Limits of Human Mobility

Jihui Nie
11/9/2017
IS LIFE RANDOM?
Human Mobility Pattern

Research subject
50,000 anonymized mobile phone users

Criteria
› Visit more than two tower vicinity during observational 3 months
› Average call frequency $f \geq 0.5 \text{ hour}^{-1}$
Entropy

- Random Entropy $S_{i}^{\text{rand}}$
- Temporal-Uncorrelated Entropy $S_{i}^{\text{unc}}$
- Actual Entropy $S_{i}$
Random Entropy

\[ S_{i}^{\text{rand}} \equiv \log_2 N_i \]

Number of distinct locations visited by user \( i \)
Temporal-Uncorrelated Entropy

\[ S_{i}^{\text{unc}} \equiv -\sum_{j=1}^{N} p_{i}(j) \cdot \log_{2} p_{i}(j) \]

\[ p_{i}(j) \quad \text{Historical probability that location} \ j \ \text{was visited by user} \ i \]
Actual Entropy \( S_i \equiv - \sum_{T_i' \subset T_i} P(T_i') \cdot \log_2[P(T_i')] \)

\( T_i = \{ X_1, X_1, \ldots, X_L \} \)

The sequence of towers at which user \( i \) was observed at each consecutive hourly interval

\( P(T_i') \) The probability of finding a particular time-ordered subsequence \( T_i' \) in \( T_i \)
Incompleteness ($q$)

The fraction of hour-long intervals when the user’s location is unknown to us
Extrapolation from $S_i(q)$ to $S_i(0)$ works well for $q < 0.8$

Retain 45,000 out of 50,000 users
Entropy

- **Random Entropy**
  The varieties of the visited locations

- **Temporal-Uncorrelated Entropy**
  The heterogeneity of visitation frequencies

- **Actual Entropy**
  The full spatiotemporal order

\[ S_{i}^{\text{rand}} \equiv \log_{2} N_{i} \]

\[ S_{i}^{\text{unc}} \equiv -\sum_{j=1}^{N} p_{i}(j) \cdot \log_{2} p_{i}(j) \]

\[ S_{i} \equiv -\sum_{T_{i}' \subset T_{i}} P(T_{i}') \cdot \log_{2}[P(T_{i}')] \]
$S_i \leq S_i^{\text{unc}} \leq S_i^{\text{rand}}$
\[ S = -\Pi^\text{max} \log_2 \Pi^\text{max} - (1 - \Pi^\text{max}) \log_2 (1 - \Pi^\text{max}) + (1 - \Pi^\text{max}) \log_2 (N - 1) \]

\[(S, N) \rightarrow \Pi^\text{max} \quad \text{Upper limit of predictability}\]

93% potential predictability in human mobility
Regularity (R)

The probability of finding the user in his/her most visited location during each hour

\[ \langle R \rangle \approx 0.7 \]

\[ \langle R_{\text{rand}} \rangle = \frac{1}{\langle N \rangle} \approx 0.016 \]

Life is far from random
Are long-distance travelers more difficult to predict?

No!
Does gender matter?

- Age
- Rural/Urban
- Weekday/Weekend
- Population density
- Gender
- Language
THANK YOU