Friendship and Mobility: User Movement in Location-Based Social Networks

Eunjoon Cho, Setha A. Myers, Jure Leskovec

Stanford University

Introduction

- Human movement/mobility high degree of freedom and variation
- Also exhibit structural patterns due to geographic and social constraints
- Understanding what basic laws govern human motion and dynamics
  - Cell phone location data (trace of 2 million phone users from Euro country)
  - Two online location-based social networks (Gowalla and Brightkite)

Motivation

- There has been a lack of data to understand this
- Cell phone data provides coarse location accuracy
- Location-based social networks (e.g., check-ins via Foursquare, Facebook, etc.) provide pinpoint location data
- Understand the following
  - Geographic movement (where do we move?)
  - Temporal dynamics (how often do we move?)
  - Social network (how do social ties interact with movement?)
- Role of geography and daily routine on mobility patterns as well as the effect of social ties

Gowalla

- Public check-in data between Feb. 2009 and Oct. 2010
- 6.4 million check-ins
- Explicit social network, friendships undirected
- 196,591 nodes, 950,327 edges

brightkite

- Public check-in data between Apr. 2008 and Oct. 2010
- 3.5 million check-ins
- Explicit social network, friendships directed, though paper considers them as undirected by only considering bi-directional edges
- 58,228 nodes, 214,078 edges

Cell phone data

- Two million users in European country
- 450 million phone calls over 455 days
- Nearest cell phone tower of both people on call recorded
- 900 million "check-ins" with spatial accuracy of 3km
- Social network tie between people that called each other at least five times
    - 2 million nodes, 4.5 million edges

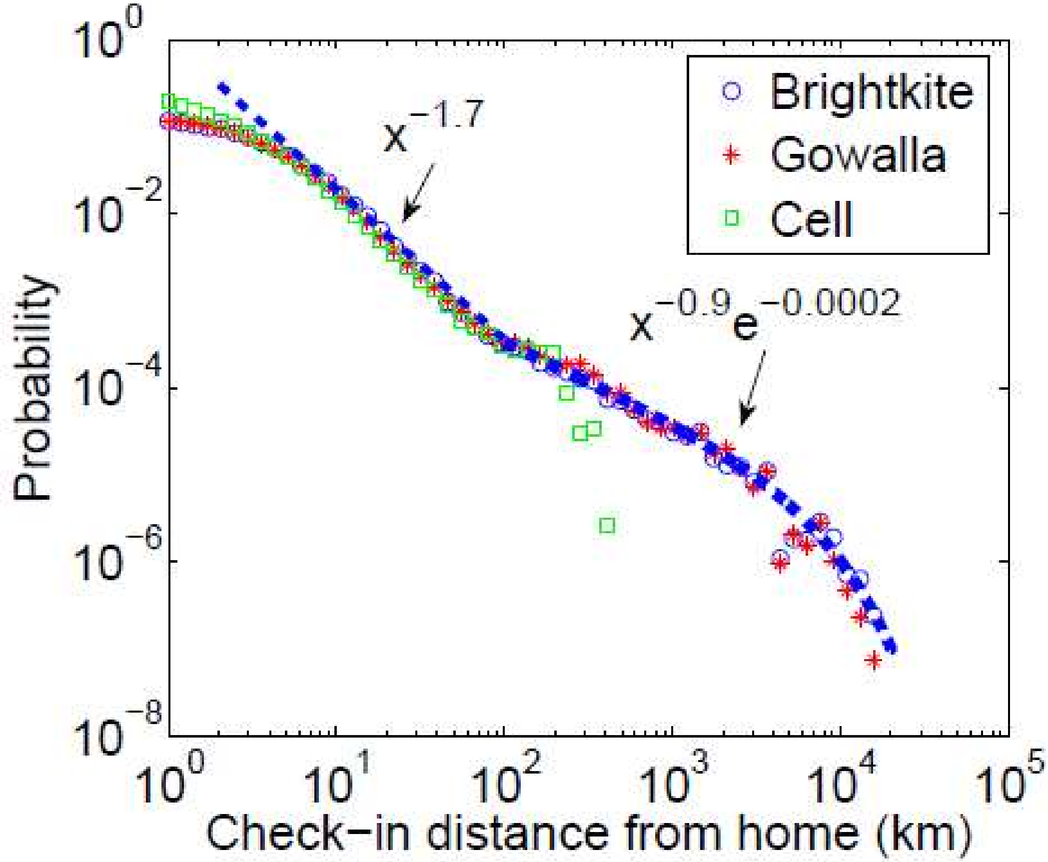


Figure 1: Fraction of check-ins as a function of distance traveled from home. Note the change in slope at around 100km.

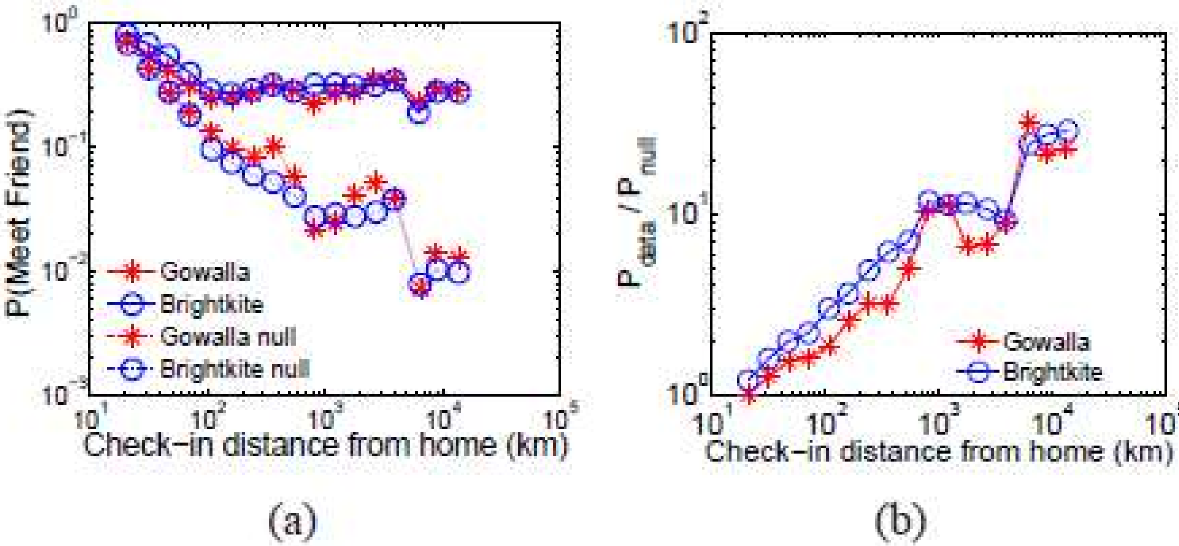


Figure 3: (a) Probability that a user will travel to friend's home as a function of distance traveled. (b) Influence of a friend relative to the null model.

Methods

Predicting geographic locations

Periodic Mobility Model (PMM)

$x_u(t)$ denote the geographic position of user u at time t.

$c_u(t)$ be the "state" at time t.

$c_u(t) = H$ denotes that the user is in the "home" state at time t.

$c_u(t) = W$ indicates the user is in the "work" state

(a) Spatial model

$$P[x(t) = x] = P[x_u(t) = x | c_u(t) = H] \cdot P[c_u(t) = H] \\ + P[x_u(t) = x | c_u(t) = W] \cdot P[c_u(t) = W]$$

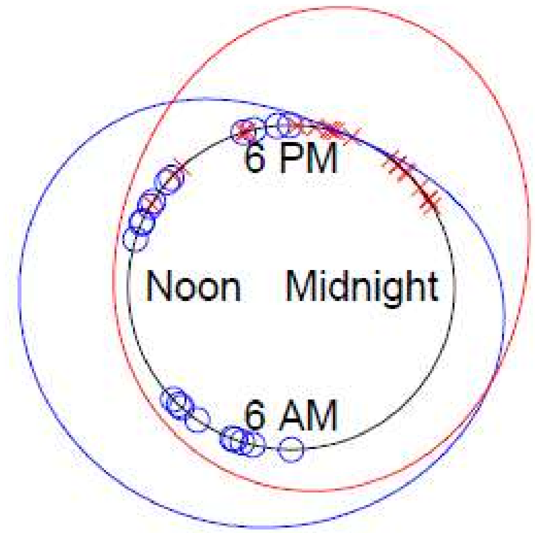
Temporal component

τ_H is the average time of the day
σ_H is the variance in time of day,
P_{c_H} is the time-independent probability

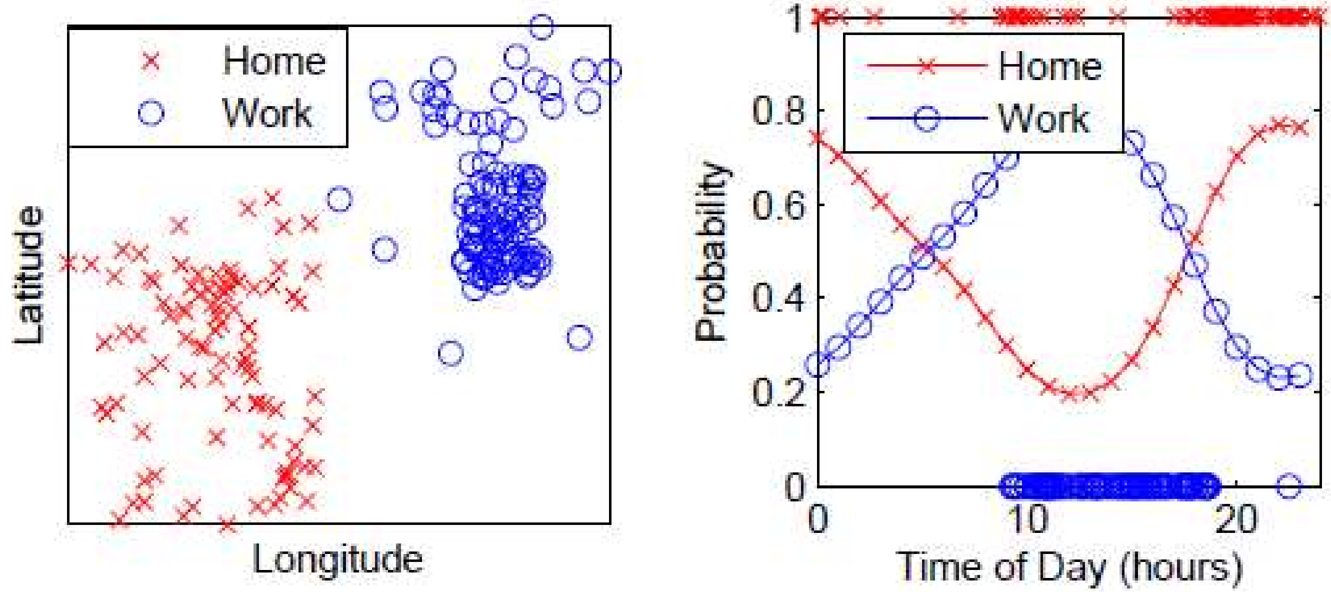
$$N_H(t) = \frac{P_{c_H}}{\sqrt{2\pi\sigma_H^2}} \exp\left[-\left(\frac{\pi}{12}\right)^2 \frac{(t-\tau_H)^2}{2\sigma_H^2}\right]$$
$$N_W(t) = \frac{P_{c_W}}{\sqrt{2\pi\sigma_W^2}} \exp\left[-\left(\frac{\pi}{12}\right)^2 \frac{(t-\tau_W)^2}{2\sigma_W^2}\right]$$

and then

$$P\left[c_u(t) = H\right] = \frac{N_H(t)}{N_H(t) + N_W(t)}$$
$$P\left[c_u(t) = W\right] = \frac{N_W(t)}{N_H(t) + N_W(t)}$$



(b) Temporal model



(a) Location of user check-ins (b) Home/work state

Figure 7: ***Periodic Mobility Model*****. (a) Check-in locations generated by home/work state. (b) State distribution over time.**

Spatial component

$$P\left[x_u(t) = x_i | c_u(t)\right] = \begin{cases} \sim \mathcal{N}(\mu_H, \Sigma_H) & \text{if } \; c_u(t) = H \\ \sim \mathcal{N}(\mu_W, \Sigma_W) & \text{if } \; c_u(t) = W \end{cases}$$

Means of user's check-in location in home and work states

Home and work check-in position covariance matrices

What does this mean for PMM?

- Two-state mixture of Gaussians with time-dependent state prior
- Temporal aspect governs transition between states
- Depending on state, geo location of check-in is generated

Periodic and Social Mobility Model (PSMM)

$z_u(t) = 1$ implies the check-in is social (non-periodic)
$z_u(t) = 0$ implies that it is periodic.

$$\boxed{\begin{aligned} P\left[x(t) = x\right] = & P\left[x_u(t) = x | c_u(t) = H\right] \cdot P\left[c_u(t) = H\right] \\ & + P\left[x_u(t) = x | c_u(t) = W\right] \cdot P\left[c_u(t) = W\right] \end{aligned}}$$

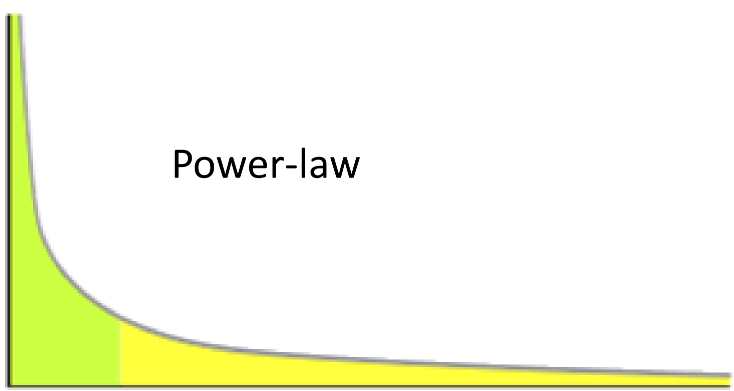
$$\begin{aligned} P_u[x(t) = x] = & P[x(t) = x | z_u(t) = 1] \cdot P[z_u(t) = 1] \\ & + \boxed{P[x(t) = x | z_u(t) = 0]} \cdot P[z_u(t) = 0] \end{aligned}$$

where $P[x(t) = x | z_u(t) = 0]$ is the *Periodic Mobility Model*.

$$P[x_u(t) = x_i | z(t) = 1] \sim \sum_{\boxed{(t_j, x_j)} \in \boxed{J_u}} |t_j - t|^{-\alpha} \cdot ||x_i - x_j||^{-\beta}$$

Time and location of *j*-th check-in

Set of check-ins by user *u*'s friends made on same day



Fitting model

- Expectation-Maximization
- PMM
  - 4 param. Temporal model
  - 12 for spatial model
  - 2 for social model (time and distance decay)
- PSMM
  - Train PMM
  - Classify as home, work, or outlier (social) check-in
  - Fit PSMM on outliers

Experimental Setup

- Users with at least 10 check-ins on each day of the week
- 80/20 split
- 7 models for each user, one for each day of the week

Baseline Models

- Most Frequent Location Model
  - Fraction of previous check-ins during that hour at that location
- Gaussian Model
  - Stochastic process centered around a single point
- RW model
  - Next location predicted to be location of last known check-in

Evaluation

- Average Log-likelihood of check-ins
- Accuracy
  - Exact location
- Expected distance Error
  - Soft version of accuracy
  - Spatial proximity of prediction to actual check-ins

Results

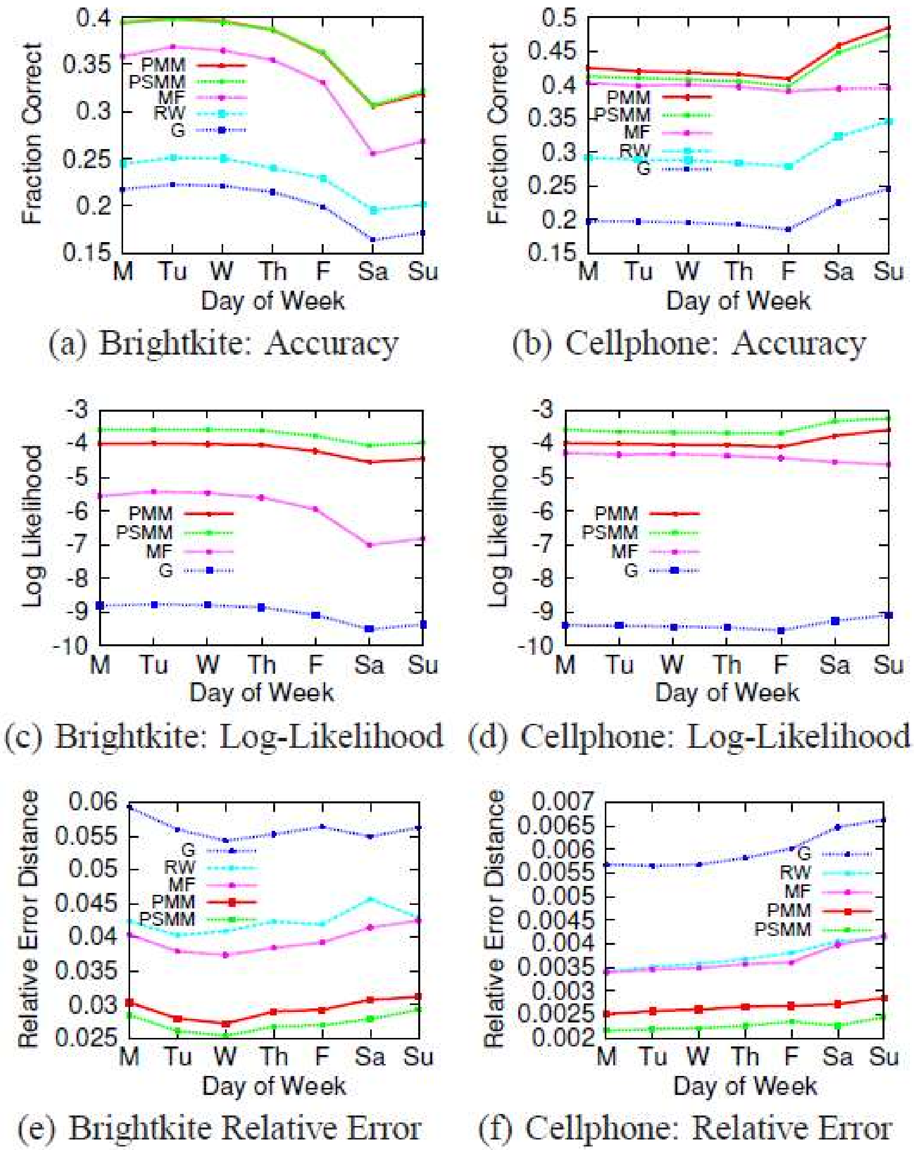


Figure 10: Performance of the *Periodic Mobility Model* (*PMM*) and the *Periodic & Social Mobility Model* (*PSMM*), compared to three baseline models. (a,b) Accuracy of check-in location prediction; (c,d) Log-likelihood of check-ins in the test set; (e,f) Expected Distance Error of predicted check-in location.

Conclusion

- Even though cell phone data and LBSN are different data, common patterns of human mobility found
- Social network structure has little impact on short range spatially and temporally periodic movement
- Long-distance travel more influenced by the social network ties
- Model reliably captures and predicts human mobility patterns
- Future work could include using more than two latent states