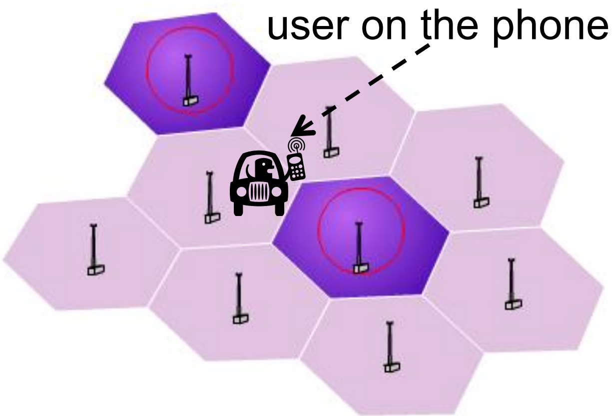
Using Complex Networks for Mobility Modeling and Opportunistic Networking

Many domains

Traffic Regulation

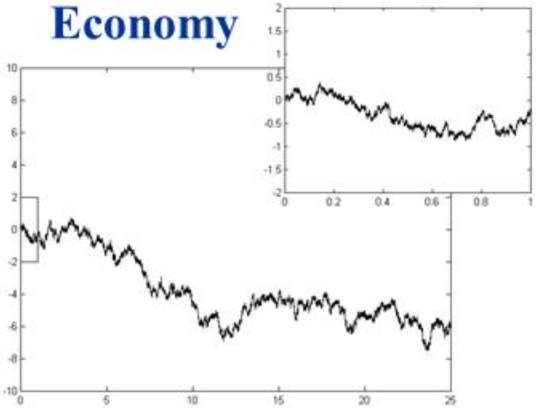
Public Transport Design

Network Design



Architecture/ Urban Planning

Economy



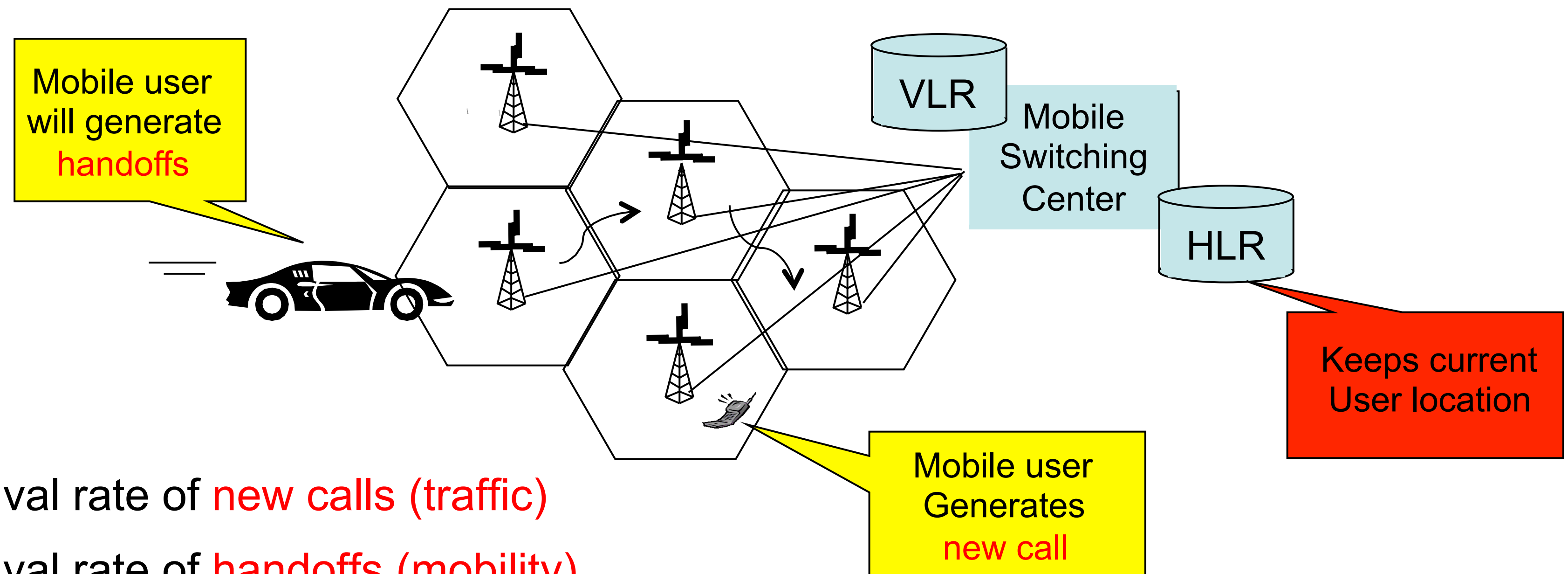
Gaming

Rationale for Mobility Models

- Mobility Models are required for
  - Performance evaluation
    - Analytical
      - A system dynamics must be tractable in order to derive characteristics of interest
    - Simulations
      - Often used as an alternative when models are too complex (no analytical derivation)
      - But still complementary to the analytical approach
    - Trace-replaying and experiments
  - Solution design
    - Networking solutions should be designed according to their *in situ* environment (i.e., mobility context and characteristics)

Evaluation of Cellular Networks

- Aim at providing integrated communications (i.e., voice, video, and data) between nomadic subscribers in a seamless fashion

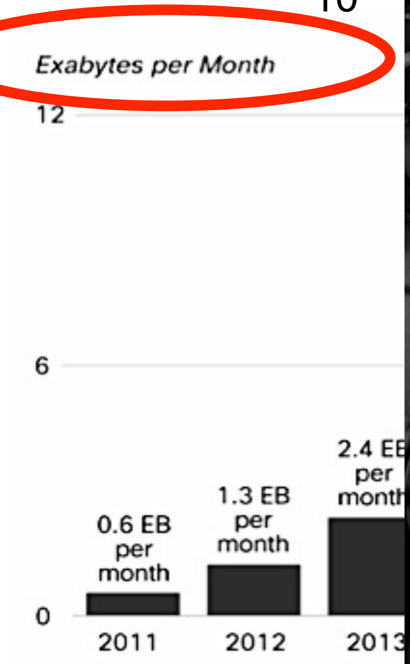


- Input
  - Arrival rate of new calls (traffic)
  - Arrival rate of handoffs (mobility)
- Output
  - HLR load, probability of call rejection

The Mobile Internet

Cisco: Global Mobil

10^{18}



demand in

other 4G)

its to capacity!

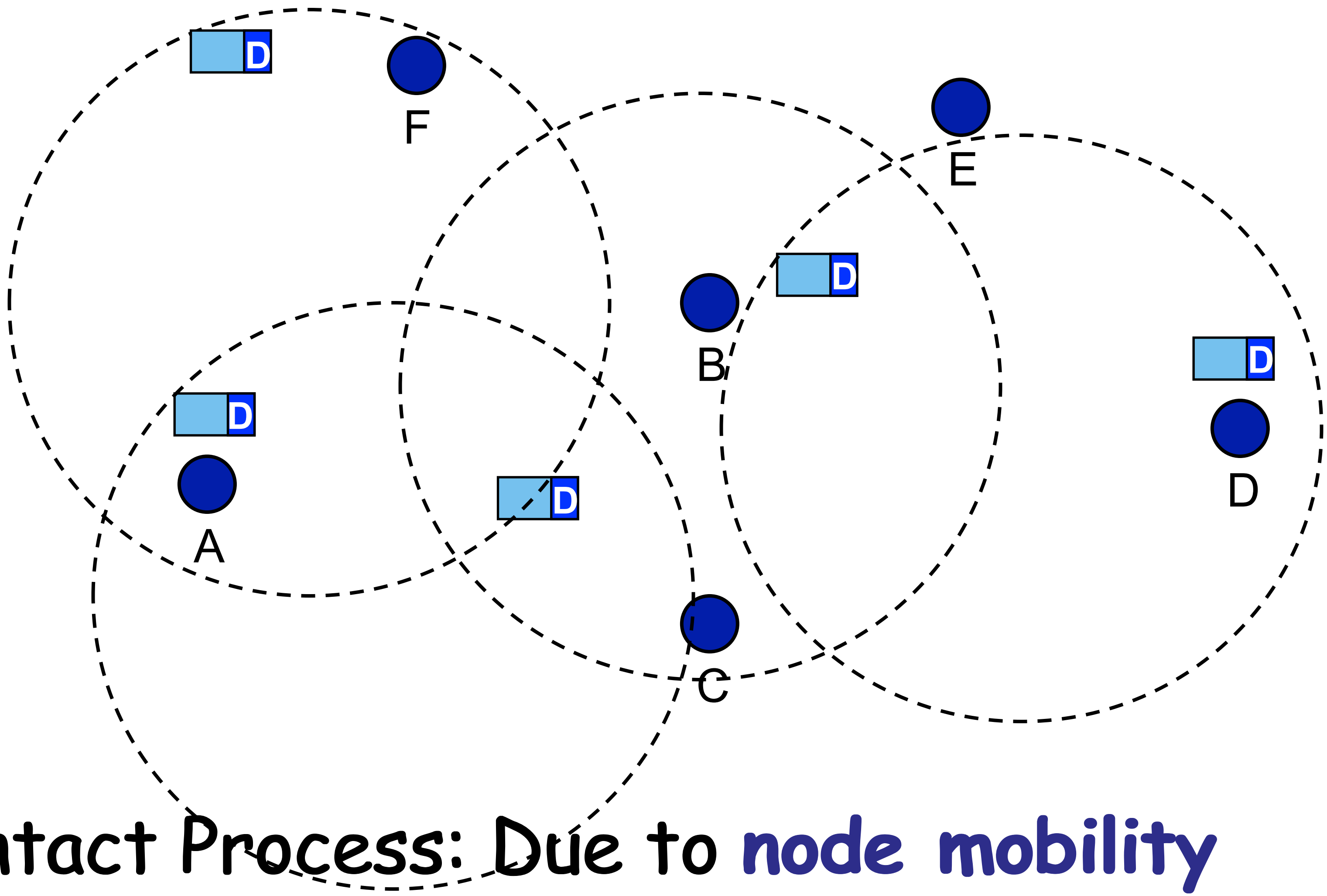
s)

Cost?

Spars... regions

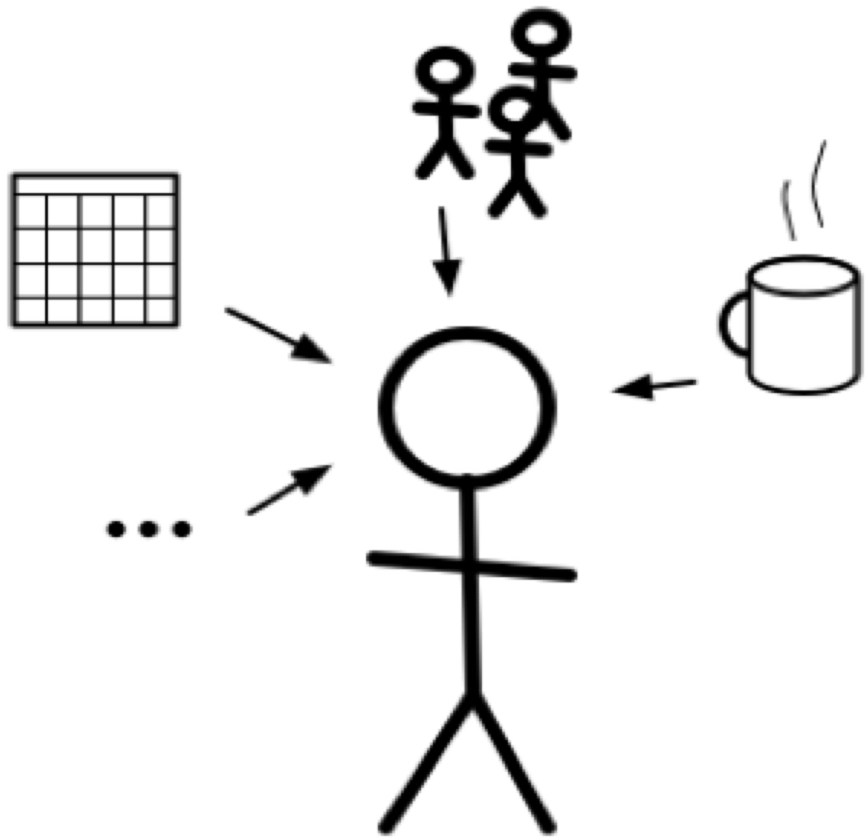
Device-to-Device Communication (e.g. Bluetooth or WiFi Direct)

Data/Malware Spreading Over Opp. Nets



- Contact Process: Due to **node mobility**
- Q: How long until X% of nodes "infected"?

Understanding mobility is complex



Classification of Mobility Models

- Scale
  - Microscopic
    - accurately describes the motion of mobile individuals
  - Macroscopic
    - considers the displacement of mobile entities (e.g., pedestrians, vehicles, animals) at a coarse grain, for example in the context of large geographic areas such as adjacent regions or cells

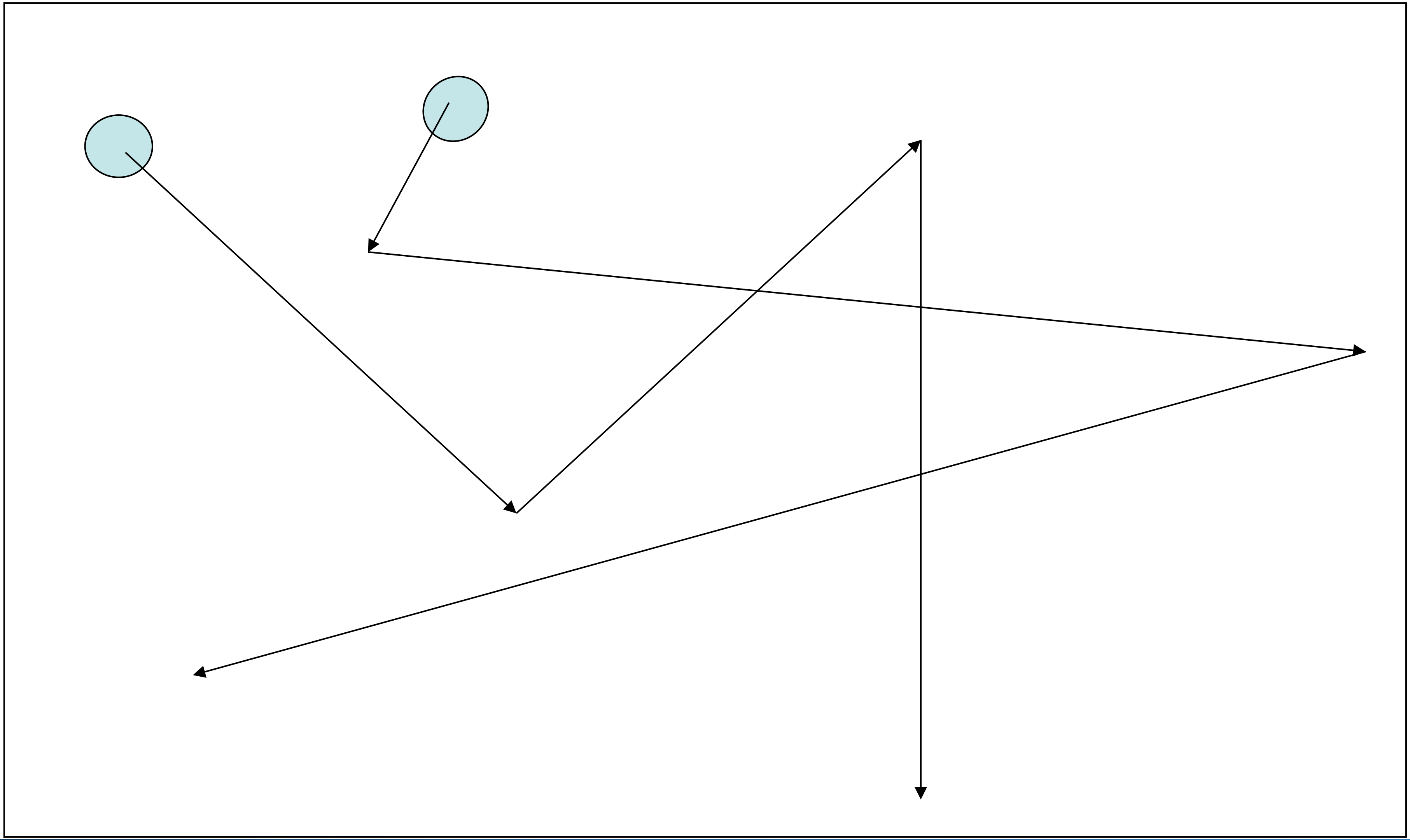
- Inputs
  - Standard Parameters: speed, direction, …
  - Additional Inputs: map, topology, preferred/popular locations…
  - Behavioral: intention, social relations, time-of-day schedule,…
  - Inherent Randomness: stochastic models (Markov, ODEs, Queuing)

I. Random Waypoint (*RWP*) Model

1. A node chooses a random destination anywhere in the network field
2. The node moves towards that destination with a velocity chosen randomly from [0, *Vmax*]
3. After reaching the destination, the node stops for a duration defined by the "pause time" parameter.
4. This procedure is repeated until the simulation ends

- Parameters: Pause time T, max velocity *Vmax*
- Comments:
  - Speed decay problem, non-uniform node distribution
  - Variants: random walk, random direction, smooth random, ...

Random Way Point: Basics



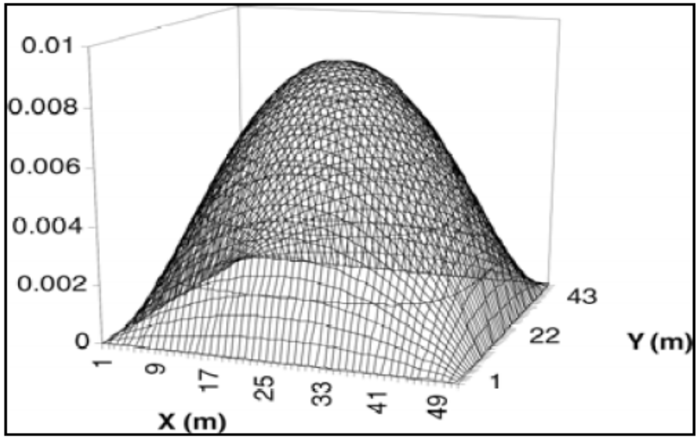


Figure 1-3. Node Spatial Distribution (Square Area)

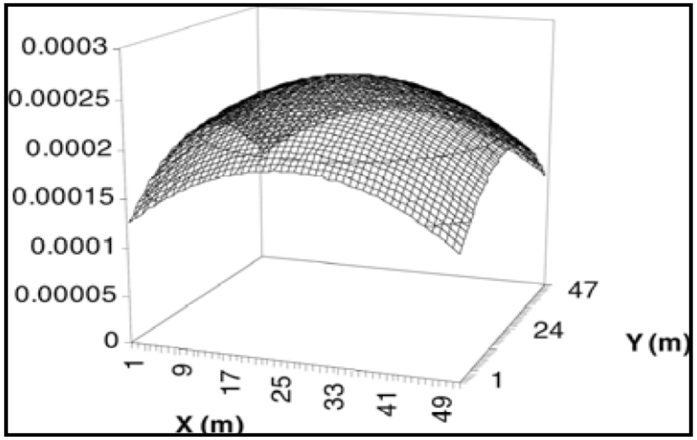


Figure 1-4. Node Spatial Distribution (Circular Area)

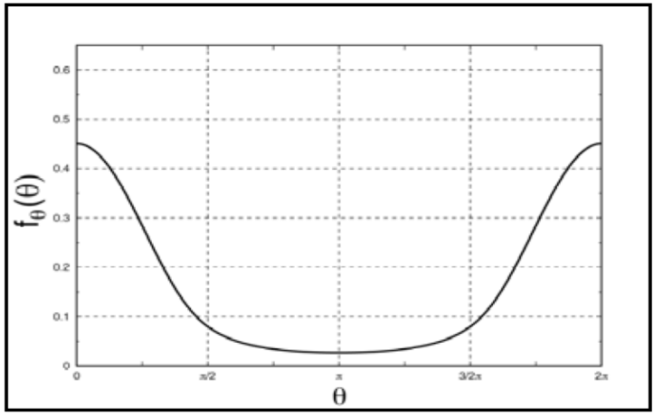


Figure 1-5. The probability distribution of movement direction

-1- *RWP* leads to non-uniform distribution of nodes due to bias towards the center of the area, due to non-uniform direction selection. To remedy this the "random direction" mobility model can be chosen.

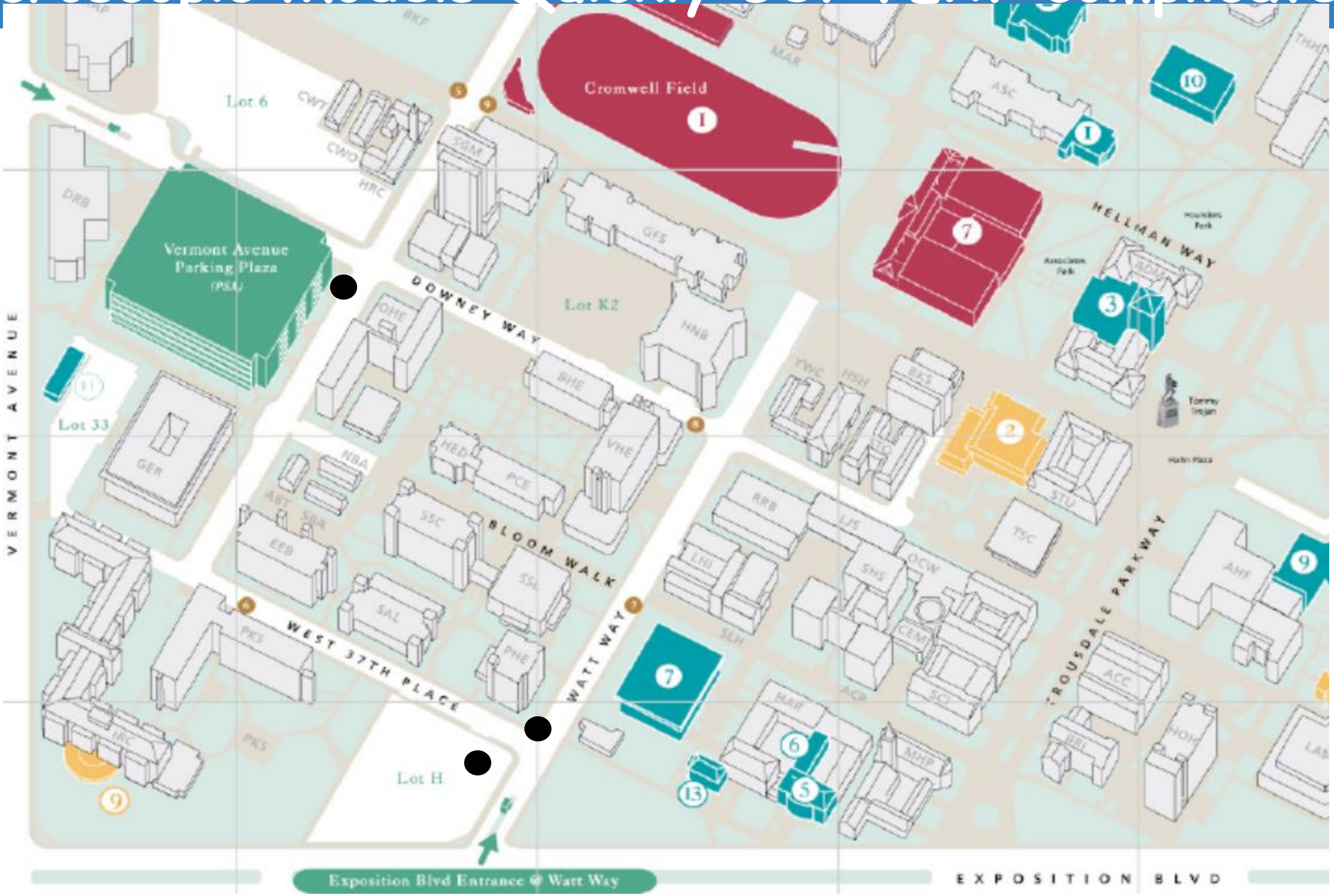
-2- Average speed decays over time due to nodes getting 'stuck' at low speeds

II. Random (RWK) Walk Model

- Similar to RWP but
  - Nodes change their speed/direction every time slot
  - New direction θ is chosen randomly between $(0,2\pi]$
  - New speed chosen from uniform (or Gaussian) distribution
  - When node reaches boundary it bounces back with $(\pi-\theta)$

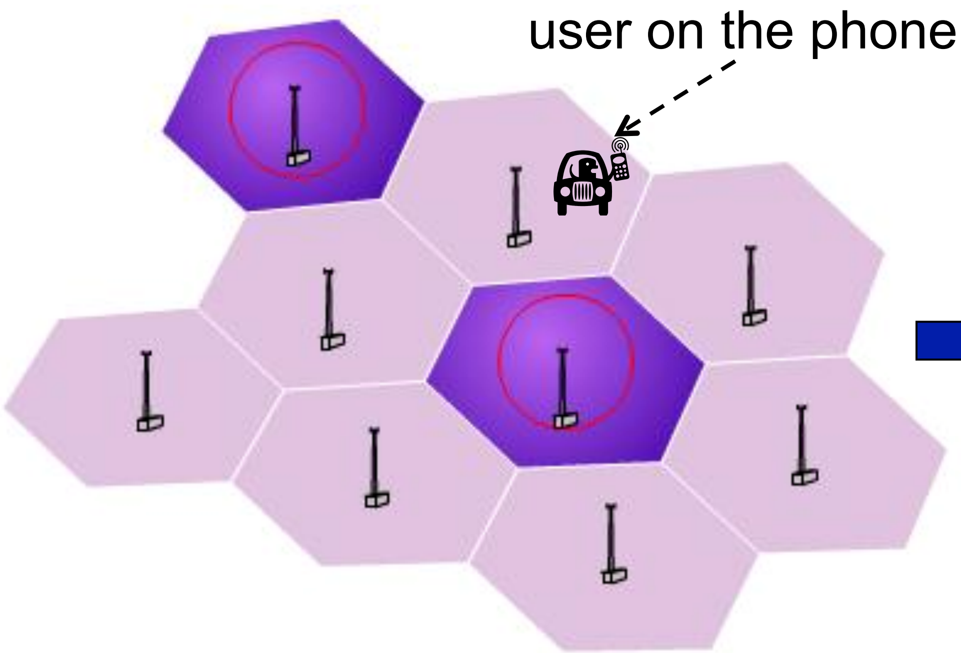
Random Walk

Microscopic Models Quickly Get VERY complicated!

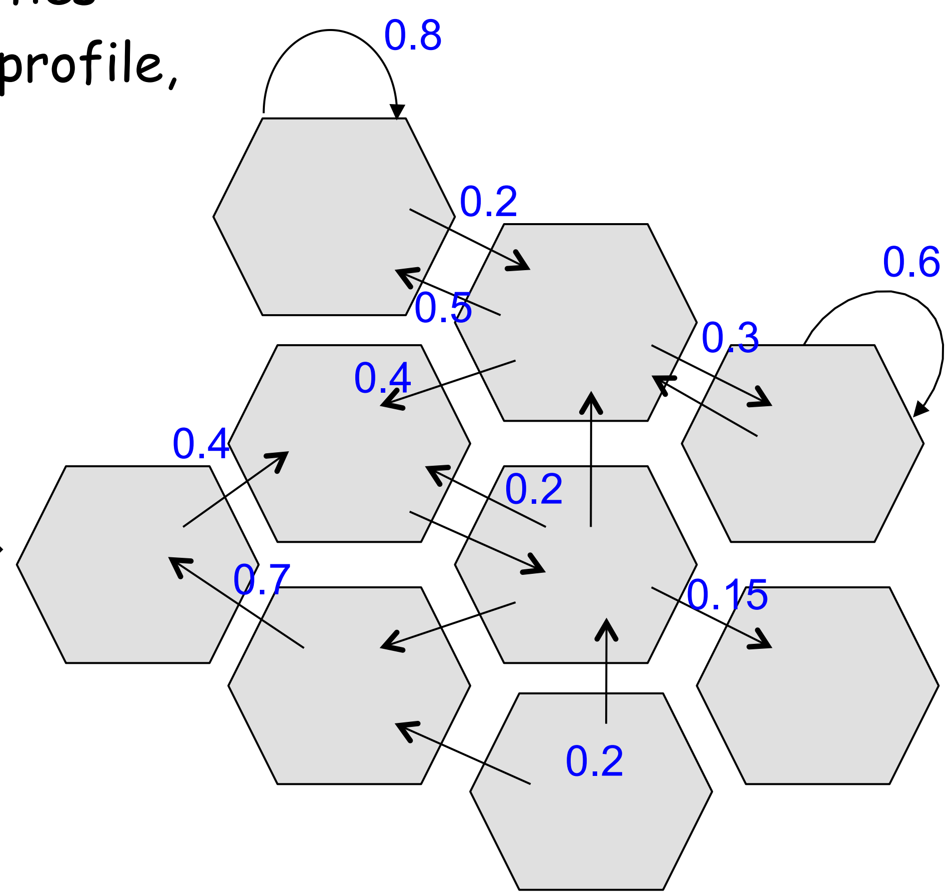


A Macroscopic Mob. Model for Cell. Networks

- A simple handover model
  - cell -> state
  - need to find transition probabilities
  - depend on road structure, user profile, statistics



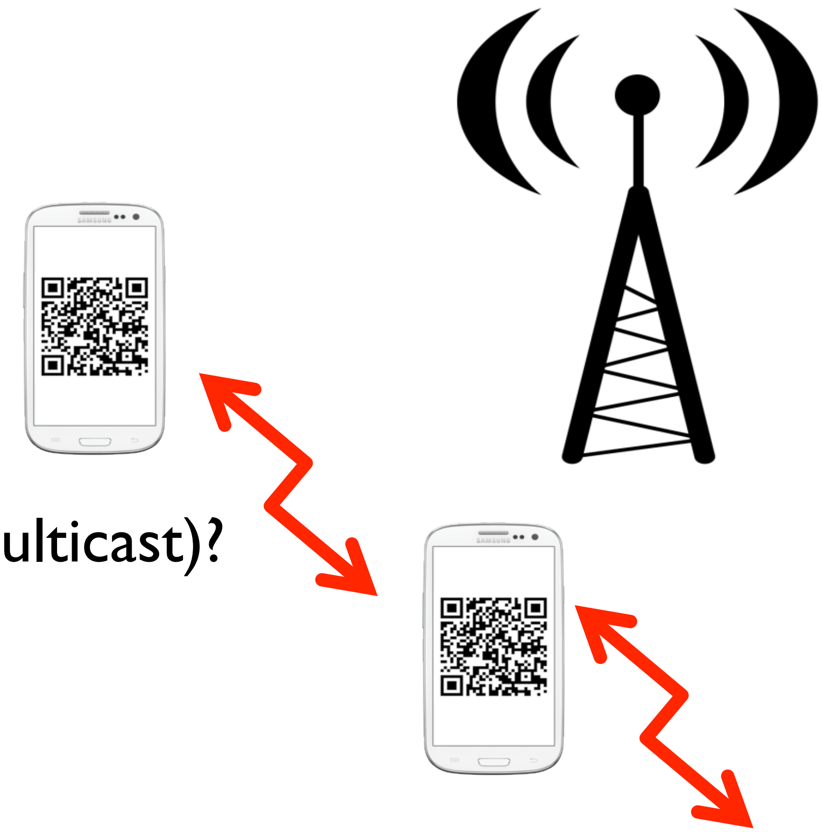
Cellular Network



Markov Chain

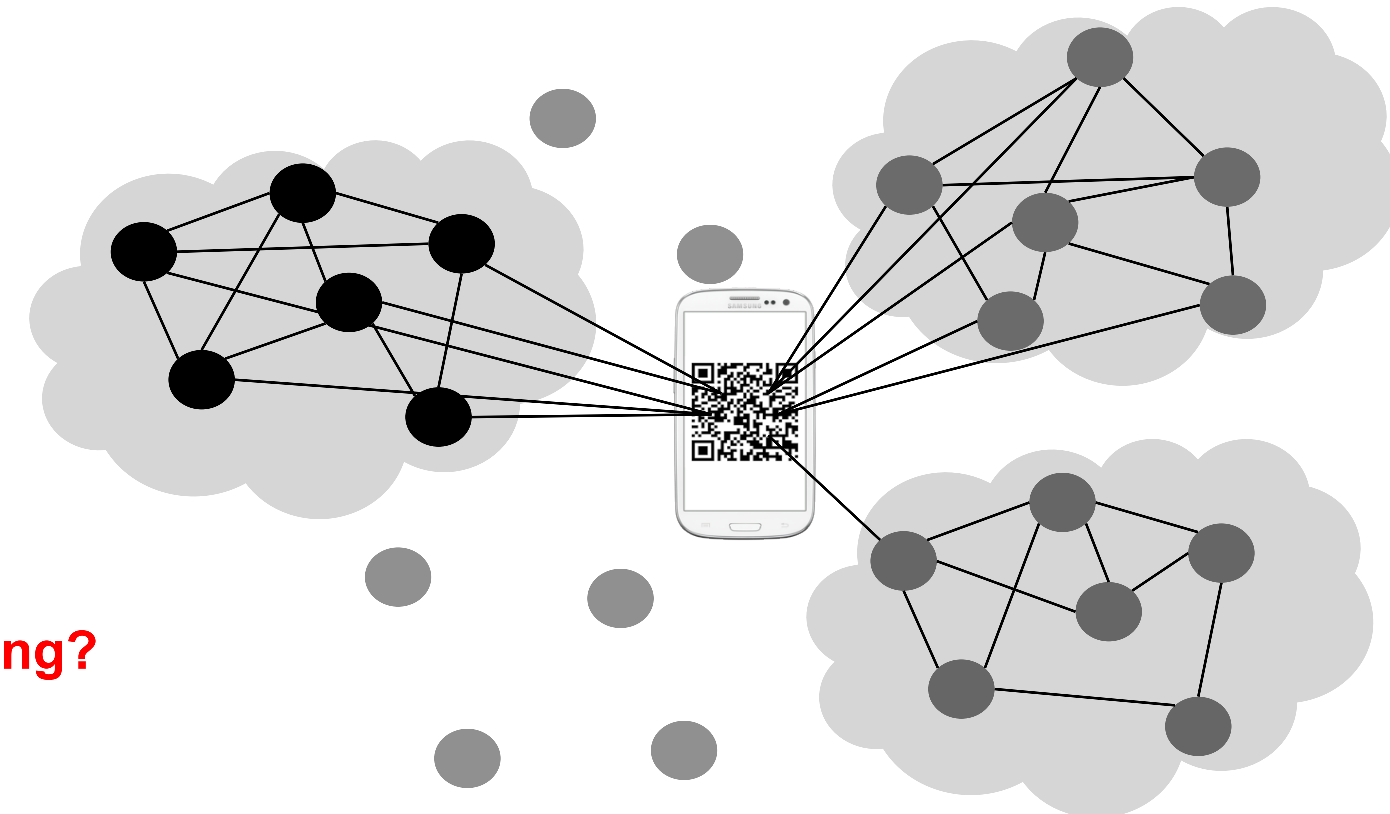
Opportunistic networks: Macroscopic View

- Connect devices to each other
  - Bluetooth, WiFi direct
- It's all about Contacts!
  - Opportunities to exchange data
- How do we route messages (unicast/multicast)?
- Whom do we trust?
- Where do we place services?



Contacts are driven by our mobility in a social context!

The Contact Graph



Routing?

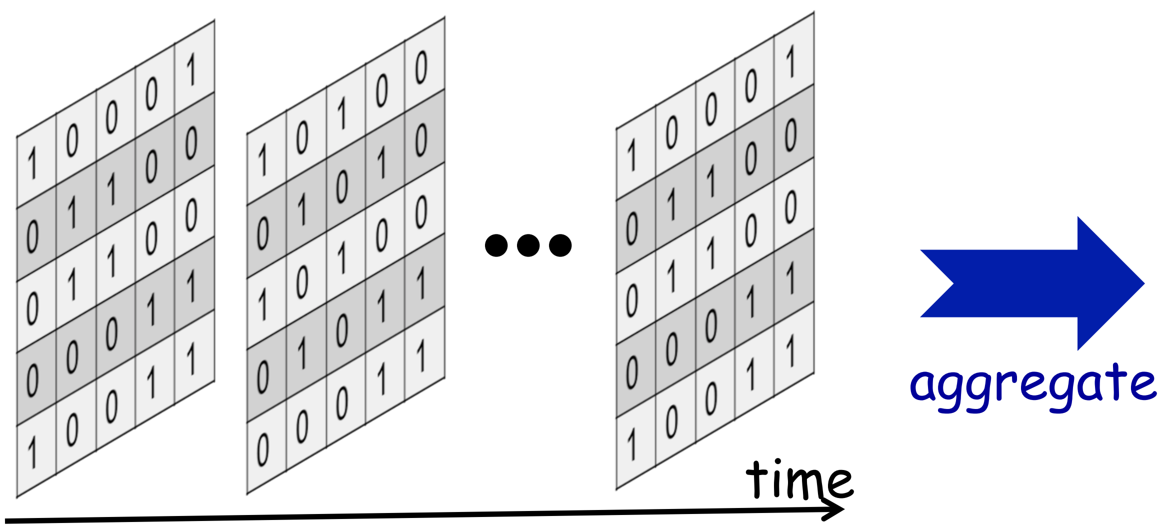
Protocol performance?

Trust?

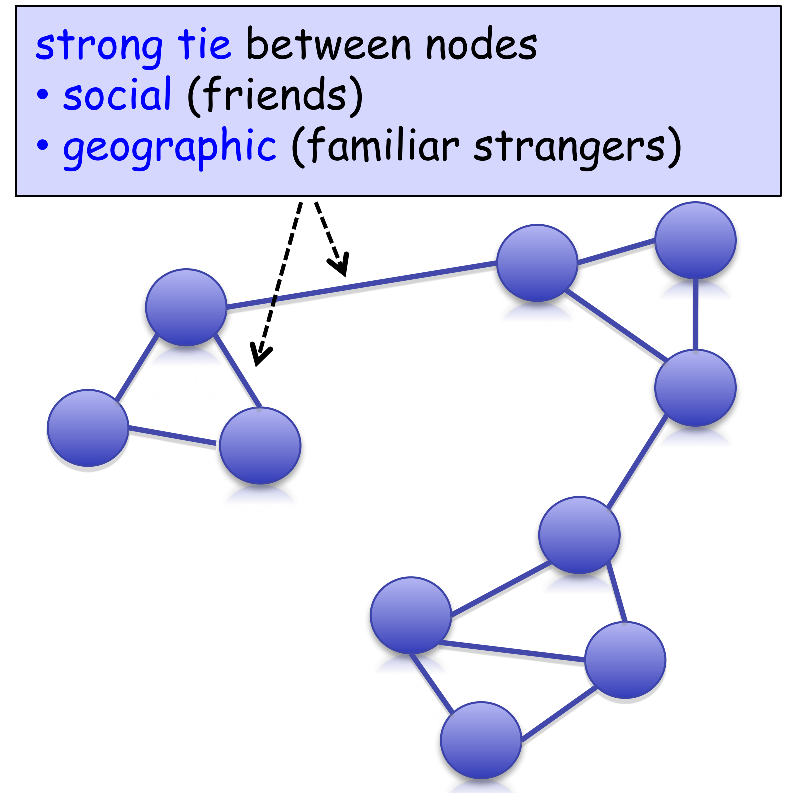
Content Placement?

Contact Graph Representation of Mobility

- Represent Mobility using a Social/Complex Graph
  - Physics, Sociology discipline
  - study of large graphs
  - scale-free, small-world, navigation, etc,



Actual Graph (over time)



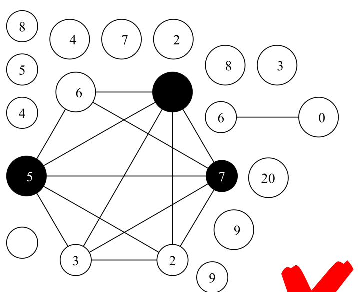
Conceptual Graph

Time-based Aggregation

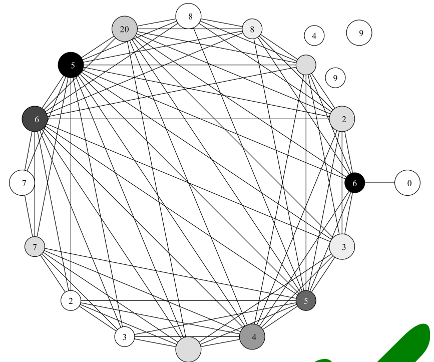
(used by SimBet, BubbleRap)

- *„Growing Time Window“*: Edges for all contacts in [0, T]
- *„Sliding Time Window“*: Edges for all contacts in [T-ΔT, T]

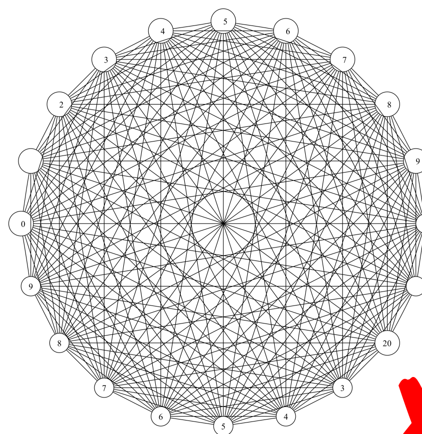
Example: ETH trace, 20 nodes on one floor



ΔT = 1h ✗



ΔT = 2h ✔



ΔT = 72h ✗

! Different networks need different time windows !

Social Properties of Real Mobility Datasets

- Different origins: AP associations, Bluetooth scans and self- reported

	DART	ETH	GOW	MIT
# People and context	1044 campus	285 campus	473 Texas	92 campus
Period	17 weeks	15 weeks	6 months	3 months
Type	AP associations	AP associations	Self-reported location	Bluetooth scanning
# Contacts total	4′200′000	99′000	19′000	81′961
# Contacts per dev.	4′000	350	40	890

- Gowalla dataset
- ~ 440’000 users
- ~ 16.7 Mio check-ins to ~ 1.6 Mio spots
- 473 “power users” who check-in 5/7 days

It's a "small world" after all!

	Clustering Coefficient				Avg. Path Length			
	1%	2%	3%	4%	1%	2%	3%	4%
DART	0.71	0.63	0.57 (0.03)	0.54	7.4	3.7	2.9 (2.0)	2.6
ETH	-	0.66	0.57 (0.03)	0.53	-	6.1	5.6 (3.0)	4.0
GOW	0.28	0.27	0.27 (0.03)	0.26	4.5	3.4	3.0 (2.3)	2.8
MIT	-	-	0.56 (0.03)	0.57	-	-	4.6 (4.5)	3.8

- Small numbers (in parentheses) are for random graph
- Clustering is high and paths are short!

Community Structure

- Louvain community detection algorithm

Trace/Model	# Comm.	Q
DART	23	0.84
ETH	21	0.81
GOW	29	0.7
MIT	6	0.52

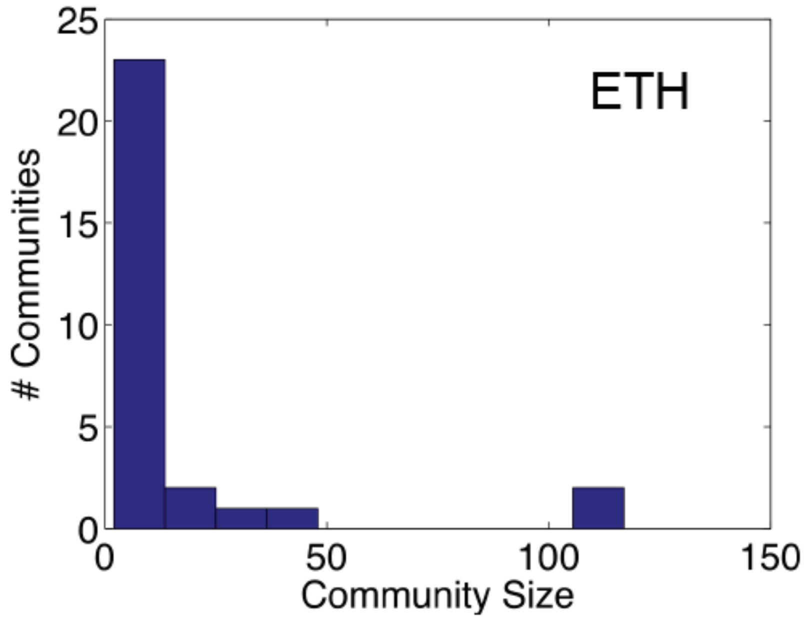
$$Q = \frac{1}{2m} \sum_{ij} \left(w_{ij} - \frac{d_i d_j}{2m} \right) \delta(c_i, c_j)$$

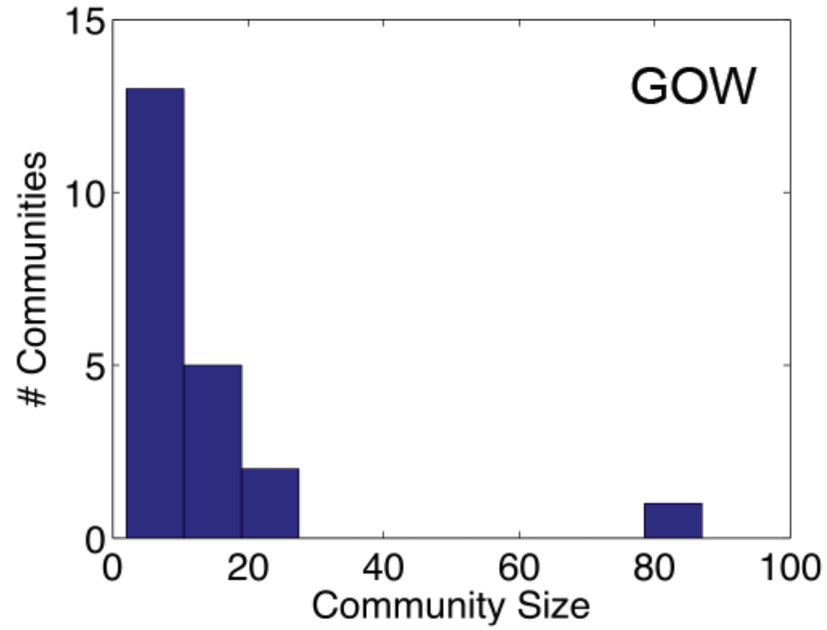
Q = 0 No communities

Q > 0.3 for many networks

- All datasets are strongly modular! ➔ clear community structure exists

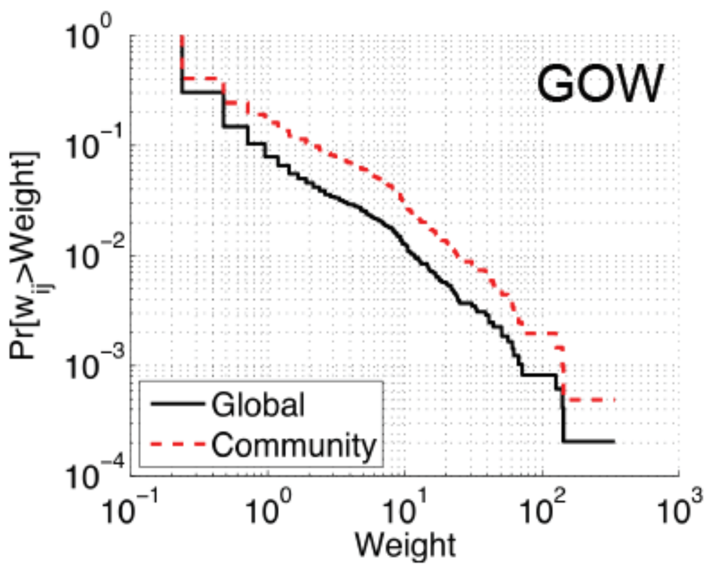
Community Sizes

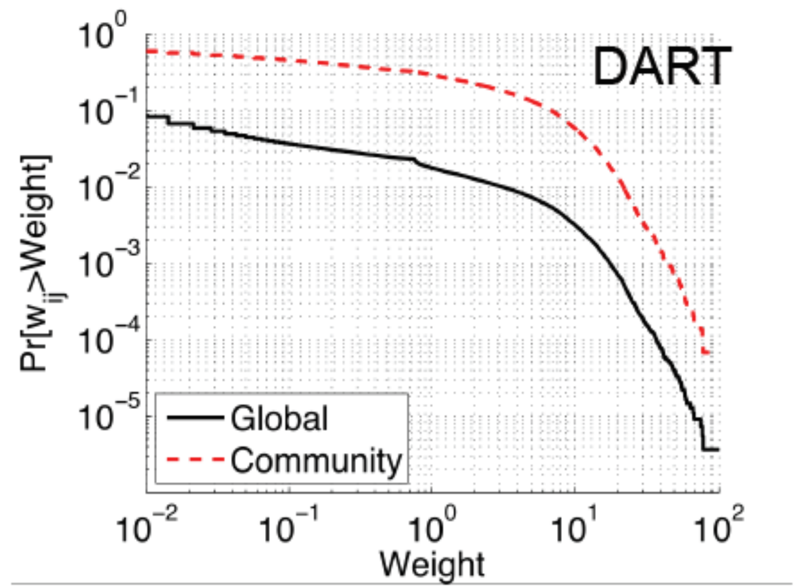




Many small, few large communities

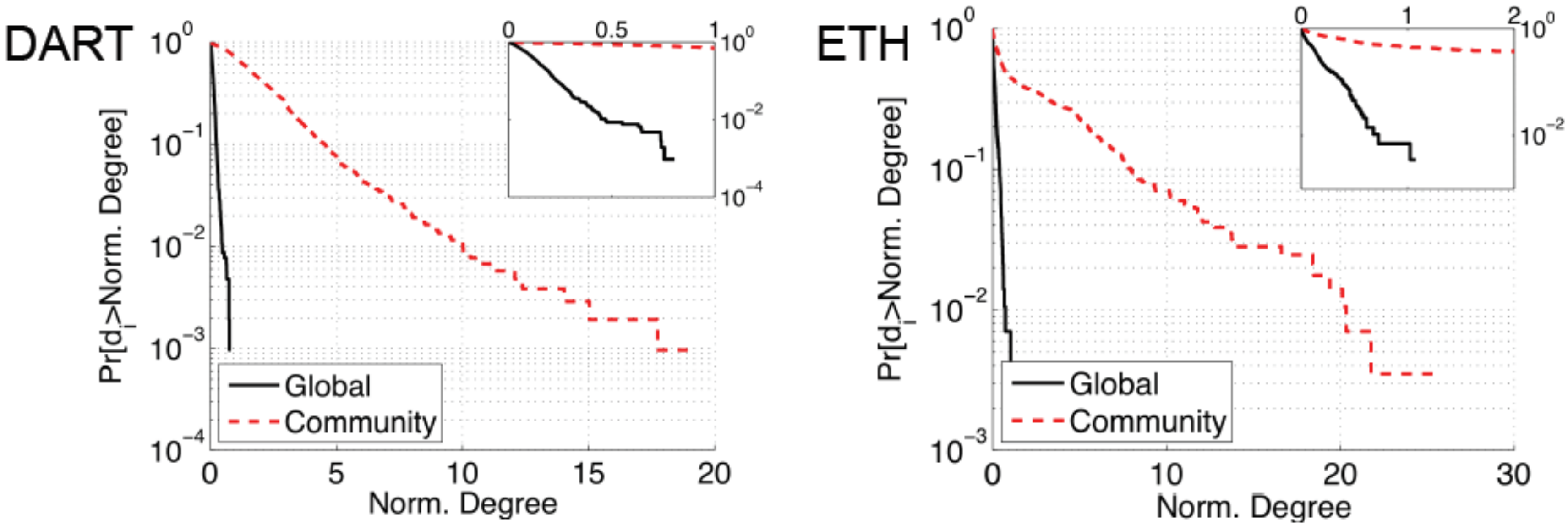
Contact Edge Weight Distribution





- Heavy tail, but not pure power law
- Heterogeneity even within communities
- Similar distributions within and across communities

Degree Distribution



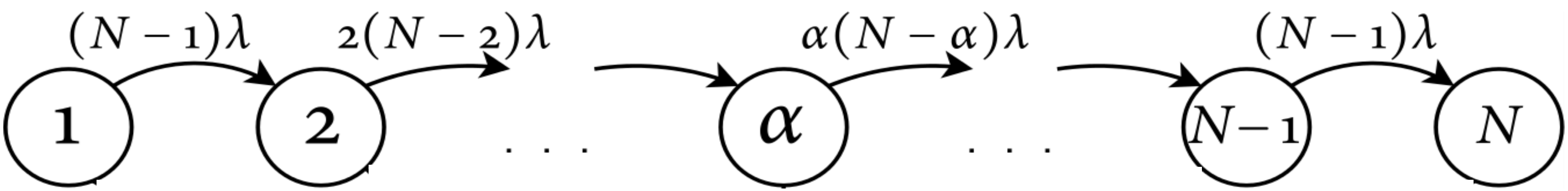
- Normalize degree by number of nodes (N global, $|c_i|$ comm.)
- Exponential(-ish), no "hubs"
- Similar distribution within community and globally

Analysis of Epidemics: The Usual Approach

Assumption 1) Underlay Graph ➔ **Fully meshed**

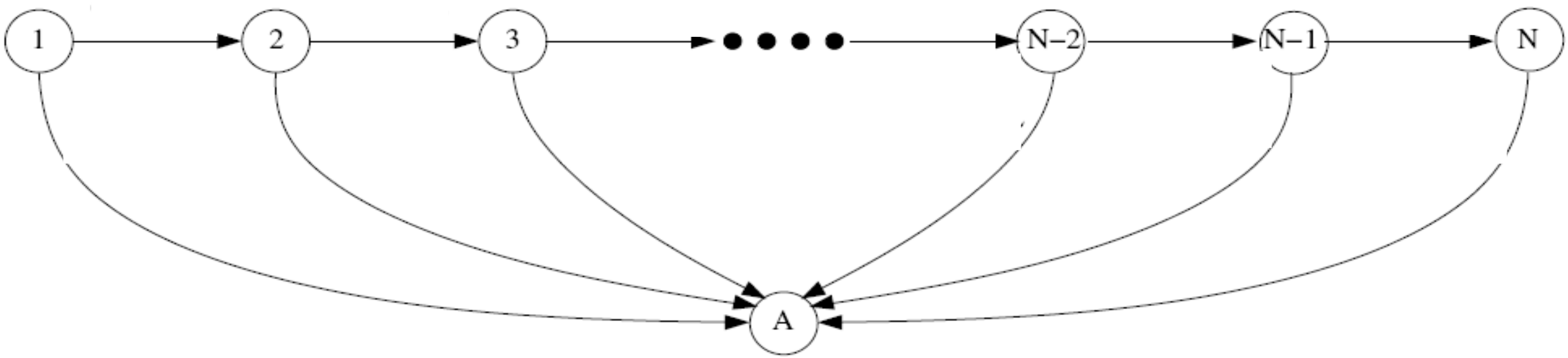
Assumption 2) Contact Process ➔ **Poisson(λ_{ij}), Indep.**

Assumption 3) Contact Rate ➔ **$\lambda_{ij} = \lambda$ (homogeneous)**



$$\boxed{ET_{dst} \approx \frac{1}{\lambda}\frac{\ln(N)}{N}}$$

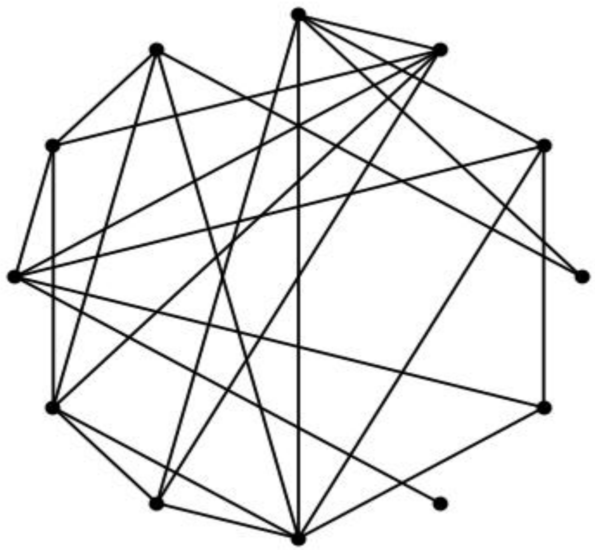
Modeling Epidemic Spreading: Markov Chains (MC)

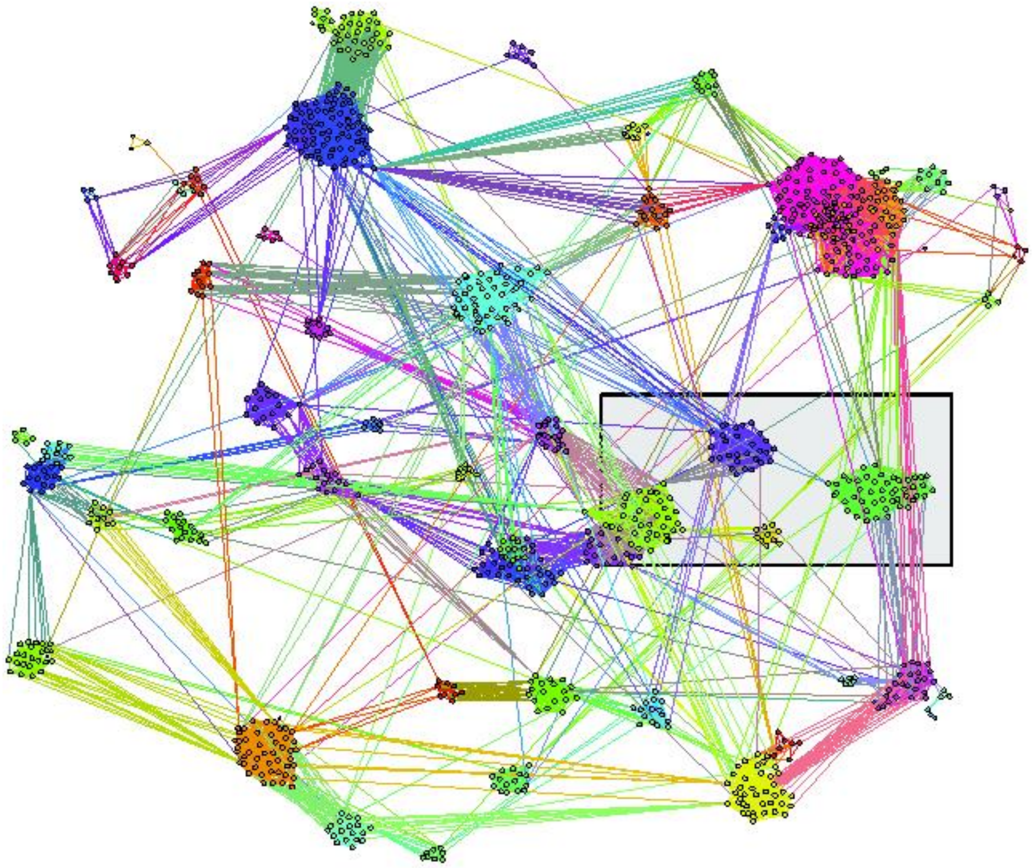


2-hop infection

How realistic is this?

A Poisson Graph

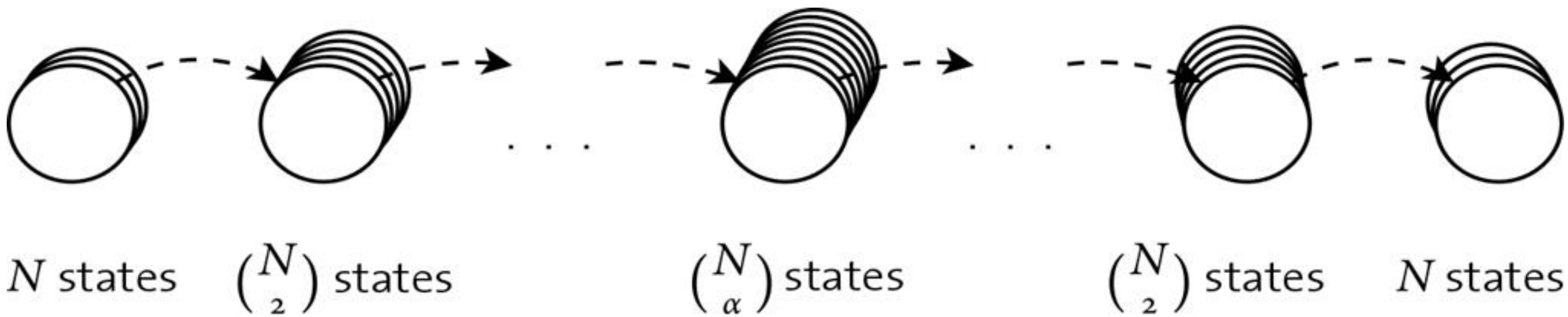




A Real Contact Graph

(ETH Wireless LAN trace)

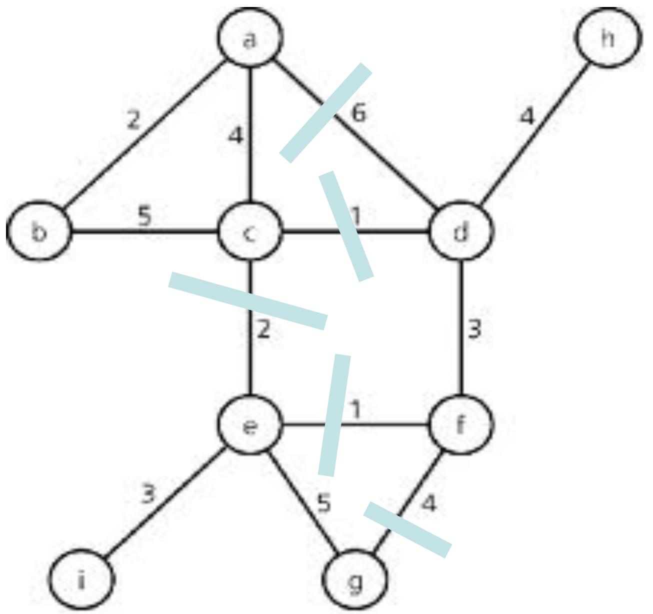
Arbitrary Contact Graphs



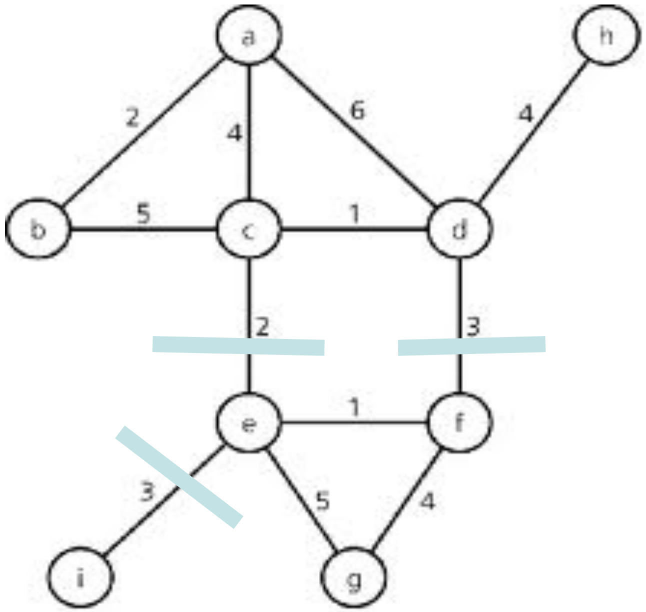
Bounding the Transition Delay

$$E[T_{k,k+1}|C_a] = \frac{1}{\sum_{i\in C_a, j\notin C_a} \lambda_{ij}} \leq \max_{C_a}\left\{\frac{1}{\sum_{i\in C_a, j\notin C_a} \lambda_{ij}}\right\} = \frac{1}{\min_{C_a}\left\{\sum_{i\in C_a, j\notin C_a} \lambda_{ij}\right\}}$$

- What are we really saying here??
- Let a = 3 → *how can split the graph into a subgraph of 3 and a subgraph of N-3 node, by removing a set of edges whose weight sum is minimum?*



A 2nd Bound on Epidemic Delay



$$\Phi = \min_{a} \frac{\left\{ \min_{C_a} \sum_{i \in C_a, j \notin C_a} \lambda_{ij} \right\}}{a(N-a)}$$

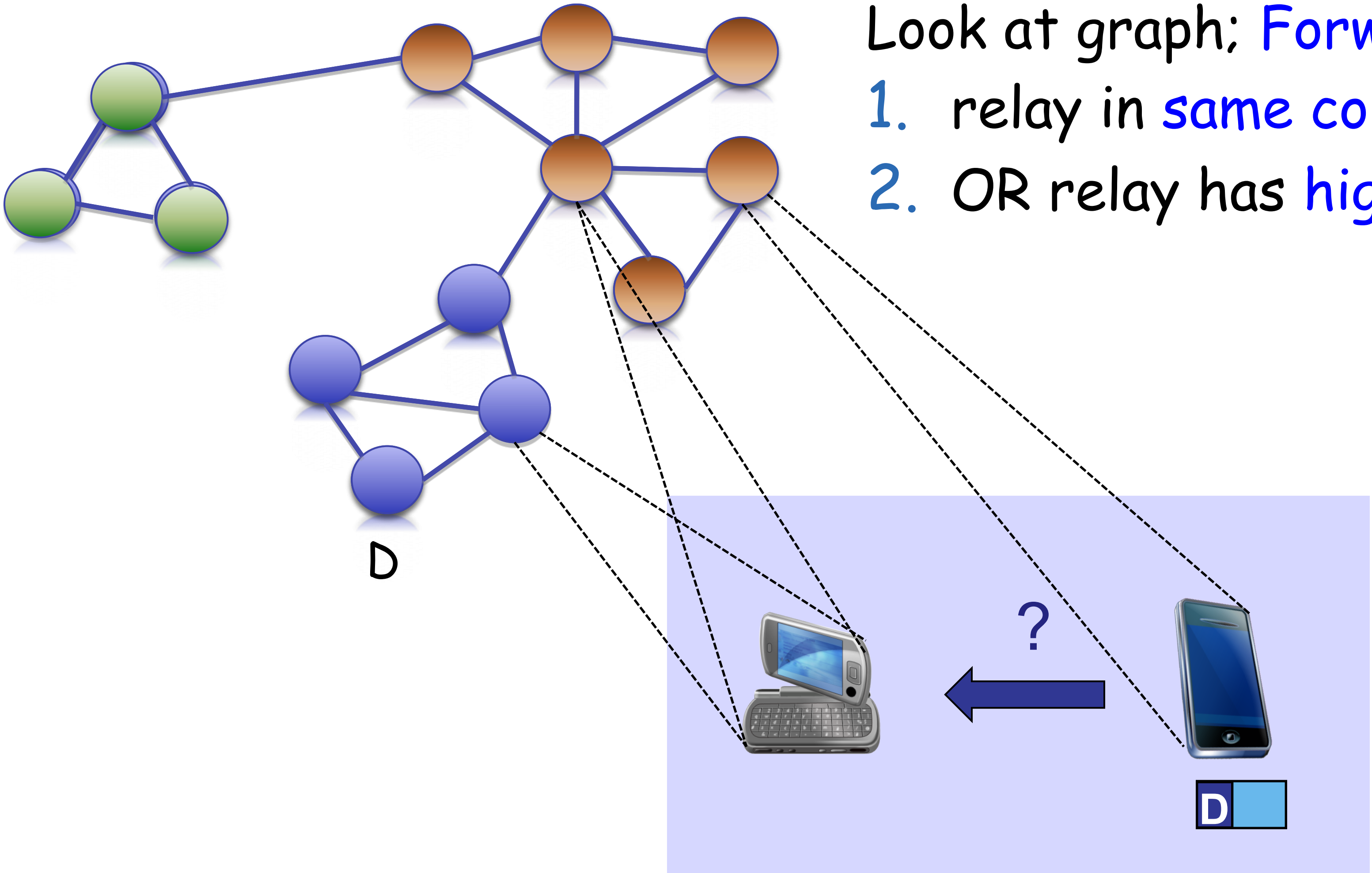
$$E[D_{epid}] \le \sum_{a=1}^{N} \frac{1}{a(N-a)\Phi_a} \le \sum_{a=1}^{N} \frac{1}{a(N-a)\Phi} \approx \frac{1}{\Phi} \frac{\ln N}{N}$$

- **Φ** is a fundamental property of a graph
- Related to graph spectrum, community structure

Opportunistic Routing: Contact Prediction

- Human mobility not fully random: Patterns, Recurrence
  - e.g. recent Barabasi's Nature papers
- Human mobility is heterogeneous
  - Different neighbors, different numbers of neighbors
- Infer Contact Pattern => Predict Future Contacts => Forward to node with Highest Delivery Probability
- HOW???
- (maybe?) recent contact with X => high prob. of future contact with X [Lindgren et al. '03, Dubois-Ferriere et al., '03]
- (maybe?) frequent contact with X => high prob. of future contact with X [Burgess et al '06]
- (maybe?) many total contacts (with anyone) => high prob of future contact with any X [Spyropoulos et al. '07, Erramilli et al. '08]

SNA-based Forwarding (SimBet, BubbleRap)



Look at graph; Forward IFF

1. relay in same community as D
2. OR relay has higher centrality

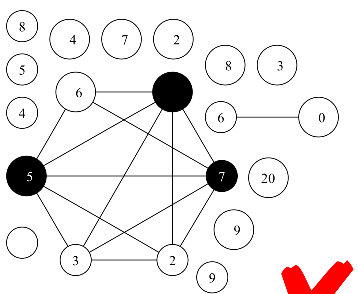
SNA-based Forwarding shows promising performance!

Time-based Aggregation

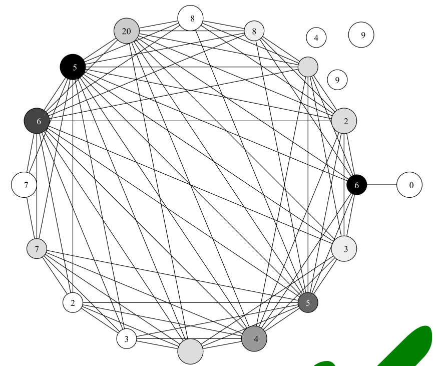
(used by SimBet, BubbleRap)

- *„Growing Time Window“*: Edges for all contacts in [0, T]
- *„Sliding Time Window“*: Edges for all contacts in [T-ΔT, T]

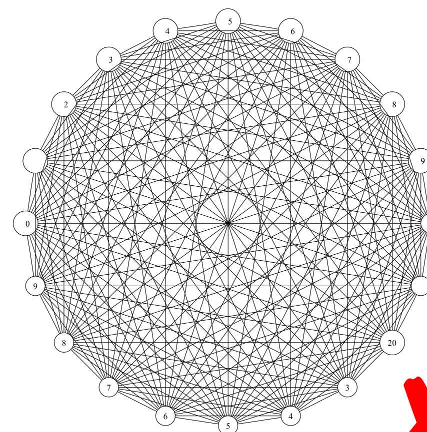
Example: ETH trace, 20 nodes on one floor



ΔT = 1h ✗



ΔT = 2h ✔



ΔT = 72h ✗

! **Different networks need different time windows** !

Density-based Aggregation

Aggregate to a certain density of the graph

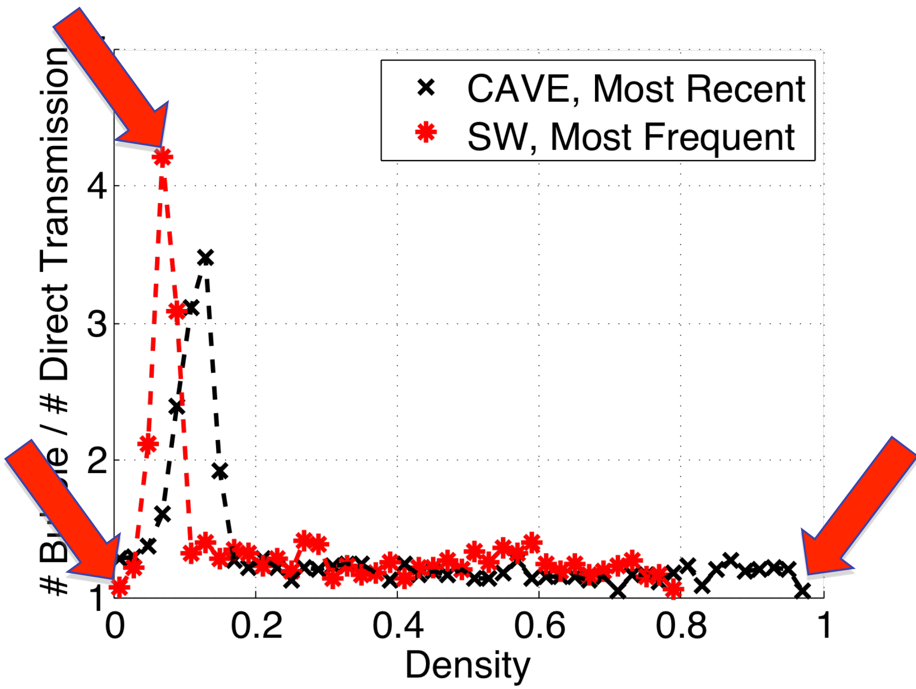
- Density $d = \frac{|E|}{\frac{N(N-1)}{2}}$ N: # nodes, E: edges included
- Easier to compare between scenarios (0 ≤ d ≤ 1)
- How to „fill" the social graph to this density?
  - *„Most recent"*: Create an edge for the x most recent contacts
  - *„Most frequent"*: An edge for the x most frequent contacts
- What is the right density???

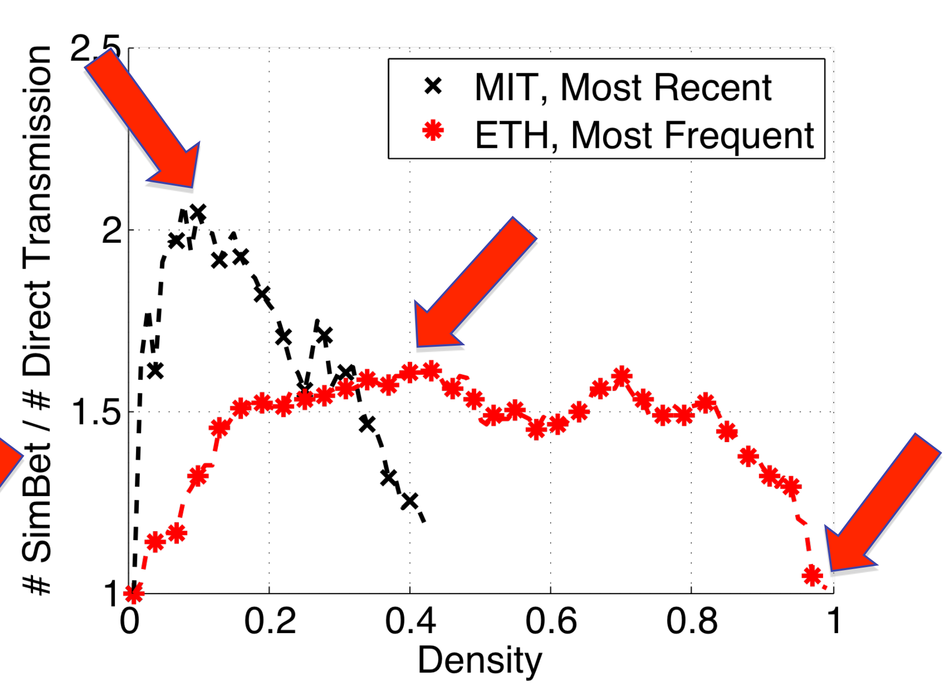
Sensitivity of SNA-based Routing Performance

? How to evaluate the social graphs ?

- Sensitivity of routing to graph density
- Good graphs => good routing performance
- Simulation using SimBet and Bubble Rap
  - Synthetic contact processes
    - Small-world, cavemen
  - Contact traces
    - ETH (20 nodes, students and staff working on 1 floor)
    - INFO (41 Infocom 2005 participants)
    - MIT (97 students and staff)

Contribution 1 - Sensitivity of Routing Performance





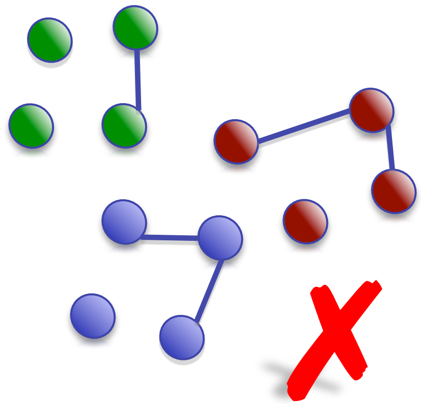
Bad performance in „extreme“ cases!

There is an optimal density range!

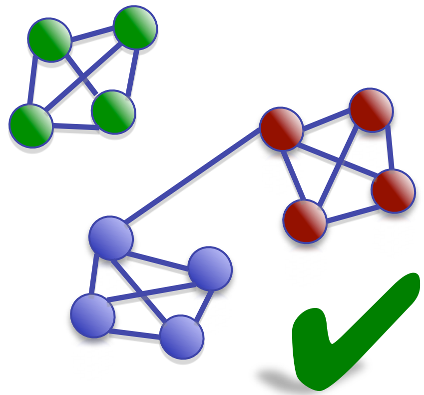
Contribution 2 - Online Algorithm

Similarity(u,v) = Number of common neighbors of u and v

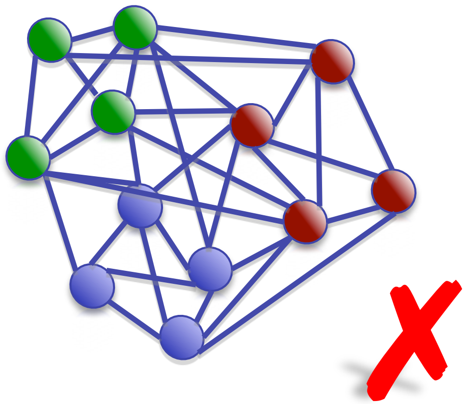
- Assumption: Two types of contacts
  - Regular, with nodes of same community -> high similarity
  - Random, with nodes of different communities -> low similiarity
- We want all regular links but no random links => Predictive!



Regular: low similarity
Random: low similarity



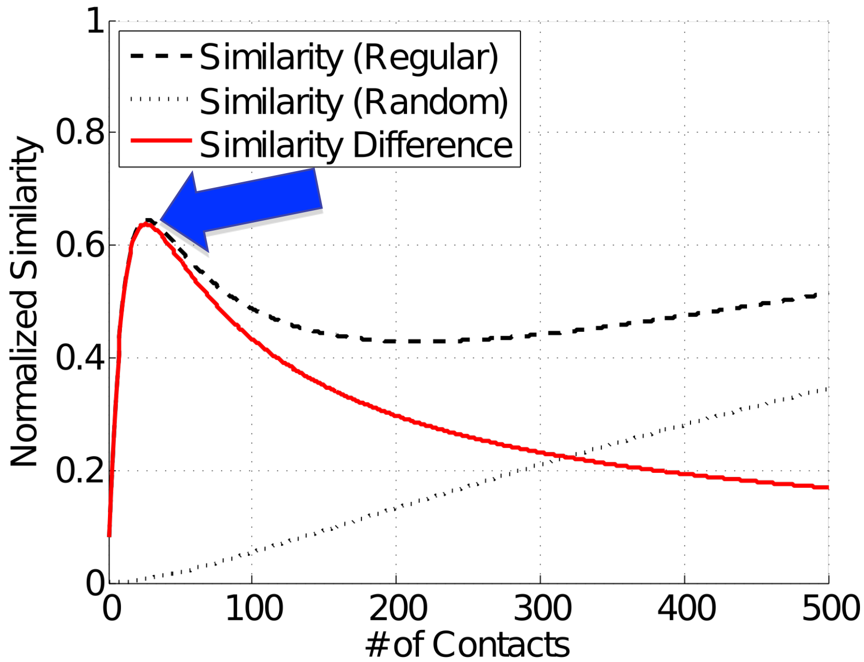
Regular: high similarity
Random: low similarity



Regular: high similarity
Random: high similarity

Doing the Math

- Based on cavemen graph model

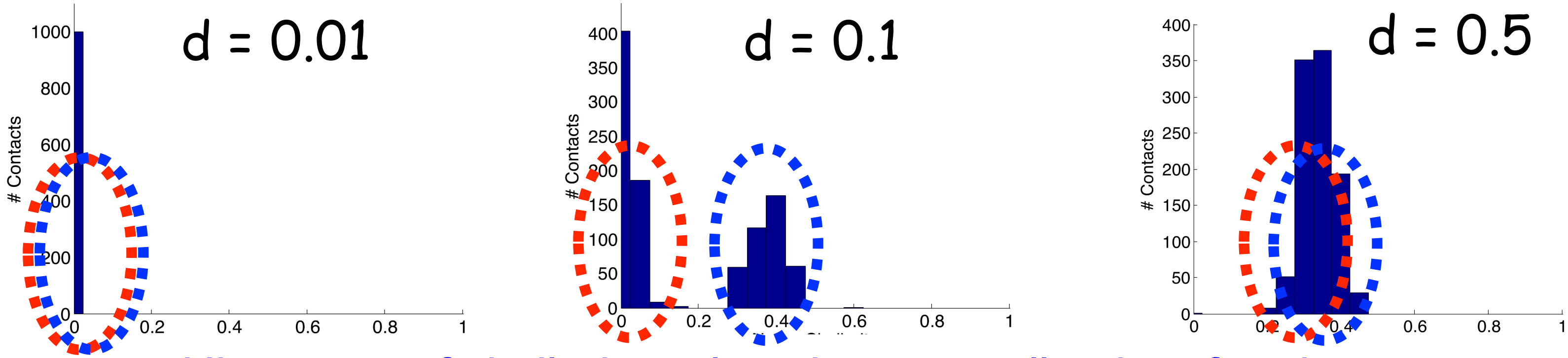


- Maximize avg similarity of Regular links
- Minimize avg similarity of Random links

Maximize Difference

Maximizing Modularity (I)

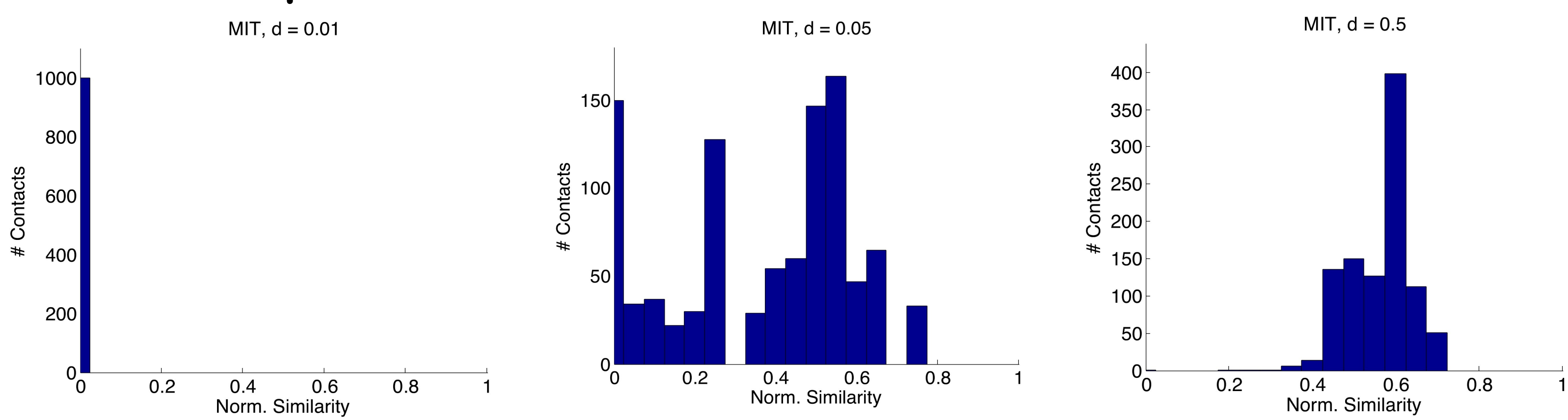
cavemen model



Histogram of similarity values between all pairs of nodes

- Clustering to distinguish Random and Regular links
- Synthetic models: 2-means clustering
  - Density with maximal cluster distance is optimal
- Real world requires more robust solution

MIT



Maximizing Modularity (II): Spectral Analysis

- Arrange observed similarity values (s_i) into a matrix W

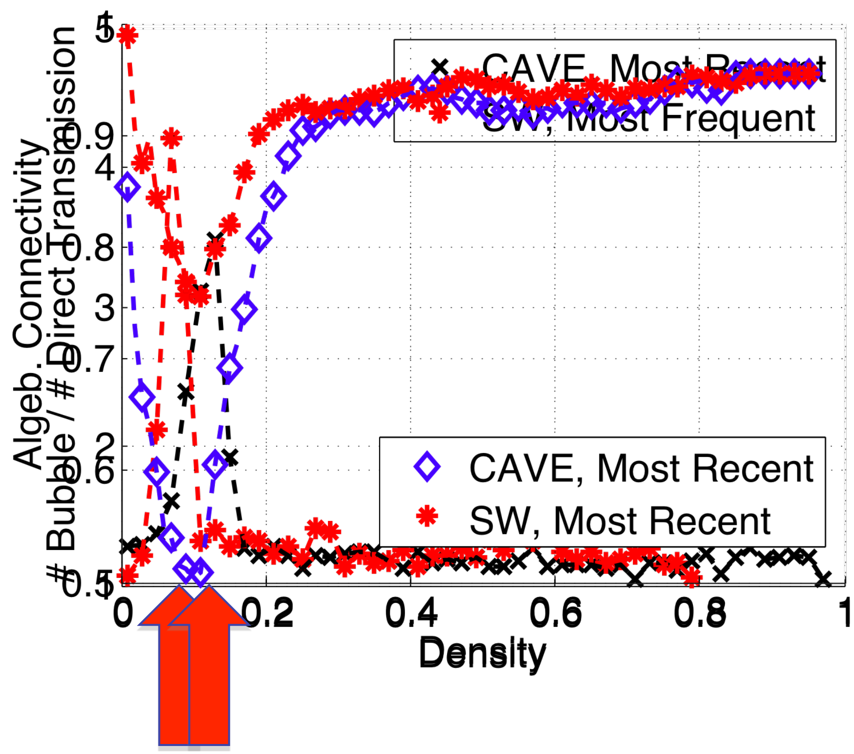
$$w_{ij} = exp\left(-\frac{\left\|s_i - s_j\right\|^2}{2\sigma^2}\right)$$

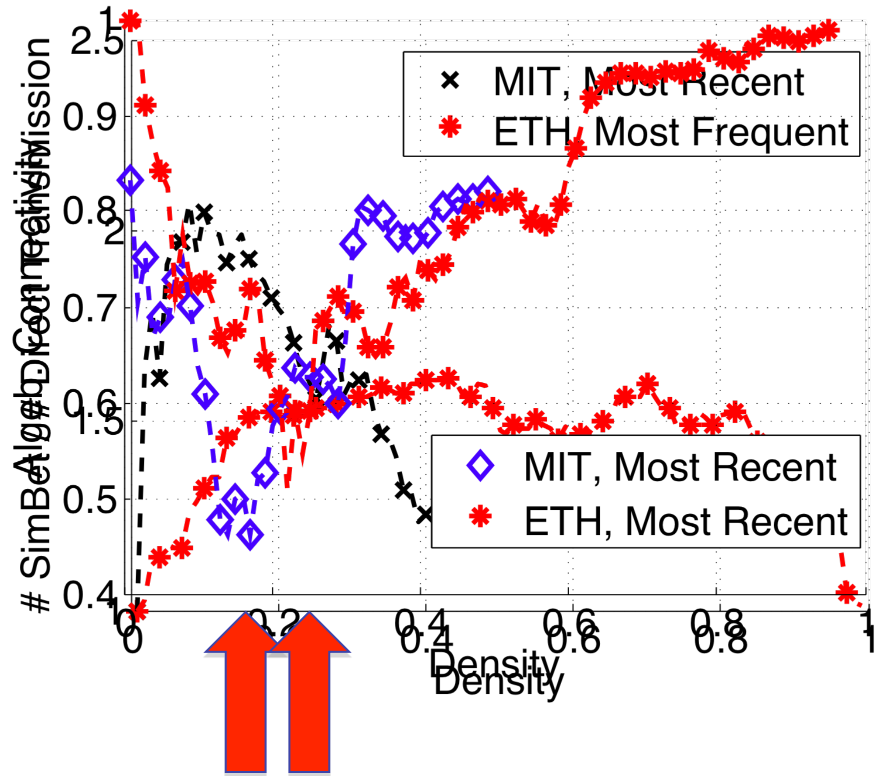
- Spectral Graph Theory
  - Calculate Laplacian L of W
  - D: diagonal normalization matrix

$$L = D^{-1/2} W D^{-1/2}$$

- Eigenvalue decomposition of L: $L \cdot e_i = \lambda_i e_i$
  - $\lambda_1 = 0 \leq \lambda_2 \leq \ldots \leq \lambda_n$
  - $\lambda_2 = 0$ if two clusters are perfectly seperable (2 connected components)
  - λ_2 (Algebraic Connectivity): small for highly modular data
- Minimizing λ_2 -> max. "distance" between Regular and Random

Maximizing Modularity (III)





Minimum correlates with optimal density

Performance of Online Algorithm (II)

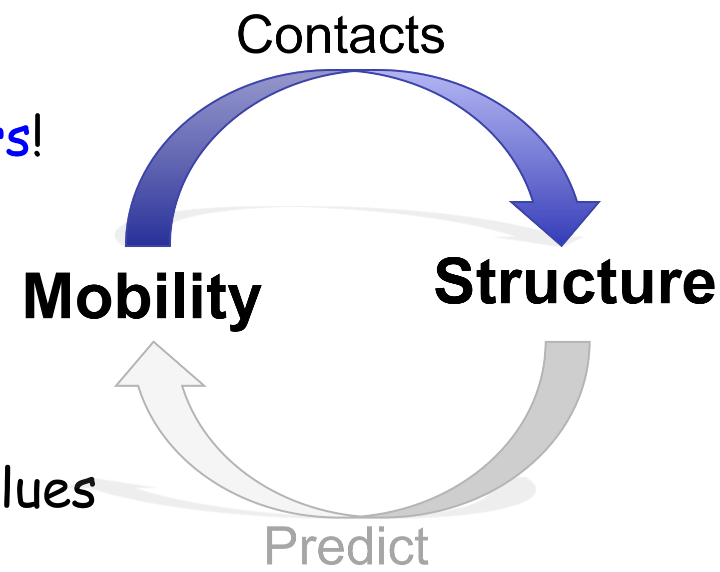
- Delivery ratio relative to Direct Transmission using
  - optimal fixed density / online algorithm

Protocol	SW	CAVE	MIT	ETH	INFO
SimBet MF	**4.3/4.1**	**3.3/3.0**	**1.8/1.8**	**1.6/1.5**	**1.3/1.2**
Bubble MF	**4.2/2.9**	**4.5/3.6**	**2.5/2.1**	**1.5/1.5**	**1.4/1.3**

Online Algorithm performance is close to optimal

Opportunistic Routing Using SNA: Summary

- **Contribution 1 - Sensitivity analysis**
  - Choosing the right aggregation density matters!
  - More than specific routing algorithm!
- **Contribution 2 – Optimal density algorithm**
  - Maximize modularity of observed similarity values
  - Spectral Graph Theory techniques
  - Performance close to that of optimal density



! General Applicability (not just routing) !