



A Study of Brightkite

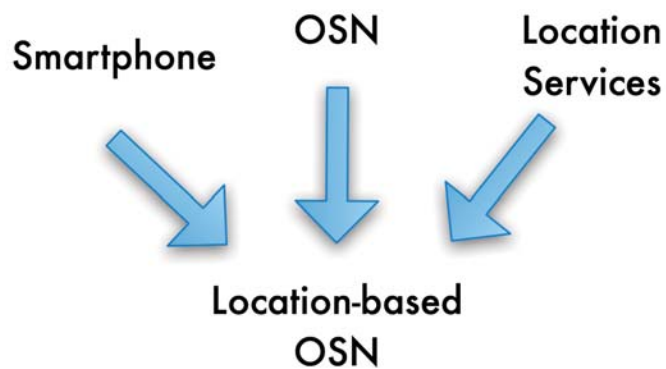
connecting people through their visited places

Guanling Chen
University of Massachusetts Lowell

glchen@cs.uml.edu

6/14/2010, WPI
Workshop on Opportunistic RF Localization for Next
Generation Wireless Devices

Emerging LSN



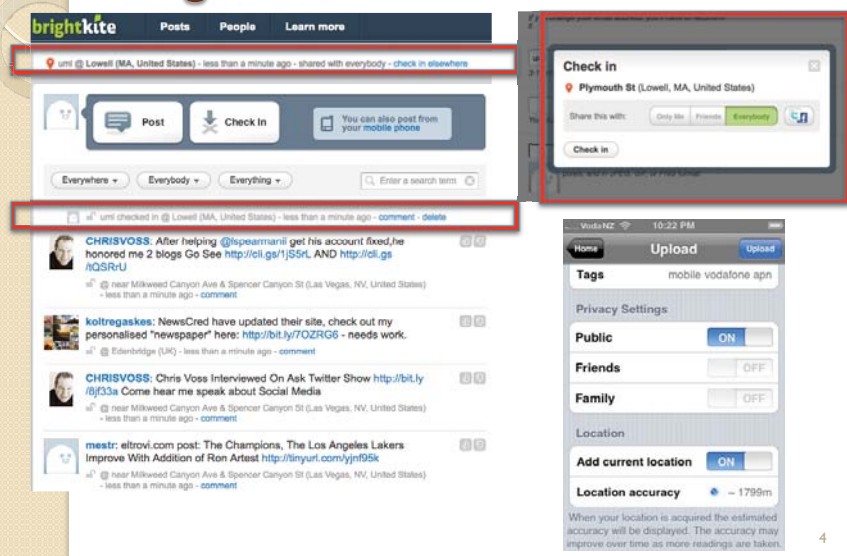


LSN Features

- Users sharing their location
- Discover users and content based on location
- Check-in models
 - Foursquare, Gowalla, Brightkite
- Tracking models
 - Loopt, Latitude

3

Brightkite



4



Measurement

- Data Collection from Brightkite
 - About two years updates 4,460,161
 - Checkin update 42.3%
 - Majority Scope is address level 56.2%
 - Daily friend lists 168 days
 - User profiles 70,337
 - Male 57.7%
 - Female 15.2%

5

Tag Cloud

apple art beer biking blog blogger books business
christian coffee computers css design
developer entrepreneur film food games geek
graphic guitar internet iphone linux mac
marketing media mobile movies
music networking photographer photography
php podcast politics programming reading rock san social software tech
technology travel twitter video web writer writing

6



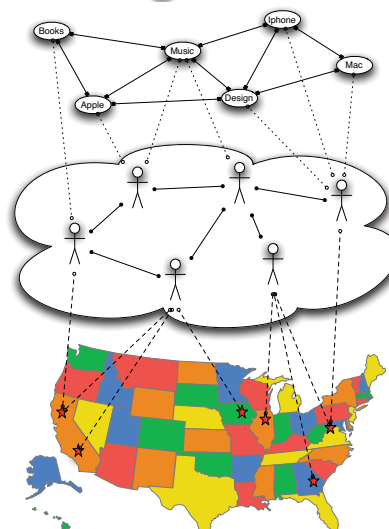
User Classification

Cluster	Prop %	Total Updates	Active Days	Loyalty	Unique Places	Friends Number	PoPU
average	N/A	95.6	27.0	0.25	16.3	7.35	0.09
trial	41	2.8	1.8	0.00	1.7	2.11	0
temporary	24	15.9	8.3	0.42	6.0	2.50	0.09
normal	23	72.0	32.7	0.35	20.9	9.58	0.12
active	11	658.3	153.3	0.70	85.6	32.90	0.15
privacy	2	35.3	3.1	0.01	2.5	3.01	0.45

7

Friendship Modeling

- Three-layer
 - Tag layer -- interests
 - Social graph layer
 - Location layer



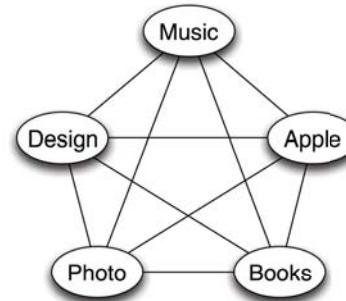
8



Tag Distance

- Top 1000 tags

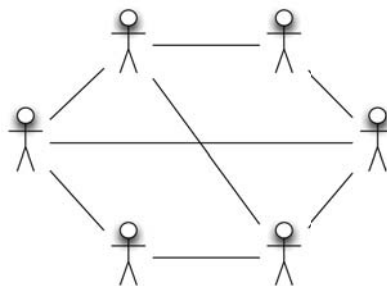
$$m_{t(i,j)} = \sum_{n=1}^N \sum_{m=1}^M w_{t_{im}, t_{jn}}$$



9

Social Distance

$$DS_{i,j} = \text{geodesic_distance}(v_i, v_j) \text{ in } G'_s(V_s, E_s - e_{i,j})$$



10

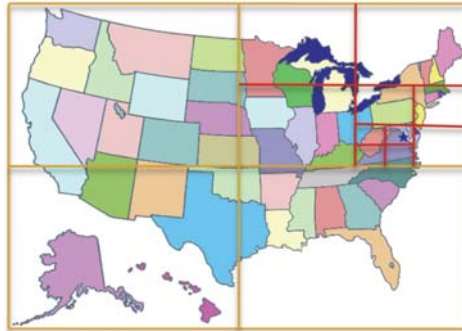


Location Distance

- Grid-based Location Metric

$$m_l(i, j) = \sum_n \text{sng}(c_{in}) \times \text{sng}(c_{jn}) \times (c_{in} + c_{jn})$$

- 36,269 cities
- Thresh hold
50,000



11

Evaluation Results

Multi-layer	NaïveBayes	J48 Tree
0.906	0.882	0.869

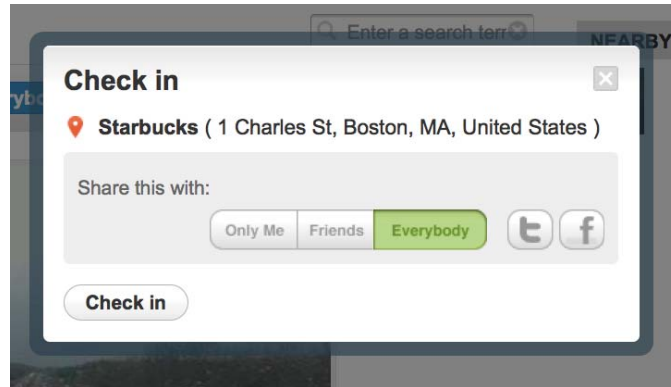
Top ^a	Friend Pairs Number and Percentage					
	Multi-layer		NaiveBayes		J48 Tree	
10,000	8,672	86.7%	6,498	65.0%	8,380	83.8%
5,000	4,804	96.1%	3,277	65.5%	4,406	88.1%
1,000	982	98.2%	639	63.9%	905	90.5%
500	487	97.4%	324	64.8%	449	89.8%
100	98	98.0%	68	68.0%	90	90.0%
50	49	98.0%	37	74.0%	48	96.0%

Attribute s	Social Dist.	Loc. Metric	Tag Metric	Gender Diff	Age Diff
Info. Gain	0.549	0.405	0.034	0.020	0.011

12



Privacy Protection



13

Privacy Metrics

- Proportion of protected updates (PoPU)

$$PoPU = \frac{N_p}{N_t}$$

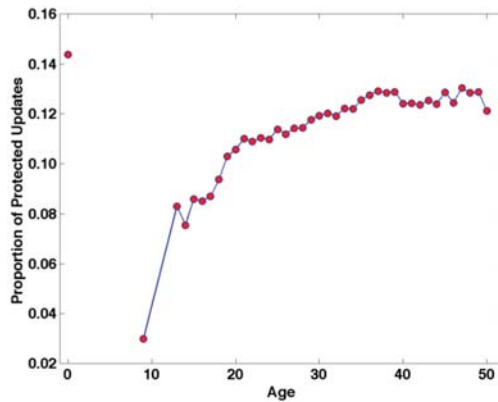
- Scopes
 - address, street, zip+4, zip, city, state, country

14



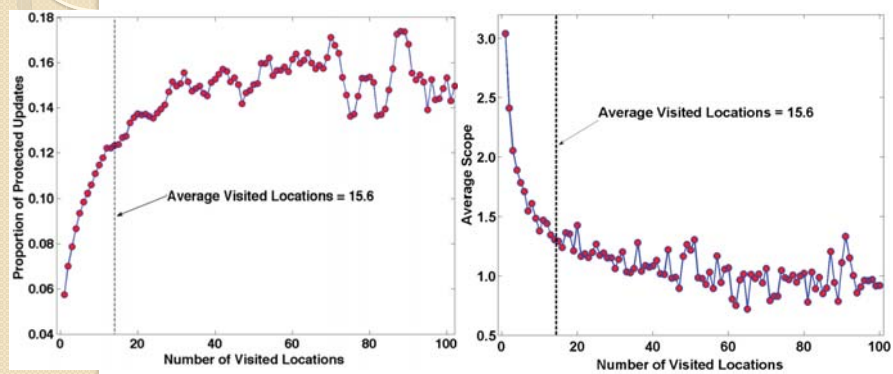
Gender and Age

	Male	Female	Unspecified
Proportion (%)	57.7	15.2	27.1
Average PoPU	0.1226	0.1502	0.1464



15

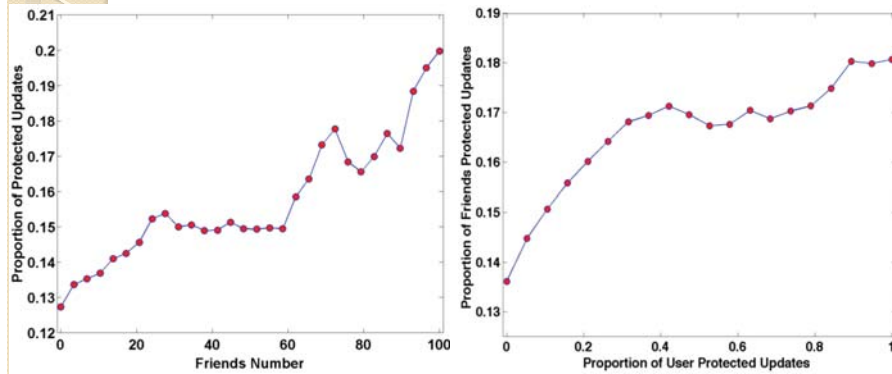
Mobility



16

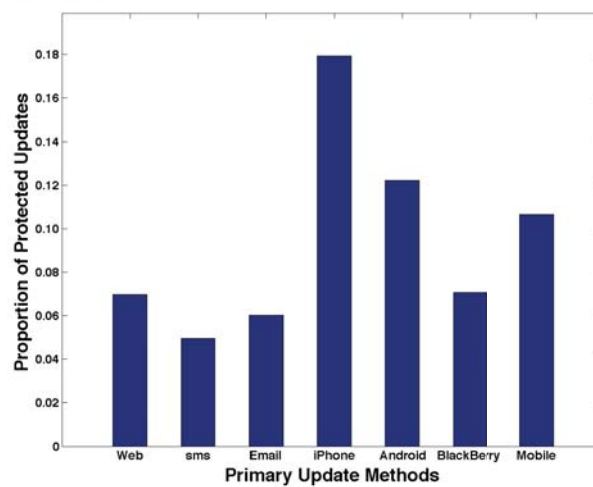


Friends



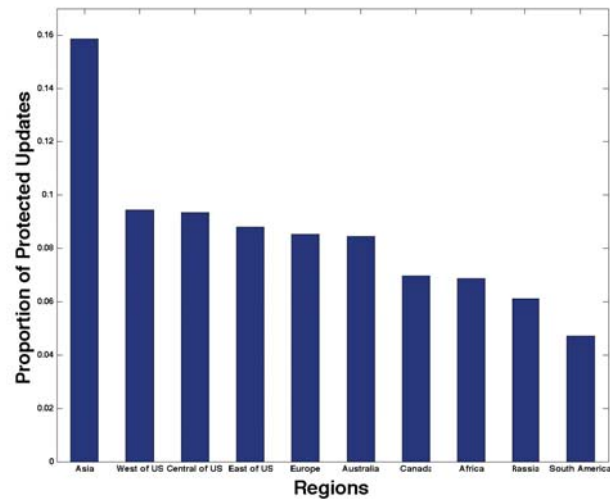
17

Sources

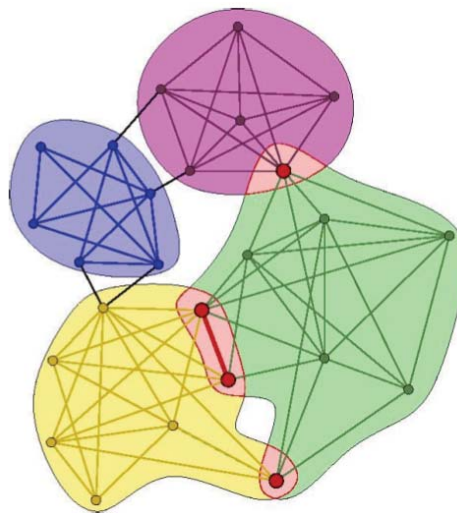


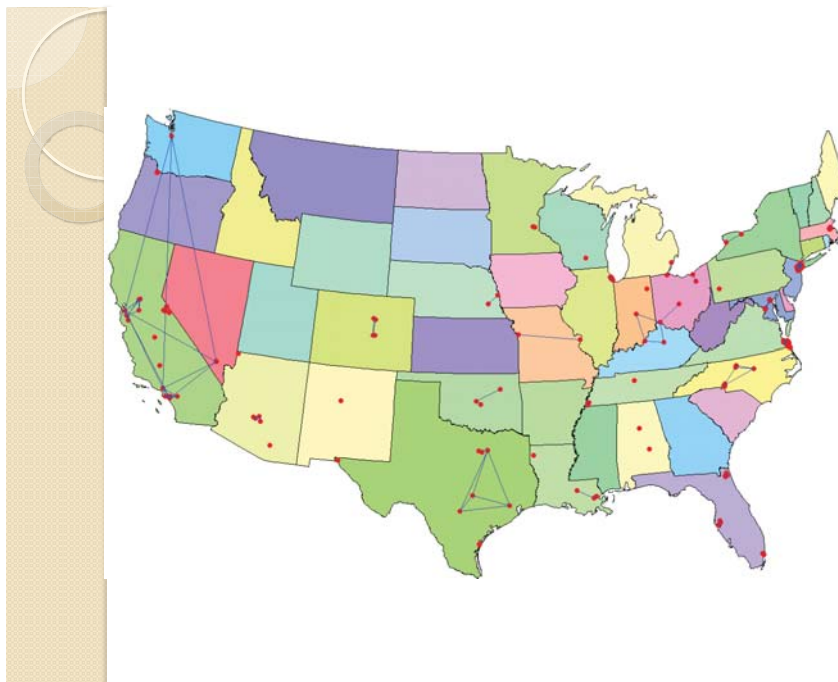


Regions



Community Analysis





More Information

- <http://www.cs.uml.edu/~glchen/>
- glchen@cs.uml.edu
- PS. Nan Li is looking for jobs.