你所在的专业“火”吗

姓名：于沁涵 学号：1910227

摘 要

目前很多研究人员认为转专业申请情况与专业热度挂钩，因此研究转专业申请数据可以有效的帮助他们研究大学专业热度情况。

本文旨在研究某大学校内专业热度情况，考虑利用已有的相关数据 2018、2019、2020 学校转专业情况对数据进行描述性分析，基于 2013 级毕业去向率研究转入专业和转出专业的差异，并通过 PageRank 算法进行排序，得到专业热度排序。研究得到：转入专业和转出专业差异化不明显，专业热度较火的专业有经济学类、法学、计算机、数据科学与大数据技术，与目前社会需求相符。

一、背景介绍

“天道酬勤。”这是孩子们从小就经常听到的话，大人们都说，“要好好学习，考上一所好大学，将来才能出人头地！”直到孩子们真的到了十八岁的年华，才知道要面临的事不止“考上一所好大学”那么简单。

高考结束的那个假期漫长且充实，除了尽情的享受假期，高三毕业生面临的第一件事就是接受自己的高考成绩，然后填报志愿。看似简单，实则让不少家庭焦头烂额。有的同学早已想好自己感兴趣的东西以及要学习的专业，因此不费吹灰之力，根据专业的排名，结合高考分数，很快的选好了自己的志愿；有的同学却没什么想法，对大学专业的未知充满了恐惧和好奇，因此，选择听从爸妈的安排；有的同学则十分严谨，他们考虑各种各样的因素，学校的排名、地理位置、专业的前景、住宿条件等，因此，在他们看来，填报志愿简直就是“折磨”。好在最终，同学们都选好了各自的志愿，等待结果到来的那一刹。

然而，很多同学上了大学后发现，自己学的专业和曾经自己幻想中的专业天差地别，可能自己擅长的学科在大学也变成了不擅长的学科，或者是在录取学校的时候意外的被调剂到了自己不感兴趣的专业，因此，很多同学在进入大学学习的时候都十分痛苦。幸运的是关于兴趣和前途的选择并不只有这一次机会，在大学期间，同学们还可以通过转专业前往自己真正感兴趣的专业，但是这样也会有一定的风险，有可能转完专业后仍然不是心中喜欢的，并且转专业需要把之前的课都学习一遍，会增加同学们的课业压力。因此，在转专业中，同学们会考虑更多的因素来进行决策。这时候，研究转专业申请数据可以帮助我们解决很多相关方面的内容，进一步研究校内专业热度情况。

本文旨在研究校内专业热度情况，考虑利用已有的相关数据 2018、2019、2020 学校转专业情况对数据进行描述性分析，并通过 PageRank 算法进行排序，得到最终结论。

二、数据来源和说明

本文数据来自 2018、2019、2020 年某大学转专业申请情况以及 2013 年各专业毕业生去向比率，数据分别有 363、459、279 以及 66 个。其中 2018 年转专业申请包括 5 个变量、2019 年转专业申请包括 8 个变量、2020 年转专业申请包括 5 个变量、2013 年各专业毕业生去向比率包括 5 个变量，涉及的变量包括年级、所在院系、所在专业、转入院系、转入专业、ABC、ABCD、ABCDE、出国出境率、国内升学率、就业率。具体变量说明如表 1:

表 1 数据变量说明

变量名称	备注
年级	学生入学时的年份，例如 2018 级表示 2018 年入学的学生
所在院系	学生所在的学院
所在专业	学生所就读的专业，即转出专业
转入院系	学生即将转入的专业所在的院系
转入专业	学生即将转入的专业
ABC	学生修过的公共、专业必修课的平均成绩和排名
ABCD	学生修过的公共、专业必修课以及专业选修课的平均成绩和排名
ABCDE	学生修过的所有已上过的课程的平均成绩和排名
出国出境率	毕业后出国的学生占所在专业的比例
国内升学率	毕业后国内读研的学生占所在专业的比例
就业率	毕业后直接工作的学生占所在专业的比例

三、描述性分析

（一）按年分析各专业转入转出情况

首先，我们对各专业申请转入人数情况，各专业申请转出人数情况进行分析。对 2018、2019、2020 三年各专业转入和转出情况绘制柱状图，如图 1、2、3 所示:

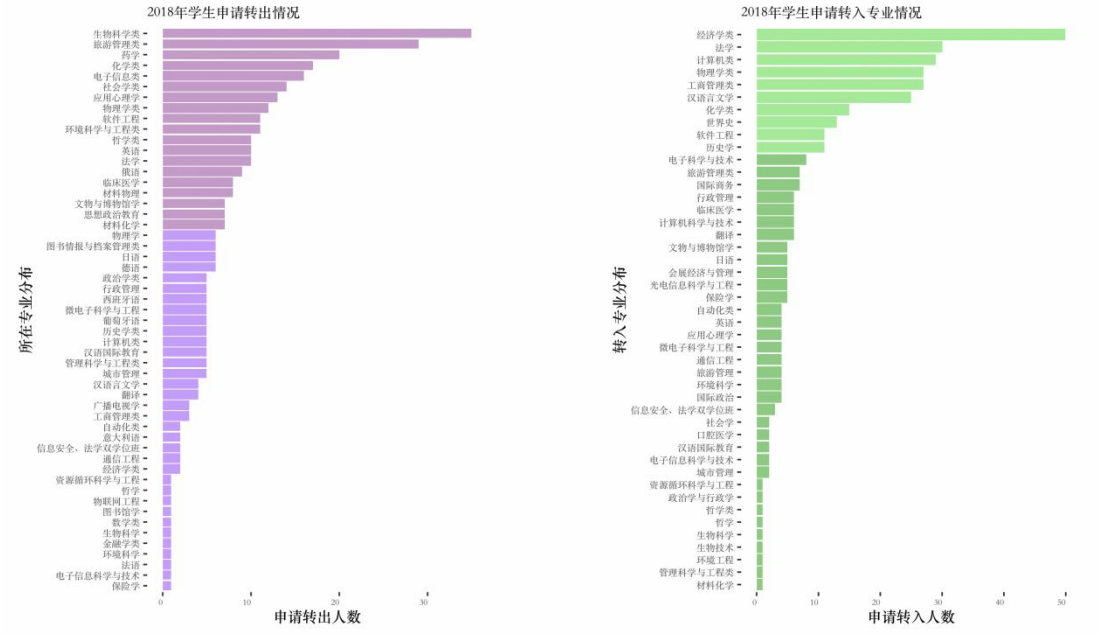

图 1 2018 年各专业转入和转出情况

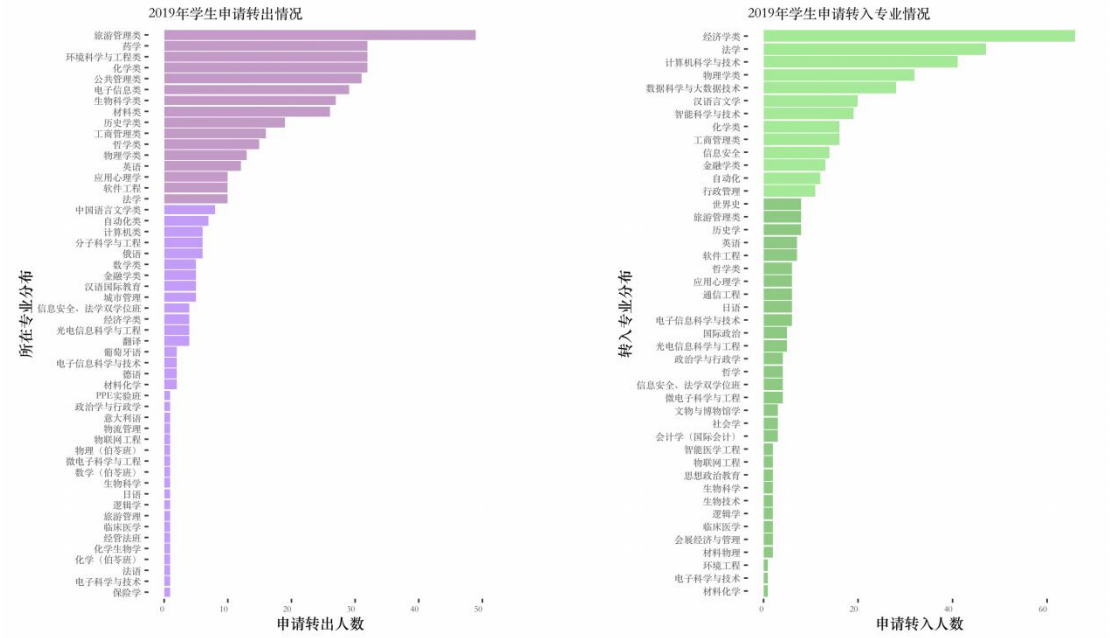

图 2 2019 年南各专业转入和转出情况

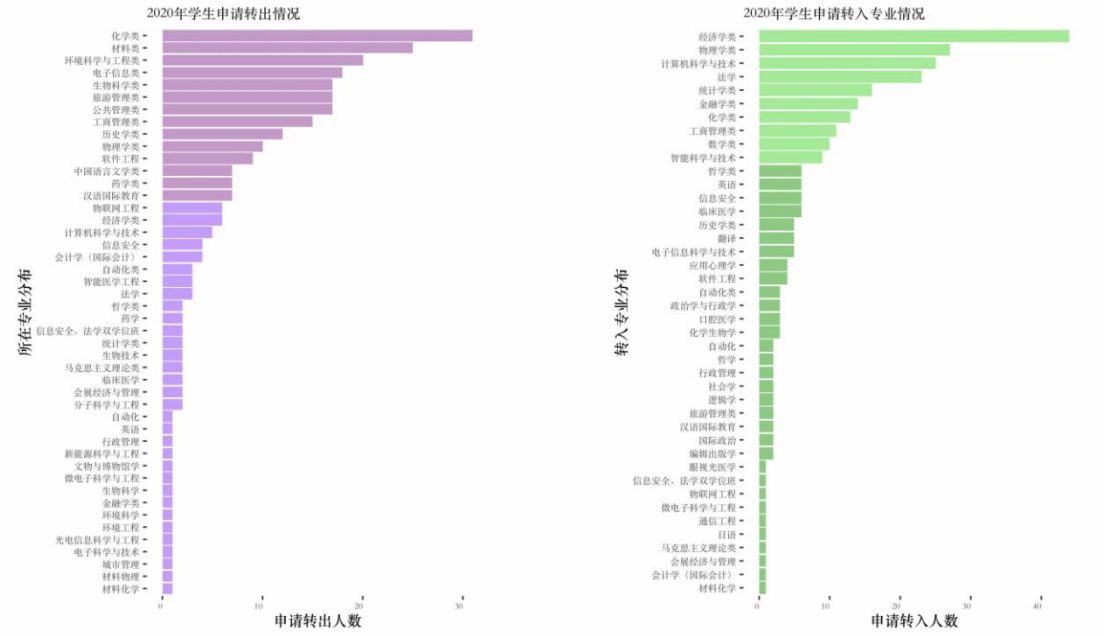

图 3 2020 年各专业转入和转出情况

从三张图中可以看出，2018 年转出专业最多的前三名是生物科学类、旅游管理类以及药学，转入专业最多的分别是经济学类、法学以及计算机类；2019 年中旅游管理类和药学专业仍在转出专业中位居前三，而转入专业最多的依然是经济学类、法学和计算机科学与技术；2020 年的转出情况有所变化，化学类成为转出人数最多的专业，而转入专业最多的仍然是经济学类。

由此可以看出，转入和转出专业的情况在这三年整体波动并不是很大，流失人数较多的专业基本在旅游管理、药学、环境科学与工程、化学之间，这可能是因为这几个专业的高考录取分数线较其他专业略低，因此很多学生们被调剂到了这专业，同时，专业的就业前景也不如一些热门专业好，因此很多学生想转出。

在转入专业中，经济学类是三年的卫冕冠军，法学和计算机类专业也都一直名列前茅，这是因为经济学、法学、计算机类专业比较符合当今社会的就业发展，因此在近几年成为炙手可热的热门专业，同时这些学院基数较大，因此招收的人数也比较多。

此外，一些新兴专业的崛起也是可圈可点的。2019 年，数据科学与大数据技术专业的增设让此专业一跃成为第 5 名，而后的 2020 年，同一院系的统计学类专业转入人数仍然保持在第 5 名，转出人数也较少，这说明，统计学在学生心中也被列入了最喜欢的专业之一。

为了进一步对比三年内的各专业转入转出情况，我们将三年数据进行合并，分别画出转入和转出人数最多的 15 个专业的 stack 图，如图 4 和图 5。

由图 4 可以看出转入总人数最多的仍然是经济学类、法学专业。在 2019 年后经济学类、法学、物理学类、金融学类、工商管理类、汉语言文学类、行政管理专业转入的人数大幅度增加，而智能科学与技术、信息安全、数据科学与大数据技术、计算机科学这些在 2019 年后才有转入学生的专业在两年中均保持比

较良好的转入情况，而软件工程、化学类专业则相对稳定，三年内转入的人数基本没有变化。

对于转出情况图 5，转出总人数最多的仍然是旅游管理类，化学类位居第二，其中，2019、2020 两年各专业转出人数波动不大。

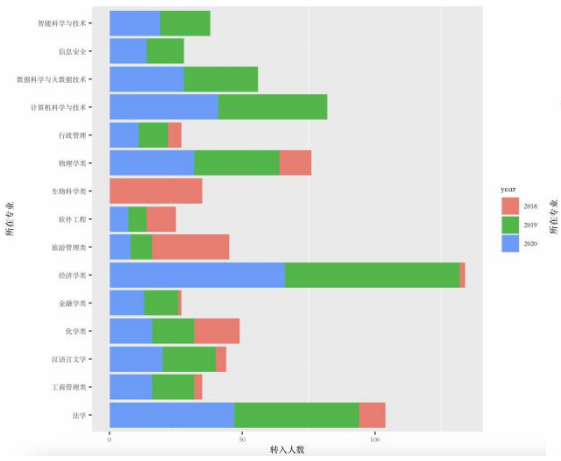

图 4 转入专业情况的 stack 图

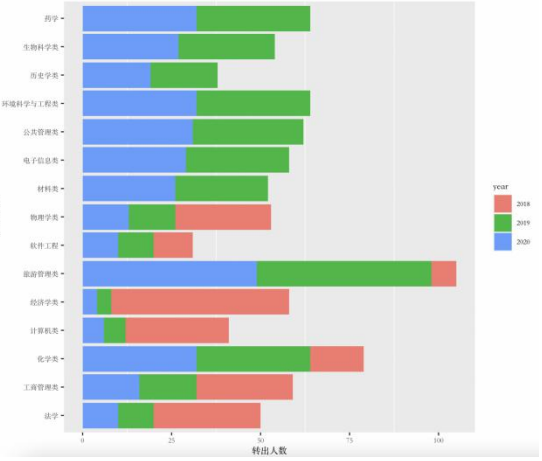

图 5 转出专业情况的 stack 图

（二）转出专业与转入专业在各指标上的差距

从上述转入转出情况来看，转专业可能和就业或者读研的前景有关，大学生一般毕业之后的去向有三种：国内读研、出国读研、就业，这与他们的前途牢牢挂钩。因此，分析毕业去向相关指标也是了解各专业情况的一个有效途径之一。根据 2013 级毕业生毕业去向比率数据，我们研究转出专业与转入专业在出国出境率、国内升学率、就业率三个指标在 2018～2020 年上的差异。

我们分别记录三年转专业数据中每个转专业学生所对应的转入专业和转出专业，将两个专业对应于 2013 年毕业去向率相应专业的指标，如果没有则舍弃，最终算出每个学生的转出专业和转入专业的出国出境率差、国内升学率差和就业率差，绘制出箱线图，比较三个指标的差距，得到结果图，如图 6～8。

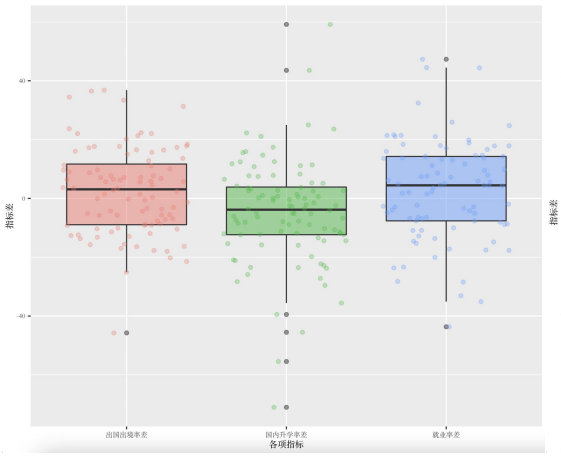

图 6 2018 年转入和转出专业指标差箱线图

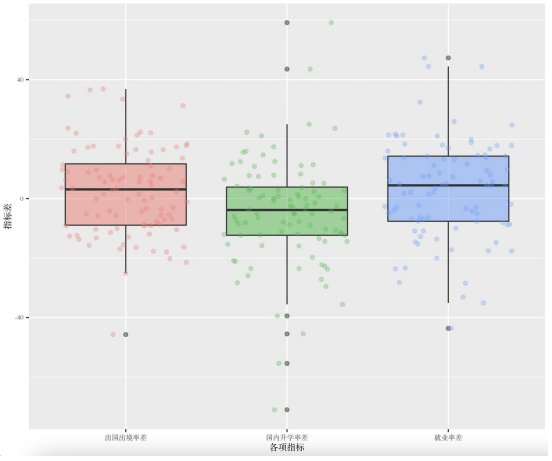

图 7 2019 年转入和转出专业指标差箱线图

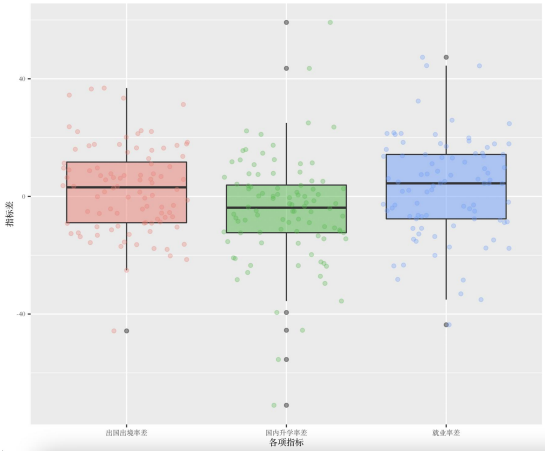

图 8　2020 年转入和转出专业指标差箱线图

可以看出，2018、2019、2020 年中出国出境率和就业率总体为正值，而国内升学率为负值，但基本都在 0 附近，这说明转入专业和转出专业在各指标上差异并不明显，但是转出专业在国内升学率上会高一些，这说明转专业的同学更多的去往了出国出境升学率和就业率较高的专业，从国内升学率较高的专业转出，但二者差异并不明显。

二、PageRank 算法

PageRank，网页排名，又称网页级别，是一种由根据网页之间相互的超链接计算的技术，而作为网页排名的要素之一，它简称为 PR，可以用来体现网页的相关性和重要性。事实上，PageRank 可以定义在任意有向图上，后来被应用到社会影响力分析、文本摘要等多个问题。

PageRank 算法的基本想法是在有向图上定义一个随机游走模型，即一阶马尔可夫链，描述随机游走者沿着有向图随机访问各个结点的行为。在一定条件下，极限情况访问每个结点的概率收敛到平稳分布，这时各个结点的平稳概率值就是其 PageRank 值，表示结点的重要度。PageRank 是递归定义的，它的计算可以通过迭代算法进行。

在本文中，我们可以通过 PageRank 算法研究各专业的热度，同理，以学院对标网页，有一个学生转入即表示该学院对相应学院有指向链接，以此建立专业向往矩阵 A。最终，利用 PageRank 算法进行排序，得到各专业的得分，比较各专业的热度。利用 R 语言，我们分年份进行分析，得到 2018 年、2019 年、2020 年的专业热度的前 15 名排序、文科专业排序以及理科专业排序，如图 9～11。在此，我们认为理科专业包括材料化学，材料物理，电子信息科学与技术，电子信息类，化学类，环境科学，环境科学与工程类，计算机类，金融学类，软件工程，生物科学，生物科学类，数学类，通信工程，微电子科学与工程，物理学，物理学类，物联网工程，信息安全、法学双学位班，药学，应用心理学，资源循环科

学与工程，自动化类；文科专业包括保险学，城市管理，德语，俄语，法学，法语，翻译，工商管理类，管理科学与工程类，广播电视学，汉语国际教育，汉语言文学，经济学类，历史学类，临床医学，旅游管理类，葡萄牙语，日语，社会学类，思想政治教育，图书馆学，图书情报与档案管理类，文物与博物馆学，西班牙语，行政管理，意大利语，英语，哲学，哲学类，政治学类，世界史，国际政治，社会学，会计学（国际会计），中国语言文学类，公共管理类，政治学与行政学，会展经济与管理，逻辑学，编辑出版学，马克思主义理论学。

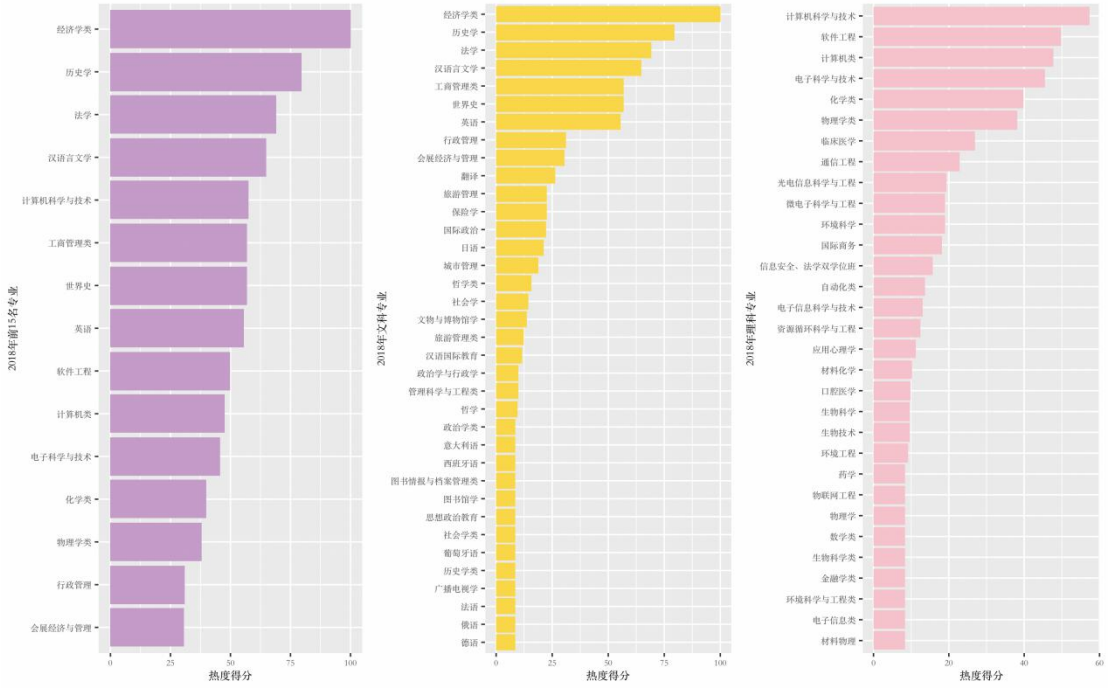

图 9　2018 年 Pageranke 算法下的专业热度图

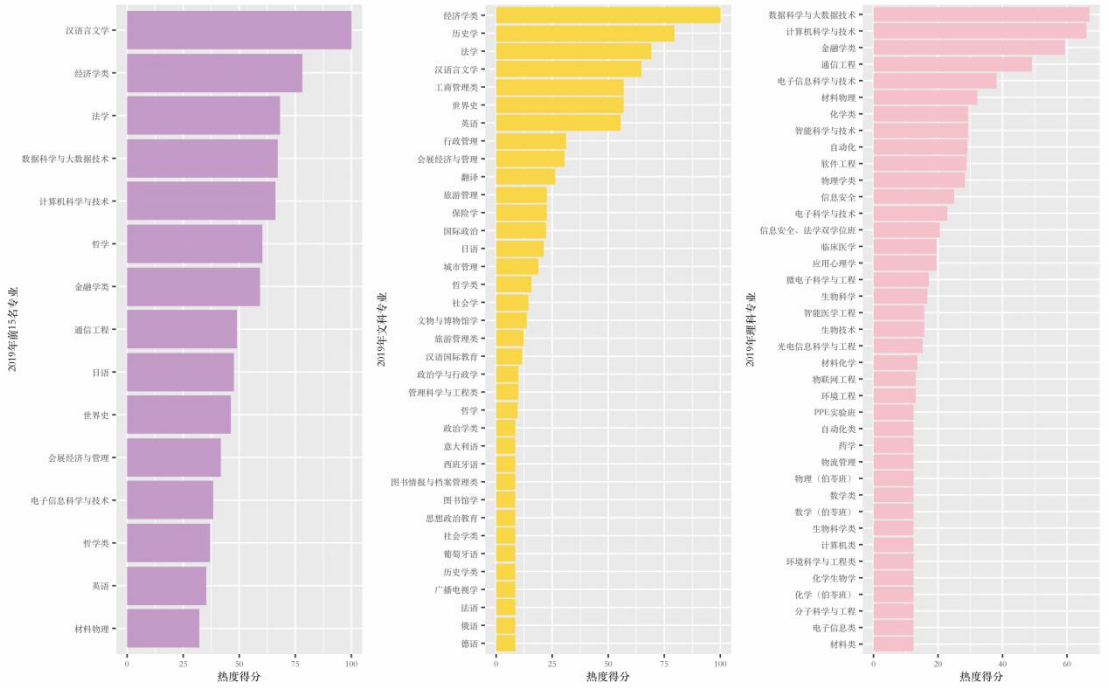

图 10　2019 年 Pageranke 算法下的专业热度图

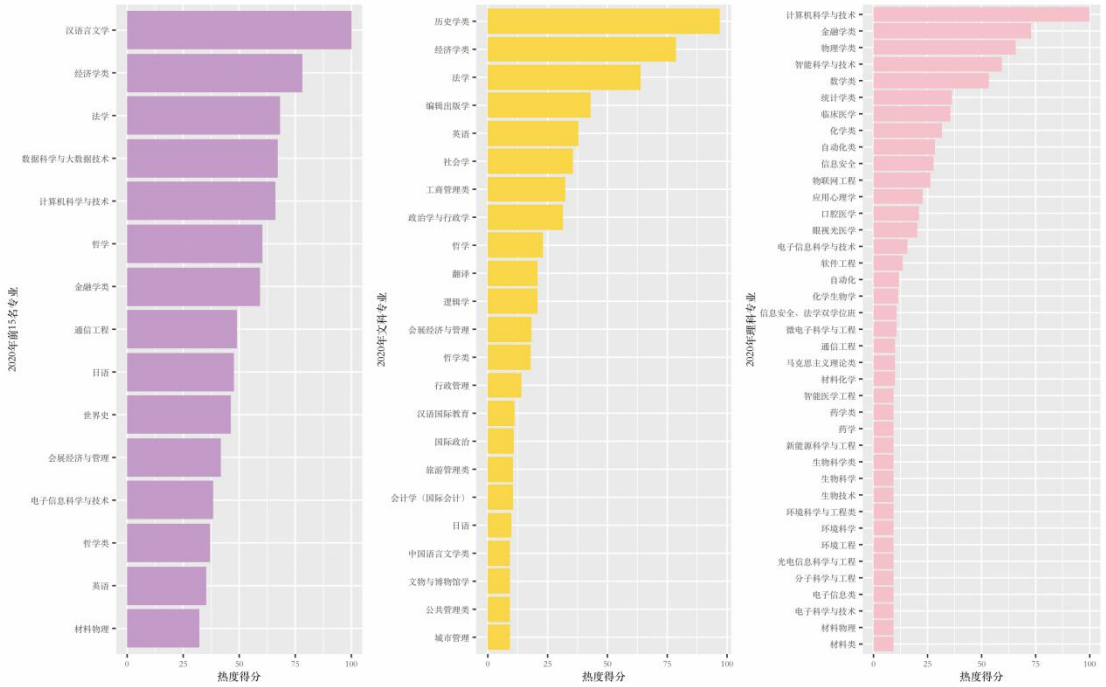

图 11　2018 年 Pageranke 算法下的专业热度图

由图 9～11 可以看出，2018、2019、2020 年中专业热度最高的几个都是文科类专业，例如经济学类、汉语言文学、法学等，而理科中计算机热度最高，2019 年后，数据科学与大数据技术新兴专业也变得火热起来，此外，世界史、哲学也保持较火热的地位。从文科分类图中看出，小语种专业（除英语）的热度相对较低，这是因为英语专业作为国际通用语言用途较广，而法语、德语、俄语等专业学习难度较大切前景不如经济、计算机等专业好，因此热度相对较低。而理科专业中，环境科学、化学、材料学专业热度相对较低，计算机、数据科学类专业热度高，这也与现在的大数据时代相对应。

七、结论与展望

本文基于大学转专业数据，试图从转专业申请数据查看该大学校内的专业热度。基于 2013 年毕业去向率指标，分析出转入专业和转出专业差异化不明显，通过 Pagerank 算法得到专业热度较火的专业有经济学类、法学、计算机、数据科学与大数据技术，与目前社会需求相符。

此外，本文还可以通过一些其他研究进一步研究专业热度，例如，可以根据转专业时的成绩排名来分析各专业转入的申请难度，由此可以探究学习能力强的学生就读专业的喜好，以此来探究学生学习能力是否对专业热度的有影响。

附录 1　R 语言程序代码

```
##批量加载包
library(ggplot2)
library(stringr)
library(reshape2)
library(igraph)
library(xlsx)
library(RColorBrewer)
library(gridExtra)

##读入数据
eighteen <- read.xlsx("/Users/yuqinhan1229/Desktop/2018.xls",1)
dt2018 <- as.data.frame(table(eighteen$所在专业))
##按人数降序排序
order18_in<-dt2018[order(dt2018$Freq,decreasing=F),]
##2018 年学生申请转出情况
g1 <- ggplot(data = order18_in,aes(x = Freq, y = reorder(Var1,Freq))) +
  geom_bar(stat = 'identity',
           fill =
ifelse(order18_in$Freq>mean(order18_in$Freq),'#CC99CC','#CC99FF'), # 根据 y 值
相对均值的大小设置颜色
           width = 0.9) +
  theme(text=element_text(size=8, family="Songti SC"))+
  # 横纵坐标轴名称
  labs(x = '申请转出人数', y = '所在专业分布', title='2018 年学生申请转出情况')+
  theme(panel.grid.major =element_blank(),   # 去除边框
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.title =    element_text(size=10,face = "bold"),
        axis.text.x =      element_text(angle=0,
                                        hjust = 1, # 调整横坐标文字位置
                                        size=6),   # 调整横坐标文字字号
        plot.margin=unit(rep(3,4),'lines'))        # 设置图片边距

##2018 年学生申请转入情况
dt2018 <- as.data.frame(table(eighteen$转入专业))
```

```
##按人数降序排序
order18_out<-dt2018[order(dt2018$Freq,decreasing=F),]
g2 <- ggplot(data = order18_out,aes(x = Freq, y = reorder(Var1,Freq))) +
  geom_bar(stat = 'identity',
           fill =
ifelse(order18_out$Freq>mean(order18_out$Freq),'palegreen2','palegreen3'), # 根据
y 值相对均值的大小设置颜色
           width = 0.9) +
  theme(text=element_text(size=8, family="Songti SC"))+
  # 横纵坐标轴名称
  labs(x = '申请转入人数', y = '转入专业分布', title='2018 年学生申请转入专业情
况')+
  theme(panel.grid.major =element_blank(),  # 去除边框
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.title =  element_text(size=10,face = "bold"),
        axis.text.x =  element_text(angle=0,
                                    hjust = 1, # 调整横坐标文字位置
                                    size=6),  # 调整横坐标文字字号
        plot.margin=unit(rep(3,4),'lines'))      # 设置图片边距
grid.arrange(g1,g2,ncol =2)

##读入数据
nineteen <- read.xlsx("/Users/yuqinhan1229/Desktop/2019.xls",1)
dt2019 <- as.data.frame(table(nineteen$所在专业))
##按人数降序排序
order19_out<-dt2019[order(dt2019$Freq,decreasing=F),]
##2019 年学生申请转出情况
g1 <-ggplot(data = order19_out,aes(x = Freq, y = reorder(Var1,Freq))) +
  geom_bar(stat = 'identity',
           fill =
ifelse(order19_out$Freq>mean(order19_out$Freq),'#CC99CC','#CC99FF'), # 根据 y
值相对均值的大小设置颜色
           width = 0.9) +
  theme(text=element_text(size=8, family="Songti SC"))+
  # 横纵坐标轴名称
  labs(x = '申请转出人数', y = '所在专业分布', title='2019 年学生申请转出情况')+
```

```
  theme(panel.grid.major =element_blank(),   # 去除边框
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.title =    element_text(size=10,face = "bold"),
        axis.text.x =      element_text(angle=0,
                                        hjust = 1, # 调整横坐标文字位置
                                        size=6),   # 调整横坐标文字字号
        plot.margin=unit(rep(3,4),'lines'))        # 设置图片边距

##2019 年学生申请转入情况
dt2019 <- as.data.frame(table(nineteen$转入专业))
##按人数降序排序
order19_in<-dt2019[order(dt2019$Freq,decreasing=F),]
g2 <- ggplot(data = order19_in,aes(x = Freq, y = reorder(Var1,Freq))) +
  geom_bar(stat = 'identity',
           fill =
ifelse(order19_in$Freq>mean(order19_in$Freq),'palegreen2','palegreen3'), # 根据 y
值相对均值的大小设置颜色
           width = 0.9) +
  theme(text=element_text(size=8, family="Songti SC"))+
  # 横纵坐标轴名称
  labs(x = '申请转入人数', y = '转入专业分布', title='2019 年学生申请转入专业情
况')+
  theme(panel.grid.major =element_blank(),   # 去除边框
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.title =    element_text(size=10,face = "bold"),
        axis.text.x =      element_text(angle=0,
                                        hjust = 1, # 调整横坐标文字位置
                                        size=6),   # 调整横坐标文字字号
        plot.margin=unit(rep(3,4),'lines'))        # 设置图片边距
grid.arrange(g1,g2,ncol =2)

##读入数据
twenty <- read.xlsx("/Users/yuqinhan1229/Desktop/2020.xls",1)
dt2020 <- as.data.frame(table(twenty$所在专业))
##按人数降序排序
```

```
order20_out<-dt2020[order(dt2020$Freq,decreasing=F),]
##2020 年学生申请转出情况
g1 <-ggplot(data = order20_out,aes(x = Freq, y = reorder(Var1,Freq))) +
  geom_bar(stat = 'identity',
           fill =
ifelse(order20_out$Freq>mean(order20_out$Freq),'#CC99CC','#CC99FF'), # 根据 y
值相对均值的大小设置颜色
           width = 0.9) +
  theme(text=element_text(size=8, family="Songti SC"))+
  # 横纵坐标轴名称
  labs(x = '申请转出人数', y = '所在专业分布', title='2020 年学生申请转出情况')+
  theme(panel.grid.major =element_blank(),  # 去除边框
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.title =  element_text(size=10,face = "bold"),
        axis.text.x =  element_text(angle=0,  # 横坐标文字旋转九十度
                                    hjust = 1, # 调整横坐标文字位置
                                    size=6),  # 调整横坐标文字字号
        plot.margin=unit(rep(3,4),'lines'))      # 设置图片边距

##2019 年学生申请转入情况
dt2020 <- as.data.frame(table(twenty$转入专业))
##按人数降序排序
order20_in<-dt2020[order(dt2020$Freq,decreasing=F),]
g2 <- ggplot(data = order20_in,aes(x = Freq, y = reorder(Var1,Freq))) +
  geom_bar(stat = 'identity',
           fill =
ifelse(order20_in$Freq>mean(order20_in$Freq),'palegreen2','palegreen3'), # 根据 y
值相对均值的大小设置颜色
           width = 0.9) +
  theme(text=element_text(size=8, family="Songti SC"))+
  # 横纵坐标轴名称
  labs(x = '申请转入人数', y = '转入专业分布', title='2020 年学生申请转入专业情
况')+
  theme(panel.grid.major =element_blank(),  # 去除边框
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
```

```
        axis.title =    element_text(size=10,face = "bold"),
        axis.text.x =     element_text(angle=0,   # 横坐标文字旋转九十度
                                       hjust = 1, # 调整横坐标文字位置
                                       size=6),   # 调整横坐标文字字号
        plot.margin=unit(rep(3,4),'lines'))        # 设置图片边距
grid.arrange(g1,g2,ncol =2)

##年份合并
#对转出数据进行年份合并绘制 stack 图
#转出数据的合并
dtall_out <- merge(order18_out,order19_out,by="Var1",all=T)
dtall_out <- merge(dtall_out,order19_out,by="Var1",all=T)
colnames(dtall_out) <- c("所在专业","2018","2019","2020") #改列名
dtall_out[is.na(dtall_out)] <- 0 #缺失值填充为 0
dtall_out$sum <- dtall_out$`2018`+dtall_out$`2019`+dtall_out$`2020`
dtall_out <- dtall_out[order(dtall_out$sum,decreasing=T),]
#取前 15 名进行合并
dtall_out <- dtall_out[1:15,1:4]
dtall_out <- melt(dtall_out,id.vars = "所在专业",variable.name = "year",
                  value.name = "转出人数")
#画 stack 图
ggplot(dtall_out,aes(x=所在专业,weight=转出人数,fill=year))+
  geom_bar(position = 'stack')+
  labs(y="转出人数")+
  coord_flip() +
  theme(text=element_text(size=8, family="Songti SC"))+
  theme(panel.grid.major =element_blank())   # 去除边框

##转入数据同理
dtall_in <- merge(order18_in,order19_in,by="Var1",all=T)
dtall_in <- merge(dtall_in,order19_in,by="Var1",all=T)
colnames(dtall_in) <- c("所在专业","2018","2019","2020") #改列名
dtall_in[is.na(dtall_in)] <- 0 #缺失值填充为 0
dtall_in$sum <- dtall_in$`2018`+dtall_in$`2019`+dtall_in$`2020`
dtall_in <- dtall_in[order(dtall_in$sum,decreasing=T),]
#取前 15 名进行合并
dtall_in <- dtall_in[1:15,1:4]
```

```
dtall_in <- melt(dtall_in,id.vars = "所在专业",variable.name = "year",
                                value.name = "转出人数")
#画 stack 图
ggplot(dtall_in,aes(x=所在专业,weight=转出人数,fill=year))+
  geom_bar(position = 'stack')+
  labs(y="转入人数")+
  coord_flip() +
  theme(text=element_text(size=8, family="Songti SC"))+
  theme(panel.grid.major =element_blank())   # 去除边框

#任务 2
##结合 2013 级毕业生毕业升学率、就业率、出国率，考察转专业的同学，其转出专业与转入专业在几个指标上是否存在明显差距？
thirteen <- read.csv("/Users/yuqinhan1229/Desktop/2013.csv")[2:5]
#2018 年转入和转出在各指标上的差异
eighteen$freq <- 1
dt2018 <- aggregate(freq~所在专业+转入专业,data=eighteen,sum)
dt2018$出国出境率差 <- 0
dt2018$国内升学率差 <- 0
dt2018$就业率差 <- 0
##如果 2018 年的转入转出专业均有在 2013 年专业的数据则记录
for(i in 1:nrow(dt2018)){
  if((nrow(thirteen[thirteen$专业== dt2018[i,1],])==1)&&(nrow(thirteen[thirteen$专业== dt2018[i,2],])==1)){
     dt2018[i,4:6] <- thirteen[thirteen$专业== dt2018[i,1],][1,2:4]-thirteen[thirteen$专业== dt2018[i,2],][1,2:4]
  } #计算两个专业均有毕业数据的差异情况

}
#删掉没有数据的转入转出专业
dt2018_1 <- dt2018[dt2018$出国出境率差!=0,]
#将后三列进行 melt
dt2018_2 <- melt(dt2018_1[,-3])
dt2018_2$value <- dt2018_2$value*100
ggplot(data=dt2018_2,aes(x=variable,y=value,fill=variable))+
  geom_boxplot(alpha=0.5)+
  geom_jitter(aes(color=variable), alpha=.3)+
```

```
  labs(x="各项指标",y="指标差")+
  guides(fill = "none",color= "none")+
  theme(text=element_text(size=8, family="Songti SC"))


#2019 年转入和转出在各指标上的差异
nineteen$freq <- 1
dt2019 <- aggregate(freq~所在专业+转入专业,data=nineteen,sum)
dt2019$出国出境率差 <- 0
dt2019$国内升学率差 <- 0
dt2019$就业率差 <- 0
##如果 2019 年的转出转入专业均有在 2013 年专业的数据则记录，算的数据差
for(i in 1:nrow(dt2019)){
  if((nrow(thirteen[thirteen$专业== dt2019[i,1],])==1)&&(nrow(thirteen[thirteen$专
业== dt2019[i,2],])==1)){
    dt2019[i,4:6] <- thirteen[thirteen$专业== dt2019[i,1],][1,2:4]-thirteen[thirteen$
专业== dt2019[i,2],][1,2:4]
  } #计算两个专业均有毕业数据的差异情况

}
dt2019_1 <- dt2019[dt2019$出国出境率差!=0,]
#将后三列进行 melt
dt2019_2 <- melt(dt2019_1[,-3])
dt2019_2$value <- dt2019_2$value*100
ggplot(data=dt2019_2,aes(x=variable,y=value,fill=variable))+
  geom_boxplot(alpha=0.5)+
  geom_jitter(aes(color=variable), alpha=.3)+
  labs(x="各项指标",y="指标差")+
  guides(fill = "none",color= "none")+
  theme(text=element_text(size=8, family="Songti SC"))


#2020 年转入和转出在各指标上的差异
twenty$freq <- 1
dt2020 <- aggregate(freq~所在专业+转入专业,data=twenty,sum)
dt2020$出国出境率差 <- 0
```

```
dt2020$国内升学率差 <- 0
dt2020$就业率差 <- 0
##如果 2019 年的转入转出专业均有在 2013 年专业的数据则记录
for(i in 1:nrow(dt2020)){
  if((nrow(thirteen[thirteen$专业== dt2020[i,1],])==1)&&(nrow(thirteen[thirteen$专业== dt2020[i,2],])==1)){
    dt2020[i,4:6] <- thirteen[thirteen$专业== dt2020[i,1],][1,2:4]-thirteen[thirteen$专业== dt2020[i,2],][1,2:4]
  } #计算两个专业均有毕业数据的差异情况

}
dt2020_1 <- dt2020[dt2020$出国出境率差!=0,]
#将后三列进行 melt
dt2020_2 <- melt(dt2020_1[,-3])
dt2020_2$value <- dt2020_2$value*100
ggplot(data=dt2020_2,aes(x=variable,y=value,fill=variable))+
  geom_boxplot(alpha=0.5)+
  geom_jitter(aes(color=variable), alpha=.3)+
  labs(x="各项指标",y="指标差")+
  guides(fill = "none",color= "none")+
  theme(text=element_text(size=8, family="Songti SC"))

# 任务三 pagerank
##2018 年
##建立 2018 年带权重的有向图 A
A18 <- graph.data.frame(dt2018[,1:3])
##获取排序情况
PR18 <- as.data.frame(page.rank(A18)$vector)
PR18$专业 <- rownames(PR18)
## 由大到小排序
PR18 <- PR18[order(PR18$`page.rank(A18)$vector`,decreasing=T),]
colnames(PR18) <- c("PageRank 值","专业")
## 标准化
PR18$PageRank 值 <- PR18$PageRank 值*100/PR18[1,1]
PR18_top <- PR18[1:15,]
art <- c('保险学','城市管理','德语','俄语','法学','法语','翻译','工商管理类',
         '管理科学与工程类','广播电视学','汉语国际教育','汉语言文学','经济学
```

```
类',
            '历史学类','旅游管理类','葡萄牙语','日语','社会学类','思想政治教育',
            '图书馆学','图书情报与档案管理类','文物与博物馆学','西班牙语','行政
管理',
            '意大利语','英语','哲学','哲学类','政治学类','历史学','世界史','国际政治','
旅游管理',
            '社会学','会计学（国际会计）','中国语言文学类','公共管理类','政治学
与行政学'
            ,'会展经济与管理','逻辑学')
PR18_art <- PR18[PR18$专业  %in% art,]
PR18_sci <- PR18[!(PR18$专业  %in% art),]
##绘制专业热度排名柱状图
##前 15 名
g1 <- ggplot(data=PR18_top,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
   geom_bar(stat="identity",fill="#CC99CC")+
   labs(x="2018 年前 15 名专业",y="热度得分")+
   theme(legend.title=element_blank())+
   theme(text=element_text(size=8, family="Songti SC"))+
   coord_flip()

##文科
g2 <- ggplot(data=PR18_art,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
   geom_bar(stat="identity",fill="gold")+
   labs(x="2018 年文科专业",y="热度得分")+
   theme(legend.title=element_blank())+
   theme(text=element_text(size=8, family="Songti SC"))+
   coord_flip()

##理科
g3 <- ggplot(data=PR18_sci,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
   geom_bar(stat="identity",fill="pink")+
   labs(x="2018 年理科专业",y="热度得分")+
   theme(legend.title=element_blank())+
   theme(text=element_text(size=8, family="Songti SC"))+
   coord_flip()
grid.arrange(g1,g2,g3,ncol =3)
```

```
##2019 年
##建立 2019 年带权重的有向图 A19
A19 <- graph.data.frame(dt2019[,1:3])
##获取排序情况
PR19 <- as.data.frame(page.rank(A19)$vector)
PR19$专业  <- rownames(PR19)
##  由大到小排序
PR19 <- PR19[order(PR19$`page.rank(A19)$vector`,decreasing=T),]
colnames(PR19) <- c("PageRank 值","专业")
##  标准化
PR19$PageRank 值  <- PR19$PageRank 值*100/PR19[1,1]
PR19_top <- PR19[1:15,]
art <- c('保险学','城市管理','德语','俄语','法学','法语','翻译','工商管理类',
         '管理科学与工程类','广播电视学','汉语国际教育','汉语言文学','经济学
类',
         '历史学类','旅游管理类','葡萄牙语','日语','社会学类','思想政治教育',
         '图书馆学','图书情报与档案管理类','文物与博物馆学','西班牙语','行政
管理',
         '意大利语','英语','哲学','哲学类','政治学类','历史学','世界史','国际政治','
旅游管理',
         '社会学','会计学（国际会计）','中国语言文学类','公共管理类','政治学
与行政学'
         ,'会展经济与管理','逻辑学','经管法班')
PR19_art <- PR19[PR19$专业  %in% art,]
PR19_sci <- PR19[!(PR19$专业  %in% art),]
##绘制专业热度排名柱状图
##前 15 名
g1 <- ggplot(data=PR19_top,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
  geom_bar(stat="identity",fill="#CC99CC")+
  labs(x="2019 年前 15 名专业",y="热度得分")+
  theme(legend.title=element_blank())+
  theme(text=element_text(size=8, family="Songti SC"))+
  coord_flip()

##文科
```

```
g2 <- ggplot(data=PR18_art,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
   geom_bar(stat="identity",fill="gold")+
   labs(x="2019 年文科专业",y="热度得分")+
   theme(legend.title=element_blank())+
   theme(text=element_text(size=8, family="Songti SC"))+
   coord_flip()

##理科
g3 <- ggplot(data=PR19_sci,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
   geom_bar(stat="identity",fill="pink")+
   labs(x="2019 年理科专业",y="热度得分")+
   theme(legend.title=element_blank())+
   theme(text=element_text(size=8, family="Songti SC"))+
   coord_flip()
grid.arrange(g1,g2,g3,ncol =3)




##2020 年
##建立 2020 年带权重的有向图 A20
A20 <- graph.data.frame(dt2020[,1:3])
##获取排序情况
PR20 <- as.data.frame(page.rank(A20)$vector)
PR20$专业  <- rownames(PR20)
##  由大到小排序
PR20 <- PR20[order(PR20$`page.rank(A20)$vector`,decreasing=T),]
colnames(PR20) <- c("PageRank 值","专业")
##  标准化
PR20$PageRank 值  <- PR20$PageRank 值*100/PR20[1,1]
PR20_top <- PR19[1:15,]
art <- c('保险学','城市管理','德语','俄语','法学','法语','翻译','工商管理类',
         '管理科学与工程类','广播电视学','汉语国际教育','汉语言文学','经济学
类',
         '历史学类','旅游管理类','葡萄牙语','日语','社会学类','思想政治教育',
         '图书馆学','图书情报与档案管理类','文物与博物馆学','西班牙语','行政
管理',
```

```
        '意大利语','英语','哲学','哲学类','政治学类','历史学','世界史','国际政治','
旅游管理',
        '社会学','会计学（国际会计）','中国语言文学类','公共管理类','政治学
与行政学'
        ,'会展经济与管理','逻辑学','编辑出版学','马克思主义理论学')
PR20_art <- PR20[PR20$专业  %in% art,]
PR20_sci <- PR20[!(PR20$专业  %in% art),]
##绘制专业热度排名柱状图
##前 15 名
g1 <- ggplot(data=PR20_top,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
  geom_bar(stat="identity",fill="#CC99CC")+
  labs(x="2020 年前 15 名专业",y="热度得分")+
  theme(legend.title=element_blank())+
  theme(text=element_text(size=8, family="Songti SC"))+
  coord_flip()

##文科
g2 <- ggplot(data=PR20_art,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
  geom_bar(stat="identity",fill="gold")+
  labs(x="2020 年文科专业",y="热度得分")+
  theme(legend.title=element_blank())+
  theme(text=element_text(size=8, family="Songti SC"))+
  coord_flip()

##理科
g3 <- ggplot(data=PR20_sci,aes(x=reorder(专业,PageRank 值),y=PageRank 值))+
  geom_bar(stat="identity",fill="pink")+
  labs(x="2020 年理科专业",y="热度得分")+
  theme(legend.title=element_blank())+
  theme(text=element_text(size=8, family="Songti SC"))+
  coord_flip()
grid.arrange(g1,g2,g3,ncol =3)
```