选择出具有旅游行为的游客：

SELECT distinct userid FROM suzhou.suzhou\_weibos\_sq\_0110 where tourism\_m\_new3 = 1;

共230577个独立个体，共发微博数为1096225，均发4.75条微博。

生成新表tourism\_id：

insert into suzhou.tourism\_id SELECT distinct userid FROM suzhou.suzhou\_weibos\_sq\_0110 where tourism\_m\_new3 = 1;

这些个体在14737个地点发送微博，

SELECT count(distinct poiid) FROM suzhou.suzhou\_weibos\_sq\_0110 where tourism\_m\_new3 = 1;

和distinct\_pois\_suzhou中一共由14695个poi重合。

SELECT count(distinct poiid) FROM suzhou.suzhou\_weibos\_sq\_0110 where tourism\_m\_new3 = 1 and suzhou\_weibos\_sq\_0110.poiid in (SELECT poiid FROM suzhou.distinct\_pois\_suzhou);

为什么suzhou.suzhou\_weibos\_sq\_0110中的poiid很多为null?

个体发微博数

update suzhou.tourism\_id set weibo\_count = (select count(\*) from suzhou.suzhou\_weibos\_sq\_0110 where userid = uid);

个体发微博poi数

update suzhou.tourism\_id set place\_count = (select count(distinct poiid) from suzhou.suzhou\_weibos\_sq\_0110 where userid = uid);

非空唯一的poi数目

insert into suzhou.tourism\_poi (select distinct suzhou\_weibos\_sq\_0110.poiid, null from suzhou.suzhou\_weibos\_sq\_0110 where suzhou\_weibos\_sq\_0110.poiid != "" and tourism\_m\_new3 = 1);

获取poi名字

update suzhou.tourism\_poi set title = (select distinct\_pois\_suzhou.title from distinct\_pois\_suzhou where tourism\_poi.poiid = distinct\_pois\_suzhou.poiid) where tourism\_poi.poiid in (select distinct poiid from distinct\_pois\_suzhou);

获取poi街道

update suzhou.tourism\_poi set poi\_street\_address = (select distinct\_pois\_suzhou.poi\_street\_address from distinct\_pois\_suzhou where tourism\_poi.poiid = distinct\_pois\_suzhou.poiid) where tourism\_poi.poiid in (select distinct poiid from distinct\_pois\_suzhou);

poi微博数

UPDATE suzhou.tourism\_poi SET weibo\_counts = (SELECT COUNT(\*) FROM suzhou.suzhou\_weibos\_sq\_0110 WHERE suzhou\_weibos\_sq\_0110.poiid = tourism\_poi.poiid);

大于1000条微博的poi有281个

SELECT \* FROM suzhou.tourism\_poi where weibo\_counts > 1000 order by weibo\_counts desc;

生成用于测试的随机1000个游客

Create table tourism\_id\_r\_1000 (SELECT \* FROM suzhou.tourism\_id where place\_count < 7 and place\_count > 0 order by RAND() LIMIT 1000);

流程：

1. 提取研究群体（游客）
2. 提取获取用户的访问点
3. 将访问点与所属的上层研究层次（景点、旅游区、市、省等）关联
4. 建立用户在研究层次上的轨迹，如景点的轨迹，旅游区的轨迹等，形成复杂网络
5. 利用寻找motif的算法，如ESU、Rand-ESU等精确网络motif发现算法来寻找子图。
6. 得到子图以及概率。
7. 后续分析，利用研究层次（景点、旅游区、市等）在复杂网络中搜索，查询在各个子图中出现的频率以及在子图中承担的功能（过渡点等）的分析