Journal of Universal Computer Science, vol. 20, no. 15 (2014), 2043-2053
submitted: 15/1/14, accepted: 30/9/14, appeared: 28/12/14

Leveraging Non-explicit Social Communities for Learning Analytics in Mobile Remote Laboratories

Pablo Orduña, Aitor Almeida
(DeustoTech - Deusto Institute of Technology, University of Deusto
Bilbao, Spain
{pablo.orduna,aitor.almeida}@deusto.es)

Salvador Ros
(Control and Communication System Department
Spanish University for Distance Education (UNED)
Madrid, Spain
sros@scc.uned.es)

Diego López-de-Ipiña, Javier Garcia-Zubia
(DeustoTech - Deusto Institute of Technology, University of Deusto
Bilbao, Spain
{dipina,zubia}@deusto.es)

Abstract: When performing analytics on educational datasets, the best scenario is where the dataset was designed to be analyzed. However, this is often not the case and the data extraction becomes more complicated. This contribution is focused on extracting social networks from a dataset which was not adapted for this type of extraction and where there was no relation among students: a set of remote laboratories where students individually test their experiments by submitting their data to a real remote device. By checking which files are shared among students and submitted individually by them, it is possible to know who is sharing how many files with who, automatically extracting what students are bigger sources. While it is impossible to extract the full real social network of these students, all the edges found are clearly part of it. These relations can indeed be used as a new input for performing the analytics on the dataset.
Key Words: Social Network Analysis, Remote Laboratories, Data Mining, Learning Analytics
Category: K.3.1, H.2.8, K.3.2

1 Introduction

Learning is a social process. People learn from one another [Bandura and McClelland 1977, Vigotsky et al. 1978], acquiring new information from those with more advanced knowledge. Social interactions help learners to cement their knowledge and to compare their ideas with others [Scardamalia and Bereiter 2006]. Identifying the relations created during those social interactions can help us to better understand what is going on in the learning process of the students. Siemens in [Siemens 2005] describes the learning process as an activity where connections are made between persons and resources.

Taking this in mind, Social Network Analysis (SNA) is an important part of the Learning Analytics process [Sie et al. 2012b]. In those systems that explicitly provide a social network (forums, connections between learners...) identifying the social interactions is a straightforward process. But it those system with no explicit social interactions (as in the WebLab-Deusto Remote Laboratory Management System) extracting the network data is a more complicated issue. Having not the possibility of on-line interaction provided by the system, the students rely on the old and proved off-line interaction. In these paper we present an analysis of that off-line interaction that happens among the users of WebLab-Deusto, a remote laboratory available for mobile devices [Almeida et al. 2012].

To be able to map these off-line interactions we have centred our analysis on the results of those interactions. In the case of Weblab-Deusto, the results are the exercises uploaded to the system. Students work together to solve the exercises proposed by the teachers and end up uploading similar programs. We have analysed the temporal relations of the uploaded files and identified those files that are identical to construct the network of collaborations between the students. It has to be taken into account that WebLab-Deusto is not used to evaluate the students and that the proposed exercises only have a pedagogical purpose. The proposed method is not intended to serve as a plagiarism detection mechanism (as is done for example in [Merlo et al. 2010]), but as a tool for the teachers to understand the dynamics of the analysed courses. Using the inferred network we are able to detect communities that are dynamically formed among the students and to identify those students that take a more active role in the course.

In this paper we will present the analysed dataset and the resulting social network. In Section 2 we analyse the current state of the art, in Section 3 we describe the analysed dataset and the WebLab-Deusto storage infrastructure, in Section 4 we study the resulting social network, explaining how it can be used by the teachers to better understand the dynamics of the courses and finally in Section 5 we discuss the conclusions and future work.

2 Related Work

A remote laboratory, as more deeply described in 3.1, is a software and hardware system that enables students to access real equipment located in universities. Different types of evaluations (related to costs, to usefulness or immersion) have been addressed in the literature [Bright et al. 2008, Nickerson et al. 2007, Jona et al. 2011, Garcia-Zubia et al. 2011]. They were based on students grades, on comparisons with students in traditional hands-on-lab sessions, on dedicated surveys, or on the data recorded of the activities performed by students in the laboratories.

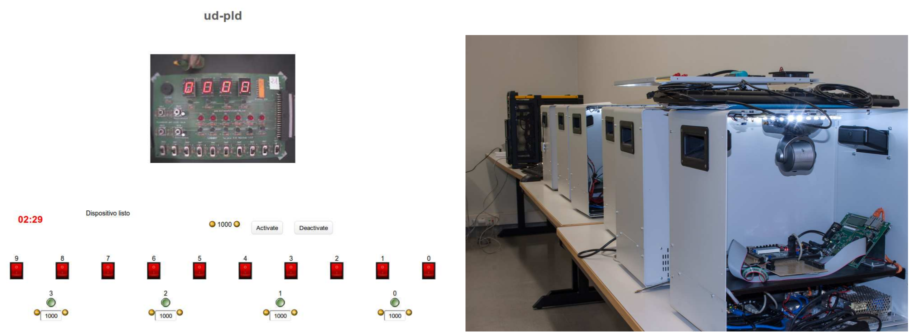
When using recorded activities, the remote laboratory must support some type of user tracking mechanism that enables researchers to gather an activity record and use it. There is a wide range of support degree for this features (from not supporting it at all to knowing all the fine grained interactions associated to each user and group [Harward et al. 2008, Orduña et al. 2011]). Even in the case of fine grained interactions, it is possible to aggregate data in periods of time or in teams that were working together. But real social networks (as opposed to assigned or self-assigned small teams) are easily not taken into account given the complexity of finding them in these mainly individual systems.

Several authors have analysed the importance of Social Networks Analysis in Learning Analytics. [Sie et al. 2012b] identify four main areas in learning related SNA: *network visualization*, *network analysis*, *simulation* and *network interventions.*

In the *network visualization* area authors have worked both on building the networks from the sociocentric perspective using the data provided by log files and forum interactions or from the egocentric perspective constructing the social networks based on the students reported interactions. The first approach has been used by [Brooks et al. 2006] in the iHelp system to build undirected networks, by [Chatti et al. 2009] in the Plone system to build undirected networks and by [Dawson 2010] to build directed networks. The second approach has been used by [De Laat et al. 2007] using WebCT logs to build directed networks and by [Martınez et al. 2003] using BSCW logs to build undirected networks.

In the *network analysis* area authors have worked on identifying common interaction patterns that take place on the social networks, analysing the structure of the communities and identifying groups. To do this authors have used different SNA metrics. The centralities (degree, betweenness, closeness, eigenvector...) have been used by [Dawson et al. 2011], [Capuano et al. 2011] and [Chatti et al. 2009]. Other authors have used the density, centralisation, share and reciprocity metrics to analyse the social network like [Yao 2010], [An et al. 2009] and [Martınez et al. 2003].

Authors have used *simulations* to evaluate the outcome of new social connections [Sie et al. 2012a], to predict the reactions of changes in the infrastructure [Wild et al. 2010] and to predict network behaviour [Steglich et al. 2006]. Finally in the *network interventions* area the performed analysis have been used to improve organizational management [Steiny and Oinas-Kukkonen 2007], improve the social interactions between learners [Jo 2009], pro-actively offer information [Su et al. 2010] and create groups of teachers with similar interests [Fetter et al. 2011].



(a) User interface (b) Equipment

Figure 1: CPLD laboratory

3 Analyzed dataset

This section explains the characteristics of the dataset used in this contribution.

3.1 WebLab-Deusto Remote Laboratory Management System

WebLab-Deusto is a remote laboratory management system. A remote laboratory is a hardware and software solution that enables students to interact with real equipment located typically in a university. Users access this equipment as if they were in a traditional hands-on-lab session, but through the Internet.

To show a clear example used in this contribution, Figure 1(a) shows the CPLD remote laboratory. On it, there is a Complex Programmable Logic Device (CPLD) with a set of inputs (switches, buttons, a clock) and outputs (LEDs and small displays). Students learn to develop designs in the VHDL language. They compile their code in their computer, submit the binary file to the real CPLD through WebLab-Deusto, and then they can interact with the device during a short fixed amount of time (e.g., 200 seconds). Other students will remain in a queue shared among the different CPLDs deployed Figure 1(b).

While there are many types of remote laboratories (e.g., chemistry, physics, electronics), they all share certain features, such as authentication, authorization, scheduling, sharing through federation or, especially interesting for this contribution, user tracking. So as to implement these common layers, open source Remote Laboratory Management Systems (RLMSs) emerged to implement these layers. Examples of these are WebLab-Deusto, the iLab Shared Architecture or Labshare Sahara. These systems essentially provide software toolkits that can be used to implement remote laboratories, without dealing with those features. For example, any remote laboratory implemented using the WebLab-Deusto APIs can automatically be shared with other institutions using this system, or rely on the authentication protocols provided by the system (database, OpenID, OAuth 2). And all the interations (which include files and commands submitted by the client) are stored and instructors can access this data.

3.2 WebLab-Deusto storage infrastructure

WebLab-Deusto provides two models when designing laboratories: one where all the interactions are submitted through WebLab-Deusto (and stored in middle layers), and other where the communications are direct between the client and the final laboratory code. In both approaches the following information is stored:

- Basic usage information (user identifier, course identifier, laboratory identifier).
- Location information (user's IP address -so we could identify who is using it from the University-, when the students starts using it and when finishes the session).
- Additional contextual information (whether it was accessed from a mobile device, or embedded Facebook).

Additionally, for those laboratories in the first model (which are the ones in this paper), it is also stored:

- All the commands submitted by students. These are control commands (e.g., whether the system is still submitting the uploaded file to the CPLD) and user interactions (e.g., turn switch 1 on), including when they were sent and the response and response time.
- All the files submitted, as well as a hash of these files is also stored, and the timestamps of the submission and response.

Previous analysis have been performed relying on this data. On [Garcia-Zubia et al. 2011], several factors have been studied, grouping certain fields in different time periods (e.g., morning, afternoon, night) or locations (e.g., university or at home).

At the time of this writing, there are over 60k uses registered in the database (including demos and tests), by over 1400 users in 6 laboratories used in the curricula. These logs only include data since 2009, when the storage engine was put in production. Before that year, there are log files not analyzed in this contribution, but there are submitted files since 2007 which include who submitted them and when was it used, useful for the contribution. Other deployments of

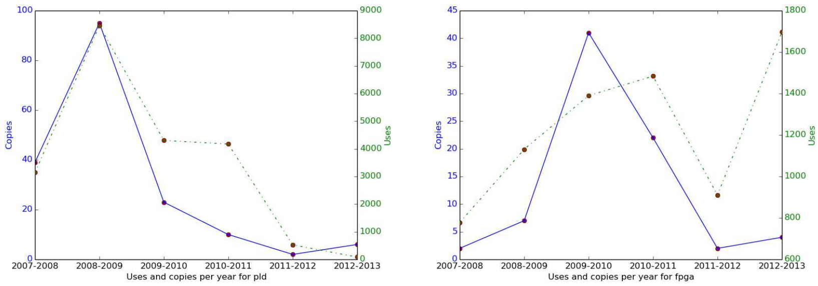


Figure 2: Uses (dotted) of the analyzed laboratories and copies (line) in those courses

WebLab-Deusto in other universities (STU[1], UNED[2]) are not considered in this contribution.

3.3 CPLD and FPGA datasets

In this contribution, only two laboratories are considered: CPLD (explained before) and FPGA (conceptually similar, and same set of inputs and outputs, but using a different device). Both systems have the same interaction.

One of the interesting features of the submitted files is that it includes not only the code, but certain metadata added by the compiler. That includes a timestamp of when the file was compiled and the name of the source file. This is convinient for the analytics since we can know that if two students send the same file (including those two fields), those files have surely been shared (unless both students had compiled the same file with the same name in the same exact second). Additionally, removing these fields, it is possible to find other coincidences that suggest that the users have shared the source file and compiled them several times. However this is only a suggestion since since they could have programmed the same code by their own, especially in the first lessons.

It is important to remark that this sharing does not mean plagiarism, since students are not evaluated by their use in the lab, but later by the knowledge they have acquired. Additionally, it must be taken into account that instructors leave example files in the Learning Management Systems, so it is important to not take into account those files previously submitted by instructors or administrators.

For this contribution, courses 2007-2008 to 2012-2013 have been considered. This is, 20,693 uses of the CPLD laboratory used by 295 students, and 7,390 uses of the FPGA laboratory used by 138 students.

[1] http://kirp50.chtf.stuba.sk
[2] http://weblab.ieec.uned.es

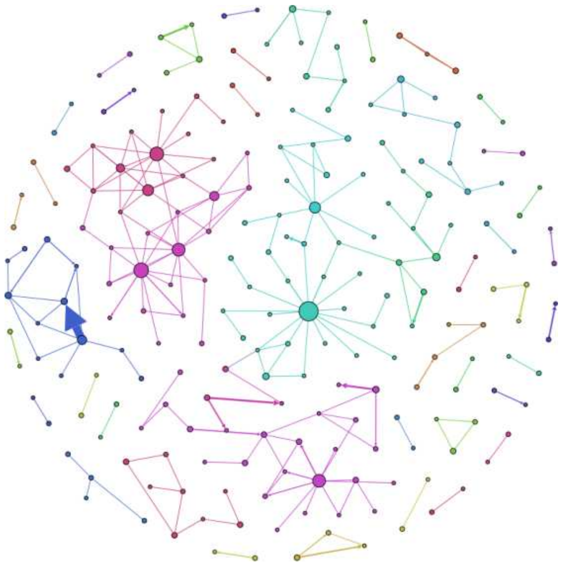


Figure 3: Social Networks created using the exercise relations.

4 Inferring social structures from the available data

[Sie et al. 2012b] identify two methods for data collection in social network analysis:

1. Data collection in *sociocentric* networks, whose main focus are the interactions between users on a network level. According to [Sie et al. 2012b], these networks are build monitoring real data and creating the network topology based on that data.

2. Data collection in *egocentric* networks, whose main focus are the relations of specific individuals. According to [Sie et al. 2012b], these networks are build asking the users to identify their contacts.

In our case we have used the first approach. As explained in the previous section, WebLab-Deusto does not store any specific information about social interactions. In order to infer the social connections created outside the system, we have used the results of those relations, the uploaded exercises. We have compared those exercises to see which ones where not only identical, but also has the same name and where compiled on the same exact second. We have then used the timestamps of those exercises to build a directed graph with the relations between users (see Figure 3). The created network has the following characteristics:

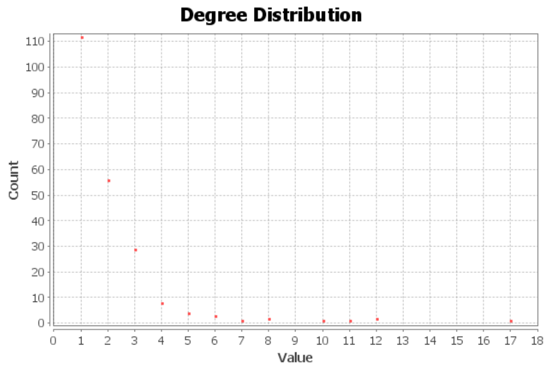
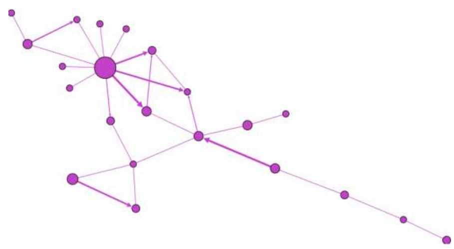
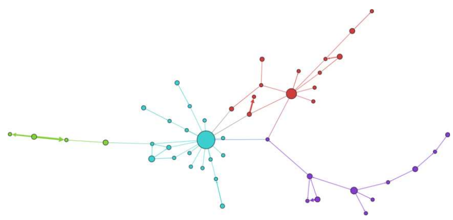


Figure 4: Degree distribution of the created social network.



(a) Size depends on outdegree



(b) Color depends on modularity

Figure 5: Different representations of connected components in different groups

- 220 *nodes*. There are several students that do not have a node in the social network because they have not any shared file. This reflects the reality where some of the students prefer to work individually or do not have any acquittance on the class.
- 233 *edges*.
- 42 *weakly connected components*. As can be expected there are several components of 2 (28 in total) or 3 nodes (4 in total), representing the isolated groups that only work with each other and do not interact with the rest of the class. The biggest connected component is composed by 48 nodes.
- An *average degree* of 1.059 (see Figure 4). Highest outdegree of 17 and highest indegree of 5.
- An *average path length* of 1.229.
- A network *diameter* of 3.

– An *average clustering coefficient* of 0.076.

4.1 Analysing the network

By analysing the degree centralities of each node we can ascertain the role of each student in the social network. Those students with a high outdegree the origin of the shared exercises. For example, in Figure 5(a) the central node has an outdegree of 10, which means that she has shared files with other 10 students. The relationship weight with her connections depend on the number of files shared with each of the students. In the same example the most important relation of the central node has a weight of 5, meaning that she has shared 5 files with that student. On the other hand we can use the indegree to identify those students that are receiving the shared files.

Teachers can use this centralities to understand the roles of the students in the class. Students with higher outdegrees work in the proposed exercises and resolve them, sharing those exercises with other students. Students with a higher indegree wait until other student has resolved the proposed exercise. Teachers can use this information to identify which students need more help and motivation to be able to complete the exercises. Also studying the correlation between the indegree and outdegree using the Spearman rank-order correlation coefficient there is an strong negative correlation of -0.597 with a significance of 1.321740044721179e-22. This indicates that students that share their exercises are not usually on the receiving end of the process.

Finally (as can be seen in Figure 5(b)) the modularity can be used to identify the different student communities that are created in the social network. In the example, the connected component has four different communities, the smallest one with 4 students and the biggest one with 19. These communities can be used by the teacher to discover those groups that like working together.

5 Conclusions and future work

In this paper we have analysed the WebLab-Deusto dataset, describing the information it contains and the difficulties that it presents for the identification of social connections between the students. We have described a method to identify the underlying social network created by the off-line interactions of the students, based on the results of those interactions, the exercises uploaded to the system. Finally we have presented an analysis of this social network and explained how the teachers can use the that information to identify the social dynamics in each course.

As future work we plan to integrate the presented work in the WebLab-Deusto dashboard, providing new tools to the teachers using the system in their classes. We also would like to cross-reference the social network analysis with the academic results of the students to analyse the correlations between the obtained marks and the usage of WebLab-Deusto.

References

[Almeida et al. 2012] Almeida, A., Orduña, P., Castillejo, E., López-de Ipiña, D., Sacristán, M.: "A method for automatic generation of fuzzy membership functions for mobile devices characteristics based on google trends"; Computers in Human Behavior; (2012).

[An et al. 2009] An, H., Shin, S., Lim, K.: "The effects of different instructor facilitation approaches on students' interactions during asynchronous online discussions"; Computers & Education; 53 (2009), 3, 749–760.

[Bandura and McClelland 1977] Bandura, A., McClelland, D. C.: "Social learning theory"; (1977).

[Bright et al. 2008] Bright, C., Lindsay, E., Lowe, D., Murray, S., Liu, D.: "Factors that impact learning outcomes in both simulation and remote laboratories."; World Conference on Educational Multimedia, Hypermedia and Telecommunications; volume 2008; 6251–6258; 2008.

[Brooks et al. 2006] Brooks, C., Panesar, R., Greer, J.: "Awareness and collaboration in the ihelp courses content management system"; Innovative Approaches for Learning and Knowledge Sharing; 34–44; Springer, 2006.

[Capuano et al. 2011] Capuano, N., Laria, G., Mazzoni, E., Pierri, A., Mangione, G. R.: "Improving role taking in cscl script using sna and semantic web"; Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on; 636–637; IEEE, 2011.

[Chatti et al. 2009] Chatti, M. A., Jarke, M., Indriasari, T. D., Specht, M.: "Netlearn: social network analysis and visualizations for learning"; Learning in the Synergy of Multiple Disciplines; 310–324; Springer, 2009.

[Dawson 2010] Dawson, S.: "'seeing'the learning community: An exploration of the development of a resource for monitoring online student networking"; British Journal of Educational Technology; 41 (2010), 5, 736–752.

[Dawson et al. 2011] Dawson, S., Macfadyen, L., Lockyer, L., Mazzochi-Jones, D.: "Using social network metrics to assess the effectiveness of broad based admission practices"; Australasian Journal of Educational Technology; 27 (2011), 1, 16–27.

[De Laat et al. 2007] De Laat, M., Lally, V., Lipponen, L., Simons, R.-J.: "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 (2007), 1, 87–103.

[Fetter et al. 2011] Fetter, S., Rajagopal, K., Berlanga, A., Sloep, P.: "Ad hoc transient groups: instruments for awareness in learning networks"; (2011).

[Garcia-Zubia et al. 2011] Garcia-Zubia, J., Orduna, P., Angulo, I., Hernandez, U., Dziabenko, O., Lopez-Ipina, D., Rodriguez-Gil, L.: "Application and user perceptions of using the weblab-deusto-pld in technical education"; Frontiers in Education Conference (FIE), 2011; IEEE, 2011.

[Harward et al. 2008] Harward, V. J., Del Alamo, J. A., Lerman, S. R., Bailey, P. H., Carpenter, J., DeLong, K., Felknor, C., Hardison, J., Harrison, B., Jabbour, I., et al.: "The ilab shared architecture: A web services infrastructure to build communities of internet accessible laboratories"; Proceedings of the IEEE; 96 (2008), 6, 931–950.

[Jo 2009] Jo, I.-H.: "The effect of social network diagrams on a virtual network of practice: a korean case"; Asia Pacific Education Review; 10 (2009), 4, 525–534.

[Jona et al. 2011] Jona, K., Roque, R., Skolnik, J., Uttal, D., Rapp, D.: "Are remote labs worth the cost? insights from a study of student perceptions of remote labs"; International Journal of Online Engineering (iJOE); 7 (2011), 2, pp–48.

[Martınez et al. 2003] Martınez, A., Dimitriadis, Y., Rubia, B., Gómez, E., De La Fuente, P.: "Combining qualitative evaluation and social network analysis for the study of classroom social interactions"; Computers & Education; 41 (2003), 4, 353–368.

[Merlo et al. 2010] Merlo, E., Ríos, S. A., Álvarez, H., LHuillier, G., Velásquez, J. D.: "Finding inner copy communities using social network analysis"; Knowledge-Based and Intelligent Information and Engineering Systems; 581–590; Springer, 2010.

[Nickerson et al. 2007] Nickerson, J. V., Corter, J. E., Esche, S. K., Chassapis, C.: "A model for evaluating the effectiveness of remote engineering laboratories and simulations in education"; Computers & Education; 49 (2007), 3, 708–725.

[Orduña et al. 2011] Orduña, P., Irurzun, J., Rodriguez-Gil, L., Zubía, J. G., Gazzola, F., López-de Ipiña, D.: "Adding new features to new and existing remote experiments through their integration in weblab-deusto."; iJOE; 7 (2011), S2, 33–39.

[Scardamalia and Bereiter 2006] Scardamalia, M., Bereiter, C.: "Knowledge building: Theory, pedagogy, and technology"; The Cambridge handbook of the learning sciences; (2006), 97–115.

[Sie et al. 2012a] Sie, R. L., Drachsler, H., Bitter-Rijpkema, M., Sloep, P.: "To whom and why should i connect? co–author recommendation based on powerful and similar peers"; International Journal of Technology Enhanced Learning; 4 (2012a), 1, 121–137.

[Sie et al. 2012b] Sie, R. L., Ullmann, T. D., Rajagopal, K., Cela, K., Bitter-Rijpkema, M., Sloep, P. B.: "Social network analysis for technology–enhanced learning: review and future directions"; International Journal of Technology Enhanced Learning; 4 (2012b), 3, 172–190.

[Siemens 2005] Siemens, G.: "Connectivism: A learning theory for the digital age"; International Journal of Instructional Technology and Distance Learning; 2 (2005), 1, 3–10.

[Steglich et al. 2006] Steglich, C., Snijders, T. A., West, P.: "Applying siena: An illustrative analysis of the coevolution of adolescents' friendship networks, taste in music, and alcohol consumption."; Methodology: European Journal of Research Methods for the Behavioral and Social Sciences; 2 (2006), 1, 48.

[Steiny and Oinas-Kukkonen 2007] Steiny, D., Oinas-Kukkonen, H.: "Network awareness: social network search, innovation and productivity in organisations"; International Journal of Networking and Virtual Organisations; 4 (2007), 4, 413–430.

[Su et al. 2010] Su, C., Huang, M., Contractor, N.: "Understanding the structures, antecedents and outcomes of organisational learning and knowledge transfer: a multi-theoretical and multilevel network analysis"; European Journal of International Management; 4 (2010), 6, 576–601.

[Vigotsky et al. 1978] Vigotsky, L. S., et al.: "Mind in society"; (1978).

[Wild et al. 2010] Wild, F., Ullmann, T. D., Scott, P.: "The stellar science 2.0 mashup infrastructure"; Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on; 621–623; IEEE, 2010.

[Yao 2010] Yao, Y.: "Comparing two discussion designs in terms of student online interactions"; Education Technology and Computer (ICETC), 2010 2nd International Conference on; volume 1; V1–219; IEEE, 2010.