可视化

一、想实现的功能

1.一张Echarts柱状图，显示TED视频中观看量top 10的视频，并在点击柱体后跳转到相应的视频播放界面；

2.一张Echarts柱状图，显示TED视频中出现次数top 10的话题；

3.一张Gephi图，以所有话题为节点，将话题出现的总次数作为节点的值。若两个话题出现在同一个视频中，则认为两个相应节点之间有边，每出现一次，边的权重加一。展现出各个话题之间关系的强弱和所属的不同社区；

4.一张Echarts图，以所有话题为节点，将所有权重不小于4的边展示在图中，将鼠标放置在某个节点上时，可以显示出与该节点相关的所有节点。

二、实现过程的重难点

1.如何实现在点击柱体之后跳转相应播放界面；

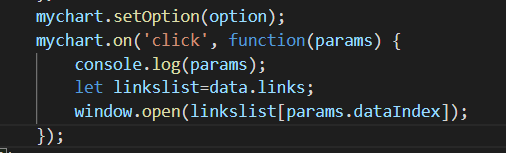
2.如何调整过滤信息使得Gephi图的节点信息不相互干扰；如何分区；如何体现节点之间的关系强弱；

3.如何获得边的权重；

4.Echarts节点关系图如何将节点和边的数据导入，具体地说，如何给出节点的坐标信息（因为节点原本没有坐标属性）；如何控制节点大小使得节点之间不相互遮盖；如何设置参数使得边不至于太过密集以至于遮挡节点。

三、解决方法

1.在放置播放量top 10的视频的echarts\_views.html中添加函数mycharts.on()用于打开新的播放界面，在views.json文件中添加一个属性links，里面放置前十名的链接。代码如下：



2.防止相互干扰的方法：

在Gephi中使用Force Atlas布局，将斥力强度调整到6000，选中“由尺寸调整”，过滤时采用范围里度的过滤，过滤掉度在30以下的节点，防止节点由于太过密集而重合；将节点的值作为节点相对大小的衡量标准，最小尺寸是15，最大尺寸是50；在预览设置里选中显示节点标签和缩短标签，防止标签之间相互遮盖；将边的厚度调整到0.1，透明度200.0，选中“弯曲“，防止由于边分布太密集或者边颜色太重导致节点信息被遮挡。

分区：

在统计时计算模块化，给节点着色时使用Modularity Class，软件会自动给出颜色和分区，可以调整区块的颜色防止颜色冲突；

显示关系强弱的方法：

在预览设置中边的栏目中颜色设置为mixed，显示出的边的颜色是它所连接的两个节点的颜色的过渡，显示出了节点之间的远近关系。

3.获得边的权重的方法：

新建sorted\_data.py文件，用python方法获取数据库数据进行筛选。

4.节点和边的数据导入：

节点属性有：x, y, id, name, symbolSize, itemstyle(color); 边的属性有：sourceID, targetID, weight.

节点并没有坐标属性，为了赋予节点坐标属性，根据数学规律，用库函数生成两组服从正态分布的随机数，分别作为节点的横纵坐标；

节点的id是按照出现次数降序排列后的序号；

节点的name是话题的名称；

节点的symbolSize是节点的值；

节点的itemStyle的color属性是按照colorList循环设置的；

边的sourceID和targetID都是节点的id；

边的weight是边本身的weight；

在图中，节点按照给出的横纵坐标在画布上显示；节点的标签内容是节点的name；结点的大小是节点的symbolStyle；

参数的设置：

节点的symbolStyle不等于它的值，而是值统一整除15，可以有效防止节点太大导致重合；

正态分布的坐标的参数为期望为0，方差为1000000，这组参数可以有效分散节点，坐标数量为节点数量348个；

将结点的标签位置设置成置右，防止节点标签重合。

输入数据时进行筛选，剔除权重小于4的边。

Visualization

I. Functions I want to achieve

1. An Echarts bar chart showing the top 10 most viewed videos in the TED videos and jumping to the corresponding video playback screen when the bar is clicked.

2. An Echarts bar chart showing the top 10 topics in the TED video in terms of number of appearances.

3. A Gephi chart with all topics as nodes and the total number of times the topic appears as the value of the node. If two topics appear in the same video, the two corresponding nodes are considered to have edges between them, and the weight of the edge is added by one for each occurrence. Exhibit the strength of the relationships between the topics and the different communities to which they belong.

4. An Echarts node-relationship graph with all topics as nodes, showing all edges with a weight of not less than 4 in the graph, and showing all nodes related to a node when the mouse is placed over it.

II. The implementation process of the important and difficult points

1. How to realize jumping to the corresponding play screen after clicking on a bar.

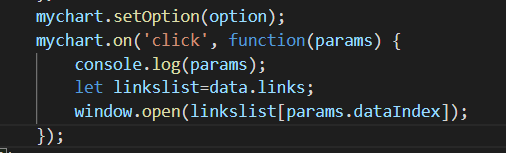
2. How to adjust the filter information so that the node information of the Gephi diagram does not interfere with each other; how to partition; How to reflect the strength of the relationship between the nodes.

3. How to obtain the weights of the edges.

4. How to import the data of nodes and edges into the Echarts node-relationship graph, Specifically, how to obtain the coordinate information of the nodes (since they do not have coordinate properties); how to control the size of the nodes so that they do not cover each other; how to set the parameters so that the edges are not so dense that they cover the nodes.

III. Solutions

1. Add the function mycharts.on() to the echarts\_views.html where the top 10 videos graph are placed, and add a property links in the views.json file where the top 10 links are placed. The code is as follows.



2. Methods to prevent interferences:

Use the Force Atlas layout in Gephi, adjust the repulsion strength to 6000, check "Adjust by size", filter by the degree of the range, and filter out nodes with a degree below 30 to prevent nodes from overlapping due to too dense; use the value of the node as a measure of the relative size of the nodes, the minimum In the preview settings, choose Show node labels and Shorten labels to prevent labels from covering each other; adjust the thickness of the edges to 0.1 and the transparency to 200.0, and choose Bend to prevent node information from being obscured due to too dense a distribution of edges or too heavy an edge color.

Partitioning.

Calculate modularity in statistics, use Modularity Class when coloring the nodes, the software will automatically give the colors and partitions, you can adjust the colors of the blocks to prevent color clashes.

Method to show relationship strength:

The color is set to mixed in the column of edges in the preview settings, the color of the displayed edge is a transition of the color of the two nodes it is connected to, showing the distance between the nodes.

3. Method of obtaining edge weight:

New sorted\_ data.py file, using Python method to obtain database data for filtering.

4. Data importation of nodes and edges:

Node attributes include: x, y, ID, name, symbolSize, itemStyle (color); edges attributes include: sourceID, targetID, weight.

Node has no coordinate attribute. In order to give node coordinate attribute, two groups of random numbers obeying normal distribution are generated by library function according to mathematical law, which are used as the horizontal and vertical coordinates of node respectively;

The node id is the sequence number arranged in descending order according to the occurrence times;

The name of the node is the name of the topic;

The symbolSize of the node is the value of the node;

The color attribute of the itemStyle of the node is set according to the colorList loop;

The sourceIF and targetID of the edge are the ID of the node;

The weight of the edge is the weight of the edge itself;

In the graph, the nodes are displayed on the canvas according to the given horizontal and vertical coordinates; the label content of the node is the name of the node; the size of the node is the symbolSize of the node;

Parameter setting:

The symbolSize of a node is not equal to its value, but is equal to the unified division of the value by 15, which can effectively prevent coincidence caused by too large nodes;

The parameter of normal distribution coordinates is expected to be 0 and the variance is 1000000. This group of parameters can effectively disperse nodes, and the number of coordinates is 348 nodes;

Set the label position of the node to the right to prevent the node labels from overlapping.

Filter the input data and eliminate the edges with weight less than 4.