Location Influence in Location-based Social Networks

Muhammad Aamir Saleem[1,2], Rohit Kumar[1], Toon Calders[1,3], Xike Xie[2,4] and Torben Bach Pedersen[2]

[1]Department of Computer and Decision Engineering
Université Libre de Bruxelles, Belgium

[2]Department of Computer Science
Aalborg University, Denmark

[3]Department of Mathematics and Computer Science
Universiteit Antwerpen, Belgium

[4]Suzhou Institute for Advanced Study
University of Science and Technology of China, China

1. Motivation

Out-of home (OOH) marketing covers 9% of marketing budget: 2.96 billion USD in year 2015.

Example:

- Locations display banners.
- Distribute promotional free gifts e.g., T-shirts.

How to decide where?

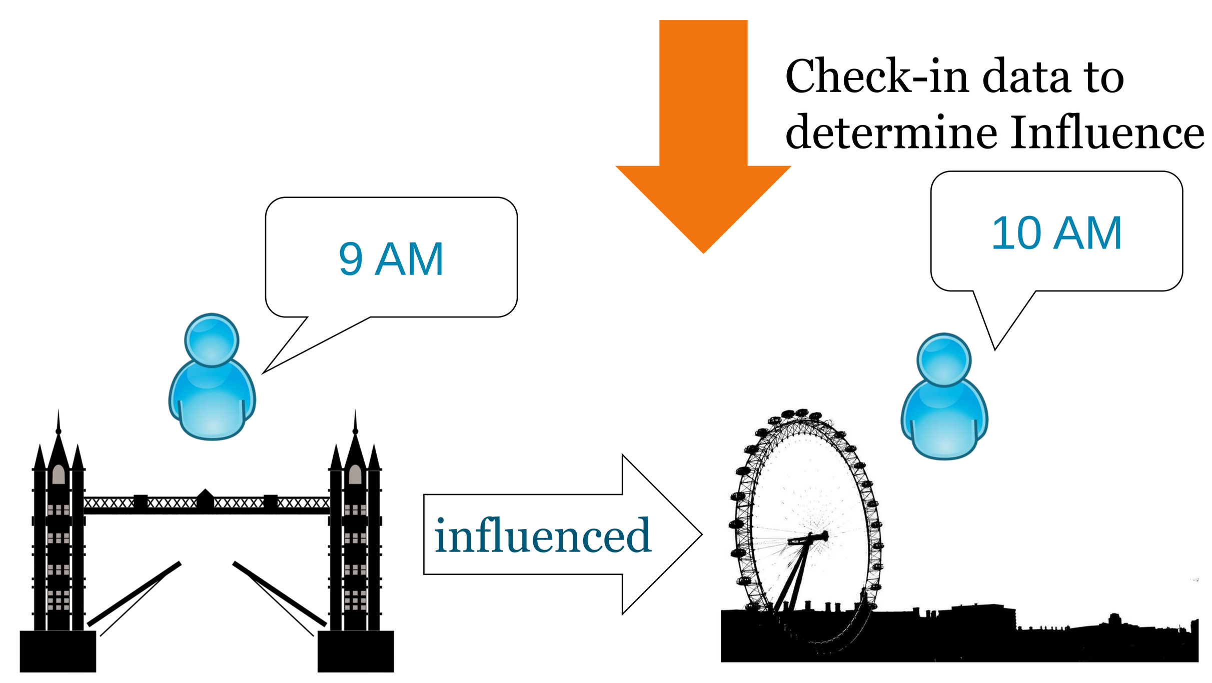
"How to increase geographical spread of the message using OOH on LBSN data!! "

2. Location Influence Definition

Location Influence: Capacity to spread its visitors to other locations.

Influence Strength: Number of users travelling between the locations.

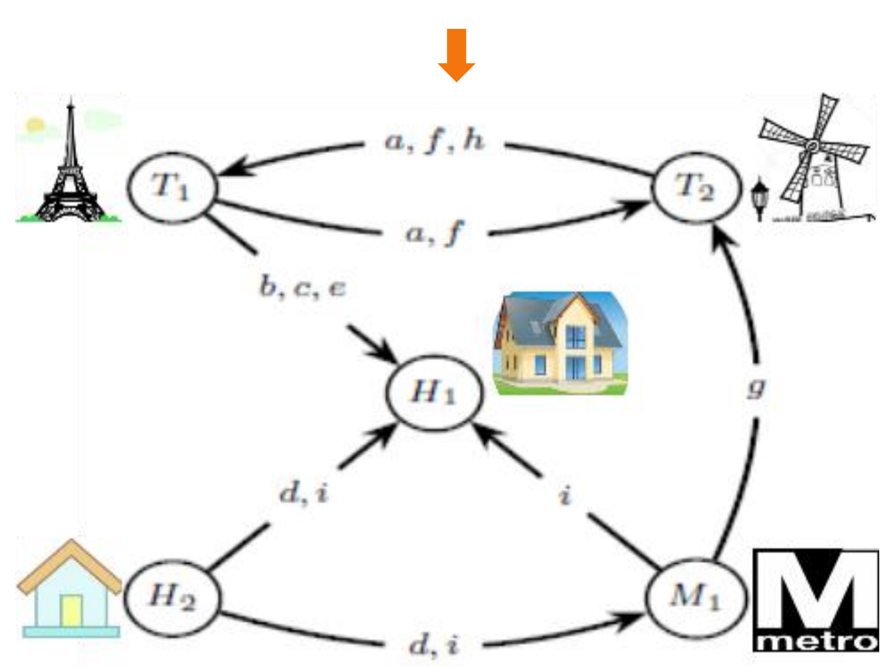
- **Absolute Influence Model:**
  - Influence exists if bridging visitors within a given time are greater than threshold
  - Example: T1 => T2 := |VB(T1,T2)| >= 2
- **Relative Influence Model:**
  - Biasness of popular locations, consider relative influence
  - Example: T1=> H1 := |b,c,e| / | b,c,e,i,d | >=0.4
- **Friendship-based Influence:**
  - Handle sparsity.
  - Predict future influence.



3. Algorithm

Check-ins

loc	Users t=1	t=2	t=3
T_1	b, c, e, f	a, h	f
T_2	a, h	f, g	a
M_1	g	i	d
H_1	$-$	b, c, d, e	i
H_2	d, i	$-$	$-$



Get influence graph from check-in data

Exact version:

Simple 3 step Algorithm for every activity:

- Update Location Summary φ(l)
- Update user history H(u).
- Prune user history for a ω.

	$t=1$	$t=2$	$t=3$	t=5
Activity:	$(i, H_2, 1)$ $(d, H_2, 1)$	$(i, M_1, 2)$ $(d, H_1, 2)$	$(i, H_1, 3)$ $(d, M_1, 3)$	$(d, H_2, 5)$
$\mathcal{H}(i)$:	$\{(H_2, 1)\}$	$\{(H_2, 1), (M_1, 2)\}$	$\{(H_2, 1), (M_1, 2), (H_1, 3)\}$	$\{(H_1, 3)\}$
$\mathcal{H}(d)$:	$\{(H_2, 1)\}$	$\{(H_2, 1), (H_1, 2)\}$	$\{(H_2, 1), (H_1, 2), (M_1, 3)\}$	$\{(M_1, 3), (H_2, 5)\}$
$\varphi(H_1)$:	$\{\}$	$\{\}$	$\{(M_1, \{d\})\}$	$\{(M_1, \{d\})\}$
$\varphi(H_2)$:	$\{\}$	$\{(H_1, \{d\}), (M_1, \{i\})\}$	$\{(H_1, \{d\}), (M_1, \{i, d\})\}$	$\{(H_1, \{d\}), (M_1, \{i, d\})\}$
$\varphi(M_1)$:	$\{\}$	$\{\}$	$\{(H_1, \{i\})\}$	$\{(H_1, \{i\}), (H_2, \{i\})\}$

Approximate version:

Compress the set using HyperLogLog*(HLL)!

Remains exactly the same.

Time Complexity improves:

O(ω log(|U|)) -> O(ω)

Memory Improves :

O(|L||U|) -> O(|L|b)

Standard error : ~ 1.04 / √b

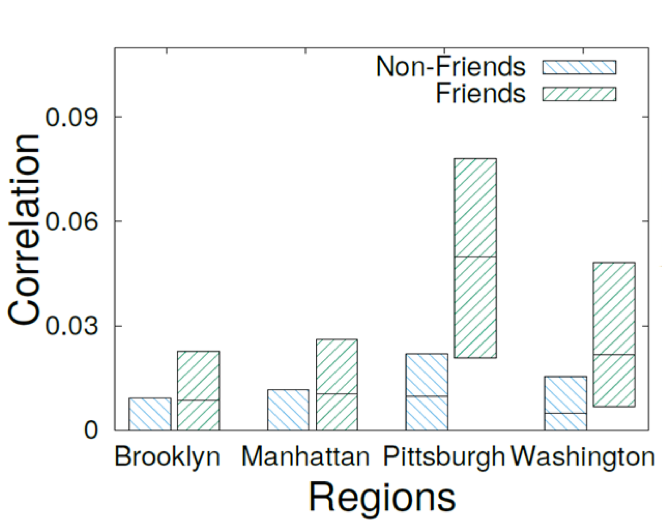
b : Number of buckets in HLL sketch

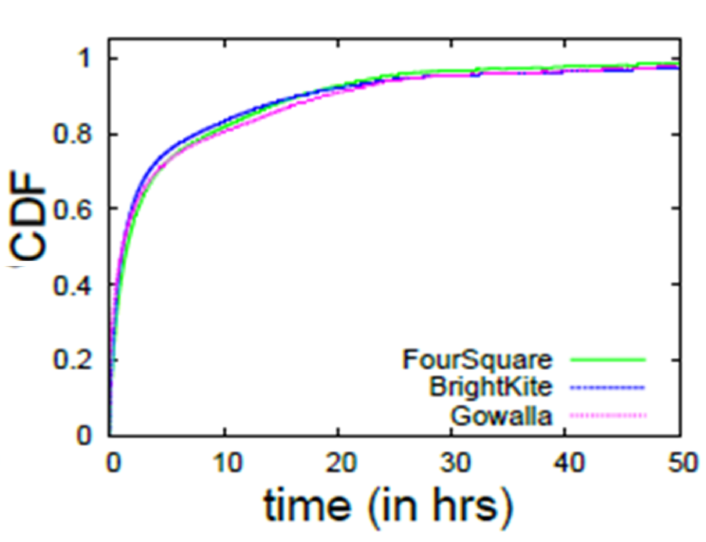
* P. Flajolet, E. Fusy, O. Gandouet, and F. Meunier. Hyperloglog: the analysis of a near-optimal cardinality estimation algorithm. In DMTCS, 2008.

4. LBSN Analysis

a. Users follow Friends

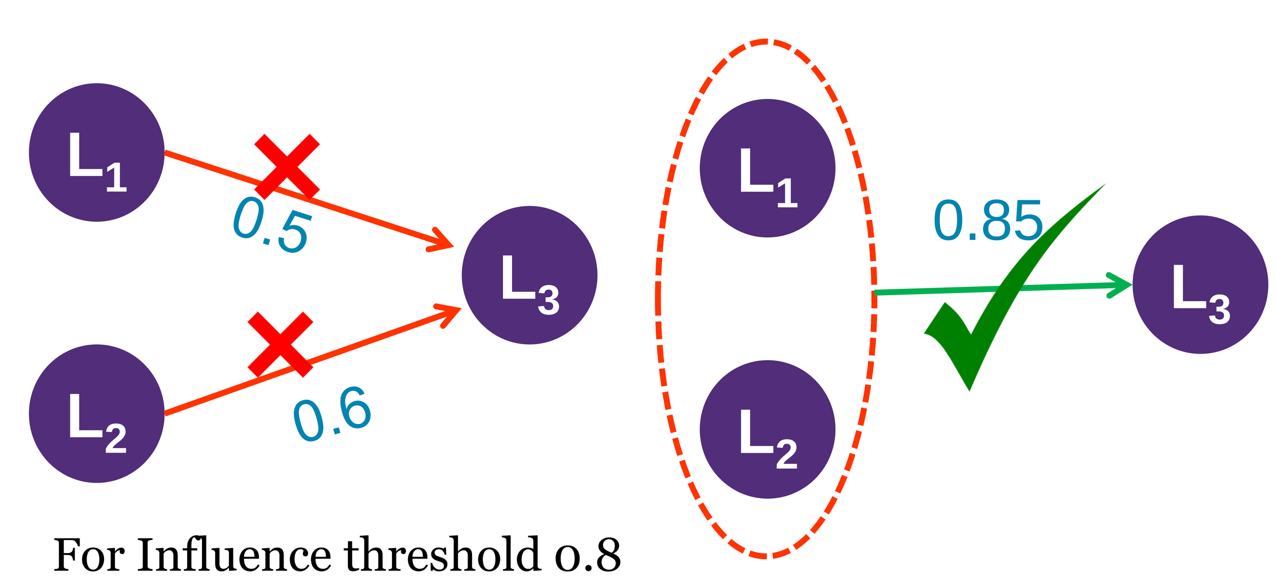
b. Most followed ≤ 8hr.





5. Influence Maximization

Combined Influence



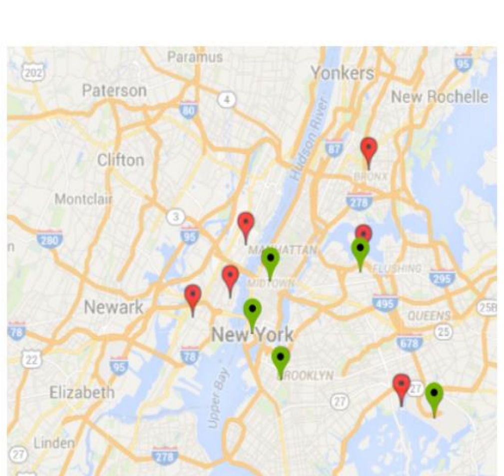
For Influence threshold 0.8

Top K location: Greedy approach to combine location with highest potential.

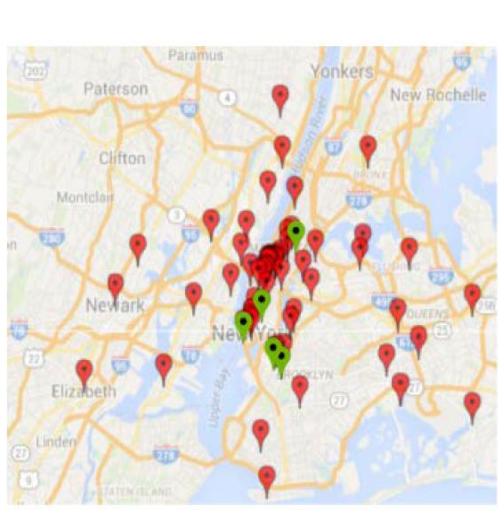
$$LI(S) = (1-\alpha) \times |\phi(S)| + (\alpha) \times \sum_{d \in L-S} (\min\{|V_B(S,d)|, \tau\})$$

6. Result: Influence spread

More than 4 times better spread!



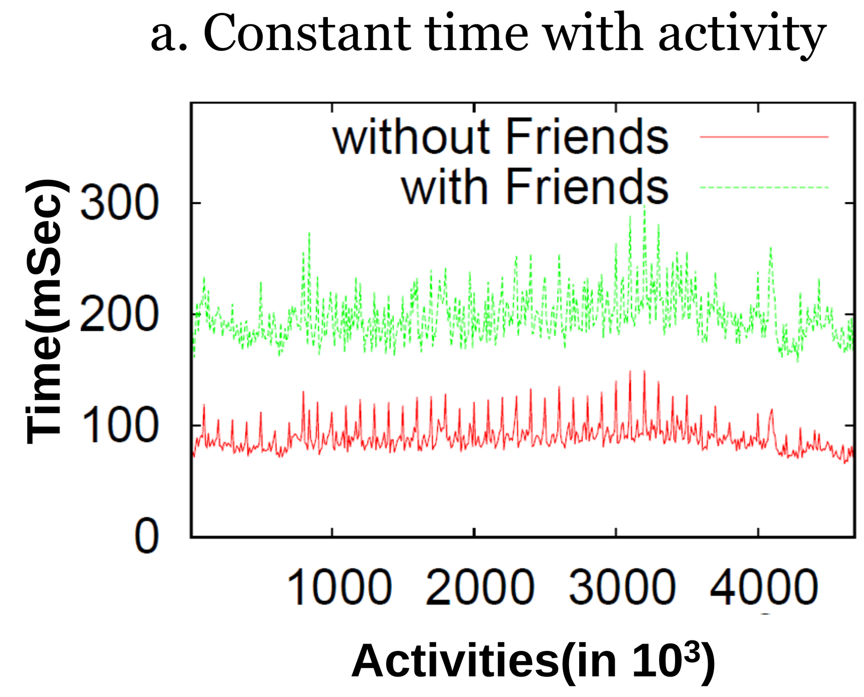
Naive BrightKite
(16 locations)



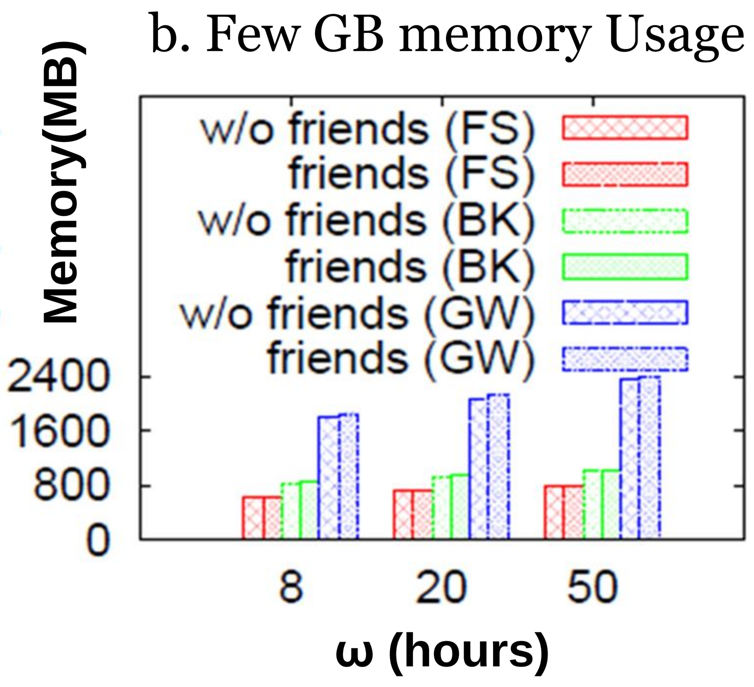
Our BrightKite
(72 locations)

7. Result : Scalability w.r.t computation time and memory requirements

a. Constant time with activity



b. Few GB memory Usage



c. Constant α=0.02 for all dataset