

社会网络中的信息传播优化问题

问题背景

1. 社会网络

社会网络是指社会个体成员之间因为互动而形成的相对稳定的关系体系。社会网络以个人为节点（node）构成社会结构，人与人之间通过相互依赖关系联结起来。相互依赖关系可能是朋友关系、同学关系、生意伙伴关系、种族信仰关系等。

一个社会网络可以用一张网络图来表示，其中节点（node）代表人，边（edge）代表人与人之间的关系。如果两节点之间的关系是双方对等的（例如朋友关系、同学关系等），则边为无向边；如果两节点之间的关系是不对等的（例如微博的关注关系、论文的引用关系等），则边为有向边，从一个节点指向另一个节点。

无向图与有向图的示例见图 1。

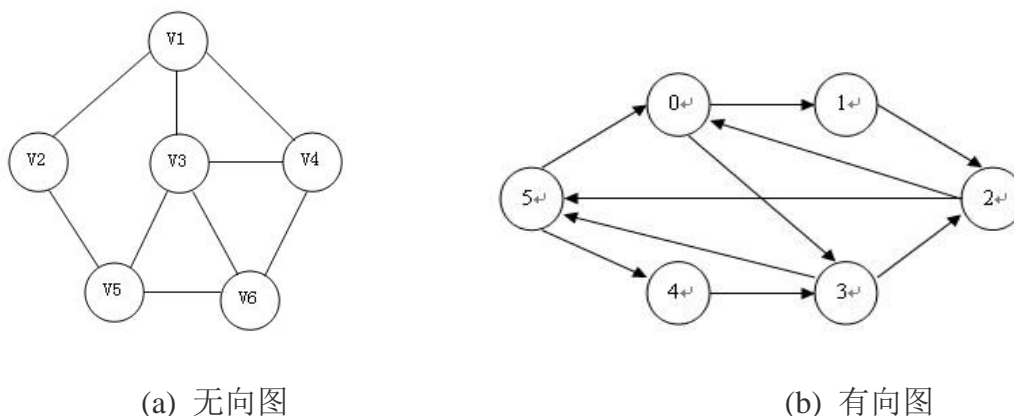


图 1

以下以微博网络为例，讨论信息在社会网络中的传播问题。微博网络结构示例见图 2。

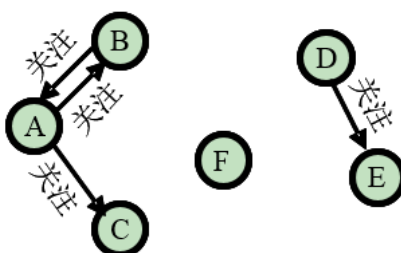


图 2

2. 社会网络中的信息传播

信息在社会网络中以个人节点为载体，沿着节点之间的边进行传播。信息传播的方向与边的指向一致。在网络中，从不同节点开始传播的信息，其传播效果可能大不相同。社会网络中的信息传播优化问题所要讨论的就是如何选择起始的

传播节点，使得信息能获得最大范围的传播（或达到指定的范围）。

3. 独立级联模型（Independent Cascade Model）

社会网络中的信息传播过程可以用独立级联模型来描述。该模型将整个社会网络看做一个有向图 $G(V, E)$ ，其中 V 是所有节点的集合， E 是所有边的集合。每条边有一个传播概率 $p, 0 \leq p \leq 1$ ，即信息有概率 p 可以沿着某条边从一个节点传播到另一个节点。在此假设所有边的传播概率都相同。

信息传播的过程如下：

在 $t = 0$ 时刻，信息从某些节点开始第一次传播。这些初始节点被认为是处于激活状态，构成初始激活集合 S_0 。

在 $t = 1$ 时刻，集合 S_0 中的节点 $v \in S_0$ 可以将信息以概率 p 传播给它们未被激活的邻居节点 $u \notin S_0$ （邻居节点即有边与之相连的节点，信息传播方向与边的指向一致）。如果传播成功，即 u 被激活，则 u 将被加入 $t = 1$ 时刻的激活集合 S_1 。集合 S_1 包含 S_0 中的所有节点，以及在 t_1 时刻被激活的所有节点。

.....

在 t 时刻，集合 S_{t-1} 中的节点 $v \in S_{t-1}$ 可以将信息以概率 p 传播给它们未被激活的邻居节点 $u \notin S_{t-1}$ 。如果传播成功，即 u 被激活，则 u 将被加入 t 时刻的激活集合 S_t 。集合 S_t 包含 $u \notin S_{t-1}$ 中的所有节点，以及在 t 时刻被激活的所有节点。

不同节点的激活过程相互独立，互不影响。已被激活的节点将永远处于激活状态，未被激活的节点不具备记忆性，下一次有同样的概率 p 被激活。

当网络中没有新的节点被激活时，传播过程停止。

优化问题

给定一个弱连通的有向网络，包含 1377 个节点（network_nodes.txt）和 2279 条边（network_edges.txt）。网络如图 3 所示。

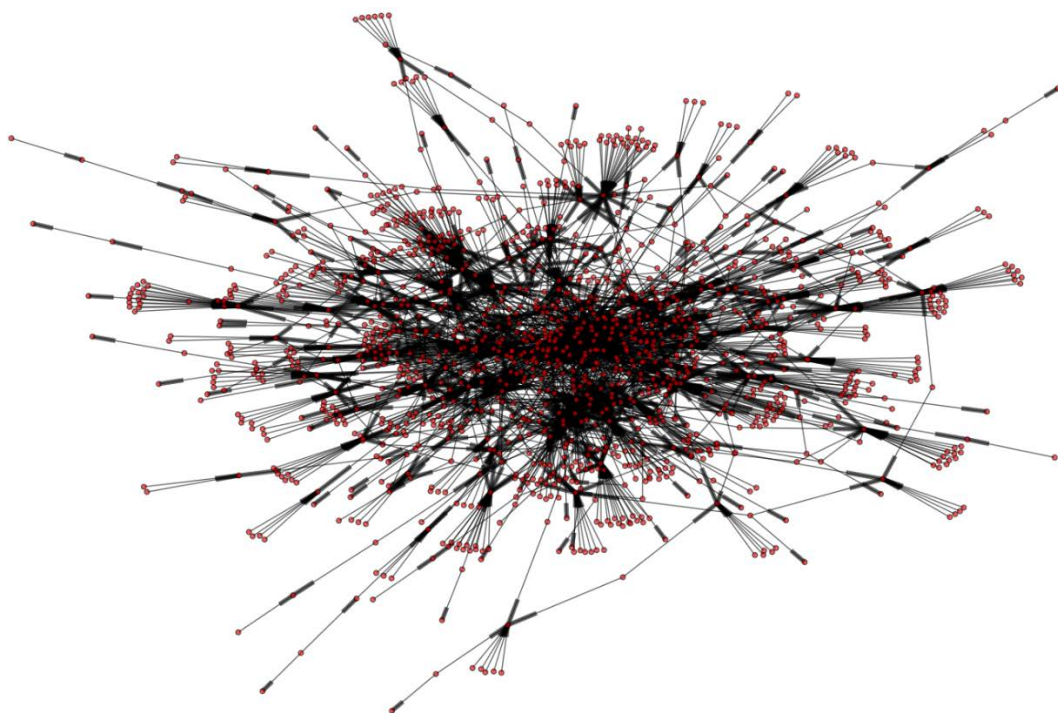


图 3

此次的信息传播优化问题描述如下：

1. 如何选择 10 个初始节点，使得信息的传播范围最广？
2. 如果希望信息的传播能覆盖 800 个以上的节点，则最少应该选择哪些用户做为传播的起始节点？

* `network_edges.txt` 文件的格式：每一行代表一条边，例如 `1381601997,1662848227` 代表如图 4 的一条边。

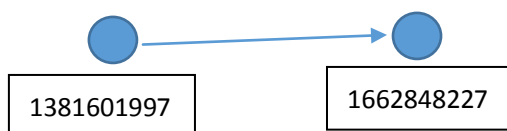


图 4