

Related TED Talks Recommendations Using Network Visualizations According to Topic Features

Mingjie Zeng, Yufei Lin

1 Introduction

Network graphs can show interconnections between a set of entities to a certain extent. Therefore network-based diagrams are very effective for the purpose of showing the correlations between data. In addition, we observe that when watching a video, users will want to find simultaneously other videos with strong content relevance in order to have a more systematic content input. This demand will be stronger when watching intellectual or general scientific kinds of videos. So we were wondering if we could combine network-based graphs which reflect the relationships between different data entities and TED talks video recommendations and use it for pushing relevant videos? Further, perhaps a visualization of the TED talks "knowledge graph" could be constructed.

In addition, the results of the network graph may differ when using different features. The themes associated with the talks are other important features that can be used to build a video recommendation network. The number of first level comments made on the talk, the number of views on the talk, the duration of the talk in seconds, the Unix timestamp for the publication of the talk on TED.com, and even the main speaker of the talk will have an impact on the prioritization of video relevance. We also hope to further investigate the impact of these features on the final visualization presentation.

In this project, to minimize the number of nodes in each network graph for faster rendering and better user experience, we will first make a preliminary classification of the videos according to the tags involved and represent the specific content of the tags in the form of a bubble chart. Next, in each tag's category, list all the videos. After continuing to select a specific video, a network diagram of that video and the videos associated with that video will be displayed. By selecting a node which represents a video, rough information of that video, such as title, tag, author,

etc., will be displayed floating.

2 One-sentence description

Based on the TED talks dataset, we will display all the videos by tags, and furthermore, generate a small range of network graphs for each selected video to show the correlation with other videos and finally achieve the relevant video recommendations goal.

3 Project Type

Webpage / Application / Database

4 Audience

This project is useful for users who want to watch TED talks systematically, reducing the difficulty of selecting videos to a certain extent and recommending videos according to the relevance of the content. The network diagrams will show the relationship between the videos clearly to help users find the video they want.

This project will also graphically display an entire dataset in a combination of several visualization ways, which is also useful for people who want to explore data visualization.

5 Approach

5.1 Details

1) Data

- Dataset: Integrate multiple datasets into a single dataset containing all the required features.

- Data preprocessing: Pre-processing of data according to the form of data needed for the project, including extraction of tags, selection of relevant videos, etc.

2) Visualization

- Bubble chart: Represent all the videos by tags to classify videos by tags initially to break the whole dataset into several smaller datasets which will make the visualization more clear and refined.
- Network graph: When users finally select a video, a network graph of this video and its related videos will be generated. Here are two points: one is that we may need to dynamically generate a partial dataset to implement a network graph centered on a selected video, the other is that there are many different kinds of network graph and we need to find a suitable one. Our current preference is to use radial network.

5.2 Evidence for Success

- We found a relatively complete dataset, and although it requires a lot of processing, we will be able to get all the information we needed.
- The frameworks of bubble chart and network are provided by D3. We need to use them flexibly and make visualization as beautiful and intuitive as possible.

6 Best-case Impact Statement

In the best-case, we will implement a gorgeous website that will provide users with a TED talks video discovery service, allowing them to easily find a range of themed videos. Ideally, the users can interact with the visual elements and be directed to the URL where the video is located. In addition, the time to generate the network diagram should be small enough to provide a better experience for the users.

7 Major Milestones

- Correct data processing to obtain the right data information.
- Determine a way to extract the related videos information and make it a new dataset which generates the network graph.

- Determine the proper animations to use.
- Choose the suitable graphs and implement the visualizations properly.

8 Obstacles

8.1 Major obstacles

- Process the data correctly and get the right features we need.
- Proper implementation of web pages and visualizations.
- The right and comfortable interaction with data.

8.2 Minor obstacles

- Get the related video information using different features and explore the impact of the features to the network visualization results.
- Give the webpages and visualizations a good sense of design.

9 Resources Needed

- Code to integrate and process the dataset.
- Code to generate the related videos dataset after selecting a specific video for network graph generation.
- Some suitable web frameworks.

10 5 Related Publications

- Propose an interactive data repository with a web-based platform for visual interactive analysis in [4]. This paper and the built platform give us a sense of the network datasets and their visualizations to better know the structure of the network graphs.
- A tutorial for simple algorithms for network visualization [3] from which we can systematically learn how to implement the network visualization.
- In this paper, we get a sense of network sampling to avoid massive data [1]. For a webpage application, it's very important to reduce the users' waiting time for the graph generation. Turning static graphs into streaming graphs can be an effective way.

- Animated exploration of graphs with radial layout [5]. Our current preference is to use radial network to present the data and this paper concludes some basic knowledge about radial layout which broadens our ideas.
- User-guided matrix reordering [2] shows the relationship between different matrix reordering algorithms and the resulting network graphs. Different used algorithms would cause very inconsistent graphs. This is an inspiration about the assumption that choosing different features for network data construction will cause the generation very different.

11 Define Success

Our project will be successful if we establish the web-pages and the visualizations properly display. And users should successfully find the video network graph and be directed to the original website of the video.

References

- [1] N. K. Ahmed, J. Neville, and R. Kompella. Network sampling: From static to streaming graphs. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 8(2):1–56, 2013.
- [2] M. Behrisch, T. Schreck, and H. Pfister. Guiro: user-guided matrix reordering. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):184–194, 2019.
- [3] M. J. McGuffin. Simple algorithms for network visualization: A tutorial. *Tsinghua Science and Technology*, 17(4):383–398, 2012.
- [4] R. Rossi and N. Ahmed. The network data repository with interactive graph analytics and visualization. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.
- [5] K.-P. Yee, D. Fisher, R. Dhamija, and M. Hearst. Animated exploration of graphs with radial layout. In *Proc. IEEE InfoVis 2001*, pages 43–50, 2001.