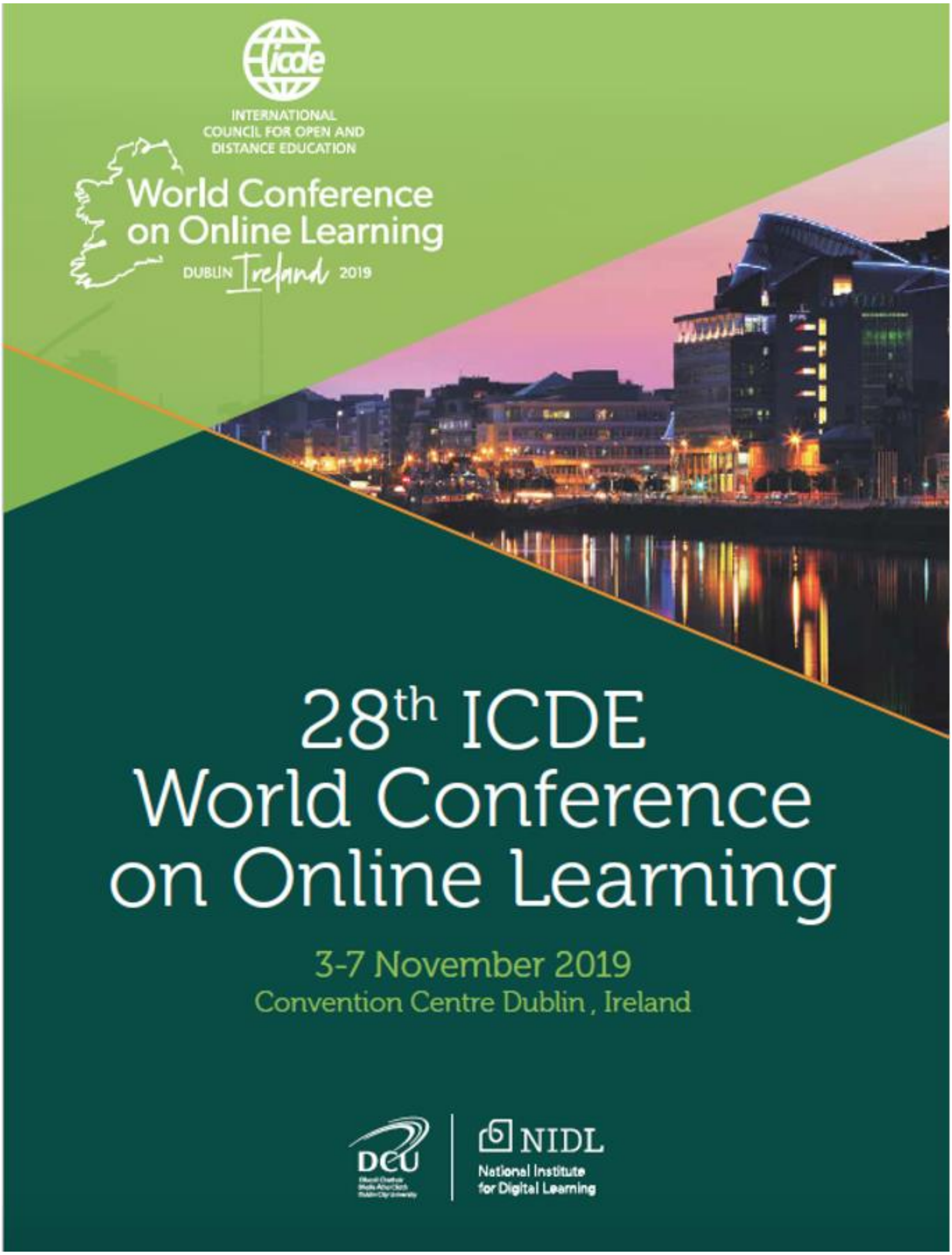
2020

Employing Social Learning Analytic Methods (SLAMs) to Reimagine the Social Dynamic of Online Learning Collaborations

Damien Sanchez
University of New Mexico

Nick Flor
University of New Mexico

Charlotte Nirmalani Gunawardena
University of New Mexico, lani@unm.edu



Conference Proceedings

Volume 1: Full Papers

2020

Please cite as:

Brown, M., Nic Giolla Mhichil, M., Beirne, E., & Costello, E. (eds.) (2020). *Proceedings of the 2019 ICDE World Conference on Online Learning, Volume 1,* Dublin City University, Dublin. http://dx.doi.org/10.5281/zenodo.3804014

ISBN: 978-1-911669-10-4

https://wcol2019.ie

@WCOL2019

Acknowledgements
The Editors would like to acknowledge the contributions of both Denise Brown and Sandra Forster to the 2019 ICDE World Conference Proceedings. We would also like to acknowledge the invaluable contribution members of our own team in the National Institute for Digital Learning (NIDL) made in helping to stage such a major international event. The success of the event would not have been possible without the efforts of Happening Conferences and Events and we are grateful to Anne Doherty and her amazing team for all of their work. Lastly, we would like to thank the ICDE team for their support in helping Dublin City University (DCU) to host the World Conference.

happening
conferences & events

TABLE OF CONTENTS

Employing Social Learning Analytic Methods (SLAMs) to Reimagine the Social Dynamic of Online Learning Collaborations

Damien Sanchez[1], Nick Flor[1], Charlotte "Lani" Gunawardena[1],

[1]*University of New Mexico, USA*

Abstract

As online networks expand, learning collaborations will occur across disciplines, countries, and people. A particular challenge is determining the process and progress of these collaborations, and how the social dynamics of interacting groups support knowledge construction. Traditional methods that rely predominantly on content analysis of transcripts to determine social construction of knowledge are time consuming to conduct and often do not provide actionable data to improve the process before it is completed. One answer to this challenge is emerging Social Learning Analytic Methods (SLAMs) that offer robust and expedient means to analyze the performance of interacting groups online. The purpose of this study is to explore the social dynamic that supports knowledge construction in interacting groups by employing SLAMs. We will use a twofold approach. First, we will demonstrate how SLAMs can be utilized in a formal learning environment during the rollout of an online university course so that findings can be used to improve the course as it is being offered. Second, SLAMs will be applied to Twitter which supports informal online learning to determine social construction of knowledge with a limited character set. SLAMs examined include, but are not limited to, sentiment analysis and social network analysis. These analyses can provide a valuable snapshot during roll out of collaborations whether in online courses or on social media so that they can be improved before they are concluded. By integrating SLAMs into online learning experiences, digital scholarship can reimagine online design, teaching, and evaluation to help meet the future needs of online collaborations in a variety of contexts. Using SLAMs, this study found that in a formal learning environment, higher phases of knowledge construction may be associated with postings containing high levels of positive sentiment and social presence. Finally, this study found that in social media informal learning environments, the social construction of knowledge occurred primarily within various stages of PI of the Interaction Analysis Model.

Keywords: Social Network Analysis, Social Learning Analytics Methods, Sentiment Analysis, Clustering Analysis, OILS Twitter Scraper

Introduction

This study uses data from both a formal and informal learning environment in order to illustrate the fidelity of Social Learning Analytics Methods (SLAMs) in both domains. Formal learning environments encompass a broad range of contexts including face-to-face classroom instruction and online learning. The common chord being an association with a traditional learning experience. This study uses a discussion board from an online course as its formal learning environment of interest. On the other hand, informal learning environments include just about everything outside of the traditional classroom as well as learning that comes from everyday experiences (Merriam & Caffarella, 1999). One of the most promising frontiers of informal learning is social media because it “is clearly the largest, richest and most dynamic evidence base of human behavior, bringing new opportunities to understand individuals, groups and society” (Batrinca & Treleaven, 2015, p. 90).This study uses Twitter as its informal learning environment of interest.

Twitter is a social media tool that allows users to compose single messages, called Tweets, to discussions organized according to thematic hashtags. Once a user has created a Tweet, users can take a variety of actions to share the information including creating a reply by writing someone’s username at the beginning of the Tweet or retweeting content by forwarding it to people within their networks (Twitter, 2016). Part of what makes Twitter a strong informal learning resource is its ability to help users organize knowledge. “Twitter hashtags help us to find discussions, snippets of knowledge, and hyperlinks to further resources from which we may learn” (Dron & Anderson, 2014, p. 181).

This study uses Social Construction of Knowledge (SCK) as its framework to identify learning because both formal discussion forums and informal Twitter dialogue share a need for strong social interaction. SCK is a subset of social constructivist theory, along with the zone of proximal development, which establishes the vital role socialization plays in the learning process (Vygotsky, 1978). Pea (1993) observes that "Knowledge is commonly socially constructed, through collaborative efforts toward shared objectives or by dialogues and challenges brought about by differences in persons' perspectives" (p. 48). The shared objectives aspect is especially pertinent to the formal learning environment as all students seek to participate in the activity. The dialogues based on various perspectives is central to the innumerable voices that contribute to Twitter hashtags.

A variety of different methods have been used to measure SCK. Notable among these are the approaches proposed by Henri (1992) and Gunawardena, Lowe, and Anderson (1997). Both of these works focus on types of Computer Mediated Communication (CMC), specifically online discussion forums. Henri (1992) recommends content analysis as one of the most important methods that should be used to understand SCK because of the complexities in assessing social and cognitive processes that underlie online discussions. Content analysis should be conducted according to 1) social dimension, 2) interactive dimension, 3) cognitive skills, and 4) metacognitive skills. One drawback to this approach is that it is difficult to distinguish between cognitive and metacognitive application of skills. In addition, interaction is descriptive and only one category is provided to assess this critical aspect of online learning.

Gunawardena et al. (1997) developed the Interaction Analysis Model (IAM) to qualitatively examine interactions among a collaborative group during the process of knowledge construction. The IAM was employed to examine the interaction that occurred in an online global debate to determine whether knowledge was constructed within the group through dialogue and discourse, and whether participants changed their understanding or developed new knowledge as a result of group interaction. Based on social constructivist and sociocultural views of learning, the model describes five phases of knowledge co-construction that are identified via content analysis: sharing and comparing constitute Phase I; dissonance is the focus of Phase II; negotiation and co-construction comprise Phase III, testing tentative constructions is incorporated in Phase IV, and statements and application of newly co-constructed knowledge are at the heart of Phase V.

Regardless of the method used to analyze SCK, both of these approaches agree that some content analysis is required to determine the context and meaning of the messages composed in online spaces. Unfortunately, reading through mountains of messages is painstaking and time-consuming work. Therefore, there is a need for additional tools that can facilitate the work of analysts who seek to understand SCK.

The Utility of SLAMs to SCK Research

Incorporating SLAMs into studies focused on identifying SCK can address the shortcomings identified by researchers using existing methods such as Lucas and Moreira (2015) who recognize “content analysis per se disregards the temporal dimension of interactions and should, therefore be complemented by other methods that can help researchers better understand such processes and facilitate additional in-depth analysis” (p. 1505). For example, time is a primary focus of the analysis conducted when using sentiment analysis and social network analysis (SNA). Datapoints are collected over time in both methods to clearly illustrate the growth of knowledge (sentiment analysis) and how frequent interactions between people gradually develop (SNA). In Gunawardena, Flor, Gómez, and Sánchez (2016) we argue that SLAMs should be used to assist researchers who are working with large volumes of qualitative data. SLAMs are powerful methods that can help researchers quickly understand trends in large and unstructured datasets, but they are no replacement for the rigor afforded by content analysis. This study builds on this assertion by providing further details regarding how SLAMs and IAM-oriented content analysis compliment one another.

By reimagining how mixed method research can be conducted using SLAMs along with traditional content analysis we suggest a way to transform the current practice of teaching online. For example, it is very challenging for online instructors to determine the quality of the interactions that take place within course discussion boards formatively because of the time and effort it takes to assess them. In many cases such assessments are often never completed even after courses have concluded. SNA offers a solution to this common challenge as the diagrams produced by this method can be used to incrementally ascertain the frequency and quality of interactions. Even more importantly, SNA can be conducted using existing data from discussion boards and there are a host of free tools like R, NodeXL, and Gephi to support the implementation of this method. Instructors who make SNA a part of their courses can display graphs that show which student(s) are most important to the discussion and which student(s) are lurking on the periphery. This can then serve as a motivating factor for students to focus on their online interactions and make sure they are more engaged in course activities. Using SNA to track student interactions has to potential to result in increased student engagement that will ultimately lead students to better futures and instructors to more robust teaching practices.

Purpose

The purpose of this study is to explore how SLAMs can serve as a means to analyze the social dynamic that supports knowledge construction in both formal and informal online discussions between collaborative groups.

Research questions

The research questions posed in this study are as follows:

1. How can SLAMs be used to assess the social dynamic that supports knowledge construction in formal online discussions?
2. How can SLAMs be used to assess the social dynamic that supports knowledge construction in an informal learning environment to determine how voluntary participation can lead to learning?

Social Learning Analytic Methods (SLAMs)

SLAMs refer to a host of analytics methods including but not limited to cluster analysis, sentiment analysis, and social network analysis.

Cluster Analysis

Cluster analysis is a method for determining the similarity between a collection of individuals based on a set of features (Romesburg, 2004). In the context of a discussion thread, cluster analysis groups postings in terms of similar word usage. It is useful for determining similarities between postings between different users.

Sentiment Analysis

Sentiment analysis is defined by Batrinca and Treleaven (2015) as “the application of natural language processing, computational linguistics and text analytics to identify and extract subjective information in source materials” (p. 90). Sentiment analysis is performed by creating a list of positive (good, wonderful, fantastic, etc.) and negative words (bad, awful, terrible, etc.), known as a lexicon, and using automated approaches to calculate sentiment orientation as either positive, negative, or neutral (Liu, 2012). Note that a lexicon is the most crucial resource for sentiment analysis because it establishes the words and statements that are used to determine sentiment orientation (Feldman, 2013).

Lexica Development

To conduct a sentiment analysis, data needs to be coded for positive and negative words. Coding is done automatically according to lists, called lexica, of positive and negative words. These words are then included in an automatic parser which produces scores for a given transcript that indicate whether the overall sentiment was positive or negative. The words included in lexica can expand this technique beyond positive and negative sentiment. For example, the lexicon used in Gunawardena et al. (2016) was created to assess social presence in an online course.

To assess social presence, instead of positive and negative words, words that contributed to or detracted from the creation of social presence were used. Words appropriate for each of these categories were identified by conducting a content analysis of the transcript. The resulting scores indicated how much social presence was being created in the course.

Social Network Analysis (SNA)

SNA can be defined as a method of identifying the relationships among social entities (e.g. dyads, triads, and larger groups) and analyzing the implications of interaction patterns (Wasserman & Faust, 1994). The key output of SNA is the sociogram which was first introduced in the early 1930s by Moreno (1953). The intent behind the sociogram is to visually illustrate the relationships between people mapping out the interactions between people or groups in a network.

In this study SNA will only be used to answer the first research question because the dataset provides the clearest example of how SNA can be used to track and enhance online interactions.

Method of analysis

SLAMs in a Formal Online Environment

In a formal online learning environment, the input for SLAMs is a table (see Figure 17) representing a discussion, which contains at least the names of the learners (e.g., *Name* column) and their postings (e.g., *Thread* column).

Count	Name	Date	Time	Thread	Initial	PhI/A	PhI/B	PhI/C	PhI/D	PhI/E	PhII/A	PhII/B	PhII/C	PhIII/A	PhIII/B	PhIII/C	PhIII/D	PhIII/E	PhIV/A	PhIV/B	PhIV/C	PhIV/D	PhIV/E	PhV/A	PhV/B	PhV/C
1	Jack	08/27/08	20:49:00	My own definition	JG	2																				
2	Jan	08/28/08	8:32:00	I agree that there	JL	2	2			1																
3	Marta	08/28/08	21:16:00	When I think of	MM	1			1																	
4	Leslie	08/29/08	9:02:00	I, too , am	LW	1	1						1													
5	Alex	08/29/08	12:34:00	Jack, These are	AJ		1					1														1
6	Alex	08/29/08	12:50:00	Hi Leslie, You	AJ																1					
7	Emily	08/29/08	16:53:00	Defining culture	EP	1	2														1				1	
8	Emily	08/29/08	17:16:00	Hello Marta, I	EP		1	1													1					
9	Emily	08/29/08	17:43:00	Hi Leslie, Its great	EP	1															1					
10	Alex	08/29/08	13:04:00	My personal	AJ	2					1	1								1	1					
11	Betty	08/29/08	22:22:00	What a devilishly	BS																					
12	Leslie	08/30/08	8:00:00	Hi Alex, I just	LW																					
13	Leslie	08/30/08	8:03:00	I agree, Emily, and	LW	1	1																			
14	Cassidy	08/30/08	19:00:00	Hello class- great	CJ	1	1	1																		
15	Jack	08/31/08	12:38:00	Hi Jan, I liked in	JG	1	1																			
16	Jack	08/31/08	12:45:00	Hi Cassidy, I	JG		2																			1
17	Jack	08/31/08	13:00:00	Hi Alex! Thank	JG														1					1		
18	Jack	08/31/08	13:13:00	Hi Marta, I liked	JG	1	2														1					
19	Tammy	08/31/08	14:43:00	I define Culture as	TK																					
20	Tammy	08/31/08	14:53:00	Jan, I love your	TK		1														1				1	
21	Kerry	09/01/08	7:23:00	I really enojoyed	KB																			1	2	1
22	Jan	09/01/08	9:29:00	Hi Kerry, Your	JL								1													1
23	Jan	09/01/08	9:43:00	Hi Jack, yes, I	JL		1	1	1																	
24	Jan	09/01/08	9:51:00	Hi Jack, ok, I see	JL								1								1					
25	Lex	09/01/08	10:44:00	I have read	LM	1																				
26	Kerry	09/01/08	12:48:00	Jan, I really didn't	KB				1																	1
26	Kerry	09/01/08	12:55:00	Lex, I really	KB		1														1					
28	Cassidy	09/02/08	12:33:00	Hi Jack, Thanks	CJ							1													1	1
29	Cassidy	09/02/08	12:48:00	Hi Kerry/Jan	CJ		1														1					
30	Cassidy	09/02/08	12:58:00	Hi Lex, Enjoyed	CJ																					1
31	Angela	09/02/08	15:58:00	After reading the	AN	1	3					1								1						

Figure 17. A sample datafile.

This file is exported in a .csv format, e.g., *IAM-RAW-Data.csv*, and read into R with code similar to *Table 2.* The code also collapses IAM subphases into the five major phases.

```
df=read.csv("IAM-RAW-DATA.csv", stringsAsFactors = F)
#
# Grab all the phase headings column numbers
#
phI.Headings=which(grepl("^phI\\.+",colnames(df),ignore.case=T))
phII.Headings=which(grepl("^phII\\.+",colnames(df),ignore.case=T))
phIII.Headings=which(grepl("^phIII\\.+",colnames(df),ignore.case=T))
phIV.Headings=which(grepl("^phIV\\.+",colnames(df),ignore.case=T))
phV.Headings=which(grepl("^phV\\.+",colnames(df),ignore.case=T))
#
# Collapse headings
#
phI=rowSums(df[,phI.Headings],na.rm=T)
phII=rowSums(df[,phII.Headings],na.rm=T)
phIII=rowSums(df[,phIII.Headings],na.rm=T)
phIV=rowSums(df[,phIV.Headings],na.rm=T)
phV=rowSums(df[,phV.Headings],na.rm=T)
#
# Create a valid time object
#
Date.Time=as.POSIXct(paste(df$Date,df$Time),format="%m/%d/%y %H:%M:%S")
#
# Create a Data Frame
#
df2=data.frame(df$Thread, Date.Time,df$Name,df$Thread,phI,phII,phIII,phIV,phV)
```

Table 2. R-code to import a table and perform data pre-processing.

Four analyses are typically run on the data: word-frequency, cluster, sentiment, and social network. We focus on last three types in this paper.

Cluster Analysis

Hierarchical clustering is performed using the default “Euclidean” distance algorithm, and plotted as a dendrogram. Table *3* shows the R-code for clustering and generating this dendrogram.

```
#
# Hierarchical Clustering: Dendrogram
#
library(tm)
thread=df$Thread
dtm=DocumentTermMatrix(Corpus(VectorSource(thread)))
mat=as.matrix(dtm)
d=dist(mat) # default is Euclidean distance
h=hclust(d) # perform hierarchical clustering
plot(h)     # plot dendrogram
```

Table 3. R-code to visualize hierarchical clustering as a dendrogram.

Sentiment Analysis

R contains built-in libraries for doing lexical-based sentiment analysis. However, this study used three custom lexica representing positive words, negative words, and social presence words. The R-code for importing these three different lexica, counting their occurrences in each posting, and generating a frequency bar graph, is shown in *Table 4.*

```
Positive.Lexicon=readLines("P-Words.csv")
Negative.Lexicon=readLines("N-Words.csv")
Presence.Lexicon=readLines("S-Words.csv")
thread=df$Thread
lexicon.scores=matrix(ncol=3,nrow=0)
for (post in thread) {
  positive.score=0
  for (phrase in Positive.Lexicon) {
    phrase=gsub("[*]", "\\\\*", phrase)
    phrase=gsub("[+]", "\\\\+", phrase)
    phrase=gsub("[-]", "\\\\-", phrase)
    phrase=paste("\\b",phrase,"\\b",sep="")
    if (grepl(phrase,post,ignore.case=T)) positive.score=positive.score+1
  }
  negative.score=0
  for (phrase in Negative.Lexicon) {
    phrase=gsub("[*]", "\\\\*", phrase)
    phrase=gsub("[+]", "\\\\+", phrase)
    phrase=gsub("[-]", "\\\\-", phrase)
    phrase=paste("\\b",phrase,"\\b",sep="")
    if (grepl(phrase,post,ignore.case=T)) negative.score=negative.score+1
  }
  presence.score=0
  for (phrase in Presence.Lexicon) {
    phrase=gsub("[*]", "\\\\*", phrase)
    phrase=gsub("[+]", "\\\\+", phrase)
    phrase=gsub("[-]", "\\\\-", phrase)
    phrase=paste("\\b",phrase,"\\b",sep="")
    if (grepl(phrase,post,ignore.case=T)) presence.score=presence.score+1
  }

lexicon.scores=rbind(lexicon.scores,c(positive.score,negative.score,presence.score))
}

dat=rbind(lexicon.scores[,"positive"],lexicon.scores[,"negative"],lexicon.scores[,"pr
esence"])
rownames(dat)=c("Positive", "Negative", "Presence")
barplot(dat,beside=T, names=1:length(df$Thread), legend=TRUE)
title("Positive, Negative, and Social Presence Scores")
```

Table 4. R-code to do a lexicon-based sentiment analysis. Three lexica are imported: Positive words, negative words, and social-presence words.

Social Network Analysis (SNA)

SNA was performed in two steps. Table *5* contains the R code for generating a table of social edges, where an edge represents who a speaker received information from.

```
names=tolower(df$Name)
thread=tolower(df$Thread)
# Extract first names
first.last=strsplit(names, " ")
first.name=sapply(first.last,function(item){tolower(unlist(item)[1])})
# Go through each name and find mentions of other names
# Edge becomes othername->name
#patt=""
#for (n in first.name) patt=paste(patt,"|","\\b",n,"\\b",sep="")
#patt=substr(patt,2,nchar(patt))

patt=paste("\\b",first.name,"\\b",sep="",collapse="|")
tbl=matrix(ncol=3,nrow=0)
for (row in 1:length(thread)) { # Go through each post
    t=thread[row]
    # find mentions
    mentions=unlist(regmatches(t,gregexpr(patt, t))) # Might want to make unique
    # create edges
    for (n in mentions) {
     label=paste(df2[row,which(grepl("ph",colnames(df2)))],collapse=",")
      tbl=rbind(tbl,c(n,first.name[row],label))
    }
}
colnames(tbl)=c("from","to", "label")
write.csv(tbl,"edges.csv",row.names=F)
```

Table 5. R-code to generate the social edges for a discussion.

Given a table of edges, *Table 6* is the R-code for depicting these edges as a sociogram, and annotating the edges with the IAM phases.

```
library(igraph)
edf=read.csv("edges.csv")
g=graph_from_data_frame(edf) # Read graph

in.degree=degree(g, mode="in") # Calculate degree
out.degree=degree(g, mode="out")
between=betweenness(g)

ug=as.undirected(simplify(g))
gr=cluster_fast_greedy(ug)
#gr=cluster_edge_betweenness(ug)

V(g)$size=strength(g) # plot
V(g)$color=groups$membership

set.seed(777)
plot(g,layout=layout_with_fr(g),vertex.label.cex=.75,edge.label.cex=.5,edge.arrow.siz
e=.5)
set.seed(777)
plot(gr,g,layout=layout_with_fr(g),vertex.label.cex=.75,edge.label.cex=.5,edge.arrow.
size=.5)
```

Table 6. R-code for generating sociograms.

SLAMs in an informal online environment

This portion of the study will illustrate how SLAMs apply to an informal learning environment using a #BlackLivesMatter Twitter dataset collected during the Freddie Gray demonstrations from April 21 – April 28, 2015. 45,646 tweets were collected during this timeframe.

OILS Twitter Scraper

Analyzing the informal learning environment begins by using the OILS Twitter Scraper to scrape unstructured tweets from Twitter from April 21-28, 2015 using the Twitter API. The OILS Twitter Scraper is an Excel-based tool written in Visual Basic by Flor (2014).

SCK Lexica Development

After scraping the data, the process moved to the development of lexica. Two common methods for creating lexica are a manual approach in which words are coded and added to the appropriate list by the researcher and a dictionary approach in which words are added according to their relationship to established words with positive or negative sentiment in the dataset as determined by an online dictionary (Liu, 2012). The manual approach involves a content analysis (Krippendorff, 1980) which is performed by manually reading text and classifying the text by hand according to a research framework. Content analysis was performed on all of the tweets from April 21st and 25th to establish a foundational lexicon. Coding was performed using an Excel spreadsheet with columns for all of the IAM phases and sub-phases like the one presented in *Figure 1.* The words identified in the manual approach enable the dictionary approach in which all words in the initial lexicon were found in the online dictionary WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). During the search, synonyms that did not fit the original context of a given word were removed while those that fit the context were added to the lexica. The final product of this process is a list of words and phrases according to each IAM phase that originate in the Twitter data used for this study.

Sentiment Analysis

The complete lists of words and phrases was added to routines in the OILS Twitter Scraper to automatically produce counts of the most frequently occurring words and phrases for each day in the dataset and to determine the presence of IAM phases in the data.

Results

Research Question 1

Cluster Analysis

Figure 18 depicts the dendrogram from the hierarchical clustering of the discussion thread. It is apparent that posting 25 is the most dissimilar. Similar postings include (10, 4, 31; leftmost cluster), (7, 8; rightmost cluster), and (13, 35; lowest cluster).

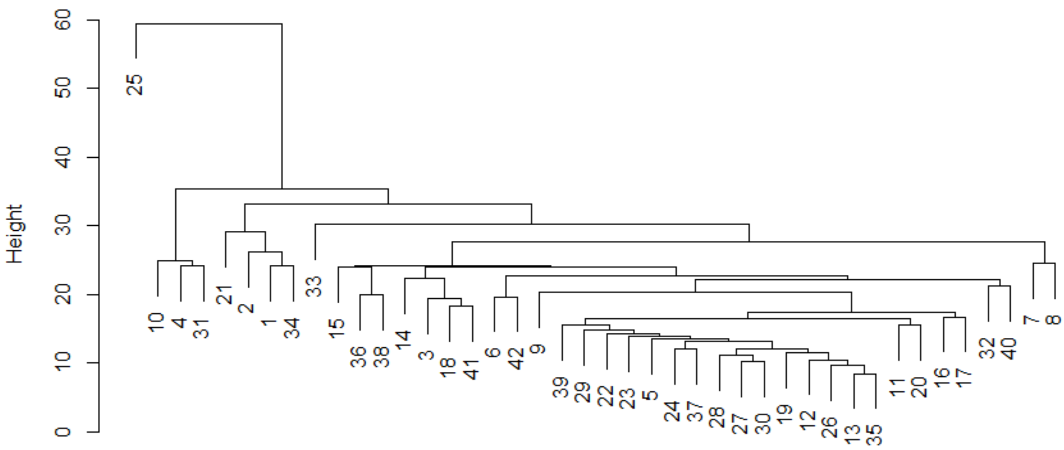


Figure 18

Figure 18. Dendrogram for the dataset.

Sentiment Analysis

Figure 19 depicts a bar chart for the sentiment and social presence analysis. High positive, negative, or presence scores indicate areas of focus for the analyst. The bar chart shows that posting 25, which was the most dissimilar in the clustering analysis, was also the posting with the highest positive, negative, and presence scores.

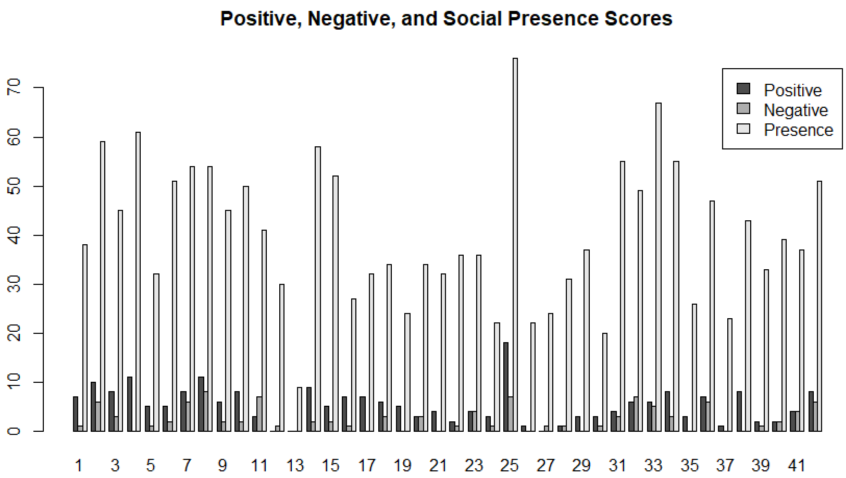


Figure 19. Sentiment analysis and social presence scores.

Social Network Analysis (SNA)

The SNA (see Figure 20) found three main cliques of users depicted by three differently-shaded clouds. Within the cliques, the most influential users (shown as bigger circles) appear to be associated with higher levels of IAM as denoted in the edge labels. By annotating the edge labels with IAM phases, the analyst can explore relationships between influence, interactions, and learning.

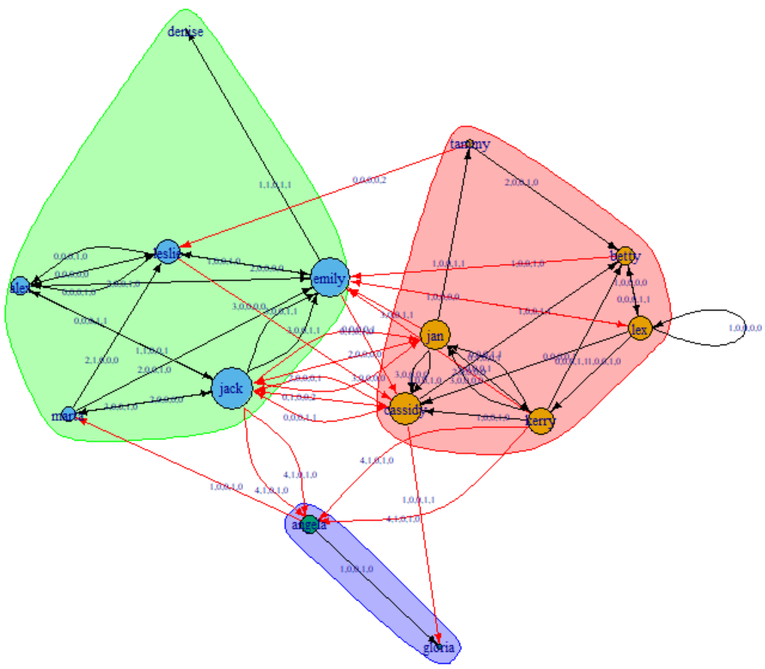


Figure 20. A sociogram depicting who received information from whom in the discussion. The edge labels represent the IAM phases associated with the information.

Research Question 2

Sentiment Analysis

The sentiment analysis results for the #Blacklivesmatter dataset are presented in *Figure* **21** organized by each day in the dataset.

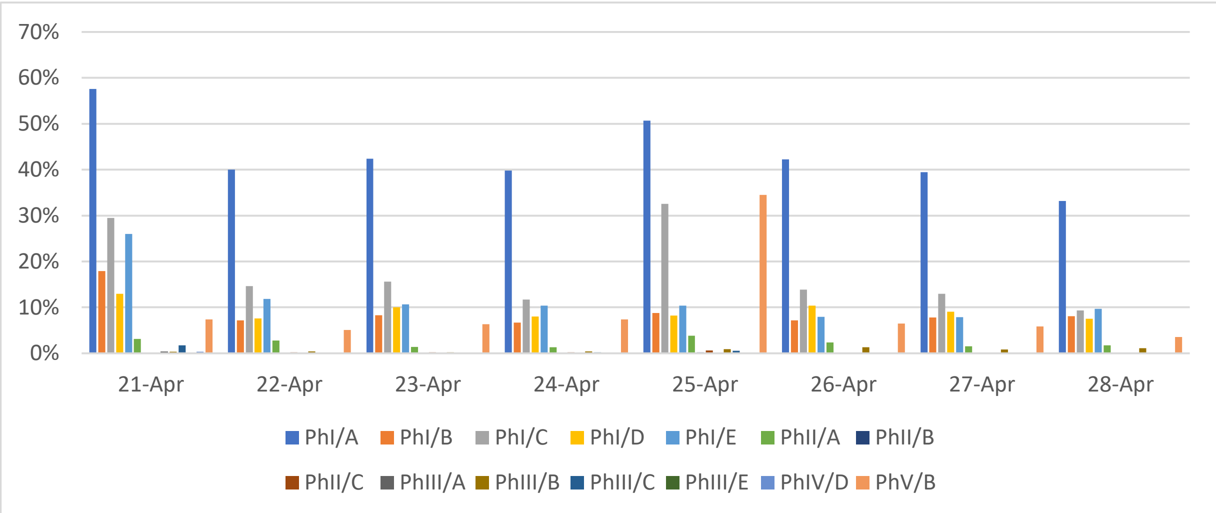


Figure 21. SCK Analysis Results.

The sentiment analysis counted the number of unigrams and bigrams that were exemplars of the given IAM phase in each sample. Note that each sample tweet can include more than one instance of a given phase or phases. The sentiment analysis shows that phase I/A was the most frequent of all the IAM phases. According to percentages, the proportion of phase I/A codes remains fairly consistent over the course of the 8-day dataset. Phase I/C and phase I/E were the most frequent following phase I/A. Excluding the content analysis days (21^{st} and 25^{th}), levels of phase I/C and phase I/E also illustrated a large degree of consistency ranging between 16% and 9% for phase I/C and 12% and 8% for phase I/E. Such consistency suggests that although the lexicon analysis does not account for all of the data, it consistently identifies SCK in this dataset.

Discussion

An important goal for researchers studying SCK — in either formal or informal online learning environments — is to understand the dynamics that support SCK. These dynamics include the changes in the phases of knowledge construction over time, and the social groups that form serendipitously during extended discourse. These changes are difficult or tedious to characterize manually, and SLAMs can help automate both the detection and visualization of these social dynamics. In the first study, cluster analysis, sentiment analysis, and social network analysis were performed on a discussion board thread. The cluster analysis showed postings which were similar based on word choice. The sentiment analysis portrayed changes in attitude towards the discussion topic, and changes in social presence. Finally, the social network analysis depicted the exchange of information, the implicit groups that formed based on this exchange, and the phases of knowledge construction associated with the exchanges.

One can combine the findings from these three analyses to form hypotheses for further analysis, or to design instructional interventions prior to the end of a course. An example of a hypothesis based on these analyses would be: *postings that are different* (e.g., posting 25 in Figure 18), *with high positive sentiment and high social presences scores* (e.g., posting 25 by subject Lex in Figure 19) *can help others reach higher phases of knowledge construction* (e.g., subjects Betty & Kerry, rightmost group in Figure 20). An example of an instructional intervention based on these analyses would be to place students in three groups, as indicated by the sociogram in Figure 20, but to put the top influencers in different groups.

The utility of SLAMs is particularly evident when studying SCK and social dynamics in massive online groups — where traditional content analyses would be prohibitively expensive in both time and money. The second study demonstrated that SCK can be found in the #BlackLivesMatter network of practice. Sentiment analysis was able to identify SCK in the dataset at varying levels. Overall, the IAM framework identified that SCK occurred mostly in phase I as many examples of sharing opinions and providing examples were found. Interestingly, the rates at which data were coded by the SCK lexicon were fairly consistent suggesting that as people voluntarily contributed to the discussion, they continued common threads and described them using similar language. The SCK lexicon performed well for specific topics that were present on both of the content analysis days.

Many of the previous studies conducted using the IAM have failed to identify higher phases of knowledge construction (Paulus, 2007). For the most part, this study is no exception because the majority of the samples were either in phase 1/A or phase 1/C indicating that higher phases of knowledge construction did not take place. One potential explanation for this is provided by (Gunawardena, Lowe, & Anderson, 1998, August) who write that dissonance is not always needed to build knowledge because the people coming into a given discussion realize they are on the same page conceptually speaking and therefore accept statements made by others instead of disagreeing and sparking productive conflict.

Conclusion

Traditional content analysis can be prohibitively time consuming when applied to massive amounts of online data, or to data rich in higher phases of knowledge construction. SLAMs provide researchers a host of semi-automated techniques for making explicit many of the tacit structures and processes that underlie the social construction of knowledge.

References

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *AI & Society: Journal of Knowledge, Culture and Communication, 30*(1), 89-116. doi: 10.1007/s00146-014-0549-4

Dron, J., & Anderson, T. (2014). Teaching crowds: Learning and social media. Issues in distance education series. Edmonton, Alberta: AU Press. Retrieved from desLibris e-book: http://www.deslibris.ca/ID/448837.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82.

Flor, N. V. (2014). OILS Twitter Scraper. Albuquerque, NM: Creative Commons Attribution-ShareAlike 4.0 International License.

Gunawardena, C. N., Flor, N. V., Gómez, D., & Sánchez, D. (2016). Analyzing social construction of knowledge online by employing interaction analysis, learning analytics, and social network analysis. *Quarterly Review of Distance Education, 17(3)*, 35-60.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4), 397-431.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1998, August). *Transcript analysis of computer-mediated conferences as a tool for testing constructivist and social-constructivist learning theories.* Paper presented at the Distance Learning '98: Proceedings of the 14th Annual Conference on Distance Teaching & Learning, Madison, WI.

Henri, F. (1992). Computer conferencing and content analysis. In A. Kaye (Ed.), *Collaborative learning through computer conferencing: The Najaden papers* (pp. 117-136). Berlin: Springer-Verlag.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

Liu, B. (2012). *Sentiment analysis and opinion mining*. San Rafael, CA: Morgan & Claypool.

Lucas, M., & Moreira, A. (2015). A visual representation of online interaction patterns. *Journal of Universal Computer Science, 21*(11), 1496-1507.

Merriam, S. B., & Caffarella, R. S. (1999). *Learning in adulthood a comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass Publishers.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography, 3*(4), 235-244.

Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama* (2nd ed.). Beacon, NY: Beacon House, Inc.

Paulus, T. M. (2007). CMC Modes for Learning Tasks at a Distance. *Journal of Computer-Mediated Communication, 12*(4), 1322-1345.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). New York, NY: Cambridge University Press.

Romesburg, C. (2004). *Cluster analysis for researchers*. North Carolina: Lulu Press.

Twitter. (2016). Twitter Glossary. Retrieved March 4, 2016, from https://support.twitter.com/articles/166337

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

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