Modeling User Mobility for Location Promotion in Location-based Social Networks

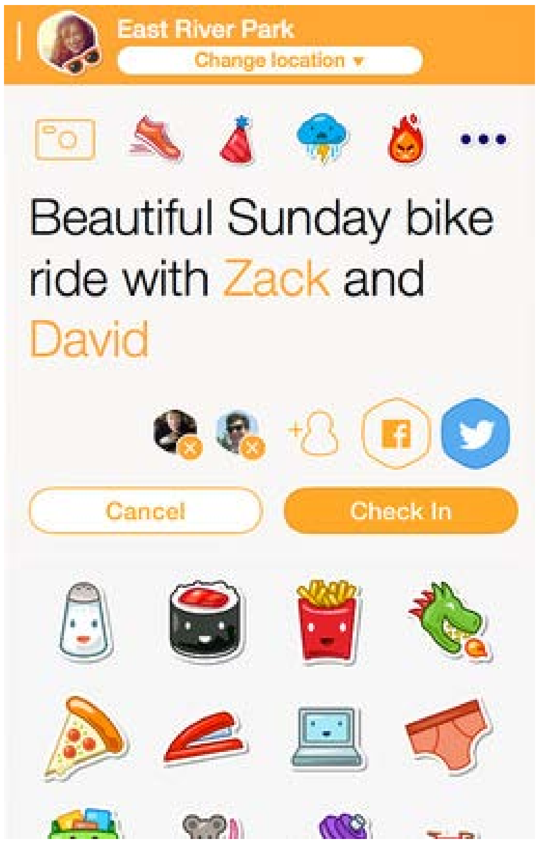
Wen-Yuan Zhu, Wen-Chih Peng, Ling-Jyh Chen, Kai Zheng and Xiaofang Zhou

National Chiao Tung University, Taiwan

Academia Sinica, Taiwan

The University of Queensland, Australia

Location-based Social Networks (LBSNs)



check-in at a location

friends activity

Foursquare

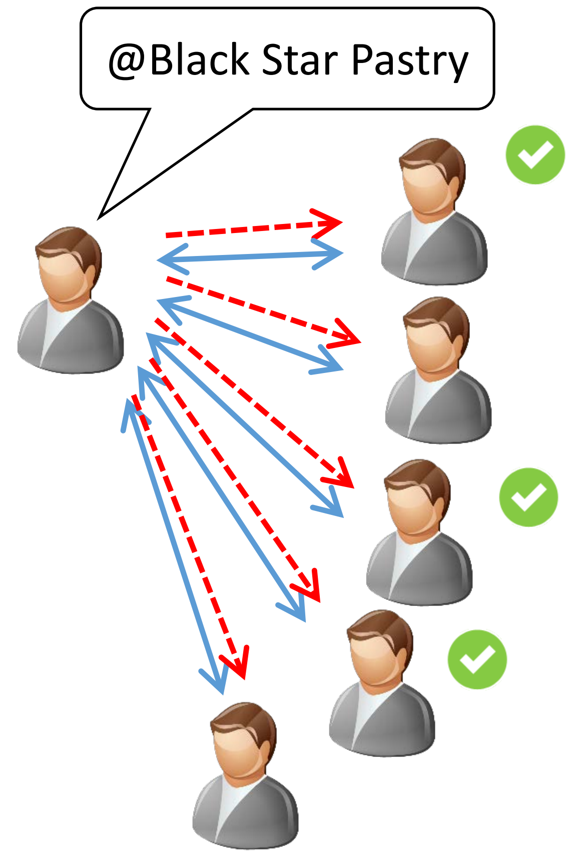
facebook

Twitter

Figures in this page are from App Store

Location Promotion in LBSNs

- Exploring check-in data to promote one location and attract users to visit

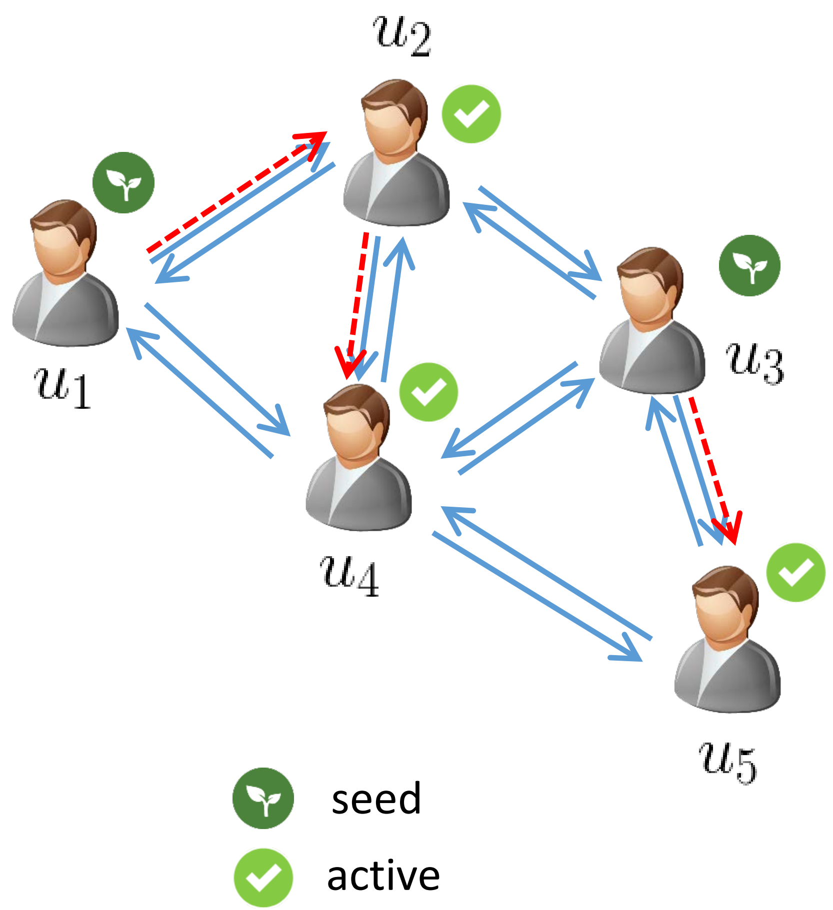


http://blackstarpastry.com.au/
http://www.howsocial.co.uk/2012/05/smallbusinesstip1-checkingin/

Promote Location via Viral Marketing

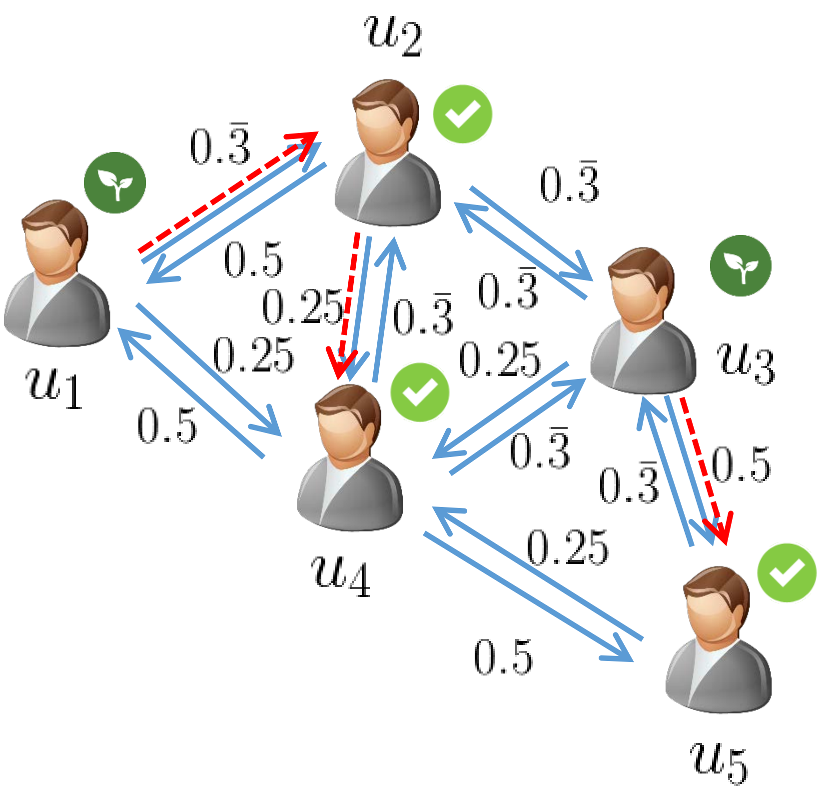
- Viral Marketing
  - Users are more likely to accept their friends' messages
- Influence maximization
  - Select *k* seeds to maximize expected # of active users (influence spread, *σ*(*S*))

$$\arg\max_{S,|S|=k} \sigma(S)$$



Promote Location via Viral Marketing (2)

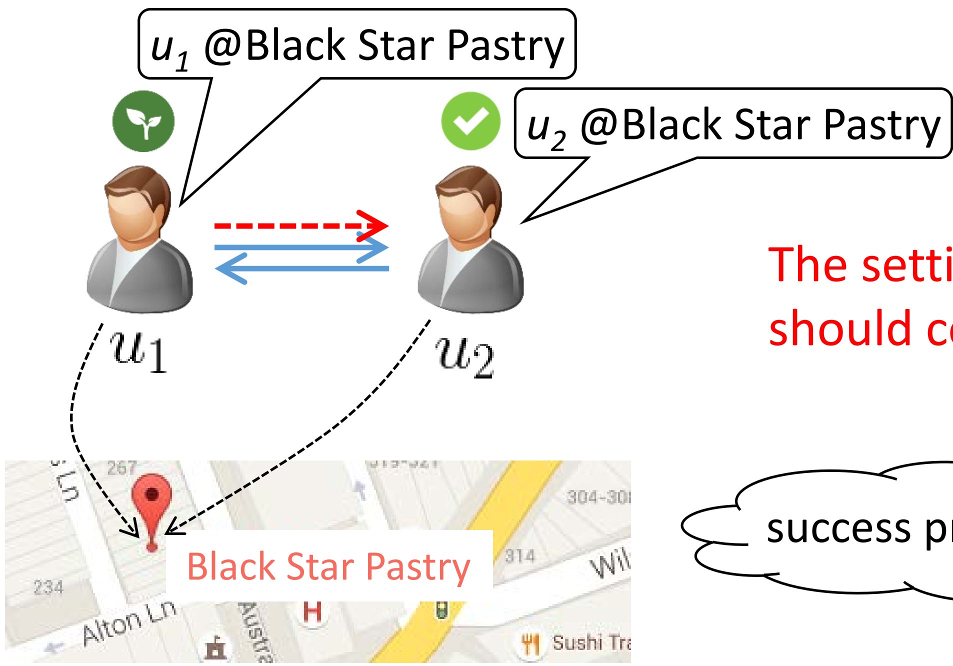
- Propagation probability
  - *pp*(*u*, *v*): the prob. of *v* becoming active influenced by *u*
  - It reflects the propagation in a social network



One issue is how to determine the propagation probability in LBSNs

Propagation in LBSNs

- Information propagate is triggered by check-in in LBSNs
  - $pp(u_1, u_2 | \textit{Black Star Pastry})$
    $= \mathrm{P}(u_2 \text{ check-in at } \textit{Black Star Pastry})$
    $= \mathrm{P}(\textit{Black Star Pastry} | u_2)$

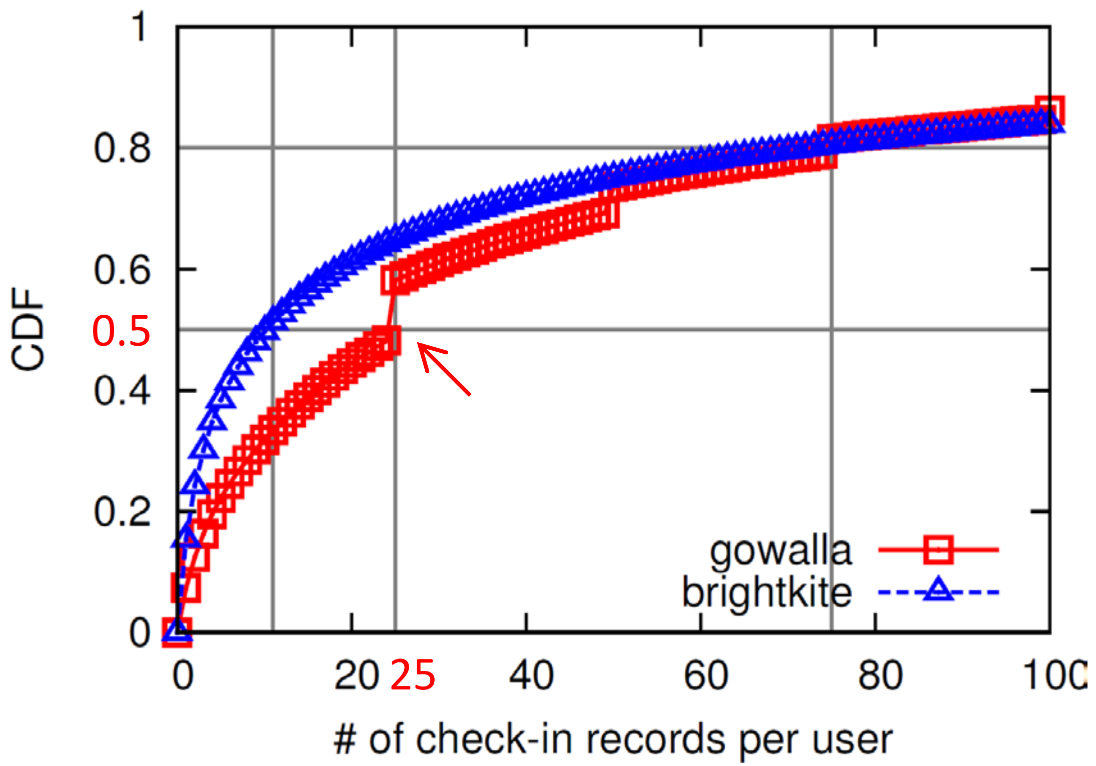


The setting of propagation probability in LBSNs should consider user's check-in behavior

success propagation

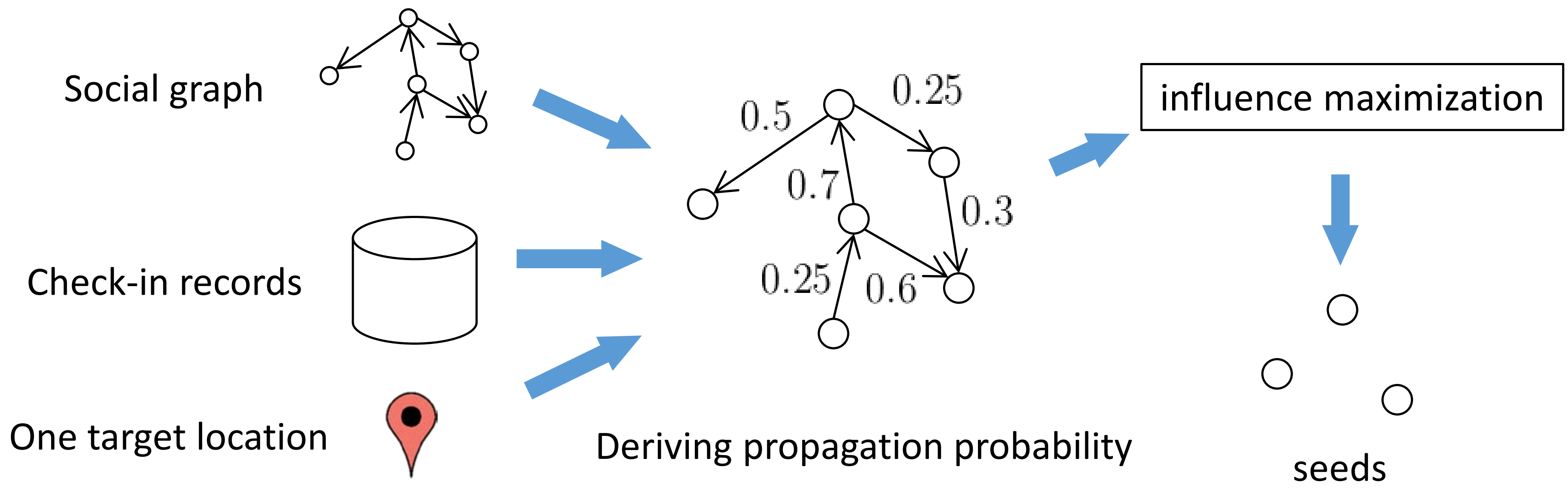
Challenge Issue

- Infer whether a user check-in at the target location
  - About 50%/80% users have 25/75 check-in records or less



Location Promotion Problem

- Given an LBSN and a target location $\ell=(x,y)$, select k seeds to maximize the expected # of users who will check-in at ℓ



Contribution

- Formulate location promotion problem in an LBSN as an influence maximization problem
- Propose Gaussian-based mobility models (GMMs) and distance-based mobility models (DMMs) to capture individual check-in behavior
- Derive the propagation probability based on GMMs and DMMs

Static Approaches

- Some common static approaches
  - Non-history related
    - `Uniform`: Uniformly random from [0,1]
    - `Trivalency`: Uniformly random from {0.1, 0.01, 0.001}
    - `Jaccard_F`: Jaccard of friends
    - `Jaccard_L`: Jaccard of locations
    - `In-degree`: $pp(u, v)$=1/indeg(v)
  - History related
    - `Bernoulli`: $pp(u, v)$=proportion of success influence attempts

[2] P. Bouros et al. Regionally influential users in location-aware social networks. GIS 14
[4] W. Chen et al. Scalable influence maximization for prevalent viral marketing in large-scale social networks. KDD 10
[9] A. Goyal et al. Learning influence probabilities in social networks. WSDM 10
[13] D. Kempe et al. Maximizing the spread of influence through a social network. KDD 03

Gaussian-based Approach

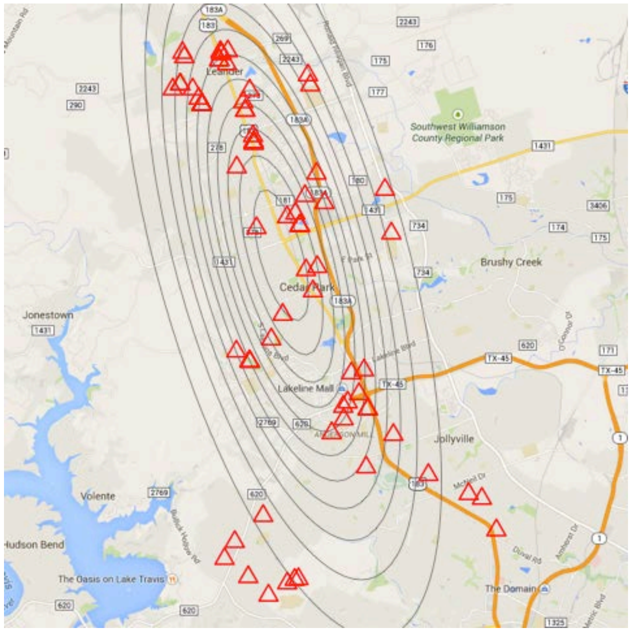
- Gaussian distribution is selected to capture individual check-in behavior
  - GMM-Basic
    - One Gaussian distribution
  - GMM-Spatial
    - Mixture Gaussian model
  - GMM-Temporal
    - Mixture Gaussian model, where each point is classified on temporal domain

$$p(\ell|u) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\ell-\mu)^T\Sigma^{-1}(\ell-\mu)}$$

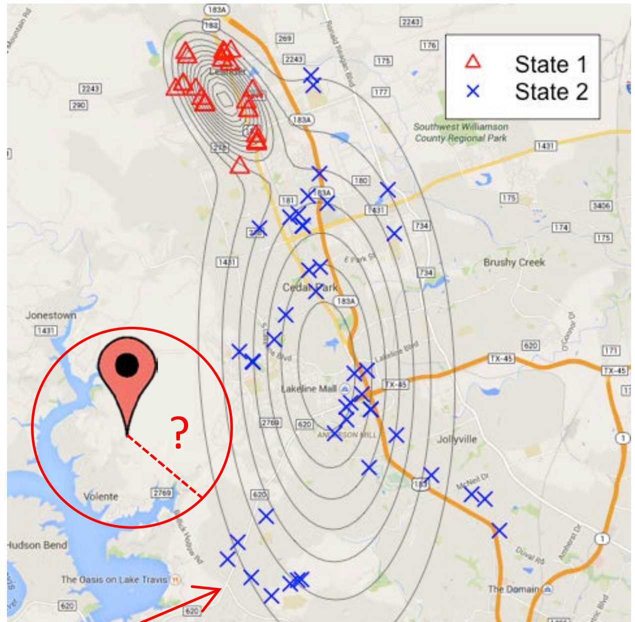
$$p(\ell|u) = \sum_{i=1}^{K} p(\ell|u, Z_i)p(u, Z_i)$$

[5] E. Cho et al. Friendship and mobility: user movement in location-based social networks. KDD 11
[6] H. Gao et al. Modeling temporal effects of human mobile behavior on location-based social networks. CIKM 13
[17] M. Lichman et al. Modeling human location data with mixtures of kernel densities. KDD 14

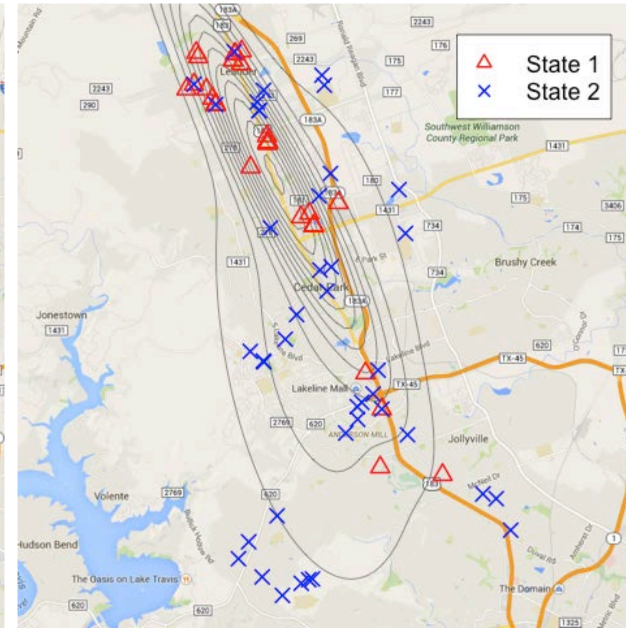
Gaussian-based Approach (2)



GMM-Basic

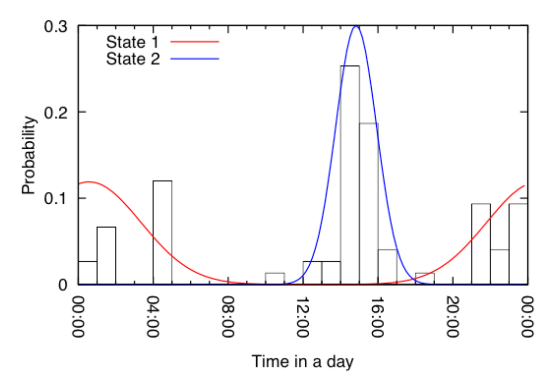


GMM-Spatial



GMM-Temporal

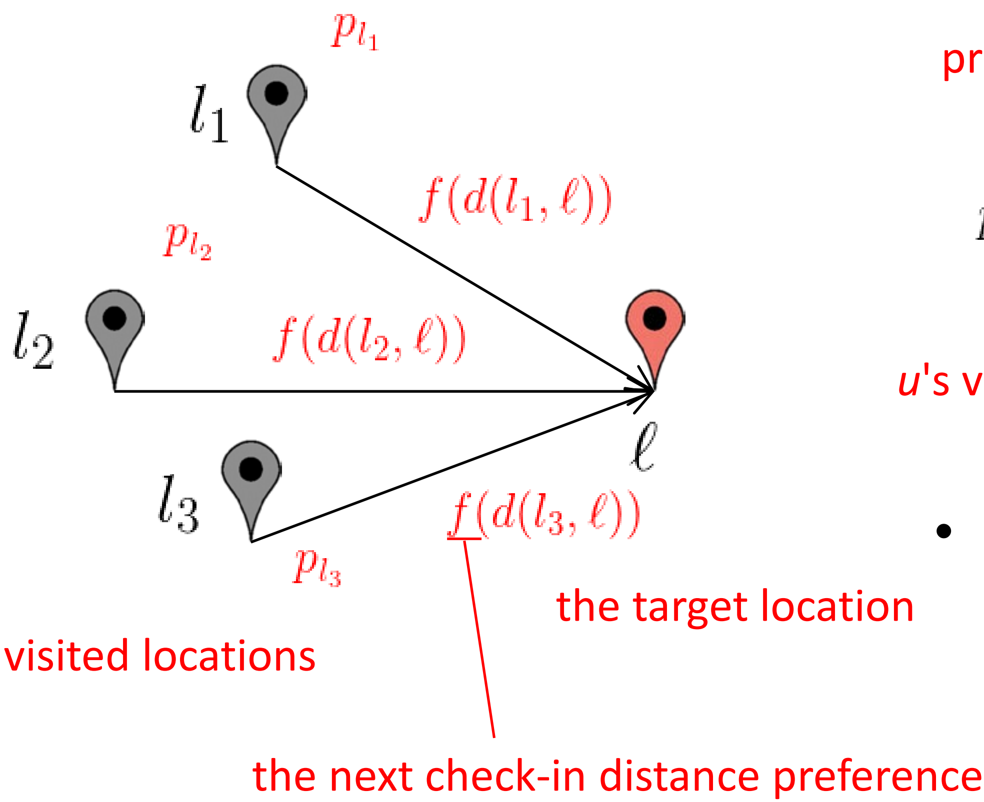
- Rough description
- Order independent
- Hard to calculate the probability of visiting a location



Gaussian mixture model, *K*=2

Distance-based Approach

- Distance-based mobility model (DMM) is a distribution of location ℓ to capture individual check-in behavior



prob. of *u* at *l*

distance between *l* and ℓ

$$p(\ell|u) = \sum_{l} p_l^{(u)} f^{(u)}(d(l, \ell))$$

u's visited location

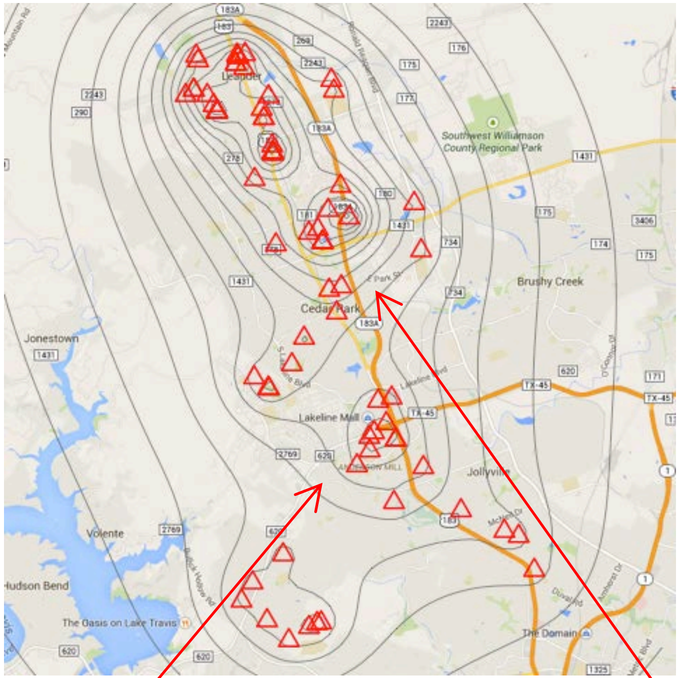
pdf of movement distance

- Select Pareto distribution for *f*

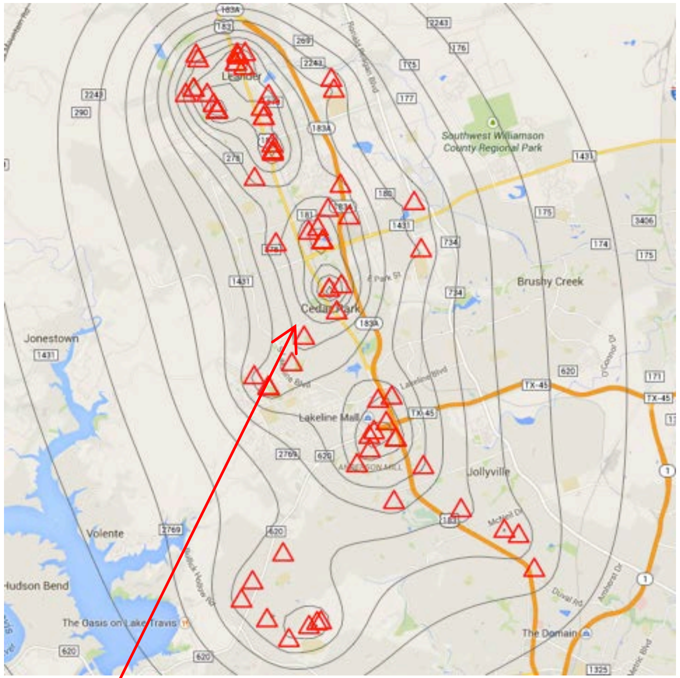
$$f(x; \alpha) = \alpha/(x+1)^{\alpha+1}, x \in [0, \infty)$$

$$\text{where } \hat{a} = N / \sum_{i} \ln(x_i + 1)$$

Distance-based Approach (2)



DMM-Basic (1)



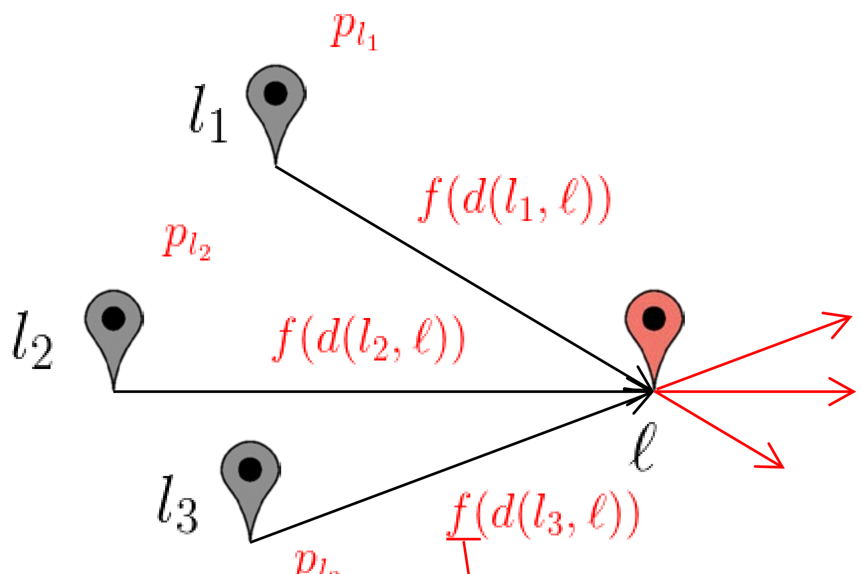
DMM-Basic (2)

more detailed description

different ordering, different behavior

Distance-based Approach (3)

- The propagation prob. represents the user prefers to move farther than the selected location



$$P(\ell|u) = \sum_l p_l^{(u)} \int_{d(l,\ell)+1}^{\infty} f^{(u)}(x)\mathrm{d}x$$

$$= \sum_l \frac{p_l^{(u)}}{(d(l,\ell)+1)^{\alpha}}$$

scale parameter of Pareto distr.

the target location

visited locations

the next check-in distance preference

Experiments

- Two real datasets in SNAP

	gowalla	**brightkite**
# of users	196,591	58,228
# of check-ins	6,442,890	4,491,143
# of edges	950,327	214,078
Period	Feb. 2009-Oct. 2010	Apr. 2008-Oct. 2010

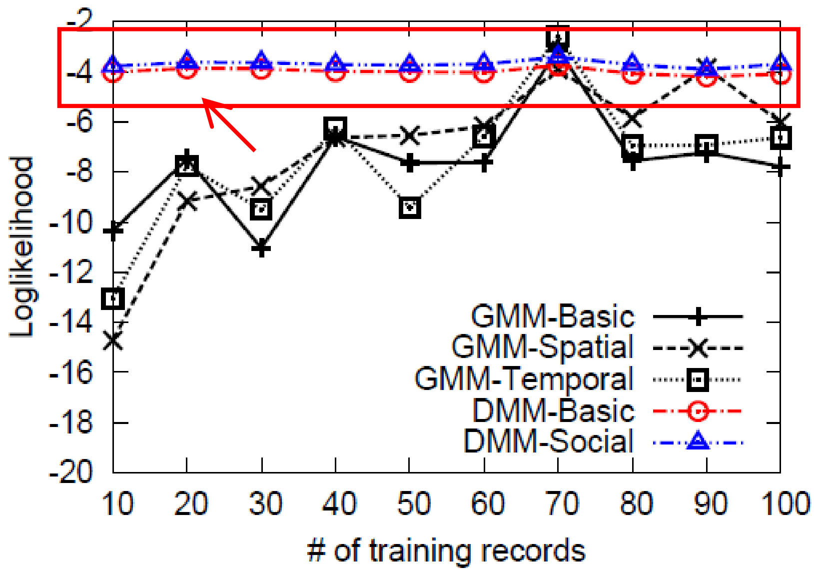
https://snap.stanford.edu/data/

Different Mobility Models

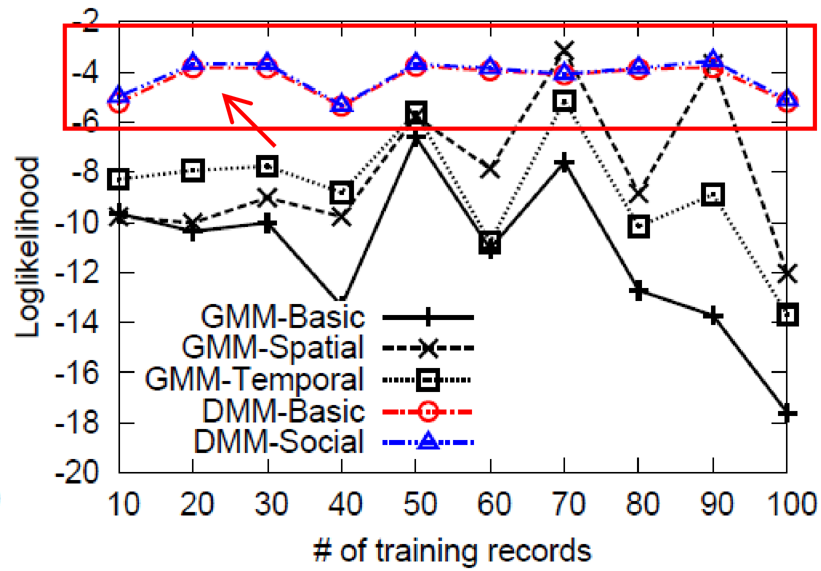
- Setting
  - 80% training, 20% testing
- Comparisons
  - `GMM-Basic`, `GMM-Spatial`, `GMM-Temporal`
  - `DMM-Basic`, `DMM-Social`
- Metric
  - Loglikihood

Different Mobility Models (2)

- Loglikelihood vs different training sizes



gowalla



brightkite

DMMs have higher performance than GMMs

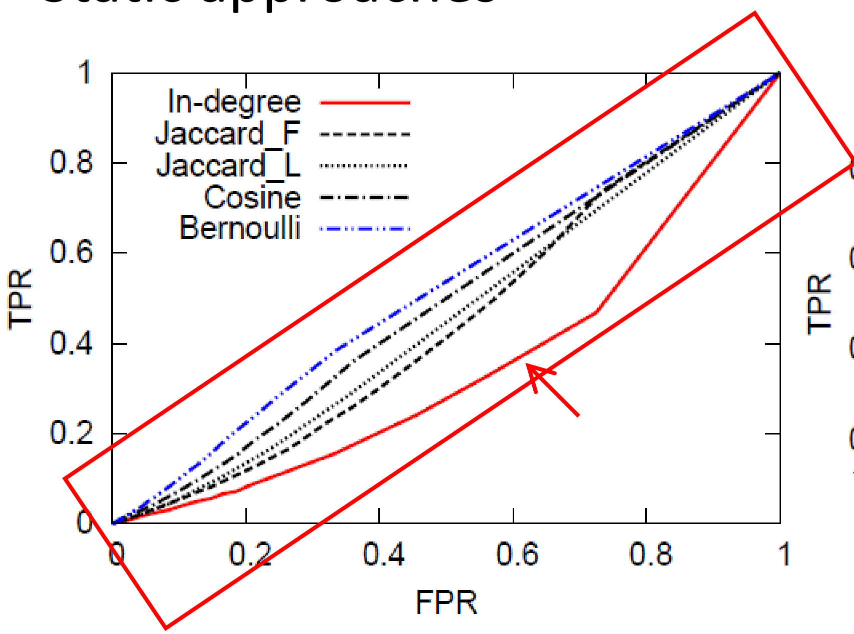
DMMs also have higher performance even when there are only 10 training records

Propagation Probability

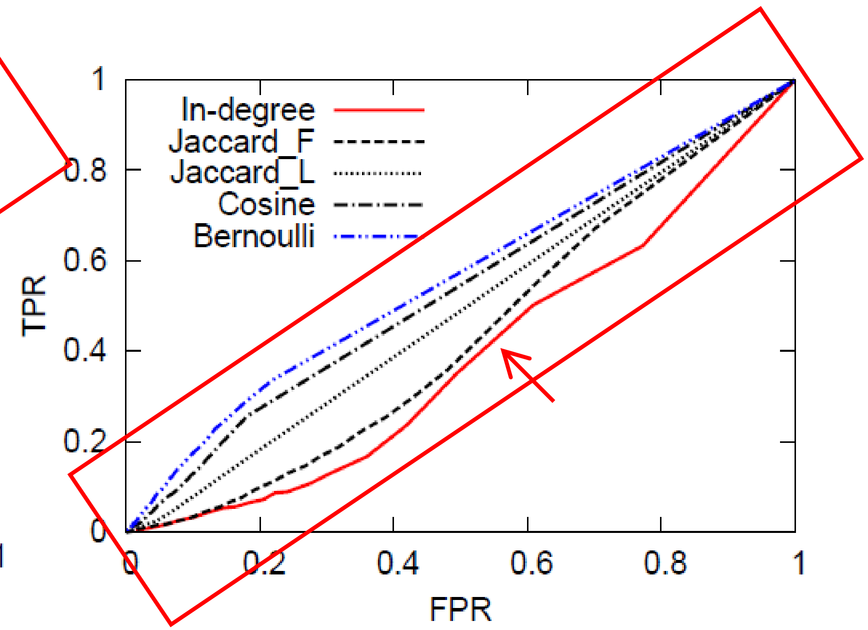
- Setting
  - Users: the users who have 10 check-in records or above
  - The target location: Center Park, NYC
  - Active range: 500m
- Comparisons
  - Static: `In-degree`, `Jaccard_F`, `Jaccard_L`, `Cosine`, `Bernoulli`
  - Location-aware: `GMM-Basic`, `GMM-Spatial`, `GMM-Temporal`, `DMM-Basic`, `DMM-Social`
- Metric
  - ROC curve/AUC

Propagation Probability (2)

- Static approaches



gowalla



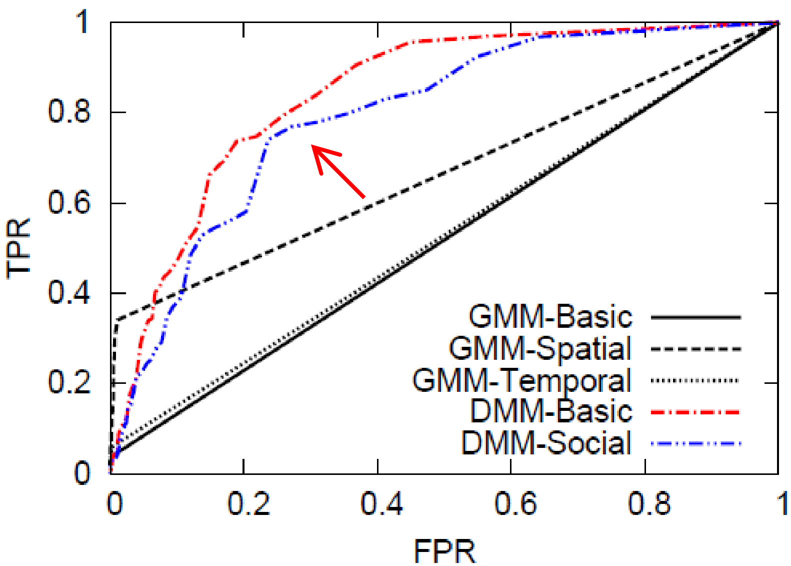
brightkite

static approaches can not reflect real information propagation in LBSNs

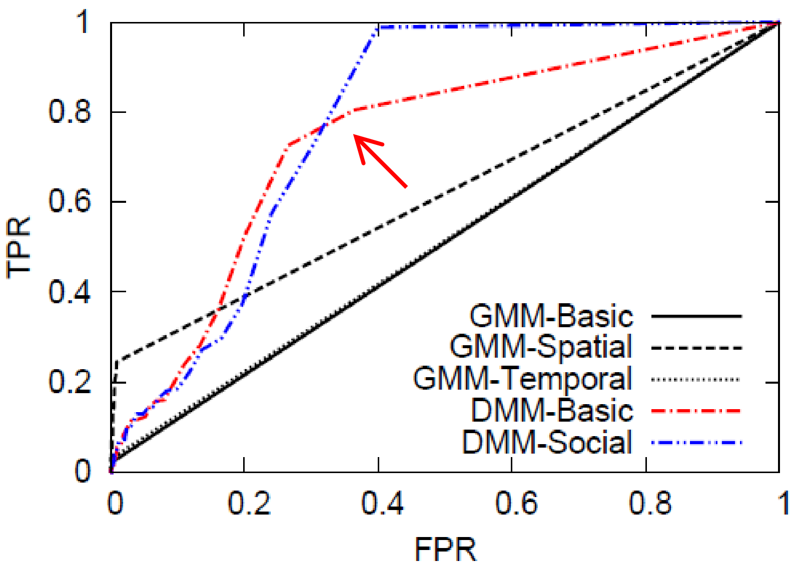
the most common setting, In-degree, has the lowest AUC

Propagation Probability (3)

- Location-aware approaches



gowalla



brightkite

DMMs can truly reflect the information propagation in LBSNs

Conclusion

- Formulate the location promotion problem as influence maximization in LBSNs
  - Propose that the propagation probability is location-aware in LBSNs
  - Propose GMMs and DMMs to capture individual check-in behavior
  - Utilize GMMs and DMMs to derive the propagation probability in LBSNs

Thank You!