Emerging Research Topics in Social Learning

Ralf Klamma

Information Systems & Databases, RWTH Aachen University, Ahornstr. 55, 52056 Aachen, Germany, klamma@dbis.rwth-aachen.de

Abstract

Despite the fact, that an uncountable crowd of technology affine learners has entered the Internet with fast connections lines, little support exists for scalable learning support technologies. In this paper, we explore research topics which address the new needs of long tail learning and technologies which may help to overcome the current situation. We have combined social network analysis, visualization and recommendation technologies supporting social navigation, positioning and reflection in learning networks. We demonstrate first results from recent research in case studies. In the future we will add game theoretic multi agent simulation.

Keywords

Learning networks, social network analysis, network visualization, recommender systems, game theory, multi agent simulation

Introduction

Learning networks are becoming of major interest in Europe. Connected with the Lisbon and Stockholm meetings, the European Commission and the national governments are trying to strengthen the competences of Europeans in the emerging knowledge society. Driven by external competition with other nations, e.g. China and the US, societal developments, e.g. obsolescence, and internal crises of the formal educational system, e.g. the Pisa shock, a lot of initiatives are on their way to support life-long learning processes in the knowledge society, such as the Life-long Learning programs Comenius, Erasmus, Leonardo da Vinci and Grundvik of the European Commission (Bienzle et al. 2007). Web 2.0 (O'Reilly 2005) and social software often plays a role in those programs because it lowers the barriers to take part in those initiatives which naturally emerge in the form of networks. This situation leads to two new situations. First, learners now are facing an abundance of available learning materials. The learning process becomes also a decision making process which learning materials or learning partners to choose. Connectivism[1] (Network Pedagogy) is one of the new social learning theories focusing on those emergent forms of learning networks. So, social learning is not equal to learning networks but by definition learners have to be connected. Connectivism competes with theories like communities of practice, social constructivism, socio-cultural theory, situated learning (Bienzle et al. 2007). Key theoreticians of connectivsm are George Siemens (2006), Stephen Downes (2005), Jay Cross (2007) and Graham Atwell (2007). Principles of connectivism are according to Stephen Downes:

- Learning is a process of *connecting entities*
- *Nurturing and maintaining connections* is needed to continual learning
- Ability to *see connections* between fields, ideas, and concepts is a core skill
- *Capacity to know more* is more critical than what is currently known
- *Decision-making* is itself a learning process

Another theoretical basis for the new forms of social learning is the musing about the nature of digital goods in general. The seminal marketing book by Chris Anderson (2006) about "The long tail" created also a debate in technology enhanced learning since similar prerequisites like the almost universal access to learning resources with together with a limited access to learning partners. This coined the term "The long tail of learning" (Brown & Adler 2008). In the long tail of learning there is a shift of focus away from the learning resources to the learners and their limitations. Decision making about connections becomes critical in the long tail of learning. Consequently, learning repositories which contain the learning resources and learning management systems

[1] http://www.elearnspace.org/Articles/connectivism.htm

ISBN 978-1-86220-225-2

which are taking care about the management of formal aspects of learning are challenged by personal learning environments (PLE) helping learners take control of and manage their own learning (van Harmelen 2006). But only PLE which support learners in the decision making processes needed to get access to appropriate learning resources and learning communities will be of use.
With the broadening use of social software many interesting datasets have been created lately which can be used for studying traces left by learners on the Web. Thus, we can study decision making processes from historical data. Because of the still high costs and the expert knowledge needed, it is only feasible for institutions to analyze and learn from historical data representing connections between learners. Learners or the media created by the learners are represented as nodes in a network while the relationships between learners or the relationships between the media created by the learners are represented by links. The learning network can be derived explicitly by numerous techniques mainly from sociology (cp. Wassermann and Faust 1994 for an overview) or implicitly by the analysis of the media networks. One of the questions followed in this paper how well established techniques can be applied and developed further not only to support experts like sociologist and money-backed institutions but also self-regulated learners with little knowledge about analytic techniques and the management of large, complex datasets. But first of all, this paper tries to collect some research topics for social learning theories which are based on a different set of observation techniques like ethnography or are deeply rooted in practice while the proponents are educational scientists of even teachers. It is not the aim of the paper to replace well established methods of data collection. But the new social software tools in technology enhanced learning not only enable new forms of connected learning but also produce heavy datasets. These datasets are ideal research data but demand new methods for data management, data analysis, data visualization and forecasting.
The rest of the paper is organized as follows. In the next section we introduce the emergent research topics in social learning in more detail. In the section following, we present three case studies which have been already conducted using the introduced topics before we discuss some lessons learnt and future work.

Research Topics

Analytic challenges for social learning theories in general and connectivism in particular involve the empirical evidence of claims and testing out success factors. If connectivism is claiming that decision making in making connection is success critical for learning, we have to investigate how we can do research about it. We are deeply convinced that social network analysis (SNA) is the foundation for analyzing the success or failure of social learning. But as we are convinced that there is not analysis possible we are convinced that SNA is only part of the solution. We may not want to determine the ratio of the impact of learning networks structure and the impact of content, but in the end only combining structural effects of networks with psycho-pedagogical insights about content, motivation, etc. will create an evidence-based theory of social learning. Visualization of networks will play as such important role as social presence, navigation, and filtering. Recommendation technologies and game-theoretic support for taking decisions in the self-regulated learning process are helping learners to define their own ways since informal learning will rather increase the cognitive load of learners instead of decreasing it. For learning communities tools for self-reflection, self-assessment and agenda-setting are needed. Here, we cannot discuss all research approaches in depth. Therefore, we will just point out in which directions research in learning networks may go.
The goal of SNA is discovering information about social relationships (Klamma et al. 2007). Beside graph theory, mathematical and statistical aspects, there are other phenomena, hardly detectable outside the network analysis (see Barabási 2007 for an overview). A typical network characteristic is the small world phenomenon, which has been discovered by Stanley Milgram (1967) and got famous as “Six Degrees Of Separation”. In scale-free networks (Newman 2003) the nodes are connected to k other nodes with a probability of p(k) $\sim k^{-\gamma}$, which is in accordance to a power law. This means for social networks that the distribution of activity is unequally distributed. Few actors hold a large extent of connections whereas many actors are linked to only few others (long tail). For SNA, graph theoretic connectivity is important, because it influences the robustness, cohesiveness, etc. (Brandes and Erlebach 2005). Most important analysis aspects in social networks are centralities (Freeman 1979). Depending on her position, an actor can exert more or less influence on other actors; thus, she holds a certain amount of power and importance. The following centralities are the most relevant:
The **degree centrality** is just the degree of a node, the number of adjacent edges. This is significant, because an actor with many connections may influence many others. The **closeness centrality** is distance-based and denotes how far the node is from all other actors in the network. The **betweenness centrality** is a measure for the influence an actor can exert - for example on the communication between others. For this, the number of

shortest paths unavoidably containing node u must be calculated. With this fundamental methodology (see Wasserman and Faust 1994 for an introduction), a meaningful and detailed analysis of learning networks is possible. But the analysis is not the only task; the visualization makes its own demands going far beyond formal aspects dealing with cognitive and social issues.

Recommender Systems (RS) are a class of systems designed to help people deal with information overload, incomplete information and their capacities to make the evaluative decisions. In commerce, many applications have applied it successfully, e. g Amazon and MovieLens. Recently, the RS research community focus on how to make RS applications effectively support personal decision making. Schafer et al. (2001) listed six techniques which are used in most of recommender systems: raw retrieval, manual selection, statistical summarization, attribute-based, item-to-item correlation (so called content-based recommendation, Sawar et al. 2001), and user-to-user correlation (so called collaborative filtering, Burke 2002). Collaborative filtering (CF) and content-based are not mutually exclusive to each other, and there are many efforts to integrate them in so-called hybrid approaches in order to improve the quality of recommendations.

Game Theory is the formal study of decision-making where player's decisions are strategic reactions to other players' actions. In 1928 John von Neumann introduced the maximin criterion to solve simple strategic, non-zero games. He showed if a player benefits one payoff than it is necessary the other lose one payoff, i.e. the own strategy depends on the rival's strategy. Game theory was established as a field in its own right after the 1944 publication of Theory of Games and Economic Behavior by von Neumann and Oskar Morgenstern. In 1950 John Nash introduced a new concept, called Nash Equilibrium later. Here, all players choose actions which are best for them given knowledge about their opponents' choices. Since the 1970s, Nash equilibrium has driven a revolution in economic theory.

An agent is a computational mechanism that exhibits a high degree of autonomy, performing actions in its environment based on information (sensors, feedback) received from the environment (Panait and Luke 2005). A Multi-Agent System (MAS) is a collection of heterogeneous and diverse intelligent agents that interact with each other and their environment (Siebers and Aickelin 2008). Because of the complex nature of many networks identifiable in technology enhanced learning the application of purely analytic techniques is often leading only to limited results or is computationally too costly. Therefore, multi agent simulation is an appropriate alternative to find trends from datasets and engineer the desired behavior of agents to track the consequences in the simulation runs.

In the case studies presented in the following these techniques have been used to research datasets from technology enhanced learning. While all case studies include some basic SNA, the second case study is about recommender systems and the third case study is about multi agent simulation.

Case Studies

Since learning networks have gained so much attention the need for evaluation of networking activities and the elaboration of success factors have become of big importance. In the following we present in short three case studies conducted from or with us to demonstrate the benefits of the research approaches mentioned in the last section.

PROLEARN Academy – A TEL PhD network for Europe

The PROLEARN Academy was an EU funded initiative of the PROLEARN network of excellence (2004-2007). Among other aims, the goal of the Academy was to establish a European network of PhD students in the area of Technology Enhanced Learning. Main means for this process were Summer Schools, Doctoral Consortia and a Mentorship Database. All these events and tools can be tracked at http://www.prolearn-academy.org. In the questionnaires we asked for existing and newly established relationships between PhD students at the events. We analyzed the results and applied the measure described above on the datasets created to identify well-connected PhD students and students brokering information in the network of students. For a similar purpose we analyzed the co-authorship relationships among the PhD students and the comparison of results from the different years 2005-2008 revealed that the clustering of students according to their topics became very obvious. Also, the overall density of the co-authorship network was increased. Data were assembled from the literature database at the website. Figure 1 presents PROLEARN Academy screenshot.

ISBN 978-1-86220-225-2

Figure 1: PROLEARN Academy Portal for training of PhD students (screenshot)

One of the successes of the network is that all events and tools survived the funding phase of the network. In 2009 there will be again a summer school funded by a consortium of almost every funded TEL project in Europe and a doctoral consortium at the EC-TEL 2009. Also the mentorship database was recently updated to cover new Web 2.0 tools and thus enabling better connectivity among the former PhD students. Moreover, students now can enter links to their published thesis work. This helps distributing research results in the European TEL research community. For a deeper analysis of the impact of the events organized by the PROLEARN academy we constantly monitored the use of the portal and all services mediated through the portal.

To summarize, we did SNA on the summer schools and a co-authorship analysis of all summer school participants. The feedback on the questionnaires we used for the SNA revealed that students coming to the summer schools were enabling connections to other PhD students and that the intenseness of collaboration among the PhD students were heavily influenced by the summer schools. Analysis of later summer schools revealed also that repeated stays at the summer school even more broadened and deepened the relationships among the students' community. Details about the analysis can be found in the public deliverables of the PROLEARN network of excellence at http://www.prolearn-project.org. For good or for bad, the summer schools and the follow up doctorial consortia shaped the TEL community by establishing a network of young researchers. Still, recommendations for learning and research were facilitated by experienced researchers. Based on these positive experiences, in the succeeding network of excellence STELLAR http://www.stellarnet.eu longitudinal SNA has become a cornerstone of monitoring and evaluating the quality of the network connections.

Analysis and Visualization of the European Schoolnet – EVA

The European Schoolnet – responsible for the support of thousands of schools in Europe - is conducting the eTwinning project. The eTwinning project was founded in 2005 and belongs to the eLearning Program of the European Commission. The concrete goal is an advancement of cooperation between European schools by the usage of ICT. All states of the European Union (Norway and Iceland inclusive) are able to participate; their school and join at any time and form partnerships with as many other schools as desired. Thereby, neither the type of school nor the engaging teaching staffs is constrained. An important concept of eTwinning is the creation of collaboration opportunities among European schools. Therefore, it creates so-called TwinSpaces, collaboration tools for project between at least two European schools from at least two different countries. Administrated by the founding teachers, the members of the project can communicate, exchange resources, document the progress and publish the outcomes. There are numerous auxiliary tools achievable at the TwinSpace, for example chat rooms, a shared calendar or the possibility to create web pages or forums. The eTwinning network has not been analyzed yet. Although the portal contains information about the size of the network, a detailed and methodical analysis was missing. In the first phase of funding more than 45.000 schools in Europe used the opportunity to conduct a joint project. This left a huge database with schools, project and teacher data. EVA – the social network analysis visualization tool for eTwinning was created to visualize the different networks in the database to give feedback to the teachers about their connections and to facilitate future collaboration by recommending possible partner schools in other countries. In Figure 2 one can see the EVA prototype[2]. The prototype is especially designed for teachers which are not experienced in social network analysis. It is offering very simple visualization tools in a browser without further requirements of installing additional software.

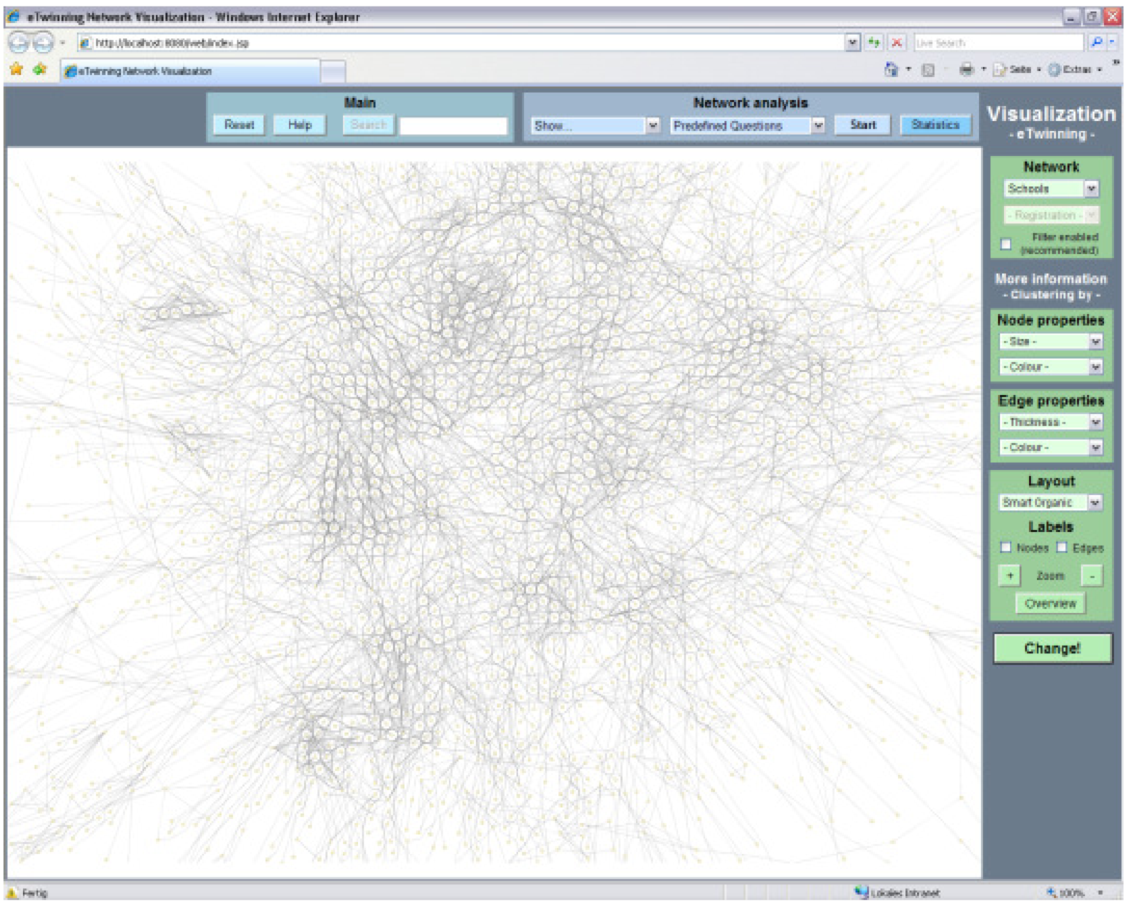

Figure 2: Screenshot of EVA (eTwinning Visualization and Analysis)

While we will publish detailed results of the study in (Breuer et al. 2009), it can be said that while the cooperation arranged by European countries are numerous, the activity of participating schools and teachers still leaves room for improvement. The connectivity of the project networks demonstrates that a rather small fraction of teachers initializes the biggest fraction of the projects, thus discloses a long tail of learning. Recommendations for teachers and schools can be taken from the statistics, the visualization and concrete pre-defined queries. However, there is no explicit recommendation engine, yet.
In 2009 we managed to get a grant by the Life Long Learning Programme of the EU to study interventions in the eTwinning networks. The new joint project by EUN, RWTH Aachen University and the Open University of

[2]http://vermeer.informatik.rwth-aachen.de:9080/eVA

ISBN 978-1-86220-225-2

the Netherlands will start in December 2009. With the datasets and the experiences gained in our case studies, we can now add game theoretic support for studying and analyzing decision making processes in learning networks. We combine game theories and MAS for recommending decisions in learning networks. We discuss it here for the example of datasets taken from eTwinning.

A network is a Nash network if each agent is playing Nash equilibrium. In a Nash network, a link between agents allows access. In eTwinning the game played by the agent is a network formation game. Assume the costs of link formation are incurred only by the teacher who initiates the link, than the network could be modeled as a one-sided link formation non-cooperative game. The strategy of an agent is the set of agents with whom he forms links. In the first phase, we choose part of all data of teachers (e.g. one third) as training data that will be simulated as agents to find rules that predict best strategies with Nash equilibrium and stochastic approaches in recommender systems. After confirmation of prediction rules, the data not used so far are used to experiment and perfect the rules. Every teacher has a goal function she tries to maximize her own utility. Maximization of utility depends on which strategies are used from a teacher. The recommender agent can make recommendations for new data (testing data). We use CF for finding out the groups of similar teachers. However, similarity is not assuring the win-win situation in collaboration. Payoffs with all other agents calculated from the agent most similar to the new teacher, serve as rating in regular recommender systems. The teacher with highest rating has the properties of the most suitable partner P. Since an agent represents a group of teacher, we can go a step further computing the correlation between T and each teacher in this group. Then a particular recommendation of a teacher can be made, instead of description of a suitable partner. To evaluate the recommender system is to evaluate the suitability of recommendation again with the data taken from the real world dataset. The historical activities of the teachers are used to assess precision of recommended partner. First and foremost, the outcomes of projects that a teacher attended provide evidence. First results from simulations in the REPAST MAS showed that threshold strategies simulate better than best n strategies and that more general payoff functions lead to better approximation of Nash networks while more specific functions modelling the domain knowledge fit better to the real data in the eTwinning database.

Identifying Communities in Computer Science – AERCS

Especially for young researchers finding a research community and establishing a stable relationship is extremely challenging. In an ideal world (like in PROLEARN), senior advisers guide young researchers finding their place in the research community. But, since speed of research and research output is getting faster year by year just be the growing number of researchers in the world, science itself has changed. Everyday a number of new workshops, conferences, journals and other events is showing up, rapidly communicated through the Internet. Senior researchers still keep their own communities alive but just as in a bee hive young researchers establish new communities seeking new opportunities. Recommending communities fitting the needs and the interests of young researcher may become an interesting challenge for research itself. A number of digital libraries like ACM Digital Library, IEEE Explore, and DBPL have been established in the last years. Recently, scientific event announcements are also collected and distributed, e.g. eventseer.net is used in the computer science community to track events.

AERCS[3] (Klamma et al. 2009) is using those services as data sources and mashing them up with network visualization. In Figure 3 one can see the visualization of the VLDB research community developing over 16 years (1990-2996). Nodes are VLDB authors and links means co-authorship of a VLDB paper. Just from the visual impression one see the development of a stable core and a vivid periphery. These are clear indicators for a high quality research community.

Our experiments show that applying a community based recommendation algorithm supports researchers in event finding. By using event participation history as background data for a CF based algorithm, we are able to recommend the most relevant academic events to researchers. The algorithm works on the dataset which can be easily extracted from references in papers documented in digital libraries like DBLP, ACM or EventSeer.net. We can derive from the data a certain likeliness that authors will continue to publish papers at a certain conference and therefore meeting these authors in future conference is not unlikely. We use this information recommending young researchers publishing their research at those conferences which are linked to their personal interests and can refine the search with name of authors. Content-based recommendation and the combination of content-based with CF and other recommendation techniques is a promising direction. In the moment we preparing the dataset by applying community mining algorithms first, to tackle typical cold start

[3] http://bosch.informatik.rwth-aachen.de:5080/AERCS

problems and sparse relations in the dataset. This very promising research lead to new visualizations of journal and conference clusters which can be accessed via the website.

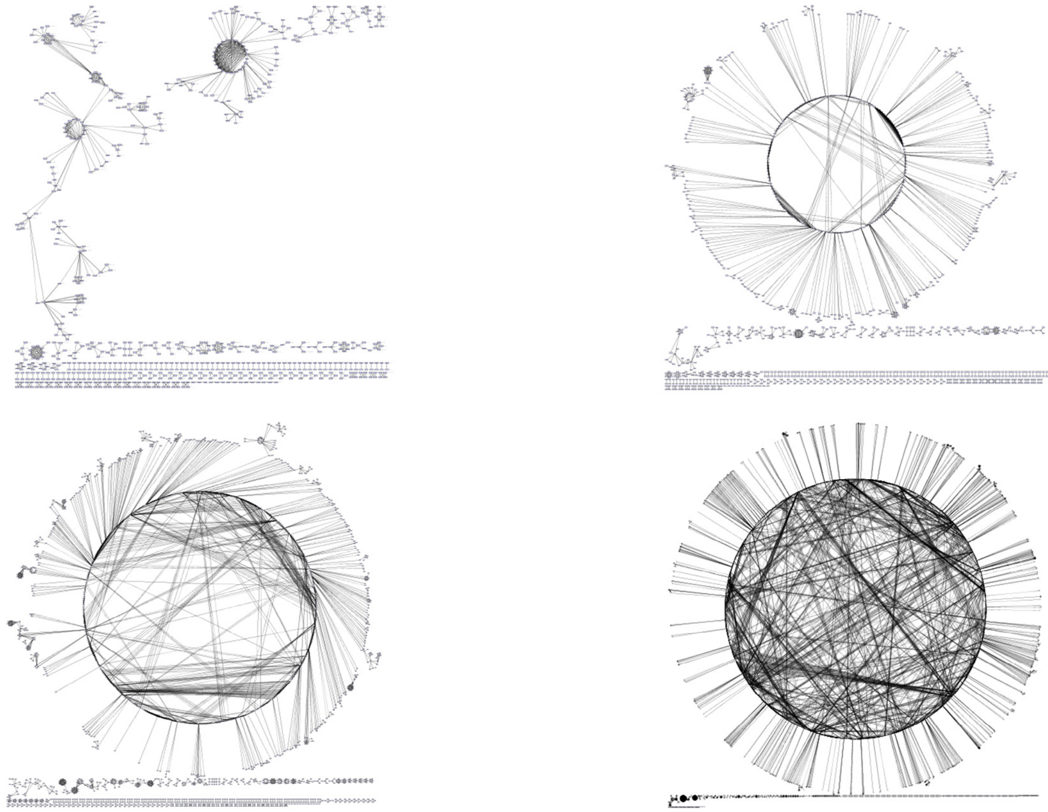

Figure 3: Development of the VLDB research community

Conclusions and Outlook

In the case studies we have applied SNA and used recommendation technologies widely. In this paper we have introduced research directions for supporting the advance of contemporary social learning theories with the goal to create empirical evidence for their claims and to support engineer in designing future personal learning environments for self-regulated learning. We have presented three case studies with examples which may fit the definition of learning networks. All the tools are available online and can be used to conduct further experiments. The availability of tools is a major prerequisite for the establishment of the new research methods in technology enhanced learning. Major lessons learnt from the case studies are:

- Longitudinal SNA studies with small but self-created network datasets help to facilitate the understanding of dynamic aspects of learning networks on a very coarse level giving you yearly or biannual updates on the changes taken from snapshots of network datasets. However, creating and maintaining network datasets in extremely expensive and needs financial support, e .g. from a network of excellence.
- Analyzing existing datasets from digital libraries (DBLP and citeseerX for example) or learning networks (eTwinning for example) can be done very effectively and efficiently by experts using a handful of good analysis and visualization tools. However, when non-experts and stakeholders (like school teachers for example) need support, we have to design special environments for doing so.
- If not only understanding of historical data is aimed at but also forecasting future behavior of learning networks some general laws about complex networks can be applied. Still, if the domain knowledge can be modeled into game theoretic payoff functions, multi agent simulation is a way to predict at least trends in the future.

In our newly started EU Integrating Project ROLE (responsive open learning environments) we try to incorporate analyzing and simulation methods described above to support self-regulated learning in personal learning environments. Learners and learning communities deserves new tools and methodologies to assess learning progress in a new way.

References

Anderson C. (2006). *The Long Tail: Why the Future of Business Is Selling Less of More*, New York: Hyperion, 2006.

Attwell, G. & Pumilia, P. (2007) The New Pedagogy of Open Content: Bringing Together Production, Knowledge, Development, and Learning , *Data Science Journal, Vol 6,* April.

Barabási, A.-L. (2007). From Network Structure to Human Dynamics, *IEEE Control Systems Magazine, 27(4):33–42*, August.

Bienzle, H. & E. Gelabert, E. & Jütte, W. & Kolyva, K. & Meyer, N. & Tilkin G. (2007). *The Art of Networking. European Networks in Education*, Die "Berater", Vienna, 2007.

Brandes, U. & Erlebach, T. (2005). Network Analysis: Methodological Foundations, *volume 3418 of Lecture Notes in Computer Science, Springer*.

Breuer, R. & Klamma, R. & Cao, Y. & Vuorikari, R. (2009). Social Network Analysis of 45,000 Schools: A Case Study of Technology Enhanced Learning in Europe, *in Cress, U., Dimitrova, V. and Specht, M. (Eds.) Learning in the Synergy of Multiple Disciplines, Proceedings of 4th European Conference on Technology Enhanced Learning, EC-TEL 2009, Nice, France, September/October 2009, LNCS 5794, Springer, pp. 166-180.*

Brown, J. S. & Adler, R. P. (2008). Minds on fire: Open education, the long tail, and learning 2.0, *EDUCAUSE Review, 43(1):16–32*.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments, *User Modeling and User-Adapted Interaction, 12(4):331–370*.

Cross, J. (2007). *Informal Learning*, Pfeiffer.

Downes, S. (2005). E-learning 2.0, *ACM eLearn Magazine*.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification, *Social Networks, 1(3):215–239*.

van Harmelen, M. (2006). Personal Learning Environments, *Proceedings of the Sixth International Conference on Advanced Learning Technologies (ICALT'06), pp. 815-816*.

Klamma, R. & Chatti, M. A. & Duval, E. & Hummel, H. & Hvannberg, E. T. & Kravcik, M. & Law, E. & Naeve, A. & Scott, P. (2007). Social Software for Life-long Learning, *Journal of Educational Technology & Society, 10(3):72-83*.

Klamma, R. & Pham, M. C., & Cao, Y. (2009). You Never Walk Alone: Recommending Academic Events Based on Social Network Analysis, *in Proceedings of the First International Conference on Complex Science (Complex'09), Feb. 23-25, 2009, Shanghai, China.*

Milgram, S. (1967). The Small World Problem, *Psychology Today, 1:60–67*.

Newman, M. E. J. (2003). The structure and function of complex networks, *SIAM REVIEW, 45(2):167–256*.

O'Reilly T. (2005). What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software", [Read July 3, 2006 from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html].

Panait, L. & Luke, S. (2005). Cooperative multi-agent learning: The state of the art, *Autonomous Agents and Multi-Agent Systems, 11(3):387–434*.

Sarwar, B. & Karypis, G. & Konstan, J. & Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms, *Proceedings of the 10th international conference on World Wide Web, New York, NY, USA, pp. 285–295*.

Schafer, J. B. & Konstan, J. A. & Riedl, J. (2001). E-commerce recommendation applications, *Data Mining & Knowledge Discovery, 5(1-2):115–153*.

Siebers, P.-O. & Aickelin, U. (2008). Introduction to multi-agent simulation, *Computing Research Repository, abs/0803.3905*.

Siemens, G. (2006). *Knowing Knowledge*, Lulu.com.

Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*, volume 8 of Structural Analysis in the Social Sciences, Cambridge University Press.