

## **An Approach for Assessing Large Online Communities in Informal Learning Environments**

**Rogério F. da Silva<sup>1</sup>, Itana M. S. Gimenes<sup>2</sup>, José C. Maldonado<sup>1</sup>**

<sup>1</sup>Institute of Mathematics and Computer Science – University of São Paulo (USP)

<sup>2</sup>Department of Informatics – State University of Maringá (UEM)

rogeriofrr@usp.br, imsgimenes@uem.br, jcmaldon@icmc.usp.br

**Abstract.** *Online Learning Communities (OLC), supported by social web technologies, have proved to be beneficial for collaborative knowledge building, mainly in informal environments. There is an increasing interest in assessing online Social Learning (SL) in such communities. However, there is no agreement on how their performance can be measured. This paper presents an approach which combines structure and discourse analyses to assess large online communities used in SL. Its objective is to identify conditions and behavioral patterns associated to learning. The results point out a set of quantitative measures which shows that participation and ongoing collaboration have a fundamental role for knowledge creation and sharing.*

### **1. Introduction**

Learning in a social context is a process of meaning-making where this meaning can be based upon prior experiences as well as more immediate context [De Laat and Prinsen 2014]. Social Learning (SL) is an interactive and dynamic process which takes place in a multi-actor setting where actors learn and co-create new knowledge in ongoing collaboration [Garrison et al. 2010]. Online learning environments promote SL mainly through discussion threads. Analyzing the structure and the content of discussions make it possible to identifying patterns of activity and discourse styles that indicate meaningful learning and knowledge construction [Haythornthwaite et al. 2018].

Academics and professionals have been investigating the potential of Online Learning Communities (OLC) to facilitate learning and to foster knowledge constructed in informal settings [Hafeez et al. 2018]. Online SL in informal learning environments occurs outside the conventional educational environments, such as social media or OLC. This paper presents an approach to assess SL within large online communities (more than one million users enrolled). Our approach combines structure and discourse analyses to assesses online SL through 16 SL elements. In the rest of the paper, we use the term online communities or OLC referring to large online learning communities for informal settings.

The overall aim of this paper is to present a general approach that can be applied to different informal learning settings. We have initially defined and refined our approach working with communities of online news sharing site Reddit [Weninger 2014]. Our main contribution is to systematize the assessing of social learning within OLC, thus identifying quantitative measures that are most strongly correlated with behaviors and conditions associated to learning. The findings presented in this paper can be used to extract useful conclusions about the SL process, with the aim of helping learners to become more aware of the productivity of their contributions to the community.

The next sections are described as follows: Section 2 introduces the theoretical background and related works; Section 3 describes our approach development process; Section 4 presents the empirical evaluation and results; finally, Section 5 describes our conclusions and future works.

## 2. Background and Related Work

A criticism often voiced about assessing SL or analyzing educational data in general is its atheoretical nature. A challenging aspect of educational research lies in the theoretical framework it is embedded in. However, many studies provide significant results without being explicitly positioned in any theoretical background [Nistor et al. 2015]. Based on this perspective, we have used the theoretical frameworks of value creation [Wenger et al. 2011] and social presence [Garrison et al. 2010] to support our approach.

### 2.1. Value Creation Framework and Structure Analysis

Value creation means the value of the learning enabled by community involvement and networking [Wenger et al. 2011]. The Value Creation Framework focuses on the value that communities create when they are used for SL activities, such as sharing information or tips, asking questions, learning from each other's experience, creating knowledge together, helping each other with challenges and creating knowledge in ongoing collaboration.

Wenger et al. (2011) suggest that analysis of structural aspects provide a good basis for talking about the networking value for community users. Regarding structure analysis, our approach encompasses two analytical methods: *(i) Social Network Analysis (SNA)* that helps in understanding how learners are connected and how they interact with each other; and *(ii) Discussion Structure Analysis* which investigates the post authorship and replying relationships. This helps to illustrate the learners' communication patterns that may improve the effectiveness of knowledge sharing. In general, only structure analysis is not enough for deeply understanding patterns of interactions in online discussions [De Laat and Prinsen 2014]. Therefore, we argue that structure and discourse analyses may be applied as complementary methods.

### 2.2. Discourse Analysis and Social Presence

Discourse analysis involves a set of methods which analyzes large amounts of text generated in online environments. Garrison et al. (2010) presented the Community of Inquiry (CoI) theoretical framework to evaluating the nature and quality of critical discourse created in online discussions. CoI proposes three key dimensions, known as presences [Ferreira et al. 2020]: *(i) Social presence* measures the ability to humanize the relationships among participants in a discussion; *(ii) Cognitive presence* is highly related to capture the progress of students' cognitive process; *(iii) Teaching presence* concerns teaching role during online courses.

The learners' ability to project themselves within an online community is initially identified in the social presence. We have used the indicators of social presence in discourse analysis because their purpose is to create a healthy support that provides a comfortable place for students to exchange ideas freely [Joksimovic et al. 2015]. Our approach centers around the measures extracted by using the Linguistic Inquiry and Word Count (LIWC) framework [Pennebaker et al. 2015].

### 2.3. Related work and research gaps

This section describes some observable research gaps of available studies related to online SL, in order to point out key aspects of our approach.

**Theoretical background.** Educational research widely accepts that the explicit and coherent application of learning theories supports empirical studies [Nistor et al. 2015]. However, large part of the studies about online SL positioned not provide the theories to which their results apply [Gruzd et al. 2016; Palazuelos et al. 2013].

**Combination of analytical methods.** The connection between social interactions and discourse analysis within a network is well established in numerous sociological studies. However, the combination of these two analytical methods is notably a research gap in educational research [Joksimovic et al. 2015].

**Small sample sizes.** The sample size of majority of the studies which investigates online learning communities is small. Some authors explicitly state this as a limitation [Ferreira et al. 2020; Jan 2019]. For example, Gruzd et al. (2016) analyzed a sample of public tweets posted by approximately 400 participants in a course and they determined the most connected ones, showing who is influencing the information flow.

**Differential of this work.** Our approach aims to bridge the gaps above described. We have applied the approach in an OLC with more than one million registered members, because large communities on the Internet are nowadays one of the most important producers of Big Data. However, there is still a shortage of research related to the evaluation of such communities. The number one million enrolled users can be understood as an indicator of the context in which our study is addressed to.

## 3. Development process

Our development process shows how the assessment of different aspects of SL within online communities, such as the level of participation, activity and engagement, the quality of interactions and characteristics of participants is obtained. It has three stages described in the following subsections. The participation in an OLC is usually structured as discussion topics. Two data structures were used to represent each discussion topic: post tree and network of interactions. A schematic representation of a typical online discussion is shown in Figure 1(a), where circles represent the learners who posted or replied to discussion topics (rectangular boxes), and arrows represent the response relation between posts. Figure 1(b) shows a post tree that represents the discussion, and Figure 1(c) shows a network of interactions where numbers (weights) in arrows represent interaction frequency. These data structures are analyzed throughout the stages of the approach development process, as suggested in Figure 2.

### 3.1. Stage 1: Defining the SL elements and their measures

Stage 1 is supported by Table 1 and 2 to systematize the assessment of online SL. Table I describes the indicators and SL elements suggested in the value creation framework to support the structure analysis. Table II reports the SL elements based on social presence framework to perform the discourse analysis. Beyond the indicators, both tables contain the following columns: *(i)* the SL elements based on SL activities; *(ii)* the measures used to assess such elements; *(iii)* Source represents the data structure used to compute the measure; and *(iv)* description states a brief comment about the measure.

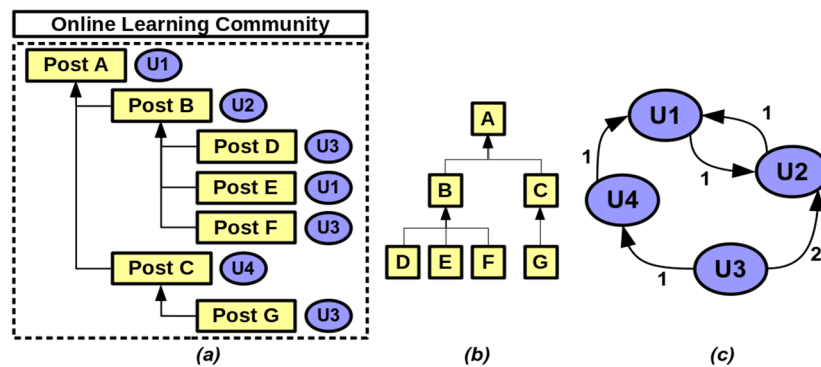


Figure 1. Typical online discussion and their data structures

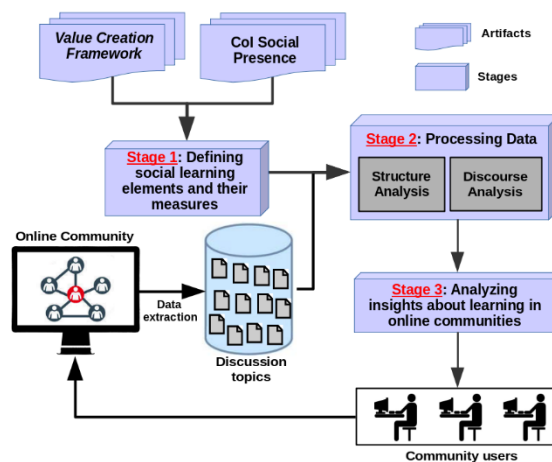


Figure 2. Approach Development Process

Although value creation and social presence frameworks suggest the indicators that enable to assess the SL elements within OLC, their authors do not offer details on how to use them in practice. Thus, we have used the empirical evidence pointed out in the educational research [Antonacci et al. 2017; Cowan and Menchaca 2014; Ferreira et al. 2020; Joksimovic et al. 2015] to propose the measures used to operationalize the assessment of SL elements in Table I and Table II.

### 3.2. Stage 2: Processing Data

Stage 2 uses computational techniques in order to both extract the files that represent the discussion topics of OLC and assesses SL elements, by using structure and discourse analyses, described as follows.

**Structure analysis** has used SNA to analyze the network of interactions and compute the measures in Table 1 whose source is the data structure network. In addition to SNA, we have used search tree algorithms to analyze the discussion structure and compute the measures which have their origin in post tree.

**Discourse analysis** has used the LIWC measures presented in the Table 2 in order to identify which word categories may indicate discourse styles associated to learning. The approach described in this paper is based on the word count strategies, which is geared toward revealing the psychological meaning of words, independently from their literal and semantic contexts. The unit of discourse analysis has been the discussion topic. Thus, our strategy to compute the discourse measures is described as follow:

- 1) first, for all discussion topics, we have calculated the average of LIWC measures presented in Table 2; for example: considering the discussion illustrated in Fig. 1(a), the value of each measure has been the average of Post A to Post G;
- 2) second, we have used the clustering algorithm *Kmeans* [Hartigan and Wong 1979] and the *Elbow* method [Bholowalia and Kumar 2014] to divide the discussion topics in four clusters (groups), based on the values of LIWC measures;
- 3) lastly, we have analyzed the clusters in order to investigate which of them presented more behaviors and conditions associated to learning; thus, we could identify which LIWC measures have been more relevant in such clusters.

**Table 1. Indicators and SL Elements of Structure Analysis**

Indicators	SL Elements	Source	Measure	Description
Level of participation	Characteristics of participants	Post tree	Number of Threads	Number of discussion topics created by user.
		Network	In-degree	Sum of incoming links.
		Network	Out-degree	Sum of outgoing links.
		Network	Betweenness centrality	Users who connect different groups of users in the network.
		Network	Closeness centrality	The degree to which an user is close to all other users.
		Network	Eigenvector centrality	The user's importance considering the importance of its neighbors.
		Network	HITS Hub	A good hub is an user (node) which has many incoming links.
		Network	HITS Authority	A good authority is an user (node) which has many outgoing links.
	Active participants	Network	PageRank	It considers the number of node's links ingoing and outgoing.
		Network	Number of participants	Number of active participants (nodes) in the network.
Level of activity	Post tree	Discussion topics	Number of discussion topics created.	
	Post tree	Size of discussions	Amount of replies in each discussion topic. <b>In Fig. 2(b), size of discussion = 6.</b>	
	Network	Time of first reply	Time until the first reply (in hours).	
Level of engagement	Network	Network density	Number of different links compared to highest possible.	
	Post tree	Width	The number of branches that post tree's root node has. <b>In Fig. 2(b) width = 2.</b>	
		Depth	The length of the root to the deepest node in the post tree. <b>In Fig. 2(b) depth = 3.</b>	
	Network	Reciprocity coefficient	The tendency to form a mutual connection between user pairs.	
	Post tree	Discussion intensity	Total number of replies / number of users.	
Quality of interactions	Post tree	Discussion duration	Total discussion period for each discussion topic (in hours).	
	Post tree	Score of discussion topics	Average score of replies.	
	Network	Sub-communities	Number of the sub-communities in the network.	
Structural shape of network	Network	Avg shortest path	The average of the shortest paths between each pair of nodes.	
		Clustering coefficient	It measures how well users' neighbors are connected to be a clique.	
		Number of Triads	Number of cliques with three users.	
	Network	Bottlenecks	Number of users that connect different sub-communities.	

**Table 2. Indicator and SL Elements of Discourse Analysis**

Indicator	SL Elements	Source	Measures	Description	Examples
Social Presence	Affective	Post tree	liwc.pronoun	Number of pronouns	he, they, itself
		Post tree	liwc.ppron	Number of personal pronouns	them, she, thou
		Post tree	liwc.i	First-person singular pronouns	I, me, mine
		Post tree	liwc.we	First-person plural pronouns	we, us, our
		Post tree	liwc.ipron	Number of impersonal pronouns	it, it's, those
		Post tree	liwc.affect	Affective processes	happy, cried
		Post tree	liwc.posemo	Positive emotion	love, nice, sweet
		Post tree	liwc.negemo	Negative emotion	hurt, ugly, nasty
		Post tree	liwc.work	Personal concerns: work	job, majors, xerox
		Post tree	liwc.power	Words related to power	superior, bully
		Post tree	liwc.drives	Words related to drives	above, lose, helps
		Post tree	liwc.percept	Perceptual processes	look, heard, feeling
		Post tree	liwc.negate	Negations	no, not, never
	Post tree	liwc.interrog	Interrogatives	how, when, what	
	Interactive	Post tree	liwc.focuspresent	Focus on the present	today, is, now
		Post tree	liwc.auxverb	Auxiliary verbs	am, will, have
		Post tree	liwc.you	Second-person pronouns	you, your, yourself
		Post tree	liwc.assent	Words related to assent	agree, OK, yes
		Post tree	liwc.focuspast	Focus on the past	ago, did, talked
	Cohesive	Post tree	liwc.affiliation	Affiliation	ally, friend, social
Post tree		liwc.social	Social processes	mate, talk, daughter	

### Stage 3: Analyzing insights about learning

Based on the results analyzed previously, this stage seeks insights about online SL in large online communities. These insights provide to researchers, learners and moderators of OLC valuable information about student participation and discourse style in informal

settings. In addition, informing learners of their level of interaction and increasing awareness of the status of collaboration with their peers, may lead to enhanced self-regulation of social interaction and knowledge sharing in OLC.

#### 4. Empirical evaluation and results

Our approach has been evaluated in the online news sharing site Reddit, the third most visited USA site in 2018 [Haythornthwaite et al. 2018]. Participants, known as *redditors*, can rate the discussion topics and their replies, thus creating the discussion *score*. The *score* can be considered a popularity counter, because the more votes a discussion has received, the more attractive it is. *Redditors* are also able to assign points to each other responses. These points, named *karma*, indicate the extent to which members are active, respected and therefore reflect their invested time and effort in preparing high-quality responses before online posting [Weninger 2014].

##### 4.1. Data Collection

We have analyzed the data of a *subreddit* (Reddit online community) about learning programming, named *r/learningprogramming*. We have extracted all discussion topics and their replies posted between 2018-Jan-01 and 2019-Dec-31. At the evaluation snapshot (March 2020), the *subreddit* had 1,292,842 members enrolled. The period under investigation (two years) had 59,787 discussion topics (tree posts), 82,077 unique active participants (network nodes) and 395,556 interactions (network links).

##### 4.2. Applying the approach

It is important to clarify two key terms used henceforth: cross-sectional and aggregated networks. A cross-sectional network represents a snapshot of all interactions between nodes in a network at a certain point in time. We need to look at changes in the structure of the successive cross-sectional networks to explore temporal dynamics of community. We divided the analysis period of two years into 24 cross-sectional networks, each corresponding to one month, in order to analyze the evolution of measures related to time. In addition, we need to examine the aggregated network which captures cumulative interactions over the entire period (two years) to discover the global orientation of the community.

**Analyzing aggregated network.** The SL element “Characteristics of participants” in Table 1 represents the measures which assess social learning individual perspective. We have used these measures as predictor variables on the user's *karma* through a multilevel linear regression model (see Table 3), in order to identify the most significant measures which reveal the user expertise.

**Analyzing cross-sectional networks.** The others SL elements provide a whole network view of communal learning. The temporal analysis of cross-sectional networks reveals the evolution of learner's behavior in the period under investigation. However, before analyzing the evolution of structure and discourse measures, we have used a linear regression model and the clustering algorithm *Kmeans* to identify the most relevant ones in three steps, described as follow. **(I)** Identifying the most relevant structure measures: first, we have analyzed the data structures (network and post tree) of each discussion topic individually. The structure measures were correlated with its *score* in a multilevel linear regression model, in order to identify the most significant ones (see Table 4). **(II)**

Identifying the most relevant discourse measures: second, we have applied the clustering algorithm *Kmeans* in order to identify which discourse measures were more strongly related to behaviors and conditions associated to learning (see Table 5 and Figure 3). **(III)** Analyzing the evolution of measures: lastly, we have evaluated the evolution of the most relevant structure and discourse measures in the cross-sectional networks along the time (see Figure 4 and Figure 5).

The results of analyzing aggregated and cross-sectional networks are described in next subsections.

### 4.3. Analyzing the aggregated network

Table 3 shows the Pearson's correlation coefficient between user's *karma* and the measures of SL element "Characteristics of participants". They indicate which structure measures are more relevant, marked in bold, to identify expert users.

**Table 3. Pearson's correlation coefficient to user's *karma***

<i>Karma</i>	1.00	<b>Out-degree</b>	<b>0.64***</b>	Closeness	0.01	Authority	0.20
<b>In-degree</b>	<b>0.81***</b>	<b>Betweenness</b>	<b>0.69***</b>	Hub	0.07	<b>PageRank</b>	<b>0.77***</b>

Note: \*\*\* means *p-value* < 0.001

### 4.4. Analyzing cross-sectional networks

The next subsections describe the three steps of analyzing cross-sectional networks. The results of linear regression model reported in Table 4 and clustering algorithm *Kmeans* helped us to identify the most significant structure and discourse measures.

#### 4.4.1 Identifying the most relevant structure measures

Discussions topics with higher *score* present stronger conditions associated with learning, because they have the best rated responses by users. Table 4 shows the Pearson's correlation coefficient between structure measures and the discussion topic *score* (the most relevant ones are highlighted in bold). *Number of participants*, *Size* and *Width* emphasize the importance of participation and interaction between users. The measures *Number of sub-communities*, *Number of bottlenecks* and *Number of Triads* also presented a strong correlation with discussion topic *score*. This could be explained by factors as the homophily: a property which states that users interact with others who are like them in one or more ways.

**Table 4. Pearson's correlation coefficient to discussion topic *score***

Discussion topic <i>Score</i>	1.00	<b>Width</b>	<b>0.80***</b>	<b>Sub-communities</b>	<b>0.83***</b>
<b>Number of participants</b>	<b>0.88***</b>	Depth	0.22	Avg shortest path	0.00
<b>Size</b>	<b>0.85***</b>	Reciprocity	-0.03	Clustering coefficient	0.00
Time of first response	0.00	Discussion intensity	0.03	<b>Number of bottlenecks</b>	<b>0.57**</b>
Network density	0.00	Discussion duration	0.26	<b>Number of Triads</b>	<b>0.54***</b>

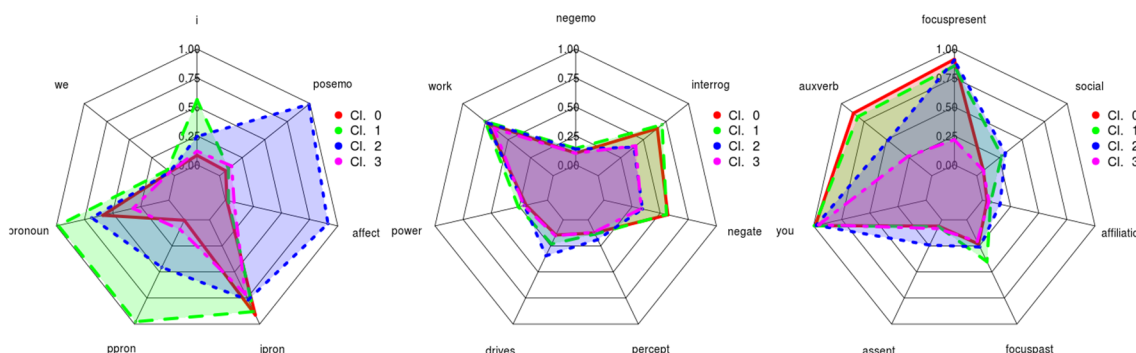
Note: \*\*\* means *p-value* < 0.001; \*\* means *p-value* < 0.01

#### 4.4.2 Identifying the most relevant discourse measures

Longer discussions were selected at this step, with size greater or equal to five (at least five replies). Consequently, 22,019 discussion topics were used as data set to the *Kmeans* clustering algorithm. Table 5 shows the results: for each cluster, we calculated the average of measures which presented strong correlation with discussion topic *score*. Cluster 2 presented the highest average for all measures (marked in bold), except *Number of bottlenecks*. Thus, we argue that such cluster has discussions whose discourse style can be more strongly associated with learning. Figure 3 shows the cluster normalized *centroids*: a value in the interval [0, 1] in which the most significant ones are the closest to one. The most relevant discourse measures in Cluster 2 were: *posemo*, *affect*, *drives*, *percept*, *assent*, *affiliation* and *social*.

**Table 5. Results of *Kmeans* clustering algorithm**

Cluster	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Size of Cluster	6,874	5,080	4,807	5,258
% of Total	31.22%	23.07%	21.83%	23.88%
Score	72.81	75.59	<b>154.84</b>	67.42
Participants	7.80	7.67	<b>9.49</b>	7.68
Size	12.73	13.24	<b>15.30</b>	11.20
Width	5.19	5.46	<b>6.78</b>	5.39
Sub-communities	5.42	5.10	<b>5.89</b>	5.74
Triads	0.23	0.19	<b>0.25</b>	0.18
Bottlenecks	2.13	1.95	1.96	<b>2.22</b>



**Figure 3. The cluster normalized *centroids***

#### 4.4.3 Analyzing the evolution of measures

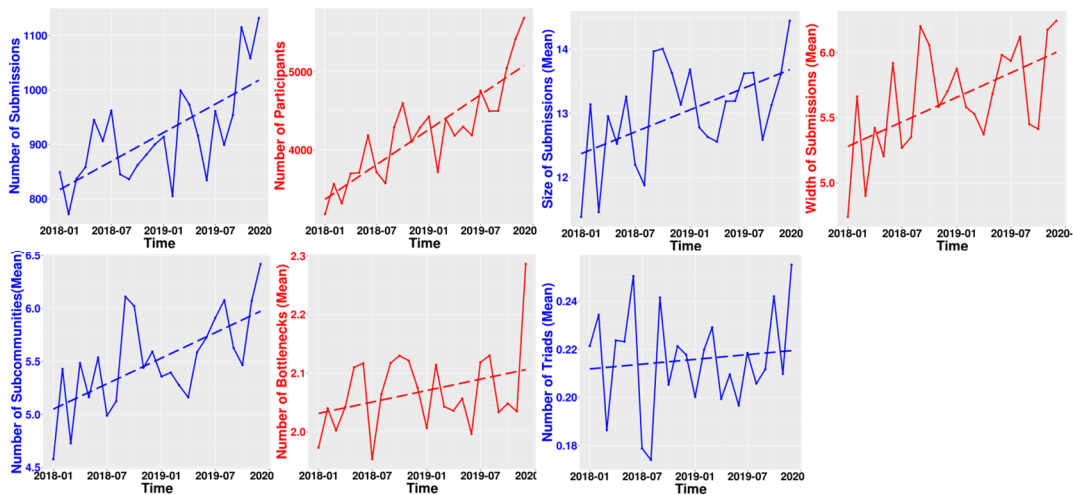
Successive cross-sectional networks and post trees were analyzed in order to reveal the evolution of measures most strongly associated with learning throughout the two years. The linear regression models depicted in Figure 4 and Figure 5 show that all measures presented positive correlation, indicating growing trend related to time. Thus, structure measures (Figure 4) suggest the increasing of user participation in *subreddit r/learnprogramming*, and discourse measures (Figure 5) suggest that discourse style related to learning has occurred in a sustainable way along the period.

### 5. Conclusions and future work

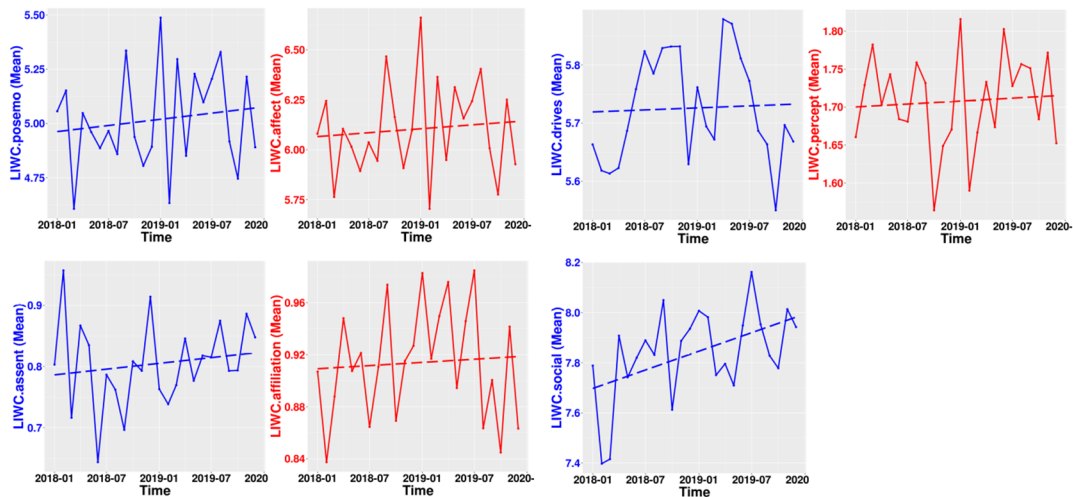
This paper presents a holistic approach that combines structure and discourse analyses, with the purpose of providing to researchers, users and moderators of online communities valuable information about student participation and discourse style in informal settings.



Our approach can be applied to online communities which users can rate each other's responses and discussion topics. We have applied it to a subreddit from the online news sharing site Reddit. The results pointed out that the most important factors fostering the behavior and conditions associated with learning in online communities are the measures related to the amount of participation. Regarding structure analysis, the growing trend detected in the linear regression models of cross-sectional networks suggests a virtuous circle. This circle shows that an increasing participation in SL activities related to knowledge-sharing leads to greater participation and so forth. The clustering algorithm used in discourse analysis supported the identification of common groups of discourse styles emerging from discussion topics. Such groups revealed a set of discourse measures associated with learning, according to the results of structure analysis. These measures indicated the conservation along the time of a discourse style related to behavioral patterns which suggest knowledge creation and sharing. Future works aim to improve the discourse analysis in order to investigate how the expression of positive (or negative) sentiment and emotional states in the messages may affect the level of participation or the discourse style in informal learning settings.



**Figure 4. Linear regression models of structure measures**



**Figure 5. Linear regression models of discourse measures**

## References

- Antonacci, G., Colladon, A. F., and Stefanini, A. (2017). It is rotating leaders who build the swarm: Social network determinants of growth for healthcare virtual communities of practice. *Journal of Knowledge Manag.*, v. 21, n. 5, p. 1218–1239.
- Bholowalia, P. and Kumar, A. (2014). Ebc-means: A clustering technique based on elbow method and k-means in wsn. *Int. J. of Computer App.*, v. 105, n. 9, p. 79.
- Cowan, J. E. and Menchaca, M. P. (2014). Investigating value creation in a community of practice with social network analysis in a hybrid online graduate education program. *Distance Education*, v. 35, n. 1, p. 43–74.
- De Laat, M. and Prinsen, F. (2014). Social Learning Analytics: Navigating the Changing Settings of Higher Education. n. 2014, p. 51–60.
- Ferreira, M., Rolim, V., Mello, R. F., et al. (2020). Towards Automatic Content Analysis of Social Presence in Transcripts of Online Discussions. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge*.
- Garrison, D. R., Anderson, T. and Archer, W. (2010). The first decade of the community of inquiry framework. *Internet and Higher Ed.*, v. 13, n. 1–2, p. 5–9.
- Gruzd, A., Paulin, D. and Haythornthwaite, C. (2016). Analyzing Social Media and Learning Through Content and Social Network Analysis: A Faceted Methodological Approach. *Journal of Learning Analytics*, v. 3, n. 3, p. 46–71.
- Hafeez, K., Alghatas, F. M., Foroudi, P., Nguyen, B. and Gupta, S. (2018). Knowledge sharing by entrepreneurs in a virtual community of practice (VCoP). *Information Technology and People*,
- Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society*, v. 28, n. 1, p. 100–108.
- Haythornthwaite, C., Kumar, P., Gruzd, A., et al. (2018). Learning in the wild: coding for learning and practice on Reddit. *Learning, Media and Technology*, v. 9884.
- Jan, S. K. (2019). Investigating VCoP with social network analysis: guidelines from a systematic review of research. *Int. J. of Web Based Communities*, v. 15, n. 1, p. 25.
- Joksimovic, S., Gasevic, D., Kovanovic, V., Riecke, B. E. and Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, v. 31, n. 6, p. 638–654.
- Nistor, N., Derntl, M. and Klamma, R. (2015). Learning Analytics: Trends and Issues of the Empirical Research of the Years 2011 – 2014. *Design for Teaching and Learning in a Networked World*, v. 4, p. 453–459.
- Palazuelos, C., García-Saiz, D. and Zorrilla, M. (2013). Social network analysis and data mining: An application to the E-learning context. *Lecture Notes in Computer Science*, v. 8083 LNAI, p. 651–660.
- Pennebaker, J. W., Boyd, R. L., Jordan, K. and Blackburn, K. (2015). The development and psychometric properties of LIWC2015.
- Wenger, E., Trayner, B. and De Laat, M. (2011). *Promoting and assessing value creation in communities and networks: a conceptual framework*. v. 18
- Weninger, T. (2014). An exploration of submissions and discussions in social news: Mining collective intelligence of Reddit. *Social Network Analysis and Mining*, v. 4, n. 1, p. 173–192.