Topological Parameters of Networked Learning

Carmel Kent[a], Amit Rechavi[b], Sheizaf Rafaeli[a]

[a]*Haifa University, Haifa, Israel*
[b]*Tel-Aviv University, Tel-Aviv, Israel*

Abstract

This paper proposes a methodological linking of learning theories and social learning analytics, applicable in real-life settings. We view online discussions in learning communities as networks in which human agents and posts of content are the nodes, while the interactions between nodes are represented as edges. The collaborative learning process is thus viewed as the construction and growth of a network involving various types of interactions among learners and content items. We contend that online learning assessment should be grounded by a theoretical framework based on quantitative analysis of the collaborative learning process as reflected by learning communities' online discussions. This paper demonstrates our proposed framework through a case study of a single community's online discussion, which took place throughout one academic semester. We use social network analysis (SNA) and other log-based techniques to determine topological parameters or network characteristics, assessing both collective and individual learning process which took place as reflected by the community online discussion.

1. Introduction

This paper proposes theoretical foundations for a networked-based modelling of collaborative learning processes. We ask what the topology parameters of a learning network are, and how they can be used to assess learning.

As suggested by educational research (e.g. Vygotsky, 1978; Barker, 1994), *interaction* is one of the most important tools for learning. Asynchronous online discussions provide learners with the opportunity to interact when reading and responding to peers' and teachers' postings. Therefore, we propose that interactivity is an essential metric when evaluating learning communities. Conceptualizing interactivity as a process of relating to each other's postings by taking conversational turns (Rafaeli & Sudweeks, 1997) emphasizes the social foundations of knowledge construction. We argue that modelling a collaborative learning process must emphasize not only interactions among community participants, but also interactions among content items such as posts and other information resources (Siemens, 2005) that are implemented as hyperlinks and actively curated by community participants (Kent & Rafaeli, 2016).

Learning analytic tools and practices are central to e-learning. They are of concern to both industry and research, as they offer the opportunity for log-based, behaviorally driven evaluations of learning. Social learning analytics focus on how learner communities create knowledge together. MOOCs (massive open online courses) and other web-based learning endeavors are based on online conversations, and enable rich logging data collection and data mining based on learners' behavior patterns (Sinha, 2014; Wu, Yao, Duan, Fan, & Qu, 2016). At the same time, the lack of face-to-face interaction in online learning calls for further analysis and research.

Despite the wealth of data on formal learning communities and social platforms, the theoretical basis of social learning analytics is still in its infancy (Deboer, Ho, Stump, & Breslow, 2014; Shum & Ferguson, 2012). We propose using social learning big data to explain and assess the online learning process based on theories of learning, information and computer-mediated communication. We present a theoretical methodology for quantitative analysis of online discussions in learning communities and suggest using network analysis to assess the learning processes experienced both individually and collectively (Haythornthwaite, 2008; Reffay & Chanier, 2003; Shea et al., 2010, 2013; Lefebvre et al., 2016).

We view online discussions in learning communities as networks in which human agents and posts are the nodes, while interactions among nodes are represented as edges. The collaborative learning process is thus viewed as the construction of a network containing various types of interactions among learners and content items (AlDahdouh, Osório, & Caires, 2015). In deriving a network of interactions from an online discourse, we use the theoretical frameworks of social constructivism and connectivism to define the node set. Social constructivism (Vygotsky, 1978) views the participants as the building blocks of collaborative learning. Connectivism (Siemens, 2005) reframes learning in the digital ecosystem, adding the notion of a hybrid network where information resources and devices act as nodes in addition to human agents. Complementarily, we use interactivity theory and assimilation theory to define the edges in the learning network. Interactivity theory (Rafaeli, 1988) focuses on interactions between community members to explain knowledge forming processes. In turn, assimilation theory (Ausubel, 1968) emphasizes the importance of forming relations between information resources for meaningful learning. We contend that knowledge cannot be conceived as constrained to a certain entity (a single object or the mind of a single individual), nor a given point in time. Instead, knowledge construction is an *ongoing* process which takes place in a *hybrid* network of human and information agents.

As distant and blended learning play an increasingly central role in our lives, policies have to be formed regarding the evaluation and assessment both of individual and of collective online learning. A data driven, quantifiable framework for collaborative online learning is essential for cogent design and evaluation of mass learning. This paper makes a step towards such a framework. We propose a set of quantifiable learning indicators that we believe will complement traditional assessment tools. We demonstrate this proposition based on a case study of a single higher education learning community which used online discussion in blended mode during the entire course of an academic semester.

In the remainder of this section, we briefly review the theoretical and research basis for social learning analytics as reflected in online discussions, laying the groundwork for our proposed units of analysis. Based on those units, we then raise preliminary research questions to demonstrate our approach. In the Method section, we present our case study, followed in the Results section by empirical evidence related specifically to the research questions. Preliminary insights regarding the validity of network analysis to the study and systematic assessment of collaborative online learning are then discussed.

1.1 Learning analytics in online discussions

Online discussions hold promise for collaborative knowledge construction: online community members share ideas, learn from peers and build knowledge collectively, while reading and reflecting on each other's thoughts. The virtual setting enables less-assertive participants to compose their thoughts (Hewitt, 2001), while allowing more time for all participants to reflect on and respond to others' contributions (Poole, 2000). Pedagogically, rationales for learning by online discourse typically make references to the collaborative construction of meaning (Lander, 2015; Fried, 2016). When learners build on the comments of others, a higher flow of communication and inference is achieved, compared to traditional turn-taking face-to-face settings (Garrison, 2006).

Previous research involving learning analytics in online discussions has referred to content analysis methods (Kovanovic, Joksimovic, Gasevic, & Hatala, 2014), visualization of interactions (Wu, Yao, Duan, Fan, & Qu, 2016), and social network analysis (Reffay & Chanier, 2003; Pedro et al., 2016), all designed mainly to provide managerial and instructional decision making tools. In these contexts, learning analytics require combining data science methods with traditional educational research methods to provide mechanisms for analyzing digital learning experiences with high-resolution and time-sensitive data (Gibson & de Freitas, 2015).

We argue that an online discussion is not analogous to non-virtual face-to-face discussion (Herring, 1999), and that it is even informed by a different set of values (Shum & Ferguson, 2012). Thus, not only needs to be designed differently, but also to be conceptualized and thus assessed differently (Deboer et al., 2014).

Specifically, smooth scale-up of assessment approaches from the traditional learning world to the online world should not be taken for granted. A theory-grounded attempt should be made to reframe learning assessment in the collaborative online setting. Similar attempts have been made with regard to the Community of Inquiry model, for example (Swan & Ice, 2010).

1.2 Learning analytics as a theory-based research and assessment tool for collaborative learning

Our main goal in this study is to learn about the collaborative learning process online communities undergo from the structure and dynamics of interactions occurring in their online discussions. As opposed to content analysis methods such as categorizing demonstrated thinking skills (Bloom, 1974), and traditional learning outcome assessment such as examination grading, we focus here on revealing the learning analytics which can be quantitatively extracted from the log-based structure of the persistent conversation alone (Erickson & Herring, 2005).

We suggest conceptualizing an online discussion of a learning community as a network and to analyze it accordingly. Social network analysis (SNA) traditionally assumes nodes in the network model human agents, while edges represent human interactions among them. Thus, SNA is usually concerned with the role and relationships of individuals within the network. When thinking about a network derived from the set of interactions within a learning community, the learners themselves are conceived as the scaffolds of learning. Social constructivism (Vygotsky, 1978) emphasized the central role played by co-learners and mediators in individual knowledge construction and viewed learning as the process by which learners were integrated into a knowledge community. Clearly, such a view supports the validity of the social network's traditional construction. Subsequent theories from information sciences supported this view. For example, interactivity theory (Rafaeli, 1988) focuses on interactions between community members to explain knowledge forming processes. However, this theory, developed in light of the emergence of the internet and distant learning, refers to an additional type of nodes beyond the scope of this human learning network: the discussion's set of messages (or resources) through which human interactions takes place (Rafaeli & Sudweeks, 1997). Accordingly, text messages and other information resources are another persistent entity in the learning network, and human interactions are modeled within the context of externalizing knowledge through these messages.

This learning concept of learning modeled as a network of interactions between knowledge items, concepts and ideas was not new by itself. Cognitive learning theories had been based on that notion long before the advent of the internet. Assimilation theory (Ausubel, 1968) emphasized the relations between information items in the process of meaningful learning. Novak (1990) used concept maps as a common learning tool, based on the principle of linking information items through semantic links into a network. Vygotsky (1986) also viewed the environment as part of the interactional scene of learning. The subsequent introduction of the internet and its increasingly central role in collaborative learning led Siemens (2005) to theoretically reframe learning in the digital ecosystem. Siemens introduced Connectivism, with its notion of a *hybrid network of learning*: now information items and devices act as nodes in addition to human agents. The rise of online communities led social scientists to examine the interactivity of hybrid networks. Stromer-Galley (2000) pointed to the distinction between user-to-user and user-to-medium interaction. Many others also considered interactivity as occurring between the user and the text (Williams, Rice, & Rogers, 1988). Moore (1989) proposed three types of interactivity: learner-content, learner-instructor and learner-learner.

Thus, when evaluating the process of knowledge construction within an online discussion, both human agents and content objects are present need to be taken into account, almost on an equal footing. Similarly, as interactions within a learning community involve such activities as knowledge exchange, questioning, associative learning, and learning by browsing, there is little point in the dichotomy of conceived interaction between two co-learners on the one hand and between a single learner and the content produced by another on the other. Accordingly, in order to evaluate learning within online communities, we model the learning network as composed of two types of nodes and three types of edges (interactions). Nodes represent (1) human agents (Haythornthwaite, 2008); and (2) information agents (posts or other media types contributed). Edges represent (1) interactions between human and information agents (such as view, vote, relate to, write); (2) interactions among human agents (such as user-follows-user); and (3) interactions among information agents (semantic links which emphasize relations between posts, such as “example of”, “contrary to”, or “makes me think of”) (Novak, 2010).

1.3 Unit of analysis: Collective vs. individual learning

Elsewhere (Kent & Rafaeli, 2016) we presented a framework for the quantitative analysis of interactivity in online learning discussions. Although these measures were extracted within a networked context (for example, the depth of threads, out-degree and in-degree of posts), the unit of analysis was the individual learner.

According to activity theory, learning is a collective participatory process of active knowledge construction emphasizing the interplay between the individual and the group's cognition in terms of both context and interaction (Cole & Engeström, 1993). In this study we add the analysis of an entire community's learning (Haythornthwaite & De Laat, 2010). Distributed cognition is defined as an extension of our internal cognition in the outside world through artifacts and other people. In teams, distributed cognition is concerned with representation of and access to other people’s knowledge (Hutchins, 1990). Klimoski and Mohammed (1994) applied this understanding of an individual’s sense-making to conceptualizing cognition as a shared or collaborative group-level phenomenon. Shared mental models refer to the overlapping mental representation of knowledge by team members (Bossche, Gijselaers, Segers, Woltjer, & Kirschner, 2010). Collaboration is a coordinated synchronous activity that is the result of an ongoing attempt to construct and maintain a shared conception of a problem (Pena, 2005). Thus, collaborative learning in online communities may be analyzed from the perspective of an enabler of individual learning, but should be complemented by analysis of the learning of the community as a whole. The more conceptual coverage and enrichment are provided (Stoyanova & Kommers, 2001), the more new and meaningful relations among bits of knowledge are revealed (Ausubel, 1968) , the more opportunities community participants have to interact, reflect, and build on others' ideas (Rafaeli & Sudweeks, 1997; Vygotsky, 1978) the more central social construction of knowledge becomes. The discussing community becomes a learning community.

This study begins by correlating individual network parameters with a learning outcome measure on our hybrid learning network. If a significant relationship is found, this will allow us to infer that network analysis is relevant to learning, at least as far as traditional assessment of collaborative learning is concerned. We will then proceed to network analysis of the entire community as the unit of analysis, and identify the topology parameters of the hybrid learning network significant for collaborative learning.

2. Research Hypotheses

Each member of an online learning community acts to gain values such as interacting, acquiring knowledge or achieving a high grade. In doing so, they act within the social community and thus contribute to changing the online network structure (posts are added, links are created, etc.). These changes in the community's structure reflect the network’s actual learning process and in that impact not only the single learner who has initiated them, but also the entire network community.

Our main research question is: How can collaborative learning be measured or assessed using network analysis techniques? When viewing the collaborative learning process via a network structure, we combine a set of network static and dynamic analyses (Reda et al., 2011; Wei & Carley, 2015). Research and best practices concerning social online learning are in their infancy. The learning of a community as a whole is not yet customary or assimilated into policies and best practices. Thus, a standard assessment tool is not yet established as a baseline.

One of our research's units of analysis is individual learning. We collect (1) topological parameters of learners such as *in-degree, out-degree, eigenvector degree, betweenness and closeness*, and (2) learning outcome indicators (pre- and post-learning examination grades) and look for correlations between them. Our research follows Russo and Koesten (2005), who used network prestige and centrality as robust predictors of cognitive learning outcomes; Toikkanen and Lipponen (2011), who found the out-degree of replies and in-degree of reading to be most strongly correlated with the meaningfulness of the learning experience; Romero et al.'s (2013) finding that the number of messages sent and the number of words written together with the degree of centrality and prestige were the most important attributes for predicting final student performance; Vequero and Cebrain (2013), who found that more frequent and intensive social interactions generally implied higher student grades; and Joksimović et al.'s (2016) research on centrality parameters and online learning. In our study, we explore further topological parameters such as node exclusivity to study the network's dynamics. We assume that the change in the student's knowledge is correlated with the topological parameters of her associated node and measure the change in the student's knowledge as reflected by the change in her course grade (pre- and post-learning examination grades). Therefore, our first hypothesis is:

H1a: At the individual level, positive correlation will be found between students' grades and their main topological parameters,.

In an online community, learners can write new posts, comment, vote, etc., but they can also create new links between existing information resources (posts). Relations created by a third-party learner between two existing posts of content are called *cross-references.* Ausubel (1968) found that the act of curating new cross-references between existing content items is itself an indicator of learning. Since we assume that cross-referencing contributes to learning and that the learning process in an online community results (even inadvertently) in topological changes, we hypothesize as follows:

H1b: Network centrality parameters are positively correlated with cross-reference activity.

We assume that individual knowledge contributes to the community's collective knowledge. However, individual knowledge construction cannot be measured or conceptualized using the same assessment framework as the community's collective knowledge construction. In this study, we propose a quantitative method based on SNA and web analytics to assess social learning at the community level. Since learning and knowledge construction are formative processes, we use dynamic analysis to assess their characteristics. Our analytic framework examines learning communities' online discussions in an attempt to identify the occurrence of a learning process based on *structural changes* in the learning network as a whole, as well as changes in its inner topological formations. We also analyze the network's end-product snapshot at the end of the academic semester as an artificial milestone representing the end of the learning process and extend Rafaeli and Sudweeks' (1997) findings by identifying static topological indicators for the learning community. In general, we look for changes in the community's network structure during the course that reflect the learning process and have the potential to further support it.

One of the topological parameters which depict such structural change is the network distance. The distance between nodes is the number of links which connect them (either by a common learner they follow or by a common post they interact with). According to Leskovec et al. (2005), the average distance between nodes often shrinks over time. In a learning community, when learners get closer to each other, they are able to exchange new thoughts and ideas and thus learning can occur effectively. Our next hypothesis is therefore:

H2: In the learning community as a whole, the average distance between nodes will decrease as the learning process evolves.

A *learning clique* is a sub-network of nodes, where the number of inner interactions is higher than the number of interactions between them and all network nodes. In Aviv et al.'s (2003) research, clique structure boosts the creation of knowledge and students who belong to more than one clique are those who transfer the information between cliques. However, Vequero and Cebrain (2013) found that during the first weeks of the course, persistent interactions among high-performing students occur, while low-performing students fail to produce reciprocity in their interactions with these closed groups. Toikkanen and Lipponen (2011) found that meaningful learning was enabled only by reading and writing to many community members (rather than settling for a few). Another interesting aspect of this matter is reciprocity. In Toikkanen & Lipponen's (2011) research, reciprocity reduced the value of an online conversation, while better conversation was achieved when several students participated and replied to others. We suggest that closed and small learning cliques in a learning network represent sub-networks with constrained and "ideal" social-constructivist learning. We suggest that learning processes go hand-in-hand with the formation of multiple connections to different learners. Our final hypothesis is therefore:

H3: In the learning community as a whole, we expect to find less learning cliques as the network evolves.

3. Method

3.1 Ligilo

Ligilo is a hyperlinked discussion platform used in blended as well as purely online asynchronous learning contexts. In Ligilo, each post is expressed as a node in a semantic network of posts. This way, learning communities actually create collective concept maps while conducting online discussions. The relations among posts are semantically tagged by community members (examples for such tags are "reminds me of", "makes me ask", "for example", or "as opposed to"). The tagged relations enable clearer comprehension of the information structured by community peers. In addition, Ligilo enables a zoomable map view of the emergent knowledge base in order to better grasp the high-level context of the constructed model, in addition to voting, following, notifying and other common social features (see Kent, Laslo, Rafaeli, & Baram-Tsabari, 2015, for further discussion and screenshots).

In terms of collected data, Ligilo is implemented as a twofold network: a network of hyperlinked posts, which underlines a social network of learners interacting based on the network of posts. In Ligilo, learners contribute to their community not only by posting, voting and reading, but also by relating to other learners' postings and tagging these relations. This is based on the known theoretical notion that meaningful learning is located in contexts and relationships rather than merely in the minds of individuals (Ausubel, 1968; Novak, 2010). This feature enabled us to collect not only social interactions, but also explicit content interactions (hyperlinks curated by the human learners).

3.2 Datasets

Twenty-eight graduate students used Ligilo for an online discussion accompanying an entire eight-week course on Virtual Communities, given by the Information and Knowledge Management department, at the University of Haifa, Israel. Participation in the discussion was rewarded by the instructor, who was also the community moderator, with a portion of 15% out of the final grade, but no minimum obligation was specified and the students were free to choose the extent and nature of their participation. The discussion's pedagogic goal was to enable conceptual coverage and deepening using online collaboration. It was initially constructed by the nine syllabus subjects, and students were asked to write about concepts raised in the course and relate them to the relevant syllabus subjects. In addition, they were advised to add and relate reading materials and to freely discuss them and the subjects in general. The instructor's active involvement involved setting up the initial framing (the nine syllabus subjects), and referring in class to some of the issues raised by the online discussants.

Individual learning was assessed using by two multiple choice examinations, before and after the semester, which focused on conceptual coverage and inferences (Haladyna, 2012). The assessment goal was to measure the delta of knowledge gained by each learner during the course, instead of just their final level of knowledge, a summative assessment which ignores the learners' initial knowledge on the subject (Campbell & Stanley, 1963).

3.3 Topological parameters

Two units of analysis were used to measure collaborative learning in this study: the learning process of an individual learner and the learning process as reflected in the structure of the entire community. In this subsection we list the parameters used to extract and reflect on their assessment as proxies of learning.

3.3.1 Learning parameters at the individual level

In Kent and Rafaeli (2016) we presented an operationalization framework for measuring interactivity patterns in individual learning within online communities. Our parameters were based on interactions and turn-taking, as well as on network analysis characteristics such as a post's level of relatedness to other posts in a discussion and out-degree. In this paper, we add the parameters such as betweenness, Bonancich's Centrality, Burt's effective network size, and exclusivity to extend the individual learning measurement (H1a). We also add an indicator for the number of cross-references a learner adds to the learning network when she adds a new relation between two existing posts (see the Ligilo screenshot in *Figure 1*).

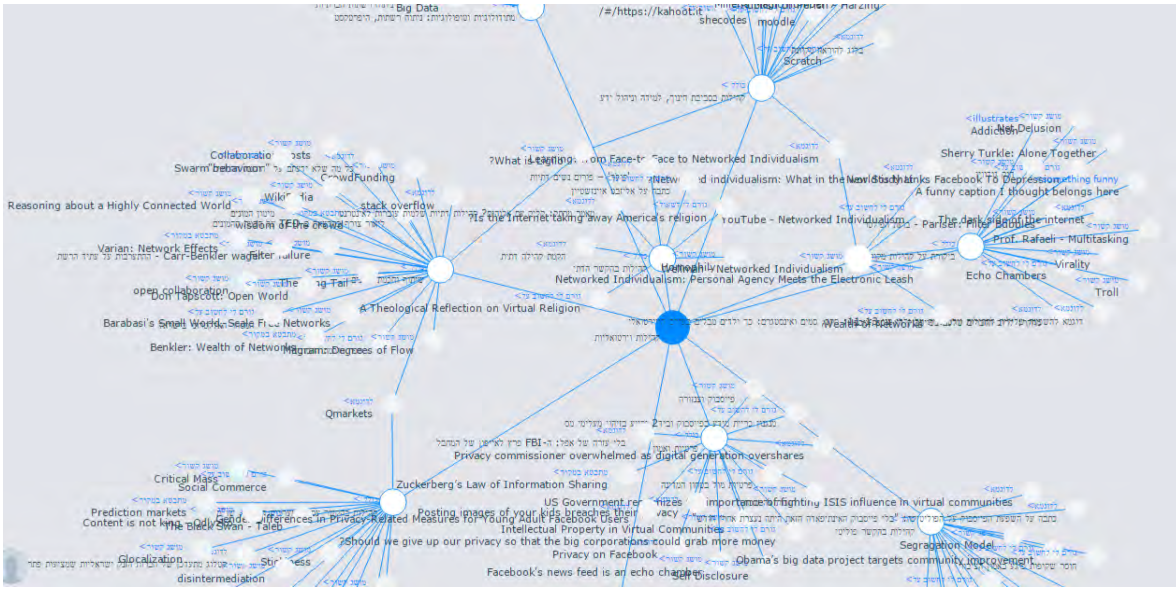
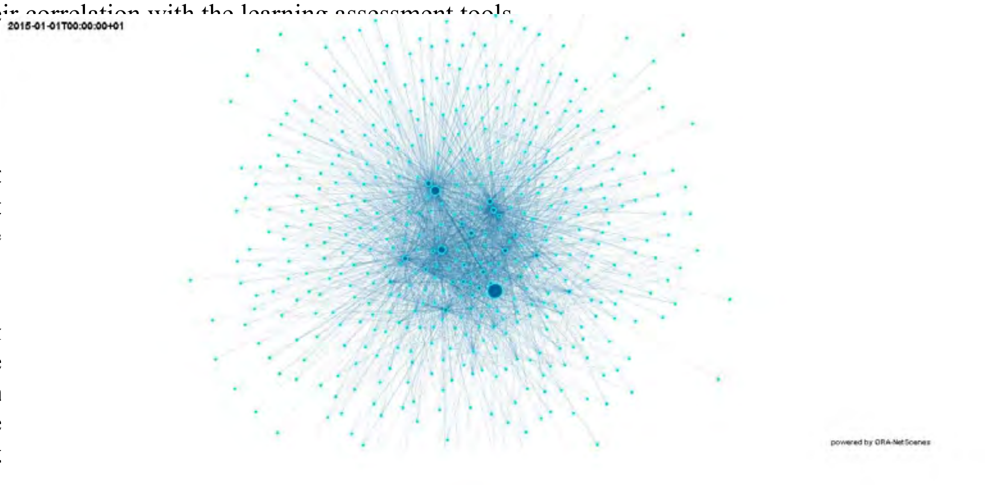


Figure 1: Cross-references (in both English and Hebrew) connecting network clusters. A learner who adds such cross-references contributes to the community not by adding new content, but by restructuring the network and bringing posts closer.

We argue that the tendency or ability to add cross-references to the network is an indicator of individual learning. Such brokers (Haythornthwaite, 2008) or bridge makers enable knowledge transfer between different parts and cliques in the network (Aviv et al., 2003). Moreover, forming new relations between already existing content items is known to implement meaningful learning (Ausubel, 1968) (H1b).

Learning outcomes are assessed or measured variously according to learning goals and educational approaches (Shavelson & Huang, 2003). In (Kent, Laslo, & Rafaeli, 2016) we suggested a positive correlation to associate individual interactivity measures to traditional learning assessments tools, such as grading and expert classification. In the current study we add our SNA measures to this equation and explore their correlation with the learning assessment tools.

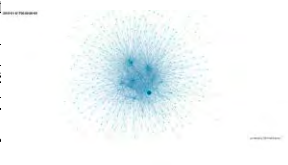


pro
net
dyn
we

con
the
am
me
dec

4.

Th
is
nun
tha
deg
and
stru



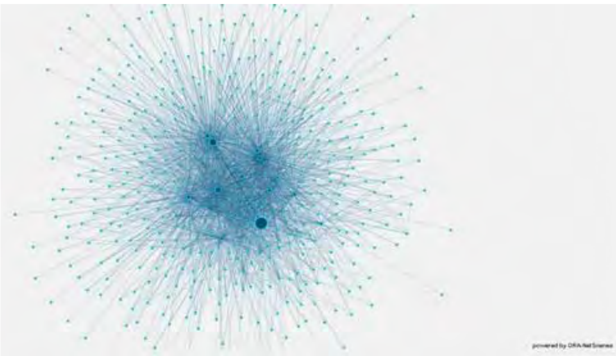


Figure 2: The hybrid learning network's final structure. Nodes are users and posts. Their size indicates the number of links to other nodes. The main nodes are User 1 and User 4.

Analyzing the **main nodes** is vital to understanding the network's structure. The nodes are either posts (P) or users (U). The most topologically important post in the network is P1, the first and central post of the discussion, created by the community moderator. P1 has the highest in-degree (.015), the higher total degree (.008), and the highest betweenness centrality. Interestingly enough, P1 also has the highest Gini index, which means it connects to multiple nodes that are different from each other in terms of their connectedness.

The most central user in the network is student U4. This node has the highest closeness centrality, the highest hub centrality, the highest out-degree centrality, the highest distinctiveness, the highest exclusivity and the largest effective network (Burt effectiveness). This means that U4 is, on average, the closest node to all others, has the largest number of outbound links to other nodes, has a unique link structure, connects to multiple nodes no other node is connected to, and reaches the highest number of nodes in the network.

U16 is another student highly involved in connecting groups. This node has the higher Simmelian ties, the higher Triad Count and the highest number of clique memberships

Three other users (U4, U19, and U37) have a star formation with sizes of 4, 5 and 10, respectively. A star pattern consists of a central node whose direct neighbors are not interconnected.

4.1 H1a – Correlation between students' grades and main topological parameters

When correlating learning as operationalized by the improvement in individual learners' grades with their topological parameters we found a medium positive correlation (0.39; $p<0.05$) between a node’s hub-centrality index and the respective user’s grade improvement. In Ligilo, users can choose to either tag the relations between posts from a list of relations provided by the moderator or create their own semantic tagging. We found a medium positive correlation (0.41; $p<0.05$) between the relative number of originally tagged relations (out of all relations created by a learner) and the improvement in the respective grade.

4.2 H1b - Network centrality parameters are positively correlated with cross-reference activity

Cross-referencing is the act of connecting two or more existing posts not previously connected. Community members engaged in cross-referencing contribute to the community’s overall learning by paving logical paths between seemingly distant issues. Some significant positive correlations were found between the amount of cross-references curated by individual learners and their topological parameters: high positive correlation (0.62; $p<0.01$) with the betweenness-centrality index of a node representing the learner; high positive correlation (0.55; $p<0.01$) with the node’s out-degree centrality index; high positive correlation (0.56; $p<0.01$) with the exclusivity index; high positive correlation (0.48; $p<0.05$) with Simmelian ties; and high positive correlation (0.55; $p<0.01$) with the network’s effective size.

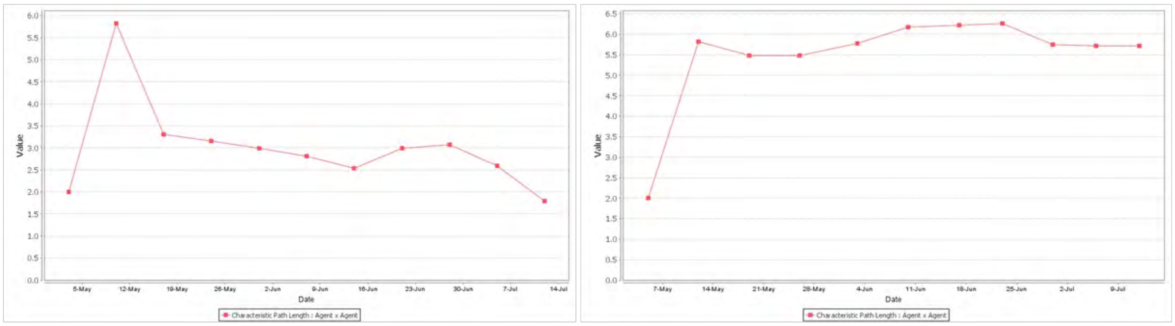
Lastly, we found significant positive correlations between the number of cross-references and several log-based counts: high positive correlation (0.6; $p<0.01$) with active behavior log counts of number of contributed posts and number of voting for posts; and a medium positive correlation (0.36; $p<0.05$) with passive behavior logs of viewing posts. The active and passive behaviors and their network-related results are analyzed and discussed in a forthcoming paper.

All the above-mentioned findings support the assumption that learning operationalized either as grade improvement or as cross-referencing is correlated with major changes in nodes' topological parameters.

4.3 H2 – The average distance decreases as the learning network evolves

The CPL of the entire hybrid network is the average distance between all nodes in the community. It depicts the distance in terms of connecting nodes between nodes and how they interact with close or far nodes.

The graph in *Figure 3*(a) presents the network's CPL on a weekly basis, and shows a decline in the average distance. During the semester, the average CPL dropped from 3 to 2.5 until Week 7. It then rose in Weeks 8 and 9 only to drop again in Weeks 10 and 11, to between 2 and 2.5. However, in order to see the evolution of the community's CPL one should look at the accumulated graph in *Figure 3*(b). The accumulated graph shows two periods: a significant increase from Week 1 to Week 2 as large numbers of nodes joined the community (both users and posts). From then until the end of the semester a more or less steady distance of 5.5 and 6 links between the nodes in the community was maintained.



(a) (b)

Figure 3: (a) Network CPL on a weekly basis; (b) Accumulated change in CPL in the entire learning network

The dynamics of the CPL can be easily explained. It rose initially since many nodes joined the network, most links being between close friends and not between strangers. As time went by and towards the end of the semester, the students began participating in the community, more links were created and the CPL slowly shrank.

4.4 H3 – As the network evolves, there are less cliques

Tightly connected nodes which are not strongly connected to other nodes can lead to a learning clique. Potentially, within learning cliques, the same opinions are heard over and over again and are not shared with others in the network. We suggest that as a part of the learning process, we should see a decrease in the number of small cliques (up to three nods). We expect to find more multi-nodal structures with multiple posts and users who create and share potential knowledge.

In order to map the evolution of cliques we first present the evolution of all nodes and links in the network. The graph in *Figure 4* presents the rise in the numbers of nodes and edges. Around Week 8, the number of nodes stabilizes while the number of links keeps climbing. This phenomenon reflects the growing density of the network with the creation of new links between exiting nodes, and in general, it presents a growing potential for exchanging opinions between learners and collaborative learning.

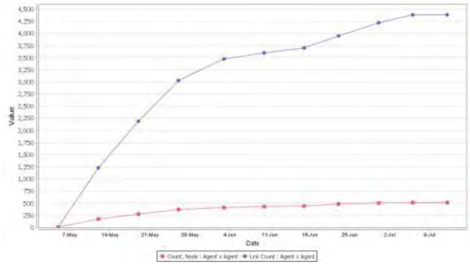
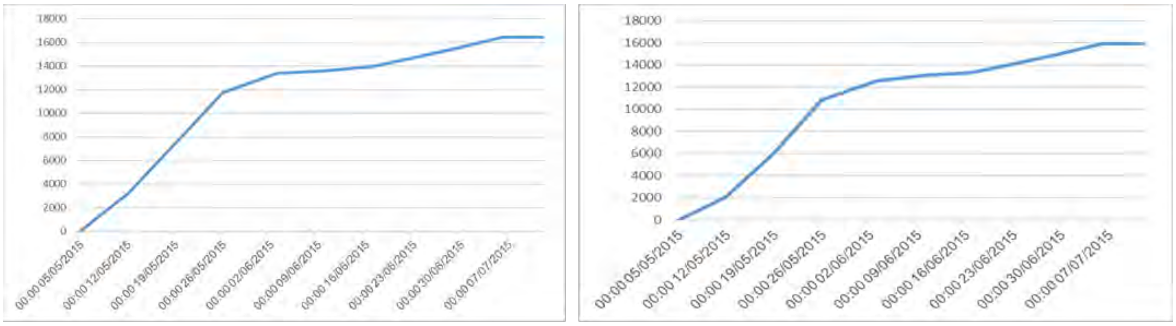


Figure 4: Growing number of nodes and edges during the learning process

Figure 5(a) presents the evolution of the total number of cliques during the semester. It is clear that the number of cliques keeps growing gradually. Almost all cliques are triads – a topological structure of three interconnected nodes. *Figure 5*(b) presents the evolution of the triads in the semester. Note that this graph is extremely similar to that depicting all cliques.



(a) (b)

Figure 5: (a) Growth in number of all cliques during the learning process; (b) Growth of triads

Next, we investigate the evolution of the number of cliques in comparison to the growth in the number of nodes and edges. The number of cliques per node and per link is relevant to estimating the cliques' dynamics and role in the learning community (see *Figure 6* below).

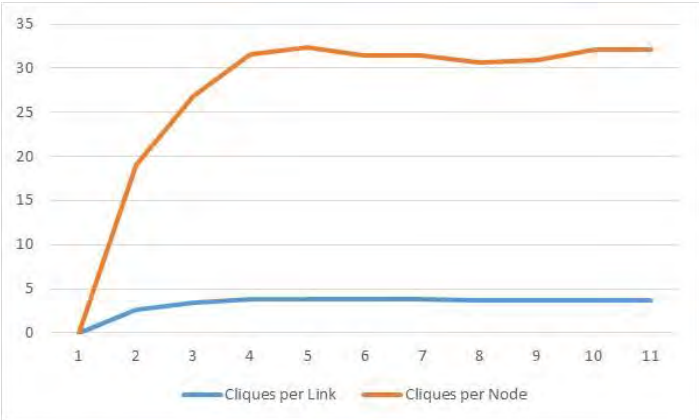


Figure 6: Starting in Week 4, the number of cliques per node stabilizes and thus the number of cliques is actually not growing but remains the same till the end of the semester

Another important type of a clique is the Simmelian tie. Simmelian ties are considered the strongest cliques, since they must contain three or more reciprocal ties between nodes in the clique. In our learning community, only 11 learners created Simmelian ties. *Figure 7*(a) presents the Simmelian ties' evolution in the learning community. *Figure 7*(b) shows the decline in the overall Simmelian index. The graph presents the early growth of tightly reciprocal ties between three or more users in the community. Then, after a short while, users stopped this kind of communication and started to relate to other users (except for triad members). This phenomenon depicts the evolution of sharing the knowledge and ideas between all users instead of talking and thinking in a closed clique.

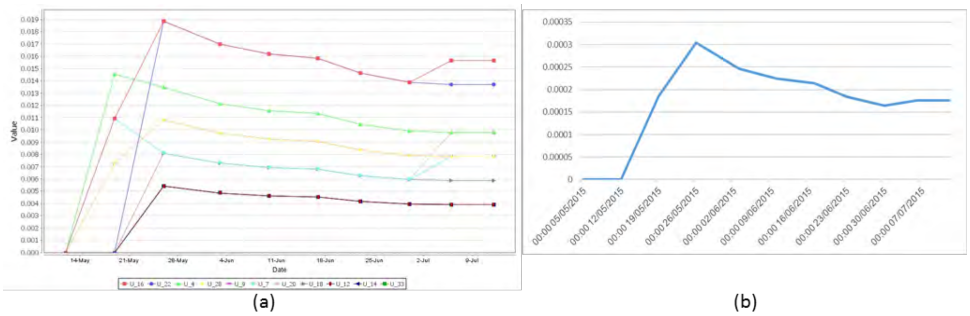


Figure 7: (a) Growth in the community's Simmelian ties; (b) Decline in the overall Simillian index

To summarize the results regarding cliques, the absolute number of cliques in the community rises during the semester. The majority of cliques are triads and their number increases. However, starting from Week 4 the number of cliques, and specifically the number of triads per node, does not change. On the whole, we identify two contradictory trends. The rise in the number of cliques suggests that knowledge is kept in close circles of friends and there is no growth in knowledge flow, and thus community learning does not evolve. On the other hand, the steady number of cliques per node suggests that there is no change in the structure that can suggest some community evolvement. Nevertheless, the decrease in number of the reciprocal triads (Simmelian ties) suggests that the community does open up its closed structures to let information flow as learning evolves.

5. Discussion

We explored a case study of a learning community's online discussion. Each learner in the community acts to gain some value interacting, acquiring knowledge or getting a high grade. In striving for these values, the learner acts as a node in her social community and her actions in the online community platform change the network's topology (posts are added, links are created, etc.). These topological changes depict the community's learning process. Interactions of a single node continuously change the network structure and thus affect the entire learning community.

5.1 Main conclusions

5.1.1 H1a: At the individual level, positive correlation will be found between students' grades and their main topological parameters

Learning can be depicted in many ways and it is hard to come up with a golden rule for learning assessment. H1a is based on the notion that learning is a process of building a product of knowledge, and its effectiveness can be measured by the improvements in learners' grades. Heo et al (2010) suggest that in an online learning group, the content of discussion posts – and not their number – is the best predictor of the group's grades. In Kent et al. (2016), the average depth of a learner's posts was found to be positively correlated with grade improvement, specifically in the specific community's discussion examined here. The depth of posts may indicate on the reading depth (before submitting a new post).

The current study found that a high number of nodes' outgoing links (hub centrality) and a high portion of original relation tagging (as opposed to readymade tagging) are positively correlated with grade improvement. These results suggest that in the course of online learning, creating connections with surrounding resources and co-learners is correlated with improved grades. The findings clearly show that a learner should maintain two types of relations in a learning community: (1) to the human colleagues and (2) to the learning materials.

5.1.2 H1b: Network centrality parameters are positively correlated with cross-reference activity

Novak (1990) suggested that learning can be reflected in the activity of cross-referencing. We explored potential correlation between topological parameters and cross-referencing. Cross-referencing characterizes learners who bridge otherwise disconnected subnetworks, contributing to overall connectivity and thus to community learning. Using SNA techniques, we found a significant positive correlation between high level of cross referencing and topological parameters such as out-degree centrality, betweenness centrality, exclusivity and Burt's effective network (. The positive correlation between cross- referencing and those parameters suggests that nodes involved in connecting detached information resources have special topological parameters and a high potential to contribute to community learning. Bridges are important in dealing with structural holes (Burt, 1992) and in the online learning community we found bridging two concepts to correlate with improved grades. The fact that someone has made the conceptual connection between two learning items does not guarantee that the network will benefit – it is only the beginning and other students should follow this newly created path. However, even if no one in the community uses this link, the creator herself learned from this bridge building. Its overall usage and specific benefits should be explored in advanced analysis of learning community dynamics.

5.1.3 H2: In the learning community as a whole, the average distance between nodes will decrease as the learning process evolves

Looking at the network as a whole, we detect at first a clear rise in the CPL (of around 6 degrees of separation between nodes) and then a drop to 5.5. This dovetails with the known research showing that various social network communities experience a decrease in their diameter and CPL (Leskovec et al., 2005; Mislove et al., 2007; Leskovec et al., 2008; Kumar, Novak, & Tomkins, 2010). In various social networks people become network-friends. This acquaintance is depicted through the shrinkage of the average distance between network nodes. In our learning community, the decrease from 6 to 5.5 means that students are better connected to each other and to their posts' contents. These tighter connections can contribute to the sharing of ideas and knowledge between users. This finding supports the accepted theoretical view that content-based interaction among community participants is the building block of collaborative knowledge construction (Vygotsky, 1978; Siemens, 2005).

5.1.4 H3: As the learning network evolves, there are less cliques in its topology

We followed Toikkanen and Lipponen (2011) who showed that when learners form closed discussion groups, the overall value of community conversations suffers. Our assumption here is that as learning evolves there will be an increase in the sharing of data / ideas / knowledge over a greater portion of the learning community. We found no evidence to support our hypothesis that the learning process should be manifested in a decrease in the number of learning cliques. We did find that (1) the ratio of cliques per user remained constant and that (2) the number of Simmelian ties dropped. These findings may indicate possible evolution in network structure and hence learning processes, despite the overall increase in the number of cliques.

5.2 Limitations

The findings reported above are based on a single case study. It is a preliminary step in an extensive methodological work that connects social learning theories and learning analytics. The online discussion examined was constructed for a specific pedagogical purposed, affected by a specific background and motivations of both learners and instructor, and used a specific platform (Ligilo). There was no minimal number of contributions to the platform per user, and in general there were minimal requirements regarding the nature and structure of contributions. Thus, our findings may not be generalized to other communities whose discussions were designed for other purposes, with different discussion structures and formal requirements. Nevertheless, the methodological aspects of this study are generalizable and may contribute to future work.

Kent, et al. (2016) used the same community, along with other communities moderated in a more structured way (for example, in order to build collaborative knowledge bases, Q&A bases and more). This specific community showed a significantly lower number of correlations relative to the more structured discussions of our log-based measures with the outcome assessment, assumingly due to its freeform discussion structure. Indeed, their studies, based on content analysis, suggested that a structured discussion involves substantial participation in the first phases of knowledge construction while the unstructured discussion was stalled (Fried, 2016). Hence, correlations with grades, as well as the level or quality of interactions may well be higher in the more structured discussions.

5.3 Summary and future research

Qualitative and traditional assessments methods, side by side with contemporary pedagogy and theories, are the basis for learning analytic methodologies. In this study we have harnessed big data methods and social network analysis to scale up collaborative online learning assessment. In our study, the network community was shaped and developed by creating links between ideas and participants and interacting as learning instruments. A single learner's motivation to learn leads to the creation of digital footprints and in that changes the topological parameters of the entire learning network.

The assessment of collaborative online learning is different than the assessment of traditional learning and is still subject to the development of evidence-based, both conceptual and technical frameworks. Such assessment should strongly rely on well-established learning theories and at the same time, be cogently informed by socio-technical trends, analytic tools and contemporary changes of values in learning. Collaborative learning and knowledge sharing play a central role in our lifelong learning experience. Computational tools should be developed to capture the entire community's knowledge and its learning process, separately from the individual learner's knowledge and its acquisition process. We believe that capturing the collective formation of knowledge can only be done by analyzing networks of learning interactions, and that assessing those collective processes can, in return, feed back into the learning process of the individual, and by extension, to that of the entire community.

Analyzing further online communities and their network dynamics is required to fully understand the network topology implications of learning attributes such as group size, instruction and moderation style, and the discussion's structure.

Conceptualizing the learning community as a network of users and knowledge resources paves the way to additional research directions, such as studying each of the network elements and assess its contribution to the learning process. For example, filtering out only the content nodes and the semantic relations between them creates a knowledge map of the online community. Similarly, exploring only human nodes and their social interactions results in charting the online community's social network. Moreover, studying the active interactions (only network edges representing commenting or writing as opposed to viewing) as opposed to only passive interactions reveals the nature of the active learning in the community, and the difference in the processes of consumption versus contribution of content within a learning discussion.

6. Acknowledgements

We are grateful to the Centre for Internet Research (http://infosoc.haifa.ac.il) and the LINKS I-CORE at the University of Haifa for their support of this study.

7. References

AlDahdouh, A. A., Osório, A. J., & Caires, S. (2015). Understanding knowledge network, learning and connectivism. . *International Journal of Instructional Technology and Distance Learning*, *12*(10), 3–21.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks*, *7*(3), 1–23.

Barker, P. (1994). Designing interactive learning. In *Design and production of multimedia and simulation-based learning material* (pp. 1–30). Springer Netherlands.

Bloom, B. S. (1974). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1-2.*⌉. Longmans: McKay.

Bossche, P., Gijselaers, W., Segers, M., Woltjer, G., & Kirschner, P. (2010). Team learning: building shared mental models. *Instructional Science*, *39*(3), 283–301. doi:10.1007/s11251-010-9128-3

Campbell, D. T., Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. *NIDA Research Monograph*. Boston: Houghton Mifflin. doi:10.1016/0306-4573(84)90053-0

Cole, M., & Engeström, Y. (1993). A cultural historic approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 1–46). New York: Cambridge University Press.

Deboer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing “Course”: Reconceptualizing Educational Variables for Massive Open Online Courses. *Educational Researcher*, *XX*(X), 1–11. doi:10.3102/0013189X14523038

Erickson, T., & Herring, S. C. (2005). Persistent Conversation : A Dialog Between Research and Design, *00*(C), 7695.

Fried, A. (2016). *Social Network Analysis of Asynchronous Discussion in Online Learning Social Network Analysis of Asynchronous Discussion in Online Learning*.

Garrison, D. R. (2006). Online collaboration principles. *Journal of Asynchronous Learning Networks*, *10*(1), 25–34. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.4536&rep=rep1&type=pdf

Gibson, D., & de Freitas, S. (2015). Exploratory Analysis in Learning Analytics. *Technology, Knowledge and Learning*, 5–19. doi:10.1007/s10758-015-9249-5

Haladyna, T. M. (2012). *Developing and validating multiple-choice test items.*⌉. Routledge.

Haythornthwaite, C. (2008). Learning relations and networks in web-based communities. *International Journal of Web Based Communities*, *4*(2), 140. doi:10.1504/IJWBC.2008.017669

Haythornthwaite, C., & De Laat, M. (2010). Social Networks and Learning Networks : Using social network perspectives to understand social learning. In *7th international conference on networked learning* (pp. 183–190). Aalborg, Denmark.

Heo, H., Lim, K. Y., & Kim, Y. (2010). Exploratory study on the patterns of online interaction and knowledge co-construction in project-based learning. *Computers & Education*, *55*(3), 1383–1392.

Herring, S. (1999). Interactional Coherence in CMC. *Journal of Computer-Mediated Communication*, *4*(4), 0. doi:10.1111/j.1083

Hewitt, J. (2001). Beyond threaded discourse. *International Journal of Educational Telecommunications*, *7*(3), 207–221.

Hutchins, E. (1990). The social organisation of distributed cognition. In L. Resnik & j Levine (Eds.), *Perspectives on socially shared cognition.* Washington, DC: APA Press.

Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & De Kereki, I. F. (2016). Translating network position into performance: importance of centrality in different network configurations. In *In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 314–323). ACM.

Kent, C., Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes.

Computers & Education Journal, *97*, 116–128. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360131516300537

Kent, C., Laslo, E., Rafaeli, S., & Baram-Tsabari, A. (2015). Poster: Designing network support for online discourse based on ambient group communication studies. In *CSCL*. Gothenburg, Sweden. Retrieved from https://sites.google.com/a/edtech.haifa.ac.il/ambienttodesigned/study3

Kent, C., & Rafaeli, S. (2016). How interactive is a semantic network ? Concept maps and discourse in knowledge communities. In *HICSS*. Kauai, Hawaii.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Man-Agement*, *20*(2), 403–437.

Kovanovic, V., Joksimovic, S., Gasevic, D., & Hatala, M. (2014). Automated cognitive presence detection in online discussion transcripts Automated Cognitive Presence Detection in Online Discussion Transcripts. In *In LAK Workshops.*

Kumar, R., Novak, J., & Tomkins, A. (2010). Structure and evolution of online social networks. In *Link mining: models, algorithms, and applications* (pp. 337–357). New York: Springer.

Lander, J. (2015). Building community in online discussion: A case study of moderator strategies. *Linguistics and Education*, *29*, 107–120. doi:10.1016/j.linged.2014.08.007

Lefebvre, V. M., Sorenson, D., Henchion, M., & Gellynck, X. (2016). Social capital and knowledge sharing performance of learning networks. *International Journal of Information Management*, *36*(4), 570–579. doi:10.1016/j.ijinfomgt.2015.11.008

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*.

Leskovec, J., Lang, K. J., Dasgupta, A., & Mahoney, M. W. (2008). Statistical properties of community structure in large social and information networks. In *Proceedings of the 17th international conference on World Wide Web* (pp. 695–704). ACM.

Milgram, S. (1967). The small world problem. *Psychology Today*, *2*(1), 60–67.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (pp. 29–42). ACM.

Moore, M. G. (1989). Editorial: Three types of interaction.. *The American Journal of Distance Education*, *3*(2), 1–6.

Novak, J. (2010). *Learning, Creating, and Using Knowledge. Concept Maps as Facilitative Tools in Schools and Corporations. 2nd Edition.* Taylor & Francis.

Pedro, L., Santos, C., Batista, J., Cabral, G., Pais, F., & Costa, C. (2016). Social network analysis and digital learning environments: a framework for research and practice using the Sapo Campus Platform. In *NTED2016 Proceedings* (pp. 1061–1070).

Pena, A. (2005). Collaborative Student Modeling by Cognitive Maps. In *International Conference DFMA'05 Distributed Frameworks for Multimedia Applications, IEEE Computer Society* (pp. 6–9). Besançon, France.

Poole, D. M. (2000). Student Participation in a Discussion-Oriented Online Course: A Case Study. *Journal of Research on Computing in Education*. Retrieved from http://web.ebscohost.com/ehost/detail?sid=df477976-7238-4d41-a31a-c2886d436d03@sessionmgr4&vid=4&hid=21&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ==#db=aph&AN=3893999

Rafaeli, S. (1988). Interactivity: From new media to communication. In R. P. Hawkins, J. M. Wieman, & S. Pingree (Eds.), *Advancing communication science: Merging mass and interpersonal processes* (pp. 110–134). Newbury Park, CA: Sage.

Rafaeli, S., & Sudweeks, F. (1997). Networked Interactivity. *Journal of Computer-Mediated Communication*, *2*(4). doi:10.1111/j.1083-6101.1997.tb00201.x

Reda, K., Tantipathananandh, C., Johnson, A., Leigh, J., & Berger-wolf, T. (2011). Visualizing the Evolution of Community Structures in Dynamic Social Networks, *30*(3). doi:10.1111/j.1467-8659.2011.01955.x

Reffay, C., & Chanier, T. (2003). How Social Network Analyses Can Help to Measure Cohesion in Collaborative Distance-Learning. *Proceedings of Computer Supported Collaborative Learning*, 343–352.

Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In *Designing for change in networked learning environments* (pp. 343–352). Netherlands: Springer.

Russo, T. C., & Koesten, J. (2005). Prestige, centrality, and learning: A social network analysis of an online class. *Communication Education*, *54*(3), 254–261.

Shavelson, R. J., & Huang, L. (2003). Responding Responsibly. *Change: The Magazine of Higher Learning*, *35*(1), 10–19. doi:10.1080/00091380309604739

Shea, P., Hayes, S., Smith, S. U., Vickers, J., Bidjerano, T., Gozza-Cohen, M., … Tseng, C.-H. (2013). Online learner self regulation: Learning presence, viewed through quantitative content-and social network analysis. *The International Review of Research in Open and Distance Learning*, *14*(3), 427–461. Retrieved from http://www.sunyresearch.net/hplo/wp-content/uploads/2012/04/April-12-Latest-Revs-to-AERA-March-Final-PJS.docx

Shum, S. B., & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology & Society*, *15*(3), 3–26. doi:10.1145/2330601.2330616

Siemens, G. (2005). A learning theory for the digital age. *Instructional Technology and Distance Education*, *2*(1), 3–10.

Sinha, T. (2014). Supporting MOOC instruction with social network analysis. *arXiv Preprint*.

Stoyanova, N., & Kommers, P. (2001). Learning Effectiveness of Concept Mapping in a Computer Supported Collaborative Problem Solving Design. In *First European International Conference on Computer-Supported Collaborative Learning* (Vol. 13). Maastricht, Netherlands: Euro-CSC. doi:10.1145/1015579.810983

Stromer-Galley, J. (2000). On-line interaction and why candidates avoid it. *Journal of Communication*, *50*(4), 111–132. doi:10.1093/joc/50.4.111

Swan, K., & Ice, P. (2010). The community of inquiry framework ten years later: Introduction to the special issue. *The Internet and Higher Education*, *13*(1-2), 1–4. doi:10.1016/j.iheduc.2009.11.003

Toikkanen, T., & Lipponen, L. (2011). The applicability of social network analysis to the study of networked learning. *Interactive Learning Environments*, *19*(4), 365–379.

Vygotsky, L. (1986). Thought and language. Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). Mind in Society: the Development of Higher Psychological Process. Harvard University Pres.

Wei, W., & Carley, K. M. (2015). Measuring Temporal Patterns in Dynamic Social Networks. *ACM Transactions on Knowledge Discovery from Data*, *10*(1), 1–27. doi:10.1145/2749465

Williams, F., Rice, R., & Rogers, E. . (1988). *Research methods and the new media*. New York: Free Press.

Wu, T., Yao, Y., Duan, Y., Fan, X., & Qu, H. (2016). NetworkSeer : Visual Analysis for Social Network in MOOCs. In *In 2016 IEEE Pacific Visualization Symposium* (pp. 194–198). IEEE.

Appendix A: Topological parameters short index

- Authority and Hub Degree - For authorities it is important also from whom the links are coming from. A node is a good authority, in case many good hubs point at it and it is a good hub, in case it points good authorities.
- Betweenness Centrality - The number of shortest paths from all nodes to all others that pass through it.
- Clustering coefficient - is the ratio of existing links connecting a node's neighbors to each other to the maximum possible number of such links
- Burt's effective network - the number of nodes a node can reach in the network.
- Clique - A group of nodes which interact with each other more intensely than with others in the network.
- Eigenvector Centrality - Based on the assumption that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes it assigns relative scores to all nodes in the network to measure the influence of a node in a network
- Exclusivity - the degree to which the node is the only node which connects other (less popular) ones
- Out-degree - The number of out-going links of a node to other nodes
- Simmelian ties – A Clique with strong reciprocal links
- Total Degree – The sum of in and out degree