

Ted Talks visualization

COM-480 Data visualization

Antti Hemilä, Erik Husgafvel, Maria Pandelev

April 2020

1 Introduction

The main goal of our project is to provide users with insights about different Ted Talks, to understand how they are perceived by the public and how they evolved throughout the years. We want users to be able to explore this vast data by zooming in from general features, such as where and when these talks are being given, to how different characteristics correlate and finally understanding relationships between them.

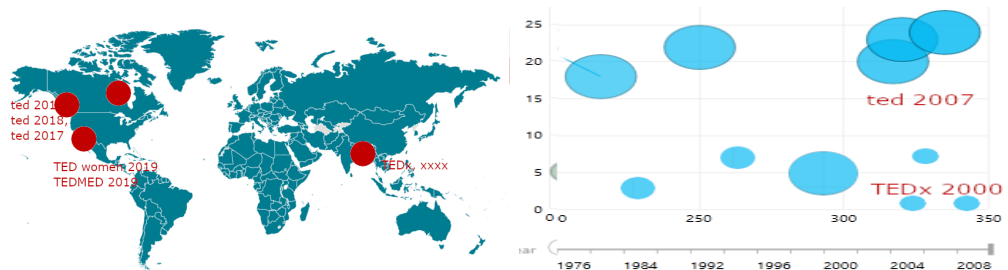
2 Main visualizations

Here we list the main visualization to reach our goal. For each visualization we list a description, sketches, desired functionalities and extra ideas.

- **Map of events and time series**

At the beginning we want to make the user gets a general idea about TED events. We will show a map with all the TED events present in this dataset, like in the following sketches. Then we will show the user a timeseries of events to help them understand the evolution of the talks.

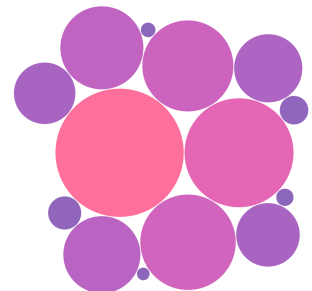
Extra ideas: animate the map and timeline (starting from 2000 and ending to 2017 for example), add a graph to show the distinction between TED and TEDx.



- **Bubble chart of ratings**

Our bubble chart presents ratings of the videos and each bubble's size is proportional to the number of times a rating has been given. Due to the different sizes, displaying text on a bubble is quite a challenge. With this presentation, user can easily see, which of the ratings are given more often. With that information one can conclude which of the ratings are rarer or less general than others. Titles and the value will appear when hovering over the bubble.

Extra ideas: add user interaction: User could filter some ratings / explore the different sets that make the bubble chart by clicking on a bubble / use another mapping for the bubble chart (100 most popular tags).



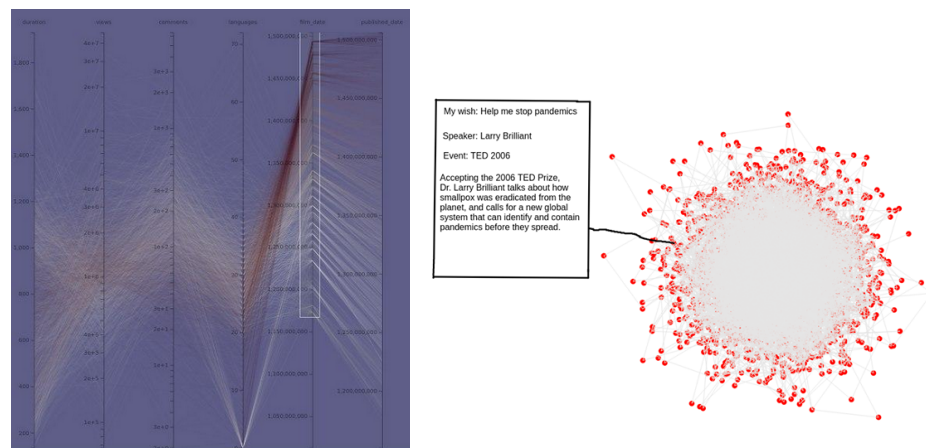
- **Speakers occupation and frequency in talks**

In this chart we would put the most frequent speakers based on the number of talks they given. At this point we are not sure what this chart would like but it would display general stats about how many talks have the first best 5 speakers given: subject of the talks, frequency of events over the years. We also want to group speakers by occupation and gender. Overall, this would be a static visualization meant to further introduce the user before jumping in analyzing the various features we have for each talk.

- **Parallel coordinates**

A Parallel coordinates graph with all interesting variables (continuous or discrete). Can be used to find links between neighboring variables. Coloring is based on one of the variables, which helps exploring the data. Further, user can drag the dimensions and use d3 brushes to filter a part of the data. Number of comments and number of views are given in logarithmic scale. The required dimensions: views, comment, duration, count of words, **words per minute**, **sentiment** (degree of positivity), number of languages.

Extra ideas: show by default the most interesting correlations.



- **Network visualizing related Ted Talks**

Since relationships can be easily expressed in terms of graphs, we are using a force-directed layout to build a network of ted talks. Right now, we are drawing edges between them by looking at the recommended talks to watch next which results in a hairball. We plan to experiment with another metric and draw edges between talks by looking at how many common tags they have. User will be able to click on a node and find out the title and speaker of the talk.

Extra ideas: make the graph "zoom-able" (since we are dealing with more than 2500 nodes applying a raw force-layout is not feasible).

2.1 Technologies and lessons

To implement these we will use the D3 library and the official W3 CSS downloaded from the official [W3SCHOOLS](https://www.w3schools.com/) website. For coloring we want to stick with the original red theme and we will use Lecture 6 to find the best way. For graphs visualizations we are using Exercise 4 for scatter plotting, Lecture 8 for maps, Lecture 9 for figuring out how to best display rating/tag names, Lecture 10 for Ted Talks network visualization and Lecture 11 for parallel coordinates.

3 Functional project prototype

You can find a functional project prototype here: tedtalksviz.github.io