

V 是所有用户集合， E_f 是在线社交网络边集， E_p 是物理世界边集。假设在物理世界某个位置发生活动，即这个位置被激活，那么这个位置周围的用户将会受到影响，受影响的用户可能会在在线社交网络上分享这个活动，它的朋友们因此可能受到影响。

Measurements

两个基于位置的社交网络数据集Brightkite和Gowalla。

BRIGHTKITE DATASET

property	value	property	value
# users	3551	# logins	430657
# edges	9317	average login	121.278
average degree	5.248	# triangles	6738
# CC	569	average CC size	6.241
# nodes in largest CC	2907	# edges in largest CC	9228

GOWALLA DATASET

property	value	property	value
# users	5231	# logins	297104
# edges	10134	average login	56.797
average degree	3.875	# triangles	11580
# CC	1778	average CC size	2.942
# nodes in largest CC	3114	# edges in largest CC	9676

作者通过分析测量数据，得出以下结果：

1. 用户的位置和朋友的数量
 - a. 物理世界中用户的登录位置在很多情况下基本保持不变。
 - b. 少量用户的朋友数量占总用户的大部分，这表明在线社交网络中可以通过初始的少数用户最终影响大量的用户。
 - c. 在物理世界中一个用户有很多的邻居，这表明用户可以很容易的在物理世界中传播事件消息。
2. 位置和友谊
 - a. 如果在在线社交网络中用户互为朋友，那么在物理世界中他们的距离将会很近。
 - b. 在在线社交网络的帮助下，尽管现实用户之间距离很远，但仍有可能相互影响；同理在线上不是朋友的用户若在线下距离很近，仍有可能相互影响。
 - c. 在线上是朋友的用户，在线下他们的行动轨迹相似度很高。

解决方法

影响传播模型

1. 基础影响传播模型

用函数 $I(x, I_{max}) = (I_{max} - 1)\sqrt{1 - (1 - x)^2} + 1$ 来衡量影响传播增量。如果活动类型和用户兴趣之间相匹配，那么用户有很大概率被影响，用 $I_1(E, v)$ 来衡量该匹配后影响和分享概率的增量，将其定义为：

$$I_1(E, v) = I(J(E.type, v.intersect), I_{max1})$$

J 为 Jaccard 相似度。

假设初始传播时间为 $E.t_{init_pro}$ 和额外传播时间 $E.t_{add_pro}$ ，事件 E 在物理世界中在初始传播周期 $[E.t_0, E.t_0 + E.t_{init_pro}]$ 被激活，在额外传播周期 $[E.t_0 + E.t_{init_pro}, E.t_0 + E.t_{init_pro} + E.t_{add_pro}]$ 上，在在线社交网络和物理世界中同时传播。

2. 初始化传播模型

$R(x, y, r)$ 表示位置 (x, y) 为中心，半径为 r 的圆形区域。

$S(x, y, r, t)$ 表示在 t 时刻，在区域 R 内的用户。

$S(x, y, r, t_1, t_2)$ 表示在 $[t_1, t_2]$ 时间内在区域 R 中的用户。

如果事件 E 被激活，那么任意节点

$v \in S(E.x_0, E.y_0, E.r_0, E.t_0 + E.t_{init_pro})$ 将会以概率

$p(E, v, init_inf) = \min(p_1 I_1(E, v) I_2(E, v), 1)$ 受到影响。其中

$I_2(E, v)$ 依赖于 v 在 R 中呆的时间：

$$I_2(E, v) = I(T(E, v) / E.t_{init_pro}, I_{max2})。$$

3. 在线社交网络上的传播模型

如果 v 被影响了，那么他将以 $p(v, osn_share) = \min(p_2 I_1(E, v), 1)$ 的概率分享给朋友 u 。 u 收到影响的概率是

$p(E, u, osn_inf) = \min(p_3 I_1(E, u) I_3(u, t), 1)$ ， $I_3(u, t)$ 表示基于接受 E 事件的用户数量的影响概率增量：

$$I_3(u, t) = I(\min(\frac{n_r(E, u, t) - 1}{n_{max}}), I_{max3})$$

4. 物理世界上的传播模型

如果 v 被影响了，那么 v 将会在物理世界中以概率

$p(v, pw_share) = \min(p_4 I_1(E, v), 1)$ 继续传播。

通过轮盘赌方式来确定哪一个邻居 u 被选择了，选择之后其被影响概率为

$$p(E, u, phy_inf) = \min(p_5 I_1(E, u) I_4(v, u), 1)$$

5. 交叉传播

- 影响可能从在线社交网络传播到物理世界
- 影响可能从物理世界传播到在线社交网络。

Algorithm

1. 问题定义

假设初始化传播时间和额外传播时间以及影响半径保持不变。将问题重新定义为给定候选位置集合 C ，事件 E ，两层图 $G^t = (V, E_f, E_p^t)$ ，最大化影响用户数量 $F(pos)$ 。

2. 启发式算法

- 从候选位置集合 C 中选择 N_s 个位置。
- 用轮盘赌方法从 N_s 个位置中选择一个。用有更好的 $F(pos)$ 值的新位置代选择位置，这个过程实现 N_i 次。
- 返回最大的 F 值的位置。

从已选位置 pos_1 计算新位置 pos_2 :

- $N_c[pos_1]$ 记录选择 pos_1 的次数。
- 计算距离 $dis(pos_1, pos_2)$ 的上界。
- 随机从 C 中选择满足与 pos_1 距离小于上界的位置作为 pos_2 。

为了改进算法效率，作者提出了新的 F :

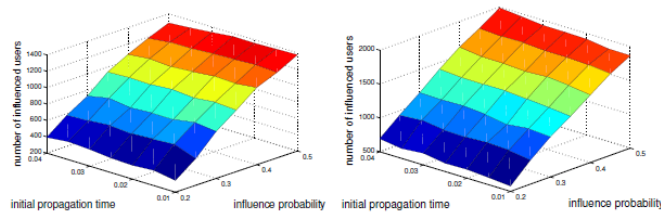
$$F'(pos) = \sum_{w \in U} |u.friends| + \sum_{v \in Neg(u,t)} \frac{|v.friends|}{Neg(u,t)}$$

实验分析

1. 影响用户的数量

设置了4组实验:

- 初始传播时间和影响概率对影响用户数量的影响

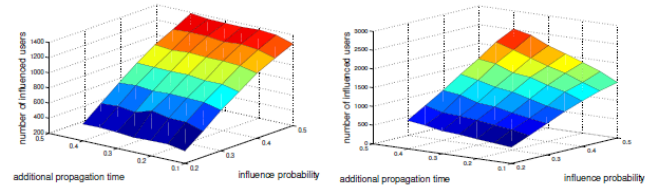


(a) Brightkite

(b) Gowalla

随着影响概率或者初始传播时间的增加，影响用户数量几乎是上升趋势。

- 额外传播时间和影响概率对影响用户数量的影响

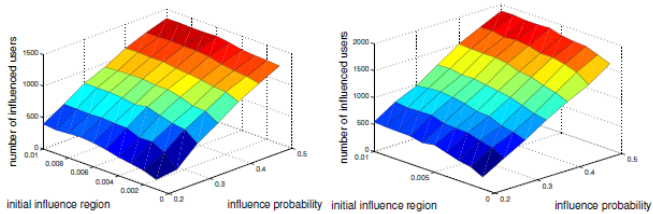


(a) Brightkite

(b) Gowalla

随着影响概率或者额外传播时间的增加，影响用户数量几乎是上升趋势。

- 初始影响区域和影响概率对影响用户数量的影响

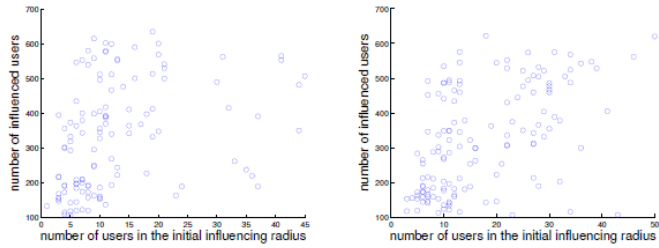


(a) Brightkite

(b) Gowalla

随着影响概率或者初始影响区域的增加，影响用户数量几乎是上升趋势。

- 在初始影响区域的用户数和最终影响用户数量无关:



2. 影响用户数量分布

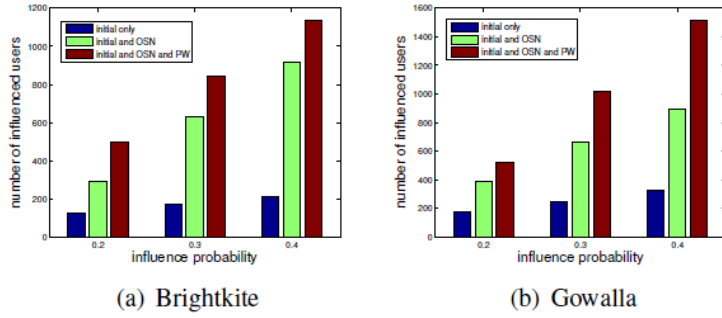


Fig. 14. Comparison of different propagation models.

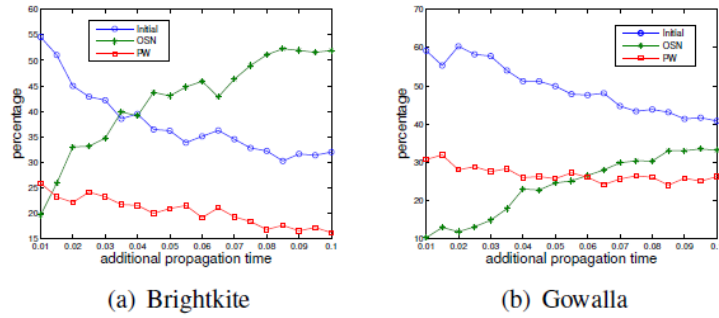
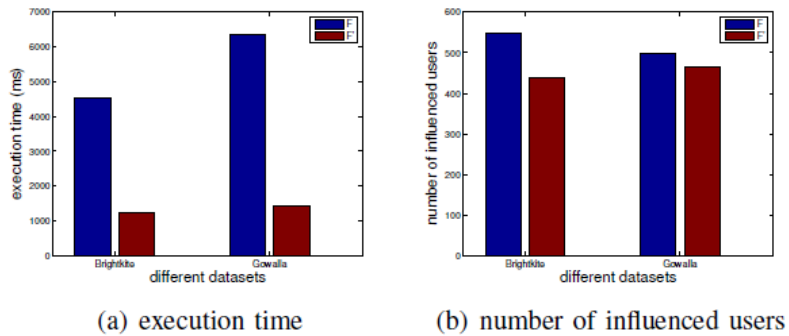


Fig. 15. Comparison of different influence manners.

在交叉影响传播模型下，影响的用户数量最多，随着额外时间的增加在线社交网络中影响的用户数量变多，在物理世界影响的用户数量变少，总体呈增加趋势。

3. F 和 F'



显然 F' 的算法要比 F 的算法运行时间要快，但是最终结果要略差一点。

创新之处

1. 基于问题的创新：提出了新的基于位置的关于在线社交网络和物理世界交叉影响传播的影响力最大化问题。

2. 算法上的创新：提出了解决基于位置的交叉传播影响力最大化问题的启发式解决算法，并取得了较好的效果。