Computer and Information Science and Engineering Department
UNIVERSITY OF FLORIDA

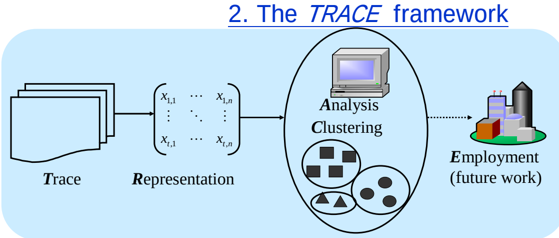
Mining Behavioral Groups in Large Wireless LANs

Wei-jen Hsu[1], Debojyoti Dutta[2], Ahmed Helmy[1]
[1]Dept. of Computer and Information Science and Engineering, U. of Florida, [2]Cisco Systems, Inc.
wjhsu@ufl.edu, dedutta@cisco.com, helmy@ufl.edu http://nile.cise.ufl.edu/MobiLib

1. Introduction

- Wireless devices and infrastructures are ubiquitous and have wider impact than mere advances in technology on our lives
- Portable personal wireless devices provide an opportunity for the study of human behavior
- We leverage WLAN traces to discover distinct behavioral groups (in terms of mobility)
  - USC WLAN trace[2] (2006 spring, 94 days, 5000 users)
  - Dartmouth WLAN trace[3] (2004 spring, 61 days, 6582 users)

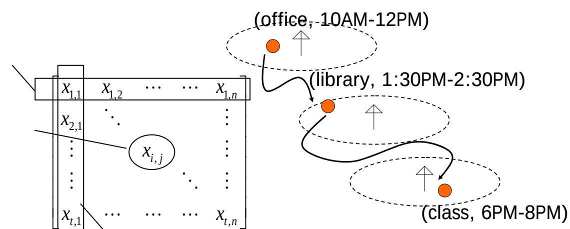
2. The *TRACE* framework



Contribution: Application of unsupervised learning techniques (clustering) with carefully selected features to identify groups

3. Preliminaries: Representing User Associations

- Use a *normalized association vector* to represent summary of user mobility in each day
- Elements in the vectors quantify the relative importance of a location to the user
- Location preference is one of the distinguishing feature of social groups
- *Association matrix* keeps long-run behavior of a user



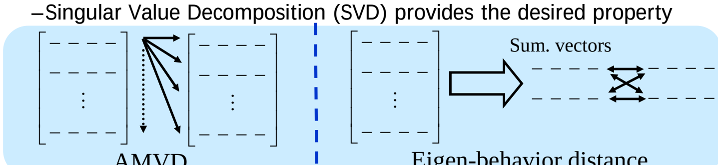
Association vector:
(library, office, class) =(0.2, 0.4, 0.4)

4. Comparing Similarity between Users' Association Matrices

- Average Minimum Vector Distance (AMVD)
  - For each association vector (row) of user *i*, find the closest vector of user *j* and take average of $|a_{jd'} - a_{id}|$ over all days *d*
  - Intuition: for every daily association vector of *i*, if there is a similar association vector for *j*, then (*i,j*) have similar behavior
- Drawback: expensive to compute $O(N^2d^2)$ and exchange. Includes noises.
- Require meaningful summary for the association matrix
  - Asso. matrices: multi-modal row vectors, but low dimensionality
  - Summary vector Y that captures the most variation in row vectors Xi's

$$SIG(Y) = \sum_{i=1}^{d} |X_i \cdot Y| \Big/ \sum_{i=1}^{d} \|X_i\|_1$$

  - Singular Value Decomposition (SVD) provides the desired property

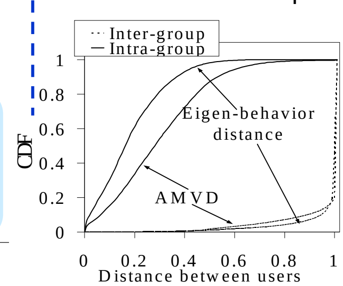


- Eigen-behaviors: The vectors that describe the maximum remaining power in the association matrix

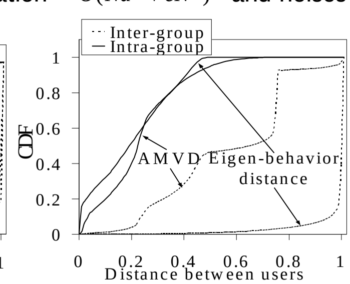
$$u_1 = \arg\max_{\|u\|=1} \|X \cdot u\|,\quad u_k = \arg\max_{\|u\|=1} \left\| (X - \sum_{i=1}^{k-1} X u_i u_i') \cdot u \right\| \quad \forall k \geq 2$$

with quantifiable importance $w_k = \sigma_k^2 \Big/ \sum_{i=1}^{Rank(X)} \sigma_i^2$

- Eigen-behavior Distance calculates similarity of users by weighted inner products of eigen-behaviors.
  - $Sim(U,V) = \sum_{\forall i,j} w_i w_j |u_i \cdot v_j|$
- Benefit: Reduced computation $O(Nd^2 + cN^2)$ and noises



USC



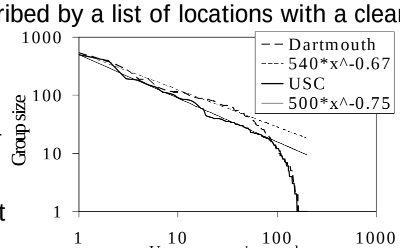
Dartmouth

5. Interpretation of Behavioral Groups

- The existence of distinct behavioral groups (~*hundreds*)– each group has unique group eigen-behavior

Significance score of top eigen-behavior for	USC	Dartmouth
Its own group	0.779	0.727
Other groups	0.005	0.004

- Skewed group size distribution – the largest 10 groups account for more than 30% of population on campus. Power-law distributed group sizes.
- Most groups can be described by a list of locations with a clear ordering of importance
- We also observe groups visiting multiple locations with similar importance – taking the majority or average of association vectors from a user is not sufficient



This work is supported by NSF CAREER Award 0134650 and partially completed when the authors were with University of Southern California.

6. Potential Usage and Future Direction

- Profile-casting[4] - Fast evaluation of similarity between nodes facilitates de-centralized packet forwarding decisions targeted at specific groups
- Better mobility model (with group size distribution and vectors describing their mobility preferences)
- Establishing norm of mobility characteristics of users
  - Abnormality detection, targeted advertisement
- The *TRACE* framework could be applied to different *representations* (e.g. encounter vector)

[1] Longer version of technical report available at http://arxiv.org/abs/cs/0606002
[2] W. Hsu and A. Helmy, MobiLib USC WLAN trace data set. Download from http://nile.cise.ufl.edu/MobiLib/USC_trace/
[3] D. Kotz, T. Henderson and I. Abyzov, CRAWDAD data set dartmouth/campus/movement/01_04 (v. 2005-03-08). Downloaded from http://crawdad.cs.dartmouth.edu/dartmouth/campus/movement/01_04
[4] W. Hsu, D. Dutta, and A. Helmy, "Profile-cast: Behavior-Aware Mobile Networking," MOBICOM 2007 poster and SRC.