1,1}, BC_{1,2}, \ldots\}, \Theta_2 = \{BC_{2,1}, BC_{2,2}, \ldots\}, \ldots, \Theta_k = \{BC_{k,1}, BC_{k,2}, \ldots\})$. As explained in section 3.2, each of the bipartite clusters $BC_{i,j} = \{A_{i,j}, R_{i,j}\}$ contains an actor group and a resource group. In the context of the tracing algorithm, the actor groups and resource groups of particular time slices are called "step-actor groups" and "step-resource groups" respectively.
Similar step-actor/resource groups occurring in subsequent time slices are assumed to be instances of the same dynamic group. Thus, a "dynamic actor group" is a timeline of similar actor groups and a "dynamic resource group" is a timeline of similar resource groups. Figure 3 depicts the traces of a dynamic actor and a resource group with step-groups as particular instances. By tracing step-groups of actors and resources over several time slices, different lifecycle events, or evolution patterns, of bipartite clusters can be identified (see Table 1).

To identify the described events, a method is needed that matches the step-groups of the time slices with step-groups of previous time slices. Although there are methods for tracing the evolution of dynamic subcommunities in one-mode social networks (Asur, Parthasarathy, & Ucar, 2009; Greene, Doyle, & Cunningham, 2010; Palla, Barabasi, & Vicsek, 2007), to our knowledge there are no such methods for bipartite networks that allow the identification of the evolutionary events in Table 1. In the following, an adaption of the method of Greene, Doyle, & Cunningham (2010) is introduced, which was originally designed for one-mode networks to trace the subgroup structure in bipartite networks like learner resource networks:

Initialization:
The method maintains two sets of not matched groups, one for not matched step-actor groups (NMA) and one for not matched step-resource groups (NMR). At the beginning of the process, the set NMA is the set of step-actor groups of the first time slice — $T_1(NMA = \{A_{1,1}, A_{1,2}, \ldots\})$ — and NMR contains all step-resource groups — $(NMR = \{R_{1,1}, R_{1,2}, \ldots\})$.

Matching phase:
After the initialization, the algorithm proceeds with the step-groups of the next time slice, T_i. The next step is to compute the similarity of the step-actor groups of T_i with all step-actor groups in NMA and the similarity of the step-resource groups in T_i with the step-resource groups in NMR. As in Greene, Doyle, and Cunningham (2010) and Palla et al. (2007) the Jaccard coefficient is used as measure of similarity between two groups:

$$sim(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

Here, X and Y are either two step-actor groups or two step-resource groups. If $sim(X, Y)$ exceeds a certain threshold, the groups will be matched.

There are four different situations that can occur when matching the current step-groups in T_i to the not

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

41

matched step-groups in NMA and NMR in T_{i-1}.

1. One-to-one match between a step-group of the current time slice and a previously not matched group in NMA or NMR. In this case, the two groups belong to the same timeline of a dynamic group. In Figure 3 this is the case for the dynamic resource group R_2, which occurs in T_2 and T_3.
2. One-to-many match: One not previously matched group in T_{i-1} matches to more than one group of the current time slice T_i. This indicates that the not previously matched group splits up into the matched step-groups of T_i. The matched step-groups of T_i are than considered as the first occurrences of a new dynamic group. In Figure 3, the actor group $A_{1,1}$ in timeline 1 matches with the two actor groups $A_{2,2}$ and $A_{3,2}$, which will then be considered instances of new dynamic actor groups A_2 and A_3.
3. Many-to-one match: More than one not previously matched group in NMA or NMR match to the same step-group of T_i. This means that these groups merge to the matching group of T_i. The forth column of Table 1 is an example of such a situation.
4. A step-actor or step-resource group of T_i cannot be matched to any of the groups in NMA or NMR. In this case, such a group will be the first instance of a new dynamic group. An example for this is the resource group $R_{2,2}$ in Figure 3.

After all step-groups of the currently investigated time slice T_i were checked, all groups from NMA and NMR that could be matched with one or more step-groups of the currently investigated time slice T_i will be removed from these sets. All step-groups of the current time slice T_i are then added to NMA and NMR respectively.

Then the algorithm moves on to the set of groups in the next time slice. The algorithm stops when the last time slice is processed. The result can then be visualized as a swim lane graph, as in Figure 3. In this visualization, a student-resource cluster is represented by two nodes. The square nodes represent the resource groups of the bipartite clusters while the circle nodes represent the student (or actor) groups. A student group node is always connected to a resource group node by a vertical line. This indicates that the particular student group forms a bipartite cluster together with the resource group in a corresponding time slice. As described above, the different groups found in one time slice of the evolving student-resource network can be re-identified in future time slices by the described similarity-based matching condition. Visually, groups matched across subsequent time steps are connected by a horizontal arrow. Hence, these arrows always point forward in time and allow the depiction of merge, split, and continuing student and resource groups.

Table 1: Possible events in bipartite cluster evolution over time.

	A group of actors and their affiliations to resources stays stable over subsequent steps in time.
	A group of actors shift their interest to new resources.
	A new group of actors (A2) forms around an existing set of resources.
	Groups A1 and A2 merge and share their affiliations to resources.
	A group of actors (A1) splits into two new groups and no longer shares their affiliations to resources.

3.5 Fostering Cluster Interpretability through Content Analysis

The method described above can be used to track the general resource access behaviour of students over time by clustering students and learning resources simultaneously. In order to interpret the results, one must investigate the content of the clusters. This step can also be supported by computational methods. Cluster labelling (Treeratpituk & Callan, 2006) is one technique commonly used in document clustering to assign keywords, which helps to interpret the content of the clusters. This idea can be adapted to label the resource groups of student-resource clusters by utilizing the metadata of the resources. The cluster labelling in this work follows a semi-automatic approach. Most of the learning resources in the online courses under investigation have assigned metadata such as keywords. Videos can be annotated with tags. Most scientific literature is also annotated with keywords. However, quizzes, lecture slides, and wiki articles are not necessarily described by keyword lists. In these cases, the course managers need to annotate these resources manually.

The resource groups of the student-resource clusters can then be labelled using a simple frequency-based approach. A keyword is assigned to a resource group if it appears in the metadata of at least 70 percent of the resources in the group. In a case where no keyword has support above the threshold of 70 percent, the most frequent keywords are used instead.

Given an example of a student-resource cluster $BC = \{A, R\}$ with a resource group $R = \{r1, r2, r6, r7\}$ from the blended learning course, which will be described in more detail in section 4.1, each resource in the cluster can be interpreted as a list of keywords (see Table 2). Keywords can either be automatically extracted from the resource metadata or added by hand by the course manager. The result for the given example would be the cluster label "ITS" because three of the four resources have this keyword in common. If needed, an analyst can refine this initial labelling by hand.

Table 2: Example of resources and allocated keywords. Italic keywords indicate manually assigned keywords. Other keywords could be directly extracted from resource metadata.

Resource ID	Resource name	Keywords
r1	ITS slides	ITS, *Pseudo Tutors*
r2	Constrained based tutors (wiki)	*Cognitive models, student modelling*
r6	Intelligent Tutoring Systems (part 3)	ITS, constraint based tutors
r7	Ragnemalm — "Student diagnosis in practice; Bridging a gap"	Student diagnosis, student modelling, ITS

This semi-automatic cluster labelling usually has the advantage of being more accurate than fully automatic labelling. The keyword table can be initially filled in by the extracted keywords from the resource metadata. A teacher or course manager knows the resources used in the course very well and thus can refine the keyword tables by adding keywords and unifying synonyms like "ITS" and "Intelligent Tutoring Systems."

4 CASE STUDIES

4.1 The GILLS Blended Learning Course

GILLS ("Gestaltung interaktiver Lehr- / Lernsysteme," or "Design of interactive teaching and learning systems") is a master level course in the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen (Germany) that focuses on user/task modelling and models of interactive and collaborative learning environments. It targets students from the three Master's programmes of "Applied Computer Science," "Interactive Media and Cognition," and "Computer Engineering" and also higher education teachers/students of computer science. The course is taught once per year with about forty participants. It consists of a weekly three-hour lecture accompanied by a one-hour exercise with weekly assignments.

Probably due to the given schedule of the course, attendance at the exercise was typically lower than at the lecture and the willingness for active participation (i.e., presentation of results on the part of students) was also low. Thus, in summer semester of 2013, the exercise was "virtualized" when the presence-based activity was replaced by a combination of individual and group assignments with reporting back incorporated into the Moodle platform (Ziebarth & Hoppe, 2014). The plenary face-to-face setting of the lecture was used as a forum to take up and discuss results from the virtual exercise activities. The role of the lecture was redefined in the sense of providing orientation knowledge ("advance organizer") and core definitions with a limited number of characteristic examples, leaving specific extensions to the virtual exercise activities. These lecture elements were provided as smaller thematic video clips, available on the platform. Evidently, this new approach was inspired by the concept of MOOCs and takes up elements of both xMOOCs and cMOOCs while being neither massive nor open.

The Moodle interface contained one forum for announcements and one for questions and discussion for the participants. The students were asked to post questions relevant to the whole course on the forum rather than to write emails to the supervisors. Additionally, the forum was used to support group exercises. Furthermore, a wiki was introduced to build a glossary for repetition and exam preparation. A typical exercise regarding a wiki article consisted of one week of creating the initial version in a small group, one week of individually reading the articles of other groups and writing peer feedback, and one week of revising the article in the original small group. Each group handled a different topic, enabling students to specialize regarding their interests. Thus, groups were formed by topic using Moodle's choice activity, which was set up to ensure groups of similar size. Furthermore, students could see the choices of their peers to enable group formation and communication.

Overall, the online exercises included (per student):
- Creating and revising four wiki articles in small groups
- Two tasks regarding constructive problem solving and modelling (partially using external tools) in small groups
- Writing peer reviews (as wiki comments) regarding the group results above (individual task)

- Creating and extending a concept map regarding the topics of the first chapter using external tools (group task)

- Three quizzes regarding the three main topics for self-testing, which could be completed several times (individual task)

- Creating questions and sample solutions in a wiki regarding the lecture topics (individual task)

- Playing the serious game Matchballs (Ziebarth, Malzahn, & Hoppe, 2012) for training relations between important concepts of the lecture (individual task).

To motivate students for participating in the online exercises, bonus points were introduced.

The Moodle course contained the following types of resources:
- Lecture videos

- Lecture slides

- Additional reading material like scientific papers

- Assignments including self-test quizzes

- Wiki articles

- Forum discussions

4.1.1 General analysis
Forty-eight students visited the first lecture of the GILLS course; of the 44 students doing the first exercise, 40 regularly did their exercises until the end of the course and 36 did the oral exam.
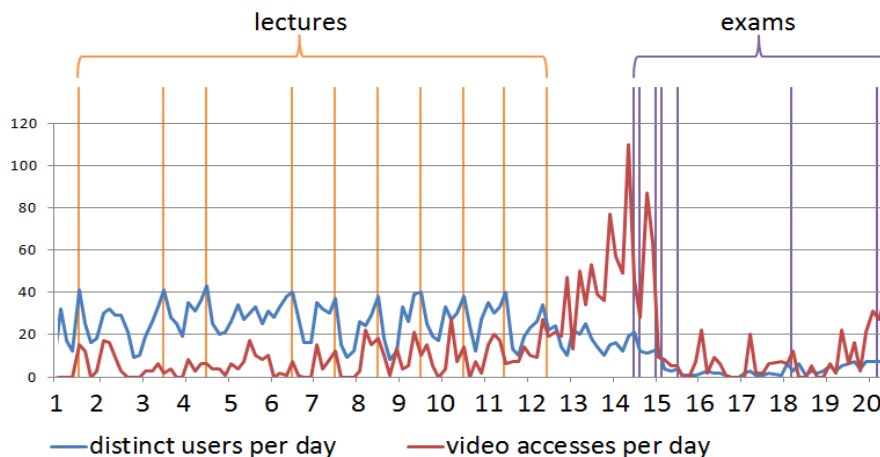

Figure 4: Moodle usage over time.

The Moodle course was frequently visited: There were 24 distinct students per day and 45 per week on average (see also Figure 4). The highest number of distinct users per day was on lecture days, the lowest was on weekends. The deadline for and assignment of new exercises was typically on lecture days, which may explain the high number of accesses on these days.

Figure 5 shows the relations of students in the course based on group choices. The edge width indicates

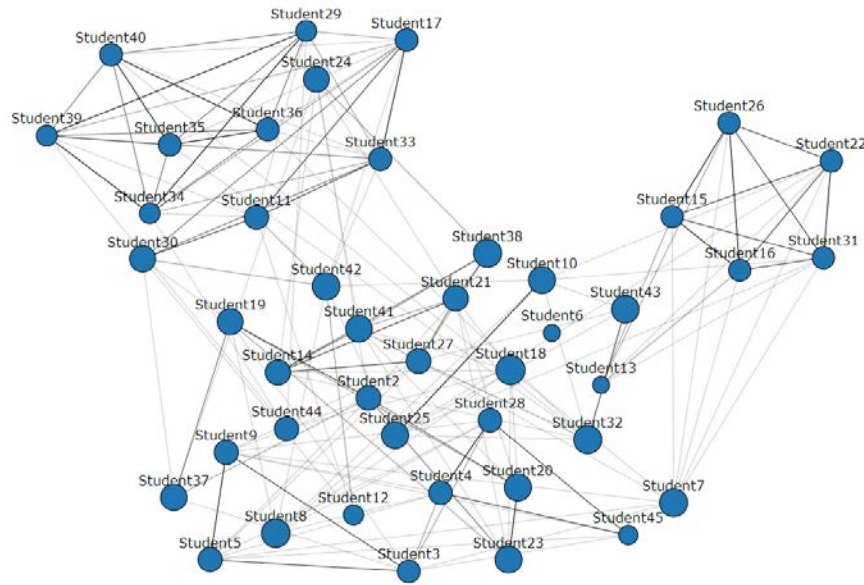how often students were in the same group; the darker the edge, the stronger the relation.

Figure 5: Student relations based on group choices.

The node size represents the centrality of the students; the bigger a node, the more relations to other students. On the one hand, the network shows some groups of students often working together in the same constellation (in the periphery of the graph). These groups often consist of students of the same study programme. On the other hand, there are students without a stable affiliation who work with a high number of different students (in the centre of the graph). These might be the ones choosing an exercise/group by topic rather than for social reasons.

4.1.2 Overlapping clusters of student and learning resources

By applying the Biclique Community method (see section 3.2) to the student resource networks generated from the resource access logs of the students, overlapping bipartite student-resource clusters were retrieved. Learning resources that frequently appear in overlaps between bipartite clusters may be of particular importance because students of different groups have relations to them.

As a prototypical example for the bipartite clustering of students and resources, Figure 6 depicts the subgroup structure of the student–resource network in the 10th week of the lecture period, where the topics were "Intelligent Tutoring Systems" and "Cognitive User Modelling." The exercise assignment for this week was to create wiki articles on these topics. It is worth mentioning that the applied Biclique Communities method for clustering does not cluster the nodes in a network exhaustively. Nodes that do not belong to a $K_{a,b}$ clique (see section 3.2) cannot belong to a cluster. Students who cannot be assigned to bipartite clusters uniquely either have a broader interest in learning resources than other students or do not access enough resources to be part of a $K_{a,b}$ biclique. In this case, $K_{3,5}$ bicliques were used to construct the clusters. Hence, Figure 6 only shows those students and their relations to resources that are part of detected bipartite clusters. The node sizes in Figure 6 correspond to the number of connections of a student or a resource respectively.
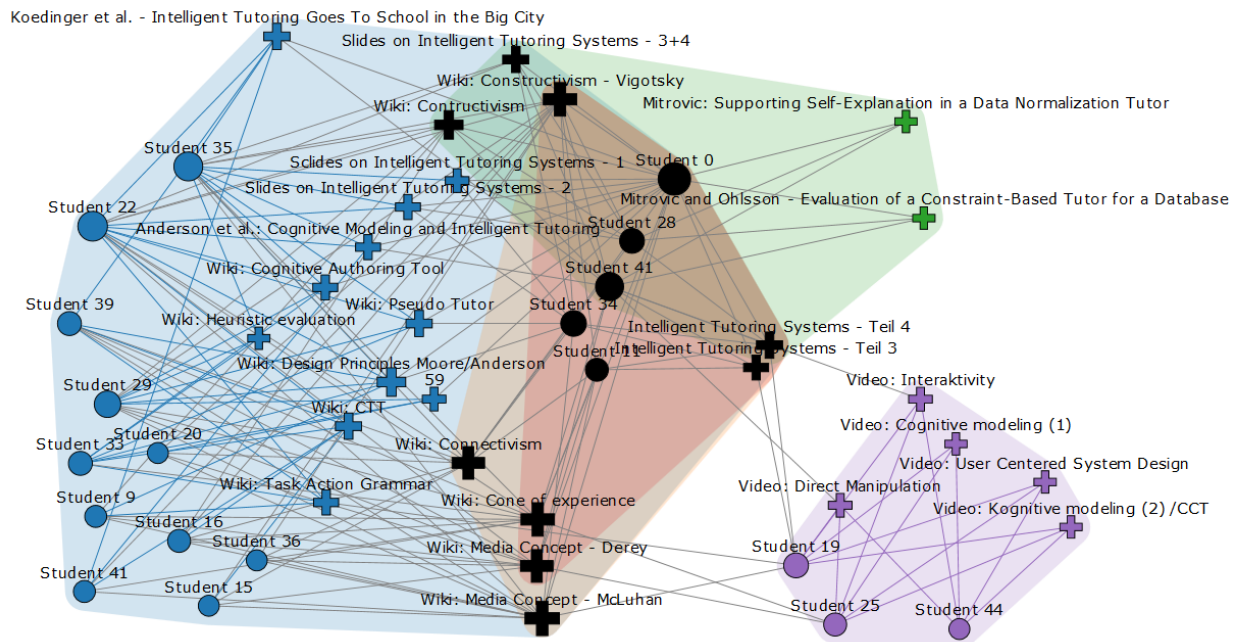
Figure 6: Overlapping bipartite subgroups in the 10th week of the lecture period. Black nodes are in multiple subgroups.

The light red cluster mainly contains the wiki articles created in the previous week, thus, the included students can be considered to have accessed them for creating peer reviews. This cluster overlaps completely with other clusters, since writing peer reviews was only part of the exercise. The second part was writing wiki articles on ITS (two-week task). There were two articles on "Constraint Based Tutoring" (CBT) and two on "Model Tracing Tutors" (one on "Cognitive Tutors" and one on "Pseudo Tutors"). The green cluster focuses on resources (papers, slides, and videos) regarding CBT. The blue cluster, on the one hand, contains resources regarding Model Tracing Tutors (papers, slides, videos) as well as some of the wiki articles that should be created (Cognitive Authoring Tool, Pseudo Tutor). On the other hand, it contains some of the wiki articles (CTT, Heuristic Evaluation, Task Action Grammar) that had to be corrected in the previous week. Since the deadline of the old exercises and the assignment of new ones are on the same day, these old assignments are also considered in the data.

The only resources in the purple cluster are videos of the first weeks of the lecture that are not directly connected to the topics of the considered time interval (week 10). This cluster does not correspond to any *a priori* expectation. Since video resources play an important role for exam preparation, it might indicate an early activity of this type.

4.1.3 Resource access during the lecture period

The resource access of the students is modelled as subsequent time slices of an evolving bipartite network. The size of the time slices is one week so that the resource access of the students during one lecture week is aggregated in one time slice. The first four weeks of the lecture period were skipped, because at this time there are not many resources accessible in the course. To identify bipartite subgroups in the time slices, the Biclique Communities method described in section 3.2 was applied with the parameters $a = 3$ and $b = 5$. This means that an uncovered subgroup contains at least a students

and b resources. This choice is reasonable, since most students share some common resources (e.g., lecture slides) and consequently smaller values lead to one big cluster in most of the time slices. Larger values for a and b have the effect that most nodes cannot be clustered because they do not belong to a corresponding $K_{a,b}$ biclique.

The tracing algorithm described in section 3.4 requires a similarity threshold for the matching of subsequent step-groups. This threshold was set to 0.3. This threshold is recommended in Greene, Doyle, and Cunningham (2010) for tracing one-mode clusters, but is also suitable for our method because it allows the detection of merges and splits of dynamic groups but is not too low that matching between step-groups occurs at random.

The results for tracing the bipartite clusters during the lecture period are depicted in Figure 7. For subgroup traces, the swim lane visualization introduced in Halatchliyski et al. (2013) was applied. The horizontal lines connect step-actor and step-resource groups that match between two time steps respectively. The vertical lines connect a step-actor group (circle) with a step-resource group (square), if they belong to the same bipartite cluster in the corresponding week. The node sizes correspond to the contained number of resources or students respectively.
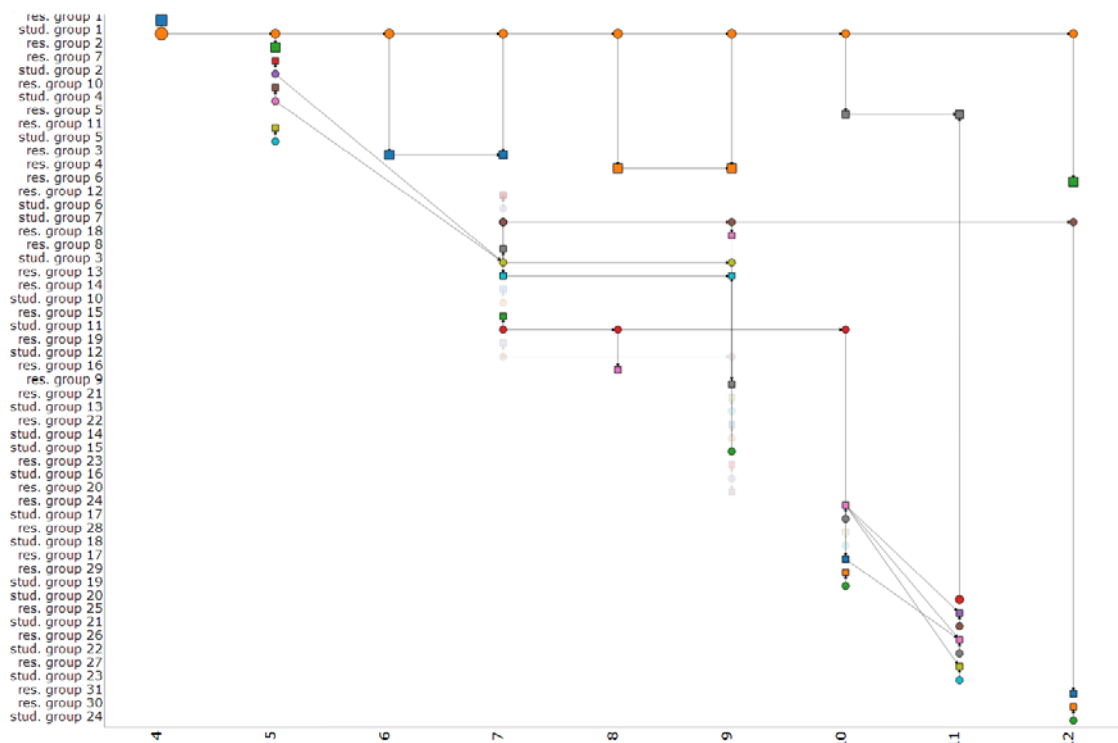

Figure 7: Visualization of the subgroup evolution of the last 9 weeks of the lecture period with interesting traces highlighted.

During the lecture period, there is one large, stable actor group (stud. group 1). This actor group forms different bipartite clusters every week with resource groups containing the core lecture material for the particular week as well as the wiki pages created by the exercise groups. This indicates that the students

of this group are not only affiliated with the core material and the wiki content created by their own group, but also access the content created by others. This can be seen in Figure 6, where the big blue bipartite cluster depicts student group 1 in week 10 and their affiliations to course resources. Many students of the major student group 1 are also members of smaller actor groups at the same time. Not all of the smaller student groups correspond to the exercise groups that collaboratively edit wiki pages. Moreover, the majority of students use the wiki pages across exercise groups, most likely an effect of the peer review exercises. The smaller groups mostly contain students with close relations to the additional learning material, as well as lecture videos. The patterns described in Table 1 correspond to the "stable group" pattern, described in the first row of Table 1. Student group 1 accesses similar sets of resources in weeks 6 and 7 and also in weeks 8 and 9. In these cases, learning resources of the previous week are still of importance. This shows that the peer reviewing and wiki article revising exercises indeed encourage many students to occupy themselves with the course material of previous weeks. However, the patterns of simultaneous merging or splitting of the actor groups and resource groups of bipartite clusters, described in rows 4 and 5 of Table 1, could not be found.

4.1.4 Resource access during the exam preparation phase

In contrast to the lecture period, exam preparation is self-directed. Thus, in this period, students choose the resources they need for their exam preparation based on their own schedule and assessment. Relevant questions regarding exam preparation are:

- Which resources are often used by certain groups of students?

- What are typical traces of actor groups and resource groups in such informal settings?

Our method can help to answer these questions in order to provide students with additional resources before the exam and to identify those combinations of resources considered relevant by the students. Figure 8 depicts the traces of bipartite student resource clusters during the exam phase. The parameter settings for the clustering and the tracing algorithm are the same as in section 4.1.3. However, the size of the time slices is only four days instead of one week as in 4.1.3. This is due to the more dynamic resource access during the exam preparation phase.

There is one big cluster of learning resources, namely resource group 1. This dynamic resource group is relatively stable, which indicates a set of learning resources of particular importance to students for exam preparation. The different step-resource groups of this dynamic group contain primarily lecture slides, especially on the topics of CSCL, CSCW, and Intelligent Tutoring Systems, which are the last lecture topics relevant for the exam. The majority of these step-groups also contain self-test quizzes. A relatively stable actor group with step-actor groups often forms a bipartite cluster with particular instances of resource group 1. This indicates that a large group of students considers the resources within this group most relevant in the days and weeks before the exam.
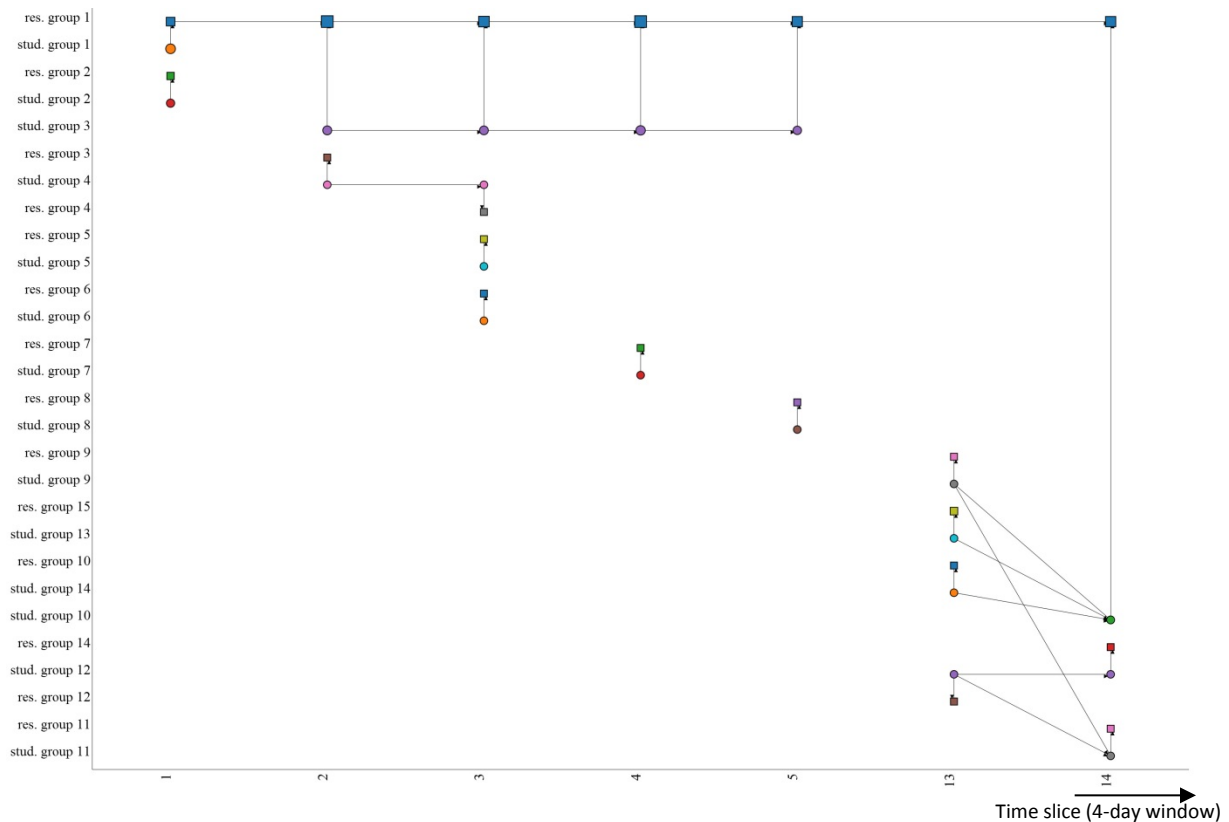
Figure 8: The evolution of bipartite subgroups during the oral examination phase.

As in the lecture period, the found clusters overlap pervasively. Learning resources that occur more than two times in more than one cluster include the following:

- Lecture slides "CSCL"

- Lecture slides "Knowledge Diagnosis"

- Wiki article "Methods and Models for Human Computer Interaction"

- Wiki article "Fitts' Law"

These resources were used by students irrespectively of the other resources they used for exam preparation.

Between points in time 6 and 12, no clusters can be found, meaning that single students only accessed resources sporadically. This can also be observed in Figure 4 between week 15 and 20, which can be explained by the scheduling of oral exams. While most students had their oral exams in the four weeks after the lecture period, a smaller group in another study programme had exams after eight weeks. Students in the second study programme seem to begin exam preparations eight days before the exam at the earliest, which is much less time than students prepare in the first exam period. Regarding exam results, students from the first group tended to get better grades (1.3 on average) than students from the second group (2.1 on average). Furthermore, the resources accessed by the second group were different at the beginning of their active phase. Instead of a dominating group of students who mainly accessed the lecture slides for exam preparation, as did the students in the beginning of the exam phase, the active students in the late phase of the exam period can be clustered into small overlapping

groups affiliated either with wiki articles or lecture videos (see Figure 9). In the last time slice, most of the students merge to a single group, but instead of forming a bipartite cluster with mostly wiki and video resources, which would correspond to the merging pattern of the 4th column of Table 1, they refocus their learning interest on the "core" preparation material (res. group 1), as can be seen in Figure 9.
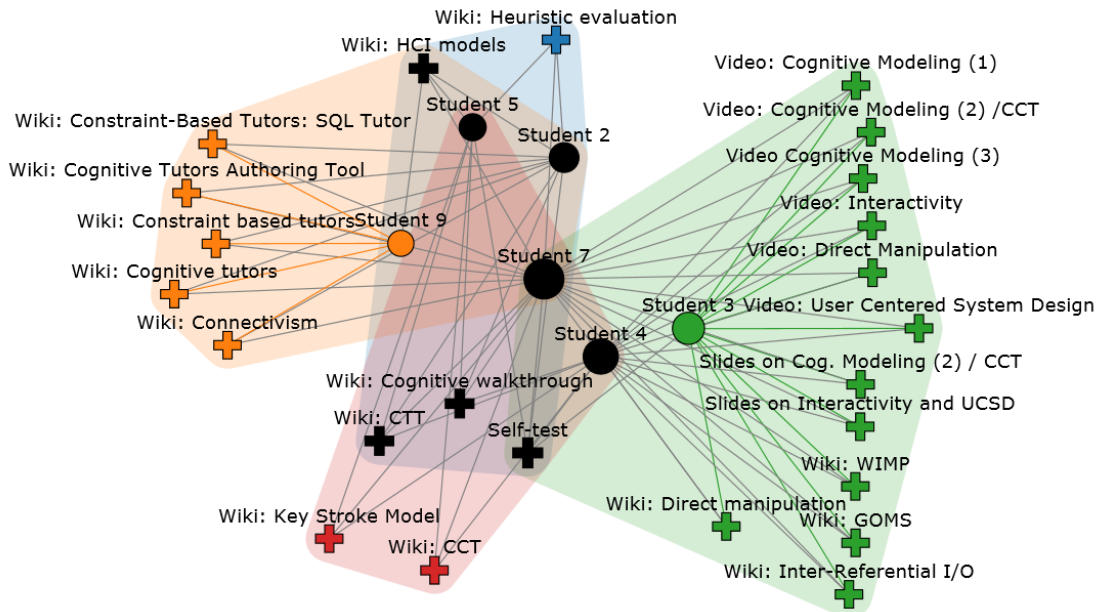

Figure 9: The student–resource clusters in time slice 13 of the exam preparation phase.

4.2 The Computer-Mediated Communication MOOC

The course "Computer-mediated communication" (CMC) was originally designed as a joint blended learning seminar for students from the courses "Applied Communications and Media Science" at the University of Duisburg-Essen and "Pedagogics" at the University of Bochum. In the winter term 2013/14, the obligatory presence sessions were replaced and the course on computer-mediated communication was redesigned as a MOOC, which was also open to participants outside the two universities. With 173 participants in total and a maximal number of 134 active participants per week, the course was not really "massive," but it had other characteristics of a MOOC, such as video based lectures and extensive online activities. As in the GILLS course, Moodle was the chosen platform for online activities. A distinctive feature of this course was that regular Master's level students from "Applied Communications and Media Science" at the University of Duisburg-Essen and "Pedagogics" at the University of Bochum had the opportunity to receive regular credits for successful completion of the course. Videos were one important type of learning resource with video messages introducing the main course objectives for each week as well as videos of expert talks and interviews. Additionally, weekly reading assignments included a variety of scientific literature. There were also self-tests and three workshops. For workshop assignments, students had to solve tasks in small groups. In contrast to the GILLS course described in section 4.1, students were not supposed to build groups themselves. Instead, they were grouped automatically by the system.

Figure 10 depicts the number of active participants per week. The Christmas gap in course activity occurs between weeks 10 and 12. Course activity shows a decrease over time during the progress of the course. This effect is commonly observed in MOOCs (Clow, 2013); however, in this course the decline of active users is not as steep as in large MOOCs since only 14 days of inactivity were tolerated in order to be allowed to participate in the final exam and to receive course credit.
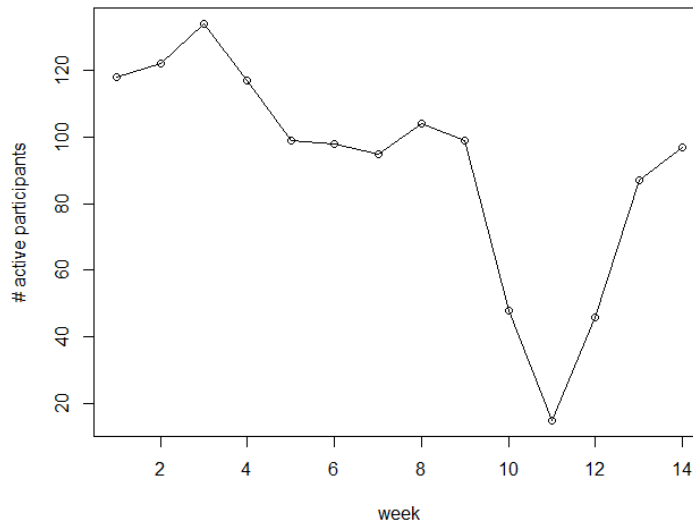

Figure 10: Number of active participants in the CMC MOOC.

4.2.1 Resource access during the course period
As in the analysis of the lecture period of the GILLS course (section 4.1), the bipartite student-resource network was sampled into time slices corresponding to each week of the lecture. Since the CMC dataset contains by far more students but fewer learning resources compared to the GILLS course analysed in section 4.1, the biclique percolation algorithm was applied to the CMC dataset with a different parameterization. The clusters were constructed from $K_{5,4}$ bicliques. This means that only students who used at least four different resources and resources used by at least five different students can appear in a bipartite cluster.

Figure 11 depicts the trace of the identified clusters during the time of the course. Similar to the GILLS course, there is a large relatively stable group of students who accessed a common set of resources over a longer period of time (see student group 1). However, especially in the beginning, many fewer clusters can be found in comparison to the blended learning course. In the first three weeks of the course, the resource access behaviour of active students was very stable such that only one student-resource cluster can be identified. In addition, the set of resources accessed during the first three weeks is very stable. The content analysis of the resources in resource group 1 in the first three weeks of the lecture period reveals that the students in this phase are mainly focused on videos and reading material on general theories of computer-mediated communication. This corresponds to the continuous interest pattern described in the first row of Table 1. The fact that a large group of students used similar resources over several weeks, which leads to the re-identification of a resource group by the proposed algorithm, repeats also in preceding weeks. Thus, a characteristic of the course is that the participants

53

deal with the learning material from previous weeks instead of focusing exclusively on the resources of the current week.
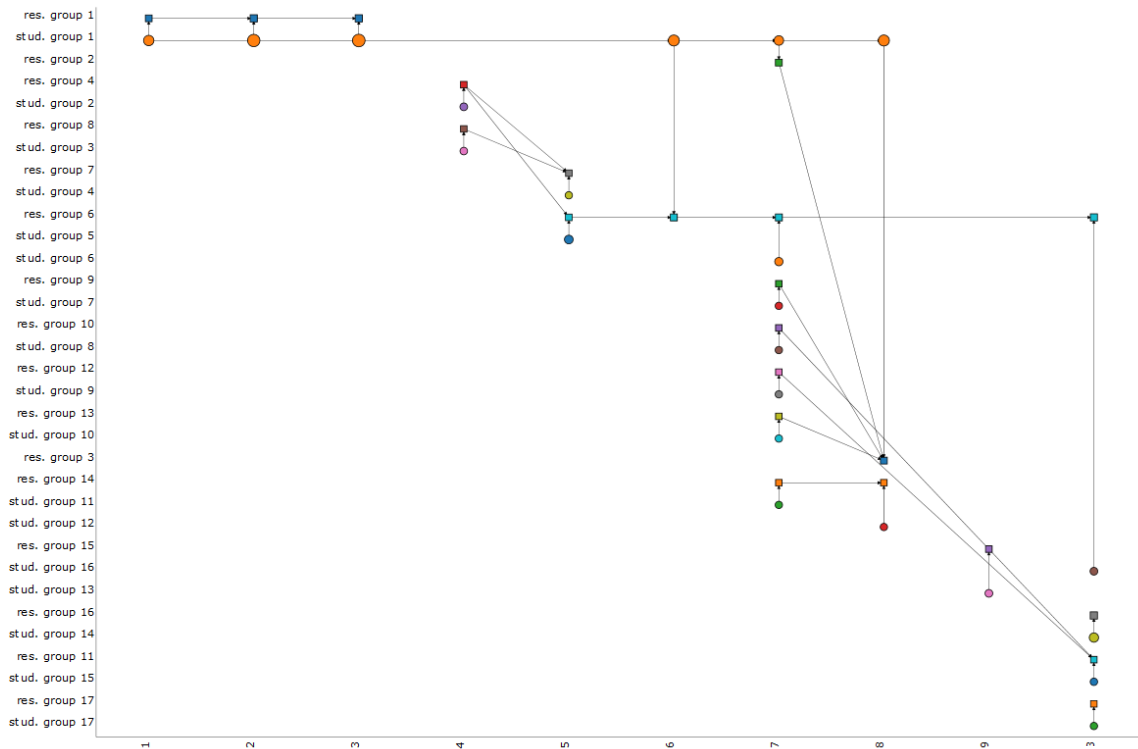

Figure 11: Visualization of the subgroup evolution CMC MOOC.

As shown in Figure 10, there is a continuous decline in the number of active students between weeks 2 and 7, which is also reflected in the swim lane diagram in Figure 11. In weeks 4 and 5, the number of resources used by each student is very low. Many of the students cannot be assigned to clusters because they do not access enough resources. This circumstance will be discussed in more detail in the next section, 4.2.2. In this phase, the large student group 1 disappears. In week 4, only two relatively small and overlapping student-resource clusters can be identified. The topic of the week was about trust and group work. The first cluster, comprising student group 2 and resource group 4, can be labelled with the keywords "distributed groups," "trust," and "CMC." The other cluster includes students and mainly resources on general topics in CMC. Resource group 4 splits into two other resource groups in week 5, namely resource groups 6 and 7. Resource group 7 is labelled again with the keyword "CMC" and resource group 6 with "distributed groups" and "trust."

Whereas in the first 6 weeks the evolution of the student-resource clusters could be characterized as stable and not very diverse, in week 7 the number of identified clusters increased and the clusters became much more diverse in terms of the resource groups. This may result from the "assignment of the week" where students were asked to create their own exam questions and therefore needed to review the material of the previous weeks. There is also a correspondence to the number of active participants depicted in Figure 10. Beginning with week 7, course activity rose slightly again. This was exactly the point in time where the clustering algorithm detected a higher number of clusters and, at the

same time, less stable clusters. Apart from the special course assignment, another reason for this is that in this late phase of the course, some participants were already preparing for the final exam and therefore accessed additional learning resources apart from the course material of the corresponding week. This can be seen especially in week 13, the last week of the course.

4.2.2 Further analysis of the clusters

As in the results for the GILLS course, the biclique percolation community-detection-method identifies highly overlapping clusters of students and resources. Figure 12 depicts a snapshot of the student-resource clusters of the 7th week of the lecture period where resource access was at its most diverse. It is worth mentioning that this is only a subset of all active students in this week. Since the clusters were built from $K_{5,4}$ bicliques, the clusters contain only those students who were active enough to appear in at least one student-resource cluster. In this case, clustered students have connections to a minimum of four resources and clustered resources to at least five students. Nodes that belong to more than one cluster are colored in black. Hence, even if six clusters were detected, only four can be visually distinct by node colour because two are completely overlapping with one or more of the others. Hence, the resource access behaviour of students in different clusters is not completely distinctive. Moreover, a set of core learning resources are of interest to the majority of the learners in the course.

The size of a node corresponds to the number of its connections, which is a useful visual representation to combine connection statistics with the clustering approach. This shows clearly that the students who access many different resources and the resources that attract many students can be found in the overlaps between two or more clusters (black nodes). Nearly every active student is interested in the quiz on expert–layman communication, the video message of week 7, and the corresponding reading material independently from the cluster allocation. An overlapping subset of students has additional affiliations to given video messages.
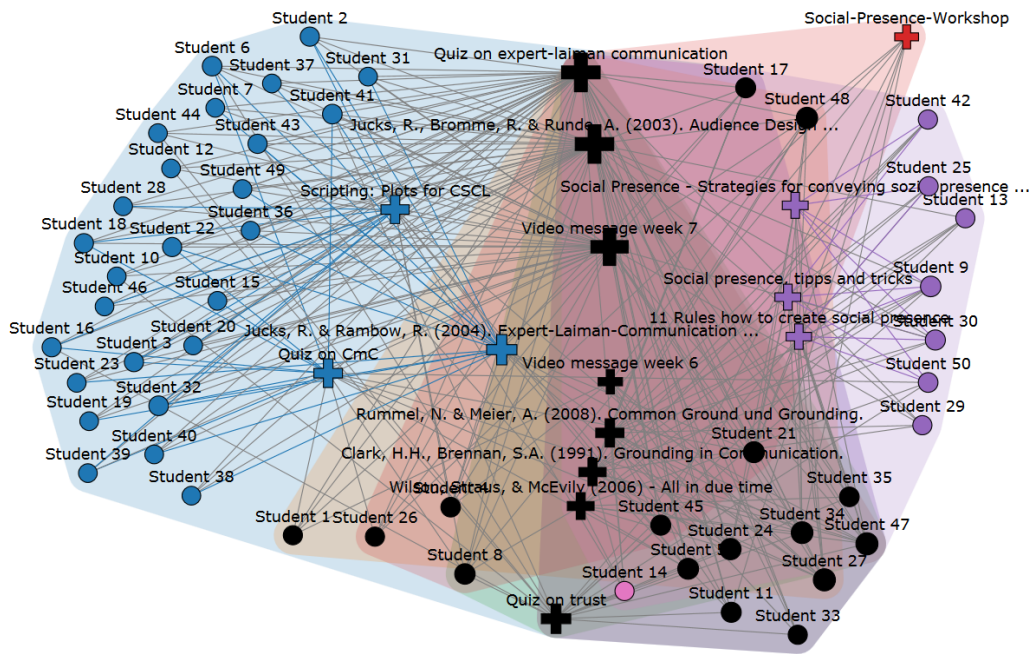

Figure 12: Student-resource clusters in the 7th week of the course.

In order to analyze the fraction of students and resources that could be assigned to bipartite clusters and the fraction of nodes that appear in the overlap between clusters, Table 3 lists basic statistics of the student-resource cluster of each week.

Table 3: Statistics of the clusters in each week of the CMC course.

Week	# Active students	# Students in clusters	# Students in overlaps	# Resources used	# Resources in clusters	# Resources in overlaps
1	118	52	0	12	8	0
2	122	92	0	17	14	0
3	134	97	0	18	14	0
4	117	16	2	20	6	3
5	99	27	2	17	6	3
6	98	64	0	23	11	0
7	95	50	16	28	15	8
8	104	67	3	26	11	2
9	99	10	0	20	4	0
10	48	0	0	10	0	0
11	15	0	0	9	0	0
12	46	0	0	34	0	0
13	87	39	10	41	20	9

This shows that whenever more than one cluster could be detected, there were always a large number of students and resources in the overlap of different clusters. Although large overlaps of the clusters are found in particular weeks of the course and the students often reuse material from previous weeks (section 4.2.1), only a few learning resources occur in cluster overlaps in multiple weeks. These are mainly workshop resources and scientific literature on trust in computer-mediated group work.

Regarding the number of students and resources that could be assigned to clusters, it is interesting to see that even if there were around 100 different learners visiting the course in weeks 4 and 5, only a minority were part of a $K_{5,4}$ biclique. That means that the majority of students accessed only a few resources in this phase. By looking at the swim lane diagram in Figure 11 in conjunction with Table 3, one can see that in weeks where a large percentage of students were accessing enough resources to be allocated to a cluster, the algorithm detects only one big student-resource cluster. This means that when the quantity of different resources used became more diverse, the concrete resource usage of the students also became more diverse, such that the algorithm detected more clusters.

5 CONCLUSION

The analysis of the two online courses using the proposed method has provided some useful insights concerning the patterns of resource usage by students. The courses, namely the GILLS course and the MOOC on computer-mediated communication (CMC), not only differ in openness and number of participants but also in course design. The GILLS course was conducted as a blended learning course with presence lectures but extensive online activities oriented towards content production in the form of wiki articles peer reviewed by other participants. In the CMC MOOC, reading assignments and self-tests were the main tasks for participants. Since, in the CMC course, there was no break between the

lecture and the exams, exam preparation started in parallel with the lecture. This also led to different resource access behaviour in the two courses.

In the GILLS course, the majority group accessed the lecture slides assigned to the topic of each particular week and the wiki articles written by other students. The continuous interest in wiki articles created by other students indicates that the peer review concept does work appropriately.

As described in section 4.1, students could earn bonus points for the exam when they did the group exercises, namely writing and reviewing wiki articles. Our findings that the articles of the glossary wiki were used across exercise groups and in addition to the lecture slides for reworking the lecture indicates that the self-generated content is important to students. Some groups of students who often overlap with the major student group are also interested in additionally provided scientific articles or the lecture videos. Lecture videos are often found in small bipartite student-resource clusters as in Figure 6. Thus, they seem to be used on demand by small groups of students as additional learning resources or to rework a presence lecture they had not attended.

During the exam preparation phase, we discovered a set of core learning resources accessed by a more or less stable group of students. By investigating the overlaps between the student–resource clusters, it is possible to identify a small set of learning material also used by more than one group of students, see section 4.1.4. For future instances of the GILLS course, the course designers should further investigate the role of these learning materials, particularly for exam preparation.

As described in section 4.1.1, students tend to work with others in their own study programme. In section 4.1.4, the analysis shows that students whose oral exams were scheduled later than those of other programmes also had different patterns of resource access during the exam preparation phase. They began preparation late and focused more on lecture videos and wiki articles than other material. In future, we can tailor the course more specifically for those distinct learner types.

In comparison to the GILLS course, the student-resource clustering in the CMC MOOC yielded fewer clusters. Since the MOOC had not such a strong emphasis on group work, the students seemed to focus more on the same resources. Another interesting insight gained by the tracing of bipartite clusters during the time of the CMC course was that the continuous interest pattern (first row of Table 1) was dominant in the first weeks of the course. This clearly indicates that the resources provided were important to students over several weeks during the lecture period. It is interesting to see that from week 7 on the resource access of course participants became more diverse as the course activity started to rise again after weeks of decline. By a deeper investigation of the clusters, it could be shown that there were periods in the course when students tended to access only a few resources and hence could not be allocated to a cluster. This is a good example of the ability of the proposed method to focus on the most active users within the MOOC. In more active periods, large numbers of students and resources are allocated to more than one cluster.

From a methodological point of view, this paper presented a modification of the subcommunity tracing method for one-mode social networks described in Greene, Doyle, and Cunningham (2010), so that it can be used to trace clusters in bipartite networks. The utility of this kind of analysis for the discovery of characteristic properties of the relations between students and learning resources was demonstrated. A major advantage of the proposed analysis in contrast to purely descriptive analysis of resource access is that one can identify not only popular and hence potentially useful resources, but also who is affiliated

with the resources, and how such affiliations change over time. This allows for the identification of common student interest in resources.

One can think of several other applications of the method in different learning environments: In cMOOC-like environments or open learning platforms such as Khan Academy[2] where users are more or less free in the resources they choose, the tracing of dynamic learner-resource networks can help to identify learners with similar learning traces and similar interests. The results could be used for peer recommendation and the formation of learning groups (cf. Harrer et al., 2007). In blended learning scenarios, where learning activities take place in real-world as well as virtual environments, student groups based on resource access that are stable over time can be an indicator for group work in the real world. Research on this can help to develop tools for better course management in such scenarios. The described method not only provides structural information about learners and their relation to learning resources, it also reflects the underlying learning processes. In environments with predefined resource access, such as sequential lectures or scripted inquiry learning settings, the method can help to verify if students use resources as intended. Unintended usage can be detected and may trigger corresponding adjustments on the part of instructors.

ACKNOWLEDGEMENTS

REFERENCES

Asur, S., Parthasarathy, S., & Ucar, D. (2009). An event-based framework for characterizing the evolutionary behavior of interaction graphs. *ACM Transactions on Knowledge Discovery from Data, 3*(4), 16:1–16:36.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. *Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining* (pp. 44–54)*,* Philadelphia, PA: ACM.

Belanger, Y., & Thornton, J. (2013). Bioelectricity: A quantitative approach Duke University's first MOOC. Technical Report, Duke University, NC (2013).

Bruff, D. O., Fisher, D. H., McEwen, K. E., & Smith, B. E. (2013). Wrapping a MOOC: Student perceptions of an experiment in blended learning. *Journal of Online Learning & Teaching, 9*(2), 187–199.

Clow, D. (2013). MOOCs and the funnel of participation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 185–189)*,* Leuven, Belgium: ACM.

Fini, A. (2009). The technological dimension of a massive open online course: The case of the CCK08 course tools. *The International Review of Research in Open and Distance Learning, 10*(5). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/643/1402.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3), 75–174.

---

[2] https://www.khanacademy.org/

58

Fox, A. (2013). From MOOCs to SPOCs. *Communications of the ACM, 56*(12), 38–40.

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education, 7*(2), 95–105.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*(12), 7821–7826.

Greene, D., Doyle, D., & Cunningham, P. (2010). Tracking the evolution of communities in dynamic social networks. *Proceedings of the International Conference on Advances in Social Network Analysis and Mining* (pp. 176–183)*,* Odense, Denmark: IEEE.

Grünewald, F., Meinel, C., Totschnig, M., & Willems, C. (2013). Designing MOOCs for the support of multiple learning styles. In D. Hernández-Leo, T. Ley, R. Klamma, & A. Harrer (Eds.), Scaling up Learning for Sustained Impact, Lecture Notes in Computer Science 8095 (pp. 371–382). Springer Berlin Heidelberg.

Halatchliyski, I., Hecking, T., Göhnert, T., & Hoppe, H. U. (2013). Analyzing the flow of ideas and profiles of contributors in an open learning community. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 66–74)*,* Leuven, Belgium: ACM.

Harrer, A., Malzahn, N., Zeini, S., & Hoppe, H. U. (2007). Combining social network analysis with semantic relations to support the evolution of a scientific community. *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning* (pp. 270–279)*,* New Brunswick, NJ: ISLS.

Hill, J., & Hannafin, M. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*(3), 37–52.

Hoppe, U., Pinkwart, N., Oelinger, M., Zeini, S., Verdejo, F., Barros, B., & Mayorga, J. I. (2005). Building bridges within learning communities through ontologies and thematic objects. *Proceedings of the Conference on Computer Support for Collaborative⸺ Learning* (pp. 211–220), Taipei, Taiwan: ISLS.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 170–179)*,* Leuven, Belgium: ACM.

Lehmann, S., Schwartz, M., & Hansen, L. K. (2008). Biclique communities. *Physical Review E, 78*(1), 016108.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: Densification laws, shrinking diameters and possible explanations. *Proceedings of the Eleventh International Conference on Knowledge Discovery in Data Mining* (pp. 177–187)*,* Chicago, IL: ACM.

Nachmias, R., & Segev, L. (2003). Students' use of content in web-supported academic courses. *The Internet and Higher Education, 6*(2), 145–157.

Pahl, C., & Donnellan, D. (2002). Data mining technology for the evaluation of web-based teaching and learning systems. *Seventh International Conference on E-Learning in Business, Government and Higher Education* (pp. 1–7)*,* Montreal, Canada*.*

Palla, G., Barabasi, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*, 664–667.

Palla, G., Derenyi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature, 435*, 814–818.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering, 21*(6), 759–772.

Rodriguez, O. (2013). The concept of openness behind c and x-MOOCs (massive open online courses). *Open Praxis, 5*(1), 67–73.

Romero, C., Gutiérrez, S., Freire, M., & Ventura, S. (2008). Mining and visualizing visited trails in web-based educational systems. *First International Conference on Educational Data Mining* (pp. 182–186), Montréal, Canada.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368–384.

Treeratpituk, P., & Callan, J. (2006). Automatically labeling hierarchical clusters. *Proceedings of the 2006 International Conference on Digital Government Research,* San Diego, CA. 167–176.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (1st ed.) Cambridge: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440–442.

Ziebarth, S., Malzahn, N., & Hoppe, H. U. (2012). Matchballs: A multi-agent-system for ontology-based collaborative learning games. *Proceedings of the 18th International Conference on Collaboration and Technology* (pp. 208–222)*,* Raesfeld, Germany: Springer.

Ziebarth, S., & Hoppe, H. U. (2014). Moodle4SPOC: A resource-intensive blended learning course. *Proceedings of the 9th European Conference on Technology Enhanced Learning* (pp. 259–372)*, Graz, Austria*: Springer.