

CS 6630: Project Process Book  
TED talks topic trend visualization

Hsuan Lee and Chien-Wei Sun

School of Computing, University of Utah

December 1, 2017

# Contents

|   |           |
|---|-----------|
| <b>1 Overview</b>                       | <b>3</b>  |
| 1.1 Overview and Motivation . . . . .   | 3         |
| 1.2 Related Work . . . . .              | 3         |
| 1.3 Questions . . . . .                 | 4         |
| <b>2 Data</b>                           | <b>5</b>  |
| 2.1 Dataset . . . . .                   | 5         |
| 2.2 Exploratory Data Analysis . . . . . | 6         |
| <b>3 Design Evolution</b>               | <b>9</b>  |
| 3.1 Prototype . . . . .                 | 9         |
| 3.2 Evolution . . . . .                 | 9         |
| 3.2.1 Network Chart . . . . .           | 9         |
| 3.2.2 Word Cloud Chart . . . . .        | 12        |
| 3.2.3 Line Chart . . . . .              | 14        |
| 3.2.4 Button of Line Chart . . . . .    | 14        |
| 3.2.5 Video List . . . . .              | 14        |
| <b>4 Implementation</b>                 | <b>17</b> |
| 4.1 Interaction . . . . .               | 17        |
| 4.1.1 Network Chart . . . . .           | 17        |
| 4.1.2 Word Cloud . . . . .              | 17        |
| 4.1.3 Line Chart . . . . .              | 17        |
| 4.1.4 Button of Line Chart . . . . .    | 17        |
| 4.1.5 Video List . . . . .              | 17        |
| 4.2 Flow-Chart . . . . .                | 17        |

|  |           |
|--|-----------|
| <b>5 Evaluation</b>                    | <b>18</b> |
| 5.1 Solution in our question . . . . . | 18        |
| <b>Bibliography</b>                    | <b>19</b> |

# Chapter 1

## Overview

### 1.1 Overview and Motivation

TED is a leading organization which provides influential and understandable talk to the world. These talks cover a lot of fields, from anthropology to machine learning, and also from biology to sociology. We are interested in the relationship between technology and world market, and we want to know if TED somehow shows the trend of popular technology or it provides a platform for topics which do not get much attention in the world.

The relevance of different categories is also what we want to discover. For example, several years ago, it was popular that researchers tried to innovate theory according to the behavior of insects, like ants and bees. There are many theories developed based on the cooperation pattern of those animals. In the past, people did not consider that there is a strong relevance between insects and learning theory. We also wonder if we can find situation which is similar to the example.

### 1.2 Related Work

When we are searching useful data for this project, we found the TED talks dataset and also a visualization by Sean Miller[3]. In this visualization, it shows statistics of the dataset and also allow user to search video by one tag. However, this visualization does not answer the questions we mention in the above. That's why we decide to build our own visualization of the dataset.

### **1.3 Questions**

Here are questions we expect to answer at the end of this project:

- What are the trend of category tags appeared on TED talks?
- Is there any relationship between the TED talks and the big events happened in the world?
- Is there a strong relevance between two topics that in general people will not think they are related?
- Can we learn the trend of research on a specific field by analyzing the popularity of keywords? Or it shows the topics which people do not put attention on for now but will become important in the future?

# Chapter 2

## Data

### 2.1 Dataset

We find the dataset from Dataset Distribution Portal[4]. This dataset include the video recording from the TED website from 1972 to 2017. For each video, its data contains the following attributes:

| <b>id</b> | <b>speaker</b>   | <b>URL</b>                    | <b>URL</b>        | <b>description</b>        | <b>transcript_URL</b> | <b>month</b> | <b>year</b> | <b>film</b> | <b>event</b> | <b>duration</b> | <b>date_publis</b> | <b>tags</b>  |
|-----------|------------------|-------------------------------|-------------------|---------------------------|-----------------------|--------------|-------------|-------------|--------------|-----------------|--------------------|--|
| 1         | Al Gore          | Averting the climate crisis   | http://www.ted.co | With the same humor       | http://www.ted.i      | 2            | 2006        | TED2006     | 0:16:17      | 6/27/06         |                    | cars,alternative energy,culture,politics,science,climate change,environment,su   |
| 2         | Amy Smith        | Simple designs to save a life | http://www.ted.co | Fumes from indoor cd      | http://www.ted.i      | 2            | 2006        | TED2006     | 0:15:06      | 8/15/06         |                    | MacArthur grant,simplicity,industrial design,alternative energy,invention,engine |
| 3         | Ashraf Ghani     | How to rebuild a broken state | http://www.ted.co | Ashraf Ghani's passus     | http://www.ted.i      | 7            | 2005        | TEDGlobal_1 | 0:18:45      | 10/18/06        |                    | corruption,poverty,economics,investment,military,culture,politics,policy,global  |
| 4         | Burt Rutan       | The real future of space expl | http://www.ted.co | In this passionate talk   | http://www.ted.i      | 2            | 2006        | TED2006     | 0:19:37      | 10/25/06        |                    | aircraft,flight,industrial design,NASA,rocket science,invention,engineering,entr |
| 5         | Chris Bangle     | Great cars are great art      | http://www.ted.co | American designer Cr      | http://www.ted.i      | 2            | 2002        | TED2002     | 0:20:04      | 2004/05/07      |                    | cars,industrial design,transportation,invention,design,technology,business,art   |
| 6         | Craig Venter     | Sampling the ocean's DNA      | http://www.ted.co | Genomics pioneer Cr       | http://www.ted.i      | 7            | 2005        | TEDGlobal_1 | 0:16:51      | 2004/05/07      |                    | biootech,invention,oceans,genetics,DNA,biology,science,entrepreneur,biodivers    |
| 7         | David Pogue      | Simplicity sells              | http://www.ted.co | New York Times coll       | http://www.ted.i      | 2            | 2006        | TED2006     | 0:21:34      | 6/27/06         |                    | simplicity,computers,software,interface design,music,media,entertainment,per     |
| 8         | David Rockwell   | A memorial at Ground Zero     | http://www.ted.co | In this emotionally ch    | http://www.ted.i      | 2            | 2002        | TED2002     | 0:24:37      | 2006/12/07      |                    | New York,memory,interview,death,culture,architecture,disaster relief,cities,urb  |
| 9         | Dean Kamen       | To invent is to give          | http://www.ted.co | Inventor Dean Kamen       | http://www.ted.i      | 2            | 2002        | TED2002     | 0:20:07      | 2004/05/07      |                    | robots,cars,industrial design,transportation,invention,education,innovation,soc  |
| 10        | Dean Ornish      | The killer American diet that | http://www.ted.co | Forget the latest diets   | http://www.ted.i      | 2            | 2006        | TED2006     | 0:03:18      | 12/14/06        |                    | obesity,disease,health,health care,science,food,science,global issues            |
| 11        | Jane Goodall     | What separates us from chm    | http://www.ted.co | Jane Goodall hasn't fr    | http://www.ted.i      | 2            | 2003        | TED2002     | 0:27:25      | 2004/05/07      |                    | primates,Africa,culture,science,environment,animals,nature,global issues         |
| 12        | Eva Verets       | Meet the future of cancer res | http://www.ted.co | Eva Verets -- only 19     | http://www.ted.i      | 2            | 2005        | TED2005     | 0:18:49      | 10/02/06        |                    | wunderkind,cancer,disease,health,science,technology                              |
| 13        | Frank Gehry      | A master architect asks, Now  | http://www.ted.co | In a wildly entertaining  | http://www.ted.i      | 2            | 2002        | TED2002     | 0:22:36      | 1/17/08         |                    | invention,interview,culture,architecture,design,creativity,business              |
| 14        | Golan Levin      | Software (as) art             | http://www.ted.co | Engineer and artist G     | http://www.ted.i      | 2            | 2004        | TED2004     | 0:14:53      | 2004/05/07      |                    | invention,software,music,entertainment,performance,technology,art                |
| 15        | Helen Fisher     | Why we love, why we cheat     | http://www.ted.co | Anthropologist Helen      | http://www.ted.i      | 2            | 2006        | TED2006     | 0:23:27      | 2009/06/06      |                    | gender,relationships,cognitive science,psychology,evolution,culture,scienc       |
| 16        | Janine Benyus    | Biomimicry's surprising less  | http://www.ted.co | In this inspiring talk at | http://www.ted.i      | 2            | 2005        | TED2005     | 0:23:19      | 2004/05/07      |                    | biomimicry,DNA,evolution,biology,fish,science,environment,animals,design,tec     |
| 17        | Kevin Kelly      | How technology evolves        | http://www.ted.co | Tech enthusiast Kevin     | http://www.ted.i      | 2            | 2005        | TED2005     | 0:20:08      | 11/14/06        |                    | philosophy,evolution,culture,choice,history,science,future,technology            |
| 18        | Malcolm Gladwell | Choice, happiness and spag    | http://www.ted.co | "Tipping Point" autho     | http://www.ted.i      | 2            | 2004        | TED2004     | 0:17:30      | 9/19/06         |                    | consumerism,marketing,economics,culture,media,food,choice,storytelling,bus       |
| 19        | Mena Trott       | Meet the founder of the blog  | http://www.ted.co | The founding mother i     | http://www.ted.i      | 2            | 2006        | TED2006     | 0:16:46      | 8/25/06         |                    | software,culture,design,entertainment,storytelling,business,communication,co     |
| 20        | Michael Shermer  | Why people believe weird th   | http://www.ted.co | Why do people see th      | http://www.ted.i      | 2            | 2006        | TED2006     | 0:13:25      | 11/08/06        |                    | faith,illusion,culture,religion,science,entertainment                            |
| 21        | Peter Gabriel    | Fight injustice with raw vide | http://www.ted.co | Musician and activist     | http://www.ted.i      | 2            | 2006        | TED2006     | 0:14:08      | 12/06/06        |                    | TED Brain Trust,film,culture,music,activism,social change,storytelling,global is |
| 22        | Pilobolus        | A dance of "Symbiosis"        | http://www.ted.co | Two Pilobolus dancer      | http://www.ted.i      | 2            | 2005        | TED2005     | 0:13:45      | 2002/09/07      |                    | dance,science and art,science,nature,entertainment,performance                   |
| 23        | Richard Baranuk  | The birth of the open-source  | http://www.ted.co | In 2006, open-learning    | http://www.ted.i      | 2            | 2006        | TED2006     | 0:18:34      | 8/21/06         |                    | open-source,library,education,culture,global issues,technology,business,collat   |
| 24        | Rivera           | If I controlled the Internet  | http://www.ted.co | How many poets cool       | http://www.ted.i      | 11           | 2006        | TEDSalon 2  | 0:04:07      | 12/14/06        |                    | love,poetry,philosophy,culture,entertainment,performance                         |
| 25        | Ross Lovegrove   | Organic design, inspired by i | http://www.ted.co | Designer Ross Loveg       | http://www.ted.i      | 2            | 2005        | TED2005     | 0:19:30      | 8/15/06         |                    | industrial design,invention,product design,science and art,DNA,biology,nature    |
| 26        | Seth Godin       | How to get your ideas to spr  | http://www.ted.co | In a world of too manj    | http://www.ted.i      | 2            | 2003        | TED2003     | 0:17:01      | 2004/05/07      |                    | TED Brain Trust,marketing,culture,choice,storytelling,business                   |
| 27        | Shawn I aurit    | The freshmining of rank       | http://www.ted.co | "Freshmining" autho       | http://www.ted.i      | 2            | 2004        | TED2004     | 0:21:14      | 9/19/06         |                    | mining,rare earth,minerals,mining,environment,nature,resource,business           |

**Figure 2.1.1:** Data get from idiap.ch

|                |                |
|----------------|----------------|
| id             | month filmed   |
| Speaker        | year filmed    |
| headline       | event          |
| URL            | duration       |
| description    | date published |
| transcript URL | tags           |

To better understand the impact of TED videos, we develop web crawlers to collect attributes like **rates**(how do people feel after watching a video), **views**(how many time a video has been played), and some potentially valuable data like datetime, redirected urls, and transcripts. We use **Scrapy**[2] as our crawler. Figure 2.1.2 displays the rating options on TED website.

Rate this talk X

How would you describe this talk? Tell us by choosing up to three words. (If you choose just one, it will count three times.)

|   |   |
|---|---|
| <input type="checkbox"/> Informative<br><input type="checkbox"/> Obnoxious<br><input type="checkbox"/> Persuasive<br><input type="checkbox"/> Unconvincing<br><input type="checkbox"/> Beautiful<br><input type="checkbox"/> OK<br><input type="checkbox"/> Confusing | <input type="checkbox"/> Inspiring<br><input type="checkbox"/> Fascinating<br><input type="checkbox"/> Longwinded<br><input type="checkbox"/> Courageous<br><input type="checkbox"/> Ingenious<br><input type="checkbox"/> Funny<br><input type="checkbox"/> Jaw-dropping |
|---|---|

Submit

[See all ratings](#)

**Figure 2.1.2:** How people rate one video in TED website

Furthermore, in order to load data easily, we transfer our data from csv file to json form. We found this preprocessing can be accomplished painlessly by using **Pandas**[1] toolkit. Figure 2.1.3 shows what kind of data one video contains.

We plan to visualize the data according to the tags/keywords of the video. It is not efficient to search all the data to find which videos are related with one specific tag on javascript. For practical implementation, we will preprocess the dataset based on tags, which means to use tag as key to create input data.

## 2.2 Exploratory Data Analysis

In our design, the main chart user interact with is the network chart, which present the co-occurrence of tags. Hence, after we finish the job of collecting data, we move forward to build the co-occurrence matrix of tags. During this procedure, we observe that some tags

```

{
  "id": 7,
  "speaker": "David Pogue",
  "headline": "Simplicity sells",
  "URL": "http://www.ted.com/talks/view/id/7",
  "description": "New York Times columnist David Pogue takes aim at technology's worst",
  "transcript_URL": "http://www.ted.com/talks/view/id/7/transcript?language=en",
  "month_filmed": 2,
  "year_filmed": 2006,
  "event": "TED2006",
  "duration": "0:21:26",
  "date_published": "6/27/06",
  "tags": "simplicity,computers,software,interface design,music,media,entertainment,perf",
  "newURL": "https://www.ted.com/talks/david_pogue_says_simplicity_sells",
  "date": "2006-06-27",
  "views": "1646773",
  "rates": [
    {
      "id": 7,
      "name": "Funny",
      "count": 968
    },
    {
      "id": 3,
      "name": "Courageous",
      "count": 46
    },
    {
      "id": 9,
      "name": "Ingenious",
      "count": 186
    },
    {
      "id": 1,
      "name": "Beautiful",
      "count": 60
    },
    {
      "id": 21,
      "name": "Unconvincing",
      "count": 104
    },
    {
      "id": 11,
      "name": "Longwinded",
      "count": 78
    }
  ]
}

```

**Figure 2.1.3:** Data of one video in JSON

appear in too many videos so that their existences are not meaningful to the matrix. These tags are ‘science’, ‘technology’, ‘global issue’. Since they show up in most of talks, we remove them from the matrix so that the network chart will look clear.

To create groups of tags, we apply k-means to divide them into 11 clusters, and one of them restore the outliers. Figure 2.2.1 are the results of two groups. One is the group whose center is tag ‘computers’, the other is the group whose center is ‘universe’. Color is used to distinguish the group in our design.

```
"computers": [
    "simplicity",
    "computers",
    "software",
    "interface design",
    "robots",
    "library",
    "one laptop per child",
    "complexity",
    "intelligence",
    "code",
    "math",
    "web",
    "ai",
    "moon",
    "literature",
    "drones",
    "hack",
    "programming",
    "prediction",
    "data",
    "internet",
    "iran",
    "bullying",
    "algorithm",
    "machine learning",
    "augmented reality",
    "surveillance",
    "sexual violence"
],
"universe": [
    "nasa",
    "cosmos",
    "universe",
    "astronomy",
    "time",
    "physics",
    "exploration",
    "planets",
    "space",
    "solar system",
    "string theory",
    "big bang",
    "extraterrestrial life",
    "dark matter",
    "mars",
    "nobel prize",
    "asteroid",
    "telescopes"
]
```

**Figure 2.2.1:** Clustering result

# Chapter 3

## Design Evolution

### 3.1 Prototype

Our design is based on the network layout, as shown in Figure 3.1.1. This network is composed of tags, and user can choose several tags they are interested in to through interaction with the network node. Next, the line chart in the middle of Figure 3.1.1 will display the tendency of chosen tags versus time/year. The last part help user to search for TED talks including these tags. User can decide the result is sorted by views or popularity.

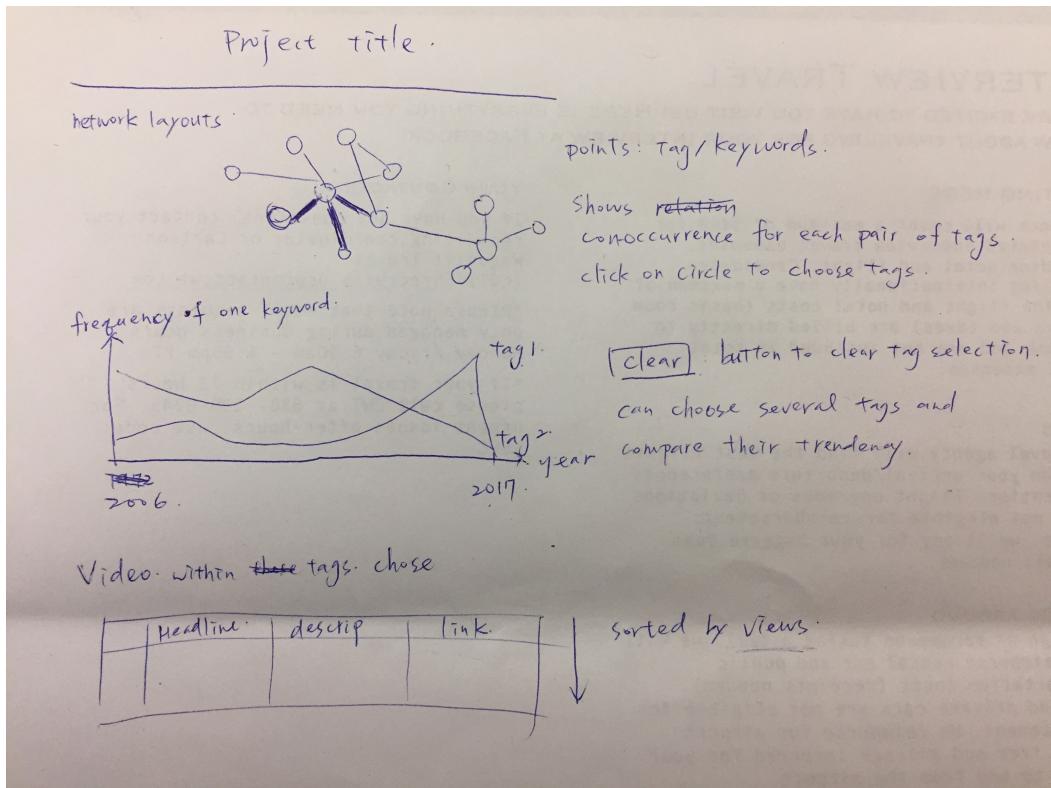
We also want to compare the statistic of the tags between years, so we design a bar chart as shown in Figure 3.1.2. By making use of the sliding bar on the top, the statistics of two years is displayed. Figure 3.1.3 helps us to figure out what attributes are needed in each chart. It also shows the relationship of charts.

### 3.2 Evolution

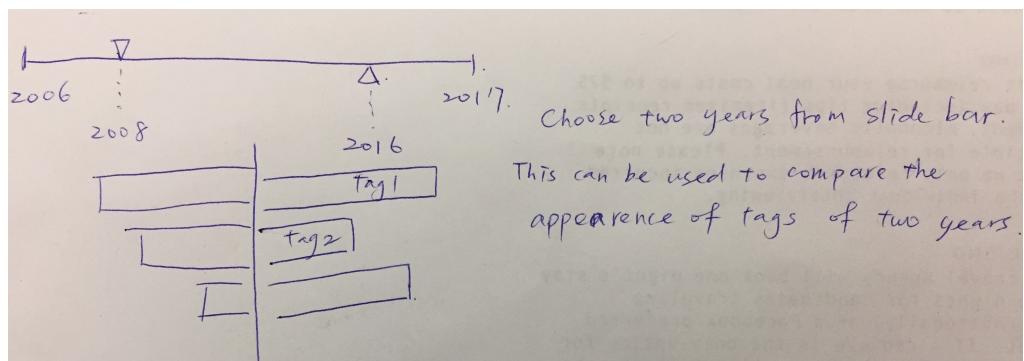
#### 3.2.1 Network Chart

First, we generate the network chart accoring the co-occurrence matrix. Each node represent a tag, and the thickness of one link is decided by the co-occurrence value between two tags. However, there are 403 tags and 19488 links on this chart, which make the network look crazy and take a lot of time to draw these lines, as shown in Figure 3.2.1.

We discuss how to fix this issue and propose two solution for that. One is to draw chord layout in the beginning. Chord layout help people understand the relationship between two groups. We can let user to click ribbon to then show the network layout of tags in these two



**Figure 3.1.1:** Design draft



**Figure 3.1.2:** Design of Optional features

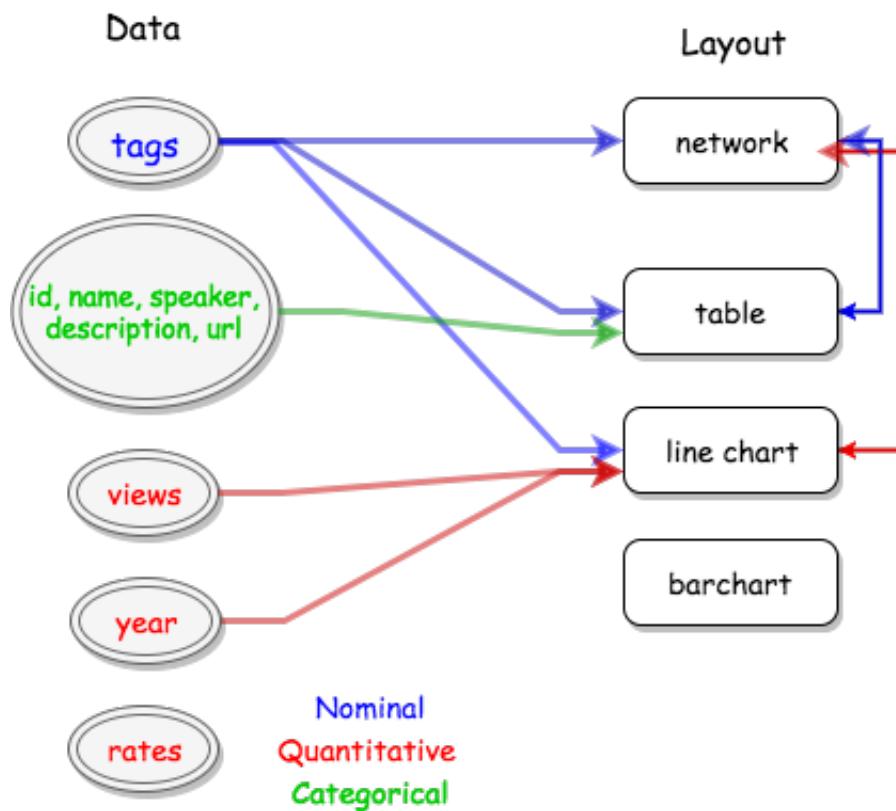
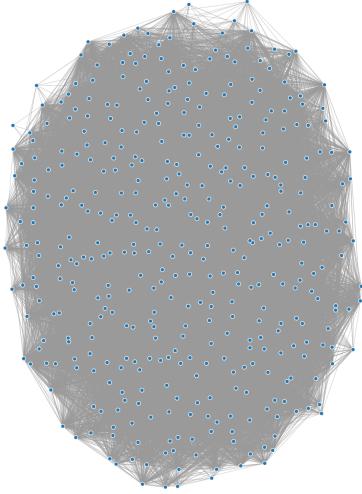


Figure 3.1.3: Category of data, and its relationship with the layout



**Figure 3.2.1:** Network chart with Over 19,000 links

groups. However, this design does not allow us to observe all the related tags of one tag we choose. The other solution is to reduce the amount of links and nodes. We can provide an overview of network chart with nodes and links whose frequencies and value of co-occurrence are bigger than threshold. Then, to zoom in on this chart, user can double-click on the tag they interested in to find all the other tag which is related to the chosen one. After applying the second method, our network chart looks better, as you can find in Figure 3.2.2.

Next, we want to observe the relationship between a chosen tag and other co-occurrence tags, so we implement a function that the network chart shows all the links whose edges include a specific tag after double-clicking on one the corresponding node. The image was not easy to understand and hard to find the most relative tag, as shown in Figure 3.2.3.

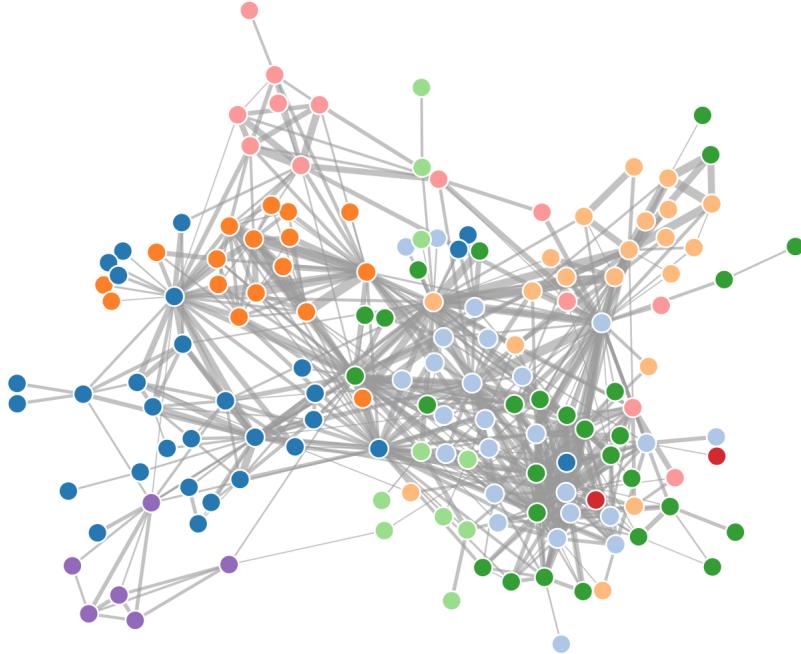
Figure 3.2.4 is the modified design to solve the above issue. This time, we grouped all the related tags by their categories. Instead of connecting the center with relative node, we connect it to an invisible group center and then link node with the center. This design help user to learn which category has strong co-occurrence within the specific, and also it is obvious to observe the most related tag.

### 3.2.2 Word Cloud Chart

After we draw the network chart and color each node according to category, we suddenly find that we did not explain the category and which tags are classified to. Therefore, we decide to add a word cloud chart on the top of page, which provides the information about

## TEDmap TED talks topic trend visualization

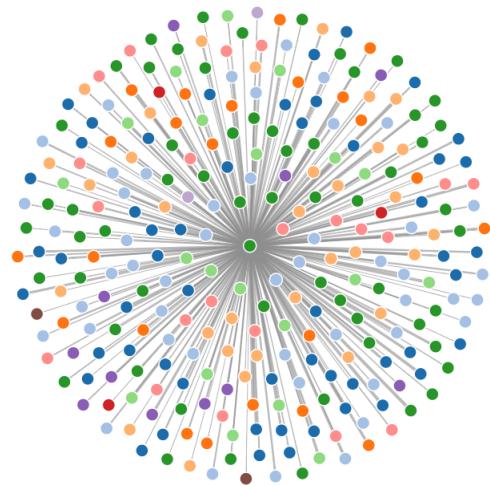
---



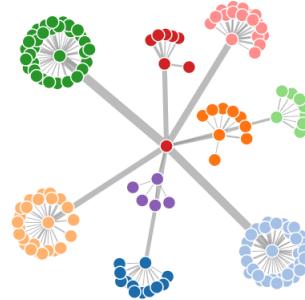
**Figure 3.2.2:** Network chart with link vale bigger than 15

## TEDmap TED talks topic trend visualization

---



**Figure 3.2.3:** Zoom in for one tag



**Figure 3.2.4:** Zoom in for one tag with grouping

the members each category includes, and the color of text follow the ordinal color scale we define in network chart.

In the beginning, the word cloud use a list of buttons for user to select the category, as Figure 3.2.5

### 3.2.3 Line Chart

### 3.2.4 Button of Line Chart

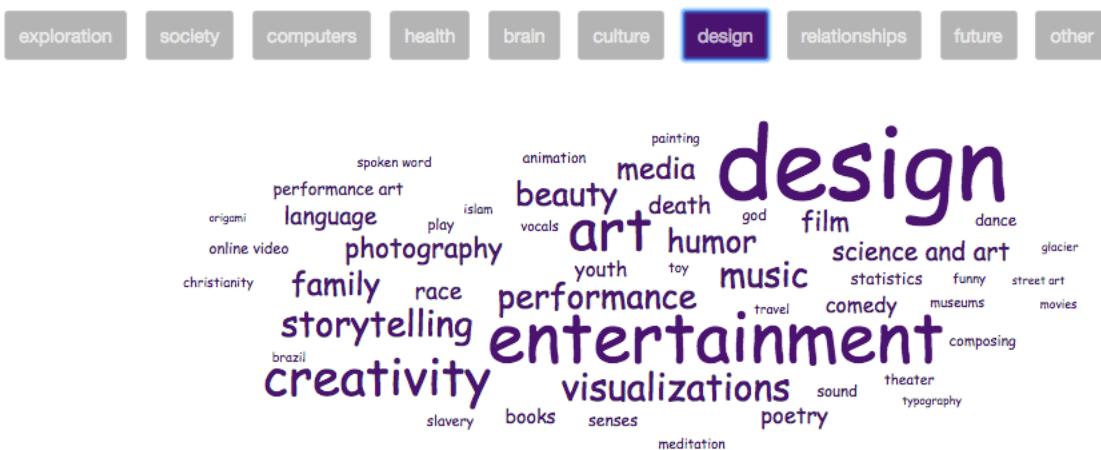
### 3.2.5 Video List



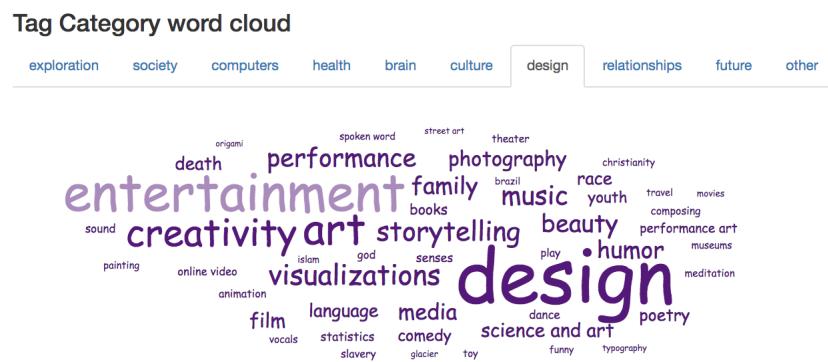
**Figure 3.2.5:** Word Cloud button design, version 1

**TEDmap** TED talks topic trend visualization

## Tag Category word cloud



**Figure 3.2.6:** Word Cloud button design, version 2



**Figure 3.2.7:** Word Cloud button design, version 3

# **Chapter 4**

## **Implementation**

### **4.1 Interaction**

#### **4.1.1 Network Chart**

#### **4.1.2 Word Cloud**

#### **4.1.3 Line Chart**

#### **4.1.4 Button of Line Chart**

#### **4.1.5 Video List**

### **4.2 Flow-Chart**

# **Chapter 5**

## **Evaluation**

### **5.1 Solution in our question**

# Bibliography

- [1] Pandas. <http://pandas.pydata.org/>.
- [2] Scrapy. <https://scrapy.org/>.
- [3] MILLER, S. Visualization of ted talks on tableau. [https://public.tableau.com/views/TEDTalks-IdeasWorthSharing/TED-IdeasWorthSpreading?:embed=y&:display\\_count=yesTED-IdeasWorthSpreading?:embed=y&:display\\_count=yes&:showVizHome=no](https://public.tableau.com/views/TEDTalks-IdeasWorthSharing/TED-IdeasWorthSpreading?:embed=y&:display_count=yesTED-IdeasWorthSpreading?:embed=y&:display_count=yes&:showVizHome=no).
- [4] PAPPAS, N. Ted dataset. <https://www.idiap.ch/dataset/ted>.