Data-driven Modeling and Design of Networked Mobile Societies:

A Paradigm Shift for Future Social Networking

Ahmed Helmy

Computer and Information Science and Engineering (CISE) Department

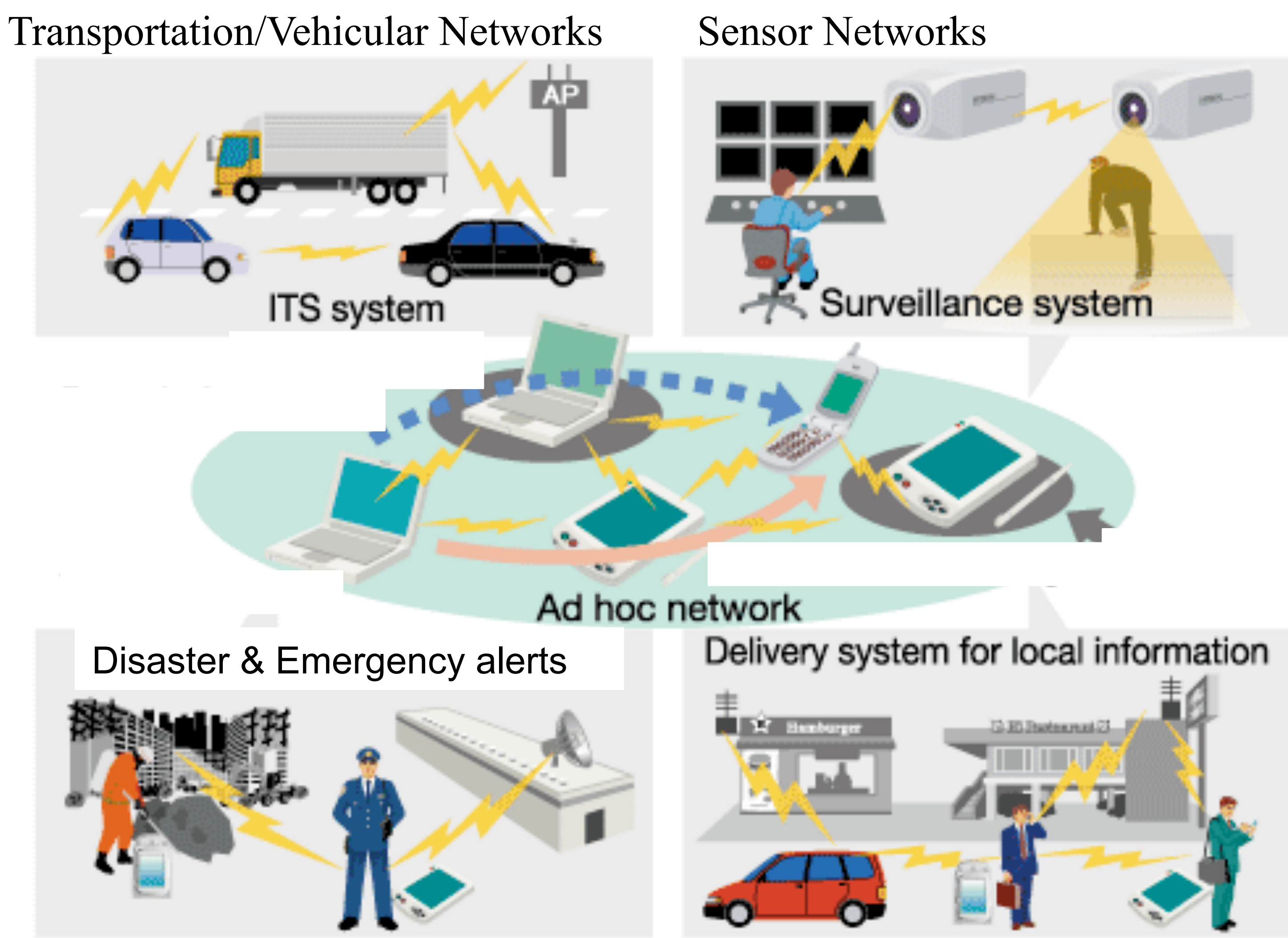
University of Florida

helmy@ufl.edu , *http://www.cise.ufl.edu/~helmy*

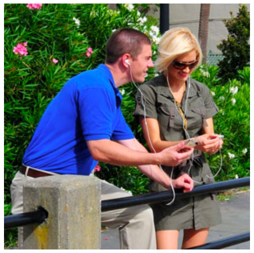
Founder & Director: Wireless Mobile Networking Lab *http://nile.cise.ufl.edu*

Funded by:

Networked Mobile Societies Everywhere, Anytime



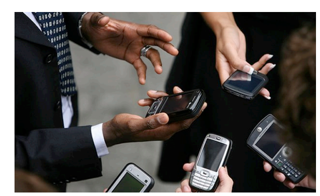
Mobile Ad hoc, Sensor and Delay Tolerant Networks



Emerging Behavior-Aware Services

- Tight coupling between users, devices
  - Devices can infer user preferences, behavior
  - Capabilities: comm, comp, storage, sensing
- New generation of behavior-aware protocols
  - Behavior: mobility, interest, trust, friendship,…
  - Apps: interest-cast, participatory sensing, crowd sourcing, mobile social nets, alert systems, …

New paradigms of communication?!



Paradigm Shift in Protocol Design

Used to:

Design general purpose protocols
Evaluate using models (random mobility, traffic, …)
Deployment context: Modify to improve performance and failures for specific context

– May end up with suboptimal performance or failures due to lack of context in the design

Propose to:

Analyze, model deployment ***context***
Design *'application class'-specific* parameterized protocols
Utilize insights from context analysis to fine-tune protocol parameters

Problem Statement

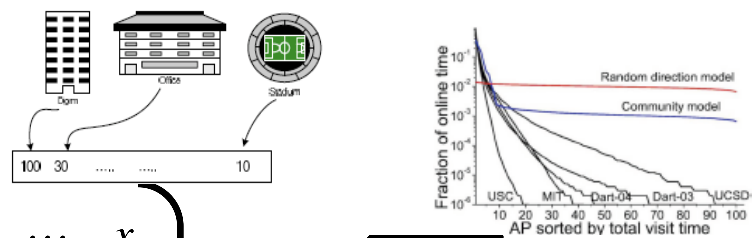
- How to gain insight into deployment context?
- How to utilize insight to design future services?

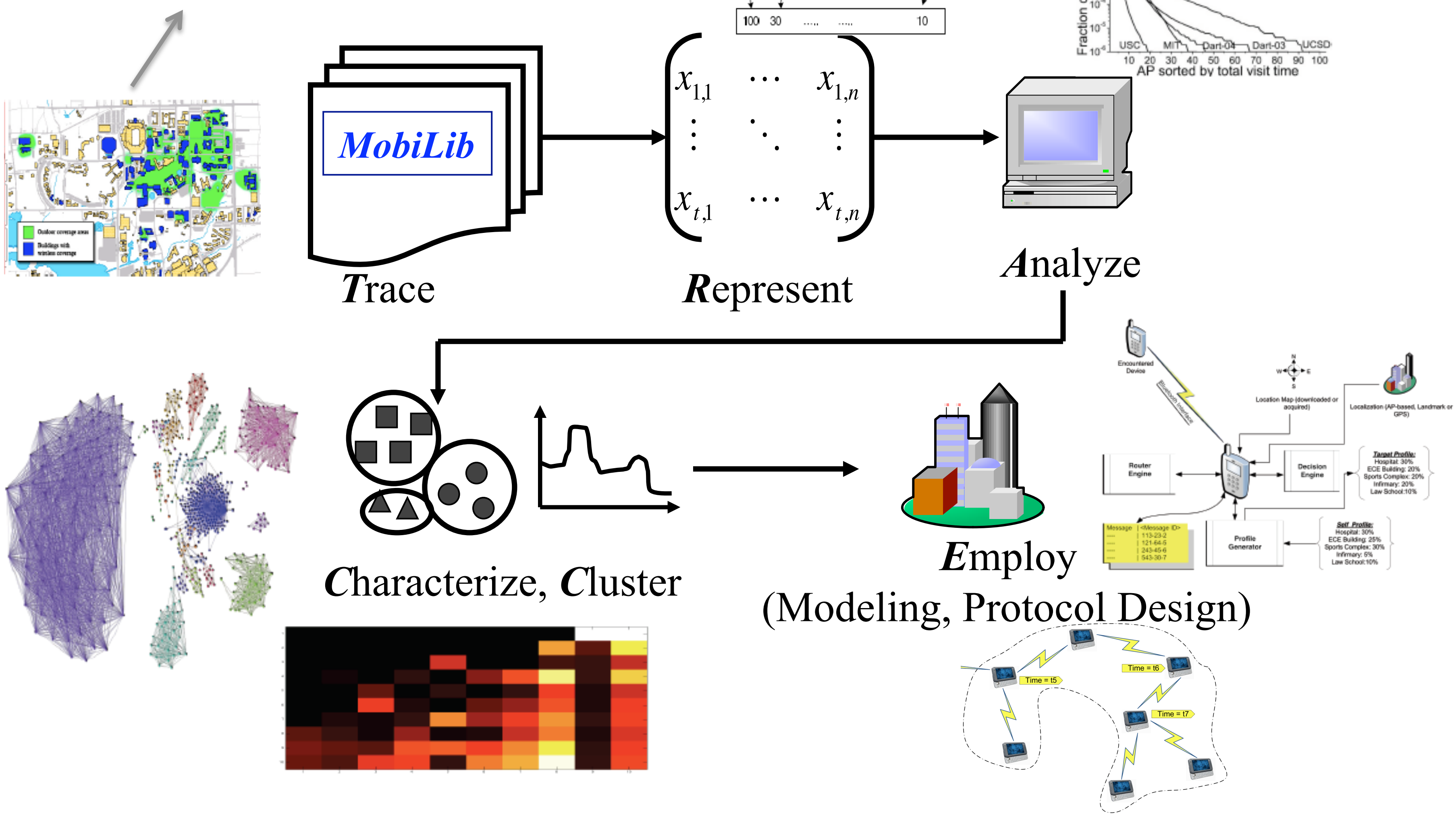
Approach

- Extensive trace-based analysis to identify dominant trends & characteristics
- Analyze user behavioral patterns
  - Individual user behavior and mobility
  - Collective user behavior: grouping, encounters
- Integrate findings in modeling and protocol design
  - I. User mobility modeling – II. Behavioral grouping
  - III. Information dissemination in mobile societies, *profile-cast*

The *TRACE* framework

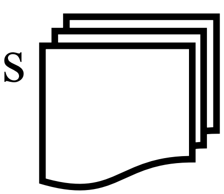
	Duration	Records	Total Users	Access points (or bldgs)
USC-WLAN	Dec 03-Jun 08	50 M	55,500	79 ports (03), 161 (08)
USC-DHCP	Dec 03-Jun 08	60 M	55,500	79 ports (03), 161 (08)
USC-netflow	Apr 05-Jun 08	50 B	50,000	161 ports
UF-WLAN	Jun 07-Current	45 M	105,500	784 Access points
UF-DHCP	Jun 07-Current	10 M	105,500	784 Access points
UF-netflow	to start Sep 09	n/a	n/a	784+ Access points





Community-wide Wireless/Mobility Library

- Library of
  - Measurements from Universities, vehicular networks
  - *Realistic* models of behavior (mobility, traffic, encounters)
  - Simulation benchmarks - Tools for trace data mining
- Available libraries:
  - CRAWDAD (Dartmouth, ‘05-) *crawdad.cs.dartmouth.edu*
    MobiLib (USC & UFL, ’04-) *nile.cise.ufl.edu/MobiLib*
    - 60+ Traces from: USC, Dartmouth, MIT, UCSD, UCSB, UNC, UMass, GATech, Cambridge, UFL, …
    - Tools for mobility modeling (IMPORTANT, TVC), data mining
- Types of traces:
  - Campuses (WLANs), Conference AP and encounter traces
  - Municipal (off-campus) wireless APs, Bus & vehicular



***T*race**

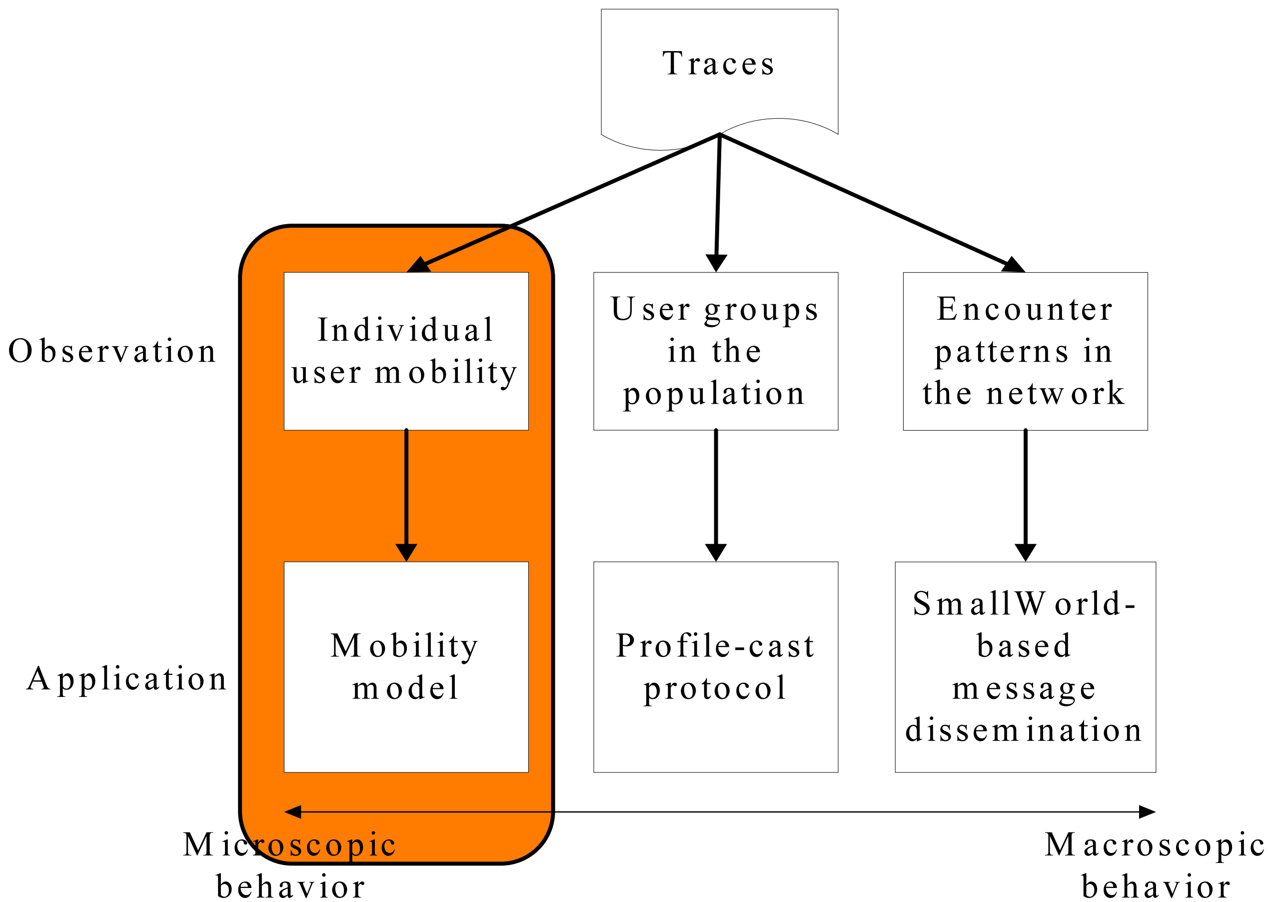
IMPACT: Investigation of Mobile-user Patterns Across University Campuses using WLAN Trace Analysis*

- 4 major campuses – 30 day traces studied from 2+ years of traces
- Total users > 12,000 users - Total Access Points > 1,300

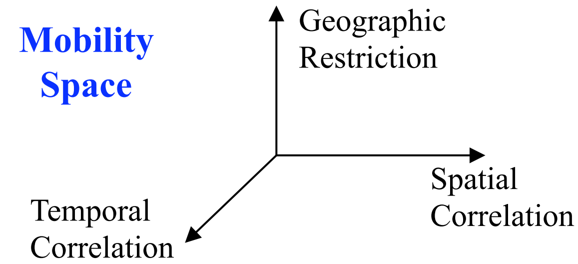
Trace source	Trace duration	User type	Environment	Collection method	Analyzed part
MIT	7/20/02 – 8/17/02	Generic	3 corporate buildings	Polling	Whole trace
Dartmouth	4/01/01 – 6/30/04	Generic w/ subgroup	University campus	Event-based	July '03 April '04
UCSD	9/22/02 – 12/8/02	PDA only	University campus	Polling	09/22/02-10/21/02
USC	4/20/05 – 3/31/06	Generic	University campus	Event-based (Bldg)	04/20/05-05/19/05

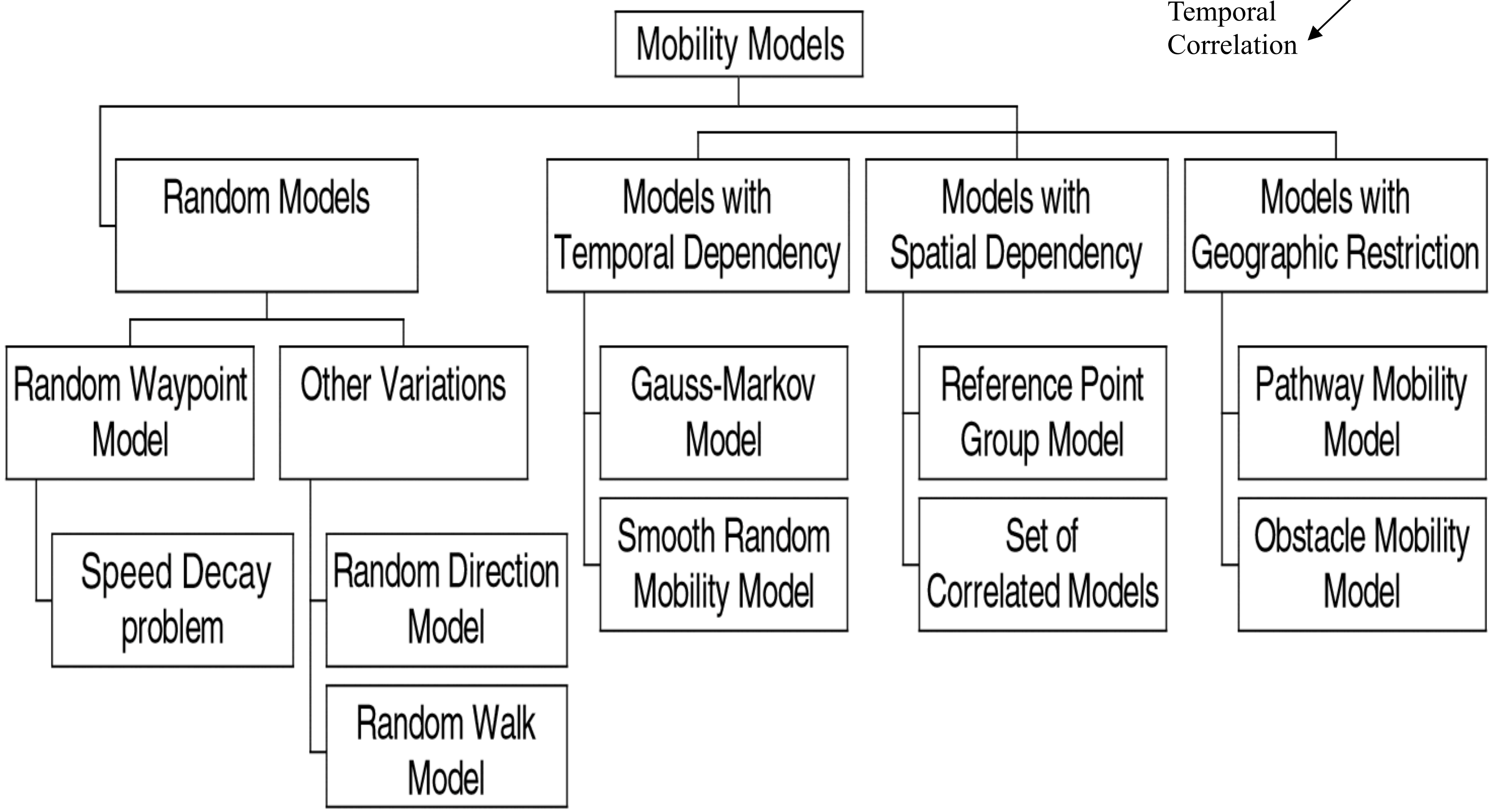
* W. Hsu, A. Helmy, "***IMPACT***: Investigation of Mobile-user Patterns Across University Campuses using WLAN Trace Analysis", two papers at *IEEE Wireless Networks Measurements (WiNMee)*, April 2006 and *IEEE Transactions on Mobile Computing*, 2010 (To appear).

Case study I – Individual Mobility



Classification of Mobility Models

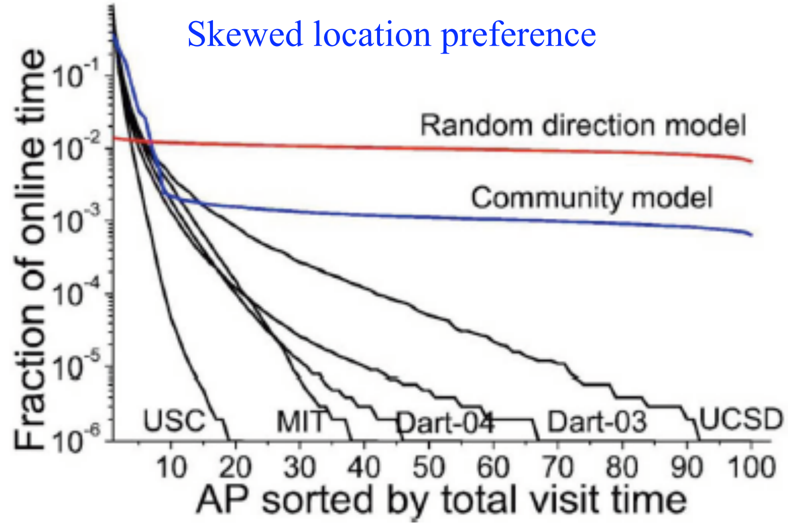




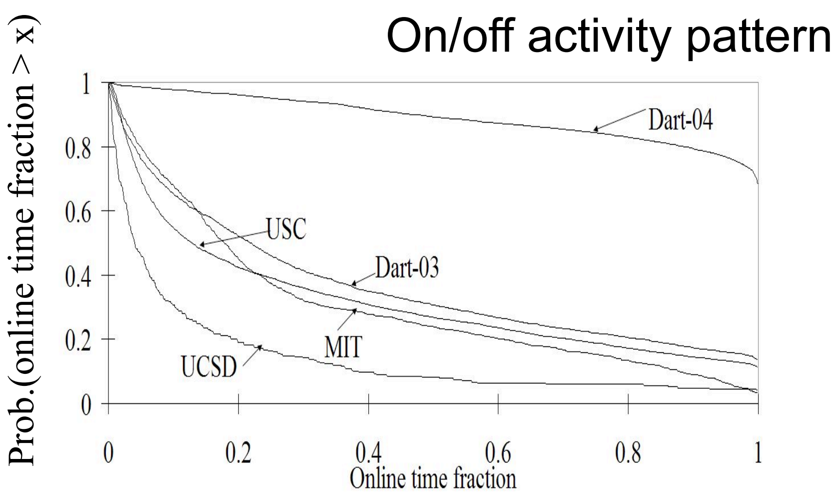
* F. Bai, A. Helmy, "A Survey of Mobility Modeling and Analysis in Wireles Adhoc Networks", Book Chapter in the book "Wireless Ad Hoc and Sensor Networks", Kluwer Academic Publishers, June 2004.

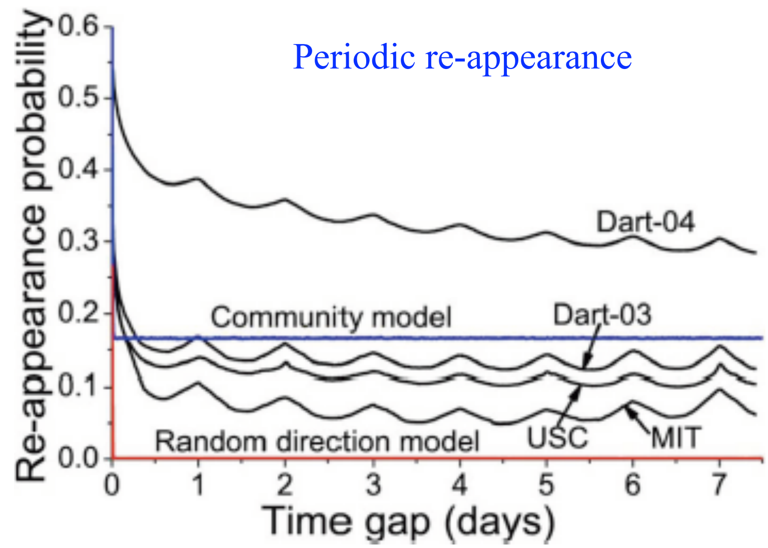
Spatio-temporal Mobility in WLANs

- Simple existing models are very different from the spatio-temporal characteristics in WLANs

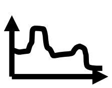


95% on-line time at 5 most visited APs





Periodic repetition peaks daily/weekly



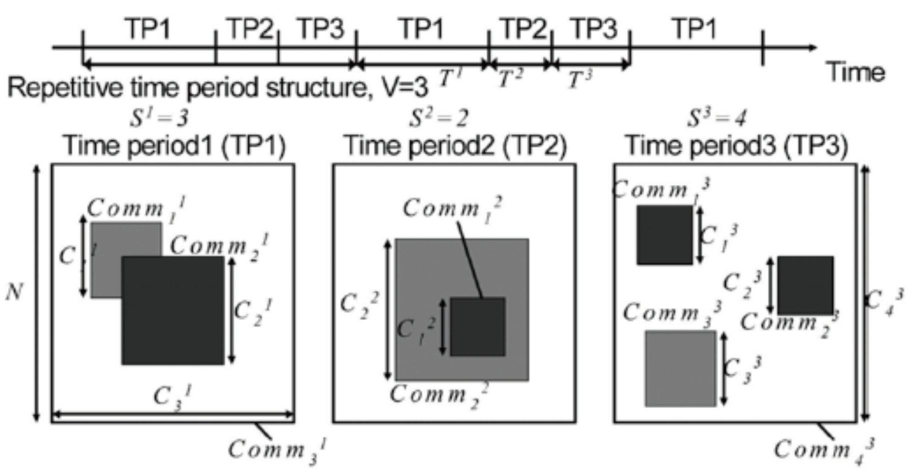
*C*haracterize

The *TVC* Model: Reproducing Mobility Characteristics

Time-Variant Community (***TVC***) Model:

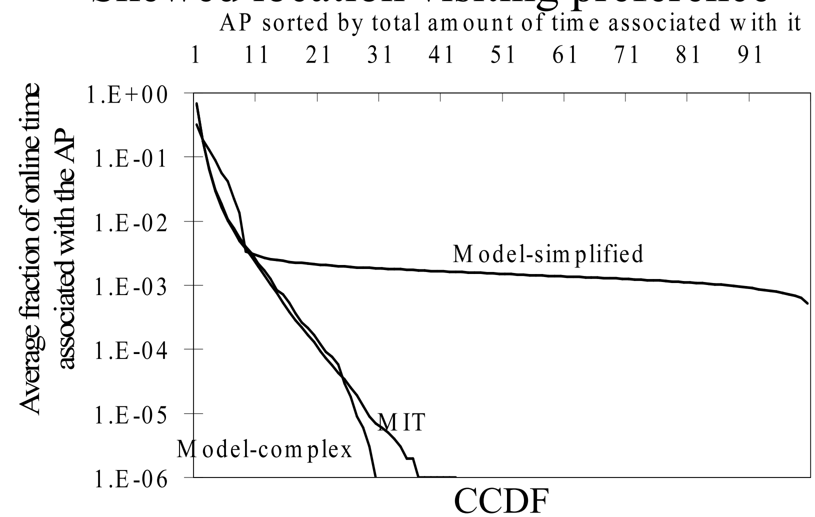
1- Assigns communities (locations) to users to re-produce location visiting preference

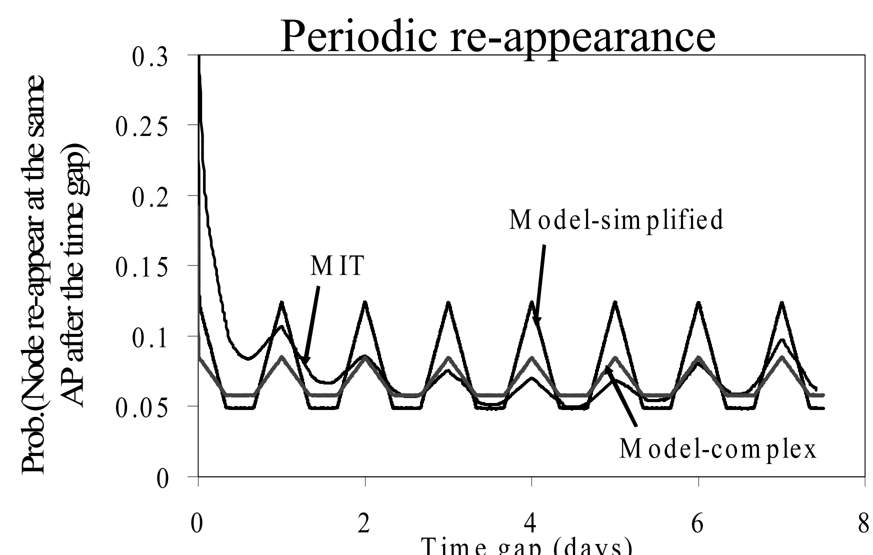
2- Varies temporal assignment of communities to re-produce the periodic re-appearance



Time-variant mobility model, with three time periods and different numbers of communities in each period.

Skewed location visiting preference





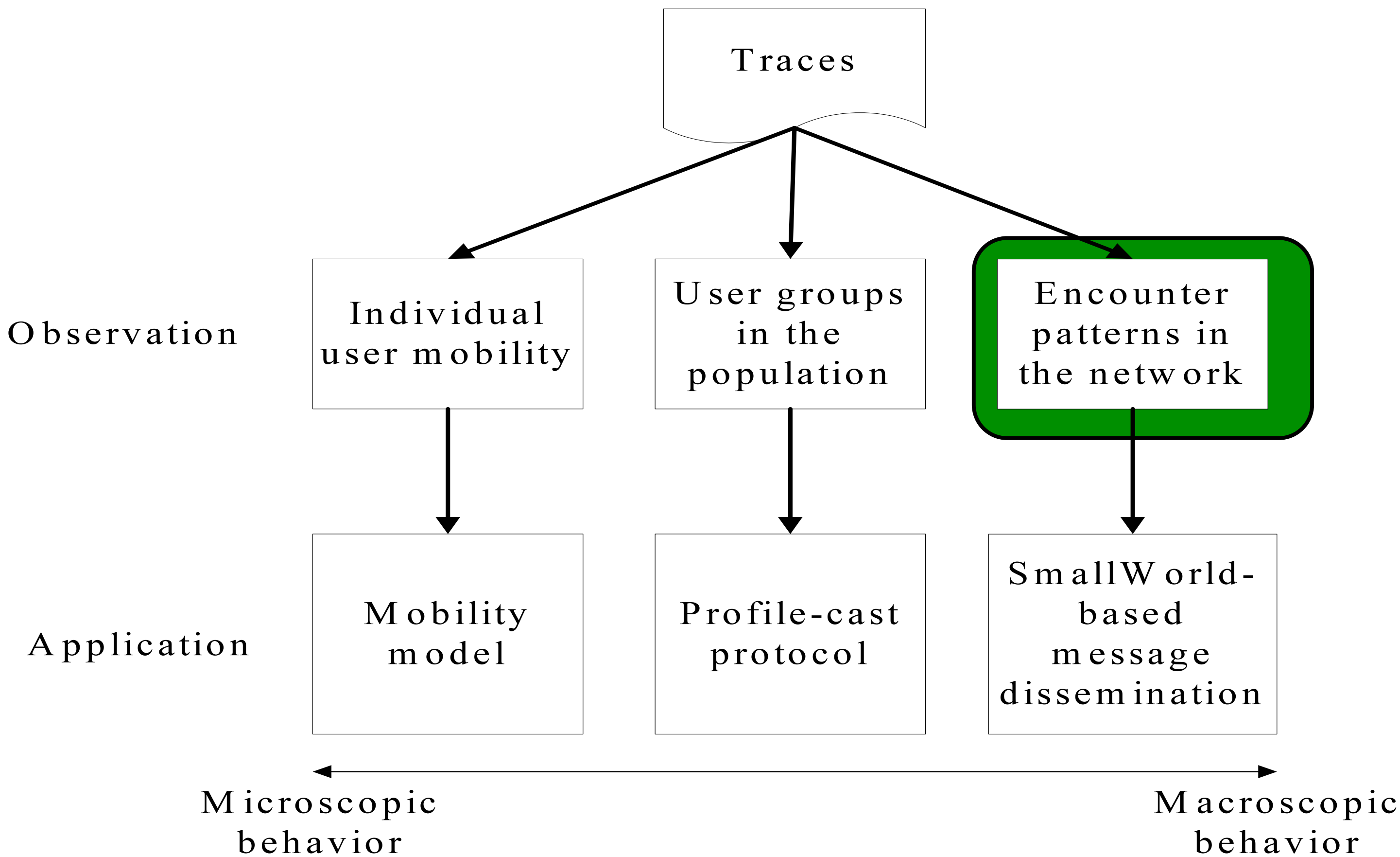
IEEE INFOCOM 2007

IEEE/ACM Trans. on Networking 2009

* Model-simplified: single community per node. Model-complex: multiple communities

** Similar matches achieved for USC and Dartmouth traces

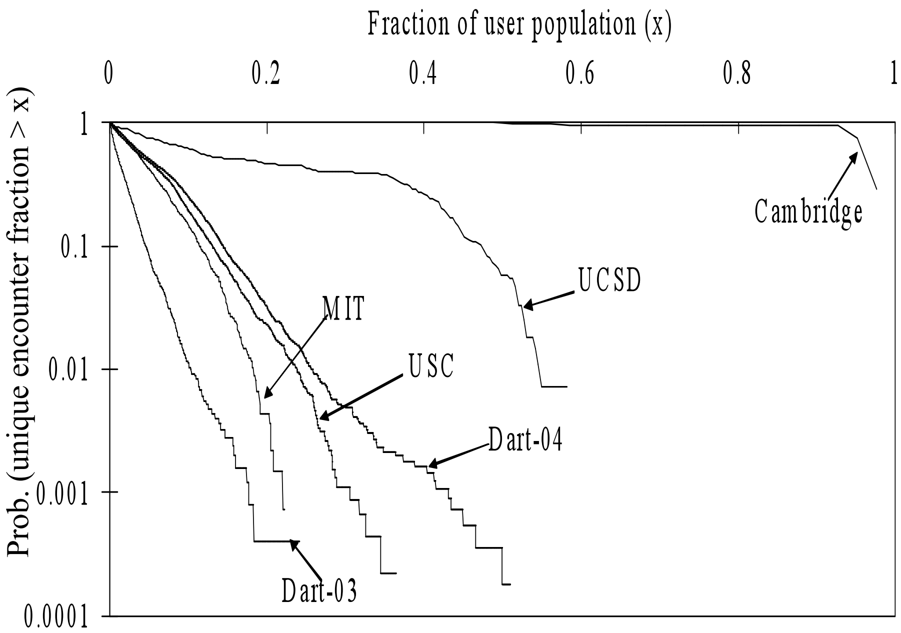
Case study II – Encounter Patterns



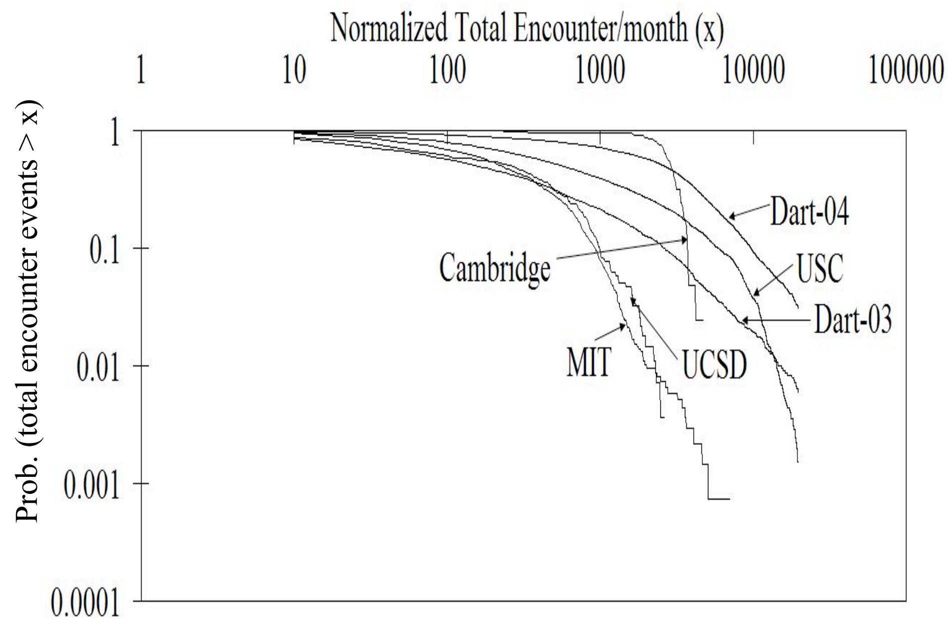
Case Study II: Goal

- Understand inter-node encounter patterns from a global perspective
  - How do we represent encounter patterns?
  - How do the encounter patterns influence network connectivity and communication protocols?
- Encounter definition:
  - In WLAN: When two mobile nodes access the same AP at the same time they have an ‘encounter’
  - In DTN: When two mobile nodes move within communication range they have an ‘encounter’

Observations: Nodal Encounters



CCDF of unique encounter count



CCDF of total encounter count

- ***In all the traces, the MNs encounter a small fraction of the user population.***
- ***A user encounters*** 1.8%-6% ***on average of the user population***
- ***The number of total encounters for the users follows a BiPareto distribution.***

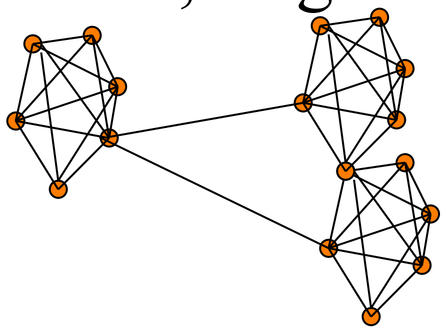
W. Hsu, A. Helmy, "On Nodal Encounter Patterns in Wireless LAN Traces", *IEEE Transactions on Mobile Computing (TMC)*, To appear

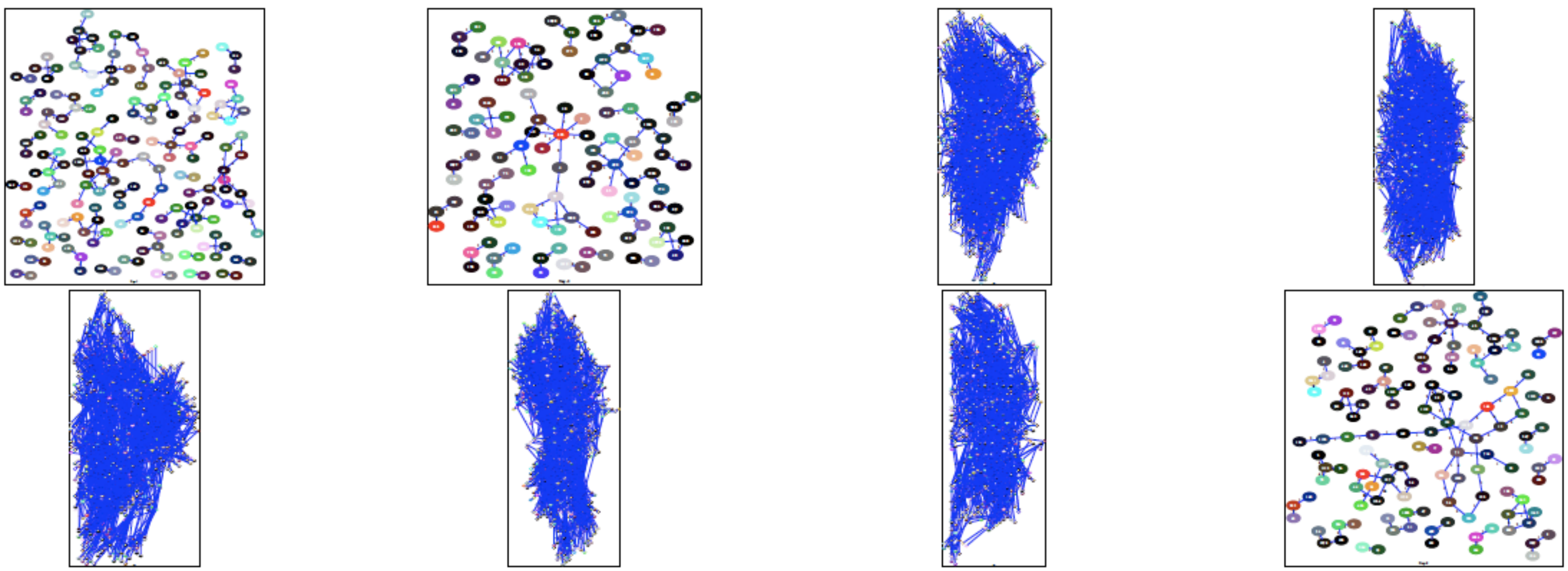
The Encounter graph

$$\begin{pmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{t,1} & \cdots & x_{t,n} \end{pmatrix}$$

Represent

- Vertices: mobile nodes, Edges: node encounters

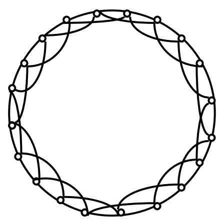




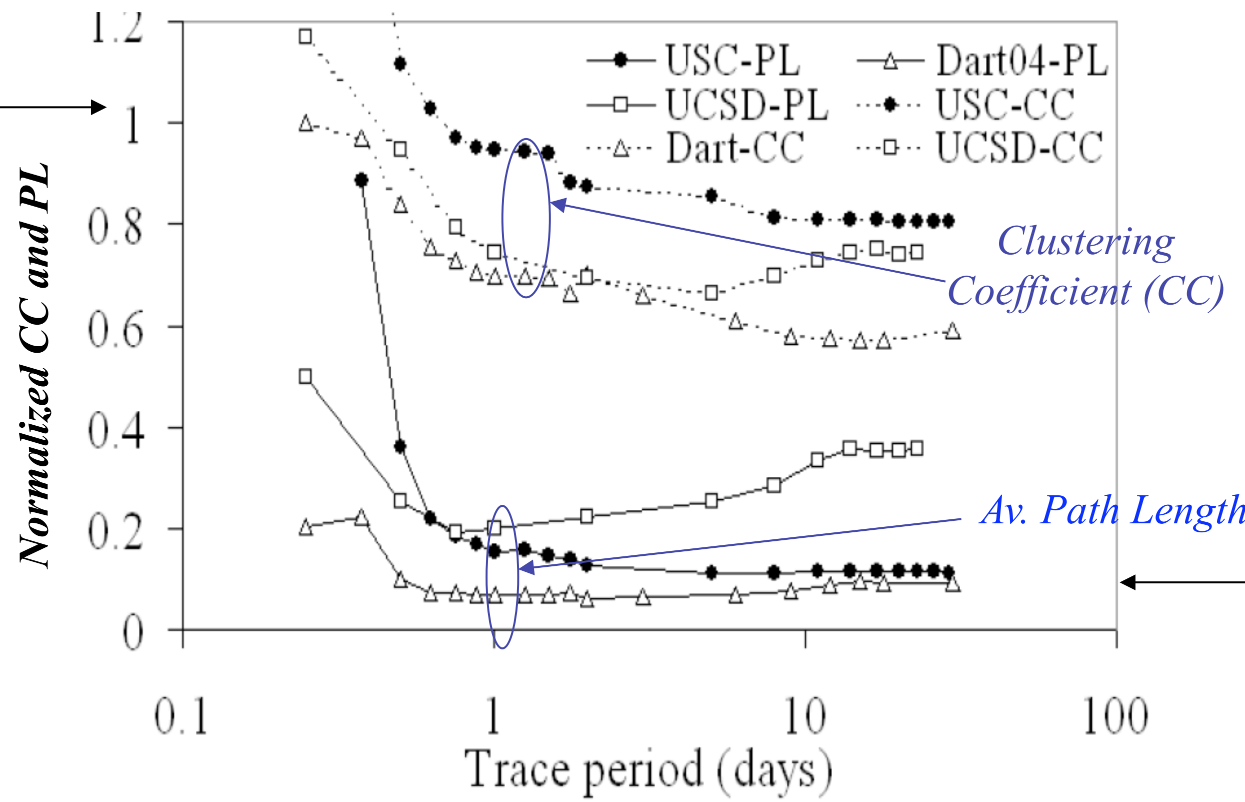
Daily encounter graphs for MIT trace

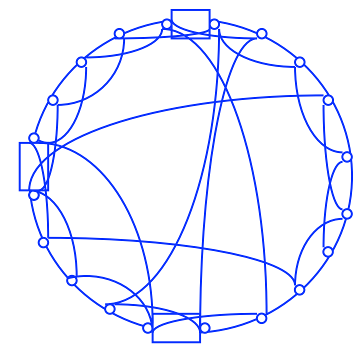
Small Worlds of Encounters

- **Encounter graph: nodes as vertices and edges link all vertices that encounter**

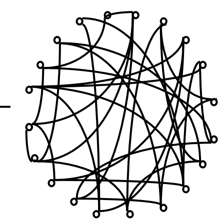


Regular graph





Small World

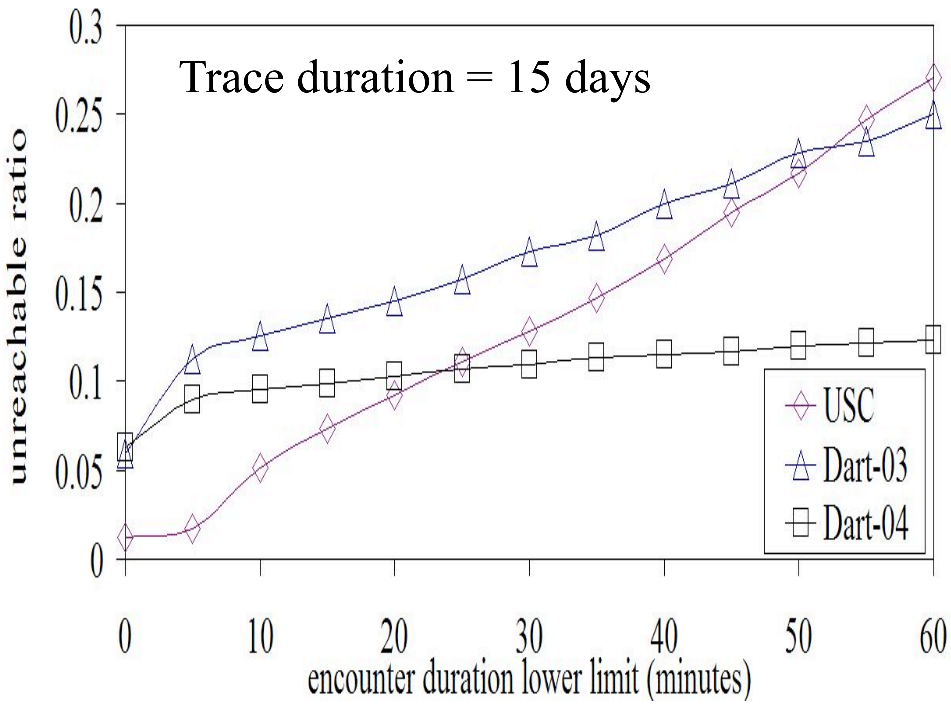


Random graph

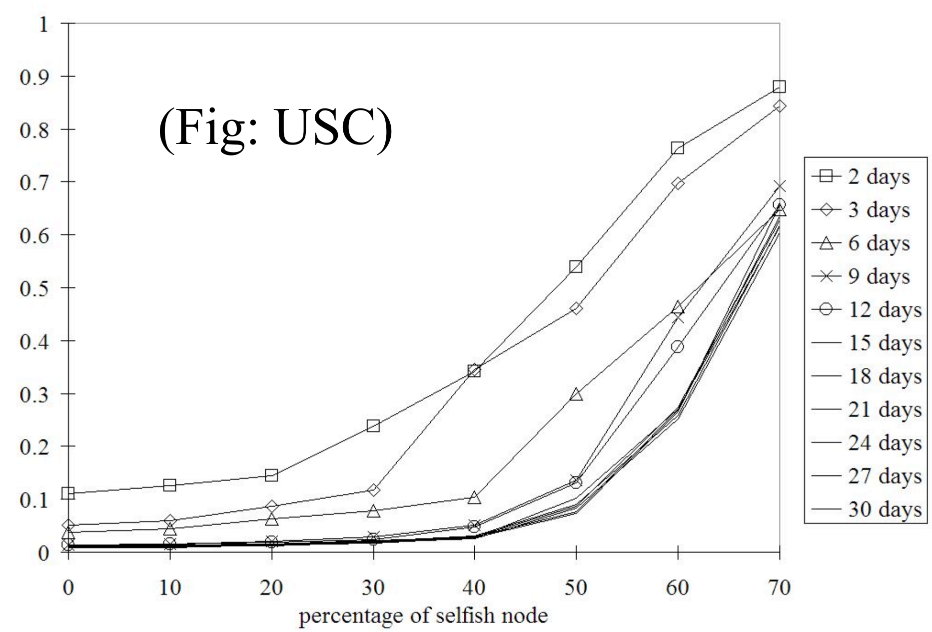
- **The encounter graph is a *Small World* graph (high *CC*, low *PL*)**
- **Even for short time period (1 day) its metrics (*CC*, *PL*) almost saturate**

Information Diffusion in DTNs via Encounters

- Epidemic routing (spatio-temporal broadcast) achieves almost complete delivery



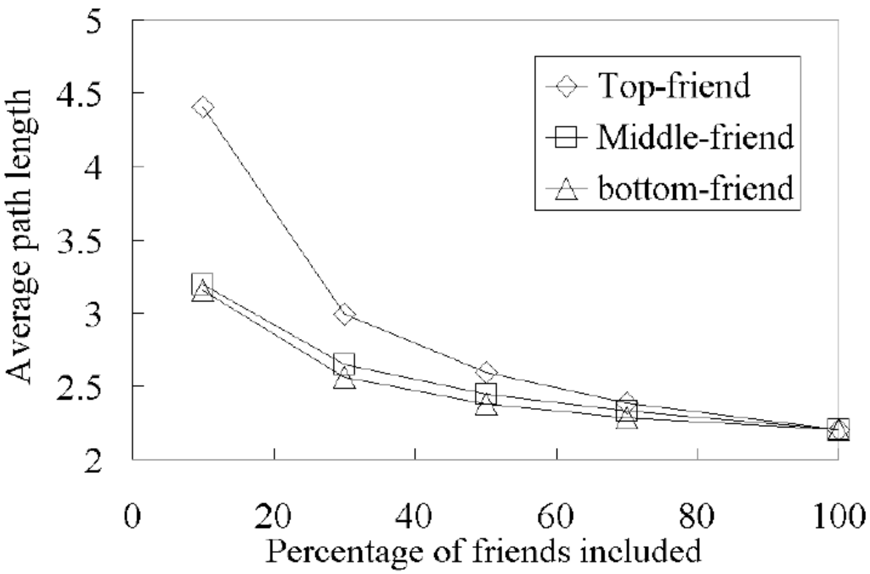
Robust to the removal of short encounters



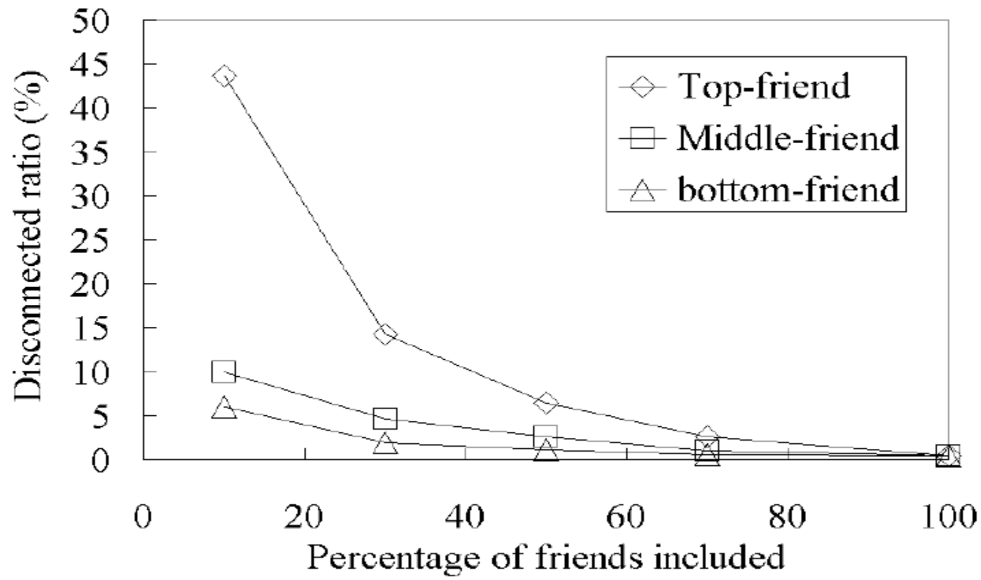
Robust to selfish nodes (up to ~40%)

Encounter-graphs using Friends

- **Distribution for friendship index *FI* is exponential for all the traces**
- **Friendship between MNs is highly asymmetric**
- **Among all node pairs: < 5% with *FI* > 0.01, and <1% with *FI* > 0.4**



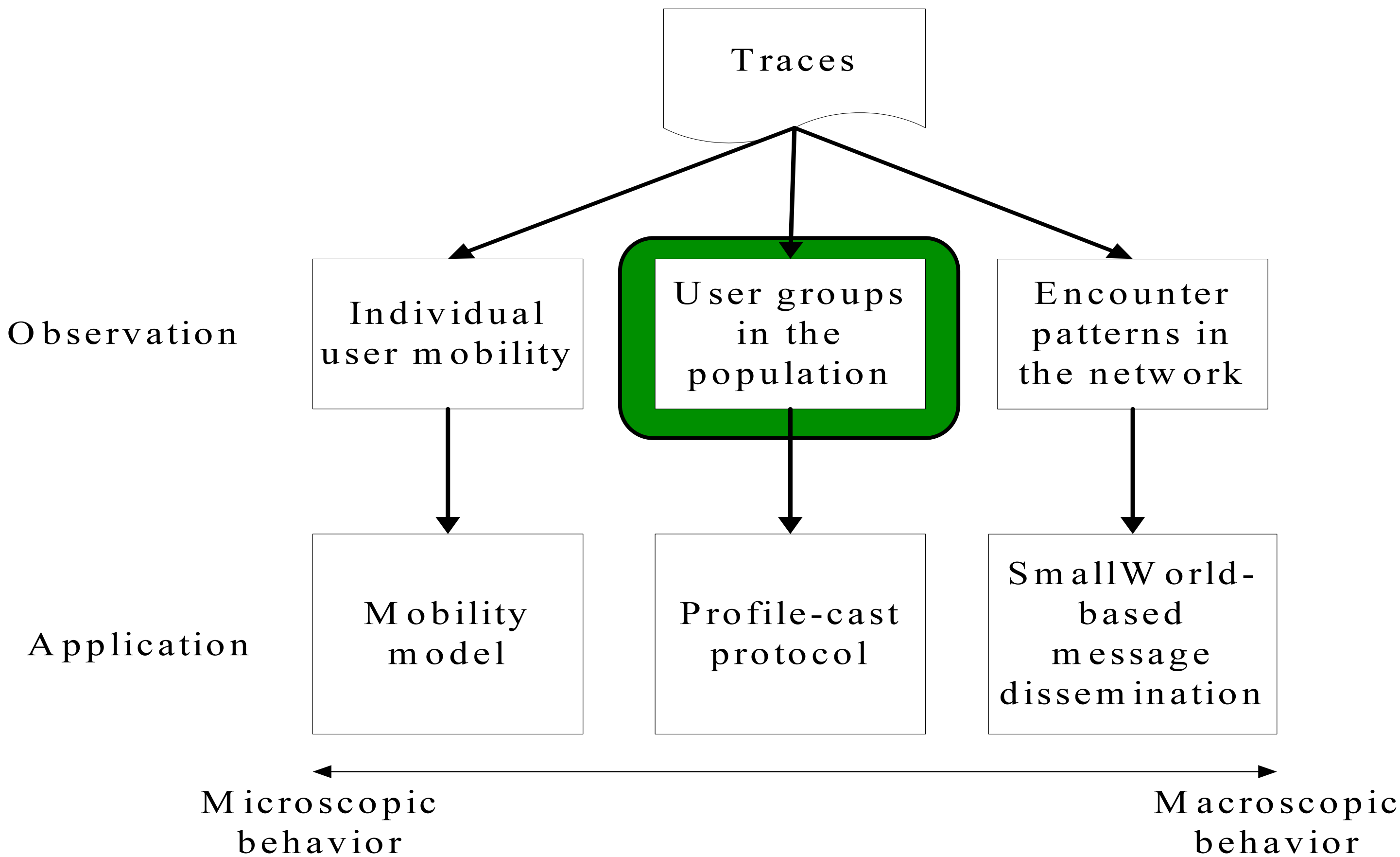
(b) Average path length



(c) Disconnected ratio

•Top-ranked friends form cliques and low-ranked friends are key to provide random links (short cuts) to reduce the degree of separation in encounter graph.

Case study III – Groups in WLAN



Case Study III: Goal

- Identify similar users (in terms of long run mobility preferences) from the diverse WLAN user population
  - Understand the constituents of the population
  - Identify potential groups for group-aware service
- Classify users based on their mobility trends and location-visiting preferences
  - Traces studied: semester-long USC trace (spring 2006, 94days) and quarter-long Dartmouth trace (spring 2004, 61 days)

Representation of User Association Patterns

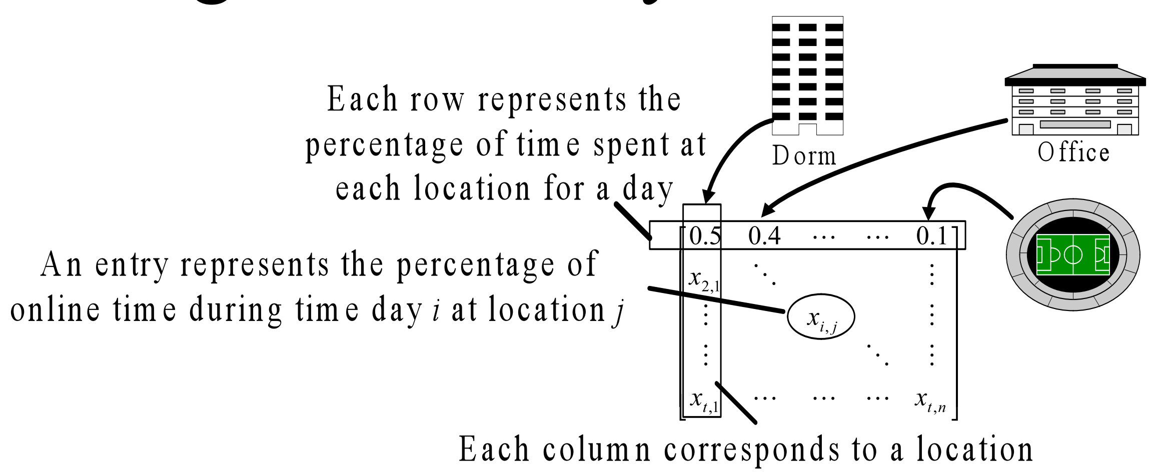
W. Hsu, D. Dutta, A. Helmy, "Mining Behavioral Groups in WLANs", *ACM MobiCom* '07

- Summarize user association per day by a vector
  - *a* = {a_j : fraction of online time user *i* spends at AP_j on day *d*}

-Office, 10AM -12PM
-Library, 3PM – 4PM
-Class, 6PM – 8PM

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

- Sum long-run mobility in "association matrix"



Example association matrix to describe a given user's location visiting preference

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{t,1} & \cdots & x_{t,n} \end{bmatrix}$$

***R**epresent*

Eigen-behaviors & Behavioral Similarity Distance

- Eigen-behaviors (*EB*): Vectors describing maximum remaining power in assoc. matrix *M* (through *SVD*):

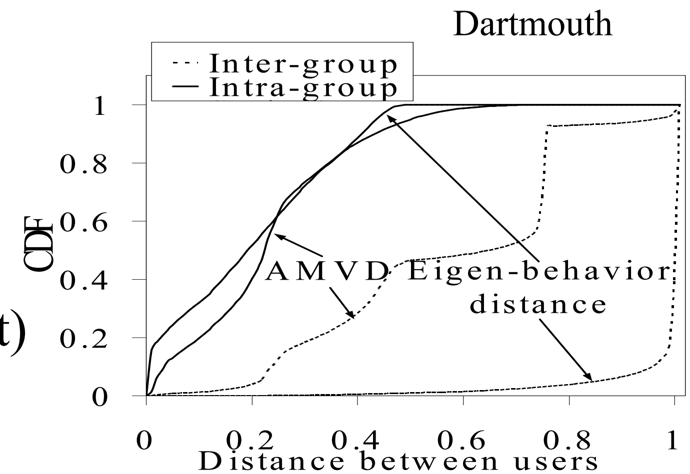
$$M = U \cdot \Sigma \cdot V^T$$

  - Get Eigen-vectors: $v_1, v_2, ..., v_{rank(M)}$ - Get Eigen-values: $\sigma_1,\ \sigma_2,\ ...,\ \sigma_{rank(M)}$
  - Get relative importance: $w_i = \frac{\sigma_i^2}{\sum_{j=1}^{Rank(M)} \sigma_j^2}$

- Eigen-behavior Distance weighted inner products of *EBs*
  - $Sim(U,V) = \sum_{\forall i,j} w_i w_j \left| u_i \cdot v_j \right|$
- Assoc. patterns can be re-constructed with low rank & error
- For over 99% of users, < 7 vectors capture > 90% of *M*'s power

Similarity-based User Classification

- Hierarchical clustering of similar behavioral groups
- High quality clustering:
  - Inter-group vs. intra-group distance
  - Significance vs. random groups
    - 0.93 v.s. 0.46 (USC), 0.91 v.s. 0.42 (Dart)



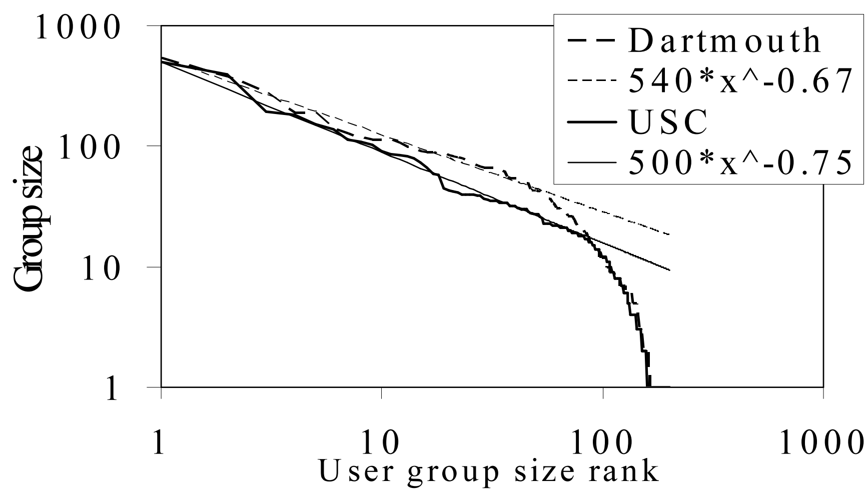
*AMVD = Average Minimum Vector Distance

  - Unique groups based on *Eigen Behaviors*

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

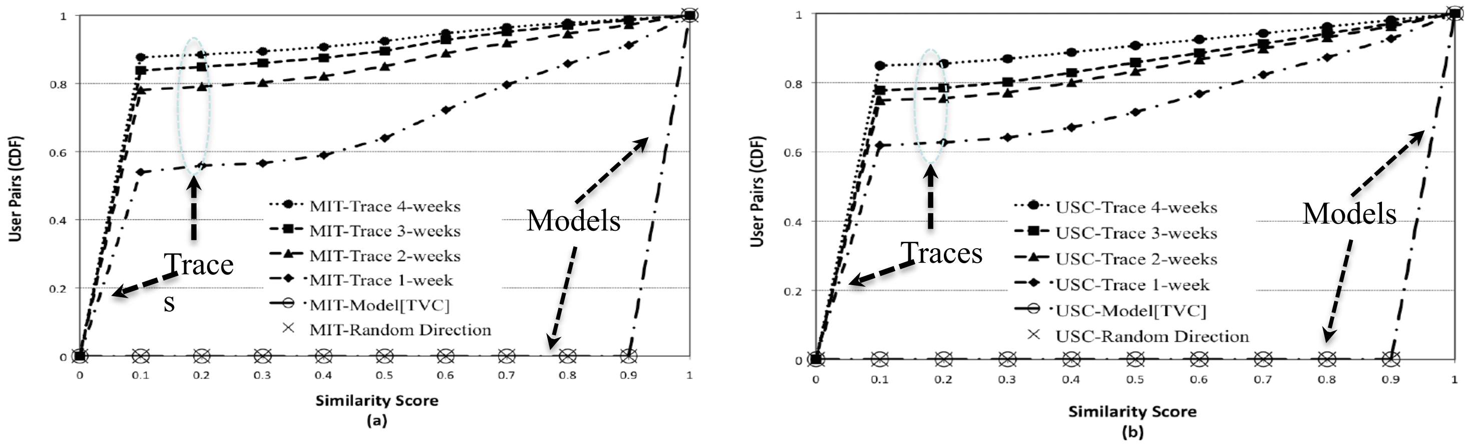
User Groups in WLAN - Observations

- Identified hundreds of distinct groups of similar users
- Skewed group size distribution –
  - the largest 10 groups account for more than 30% of population on campus
  - Power-law distributed of group sizes
- Most groups can be described by a list of locations with a clear ordering of importance
- Some groups visit multiple locations with similar importance –
  - taking the most important location for each user is not sufficient



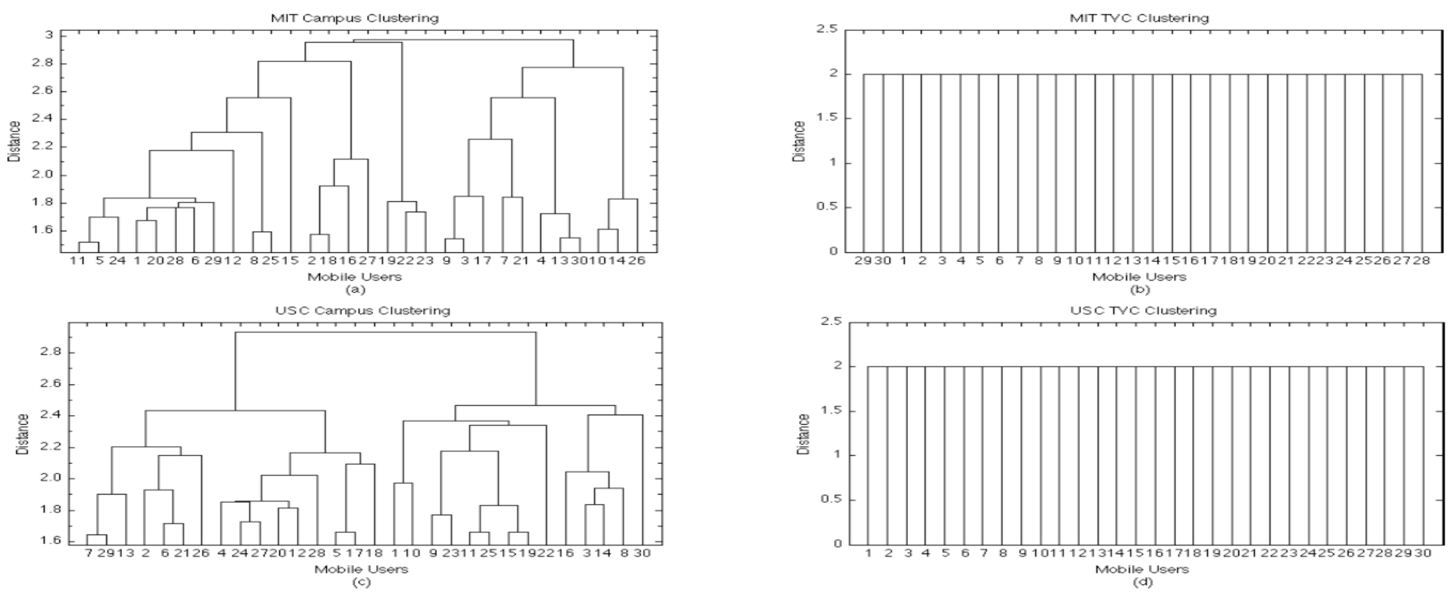
Videos

Behavioral Similarity: The Missing Link



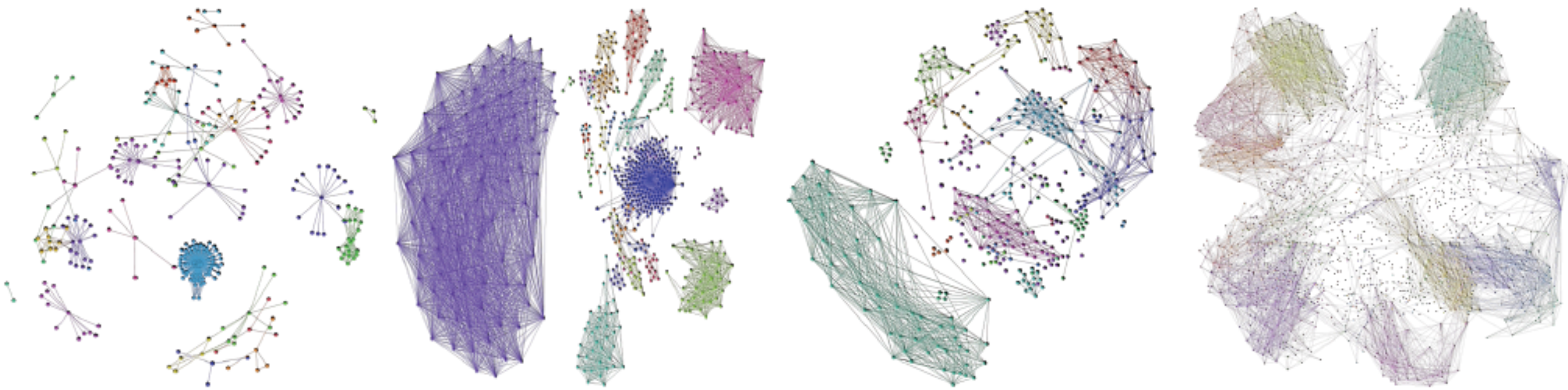
Traces

Models

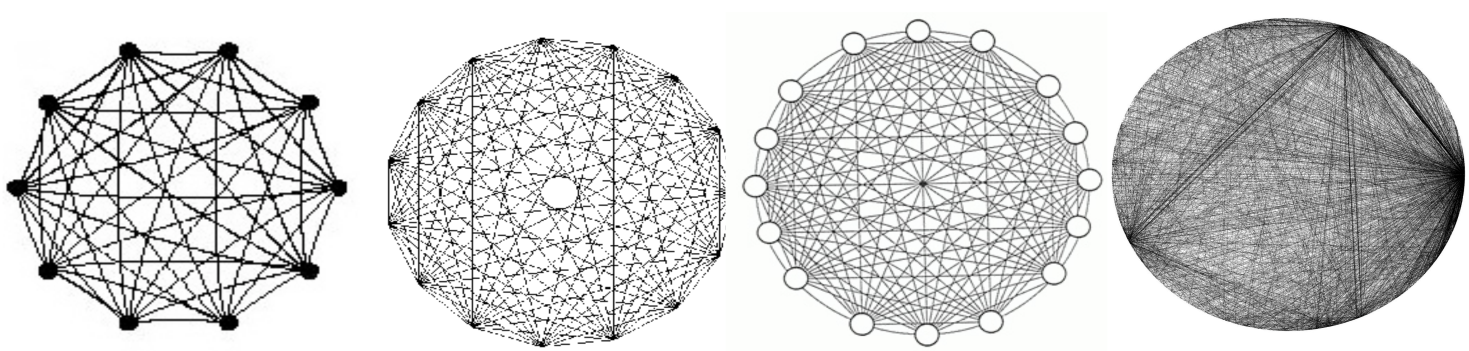


Existing models produce behaviorally homogeneous users and lack the richness of behavioral structure in real traces. Richer models are needed !

Behavioral Similarity Graphs



(a) Dartmouth Campus (b) MIT Campus (c) UF Campus (d) USC Campus



Random and community models produce fully connected similarity graphs

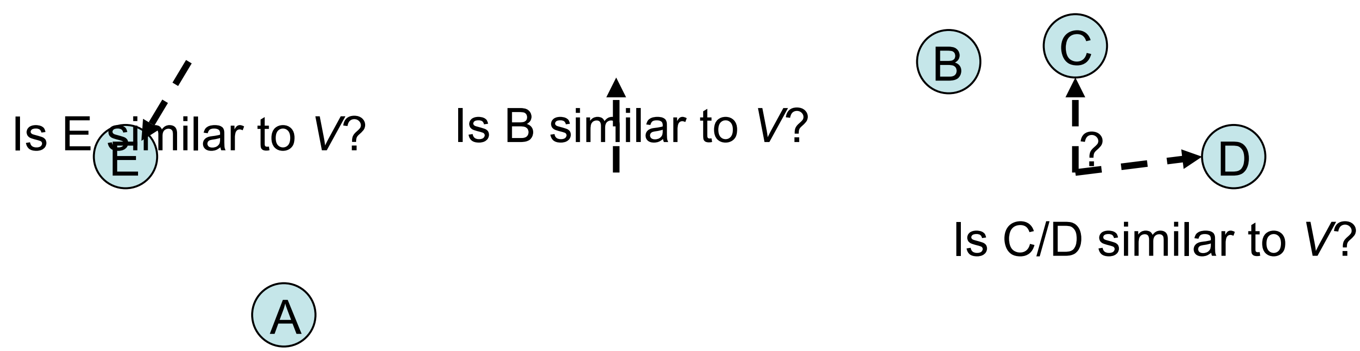
G. Thakur, A. Helmy, W. Hsu, “Similarity analysis and modeling of similarity in mobile societies: The missing link”, UF Tech Report, Jun 2010

Profile-cast: A New Communication Paradigm

W. Hsu, D. Dutta, A. Helmy, *ACM Mobicom* 2007, *WCNC* 2008, *Trans. Networking* To appear

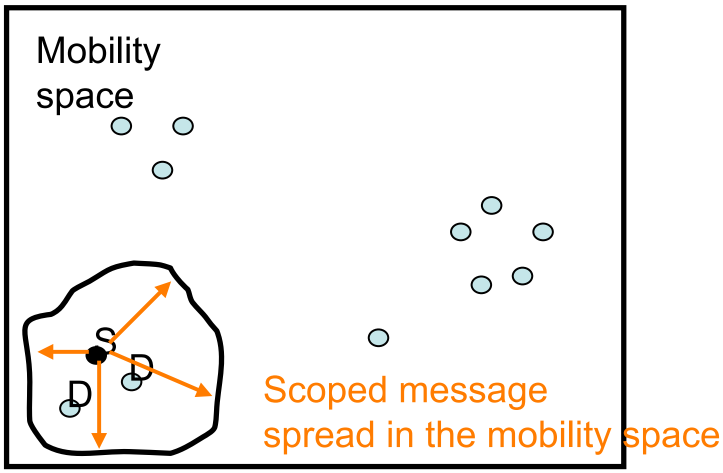
Payload	Dest Address	→	Payload	Target Profile

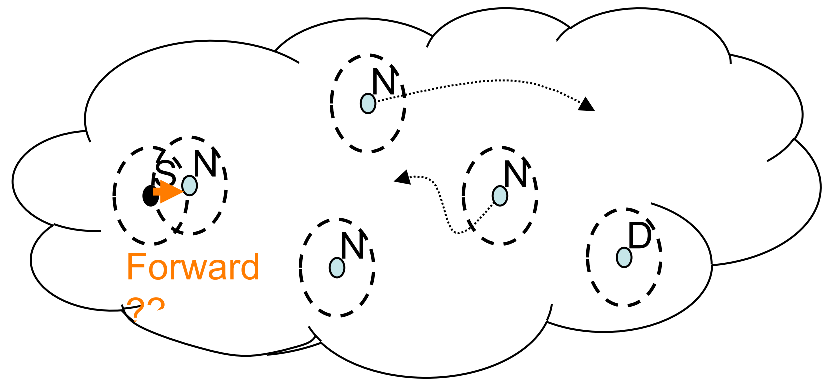
- Sending messages to others with similar behavior, without knowing their identity
  - Announcements to users with specific behavioral profile *V*
  - Interest-based ads, similarity resource discovery
- For Delay Tolerant Networks (DTNs)



Profile-cast Use Cases

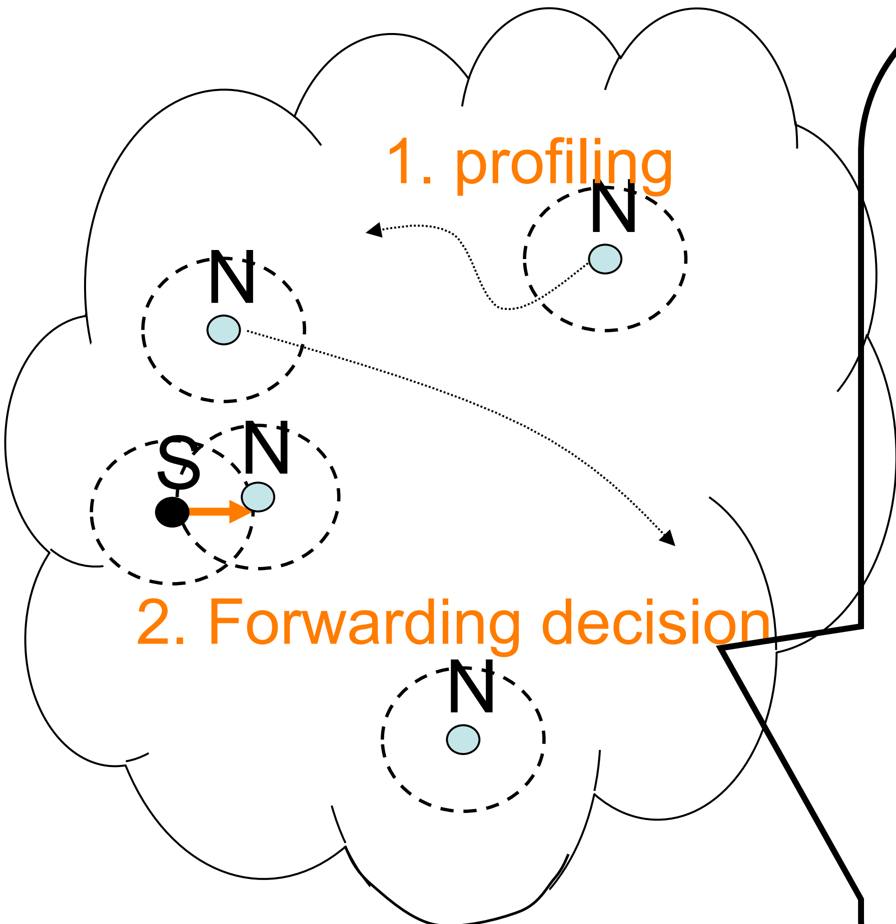
- Mobility-based *profile-cast* (Target mode)
  - Targeting group of users who move in a particular pattern (lost-and-found, context-aware messages, moviegoers)
  - Approach: use “similarity metric” between users





- Mobility-independent *profile-cast* (Dissemination mode)
  - Targeting people with a certain characteristics independent of mobility (classic music lovers)
  - Approach: use “Small World” encounter patterns

Profile-cast Operation

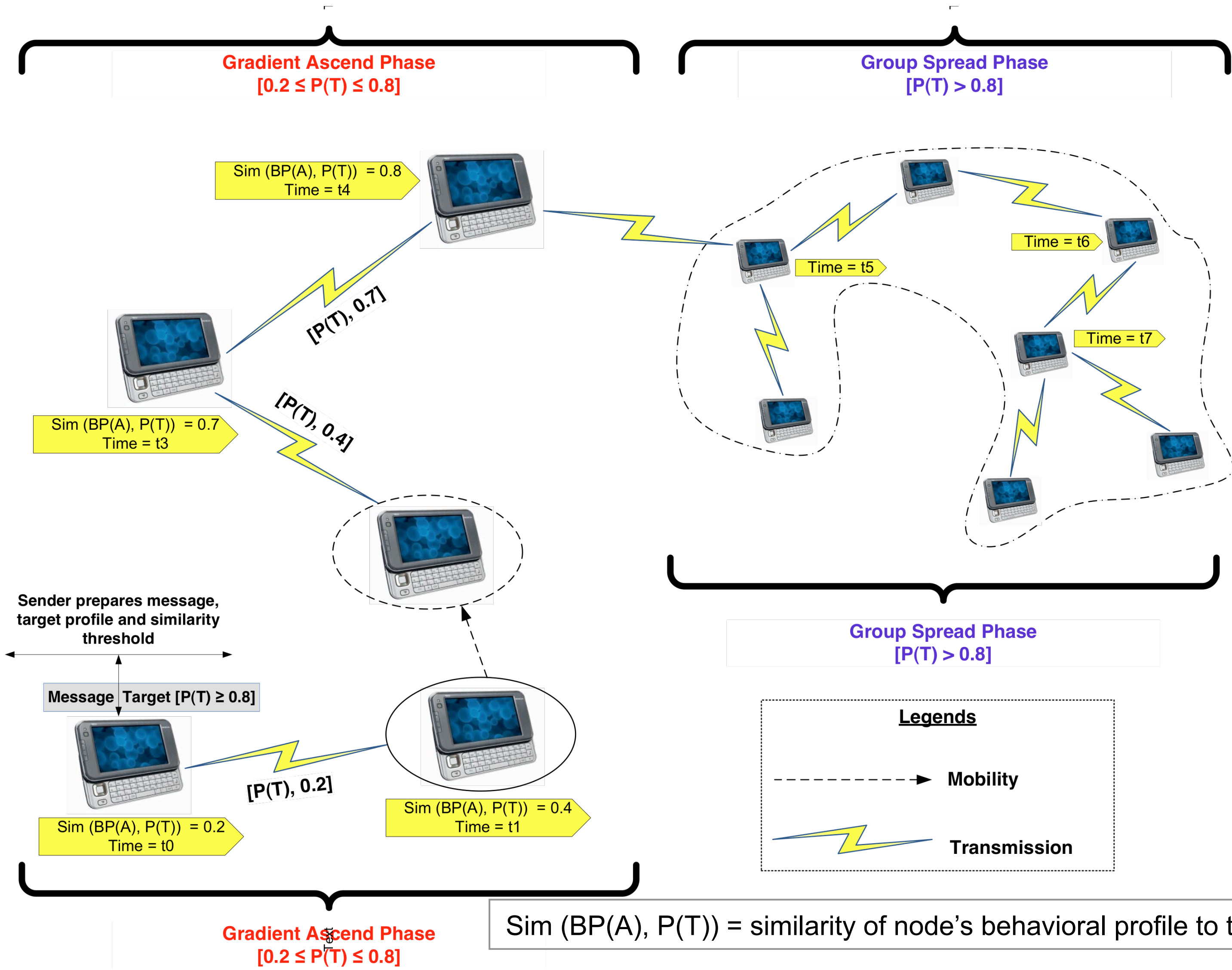


- Determining user similarity
  - ***S*** sends Eigen behaviors for the *virtual* profile to ***N***
  - ***N*** evaluated the similarity by weighted inner products of Eigen-behaviors

$$Sim(U,V) = \sum_{\forall i,j} w_i w_j \left| u_i \cdot v_j \right|$$

  - Message forwarded if Sim(U,V) is high (the goal is to deliver messages to nodes with similar profile)
  - Privacy conserving: ***N*** and ***S*** do not send information about their own behavior

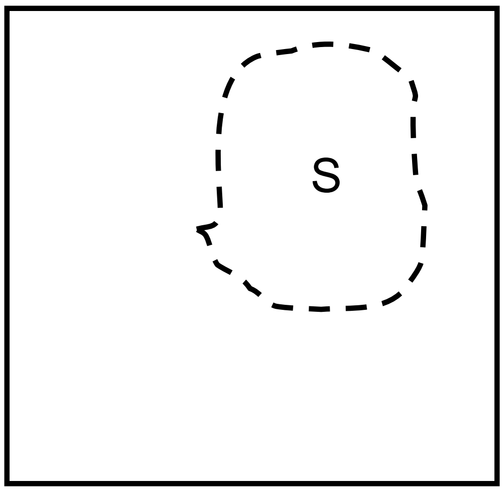
Profile-cast CSI protocol: Target-mode



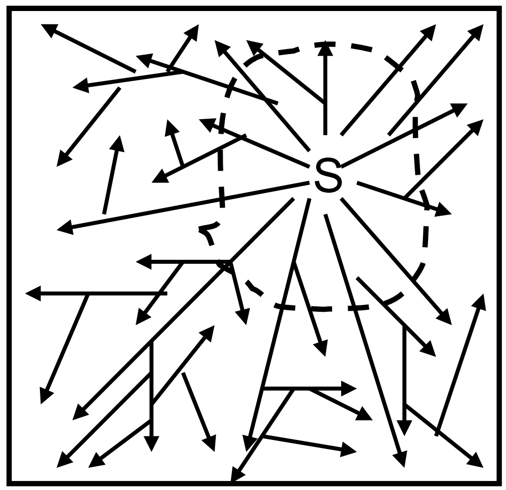
Sim (BP(A), P(T)) = similarity of node's behavioral profile to the target profile

Mobility Profile-cast (intra-group)

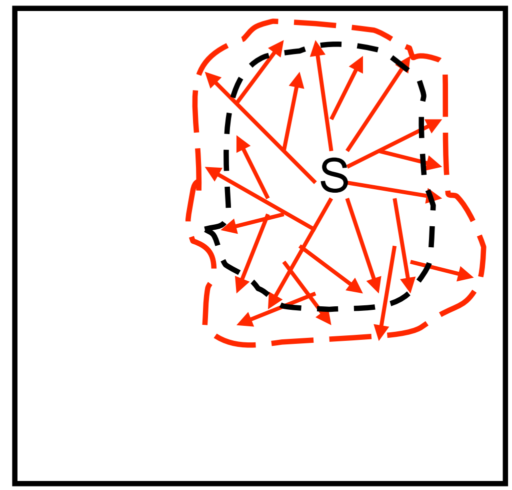
Goal



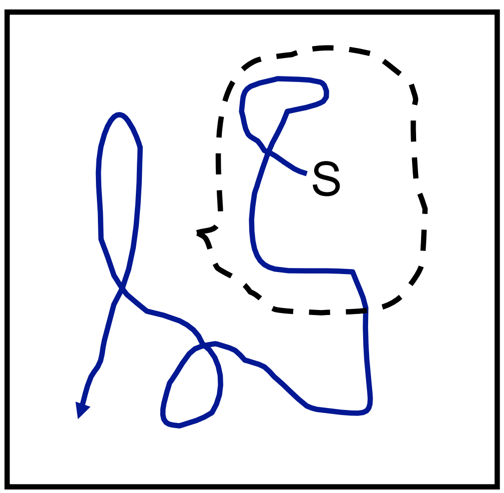
Epidemic



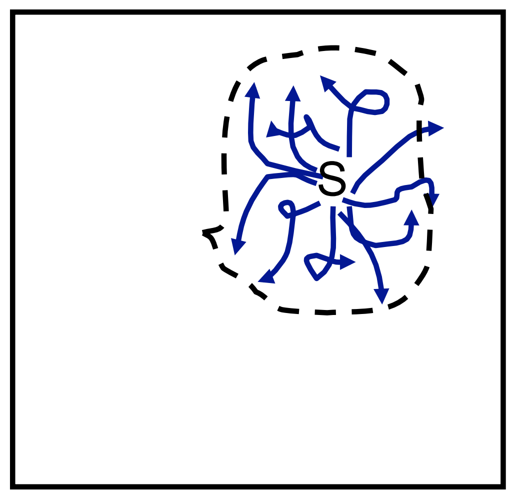
Group-spread



Single long random walk

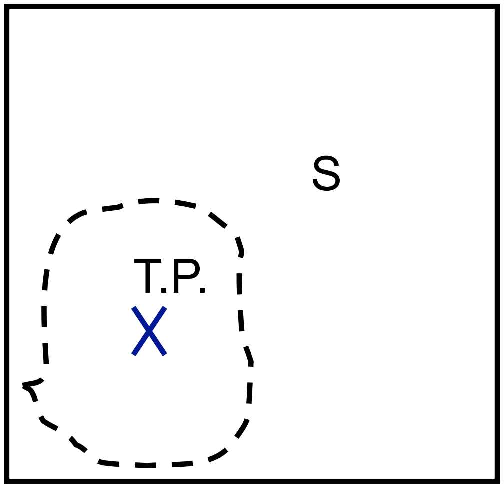


Multiple short random walks

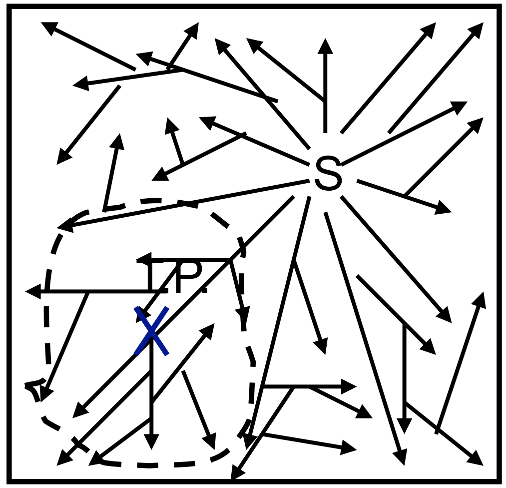


Mobility Profile-cast (inter-group)

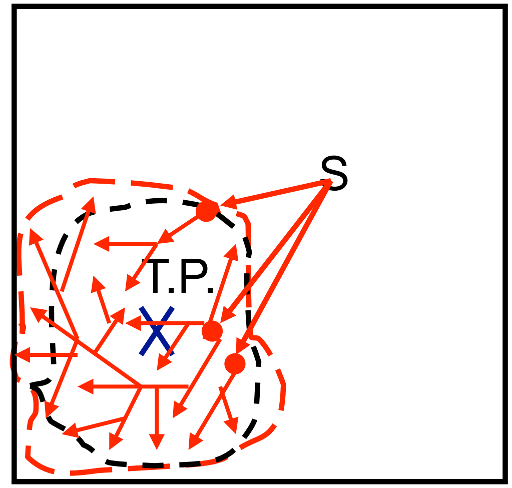
Goal



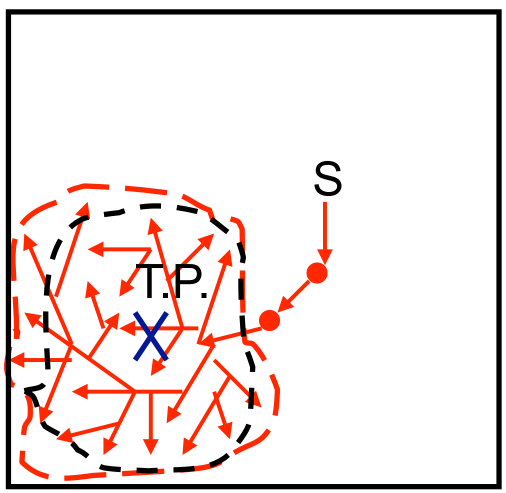
Epidemic



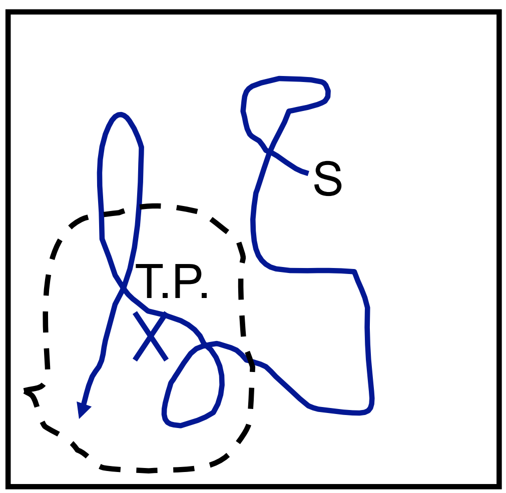
Group-spread



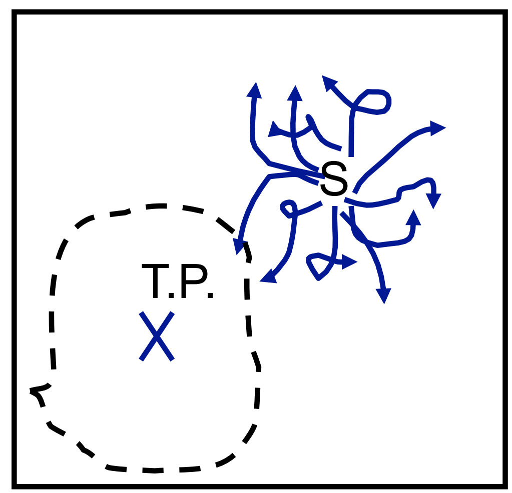
Gradient-ascend



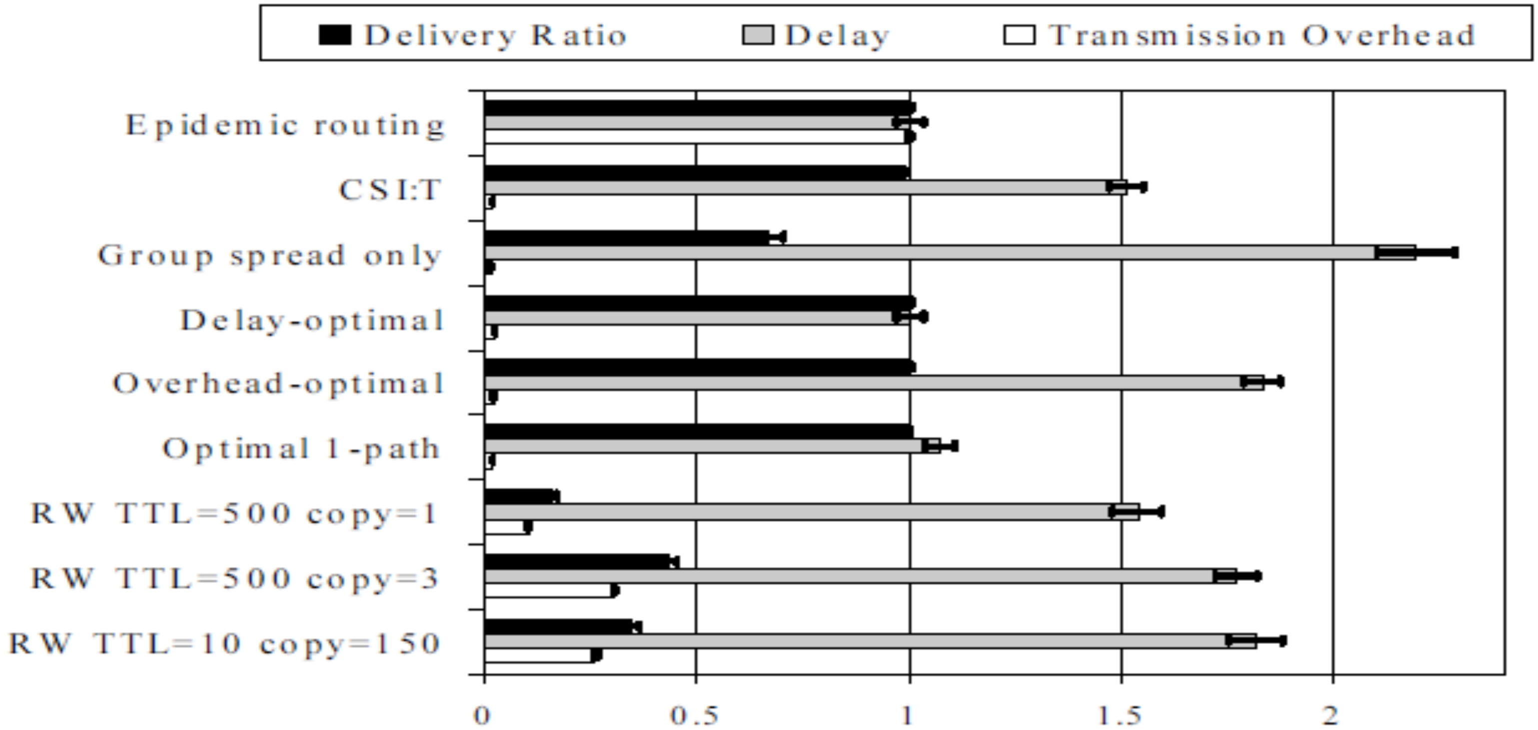
Single long random walk



Multiple short random walks



Profile-cast Evaluation



(a) USC.

* Results presented as the ratio to epidemic routing

- Over 96% delivery ratio – Over 98% reduction in overhead w.r.t. Epidemic
- RW < 45% delivery
- Strikes a near optimal balance between delivery, overhead and delay
- Other variants (e.g., multi-copy, simulated annealing) under investigation

Video

Extending Interest, Behavior Beyond Mobility

- In addition to mobility, user's web access and traffic patterns, applications used (among others) represent other dimensions of interest and behavior
- Further analysis of network measurements (e.g., *Netflow*) can reveal behavioral characteristics in these dimensions
- Netflow traces are 3 orders of magnitude larger than WLANs (*WLANs*: dozens of millions, *Netflows*: dozens of billions)
- New challenges in mining 'big data' to get information

	Duration	Records	Total Users	Access points ports
USC-WLAN	Dec 03-Jun 08	50 M	55,500	79 ports (03), 161 (08)
USC-DHCP	Dec 03-Jun 08	60 M	55,500	79 ports (03), 161 (08)
USC-netflow	Apr 05-Jun 08	50 B	50,000	161 ports
UF-WLAN	Jun 07-Current	60 M	140,000	784 Access points
UF-DHCP	Jun 07-Current	13 M	140,000	784 Access points
UF-netflow	starting Nov 09	2.5B/month	45,000	784 Access points

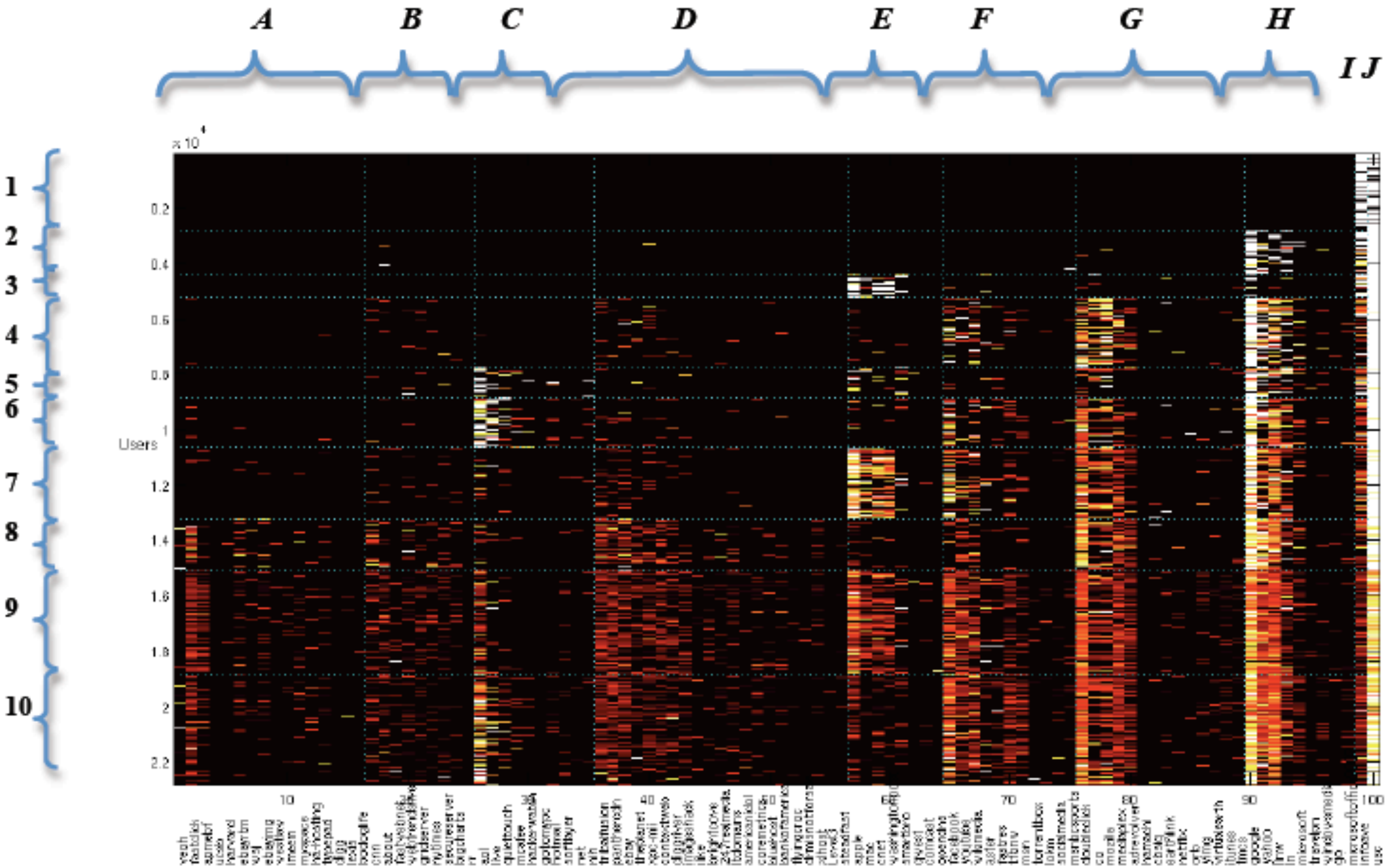


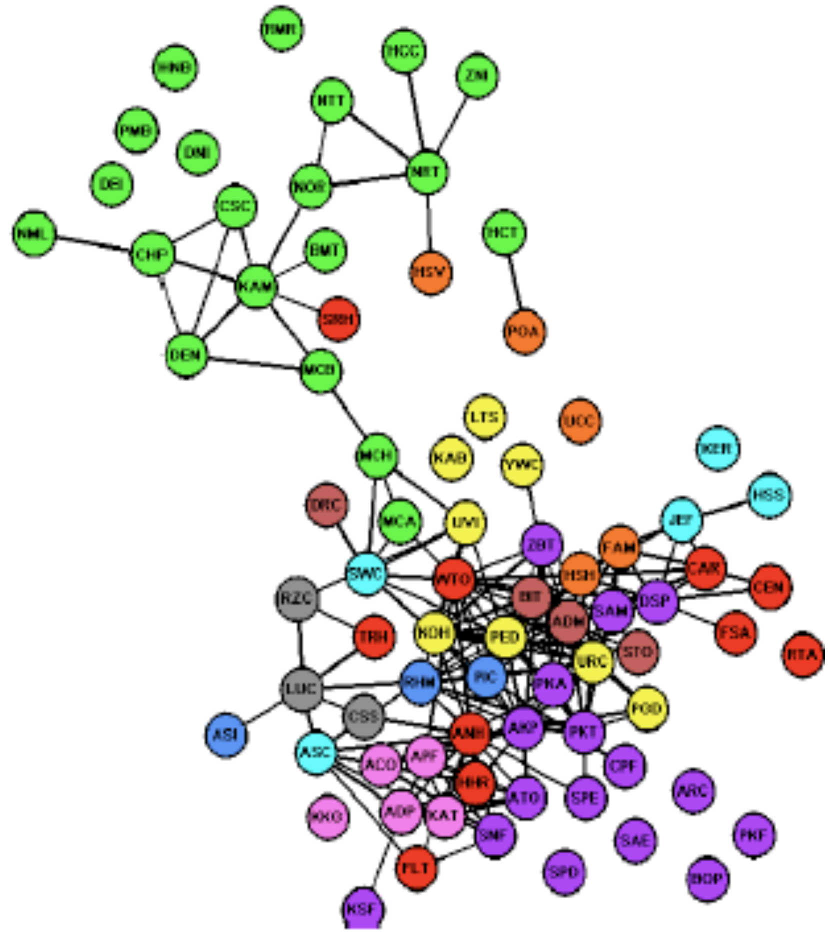
Fig. 5. Information theoretic co-clustering on user-domain matrix, March 2008. The result is given for ten clusters of users (1 through 10) and ten clusters of domains (*A* through *J*). Domain clusters *I* and *J* include one domain each.

S. Moghaddam, A. Helmy, S. Ranka, M. Somaya, “Data-driven Co-clustering Model of Internet Usage in Large Mobile Societies”, UF Tech Report, May 2010

Web-usage Spatio-temporal multi-D Clustering

Table 2 -Major related websites which are clustered together

Cluster	Domains
A	myspace – imeem (social media service) - digg (social news) – typepad (blogging service) ebayrtm - ebayimg - wsj (business news) - bodoglife(online gampling) ucsb - harward - westlaw
B	cnn - new york times
C	mcafee - hackerwatch live - hotmail
D	ebay - bankofamerica
E	apple – mac washingtonpost - cnet
F	facebook – youtube - social media msn - msnbcsports
G	netflix – itunes - orb (media cast) tmcs (social city search) - virtualearth (online map)
H	google – yahoo microsoft - windowsmedia - microsoftoffice2007

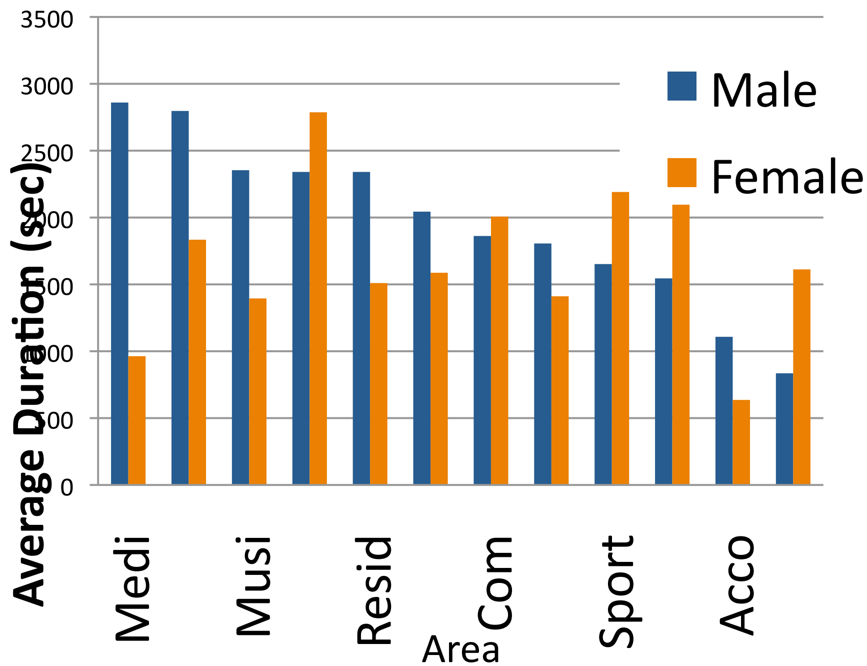


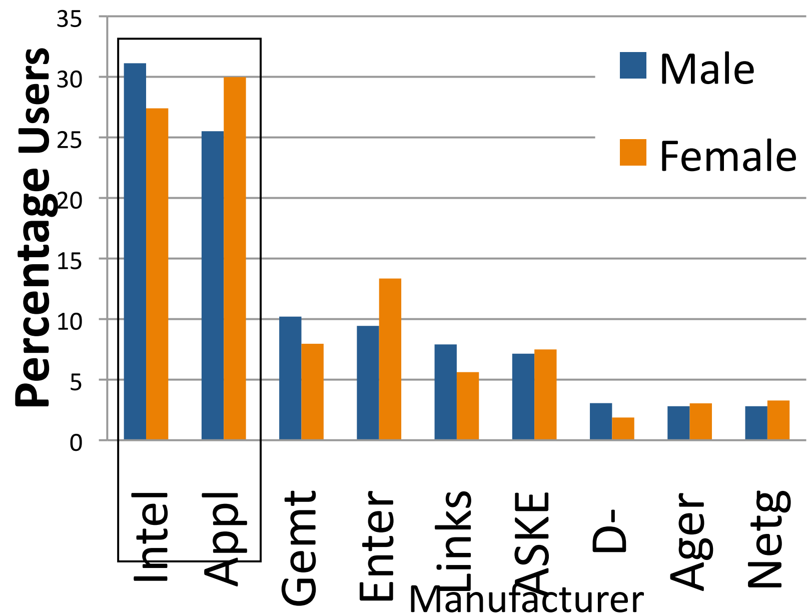
Clustering of Locations based on web access (similar locations coded with same color)

- Users can be consistently modeled using few (~10) clusters with disjoint profiles.
- Access patterns from multiple locations show clustered distinct behavior.

Gender-based feature analysis in Campus-wide WLANs

U. Kumar, N. Yadav, A. Helmy, Mobicom 2007, Crawdad 2007





- Able to classify users by gender using knowledge of campus map

-Users exhibit distinct on-line behavior, preference of device and mobility based on gender

-On-going Work

-How much more can we know?

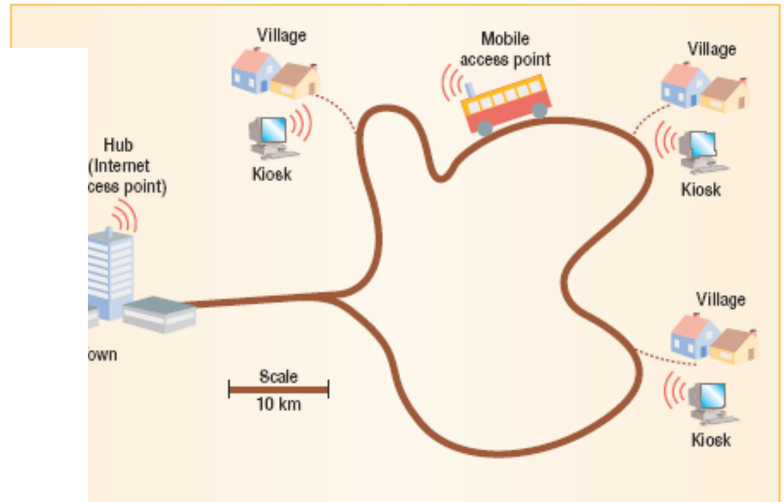
-What is the *"information-privacy trade-off"*?

Future Directions (Applications)

- Behavior aware push/caching services (targeted ads, events of interest, announcements)
- Caching based on behavioral prediction
- Detecting abnormal user behavior & access patterns based on previous profiles
- Can we extend this paradigm to include social aspects (trust, friendship, cooperation)?
- Privacy issues and mobile *k-anonymity*
- Participatory sensing, deputizing the community

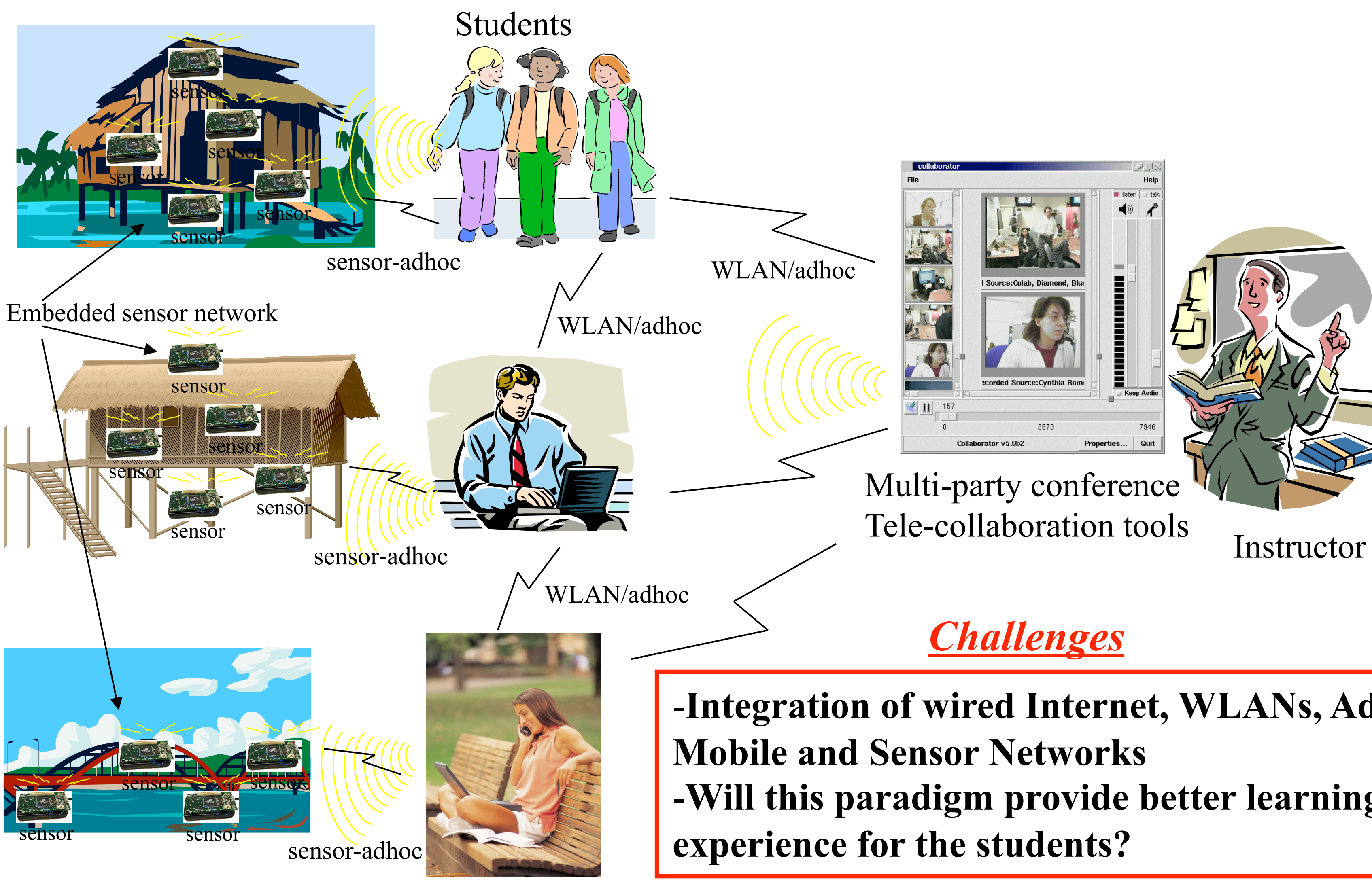
Disaster Relief (Self-Configuring) Networks

On-going and Future Directions Utilizing mobility



- Controlled mobility scenarios
  - DakNet, Message Ferries, Info Station
- Mobility-Assisted protocols
  - Mobility-assisted information diffusion: EASE, FRESH, DTN, $100 laptop
- Context-aware Networking
  - Mobility-aware protocols: self-configuring, mobility-adaptive protocols
  - Socially-aware protocols: security, trust, friendship, associations, small worlds
- On-going Projects
  - Next Generation (Boundless) Classroom
  - Disaster Relief Self-configuring Survivable Networks

The Next Generation (Boundless) Classroom



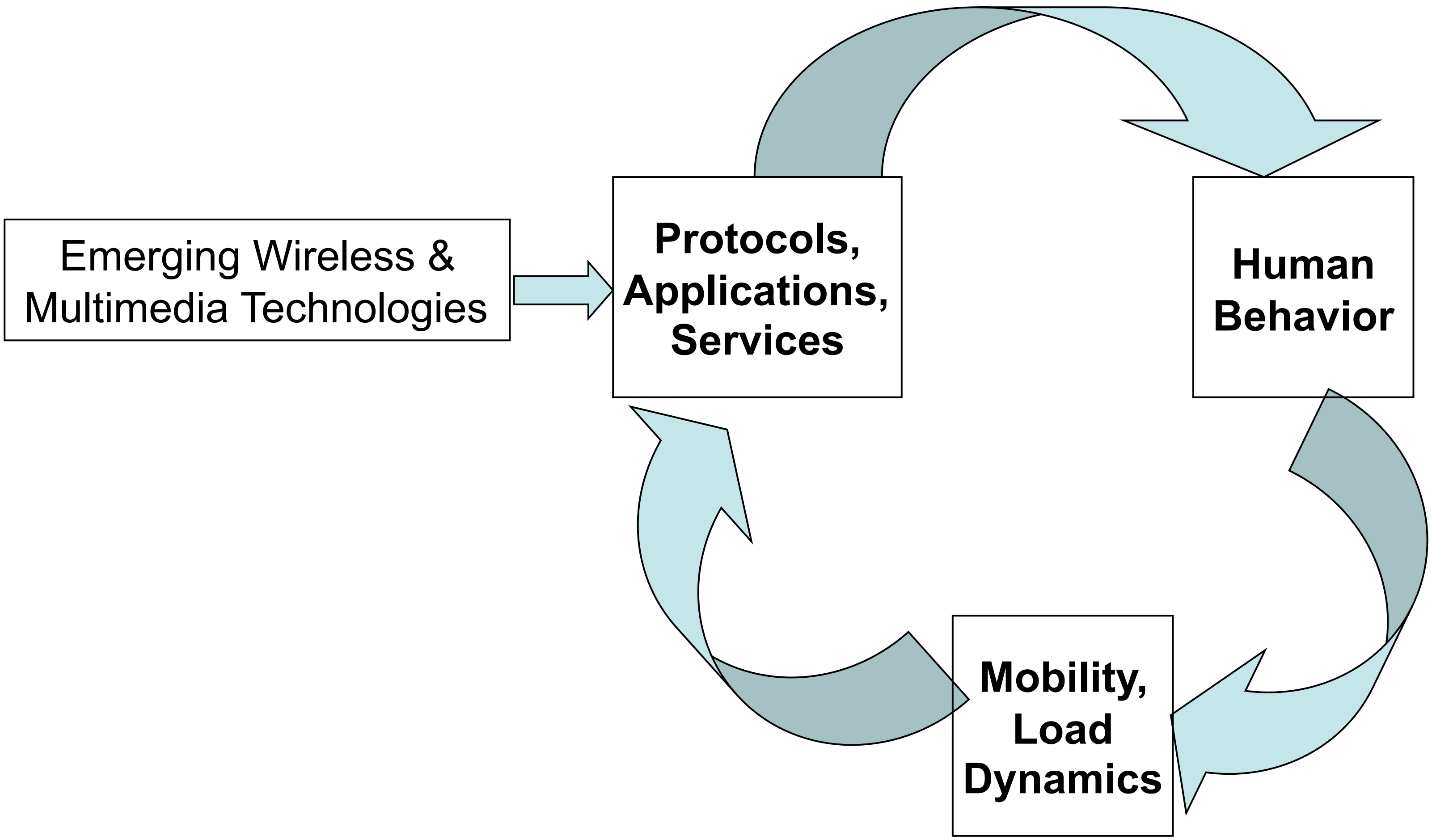
Challenges

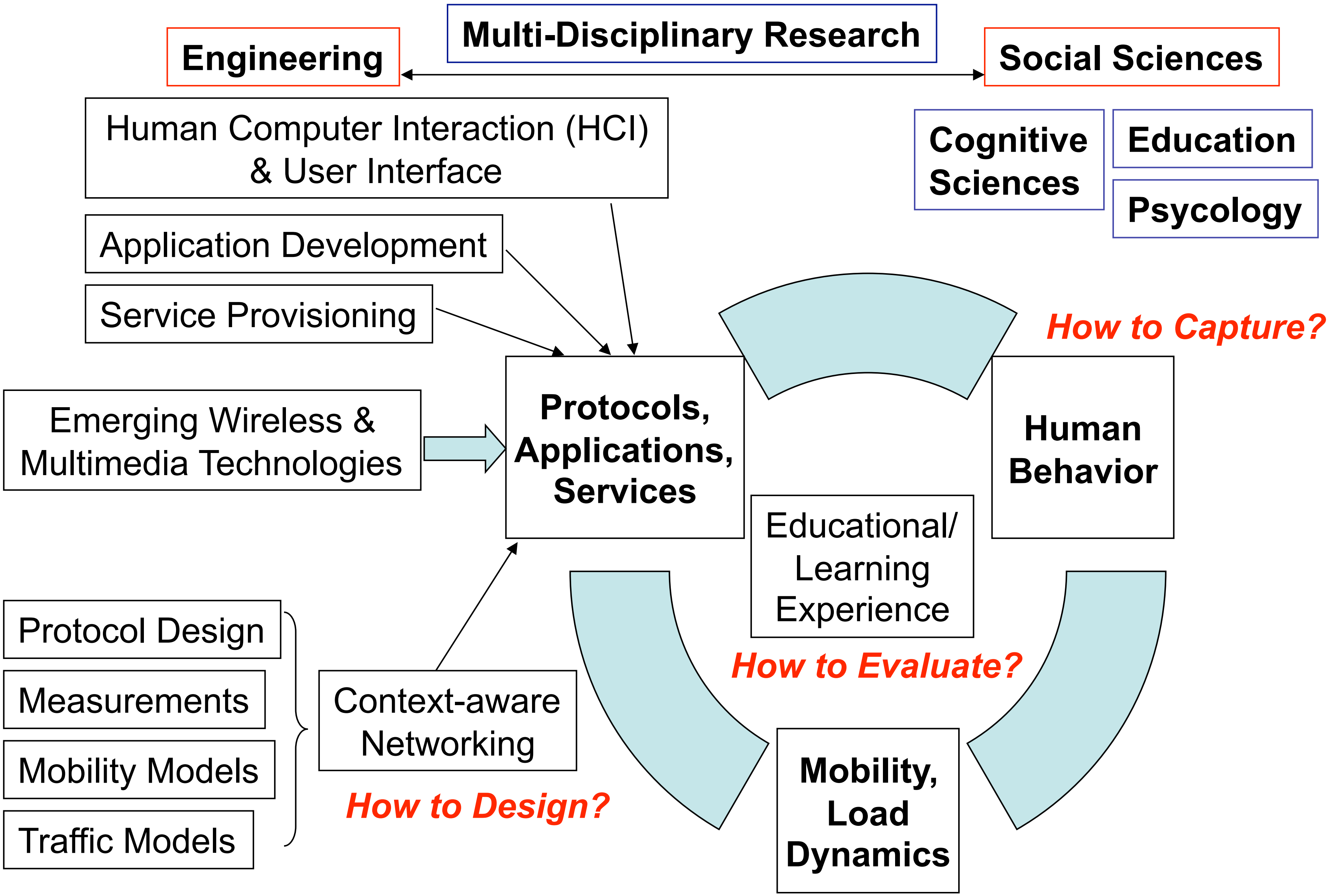
-Integration of wired Internet, WLANs, Adhoc Mobile and Sensor Networks
-Will this paradigm provide better learning experience for the students?

Real world group experiments (structural health monitoring)

Future Directions: Technology-Human Interaction

The Next Generation Classroom




Engineering
Multi-Disciplinary Research
Social Sciences
Human Computer Interaction (HCI) & User Interface
Cognitive Sciences
Education
Psycology
Application Development
Service Provisioning
How to Capture?
Emerging Wireless & Multimedia Technologies
Protocols, Applications, Services
Human Behavior
Educational/ Learning Experience
Protocol Design
Measurements
Mobility Models
Traffic Models
Context-aware Networking
How to Evaluate?
How to Design?
Mobility, Load Dynamics

Thank you!

Ahmed Helmy helmy@ufl.edu
URL: www.cise.ufl.edu/~helmy
MobiLib: nile.cise.ufl.edu/MobiLib