Social Media Analysis using Twitter for Higher Education Initiatives

Introduction

Coursolve connects courses with organizations to empower students to solve real-world problems. In other words, we match nonprofits, social enterprises, small businesses, and other organizations with relevant academic courses so that students' work in course projects can meet these organizations' needs [1].

In this regard, Coursolve is interested in implementing Social Media Analysis to

1) identify the topics that drive higher education
2) identify thought leaders and the topics they discuss on social media
3) disseminate relevant content [1] to organizations and students using the analysis
4) promote brand visibility [1]

In this project, I chose to analyze Twitter feeds that would attempt to answer the above-mentioned points. Based on the available data, my analysis indicates some of the terms such as learning, ed, tech, and education as being highly relevant. I also identified some of the top tweeters who are tweeting on these topics.

Methods

Data Collection

The project was done after discussing with a group of fellow students. First, we identified the key topics involved. The following is a list that we have currently used as a starting point for analysis. Several more could be added in the future.

"MOOC", "Problem-solving", "Anant Agarwal", "Blended learning", "Collaborative learning", "Crowd-sourcing", "Distance learning", "E-learning", "Online education", "Online learning", "Workforce development", "Higher Education Access", "Higher Education Inequality", "University", "Higher Education Affordability", "Community college", "Ed-tech", "K-12 education", "Mobile learning", "Khan Academy".

For the analysis, I used Twitter data that was collected by a fellow student from his Github project [2]. He wrote a program to collect Twitter feeds using the Twitter search API for all the relevant topics. The data collected was very limited since Twitter had changed its terms and conditions by applying rate limiting [8]

Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the observed data. Specifically, I performed visualization using Tableau Public [3]. Exploratory analysis was used to

1) identify top Tweeters and topics
2) verify the quality of the data, and
3) determine the terms used in the Tweets involved.

For the analysis, I created CSV files from the supplied raw files [2] that contained the top 100 Tweet topics, the top 100 Tweeting users, the top 100 Retweeted Tweets, and the top 100 mentions.

I had to do data munging and cleaning for every single CSV file, as there were several Unicode characters in the Tweet data that had not been correctly resolved. This resulted in some of the tweets being discarded. Since the raw files were not in any specific format, I manually created JSON files first using an online JSON editor [4] I then used a JSON to CSV utility to create CSV files. [5] I could have written a Python program but found these tools to be easy and useful.

I was also given data that was pulled from DataSift by Coursolve [6]. But I was unable to draw sufficient insights as I did not have sufficient time to understand and analyze all the data.

Finally, I uploaded all of them to the Tableau software, and constructed 5 separate charts and 2 dashboards with the data. Tableau Public provides a simple and clear way to construct graphs. The results were then uploaded to Tableau that generated a public website URL [7]. I then shared this URL with the group and discussed my findings with them.

Findings

1) The dashboard below indicates the top tweet topics and the top tweeting users.

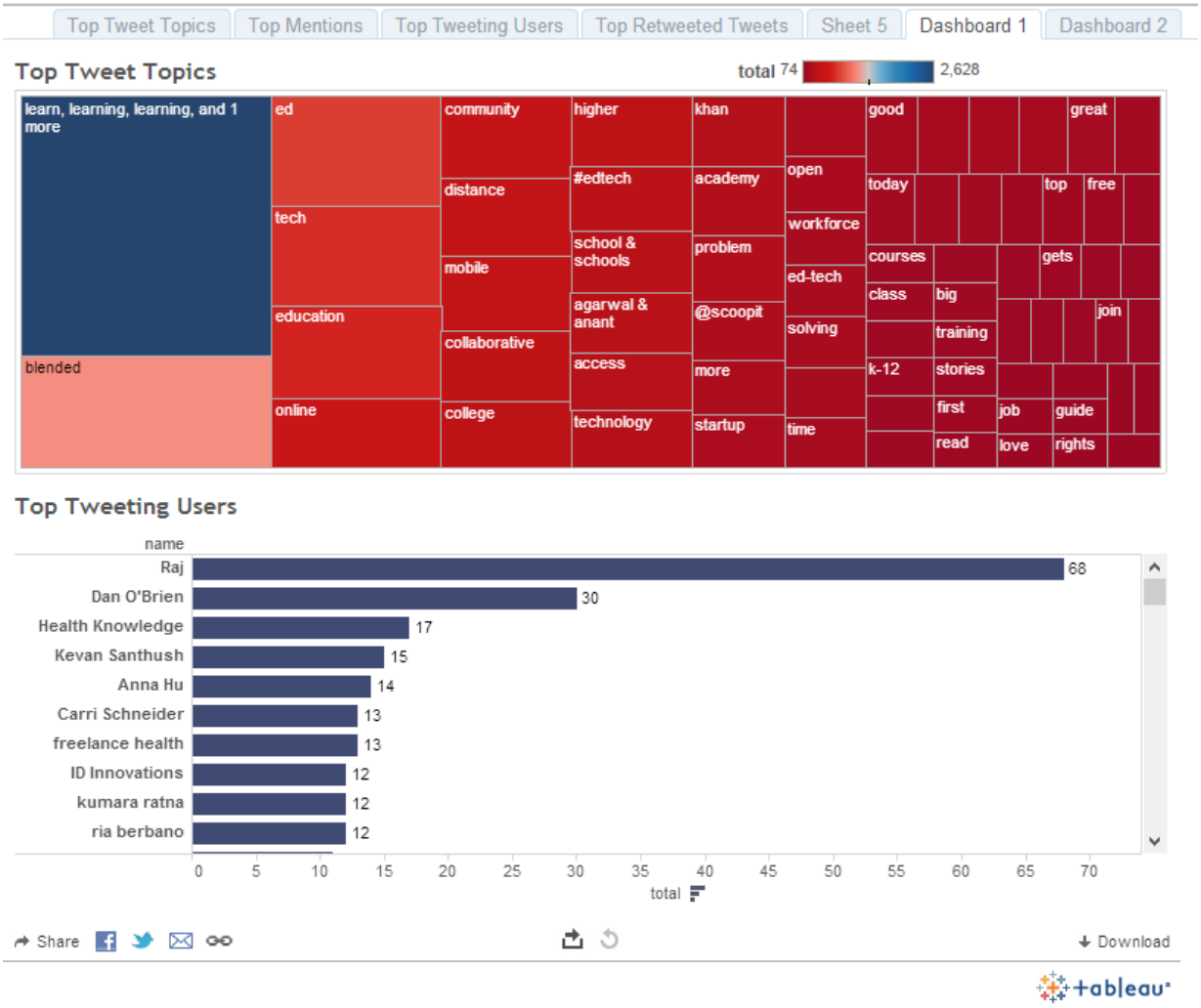

Fig 1. Dashboard showing top tweet topics and top tweeting users

From the first chart, we see that the terms learning, ed, tech, education are very relevant and tweeted the most. There are also several words like blended that do not really correspond to what we are looking for.

The Top tweeting users chart shows 68 tweets by a tweeter named Raj. All the tweeters indicated have tweeted about higher education topics using the terms in the first chart.

2) The dashboard below indicates the top retweeted tweets and the top mentions.

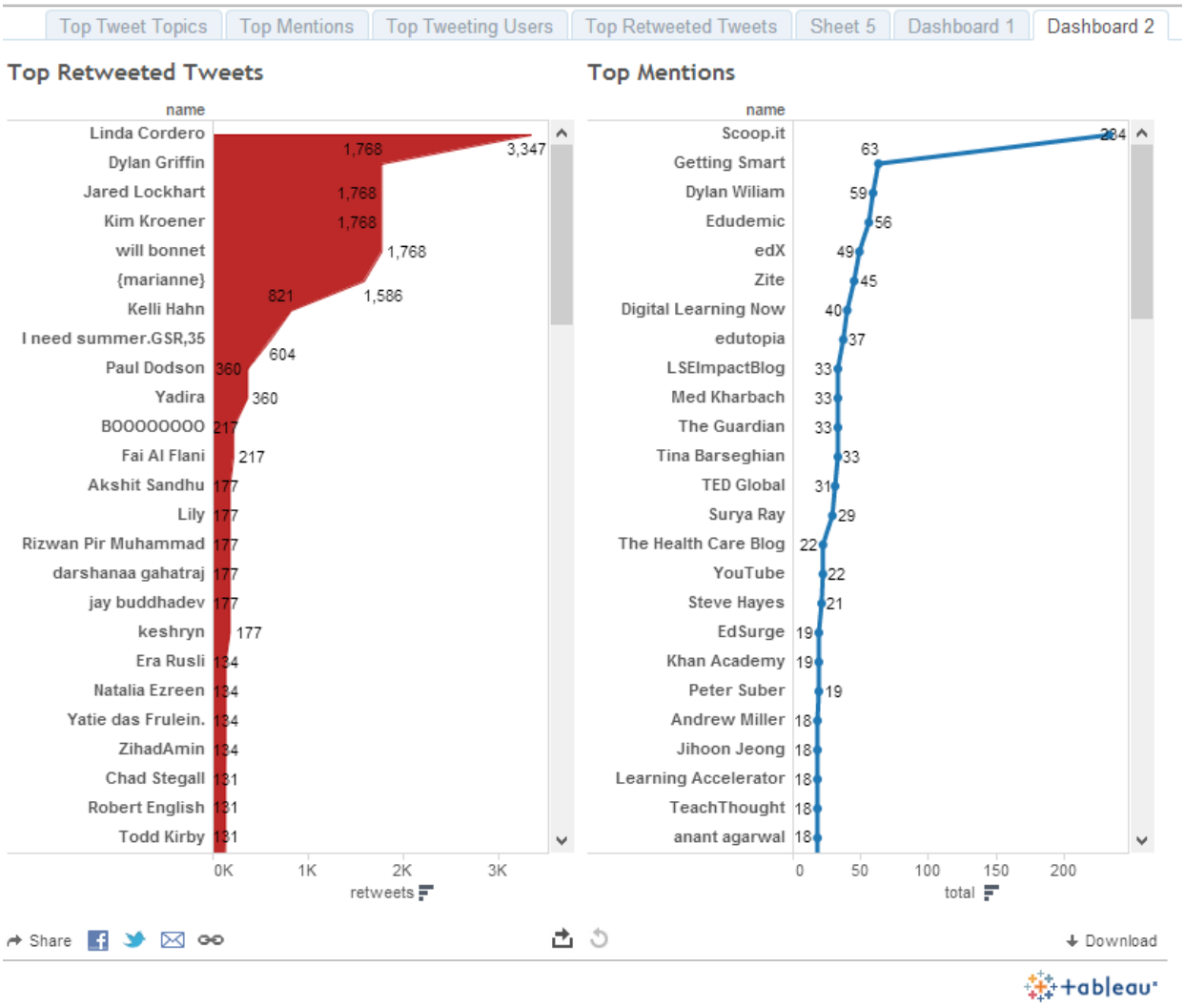

Fig 2. Dashboard showing top retweeted tweets and top mentions

The Top retweeted tweets indicate Linda Cordera with 3,347 tweets. And the top mentions include names such as Scoop.it and Getting Smart with 234 and 63 tweets respectively.

3) I tried to look at the DataSift data, and constructed a simple graph showing the sentiment scores of education-related tweets from around the world.

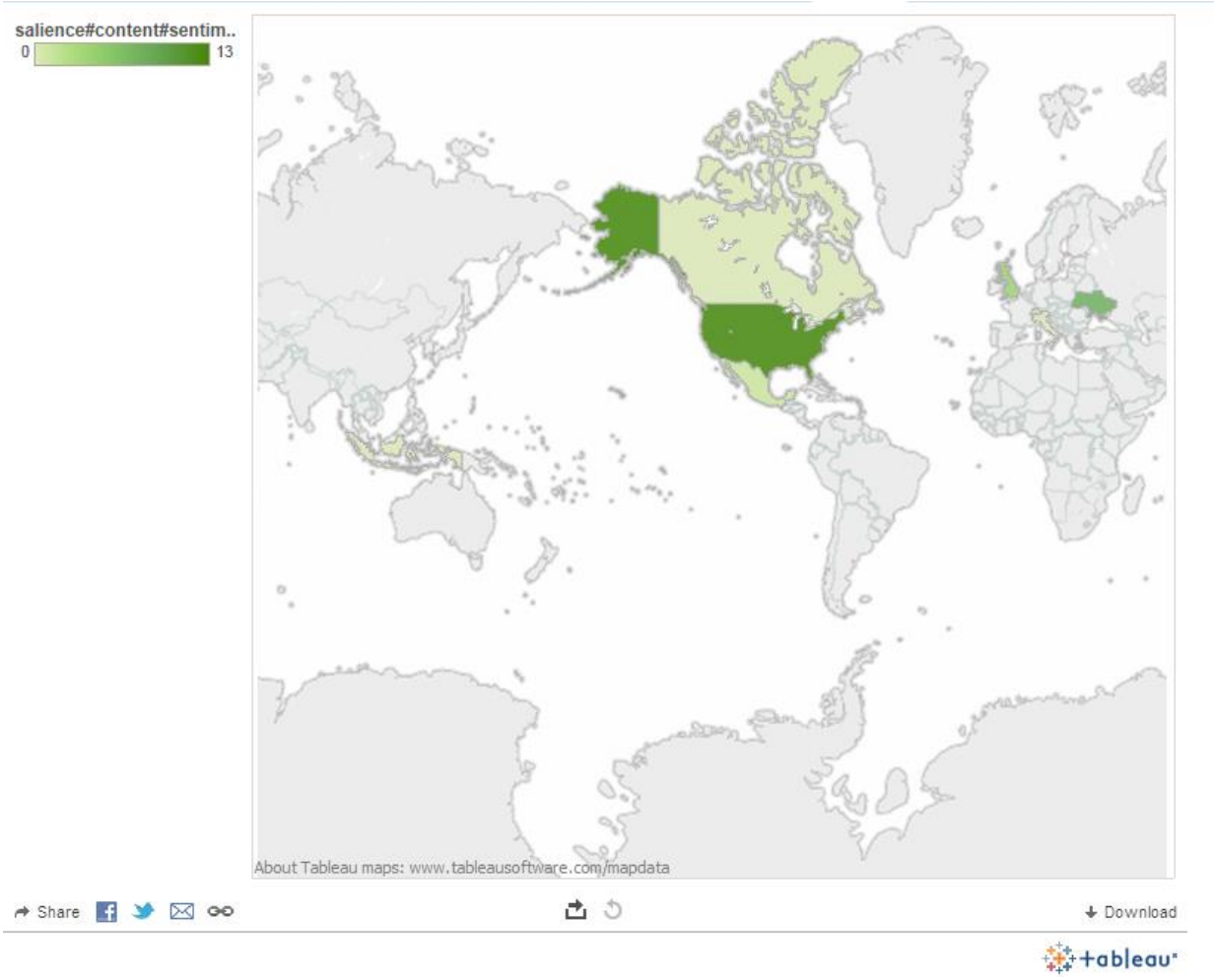

Fig 3. Sentiment scores of tweeters around the world

From the chart, it is clear that most of these tweets are from the USA.

Conclusions and Future Work

My findings suggest that the data collected was insufficient to identify correlations between the tweeters and higher education topics. While this analysis is an interesting first step, it is still based on a limited sample of tweets from the data available for the project. A larger collection of representative data may be more appropriate for understanding the relationship between higher education and social media. The analysis may involve other complex variables that have not been considered in this project.

Future work would involve setting up a system that regularly collects and unifies the data from as many sources as possible for better analysis.

References

1. Coursolve Project - Social media analysis for higher education initiative. URL: https://github.com/sgsangam/GitRep Accessed 6/30/2013.

2. Github profile of S G Sangameshwara. URL: https://github.com/sgsangam/GitRep. Accessed 6/28/2013

3. Tableau Public. URL: http://www.tableausoftware.com/public. Accessed 6/29/2013

4. Online JSON Editor. URL: http://jsoneditoronline.org/. Accessed 6/29/2013

5. JSON to CSV utility. URL: http://www.danmandle.com/blog/json-to-csv-conversion-utility/. Accessed 6/29/2013

6. DataSift data given by Coursolve. URL: https://groups.google.com/forum/#!topic/coursolvedata/tVeUnFpDN7g. Accessed 6/29/2013

7. Tableau Visualization done by Kannan Sankaran. URL: http://public.tableausoftware.com/views/Coursolve-Jun30_13/Dashboard1?:embed=y&:display_count=no. Accessed 6/30/2013

8. Twitter API Rate Limiting. URL: https://dev.twitter.com/docs/rate-limiting/1.1/. Accessed 6/29/2013