

# Chapter 2: Computational Methods for the Analysis of Learning and Knowledge Building Communities

H. Ulrich Hoppe

Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen, Germany

DOI: 10.18608/hla17.002

## ABSTRACT

Learning analytics (LA) features an inherent interest in algorithms and computational methods of analysis. This makes LA an interesting field of study for computer scientists and mathematically inspired researchers. A differentiated view of the different types of approaches is relevant not only for “technologists” but also for the design and interpretation of analytics applications. The “trinity of methods” includes analytics of 1) *network structures* including actor–actor (social) networks but also actor–artefact networks, 2) *processes* using methods of sequence analysis, and 3) *content* using text mining or other techniques of artefact analysis. A summary picture of these approaches and their roots is given. Two recent studies are presented to exemplify challenges and potential benefits of using advanced computational methods that combine different methodological approaches.

**Keywords:** Trinity of computational methods, knowledge building, actor–artefact networks, resource access patterns

The newly established field of learning analytics (LA) features an inherent interest in computational or algorithmic methods of data analysis. In this perspective, “analytics” is more than just the empirical analysis of learning interactions in technology-rich settings, it actually also calls for specific computational and mathematical approaches as part of the analysis. This line of research builds on techniques of data mining and network analysis, which are adapted, specialized, and potentially developed further in an LA context.

To better understand the potential and challenges of this endeavor, it is important to introduce some distinctions regarding the nature of the underlying methods. Computational approaches used in LA include analytics of 1) *network structures* including actor–actor (social) networks but also actor–artefact networks, 2) *processes* using methods of sequence analysis, and 3) *content* using text mining or other techniques of computational artefact analysis. This distinction is not only relevant for “technologists” who actually work with and on the computational methods, it is also important for the design of “LA-enriched” educational environments and scenarios. We should not expect LA to develop new computational-analytic techniques from scratch but to adapt and possibly extend existing approaches

in an LA context. First, the different premises and affordances of the different types of methods should be well understood. Furthermore, the combination and synergetic use of different types of methods is often desirable from an application perspective but this constitutes new challenges from a conceptual as well as a computational point of view.

The computational analysis of interaction and communication in group learning scenarios and learning communities has been a topic of research even before the field of LA was constituted, and this work is still relevant to LA. Early adoptions of social network analysis (SNA type 1) in this context include the works of Haythornthwaite (2001), Reffay & Chanier (2003), Harrer, Malzahn, Zeini, and Hoppe (2007), and De Laat, Lally, Lipponen, and Simons (2007). Process-oriented analytics techniques (type 2) have an even longer history, especially in the analysis of interactions in a computer-supported collaborative learning (CSCL) context (Mühlenbrock & Hoppe, 1999; Harrer, Martínez-Monés, & Dimitracopoulou, 2009). Although somewhat later and less numerous, content-based analyses (type 3) based on computational linguistics techniques have been successfully applied to the analysis of collaborative learning processes (e.g., by Rosé et al., 2008).

## Network-Analytic Methods

Network-analytic approaches, especially social network analysis (SNA), are characterized by taking a relational perspective and by viewing actors as nodes in a network, represented as a graph structure. In this sense, a network consists of a set of actors, and a set of ties between pairs of actors (Wasserman & Faust, 1994). The type of pairwise connection defines the nature of each social network (Borgatti, Mehra, Brass, & Labianca, 2009). Examples of different types of ties are affiliation, friendship, professional, behavioural interaction, or information sharing. The visualization of such network structures has emerged as a specific subfield (Krempel, 2005). Standard methods of network analysis allow for quantifying the importance of actors by different types of “centrality measures” and detecting clusters of actors connected more densely among each other than the average (detection of “cohesive subgroups” or “community detection” – for an overview, see Fortunato, 2010).

A well-known inherent limitation of SNA is that the target representation, i.e., the social network, aggregates data over a given time window but no longer represents the underlying temporal dynamics (i.e., interaction patterns). It has been shown that the size of the time window of aggregation has a systematic influence on certain network characteristics such as subcommunity structures (Zeini, Göhnert, Hecking, Krempel, & Hoppe, 2014). To explicitly address time-dependent effects, SNA techniques have been extended to analyzing time series of networks in dynamic approaches.

It is important to acknowledge that network analytic techniques (even under the heading of SNA) do not exclusively deal with actors and social relations as basic elements. So-called “affiliation networks” or “two-mode networks” (Wasserman & Faust, 1994) are based on relations between two distinct types of entities, namely actors and affiliations. Here, the “affiliation” type can be of a very different nature, including, for example, publications as affiliations in relation to authors as actors in the context of coauthoring networks. In general, two-mode networks can be used to model the creation and sharing of knowledge artefacts in knowledge building scenarios. In pure form, these networks are assumed to be bipartite, i.e., only alternating links actor–artefact (relation “created/modified”) or artefact–actor (relation “created-by/modified-by”) would be allowed. Using simple matrix operations, such bipartite two-mode-networks can be “folded” into homogeneous (one-mode) networks of either only actors or only artefacts. Here, for example, two actors would be associated if they have acted upon the same artefact. We would then say that the relation between the actors was mediated by the ar-

tefact. Similarly, we can derive relationships between artefacts by considering agents (one actor engaged in the creation of two different artefacts) as mediators.

We have seen an increasing number of studies of educational communities using SNA techniques related to networked learning and CSCL. Originally, networks derived from email and discussion boards were the most prominent conditions studied, as for example the early study of cohesion in learning groups (Reffay & Chanier, 2003). Meanwhile, network analysis belongs to the core of LA techniques. The classification of approaches to “social learning analytics” by Ferguson and Shum (2012), though not primarily computationally oriented, prominently mentions network analysis techniques including both actor–actor and actor–artefact networks.

## Process-Oriented Interaction Analysis

The computational analysis of learner (inter-)actions based on the system’s logfiles has a tradition in CSCL. There were even attempts to standardize action-logging formats in CSCL systems to facilitate the sharing and combination of existing interaction analysis techniques (Harrer et al., 2009). One of the earliest examples of applying intelligent computational techniques in a CSCL context (namely sequential pattern recognition) was suggested and exemplified by Mühlenbrock and Hoppe (1999). This approach was later used in an empirical context to pre-process CSCL action logs in order to automatically detect the occurrence of certain collaboration patterns such as “co-construction” or “conflict” (Zumbach, Mühlenbrock, Jansen, Reimann, & Hoppe, 2002).

Whereas these approaches were developed in a learning-related research context, there are also more general techniques that can be adapted and used, such as the scalable platforms management forum (SPMF) and library of sequential patterns mining methods (Fournier-Viger et al., 2014). In an LA context, SPMF is used by the LeMo tool suite for the analytics of activities on online learning platforms (Elkina, Fortenbacher, Merceron, 2013). In another recent study, Bannert, Reimann, & Sonnenberg (2014) have used “process mining,” a computational technique with roots in automata theory, to characterize patterns and strategies in self-regulated learning.

## Content Analysis Using Text-Mining Methods

There is a tradition of content-analysis-based human interpretation and coding often used as input to quantitative empirical research, as discussed by Srijbos, Martens, Prins, and Jochems (2006) from a CSCL perspective. In contrast, from a computational point of view, we are interested in applying informa-

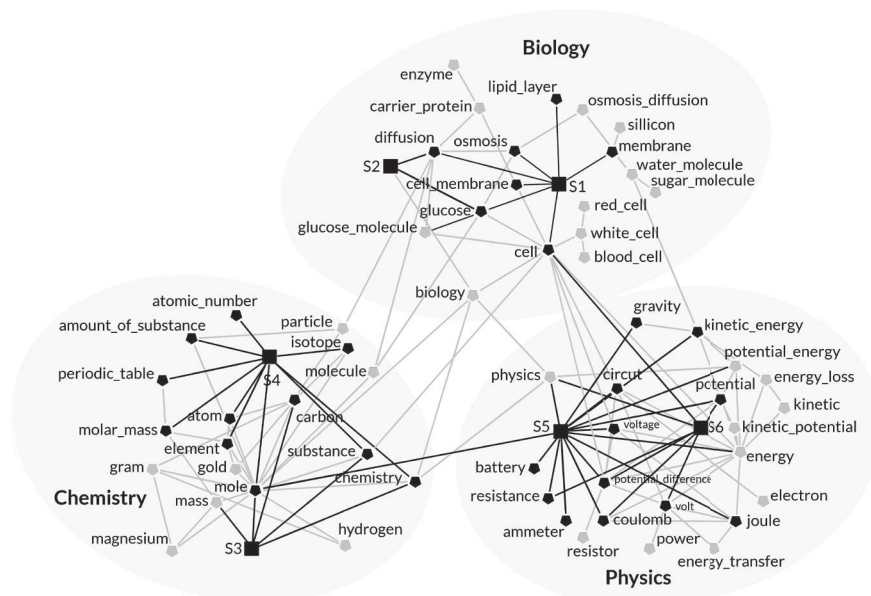
tion-mining techniques to extract semantic information from artefacts. Obviously, this is of particular interest in the case of learner-generated artefacts. Rosé et al. (2008) has demonstrated the usefulness of automatic text classification with a corpus of CSCL transcripts. Sherin (2013) used computational techniques of content analysis on student interview data to discover the students' understanding of science concepts. Content analysis techniques have also been used for the clustering of e-learning resources according to their similarity (Hung, 2012). He (2013) proposed the usage of similar techniques for grouping learners' main topics in student-to-teacher online questions and peer-to-peer chat messages in the context of online video-based learning.

Typically, these methods of textual content analysis are based on the “bag of words” model in which the given order of words in a text is of no relevance to the analysis. This is the case for a variety of probabilistic topic modelling techniques such as the currently quite popular method of latent Dirichlet allocation (LDA; Blei, 2012). A method that does take into account the positioning of words in a text is network text analysis (NTA). NTA is a text mining method that connects content analysis with network representations in that it extracts a network of concepts from given texts (Carley, Columbus, & Landwehr, 2013). Links between concepts are established if the corresponding terms co-occur with a certain frequency in a sliding window of pre-specified width that runs over a normalized version of the text. A “meta thesaurus” allows for introducing different concept categories (e.g., “person,” “location,” “domain\_concept” et cetera). On this

basis, multimode networks can be formed, in which the concept–concept relations are restricted to certain inter-category types such as *location–person* or *person–domain\_concept*. These representations can in turn be analyzed using network-analytic concepts such as centrality measures or the detection of cohesive subgroups as a network-based clustering technique.

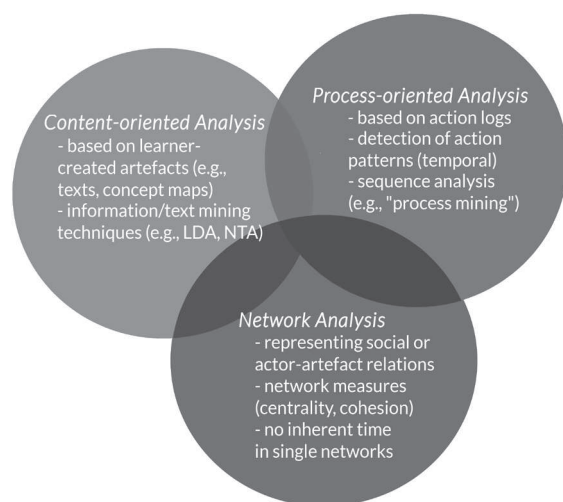
Figure 2.1 shows the result of applying NTA to transcripts from teacher–student workshops in the context of the European project JuxtaLearn (Hoppe, Erkens, Clough, Daems, & Adams, 2013). The resulting networks nicely reflect the different topics from the areas of biology, chemistry, and physics, initially presented by students and then discussed in the whole group. Here *topics* (pentagon-shaped nodes) and *topic–topic* relations are depicted in grey, whereas *persons* (square nodes) and *person–topic* relations are darker (black).

In the topic–topic network, the three different fields of science appear as more densely connected islands (or “cohesive subgroups”) although certain cross-links exist (e.g., *diffusion* in biology is linked to *molecule* in chemistry). The person–topic links allow for judging the importance of the individuals' contributions in the presentations and discussion. Most contributions stay within one subfield. Student S5 stands out in terms of degree centrality (14 connections to different topics) and with most contributions to physics but one link bridging over into chemistry. In the JuxtaLearn project, this approach has been further developed to assess students' problems of understanding from question/answer collections related to science videos (Daems, Erkens, Malzahn, & Hoppe, 2014). The extracted net-



**Figure 2.1.** Topic–topic and person–topic relations extracted from transcripts of teacher–student workshops.





**Figure 2.2.** The “trinity” of methodological approaches.

works of concepts have been contrasted with teacher-created taxonomies. This has led to an enrichment of the taxonomies and to the identification of specific problems of understanding on the part of the learners. From a pedagogical perspective, this provides empirical insights relevant to curriculum construction and curriculum revision (here specifically related to teacher-created micro-curricula).

Figure 2.2 summarizes the characteristics of the three methodological approaches in terms of their basic representational characteristics and typical techniques. Overlapping areas between the approaches are of particular importance for new integrative or synergetic applications.

The remainder of this article presents two case studies of applying specific computational techniques to the analysis of learning and knowledge building in communities. The first example shows that more sophisticated methods of network analysis may yield interesting insights in a case where “first order approaches” would fail to resolve interesting structures. In that it considers the evolution of patterns of resource access on a learning platform over time, it combines the network analytic approach with process aspects. The second case describes the adoption and revision of a scientometric method to characterize the evolution of ideas in a knowledge-building community. This network-analytic approach is then combined with content-based text mining methods. So, both examples support the general point that we can expect additional benefit from combining different methods.

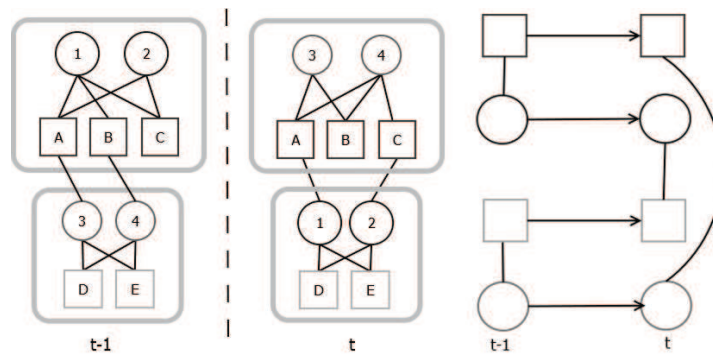
### **Example 1: Dynamic Resource Access Patterns in Online Courses**

Nowadays higher education practice is commonly supported by learning platforms such as Moodle to

distribute educational materials of various types, including lecture slides, videos, and task assignments, but also to collect exercise or quiz results and to facilitate individual or group work using forums or wikis. In this way, classical presence lectures are turned into blended learning courses or, according to Fox (2013), “small private online courses” (SPOCs). As for the traces that learners leave on such learning platforms, the most abundant actions are resource access activities that constitute actor (learner) – artefact (learning resource) relations. Only in special cases, such as the co-editing of wiki articles, such data may be interpreted in a quite straightforward way as actor–actor relations by “folding away” the mediating artifact (i.e., inter-connecting co-authors of the same wiki article). If applied to instructor-provided lecture materials, the actor–actor relation based on access to the same lecture would not be selective and would result in a dense network. Accordingly, the detection and tracing of clusters or subcommunities in such induced actor–actor networks would not be likely to provide interesting insights.

In a study based on one of the author’s regular master courses (Hecking, Ziebarth, & Hoppe, 2014), a more sophisticated technique has been used to overcome this problem. Applying a subcommunity detection algorithm for two-mode networks to the original learner-resource data leads to much more selective and differentiated results in terms of identifying groups of learners working with certain groups of materials in a given time slice. This approach is based on the network-analytic method of “bi-clique percolation analysis” (Lehmann, Schwarz, & Hansen, 2008), which is a generalization of the clique percolation method originally defined for one-mode networks (Palla, Derenyi, Farkas, & Vicsek, 2005). The clique percolation method (CPM) builds subcommunities on the presence of cliques (fully connected subgraphs) in one-mode networks. CPM is of particular interest for the analysis of collaborating communities because the resulting clusters may overlap and thus can also be used to identify potential brokers or mediators between different subcommunities. This characteristic also holds for the bi-clique percolation method (BCPM) with two-mode networks. In their original article, Lehmann et al. (2008) identified the higher selectivity of subcommunity in the two-mode network. We have been able to corroborate this in our application case.

Figure 2.3 shows how, on principle, BCPM can be used to trace cohesive clusters in two-mode networks. First, BCPM is applied to each time slice of the network (left-hand side). The diagram on the right abstracts from the individual entities and just depicts and inter-links between groups of actors (squares) and groups of resources (circles). In one particular time slice, two



**Figure 2.3.** Evolving two-mode clusters (left) and the corresponding swim lane diagram (right).

groups of different types are linked by a vertical edge, indicating that these two groups form a bipartite cluster. Horizontal edges appear across time slices and link two groups of the same type, indicating that the two groups can be considered as similar. Here, we see a situation where the connection between actors and resources is switched from one time slice to other. In general, it is not clear if the basic groups “survive” from one time slice to the other (as is the case here). Palla, Barabasi, and Vicsek (2007) have defined a complex system of transformations (such as “birth,” “merger,” “split” et cetera) that can be used to trace the evolution of subgroups over time.

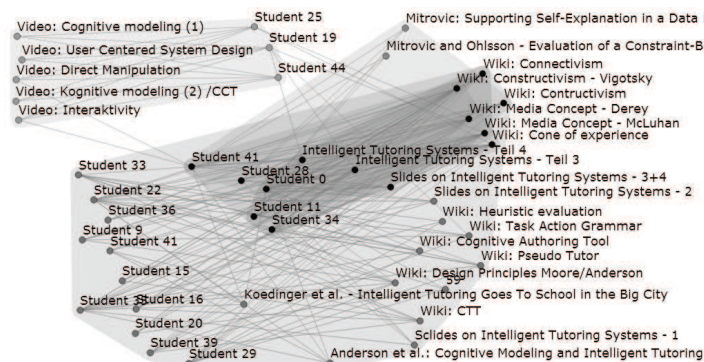
In our study (Hecking et al., 2014), affiliation networks were built based on students’ access to learning resources during a blended learning course on interactive learning and teaching technologies. The course was resource intensive in the sense that the traditional lecture was accompanied by a variety of additional learning resources like lecture videos, slides, serious games, as well as a glossary of important concepts created by the students themselves as a wiki. Students and resources were simultaneously grouped into mixed and overlapping clusters as explained above. Those clusters can be interpreted as a group of students who have a common interest in a group of learning resources but not necessarily having social connections. A typical example cluster is depicted in

Figure 2.3.

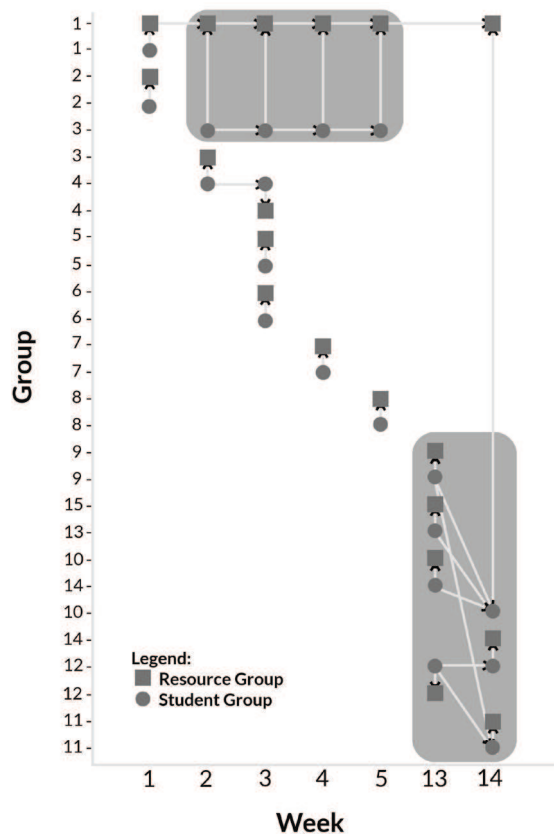
By applying the method to the student–resource networks of particular weeks during the lecture period, this analysis reflects certain groupings induced by explicit assignments but also yields some surprising insights regarding the usage materials. This can be seen, for example, in Figure 2.4 where the orange coloured cluster comprises lecture videos and students who seem to have a distinct interest in learning resources compared to the others.

In addition, the tracking of bipartite student–resource clusters was used to investigate student resource-access behaviour during exam preparation after the last lecture. This period is particular interesting because by then the entire pool of learning materials, successively added week by week to the course, was available, including the wiki articles created by the students.

The swim lane diagram in Figure 2.5 depicts the resource access patterns found in the course during this phase. Time slices were build based on a time window size of 4 days. The oral exams were distributed over two weeks for most of the students, while for another study program the examination phase began six weeks after the last lecture. One finding is that a large majority of students accessed large portions of the learning material over several time slices (highlighted blue box). Between time slices 2 and 5, there was a stable



**Figure 2.4.** Bipartite clusters of students and learning resources (black nodes belong to more than one cluster).



**Figure 2.5.** Swim lane diagram of the evolving student–resource clusters during the exam phase.

set of students (stud. group 3) using this material for exam preparation. In contrast, the students of the study program who had their oral exams later had a more diverse resource access behaviour (green box). Also, they began their exam preparation much closer to the time of the exam compared to the other study programs. In the last time slice, three of the four student groups merged to a larger group that was then more affiliated to the core learning material (res. group 1).

On the one hand, this example shows the possible expressiveness of sophisticated network analysis methods in a case where “first order methods” would not be able to resolve interesting and meaningful structural relations. On the other hand, it demonstrates that additional effort is needed to support a dynamic, evolutionary interpretation of network-based models (given that each single network is “ignorant” about time).

In an ongoing research project on supporting small group collaboration in MOOCs, we have used this approach of tracking cohesive clusters of learners and resources to distinguish “mainstreaming behaviour” from more individual or idiosyncratic patterns of resource usage on the part of learners (Ziebarth et al., 2015). Given this model-based distinction, we found that extrinsic motivation was more prevalent in the

mainstreaming group. This suggests that specific patterns in actor–artefact relations may serve as indicators for learning styles.

### Example 2: Analyzing the Evolution of Ideas in Knowledge-building Communities

Scientific production can be seen as a prototypical case of knowledge building in a community. Accordingly, methods developed to analyze scientific production and collaboration (“scientometric methods”) can plausibly also be used to analyze other types of knowledge building in networked learning communities. Hummon and Doreian (1989) have proposed the method of “main path analysis” (MPA) to detect the main flow of ideas in citation networks with scientific publications as nodes connected by citations. The original paper uses a corpus of publications in DNA biology as an example.

The MPA method relies on the acyclic nature of citation graphs. Different from other network-analytic techniques, MPA has an implicit notion of time that stems from the nature of citation networks (always the citing paper is more recent than the cited one). As a consequence of this time ordering, and given that every collection is finite, in a corpus, there are always documents not cited by others (end-points or “sinks”) as well as documents that do not cite other documents in the corpus (“sources”). The idea of MPA is to find the most used edges in terms of the information flow from the source nodes to the sink nodes. One common approach to finding these edges is the “search path count” or SPC method (Batagelj, 2003). All sources in the network are connected to a single artificial source and all sinks to a single artificial sink. SPC assigns a weight to an edge according to the number of paths from the source to the sink on which the edge occurs. The main path can then be found by traversing the graph from the source to the sink by using the edges with the highest weight, as depicted in Figure 2.6.

The idea of applying MPA to learning communities working with hyper-linked connections of wiki documents was first proposed by Iassen Halatchliyski and colleagues (2012). However, MPA cannot be applied directly to hyper-linked web documents because the premise of directed acyclic graphs (DAGs) is usually not fulfilled. Since the content of articles in a wiki is dynamically evolving, hyperlinks between two articles do not induce a temporal order between them and cycles or even bidirectional citation links are quite frequent. In Halatchliyski, Hecking, Göhnert, & Hoppe (2014), we have proposed a formal modification that allows for applying MPA also to this case. The adapted approach considers the particular revisions (successive versions) of articles instead of the articles themselves. Revisions of an evolving wiki article are



artefacts with stable content as scientific publications. In such a network based on versions as nodes, we introduce revision edges between successive revisions of the same article. The original hyperlinks between different articles connect specific revisions and also introduce new versions. This trick avoids cyclic structures and allows for applying MPA. In the context of the Wikiversity learning community, we have used the coincidence of articles with identified main paths as a basis to judge the importance or weight of contributions and to characterize author profiles in terms of specific role models (inspirator, connector, worker). These characterizations serve as supportive information for the management of knowledge building communities.

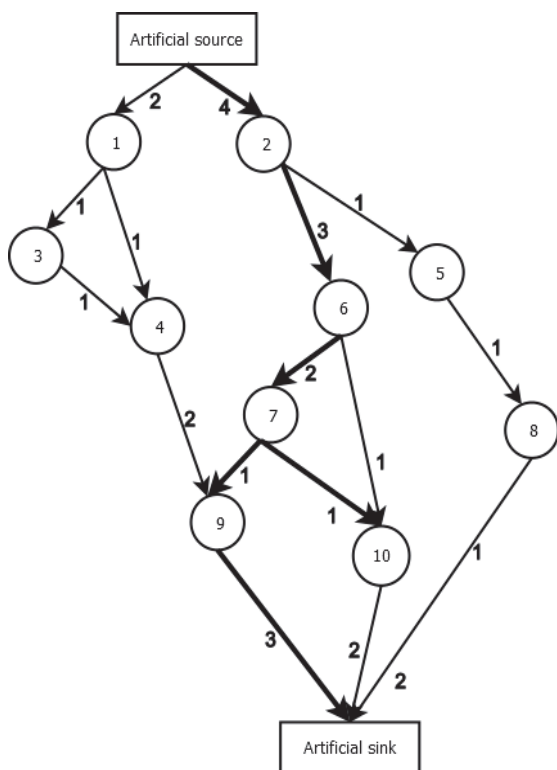
In a subsequent study (Hoppe, Göhnert, Steinert, & Charles, 2014), we have combined the network-analytic method of MPA with content analyses to analyze chat interactions in an educational community (*Tapped In* – see Farooq, Schank, Harris, Fusco, & Schlager, 2007). Here, the characteristic of chat as a synchronous communication medium, especially regarding turn taking, possible parallel threading, and interactional coherence had to be taken into account. Our work used contingency analysis (Suthers, Dwyer, Medina, & Vatraou, 2010) as theoretical background and reference to detect general dependencies based on operational

rules. We reconstructed and refined this approach by using the concept of dialogue act tagging (Wu, Khan, Fisher, Shuler, & Pottenger, 2005) to enrich the basic set of indicators. We have tested our method using several examples of chat protocols from a teacher community as benchmarks. This allowed us to assess the agreement between the contingency links generated by our method with previously hand-coded contingencies (Suthers & Desiato, 2012) based on the F-score (a measure used in information retrieval combining precision and recall). The automatically generated contingencies reached an F-score similarity of 83% to 97%, which is comparable to the pairwise F-score similarity of manually analyzed graphs. Figure 2.7 shows a fragment of a chat sequence with contingency links indicated on the right hand side, main path contributions highlighted in bold, and the message categories resulting from dialogue act tagging (e.g., “Statement” or “ynQuestion”) added in brackets.

The main path information should be interpreted as an indicator for the relevance of contributions in the evolution and progress of the overall discourse. This relevance measure for contributions can in turn be used to estimate the influence of participants in the discourse. Since we did not have human ratings for this feature, we have compared the measure “percentage of contributions on main path” (%MainPath) per actor to other influence rankings based on the well-established *PageRank* and *Indegree* measures. We applied these measures to different versions of the contingency graphs resulting from human and automatic coding. As a result, we found a 0.82 (0.82) correlation of %MainPath with *PageRank* and a 0.69 (0.88) correlation with *Indegree*. Per se, %MainPath is just another competing indicator. However it is different from the other measures since it takes into account the flow of arguments in the discourse and not only local (*Indegree*) or globally weighted prestige (*PageRank*). As can be seen in Figure 2.6, MPA also allows for filtering the discourse for main threads of argument. In this sense, MPA makes the network model more specific and meaningful. However, further investigation is needed to validate these constructs.

## DISCUSSION AND OUTLOOK

In general, we cannot expect LA to invent genuinely new computational methods of data mining and analysis. Yet, we have seen the successful adoption of a number of existing techniques. A prominent case is certainly social network analysis – to the extent that SNA concepts such as centrality measures or cohesive clusters (subcommunities) are now part of the conceptual repertoire used in LA discourse. This is still less the case for process-oriented techniques (such as



**Figure 2.6.** Example network illustrating the SPC method (edge weights are SPC counts; thick edges indicate the main paths).



**Figure 2.7.** Fragment of a chat protocol with inferred contingency links (mainpath contributions and links indicated in bold).

sequential pattern mining) or linguistics-based methods for the analysis of dialogue and textual artefacts. The examples and arguments presented in this article corroborate 1) that even SNA has more to offer than the better known “first order approaches” and 2) the most benefit can be expected from combining different types of analytic methods. Regarding network analysis techniques, moving from pure actor–actor networks to actor–artefact (or two-mode) networks provides a richer basis of information that can resolve more significant and meaningful relations. The example on analyzing resource access in online courses illustrates how this can make a difference. It also shows the inclusion of time by considering temporal sequences of networks.

This article looks at the issues and challenges primarily from a computational perspective. From a pedagogical perspective, it is important to identify the affordances but also the deficits of certain methods in order to judge their potential benefits. For example, when targeting “interaction patterns” in knowledge building communities it should be clear that pure SNA models would only reveal static actor–actor relations but not time-dependent patterns. Possible extensions would use time series of networks and/or actor–artefact relations. Network–text analysis is an example of an approach that converts textual artefacts into networks of concepts (of possibly different categories) and thus allows for combining content and network analytic approaches. On the other hand, given these computational methods, where are the potential pedagogical added values? In this respect, we have seen the following examples:

- Concept networks derived from learner-generated texts using NTA can reveal students’ mental models and misconceptions. This can be a basis for enriching domain taxonomies and for curriculum revision.
- The primary information that we get from learning platforms is about learners accessing (or possibly creating/uploading) resources. From sequences of ensuing two-mode learner–artefact networks, we can classify learner behaviour as “main-streaming” or more individually varied, possibly intrinsically motivated or curiosity-driven.
- Techniques borrowed from scientometrics allow for identifying the main lines of the evolution of ideas in knowledge building communities. On this basis, we can characterize contributions and the role of contributors to support better-informed decisions in the management of the community.

Forum participation in massive online courses has recently been the subject of several LA-inspired studies. Using a mix of analytic techniques involving SNA patterns combined with “regularity” of interactions and content assessment (based on human ratings), Poquet and Dawson (2016) have characterized success factors for productive and supportive forum interactions. Interestingly, they found an important influence of certain community members who were not themselves part of densely connected subgroups (or “cliques”) on the positive evolution of the networked community. In a similar context, Wise, Cui, and Vy-tasek (2016) have identified certain linguistic features



as predictors for distinguishing content-related from non-content-related (social or organizational) talk in such forums. In addition to domain-specific vocabulary, they also found general terms such as “understand,” “example,” “difference,” or question words among the predictors of content-related contributions. These, in turn, correspond to so-called “signal concepts” used in the network-text analysis of educational video comments by Daems, Erkens, Malzahn, and Hoppe (2014). This again shows the importance of having a

mix of modelling approaches and analysis techniques “at hand” to gain better insight and understanding of the determinants of learning and knowledge building communities.

The overarching claim and hope is to increase the awareness of the richness and variety of computational methods in the LA community, and thus to lay the ground for more synergy between LA and computationally inspired research.

## REFERENCES

- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, 9(2), 161–185.
- Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv:cs/0309023v1.
- Blei, D. M. (2012). Introduction to probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892–895.
- Carley, K. M., Columbus, D., & Landwehr, P. (2013). AutoMap user's guide 2013. Technical Report CMU-ISR-13-105. Pittsburgh, PA: Carnegie Mellon University, Institute for Software Research. [www.casos.cs.cmu.edu/publications/papers/CMU-ISR-13-105.pdf](http://www.casos.cs.cmu.edu/publications/papers/CMU-ISR-13-105.pdf).
- Daems, O., Erkens, M., Malzahn, N., & Hoppe, H. U. (2014). Using content analysis and domain ontologies to check learners' understanding of science concepts. *Journal of Computers in Education*, 1(2–3), 113–131.
- De Laat, M., Lally, V., Lipponen, L., & Simons, R. J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 87–103.
- Elkina, M., Fortenbacher, A., & Merceron, A. (2013). The learning analytics application LeMo: Rationals and first results. *International Journal of Computing*, 12(3), 226–234.
- Farooq, U., Schank, P., Harris, A., Fusco, J., & Schlager, M. (2007). Sustaining a community computing infrastructure for online teacher professional development: A case study of designing Tapped In. *Computer Supported Cooperative Work*, 16(4–5), 397–429.
- Ferguson, R., & Shum, S. B. (2012). Social learning analytics: Five approaches. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*, 29 April–2 May 2012, Vancouver, BC, Canada (pp. 23–33). New York: ACM.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3), 75–174.
- Fournier-Viger, P., Gomariz, A., Gueniche, T., Soltani, A., Wu, C. W., & Tseng, V. S. (2014). SPMF: A Java open-source pattern mining library. *Journal of Machine Learning Research*, 15(1), 3389–3393.
- Fox, A. (2013). From MOOCs to SPOCs. *Communications of the ACM*, 56(12), 38–40.
- Halatchliyski, I., Oeberst, A., Bientzle, M., Bokhorst, F., & van Aalst, J. (2012). Unraveling idea development in discourse trajectories. *Proceedings of the 10th International Conference of the Learning Sciences (ICLS '12)*, 2–6 July 2012, Sydney, Australia (vol. 2, pp. 162–166). International Society of the Learning Sciences.
- Halatchliyski, I., Hecking, T., Göhnert, T., & Hoppe, H. U. (2014). Analyzing the path of ideas and activity of contributors in an open learning community. *Journal of Learning Analytics*, 1(2), 72–93.
- 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. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.),

- Proceedings of the 7th International Conference on Computer-Supported Collaborative Learning (CSCL 2007)*, 16–21 July 2007, New Brunswick, NJ, USA (pp. 270–279). International Society of the Learning Sciences.
- Harrer, A., Martínez-Monés, A., & Dimitracopoulou, A. (2009). Users' data: Collaborative and social analysis. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (eds.), *Technology-enhanced learning: Principles and products* (pp. 175–193). Springer Netherlands.
- Haythornthwaite, C. (2001). Exploring multiplexity: Social network structures in a computer-supported distance learning class. *The Information Society*, 17(3), 211–226.
- He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29(1), 90–102.
- Hecking, T., Ziebarth, S., & Hoppe, H. U. (2014). Analysis of dynamic resource access patterns in online courses. *Journal of Learning Analytics*, 1(3), 34–60.
- Hoppe, H. U., Erkens, M., Clough, G., Daems, O. & Adams, A. (2013). Using Network Text Analysis to characterise teachers' and students' conceptualisations in science domains. In R. K. Vatrappu, P. Reimann, W. Halb, & S. Bull (Eds.), *Proceedings of the 2nd International Workshop on Teaching Analytics (IWTA 2013)*, 8 April 2013, Leuven, Belgium. <http://ceur-ws.org/Vol-985/paper7.pdf>
- Hoppe, H. U., Göhnert, T., Steinert, L., & Charles, C. (2014). A web-based tool for communication flow analysis of online chats. *Proceedings of the Workshops at the LAK 2014 Conference co-located with the 4th International Conference on Learning Analytics and Knowledge (LAK '14)*, 24–28 March 2014, Indianapolis, IN, USA. <http://ceur-ws.org/Vol-1137>.
- Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11(1), 39–63.
- Hung, J. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology*, 43(1), 5–16.
- Krempel, L. (2005). *Visualisierung komplexer Strukturen – Grundlagen der Darstellung mehr-dimensionaler Netzwerke*. Frankfurt, Germany: Campus-Verlag.
- Lehmann, S., Schwartz, M., & Hansen, L. K. (2008). Biclique communities. *Physical Review E*, 78(1). doi:10.1103/PhysRevE.78.016108
- Mühlenbrock, M., & Hoppe, H. U. (1999). Computer supported interaction analysis of group problem solving. *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning (CSCL '99)*, 12–15 December 1999, Palo Alto, California (pp. 398–405). International Society of the Learning Sciences.
- Palla, G., Barabasi, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446(7136), 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(7043), 814–818.
- Poquet, O., & Dawson, S. (2016). Untangling MOOC learner networks. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK '16)*, 25–29 April 2016, Edinburgh, UK (pp. 208–212). New York: ACM.
- Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In B. Wasson, S. Ludvigsen, & H. U. Hoppe (Eds.), *Proceedings of the Conference on Computer Support for Collaborative Learning: Designing for Change in Networked Environments (CSCL 2003)*, 14–18 June 2003, Bergen, Norway (pp. 343–352). International Society of the Learning Sciences.
- Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237–271.
- Sherin, B. (2013). A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of the Learning Sciences*, 22(4), 600–638.

- Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1), 29–48.
- Suthers, D. D., Dwyer, N., Medina, R., & Vatraou, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 5–42.
- Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. *Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS-45)*, 4–7 January 2012, Maui, HI, USA (pp. 3368–3377). IEEE Computer Society.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.
- Wise, A. F., Cui, Y., & Vytasek, J. (2016). Bringing order to chaos in MOOC discussion forums with content-related thread identification. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK '16)*, 25–29 April 2016, Edinburgh, UK (pp. 188–197). New York: ACM.
- Wu, T., Khan, F. M., Fisher, T. A., Shuler, L. A., & Pottenger, W. M. (2005). Posting act tagging using transformation-based learning. In T. Y. Lin, S. Ohsuga, C.-J. Liao, X. Hu, & S. Tsumoto (Eds.), *Foundations of data mining and knowledge discovery* (pp. 321–331). Springer Berlin Heidelberg.
- Zeini, S., Göhnert, T., Hecking, T., Krempel, L., & Hoppe, H. U. (2014). The impact of measurement time on subgroup detection in online communities. In F. Can, T. Özyer, & F. Polat (Eds.), *State of the art applications of social network analysis* (pp. 249–268). Springer International Publishing.
- Ziebarth, S., Neubaum G., Kyewski E., Krämer N., Hoppe H. U., & Hecking T. (2015). Resource usage in online courses: Analyzing learner's active and passive participation patterns. *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, 7–11 June 2015, Gothenburg, Sweden (pp. 395–402). International Society of the Learning Sciences.
- Zumbach, J., Mühlenbrock, M., Jansen, M., Reimann, P., & Hoppe, H. U. (2002). Multi-dimensional tracking in virtual learning teams: An exploratory study. *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (CSCL 2002)*, 7–11 January 2002, Boulder, CO, USA (pp. 650–651). International Society of the Learning Sciences.