Part2页面演示操作说明文档：

1. 准备环境（Eclipse + Mysql + Tomcat）
2. 在 Eclipse中加入Tomcat插件
3. 数据库中建表（执行给定的sql语句）
4. 修改项目的配置文件 c3p0-config.xml文件

<property name=*"jdbcUrl"*>jdbc:mysql://127.0.0.1:3306/XXXX</property>

<property name=*"user"*>root</property>

<property name=*"password"*>\*\*\*\*\*\*\*</property>

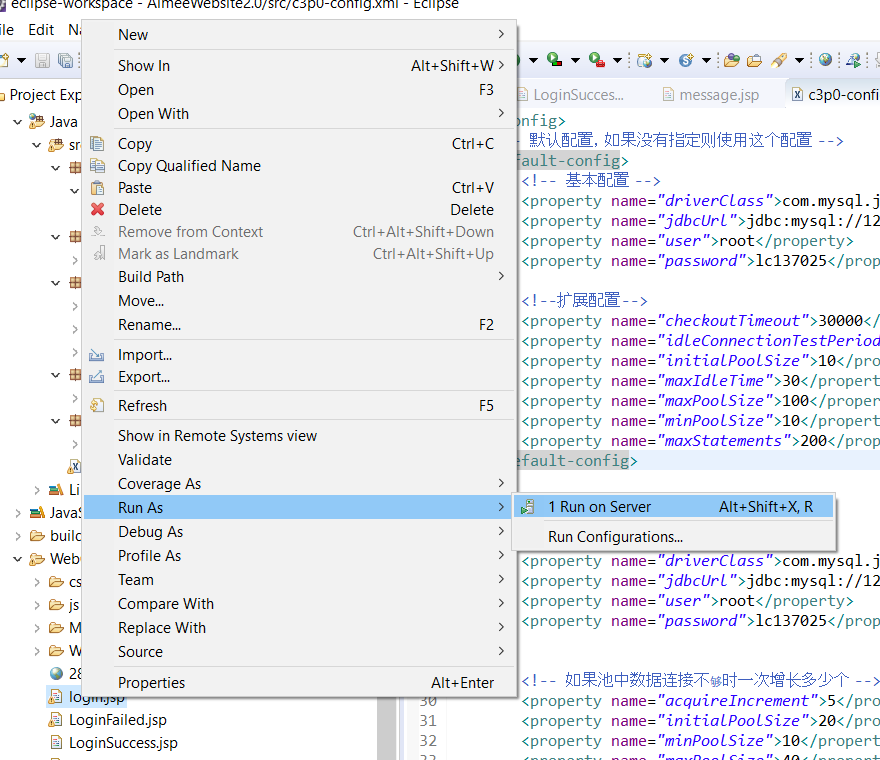
XXXX表示自己的数据库名字，user和password改为自己的用户名及密码。

1. 点击login.jsp文件, 在服务器上运行，即可以转到登陆界面。

数据库中提前插入的两条测试数据为：

用户名：tom 密码：123

用户名：sad 密码：asd



1. 其他界面同理。

Part1部分：(提取《釜山行》人物关系网络图谱)

背景：《釜山行》是一部丧尸灾难片，其人物少、关系简单，非常适合学习文本处理。本项目使用python编写代码实现对《釜山行》文本的人物关系提取，最终利用Gephi软件对提取的人物关系绘制人物关系图。实体间的共现是一种基于统计的信息提取。关系紧密的人物往往会在文本中多段内同时出现，可以通过识别文本中已确定的实体(人名)，计算不同实体共同出现的次数和比率。当比率大于某一阈值，我们认为两个实体间存在某种联系。

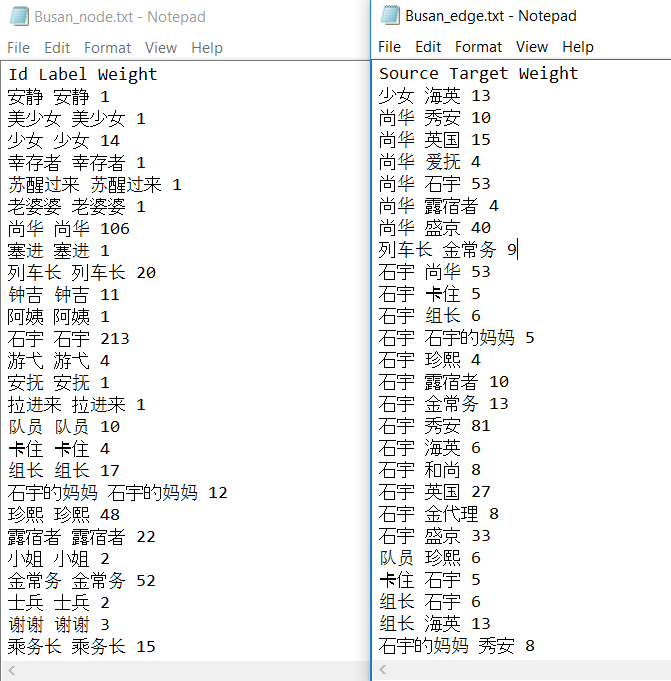
* 数据来源的说明

1. 网络下载《釜山行》的中文剧本Busan.txt
2. 网络下载字典 dict.txt

完整代码如下：

*# !/usr/bin/python  
# coding:utf-8***import** jieba  
**import** codecs  
**import** jieba.posseg **as** pseg  
  
  
names **=** {} *# 姓名字典*relationships **=** {} *# 关系字典*lineNames **=** [] *# 每段内人物关系  
  
# count names*jieba.load\_userdict("C:\Users\I332487\Desktop\dict.txt") *# 加载字典***with** codecs.open("C:\Users\I332487\Desktop\Busan.txt", "r", "utf-8") **as** f**:  
 for** line **in** f.readlines()**:** poss **=** pseg.cut(line) *# 分词并返回该词词性* lineNames.append([]) *# 为新读入的一段添加人物名称列表* **for** w **in** poss**:  
 if** w.flag **!=** "nr" **or** len(w.word) **<** 2**:  
 continue** *# 当分词长度小于2或该词词性不为nr时认为该词不为人名* lineNames[**-**1].append(w.word) *# 为当前段的环境增加一个人物* **if** names.get(w.word) **is** None**:** names[w.word] **=** 0  
 relationships[w.word] **=** {}  
 names[w.word] **+=** 1 *# 该人物出现次数加 1  
  
# explore relationships***for** line **in** lineNames**:** *# 对于每一段* **for** name1 **in** line**:   
 for** name2 **in** line**:** *# 每段中的任意两个人* **if** name1 **==** name2**:  
 continue  
 if** relationships[name1].get(name2) **is** None**:** *# 若两人尚未同时出现则新建项* relationships[name1][name2]**=** 1  
 **else:** relationships[name1][name2] **=** relationships[name1][name2]**+** 1 *# 两人共同出现次数加 1  
  
# output***with** codecs.open("C:\Users\I332487\Desktop\Busan\_node.txt", "a+", "utf-8") **as** f**:** f.write("Id Label Weight**\r\n**")  
 **for** name, times **in** names.items()**:** f.write(name **+** " " **+** name **+** " " **+** str(times) **+** "**\r\n**")  
  
**with** codecs.open("C:\Users\I332487\Desktop\Busan\_edge.txt", "a+", "utf-8") **as** f**:** f.write("Source Target Weight**\r\n**")  
 **for** name, edges **in** relationships.items()**:  
 for** v, w **in** edges.items()**:  
 if** w **>** 3**:** f.write(name **+** " " **+** v **+** " " **+** str(w) **+** "**\r\n**")

运行得到节点集合Busan\_node.txt，边集合Busan\_edge.txt，如下图：



* 图中结点、连线等说明

字典类型names保存人物，该字典的键为人物名称，值为该人物在全文中出现的次数。

字典类型relationships保存人物关系的有向边，该字典的键为有向边的起点。

值为一个字典edge，edge的键是有向边的终点，值是有向边的权值，代表两个人物之间联系的紧密程度。

即：图中的节点代表的是人物，连线代表两个人物之间联系的紧密程度。权值越大，联系越紧密，连线越明显。

* 可视化工具（Gephi）的简要操作流程分析

1）文件导入

点击菜单栏中的“文件”->“打开”后即可输入选择的文件，支持的文件类型有很多，可以在“文件类型”中选择，输入文件后产生一个输入报告，报告中有关于节点和边的信息等。点击输入报告中的“确定”后，产生一个初始图像。

若要从数据库中导入，则选择“文件”->“输入数据”->“边名单”。 若要随机生成一个随机图，则选择“文件”->“生成”->“随机图”，可 以输入点数和连线的概率。

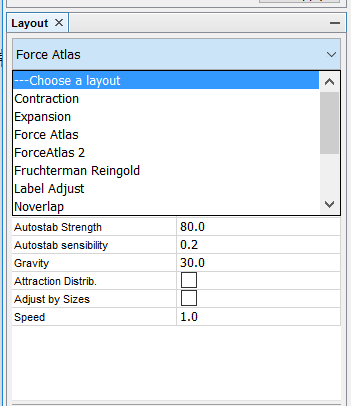
2）可视化操作

可以滚动鼠标滑轮，对图像进行放大缩小，点击鼠标右键可以将图形进行拖动。

3）布局/流程

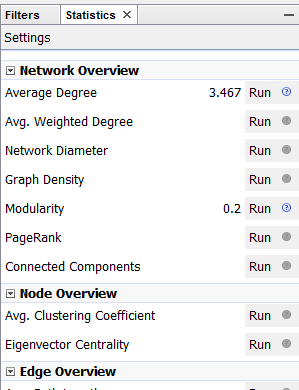
可以选择下拉框中的 12 种布局方式，前六种是主要布局工具，后面六种是辅助布局工具。

选择一种布局算法，点击运行即可看到布局效果。最常用的是力导向算法 （Force Atlas 和 Forceatlas 2）、圆形布局和胡一凡布局。



4）统计

图的特征可在统计功能模块中计算得到，其模块如下图。还可以计算相应的图的特征数值，如要查看详细内容，可单击“问号”图标产生相应的报告。



5）分割

分割也是一种归类，把值相同的节点或边用不同的颜色标示出来，还可以把值相同的节点组合成一个节点。

6）过滤

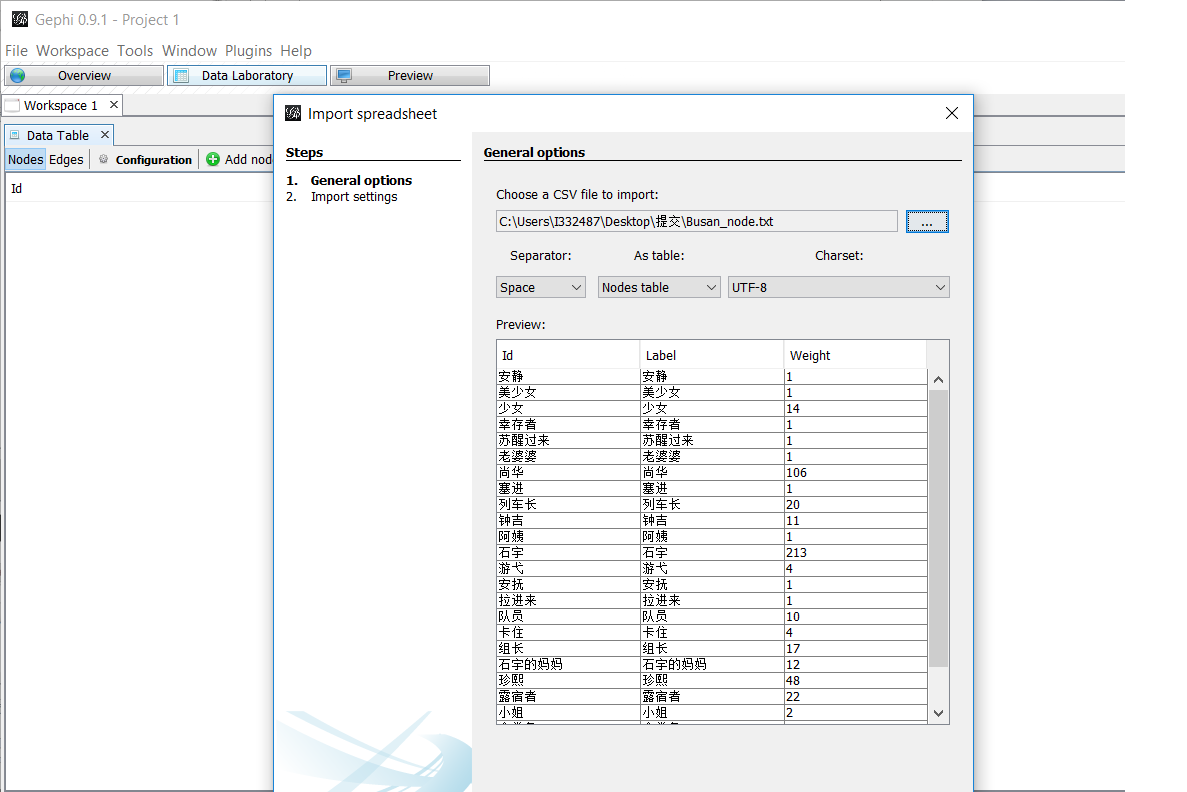
在作图过程中经常需要把一些值相同的节点或边选择出来，此时需要用到过滤工具，通过过滤功能实现选择或者将符合条件的节点和边过滤出来。

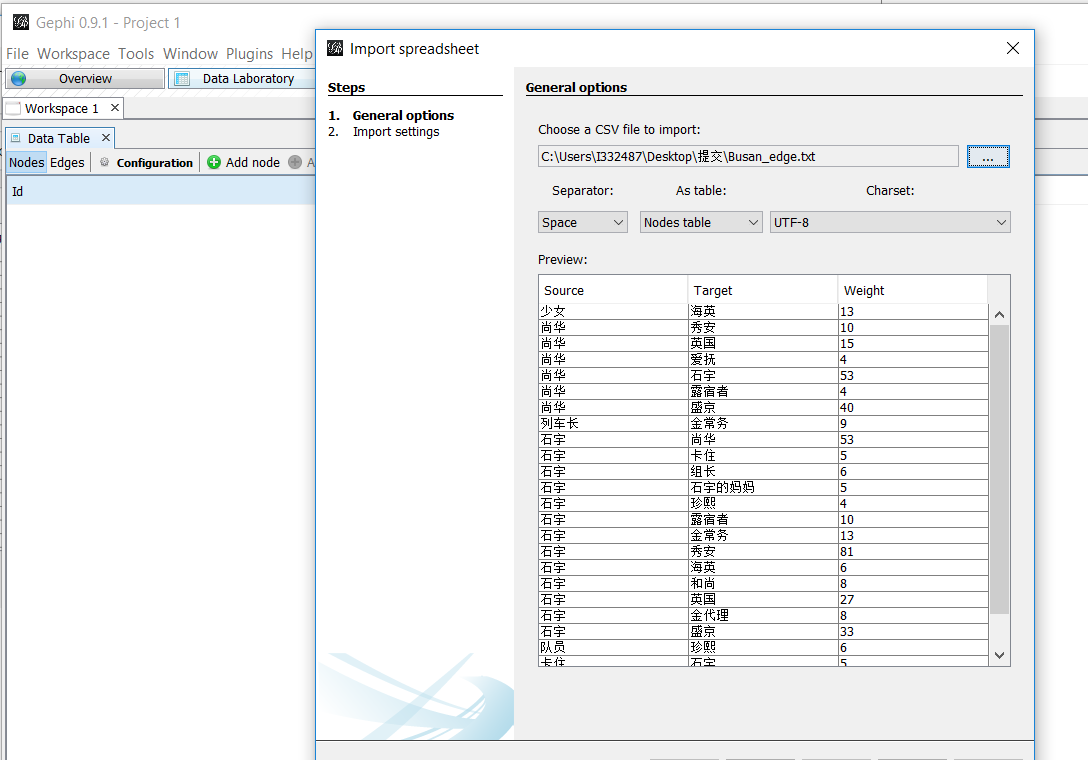
7）预览

预览是输出控制的环节，在预览界面可以对前面编辑的图形做最后的美化，包括图形外观样式的选择和显示细节的调整，之后便可导出图形。

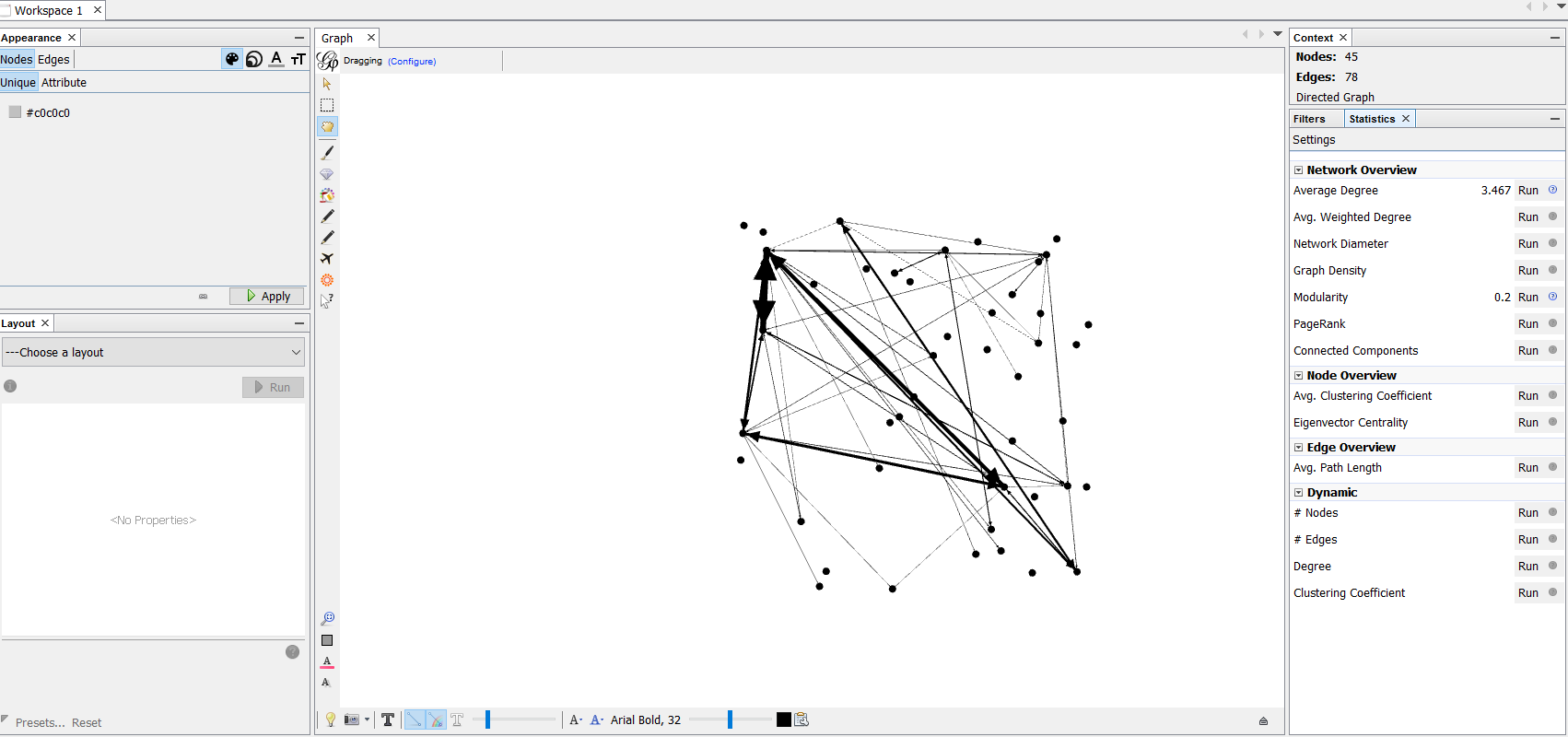
* 重要步骤及结果的截图

Step1：选择节点集合Busan\_node.txt，边集合Busan\_edge.txt，将两个文件分别导入Gephi, 分隔符选择空格，编码选择UTF -8。

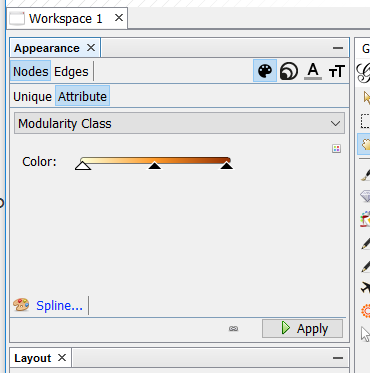




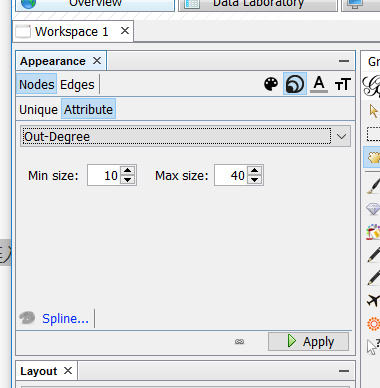
Step2: 在最上方的数据资料选项卡中查看图中所有的边和节点，对于分词不准确导致的噪音可以手动删除。分别点击右侧统计栏中平均度和模块化运行计算。模块化运算时Resolution值填写0.5。



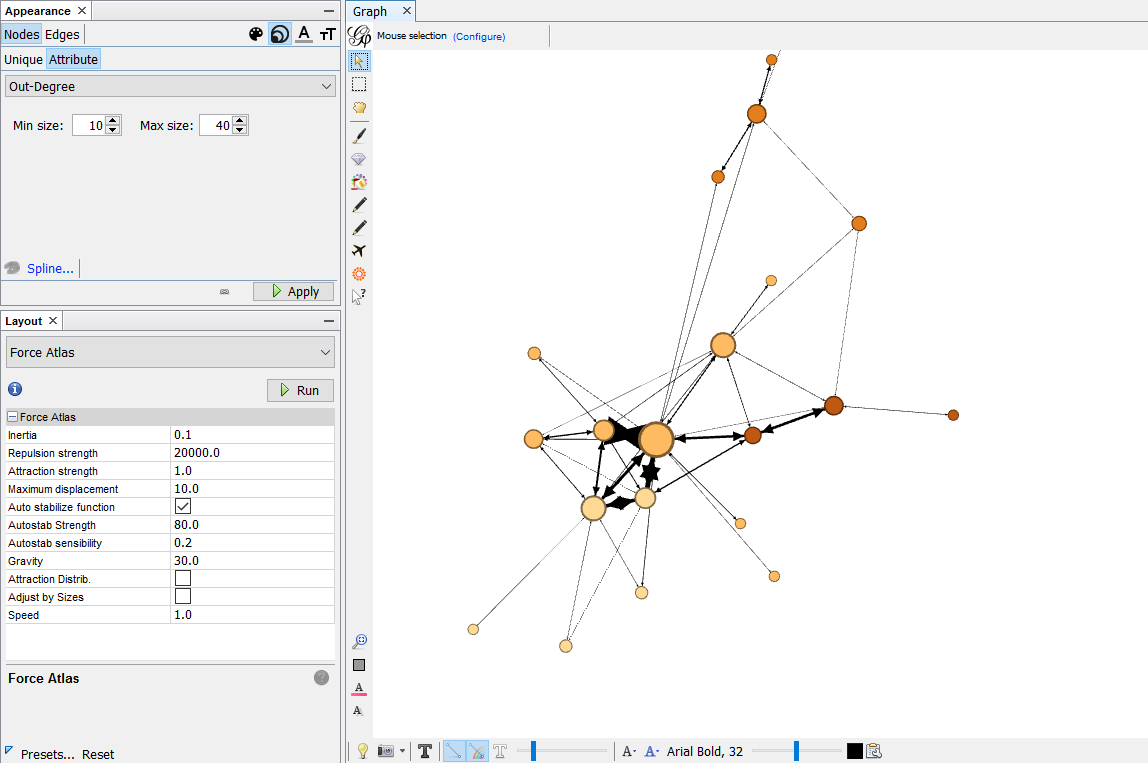
Step3: 点击左上角外观中节点第一个选项卡，选择Attribute，选择Modularity Class，点击Apply。



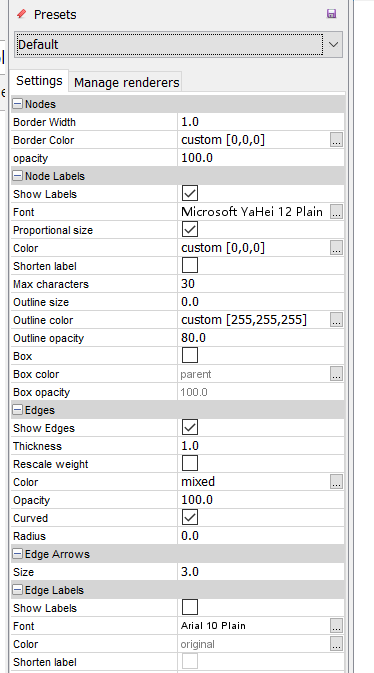
Step4: 点击左上角外观中节点第二个选项卡，选择Attribute，选择Out-Degree(连入度)，Min size填10，Max size填40，点击应用。



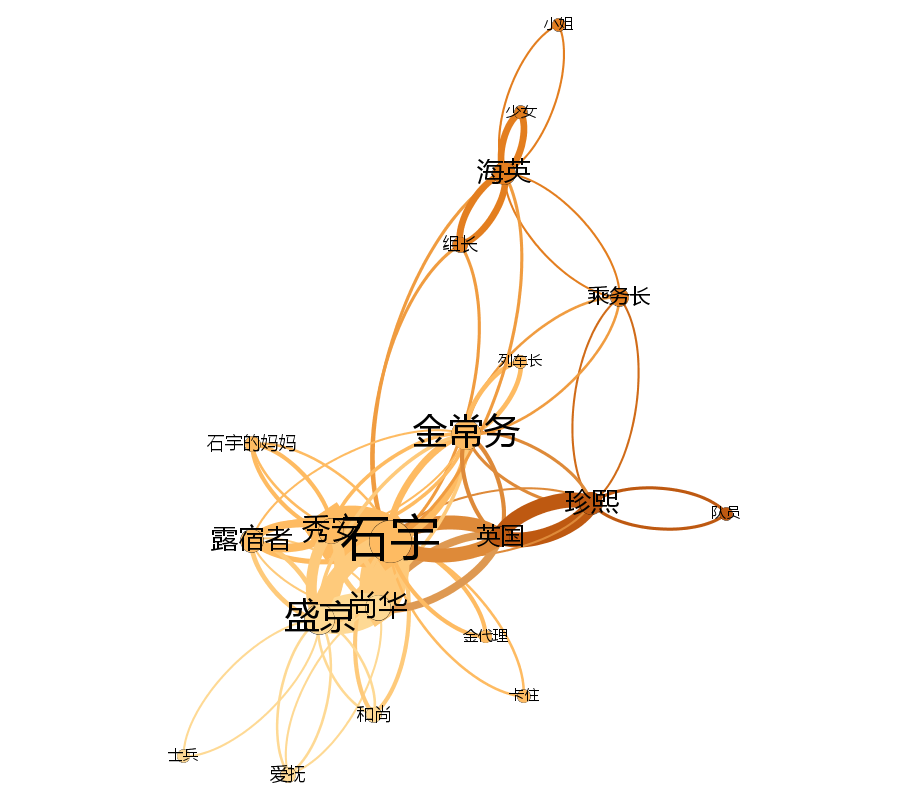
Step5: 选择左下角布局中的Force Atlas，斥力强度填写20000.0，吸引强度填写 1.0。点击运行，稍后点击停止。



Step6: 点染色根据模块化计算结果不定，但染色效果大致相同。点击最上方的预览按钮，选中左侧节点标签中显示标签选项，并选择一种字体。



Step7: 点击刷新按钮，就可以显示最终的人物关系图。



* 图谱结果的分析

对于没有看过电影的人来说，通过以上的图谱，可以很清楚的看到这部电影的主要角色以及各个角色之间的关系。

我们可以大致猜测出主要人物有：石宇、秀安、盛京、尚华、英国、珍熙、金常务。