西安工程大学 管理学院
XI'AN POLYTECHNIC UNIVERSITY School of Management

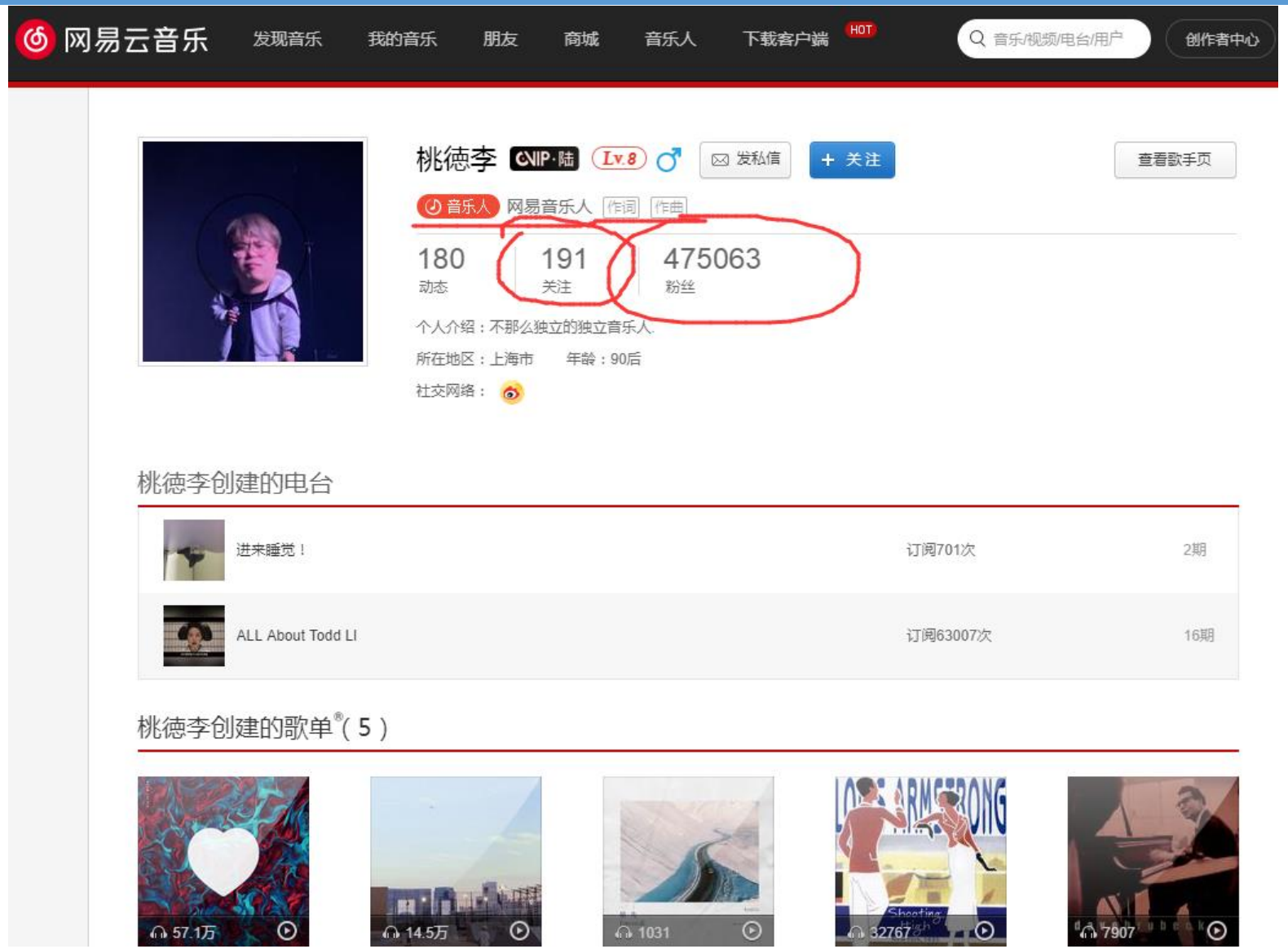
社会网络分析

案例：抓取并构建网易音乐人网络

邵鹏 博士
副教授 硕士生导师
西安工程大学管理学院
shaopengde@sohu.com
inebm.cn

问题

网易云音乐有很多原创音乐人，这些音乐人在在线上平台上社交关系。
这些音乐人的线上社交网络关系有什么特征？网络如何构建？



思路

1 确定数据爬取方案

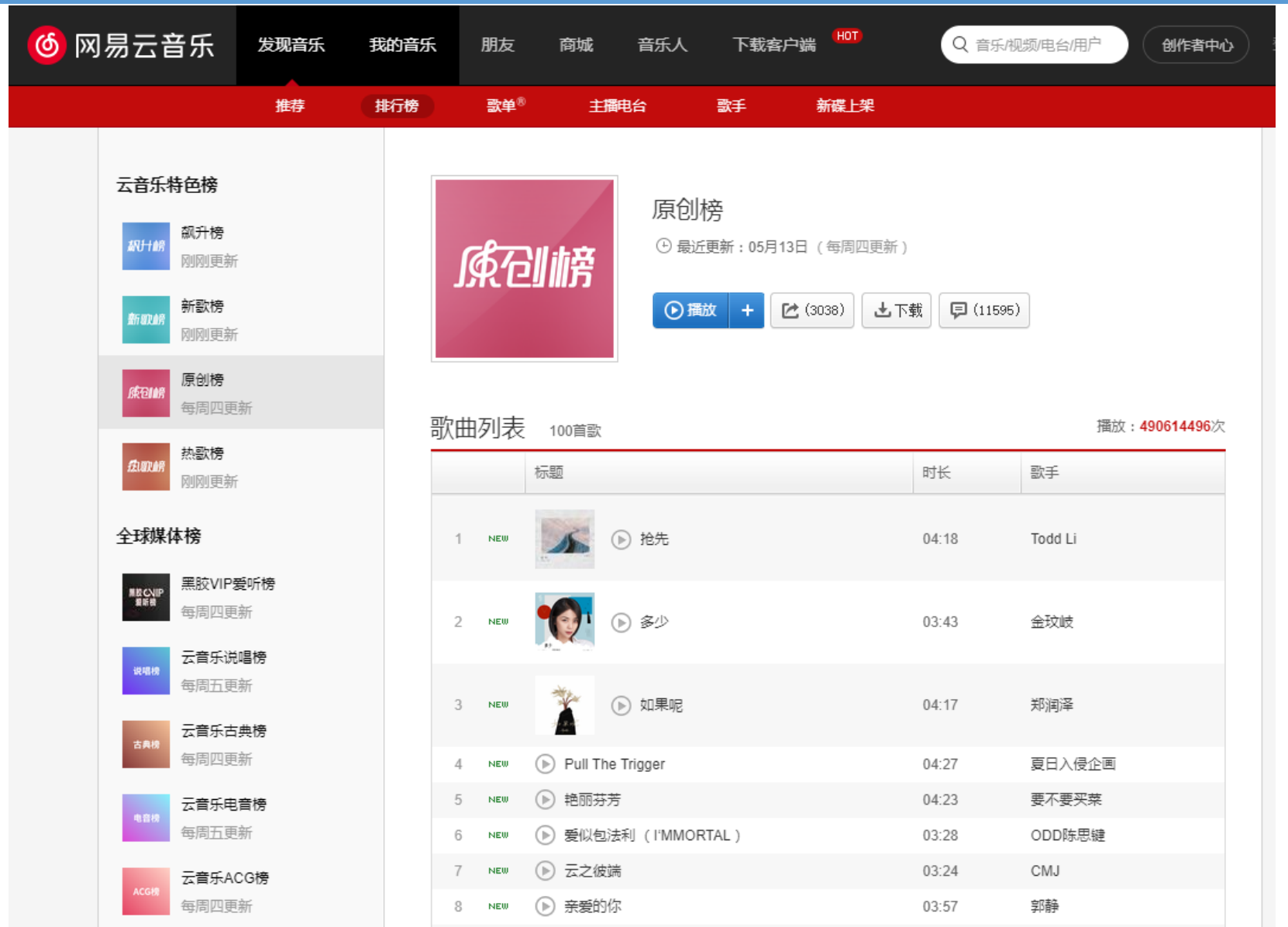
2 爬取云音乐相关数据（创作人社交信息）

3 清洗数据、构建网络

1 确定数据爬取方案

以"原创榜100歌"的音乐人为例进行研究

https://music.163.com/#/discover/toplist?id=2884035

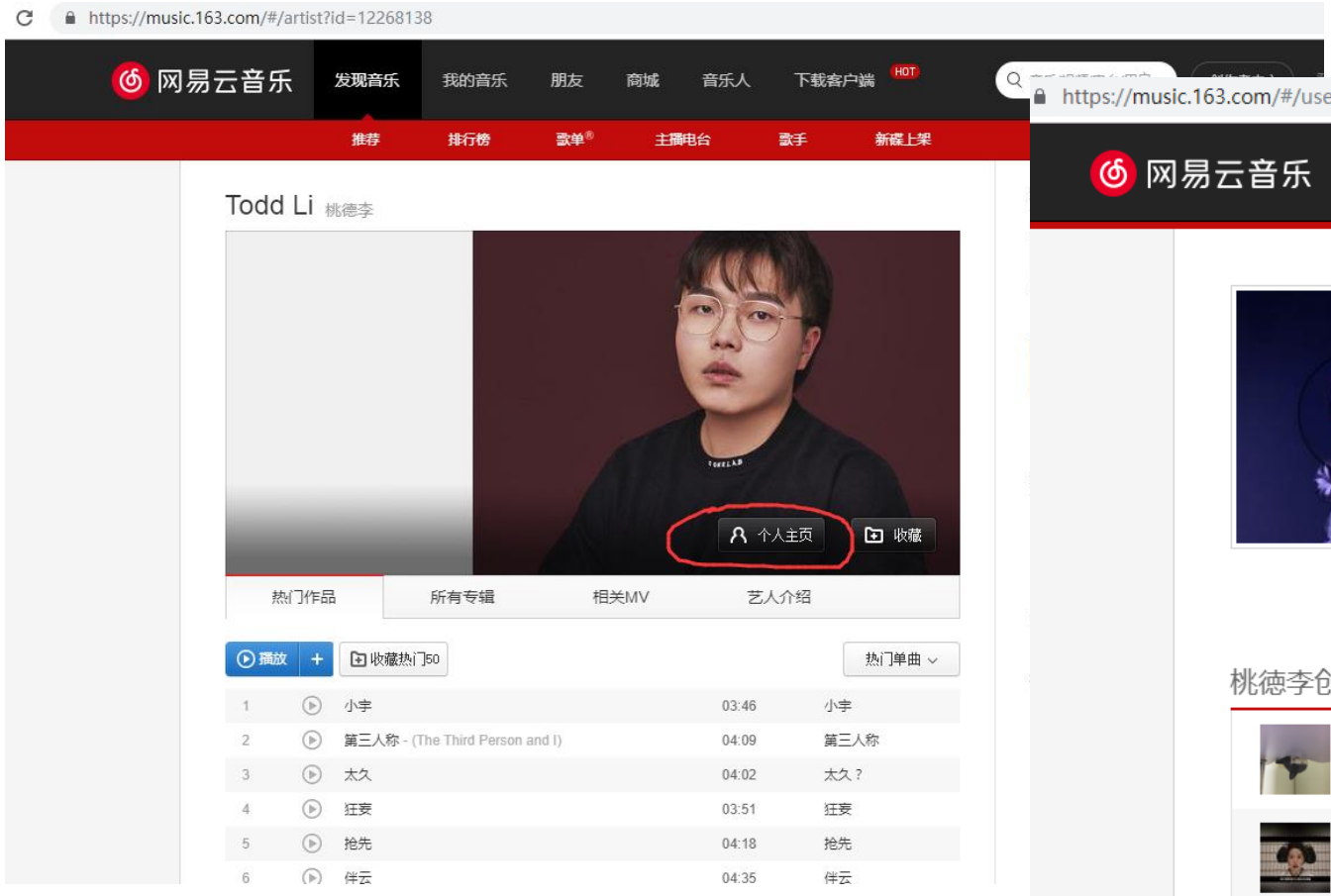


1 确定数据爬取方案

西安工程大学 管理学院
XI'AN POLYTECHNIC UNIVERSITY School of Management

获取歌手艺术家artist页面网址

https://music.163.com/#/artist?id=12268138



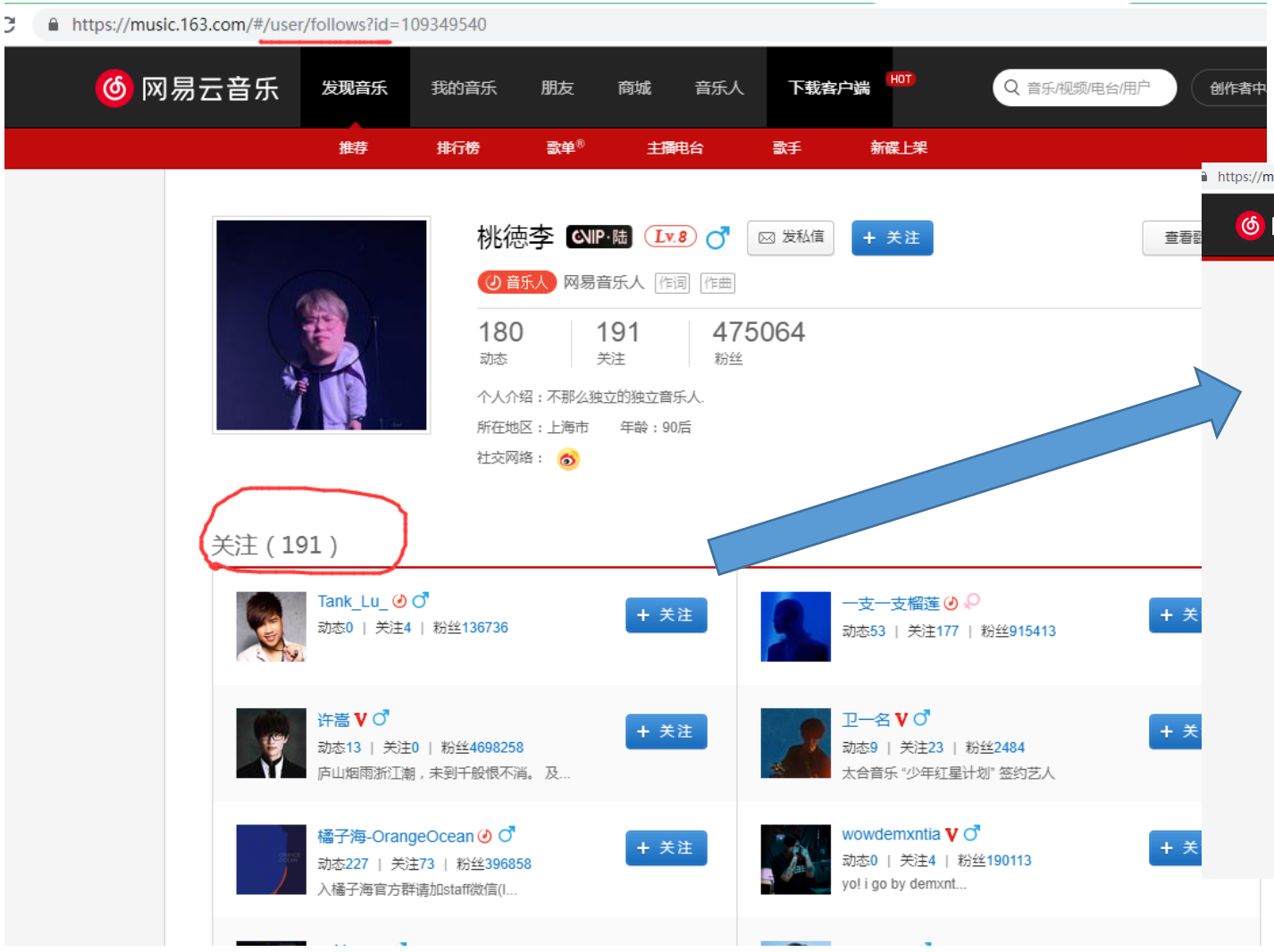
获取歌手用户user页面网址

https://music.163.com/#/user/home?id=109349540

1 确定数据爬取方案

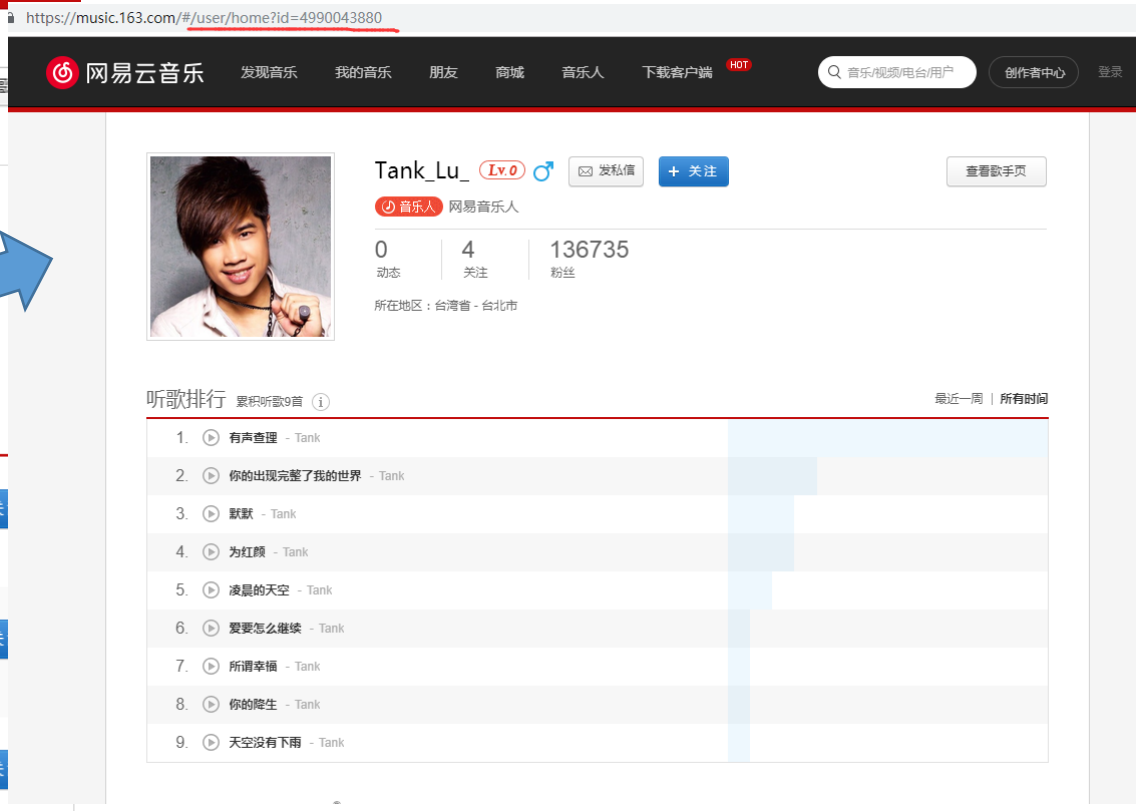
获取歌手出度follows网址

https://music.163.com/#/user/follows?id=109349540



抓取该歌手关注的用户的信息

https://music.163.com/#/user/home?id=4990043880



1 确定数据爬取方案

以"原创榜100歌"的音乐人为例进行研究

https://music.163.com/#/discover/toplist?id=2884035

获取100首歌的歌手艺术家artist页面网址

如 https://music.163.com/#/artist?id=12268138

EXCEL1，数据n小于等于100条

获取歌手用户user页面网址,出度follows网址

如 https://music.163.com/#/user/home?id=109349540

如 https://music.163.com/#/user/follows?id=109349540

EXCEL2，数据n小于等于100条

抓取每位歌手关注的用户的信息

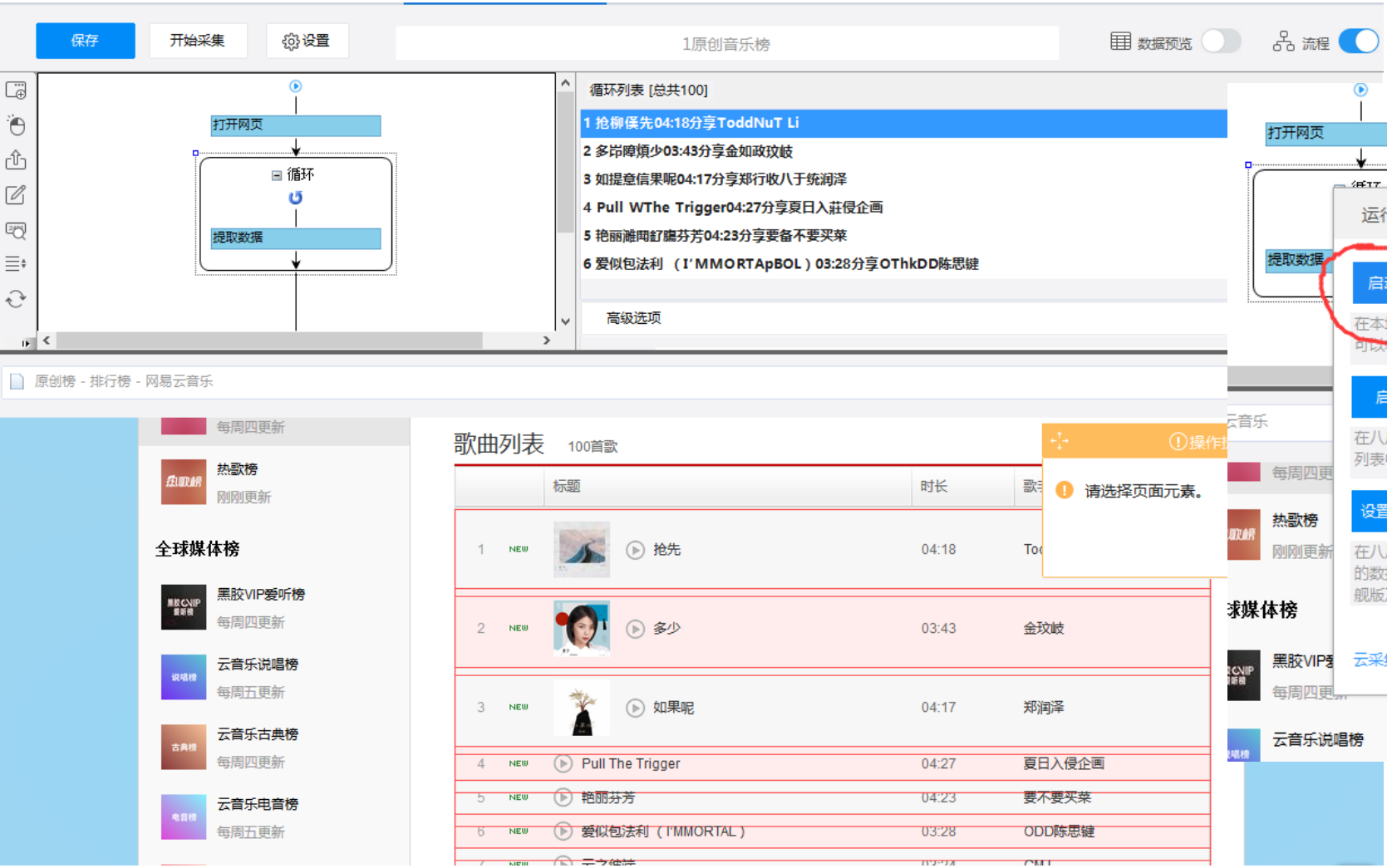
如 https://music.163.com/#/user/home?id=4990043880

EXCEL3,n位歌手及其关注的用户的网址

2 爬取云音乐相关数据

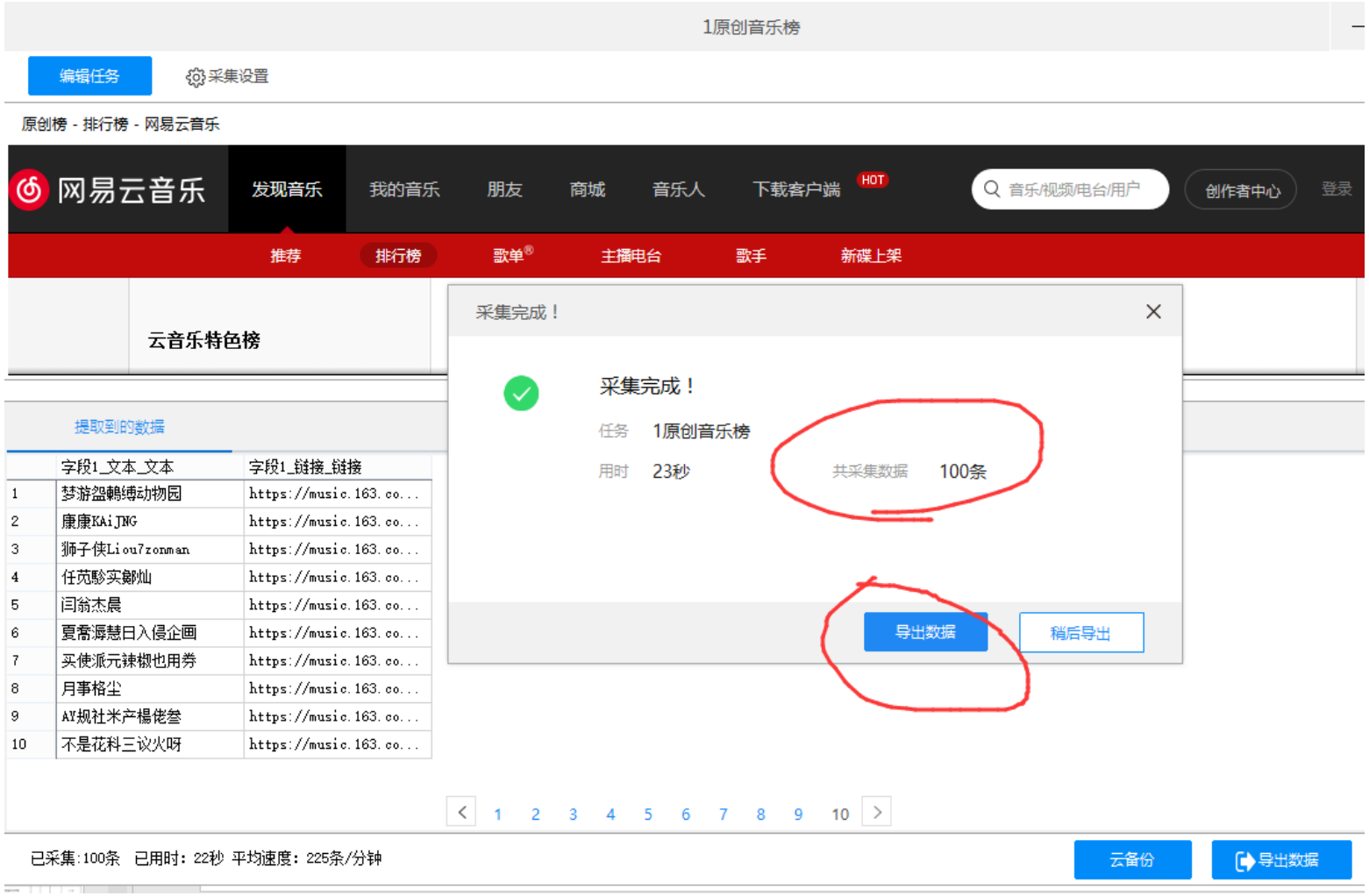
西安工程大学 管理学院
XI'AN POLYTECHNIC UNIVERSITY School of Management

规则1，获取100首歌的歌手艺术家artist页面网址



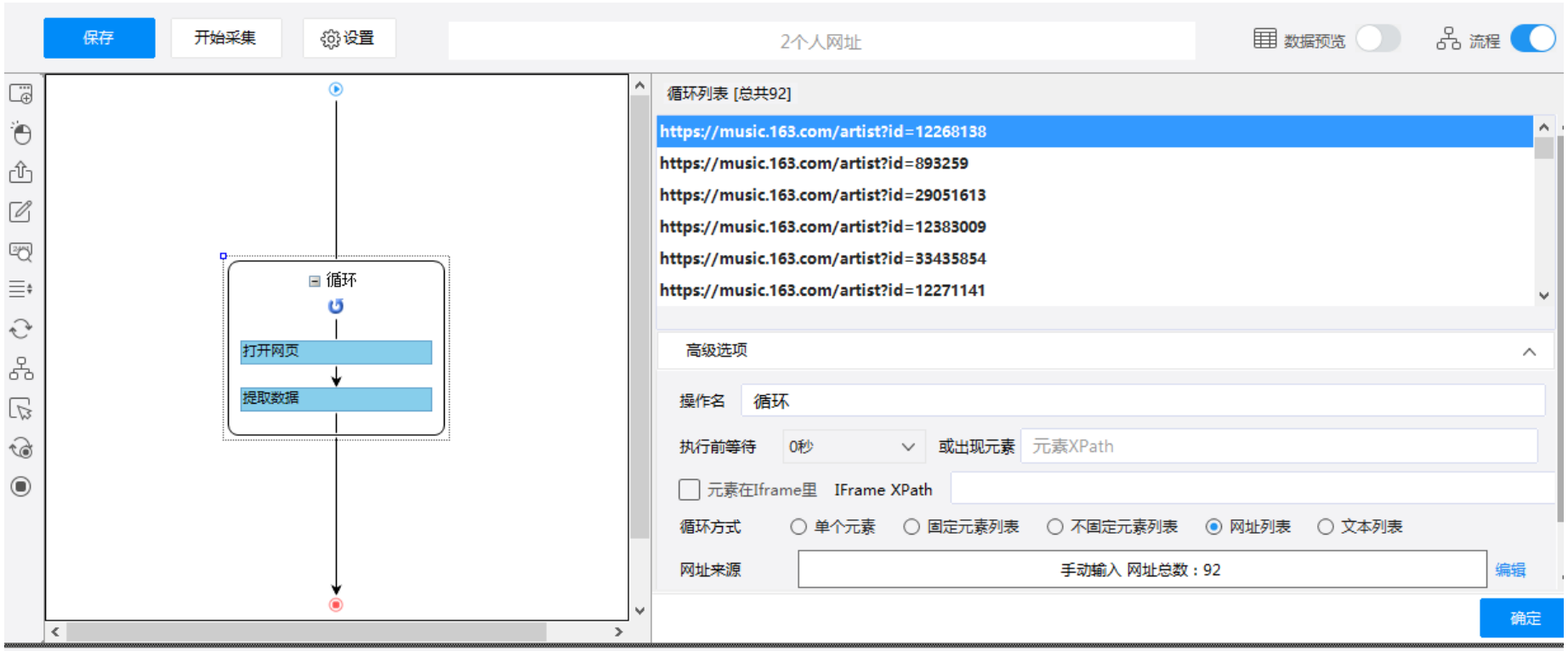
2 爬取云音乐相关数据

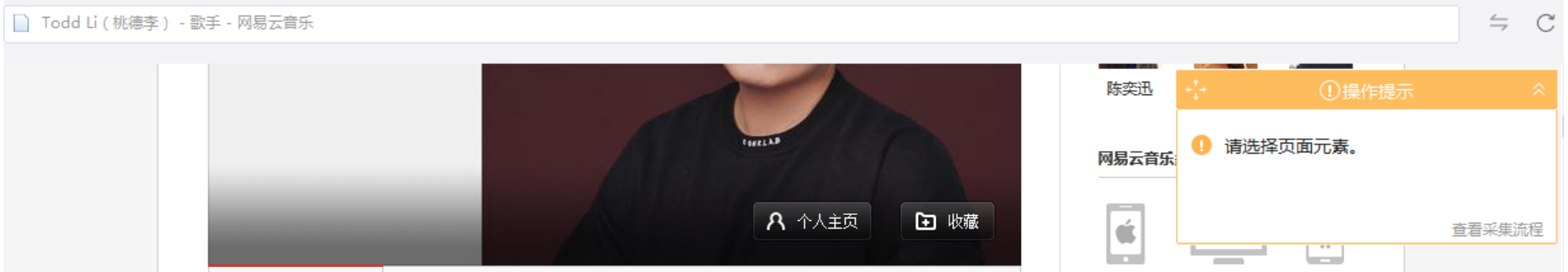
导出数据到EXCEL



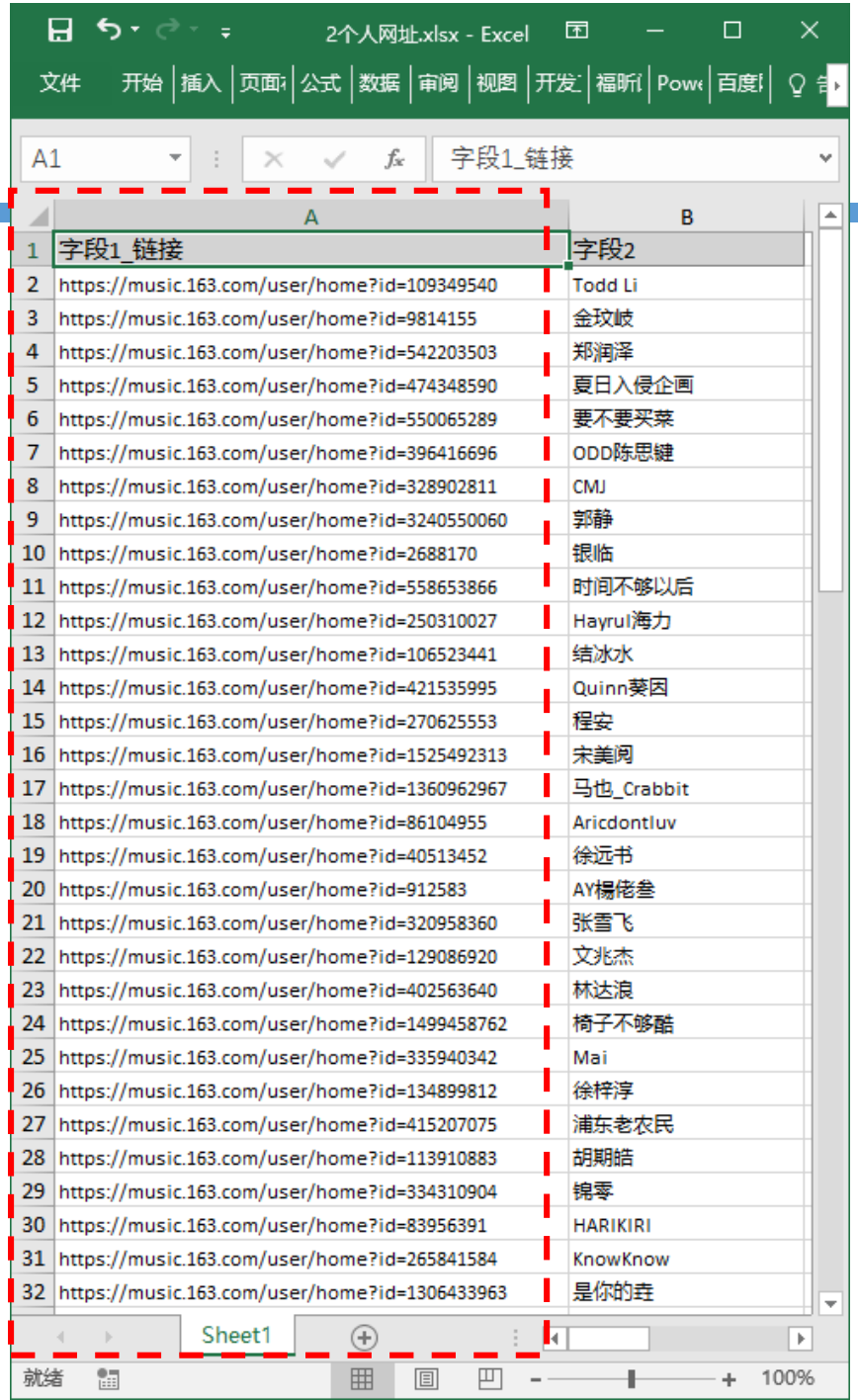
2 爬取云音乐相关数据

规则2，循环92位艺术家artist页面网址，获取用户页面网址



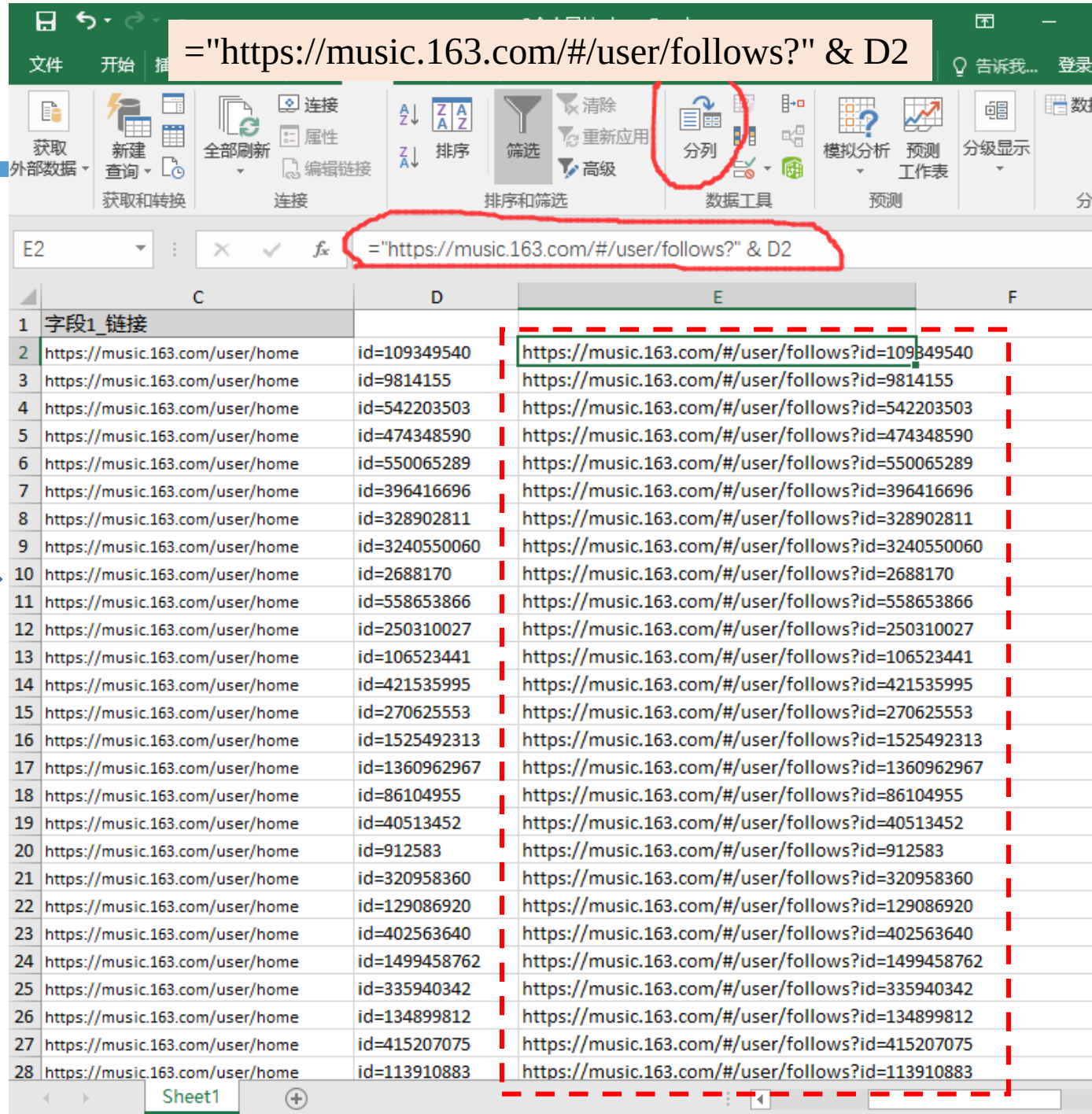


	字段1_链接	字段2
1	https://music.163.co...	Gold Child
2	https://music.163.co...	因你而在的梦
3	https://music.163.co...	NaCho
4	https://music.163.co...	孟琦
5	https://music.163.co...	小好Lil Good
6	https://music.163.co...	红白色乐队
7	https://music.163.co...	potatocake
8	https://music.163.co...	谢帝
9	https://music.163.co...	王同学Able


2个人网址.xlsx - Excel
A1
字段1_链接
字段1_链接
字段2
https://music.163.com/user/home?id=109349540 Todd Li
https://music.163.com/user/home?id=9814155 金玟岐
https://music.163.com/user/home?id=542203503 郑润泽
https://music.163.com/user/home?id=474348590 夏日入侵企画
https://music.163.com/user/home?id=550065289 要不要买菜
https://music.163.com/user/home?id=396416696 ODD陈思键
https://music.163.com/user/home?id=328902811 CMJ
https://music.163.com/user/home?id=3240550060 郭静
https://music.163.com/user/home?id=2688170 银临
https://music.163.com/user/home?id=558653866 时间不够以后
https://music.163.com/user/home?id=250310027 Hayrul海力
https://music.163.com/user/home?id=106523441 结冰水
https://music.163.com/user/home?id=421535995 Quinn葵因
https://music.163.com/user/home?id=270625553 程安
https://music.163.com/user/home?id=1525492313 宋美阅
https://music.163.com/user/home?id=1360962967 马也_Crabbit
https://music.163.com/user/home?id=86104955 Aricdontluv
https://music.163.com/user/home?id=40513452 徐远书
https://music.163.com/user/home?id=912583 AY楊佬叁
https://music.163.com/user/home?id=320958360 张雪飞
https://music.163.com/user/home?id=129086920 文兆杰
https://music.163.com/user/home?id=402563640 林达浪
https://music.163.com/user/home?id=1499458762 椅子不够酷
https://music.163.com/user/home?id=335940342 Mai
https://music.163.com/user/home?id=134899812 徐梓淳
https://music.163.com/user/home?id=415207075 浦东老农民
https://music.163.com/user/home?id=113910883 胡期皓
https://music.163.com/user/home?id=334310904 锦零
https://music.163.com/user/home?id=83956391 HARIKIRI
https://music.163.com/user/home?id=265841584 KnowKnow
https://music.163.com/user/home?id=1306433963 是你的垚
Sheet1
就绪
100%

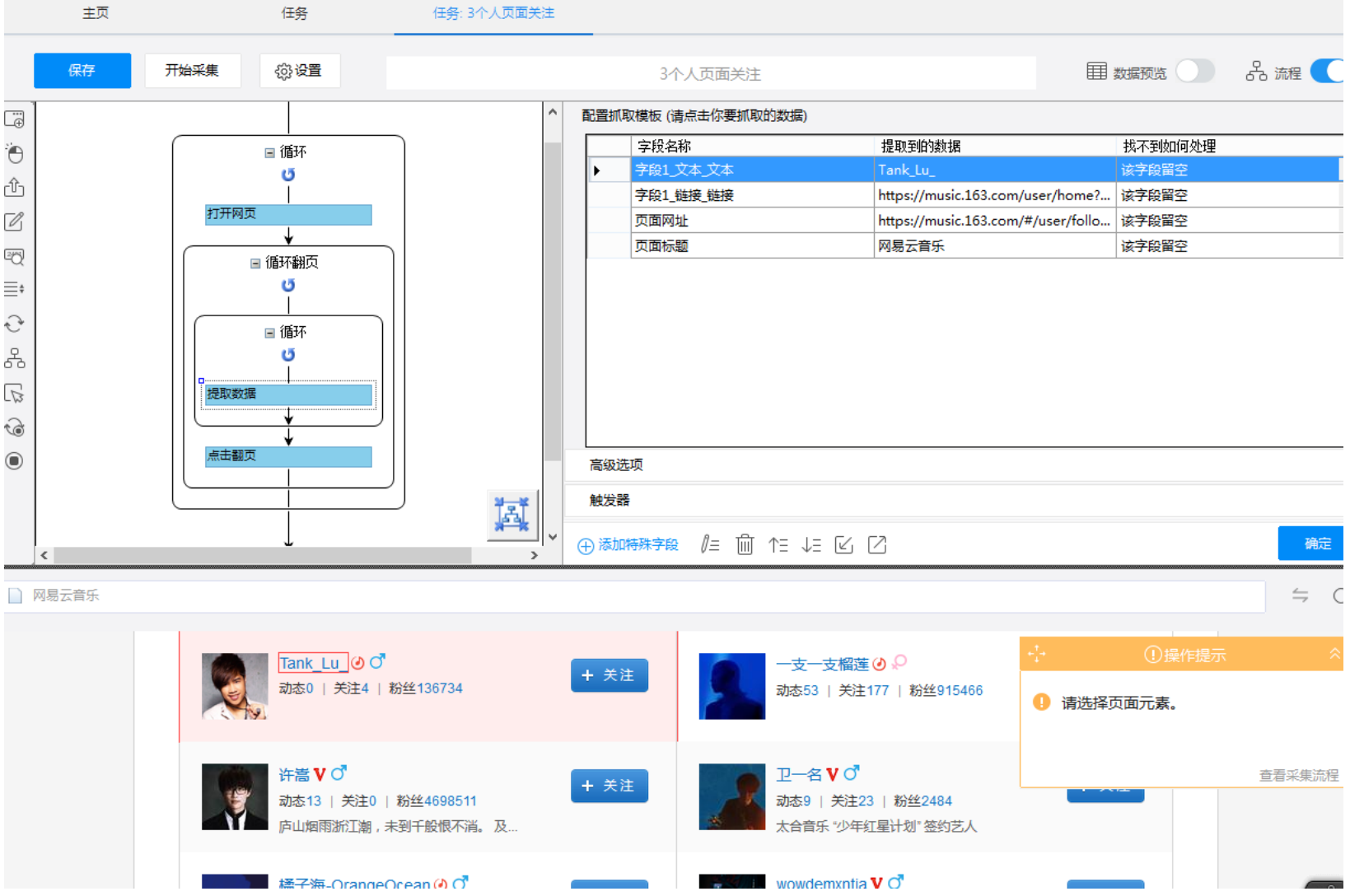
分列

连接


="https://music.163.com/#/user/follows?" & D2
获取外部数据
新建查询
全部刷新
连接
属性
编辑链接
获取和转换
连接
排序
筛选
清除
重新应用
高级
分列
排序和筛选
数据工具
模拟分析
预测工作表
预测
分级显示
E2
="https://music.163.com/#/user/follows?" & D2
字段1_链接
https://music.163.com/user/home id=109349540 https://music.163.com/#/user/follows?id=109349540
https://music.163.com/user/home id=9814155 https://music.163.com/#/user/follows?id=9814155
https://music.163.com/user/home id=542203503 https://music.163.com/#/user/follows?id=542203503
https://music.163.com/user/home id=474348590 https://music.163.com/#/user/follows?id=474348590
https://music.163.com/user/home id=550065289 https://music.163.com/#/user/follows?id=550065289
https://music.163.com/user/home id=396416696 https://music.163.com/#/user/follows?id=396416696
https://music.163.com/user/home id=328902811 https://music.163.com/#/user/follows?id=328902811
https://music.163.com/user/home id=3240550060 https://music.163.com/#/user/follows?id=3240550060
https://music.163.com/user/home id=2688170 https://music.163.com/#/user/follows?id=2688170
https://music.163.com/user/home id=558653866 https://music.163.com/#/user/follows?id=558653866
https://music.163.com/user/home id=250310027 https://music.163.com/#/user/follows?id=250310027
https://music.163.com/user/home id=106523441 https://music.163.com/#/user/follows?id=106523441
https://music.163.com/user/home id=421535995 https://music.163.com/#/user/follows?id=421535995
https://music.163.com/user/home id=270625553 https://music.163.com/#/user/follows?id=270625553
https://music.163.com/user/home id=1525492313 https://music.163.com/#/user/follows?id=1525492313
https://music.163.com/user/home id=1360962967 https://music.163.com/#/user/follows?id=1360962967
https://music.163.com/user/home id=86104955 https://music.163.com/#/user/follows?id=86104955
https://music.163.com/user/home id=40513452 https://music.163.com/#/user/follows?id=40513452
https://music.163.com/user/home id=912583 https://music.163.com/#/user/follows?id=912583
https://music.163.com/user/home id=320958360 https://music.163.com/#/user/follows?id=320958360
https://music.163.com/user/home id=129086920 https://music.163.com/#/user/follows?id=129086920
https://music.163.com/user/home id=402563640 https://music.163.com/#/user/follows?id=402563640
https://music.163.com/user/home id=1499458762 https://music.163.com/#/user/follows?id=1499458762
https://music.163.com/user/home id=335940342 https://music.163.com/#/user/follows?id=335940342
https://music.163.com/user/home id=134899812 https://music.163.com/#/user/follows?id=134899812
https://music.163.com/user/home id=415207075 https://music.163.com/#/user/follows?id=415207075
https://music.163.com/user/home id=113910883 https://music.163.com/#/user/follows?id=113910883
Sheet1

2 爬取云音乐相关数据

规则3，循环92位用户页面网址，对于每位用户循环其所有关注的用户信息



3 清洗数据、构建网络

D	E
页面网址	A的ID
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540
https://music.163.com/#/user/follows?id=109349540	109349540

关注

3个人页面关注.xlsm - Excel

A	B	C
B	B	B的ID
炸鸡少女阿肆	https://music.163.com/user/home?id=73125610	73125610
Tank_Lu_	https://music.163.com/user/home?id=4990043880	4990043880
一支一支榴莲	https://music.163.com/user/home?id=1649808342	1649808342
许嵩	https://music.163.com/user/home?id=31265745	31265745
卫一名	https://music.163.com/user/home?id=388865241	388865241
橘子海-OrangeOcean	https://music.163.com/user/home?id=109698685	109698685
wowdemxntia	https://music.163.com/user/home?id=3267561367	3267561367
谭禧龙	https://music.163.com/user/home?id=435244105	435244105
葛东琪	https://music.163.com/user/home?id=91724921	91724921
惟安娜-Vienna-	https://music.163.com/user/home?id=52829347	52829347
SuperLuckyQi	https://music.163.com/user/home?id=288477618	288477618
EdSheeran	https://music.163.com/user/home?id=416608258	416608258
張震嶽	https://music.163.com/user/home?id=5158334	5158334
BrunoMars	https://music.163.com/user/home?id=1671465495	1671465495
金玟岐	https://music.163.com/user/home?id=9814155	9814155
Keshimusic	https://music.163.com/user/home?id=3827086435	3827086435
mrbrunomajor	https://music.163.com/user/home?id=1409309284	1409309284
WeAreVarious	https://music.163.com/user/home?id=75388755	75388755
er呆酱	https://music.163.com/user/home?id=260930930	260930930
TaylorSwift	https://music.163.com/user/home?id=1880071427	1880071427
ShawnMendesCN	https://music.163.com/user/home?id=3773216955	3773216955
蛙池WaChi	https://music.163.com/user/home?id=73884324	73884324
笛呆子囚牛	https://music.163.com/user/home?id=358189202	358189202
茄子蛋EggPlantEgg	https://music.163.com/user/home?id=590126810	590126810
裘德Chiu	https://music.163.com/user/home?id=98087305	98087305
glassanimalsband	https://music.163.com/user/home?id=3718571391	3718571391
溺溺___	https://music.163.com/user/home?id=85065212	85065212
悶餅MOONBAND	https://music.163.com/user/home?id=342197871	342197871
傻子與白痴FoolandIdio...	https://music.163.com/user/home?id=1456859256	1456859256
西彬呀	https://music.163.com/user/home?id=323445735	323445735
OneRepublic_CN	https://music.163.com/user/home?id=1924705910	1924705910
棱镜乐队	https://music.163.com/user/home?id=6356356	6356356
动物园钉子户乐队	https://music.163.com/user/home?id=124108424	124108424

Sheet1 Sheet2 Sheet3

3 清洗数据、构建网络

A的ID	B的ID
109349540	73125610
109349540	4990043880
109349540	1649808342
109349540	31265745
109349540	388865241
109349540	109698685
109349540	3267561367
109349540	435244105
109349540	91724921
109349540	52829347
109349540	288477618
109349540	416608258
109349540	5158334
109349540	1671465495
109349540	9814155
109349540	3827086435
109349540	1409309284
109349540	75388755
109349540	260930930
109349540	1880071427
109349540	3773216955
109349540	73884324
109349540	358189202
109349540	590126810
109349540	98087305
109349540	3718571391
109349540	85065212
109349540	342197871
109349540	1456859256
109349540	323445735
109349540	1924705910
109349540	6356356

建立A(音乐人)→B(关注的人)的网络

大网络

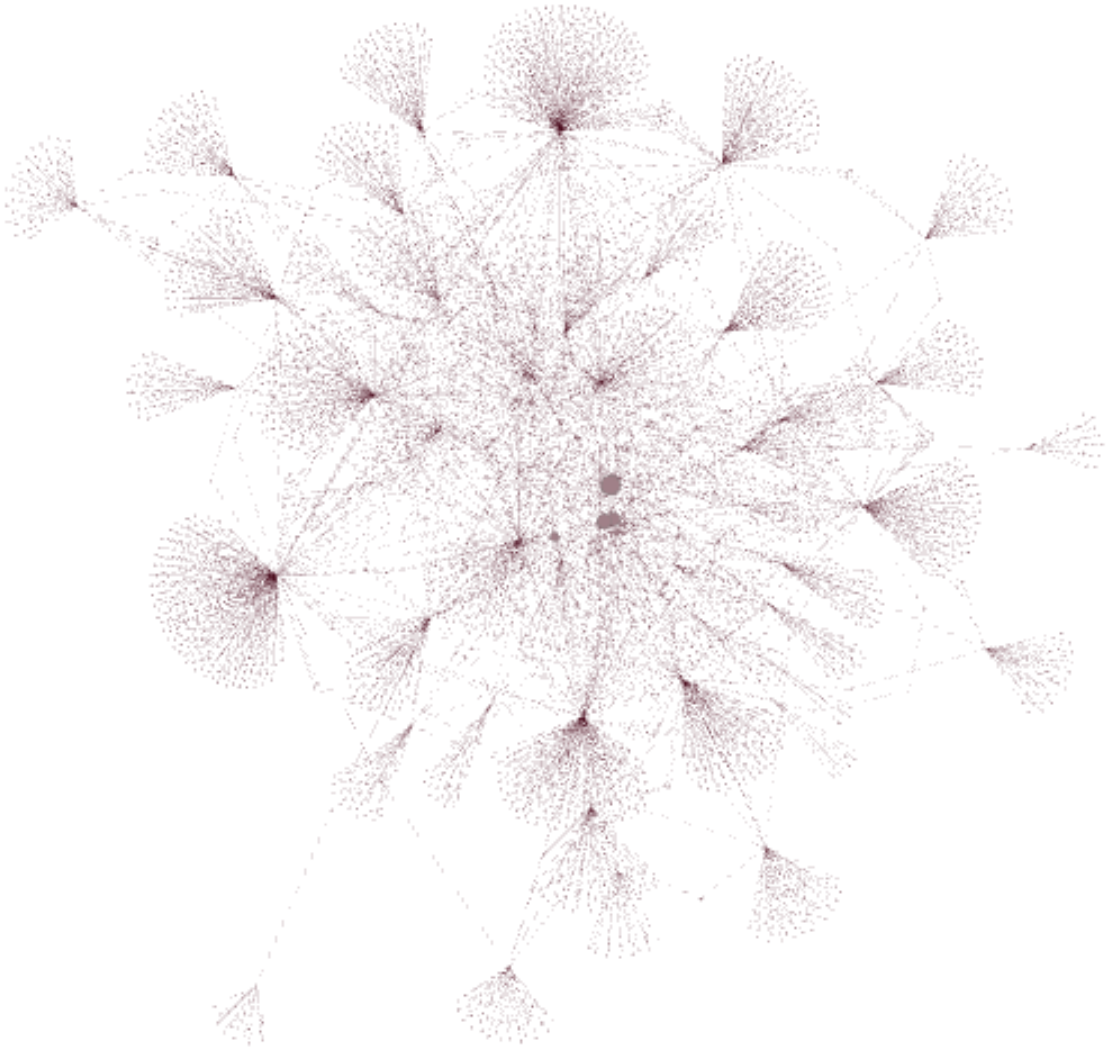
“A B 1”格式，表示A节点为源节点，B为目标节点，边的权重为1；
导入Gephi/pajek即可绘制网络。

小网络

A（源节点）、B（目标节点）的ID两列保存在Sheet2
Step1:如果Sheet2第j行第2列中ID不在Sheet2第2列中，则删除第j行；
Step2:根据关注关系，构建n(A)×n(A)的矩阵；
导入Gephi/pajek即可绘制网络

3 清洗数据、构建网络

大网络



3 清洗数据、构建网络

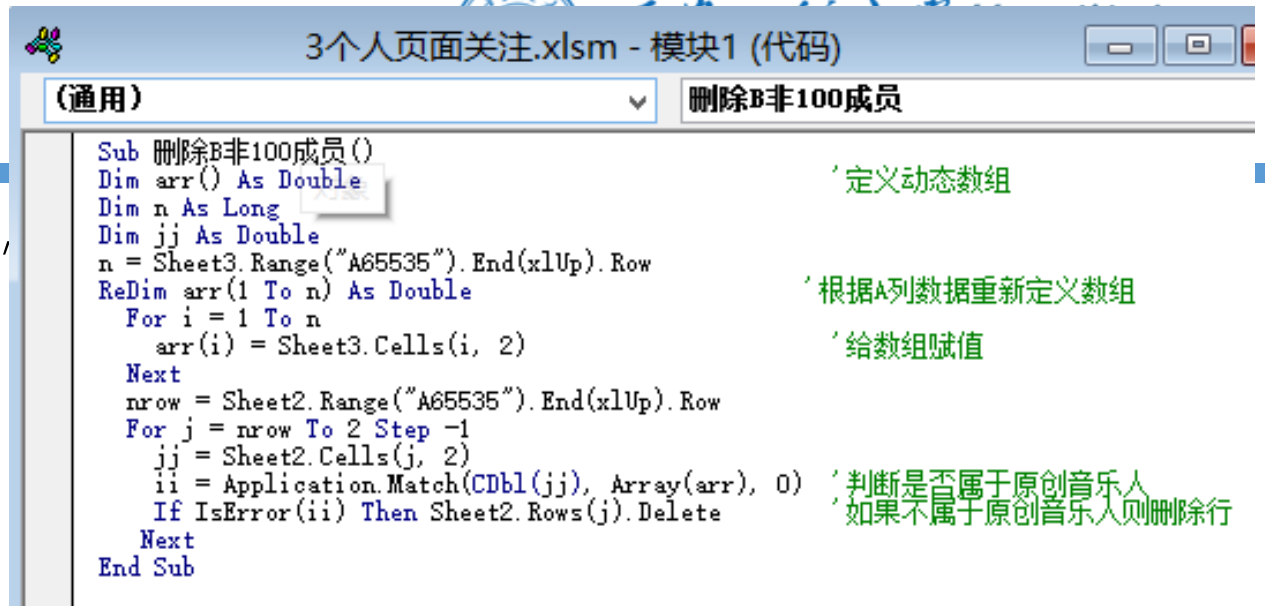
小网络

Step1: 如果B列中ID不在A列中，则删除该行；

将A (源节点，音乐人) 的信息复制到Sheet3)

	A	B	C
1		109349540	Todd Li
2		9814155	金玟岐
3		542203503	郑润泽
4		474348590	夏日入侵企画
5		550065289	要不要买菜
6		396416696	ODD陈思键
7		328902811	CMJ
8		3240550060	郭静
9		2688170	银临
10		558653866	时间不够以后
11		250310027	Hayrul海力
12		106523441	结冰水
13		421535995	Quinn葵因
14		270625553	程安
15		1525492313	宋美阅
16		1360962967	马也_Crabbit
17		86104955	Aricdontluv
18		40513452	徐远书
19		912583	AY楊佬叁
20		320958360	张雪飞
21		129086920	文兆杰
22		402563640	林达浪
23		1499458762	椅子不够酷
24		335940342	Mai
25		134899812	徐梓淳
26		415207075	浦东老农民
27		113910883	胡期皓
28		334310904	锦零
29		83956391	HARIKIRI

Sheet1 | Sheet2 | Sheet3 | 92网络

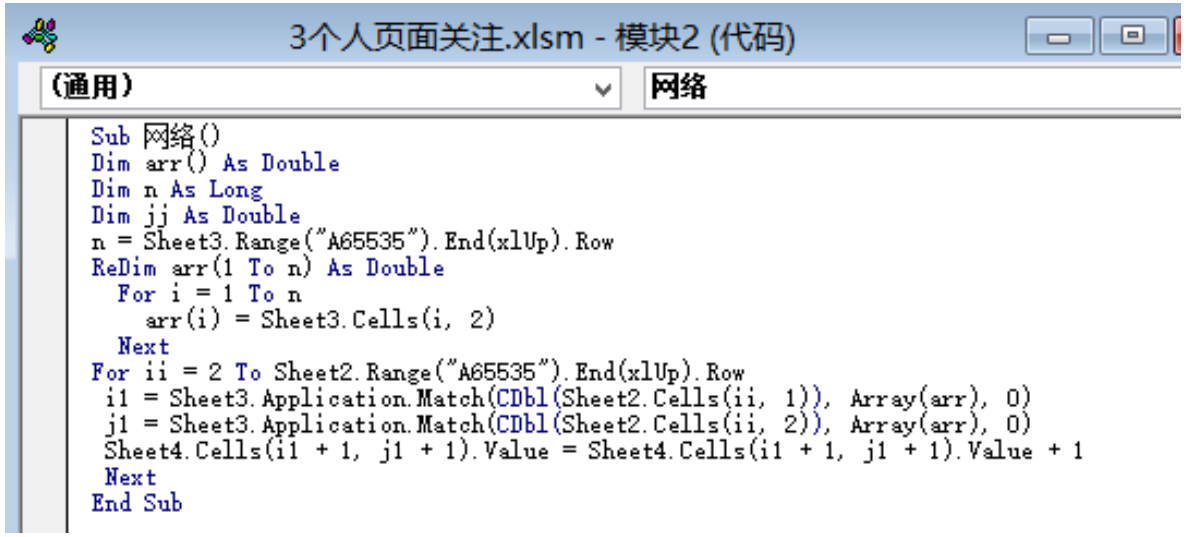



```
Dim arr() As Double
Dim n As Long
Dim jj As Double
n = Sheet3.Range("A65535").End(xlUp).Row
ReDim arr(1 To n) As Double
For i = 1 To n
   arr(i) = Sheet3.Cells(i, 2)
Next
nrow = Sheet2.Range("A65535").End(xlUp).Row
For j = nrow To 2 Step -1
   jj = Sheet2.Cells(j, 2)
   ii = Application.Match(CDbl(jj), Array(arr), 0)
   If IsError(ii) Then Sheet2.Rows(j).Delete
Next
```

3 清洗数据、构建网络

小网络

Step2:根据关注关系，构建n(A)×n(A)的矩阵；



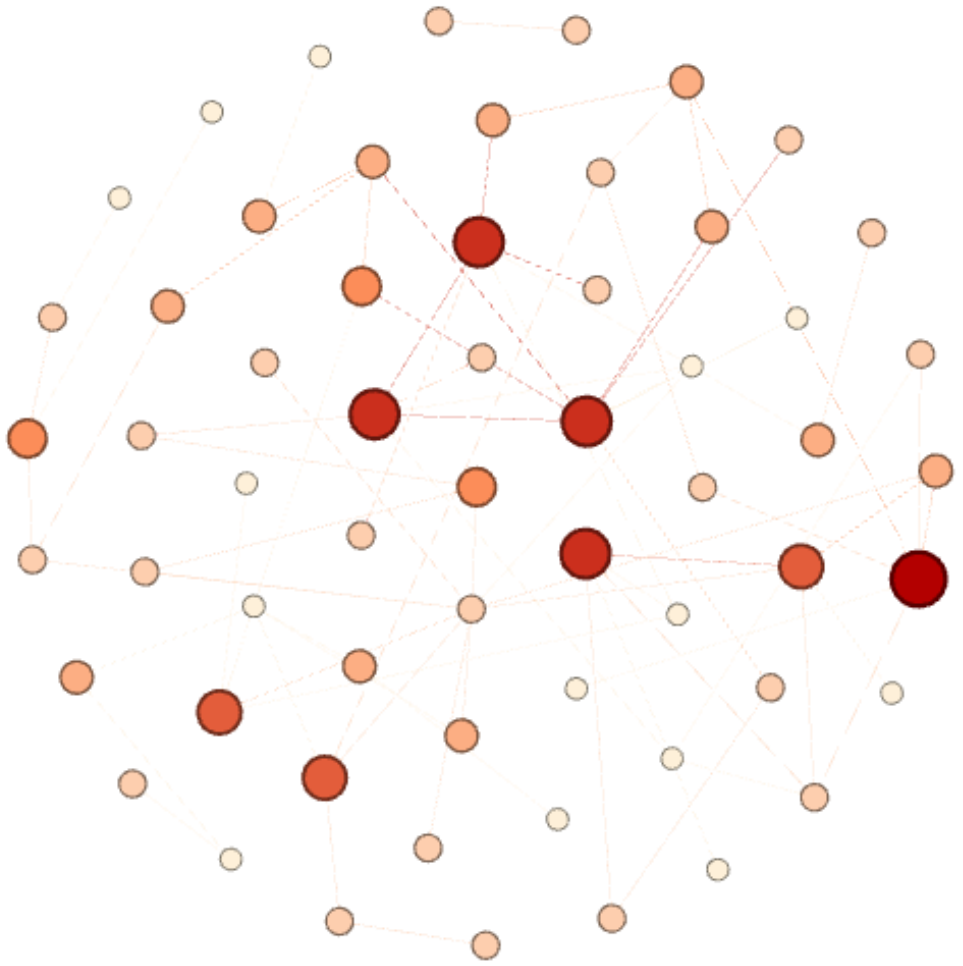

```
Dim arr() As Double
Dim n As Long
Dim jj As Double
n = Sheet3.Range("A65535").End(xlUp).Row
ReDim arr(1 To n) As Double
For i = 1 To n
    arr(i) = Sheet3.Cells(i, 2)
Next
For ii = 2 To Sheet2.Range("A65535").End(xlUp).Row
   i1 = Sheet3.Application.Match(CDbl(Sheet2.Cells(ii, 1)), Array(arr), 0)
   j1 = Sheet3.Application.Match(CDbl(Sheet2.Cells(ii, 2)), Array(arr), 0)
   Sheet4.Cells(i1 + 1, j1 + 1).Value = Sheet4.Cells(i1 + 1, j1 + 1).Value + 1
Next
```

3 清洗数据、构建网络

小网络

	1	2	3	4	5	6	7	8	9	10	11
1		1		1							
2											
3		1									
4	1										
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28							1				

Sheet1 Sheet2 Sheet3 92网络



谢谢