

通用科技文本可视化挖掘系统

Intelligent Insight

【ITGinsight】

使用手册

版本 V 2.1.0

20220522

目录

目录	I
第一章：功能与用户	1
1.1 系统简介	1
1.2 功能框架	1
1.3 同类型工具对比	4
1.4 对标的国际产品与 ITGinsight 的技术优势	5
1.5 适用用户	5
1.6 关于试用	5
1.7 关于数据处理规模	5
1.8 视频教程与技术支持	6
1.9 历史各版本新增功能	6
第二章：安装与运行	10
2.1 安装必备	10
2.2 安装过程	10
2.3 系统卸载	11
2.4 系统启动	11
2.5 系统注册	13
2.5.1 本地注册方法	13
2.5.2 网络注册方法	14
2.5.3 集团客户注册	14
2.5.4 保密版注册	14
2.6 在线升级	15
第三章：数据分析与可视化	15
3.1 数据格式转换/读取文献数据生成 itgn 文件	15
3.2 读取 itgn 文件进行可视化	19
3.3 合著关系可视化	20
3.4 同现关系可视化	22
3.5 耦合关系可视化	23
3.6 关联分析可视化	23
3.7 对应分析可视化	25
3.8 引证关系可视化	26
3.9 演化分析可视化	29
3.10 突破分析可视化	30
3.11 网络布局算法/选择合适的网络布局算法使网络图尽可能美观	31
3.12 关键信息过滤/删除不重要的连接线	31
3.13 更改图形样式/美化图形	32
1) 默认图形效果	32
2) 图形设置面板	32
3) 更改节点样式	33
4) 更改部分节点样式	36
5) 更改节点颜色	36
6) 按照关系强度、节点形状、节点名称、节点大小对节点染色	37
7) 更改节点边框颜色	37

8) 更改连线为直线或曲线.....	37
9) 更改连线颜色.....	40
10) 更改节点注释显示模式.....	40
11) 更改节点注释显示内容.....	41
12) 更改注释颜色.....	41
13) 更改节点字体.....	41
14) 更改节点文字大小写.....	41
15) 更改节点名称.....	42
16) 更改节点文字显示位置.....	42
17) 节点文字自动防重叠.....	42
18) 更改节点大小.....	43
19) 更改节点大小对比度.....	43
20) 更改聚类类别颜色.....	44
3.14 更改滑块设置.....	45
3.15 图形缩放、平移、拉伸、旋转.....	45
3.16 更改系统语言.....	46
3.17 更改背景色和背景边框.....	46
3.18 查找图形节点.....	46
3.19 删除/恢复节点.....	46
3.20 计算网络密度、节点中心度与主路径指标.....	47
3.21 屏幕截图.....	48
3.22 输出 Excel 数据表.....	49
3.23 Excel 报表输出内容设置.....	50
3.24 输出 Word 智能报告.....	50
3.25 输出 PPT 演示文稿.....	51
3.26 打开保存 Mod 图形文件.....	52
3.27 打开保存 layout 位置信息文件（位置信息的重复利用）.....	53
3.28 打开保存 graph style 样式信息文件（样式信息的重复利用）.....	53
3.29 可视化图形与文献数据交互.....	53
3.30 导出坐标.....	55
3.31 导出图例.....	55
3.32 一次性绘制所有可视化图形到 Word 中.....	55
3.33 快捷键.....	56
3.34 退出.....	57
第四章：聚类分析、热力图/地形图/密度图、世界地图、气象图、矩阵图可视化.....	58
4.1 网络图聚类分析.....	58
4.2 热力图/地形图/密度图可视化.....	59
4.3 世界地图可视化.....	64
4.4 中国地图可视化.....	64
4.5 矩阵图可视化.....	65
第五章：系统设置.....	66
5.1 使用过滤器切换分析对象.....	66
5.2 分析阈值、参数设置.....	67
5.3 停用词设置.....	68

5.4 主题词设置.....	69
5.5 人名辞典设置.....	70
5.7 国家名称辞典设置.....	71
5.8 省份辞典设置.....	71
5.9 词典内容大小写区分设置.....	71
5.10 在字典中应用正则表达式进行高级筛选和替换.....	72
5.11 第一次使用软件时如何设置字典.....	73
第六章：数据清洗	74
6.1 选择待洗数据源.....	74
6.2 数据一览.....	74
6.3 数据人工分组实现数据清洗功能.....	75
6.4 数据自动分组实现数据清洗功能.....	76
6.5 保存字典.....	77
6.6 用词典再次清洗数据、数据分析，自动分组。	77
6.7 保存清洗结果.....	78
6.8 组合分析（跨维度、跨层次的共现矩阵、引证矩阵）	79
6.9 分组统计（数据清洗后，按新分组进行统计）	81
6.10 文档聚类.....	81
6.11 数据链.....	86
第七章：辅助软件工具	87
7.1 INPADOC 同族专利可视化分析工具	87
7.2 美国专利引证可视化分析工具.....	87
7.3 美国专利权利要求解析工具.....	88
第八章：自定义结构化数据可视化	90
8.1 ima 格式数据.....	90
8.2 imb 格式数据	90
8.3 imc 格式数据.....	91
8.4 imd 格式数据	91
8.5 excel 格式数据（万能格式）	92
第九章：中英文科技术语识别（构建用户自定义主题词表）	94
第十章：与 VosViewer, Pajek, Ucinet 进行交互	96
第十一章：自动报告	97
第十二章：元数据	98
12.1 元数据导入导出.....	98
12.2 增加列.....	99
12.3 依据元数据对图形进行查找、更改节点大小、更改节点文字、绘制凸包	99
12.4 数据保存.....	99
第十三章：转化为参考文献	100
13.1 将文献按照参考文献格式输出到 WORD.....	100
附录：	101
附录 A.合著关系/同现分析/耦合分析.....	101
附录 B.对应分析	101
附录 C.关联分析	101
附录 D.引证关系.....	102

第一章：功能与用户

1.1 系统简介

通用科技文本可视化挖掘系统，英文缩写为 ITGinsight，英文全称为 Intelligent Insight，其 V1.0 版本软件著作权登记号为 2011SR075074，V1.1 版本软件著作权登记号为 2016SR171854，V1.3 版本软件著作权登记号为 2018SR745969，V1.5 版本软件著作权登记号为 2018R11L1049159，V1.9 软件著作权登记号 2021SR1062611，V2.0 软件著作权登记号 2022SR0727330，由北京正乙科技有限公司(www.zyinsight.com)设计开发，资源下载地址 www.itginsight.com。

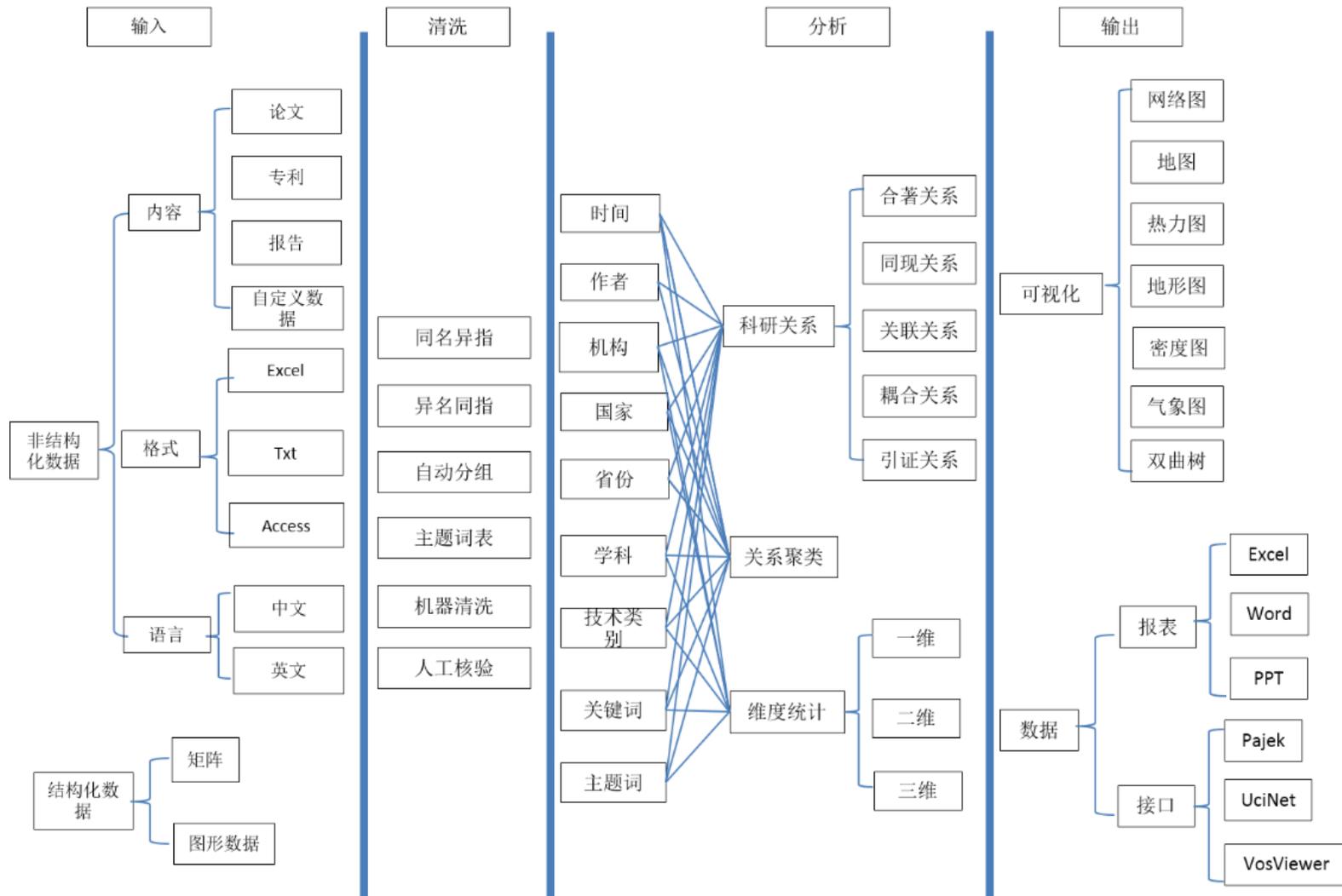
该软件是一款高级的科技文本挖掘与可视化分析工具，主要针对科技文本，如专利、论文、报告、报刊等进行可视化的分析与挖掘，也可应用于微博、微信等互联网文本数据可视化，可视化挖掘方法有合作关系可视化、同现关系可视化、耦合关系可视化、关联关系可视化、引证关系可视化、演化分析可视化，可视化输出包括网络图、热力图、密度图、世界地图、矩阵图、演化图、聚类图。该工具增强了对大规模数据的处理，将聚类分析、技术热力图、技术地形图、技术气象图整合到系统中。

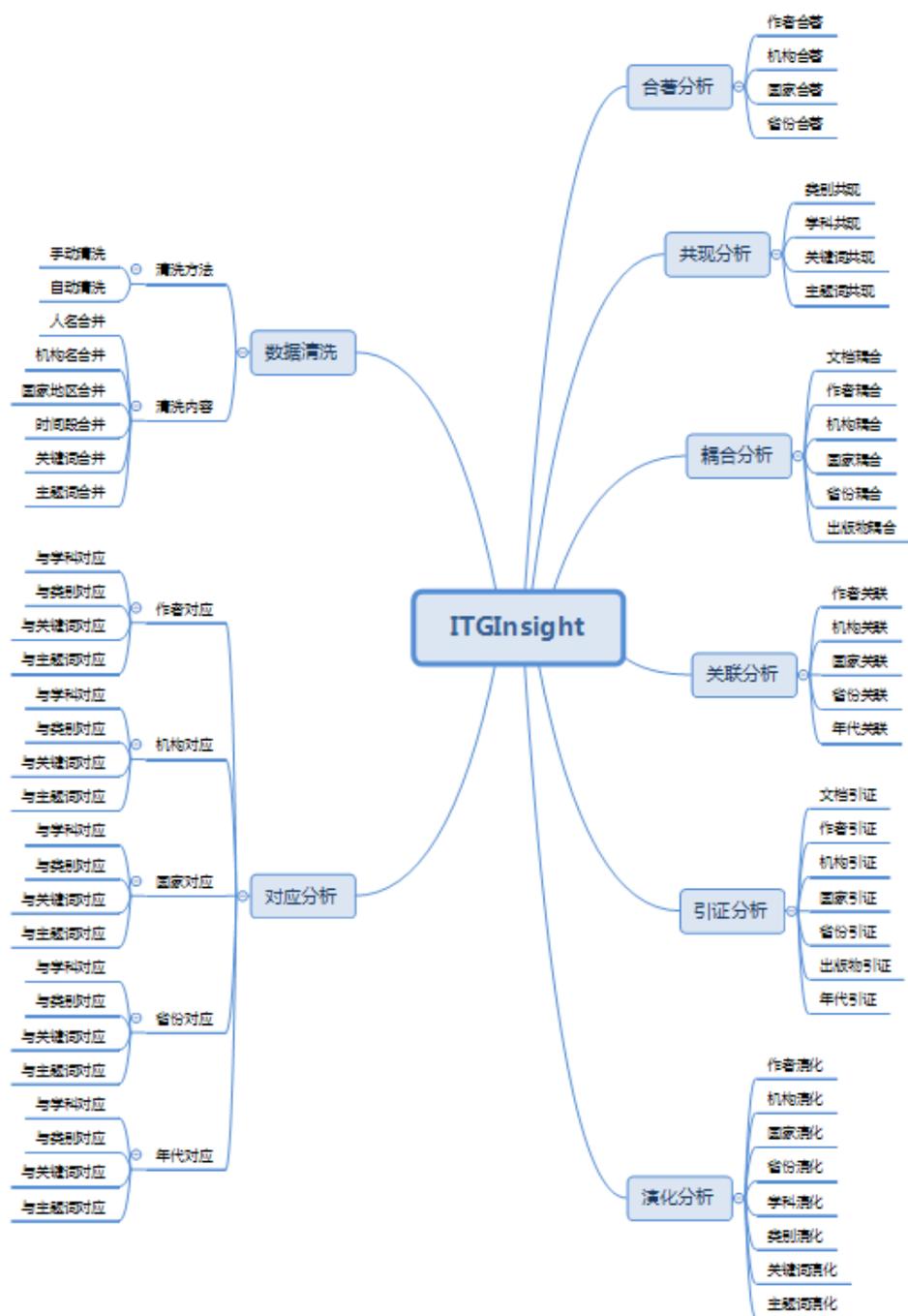
用户可应用该工具对 SCI、CNKI、万方论文数据，德温特专利、美国专利、中国专利、欧洲专利，科研项目、技术标准、产业政策等数据进行可视化挖掘，进而开展学术评价、技术监测、技术机会分析、竞争态势分析等科研管理与情报分析任务。同时，该工具也是一款综合的情报分析平台，提供除文本挖掘和可视化分析以外的基本维度统计、excel 报表、word 智能报告、ppt 可视化输出等辅助功能。

该系统支持用户自定义格式的任何文本数据、图形数据，并提供与情报分析工具 VOSviewer、复杂网络工具 Pajek、UCINET 的数据接口、使用接口。

1.2 功能框架

该系统的功能框架可参考如下两幅图，具体操作过程见第三、四章。



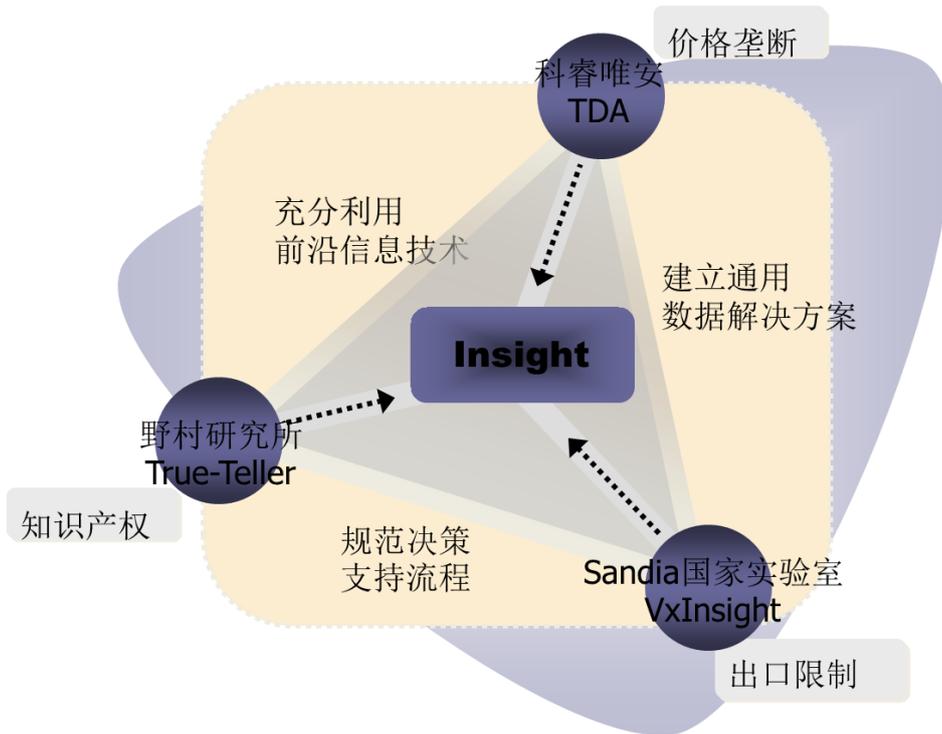


1.3 同类型工具对比

序号	软件	归属	功能类型	分析数据				分析方法						可视化输		自动报告		
				数据源	数据类型	数据清洗	用户词表	基本统计	合作分析	共词分析	引证分析	关联分析	演化分析	可视化图形	交互接口	自动报表	自动报告	
1	UCINET	美国 加州大学	可视化展示工具	任意	结构	无	无	有	无	无	无	无	无	无	统计图、网络图	弱	无	无
2	Pajek	斯洛文尼亚 卢布尔雅那大学	可视化展示工具	任意	结构	无	无	无	无	无	无	无	无	网络图、树图	强	无	无	
3	Vxinsight	美国 Sandia国家实验室	可视化展示工具	任意	结构	无	无	无	无	无	无	无	无	网络图、主题图	强	无	无	
4	CiteSpace	美国 Drexel 大学 陈超美	基于文本的可视化分析软件	任意	结构/非结构	有	无	无	有	有	有	无	无	网络图、地图	强	无	无	
5	True-Teller	日本 野村研究所	基于文本的可视化分析软件	任意	结构/非结构	有	无	无	无	有	无	无	无	热力图、网络图	弱	无	无	
6	VOSviewer	荷兰 莱顿大学的科学和技术研究中心	基于文本的可视化分析软件	任意	结构/非结构	有	无	无	有	有	无	无	无	热力图、网络图、聚类图	弱	无	无	
7	Vantage-Point	美国 GIT技术政策与评估中心	基于文本的可视化分析软件	任意	结构/非结构	有	有	强	有	有	无	有	无	统计图、矩阵图、网络图	强	有	有	
8	Thomson Data Analyzer	美国 汤姆森·路透	基于文本的可视化分析软件	任意	结构/非结构	有	有	强	有	有	无	有	无	统计图、矩阵图	强	有	有	
9	ITGinsight	中国	基于文本的可视化分析软件	任意	结构/非结构	有	有	强	有	有	有	有	有	热力图、网络图、矩阵图、聚类图、演化图、双曲树	强	有	有	

1.4 对标的国际产品与 ITGinsight 的技术优势

国内无同类产品，国外加拿大科睿唯安 DDA(原来汤姆森路透 TDA)、日本野村研究所的 True-Teller、美国 Sandia 国家实验室的 VxInsight 与 ITGinsight 具有相似的功能。相比国外产品，ITGinsight 在术语识别、中文支持、数据处理数量，以及可视化展示的美观程度上具有显著的技术优势。



1.5 适用用户

- 高校图书馆
- 科技情报研究所
- 企业工程技术人员
- 企业知识产权管理决策人员
- 高校、科研机构教师、学生
- 其它情报分析师、知识产权分析师、咨询师、代理人、律所

1.6 关于试用

软件分为保密版，企业版，教学版（学术研究版）和学生版（社区版）。学生版（社区版）下载地址为 www.itginsight.com，不用注册，仅限学生撰写论文，上传用户数据，无技术支持，其他版本均针对付费用户。

1.7 关于数据处理规模

软件分为 32 位和 64 位版本，一般情况下 64 位版本在 8G 内存的普通计算机支持至少

10 万条数据分析/清洗，16G 内存最高可支持至少 15 万条数据的分析/清洗。在实际使用过程中，用户使用 256G 内存，24 核心 CPU 可处理 250 万-450 万条文献。

文本聚类分析，普通 8G 内存计算机可支撑 2 万条专利或论文的聚类，提高计算机配置可提高聚类专利或论文数量。

1.8 视频教程与技术支持

本软件提供详尽的视频教程，<http://cn.itginsight.com/course/>

本软件根据用户级别提供技术支持，线上技术支持为 QQ 群：908179419，对于企业级最高权限用户提供现场技术支持和培训。

1.9 历史各版本新增功能

V2.1.0.0 新增功能如下：

- 1) 增加突破分析
- 2) 开放 SCI/SSCI 二级单位过滤器
- 3) 数据清洗增加分组统计
- 4) 在 Dataset 页面增加 Doc To Excel, Doc To Txt, 数据清洗读取数据的 2Excel 方式修订，速度加快
- 5) bigdataset, Sqlite 模式，单台计算可处理百万级数据
- 6) 作者、机构、关键词增加语义指纹，和快速模式数据清洗
- 7) 节点名称修改支持多个正则表达式替换

V2.0.0.0 新增功能如下：

- 1) 数据清洗读取的数据量和读取时间性能提升 50 倍以上，自动分组速度提升 5 倍以上
- 2) 数据清洗页面，元数据页面整合到主页面；增加浅色和深色外观
- 3) 增加技术竞争力指标
- 4) 增加 Recluster 功能、节点分组、聚类的修改
- 5) 增加聚类图与 DataSet 之间的交互，标签修改 ReTag 功能
- 6) 增加 Kmeans (N) 聚类
- 7) 增加聚类结果自动打标签功能
- 8) 数据清洗增加对四种模式的分词处理
- 9) 数据清洗时，增加只读第一作者/第一单位/第一国家等读取设置
- 10) 增加 PCA、KPCA 降维

- 11) 增加 ToReference 功能，即将文献著录项目转化为参考分析格式
- 12) 可视化文字增加大小写转化 2Upper, 2Lowe, 2Cameral 功能
- 13) 组合分析结果增加 txt 格式数据输出, 节点列表增加 2Txt, 避免 excel 格式输出时, 由于 office 安装问题引起的闪退
- 14) 修正演化图保存 mod 文件格式不完全的 BUG
- 15) 修复节点过小引起的闪退 BUG
- 16) 修正了技术主题图重绘后, Tag 标签位置没变化的 BUG
- 17) 修复进度条假死 bug
- 18) 白名单验证模式, 对高校范围内的 IP 地址批量授权

V1.9.1.0 新增功能如下:

- 1) 字典增加大写、小写、大小写混合、首字母大写等模式; 正则表达式筛选
- 2) 修订中介中心度计算
- 3) 增加中英文混合的数据分析, 词长度阈值分中文、英文两个参数
- 4) 增强快捷键的操作
- 5) 优化数据读取速度, 在大规模数据和大规模词典应用时, 读取速度提升 40%左右
- 6) 增加聚类主题地形图
- 7) 增加图密度指标
- 8) 英文词语全部按照单数处理, 并且增加词表的正则表达式替换功能
- 9) 增加 Scopus 数据支持
- 10) 删除世界地图和中国地图的背景, 近保留坐标布局, 删除中国省份地图布局
- 11) 集团用户登录数量提示
- 12) 修正数据清洗, 中文分组中有多余空格的问题

V1.9.0.0 新增功能如下:

- 1) 增加图例, 在 PPT 中绘制图例
- 2) 过滤器中的摘要和关键词字段支持多个字段合并, 用|分割多个字段
- 3) 自动报告增加 Patsnap、Incopat、Innovation 数据源支持
- 4) 自动报告表格居中
- 5) 当数据时间字段缺失时, 可视化去掉了 1900 的数据
- 6) 增加对高被引论文的过滤, 通过设置 Number1、Number2、Number3 阈值来过滤论文或专利中高被引、高同族数据, 使得仅超过 Number1、Number2、Number3 阈值

的数据才会被分析、被清洗

- 7) 增加针对单个机构文献数据的自动分析报告模板，多个机构文献对比分析的自动报告分析模板
- 8) 字典增加了反向筛选功能，即只有词表中含有的人名、机构、国家、省份、类别、关键词等，才进入分析
- 9) 增加绘图机器人操作，一次性绘制所有可视化图形
- 10) 自动分组部分时在状态栏增加进度提示

V1.8.0.0 新增功能如下：

- 1) 增加 class3、class4
- 2) 优化自动报告引擎，关联分析部分，报告模板组件构成
- 3) aiReview.onlie\SciReport.online 服务开通

V1.7.0.0 新增功能如下：

- 1) 项目合著分析
- 2) 分析部分，删除了孤立点相关的处理操作
- 3) 单 IP 提醒功能
- 4) 增加 LinLog 布局，并设定为首选布局
- 5) 自动报告改为全自动，生成单独系统 ezReport，并设立独立权限
- 6) 增加默认的参数最优化设定，后台计算
- 7) 过滤器中的摘要字段支持多个字段合并，用|分割多个字段

V1.6.0.0 新增功能如下：

- 1) 增加了元数据功能，类似于 GELPHI 的增加列
- 2) 增加了 6 种主题图的表现形式，类似于 VOSViewer 的主题图
- 3) 增加了聚类密度图
- 4) 增加了标签防重叠功能
- 5) 增加了节点大小对比度参数 sizevariation
- 6) 增加了面板边框大小设置，目的：截取密度图、热力图、聚类图，能够截取全部图形
- 7) 增加坐标直接导出功能
- 8) 增加数据链的功能，即在数据清洗模块增加 Go To Related 功能
- 9) 增加了 TSNE 布局

10) 增加浮动窗口

V1.5.0.9 新增功能如下:

- 1) 新增节点大小批量修改功能
- 2) 增加普通电脑进行高清截图功能
- 3) 增加 SCI 论文自动报告功能
- 4) 版本划分为: 学生版、学术版、教学版、企业版、集团用户版、军工版
- 5) 增加批量“显示或隐藏节点名称”功能
- 6) 修正 BUG: 配置文件错误强制退出
- 7) 运行用户自定义密度图颜色
- 8) 增加全英文自动报表, 全英文用户手册

V1.5 新增功能如下:

- 1) 新增三维统计
- 2) 新增数据清洗后文档聚类, 密度图可视化
- 3) 新增智能报告
- 4) 增加了 docadapter 模式, 读取数据不分析; 再次读取 docadapter 后再分析的模式
- 5) 增加网络图的凸包覆盖
- 6) 增加集团客户免注册功能
- 7) 增加.netx 格式文件的处理
- 8) 增加图形左右、上下翻转功能
- 9) 增加 excel 格式的频数矩阵、相似矩阵、皮尔森矩阵可视化

V1.3 新增功能如下:

- 1) 新增主题演化分析, 跟踪技术的产生、消亡、增强、减弱、聚合和裂变的过程
- 2) 新增机构、作者、国家、省份、关键词、技术类别演化分析, 拓展主体演化分析范围
- 3) 新增 SPC 主路径指标, 识别技术发展过程中的关键技术节点
- 4) 增加人名的同名异指和异名同指的计算机识别

V1.2 新增功能如下:

- 1) 全新报表引擎, 一键输出近百张分析报表, 全方位洞悉数据特征
- 2) 增强语义分析, 自动识别相似主题词, 相似机构名, 人名, 地名
- 3) 智能组合分析, 任意二维、三维数据矩阵, 跨维度跨层次可视化
- 4) 优化渲染技术, 增加技术云图, 知识扩散, 功效矩阵, 中国各省地图

第二章：安装与运行

2.1 安装必备

- 操作系统:

Window7 及以上操作系统, Office2013 或以上; 32 位软件对应 32 位版本 office, 64 位软件对应 64 位 office

- 硬件配置:

内存: 1G 以上; 硬盘: 100M 以上; CPU: 主频 1G 以上。

2.2 安装过程

- Installer3.1 安装, 如果计算机无 Intaller3.1, 需进行网络下载安装。系统会自动连接网络进行下载, 无需用户做特定操作。或点击安装目录 WindowsInstaller3_1 文件夹下的 WindowsInstaller-KB893803-v2-x86.exe 手动安装。
- NetFramework4.5 安装, 如果计算无 NetFramework4.5, 需进行网络下载安装。系统会自动连接网络进行下载, 无需用户做特定操作。
- ITGinsight 下载地址为 <http://cn.itginsight.com/download/>, ITGinsight 绿色版不需要安装, 解压缩后直接找到.exe 文件即可运行。非绿色版安装, 点击安装文件夹下的 setup.exe, 顺序弹出如下对话框:





2.3 系统卸载

绿色版直接删除文件夹即可。

非绿色版卸载，打开“控制面板”→“添加删除程序”→“更改或删除程序”，在“当前程序中”找到 ITGinsight，如下图：



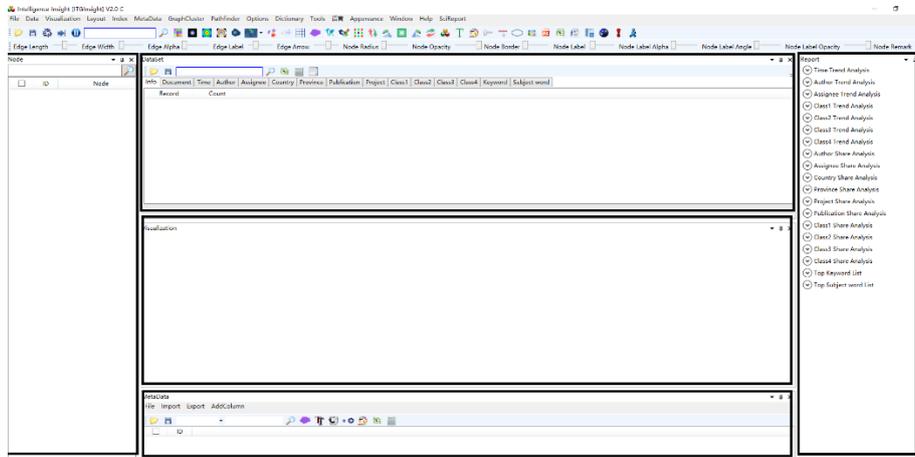
点击“删除”按钮即可。

2.4 系统启动

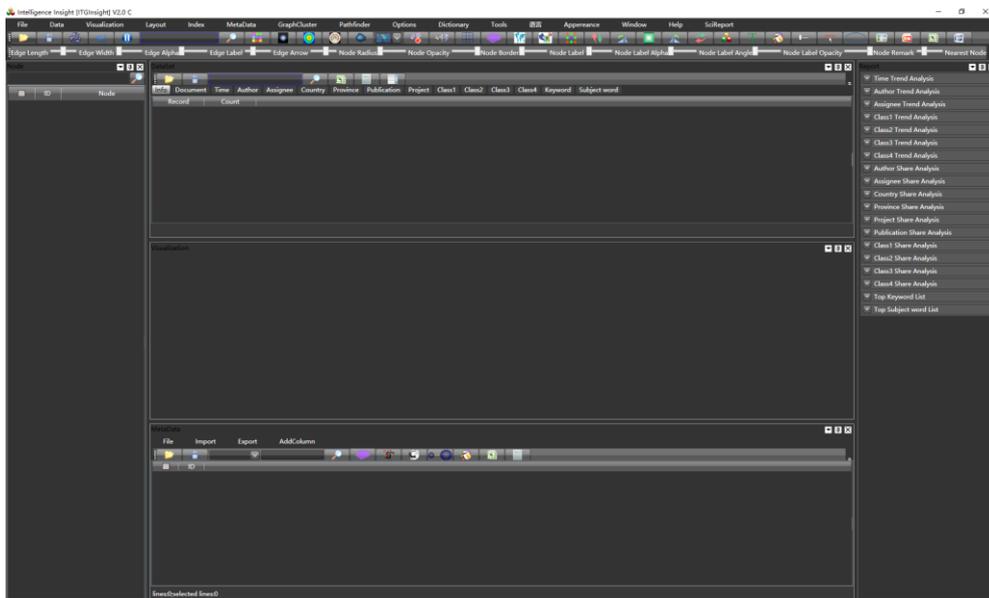
系统安装后将在“桌面”和“程序”中放置启动的快捷方式，如下图所示：

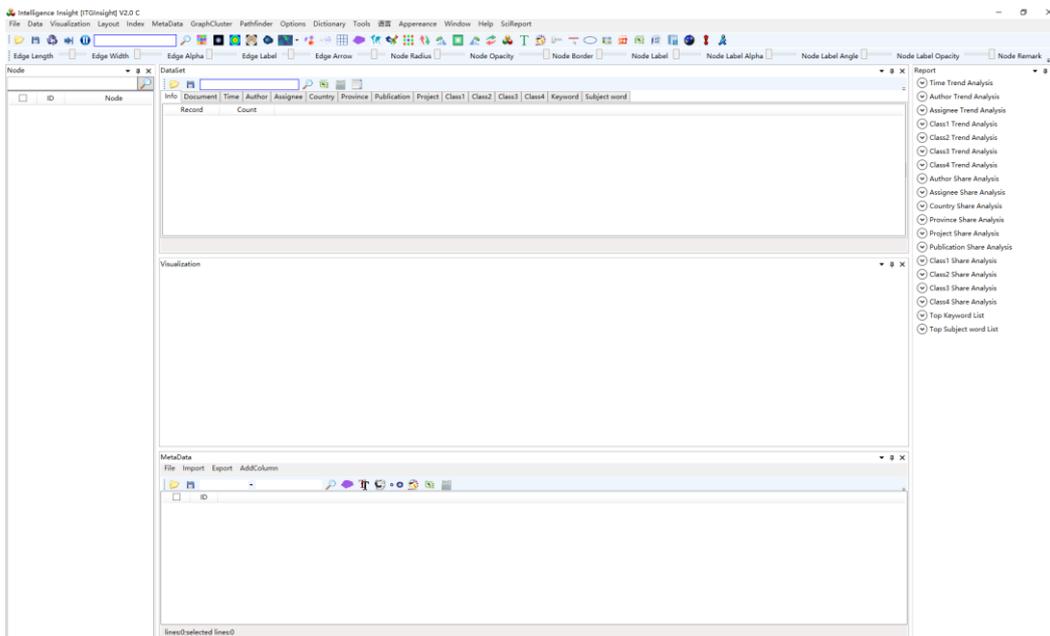


启动后，软件主要窗口区域包括可视化区域、数据集区域、元数据区域、节点区域、报告区域，如下图所示，默认只显示可视化区域、节点区域、报告区域，通过菜单栏的 Window/窗口按钮，进行各个区域的显示设置。



软件支持浅色和深色两种外观，通过菜单栏的 Appearance 外观按钮进行切换，深浅外观如下图所示。





2.5 系统注册

一般情况下，系统仅需网络注册，极特殊情况需要本地注册+网络注册。如果软件能够正常启动，说明本地注册已完成，只需要网络注册。绿色版一般都不需要本地注册。

社区版/学生版不需要任何注册。非学生版需要网络注册。

2.5.1 本地注册方法

1) 运行软件安装目录下子目录 hid 中的 HID.exe 文件，得到计算机的序列号：

 HardwareID	2015/12/24 11:00	TXT 文件	1 KB
 HID	2015/10/26 9:27	应用程序	62 KB
 HID_For_Liscence	2015/10/26 9:27	WinRAR ZIP 压缩...	29 KB

2) 将“机器码”连同“机构”、“用户”、“邮箱”等信息一起发送到客服邮箱；

3) 客服在收到注册信息并通过验证后，发给具有时间限制的授权文件到用户邮箱，时间限制一般为“一个月”，如需延长时间，需另外说明。

未进行本地注册的用户在使用时会定时弹出“授权警告”窗口，如下图，



4) 软件技术支持 QQ 群: 908179419, www.itginsight.com 会发布通用的本地注册文件，该授权不与计算机硬件绑定，任何用户均可进行本地注册。

2.5.2 网络注册方法

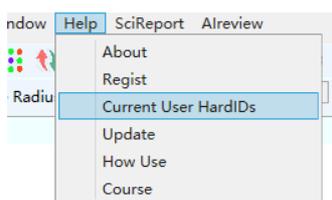
- 1) 完成本地注册
- 2) 运行软件，点击 help->register 弹出如下画面



- 3) 将“机器码”连同“机构”、“用户”、“邮箱”等信息一起发送到客服邮箱；
- 4) 由客服完成网络注册。未进行网络注册的用户，系统会在 5 分钟自动退出。

2.5.3 集团客户注册

集团客户，即不需要网络注册，也不需要本地注册，但是当同时登陆用户数量超过集团购买数量，软件启动后会弹出当前登录用户总数量情况。比如，集团购买 5 个账号，仅能保证 5 个人同时在线。在集团用户登录时，会对登录情况进行验证，如果达到用户上限，系统提示当前达到上限，并显示已经登录的硬件 ID，当前用户可以对已经登录的硬件 ID 进行强制退出，否则当前用户登录后也会由于达到用户上限被退出。集团用户登录后也可以通过帮助查看所有已经登录使用的用户硬件 ID，如下图。



Current User HardIDs/Max User Allowed:2		
HardID	LastOperate	Operate
92528D6CBFEBFF000506E3	2021/8/5 21:30:52	LogOut

2.5.4 保密版注册

保密版需本地注册，无需网络注册，一机一码，无需联网，适用于对数据敏感或者有保密资质的单位使用。保密版本不能连接互联网，否则会强制退出，但可以连接内网。

2.6 在线升级

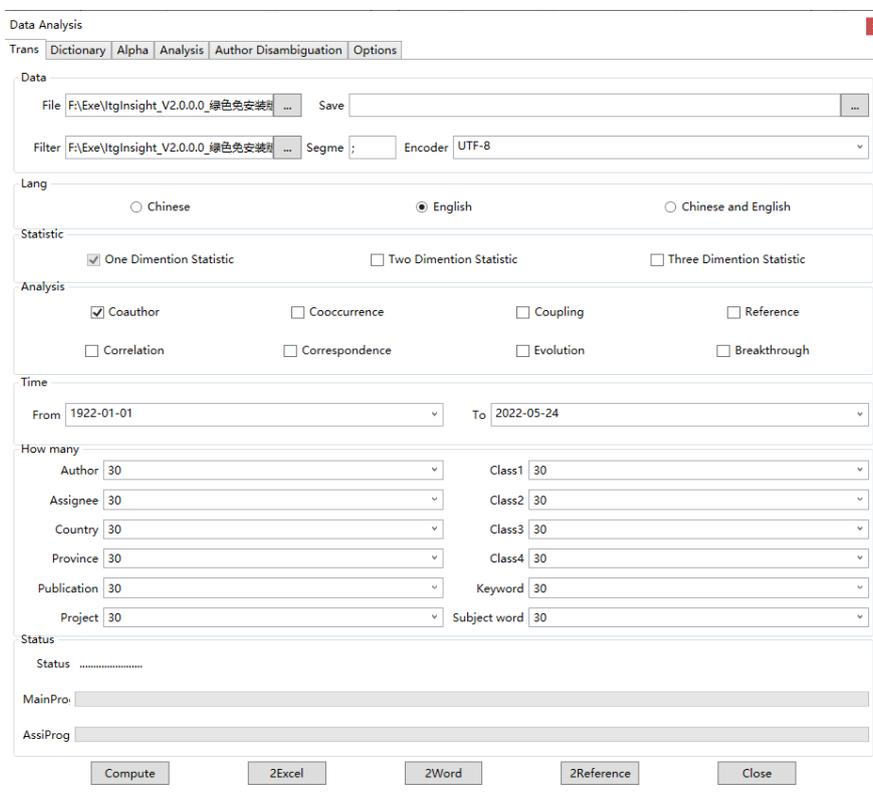
点击“help/帮助”->“update/更新”，在联网的环境下，系统会自动检查软件版本，进行系统升级，升级过程中，确保 ITGinsight 处于关闭状态。在线升级较慢，建议到技术交流群或者资源网站下载最新版。

第三章：数据分析与可视化

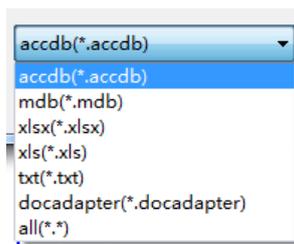
3.1 数据格式转换/读取文献数据生成 itgn 文件

应用 ITGinsight 进行数据分析，首要的工作便是将文献数据转化为 ITGinsight 的数据格式，并应用数据转化功能进行数据的分析。

点击菜单栏上的“Data/数据->Analysis/分析”，弹出数据转化页面，如下图：



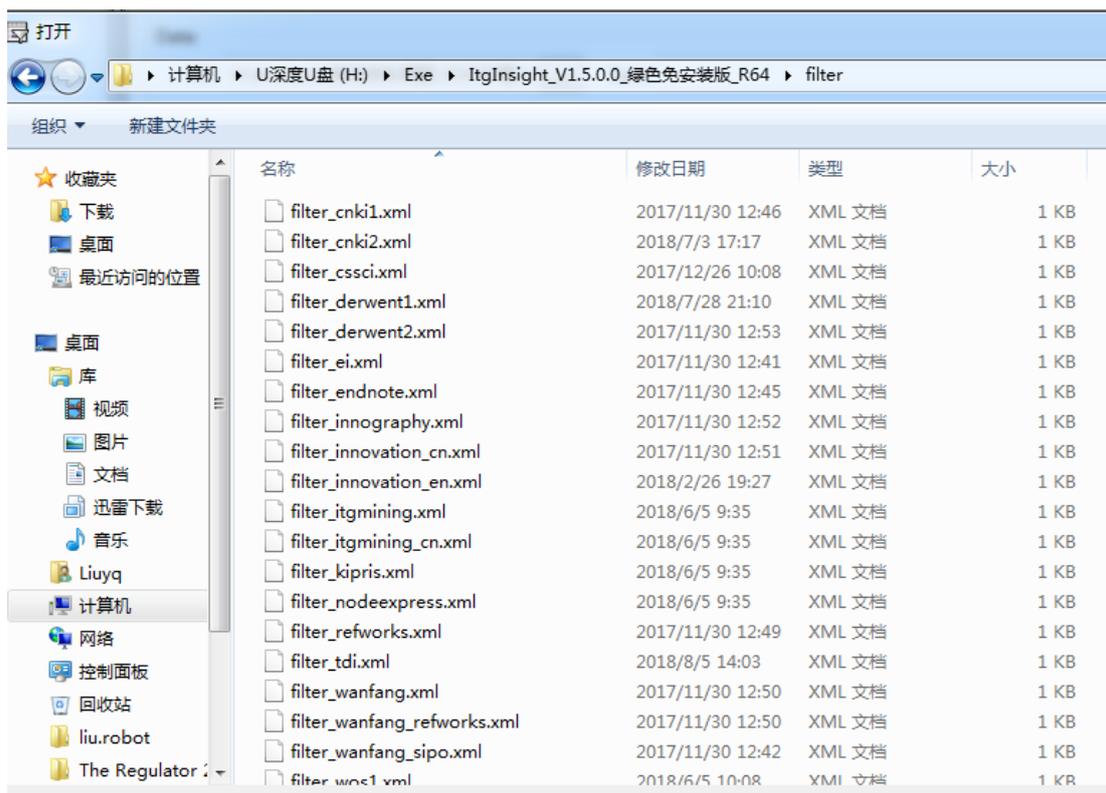
“Data 数据” 标签下的“File/文件”处，点击，弹出数据导航对话框，选择数据来源，如下图：



系统支持由 CNKI 下载的中文核心期刊数据，参考安装目录下的 example_data_cnki.txt；

由 Web Of Science 下载的 SCI 论文数据和德温特专利数据，参考安装目录下的 example_data_wos.txt；由专利分析软件 ItgMining 导出的专利数据，参考安装目录下的 example_data_itgmining.xls 或 example_data_itgmining.accdb 等样例数据。数据文件可以是 Excel03、07 及以上格式，Access03、07 及以上格式，txt 格式。同时，数据文件也可以是 docapadter 格式，该格式是由 ITGinsight 生成的数据文件。

在“Filter/过滤器”处，点击 ，弹出过滤器选择导航对话框，选择过滤器，如下图：



从中选择过滤器。比如，当数据是由 itgmining 导出的，那么过滤器就选择 filter-itgmining，这样系统就知道数据来源，从而，采用对应的数据处理规则，如果是 SCI 数据，过滤器就选择为 filter-wos，以此类推。

在“Segment/分隔符”一栏填写分割符号，系统默认为“;”，如果有多个分割符，同时在该处填写。

当被分析对象的一个记录中含有多个记录，比如“作者”，在数据库中一条记录有多个作者，并且用“;”分割，那么在分析时，系统会依据这个“;”分隔符把所有作者识别出来。

当选择的数据为文本 txt 格式时，“Encoder/编码”一栏发挥作用，系统根据编码内容进行文本的解析，如果 Encoder 的设置与数据 txt 的真实编码不一致，系统会无法正确分析文本内容。“Encoder/编码”的设置可以下拉选择，也可以手工输入。

“Save/保存”一栏，点击，填写文件保存的路径和文件名，系统默认 itgn 为文件后缀，该文件是用以进行可视化分析的项目文件。

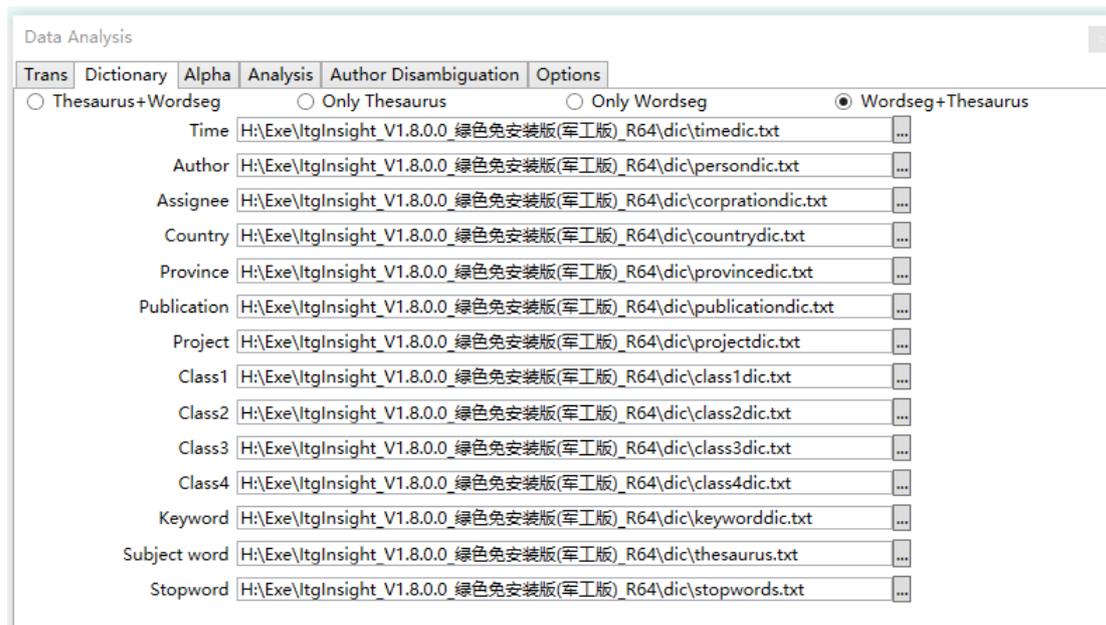
在“Statistic/统计”标签下，选择统计分析的维度，一维统计为必选项，二维统计、三维统计为可选项；当选择后面的关联分析后，二维统计自动成为必选项；当选择二维统计、三维统计后，分析的时间会有所增加。

在“Analysis/分析”标签选择要进行的分析内容，“Coauthor/合著分析”“Cooccurrence/同现分析/”“Correlation/关联分析”“Correspondence/对应分析”“Reference /引证分析”等，可多选。

在“Time/时间”标签设定被分析数据的起止时间。

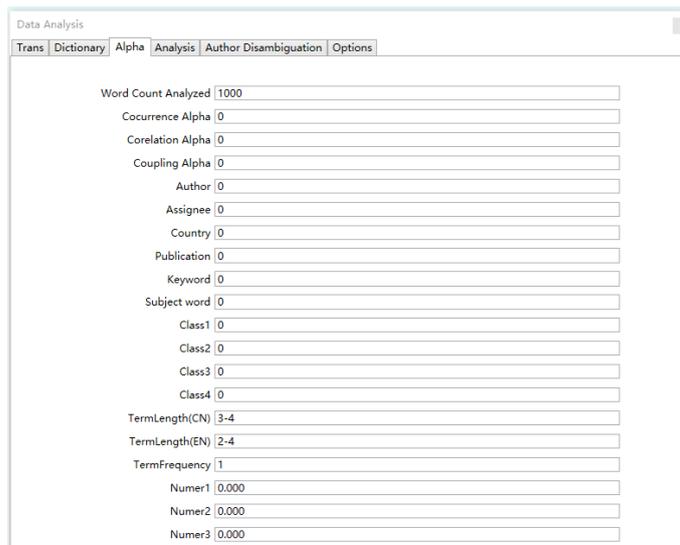
在“How many/多少项”标签下，输入将要分析的机构、作者、国家、类别、期刊、关键词、摘要词数目，分析数目按照数量多少排序。

切换到 Dictionary 标签，如下截图：



选择字典，首次使用用户可以到软件安装目录下 dic 目录中找到相关的字典文件。

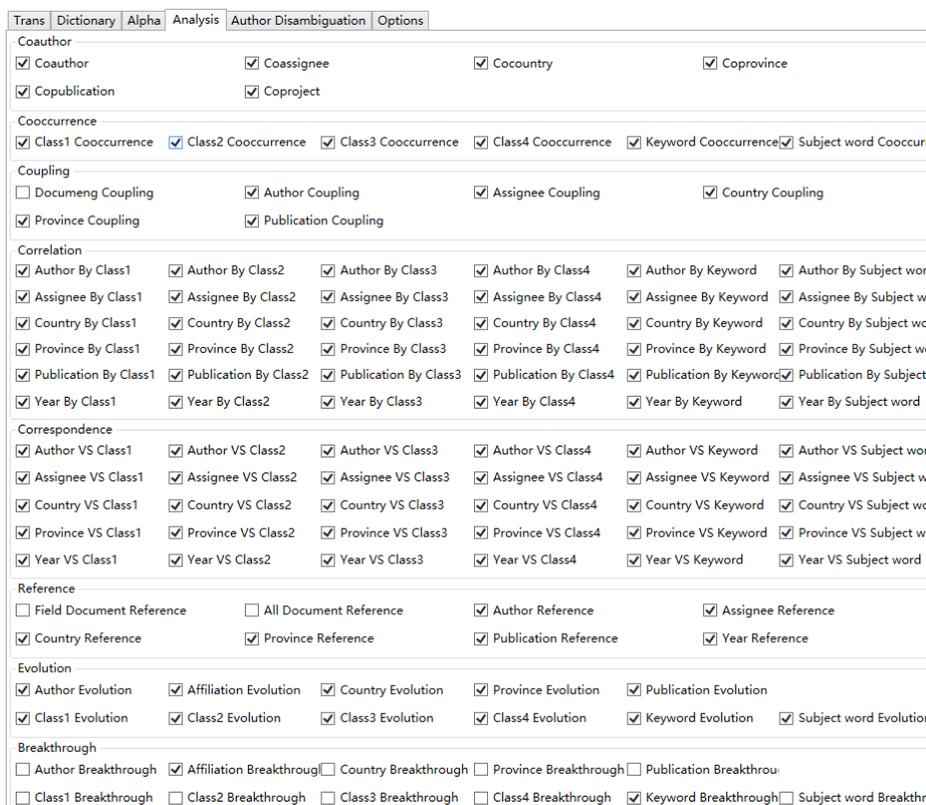
切换到 Alpha 标签下，如下截图：



首次使用用户，保存默认不变。其中，TermLength 和 TermFrequency 是提取主题词的词长、词频限制，英文建议词长取,2，中文词长取 3，当数据量比较大时提高词频阈值可以加快分析速度。

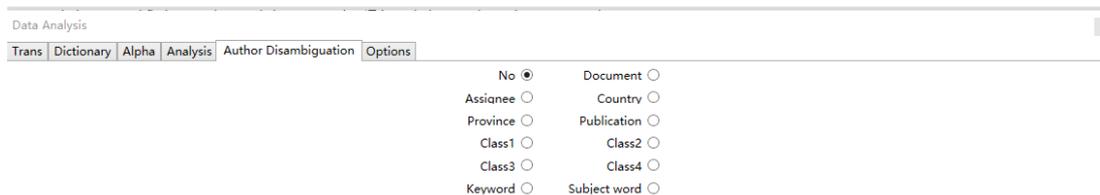
Number1 的阈值含义：以 SCI 论文为例，当 Number1 设置为 3 时，如果被引次数超过 3，参与分析；低于 3 的论文会被过滤掉，不参与分析。以此类推，设置 Number2、Number3 的阈值，但建议这两个设置 0。

切换到 Analysis 标签下，如下截图：



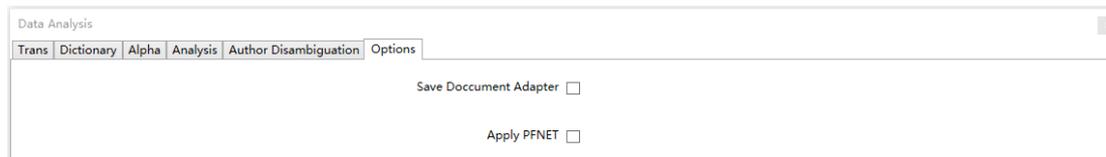
首次使用保持默认不变，如果数据量较大，Document Reference/文档引证分析会耗费较长时间，该项去掉后，分析时间会加快。

Author Disambiguation/作者消歧标签如下：



含义是说，如果遇到同名异指，即同一个名字不同人的情况，如何处理，默认 No/不区分，认为是一个人。如果选择 Assignee/机构，则进一步区分作者是否为同一个人，看文献的机构信息，其他字段选择的含义类似。

切换到 Options/选项标签，如下图：



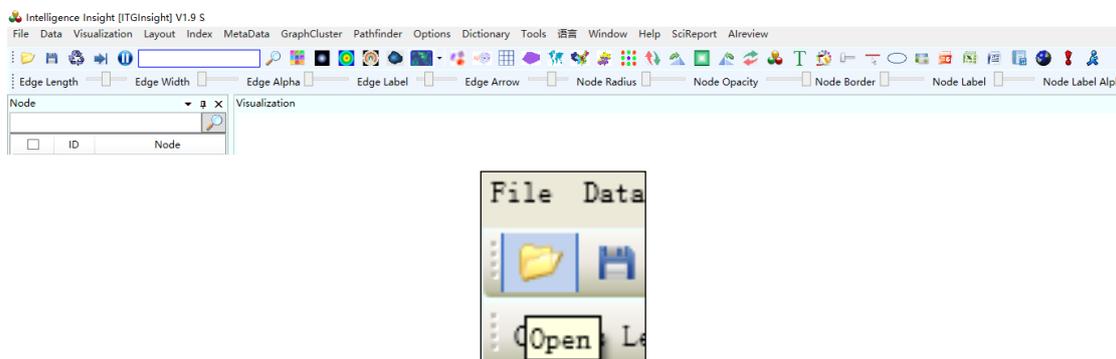
Save Document Adapter 的意思是：读取数据完成后保留数据分析结果，这个中间结果以.docadapter 为后缀，后续可以将这个文件作为输入，进行二次分析。

Apply PFNET 是分析过程中是否采用 PFNET 进行网络图压缩，默认即可。

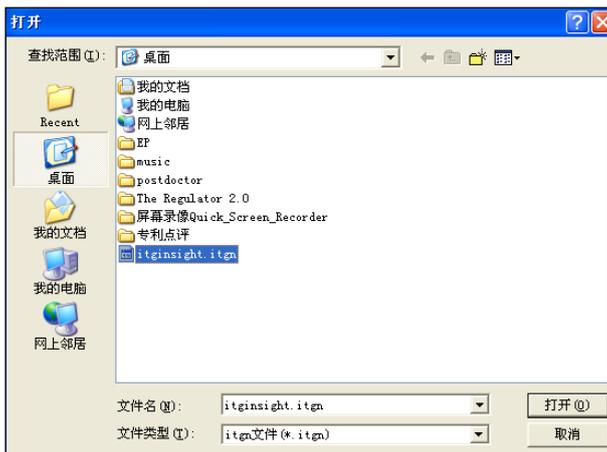
最后切换到 Trans/转换标签，点击“OK/确认”按钮，“MainProgress/主进度”、“AssiProgress/辅进度”和“Status/状态”处将显示后台数据转换的情况。

3.2 读取 itgn 文件进行可视化

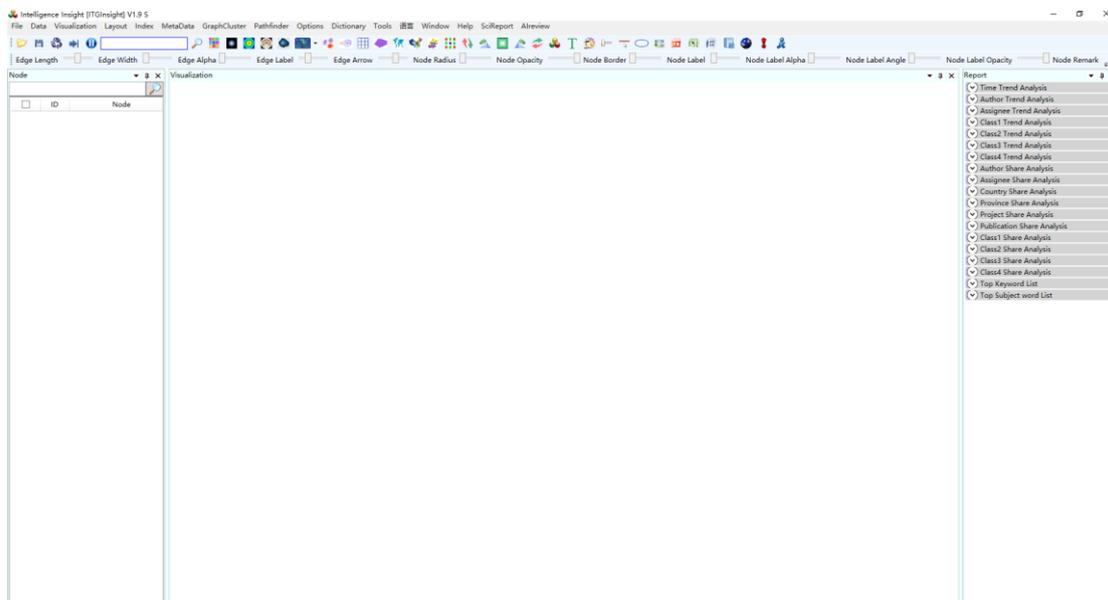
点击菜单栏上的文件菜单项，或点击工具栏上的打开按钮，如下图：



弹出文件导航对话框，导航到要分析的 itg 项目文件，进行文件读取。

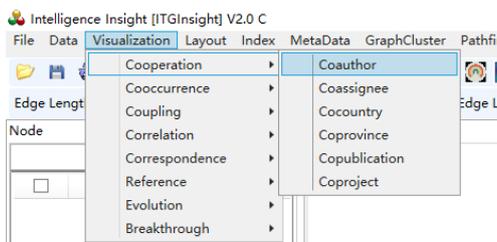


读取 itgn 项目文件后，系统主页面右侧的基本统计部分显示出一些基本的维度统计结果，可视化区域需进一步按 3.3-3.9 的操作方式指定分析内容，才能进行可视化结果的输出，如下图。



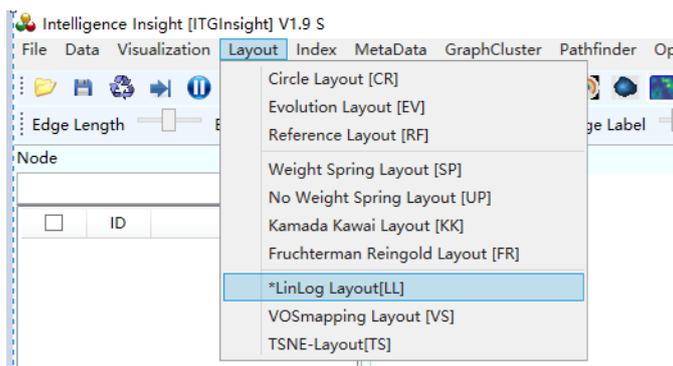
3.3 合著关系可视化

1) 点击菜单栏“Visualization/可视化”——>“Cooperation/合著网络”——>“作者合著/机构合著/国家合著/省份合著/出版物合著”，如下图。

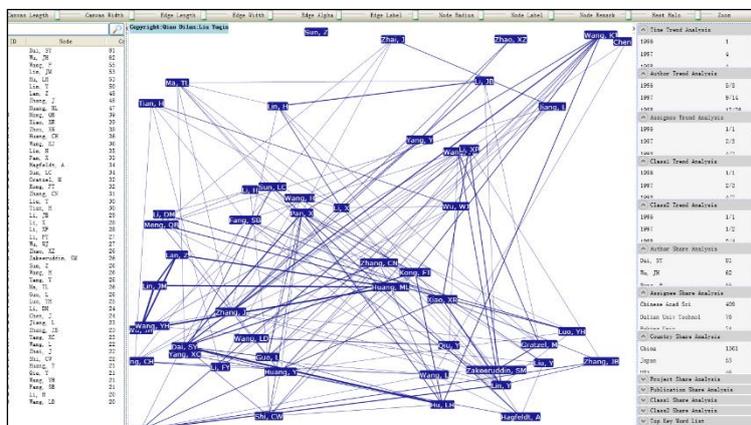


2) 点击菜单栏“Layout/布局”——>“CR 布局/EV 布局/RF 布局//UP 布局/SP 布局/KK 布局/FR 布局/LL 布局/VS 布局/TS ”，如下图。布局算法选择以图形是否美观、易读为标准，

默认选择 LL 布局算法，即可满足绝大多数情况下的可视化。

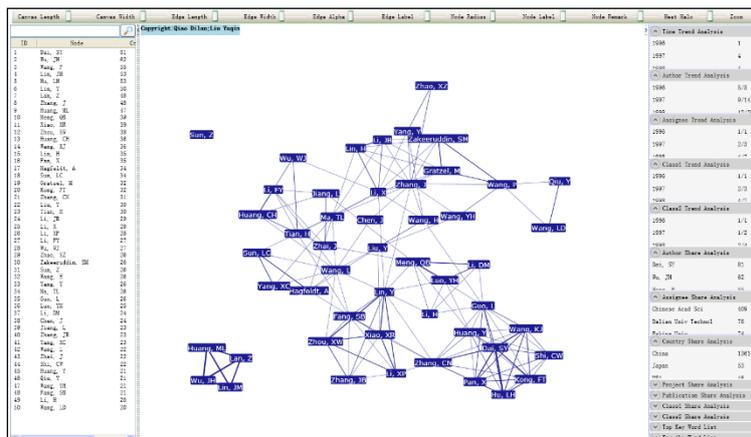


3) 点击工具栏 ，初始可视化分析图形，如下图。

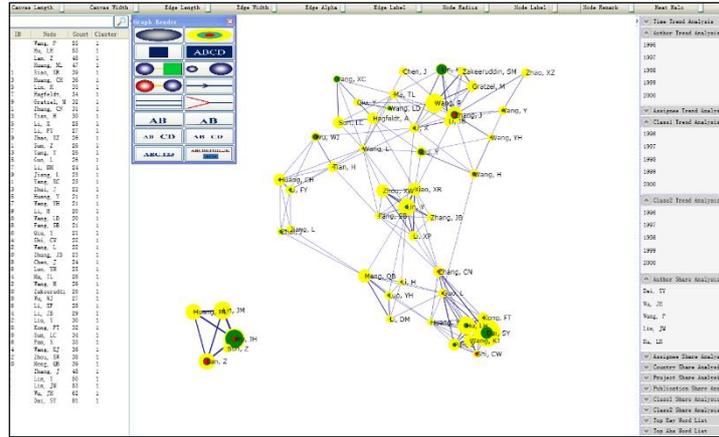


3) 点击工具栏 ，启动图形优化。

4) 在图形优化过程中，点击工具栏 ，停止图形优化，以得到更加简洁清晰的可视化分析结果，如下图。

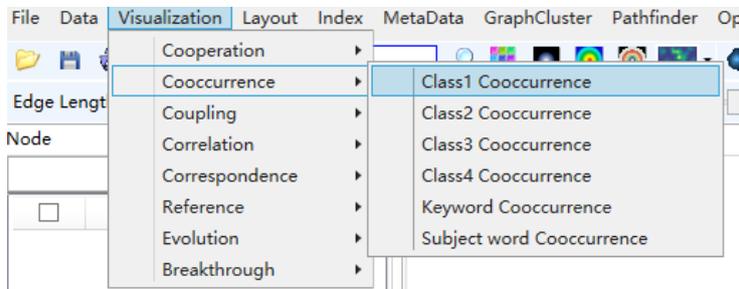


5) 应用本手册后面的“图形样式设置”、“滑块设置”进行图形的个性化设计，如下图，为合著关系的典型可视化结果。

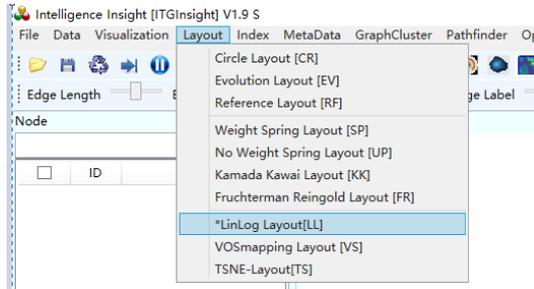


3.4 同现关系可视化

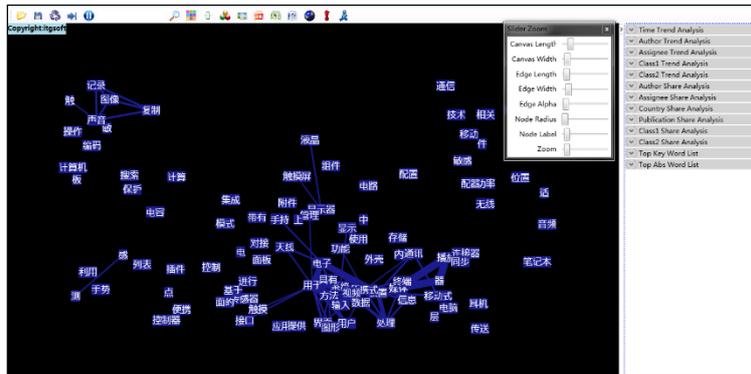
1) 点击菜单栏“Visualization/可视化”——>“Cooccurrence/同现网络”——>“类别1同现/类别2同现/关键词同现/摘要词同现”，如下图。

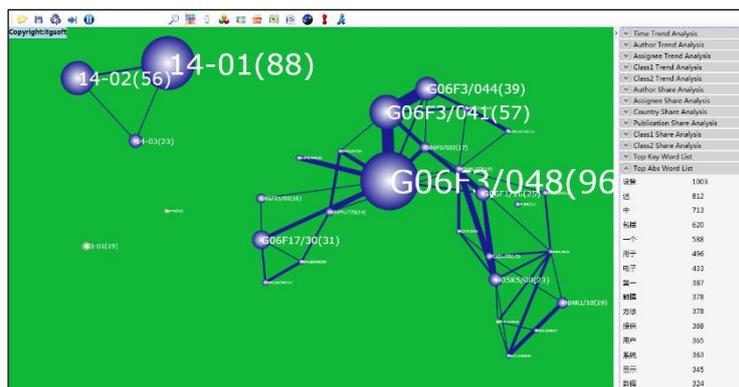


2) 点击菜单栏“Layout/布局”——>“CR 布局/EV 布局/RF 布局//UP 布局/SP 布局/KK 布局/FR 布局/LL 布局/Vs 布局/TS ”，如下图。



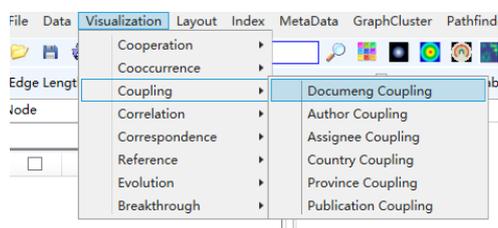
3) 其余步骤与合著分析相同。下图为同现分析的典型可视化结果。



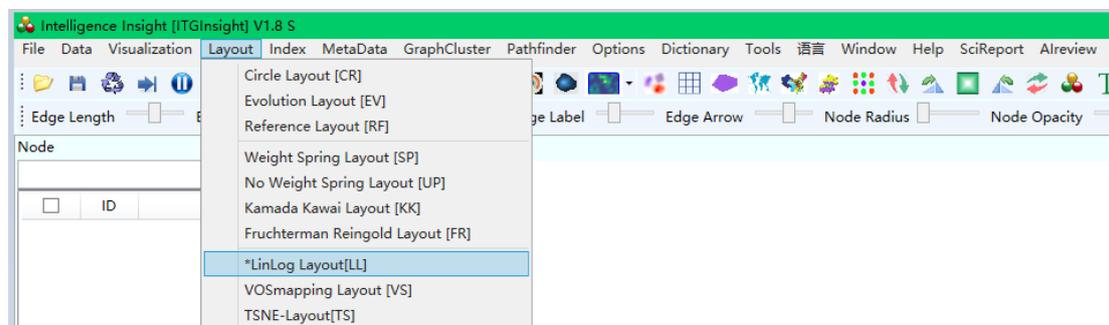


3.5 耦合关系可视化

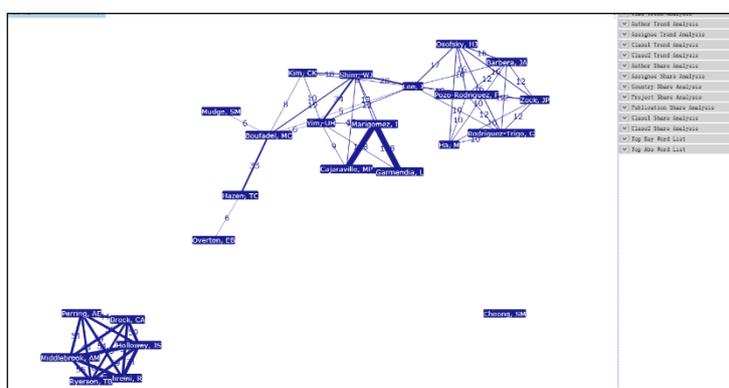
1) 点击菜单栏“Visualization/可视化”——>“Coupling/耦合网络”——>“文献耦合/作者耦合/机构耦合/国家耦合/省份耦合/出版物耦合”，如下图。



2) 点击菜单栏“Layout/布局”——>“CR 布局/EV 布局/RF 布局/UP 布局/SP 布局/KK 布局/FR 布局/LL 布局/VS 布局/TS ”，如下图。

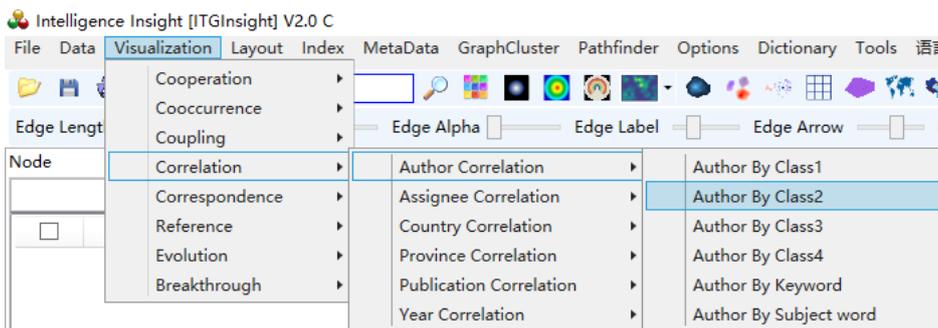


3) 其余步骤与合著分析相同。下图为耦合分析的典型可视化结果。

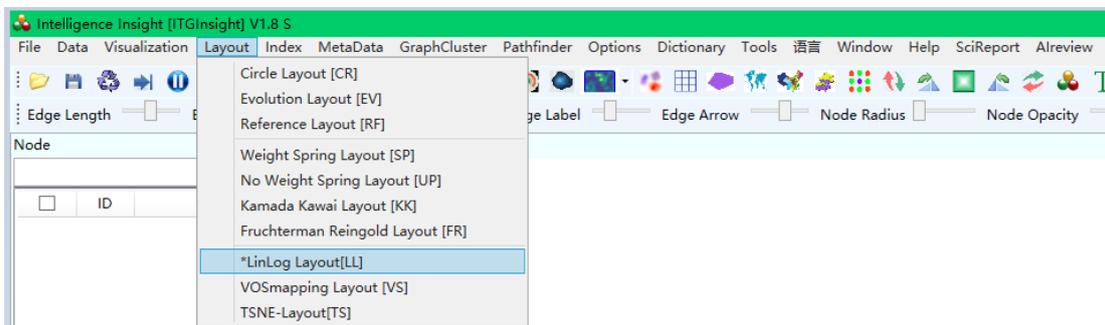


3.6 关联分析可视化

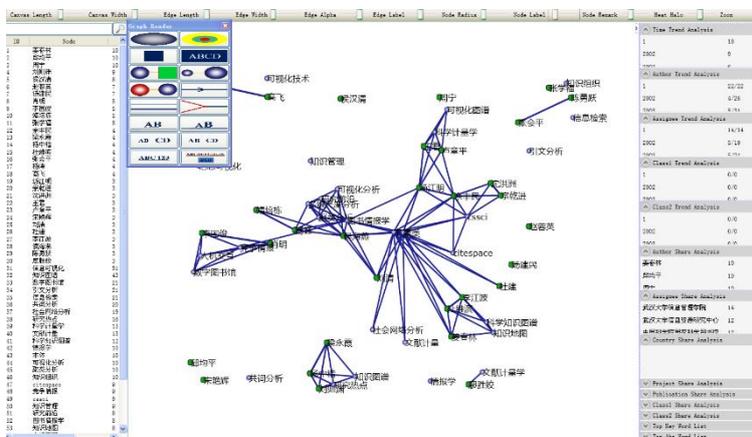
1) 点击菜单栏“Visualization/可视化”——>“Correlation/关联分析”——>“作者关联/机构关联/国家关联/省份关联/出版物关联/年代关联”——>“作者 BY 类别 1/作者 BY 类别 2/作者 BY 关键词/作者 BY 主题词”; “机构 BY 类别 1/机构 BY 类别 2/机构 BY 关键词/机构 BY 主题词”; “国家 BY 类别 1/国家 BY 类别 2/国家 BY 关键词/国家 BY 主题词”; “省份 BY 类别 1/省份 BY 类别 2/省份 BY 关键词/省份 BY 主题词”; “出版物 BY 类别 1/出版物 BY 类别 2/出版物 BY 关键词/出版物 BY 主题词”; “年代 BY 类别 1/年代 BY 类别 2/年代 BY 关键词/年代 BY 主题词”, 如下图。



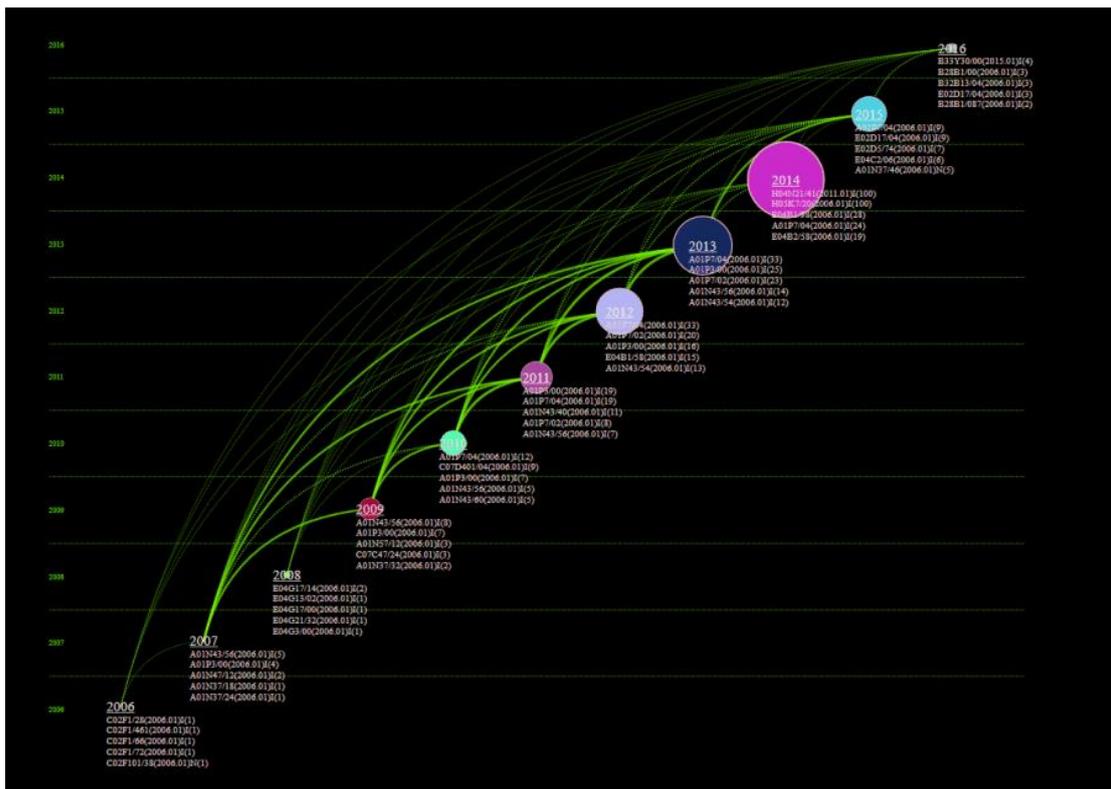
2) 点击菜单栏“Layout/布局”——>“CR 布局/EV 布局/RF 布局//UP 布局/SP 布局/KK 布局/FR 布局/LL 布局/VS 布局/TS ”, 如下图。



3) 其余步骤与合著分析相同。下图为关联分析分析的典型可视化结果。

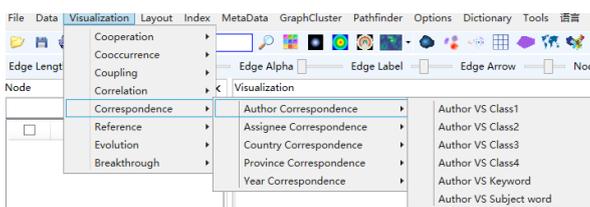


当进行时间关联分析时, RF 布局, 关联图形按如下形式进行显示。

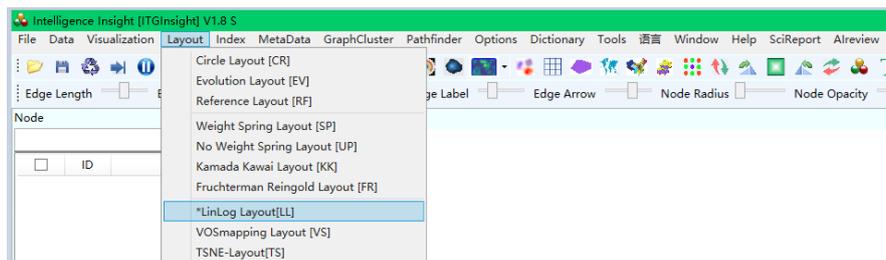


3.7 对应分析可视化

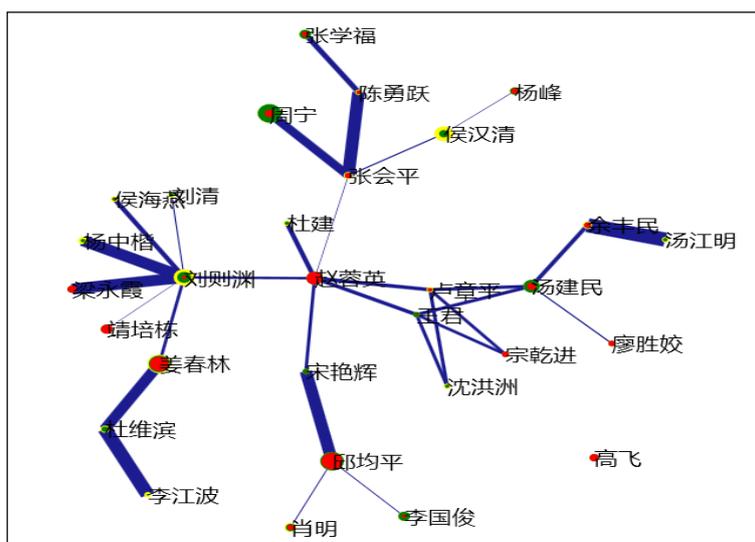
1) 点击菜单栏“Visualization/可视化”——>“Correspondence/对应分析”——>“作者对应/机构对应/国家对应/省份对应/年代对应”——>“作者 VS 类别 1/作者 VS 类别 2/作者 VS 关键词/作者 VS 主题词”; “机构 VS 类别 1/机构 VS 类别 2/机构 VS 关键词/机构 VS 主题词”; “国家 VS 类别 1/国家 VS 类别 2/国家 VS 关键词/国家 VS 主题词”; “省份 VS 类别 1/省份 VS 类别 2/省份 VS 关键词/省份 VS 主题词”; “年代 VS 类别 1/年代 VS 类别 2/年代 VS 关键词/年代 VS 主题词”, 如下图。



2) 点击菜单栏“Layout/布局”——>“CR 布局/EV 布局/RF 布局//UP 布局/SP 布局/KK 布局/FR 布局/LL 布局/VS 布局/TS”, 如下图。

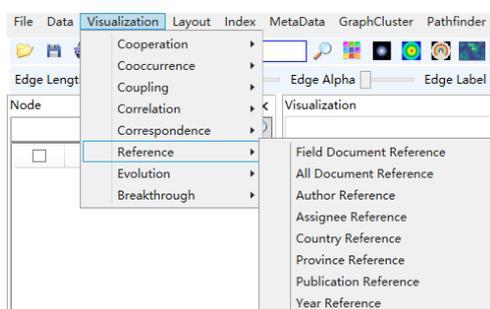


3) 其余步骤与合著分析相同。下图为对应分析分析的典型可视化结果。

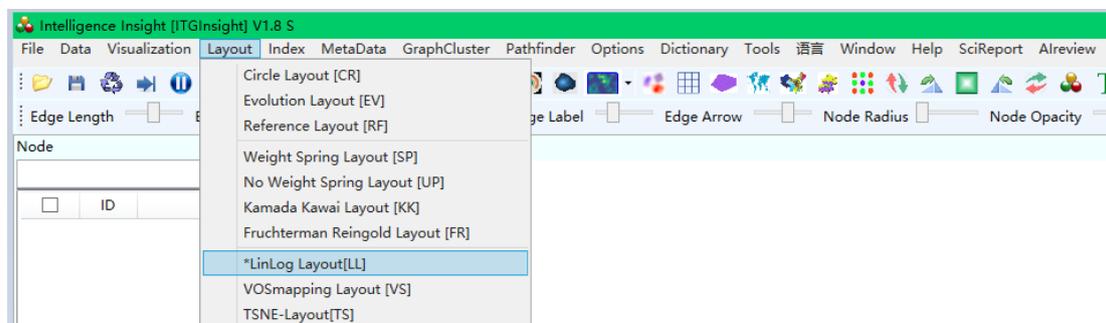


3.8 引证关系可视化

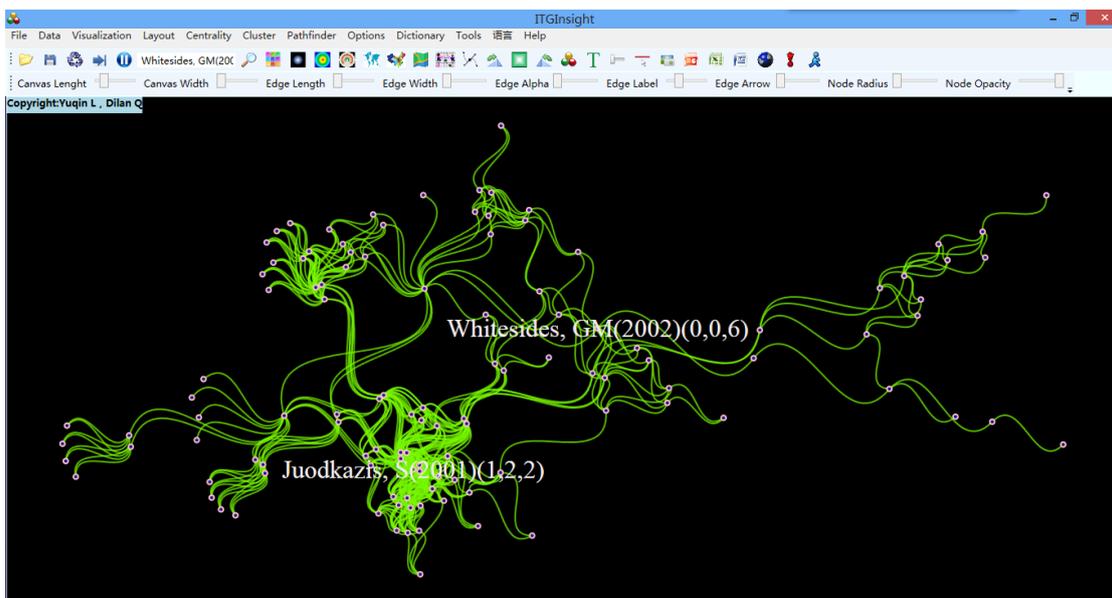
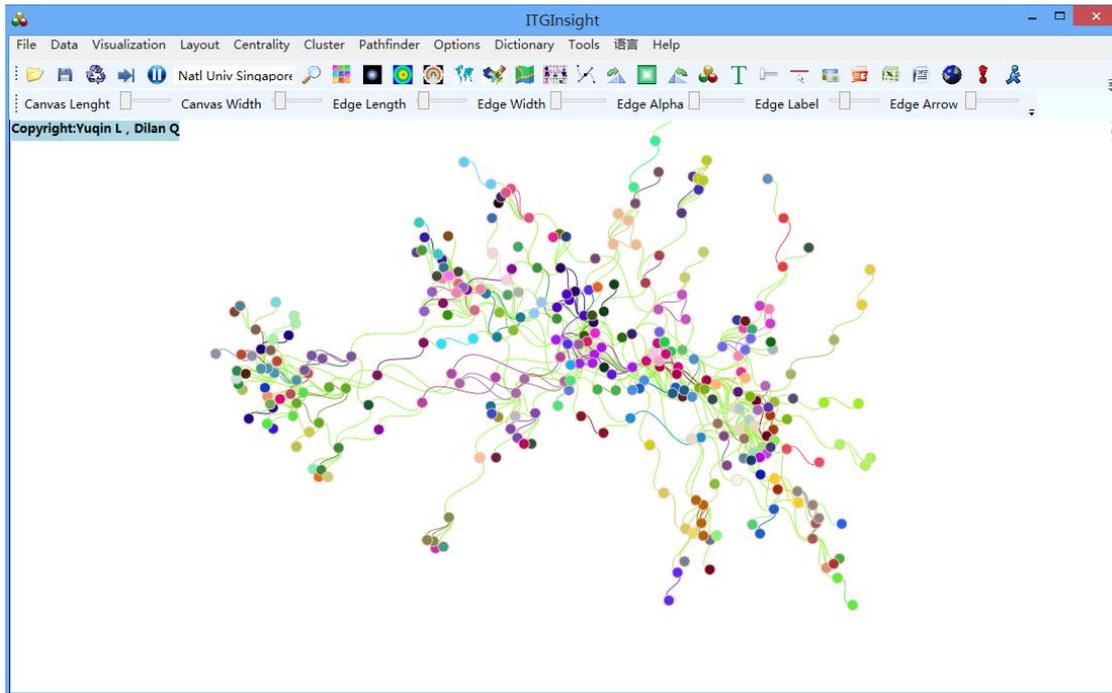
1) 点击菜单栏“Visualization/可视化”——>“Reference/引证分析”——>“领域文献引证/所有文献引证/作者引证/机构引证/国家引证/省份引证/出版物引证/年代引证”，如下图所示。

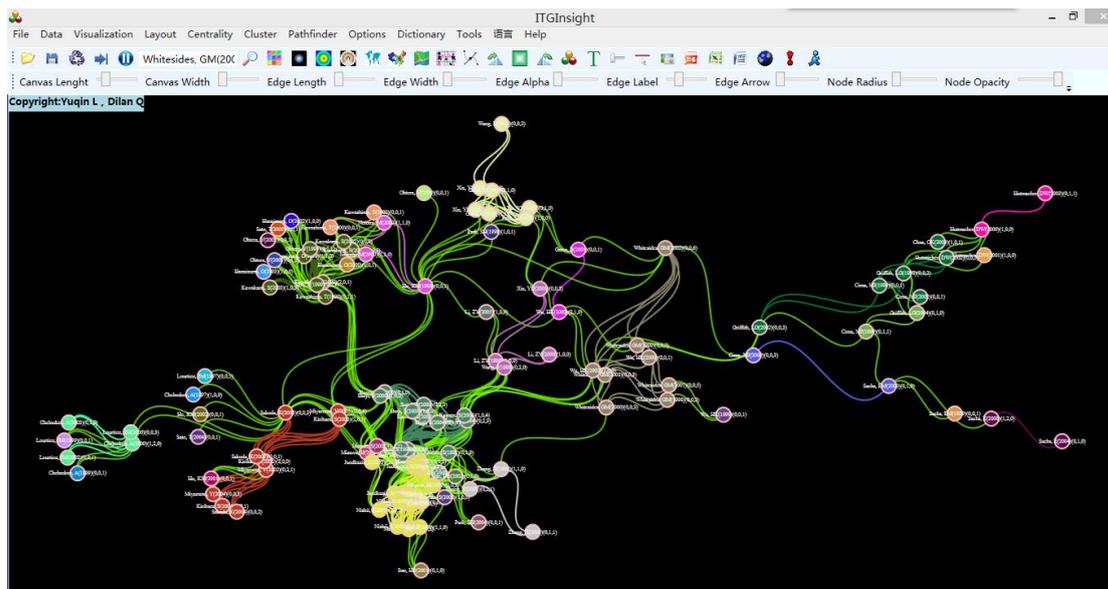


2) 点击菜单栏“Layout/布局”——>“CR 布局/EV 布局/RF 布局//UP 布局/SP 布局/KK 布局/FR 布局/LL 布局/VS 布局/TS ”，如下图所示。

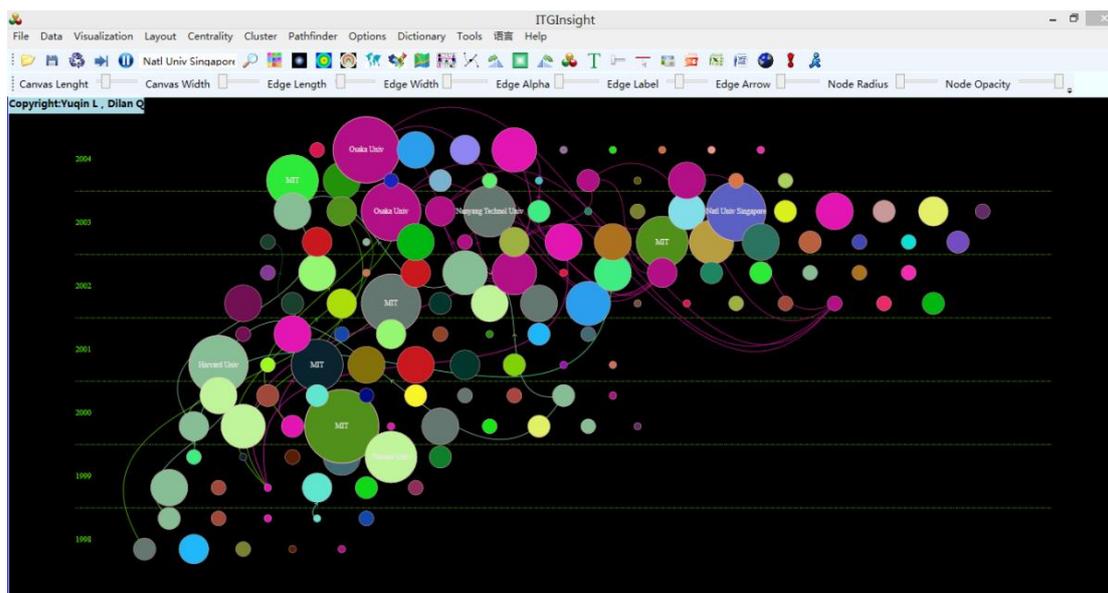


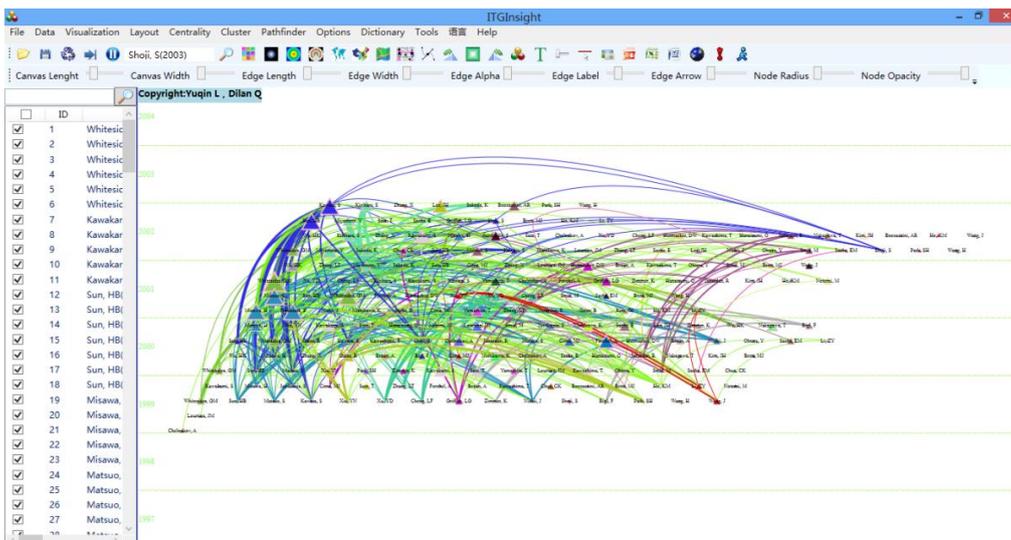
3) 其余步骤与合著分析相同，下图为引证分析的典型可视化结果。





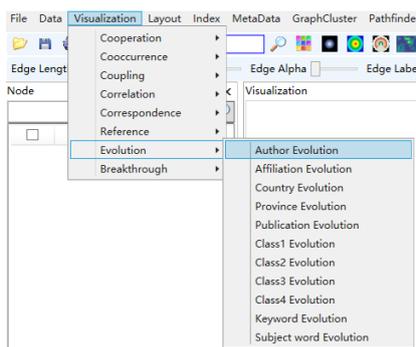
引证关系可视化，除了网络图以外，可以采用时间线进行可视化，点击工具栏的 RF 布局，后可可视化结果如下：



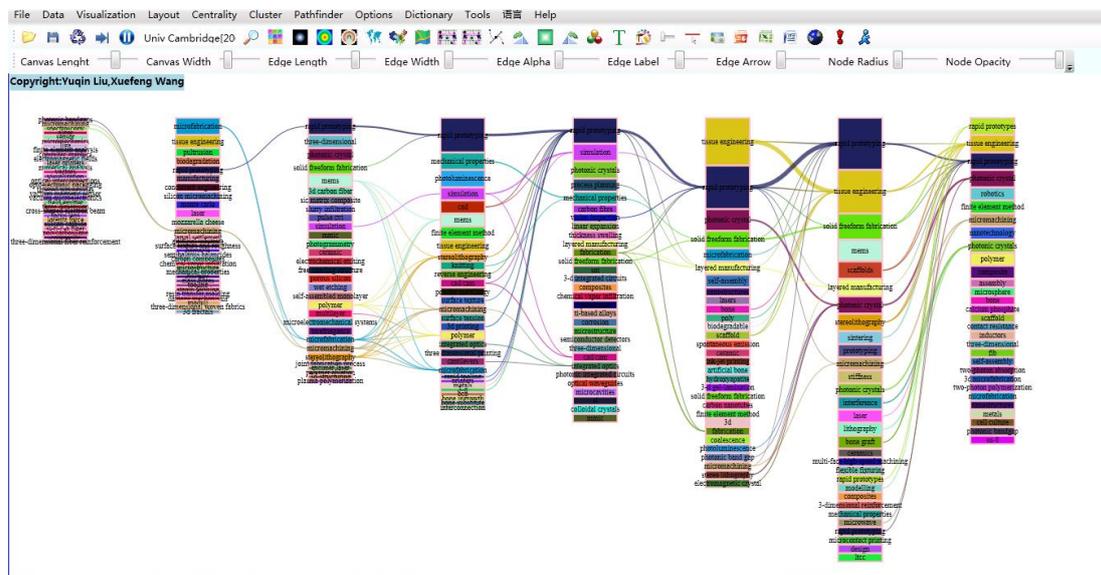


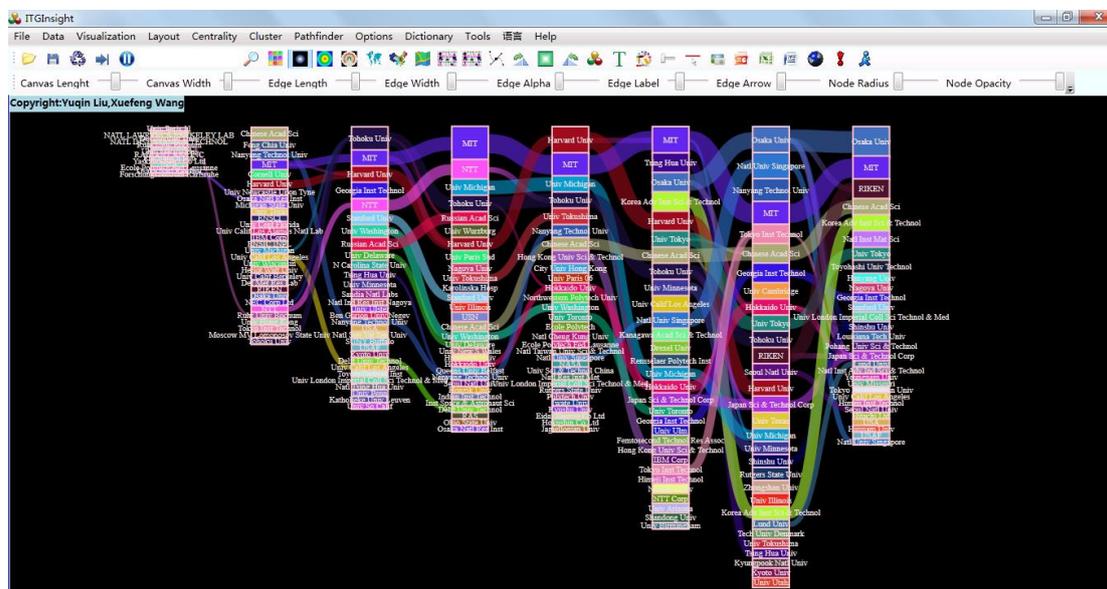
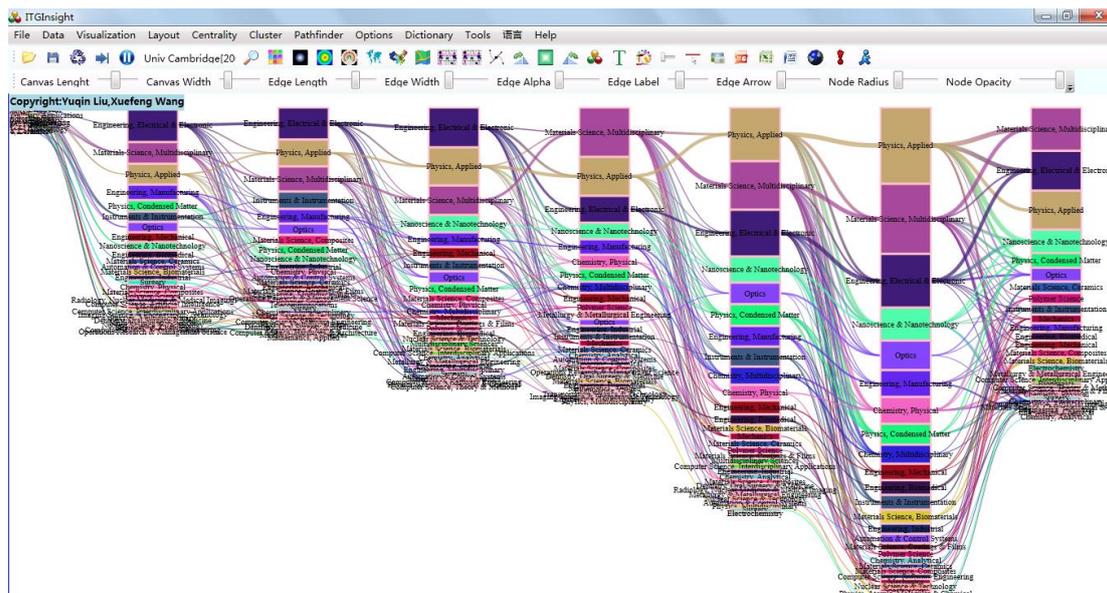
3.9 演化分析可视化

点击菜单栏“Visualization/可视化”——>“Evolution/演化分析”——>“作者演化/机构演化/国家演化/省份演化/出版物演化/类别 1 演化/类别 2 演化/类别 3 演化/类别 4 演化/关键词演化/主题词演化”，如下图。



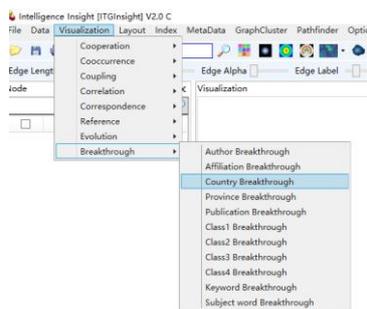
可视化区域显示演化图如下：



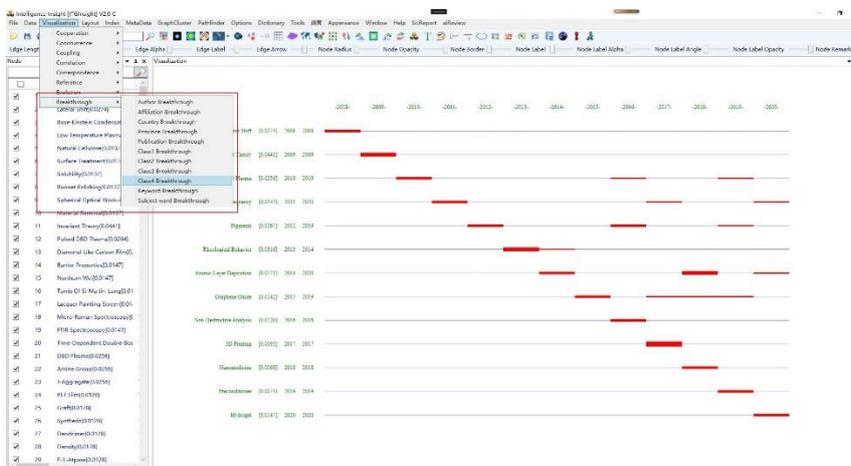


3.10 突破分析可视化

1) 点击菜单栏“Visualization/可视化”——>“Breakthrough/突破分析”——>“作者突破/机构突破/国家突破/省份突破/出版物突破/类别 1 突破/类别 2 突破/类别 3 突破/类别 4 突破/关键词突破/主题词突破”，如下图。



可视化区域显示演化图如下，其中“[]”数值为突破率，红色线条为当年出现与否，宽带可设置为与频数成正比，也可设为相同宽带，默认为相同宽带。



3.11 网络布局算法/选择合适的网络布局算法使网络图尽可能美观

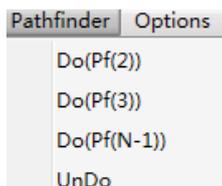
在 3.3-3.9 的分析步骤中，均涉及到布局算法的选择，默认情况下，布局算法针对整个可视化区域的网络图进行计算；如果希望算法仅对网络图某一局部内的网络节点有效，右键按住 ctrl，同时鼠标左键按下、移动，选择局部网络节点，这时对网络图的操作进行对被选中的局部有效。这种操作方式可以使同一网络图不同部分采用不同的布局算法，以便整体网络图更加清晰、可阅读。取消局部选择，放开 ctrl 键，任意点击鼠标左键即可。

在各种布局算法中，LL(LinLog)VS(VosMapping)布局算法的与其它算法的不同之处在于，通过这两种算法布局节点，节点距离与节点之间的关系数量或强度成反比，也就是说距离具有实际意义。

建议用户使用 LL(LinLog Layout)，该算法满足绝大数的网络布局。

3.12 关键信息过滤/删除不重要的连接线

在关联关系分析过程中，可以通过路径压缩技术多网络图的关键信息进行过滤，也就是删除不重要的连接线，保留相对重要的连接线。具体操作，点击工具栏的 Pathfinder/压缩，如下图：

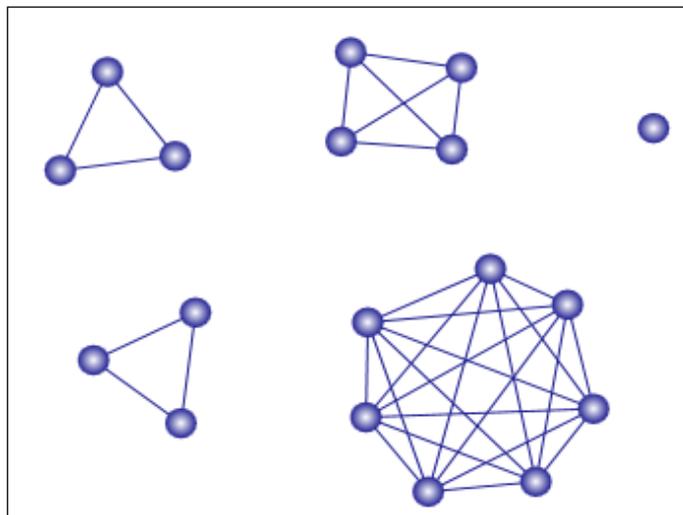


Pf (2)、Pf (3)、Pf (N-1) 三种压缩操作，压缩强度逐渐增加。如果取消压缩按 Undo/撤销按钮。如果连续两次压缩，撤销仅能恢复最后一次的压缩操作。

3.13 更改图形样式/美化图形

1) 默认图形效果

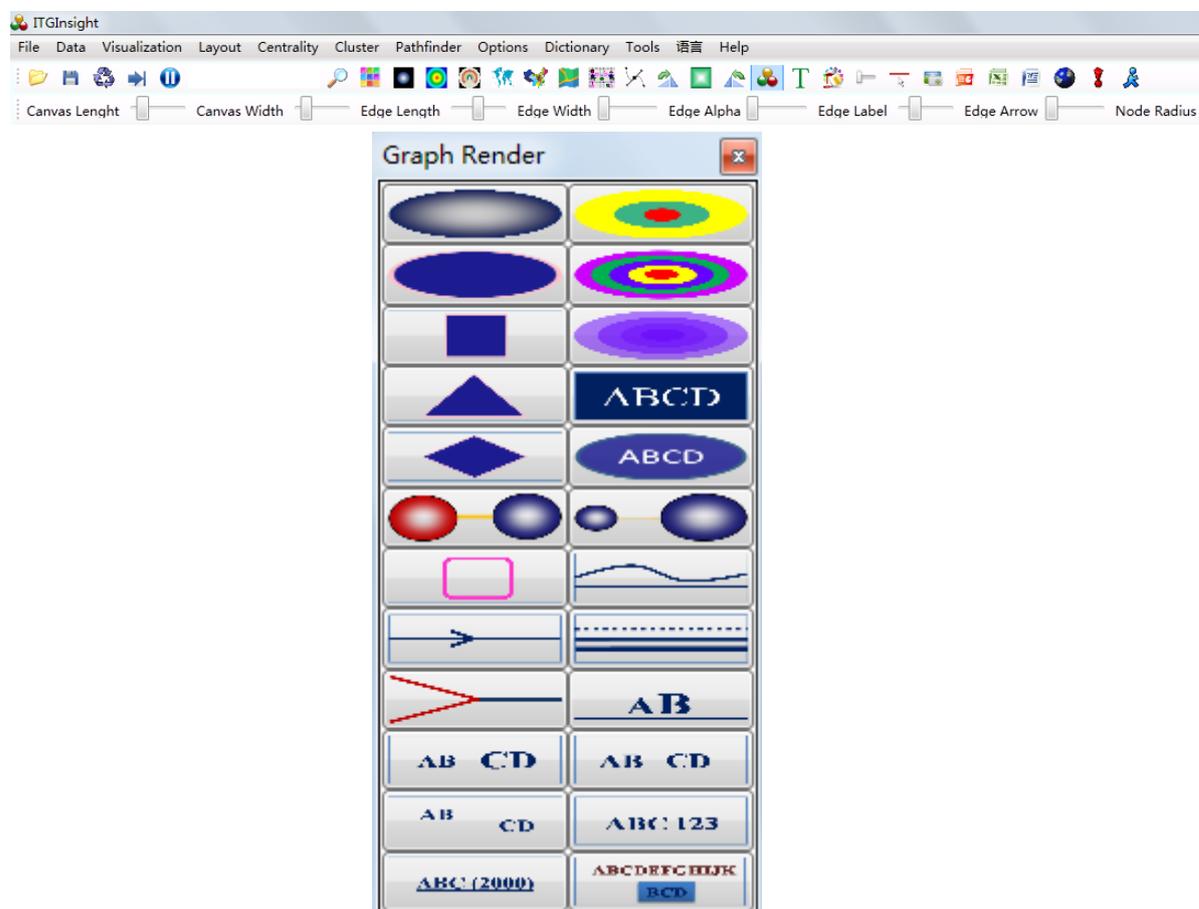
图形区默认图形效果如下：



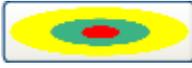
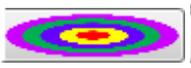
2) 图形设置面板

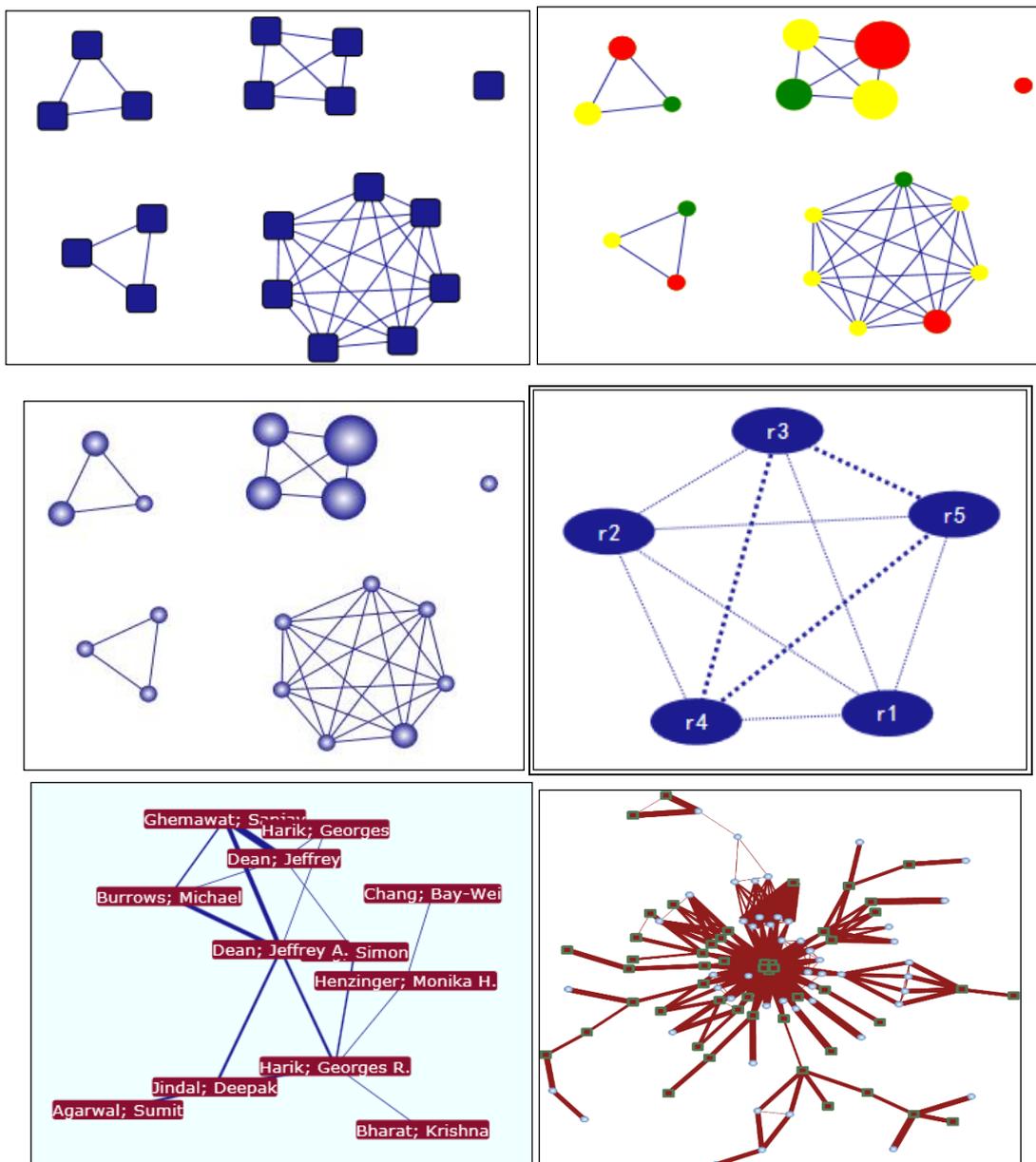
点击工具栏上，或菜单栏“Options/选项”——>“Graph Render/图形渲染”，弹出

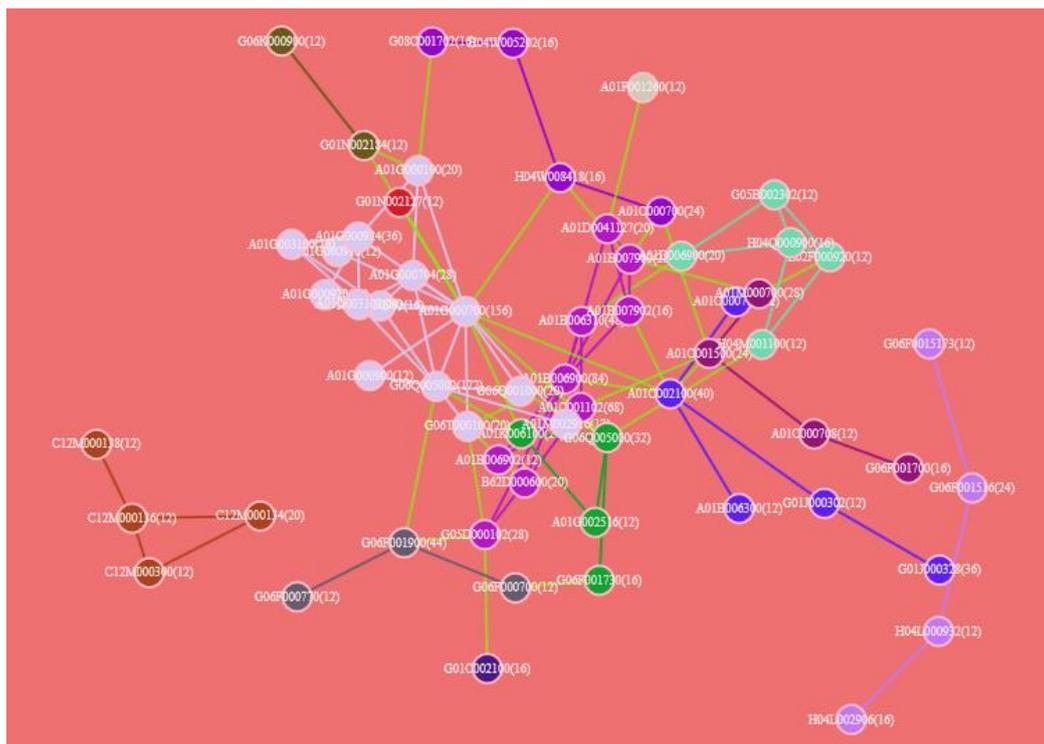
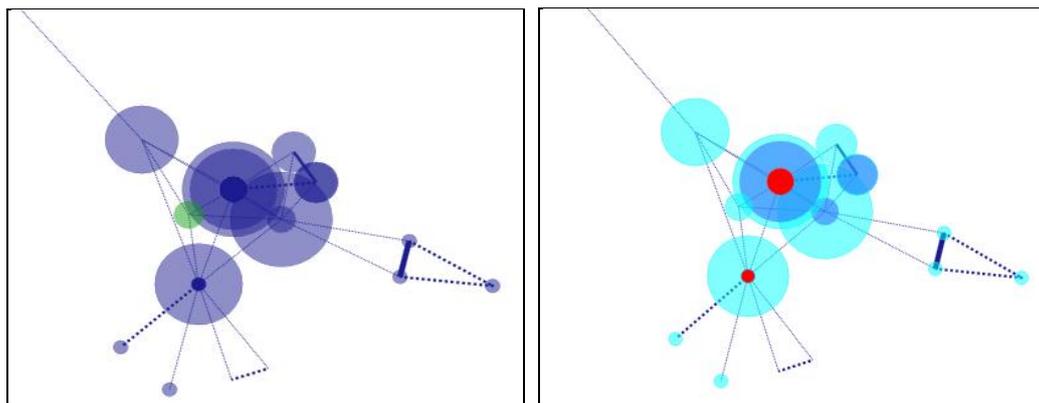
图形渲染设置工具栏或面板如下图：

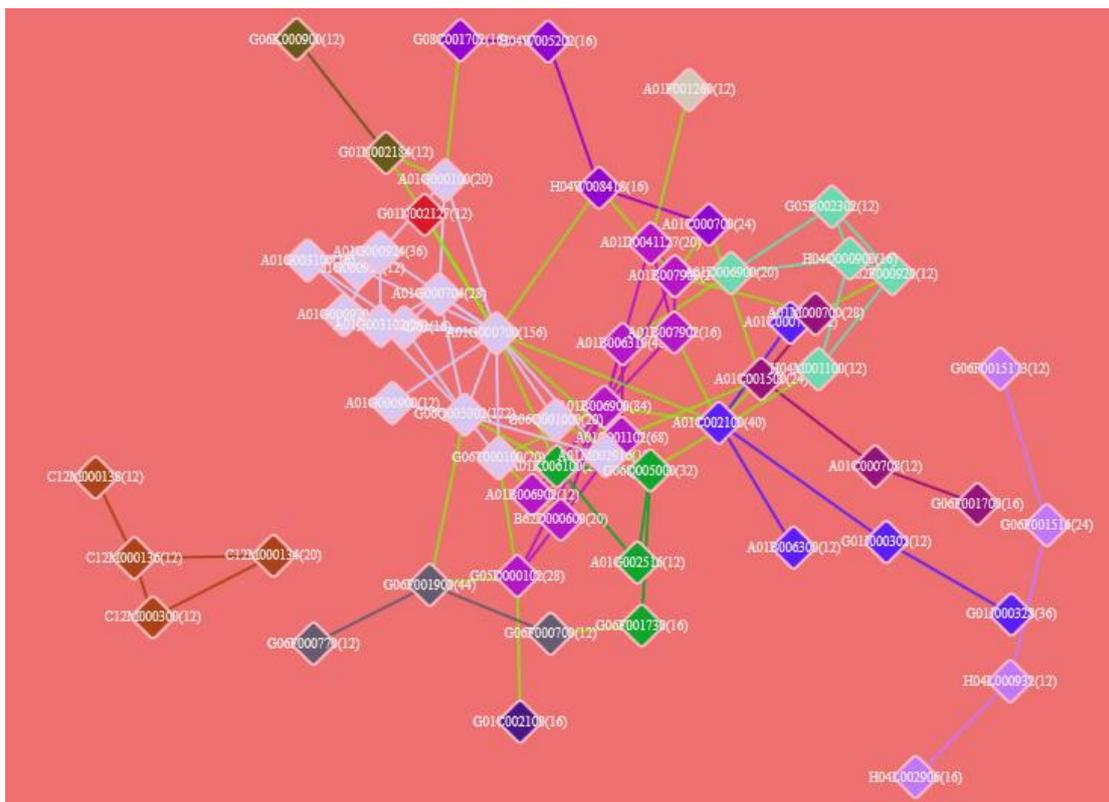
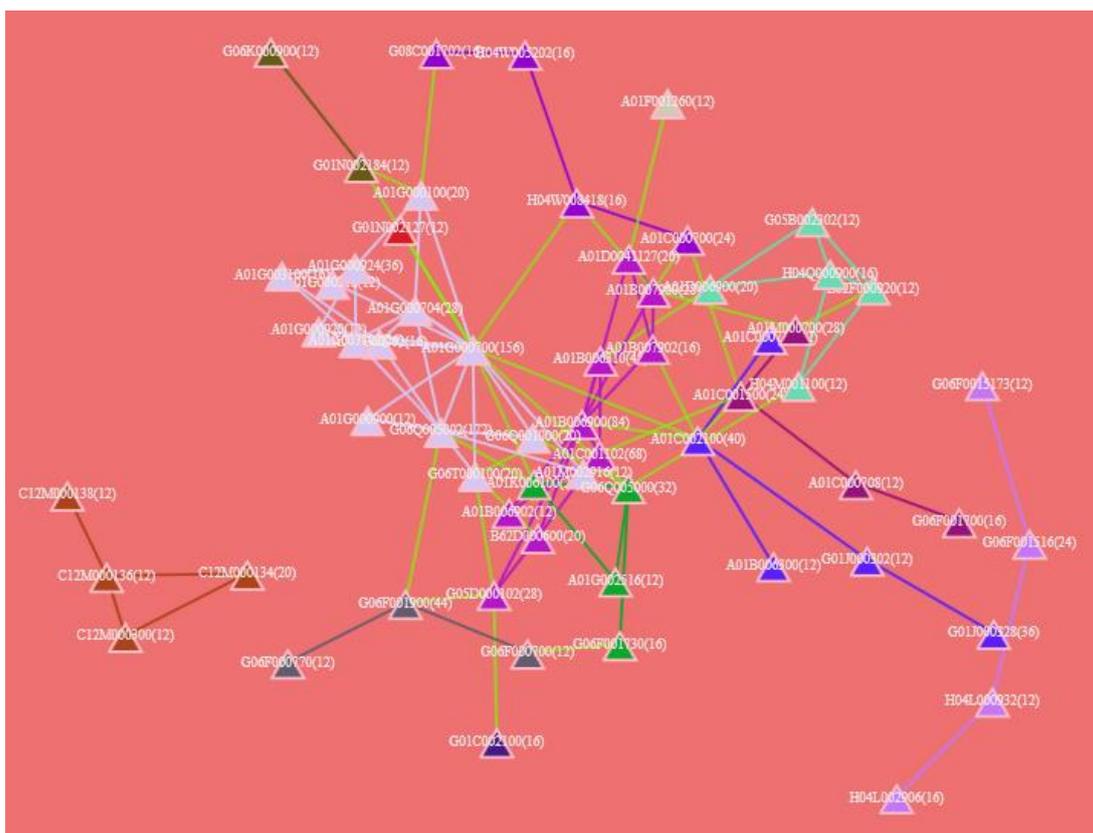


3) 更改节点样式

点击图形样式面板上的           , 切换节点的显示样式, 如下图所示。





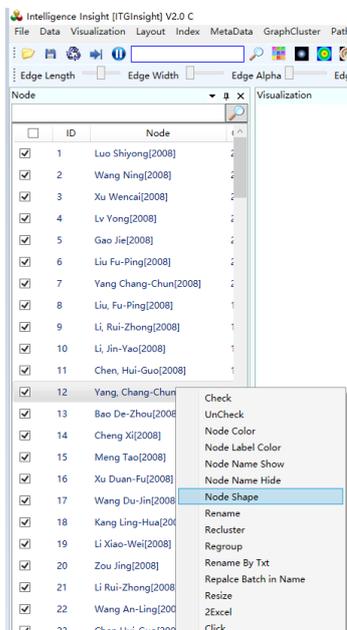


单击图形面板中的,表示所有节点大小一致,在各种分析中均可使用。
再次单击表示所有节点大小不一致,与节点所代表的数量成比例。

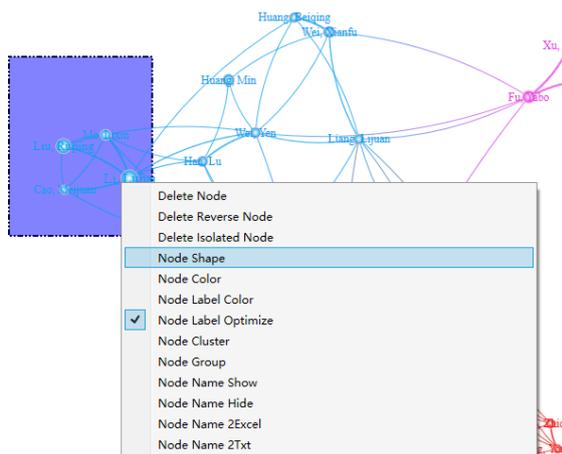
单击图形面板中的，表示用两种颜色区分被选中节点和未被选中的节点，选中的方式为鼠标单击。

4) 更改部分节点样式

软件左侧选中部分节点后，右键单击 **shape**，如下截图：



即可对部分节点的形状进行修改；也可以在图形区，用鼠标+shift 键，选中多个节点进行节点形状的修改，如下图。



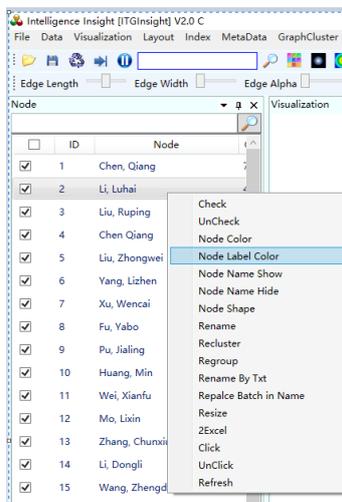
5) 更改节点颜色

在上一操作中双击样式面板中的样式选项，弹出颜色对话框，选择不同的颜色，图形区域中的节点颜色将随之改变。

也可以在主页面左侧的节点内容面板中进行操作，先左键选择一个或多个节点，然后

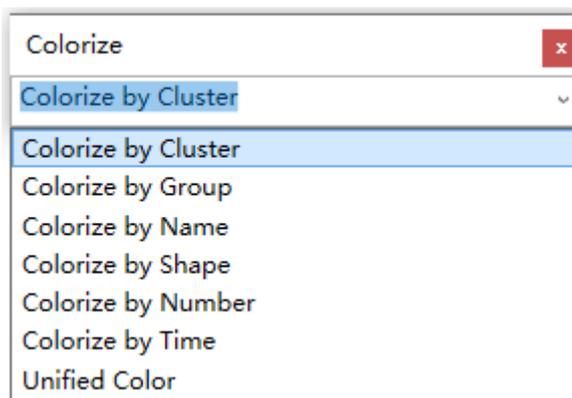
右键点击 color，如下图，进行节点颜色的个性化设置。

也可以在图形中去用鼠标+shift 键，一次选择多个节点进行颜色修改。



6) 按照关系强度、节点形状、节点名称、节点大小对节点染色

点击工具栏的 ，弹出节点着色选项，如下：



分别表示按节点之间的关系强度、节点分组、节点名称、节点形状、节点大小对节点染色，单机选中即可。

7) 更改节点边框颜色

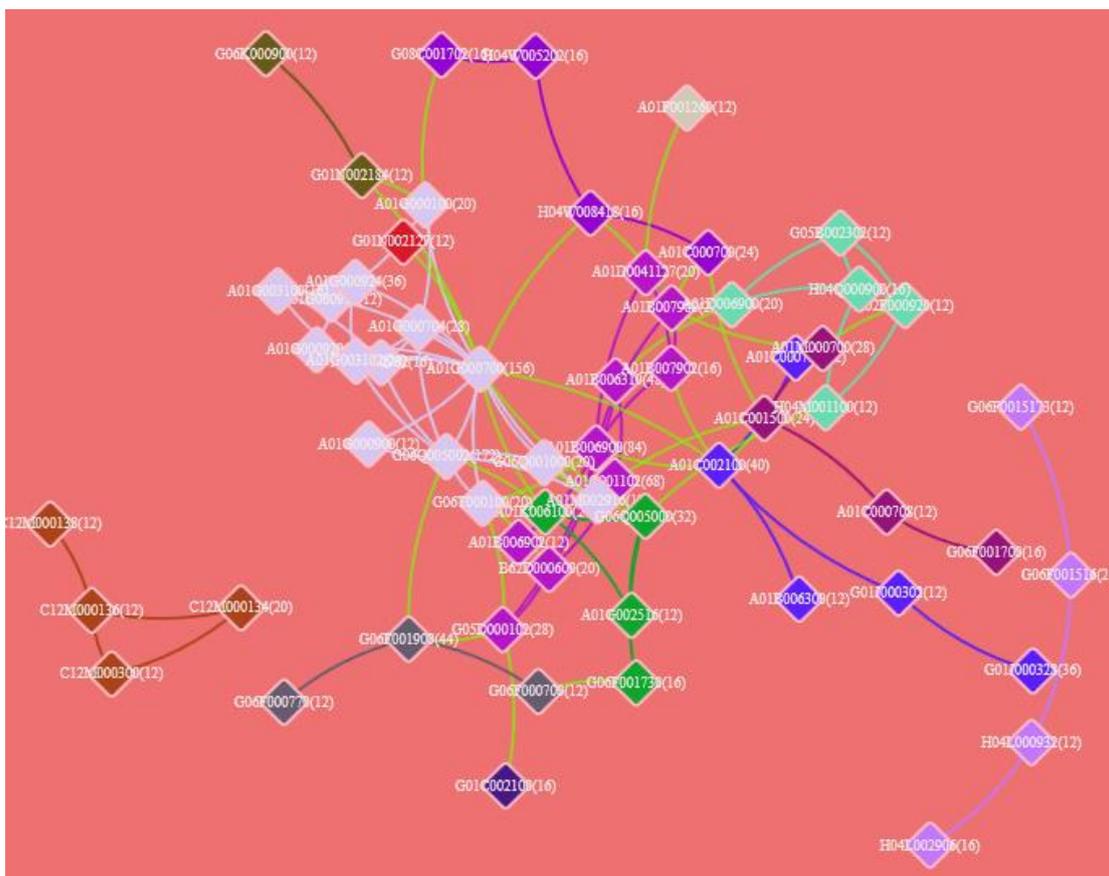
点击样式面板 ，切换节点边框显示与否，双击则更改节点边框颜色。

8) 更改连线为直线或曲线

单击图形样式面板中的 ，表示所有节点连线宽度一致，在各种分析中均可使用。再次单击表示所有节点连线宽度不一致，带有数量比较，在合著关系分析、同现关系分析时使用，表示数量，在关联关系分析时使用表示关系的强度。  表示连线的两个节点有初始结束关系，在引证关系分析时使用。单击图形样式面板中的

AB 表示在线上注明连接的数量，再次单击隐藏数量，在各种关系分析中均可使用。 表示连线为直线或是曲线，连续点击会在直线和曲线之间进行切换，切换单曲线时，会有多种曲线风格，如下。





单击图形面板中的, 表示被选中节点所拥有的连线颜色是否与其它连

线进行区别，默认为区别，第一次单击不区别显示，第二次单击区别显示，第三次单击区别显示，并且与备选节点间接相连的连线也进行区别显示。

9) 更改连线颜色

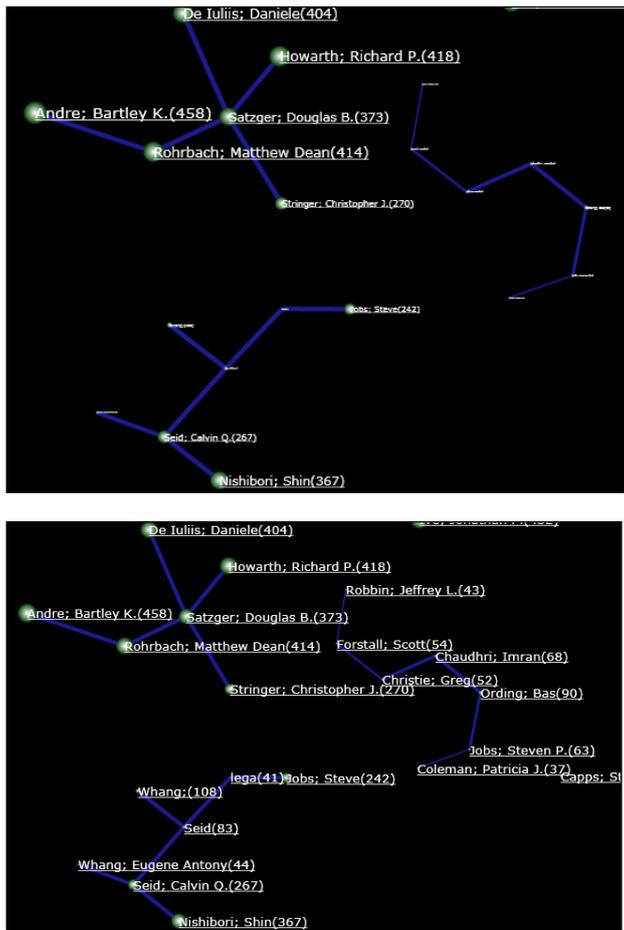
在上一操作中双击样式面板中的样式选项，弹出颜色对话框，选择不同的颜色，图形区域中的连线颜色将随之改变。

10) 更改节点注释显示模式

系统提供三种显示节点注释的模式：1、鼠标点击节点，显示被点击节点的注释；2、显示所有节点的注释；3、所有节点注释均不显示。

通过点击  或  进行三种模式的切换，默认条件下为第 1 种模式。

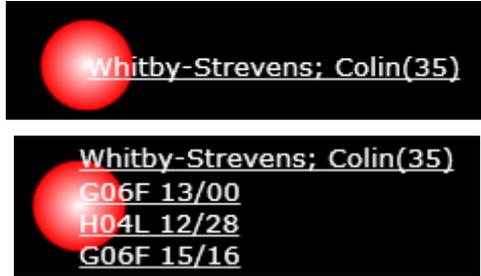
 与  的区别在于节点注释的字体大小是否按照节点所表示的数量进行比例化处理，其效果如下。



当点击  时，节点文字如果具有时间信息，则节点文字在显示节点文字信息与否之间进行切换。

11) 更改节点注释显示内容

除了一般情况下显示节点名称的注释方式外，系统还提供了 2 种备选方式：显示节点数量，显示节点的备注。如下图。



其中数量表示该节点所代表的数量，文字分别代表节点的名称和备注信息。通过单击图形样式面板中的 进行切换。

12) 更改注释颜色

在上一操作中双击样式面板中的样式选项，弹出颜色对话框，选择不同的颜色，图形区域中的连线颜色将随之改变。

13) 更改节点字体

点击工具栏上的 **T**，弹出节点字体设置窗体，如下图，可对节点文字的字体进行设置。



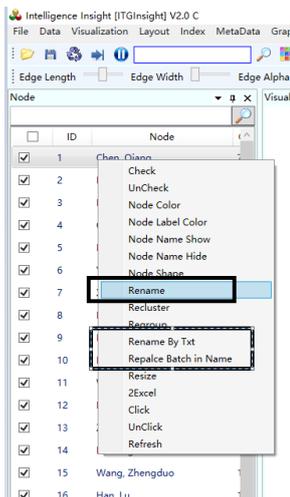
14) 更改节点文字大小写

在节点列表区域选中节点，鼠标右键切换节点文字大小写，如下图。

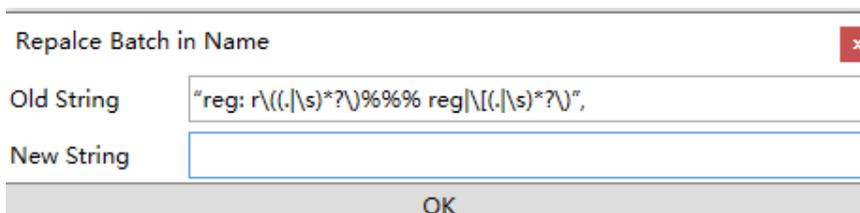


15) 更改节点名称

主页面左侧的节点内容面板中进行操作，先左键选择一个或多个节点，然后右键点击 rename（更改节点名称）或者 Rename By Txt（通过 ID 对应 Txt 文本中每一行，进行节点名称的批量修改）或者 Replace Batch In Name（批量替换节点名称里的某些文字，如空格等）进行节点名称修改，如下图，进行节点名称的个性化设置。



Replace Batch In Name 批量替换支持一次替换多个词，用%%%进行分割，同时支持正则表达式替换，例如希望批量删除节点名称里“()”和“[]”中的内容，可以输入“reg: r\((.\s)*?\)\%%% reg\[.\s]*?\)”，替换后的字符串为空

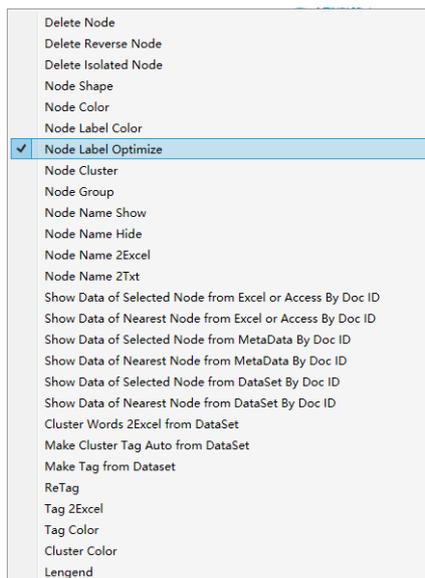


16) 更改节点文字显示位置

点击样式面板 ，使节点文字的位置显示在节点中央或者节点右侧，多次点击，则在以上两种情况下进行切换。

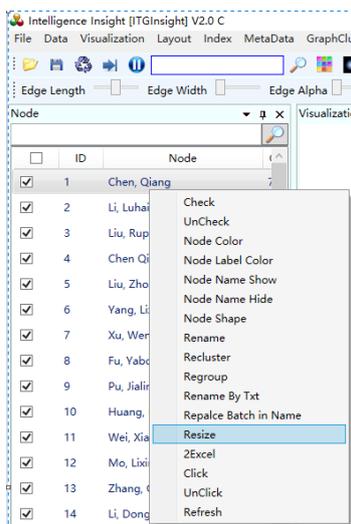
17) 节点文字自动防重叠

在可视化区域右键点击 Node Label Optimize,如下图，节点标签根据算法进行选择，隐藏部分节点文字。



18) 更改节点大小

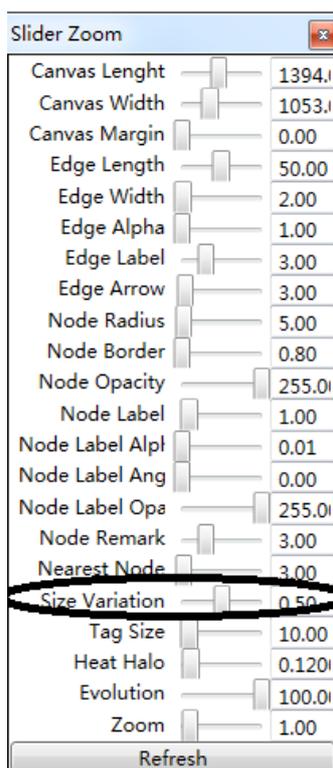
在节点内容面板右键点击 **Resize/更改节点大小**，如下图：



指定节点大小文件，一般为 txt 文件，参考 example\txt 目录下的 nodesize.txt，第一列为节点 id，第二列为更改后的节点大小，如果可视化区域的节点 id 与 txt 文件中的第一列 id 相等，节点大小更改。

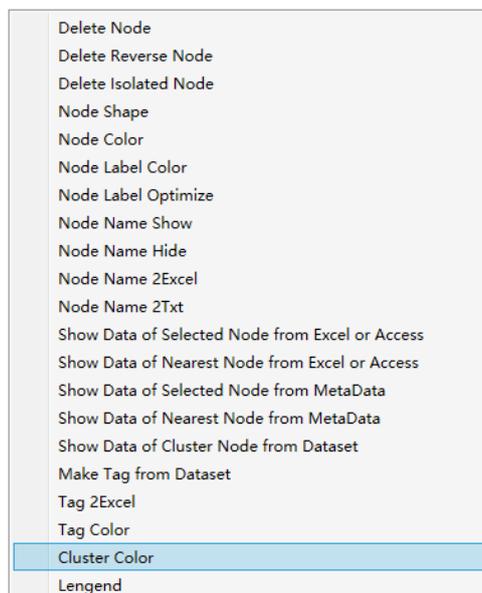
19) 更改节点大小对比度

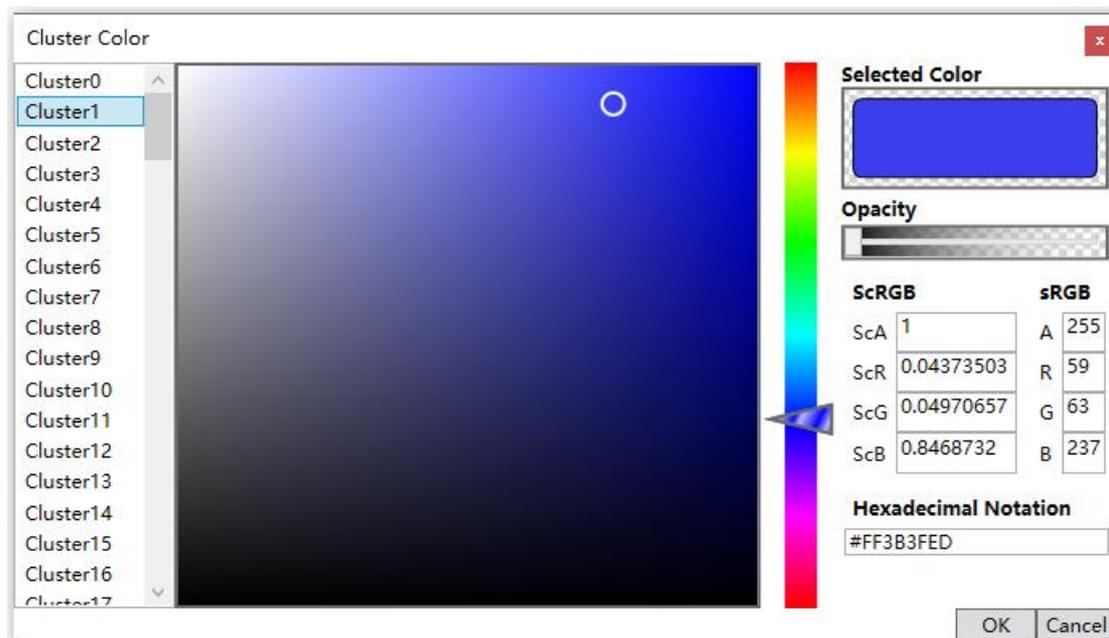
当节点按照大小进行显示时，通过滑块设置的 **Size Variation** 进行节点大小对比度调整，如下图。



20) 更改聚类类别颜色

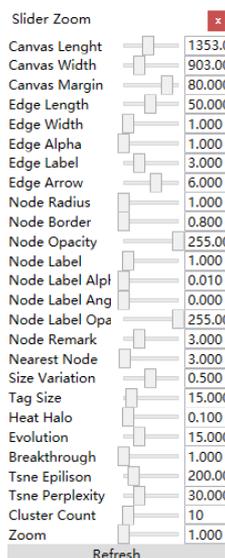
在可视化区域点击鼠标右键，点击“聚类颜色/Cluster Color，弹出类别颜色调整面板，如下图。选择对应的类别颜色进行修改，修改后的颜色存储在软件目录下 colors/clustercolors.txt 的文件中，之后在使用聚类，类别颜色将按照新的设置进行显示。





3.14 更改滑块设置

1) 点击工具栏上的 ，或菜单栏的“Options/选项”——>“Slider Zoom/滑块缩放”，弹出滑块设置工具栏目或面板，如下图。



2) 通过设置滑块值进行“画布长”、“画布宽”、“边长度”、“边宽度”、“边阈值”、“边标注”、“边箭头”、“节点半径”、“节点边框”、“节点透明度”、“节点标注”、“节点标注阈值”、“节点标注角度”、“节点标注透明度”、“节点备注”、“最近 N 个节点”、“标签大小”、“热力光圈”、“演化分析”、“图形缩放”的大小、长短、多少、比例设置、突破数量。

3.15 图形缩放、平移、拉伸、旋转

通过滑块设置可以对图形进行缩放；通过鼠标拖动可以对优化过程中的图形节点进行

位置的移动；通过鼠标拖动并按下 **shift** 键，可以实现图形的移动；通过鼠标滑轮+“左或右箭头”实现图形的横向拉伸；通过鼠标滚轮+“上或下箭头”实现图形上下拉伸。

通过工具栏  实现图形顺时针或逆时针旋转。

通过工具栏  实现图形的上下或左右翻转。

3.16 更改系统语言

系统默认提供中英文操作，点击菜单栏上的 **Lans/语言**，进行语言选择，如下图：



如果希望用日文、韩文或其它任意语言进行操作，与开发者进行联系，我们将提供任何语言的软件版本。通过上面“其它”进行非中英文语言的设置。

3.17 更改背景色和背景边框

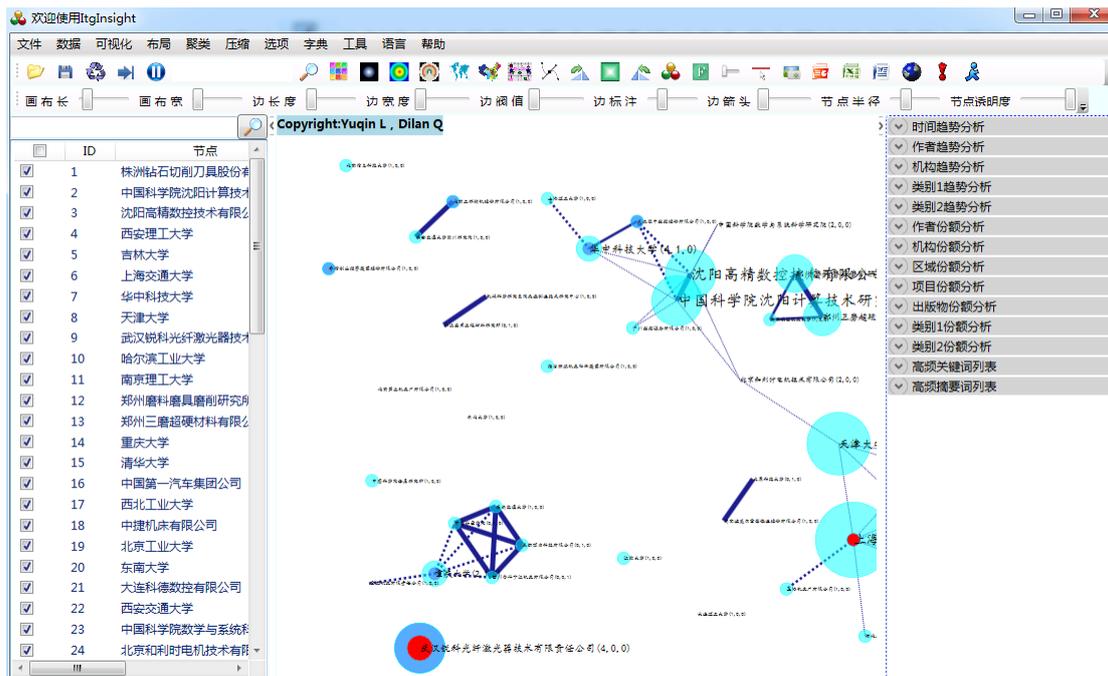
点击工具栏上的 ，弹出颜色对话框，选择颜色进行图形区域背景色的设置。如果想在黑白两种颜色间快速切换背景色，点击工具栏上的  按钮即可。单击工具栏上的 ，背景显示网格，再次单击，背景不显示网格；双击则弹出颜色对话框，选择颜色，确定网格的颜色。

3.18 查找图形节点

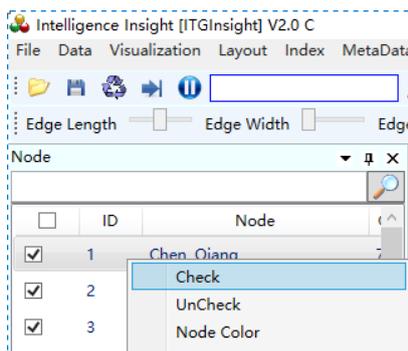
在工具栏的 中输入要查找的节点名称，点击 ，图形显示区域将突出显示被查找节点的名称。

3.19 删除/恢复节点

主页面左侧显示可视化图形节点的名称、数量、聚类结果、节点的连接边数量等信息。可以通过 控制节点的显示与否。如果希望对多个节点同时进行显示与否控制，按住 **shift** 键，再屏幕左侧依次点击节点的名称，然后右键点击。



右键点击弹出菜单如下：



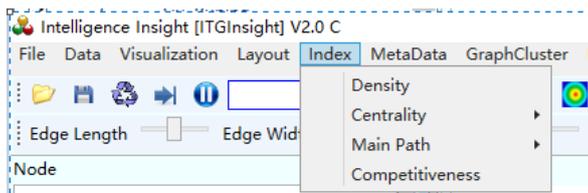
通过 check/uncheck 批量控制节点显示与否。在此需要注意，当去除某个节点后，又要恢复这个节点显示，这个过程不能更换布局算法。

如果需要删除图形中没有连接线的节点，在图形区域右键点击 **Delete Isolated Node** 即可。

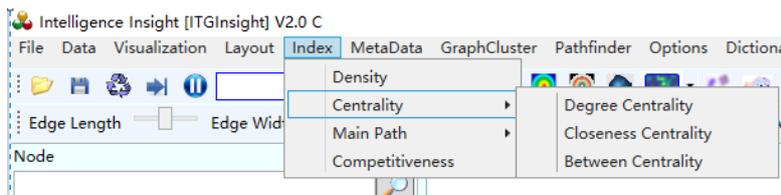
也可以在图形区域，按住 Ctrl 键的同时用鼠标选择一个区域，然后针对这个区域内的节点进行删除，或针对这个区域外的节点进行删除。分别点击 **Delete Node** 和 **Delete Reverse Node** 即可。

3.20 计算网络密度、节点中心度与主路径指标

(1)点击菜单栏的 Index/指标->Density/密度，计算网络密度，如下图。



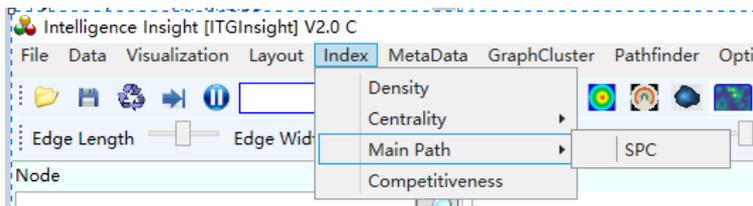
(2) 点击菜单栏的 Index/指标->Centrality/中心度，分别计算网络图中节点的点度中心度、接近中心度和中介中心度，操作如下图。相关中心度的概念和应用参见 paper 文件夹下的学术论文“网络中心度用于期刊引文评价的有效性研究”。



计算结果显示在软件左侧的节点明细，如下图。

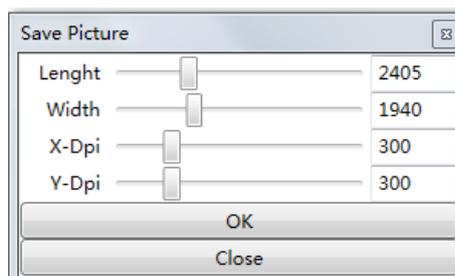
Node	Degree Centrality	Closeness	Between
<input checked="" type="checkbox"/> C-J斯特朗	0.6735	0.6125	0
<input checked="" type="checkbox"/> D-J科斯特	0.6531	0.6049	0
<input checked="" type="checkbox"/> B-K安德	0.6122	0.5904	0
<input checked="" type="checkbox"/> M-D罗尔	0.4898	0.5506	0
<input checked="" type="checkbox"/> R-霍博德	0.4694	0.5444	0
<input checked="" type="checkbox"/> D-R克	0.4898	0.6125	0
<input checked="" type="checkbox"/> R-P霍夫	0.4490	0.5385	0
<input checked="" type="checkbox"/> E-A黄	0.4490	0.5385	0
<input checked="" type="checkbox"/> J-P艾夫	0.4898	0.5506	0
<input checked="" type="checkbox"/> D-德乌利	0.4490	0.5385	0
<input checked="" type="checkbox"/> D-B伊茨	0.4898	0.5213	0
<input checked="" type="checkbox"/> S-乔布斯	0.5714	0.5632	0
<input checked="" type="checkbox"/> C-Q安德	0.4694	0.5158	0
<input checked="" type="checkbox"/> 西德信	0.4286	0.5326	0
<input checked="" type="checkbox"/> P-拉斐尔	0.3673	0.4900	0
<input checked="" type="checkbox"/> E-汉基	0.4286	0.5326	0
<input checked="" type="checkbox"/> 西德信	0.3673	0.4900	0
<input checked="" type="checkbox"/> J-巴巴光	0.4082	0.5269	0
<input checked="" type="checkbox"/> J-阿卡的	0.3878	0.5104	0
<input checked="" type="checkbox"/> P-拉斐尔	0.3878	0.5213	0
<input checked="" type="checkbox"/> B-奥丁	0.1429	0.4375	0
<input checked="" type="checkbox"/> S-P霍特	0.1020	0.4414	0
<input checked="" type="checkbox"/> D-R科	0.2653	0.4224	0
<input checked="" type="checkbox"/> S-福斯塔	0.0816	0.3121	0
<input checked="" type="checkbox"/> G-克里斯	0.0816	0.3121	0
<input checked="" type="checkbox"/> I-乔德	0.0816	0.3121	0
<input checked="" type="checkbox"/> S-O勒博	0.1020	0.3182	0

(3) 点击菜单栏的 Index/指标->Main Path/主路径，分别计算网络图中节点的 SPC 值，操作如下图。主路径概念和应用参见 paper 文件夹下的学术论文“专利引用网络主路径方法研究述评与展望_张娴”。计算结果查看方法与中心度查看方法一致。

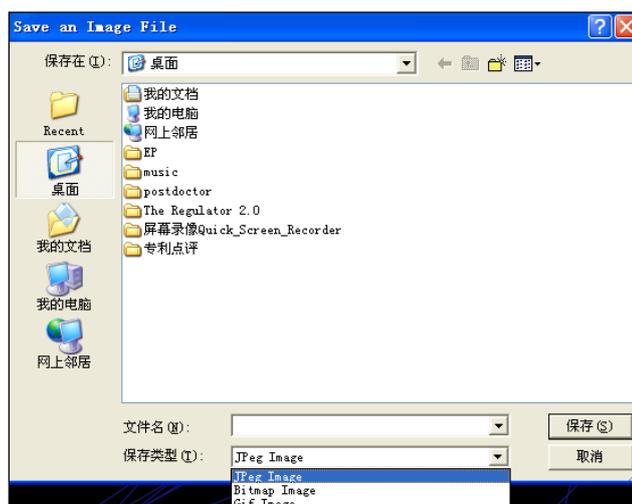


3.21 屏幕截图

点击工具栏上的 ，先后弹出截图设置和图形文件保存对话框，按提示保存分析的图形，如下图。



上图提示截图的大小和分辨率，xdpi 与 ydpi 越大，图像越清晰，大小越大，默认值 300 可满足打印需求。设置后，继续按下图保存截图后的文件。



3.22 输出 Excel 数据表

打开一个 itgn 文件后，直接点击工具栏上的 ，系统会输出各种统计数据，输出内容与输出的 word 报表内容相同，同时增加了报表目录到第一个 sheet 中。

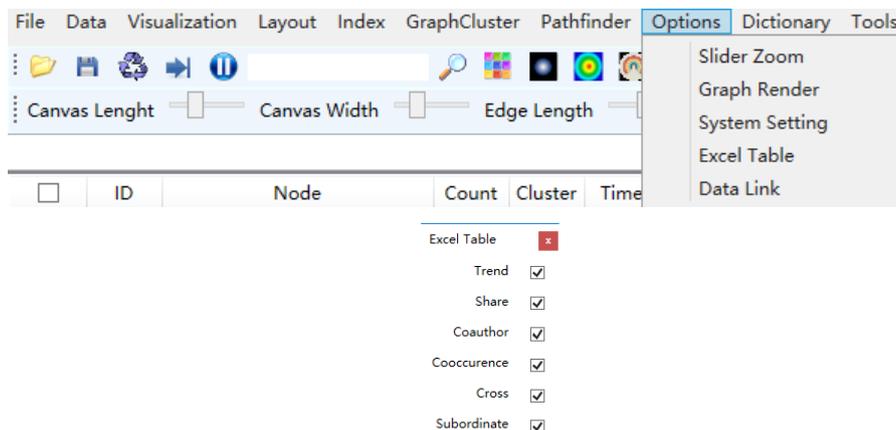
打开一个 mode 文件，或者打开 itgn 文件进行合著、同现、关联、引证分析，有图形在图形区显示后，直接点击工具栏上的 ，系统将把图形中的节点数据打印到微软 Excel 表格中，效果图如下。

	A	B	C	D	E	F
1	ID	NAME	Remark	Number	Turnnumber	Referencednumber
2	0	Ghemawat,	1/1 707/999.101	23	7;4;12	288;17
3	0	Baluja; S	382/305 707/999.104	19	10;7;2	260;0
4	0	Rasmusser	701/208 340/995.14	18	12;5;1	333;21
5	0	Tong; Sir	1/1 707/999.003	18	10;2;6	307;2
6	0	Rasmusser	701/208 340/995.14	15	1;10;4	240;22
7	0	Dean; Jef	1/1 707/999.1	13	3;3;7	226;27
8	0	Lawrence,	707/999.003 707/748	12	3;8;1	242;7
9	0	Petry; Sc	709/206 709/223	12	5;6;1	248;6

此功能需安装微软的 office 2007 或以上的版本。

3.23 Excel 报表输出内容设置

默认情况下，系统的 Excel 报表仅仅输出趋势、份额等一维的统计报表，如果希望输出更多内容，可在“Options/选项”——>“Excel Table/报表输出”中进行更细致的设置，如下图所示。



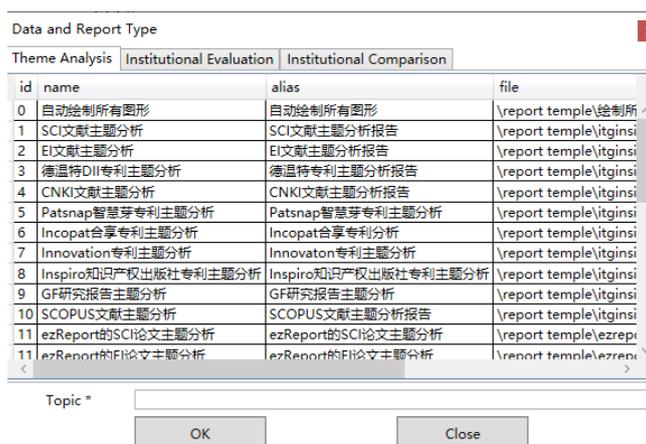
输出内容越多，打印报表时间越长，并且保证在 3.1 进行数据转换时，生成的 itgn 文件进行了有关的分析。以 SCI 论文为例，全部分析报表多达 90 个，此时建议采用 EXCEL 输出。

3.24 输出 Word 智能报告

ITGinsight 提供计算机的自动写报告功能，首先打开一个 itgn 文件，点击工具栏上的



，弹出智能报告对话框如下。



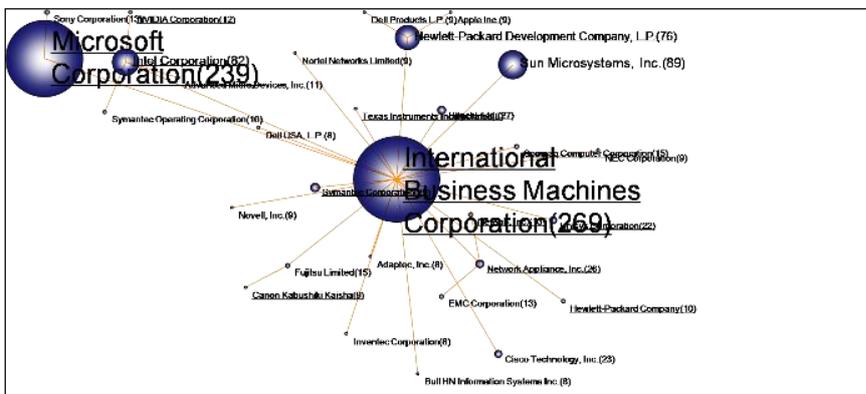
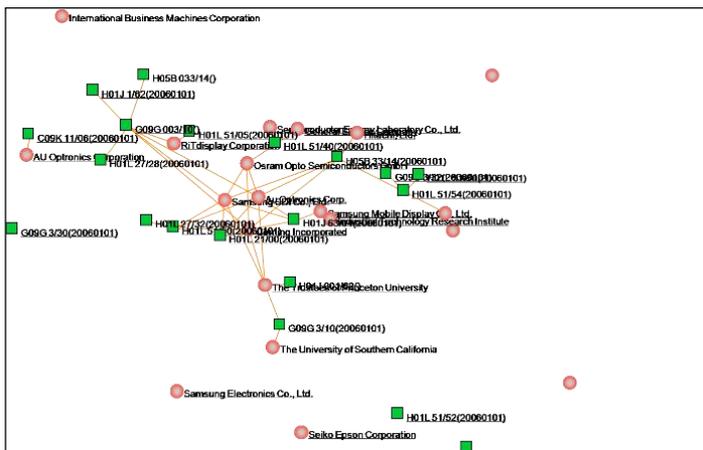
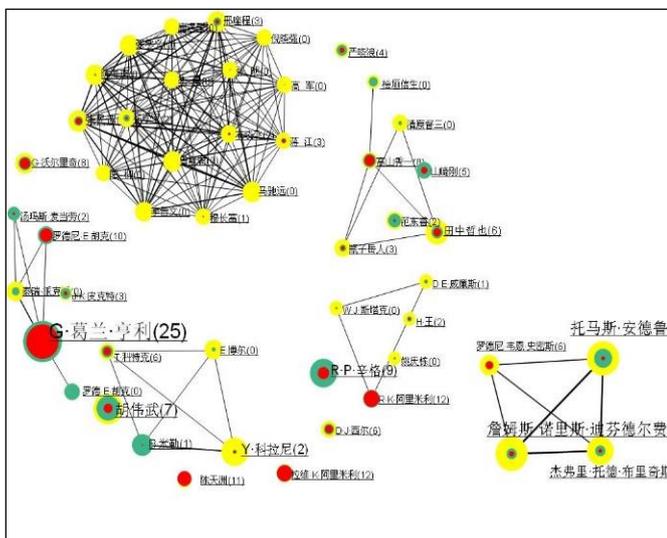
选中一个适合的报告模板，在 topic/主题部分填写技术领域，点击确定，软件自动输出相对完整的报告，根据提示对报告做适当修改。

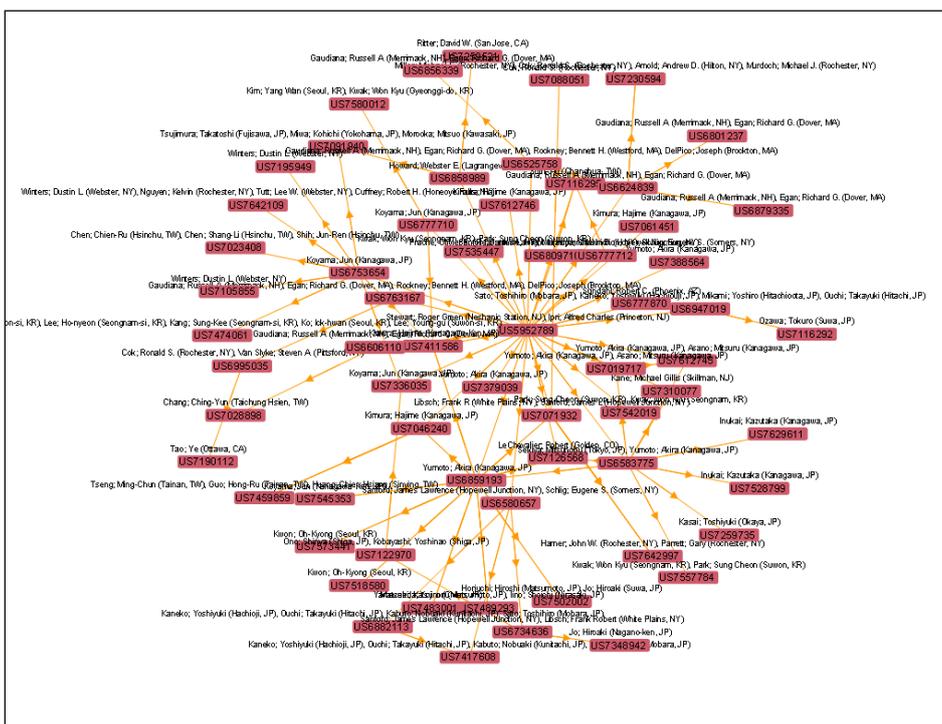
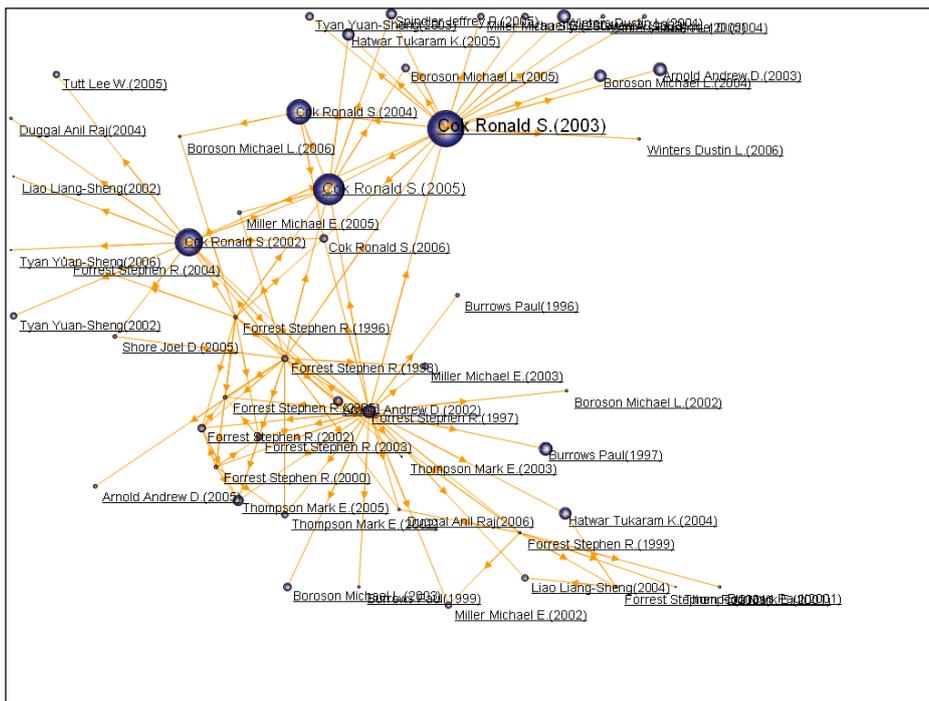
注释：系统针对企业版顶级用户提供默认的两个报告模板，普通用户无法使用该功能。如果想针对其他数据源或者其他报告的模式，ITGinsight 提供更多的模板，但是需要技术服

务费。

3.2.5 输出 PPT 演示文稿

点击工具栏上的，系统将把图形直接在微软 Power Point 上打印出来，效果图如下：





此功能需安装微软的 office 2007 或以上的版本。

3.26 打开保存 Mod 图形文件

每次分析，为了保存当前的分析结果，点击工具栏上的 ，或菜单栏的“File/文件”——>“Save/保存”，将当前分析结果保存为*.mod 文件。在下次使用时，只要点击工具栏上的 ，或菜单栏的“File/文件”——>“Open/打开”，导航到*.mod 文件进行打开即可。Mod 文件保存内容包含三方面：1、节点位置信息，2、图形样式信息（颜色、阈值、大小、长短、

粗细等相关设置等)，3、节点内容信息（节点文字，备注、数量、时间等）。

3.27 打开保存 layout 位置信息文件（位置信息的重复利用）

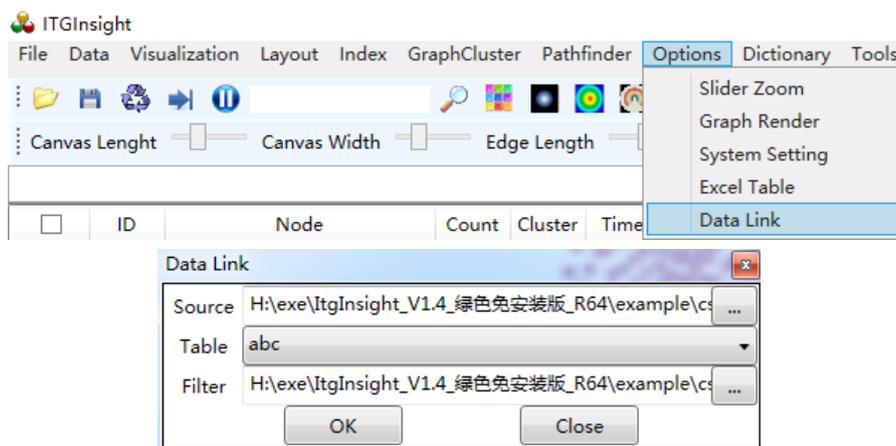
每次分析，为了保存当前的分析结果中的位置信息，即 mod 文件中的第一类信息，使得相同名称的节点位置不变，点击工具栏上的 ，或菜单栏的“File/文件”——>“Save/保存”，将当前分析结果保存为*.layout 文件。在下次使用时，只要点击工具栏上的 ，或菜单栏的“File/文件”——>“Open/打开”，导航到*.layout 文件，即可使相同名称的节点位置不变。

3.28 打开保存 graph style 样式信息文件（样式信息的重复利用）

每次分析，为了保存当前的调整后的样式信息，即 mod 文件中的第二类信息，使得样式能够重复利用，点击工具栏上的 ，或菜单栏的“File/文件”——>“Save/保存”，将当前分析结果保存为*.Graph style 文件。在下次使用时，只要点击工具栏上的 ，或菜单栏的“File/文件”——>“Open/打开”，导航到*.Graph style 文件进行打开即可。

3.29 可视化图形与文献数据交互

每一张可视化图形一定对应一个原始的数据源，如果这个数据源存储在 access 或 excel 中，则可以通过图形与原始数据进行交互，首先指定可视化图形的数据源，点击 option/选项下 data link/数据连接，如下图：



导航到可视化图形所使用的 access 或 excel 数据源，并指定应用了数据源中的那个表，分析时采用的过滤器，设定之后，则可以通过图形调取原始数据。

在可视化图形输出中，双击节点或连线弹出可视化元素对应的文献数据列表，如下图。默认情况只会弹出节点对应的文献数据，如果希望弹出连线对应的文献数据，点击工具栏上的  进行连线数据的显示。（该功能仅限于 access、excel 数据文件，对存储在 txt 中的数

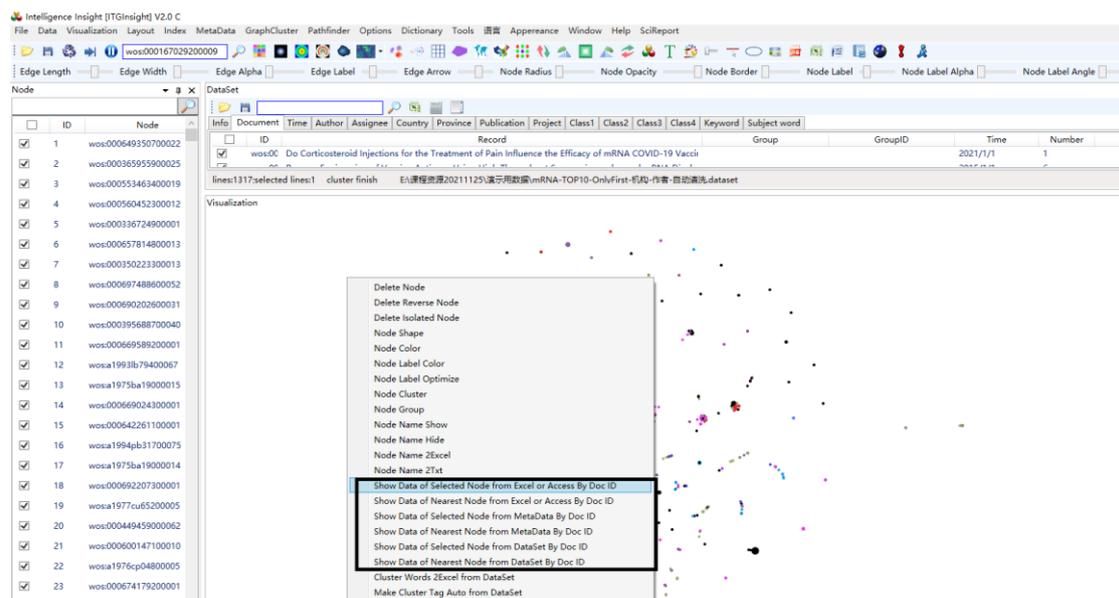
据文件不适用)

专利号	题目	申请人	国家	发明人	ID:
CN201130076502.X	USB设备	苹果公司	US	J·阿卡纳 B·K·安德烈 J·巴塔尤 D·J·科斯特	158
CN200930006608.5	包装盒	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	专利号:
CN200830005183.1	包装盒	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	CN2
CN200730350750.2	包装盒	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	007
CN200730350765.9	包装件	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	303
CN200730350751.7	包装件	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	507
CN200730350748.5	包装件	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	50.
CN201030152616.3	包装物	苹果公司	US	J·阿卡纳 B·K·安德烈 J·巴塔尤 D·J·科斯特	2
CN200730350752.1	包装用隔板	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	申请日:
CN200730350753.6	包装用托板	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	200
CN200730350749.X	包装用托板	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	7-1
CN200730299964.1	保护壳	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	2-2
CN200730299960.3	保护壳	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	0
CN200730299962.2	保护壳	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	0:0
CN200730299963.7	保护壳	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	0:0
CN200830005193.5	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	0
CN200830005195.4	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	2
CN200830005175.7	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	题目:
CN200830005177.6	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	包装盒
CN200830005181.2	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	申请人:
CN200830005179.5	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	苹果公司
CN200830005176.1	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	发明
CN200830005192.0	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	
CN200830005194.X	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	
CN200830005180.8	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	
CN200830005178.0	保护套	苹果公司	US	B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·	

将鼠标置于蓝色线条处调整显示屏的内容，双击左侧表中任意一行，右侧显示该行的更进一步信息，如下图。

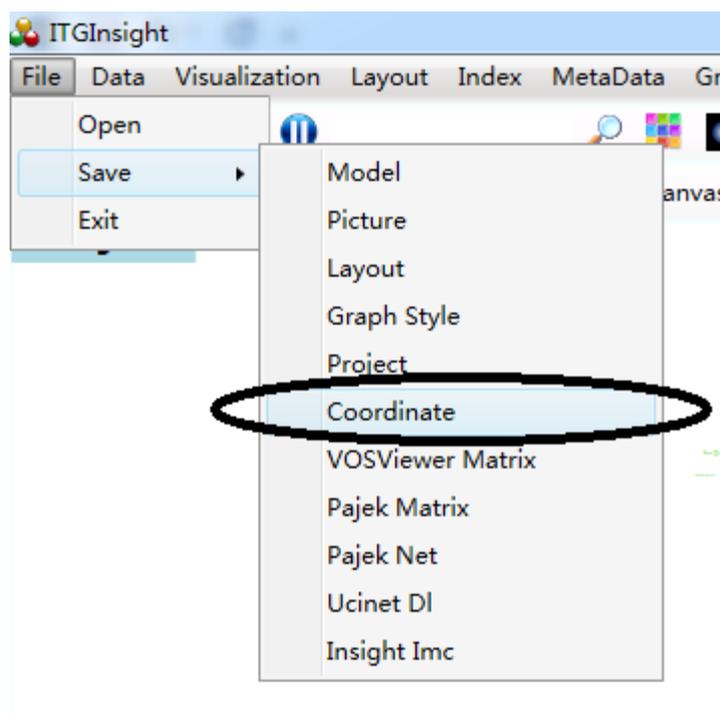
专利号	题目	申请人	国	ID:158
CN201130076502.X	USB设备	苹果公司	US	专利号:CN200730350750.2
CN200930006608.5	包装盒	苹果公司	US	申请日:2007-12-20 0:00:00
CN200830005183.1	包装盒	苹果公司	US	题目:包装盒
CN200730350750.2	包装盒	苹果公司	US	申请人:苹果公司
CN200730350765.9	包装件	苹果公司	US	发明人:B·K·安德烈 D·J·科斯特 D·德乌利斯 R·P·
CN200730350751.7	包装件	苹果公司	US	豪沃斯 J·P·艾夫 S·乔布斯 D·R·克尔 西姆信 M·D·
CN200730350748.5	包装件	苹果公司	US	罗尔巴克 P·拉塞尔-克拉克 D·B·萨茨格 C·Q·塞德
CN201030152616.3	包装物	苹果公司	US	C·J·斯特林格 E·A·黄 R·策肯德费尔
CN200730350752.1	包装用隔板	苹果公司	US	摘要:
CN200730350753.6	包装用托板	苹果公司	US	权利要求:
CN200730350749.X	包装用托板	苹果公司	US	同族数量:
CN200730299964.1	保护壳	苹果公司	US	被引用数量:
CN200730299960.3	保护壳	苹果公司	US	国家:US
CN200730299962.2	保护壳	苹果公司	US	IPC:09-03
CN200730299963.7	保护壳	苹果公司	US	UPC/EPC:
CN200830005193.5	保护套	苹果公司	US	引用专利号:
CN200830005195.4	保护套	苹果公司	US	被其它专利引用:
CN200830005175.7	保护套	苹果公司	US	关键词:
CN200830005177.6	保护套	苹果公司	US	
CN200830005181.2	保护套	苹果公司	US	
CN200830005179.5	保护套	苹果公司	US	
CN200830005176.1	保护套	苹果公司	US	
CN200830005192.0	保护套	苹果公司	US	
CN200830005194.X	保护套	苹果公司	US	
CN200830005180.8	保护套	苹果公司	US	
CN200830005178.0	保护套	苹果公司	US	

如果图形是采用 6.8 节聚类分析得到的聚类图，则需要鼠标右键，如下图解中国方式，连接到原始数据。



3.30 导出坐标

按下入操作导出坐标，坐标文件格式为 tsv 格式。



3.31 导出图例

在可视化区域，鼠标右键点击 Legend/图例，图例会导出到 PPT 中，在 PPT 中对图例进行修改。

3.32 一次性绘制所有可视化图形到 Word 中

打开 .itgn 文件后，点击工具栏 DrawingRobot/绘图机器人，软件自动绘制所有图形到 word 中，如下是绘制的 SCI 论文分析图形的图形目录。

图的目录

图 1:历年论文数量及其增长率趋势图	3
图 2:历年论文数量及其累积数量趋势图	3
图 3:技术生命周期图	4
图 4:coabswordgraph	5
图 5:cokeywordgraph	6
图 6:coaffiliationgraph	7
图 7:coauthorgraph	7
图 8:coclass1graph	8
图 9:coclass2graph	8
图 10:cocountrygraph	9
图 11:coprojectgraph	9
图 12:coprovincegraph	10
图 13:copublicationgraph	10
图 14:corelationAssigneeAbswordgraph	11
图 15:corelationAssigneeKeywordgraph	11
图 16:corelationAssigneeClass1graph	12
图 17:corelationAssigneeClass2graph	12
图 18:corelationAuthorAbswordgraph	13
图 19:corelationAuthorClass1graph	13
图 20:corelationAuthorClass2graph	14
图 21:corelationAuthorKeywordgraph	14
图 22:corelationCountryAbswordgraph	15
图 23:corelationCountryClass1graph	15
图 24:corelationCountryClass2graph	16
图 25:corelationCountryKeywordgraph	16
图 26:corelationProvinceAbswordgraph	17
图 27:corelationProvinceClass1graph	17
图 28:corelationProvinceClass2graph	18
图 29:corelationProvinceKeywordgraph	18
图 30:corelationPublicationAbswordgraph	19
图 31:corelationPublicationClass1graph	19
图 32:corelationPublicationClass2graph	20
图 33:corelationPublicationKeywordgraph	20
图 34:corelationYearAbswordgraph	21
图 35:corelationYearClass1graph	21
图 36:corelationYearClass2graph	22
图 37:corelationYearKeywordgraph	22
图 38:evolutionAbswordgraph	23
图 39:evolutionAffiliationgraph	23
图 40:evolutionAuthorgraph	24
图 41:evolutionClass1graph	24
图 42:evolutionClass2graph	25

同时，word 中所有图形的 mod 矢量图和 png 截图都可以在软件 report temple 目录下找到，用户可以对 mod 矢量图进行再加工。

3.33 快捷键

软件提供了部分快捷键操作，如下表：

快捷键操作	
功能	快捷键

以矩形的方式选择多个节点作为子图	鼠标左键按下+Ctrl 键，移动鼠标
选择与某个节点联通的节点作为子图	Ctrl+Shift 键，鼠标左键点击某一个节点
平移整个图形	鼠标左键按下+Shift，移动鼠标
平移子图	选择子图，鼠标左键按下+Shift，移动鼠标
孤立节点平均分布在页面边缘	按键 C 或 c
图形优化开始/暂停	Enter
平移标签或者演化分析的年代标签	鼠标左键按下+Alt，移动鼠标

3.34 退出

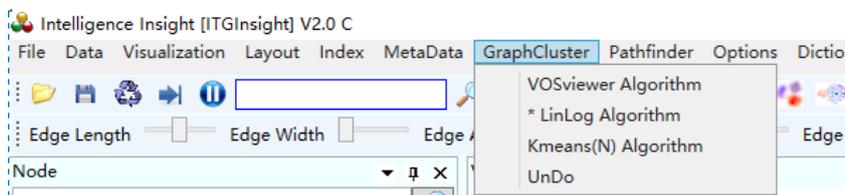
点击工具栏上的 ，或菜单栏的“File/文件”——>“Exit/退出”，系统安全退出。

第四章：聚类分析、热力图/地形图/密度图、世界地图、气象图、矩阵图可视化

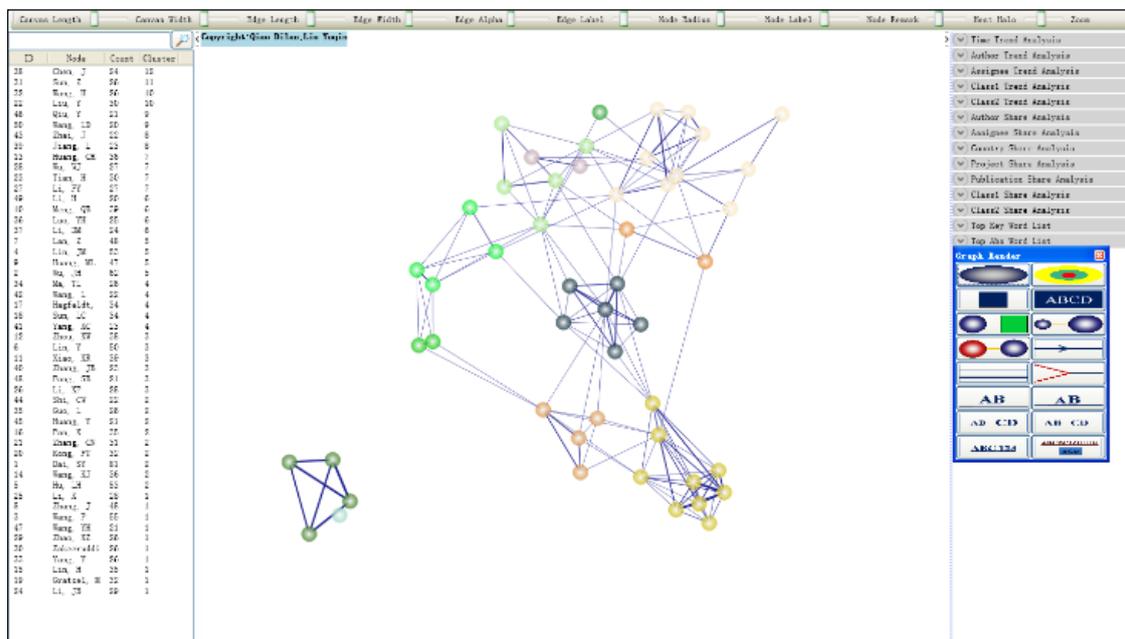
4.1 网络图聚类分析

在进行合著、同现、耦合、关联、引证分析时，如果按照 3.3-3.8 的网络图进行显示，可以进一步对网络图进行二次聚类，二次聚类在网络图节点数量较多的情况下，能够更加清晰的表示网络结构。具体操作如下：

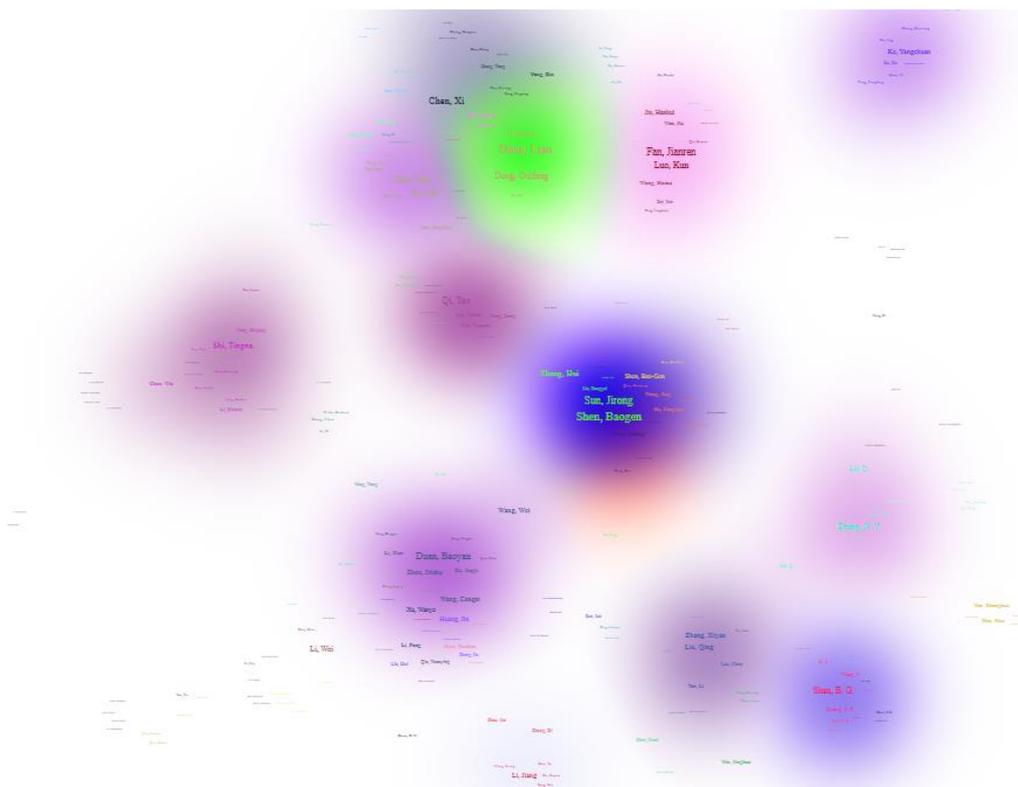
1) 点击菜单栏“GraphCluster/图聚类” ——> “Vosviewer Algorithm” 或 “LinLog Algorithm” 或 “Kmeans (N)” ——> “Do/执行” 或 “UnDo/撤销” 对网路进行聚类或取消聚类，如下图。前两种聚类数量有算法确定，不能调整，Kmeans (N) 聚类数量默认为 5，可通过 slider 面板进行调整。



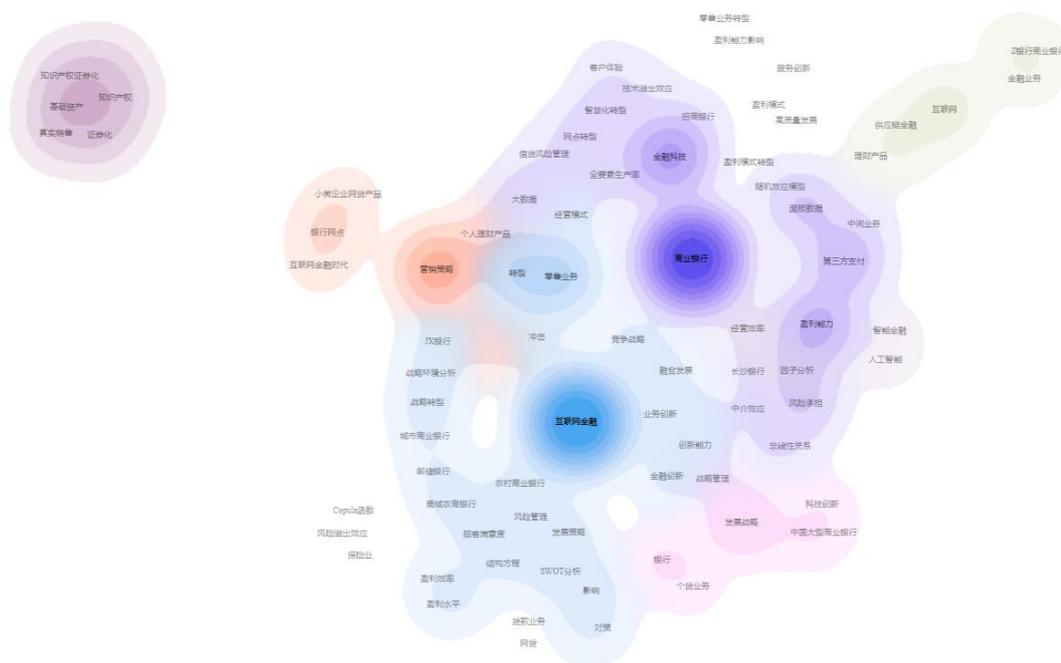
2) 在主页面左侧的节点内容面板中查看聚类结果，也可以图形显示区的节点颜色区别节点所属类别，如下图。



18) 点击工具栏的  按钮，网络聚类图显示模式如下图效果。

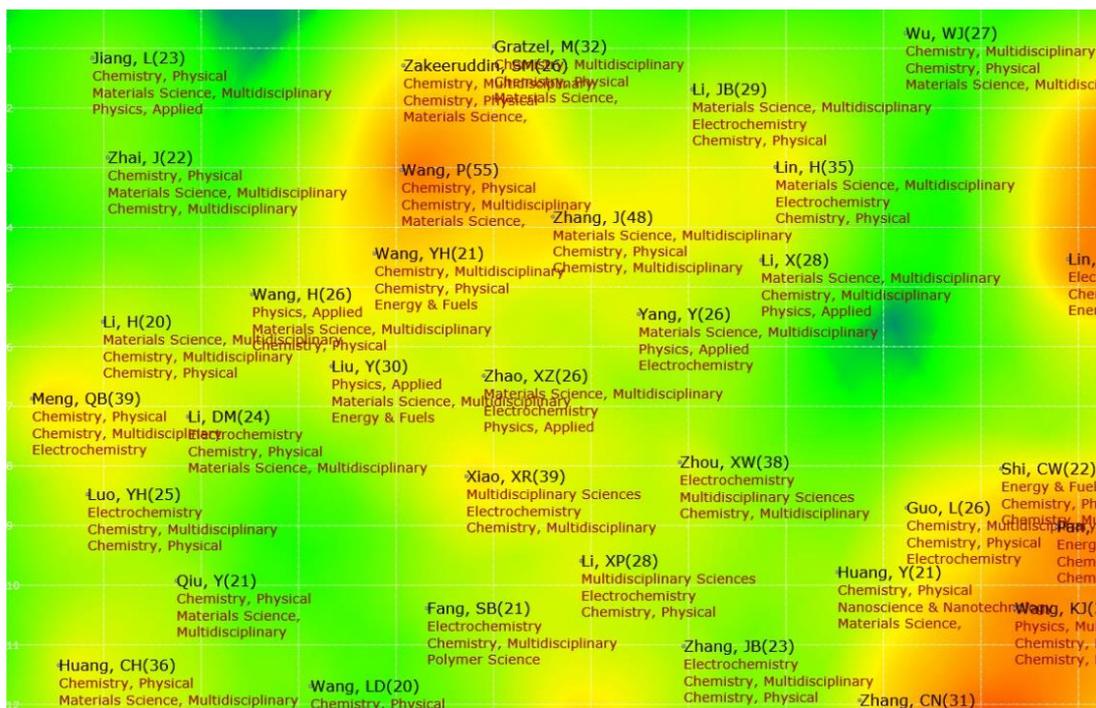


19) 点击工具栏的  按钮，网络聚类图显示模式如下图效果。

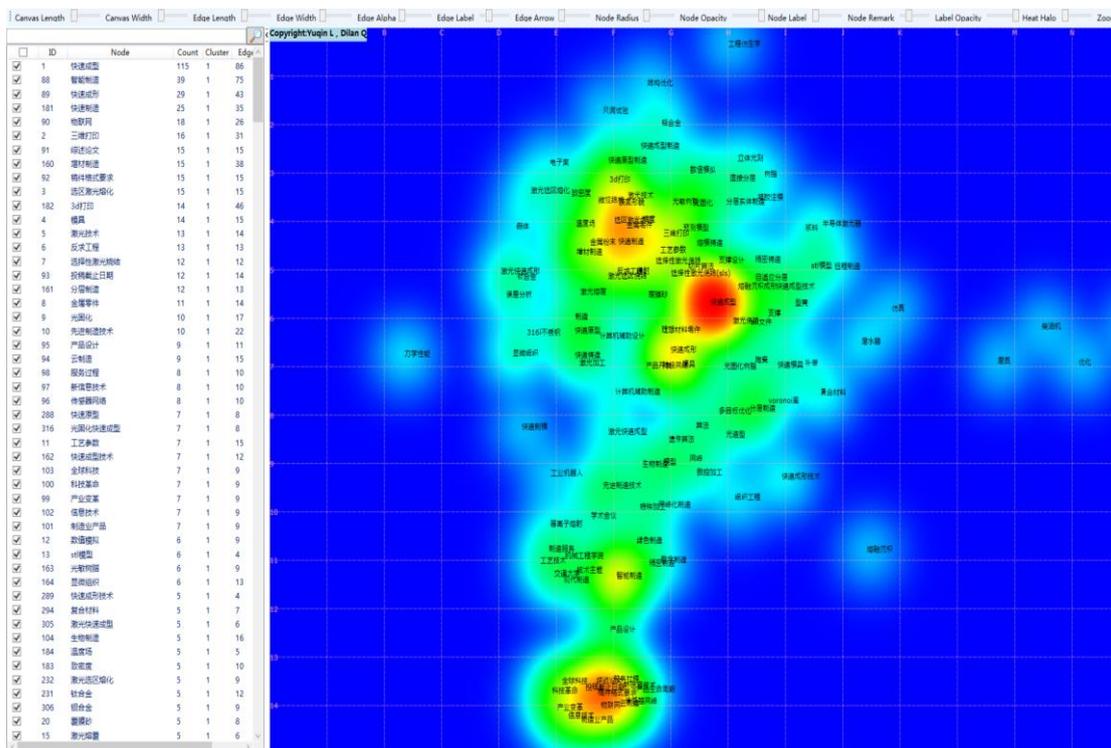


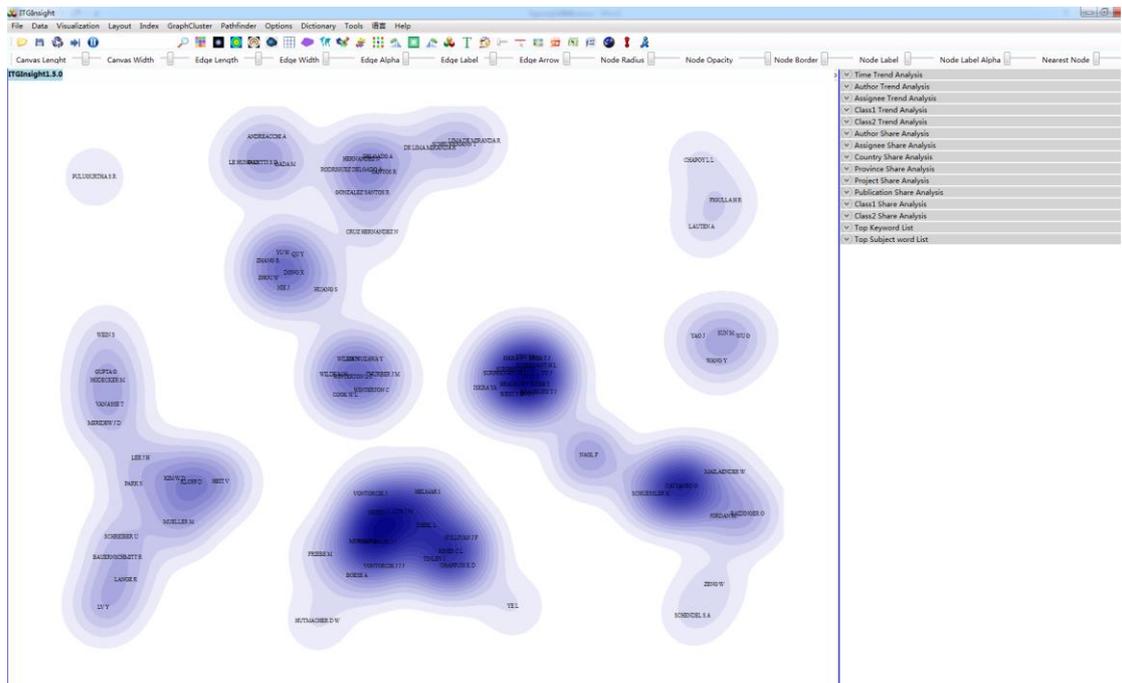
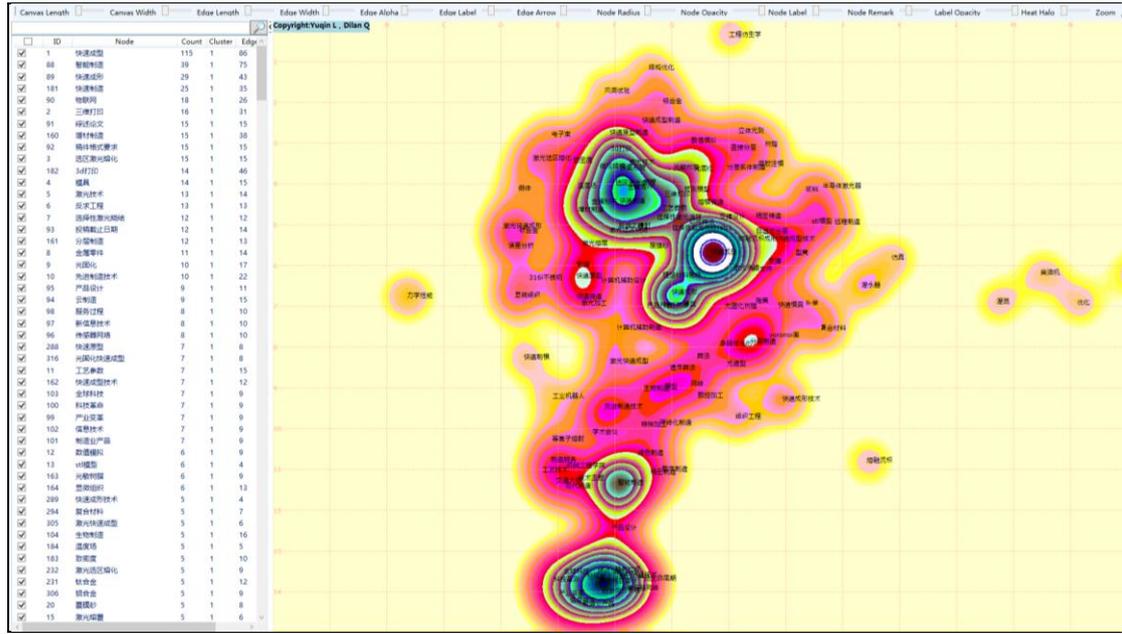
4.2 热力图/地形图/密度图可视化

系统可视化结果以网络图为主，同时提供热力图/地形图/密度图可视化。热力图是对自然界的热力成像原理的计算机模拟，通过红黄绿蓝四种颜色的深浅来区别数据的大小，颜色块区别数据的密集程度，参考下图。



具体操作如下：





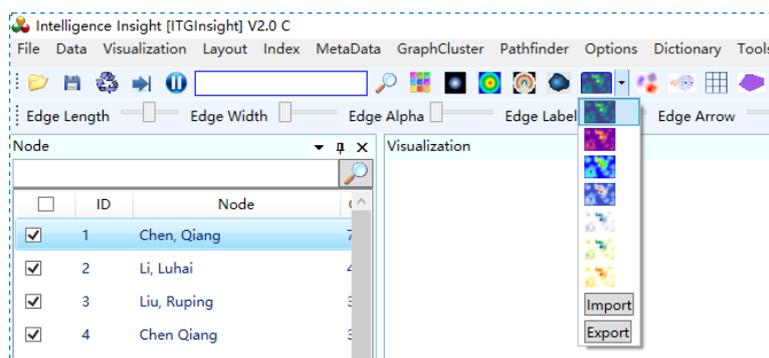
1) 点击菜单栏“Layout/布局”——>“VS 布局或 FS 布局”。

2) 对经过布局的网络图，再次点击菜单栏的  或  或  按钮，系统提示后台操作状态条，待状态条消失后，图形区即可呈现热力图结果。

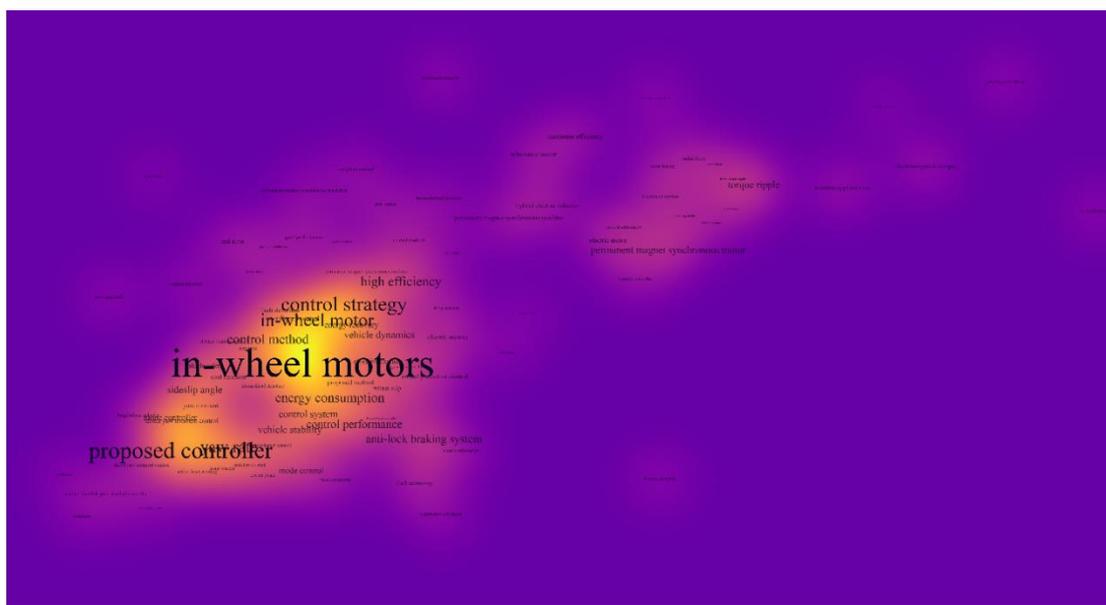
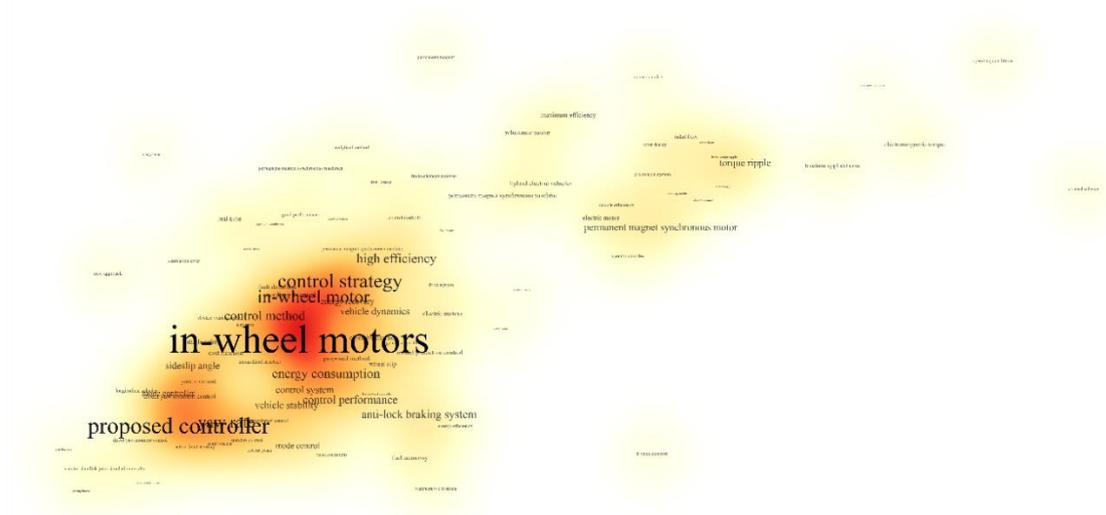
3) 按照 3.9-3.15 的操作方法，在热力图/地形图/密度图上进行相关内容的显示控制。

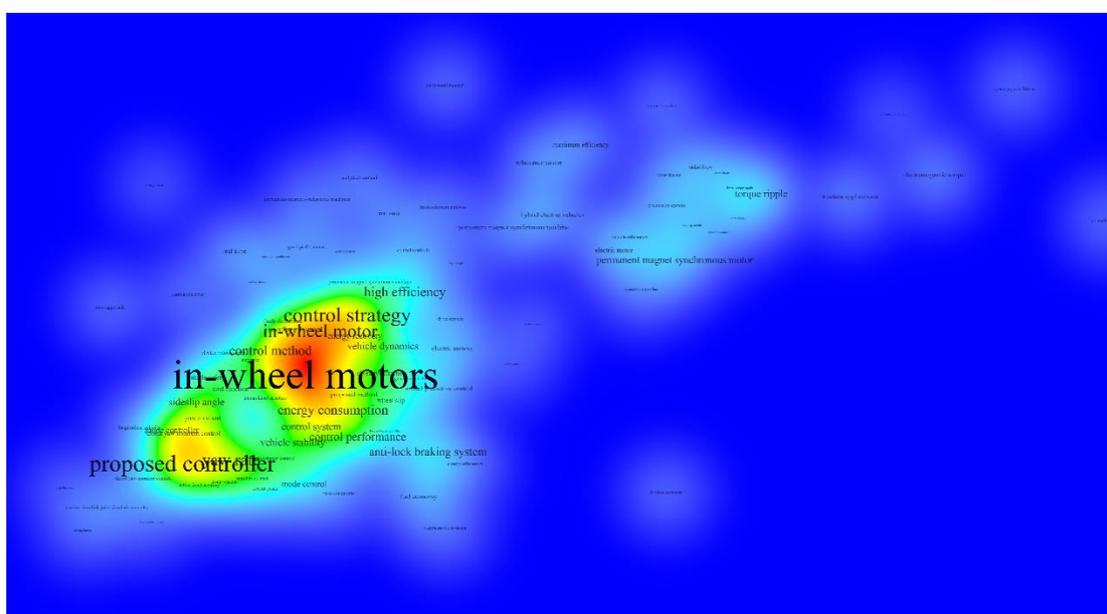
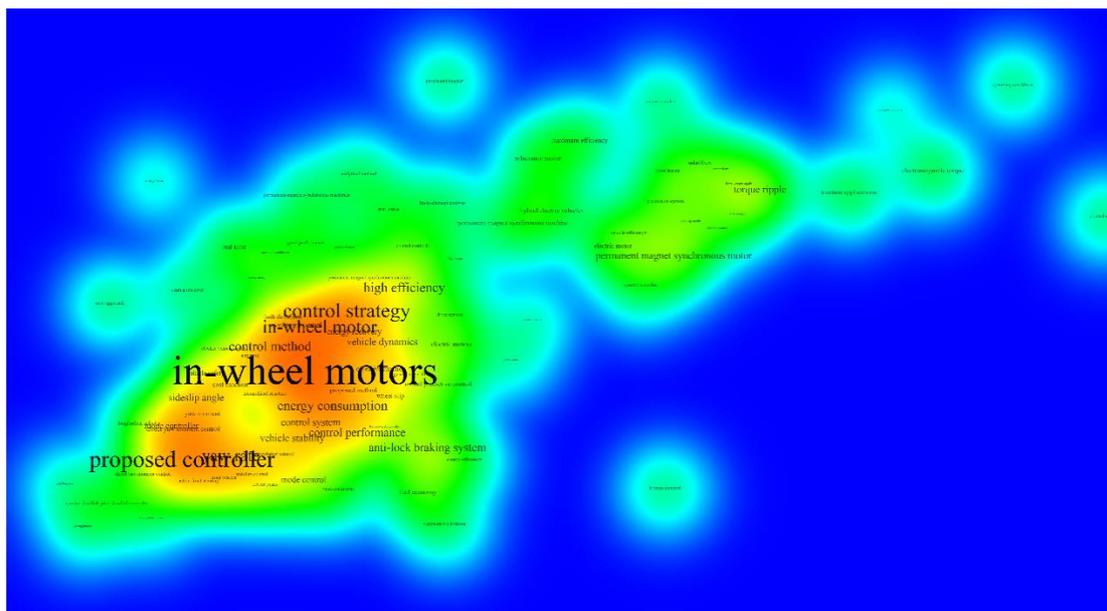
4) V1.6 版本以后增加了类似 VOSViewer 的主题图形式，操作方法是点击工具栏如下

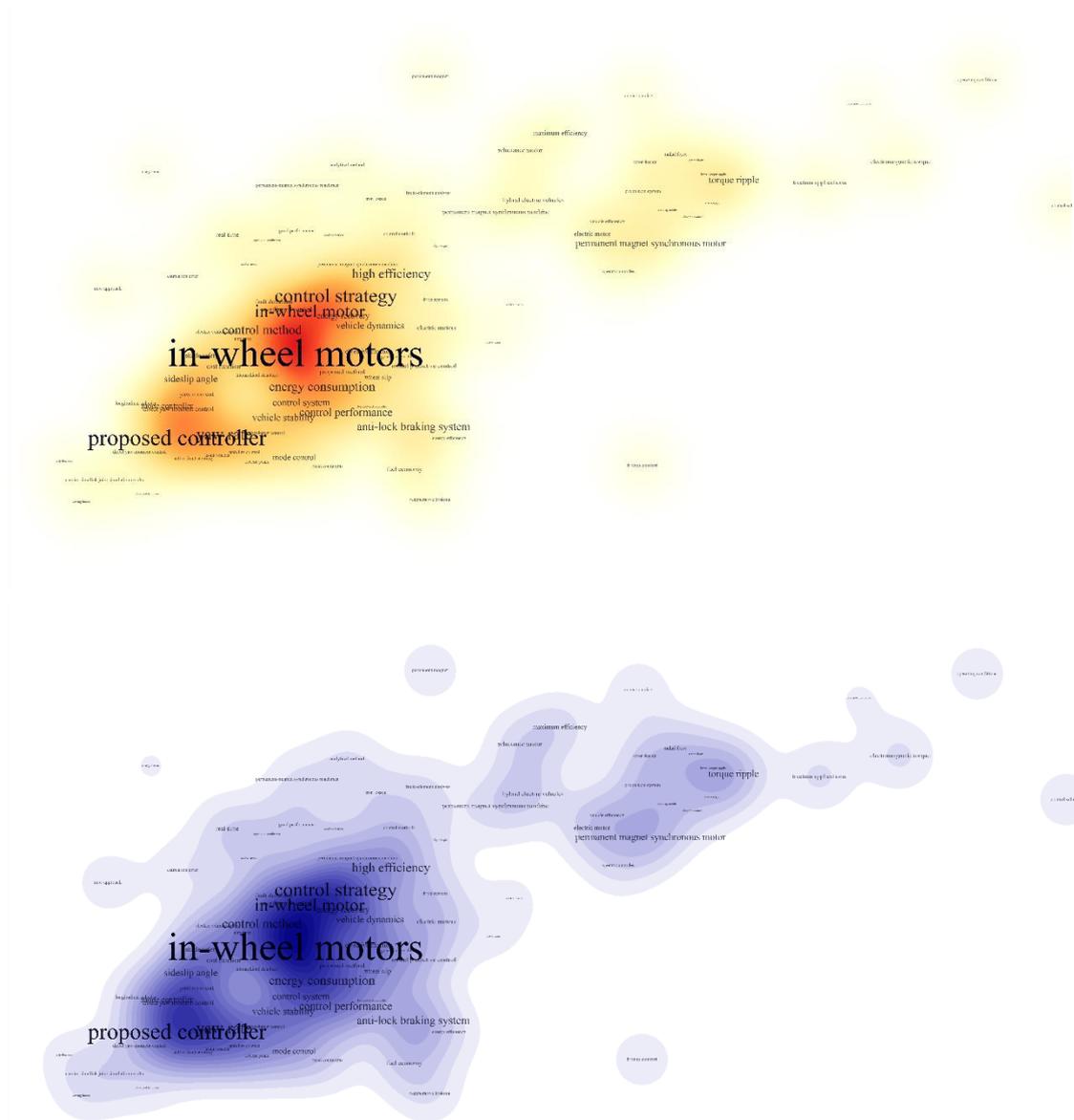
图：



可以修改主题图的样式，也可以通过导入、导出功能自定义主题图的颜色，自定义颜色格式参考可以选择一种颜色模式后，导出，观察颜色格式进行修改。以下为同一张图的不同效果：







4.3 世界地图可视化

系统提供世界地图可视化。在系统安装目录下的 lalo_world.txt 文件中（坐标信息可增加、修改，参考文件中已有的坐标格式即可），保存有世界主要国家的经纬度信息，当网络图中的节点名称出现在 lalo_world.txt 中时，点击工具栏的 ，采用世界地图布局进行可视化输出，也就是每个节点按照地理位置进行坐标确定。

4.4 中国地图可视化

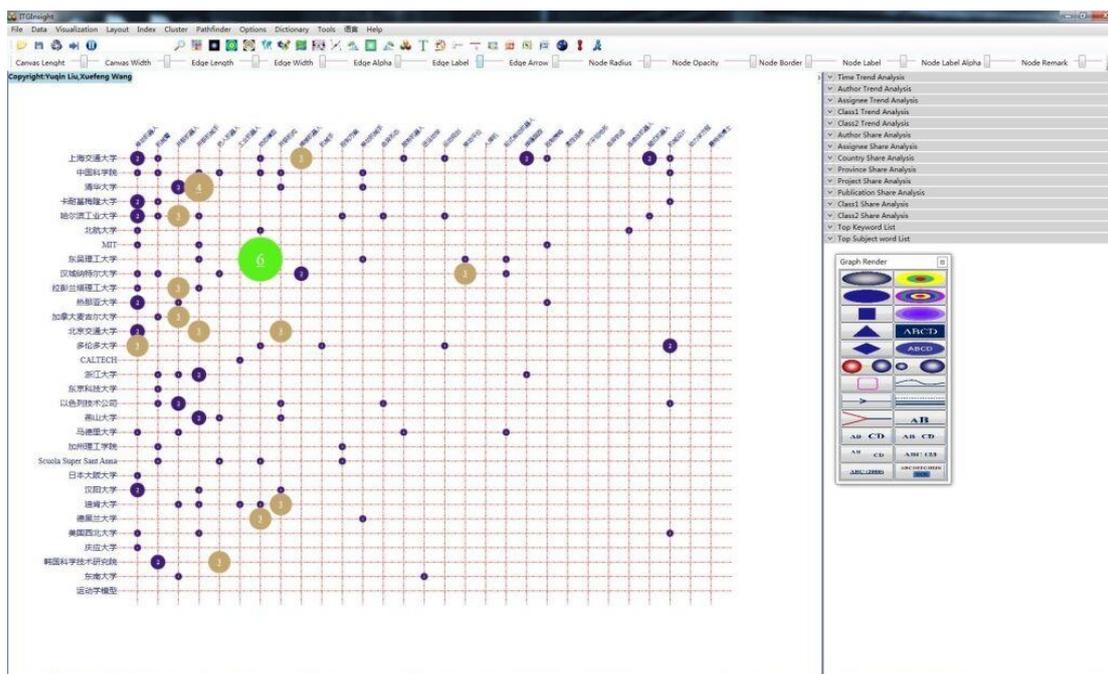
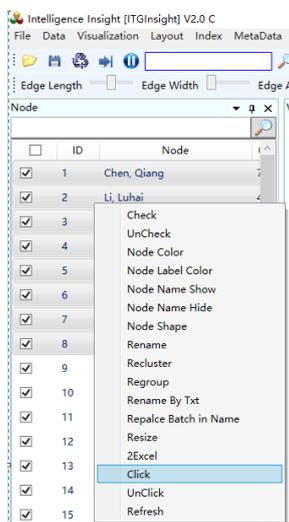
系统提供中国地图可视化。在系统安装目录下的 lalo_china.txt 文件中（坐标信息可增加、修改，参考文件中已有的坐标格式即可），保存有中国部分省市的经纬度信息，当网络图中的节点名称出现在 lalo_china.txt 中时，点击工具栏的 ，采用中国地图进行可视化输

出，也就是每个节点按照地理位置进行坐标确定。

4.5 矩阵图可视化

对于以网络图形式输出的可视化样式，通过点击工具栏的，使网络图的节点转化为矩阵的行或列，网络图的连接线转化为矩阵行列交叉部分的节点，图形转化为矩阵形式，如下图所示。其中，鼠标单击行的名称，再次点击，行变为列；反之，亦然。对于节点文字、节点文字（仅是数字）的大小、颜色调整，与网络图的操作相同。

注意：网络图能够转化为矩阵图的条件是，网络图中的节点分为两个颜色，即选中状态和非选中状态。可以在软件左侧，选中若干节点，右键点击 click，然后再点击矩阵按钮，即可出现矩阵效果，如下图所示。



第五章：系统设置

5.1 使用过滤器切换分析对象

ITGinsight 分析对象为中英文的专利、论文、报告，并定义过滤器进行分析对象的切换。系统预先设计了经常分析的 SCI、CNKI、德温特、ItgMining 数据的过滤器，下图是分析由 Web Of Science 下载的 SCI 数据所使用的过滤器。

```
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
<?xml version="1.0" encoding="utf-8"?>
<Config>
  <FieldMap>
    <Source>Wos1</Source>
    <ID>DI</ID>
    <Keyword>DE</Keyword>
    <Abstract>TI|AB</Abstract>
    <Authors>AF</Authors>
    <Affiliation>C1</Affiliation>
    <Class1>WC</Class1>
    <Class2>SC</Class2>
    <Class3>PT</Class3>
    <Class4>DT</Class4>
    <Countries>C1</Countries>
    <Provinces>C1</Provinces>
    <Founders>FU</Founders>
    <Publication>SO</Publication>
    <Description></Description>
    <Reference>CR</Reference>
    <ReferencedBy></ReferencedBy>
    <Time>PY</Time>
    <Title>TI</Title>
    <Number1>TC</Number1>
    <Number2>CR</Number2>
    <Number3></Number3>
  </FieldMap>
</Config>
```

其中<FileMap></FileMap>节点下的<ID><Abstract>...<Title>分别用来设置 access (excel 或 txt) 数据库中的对应字段名，只要设置这些节点对应的专利、报告、论文所存储的数据对应字段即可。

<Source>数据来源，可以是 wos, cnki, itgmining, 或者任意其它来源 default

<ID>对应唯一标准字段，必须存在，否则出错；

<Abstract>对应摘要字段；可对应多个字段

<Authors>对应作者字段；

<Affiliation>对应机构字段；

<Class1>对应类别字段；

<Class2>对应类别字段；

<Class3>对应类别字段；

<Class4>对应类别字段;

<Keyword>对应关键词字段; 可对应多个字段

<Countries>对应地区字段;

<Publication>对应期刊字段;

<Description>对应正文字段;

<Reference>对应引用文献 ID;

<ReferencedBy>对应被引文献 ID;

<Time>对应时间;

<Title>对应题目

<Number1>对应被引用数量

<Number2>

<Number3>

如果对应字段不存在, 对应部分不填写任何内容, 如本例中的<Description>节点。

5.2 分析阈值、参数设置

在 sysset.xml 中<System></System>节点下的内容用来设置与系统或分析有关的参数。

<SoftName>用于更改软件启动后的界面显示名称;

<SoftCopyRight>用于设置软件启动后页面左上角的版权信息;

<VSMwordMax>用于指定进行语义计算时最多采用的词语数量, 越多, 对系统的硬件要求越高、计算时间越长;

<StopWordFile>用于设定停用词文件路径, 默认为 stopwords.txt;

<ThesaurusFile>用于设定主题词文件路径, 默认为 thesaurus.txt;

<DataFile>用于设定 Access 数据文件路径;

<CoThreshold>用于指定合著分析的阈值, 如果超过该值, 才认为进行合著计算, 在图形上表示为存在“连线”;

<CoreThreshold>用于指定在关联分析中和关联分析中关联的阈值, 如果超过该值, 才认为进行关联计算, 在图形上表示为存在“连线”;

<OnlyEdgeCoauthor>用于指定在输出作者合著图形时是否只输出带有合著关系的节点;

<OnlyEdgeCoassignee>用于指定在输出机构合著图形时是否只输出带有合著关系的节点;

<OnlyEdgeCocountry>用于指定在输出区域合著图形时是否只输出带有合著关系的节点;

<OnlyEdgeCoclass1>F 用于指定在输出类别 1 同现图形时是否只输出带有同现关系的节点;

<OnlyEdgeCoclass2>用于指定在输出类别 2 同现图形时是否只输出带有同现关系的节点。

<OnlyEdgeCoword>用于指定在输出关键词同现图形时是否只输出带有同现关系的节点；

<OnlyEdgeDocumentReference>用于指定在输出文献引证图形时是否只输出带有引证关系的节点；

<OnlyEdgeAuthorReference>用于指定在输出作者引证同现图形时是否只输出带有引证关系的节点；

<OnlyEdgeAssigneeReference>用于指定在输出机构引证图形时是否只输出带有引证关系的节点；

<OnlyEdgePublicationReference>用于指定在输出期刊引证图形时是否只输出带有引证关系的节点；

<OnlyEdgeDocumentCoReference>用于指定在输出文献耦合图形时是否只输出带有耦合关系的节点；

<OnlyEdgeAuthorCoReference>用于指定在输出作者耦合图形时是否只输出带有耦合关系的节点；

<OnlyEdgeAssigneeCoReference>用于指定在输出机构耦合图形时是否只输出带有耦合关系的节点；

<OnlyEdgePublicationCoReference>用于指定在输出出版物耦合图形时是否只输出带有耦合关系的节点；

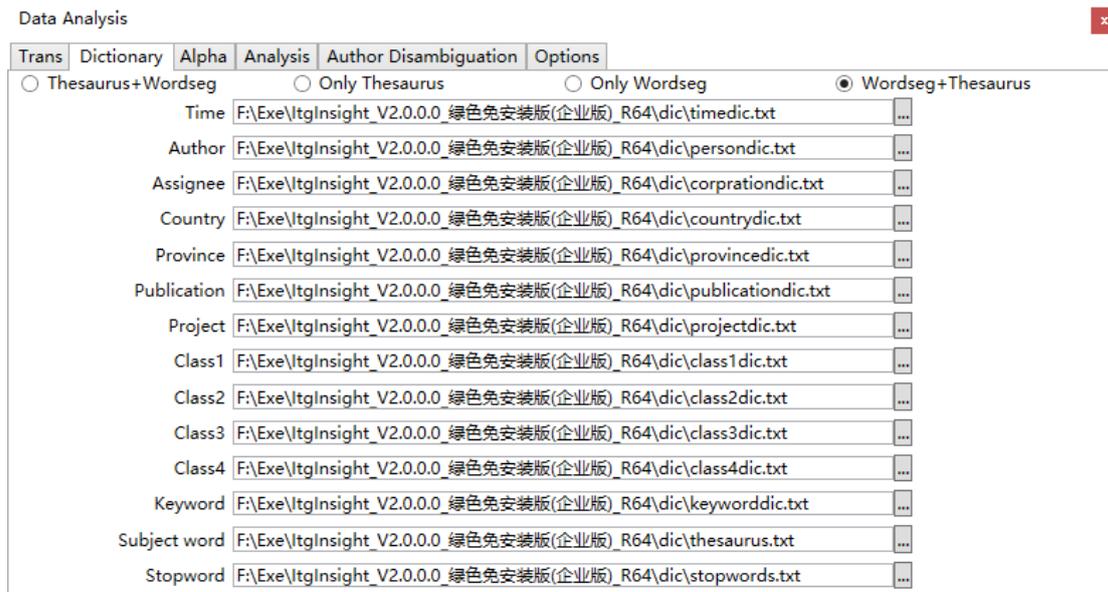
<OnlyEdgeCountryCoReference>用于指定在输出国家耦合图形时是否只输出带有耦合关系的节点；

<PFNET>用于设定是否由计算机完成网络图的压缩。

也可以通过“Options/选项”——>“System Setting/系统设置”进行有关参数的设置，如下图。或者在进行数据转换时指定以上参数的设置。

5.3 停用词设置

如果在与摘要词相关的分析中出现了没有意义的词，如“of、 in、 on”“的、地、得”等，可以通过设置停用词的方式去掉这些关键词（系统默认情况下已经去除了大部分停用词）。停用词的设置在系统安装目录下 dic/stopwords.txt 文件中，每个停用词单独一行。设置后重新启动程序，停用词表才会发挥作用。可通过“Options/选项”——>“System Setting/系统设置”——>“Dictionary/字典”指定停用词文件所在路径，如下图。



或者在进行数据转换时指定使用的停用词文件所在路径。

也可以在“字典/Dictionary”——>“停用词/Stop Words”中进行停用词的添加、删除，如下图。



停用词不区分大小写。

5.4 主题词设置

如果希望在摘要词的分析中，按照用户指定的词进行词的拆分，这里称为主题词拆分。

主题词的设置在系统安装目录下 dic/thesaurus.txt 文件中，设置格式为：

词 A1|词 A2|词 A3|词 A4|词 A5|词 A6|词 A7|词 A8——含义为用词 A1 替换词 A2-A8
|词 B1|词 B2|词 B3——含义为词 B1、B2、B3 为无意义词（类似于停用词）不在统计范围内

其设置后重新启动程序，主题词表才会发挥作用，并且对关键词的分析中同样发挥作用。也可以通过“Options/选项”——>“System Setting/系统设置”——>“Dictionary/字典”指定主题词文件所在路径。也可以在“字典/Dictionary”——>“主题词/Thesaurus”中进行主题词的添加、删除，具体方法和停用词的操作一致。对主题词的使用可设置为“应用主题词”、“仅用主题词”“仅用分词”来设置主题词的使用方式。三种方式中仅用主题词表示本软件的推荐使用方案，其优势在于处理的数据量会成倍的增加，而且主题词更有实际意义。主题词表不区分大小写。

5.5 人名辞典设置

为了规范人名书写格式差异带来的统计结果不准确，下图提供人名词典文件进行人名的修正，存在软件目录 dic/persondic.txt 中。对于该文件的修改于主题词一致。设置格式相似：

人名 A1|人名 A2|人名 A3——含义为用人名 A1 替换人名 A2-A3

|人名 B1|人名 B2|人名 B3——含义为人名 B1、B2、B3 为无意义人名不在统计范围内

+|A

+|B

+|C

+|A|B|C 只分析或筛选人名字段中含有 A 或 B 或 C 的数据，其他数据不分析

默认条件下，人名词典不区分大小写，第一行取值：0，不区分大小写，全部按小写处理、-1，不区分大小写，全部按照大写处理、1，区分大小写，按照原文大小写处理、2，第一个字母按照大写，其他按照小些，如果一个单词全部为大写，保留全部大写。

5.6 公司辞典设置

为了规范机构名称书写格式、名称修改、并购合并等带来的统计结果不准确，系统提供机构词典文件进行机构名称的修正，存在软件目录 dic/corprationdic.txt 中。对于该文件的修改于主题词、人名词典一致。设置格式相似：

机构 A1|机构 A2|机构 A3——含义为用机构 A1 替换机构 A2-A3

|机构 B1|机构 B2|机构 B3——含义为机构 B1、B2、B3 为无意义机构不在统计范围内

+|A

+|B

+|C

+|A|B|C 只分析或筛选人名字段中含有 A 或 B 或 C 的数据，其他数据不分析

默认条件下，公司词典不区分大小写，第一行取值：0，不区分大小写，全部按小写处理、-1，不区分大小写，全部按照大写处理、1，区分大小写，按照原文大小写处理、2，第一个字母按照大写，其他按照小些，如果一个单词全部为大写，保留全部大写。

5.7 国家名称辞典设置

为了规范国家名称书写格式、名称修改、并购合并等带来的统计结果不准确，系统提供国家词典文件进行地名名称的修正，存在软件目录 dic/countrydic.txt 中。对于该文件的修改与主题词、人名词典一致。设置格式相似：

地名 A1|地名 A2|地名 A3——含义为用地名 A1 替换地名 A2-A3

|地名 B1|地名 B2|地名 B3——含义为地名 B1、B2、B3 为无意义地名不在统计范围内

+|A

+|B

+|C

+|A|B|C 只分析或筛选人名字段中含有 A 或 B 或 C 的数据，其他数据不分析

默认条件下，地名词典不区分大小写，第一行取值：0，不区分大小写，全部按小写处理、-1，不区分大小写，全部按照大写处理、1，区分大小写，按照原文大小写处理、2，第一个字母按照大写，其他按照小些，如果一个单词全部为大写，保留全部大写。

5.8 省份辞典设置

为了规范省份名称书写格式、名称修改、并购合并等带来的统计结果不准确，系统提供省份词典文件进行地名名称的修正，存在软件目录 dic/province-chinese.txt 中。对于该文件的修改于主题词、人名词典一致。设置格式相似：

地名 A1|地名 A2|地名 A3——含义为用地名 A1 替换地名 A2-A3

|地名 B1|地名 B2|地名 B3——含义为地名 B1、B2、B3 为无意义地名不在统计范围内

+|A

+|B

+|C

+|A|B|C 只分析或筛选人名字段中含有 A 或 B 或 C 的数据，其他数据不分析

默认条件下，地名词典不区分大小写，第一行取值：0，不区分大小写，全部按小写处理、-1，不区分大小写，全部按照大写处理、1，区分大小写，按照原文大小写处理、2，第一个字母按照大写，其他按照小些，如果一个单词全部为大写，保留全部大写。

5.9 词典内容大小写区分设置

在对应的人名、机构、地名、类别、关键词典文本中，#开头行为注释行，第一个非注释行取值 0,1, -1,2 含义如下：

- 0，不区分大小写，全部按小写处理
- 1，不区分大小写，全部按照大写处理
- 1，区分大小写，按照原文大小写处理
- 2，第一个字母按照大写，其他按照小些，如果一个单词全部为大写，保留全部大写

5.10 在字典中应用正则表达式进行高级筛选和替换

在人名、机构名、国家、省份、类别、项目等字典中可以使用正则表达式进行数据的高级替换和筛选，以 CNKI 论文为例，CNKI 数据中机构名不规范经常会有邮政编码，省份、城市作为机构名称，例如：“有机复合温拌剂对沥青性能影响研究”这篇论文，机构信息如下：

王仕峰; 阎国杰; 郑国民; 李亮; 夏庆华; 滕振华
 上海交通大学; 上海浦东路桥建设股份有限公司; 上海交通大学; 上海浦东路桥建设股份有限公司 上海市201210; 上海浦东路桥建设股份有限公司 上海市201210; 上海浦东路桥建设股份有限公司 上海市201210

未在数据清洗前不能穷尽这些信息，导致数据分析不准确，处理方式一：数据清洗或分析后发现异常数据，用字典逐个进行替换或删除，重新进行数据清洗或者分析；处理方式二：在数据未进行清洗或分析前，在机构词典中使用正则表达式进行处理，如下图为机构词典。

```

1 #本行为注释行，当下一行为1时，词表和原始文献的处理严格区分大小写，按照小写处理；
2 #为-1时按照大写处理；
3 #为0时不区分大小写；
4 #为2时，对每一个单次按照题目对待，也就是首字母大写，其他小写，但是全是大写的单词仍然为大写
5 #建议取2。
6 #####
7 #如果想只分析某几个机构，其他机构不分析，比如只分析A、B、C三家公司，在词表中加入三行，+|A +|B +|C
8 #此时无论是数据清洗还是数据分析，哪些不含有A、B、C三家的数据会被忽略掉
9 #这么操作的好处是减少读取的数据量
10 #####
11 #字典格式必须为UTF8格式，否则字典不起作用。
12 #####
13 2
14 r|[0-9]{6}
15 r|\\s{3}$
16 r|\\s{2}$
17 r|\\s{2}\\s
18 r|^\\s{2}\\s
19 r|^\\s{2}$|\\s{2}$
20 |本刊编辑部
21 |本刊通讯员
22 |本刊讯
23 |无
24 |不详
25 |IT字
  
```

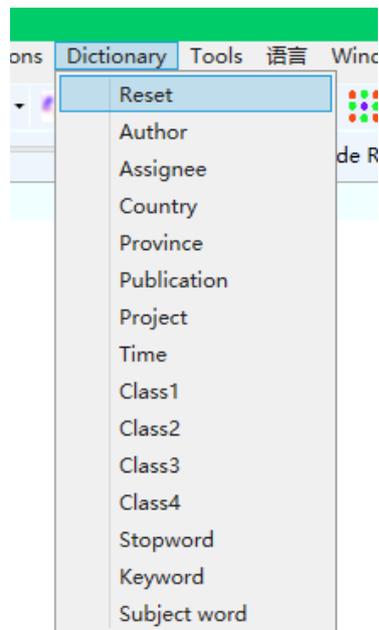
上图中以 r 开头的表示这一行是正则表达式替换，基本格式为“r|替换后的新的字符|正则表达式查找需要替换的字符”。替换后的新的字符=“”，替换变为删除。

例：r|[0-9]{6}表示将 6 位数字替换为空，也就是删除六位数字，六位的邮政编码 r|\\s{3}\$、r|\\s{2}\$两个分别表示替换长度为 3、2 的机构名称为空，也就是删除长度 3、

2 的机构名称，以此类推，可以用正则表达式进行过滤高级删除和替换。正则表达式需要对相关规则有所了解。

5.11 第一次使用软件时如何设置字典

第一次使用软件时，首先启动软件，按下图进行字典初始化后关闭软件，再重新启动，即可快速完成字典的设置。

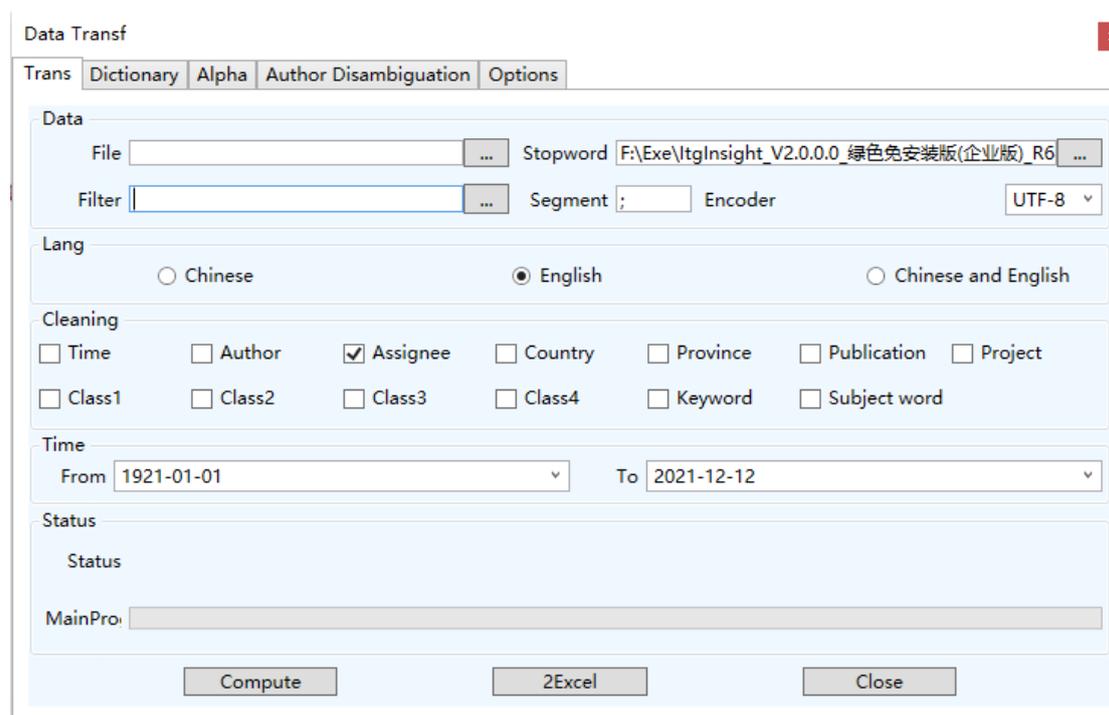


第六章：数据清洗

ITGinsight 提供数据清洗的功能,通过对数据进行清洗,生成词典,把生成的词典用于数据分析和可视化中。ITGinsight 对数据的清洗，并不改变原始数据，只是通过生成词典对分析过程进行人工干预。

6.1 选择待洗数据源

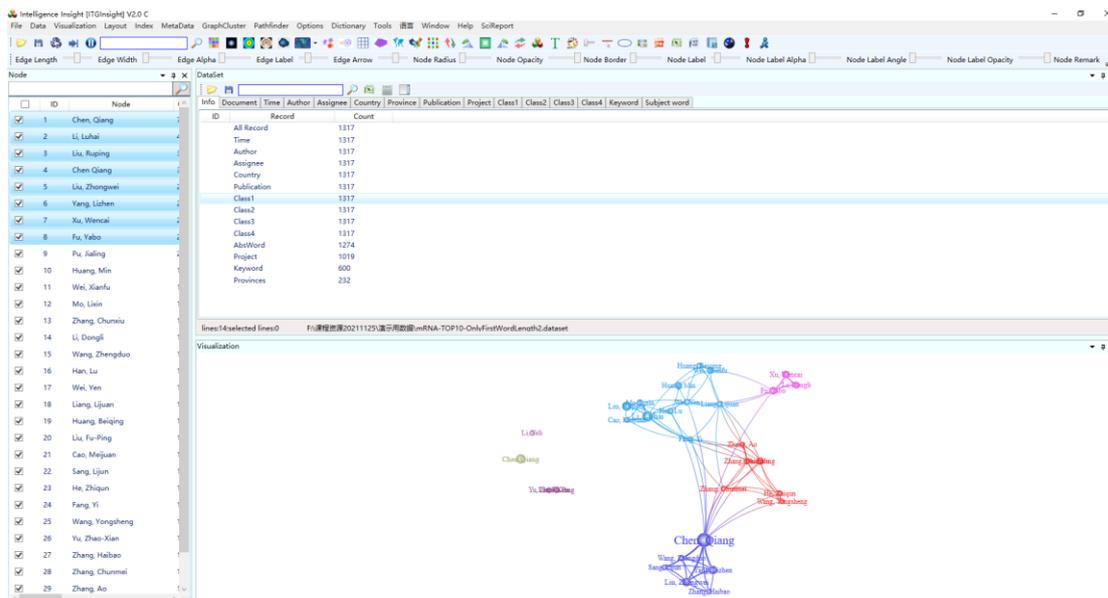
点击工具栏“Data/数据”->“清洗/Cleaning”，进入数据清洗页面，如下图。



选择要进行数据清洗的内容，操作与 3.1 数据格式转换一致。该步骤也可以选择词典，但不建议使用。

6.2 数据一览

清洗后，软件主页面会显示数据清洗读取的基本情况，数据基本情况在“信息/Info”标签下，显示，如下图。



该页面显示了共有多少条记录，其中包括时间的记录有多少条，包括作者信息的记录有多少条，包括机构、国家、省份、出版物、类别 1、类别 2、关键词、摘要词的记录都分别是多少。

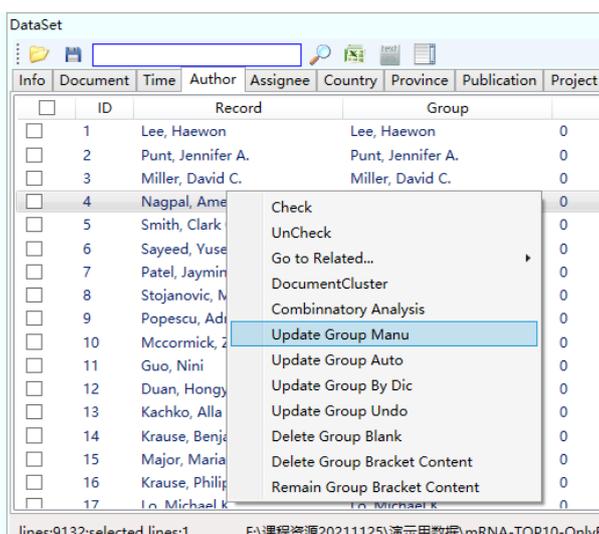
文本框支持逻辑查询，比如查询不包含 A 的结果，\$not A

查询同时包含 A BC 的结果 A \$and B \$and C

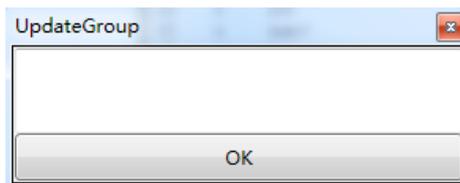
查询包含任意 A BC 的结果 A \$or B \$or C

6.3 数据人工分组实现数据清洗功能

在人名、机构、国家、省份、类别、关键词、摘要词等标签下，通过鼠标左键+shift 选择要进行数据清洗的记录，然后鼠标右键点击“Update Group Manu”，如下图。

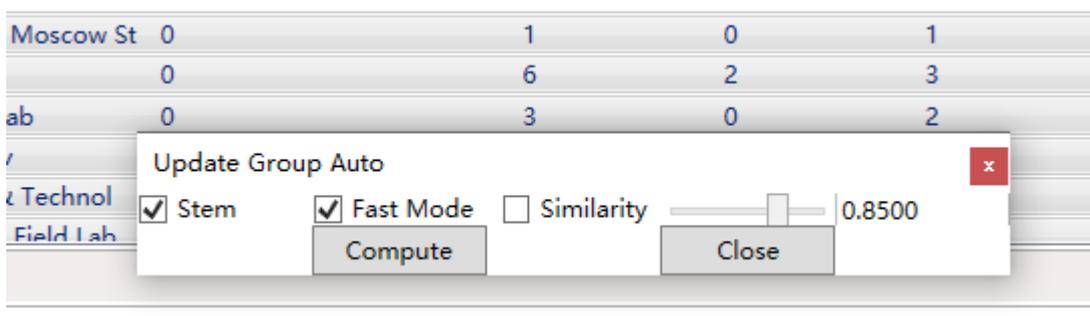


弹出修改分组窗口，如下图，输出分组名称，即可。



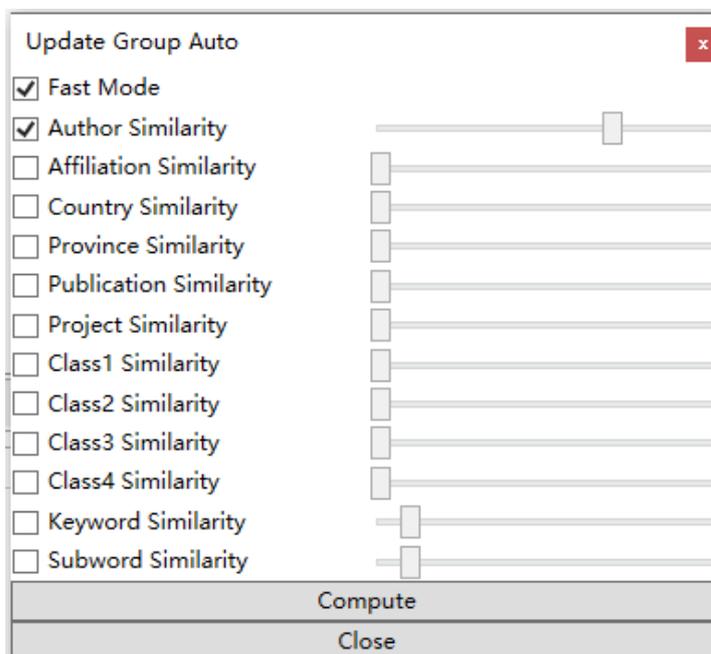
6.4 数据自动分组实现数据清洗功能

在作者、机构、国家、省份、类别、关键词、摘要词等标签下，通过鼠标左键+shift 选择要进行数据清洗的记录，然后鼠标右键点击“Update Group Auto”。弹出自动分组窗口，如下图。



Stemming 是说针对数据进行词性还原，用于英文。Similarity 是否按照相似度进行数据的合并，如果是相似度大小多少，用滑块设置相似度大小。Fast Mode 快速模式的含义与作者清洗的快速模式相同。

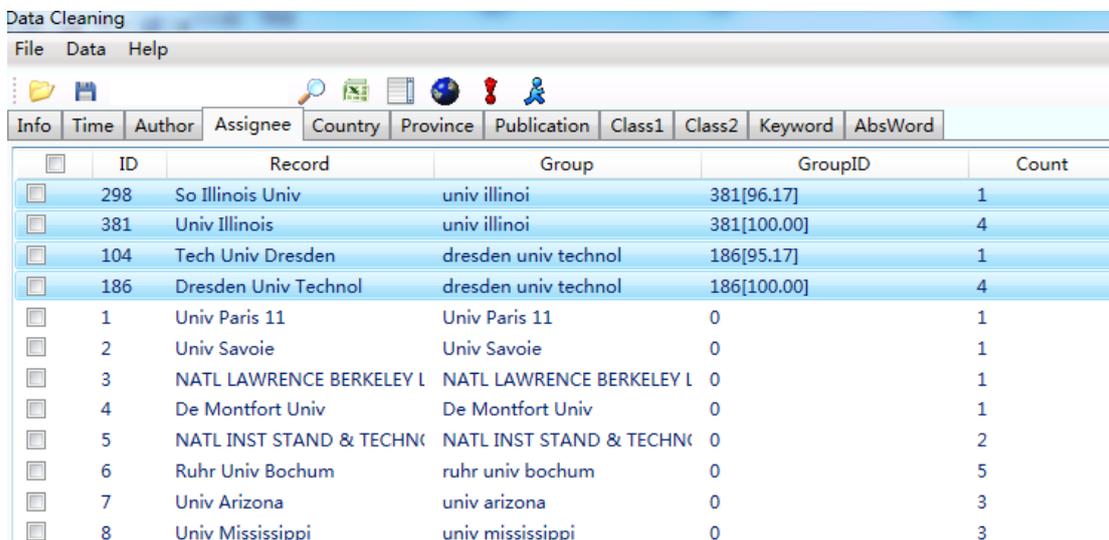
如果是人名标签，弹出的自动分组对话框略有不同，如下：



该窗口的意思是如何区分作者是否为相同作者或不同作者，依据作者名称相似度、所属机构相似度、所属国家相似度等等，该功能对于识别同名异指和异名同指具有重要的实用价

值。其中 **Fast Mode** 快速模式的意思是说：在计算相似度时，根据人名的首字母判断，如果前 N 个首字母相同，才计算相似度，这样计算相似度的情况会少很多，计算速度明显提升，建议采用快速模式。

之后点击关闭，系统自动计算记录的相似性，计算后的结果如下形式：



ID	Record	Group	GroupID	Count
298	So Illinois Univ	univ illinoi	381[96.17]	1
381	Univ Illinois	univ illinoi	381[100.00]	4
104	Tech Univ Dresden	dresden univ technol	186[95.17]	1
186	Dresden Univ Technol	dresden univ technol	186[100.00]	4
1	Univ Paris 11	Univ Paris 11	0	1
2	Univ Savoie	Univ Savoie	0	1
3	NATL LAWRENCE BERKELEY L	NATL LAWRENCE BERKELEY L	0	1
4	De Montfort Univ	De Montfort Univ	0	1
5	NATL INST STAND & TECHN	NATL INST STAND & TECHN	0	2
6	Ruhr Univ Bochum	ruhr univ bochum	0	5
7	Univ Arizona	univ arizona	0	3
8	Univ Mississippi	univ mississippi	0	3

按照 **GroupID** 进行排序，如上图记录 381 和 298 为一组，其 **GroupID** 相同，中括号里的数字为相似度，分组之后的数据存储为词典后，认为是一个记录，在之后的分析中作为用户词典。如果认为计算机分组有误，可点击右键选择取消分组，分组恢复原始状态。

6.5 保存字典

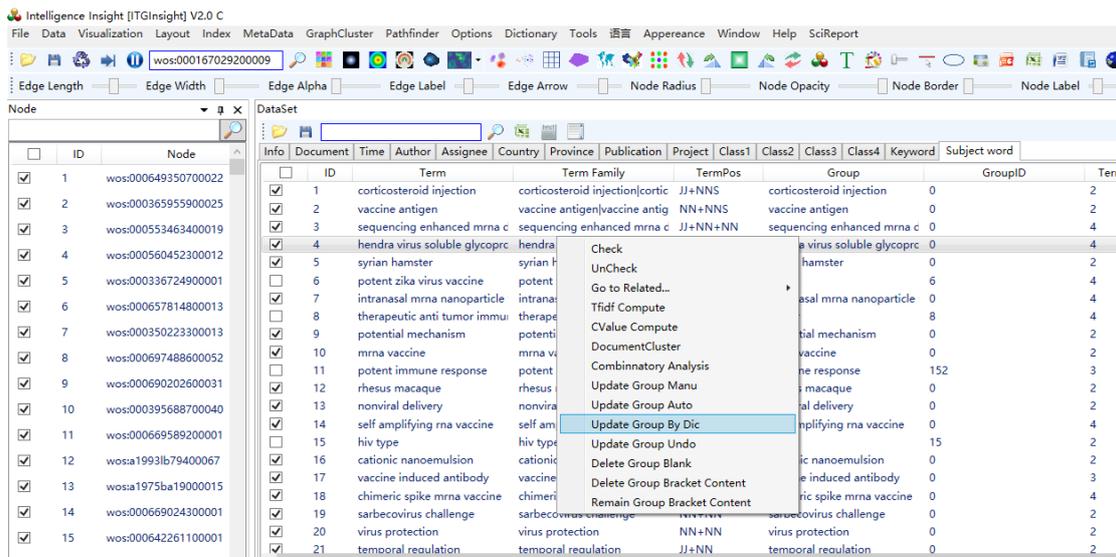
在进行数据分析时，分组后的记录统一被新的分组代替。为此，需要将记录保存为字典。

通过鼠标左键+shift 选择要保存的记录，右键点击 **check**，最后点击工具栏的  按钮，将结果输出为 txt 格式的字典。也可以将记录保存为 excel 文件，点击工具栏的  按钮即可。

6.6 用词典再次清洗数据、数据分析，自动分组。

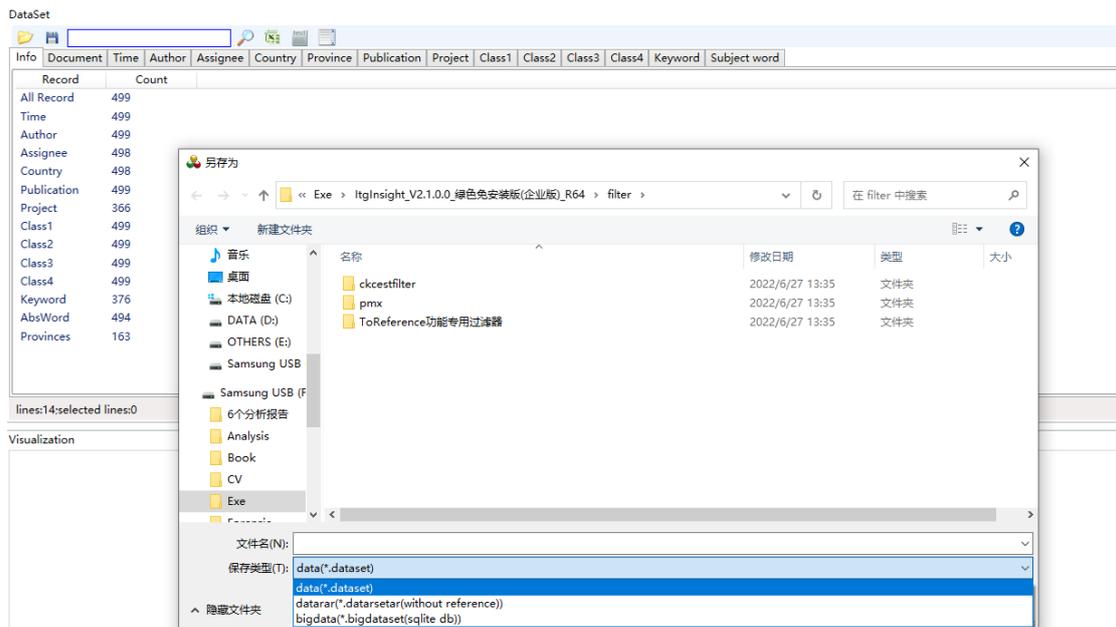
保存后的字典可以在以后的数据清洗和数据分析中使用，只需要将其设定到对一个的字典目录即可。

如果想用字典进行分组，按如下截图进行操作，可以重复利用字典，避免重复工作。



6.7 保存清洗结果

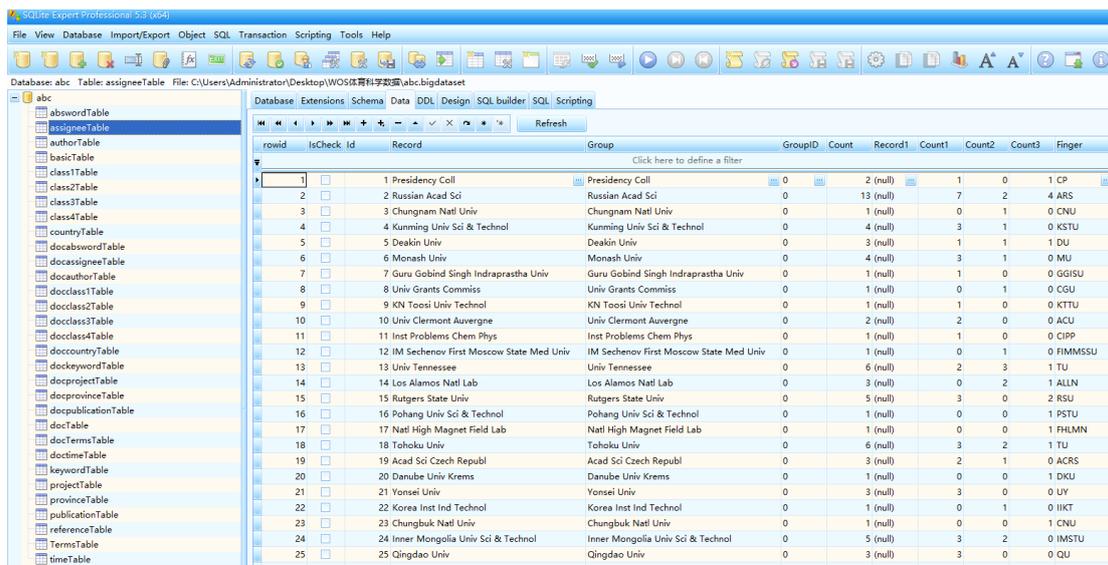
清洗后的结果，点击工具栏的保存按钮将结果保存，下次使用时通过工具栏的打开按钮打开即可，如下截图：



dataset 格式，为默认的正常保存格式，适用于数据量不超过 3 万条的情况；

datasetrar 格式，为可选的保存格式，保存过程中，去掉了引用信息，适用于数据量超过 3 万条的情况；

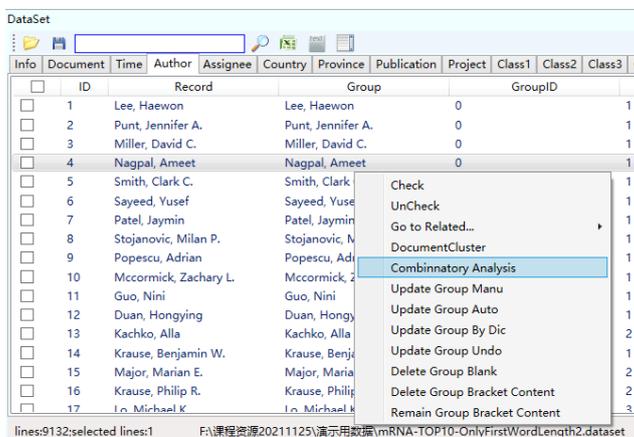
bigdataset 格式，适用大数据量，一般超过 5 万条建议保存为 bigdataset，超过 10 万条必须保存为 bigdataset； bigdataset 格式是 Sqlite 数据库文件，可以通过 Sqlite Expert 等第三方软件打开，如下图，该数据库文件支持各种自定义用户的扩展功能。



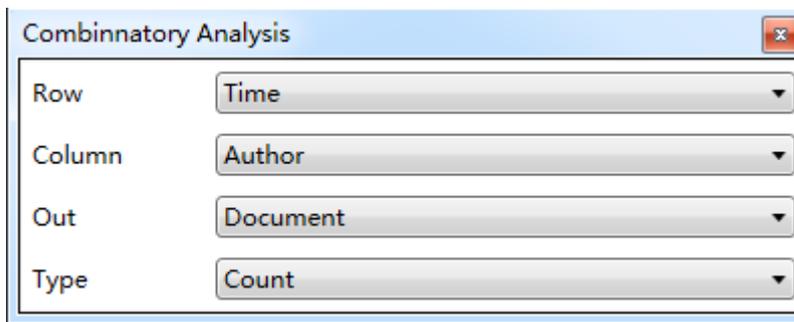
因为操作系统对于保存的数据大小由限制，而引文信息过大，导致无法保存，为此在数据清洗阶段，如果对于引文信息没有分析需求的情况下，选择保存 bigdatarar 格式。

6.8 组合分析（跨维度、跨层次的共现矩阵、引证矩阵）

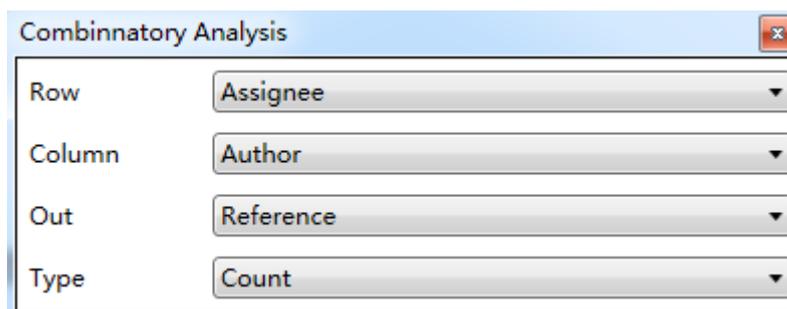
在时间、作者、机构、国家、省份、类别、关键词、摘要词中任意选择 3 个维度中的若干数据，勾选数据。后右键点击 **Combinatory Analysis**，如下图。



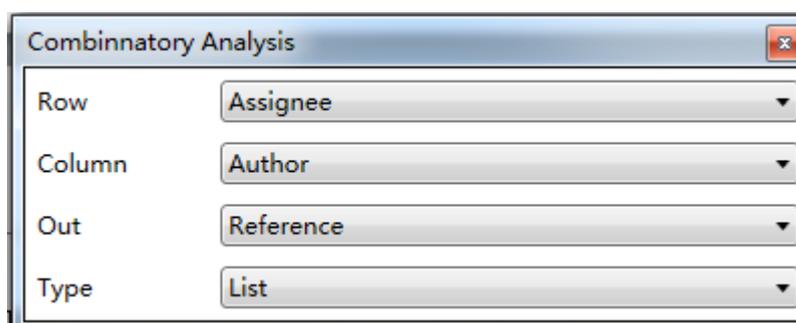
弹出组合分析对话框，选择要分析的行，列和统计的标准，类型，如下图：



上图的含义是统计时间、作者两个维度交叉后的文献数量。



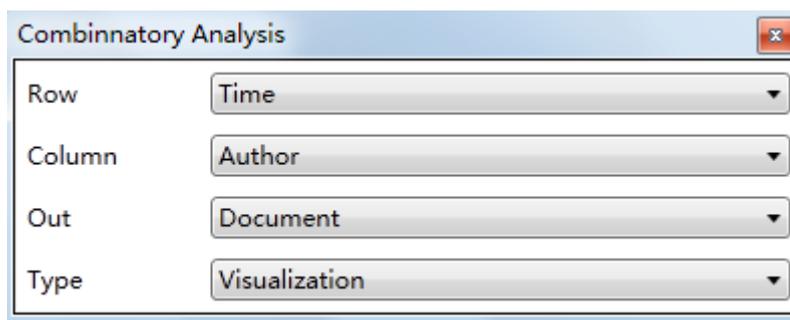
上图的含义是统计机构、作者两个维度交叉后的参考文献数量，也就是某个作者引用了某个机构的文献数量。



上图的含义是统计机构、作者两个维度交叉后的参考文献列表，也就是某个作者引用了某个机构的文献列表（文献的 ID 号）。

选择之后，关闭改窗口，系统调用 Excel 输出统计矩阵，如下格式：

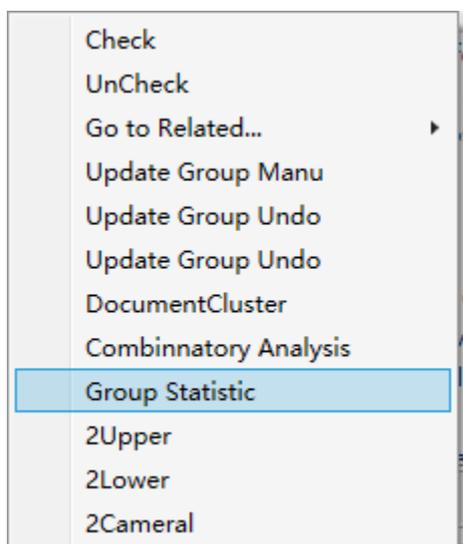
	Adalsteir	Calvert,	Chelnokov	Christens	Coutaz,	JDuvillare	Eufinger,	Gaitan,	MKajfez,	IKramer,	GL
Adalsteir	1										
Calvert, P		1									
Chelnokov, A			5		1	1					
Christensen, RG				1				1			1
Coutaz, JL			1		1	1					
Duvillaret, L			1		1	1					
Eufinger, H							3				
Gaitan, M				1				1			1
Kajfez, D									1		
Kramer, GW				1				1			1
Lees, AEU											
Lin, TL		1									
Locascio, LE				1				1			1
Lourtioz, JM			5		1	1					
MacCrehan, WA				1				1			1
Machtens, E								1			
Martin, H		1									
Martynova, L				1				1			1
Rowson, S			3		1	1					
Sethian,	1										
Wehmoller, M								3			
Wu, LH											



上图的含义是时间、作者两个维度交叉后的文献数量后，直接在软件的图形区域输出可视化图形。

6.9 分组统计（数据清洗后，按新分组进行统计）

组合分析是对数据进行二维、三维统计，如果在数据分组后，按照分组进行一维统计，勾选要统计的选项，右键点击分组统计/Group Stat 进行一维的分组统计，如下图：



无论是分组统计还是组合分析，都是按照新的分组进行计算。

6.10 文档聚类

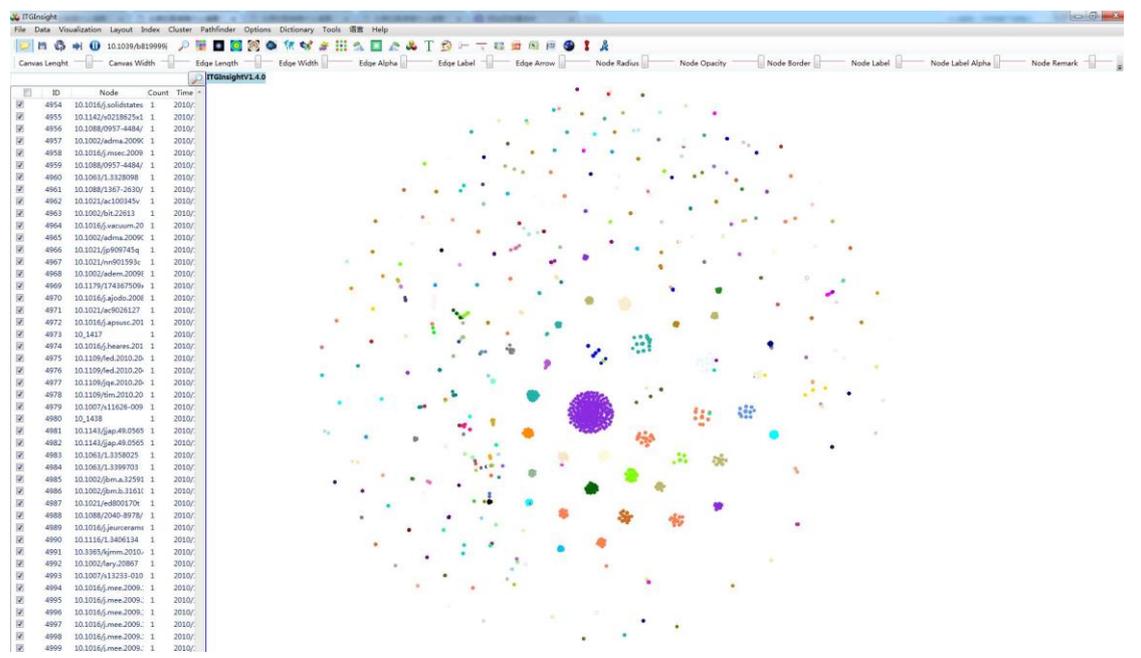
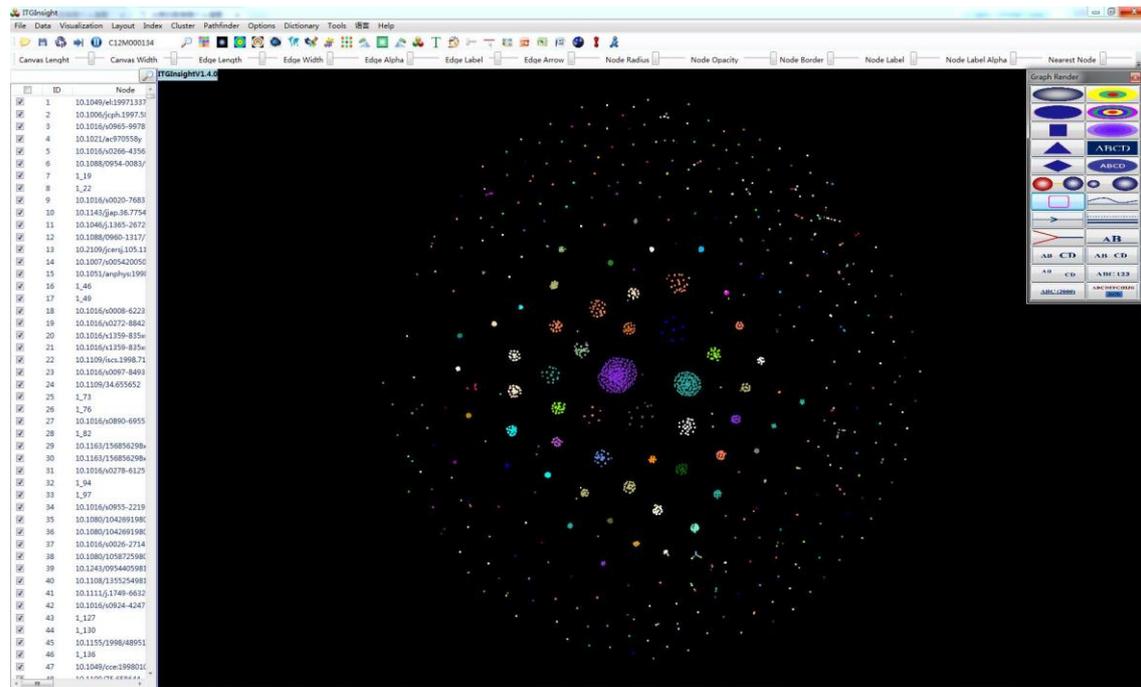
在数据清洗界面点击鼠标右键选择 cluster/聚类，如下图：

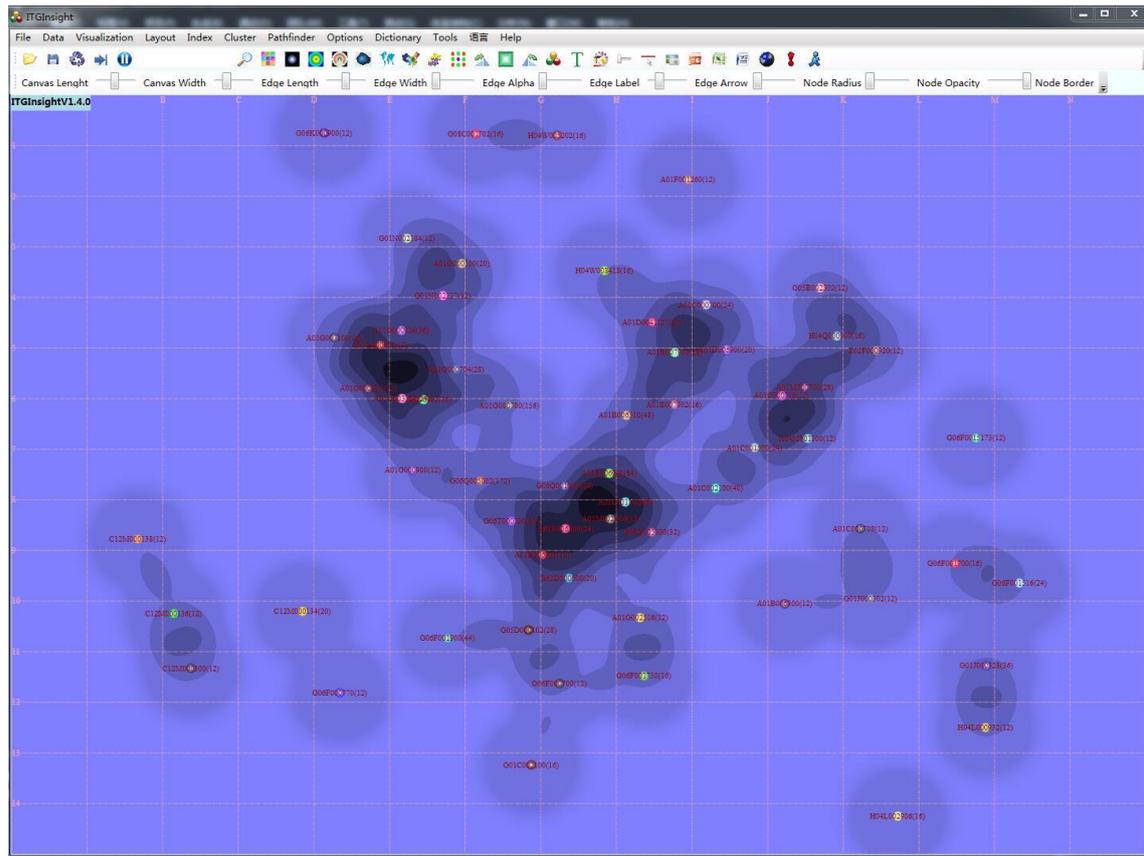
Info	Document	Time	Author	Assignee	Country	Province	Publication	Project	Class1	Class2	Class3	Class4	K	
ID	Record	Group	GroupID	Count										
<input type="checkbox"/>	1	Lee, Haewon	Lee, Haewon	0										1
<input type="checkbox"/>	2	Punt, Jennifer A.	Punt, Jennifer A.	0										1
<input type="checkbox"/>	3	Miller, David C.	Miller, David C.	0										1
<input type="checkbox"/>	4	Nagpal, Ameet	N	Check										1
<input type="checkbox"/>	5	Smith, Clark C.	S	UnCheck										1
<input type="checkbox"/>	6	Sayeed, Yusef	S	Go to Related...										1
<input type="checkbox"/>	7	Patel, Jaymin	P	DocumentCluster										1
<input type="checkbox"/>	8	Stojanovic, Milan P.	S	Combinatory Analysis										1
<input type="checkbox"/>	9	Popescu, Adrian	P	Update Group Manu										1
<input type="checkbox"/>	10	Mccormick, Zachary L.	N	Update Group Auto										1
<input type="checkbox"/>	11	Guo, Nini	G	Update Group By Dic										1
<input type="checkbox"/>	12	Duan, Hongying	D	Update Group Undo										2
<input type="checkbox"/>	13	Kachko, Alla	K	Delete Group Blank										1
<input type="checkbox"/>	14	Krause, Benjamin W.	K	Delete Group Bracket Content										2
<input type="checkbox"/>	15	Major, Marian E.	N	Remain Group Bracket Content										2
<input type="checkbox"/>	16	Krause, Philip R.	K											2
<input type="checkbox"/>	17	Lo, Michael K.	L											3

其中 cluster 为聚类对象，默认为对文档进行聚类。by 为聚类采用的字段为，默认为关键字，可选择其他字段。how many 为采用多少关键词进行文档关系的计算，这个数据越大，聚类时间越长。Algorithm 未降维算法，默认 TSNE，效果最好，但是最慢，PCA 最快，效果最差，KPCA 介于二者之间。out 为输出内容，默认为输出网络可视化图，也可以选择密度图。color 制定聚类后的网络节点颜色区分，是按照那个字段进行区分，比如按照文档所属机构或者所属国家进行颜色区分，默认为不区分颜色。注意：无论是哪个字段，必须勾选才起作用，不勾选的文档不聚类，不勾选的关键词字段或其它字段在聚类分析过程中不起作用。

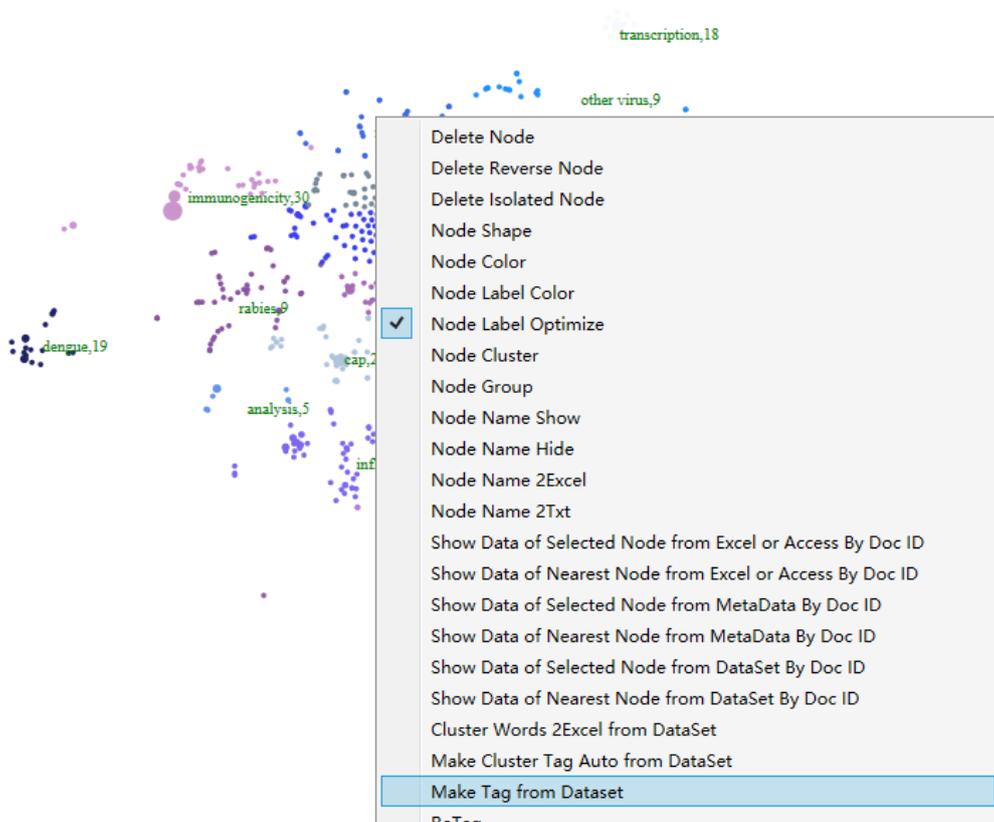
Cluster	Document
By	Keyword
How many	50
Algorithm	TSNE
Out	Visualization
Color	No
<input type="button" value="OK"/> <input type="button" value="Close"/>	

聚类计算完成后，在主窗体显示聚类结果的网络图，形式如下：

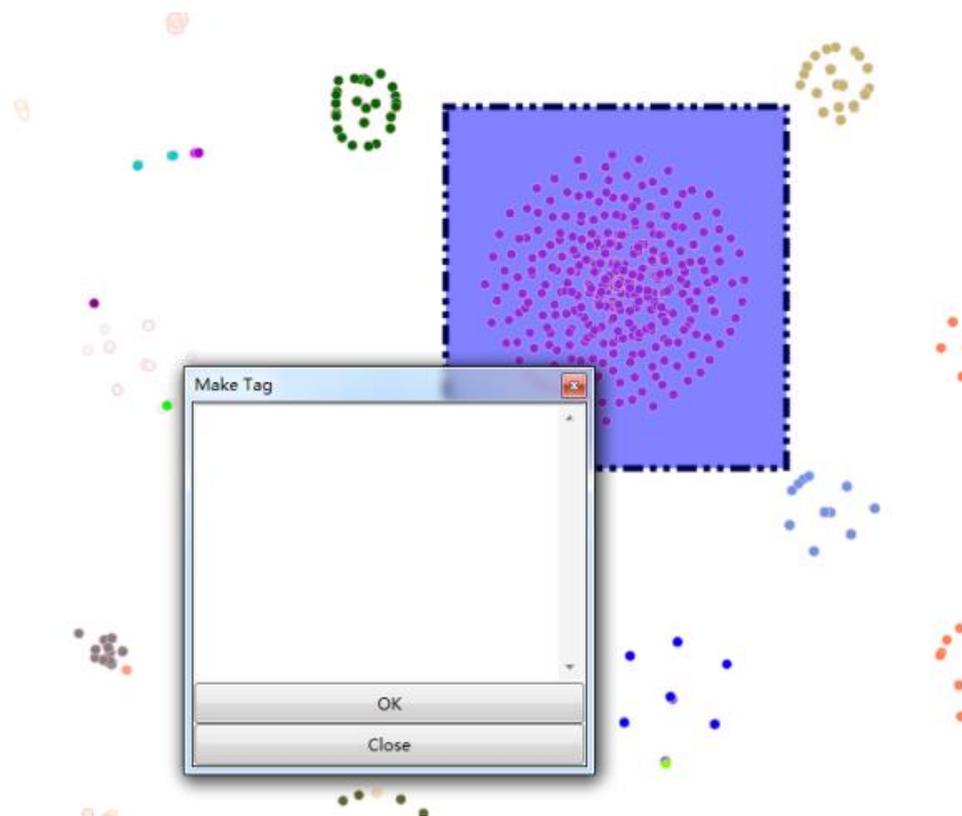




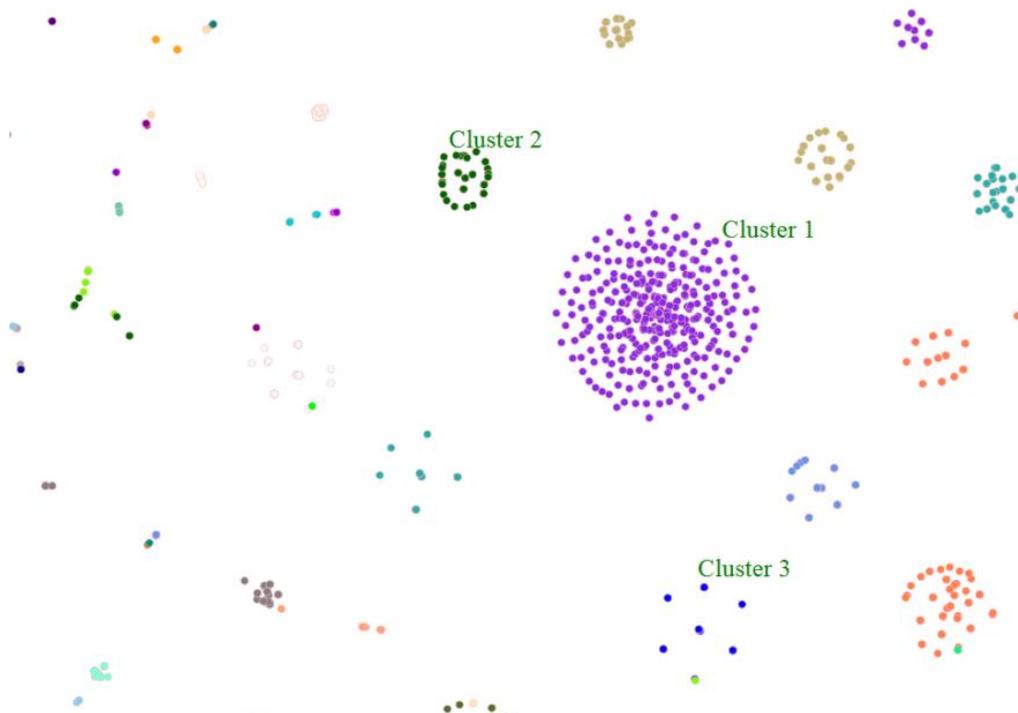
进一步对聚类图聚类结果打标签，点击鼠标右键，如下图：



弹出输入标签框：如下图：



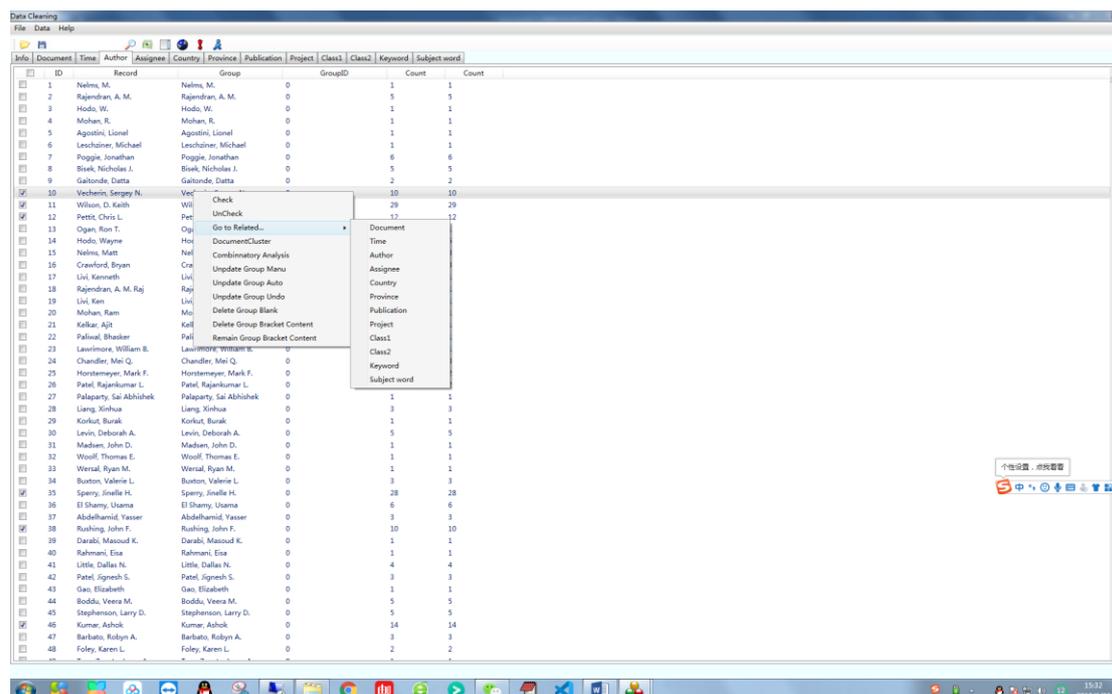
可手工输入标签，如对聚类进行命名，标注每个类别的主题词。同时，如果数据清洗界面如果打开了聚类图对应的 dataset，软件会推荐类别标签内容。通过鼠标右键的 Tag Color 可更改标签颜色，通过滑块设置 tag 滑块，更改标签大小。打过标签的聚类图如下：



手工打标签比较繁琐，可以上图右键弹出菜单的 Make Cluster Tag Auto from Dataset 功能，为每个聚类结果自动打标签。

6.11 数据链

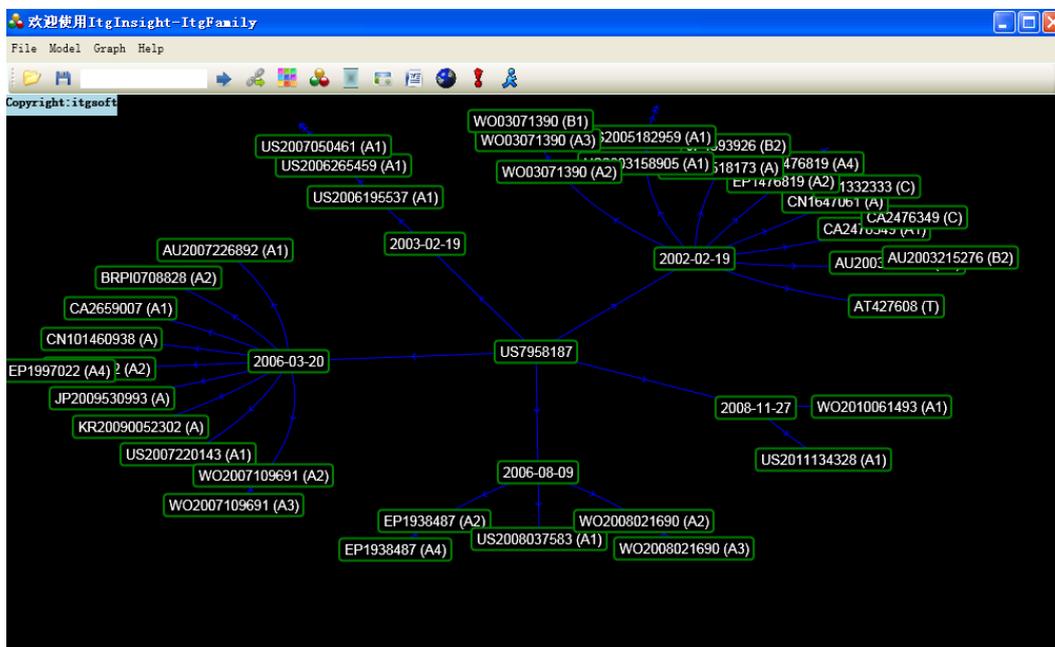
在数据一览，任何一个维度，选中几条数据，点击右键，如下图，到导航到对应的数据。如选中任何一个作者，点击 go to related, publication, 则会导航这几个作者发表论文对应的期刊，以此类推，可进行任意维度数据的链接。



第七章：辅助软件工具

7.1 INPADOC 同族专利可视化分析工具

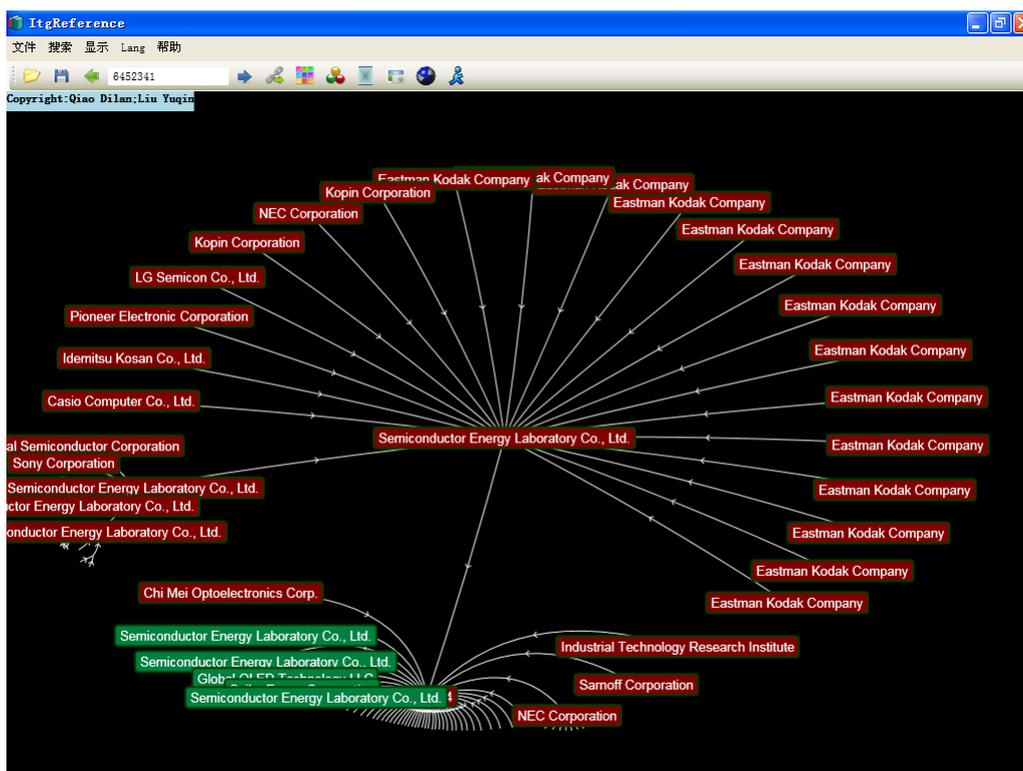
点击“Tools/工具”——>“ItgFamily/INPADOC 同族分析”调取世界专利同族实时更新与分析系统 ItgFamily，如下图：



ItgFamily 世界专利同族实时更新与分析系统是一款“企业专利战略决策”支持工具，该系统主要实现世界专利的同族数据采集以及树形呈现。其最大的特点是对于专利在不同时间，不同国家进行的同族申请进行实时更新，并以树形结构，按时间顺序排列同族专利，在屏幕上显示的专利个数没有限制。系统采用鼠标双击和拖动进行用户交互，分析图形简洁优美，结构清晰。关于该工具的使用查看其对应的帮助文件 itgfamily 帮助.pdf。

7.2 美国专利引证可视化分析工具

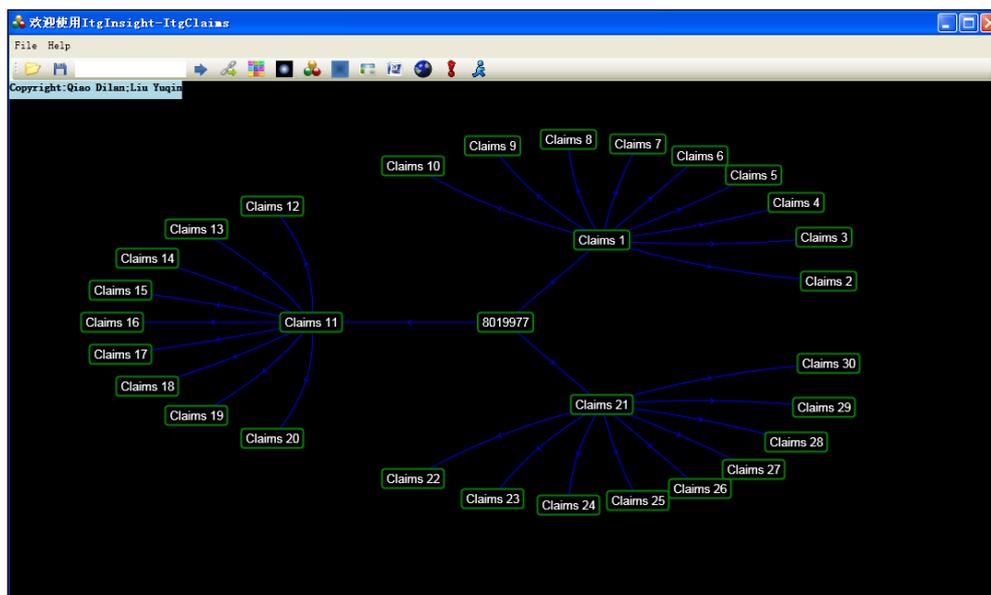
点击“Tools/工具”——>“ItgReference/美国授权专利引证分析”调取美国专利引证信息实时更新与分析系统 ItgReference，如下图：



ItgReference 美国专利引证实时更新与分析系统是一款“企业专利战略决策”支持工具，该系统主要实现美国单件专利的前引、后引数据采集以及树形呈现。其最大的特点是前引、后引的专利引证层级没有限制，在屏幕上显示的专利个数没有限制。系统采用鼠标双击和拖动进行用户交互，分析图形简洁优美，结构清晰，可以与 Aureka 的专利引证图谱相媲美。关于该工具的使用查看其对应的帮助文件 itgreference 帮助.pdf。

7.3 美国专利权利要求解析工具

点击“Tools/工具”——>“ItgClaims/美国专利权利要求分析”调取美国专利权利要求系统 ItgClaims，如下图：



ItgClaims 美国专利权利要求解析系统,作为通用科技文本可视化挖掘系统 ITGinsight 的子系统,主要实现美国单件专利权利要求中的独立权利要求和从属权利要求在线搜索、提取、结构解析与可视化呈现。其最大的特点是自动生成权利要求结构树,并以可视化的形式显示,在屏幕上显示的权利要求个数没有限制。系统采用鼠标双击和拖动进行用户交互,解析图形简洁优美,结构清晰。关于该工具的使用查看其对应的帮助文件 [itgclaims 帮助.pdf](#)。

第八章：自定义结构化数据可视化

ITGinsight 系统支持对 4 种类型的自定义格式数据进行可视化，分别为.ima, .imb, .imc 和 excel 格式。

8.1 ima 格式数据

第一类：ima 格式数据参考软件安装目录 example\ima 下的文件，格式如下：

```

      c1 c2 c3 c4 c5
r1 1 2 3 4 5
r2 8 9 10 11 12
r3 15 16 17 18 19
r4 15 16 17 18 19
r5 15 16 17 18 19

```

ima 格式数据是最简单的矩阵数据，数据以单个空格分隔，ima 格式数据仅用于数据演示，方便用户理解不同 layout 算法的区别，一般不在实际的分析项目中进行应用。

8.2 imb 格式数据

第二类：imb 格式数据参考安装目录 example\imb 下的文件，格式如下：

```

Node:ID|NAME|Number|Turnnumber|Referencednumber|Famillynumber|Cluster|Remark1|Remark2|Remark3|Remark4
Node:1|Cn|443|443;0;0;0|0|0|1|||||
Node:2|Us|61|61;0;0;0|0|0|1|||||
Node:3|Jp|11|11;0;0;0|0|0|1|||||
Node:4|De|9|9;0;0;0|0|0|1|||||
Node:5|Nl|8|8;0;0;0|0|0|1|||||
Node:6|Ch|4|4;0;0;0|0|0|1|||||
Node:7|Sg|2|2;0;0;0|0|0|1|||||
Node:8|Gb|2|2;0;0;0|0|0|1|||||
Node:9|Dk|2|2;0;0;0|0|0|1|||||
Node:10|Au|1|1;0;0;0|0|0|1|||||
Node:11|Cl|1|1;0;0;0|0|0|1|||||
Node:12|Kr|1|1;0;0;0|0|0|1|||||
Node:13|Fr|1|1;0;0;0|0|0|1|||||
Edge:nodeid,nodeid,edge
Edge:2,1,443
Edge:3,1,61
Edge:4,1,11
Edge:5,1,9
Edge:6,1,8
Edge:7,1,4
Edge:8,1,2
Edge:9,1,2
Edge:10,1,2
Edge:11,1,1
Edge:12,1,1
Edge:13,1,1

```

imb 格式数据可以定义节点和连线，进行简单的可视化展示。

Node:ID|NAME|Number|Turnnumber|Referencednumber|Famillynumber|Cluster|Remark1|Remark2|Remark3|Remark4

用以定义节点，其中 ID 为节点标示，采用 1、2、3...加以区分，不能重复；

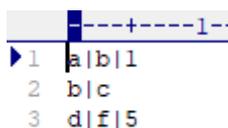
Name 为节点名称，会在可视化图形中进行显示；

Number 为节点代表的数量；

TurnNumber 为节点表示的第一、二、三作者数字，其和必须与 Number 相等；

Referencenumber 为节点的引证和被印数量，两个数字，都可以为 0；

第四类：imd 格式数据文件也是较为灵活的格式，如下：



读取后会构建一个由 a b c d f 五个节点，a 到 b 连线为 1，d 到 f 连线为 5 的网络图。imc 和 imd 格式适用于那些带有单个方向的网络数据，比如专利转移转化，也可以用于带有双向指向的数据。如上图，还可以增加 f | d | 3，表示 f 到 d 连线为 3，与 d 到 f 连线为 5，一起构成双向图。

8.5 excel 格式数据（万能格式）

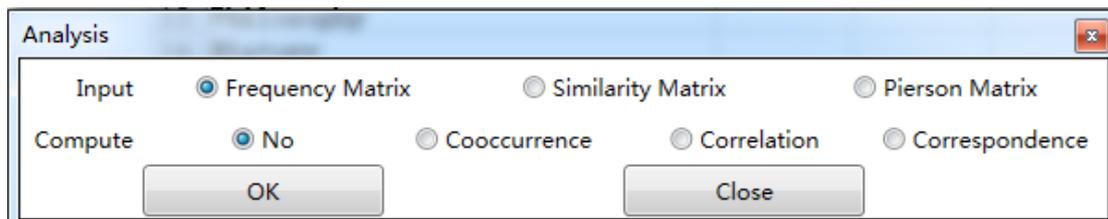
第五类：excel 格式的数据文件，参考安装目录 example\excel 中的文件，格式如下：

	A	B	C	D	E	F	G	H	I
1		practical in	information	Knowledge	USER ACC	SOCIAL MI	CITATION	research	lit social scier
2	Information Science & Library Sci	49	39	38	34	33	32	31	30
3	Computer Science	20	25	15	11	11	18	15	16
4	Business & Economics	11	10	21	7	1	1	8	
5	Communication	4			1	5			1
6	Geography								
7	Physical Geography								
8	Health Care Sciences & Services			1					
9	Medical Informatics			1					
10	Social Sciences - Other Topics				2	3			1
11	Telecommunications	1			1				
12	Government & Law			1					
13	Philosophy				1				
14	History								
15	Arts & Humanities - Other Topics								
16	Education & Educational Research								

或者如下：

	BRISTOL-M	MERCK SHA	GILEAD SC	ABBVIE IN	HOFFMANN	SCHERING	BOEHRINGE	VERTEX PF	GLAXOSMII	NOVARTIS	IDENIX PH	IST RICEF
BRISTOL-M	1	-0.038	-0.037	-0.037	-0.033	-0.033	-0.032	-0.031	-0.03	-0.029	-0.025	-0.025
MERCK SHA	-0.038	1	-0.028	-0.027	-0.025	0.287	-0.024	-0.023	-0.023	-0.022	0.032	0.213
GILEAD SC	-0.037	-0.028	1	-0.027	0.007	-0.024	-0.023	-0.022	-0.022	-0.021	-0.018	-0.018
ABBVIE IN	-0.037	-0.027	-0.027	1	-0.024	-0.024	-0.023	-0.022	-0.022	-0.021	-0.018	-0.018
HOFFMANN	-0.033	-0.025	0.007	-0.024	1	-0.021	-0.02	-0.02	-0.02	-0.019	-0.016	-0.016
SCHERING	-0.033	0.287	-0.024	-0.024	-0.021	1	-0.02	-0.02	-0.02	-0.019	-0.016	0.013
BOEHRINGE	-0.032	-0.024	-0.023	-0.023	-0.02	-0.02	1	-0.019	-0.019	-0.018	-0.016	-0.016
VERTEX PF	-0.031	-0.023	-0.022	-0.022	-0.02	-0.02	-0.019	1	-0.018	-0.018	-0.015	-0.015
GLAXOSMII	-0.03	-0.023	-0.022	-0.022	-0.02	-0.02	-0.019	-0.018	1	0.01	-0.015	-0.015
NOVARTIS	-0.029	-0.022	-0.021	-0.021	-0.019	-0.019	-0.018	-0.018	0.01	1	-0.015	-0.015
IDENIX PH	-0.025	0.032	-0.018	-0.018	-0.016	-0.016	-0.016	-0.015	-0.015	-0.015	1	-0.013
IST RICEF	-0.025	0.213	-0.018	-0.018	-0.016	0.013	-0.016	-0.015	-0.015	-0.015	-0.013	1

用户打开 excel 格式文件后，系统会询问用户打开的数据是频数矩阵、相似矩阵、皮尔森矩阵中的哪一种；是否依据输入矩阵进行矩阵运算以便得到同现矩阵、关联矩阵、对应矩阵，如下图：



如果选择 Similarity Matrix/相似矩阵、Pierson Matrix/皮尔森矩阵，系统默认不对矩阵

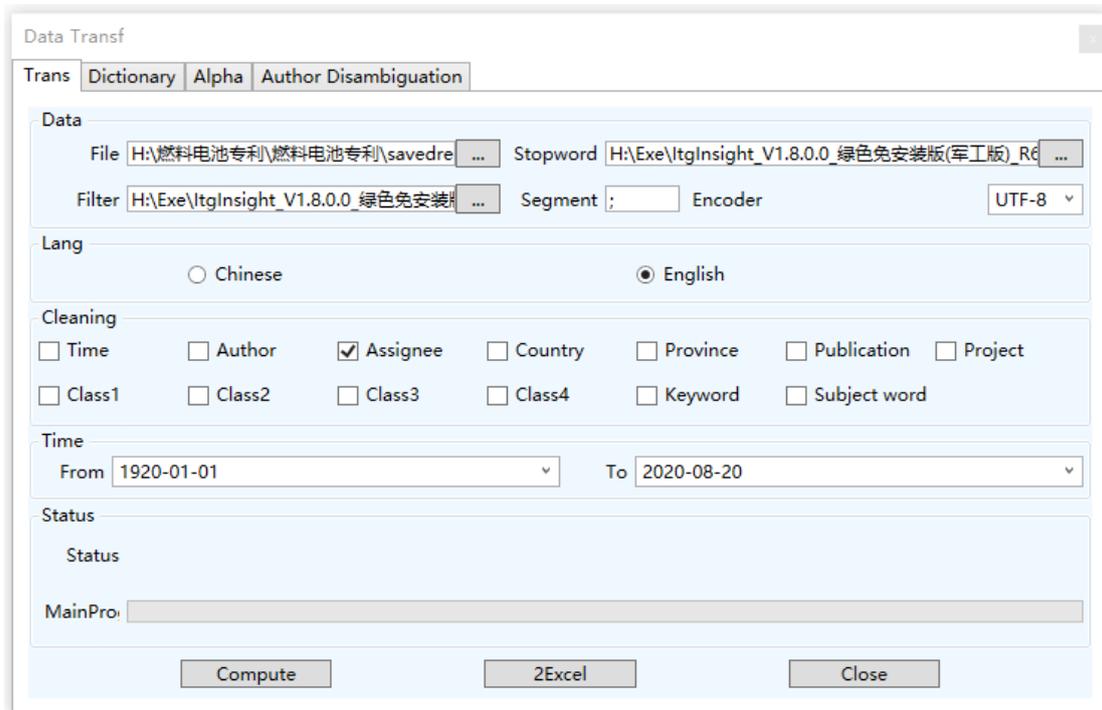
进行任何计算，直接将矩阵转化为网络图。

如果选择 **Frequency Matrix**/频数矩阵，不进行计算，如果频数矩阵行名和列名完全相同，输出网络图为 1 模网络图；如果不完全相同为 2 模网络图。

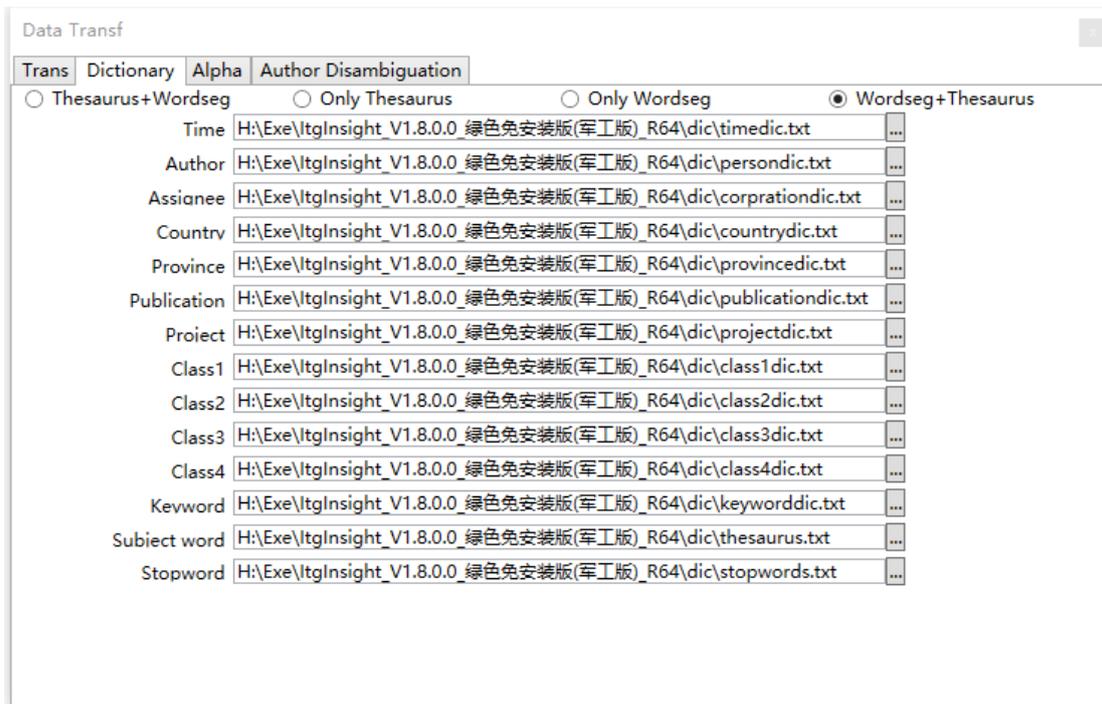
如果选择 **Frequency Matrix**/频数矩阵，按照剩余三种模式进行计算，系统默认对矩阵进行计算，分别为同现分析，对应分析，关联分析，对数据进行计算后进行可视化输出。

第九章：中英文科技术语识别（构建用户自定义主题词表）

在 V1.2 版本以后，ITGinsight 将中英文术语识别正式嵌入到软件中，具体操作在数据清洗功能中，按照 6.1 节选择数据源，勾选 Subject Word/主题词选项，如下图：



切换到 Dictionary/字典标签下，选择字典（一般选择安装目录 dic 目录下的默认字典），分词选项选择 Wordseg+Thesaurus/分词+词表，如下图：



切换到 Alpha/阈值标签下，如下图：

Data Transf

Trans Dictionary Alpha Author Disambiguation

CValue

TermLength(CN) 3-4

TermLength(EN) 2-4

TermFrequency 1

Numer1 0.000

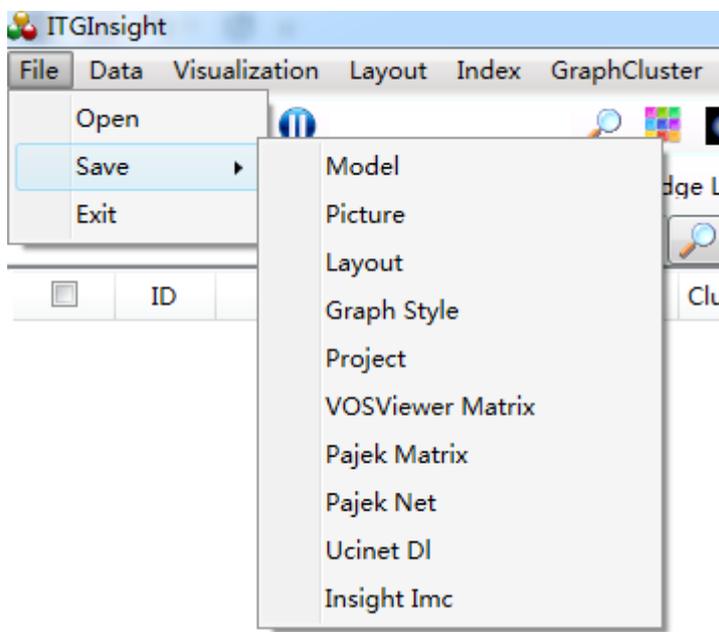
Numer2 0.000

Numer3 0.000

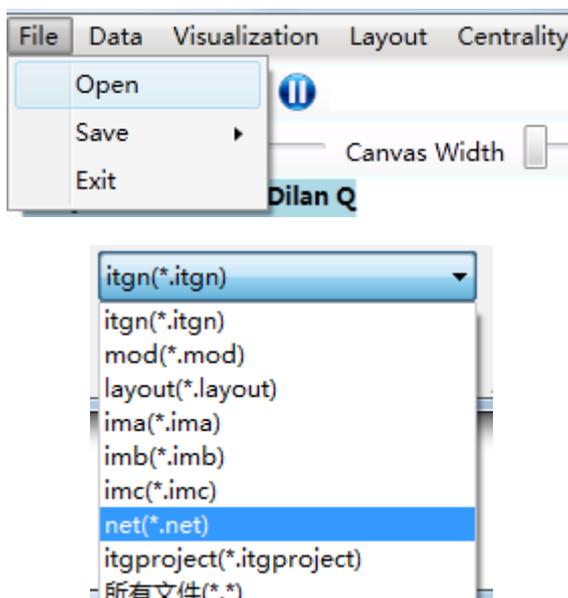
C-Value/是否进行术语度计算，术语度计算耗时较长但是可以提供“该次是否为术语”的参考，词长度和词频的选择依据用户偏好进行设置，如果数据较多，建议相关阈值设大些。主题词的合并可以采用 UpdateGroup_Auto（参见 6.4 数据自动分组实现数据清洗功能一节）进行。

第十章：与 VosViewer, Pajek, Ucinet 进行交互

系统支持将可视化图形导出为 VosViewer, Pajek, Ucinet 等可视化软件的输入格式文件，操作方法为点击：Sava/保存即可，如下图。保存后的文件可直接被 VosViewer, Pajek, Ucinet 使用。



同时，系统提供打开 Pajek 软件 net 格式的图形文件，如下图，点击打开，选择对应 net 文件即可。



本软件提供了 net 扩展格式 netx 格式的图形文件输入，netx 与 net 格式的差别是 netx 可以指定网络节点的大小，具体格式参考 example/netx/example.netx 文件。

第十一章：自动报告

ITGinsight 应用机器学习进行研究报告的自动化、智能化、模块化撰写，系统提供默认的报告撰写模板，也可以自定义报告模板。自动报告功能仅限于顶级企业级用户，由计算机进行报告的智能组织，用户仅做轻微的修改，具体操作如下。

打开 .itgn 文件，点击菜单栏的 word 图标，弹出如下界面：

Report Template				
Theme Analysis		Institutional Evaluation	Institutional Comparison	
id	name	alias	file	createtime
0	自动绘制所有图形	自动绘制所有图形	\\report template\绘制所有图形.template	2020-12-30
1	SCI文献主题分析	SCI文献主题分析报告模板	\\report template\itginsighttemplate\sci-basetemplate.template	2020-12-30
2	德温特DII专利主题分析	德温特专利主题分析报告模板	\\report template\itginsighttemplate\dii-basetemplate.template	2020-12-30
3	CNKI文献主题分析	CNKI文献主题分析报告模板	\\report template\itginsighttemplate\cnki-basetemplate.template	2020-12-30
4	Patsnap智慧芽专利主题分析	Patsnap智慧芽专利主题分析	\\report template\itginsighttemplate\patsnap-basetemplate.template	2020-12-30
5	Incopat合享专利主题分析	Incopat合享专利分析	\\report template\itginsighttemplate\incopat-basetemplate.template	2020-12-30
6	Innovation专利主题分析	Innovation专利主题分析	\\report template\itginsighttemplate\innovation-basetemplate.template	2020-12-30
7	GF研究报告主题分析	GF研究报告主题分析模板	\\report template\itginsighttemplate\gf-baogao-basetemplate.template	2020-12-30
11	ezReport的SCI论文主题分析	ezReport的SCI论文主题分析模板	\\report template\ezreport.online\ezreport.online.sci.template	2020-12-30
12	ezReport的DII专利主题分析	ezReport的DII专利主题分析模板	\\report template\ezreport.online\ezreport.online.dii.template	2020-12-30
13	ezReport的CNKI论文主题分析	ezReport的CNKI论文主题分析模板	\\report template\ezreport.online\ezreport.online.cnki.template	2020-12-30
14	ezReport的Patsnap智慧芽专利主题分析	ezReport的Patsnap专利主题分析模板	\\report template\itginsighttemplate\ezreport.online.patsnap.template	2020-12-30
15	ezReport的Incopat合享专利主题分析	ezReport的Incopat专利主题分析模板	\\report template\itginsighttemplate\ezreport.online.incopat.template	2020-12-30
21	aiReview的SCI论文主题分析	aiReview的SCI论文主题分析模板	\\report template\aireview.online\aireview.online.sci.template	2020-12-30
22	aiReview的DII专利主题分析	aiReview的DII专利主题分析模板	\\report template\aireview.online\aireview.online.dii.template	2020-12-30
23	aiReview的CNKI论文主题分析	aiReview的CNKI论文主题分析模板	\\report template\aireview.online\aireview.online.cnki.template	2020-12-30
31	CKCEST期刊论文主题分析	中国工程科技知识中心期刊论文主题分析模板	\\report template\ckcesttemplate\ckcest_paper.template	2020-12-30
32	CKCEST会议论文分析	中国工程科技知识中心会议论文主题分析模板	\\report template\ckcesttemplate\ckcest_conference.template	2020-12-30
33	CKCEST学位论文分析	中国工程科技知识中心学位论文主题分析模板	\\report template\ckcesttemplate\ckcest_degree.template	2020-12-30
34	CKCEST专利分析	中国工程科技知识中心专利主题分析模板	\\report template\ckcesttemplate\ckcest_patent.template	2020-12-30
35	CKCEST成果分析	中国工程科技知识中心成果主题分析模板	\\report template\ckcesttemplate\ckcest_achievement.template	2020-12-30
36	CKCEST报告分析	中国工程科技知识中心报告主题分析模板	\\report template\ckcesttemplate\ckcest_report.template	2020-12-30
37	CKCEST项目分析	中国工程科技知识中心项目主题分析模板	\\report template\ckcesttemplate\ckcest_project.template	2020-12-30
38	CKCEST政策分析	中国工程科技知识中心政策主题分析模板	\\report template\ckcesttemplate\ckcest_policy.template	2020-12-30
39	智IP	智IP专利主题分析模板	\\report template\pmx\pmx-patent.template	2020-12-30

Topic *

OK Close

系统支持三类模板，主题分析模板，单个机构评价模板，多个机构对比模板。例如，选择一个报告模板，在 Topic/主题文本框填写分析报告的技术领域，比如“纳米技术”，点击 OK/确定按钮，软件自动撰写报告，用户根据需要进行修改，具体操作详见视频教程“智能报告撰写”。撰写报告的中间结果，包括各种矢量图、统计表格在软件安装目录 report template 下。

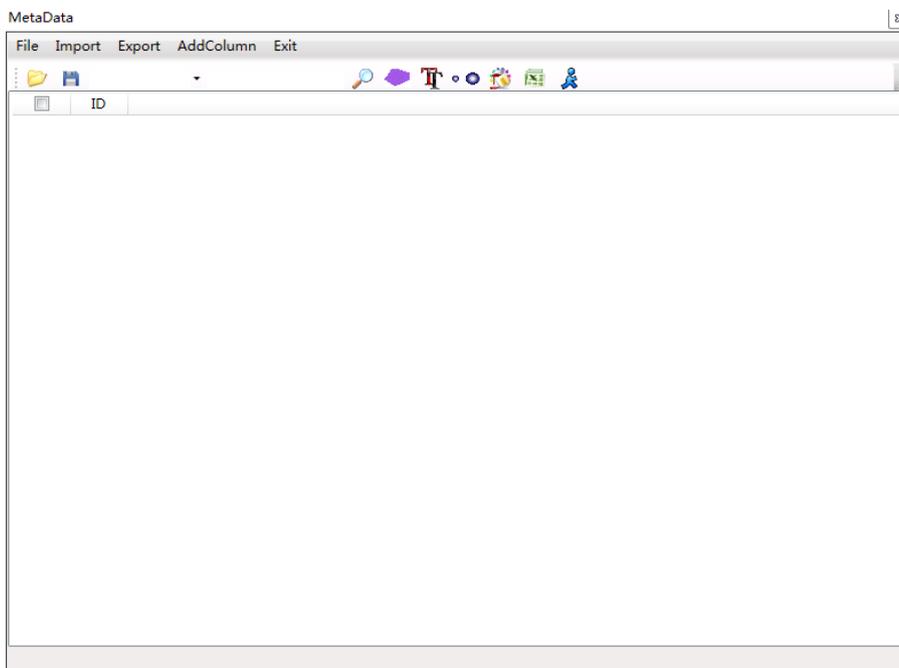
- aireview.online
- basetemplate
- ckcesttemplate
- ezreport.online
- ezreview.online
- itginsighttemplate
- pmx
- sharedimage
- sharedmacro
- tempimage
- tempxlsx
- reporttemplate.xml
- 绘制所有图形.template
- 绘制所有图形.xlsx

第十二章：元数据

12.1 元数据导入导出

元数据是指在数据分析完成、图形绘制完成后，将图形数据与后期导入的数据进行一一对应，类似于 GELPHI 的增加列功能。具体操作如下：

任何绘制好的图形后，点击工具栏的元数据按钮，如下图，弹出元数据页面：



点击导入元数据，系统提示输入 access、excel、txt 格式的元数据文件，TXT 格式的元数据文件格式如下：

```

EditPlus - [H:\Exe\ItgInsight_V1.6.0.0_绿色免安装版(企业版)_R64\example\metadata\meta.txt]
文件(F) 编辑(E) 视图(V) 搜索(S) 文档(D) 工程(P) 工具(T) 浏览器(B) 窗口(W) 帮助(H)
1 |id0|ipc_chars|ipc_2digs|meaning|ipc_lchar|ipc_1_num|ipc_2_num|title|abstract|kmeans1|kmeans2|kmeans3|kmeans4
2 |447010134|04|H04W|电通信技术|W|88|02|Reconfigurable multi-mode transceiver|Reconfiguring a transceiver design using a
3 |421953543|01|H01T|基本电气元件|T|21|02|High voltage connection sealing method for corona ignition coil|A corona ignit
4 |456796844|03|H03F|基本电子电路|F|3|30|Amplifier with triple-coupled inductors|An apparatus includes an amplifier and
5 |478609420|04|H04N|电通信技术|N|19|50|Method and device for sharing a candidate list|The present invention relates to
6 |438534158|01|H01J|基本电气元件|J|37|26|Preparation of cryogenic sample for charged-particle microscopy|A method of pr
7 |438140947|04|H04L|电通信技术|L|29|06|Aggregator node, method for aggregating data, and computer program product|Accor
8 |446540462|04|H04N|电通信技术|N|5|369|Radiographic image capturing apparatus|A radiographic image capturing apparatus
9 |451728729|02|H02P|发电、变电或配电|P|27|06|Motor driving apparatus|A motor driving apparatus includes a boosting sect
10 |48368472|01|H01R|基本电气元件|R|9|26|Relay terminal block|A relay terminal block fixed to a fixing part by a screw i
11 |425721885|04|H04N|电通信技术|N|17|02|Spectral synthesis for image capture device processing|A substantially rectangul
12 |425721887|04|H04N|电通信技术|N|5|235|Flicker detection method and flicker detection apparatus|The present invention c
13 |472252448|01|H01L|基本电气元件|L|31|18|Semiconductor device and method for manufacturing the same|A semiconductor dev
14 |472252450|02|H02J|发电、变电或配电|J|50|80|Tunable wireless power architectures|Described herein are improved config
15 |470676416|04|H04N|电通信技术|N|21|8545|Method and apparatus for managing connection between broadcast receiving devic
16 |421953573|01|H01G|基本电气元件|G|11|86|Electrodes for capacitors from mixed carbon compositions|Wet carbon paper proc
17 |422412329|04|H04M|电通信技术|M|1|02|Electronic device|An electronic device includes a first casing rotatably supporte
18 |426344490|05|H05K|其他类目不包含的电技术|K|7|00|Mounting device and electronic apparatus|A mounting device includes a
19 |424902663|04|H04N|电通信技术|N|9|80|Imaging apparatus|Provided are a recording source which outputs a stream includin
20 |450228639|02|H02M|发电、变电或配电|M|3|335|Bidirectional DC/DC converter|During a dead time, a resonance current gene
21 |471302200|04|H04L|电通信技术|L|27|38|Communication apparatus, demodulation apparatus, carrier reproduction apparatus,
22 |476414009|02|H02P|发电、变电或配电|P|27|08|Electric power steering apparatus|An electric power steering apparatus, wi
23 |438403131|04|H04L|电通信技术|L|29|12|Configuring a virtual machine|In some examples, a node in a distributed system n
24 |473379146|01|H01L|基本电气元件|L|33|64|Light emitting device|A light emitting device (1) comprising at least one light
25 |475693118|01|H01M|基本电气元件|M|10|54|Storage battery transfer support device and storage battery transfer support n
26 |471302218|05|H05B|其他类目不包含的电技术|B|41|14|Ripple based light emitting diode driving|Drivers (10) for driving a
27 |445775950|03|H03M|基本电子电路|M|1|78|Hybrid R-2R structure for low glitch noise segmented DAC|The apparatus may be a

```

第一行为表头，之后为数据，列“|”进行分割。元数据与图形数据通过 ID 列进行对应，ID 列必须为 1、2、3、4...格式的数字。或者导入元数据不包括 ID 列，软件自动添加 ID 列。

对于 EXCE 和 ACCESS 格式的元数据 ID 列的要求相同。

当导入的元数据列想以数字型导入，在表头加 /double，比如上图设置为“ipc_chars/double”后 ipc_chars 列以数字型导入。

推荐以 EXCEL 格式进行元数据的导入。

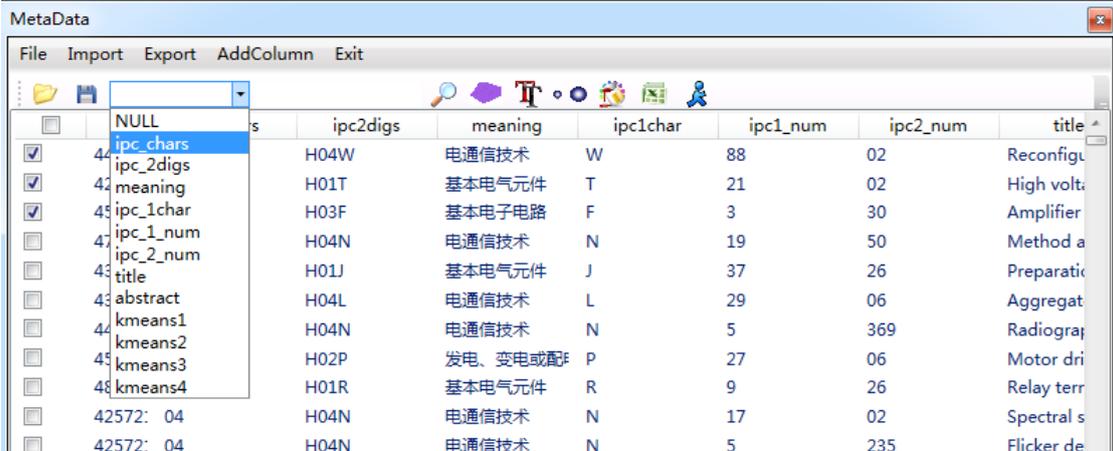
如果导出原始，点击导出按钮，系统默认导出 txt 格式的元数据，点击工具栏的 EXCEL 图标，则会导出 EXCEL 格式的元数据。

12.2 增加列

导入元数据后，如果想增加列，点击“增加列”按钮，导入方式与元数据导入一致，只是在先前导入的元数据后面再增加列。

12.3 依据元数据对图形进行查找、更改节点大小、更改节点文字、绘制凸包

根据导入的元数据对图形进行操作，点击工具栏，选择操作的列名，如下图。通过工具栏的        对图形进行修改，分别对图形进行查找、更改节点大小、更改节点文字、绘制凸包。



	ipc2digs	meaning	ipc1char	ipc1_num	ipc2_num	title
NULL						
44 ipc_chars	H04W	电通信技术	W	88	02	Reconfigu
44 ipc_2digs	H01T	基本电气元件	T	21	02	High volt
44 meaning	H03F	基本电子电路	F	3	30	Amplifier
45 ipc_1char	H04N	电通信技术	N	19	50	Method a
47 ipc_1_num	H01J	基本电气元件	J	37	26	Preparatic
43 ipc_2_num	H04L	电通信技术	L	29	06	Aggregat
43 title	H04N	电通信技术	N	5	369	Radiograp
44 abstract	H02P	发电、变电或配	P	27	06	Motor dri
44 kmeans1	H01R	基本电气元件	R	9	26	Relay terr
44 kmeans2	H04N	电通信技术	N	17	02	Spectral s
45 kmeans3	H04N	电通信技术	N	5	235	Flicker de
48 kmeans4						
42572: 04						
42572: 04						

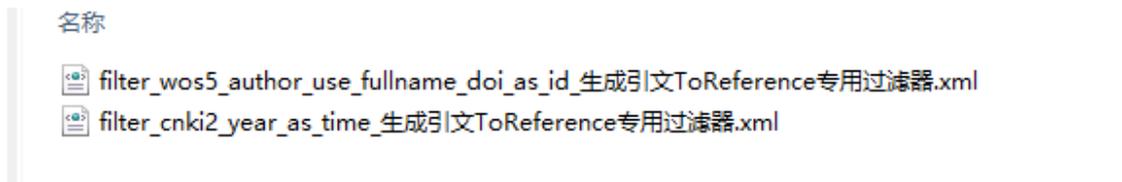
12.4 数据保存

导入元数据后点击保存按钮，提示保存为 .metadata 格式的元数据，即二进制元数据，之后可以直接打开，节省每次元数据导入时间。

第十三章：转化为参考文献

13.1 将文献按照参考文献格式输出到 WORD

点击菜单栏上的“Data/数据->Analysis/分析”，弹出数据转化页面，点击 ToReference 可将文献著录项目转化为 word 版本的参考文献格式。此处特别注意，选择的过滤器必须为 filter 目录下，标记为 ToReference 专用的过滤器，如图所示：



参考文献格式编辑器为软件目录下的 Referenceformat.json 文件，打开如下：

```

{
  "Field": "PublicationTypeMeta",
  "FieldValue": "J",
  "ReferenceFormat": "AuthorMeta$,.$AuthorMeta*TitleMeta*[J].$TitleMeta+PublicationMeta+PublicationMeta+TimeMeta,$TimeMeta+VolumeMeta+VolumeMeta+($PeriodMeta+PeriodMeta)$PeriodMeta+PageStartMeta+PageEndMeta+PageEndMeta"
},
{
  "Field": "PublicationTypeMeta",
  "FieldValue": "Article",
  "ReferenceFormat": "AuthorMeta$,.$AuthorMeta*TitleMeta*[J].$TitleMeta+PublicationMeta+PublicationMeta+TimeMeta,$TimeMeta+VolumeMeta+VolumeMeta+($PeriodMeta+PeriodMeta)$PeriodMeta+PageStartMeta+PageEndMeta+PageEndMeta"
},
{
  "Field": "PublicationTypeMeta",
  "FieldValue": "期刊",
  "ReferenceFormat": "AuthorMeta$,.$AuthorMeta*TitleMeta*[J].$TitleMeta+PublicationMeta+PublicationMeta+TimeMeta,$TimeMeta+VolumeMeta+VolumeMeta+($PeriodMeta+PeriodMeta)$PeriodMeta+PageStartMeta+PageEndMeta+PageEndMeta"
},
{
  "Field": "PublicationTypeMeta",
  "FieldValue": "期刊",
  "ReferenceFormat": "AuthorMeta$,.$AuthorMeta*TitleMeta*[C].$TitleMeta+PublicationMeta+PublicationMeta+PublicationAreaMeta+PublicationAreaMeta+PublishMeta,$PublishMeta+TimeMeta,$TimeMeta+PageStartMeta+PageEndMeta+PageEndMeta"
},
{
  "Field": "PublicationTypeMeta",
  "FieldValue": "过程型 Paper",
  "ReferenceFormat": "AuthorMeta$,.$AuthorMeta*TitleMeta*[C].$TitleMeta+PublicationMeta+PublicationMeta+PublicationAreaMeta+PublicationAreaMeta+PublishMeta,$PublishMeta+TimeMeta,$TimeMeta+PageStartMeta+PageEndMeta+PageEndMeta"
},
{
  "Field": "PublicationTypeMeta",
  "FieldValue": "期刊",
  "ReferenceFormat": "AuthorMeta$,.$AuthorMeta*TitleMeta*[J].$TitleMeta+PublicationMeta+PublicationMeta+TimeMeta,$TimeMeta+VolumeMeta+VolumeMeta+($PeriodMeta+PeriodMeta)$PeriodMeta+PageStartMeta+PageEndMeta+PageEndMeta"
}
}

```

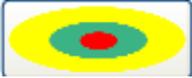
以第一个参考文献格式为例，含义如下，依据过滤器中的“PublicationType”的值，如果文献中“PublicationType”的字段值为“J”，采用 ReferenceFormat 的格式进行参考文献编辑，其中：

- AuthorMeta\$,的意思是输出参考文献时，文献作者之间用逗号分隔；
- +.\$AuthorMeta 的意思是如果存在文献作者字段，在作者后面加上符号“.”；
- 以此类推，可以自定义参考文献格式；

软件默认支持 CNKI 和 SCI 的文献著录项目转化为参考文献，也可以增加其他数据源的参考文献格式。

附录:

附录 A. 合著关系/同现分析/耦合分析

模型\图形元	节点大小	节点颜色	节点连线	节点文字
	第一、 二、三作 者数量	红色第一作者 数量、绿色第 二作者数量、 黄色第三作者 数量	合著/同现/耦合数量多少， 越多线越粗（如果选择 VS 和 UP 布局算法，那么连线 的长度与连线所代表的数 量呈反比）	作者/机构/ 国家/出版 物
	数量	颜色由内向外 逐渐变浅，颜 色环与各年份 数量成正比， 红色表示最接 近目前的时间	同上	同上

附录 B. 对应分析

模型\图形元	节点大小	节点颜色	节点连线	节点文字
	数量	无意义	关联性越大线越粗；反 之、越细（如果选择 VS 和 UP 布局算法，那么连 线的长度与连线所代表 的数量呈反比）	作者/机构/国 家/年代
	同上	颜色由内向外 逐渐变浅，颜 色环与各年份 数量成正比， 红色表示最接 近目前的时间	同上	同上

附录 C. 关联分析

模型\图形元	节点大小	节点颜色	节点形状	节点连线	节点文字
	数量	无意义	圆表示作者或机构、矩形表示技术类别	关联性越大、线越粗；反之、越细（如果选择 VS 和 UP 布局算法，那么连线的长度与连线所代表的数量呈反比）	作者/机构/国家/年代/类别

附录 D.引证关系

模型\图形元	节点大小	节点颜色	节点连线	节点文字
文献引证	无意义	无意义	引证关系	文献号 (主键文献)
作者/机构/期刊引证	被引用数量	无意义	引证关系	作者+年代/ 机构+年代/ 出版物+年代