

<图查看分析工具> 用户手册

部 门： 网络群体与市场第九组
批 准： 窦志成
日 期： 2017.12.17

中国人民大学信息学院

目 录

1 引言.....	1
1.1 编写目的.....	1
1.2 背景.....	1
1.3 术语和缩写词.....	1
2 软件概述.....	2
2.1 基本功能.....	2
2.2 工具与运行环境.....	2
2.3 配置方法.....	3
3 软件使用过程.....	3
3.1 运行过程.....	3
3.2 具体操作及实例.....	3
3.3 输入说明.....	9
4 程序清单说明.....	9

1 引言

1.1 编写目的

本工具为网络群体与市场课程的大作业，提供对于包含了约 6 万多个新闻的 json 数据的图形式的查询与分析功能。实验环境较为复杂，为了使老师能够对本作业进行测试并给予评价，以及为了方便潜在用户的使用，特编写此用户手册。

1.2 背景

本软件将从一批给定的新闻数据从中挖掘出这些新闻中的社交网络关系。

数据格式：数据文件中包含了约 6 万多个新闻的 json 数据，每个新闻给出了新闻网址、标题、正文、时间、以及其中包含的人名（Entity_Person）和机构名(Entity_Organization)。

本工具将基于该数据，建立社交网络图，我们作如下假设：

- ① 两个人如果出现在同一篇新闻中，则这两个人有联系。
- ② 两个人的联系强弱通过共同出现的文章的数目来表示。例如，假设 A 和 B 在 10 篇新闻中同时出现过，则 A-B 之间的边的权重为 10。

我们建立基于以上数据以及以上假设的社交网络图，并提供对该图的查询等功能。

1.3 术语和缩写词

基于老师给定数据按照“背景”中所描述的假设建立的社交网络图：network。

2 软件概述

2.1 基本功能

我们的软件将针对 `network` 提供以下功能：

- ① 输入人名 A 以及数字 k，找到与 A 关系最强的前 k 个邻居，若邻居数本身不到 k 个，则输出所有邻居。
- ② 输入人名 A 和人名 B，找到 A 和 B 之间最短的 10 条路径，如果有两条路径步长一致，则选取路径上所有边的权重和更大的一条。
- ③ 输入人名 A 和人名 B 以及数字 n，找到 A 和 B 之间排名第 n 的最短路径，其中 n 必须小于或等于 10。
- ④ 输入人名 A 和人名 B，A 和 B 之间所有的联系，即列出 A 和 B 所共同出现过的新闻列表。
- ⑤ 显示 `network` 中任意节点的聚集系数。
- ⑥ 显示 `network` 中任意节点的 PageRank 值。
- ⑦ 显示 `network` 中任意节点的点度中心性。
- ⑧ 显示 `network` 的节点个数。
- ⑨ 显示 `network` 的连通分量个数。
- ⑩ 显示 `network` 的最大连通分量的大小。
- ⑪ 显示在 `network` 中影响力最大的 20 个人。

2.2 工具与运行环境

- ① 使用 Django 作为 Web 框架，将前后台分开，以此为基础搭建框架。

- ② 引入 vis.js 包对 network 进行可视化。
- ③ 使用 python 3.6.3 编写后台。
- ④ 使用 Html 和 jQuery 编写前端。
- ⑤ 使用 NetworkX 对 network 进行分析和计算。

2.3 配置方法

- ① 安装 Python 环境（3.6.3）
- ② 使用 pip 导入 NetworkX 模块：pip install NetworkX
- ③ 使用 pip 安装 Django：pip install Django
- ④ vis.js 包已经放在代码中的 static 文件夹下，故不提。

3 软件使用过程

3.1 运行过程

- ① 将压缩包解压，进入 TinyWorld（即与 manage.py 同一目录下），打开命令行输入框。
- ② 输入 python manage.py sunserver 0.0.0.0:8000。
- ③ 打开浏览器输入 127.0.0.1:8000/index，按回车键即可访问工具。

3.2 具体操作及实例

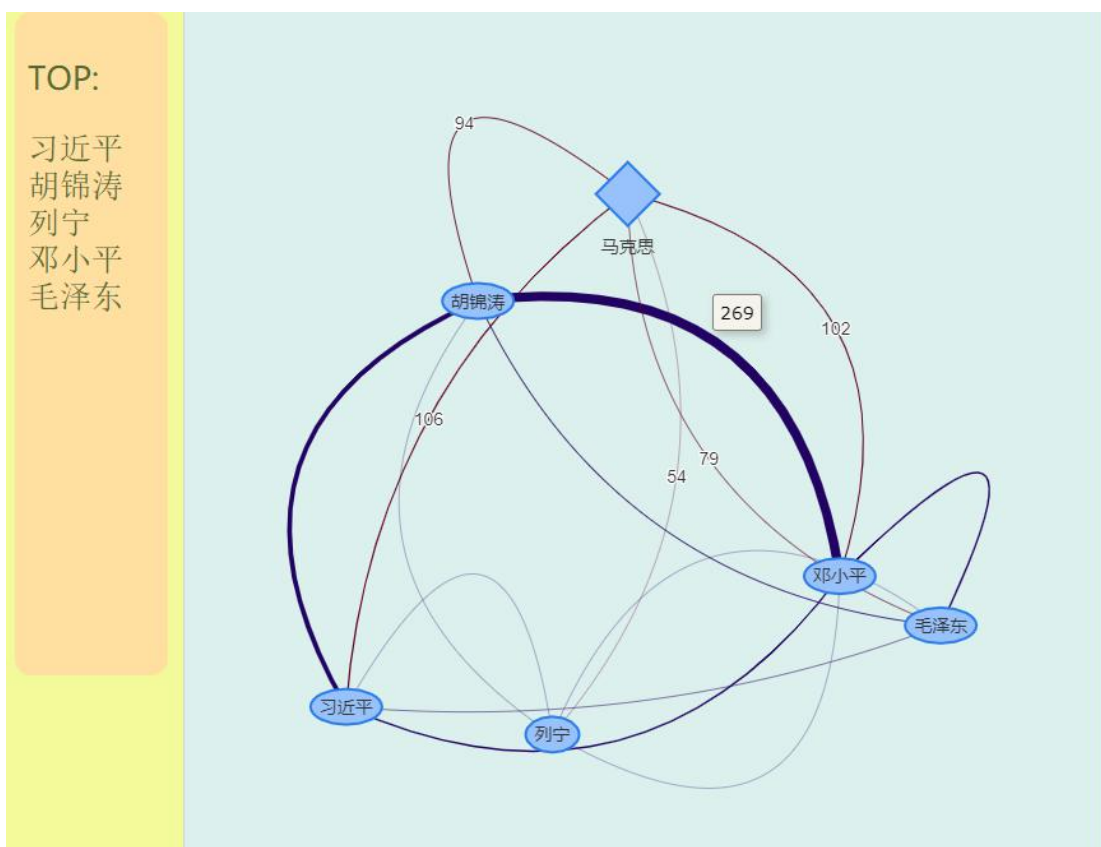
- ① 邻居查询模块具有两种显示模式，故具有“提交”和“提交 2”两个按钮：

显示模式一：在两个输入框中分别输入人名 A 和数字 k，点击按钮“提交”，即右方红色显示框中，显示 A 的 k 个关系最强的邻居，在右方蓝色显示框中显示 A 与这 k 个邻居构成的图的结构（显示它们之间所有的边）。

示例：

输入：马克思/5

输出：



【注：菱形节点表示 A，椭圆形节点表示 A 的邻居；

红色的边为一端节点为 A 的边，蓝色的边则为邻居之间的边；

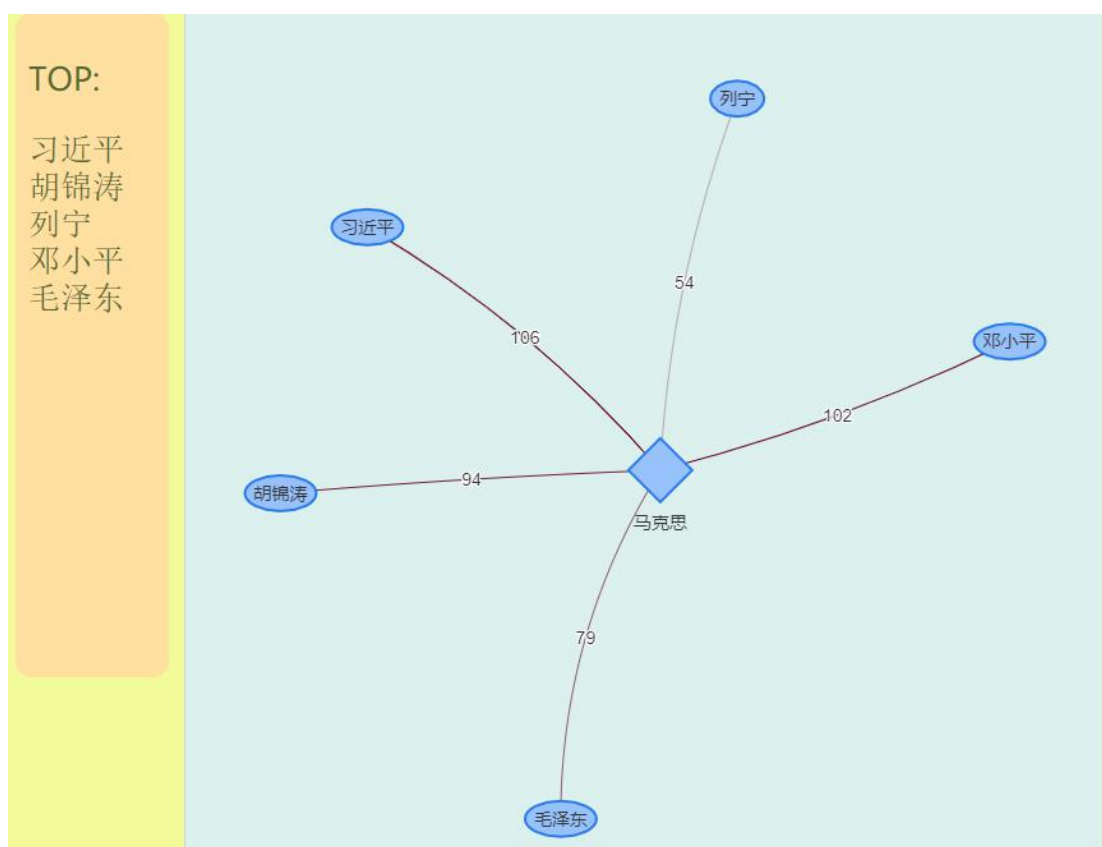
红边上的数字表示边的权重，而蓝色边上的权重需要鼠标移至边上才会显示，如图中的数字 269 即为边“胡锦涛-邓小平”的权重。】

显示模式二：在两个输入框中分别输入人名 A 和数字 k，点击按钮“提交 2”，即右方红色显示框中，显示 A 的 k 个关系最强的邻居，在右方蓝色显示框中显示 A 与这 k 个邻居构成的图的结构（只显示一端节点为 A 的边）。

示例：

输入：马克思/5

输出：



② 路径查找模块具有两种显示模式，故具有“寻找所有路径”和“寻找单条路径”两个按钮：

请输入姓名1

请输入姓名2

请输入路径排名

寻找所有路径

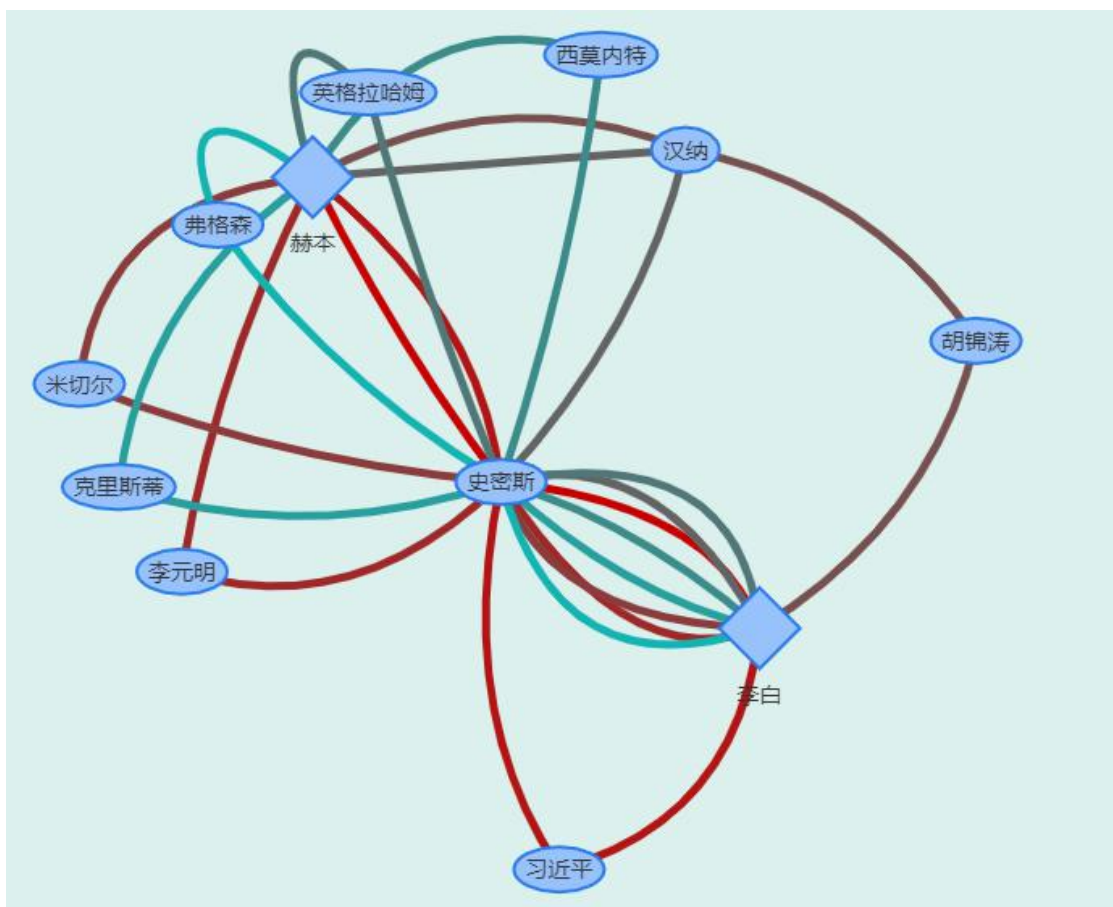
寻找单条路径

显示模式一：在前两个输入框分别输入人名 A 和人名 B，点击“寻找所有路径”按钮，则在右边的蓝色显示框中显示 A 和 B 之间的 10 条的最短路径。

示例：

输入：李白/赫本

输出：



【注：不同颜色标识不同的路径，故十条路径对应着十种不同的颜色】

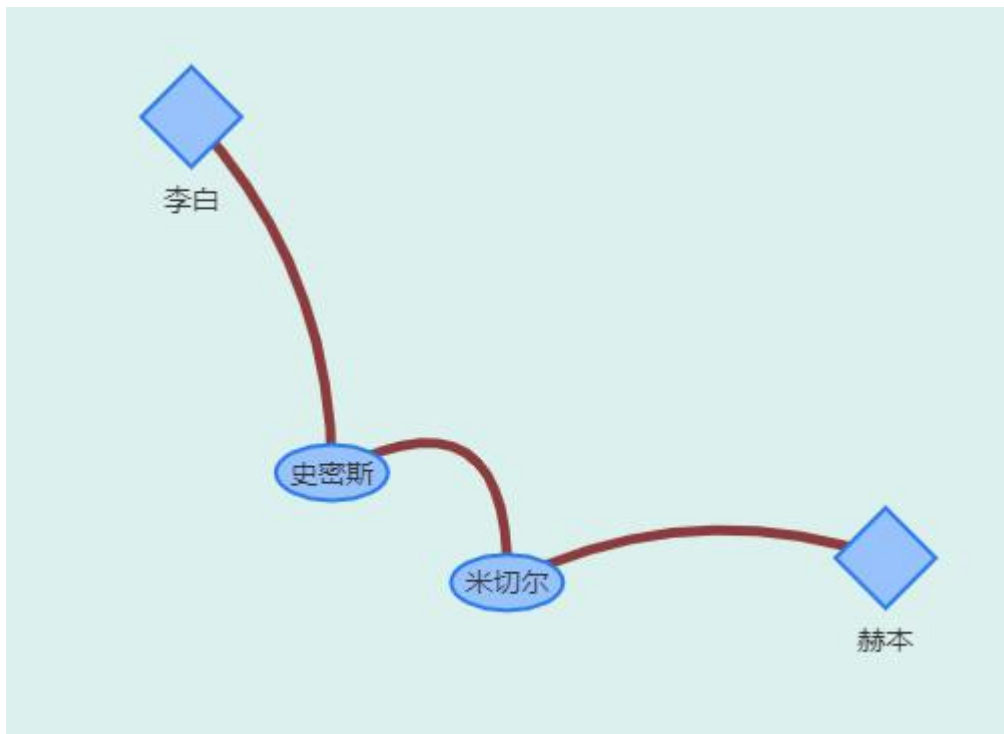
显示模式二：

在前两个输入框分别输入人名 A 和人名 B，在第三个输入框中输入数字 n，点击“寻找单条路径”按钮，则在右边的蓝色显示框中显示 A 和 B 之间的第 k 条最短路径。

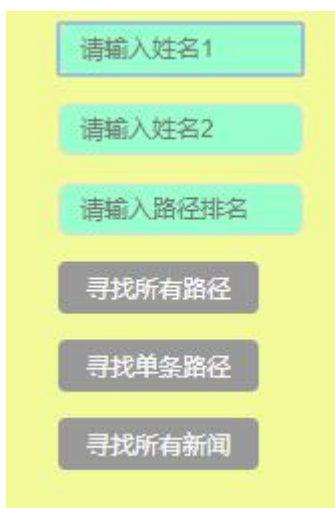
示例：

输入：李白/赫本/4

输出：



③ 寻找新闻列表模块：



在前两个输入框中分别输入人名 A 和人名 B, 点击“寻找所有新闻”按钮, 则显示 A 和 B 之间所有的联系, 即他们所共同出现过的新闻。

示例:

输入: 李白/王维

输出:

[驻札幌总领事孙振勇访问富士大学](#)2016-06-13T16:00:00Z
[驻格鲁吉亚大使陈建福出席第比利斯自由大学孔子学院揭牌仪式](#)2010-11-27T16:00:00Z

【注: 点击新闻题目可以直接访问新闻所在的原网页, 新闻题后的黑体字为新闻的时间】

④ 无论是邻居查询模块还是路径查找模块中, 在蓝色显示框中显示的图上的节点, 只要把鼠标放上去就能够显示该节点的聚集系数、PageRank 值以及中心性。

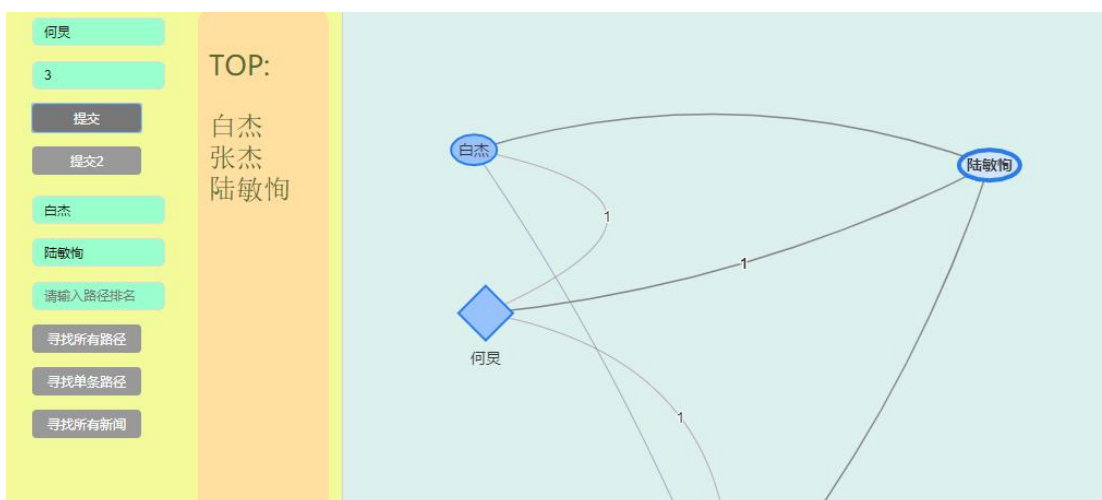
示例:



⑤ 点击蓝色显示框上图中的任意一个节点，就能够将该节点的人名写入输入框“请输入人名 1”和“请输入人名 2”中，方便用户进行查找。

示例：

单击下图中的“白杰”和“陆敏恂”节点，可以发现“白杰”“陆敏恂”被写入左边的输入框中了：



3.3 输入说明

- ① 请在输入框“请输入路径排名”输入小于等于 10 的数字。
- ② 请在姓名输入框输入 network 网络中存在的节点名称。

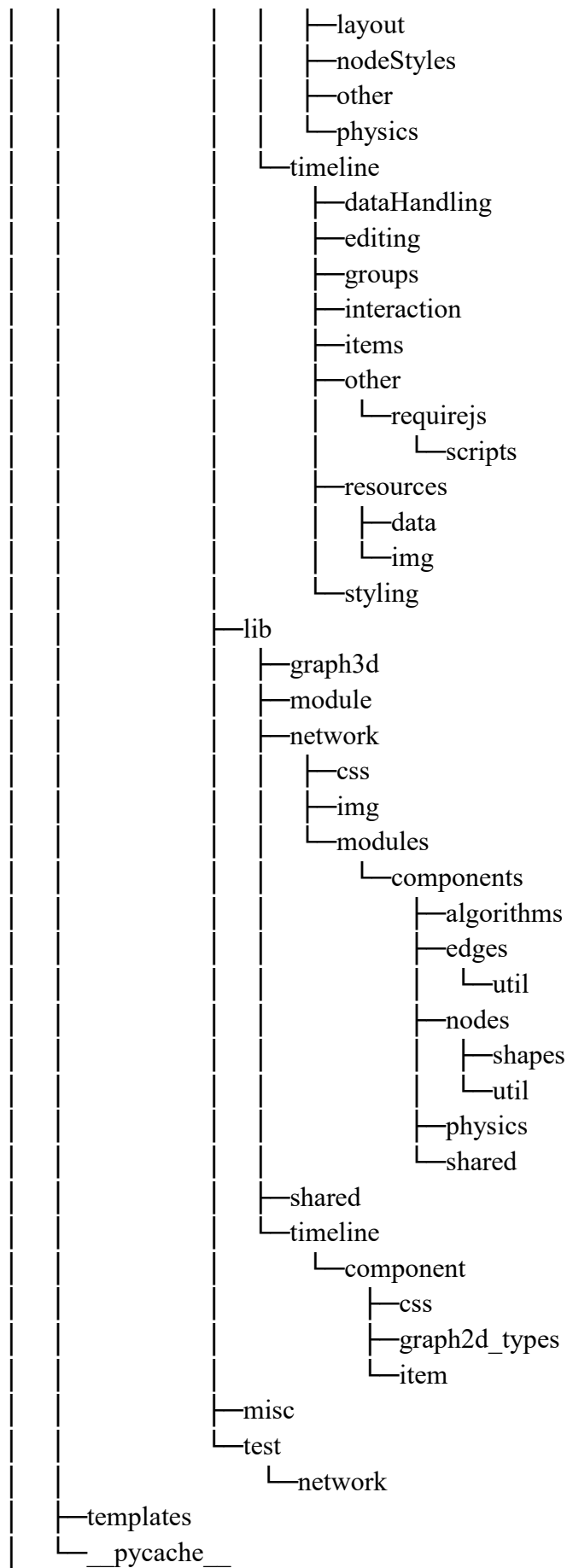
4 程序清单说明

- └templates
- └TestModel

```

├── migrations
│   └── __pycache__
├── static
│   └── vis-4.21.0
│       └── vis-4.21.0
│           ├── .github
│           ├── dist
│           │   ├── img
│           │   └── network
│           ├── docs
│           │   ├── css
│           │   ├── data
│           │   ├── fonts
│           │   ├── graph2d
│           │   ├── graph3d
│           │   ├── img
│           │   │   ├── external-link-icons
│           │   │   ├── graph
│           │   │   └── timeline
│           │   ├── js
│           │   │   └── prettify
│           │   ├── network
│           │   └── timeline
│           ├── examples
│           │   ├── graph2d
│           │   ├── graph3d
│           │   │   ├── playground
│           │   │   └── prettify
│           │   ├── network
│           │   │   ├── data
│           │   │   │   └── dotLanguage
│           │   │   │       ├── data
│           │   │   │       └── graphvizGallery
│           │   │   │           └── screenshots
│           │   ├── datasources
│           │   ├── edgeStyles
│           │   ├── events
│           │   ├── exampleApplications
│           │   ├── imageSelected
│           │   ├── img
│           │   │   ├── indonesia
│           │   │   ├── refresh-cl
│           │   │   └── soft-scraps-icons
│           │   └── labels

```



└TinyWorld
└__pycache__

以上为本工具的程序源代码清单，最主要的代码部分有三块：

- ① TinyWorld\TestModel\views.py：网页后台控制器。
- ② TinyWorld\TestModel\index.html：网页前端页面。
- ③ TinyWorld\manage.py：图处理程序。

数据部分为：

TinyWorld\clean.json

TinyWorld\network.bak

TinyWorld\network.dat

TinyWorld\network.dir