Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community

Iassen Halatchliyski
Knowledge Media Research Center, Germany

Tobias Hecking, Tilman Göhnert, H. Ulrich Hoppe
University of Duisburg-Essen, Germany

i.halatchliyski@iwm-kmrc.de

Abstract: This paper introduces the scientometric method of main path analysis and its application in an exemplary study of the paths of knowledge development and the roles of contributors in Wikiversity. Data from two scientific domains in this online learning community has been used. We see this as a step forward in adapting and adopting network analysis techniques for analyzing collaboration processes in knowledge building communities. The analysis steps are presented in detail including the description of a tool environment ("workbench") designed for flexible use by non-computer experts. By identifying directed acyclic graphs, the meaningful interconnections between developing learning resources are analyzed by considering their temporal sequence. A schema for the visualization of the results is introduced. The potential of the method is elaborated for the evaluation of the overall learning process in different domains as well as for the individual contributions of the participants. Different outstanding roles of contributors in Wikiversity are presented and discussed.

Keywords: knowledge artifacts, idea flow, scientometrics, main path analysis

1. INTRODUCTION

Nowadays, it is commonplace to perceive learning and knowledge building as closely related activities on the Web. Knowledge building is based on *epistemic artifacts* (Knorr-Cetina, 2001) created and shared in a community. Bereiter and Scardamalia (2003) point out that knowledge building is essential for learning but has a wider scope in that it is not necessarily limited to explicit learning scenarios. Scientific research is an example of a distributed knowledge building activity that takes place in scientific communities and typically is not characterized as learning. According to Scardamalia and Bereiter (1994), the knowledge building pedagogy takes scientific research as a blueprint of the collaborative learning of students that needs to be facilitated. During a knowledge-building process, students discuss ideas and develop their shared knowledge in the manner of scientists. The philosophical foundation of this view dates back to Popper (1968), who explains the development of scientific knowledge as a constant process of emergence of new ideas and their gradual improvement or abandonment after discovering contradictory evidence. In fact, any learning community defines concepts and builds its knowledge base in a similar way (Stahl, Koschmann, & Suthers, 2006).

With the present work we offer an approach to analyzing learning processes organized in the form of online knowledge building. Online knowledge building is characterized by collaborative activities and the creation of shared artifacts within a community of learners. This form of collaborative learning is becoming increasingly popular on the Web and goes beyond formal educational contexts (Halatchliyski, Moskaliuk, Kimmerle, & Cress, in press). As this is a relatively new phenomenon and it shifts the focus from the individual learner to the knowledge processes within a community, appropriate methodologies are expectedly complex and in a very early developmental stage.

Due to the relation between scientific production and learning in communities, we aim to show that both processes can be studied using the same analytical approaches. S*cientometrics* as a research field is particularly concerned with the quantitative measurement of scientific work, and so offers a variety of potentially fruitful approaches new to the area of learning analytics (Suthers & Verbert, 2013). Scientometric methods are tailored for the analysis of knowledge artifacts, most prominently publications, and their authors. One well-known method is the calculation of the h-index as a measure of scientific reputation (Hirsch, 2005). In the context of learning communities, however, individual excellence is not a primary concern. Rather more interesting would be an approach to the long-term characteristics and the dynamics of interactive learning environments.

Hummon and Doreian (1989) have proposed a method to detect the main idea flows based on citation networks using a corpus of publications in DNA biology as an exemplar. Our work reported in this paper takes the *main path analysis* technique as a starting point in the analysis of a broad range of knowledge building processes that take place in formal as well as informal collaborative settings. After an initial promising application of main path analysis to networks of knowledge artifacts created for educational purposes (Halatchliyski, Oeberst, Bientzle, Bokhorst, & van Aalst, 2012), we now want to elaborate on the adaptation and adequate formalization of the method. Our guiding question in this endeavor is: What kind of insights can be gained from the main path analysis of knowledge creation in online learning communities? We will explore this question using data from Wikiversity[1] as an example. Wikiversity is understood by its active members as an "open learning community" in which users can actively produce learning resources for a broad range of topics in the form of web-based hypermedia. In our view, it represents a challenging and yet relevant field for exploring the potential of scientometric methodology to tackle the dynamics of computer-supported learning processes.

2. BACKGROUND

2.1 Community learning

New knowledge in the world might be the accomplishment of an individual, but it is inconceivable without the body of previously existing knowledge that in turn has been established by many other individuals. Consequently, learning and development of new knowledge must be examined in the context of the community in which they take place.

Online communities like Wikiversity facilitate learning through the creation of a shared knowledge base in the form of digital artifacts such as texts, pictures, or other multimedia. Users can passively learn by making use of the existing artifacts. Users can also actively learn by participating in the creation of new

---

[1] http://www.en.wikiversity.org/

artifacts. The *knowledge building theory* suggests incorporating such activity in formal education (Scardamalia & Bereiter, 1994). Students are expected to benefit from self-motivated exploration of knowledge areas when they share and build on each other's findings in a collaborative online environment. During this long-term process, the shared community knowledge develops as ideas are constantly improved by the participants. Individual learning is an outcome of the knowledge development of the whole community.

The collaborative production of digital knowledge artifacts has become widespread since the emergence of Web 2.0. Widely and easily available tools such as wikis afford a long-term process of mass collaboration, as artifacts are built piece-by-piece and individual contributions have variable sizes. Moreover, a single contribution to an artifact can be revised or be built upon in order to produce newer versions. Every change to the shared artifacts of a wiki community can be logged as an individual contribution activity, but the ongoing development of the knowledge base is an emergent product of the community as a whole. Intersubjectivity and shared meaning-making are epiphenomena of the interaction among individuals in a community (Stahl et al., 2006). From the systemic view of the *co-evolution model* of individual learning and collaborative knowledge building (Cress & Kimmerle, 2008), a community and the participating individuals function as two different types of systems that co-evolve through mutual fertilization. Knowledge development is reflected in the changing shape and content of the artifacts.

Knowledge artifacts often hold connections among themselves that are marked by higher-level semantic structures like topical relations, problem-solution chains, discourses, etc. Regardless of whether these connections are deliberately made by the participants in a community or whether they are automatically produced by the online environment, hypermedia links bear meaning. This meaning is an integral part of the knowledge created by a community. It is also subject to change, as connections are added or deleted in parallel with the artifact development.

In sum, learning in a community represents a complex process dependent on the activities of many participants and supported by the use and development of artifacts as learning resources. The process evolves with the constant change of the shared knowledge base at the level of single resources or their interconnections.

2.2 Temporality of a Learning Process

The learning of an individual or of a whole community is a process that essentially develops over periods of time. New knowledge is built upon existing knowledge. A knowledge base develops gradually as its information content evolves. Single ideas become more concrete, they can flow together or split into independent directions, marking a convergence or divergence in the development process (Halatchliyski, Kimmerle, & Cress, 2011). At a higher level of abstraction, the interconnections within the knowledge base also develop when new ideas are added to existing content, or when already existing connections are subsequently changed.

All these changes should be studied in order to understand the corresponding learning processes. Accordingly, the temporal dimension should be regarded as a main component of learning analytics. However, the modelling of the overall process of knowledge development is challenging, as the sequential relations between all the changes in the knowledge base need to be tracked. Any aggregation across time easily leads to a biased analysis of individual and community-level variables. A longitudinal

study of different points in time is also an unsatisfactory option, as it misses out on the authorship of changes made between the chosen time points. Especially difficult to grasp is the nonlinear flow of ideas that is characteristic to any learning process.

Previous work in the area of computer-supported learning has paid attention to the interactivity of collaborative processes and thereby implicitly to learning dynamics. Environment data logs have been used to describe and map interaction patterns. Their interpretation has often been supported by additional analysis of the content in the case of discussion board messages (see, for example, Hara, Bonk, & Angeli, 2000; Schrire, 2004). Suthers, Dwyer, Medina, and Vatrapu (2010) also presented a universal framework for describing interactivity in the form of uptakes between contributors independent of the environment used. Nevertheless, the field of learning analytics still needs a method to address the temporality of learning processes quantitatively. Aspects that need to be taken into account include who influenced whom, which ideas were taken up in later stages and which were not, and how differently do the participants contribute to the overall learning process. The method should also be adaptable to the multiplicity of learning environments and communities that have emerged with Web 2.0.

Different forms of sequential analysis of learner actions have also been developed in order to detect and understand the best practices of orchestration of tools and content in the learning process (Cakir, Xhafa, Zhou, & Stahl, 2005; Jeong, 2003; Perera, Kay, Koprinska, Yacef & Zaïane, 2009). Frequently occurring sequences of actions or events reveal connections between the learning history as captured in log files and learning performance. Such analysis should help warn learners against inefficient strategies and also better adapt the environment and the learning materials to their needs (Zaïane & Luo, 2001). Although it certainly accounts for the temporal dimension and thereby gives deeper insight into the learning process, sequential analysis as a data-mining technique relies on the *a priori* definition of activity and event categories. The necessary coding scheme always represents a potential weak point in the analysis as it predetermines the level of abstraction and the scope of possible patterns that can be found. The method also lacks the possibility of utilizing information on the relations between specific participants or artifacts. The latter lend themselves to analysis with a network perspective.

Social network analysis (SNA) has been used in various areas, including computer-supported collaborative learning (Aviv, Erlich, Ravid, & Geva, 2003; de Laat, Lally, Lipponen, & Simons, 2007; Harrer, Malzahn, Zeini, & Hoppe, 2007; Reffay & Chanier, 2002). The basic approach relies on representing communication events as links between the actors in the network. Applied to networks of knowledge artifacts on the Web, SNA can be an efficient approach to knowledge and its collaborative development by analyzing the meaningful structure of connections between knowledge artifacts (Halatchliyski et al., in press). The resulting network structure will very much depend on the time span during which these events are collected (Zeini, Göhnert, & Hoppe, 2012). However, the target representation no longer represents temporal characteristics. For this reason, SNA has been criticized for eliminating time. Although advances are being made to analyze the development of networks, these rarely address true network dynamics. Process temporality represents a major dimension of online learning and should not be ignored in an analysis. In this paper we present a network analysis technique that can explicitly address learning dynamics in the context of an open learning community.

75

3. ANALYTICAL APPROACHES TO KNOWLEDGE DEVELOPMENT

3.1   Actor-Artifact Networks

The knowledge building process develops around the creation of knowledge artifacts. A specific version of a so-called two-mode-network can be built on the basis of the relation between the actor (or author) and the artifact (or product). In the SNA methodology (Wasserman & Faust, 1994), such two-mode networks are also called affiliation networks. In the pure form, these networks are assumed to be bi-partite, that is, only links alternating between actor-artifact ("created/modified") and artifact-actor ("created/modified-by") would be allowed. Using simple matrix operations, such bi-partite two-mode networks can be "folded" into homogeneous (one-mode) networks. Here, for example, two actors would be associated if they have acted upon the same artefact (Suthers & Rosen, 2011). We would then say that the relation between the actors was mediated by the artifact. A typical example of such a transformation is offered by co-publication networks based on co-authorship. Similarly, we can derive relationships between artifacts by considering agents (engaged in the creation of two different artifacts) as mediators.

The "pure" view of actor-artifact relations as bi-partite networks has a clear mathematical-operational structure. However, there are good reasons to extend this approach: Both actors and artifacts may be interrelated in other ways than by this type of cross-wise mediation. For instance, social relations between actors may operate independently of the artifact mediation. Semantic relations may be salient between knowledge artifacts, as in the "semantic web." Mika (2007) was one of the first to elaborate on methods and potential gains of blending social and semantic network structures. Other approaches allocate actors and artifacts on different layers of a multi-layer model with homogeneous relation within each layer and heterogeneous relations in between (Reinhardt, Moi, & Varlemann, 2009; Suthers & Rosen, 2011). Such multi-relational representations may appear superior in expressiveness; however, operations in such structures are more difficult to define.

As with any other network representation, actor-artifact networks also fail to capture the notion of time explicitly. However, time may be implicitly modelled in the network relations. In the scientometric field, this is the case for citation networks: If publication X cites publication Y, we can safely assume that Y is older than X. The ensuing network structure does not contain cycles (excluding specific rare cases of cross-citation). The main path analysis method builds on such acyclic citation networks and can also be adapted to the dynamics of networks of knowledge artifacts built in the process of online collaborative learning.

3.2   Main Path Analysis

The main path analysis (Hummon & Doreian, 1989) is a network analysis technique for the scientometric study of scientific citations over a period of time. Its major application is the identification of key publications in the development of a scientific field. While many scientometric methods, such as the analyses of co-citation and co-authorship networks, stress the semantic structure of scientific work, main path analysis additionally considers the temporal structure of development. Temporality is accounted for through the very definition of a *directed acyclic graph* (DAG) where nodes are single publications and directed edges represent citations between publications. The direction of an edge corresponds to the flow of knowledge from the cited and older publication to the citing and newer

publication. Therefore, these links incorporate both the dimension of content relations and the temporal order of the contributions.

A DAG always contains at least one node with no ingoing edges (i.e., a source) and at least one node with no outgoing edges (i.e., a sink). In the citation network of scientific publications within one field, often one important publication is chosen as a starting point for the development of the field. This publication represents the first source. Later on, other sources that may not have cited previous publications in the field can become prominent and highly cited. Sink nodes, in contrast, represent either unimportant or very new publications not cited at the time of analysis.

The main path can be described as the most used path among all possible paths of successive edges from the source nodes to the sink nodes in a citation network. This most used path can be found by a two-step procedure: first, the traversal counts for each edge are calculated as the number of different paths between each source and sink nodes that go through this edge and, second, an algorithm is used to identify the main path based on the edge traversal counts.

This paper employs the search path count (SPC) algorithm (Batagelj, 2003), which introduces one fictitious source node and one fictitious sink node and links these to each of the actual source and sink nodes, respectively. In the example in Figure 1 the fictitious source and sink nodes are 1 and 10. Their only purpose is to simplify the original procedure (Hummon & Doreian, 1989) of weight calculation for the edges connecting the real nodes. Starting at the fictitious source node (1), the main path is identified by successively following the edge with the maximal weight to the next node until the fictitious sink node (10) is reached. At node 7 in Figure 1, there are two possible alternatives to reach the next node, because both outgoing edges have the same traversal weight, in this case the main path branches.
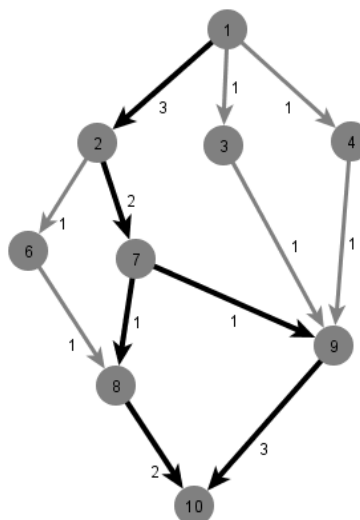

Figure 1. Example of a main path calculation

The SPC algorithm might present too strict an approach to the idea of main path, depending on the nature of the graph. For the case when the analysis requires a broader view on the main contributions in a field, Liu and Lu (2012) suggested lowering the search constraint by defining a threshold. In each step, one chooses not only the edges with the maximum weight but also edges with weight above a certain

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)

77

percentage of the maximum weight. In the present work, we applied a slightly modified procedure to identify the *multiple* main paths (Liu & Lu, 2012): After calculating the traversal weight of each node, we considered all the nodes with a weight above a certain threshold as part of the multiple main paths. This strategy facilitates the identification of multiple main paths of important but thematically disparate contributions that may not necessarily build one connected component.

Methods related to the main path analysis represent a structural approach appropriate for addressing the dynamics of online community learning. Depending on the nature of hyperlinks, a DAG may trace the flow of influences between ideas or the change in meanings that accompanies knowledge development. The technique allows identifying the most influential contributions and their authors in the course of the construction of a community knowledge base over time. It also facilitates the characterization of the overall discourse trajectory in collaborative learning (Halatchliyski et al., 2012).

4. EMPIRICAL STUDY

4.1 The Context of the Wikiversity Data

Wikiversity is an online learning environment operating on a wiki technology since 2006. Like its larger and older sister projects, Wikipedia and Wikibooks, Wikiversity is offered in many languages and directed at any Web user. It is not designed as the online version of an academic organization providing courses or exam certificates. It is rather an experimental open space for collaborative learning to be used by any groups of participants according to their learning goals. A major feature is the openness of the created artifacts and of the community practices to accept constructive suggestions and participation by any interested user. Thus, Wikiversity follows a learner-centred approach (Bonk & Cunningham, 1998).

As a constantly developing so-called *open learning community*, Wikiversity accumulates a rather diverse body of many types of learning resources loosely structured in scientific topics from accounting to zoology. The pages categorized under any one Wikiversity category are often set up by different users and may serve different purposes. There are separate articles but also pages connected as bigger projects or organized as courses. Nevertheless, there are often hyperlink interconnections between the different pages and contributors often join multiple projects, sometimes years after their initial start. Because of the openness, there is a great variety of participation modes within and between the different topic categories.

The development of participation is an essential part of the learning process for users. In fact, users who become more involved with the community extend their participation to many unrelated scientific topics. Even when experienced users stay within the borders of one scientific category, their contributions increasingly follow the dynamics of the shared online environment and go beyond the starting individual goals. Such possible starting goals might be, for example, the arrangement of materials for a clearly delineated course as a teacher or the participation in such a course as a student, often in connection with offline lectures in parallel. Similar scenarios of online learning and teaching in Wikiversity do occur but are not representative of the idea that the community envisions, because this form of participation is not particularly collaborative. In the long run, the learning of individuals should become interconnected, producing an interwoven socio-epistemic fabric of a community constantly open to new constructive contributions.

Because of the non-homogeneous learning practices and artifacts, the Wikiversity data represents a real challenge for a learning analytics specialist. In the following, we present our approach for discerning major patterns of learning activities and profiles of contributing participants.

4.2   Extraction and Preparation of Wiki Data

As explained in section 3.1, the main path analysis was originally developed as a method to investigate the main discourse structure of scientific fields, using networks of publications linked by citations. However, the analysis method is not restricted to this field of application. The first author and colleagues have already demonstrated how it can be applied in the educational context of computer-supported classroom discussions (Halatchliyski et al., 2012). Moreover, it can be applied to any kind of directed acyclic graph (DAG). In this paper we show how to employ the main path analysis approach to examine the development of interconnected learning resources related to a knowledge domain in the context of a wiki environment.

All analyses presented in this paper are based on data from an official dump file[2] of the English Wikiversity from 20 February 2012. We did not use the complete wiki data but employed the concept of MediaWiki[3] categories in order to identify the body of artifacts related to a specific knowledge domain. Each wiki page can be categorized under one or more headings. The categories are themselves structured into subcategories. The actual data gathering process usually starts with extracting the complete subcategory structure by following the hierarchy starting at a given top-level category. In a second step, all pages organized into at least one of the categories found in this structure are identified. It is not mandatory that each wiki page be categorized, but approximately 70 percent of all articles in the English Wikiversity belong to at least one category. Thus, we assume that our procedure yields a representative selection of the major learning resources in a knowledge domain. The chance of considering pages unrelated to a domain, which can happen when complete subcategory structures are extracted, also needs to be eliminated. One example is the category "electrical engineering" which contains "Wikiversity" as a subcategory with its large number of administrative pages that are factually unrelated to electrical engineering. Therefore, a list of subcategories for exclusion from the extraction process needs to be predefined.

As a next step, a directed acyclic graph is constructed, describing the complete flow of knowledge within a single domain in a wiki. Networks of hypermedia resources in a wiki are analogous to networks of publications interconnected by citations. Wiki pages can be regarded as publications connected by hyperlinks instead of citations. Both citations and hyperlinks indicate a flow of knowledge with a direction from a source (i.e., a cited paper or a hyperlinked page) to a target (i.e., a citing paper or a hyperlinking page).

The temporal stability of publications is crucial for the generation of a DAG from citation networks. Moreover, only works already published can be cited. In contrast to scientific publications and citations featured in their content, which are published once and then remain static from that point on, wiki pages evolve over time under the collaborative efforts of community members. Furthermore, it is quite natural that one wiki page is hyperlinked to a second page and, at the same time, the second page links

---

[2] http://dumps.wikimedia.org/enwikiversity
[3] http://www.mediawiki.org/

back to the first one, thus introducing a cycle. In order to overcome these problems, we used the Wikiversity revision logs and the page versions after each revision contained in the dump.

Regarding stability over time, revisions of a wiki page behave like classical publications. They are created (published) at a certain point in time and do not change later on. A change to a wiki page will result in a new revision and thus a modified content of that page but not in a modification of the former revision. This approach suggests using page revisions instead of wiki pages as nodes in a DAG extracted from wiki data. We distinguish between two types of directed edges in such graphs: update edges and hyperlink edges.

Update edges can be introduced between any two directly subsequent revision nodes that belong to the same page. Update edges are directed from the older revision to the newer, updated revision and, thus, represent knowledge flow over the course of the collaborative process on a single wiki page.

Hyperlink edges can be traced between two revision nodes that belong to different pages with a hyperlink pointing from one to the other. A wiki hyperlink almost exclusively points to a page and not to a specific revision and it can be interpreted as an inversely directed knowledge flow, so in the proposed DAG hyperlink edges go in a direction *opposite* to the direction of the hyperlinks in the wiki. A knowledge flow between two wiki pages is elicited at the moment of the hyperlink creation between them. Thus, a hyperlink edge in the DAG starts at the latest revision of a hyperlinked page relative to the creation time of the relevant hyperlink and points to the first revision of the target page containing that hyperlink.
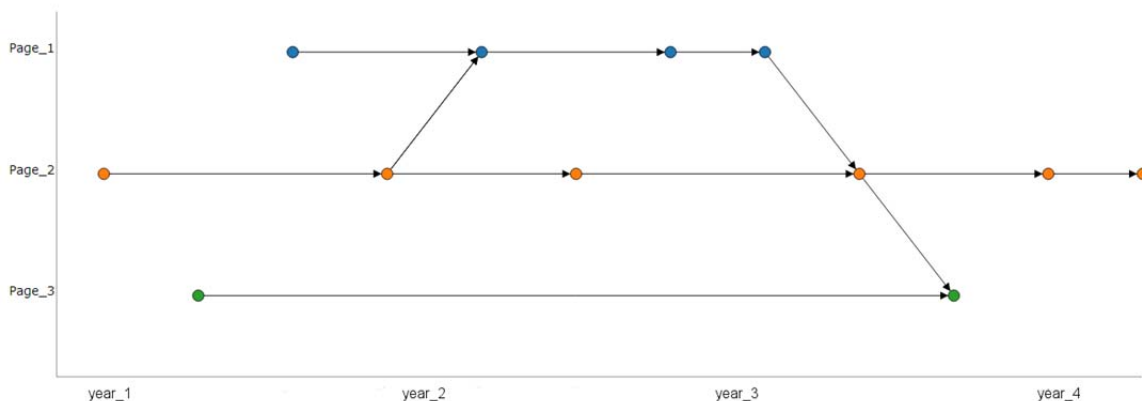

Figure 2. Swim lane diagram of a sample DAG of three articles with update and hyperlink edges

The described construction procedure results in a two-relational DAG that features update edges between revisions of a single page on the one hand and hyperlink edges between revisions of two related pages on the other hand. The procedure also guarantees that all update and all hyperlink edges are directed from a preceding revision to a succeeding revision in time. An example for such a DAG can be seen in Figure 2. In order to visualize the main paths of idea flows in a wiki, we use the visual metaphor of a "swim lane" diagram introduced in Figure 2. The page titles are shown in the left part of the diagram. All revisions of one page are represented as nodes connected by update edges and ordered in a horizontal line. The update edges of different pages are drawn parallel to one another, forming horizontal "swim lanes." Hyperlink edges between different pages are depicted as diagonal lines crossing the swim lanes. All edges point from left to right depicting the knowledge flow over time. Time

is represented on the horizontal axis along the swim lanes. For any pair of nodes that belong to the same or to different pages, the node closer to the left represents the earlier of the two revisions. Node size reflects the traversal weight of a revision as calculated by the main path analysis. The more important a revision is within the paths of ideas, the larger the node is that represents it.

4.3   Results of the Main Path Analyses

Using the described method to build a DAG from wiki data, we analyzed the main paths in the two scientific domains *biology* and *electrical engineering* in Wikiversity. Both chosen categories represent well-developed domains in Wikiversity and serve as example datasets of different scales to illustrate our analysis method. Table 1 first gives a basic description of the two domains based on the revision logs in the dump.

Table 1. Descriptive characteristics of the studied domains

Domain	Pages in total	Pages on multiple main paths (90%)	Pages on main path	Edits in total	Edits on multiple main paths (90%)	Edits on main path	Authors in total	Authors on multiple main paths (90%)	Authors on main path
Biology	1268	58	8	9404	949	111	925	118	6
El. Engin.	398	34	6	4672	442	130	687	103	42

The three data blocks in Table 1 contain the number of pages, edits, and authors in the chosen Wikiversity categories. Each block shows the total count of each variable, as well as their distribution on the main path according to the SPC method (see section 3.2) and on the multiple main paths with 90 percent threshold (i.e., containing all nodes with a traversal weight above the 90th percentile).

Although the biology domain is much larger than electrical engineering in terms of page count, the latter domain is marked by a proportionally higher number of edits and authors. A clearly higher percentage of the pages in biology seem to be peripheral to the development of this domain. A similar number of authors in biology have produced roughly double the number of edits and pages on the multiple main paths in electrical engineering. This comparison reveals a higher average productivity of the authors on the multiple main paths in the biology domain. From the reverse point of view, this means that the multiple main paths in the biology domain were developed less collaboratively than those in the electrical engineering domain. Lastly, the main path in both domains is of similar length of edits and pages, but in electrical engineering, it is created by proportionally many more authors. Next, we present in detail the main path and the multiple main paths in both domains.

4.3.1   Main paths in the biology domain

The result of the main path analysis with the SPC method is depicted in Figure 3 as a swim lane diagram. The main path consists of pages from an online course on the applications of evolutionary principles held in 2009. The articles are well orchestrated, indicating a course syllabus of topics that build on one another. With only six contributors in total (see Table 1) and only two of them contributing more than two changes to the pages, the course represents a top-down approach to the design of instructional materials for a relatively passive group of learners. The revision logs reveal that the course materials did not initiate further development of the topic, as only three edits have been made since the second half of 2009, namely to the article on applications in physics (see Figure 3).
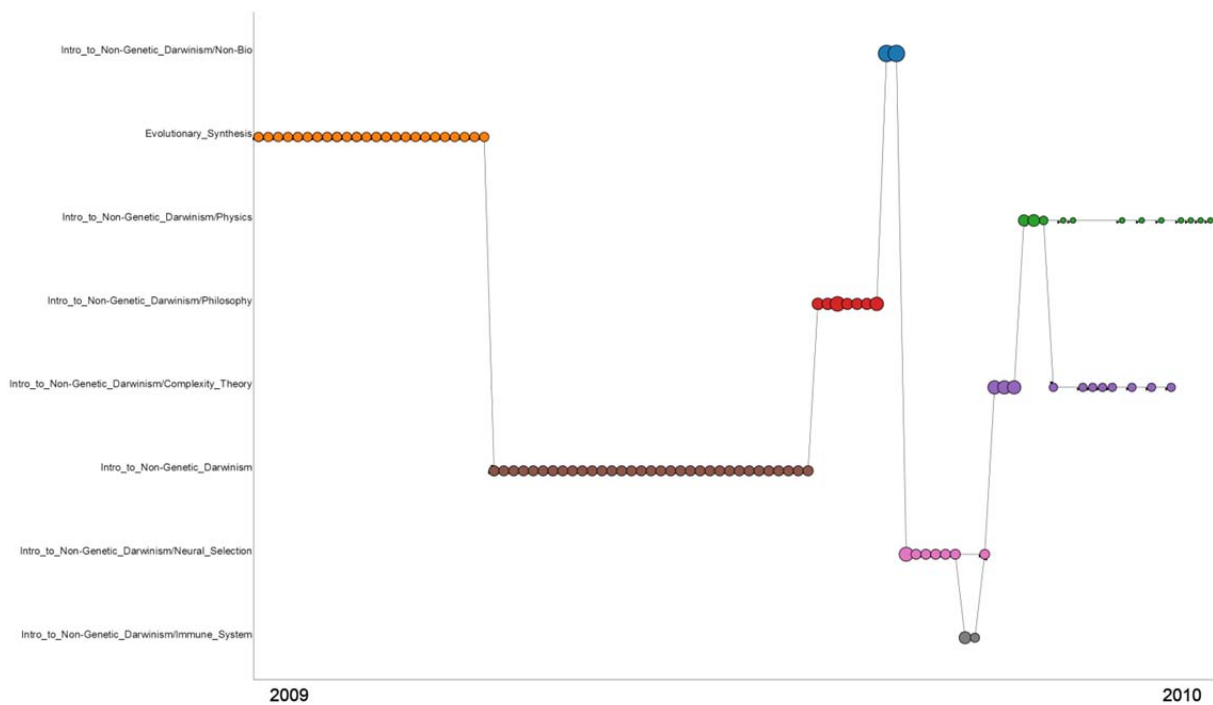
Figure 3. Simple main path in the biology domain

In order to broaden the range of important topics in the further analysis of the biology domain, we identified the multiple main paths as explained in section 3.2. Figure 4 shows the resulting swim lane diagram with additional branches of nodes and edges. Only ten percent (90th percentile threshold) of the article revisions with the highest traversal weight appear as part of the multiple main paths. Among them are all revisions presented as the main path in Figure 3.

Besides the discussed main path of the online course on evolutionary principles, several other topics appear as new separate branches: a cluster on sustainability and renewable energy from 2007 and 2008; two pages from a course on complex systems from 2011; an article about gynecological interviews gradually developed from 2007 to 2011; a small cluster on UFO research from 2006 and 2007; a larger and long-spanning cluster containing well-developed learning project pages about vitalism and consciousness, RNA interference, stem cells, life origins, human genetics, dominant group, and the connected basic biological concepts. Both branches containing the topics of vitalism and human genetics were first developed independently and later on flowed into the larger cluster. The main trajectory of that cluster starts with the topics RNA interference and cell improvement and ends with the topic dominant group.

The overall picture of the learning process in this domain suggests a heterogeneous evolution of ideas organized into separate topics. This conforms to the picture of groups of learners that followed different clearly defined interests in biology with little inter-group collaboration, except for the larger cluster of projects building on basic shared learning resources such as the general article on biology. The biology domain seems representative for the diverse and partly disconnected culture of online learning in the whole Wikiversity community.
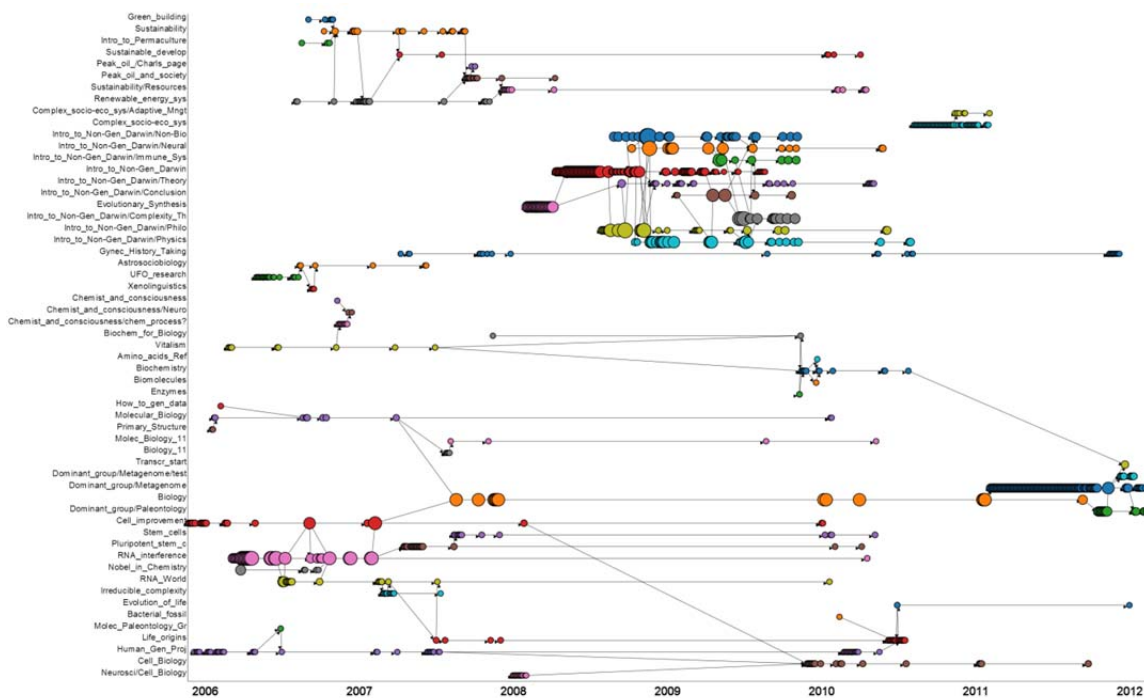
Figure 4. Multiple main paths in the biology domain

4.3.2   Main paths in the electrical engineering domain
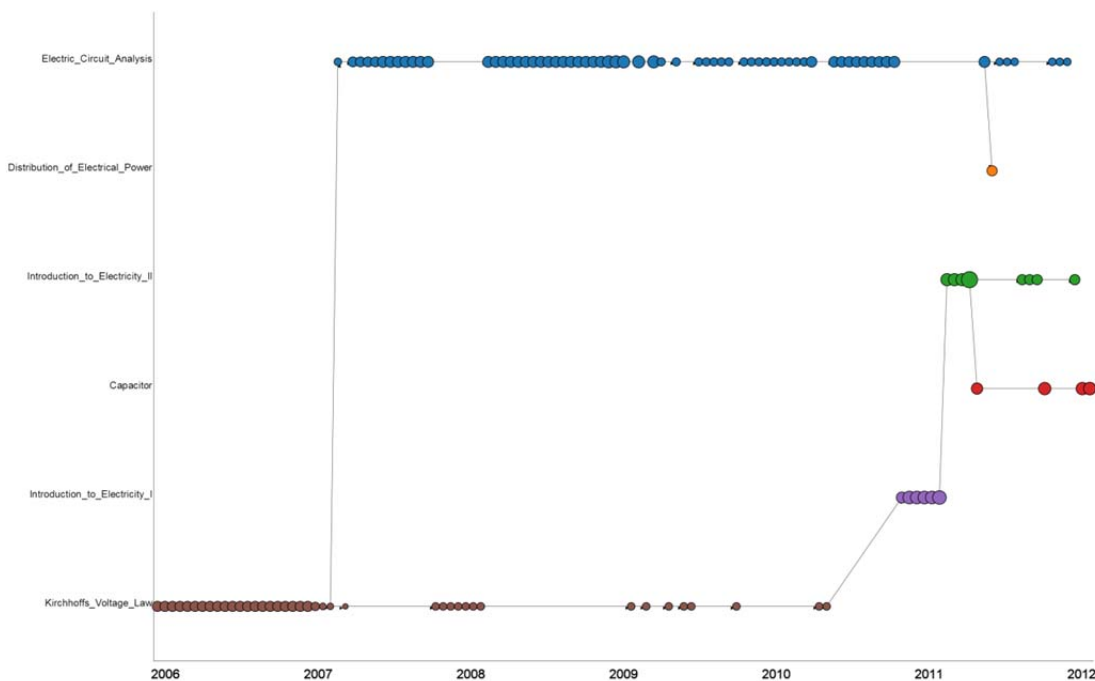Figure 5 shows the swim lane diagram of the SPC main path in the domain.

Figure 5. Simple main path in the electrical engineering domain

83

As with the main path in the biology domain, the core of the main path is the main page of an online course on electric circuits. In contrast to the course on evolutionary principles, this electric engineering course has been developed over a longer period from 2007 to 2009 and thus goes beyond the format of a course in the formal educational sense. The main path also contains an older resource from 2006 about voltage law that was later included in the course syllabus, as well as newer introductory resources on electricity from 2010 and 2011 that also referred to voltage law.

The interconnected and well-maintained articles indicate the core and narrowly interrelated topics in the domain. The creation of these core materials is an example of a truly collaborative learning process with many participating contributors (42 authors as shown in Table 1) over longer periods of time. The produced materials are structured as courses in order to facilitate any passive user encountering the topic, but the interesting learning process of the community of contributors is manifested in the collaborative creation of the study material itself.

As in the biology domain, we took a detailed look into the broader range of important topics in electrical engineering by analyzing the multiple main paths traced by ten percent of the article revisions with the highest traversal weight (90th percentile threshold). Figure 6 shows the resulting swim lane diagram that contains several new branches and additional nodes besides all the edits on the main path from Figure 5.
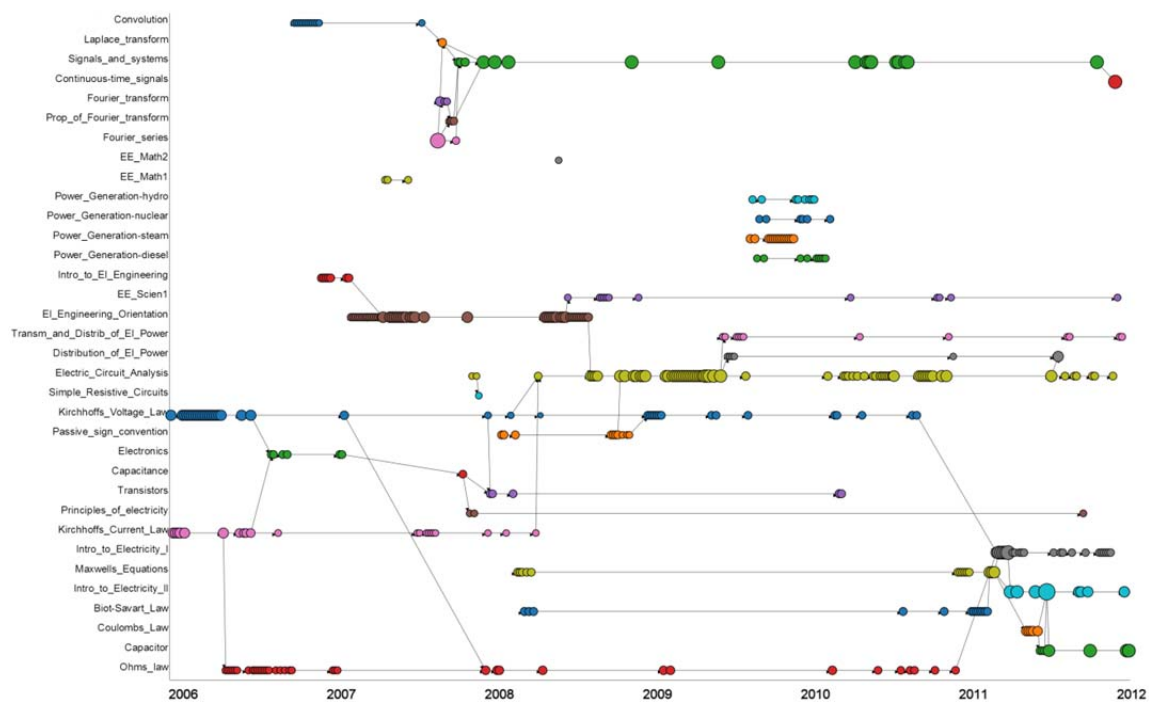

Figure 6. Multiple main paths in the electrical engineering domain

A new cluster of pages from 2006 and 2007 appearing now on the main paths covers the topic of signals and systems. The remaining separate pages of the main paths relate to mathematical tests and to a course on electrical power generation from 2008, all written by the same single author, as indicated by the revision logs.

The core cluster of the discussed main path now consists of many new articles covering basic electrical laws. On the main paths also appear pages from other topics structured as courses: on orientation to the domain and on transmission and distribution of electrical power. The important position in the DAG of the electric circuits topic in between the early orientation to electrical engineering and the later introduction to electricity explains why it is part of the main path in Figure 5. Although the enlarged core cluster consists of different courses and groups of topics, we found strong cross-participation of contributors across the pages in the cluster as we consulted the revision logs. In addition to the pages being thematically close, the cross-collaboration of authors presents an additional reason for the emergence of this large connected cluster.

Overall, our study showed that electrical engineering was a more compact and coherent domain than biology in the Wikiversity community. Many contributors collaborated over longer periods of time and a large number of pages, creating highly interrelated learning resources. Thus, materials organized as online courses were authored by a large number of people and serve general interests instead of that of a limited number of students for a limited period of time. The electrical engineering domain is an example of a self-organized learning community with enough time to build collaborative structures of practices and artifacts. Evaluated by main path analysis, the development resulted in more tightly interwoven topics than in the biology domain. Overall, the method revealed one large cluster of articles in both domains, as well as a few smaller ones, representing the core knowledge in those domains. This method allows for a subsequent analysis of the development of the topics over time and of the distribution of participation of their authors.

4.4   Author Profiles and Roles

After the overview of the main paths in the two domains, we turn to the analysis of the authors contributing to pages off as well as on the main paths. Here, we used the main path analysis results in combination with the revision logs in the dump. As explained in section 3.1, Wikiversity is an open virtual space and so there is no standard guideline on how authors should interact and use the environment. However, our data revealed differences in the contribution activity profiles of authors that can be interpreted in terms of a division of roles in the process of collaborative learning in a Wikiversity knowledge domain. We started by calculating for each author the number of edits and different edited pages and focused on the profiles of prominent authors who stood out among the large group of low contributors. Forty-six percent of the authors in the biology domain and 51% of the authors in the electrical engineering domain had minimal participation, just making a single edit without hyperlinks in the DAG. Respectively, 30% and 27% of the authors in the two domains who had at least one contribution on the multiple main paths did not make any other contribution. This highly skewed distribution of participation in online environments is a well-documented fact (Rafaeli & Ariel, 2008). More specifically, we see that authors who have a contribution on the main paths are generally less likely to make only a single contribution. According to this evidence, main path contributions can be interpreted to indicate high involvement in the community.

According to our interpretations of the profiles of active authors, we identified several categories of contributors: first, the role of *specialists*, who made many edits to only one or a few pages; second, the role of *maintainers* with a relatively high number of edited pages and a relatively low number of edits; third, the role of *leaders* with an outstanding number of edits and edited pages. As we show in the following, the interpretation of these roles was only accurate after taking the results of the main path analysis into account.

The investigated articles, and thus the contributions to them, are not of equal importance to the collaborative learning process of the community. Many articles are short stubs not interlinked with any other articles within the corresponding category. Such isolated and largely unimportant articles are not part of the main paths in a domain. Therefore, the results of the main path analyses in both domains of the study can enhance the analysis of the author roles by qualifying the number of contributions that lie on the main paths. As mentioned above, the SPC method of identifying a single main path leads to a strong focus on a small number of revisions and articles on a narrow topic. Hence, in this paper the author profiles are related to the extracted multiple main paths described in the previous subsections. Using the main path analysis in this way, a more adequate view on activity and division of roles of authors is achieved.

4.4.1   Author roles in biology

The three analyzed author roles in the biology domain are presented in the rows of Table 2 through the contribution profiles of distinctive sample authors. Each role is subdivided into type A and type B according to whether any of the contributions of an author are part of the main paths. The author activity in total and on the main paths is grouped in blocks containing the number of edits, edited pages, and edits with hyperlinks. As explained in section 3.2, hyperlinks represent knowledge flows between pages. Thus, the edits introducing a hyperlink and the edits referred through a hyperlink by another edit are important and should be regarded separately.

Table 2. Sample authors with a distinct role in the biology domain

Author profile	Author ID	Edits in total	Edits on multiple main paths	Pages in total	Pages on multiple main paths	Hyperlinked/ / hyperlinking edits	Edits with links on multiple main paths
Specialist A	278565	468	0	1	0	0 / 0	0 / 0
Specialist B	348476	10	10	1	1	0 / 0	0 / 0
Maintainer A	9357	35	0	31	0	0 / 0	0 / 0
Maintainer B	21778	43	9	41	8	0 / 1	0 / 0
Leader A	263421	1966	0	729	0	0 / 0	0 / 0
Leader B	20	552	154	112	20	31 / 35	25 / 20

The first rows, the specialist A with ID 278565 has the third highest number of edits in the domain, but these edits were all made to the same single page, moreover, none of them is part of the multiple main paths. This example shows that output quantity — the number of contributions — does not necessarily correspond to output quality — the importance for the evolution of discourse in a Wikiversity knowledge domain. The example of author 348476 adds to this finding. With ten edits in the domain in total, this is the most prolific author among the type B specialists — authors who are specialized in one single page and have at least one edit on the main paths. The low rate of activity of such specialists with important contributions would normally suggest that they should be regarded as low contributors. In

the next rows, the type A and B maintainers 9357 and 21778 similarly show a low to middle rate of contribution. Maintainers mostly make small formal changes unrelated to the content of the edited Wikiversity pages. They correct spelling mistakes, organize the categorization, and sometimes also set hyperlinks, as does author 21778. Such authors typically contribute to very different domains and topics at the same time. Most of their contributions that appear on the main paths can be regarded as coincidental as they fall within a chain of important updates of the page content made by other authors. Table 2 further shows that the most prolific contributor and a type A leader in the biology domain, author 263421, didn't make a single important contribution on the main paths. A closer look into the data revealed that this author used Wikiversity to build a database on specific genes. This voluminous project was not much related to the other core topics in biology. Type B leaders, such as author 20, whose edits sometimes appear on the main paths, seem to play the most important role in the domain. Besides having the highest number of contributions on the main paths, this author also has the highest number of edits with hyperlinks. Further analyses of the data showed that authors with edits on the main paths tend to have more contributions and especially more interlinked edits than authors without edits on the main paths. Indeed, by the design of the method itself, hyperlinked and hyperlinking edits are more likely to occur on the main paths.

4.4.2 Author roles in electrical engineering

Table 3 presents the analysis of author roles in the electrical engineering domain following the structure of Table 2.

Table 3. Sample authors with a distinct role in the electrical engineering domain

Author profile	Author ID	Edits in total	Edits on multiple main paths	Pages in total	Pages on multiple main paths	Hyperlinked/ / hyperlinking edits	Edits with links on multiple main paths
Specialist A	858	44	0	1	0	0 / 0	0 / 0
Specialist B	292570	6	6	1	1	0 / 0	0 / 0
Maintainer A	3705	19	0	17	0	0 / 0	0 / 0
Maintainer B	8437	34	8	27	4	0 / 0	0 / 0
Leader A	32	245	0	75	0	1 / 0	0 / 0
Leader B	19038	867	114	133	14	20 / 35	8 / 8

As argued in section 3.3, the two domains are marked by a number of differences. Nevertheless, the studied author roles are identifiable in the same way in both domains, so the inferences about the authors in biology made in the previous subsection also apply for the authors in electrical engineering. The only difference worth mentioning is that author 19038, a type B leader in Table 3, has the highest number of contributions among all authors in the domain and at the same time has contributed the highest number of edits on the main paths. This case still corresponds to the conclusion that important

authors are distinguished not just by a high number of edits but also by significant contributions appearing on the main paths.

5. TECHNOLOGICAL IMPLEMENTATION

The analysis processes described in this paper have been integrated into our network analytics workbench (Göhnert, Harrer, Hecking, & Hoppe, 2013). A form of this workbench was used in the recent EU project "SISOB,"[4] which had the goal of measuring the influence of science on society based on the analysis of (social) networks of researchers and created artifacts. One area of research in this project was *knowledge sharing*. Thus the analysis techniques based on main path analysis presented in this paper were also of essential value in the project context.

We conceive workbenches as a general type of software environment designed to serve active and skilled users, without assuming the users to be computer experts. We have decided to develop a network analytics workbench as a web-based environment for several reasons, such as ease of deployment, access and update, and independence of the local computing facilities and devices. An important part of our experience with network analysis and network analysis tools is the need to combine several tools even for a single analysis process. The use of several tools sometimes also results in the need for conversion between the different data formats used by these tools. Therefore one important goal behind the development of the network analytics workbench is the integration of multiple tools and conversion mechanisms into one interface.

The workbench provides readily available processing chains for known use cases and furthermore allows for setting up new ones. The user interface (UI) is built upon a pipes-and-filters metaphor for processing chains in order to reduce the complexity of the underlying system for users who are not computer experts. An example of the UI that has been created using the WireIt[5] JavaScript library can be seen in Figure 7. In using the pipes-and-filters metaphor and being web-based, the workbench is similar to mashup projects like YAHOO pipes.[6]

In contrast to these projects, the actual processing of data in our workbench is not part of the user interface code itself but is done by a multi-agent system controlled by the workbench. The multi-agent system approach allows for combining several mostly independent tools into one workflow. These tools can be either pre-existing or newly developed. Examples of existing tools successfully integrated into the workbench are the network text analysis tool AutoMap (Diesner & Carley, 2005), the network analysis tool Pajek (Batagelj & Mrvar, 1998), and a wrapper for the R language.[7] Examples for newly developed components are a MediaWiki extraction component based on the mechanism presented in this paper and a main path analysis filter also used for the analyses presented in this paper. The communication between the web-based user interface and the agents is based on the SQLSpaces (Weinbrenner, Giemza, & Hoppe, 2007), an implementation of the tuple space architecture (Gelernter, 1985). From the user interface a description of the constructed workflow is posted into the SQLSpaces server, which contains a message for each agent (filter) type that is part of the workflow. These messages contain information about the input data and the parameter configuration of that filter.

---

[4] http://sisob.lcc.uma.es/
[5] http://neyric.github.com/wireit/docs/
[6] http://pipes.yahoo.com/pipes/
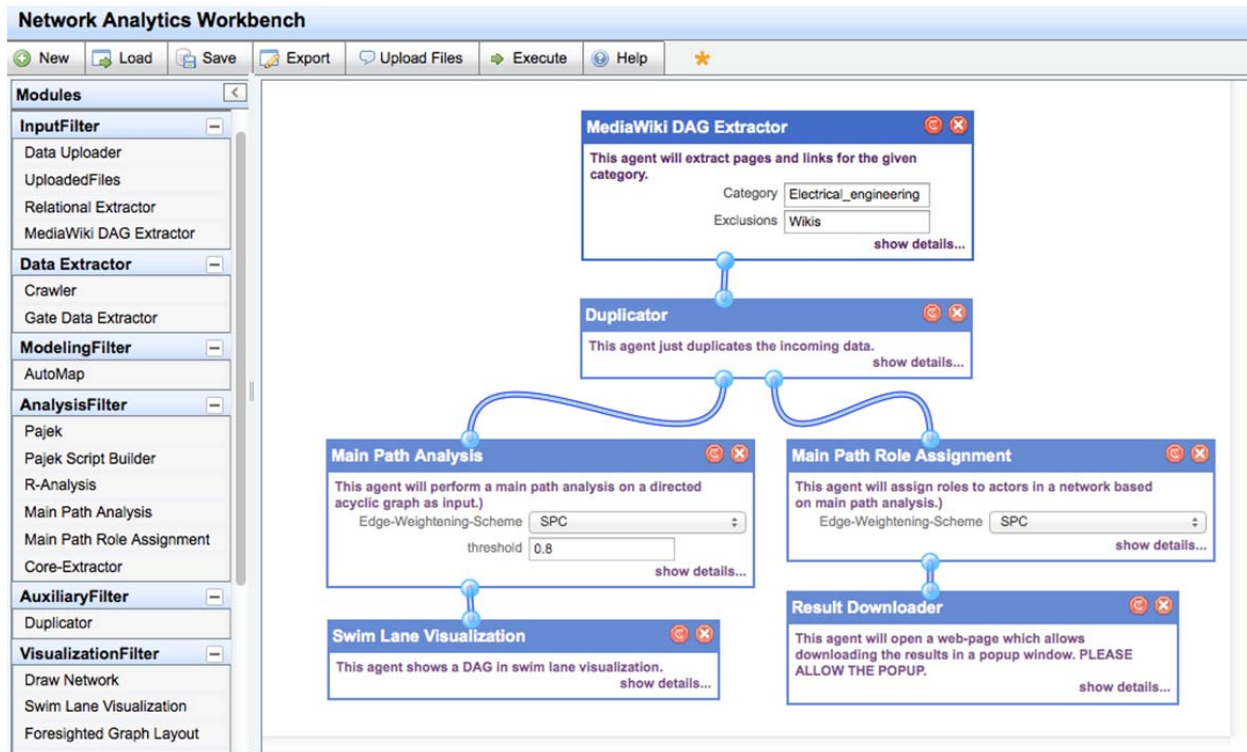[7] http://www.r-project.org/

Figure 7. Screenshot of the Network Analytics Workbench

Figure 7 shows one of the workflows used for the analyses described in this paper. The first filter is used to provide input for the following filters. In this case, the filter connects to a MediaWiki database with Wikiversity data and creates a DAG for a given category from it. The extraction process follows the approach outlined in section 3.2 of this paper. The filter accepts two parameters. The name of the category for which the DAG should be extracted is a mandatory parameter. The second parameter accepts a list of categories to be excluded from the search and is optional. The next filter in the workflow presented here just duplicates all input into two parallel outputs. Thus, it allows for performing different analyses on the same possibly preprocessed input data in one workflow. In this example, the two outputs are used to perform main path analysis and analysis of author profiles in the same category of a wiki, as presented in this paper (sections 3.3 and 3.4). On the left side, the Main Path Analysis filter allows for selecting a weighting scheme to be used in the main path analysis and for defining a threshold for the multiple main path analysis. The results of this filter are then visualized using the swim lane metaphor also used throughout this paper. The other branch of the duplicator leads into the Main Path Role Assignment filter, which generates the tables used for the author profile analysis as described in section 3.4. These tables are then fed into the Result Downloader, which allows for downloading these results onto the local machine for further usage.

6. CONCLUSION

With the help of the main path analysis, we detected the core topics in the two Wikiversity domains of biology and electrical engineering. While biology had much broader scope, the collaboration of the authors was weaker. The resulting main paths had a similar size and structure to the main paths in electrical engineering, which was a small coherent domain with a relatively large group of authors and a

higher necessity for collaboration. Thus, the small ratio of main path versus other articles in the biology domain compared to the electrical engineering domain could be explained through differences in the level of collaboration among the authors revealed by the revision logs.

The exemplary results of the presented empirical study may be useful for the Wikiversity community as a whole. As it seems, some scientific domains like biology might benefit from strengthening of collaboration. Additional analyses may be helpful to choose appropriate directions for development, but our results point to the need for better coordination of the disparate topics in this domain. The main path analysis can also orient participants by showing them the importance of the topic they are working on. It can also reveal important reference points to other core topics in the field. A beginning contributor can be aided by a presentation of the main paths with the decision to add to an existing strand of knowledge development or to start a new peripheral one. An advanced participant in the community may benefit from the analysis as a historical reconstruction of the shared knowledge-building process, in order to compare his or her own visions and goals with the actual knowledge development of the community and to discover topical gaps necessitating further efforts. With some additional work to adapt and standardize the analysis and the necessary interventions relative to the specific goals within an educational context, the main path analysis can be used to support and even take the load off a teacher or coordinator of knowledge building.

Our approach presented in this paper is the first application of scientometric methodology for analyzing the flow of ideas in the context of an open learning wiki environment. Using the examples of the biology and electrical engineering domains in Wikiversity, we showed how main path analysis can be employed to analyze the collaborative creation of various knowledge artifacts and the learning processes of the online community. Our methods have been embedded into a web-based analytics workbench that supports the definition and re-use of analysis modules in a user-friendly visual environment.

The paper presented a procedure for creating directed acyclic graphs from wiki data and for illustrating the obtained main idea flows in swim lane diagrams. Our visualization technique allows for a unified view of knowledge flows in a network of artifacts with multiple relationships. The main path analysis results were helpful in understanding the differences in the collaborative structure of two scientific domains in Wikiversity. The results further facilitated the characterization of different roles that authors have in the community. We found that the total rate of contribution was not a sufficient criterion for identifying the most important authors in a domain. But, as the role of maintainers demonstrates, some contributions on the main paths may also not testify to the importance of an author. Instead, the total number of contributions should be evaluated in combination with the number of contributions that appear on the main paths.

For our future work, we plan to elaborate on the characterization of contributions and contributors with respect to the main paths of development in other educational knowledge-building scenarios. It appears promising to provide moderators, teachers, tutors, or the productive teams themselves with results of such analyses, in order to support reflective practices (Schön, 1983). This will raise further challenges regarding visualization and cognitive ergonomics.

ACKNOWLEDGMENTS

90

REFERENCES

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal for Asynchronous Learning Networks, 7*, 1–23.

Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv: Computer Science. http://arxiv.org/abs/cs/0309023

Batagelj, V., & Mrvar, A. (1998). Pajek: A program for large network analysis. *Connections, 21*, 47–58.

Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merriënboer (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 73–78). Oxford, UK: Elsevier Science.

Bonk, C., & Cunningham, D. (1998). Searching for learner-centered, constructivist, and sociocultural components of collaborative educational learning tools. In C. J. Bonk & K. S. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse* (pp. 25–50). Mahwah, NJ: Erlbaum.

Cakir, M., Xhafa, F., Zhou, N., Stahl, G. (2005). Thread-based analysis of patterns of collaborative interaction in chat. Proceedings from: *International Conference on Artificial Intelligence in Education* (pp. 120–127). 18–22 July. Amsterdam, The Netherlands.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, *3*, 105–122.

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*, 87–103.

Diesner, J., & Carley, K. (2005). Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis. In V. K. Naraynan & D. J. Armstrong (Eds.), *Causal mapping for information systems and technology research: Approaches, advances, and illustrations* (pp. 81–108). Harrisburg, PA: Idea Group Publishing.

Gelernter, D. (1985). Generative communication in Linda. *ACM Transactions on Programming Languages and Systems (TOPLAS), 7*(1), 80–112.

Göhnert, T., Harrer, A., Hecking, T., & Hoppe, H. U. (2013). A workbench to construct and re-use network analysis workflows: Concept, implementation, and example case. Proceedings from ASONAM 2013: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1464–1466)*. Niagara Falls, Canada, 25–28 August. New York: ACM Press.

Halatchliyski, I., Kimmerle, J., & Cress, U. (2011). Divergent and convergent knowledge processes on Wikipedia. Proceedings from CSCL 2011: *Connecting Computer-Supported Collaborative Learning to Policy and Practice* (Vol. 2, pp. 566–570)*. 4–8 July. Hong Kong, China: International Society of the Learning Sciences.

Halatchliyski, I., Moskaliuk, J., Kimmerle, J., & Cress, U. (in press). Explaining authors' contribution to pivotal artifacts during mass collaboration in the Wikipedia's knowledge base. *International Journal of Computer-Supported Collaborative Learning.*

Halatchliyski, I., Oeberst, A., Bientzle, M., Bokhorst, F., & van Aalst, J. (2012). Unraveling idea development in discourse trajectories. Proceedings from ICLS 2012: *The Future of Learning: 10th International Conference of the Learning Sciences* (Vol. 2, pp. 162–166). 2–6 July. Sydney, Australia: International Society of the Learning Sciences.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science, 28*, 115–152.

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 from CSCL 2007: *Mice, Minds, and Society: The Computer Supported Collaborative Learning Conference* (pp. 267–276)*.* New Brunswick, NJ, 16–21 July. Mahwah, NJ: Lawrence Erlbaum Associates.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 16569–16572.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*, 39–63.

Jeong, A. C. (2003). The sequential analysis of group interaction and critical thinking in online. *The American Journal of Distance Education, 17*, 25–43.

Knorr-Cetina, K. (2001). Objectual practice. In T. R. Schatzki, K. Knorr-Cetina, & E. von Savigny (Eds.), *The practice turn in contemporary theory* (pp. 175–188). London/New York: Routledge.

Liu, J. S., & Lu, L. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology, 63*, 528–542.

Mika, P. (2007). *Social networks and the semantic Web*. New York: Springer.

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

Popper, K. R. (1968). Epistemology without a knowing subject. *Studies in Logic and the Foundations of Mathematics, 52*, 333–373.

Rafaeli, S., & Ariel, Y. (2008). Online motivational factors: Incentives for participation and contribution in Wikipedia. In A. Barak (Ed.), *Psychological aspects of cyberspace: Theory, research, applications* (pp. 243–267). Cambridge, UK: Cambridge University Press.

Reffay, C., & Chanier, T. (2002). Social network analysis used for modelling collaboration in distance learning groups. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *ITS 2002 Lecture Notes in Computer Science, 2363* (pp. 31–40). Berlin, Heidelberg: Springer.

Reinhardt, W., Moi, M., & Varlemann, T. (2009). Artefact-actor networks as tie between social networks and artefact networks. Proceedings from CollaborateCom 2009: *5th International Conference on Collaborative Computing: Networking, Applications and Worksharing* (pp. 1–10)*.* Washington DC, 11–14 November. IEEE Computer Society.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences, 3*, 265–283.

Schön, D. (1983). *The reflective practitioner: How professionals think in action.* London: Temple Smith.

Schrire, S. (2004). Interaction and cognition in asynchronous computer conferencing. *Instructional Science, 32*, 475–502.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*, Cambridge, UK: Cambridge University Press.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning, 5*, 5–42.

Suthers, D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. In P. Long, G. Siemens, G. Conole, & D. Gasevic (Eds.), *Proceedings of the 1st international conference on learning analytics and knowledge, LAK'11* (pp. 64-74). Banff, Alberta, Canada, 27 February–1 March. New York: ACM Press.

Suthers, D., & Verbert, K. (2013). Learning analytics as a "middle space". In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the 3rd international conference on learning analytics and knowledge, LAK'13* (pp. 1-4). Leuven, Belgium, 8–12 April. New York: ACM Press.

Wasserman, S., & Faust, K. (1994). *Social networks analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Weinbrenner, S., Giemza, A., & Hoppe H. U. (2007). Engineering heterogeneous distributed learning environments using tuple spaces as an architectural platform. Proceedings from ICALT 2007: *7th IEEE International Conference on Advanced Learning Technologies* (pp. 434–436). Niigata, Japan, 18–20 July. IEEE Computer Society.

Zaïane, O., & Luo, J. (2001). Web usage mining for a better web-based learning environment. Proceedings from: *Conference on Advanced Technology for Education* (pp. 60–64). 27–28 June. Banff, Alberta, Canada.

Zeini, S., Göhnert, T., & Hoppe, H. U. (2012). The impact of measurement time on subgroup detection in online communities. Proceedings from ASONAM 2012: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 389–394). Istanbul, Turkey, 26–29 August. IEEE Computer Society.