

微博Gephi关系图生成

自动爬取微博转发关系，用于生成Gephi下的节点图

Powered By [zy150\(一只懒虫\)](#)

repository: [zy150/-Weibo](#)

配置文件

配置文件在properties.json中,以下为样例

```
{
  "url": "https://weibo.com/ajax/statuses/repostTimeline",
  "topic": "#哈利波特海格扮演者去世",
  "sleep_time": 2,
  "mid": "P0KUL0lyE",
  "id": "5101803435725864",
  "limit": 1000,
  "pages": 1000,
  "save_path": "./Weibo_data/",
  "headers": {
    "authority": "weibo.com",
    "x-requested-with": "XMLHttpRequest",
    "sec-ch-ua-mobile": "?0",
    "cookie": "SCF=Alud2Q7aqlV5oX5tsYZdMHaf2Yiu47pn0-acCQPKtDFngGTuz2R2B4is-OtnqfDKfaIXKdVdCWH_sucdfrRux8.; SINAGLOBAL=3888515822091.3584.1731225435266; ULV=1731225435330:1:1:1:3888515822091.3584.1731225435266; PC_TOKEN=17d705145c; ALF=1734591982; SUB=_2A25KOEi-DeRhGeFL7FUT9inLyT-IHXVpNMR2rDV8PUJbkNANLU3gkw1Nfb1upkr04qq-q1RW3DmgfGiW2KzzkJG0; SUBP=0033WrsXqPxfm725ws9jqgMF55529P9D9W5Sfa8yDsMSxFC63lqwQAwb5JpX5KMhUgL.FomfSOME SomNeoe2dJLoIp7LxKML1KBLBKnlxKqL1hnLBOMNSKMNeoqNS0z0; XSRF-TOKEN=_Pxmi327oXRJpVAsm29l56JB; WBStorage=fcc86192|undefined; WBPSESS=NI5tmsiRcfXGnErssJsJXL9deUK-ry1_A8i1HVR0hPF7cvvSTq_CGe4vpzTUPZ2Vxem_E38mu-7qo_jyMXOpbtDNiATk8bKsKwvXgZE6zIdzBdmemr7CXHtcuERGvGwTmLci1MD4K-e4QhRubSOQ==",
    "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36",
    "content-type": "application/x-www-form-urlencoded",
    "accept": "*/*",
    "sec-fetch-site": "same-origin",
    "sec-fetch-mode": "cors",
    "sec-fetch-dest": "empty",
    "referrer": "https://weibo.com/1192329374/knng78Yf3?filter=hot&root_comment_id=0&type=comment",
    "accept-language": "zh-CN,zh;q=0.9,en-CN;q=0.8,en;q=0.7,es-MX;q=0.6,es;q=0.5"
  },
  "params": {
    "moduleID": "feed"
  }
}
```

必须填写

- id: 必须填写

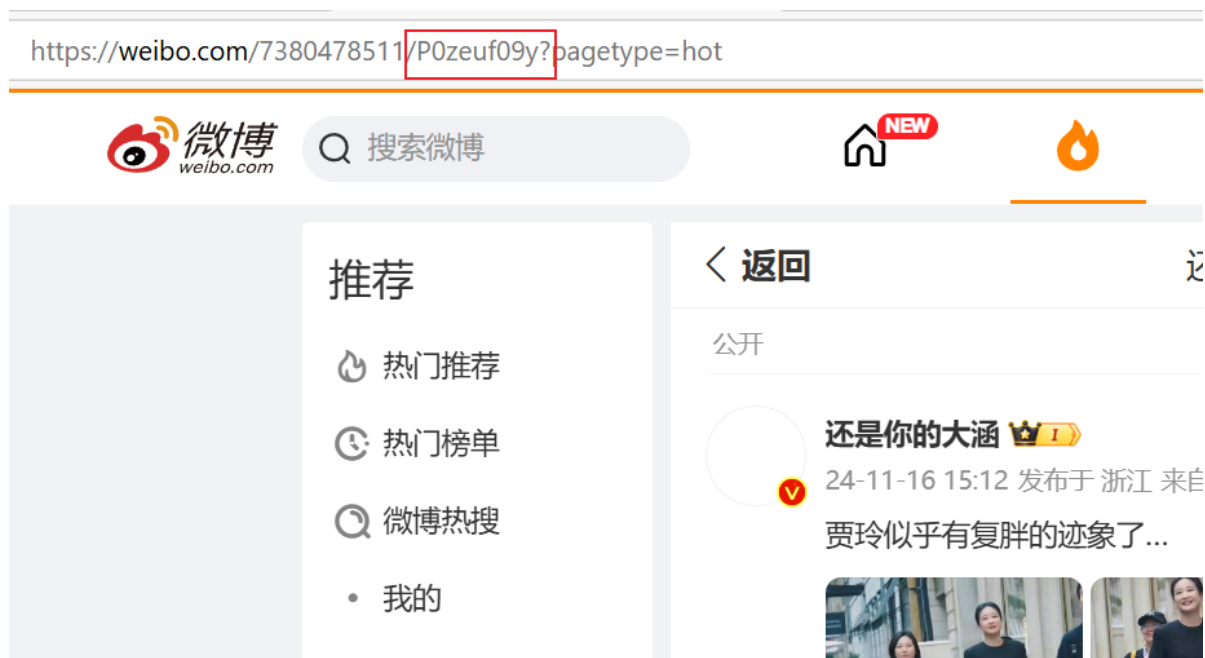
- mid: 必须填写
- topic: 必须填写
- headers: 请求头, 必须填写, 里面有cookie参数, 这里可以用下面的网站一键复制

无需填写

- sleep_time: 爬取间隔 默认为3
- pages: 爬取多少页 默认为500 每页大概15条数据, 爬虫已经自动判断最大有多少页, 到最大页就会停止
- result_file: 结果保存路径 默认为 ./WeiBo_data/mid+id.csv
- params: 请求参数 无需填写
- url: 请求url **无需填写**

参数获取

mid: 点开任意一条具体博文, 红框里就是mid



id&header:

第一步:

第三步：进入此网站[Convert curl commands to code](#)

Convert [curl](#) commands to Python, JavaScript and more

curl command

Examples: [GET](#) - [POST](#) - [JSON](#) - [Basic Auth](#) - [Files](#) - [Form](#)

```
-H 'sec-fetch-mode: cors' \
-H 'sec-fetch-site: same-origin' \
-H 'server-version: v2024.11.18.1' \
-H 'user-agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0 Safari/537.36 Edg/131.0.0.0' \
-H 'x-requested-with: XMLHttpRequest' \
-H 'x-xsrf-token: _Pxmi327oXRjpVAsm291563B'
```

复制到这里即可

[Ansible](#) [C](#) [C#](#) [ColdFusion](#) [Clojure](#) [Dart](#) [Elixir](#) [Go](#) [HAR](#) [HTTP](#) [HTTPie](#) [Java](#) [JavaScript](#) [Julia](#) [JSON](#) [Kotlin](#) [Lua](#) [MATLAB](#)

[Node.js](#) [Objective-C](#) [OCaml](#) [Perl](#) [PHP](#) [PowerShell](#) [Python](#) [R](#) [Ruby](#) [Rust](#) [Swift](#) [Wget](#) [可以自行转换代码](#)

```
import requests

cookies = {
    'SCF': 'Alud2Q7aq1V5oX5tsYzdMHaF2Yiu47pn0-acCQPKtDFngTuz2R2B4is-OtXnqfDKfaIXKdVdCwH_sucdfrRux8.',
    'SINAGLOBAL': '3888515822091.3584.1731225435266',
    'XSRF-TOKEN': '_Pxmi327oXRjpVAsm291563B',
    'WBPSESS': 'NIS5tmsiRcfXGnErssJsJXLa9deUK-ry1_A811HVR0hPF7cVvStq_CGe4vpzTUP2ZVxem_E38mu-7qo_jyMxOPbtDNiATk8bksKwvXgZE6zIdzBdmemr7CXHtcuERgVgGwT',
    'ALF': '1734592970',
    'SUB': '_2A25KOEyaDeRheFL7FUT9inLyT-IHXVpNMBSrDV8PUJbKbNANLRTnKw1Nfb1upjKhyKLTvyn-3q2gRvxJz8WYJQ8_',
    'SUBP': '0033WrsXqPxfM725Ws9jqgWF55529P9D9W55fa8yDsmSxFC63lqQAwB5JpX5K-huGL.FoMFS0MEs0MNeoe2dJLoIp7LxKML1KBLBKnlxKqL1hnLBoMNSKMNeoeqNS0z0',
    '_s_tentry': 'weibo.com',
    'Apache': '2729769732574.885.1732002333753',
    'ULV': '1732002333806:2:1:2729769732574.885.1732002333753:1731225435330',
}

headers = {
    'accept': 'application/json, text/plain, */*',
    'accept-language': 'zh-CN,zh;q=0.9,en;q=0.8,en-GB;q=0.7,en-US;q=0.6',
    'client-version': 'v2.46.33',
    # 'cookie': 'SCF=Alud2Q7aq1V5oX5tsYzdMHaF2Yiu47pn0-acCQPKtDFngTuz2R2B4is-OtXnqfDKfaIXKdVdCwH_sucdfrRux8.; SINAGLOBAL=3888515822091.3584.17312',
    'priority': 'u=1, i',
    'referer': 'https://weibo.com/1855757243/Malkoy741?refer_flag=1001030103_',
    'sec-ch-ua': '"Microsoft Edge";v="131", "Chromium";v="131", "Not_A_Brand";v="24"',
    'sec-ch-ua-mobile': '?0',
    'sec-ch-ua-platform': "Windows",
    'sec-fetch-dest': 'empty',
    'sec-fetch-mode': 'cors',
    'sec-fetch-site': 'same-origin',
    'server-version': 'v2024.11.18.1',
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0 Safari/537.36 Edg/131.0.0.0',
    'x-requested-with': 'XMLHttpRequest',
    'x-xsrf-token': '_Pxmi327oXRjpVAsm291563B',
}

params = {
    'is_reload': '1',
    'id': '4824652688130309',
    'is_show_bulletin': '2',
    'is_mix': '0',
    'count': '10',
    'uid': '1855757243',
    'fetch_level': '0',
    'locale': 'zh-CN',
}
```

cookie在这里，建议复制整个header
实际只更换配置文件中的cookie就行

id在这里

用法说明

更改id，mid，cookie之后，直接跑就行

- `crawl()`方法为爬取并存储为csv
- `process()`方法为处理`crawl`生成的csv，转为供Gephi使用的csv，保存路径 默认为 `./WeiBo_data/mid+id+_nodes.csv`
- 将最终的`_nodes.csv`导入到Gephi即可
- **注意：**建议数据量少时`auto_direct_topic`设置为true，代码有具体解释，默认为true