

The Interdisciplinarity of TED Talks, a Network Perspective

MACSS Thesis Proposal
Huanye Liu

1. Introduction and research questions

Interdisciplinarity generally means the combination or integration of knowledge from different specialties. People now believe that interdisciplinarity can bring added value as they believe the total is greater and more interesting than the sum of parts. So far we have found increasing levels of interdisciplinarity in all kinds of knowledge-intensive organizations and platforms such as high-tech companies, academic institutions and research centers. In this paper, we turn to TED, a nonprofit platform to spread ideas usually in the form of short talks[0].

TED talks are garnering public's attention these years with various interesting topics. For now, 449 topic keywords have been listed on [TED website](#) which have covered almost every aspect of human life[0]. We notice each talk published on the website was labeled with more than one topic keyword, and for some talks the list of keywords includes topics from several different domains, which implies the interdisciplinary nature of these talks. So how can we measure the interdisciplinarity of TED talks based on its list of topic keywords? What is the trend of interdisciplinarity of TED talks measured at different granularities? Is there any relationship between the talk interdisciplinarity and viewers' response?

2. Data and method

We download the dataset from [Kaggle](#) which consists of meta data about 2550 TED talks from year 2005 to 2017. Later we will try scraping data directly from [TED website](#).

The dataset has 19 columns from which we select 3 relevant columns:

‘id’: the TED talk ID

‘comments’: the number of comments made by online viewers for each TED talk

‘tags’: a list of keyword topics labeled by the publisher for each TED talk

We use column ‘comments’ to compute the average number of comments indicating talk popularity instead of the number of views, because the former is more related to the talk content and comments are often made after watching the whole video thus more indicative of the genuine popularity. We use column ‘tags’ to derive metrics for interdisciplinarity measurement.

In this thesis we propose several metrics for interdisciplinarity measurement at different granularities by means of LDA topic modeling and network analysis.

1) First, we identify a few major categories for TED talks by applying the LDA topic model to their lists of topic keywords, which generates one major category distribution for each TED talk.

2) Based on the major category distribution, we score each talk's interdisciplinarity using the pvalue from the Chi-square test.

- 3) Next we build a network of talks based on the similarity of every two talks' major category distributions and detect clusters in the network with one cluster corresponding to one major category, and then
- 4) We compute the similarity ratios define as the inter-cluster similarity over the intra-cluster similarity between every two clusters, a metric of interdisciplinarity measured for one major category related to another major category.
- 5) Using this metric we divide TED talks by year to examine the temporal trend of similarity between every pair of major categories.

We also explore the relationship between the popularity and the interdisciplinarity measured at the major category level. For another interdisciplinarity measurement at the major category level, we take a new approach as follows:

- 1) First choose a relatively large number of major categories and build one network for each year based on similarity between talk distributions on major categories, and use Louvain algorithm to detect clusters in each network as before to examine which major categories more often appear in one cluster and how the major category members in each cluster change in different years.
- 2) Based on how many times every pair of major categories falls into one cluster over years, build a new network with nodes of major category and edges weighted by number of times sharing one cluster, and then
- 3) Examine the relationship between two metrics of node centrality in the network and the average number of comments for each major category respectively.

3. Literature review

Previous quantitative analysis on interdisciplinarity were mainly related to scientific research, papers, journals and education systems. Richard Van Noorden's analysis on interdisciplinary research revealed the extent and impact of research bridging disciplines by several summative indicators such as the number of papers with "interdisciplinar*" in title, the change in average paper citations, citations from outside disciplines et al[1]. Kristine Lund et al explored the intersection of education and computer science from the MOOC literature using bibliographic coupling and network analysis[2]. Zhiya Zuo et al showed that a higher level of multidisciplinary in a academic institution not necessarily leads to a higher level of collaboration within the the institution using social network analysis and text mining[3]. Loet Leydesdorff et al proposed a method to distinguish and rank journals in terms of interdisciplinary using betweenness centrality and diversity from the citation networks of journal articles[4]. Qi Wang et al measured interdisciplinarity of research at the level of topics by measuring both the degree of dissimilarity between pairs of publications for a topic and the strength of within-topic citation relations for these pairs[5]. Filipi N. Silva et al used an entropy measurement based on the diversity of subject categories citing a specific journal to quantify the interdisciplinarity of the journal[6]. Alan L. Porter et al defined two metrics to measure the interdisciplinarity for a body of research: "Integration" measuring the extent to which diverse Web of Knowledge Subject Categories were cited in a research article, and "specialization" measuring the spread of Subject Categories the body of research was published in[7]. Among them, [5][7] measured interdisciplinarity by comparing two opposite metrics, [2][3][4][6] network analysis and [3] also text mining. In the thesis, we

borrow ideas from all the three approaches and integrate them to measure the interdisciplinarity of TED talks at different granularities.

References:

[0] www.ted.com

[1] Richard Van Noorden (2015), Interdisciplinary research by the numbers, article in Nature 525(7569):306-7, September 2015.

[2] Kristine Lund, Bodong Chen and Sebastian Grauwin(2018), The Potential of Interdisciplinarity in MOOC Research: How Do Education and Computer Science Intersect? Fifth Annual ACM Conference on Learning at Scale, London, UK, June 2018.

[3] Zhiya Zuo and Kang Zhao(2018), The more multidisciplinary the better? The prevalence and interdisciplinarity of research collaborations in multidisciplinary institutions, article in Journal of Informetrics 12(3):736-756, August 2018.

[4] Loet Leydesdorff, Caroline S. Wagner and Lutz Bornmann, Betweenness and diversity in journal citation networks as measures of interdisciplinarity—A tribute to Eugene Garfield, article in Scientometrics, Volume 114, Issue 2, pp 567–592, February 2018.

[5] Qi Wang and Per Ahlgren, Measuring the interdisciplinarity of research topics, 23rd International Conference on Science and Technology Indicators, Leiden, The Netherlands, September 2018.

[6] Filipi N. Silva, Francisco A. Rodrigues, Osvaldo N. Oliveira Jr, Luciano da F. Costa, Quantifying the interdisciplinarity of scientific journals and fields, article in Journal of Informetrics, DOI: 10.1016/j.joi.2013.01.007 , April 2013

[7] Alan L. Porter, Alex S. Cohen, J. David Roessner, Marty Perreault, Measuring researcher interdisciplinarity, article in *Scientometrics*, Vol. 72, No. 1 (2007) 117–147, July 2007