“Exploring the MOOC community on YouTube”

Project Description
Over the recent years, social networks have taken a totally different meaning. With disruptive innovation leading to new, wide and open markets, existing technology is fast being replaced with new systems \[1\]. Crowdsourcing, crowdfunding are some of the disruptive technologies that have made use of the huge networks that are available to us. This project is about one more disruptive technology -- Massive Open Online Courses (MOOC). In the recent years Massive Open Online Courses (MOOC) has gained much importance and I would like to explore the YouTube community for MOOC. By exploring the video tag MOOC on Youtube, I would like to find answers to some of the questions:

Who are the key players in the MOOC community? Are all countries adapting to this new way of learning? What are the different types of categories available in this domain? Are businesses equally likely to adapt this or is this system good for universities only?

Data
The data was imported from the Youtube video network searching for keywords whose title, keywords, description, category or authors name contains “MOOC”. NodeXL would add an edge for each pair of videos having the same category limiting the size of the data to 100.

The data was prepared by counting and merging duplicate edges and adding a new edge weight column to it. The data was as follows:
- 100 vertices
- 2425 unique edges
- 6 connected components

Other attributes that were included with the data set were:
- Title
- Author
- Created Date
- Ratings
- Views
- # of comments
- links to the videos,

The vertices were connected to each other based on the shared categories and the shared categories were identified as:
- film
- entertainment
- people
- tech
- nonprofit
- news
The Graph Metrics were then calculated and the raw data looked as follows:

Figure 1: Raw data with Harel-Koren Fast Multiscale Layout.

The graph in Figure 1 displayed a cluster of networks in between along with some around it. Those vertices without any edges or one or two edges were also displayed.
Headlines

#1: Dave Cormier\footnote{2} (who coined the word ‘MOOC’ in 2008) is among the persons who has maximum views of his MOOC video

Figure 2: Each of the five groups have the name of the authors who have views above 150,000. Dave Cormier, in the Education group is among the Authors whose YouTube video has the maximum views.

Figure 3: Author names have been replaced to show the vertices as the title of the video in this figure. The encircled title is Dave Cormier’s video titled: “What is MOOC?”

Figure 2 was created as follows:
Graph Type: Directed
Groups: Grouped by cluster using the Clauset-Newman-Moore cluster algorithm
Layout: Circle Layout
With a sudden rise in the number of free online courses on the internet, Dave Cormeir’s video “What is MOOC” is still a popular Youtube video with more than 150,000 views from people all over the world. Dave went on to create more videos offering courses online but his video “What is MOOC” still remained a popular one. The course offered by Dave Cormeir along with his colleagues later gave rise to other MOOCs.\cite{3} \cite{4} \cite{5} From the above Figure, we also infer that more titles that spoke about what MOOC is, were more popular than others.

**#2: The French are way ahead in MOOC interactions than rest of the world**

![Figure 4](image)

**Figure 4:** This figure has the vertices name as the title of the video. The figure shows a growing number of titles in French along with a growing French commenter community.

Figure 4 was created as follows:
- **Graph Type:** Directed
- **Groups:** Grouped by cluster using the Clauset-Newman-Moore cluster algorithm
- **Layout:** Circle Layout
- **Edge Width:** Based on Edge Weight
- **Edge Opacity:** Based on Edge Weight
- **Edge Label:** Shared Commenters
- **Vertex Color:** Degree
- **Vertex Size:** Degree
- **Vertex Label:** Title
Eventhough MOOC was coined and introduced in the United States in 2008, it has seen a sudden acceptance in the European community especially the French community. The French Government has also announced the creation of its first public alternative to existing solutions and the French business schools have also opened up their own MOOCs. Figure 3 shows an edge going from French MOOC communities and the thickness shows relationship which in this case is the number of shared commenters. Also, Figure 4 has vertex names as the titles of the videos and most of the education group has titles in French.

**#3 MOOC is the future of higher education**

Figure 5: This figure has been filtered by date starting December 2010 to 11th November 2012

Figure 6: This figure has been filtered by data starting from exactly last year (11th November 2012 up till 4th November 2013)
Figure 5 was created as follows:
Graph Type: Directed
Groups: Grouped by cluster using the Clauset-Newman-Moore cluster algorithm
Layout: Circle Layout
Edge Width: Based on Edge Weight
Edge Opacity: Based on Edge Weight
Edge Label: Shared Commenters
Vertex Color: Degree
Vertex Size: Degree
Vertex Label: Authors
Vertex Tooltip: Clustering Coefficient

Figure 5 and 6 show a direct comparison of the number of videos, their categories, people/businesses/organizations/universities associated with them on Youtube. Within the last year, the number has exponentially increased and there have been various types of interactions between different groups, most of them being between education and people categories. Figure 6 below also shows how the number of videos have increased with time with the Y-axis showing is the number of views.

Figure 7: The Y-axis is the time and the X-axis is the number of views of the video. Early in time, Dave Cormier’s video had received the maximum number of views and it still remains the same. Later on, the number of videos have increased drastically with a sudden increase in business school MOOCs.
NodeXL Critique:

Pros:
- NodeXL provides a variety of ways to play around with the data using the AutoFill Columns. It has many options which are very useful in analyzing the data.
- The import feature is by far the best one which allows importing from Flickr, YouTube, Facebook, Wiki, Twitter, etc. It also is pretty fast at extracting at least 100 vertices at a time.
- The various types of layouts are very rich and useful along with the layout options which also allow you to specify the distance between vertices by specifying it in the strength of the repulsive force between vertices.

Suggestions for Improvement:
- Group By different cluster algorithms allow NodeXL users to exploit the data in various ways to look for interesting findings and clusters. However, there’s no freedom in trying to modify or add your own cluster algorithm and hence there should be a way to do that.
- An UNDO option is needed as going back and forth to the AutoFill Columns is too tedious to make changes or revert back to changes.
- The data extracted by YouTube outputs the vertices as alphanumeric codes and they aren’t any helpful while trying to tweak the data in excel
- The graph window sometimes can get annoying when you are trying to look at the excel data or want to do something else like change the autofill options. Its gets in the way and can be annoying to keep moving it out of the way to get access to other data
- NodeXL window freezes if data entries are too large.

References: