

A Better Way to Recommend Educational Content

Tiffany Hsu, Kevin Jung
CS224W - Analysis of Networks - Dec 2017

1, Introduction

For generations, humans have looked outwards at others for inspiration, individuals of great intellect and accomplishment have been regarded as heroes and model citizens. However, this kind of admiration has not always been shared equally amongst the diverse set of people in the world. As such, with growing awareness, one of the greatest questions of the 21st century is the question of bias, such as gender equality. By investigating user ratings of speakers in TED talks, a popular lecture series that are widely shared for approachable discussions from various experts, we aim to explore whether even experts are affected by bias and whether a platform primarily served to distribute educational content is subject to potential unconscious biases as well. We also aim to explore whether certain types of content are systematically more recommended than others.

One form of bias that we are particularly interested in is bias in gender perceptions. Previous research in social psychology have found that in various realms and skills, men and women are perceived differently despite lack of difference in actual skill level - for example, among STEM and non-STEM students, it was found that there is a significant pro-Male-STEM bias (Farrell et al., 2016). Even among evaluations of teachers, favoritism towards male professors were found (Boring 2016).

It is well documented that recommendation systems are largely correlated with viewership (Zhou et al., 2016). We believe that one of the most important goals of education is to promote a diversity of thinking. By investigating the network

structure created by TED talk recommendation lists, we explored how the content is being curated and what kind of content is promoted to users. As stated in their organization page, the goal of TED is to promote a global community, welcoming people from every discipline and culture who seek a deeper understanding of the world. As such, we believe it is important that the recommendation network of TED and similar educational services curate recommendations that promote views towards a diverse set of ideas.

Other than gender bias, we hope that by using purely data-driven community-detection and centrality analyses, we could elucidate the features behind the viewership network and identify the key videos that may be engagement points for TED.com to look at whether talks with certain attributes are systematically emphasized or ignored. From here, we will be able to study the metrics involved that correlate amongst influential nodes with hopes to provide metrics that predict future key videos. Continuing, we hope that insight in differences between influential and non-influential nodes could inform more conscious efforts in distributing educational content.

2. Related Work

Paper 1: Boosting video popularity through keyword suggestion and recommendation systems (Zhou et al, 2016)

Viewership has become one of the most important metrics in today's internet, especially in distribution channels such as

YouTube. While the first content consumed is generally linked from a specified search or external link, the subsequent propagated content consumption is led by the recommendations, commonly found on many content aggregation websites to the side or below the main content.

By examining the videos' metadata, recommendations lists, and view statistics, the researchers were able to determine the relationship between various video contents and the view propagation in YouTube. The researchers found that more than 80% of propagated views were contributed by the top referrers, 68% directly from a referring video while 15% were indirect, meaning more than one video removed. This demonstrates the importance of recommendation lists as the key component in users accessing content when navigating YouTube. For our application, we see a parallel to the recommendation lists found in TED.com. The implication was that if the larger the number of views the referring video has, the larger the benefit to a video.

One of the challenges in working with our dataset compared to the one used in Zhou et al. is the difference in the available data. Much richer metadata is available for YouTube to extract the types of relationships between the many videos. For example, it was possible to extract the source of a view of a video, thus being able to identify the top paths used to reach the content. While we cannot directly derive this information with TED.com, we may conjecture that the recommended list on TED.com behaves in a similar manner.

The researchers also found that the top referrer videos remained constant over a 4-month period for 96% of videos.

This all demonstrates the importance of recommendation lists for increasing viewership in online content. They are hugely influential in the trajectory of viewership for a given video and more than

any other source, on average affect content engagement the most. Thus it is important to create a robust recommendation system so that meaningful content could be suggested to viewers and engagement stays high.

In order to come up with an improved model for suggestions, the researchers used clustering techniques to match videos by keywords. Then, using the pool of keywords from the clusters, a ranking algorithm was developed to evaluate the most relevant and influential keywords. High ranking keywords were ones associated with popular videos. By comparing two series of videos, uploaded at the same time, one with no associated keyword tags, and another with them, the researchers found a near 10x improvement in viewership when the keyword tags were used.

Paper 2: Fast Algorithm for Modularity-Based Graph Clustering (Shiokawa et al, 2013)

One of the key ways to characterize a network is to identify the clusters. As the number of active users grow in the web, especially through the spread of social media, a robust modularity-based graph clustering algorithm needs to be identified.

The Newman and Girvan algorithm (Newman and Girvan 2004) has been popular for extracting clusters from graphs, but increased sizes of networks have shown its limitations in speed. An alternative, the Clauset, Newman, and Moore algorithm (CNM; Clauset, Newman, and Moore 2004) is popular for its speed, but previous works have found it to have a tendency to find super-clusters. BGLL is an alternative algorithm that is considered state-of-the-art for its reliability and speed. And lastly, the researchers of this investigation proposed a new algorithm that improves upon BGLL.

For this project, we will be using CNM as it is built into SNAP.py. Do to our graph's

characteristics with only 2550 nodes, there is little benefit to implementing a more complicated algorithm to see improvements. The built-in CNM function gave us insights into the graph structure in an efficient manner. The Girvan-Newman algorithm did not produce clusters in a reasonable time frame.

3. Method

Dataset

We used the “TED Talks” dataset from kaggle.com, an aggregate platform for data-driven analytics competitions that contains a wealth of data in various domains. The dataset contains rich information from TED.com on all video and audio recordings of TED talks uploaded on the website until September 21st, 2017. This dataset includes information including speaker name, title of discussion, number of views, number of comments, related videos, and user ratings. The ratings data is especially insightful as it is subdivided into 14 distinct categories, 9 positive categories - *beautiful, inspiring, ingenious, fascinating, jaw-dropping, courageous, fascinating, funny, informative, persuasive*, and 5 negative categories - *obnoxious, confusing, long winded, OK, and unconvincing*. The dataset also contains a list of “related talks” to watch for each TED talk according to the recommended talk links on the actual website for each talk.

Analysis Methods

All network and numerical analyses were performed in Python using SNAP and other packages.

We explored two central questions:

1. How is the network structured by TED.com's current recommendation algorithm?

2. Is there a better way to curate the recommendations to better diversity viewership and exposure to different types of content?

In order to ask the first question, we assigned each talk to a unique identifier. These identifiers were used as nodes of a directed graph and connected every talk with a directed edge to each talk on its recommended list; thus, we have a recommendation network of talks linked by recommendation paths. We also classified all speakers by their genders (0=man, 1=woman, 2=other [groups, animals, etc]) as this information was not provided in the initial dataset. We also deleted self-edges in the graph, as the self-edges were not reflected on TED.com (the same video is not listed in the video's recommended list).

To answer the first question, we performed in- and out-degree analyses on the network, specifically looking at degree distribution to see if a small number of nodes are receiving a large number of recommendations (and thus potentially gaining for exposure). A preliminary analysis of the distribution included checking averages, standard deviations, maxima and minima. Analyses included comparing the amount of recommendation received by each gender on a group level.

To answer the second question, we used the CNM algorithm to divide the network into communities in order to assess the characteristics of the groups of talks (i.e. the communities) that viewers might be stuck watching if purely viewing through the recommendation lists. We then used the “tags” information provided from the dataset to look at the topics associated with each community. The tags are microtopics that each talk is associated with; each talk is associated with a subset of a total 416 tags. Because of the large number of potential

tags, we focused on the top 15 tags associated with each community.

We then analyzed the characteristics of the communities in order to assess potential discrepancies in the recommendation network. If communities are defined by unsurprising features such as topic and speaker occupation, we can confirm that the recommending algorithm is indeed pulling related talks somewhat unbiasedly. However, this could also mean that a typical viewer, if browsing the site, might be exposed to a lack of diversity in content.

4. Results

A total of 2550 talks were in the dataset, 1682 of which were by men, 791 by women, and 77 by groups or non-humans (i.e. a parrot).

Likely due to the design of the website which caps the number of recommended videos to 6 and the recommendation algorithm implemented by TED.com, the distribution of the number of recommended talks for each talk was relatively tight (range 1-6, mean 5.89, S.D. 0.62). However, the

number of times each talk was included as a recommended video (in-degree) differed largely (range 0-38, mean 5.89, S.D. 4.30). This provides the variability for us to investigate potential disparate and disproportionate patterns that determine which videos are recommended more.

Preliminary Results

Structure of the recommendation network. Log-log plot of the in-degree distribution (Figure 1) reveals a plausibly power-law distribution, with the majority of nodes having small in-degrees and few nodes with large in-degrees. This pattern shows that the few large in-degree nodes could potentially have a lot of power in directing viewership.

There was a significant positive correlation between in-degrees and number of views ($r = 0.15$, $p = 1.58 \times 10^{-14}$), indicating that as expected, talks that were recommended more were also viewed more. This confirmation of increased viewership further emphasizes the importance of detailing the characteristics behind talks that were recommended or not recom-

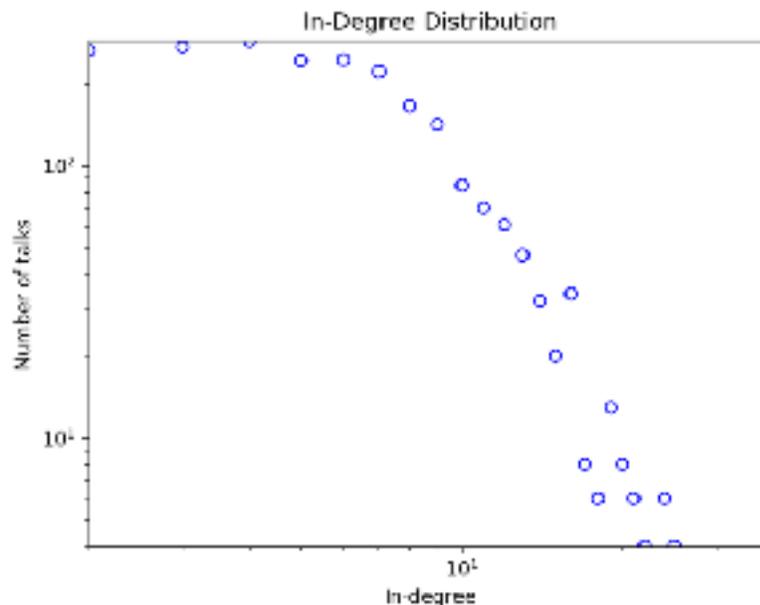


Figure 1. Log-log plot of in-degree distribution

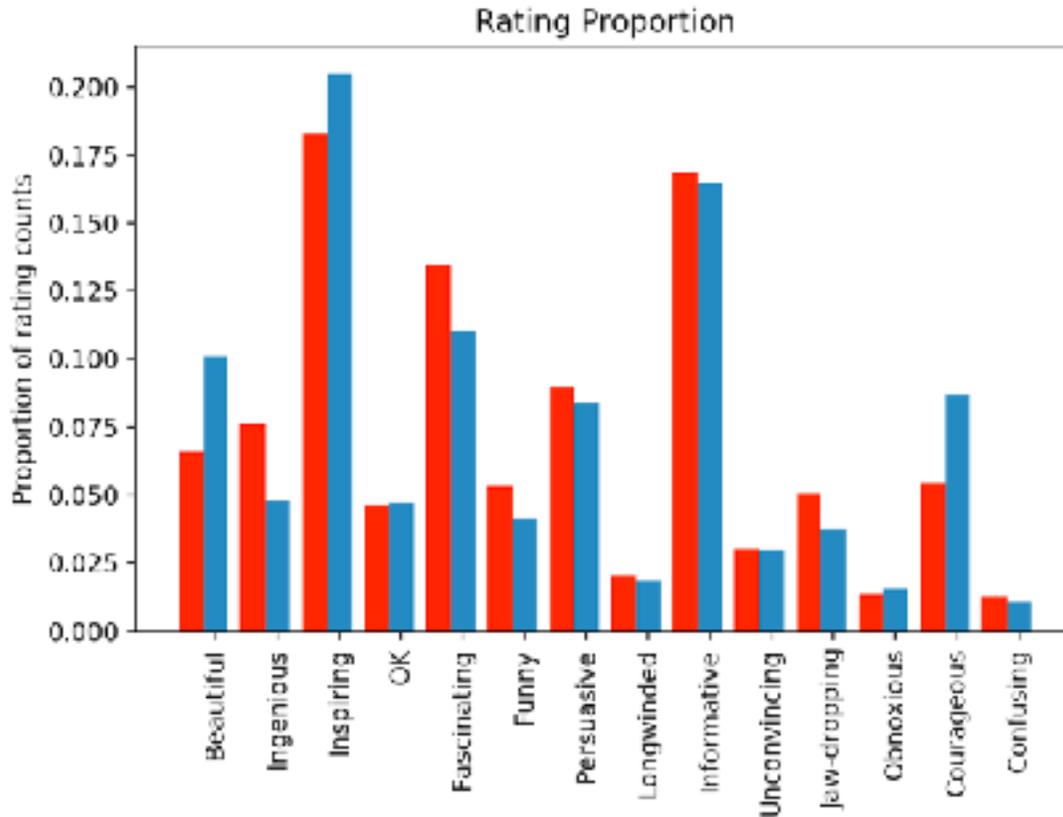


Figure 2. Proportion of counts attributed to each rating category for each gender (men = red, women = blue)

mended. However, since the correlation, though significant, was not nearly close to 1, clearly there is some variance in viewership unexplained by level of recommendation; this provided us with motivation to investigate the the mediators of this association.

Gender Descriptives. To begin exploring potential gender discrepancies, we found no statistical gender difference in in-degrees (men mean = 5.85; women mean = 6.08; two-tailed two-sample t-test $t = -1.25$, $p = 0.21$); that is, talks by men and women did not seem to be recommended disproportionately.

However, talks by men and women received different rating patterns (Figure 2). Specifically, talks by men were rated as

significantly more ingenious ($t = 8.33$, 1.34×10^{-16}), funny ($t = 2.67$, $p = 0.0076$), and jaw-dropping ($t = 2.29$, $p = 0.022$) than talks by women. In contrast, talks by women seemed to be rated as significantly more beautiful ($t = 3.17$, $p = 0.0016$) and courageous ($t = -4.59$, 5.12×10^{-6}) than talks by men. These patterns could be at least partially due to differences in topics covered by men and women, but prompted further analyses to be presented later in the paper through topical and community analyses.

Recommendation propagation by gender.

Even though there was no significant gender difference in the number of times talks were recommended, we explored the degree to which talks by each gender was linked to talks of the same or opposite gender. We

found that on average in talks by men the recommendation list was composed 71.2% by men and 26.7% were by women; in contrast, for women, 54.3% were talks by men and 43.5% were talks by women. Thus, men tended to recommend men, whereas women recommended proportionally more women. This discrepancy could in part be due to the difference in the number of male presenters to female, but this result motivated us to further assess the significance of these gender-gender links to see whether these links faction the recommendation network, potentially creating clusters in which viewers are “stuck” watching talks primarily given by people of the same gender. As part of an effort to promote diversity in thinking, we believe that being exposing viewers to multiple backgrounds and self-identity is a key goal for TED.com.

Community analysis

Using the CNM community detection algorithm, we identified ten non-overlapping communities in the recommendation network, with sizes 6, 10, 15, 27, 52, 160, 499, 501, 638, and 642. The modularity of the network was 0.542.

Community talk topics. We then explored the topics most represented in each community. We used the “tags” information from the data as a proxy for topic - tags are microtopics that each talk is associated with; each talk could be associated with multiple tags. While some were rather similar, there were a total of 416 unique topics in a wide range of areas.

Because of the large number of topics, we sought to simplify analyses by looking at broader topics. We found the proportion of each talks associated to each topic within a community by calculating ratio of raw counts of each tag. Because of the large number of potential topics, we focused on the top 15 most represented topics in each

community. The 15 most represented topics seemed to be somewhat consistent with each other; thus we assigned each community to an overarching “megatopic” based on the 25 topics, ignoring the TED-specific tags such as TEDx, TED-Ed, and TED Fellows (Table 1).

The 15 most represented topics of each community are somewhat consistent, making the identification of megatopics not difficult. But community sizes varied widely, and the variation in topics seemed to increase with increased community size; thus, it is important to note that these megatopics are rough groupings and much variation exist within each megatopic. Because of the limited time for this study, we did not have time to more objectively group each topic, but one way could be to use principle components analysis to group the topics into independent megatopics, then group the communities using the megatopic most represented in that community. The analysis below were based on our rough groupings, but we tried to refrain making strong claims and only focused on the larger trends; hopefully future studies could delve deeper into these trends using more mathematically driven grouping methods.

Gender distribution across topics and communities. We calculated the proportion of talks in each community that was given by men or women, and found that there is a significant difference in the distribution of men and women across the ten communities (two-sample t-test, $t = 8.37$, $p = 2.15 \times 10^{-07}$), indicating that gender distribution was far from uniform across megatopics.

Specifically, the three communities with the largest proportion of talks by men were: community 1 (83.3%), community 3 (80%), and community 4 (78%). Specifically, community 1 had no talks given by women (the other 16.6% was by non-human or a group).

Table 1. Topics (in decreasing order) and gender represented in each community

Cmty	Top 15 topics	Megatopic	Cmty size	% men	% women
1	science, physics, education, culture, technology, philosophy, curiosity, math, consciousness, space, nanoscale, Planets, creativity, psychology, history	futuristic science	6	83%	0%
2	compassion, charter for compassion, global issues, religion, self, science, culture, faith, brain, business, communication, humanity, community, life, fear	faith and religion	10	60%	40%
3	technology, design invention, interface design, computers, innovation, demo, creativity, future, communication, Internet, art, visualizations, potential, engineering, cities	engineering innovation	15	80%	13%
4	technology, global issues, health, disease, design, science, innovation, medicine, insects, business, invention, manufacturing, public health, health care, biotech	global wellness	27	78%	22%
5	data, culture, love, technology, society, art, relationships, LGBT, communication, humanity, humor, social change, media, design, business	love and society	52	54%	40%
6	education, children, culture, global issues, food, teaching, technology, language, science, creativity, social change, design, society, math, poetry	global social activism	160	70%	28%
7	global issues, technology, culture, business, economics, politics, social change, activism, war, government, society, communication, Internet, Africa	global political activism	499	63%	35%
8	science, technology, health, medicine, health care, brain, biology, culture, medical research, neuroscience, humanity, global issues, future, design, innovation	medicine and biology	501	65%	33%
9	science, technology, design, environment, global issues, cities, innovation, biology, exploration, nature, business, architecture, art, climate change, oceans	environment	638	72%	27%
10	entertainment, culture, music, technology, performance, art, creativity, business, psychology, live music, design, storytelling, science, society, personal growth	culture and arts	642	62%	31%

The three communities with the largest proportion of talks by women were community 5 (40.4%), community 2 (40%), and community 7 (35.1%). Notably, however, even in community 5, the community with the largest proportion of women, men nonetheless were still a larger proportion than women (54% vs. 40%).

The three communities with the largest proportion of men, communities 1, 3, and 4, were among the smallest communities and represent the hard science, technology, and engineering-focused topics, which are traditionally and stereotypically associated as male disciplines, and seen to have a

prevalent gender-gap (Beede et al., 2011). In contrast, the three communities with the largest proportion of women, communities 5, 2, and 7, encompass topics in faith and religion, love, society, and political activism, “soft sciences” traditionally and stereotypically associated with females. This trend was also represented by the top ten topics of the talks given by each gender.

Ratings and viewership distribution across communities. For each community, we aggregated the total counts for each of the 15 possible rating categories, obtaining a raw rating count distribution, then divided

Table 2. Top eight ratings for each community (in decreasing order)

Cmty	Top 8 ratings
1	informative, fascinating, inspiring, persuasive, ingenious, funny, beautiful, jaw-dropping
2	inspiring, beautiful, informative, persuasive, fascinating, courageous, OK, longwinded
3	jaw-dropping, ingenious, fascinating, inspiring, informative, beautiful, OK, persuasive
4	ingenious, informative, fascinating, inspiring, jaw-dropping, persuasive, funny, OK
5	inspiring, informative, fascinating, funny, beautiful, courageous, persuasive, ingenious
6	inspiring, informative, persuasive, fascinating, beautiful, funny, ingenious, courageous
7	inspiring, informative, persuasive, courageous, fascinating, jaw-dropping, funny, beautiful
8	informative, inspiring, fascinating, persuasive, courageous, jaw-dropping, beautiful, ingenious
9	inspiring, fascinating, informative, ingenious, beautiful, jaw-dropping, persuasive, funny
10	inspiring, fascinating, informative, beautiful, persuasive, funny, courageous, ingenious

each count by the community size to obtain normalized rating distributions. The top eight ratings for each community is presented in Table 2.

We then performed two-tailed two-sample t-tests on the normalized rating distributions of every pair of communities. We found that for almost all community pairs, the rating distributions were not significant (with $p < 0.05$ threshold), except for the following pairs: (2, 3), (2, 5), (2, 6), (2, 10), (3, 4), and (4, 10). Community 2 - the community with the second highest proportion of female speakers - seemed to differ the most from other communities in its rating distribution. However, its rating distribution did not differ only from rating distributions of male-dominated communities, as it differed as well from community 5 - the community with the largest representation of women. The same pattern was seen in men, as communities 3 and 4 differed in their rating distributions despite both having large male representation. Overall, a consistent gender difference in ratings did not seem to be found when talks were grouped into topical communities, suggesting that the group-

level gender difference in ratings that we saw in our summary statistics above were likely reflecting topical differences in talks by men and women.

Of the top 15 most viewed talks, the distribution was as followed:

- 12 in community 10
- 2 in community 8
- 1 in community 7

Of the 15 least viewed talks, the distribution was as followed:

- 8 in community 10
- 4 in community 7
- 1 in community 8

Of the top 15 highest in-degree talks, the distribution was as followed:

- 7 in community 10
- 5 in community 9
- 2 in community 7
- 1 in community 8

Of the 15 lowest in-degree talks, the distribution was as followed:

- 6 in community 7
- 5 in community 10
- 3 in community 9
- 1 in community 5

Unsurprisingly from a probabilistic standpoint, the largest community contained all four of these groups. However, the distributions, community 7 - with topics on global politics and activism - seemed to have more least viewed talks and more lowest in-degree talks, providing some evidence of association between low in-degree and low viewership. Whether the low viewership is a result of the topic itself or the low recommendations could pose interesting future studies.

5. Discussion and Future Work

TED is a unique content aggregator in that its core mission to create content is not to raise viewership for viewerships' sake or for monetization, but to spark discussion and spread ideas. While most recommendation algorithms in social media and content platforms like Facebook, Pinterest, or YouTube aim to promote user-desirable content to keep users on the platform, we believe that the goal of education goes further than promoting simply the familiar, but also diversity of thought, ideas, and background.

We hoped to explore potential deficiencies in the current structure and content on TED.com, specifically using its recommendation network, and exploring whether certain modifications could be made to "improve" the structure and content towards the goals highlighted above. Using centrality analysis and community detection, we grouped the recommendation network into ten communities that seemed to distinguish themselves through the topics represented. We thus labeled each community with a "megatopic" and explored the characteristics of each community in relation to its megatopic. From our analyses, we draw two major findings and implications:

Potentially sparse inter-topic interaction. The fact that each community could be identified by certain topics illustrates that the recommendation network is somewhat driven by topics, which is a common way of recommending related videos. However, since TED.com is first and foremost an educator, perhaps the site should utilize the recommendation network more effectively to educate the viewers instead of purely guiding them to videos they are potentially interested in from a topical standpoint.

However, as we stated in the results, the way in which we group each community into megatopics was rough and would warrant more sophisticated, objective analyses to draw stronger conclusions.

Potential perpetuation of gender stereotypes. Partly shown by our analyses on the percentages of male and female talks on the recommended talks list of male and female talks, the difference in gender distribution across communities further confirms our speculation of male-recommending-male and female-recommending-female phenomenon. Whether they were driven by topical differences between genders or not, this information is important as a purely network-driven community detection algorithm was able to "separate out" communities more with higher or lower representation from each gender. Additionally, the topics represented by each gender seemed to conform to the stereotypes associated with each gender - males with "hard science," and females with "soft science." This is important - TED.com is an educational content distributor, and this disproportionate gender representation could further perpetuate these gender stereotypes to eager learners.

These patterns could be alleviated by efforts in finding speakers that combat the gender stereotype of disciplines - ex. more

females in STEM, males in topics on love and relationships - and in combination with recommending talks across topics so that all topics - and thus both genders - gain the same amount of exposure to all viewers.

Future directions

The results from this project could warrant some potential future studies.

First, we could identify the “weak ties” between communities, investigate the characteristics of these talks, and perhaps transform these weak ties to strong ties so that the recommendation network - and thus the viewers - are better connected to the variety of topics presented on the site.

Using centrality analysis and community detection, we identified the key topics that may be engagement points for TED.com. From here, we could further study the metrics involved that correlate amongst influential nodes and topics to potentially provide metrics that predict future key videos. Continuing, we hope that insight in differences between influential and non-influential nodes could inform more conscious efforts of distributing educational content. We could also implement an influence maximization algorithm, such as the greedy hill-climbing (Kempe et al., 2003), to identify nodes with large marginal gains. These top nodes would provide a varied list of videos that are in their respective unique communities. If the user is inclined, these videos could also lead them to a huge and diverse set of content.

Second, we treated all videos on the recommended list as equal, but by the design of the website and part human psyche, the first recommended video is likely to have greater impact on user experience than the 6th. We could explore using weighted edges to generate the TED.com graph and explore how generating the graph with this type of weighted

significance changes our perception of the TED network.

Third, although outside the scope of this project, an expansion of our techniques would be to look at correlations between recommendations to other measures of bias such as race and topic. Diversity in education is important, and given the huge influence of the media and the web, the curation of online education should be especially careful in its design so that potential perpetuation of race, gender, and other biases is taken into consideration and addressed.

References

- (1) D. N. Beede, T. A. Julian, D. Laongdon, G. McKittrick, B. Khan, M. Doms. 2011. Women in STEM: A Gender Gap to Innovation. Economics and Statistics Administration Issue Brief, 04-11.
- (2) A. Boring. 2017. Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41
- (3) L. Farrell, L. McHugh. 2017. Examining gender-STEM bias among STEM and non-STEM students using the Implicit Relational Assessment Procedure (IRAP). *Journal of Contextual Behavioral Science*, 6(1), 80-90.
- (4) D. Kempe, J. Kleinberg, E. Tardos. Maximizing the Spread of Influence through a Social Network. In *Proc. KDD 2003*.
- (5) R. Zhou, S. Khemmarat, L. Gao, J. Wan, J. Zhang, Y. Yin, J. Yu. 2016. Boosting video popularity through keyword suggestion and recommendation systems. *Neurocomputing*, 205, 529-541
- (6) H. Shiokawa, Y. Fujiwara, M. Onizuka. Fast Algorithm for Modularity-Based Graph Clustering. In *AAAI*, 2013.

- (7) Newman, M. E. J., and Girvan, M. 2004. Finding and Evaluating Community Structure in Networks. *Physical Review E - PHYS REV E* 69:026113.
- (8) Clauset, A.; Newman, M. E. J.; and Moore, C. 2004. Finding Community Structure in Very Large Networks. *Physical Review E - PHYS REV E* 70:066111.