RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

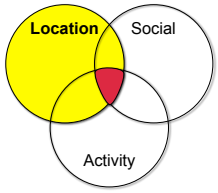
MINING DATA FROM MOBILE DEVICES

Algorithms: Location & Context

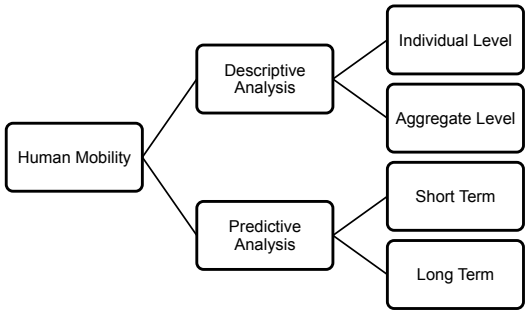
Spiros Papadimitriou, *Tina Eliassi-Rad*
{spiros.papadimtriou, tina.eliassi}@rutgers.edu

Location & Context

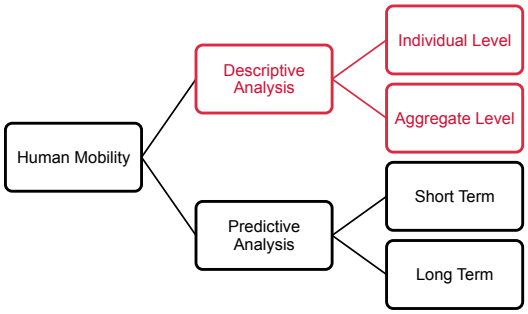
- **Location/Local**
  - What resources are nearby?
  - Where are you?
- **Context: Social**
  - Who are you with?
- **Context: Activity**
  - What are you doing?



Work on human mobility

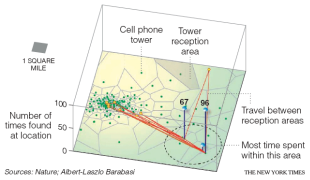


Work in human mobility



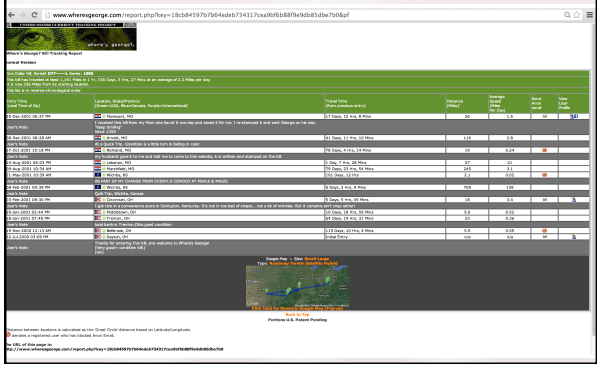
Human mobility at the individual level (descriptive)

- Human trajectories are not random!
- They have high degree of temporal and spatial regularity
- Individual humans follow simple reproducible patterns
  - **Exploration + preferential return**
- Impact: epidemic prevention, emergency response, urban planning, ...



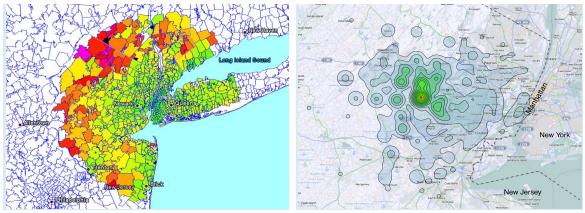
Marta Gonzalez, Cesar Hidalgo, Albert-Laszlo Barabasi: Understanding individual human mobility patterns. Nature 453, 779-782, 2008.

http://www.wheresgeorge.com/



Human mobility at the aggregate level (descriptive)

- Billions of anonymized Call Detail Records (CDRs) from a cellular network
- Characterized **daily travel**, **carbon emissions**, **number of workers** and **event goers**, and **traffic volumes** of hundreds of thousands of people

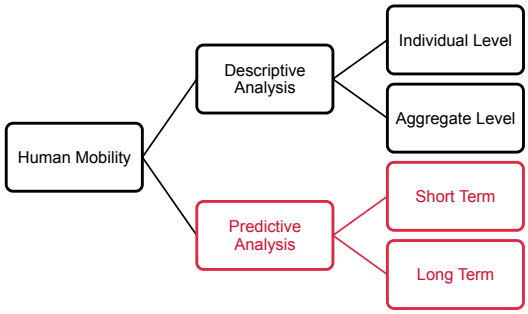


Richard A. Becker, Ramon Caceres, Karrie Hanson, Sibren Isaacman, Ji Meng Loh, Margaret Martonosi, James Rowland, Simon Urbanek, Alexander Varshavsky, Chris Volinsky: Human mobility characterization from cellular network data. Commun. ACM 56(1): 74-82 (2013).

Human mobility (descriptive)

- M. Kim, D. Kotz, S. Kim: Extracting a mobility model from real user traces. In InfoCom 2006:1-13
- K. Lee, S. Hong, S. Kim, I. Rhee, S. Chong: Slaw: A new mobility model for human walks. In InfoCom 2009, 855-863
- Z. Li, B. Ding, J. Han, R. Kays, P. Nye: Mining periodic behaviors for moving objects. In KDD 2010, 1099-1108
- M. Martino, F. Calabrese, G. Di Lorenzo, C. Andris, L. Liu, C. Ratti. Ocean of Information: Fusing Aggregate & Individual Dynamics for Metropolitan Analysis. IUI 2010: 357-360
- M. Kim, D. Kotz, D: Identifying unusual days. Journal of Computing Science and Engineering 5(1), 2011:71-84
- …

Work in human mobility



Human mobility at the individual level (predictive; short-term; GPS)

- **"Where are you going to be in the next hour?"**
  - Successful techniques: **hidden Markov models**, **random walk based formalisms**
  - Performance around **3-5 km off**; classification **accuracy low 90%**
  - Learning from GPS alone
    - D. Ashbrook, T. Starner: Using GPS to learn significant locations and predict movement across multiple users. Personal Ubiquitous Comput. 7, 2003:275-286.
    - L. Liao, D. Fox, H. Kautz: Location-based activity recognition using relational Markov networks. In IJCAI 2005.
    - J. Krumm, E. Horvitz: Predestination: Inferring destinations from partial trajectories. In UbiComp 2006: 243-260.
    - B. Ziebart, A. Maas, A. Dey, J. Bagnell: Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In UbiComp 2008:322-331.
    - B. Shaw, J. Shea, S. Sinha, A. Hogue. Learning to Rank for Spatiotemporal Search. WSDM 2013: 717-726.
    - …

Human mobility at the individual level (predictive; long-term; GPS)

- **"Where are you going to be 285 days from now at 2PM?"**
- FarOut
  - Identifies periodicity via Fourier analysis (mapping time to frequency)
  - Uses PCA for pattern extraction
  - Utilizes PCA-based classification
- Data: 32K days worth of GPS data across 703 subjects (⅓ people; ⅓ cars)
- High variance in area across subjects
  - From 30 to more than 10^8 km^2
  - Surface area of earth = 5.2 × 10^8 km^2
- Number of contiguous days = 7 to 1247 (μ = 45.9; σ = 117.8)
- Captures both continuous (raw GPS) and discretized (triangular cells) data
- Each subject has a matrix D, where each row is a day.

Figure 3: Our continuous vector representation of a day ***d*** consists of the median latitude and longitude for each hour of the day (00:00 through 23:59), binary encoding of the day of week, and a binary feature signifying whether a national holiday falls on ***d***.

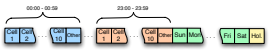


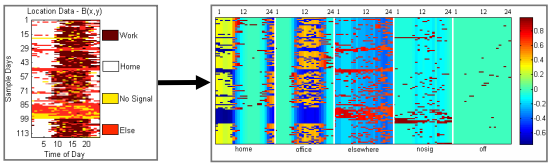
Figure 4: Our cell-based vector representation of a day ***d*** encodes the probability distribution over dominant cells conditioned on the time within ***d***, and the same day-of-week and holiday information as the continuous representation (last 8 elements).

- Performance **continuous** rep. : **1 km off**; baseline 2.5km off
- Performance **discrete** rep: **80% accuracy up to 80 weeks into the future**; baseline ~60%

Adam Sadilek and John Krumm: Far Out: Predicting Long-Term Human Mobility. AAAI 2012.

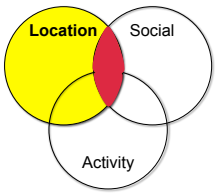
Human mobility at the individual level (predictive; long-term but < 24 hours)

- N. Eagle and A. Pentland: Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology & Sociobiology* 63(7), 2009:1057-1066
  - Predictions up to 12 hours into the future
  - Class labels: {Home, Elsewhere, Work, No Signal, Off}.
  - PCA based classification
  - 79% accuracy



Context includes…

- **Location/Local**
  - What resources are nearby?
  - Where are you?
- **Social**
  - Who are you with?
- **Activity**
  - What are you doing?



Location & Social

1. Nathan Eagle and Alex (Sandy) Pentland. 2006. Reality mining: sensing complex social systems. Personal Ubiquitous Comput. 10, 4 (March 2006), 255-268.
2. Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. KDD 2011, 1082-1090.
3. Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, Cecilia Mascolo: Socio-spatial properties of online location-based social networks. ICWSM 2011.
4. Salvatore Scellato, Anastasios Noulas, Cecilia Mascolo: Exploiting place features in link prediction on location-based social networks. KDD 2011: 1046-1054.
5. Yu Zheng: Location-based social networks - users. In Computing with Spatial Trajectories, Chapter 8, Eds: Yu Zheng, Xiaofang Zhou, Springer, 2011.
6. Yu Zheng, Xing Xie: Location-based social networks - location. In Computing with Spatial Trajectories, Chapter 9, Eds: Yu Zheng, Xiaofang Zhou, Springer, 2011.
7. Huiji Gao, Jiliang Tang, Huan Liu: gSCorr: Modeling geo-social correlations for new check-ins on location-based social networks. CIKM 2012: 1582-1586.
8. Huiji Gao, Jiliang Tang, Huan Liu: Exploring social-historical ties on location-based social networks. ICWSM 2012.
9. Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, Cecilia Mascolo: Where online friends meet: Social communities in location-based networks. ICWSM 2012.
10. Miltiadis Allamanis, Salvatore Scellato, Cecilia Mascolo: Evolution of a location-based online social network: analysis and models. Internet Measurement Conference 2012: 145-158.
11. M. Domenico, A. Lima, and M. Musolesi. Interdependence and predictability of human mobility and social interactions. In *Nokia Mobile Data Challenge: http://research.nokia.com/page/12000*, MDC 2012.
12. P.A. Grabowicz, J.J. Ramasco, B. Goncalves, V.M. Eguiluz. Entangling Mobility and Interactions in Social Media. PLOS One 9(3):e92196. March 2014.
13. J. Toole, C. Herrera-Yaque, C.M. Schneider, M.C. Gonzalez. Coupling Human Mobility and Social Ties. In arXiv: 1502.00690v1, February 2015.
14. …

Location & Social

1. Nathan Eagle and Alex (Sandy) Pentland. 2006. Reality mining: sensing complex social systems. Personal Ubiquitous Comput. 10, 4 (March 2006), 255-268.
2. Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. KDD 2011, 1082-1090.
3. Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, Cecilia Mascolo: Socio-spatial properties of online location-ba
4. Salvatore S … location-based soci
5. Yu Zheng: … Eds: Yu Zheng, Xia
6. Yu Zheng, … Chapter 9, Eds: Yu Zh
7. Huiji Gao, … on-based social netw
8. Huiji Gao, … WSM 2012.
9. Chloë Brow … e friends meet: Soci
10. Miltiadis Al … twork: analysis and models. Internet Measurement Conference 2012: 145-158.
11. M. Domenico, A. Lima, and M. Musolesi. Interdependence and predictability of human mobility and social interactions. In *Nokia Mobile Data Challenge: http://research.nokia.com/page/12000*, MDC 2012.
12. P.A. Grabowicz, J.J. Ramasco, B. Goncalves, V.M. Eguiluz. Entangling Mobility and Interactions in Social Media. PLOS One 9(3):e92196. March 2014.
13. J. Toole, C. Herrera-Yaque, C.M. Schneider, M.C. Gonzalez. Coupling Human Mobility and Social Ties. In arXiv: 1502.00690v1, February 2015.
14. …

- Long distance movements are influenced by ties in the social network.
- This is not true for short-range movements or temporally periodic movements.

Location & Social

1. Nathan Eagle and Alex (Sandy) Pentland. 2006. Reality mining: sensing complex social systems. Personal Ubiquitous Comput. 10, 4 (March 2006), 255-268.
2. Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. KDD 2011, 1082-1090.
3. Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, Cecilia Mascolo: Socio-spatial properties of online location-based social networks. ICWSM 2011.
4. Salvatore Scellato, Anastasios Noulas, Cecilia Mascolo: Exploiting place features in link prediction on location-based social networks. KDD 2011: 1046-1054.
5. Yu Zheng: Location-based social networks - users. In Computing with Spatial Trajectories, Chapter 8, Eds: Yu Zheng, Xiaofang Zhou, Springer, 2011.
6. Yu Zheng, Xing Xie: Location-based social networks - location. In Computing with Spatial Trajectories, Chapter 9, Eds: Yu Zheng, Xiaofang Zhou, Springer, 2011.
7. Huiji Gao, Jiliang Tang, Huan Liu: gSCorr: Modeling geo-social correlations for new check-ins on location-based social networks
8. Huiji Gao, Jiliang … M 2012.
9. Chloë Brown, V … nds meet: Social co
10. Miltiadis Allamanis, Salvatore Scellato, Cecilia Mascolo: Evolution of a location-based online social network: analysis and models. Internet Measurement Conference 2012: 145-158.
11. M. Domenico, A. Lima, and M. Musolesi. Interdependence and predictability of human mobility and social interactions. In *Nokia Mobile Data Challenge: http://research.nokia.com/page/12000*, MDC 2012.
12. P.A. Grabowicz, J.J. Ramasco, B. Goncalves, V.M. Eguiluz. Entangling Mobility and Interactions in Social Media. PLOS One 9(3):e92196. March 2014.
13. J. Toole, C. Herrera-Yaque, C.M. Schneider, M.C. Gonzalez. Coupling Human Mobility and Social Ties. In arXiv: 1502.00690v1, February 2015.
14. …

- Social ties affect human mobility.
- Human mobility affects social ties.

Predicting the next check-in on a LBSN

- Predicting the next FourSquare check-in
- A supervised approach
  - M5 tree and linear regression
- Tried out a bunch of features
  - User mobility features
  - Global mobility features
  - Temporal features
- How well?
  - Much lower accuracies compared to GPS
  - Much harder task

Anastasios Noulas, Salvatore Scellato, Neal Lathia, Cecilia Mascolo: Mining User Mobility Features for Next Place Prediction in Location-Based Services. ICDM 2012: 1038-1043.

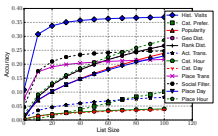


Fig. 2. Feature Predictability: Mean Accuracy for all features when they are being tested on an individual basis for different prediction list sizes N.

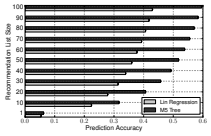


Fig. 4. Average accuracy for the supervised learning algorithms (linear regression and M5 model trees) for different recommendation list sizes.

Spatial search on a LBSN (using all available data, GPS, check-ins, *etc*)

- Best models of venue search are spatiotemporal models using the mixture of Gaussians, the timeliness feature, as well as popularity combined as a linear sum of log-likelihoods

Model	P@1
Baseline (nearest by distance)	0.130
Spatial Gaussian mixture model	0.193
Spatiotemporal models	0.277

- B. Shaw, J. Shea, S. Sinha, A. Hogue. Learning to Rank for Spatiotemporal Search. WSDM 2013: 717-726.
- R. Kumar, M. Mahdian, B. Pang, A. Tomkins, S. Vassilvitskii. Driven by Food: Modeling Geographic Choice. WSDM 2015: 213-222.

Spatial search on a LBSN (using all available data, GPS, check-ins, *etc*)

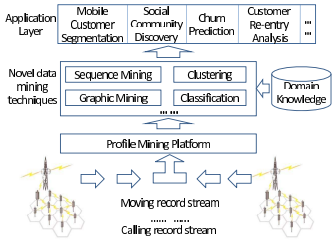
- Best models of social venue search: LambdaMart with a laundry list of features:
  - Spatial score, timeliness, popularity, here now, personal history, creator, mayor, friends here now, personal history w/ time of day

Model	P@1
Random	0.009
Spatial only	0.201
User history only	0.358
Popularity only	0.143
Linear regression: spatial + temporal	0.230
Linear regression: spatial + temporal + popularity	0.251
Linear regression: all features	0.434
Coordinate ascent: all features w/ nonlinear pairs	0.493
LambdaMART: all features	0.531

- B. Shaw, J. Shea, S. Sinha, A. Hogue. Learning to Rank for Spatiotemporal Search. WSDM 2013: 717-726.
- Q. Wu, C. J. Burges, K. M. Svore, and J. Gao. Adapting boosting for information retrieval measures. Information Retrieval, 13(3):254–270, June 2010.

Location & Activity

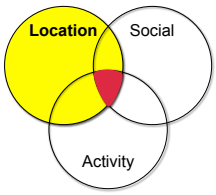
- MobileMiner [Wang et al. SIGMOD 2009]



Tengjiao Wang, Bishan Yang, Jun Gao, Dongqing Yang, Shiwei Tang, Haoyu Wu, Kedong Liu, Jian Pei: MobileMiner: A real world case study of data mining in mobile communication. SIGMOD 2009: 1083-1086

Context includes…

- **Location/Local**
  - What resources are nearby?
  - Where are you?
- **Social**
  - Who are you with?
- **Activity**
  - What are you doing?



Location & Social & Activity

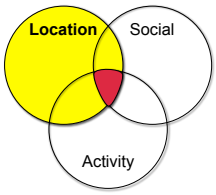
- Lots of work analyzing Twitter data in this space
- A recent best paper winner is from U. of Rochester
- *Flap* uses a dynamic Bayesian network per user to predict his/her locations given location of friends, time of day, type of day
- Experiments on over 4M tweets from users in LA and NYC
- It can correctly place a user within a **100m radius** (location prediction) with **up to 85% accuracy** (link prediction)

Figure 2: Flaps's visualization of a sample of geo-active friends in NYC. Red links between users represent friendships.

- Adam Sadilek, Henry A. Kautz, Jeffrey P. Bigham: Finding your friends and following them to where you are. WSDM 2012:723-732 (**best paper**).
- Adam Sadilek, Henry Kautz, Jeffrey P. Bigham: Modeling The Interplay of People's Location, Interactions, and Social Ties. IJCAI 2013.

Context includes…

- **Location/Local**
  - What resources are nearby?
  - Where are you?
- **Social**
  - Who are you with?
- **Activity**
  - What are you doing?



RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

MINING DATA FROM MOBILE DEVICES

Algorithms: Location & Context

Spiros Papadimitriou, *Tina Eliassi-Rad*
{spiros.papadimtriou, tina.eliassi}@rutgers.edu

License