USE OF SYNTHETIC POPULATIONS

A. Apolloni

Institute des Systems Complexes Rhone-Alpes
Laboratoire de Physique
ENS-Lyon

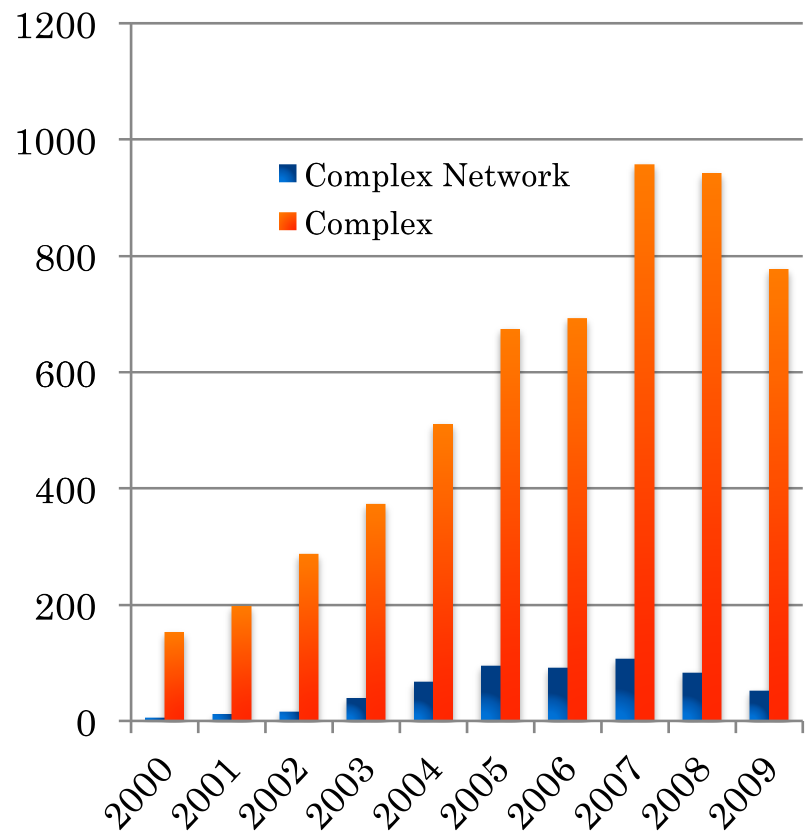
Southern Methodist University
Dallas
19/09/11

OUTLINE

- Motivation
- Synthetic population
- Epidemics and optimal vaccine allocation
- *Least Spread Line* and numerical simulation
- Use of vulnerability
- Comparison with other models
- Conclusions
- Other works going on
- Socio physics and private communication
- Conclusion

IN THE LAST 10 YEARS

Witnessed a terrific growing interest in studying complexity and network:
• More data available (e.g.Enron mail network)
•New communication and social technology
•New models for network evolution (e.g.. Barabasi Albert; Watts and Strogatz)

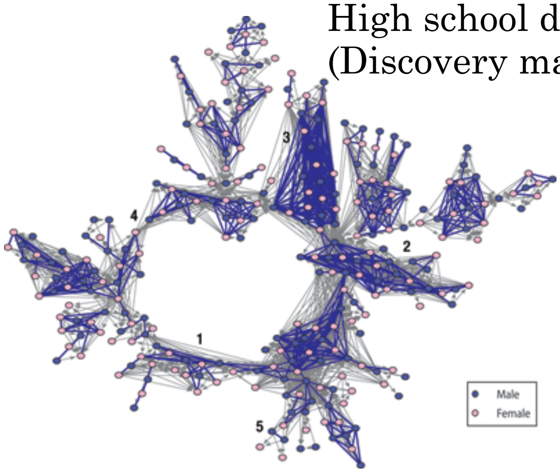


Different networks:
•Interdisciplinary approach
•Network formation and topology
•Real networks analysis
•Topological effects on diffusion

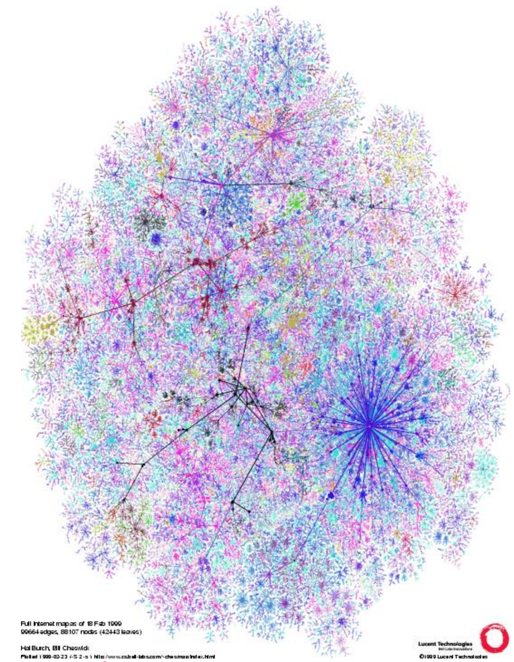
Network examples:
•Internet, WWW, Cell-phone, e-mail ,Airlines and transportation, Power grid, Protein-protein interaction, Sexual and contagions, Directory boards and actors, Linguistic, citation, collaboration

MANY EXAMPLES

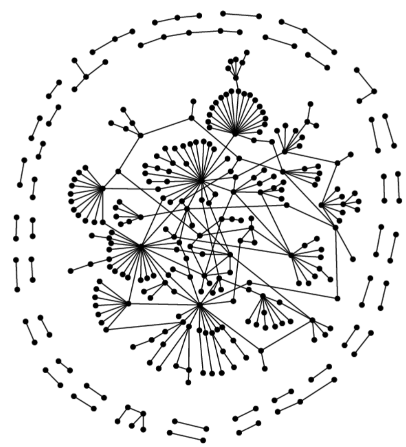
High school dating network
(Discovery magazine, 2007)



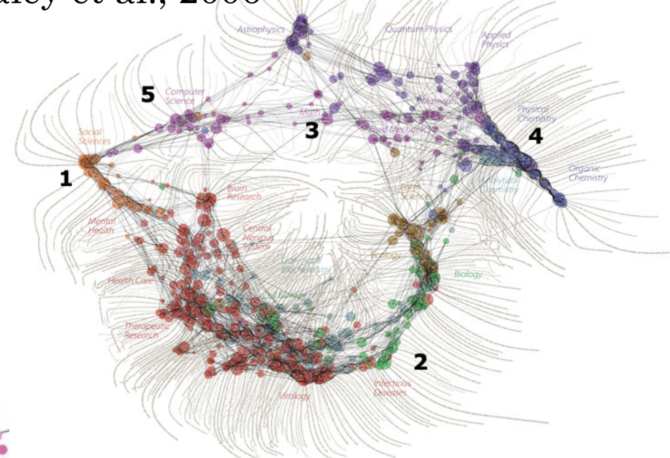
Router level graph

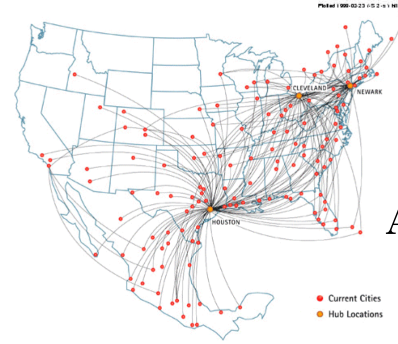


Yeast protein
Interaction network



Connections between disciplines
Paley et al., 2006





Airline routes graph

BASIC DEFINITIONS: COMPLEX NETWORKS

- Set of nodes linked by edges
- Highly unstructured system comprised of several parts
- Cannot be understood as a sum of its parts

Blogosphere
(datamining.typepad.com)

IMPORTANT PROPERTIES OF COMPLEX NETWORKS

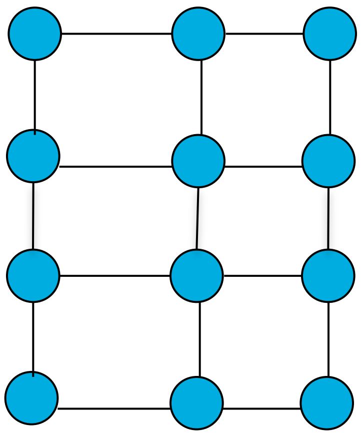
- ❑ Too complex to understand by visualization
- ❑ Need to understand using simple properties

- Degree distribution
- Closure, Assortivity, Homophily
- Centrality
- Robustness

- Need new measures for studying dynamics

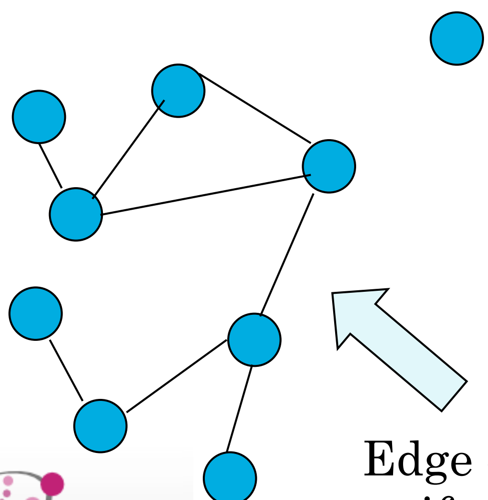
SIMPLE RANDOM GRAPH MODELS

Grid



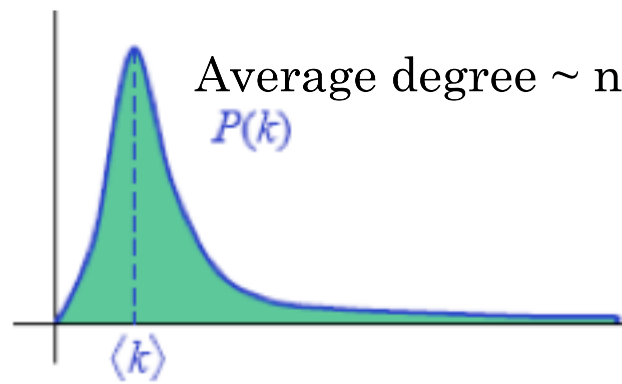
All the nodes have the same number of neighbors: *homogeneous* network; if infinite no point plays any important role

Erdos-Renyi model (1958):
First model of heterogeneous network

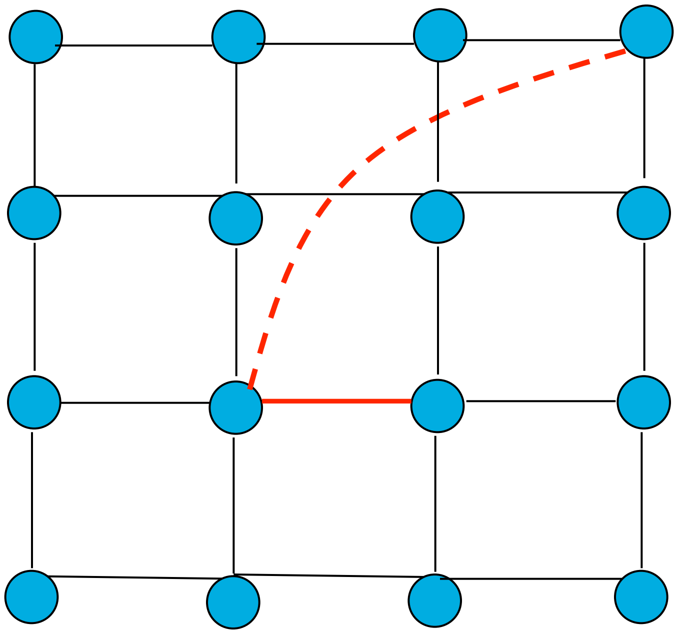


Edge created with uniform probability p

Average degree ~ np



SMALL WORLD NETWORKS: large clustering coefficient and small average path length



Watts and Strogatz, *Science*, 1998
• Consider a regular lattice
•Pick randomly a link (red) and rewire one end with probability p: p=0 grid;p=1 random
•Check the connectivity

Other version: Kleinberg, *Nature*, 2000]
Long distance neighbor: grid of nodes; links are created between nodes chosen with probability $c \cdot d(u,v)^{-k}$, where d(u,v) is distance

Model rooted in social network: most of the people have friends in close neighborhood; occasionally friends are located far away

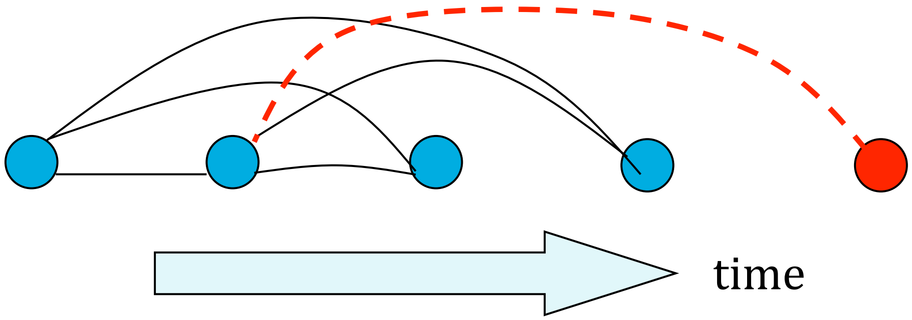
Milgram's experiment (1967): six degrees of separation
- Target: stock broker in Boston
- 296 senders from Omaha
- Each person asked to send a letter to a contact who is *closer* to the destination than they themselves are
- 20% of chains reach destination
- Mean length of chain ~ 6.5

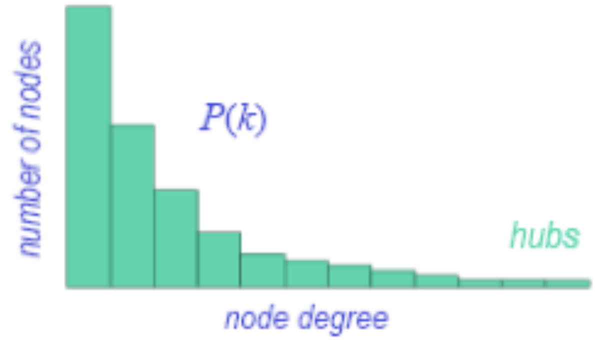
Institut des systèmes complexes
Complex Systems Institute
Rhône-Alpes
IXXI

ECOLE NORMALE SUPERIEURE DE LYON

EVOLUTIONARY MODELS

[Barabasi et al., *Science*, 1999]





- Initial set of *m* nodes
- At each time step new node is added with $m_0 < m$ links
- New node chooses edge to earlier nodes with prob. proportional to their degrees

$$P_{i,new} = \frac{k_i}{\sum_{i=1}^{m+t} k_i}$$

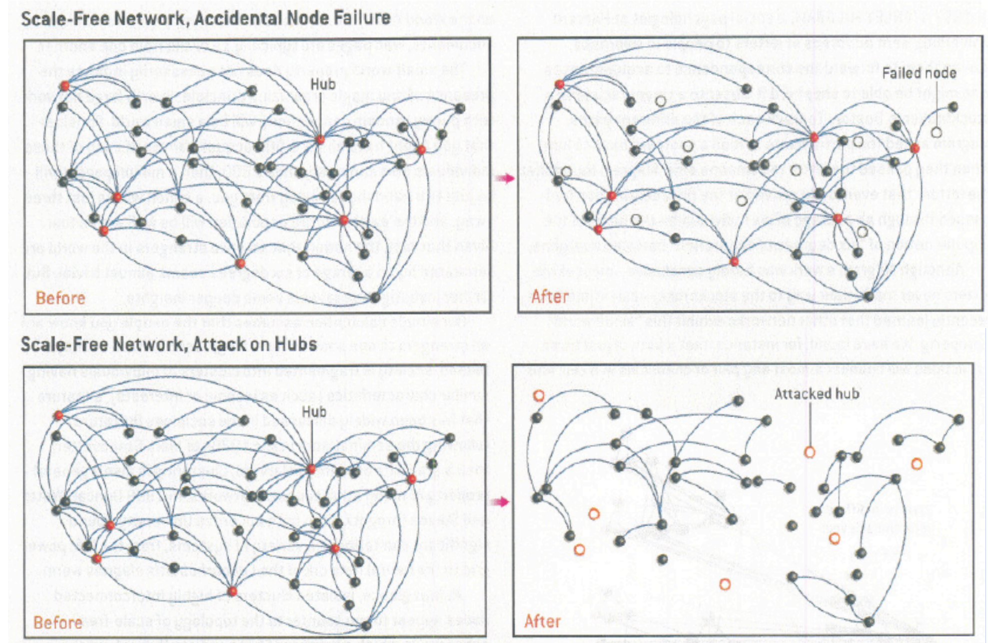
- #nodes of degree k ~ $k^{-\alpha}$ $\alpha \approx 2.0$
- Examples: internet, www, airlines network

Consequences:

- Finite fraction of nodes with large number of links *(hubs)*
- No scale parameter
- Average degree is not a good parameter, large fluctuations

SOME EFFECTS

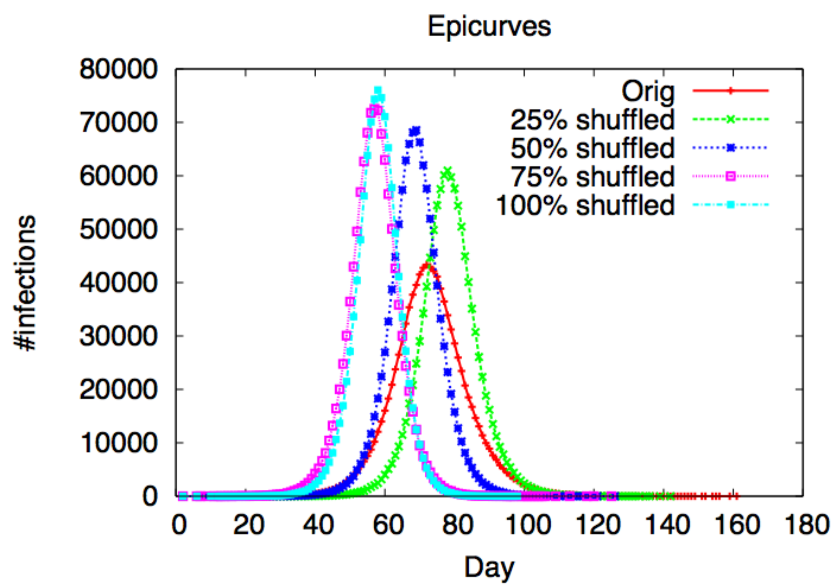
- **Robustness under deletion of links**: consider failure, network can be disconnected. Random graph is indifferent to random or targeted attacks; Scale free network is robust to random attacks but fragile to targeted ones



- **Contagions**: considered as a percolation process; degree distribution often important however transmissibility is critical e.g. S.A.R.S.; Existence of a threshold value for the transmissibility T:

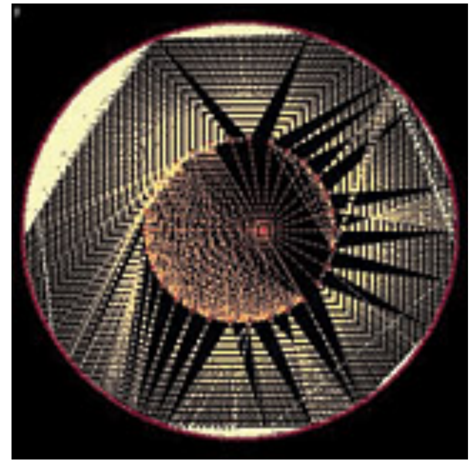
$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

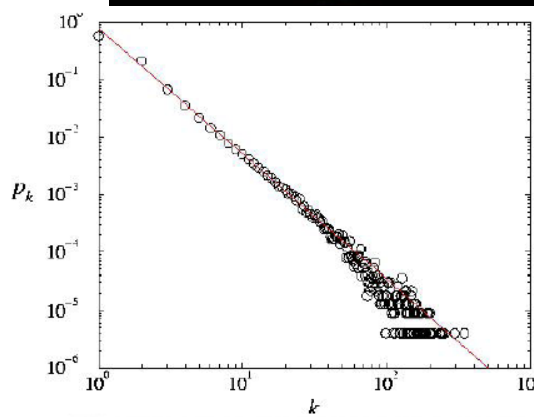
- Random graph T_c finite; power law T_c null

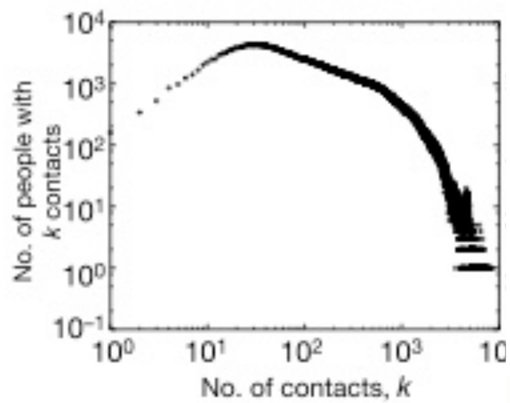


LIMITATIONS

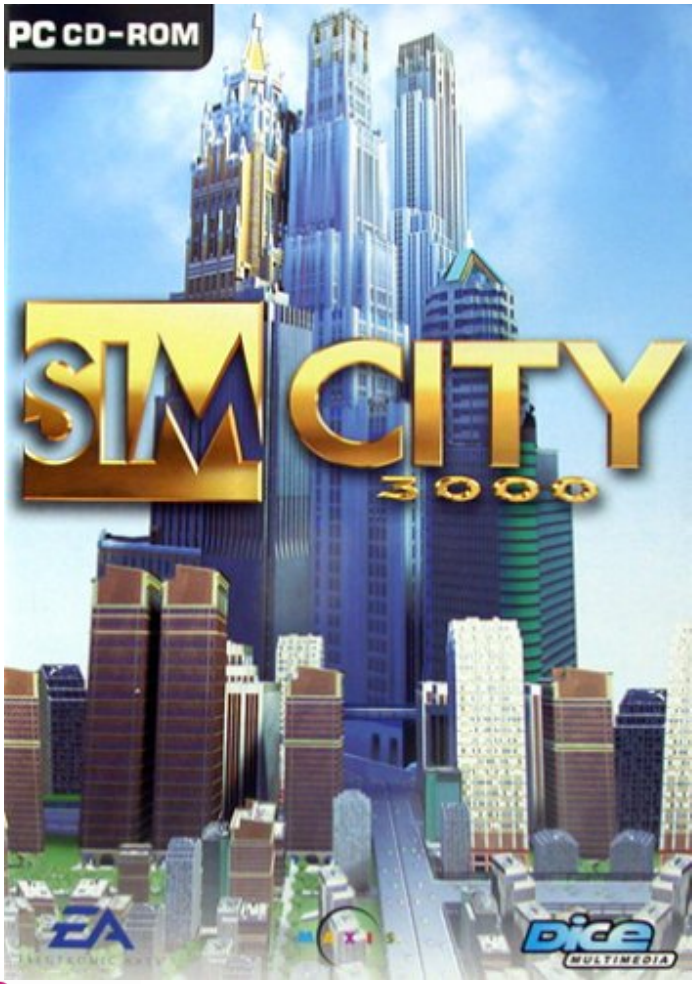
- **Network model**: realistic model of individual social-contact networks are different from theoretical ones
- **Networks evolve in time**: due to family, job constraints individual follow certain path, meeting certain subset of population. Links are created and destroyed. Individuals' mobility is not random, but they visit just a small set of locations ([Barabasi, *Physica A*,2008]) Theoretical approach is daunting.
- **Analytical approach for diffusion**: based on tree assumption, there aren't any structure in the network and any correlation between nodes

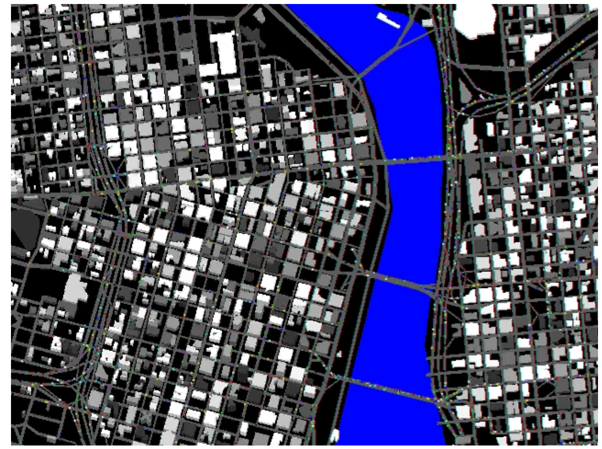






Synthetic Population



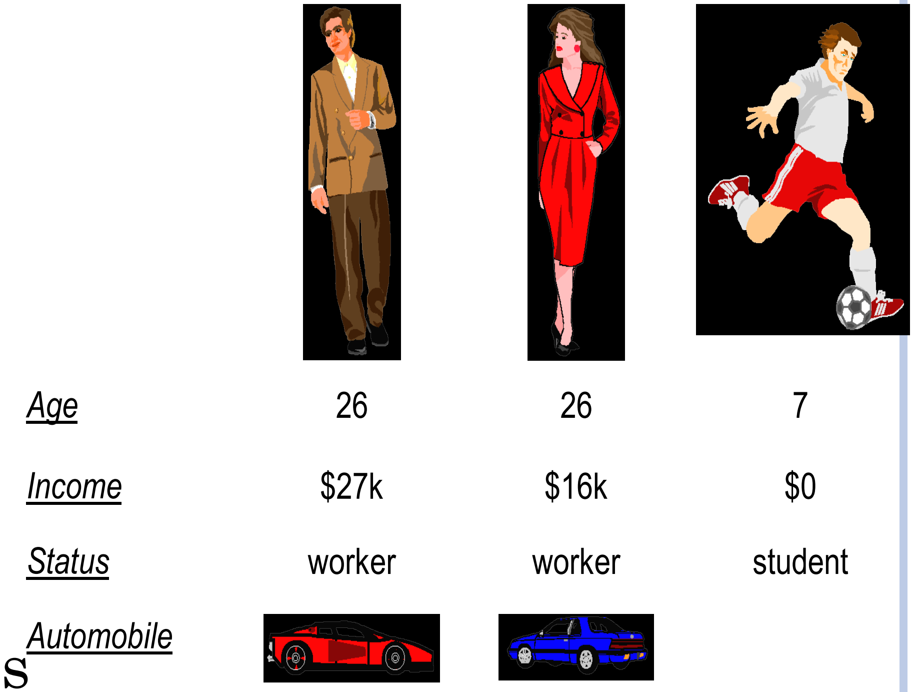


Institut des systèmes complexes
Complex Systems Institute
Rhône-Alpes
IXXI

ECOLE NORMALE SUPERIEURE DE LYON

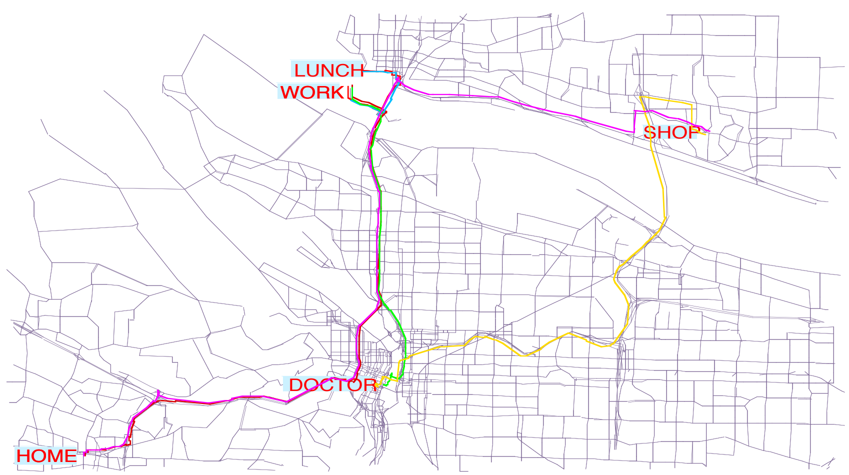
Synthetic Population : household construction

- Creation of a Synthetic Household: household are generated from census data and geo-located accordingly. Population is statistically in-distinguishable from real one.
- Each household member is assigned a routine of activities extracted from survey data.
- Activities are geo-located

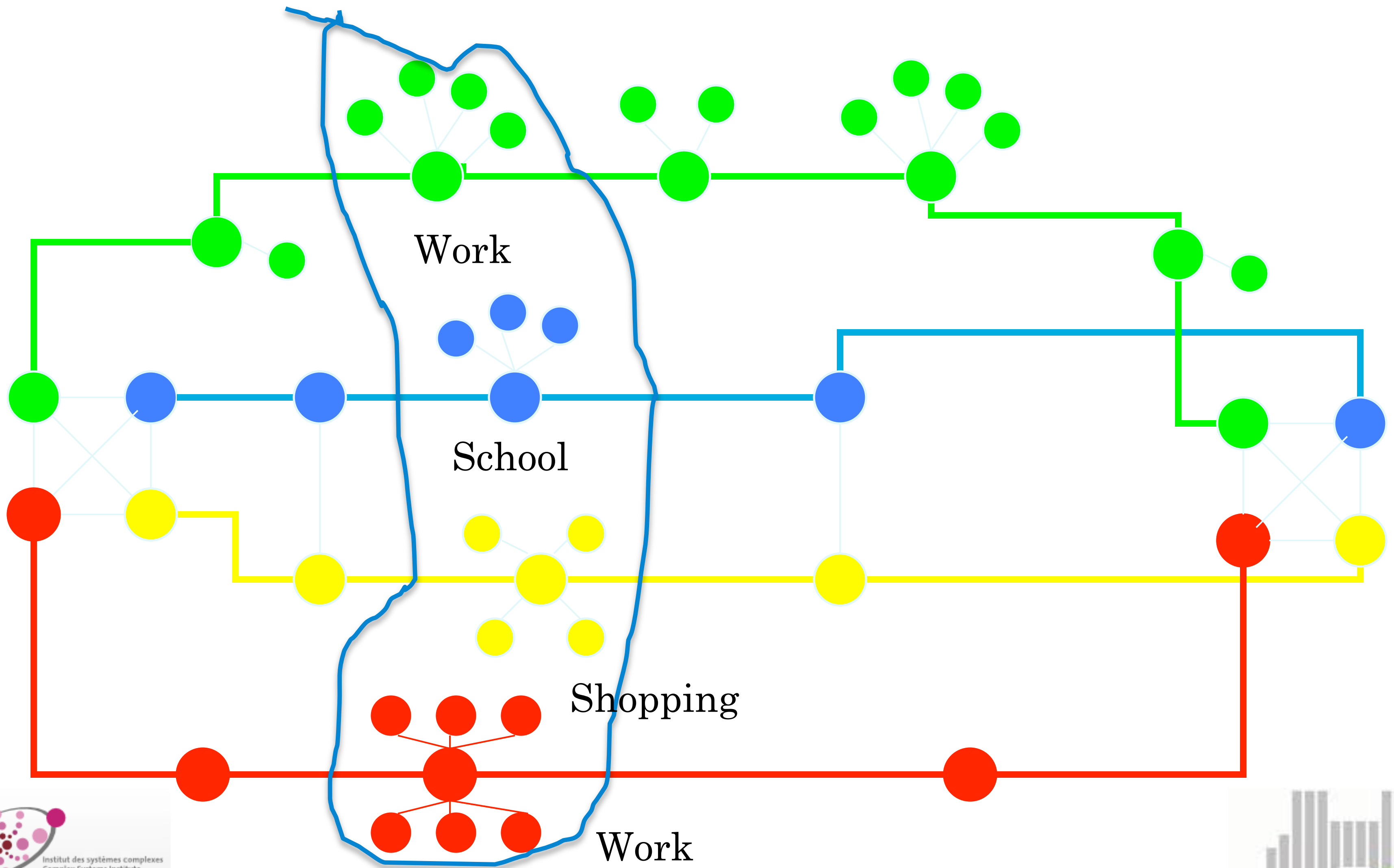


SYNTHETIC POPULATION: INTERACTION NETWORK

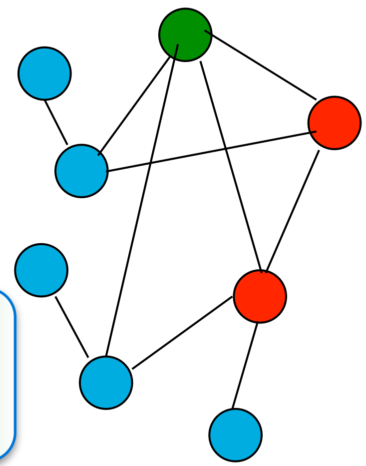
- Locations are assigned to activities sequence in two steps: choosing a zone and then a location. Assignment function depends on time to reach from the precedent activity.
- Social network is constructed dividing location in sub-locations of different sizes and maximum capability
- *People are in contact when sharing the same sub-locations.*
- Associate probabilistic timed finite state machine for each individual
- Define model for diffusion process



Synthetic Population 3)



Example : Influenza spreading

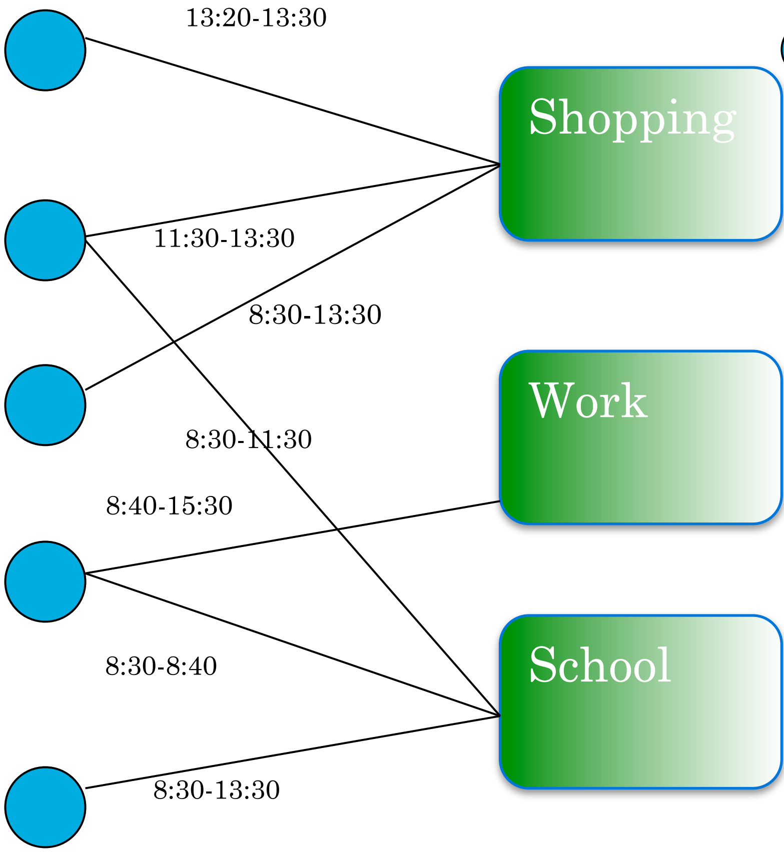


Transmission in each Sub-location

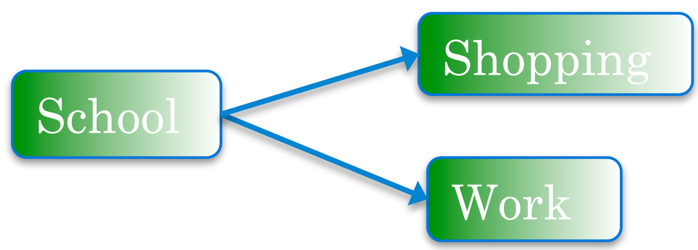
Links randomly chosen
Probability of being infected at each time step

$$P = 1 - (1 - T)^{k(t)}$$

Bernoulli trials for all the duration of links

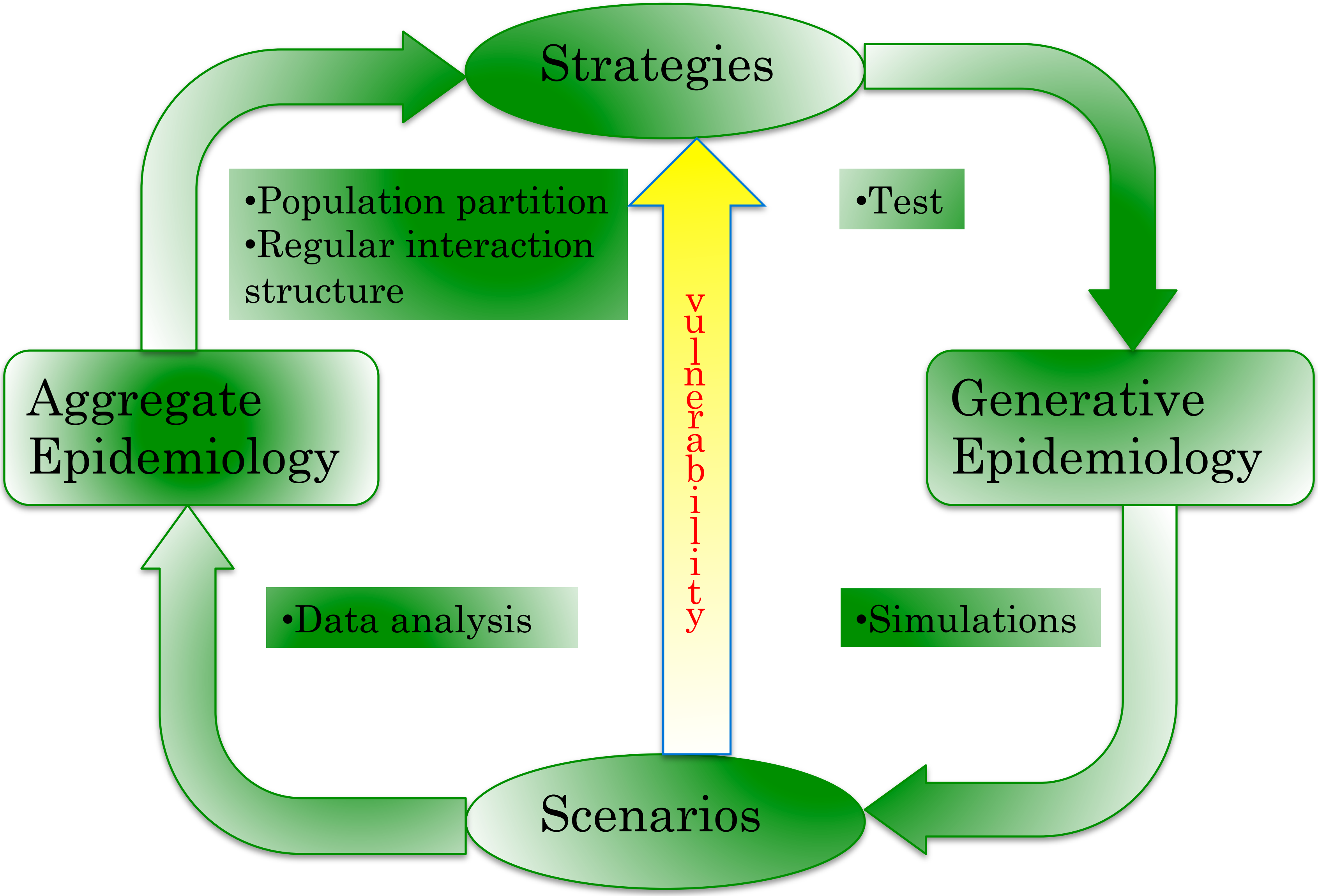


Diffusion is related to individual activity

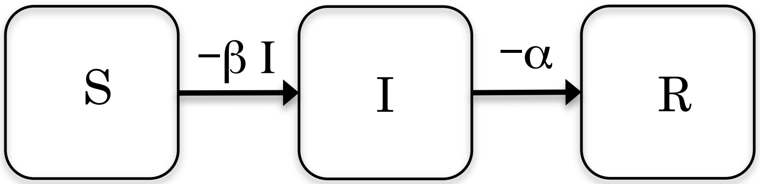


DIFFUSION OF EPIDEMICS

A. A., S. Eubank, B. Lewis
(N.D.S.S.L.)
M. McCauley (Clemson)
E. Goldstein, M. Lipsitch,
J. Miller, J. Wallinga (Harvard)


Strategies
•Population partition
•Regular interaction structure
•Test
Aggregate Epidemiology
vulnerability
Generative Epidemiology
•Data analysis
•Simulations
Scenarios

S.I.R. (Susceptible, Infectious, Removed) model



- Is a compartmental O.D.E. model
- S=#susceptibles; I=#infectious; R=#recovered; S+I+R=N
- Homogeneous mixing assumption: everybody is in contact
- No structures

$$\frac{dS}{dt} = -\beta S I$$

$$\frac{dI}{dt} = \beta S I - \alpha I$$

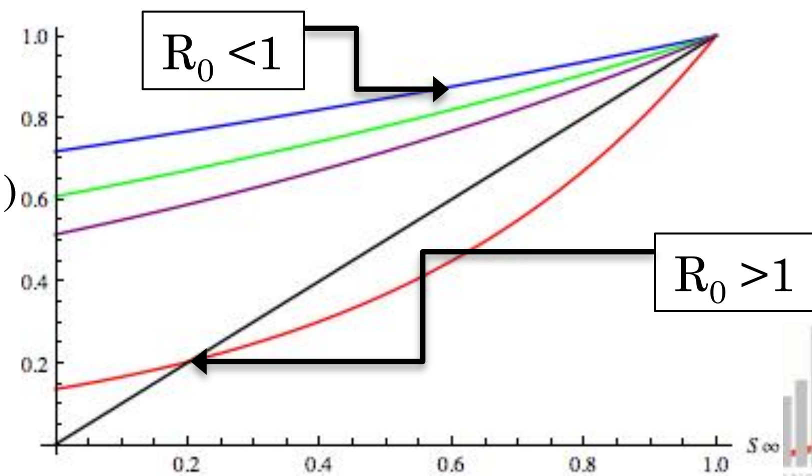
$$\frac{dR}{dt} = \alpha I$$

β=infection rate
α=recovery rate

$$S(\infty) = S(0) e^{-R_0 \left(1 - \frac{S(\infty)}{N}\right)}$$

Characteristic parameter:
$R_0 = \beta/\alpha\, N$ = **basic reproductive ratio**

- ❑ represents the average number of people infected while infectious
- ❑ $R_0 < 1$ no epidemics
- ❑ $R_0 > 1$ epidemics



• S.A.R.S. $R_0 > 2$

ADDING HETEROGENEITY

- Population divided in groups (e.g. by age).
- The average number of contacts vary between groups.
- State of the population defined by vector ψ (t): number of infected in each group.
- Next Generation matrix K: K_{ij}= average number of individuals of type *i* infected by individual of group *j*
- K is non-negative definite and has positive largest eigenvalue (λ_1) with positive eigenvector (ϕ_1)

$$\Psi(1) = K\Psi(0)$$

$$\Psi(t+1) = K\Psi(t) = KK\Psi(t-1) = K^t\Psi(0)$$

Diagonalizing

$$\Psi(t+1) = \sum_{i=1}^{n} c_i \lambda_i^t \phi_i = c_1 \lambda_1^t (\phi_1 + \sum_{i=2}^{n} \frac{c_i \lambda_i^t \phi_i}{c_1 \lambda_1^t}) \approx c_1 \lambda_1^t \phi_1$$

λ_1<1 # infected decreases
λ_1>1 # infected explodes

λ_1 defined as Reproductive number

Target of the collaboration: Asses limits for application of stratified models in real networks

Probability that a member of stratum *i* becomes infected by a member of group *j*, given that *j* is infected and *i* is susceptible

Added features of more realistic models

Depends on *i,j*, history (eg contacts of *j*), Possibly also stochastic

Reality

Small loops in transmission network (eg triangles)
WAIFW ≠WCW, even at the start of the epidemic b/c local depletion of susceptibles

Depends on *i,j* , possibly also stochastic

Traditional network model

Group structure need not be specified
Stochasticity in transmission

Deterministic function of *i,j*

Stratified mass action model

Group-specific risk factors for Infection and transmission and pairwise Behavioral interactions between groups

Independent of *i,j*

Ordinary mass action model

NEXT GENERATION MATRICES

- Static Next Generation: evaluated from duration of contacts between agents t_{ij} ,transmissibility p , average duration of infection D

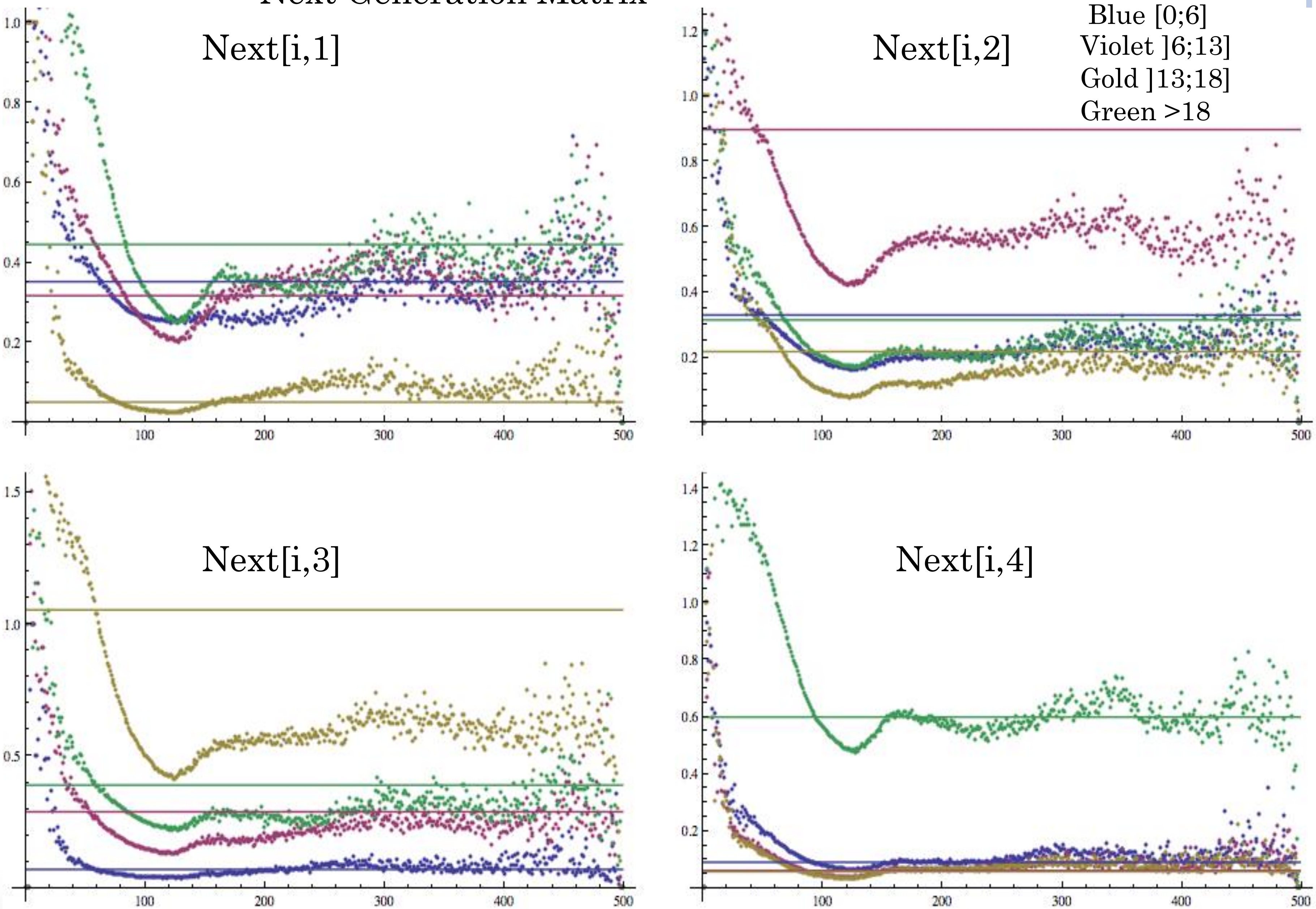
$$k_{ij}^{st} = p \cdot D \cdot \frac{t_{ij}}{s_j} = p \cdot D \cdot b_{ij} \cdot s_i$$

- Dynamic Next Generation: evaluated considering average number of people of type i infected in the future, $Y_{ij}(t)$,by individual of type j infected in day t, $X_j(t)$

$$k_{ij}^{d} = \frac{Y_{ij}(t)}{X_j(t)}$$

Takes into account activity and time

Next Generation Matrix



VACCINE ALLOCATION AS A LOGISTICAL PROBLEM

Given a certain quantity of vaccine what is the most efficient policy?
(Efficient: minimizing number of infected people)

- Random Allocation: randomly distribute
- Vulnerability allocation: vaccinate the most vulnerable nodes
- Least spread line: reducing the reproductive number

Mathematical approach: Least Spread Partition

- Consider a partition in N groups of a susceptible population: S =$\{s_1,s_2,\ldots,s_N\}$

$$\sum_{i=1}^{N} s_i = 1$$

- Allocate a certain quantity of vaccine Q with efficacy E ($0<E\leq 1$) Partition $s_i^T = s_i - Eq_i$ „ reduces the epidemic reproductive number

$$\sum_{i=1}^{N} s_T^i = T$$

- Problem reduces to finding distribution that minimize the largest eigenvalue of Next generation matrix:

$$K_{ij} = s_i a_i b_{ij} c_j$$

S = diagonal matrix
A = diagonal susceptibility matrix
B = symmetric contact matrix
C =cumulative infection force

Vaccine reduces number of susceptibles to S_T:

$$K = SABC \longrightarrow K_T = S_T ABC$$

K_T can be diagonalized:

λ largest eigenvalue

ω right eigenvector

v^T left eigenvector= $(C\,A^{-1}S^{-1}_T\,\omega)^T$

$v^T.\omega=1$

Find a partition S_T which minimizes the largest eigenvalue

Property: For a family of matrices M(t), where t is a parameter:

$$\begin{cases} \lambda(M(t)) = v(t)^T \cdot M(t) \cdot \omega(t) \\ \left.\dfrac{d}{dt}\lambda(M(t))\right|_{t=0} = v(0)^T \cdot M'(0) \cdot \omega(0) \end{cases}$$

Consider S_T is minimizing largest eigenvalue, we perturb this partition adding quantity t*U where U is a diagonal matrix of null trace. We consider the family of matrices depending on t

$$K(t) = (S_T + tU)ABC$$

$$K(0) = S_T ABC = K_T$$

$$K'(0) = \lim_{t\to 0} \frac{(S_T + tU)ABC - S_T ABC}{t} = UABC = US_T^{-1}K_T$$

We solve the equation for finding the minimum :

$$0 = \frac{d}{dt}\lambda(K(t))\Big|_{t=0} = v^T \cdot US_T^{-1}K_T \cdot \omega = \lambda v^T \cdot US_T^{-1} \cdot \omega$$

$$\sum_i \frac{c_i \omega_i^2}{a_i (s_i^T)^2} \cdot u_{ii} = 0$$

Null trace: Condition is satisfied if all the coefficient are equal to a constant value (in this case we consider 1)

$$\omega_i = \sqrt{\frac{a_i}{c_i}} s_T^i \Rightarrow \sum_j s_T^i a_i b_{ij} c_j \omega_j = \lambda \omega_i \Rightarrow \sum_j b_{ij} c_j \omega_j = \lambda \Big/ \sqrt{a_i c_i}$$

From the definition of ω and the fact that C is diagonal it follows:

$$\sum_j b_{ij}\sqrt{a_i a_j c_i c_j}\, s_j^T = \lambda \longrightarrow b_{ij}\sqrt{a_i a_j c_i c_j} = \sqrt{k_{ij}k_{ji}/(s_i s_j)} = L_{ij}$$

Definition of Next generation Matrix

So the least spread partition is given by the solution of the following linear system of N+1 equations in N+1 variables

$$L \cdot S_T = \Lambda^T$$

$$\sum_i s_T^i = T$$

Λ vector $\{\lambda, \lambda, \lambda, \ldots, \lambda\}$

In general the quantity of vaccine to distribute is unknown apriori: What is the best strategy to reach the least spread partition?

Solution linear system is proportional to $L^{-1}.u^T$, where $u=\{1,1,1,\ldots,1\}$
We indicate $s^i_1=(L^{-1}.u^T)^i$, in this way $s^i_T = T\, s^i_1$

We impose conditions on s^i_T: $(1-E)\, s^i \leq s^i_T \leq s^i$

Which are translated in the condition for T: $0 \leq T \leq s^i / s^i_1$

Necessary Condition for validity is: $0 \leq s^j / s^j_1$ for all i,j

Then we can choose T as $0 \leq T \leq min_j(s^j/s^j_1)$

Since Q=1-T, we can minimize taking the maximum allowed value for T. The vaccine allocation is given considering $s^i - T\, s^i_1$.

Vaccine allocation:
- Evaluate $s^i_1=(L^{-1}.u^T)^i$
- Evaluate $T= min_j(s^j/s^j_1)$

IXXI Institut des systèmes complexes Complex Systems Institute Rhône-Alpes

ECOLE NORMALE SUPERIEURE DE LYON

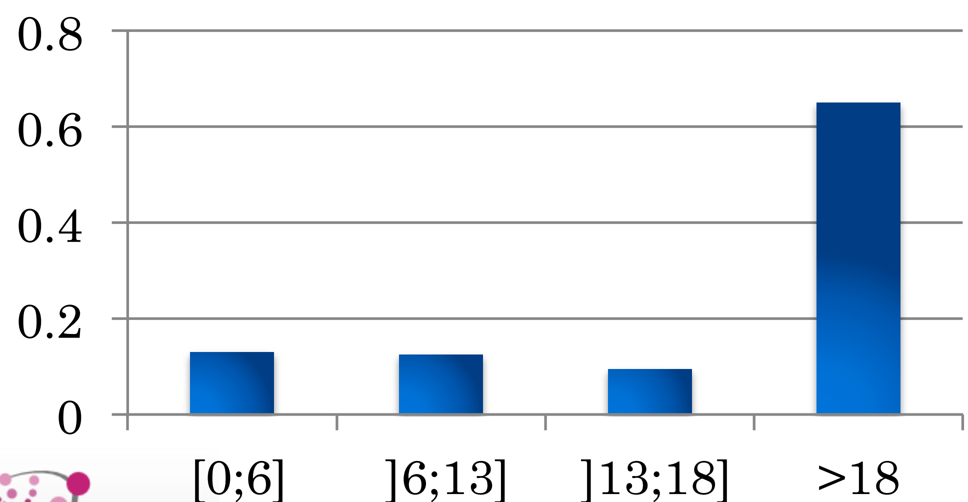
Priorities

- Procedure determines group priorities for vaccination
- Priority is given by ratio #vaccinated/(#in the group):

$$r = \frac{s_i - T s_i^1}{s_i} = 1 - T \frac{s_i^1}{s_i}$$

- Decreasing the ratio s_i^1/s_i means to vaccinate more people in the group. The lower the ratio the higher is the priority

Experiment



Population:2.2 milion
Infectivity:
Moderate strength influenza (~25% infected if unmitigated)
Seeds: 5 individuals
500 days duration

Comparisons among policies (E=1):
- Optimal allocation
- Distribution 60% of vaccine to kids and 40% adults
- Random allocation

Comparisons with larger amount of vaccine among:
- Optimal allocation
- Distribution 100% of vaccine to kids and 0% adults
- Random allocation

Consider initial partition

$$s = (0.130,\ 0.125,\ 0.094,\ 0.651)$$

Take Next generation matrix at exponential growing

$$\begin{pmatrix} 0.40, & 0.34, & 0.09, & 0.20 \\ 0.49, & 0.86, & 0.31, & 0.13 \\ 0.11, & 0.31, & 1.24, & 0.13 \\ 0.85, & 0.43, & 0.45, & 1.16 \end{pmatrix}$$

Evaluate largest eigenvalue

$$\{ 1.816,\ 1.031,\ 0.733,\ 0.091 \}$$

Evaluate matrix L

$$\begin{pmatrix} 3.08, & 3.20, & 0.90, & 1.42 \\ 3.20, & 6.88, & 2.86, & 0.83 \\ 0.90, & 2.86, & 13.19, & 0.96 \\ 1.42, & 0.83, & 0.97, & 1.79 \end{pmatrix}$$

Solve linear system and find least spread partition. T= min (β)

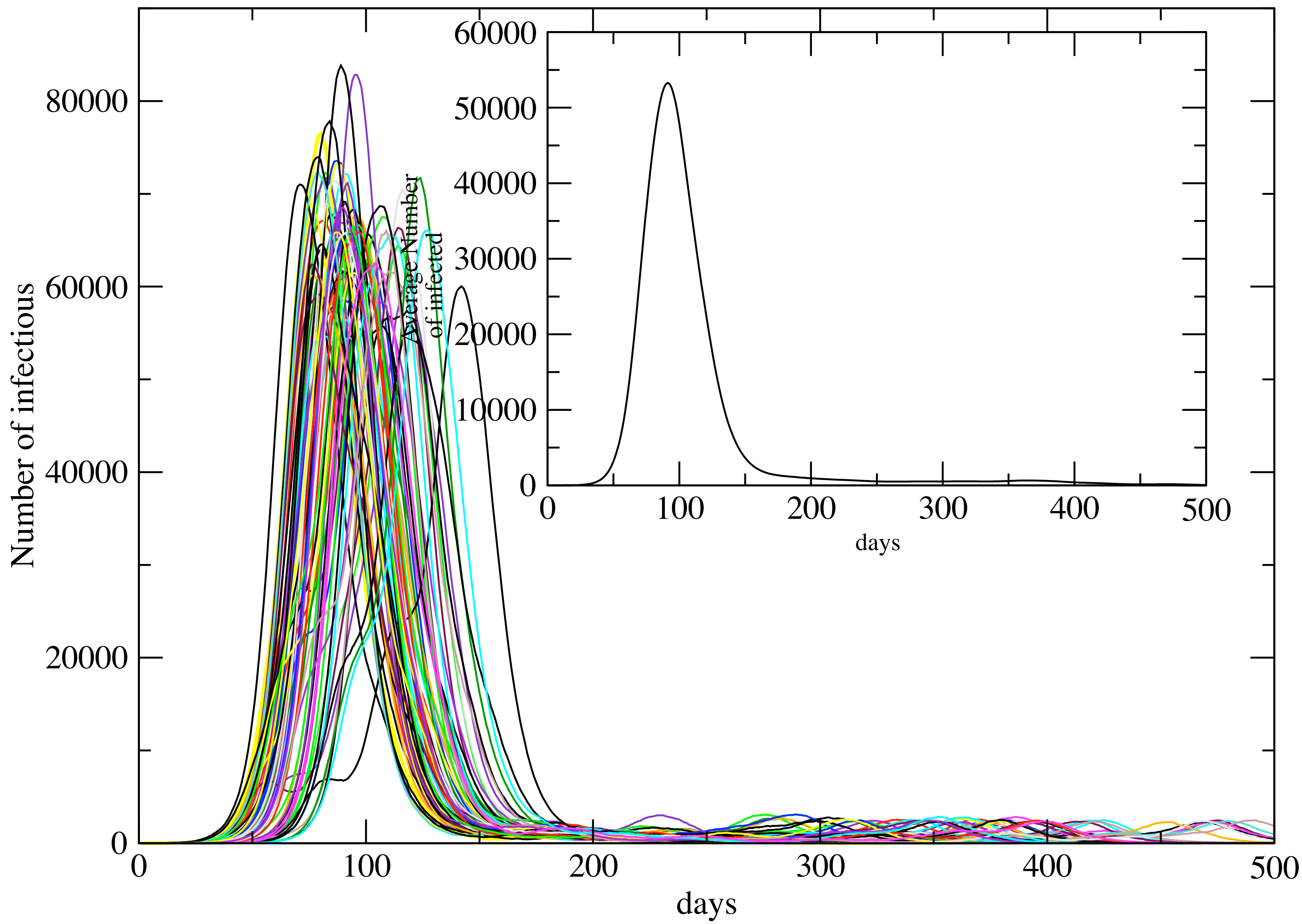
$$s^1=(0.023,\ 0.065,\ 0.024,\ 0.499)$$

$$\beta = s/s^1=(5.59,\ 1.93,\ 3.96,\ 1.31)$$

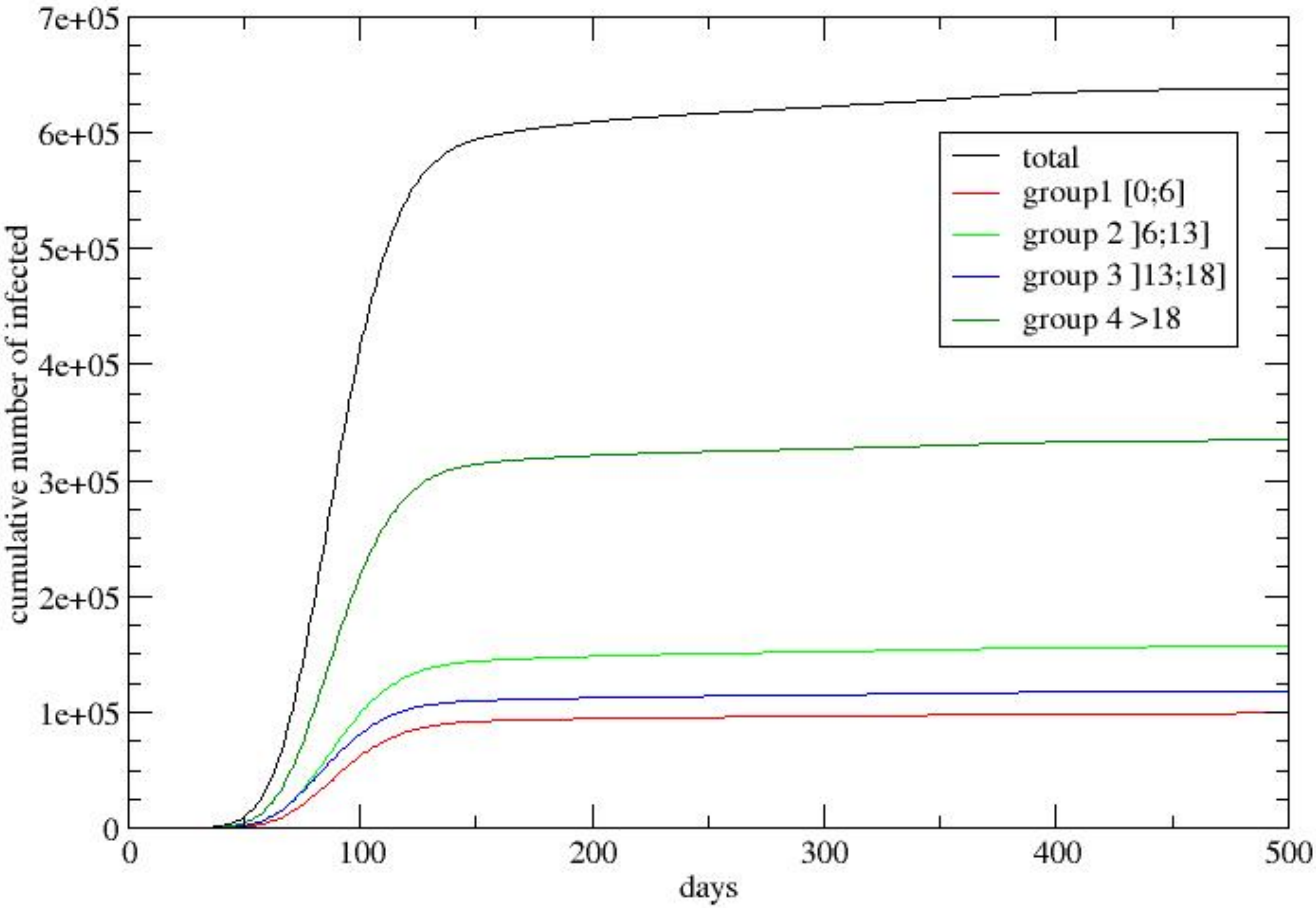
Evaluate vaccine doses and allocation

$$q=s-s^1\ T=(0.107,\ 0.033,\ 0.063,\ 0)$$

No Strategy



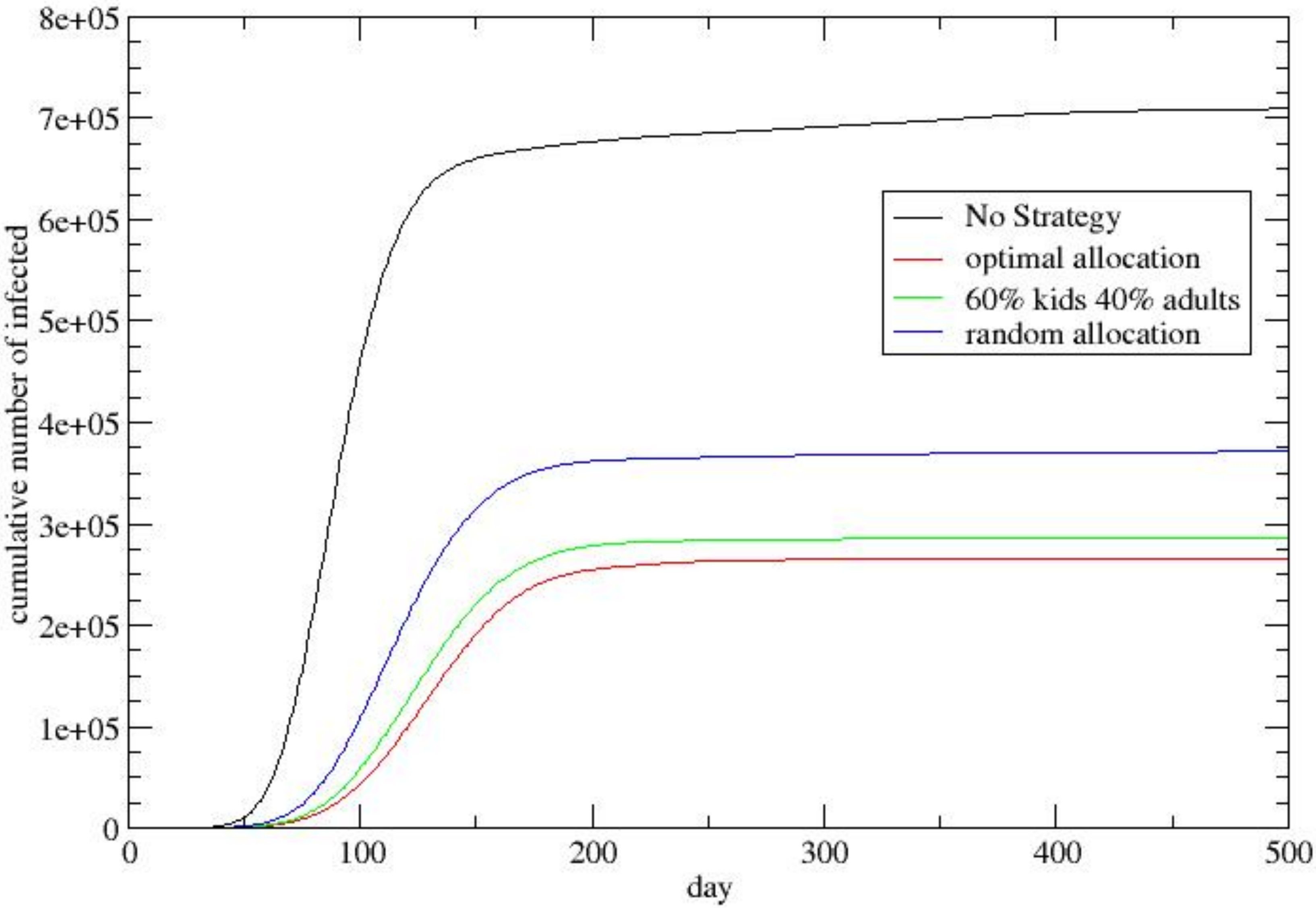
Number of infectious
80000
60000
40000
20000
0
0
100
200
300
400
500
days
Average Number
of infected
60000
50000
40000
30000
20000
10000
0
0
100
200
300
400
500
days

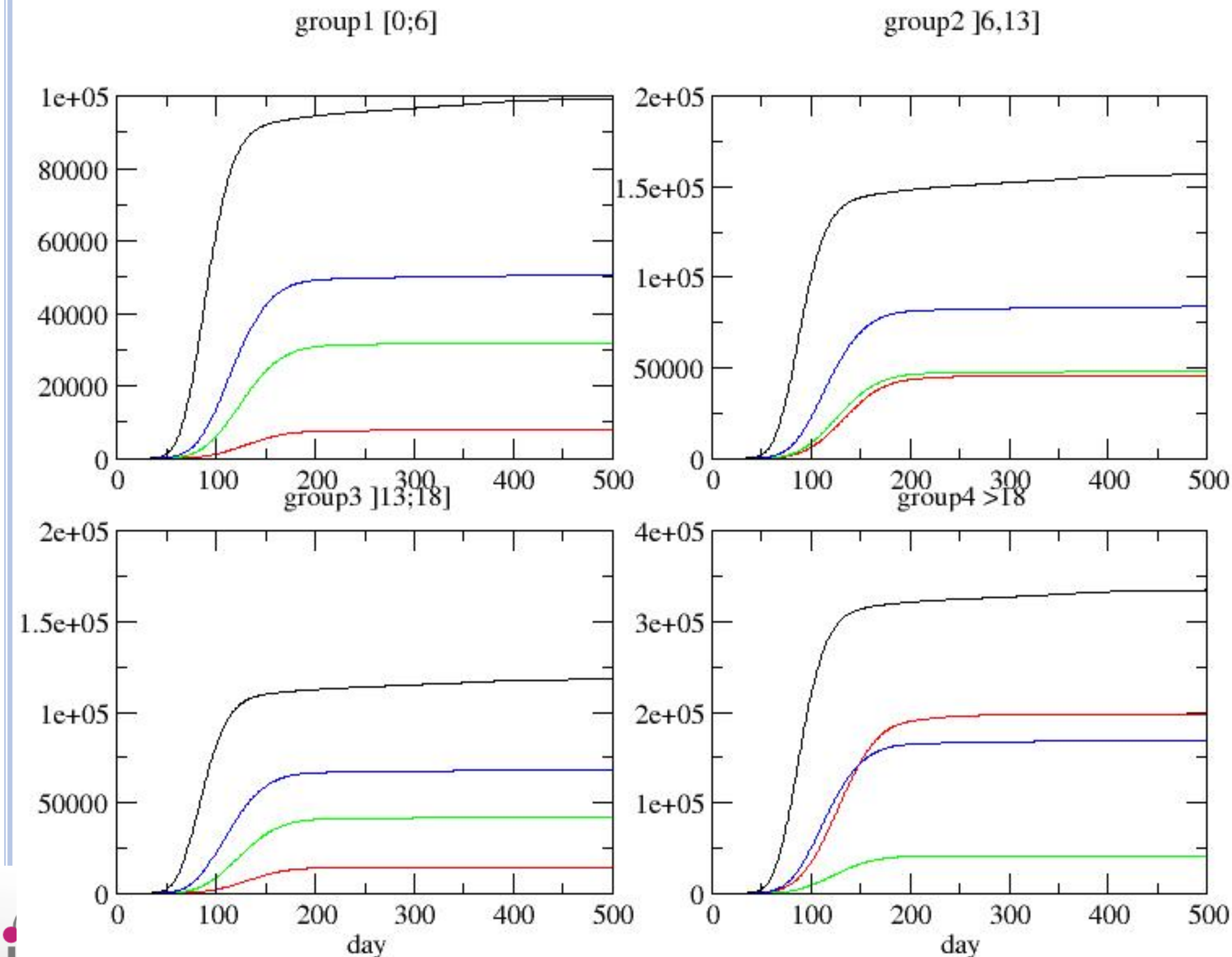

7e+05
6e+05
5e+05
4e+05
3e+05
2e+05
1e+05
0
0
100
200
300
400
500
days
cumulative number of infected
total
group1 [0;6]
group 2]6;13]
group 3]13;18]
group 4 >18

Vaccination Policies

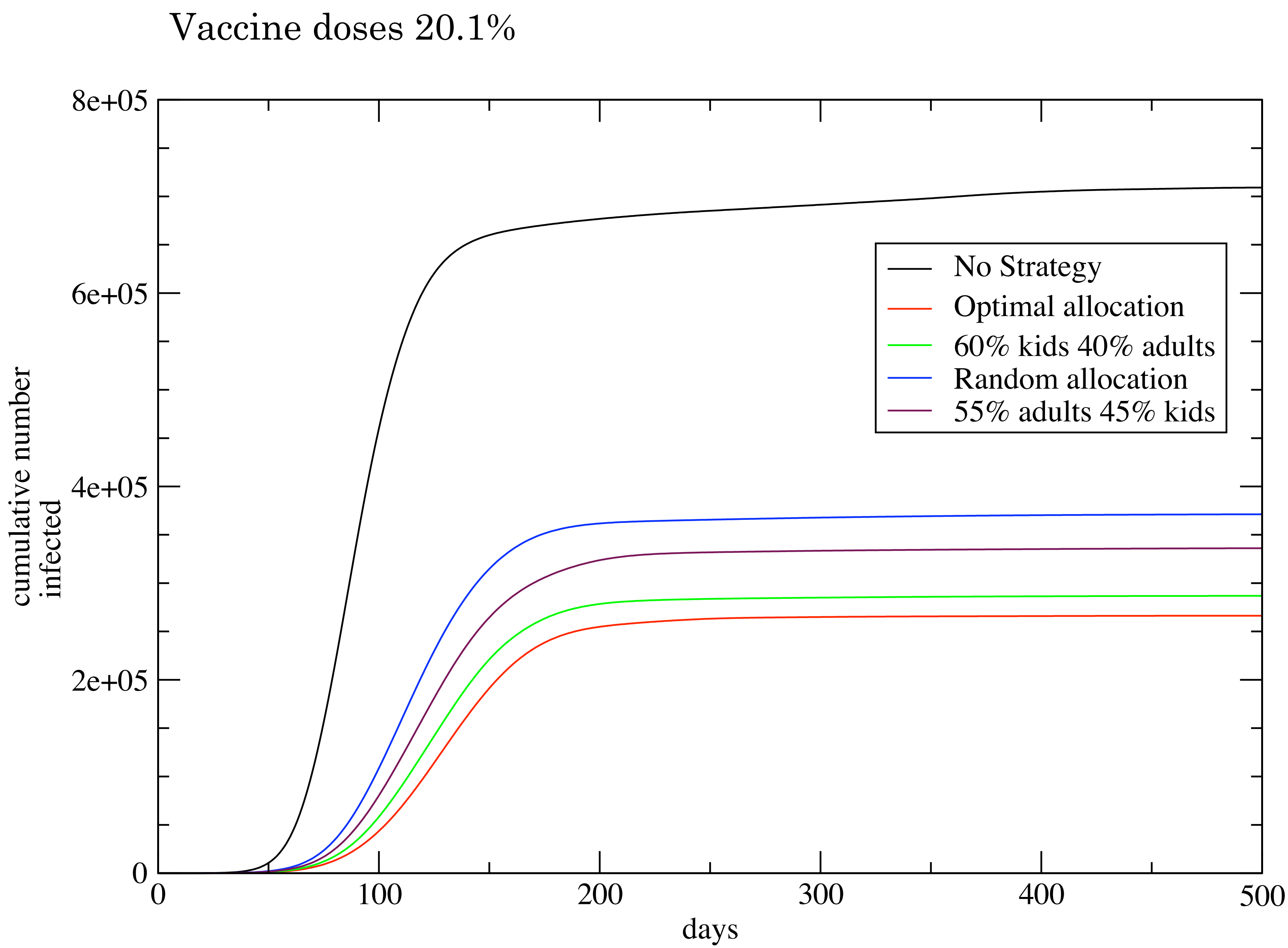
- Optimal allocation Strategy: quantity of vaccine doses is 20.1% of the population
- Strategy1: optimal allocation, doses of vaccine divided in groups:{0.107,0.033,0.061,0} which correspond to vaccinating {0.83,0.27,0.65,0} of the population in each group
- Strategy2: given 60% of doses to kids [0;18] and the rest to adults
- Strategy3:random distribution
- Strategy4:given 55% of doses to kids [0;18] and the rest to adults (CDC)

Vaccine doses 20%



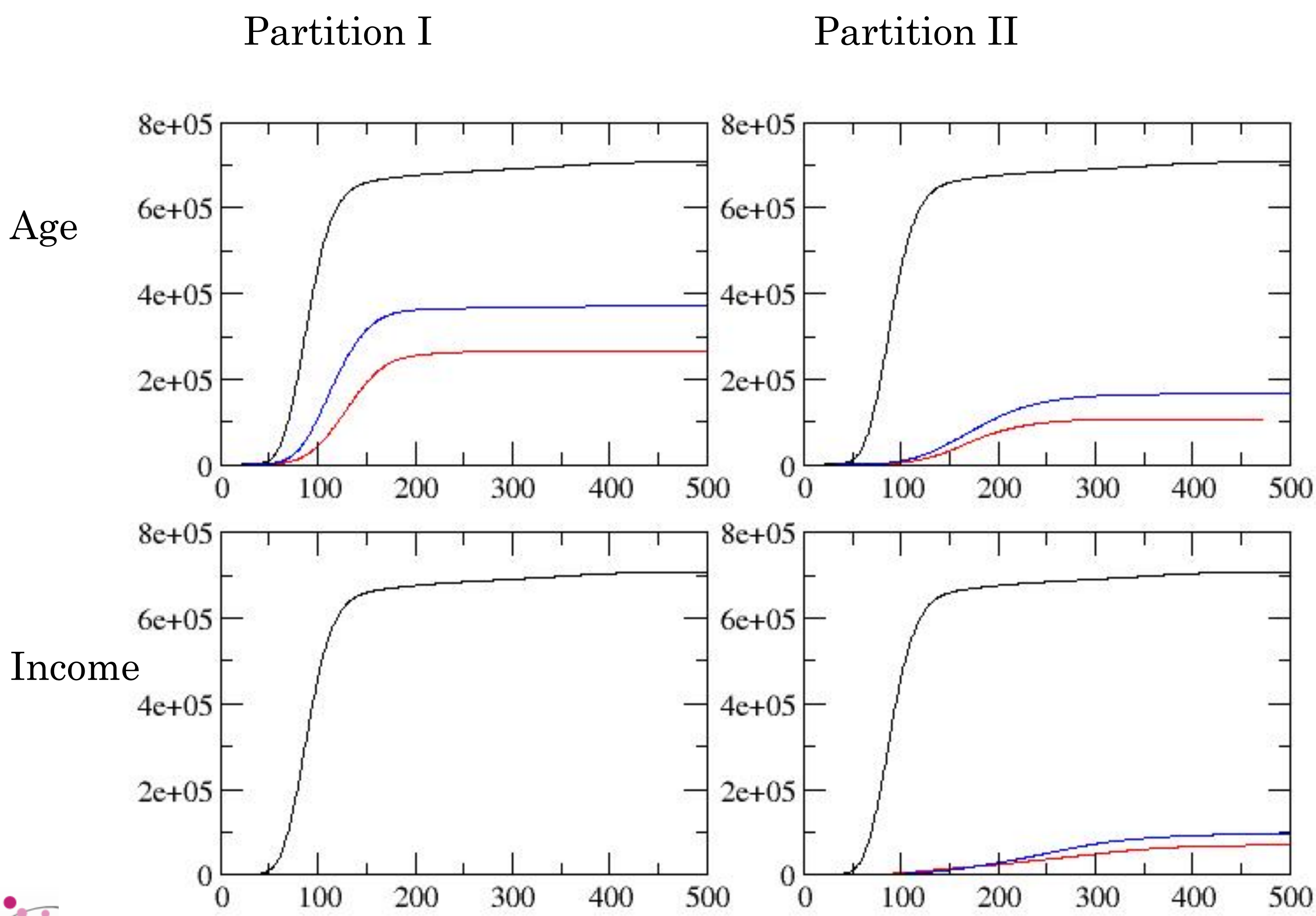


group1 [0;6]
group2]6,13]
group3]13;18]
group4 >18
day
day


Vaccine doses 20.1%
8e+05
6e+05
4e+05
2e+05
0
cumulative number
infected
0
100
200
300
400
500
days
No Strategy
Optimal allocation
60% kids 40% adults
Random allocation
55% adults 45% kids

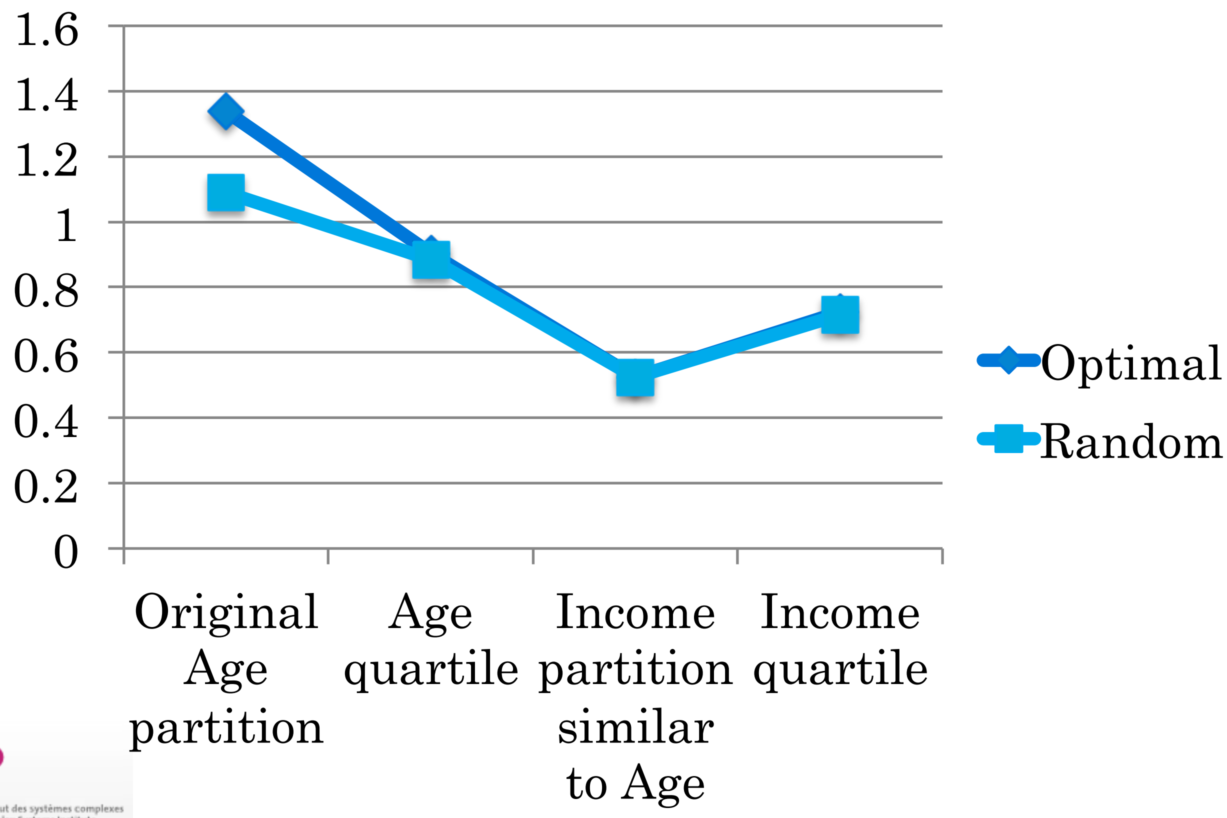
Groups and partition effect

	Partition 1	Partition 2
Age	0.130 [0;6] 0.125]6;13] 0.094]13;18] 0.651 >18	0.255 [0;13] 0.260]13;27] 0.244]27;45] 0.241 >45
	0.201 population	0.343 population
	0.107 0.032 0.061 0.	0.153 0.189 0 0.002
Income	0.130 [0;21500] 0.125 [21500;32900] 0.094]32900;40200] 0.651 >40200	0.251]0;32500] 0.252]32500;52900] 0.245]52900;78200] 0.252 >78200
	0.614 population	0.438 population
	0.041 0.050 0. 0.523	0 0.034 0.210 0.195


Partition I
Partition II
Age
Income
8e+05
6e+05
4e+05
2e+05
0
0
100
200
300
400
500

Efficiency: reduction in final size/vaccine doses

$$\text{Efficiency} = \frac{\text{Final size (without intervention)-Final size with vaccination}}{\text{Vaccine doses}}$$

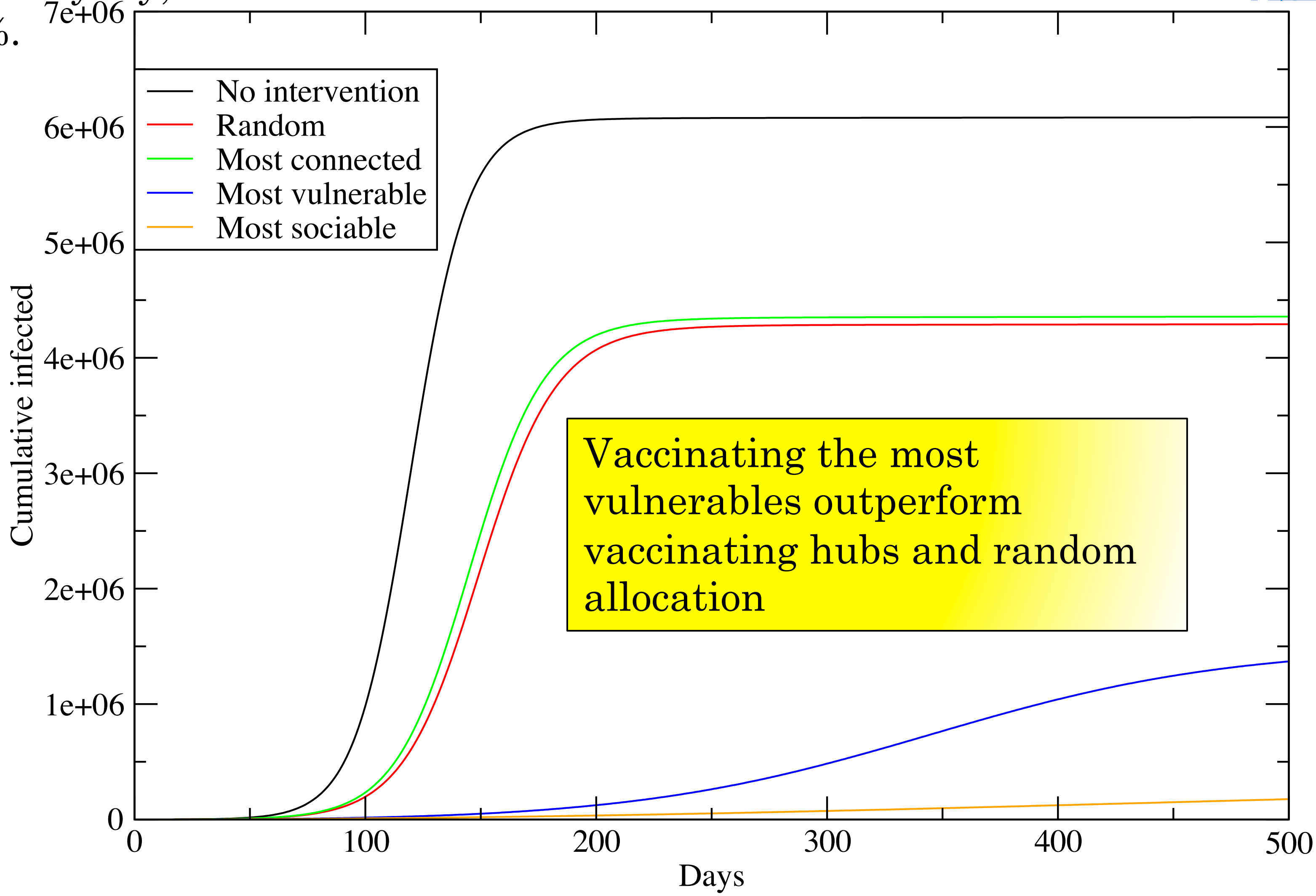


What is Vulnerability?

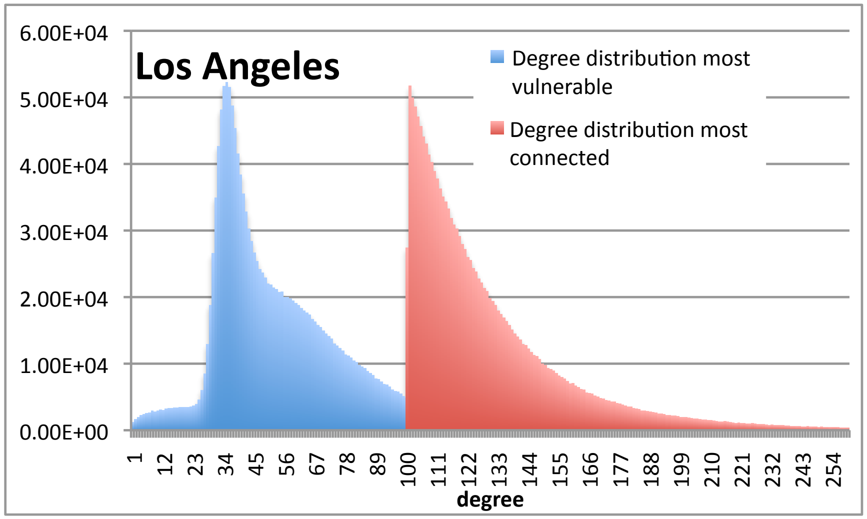
- Vulnerability : fraction of times a node is infected when we consider several epidemics outbreaks; frequency of a node gets infected when many simulations are run
- Is a "dynamical" variable: changes with transmissibility; takes account of global properties of the network; takes account of position in the network.
- Vulnerables ≠ from superspreaders: we are vaccinating individuals that are most likely to be hit, independently on their begetting (passive measure)

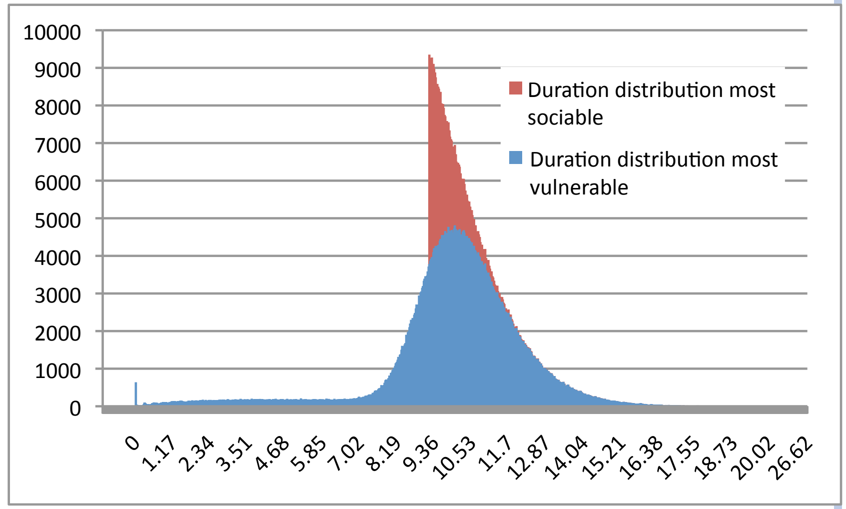
Example: Los Angeles

16 millions inhabitants
Influenza : T=4 *E(-5);
10 infected everyday;
Vaccine 10%.



WHO ARE THE MOST VULNERABLES AND SOCIABLE ?





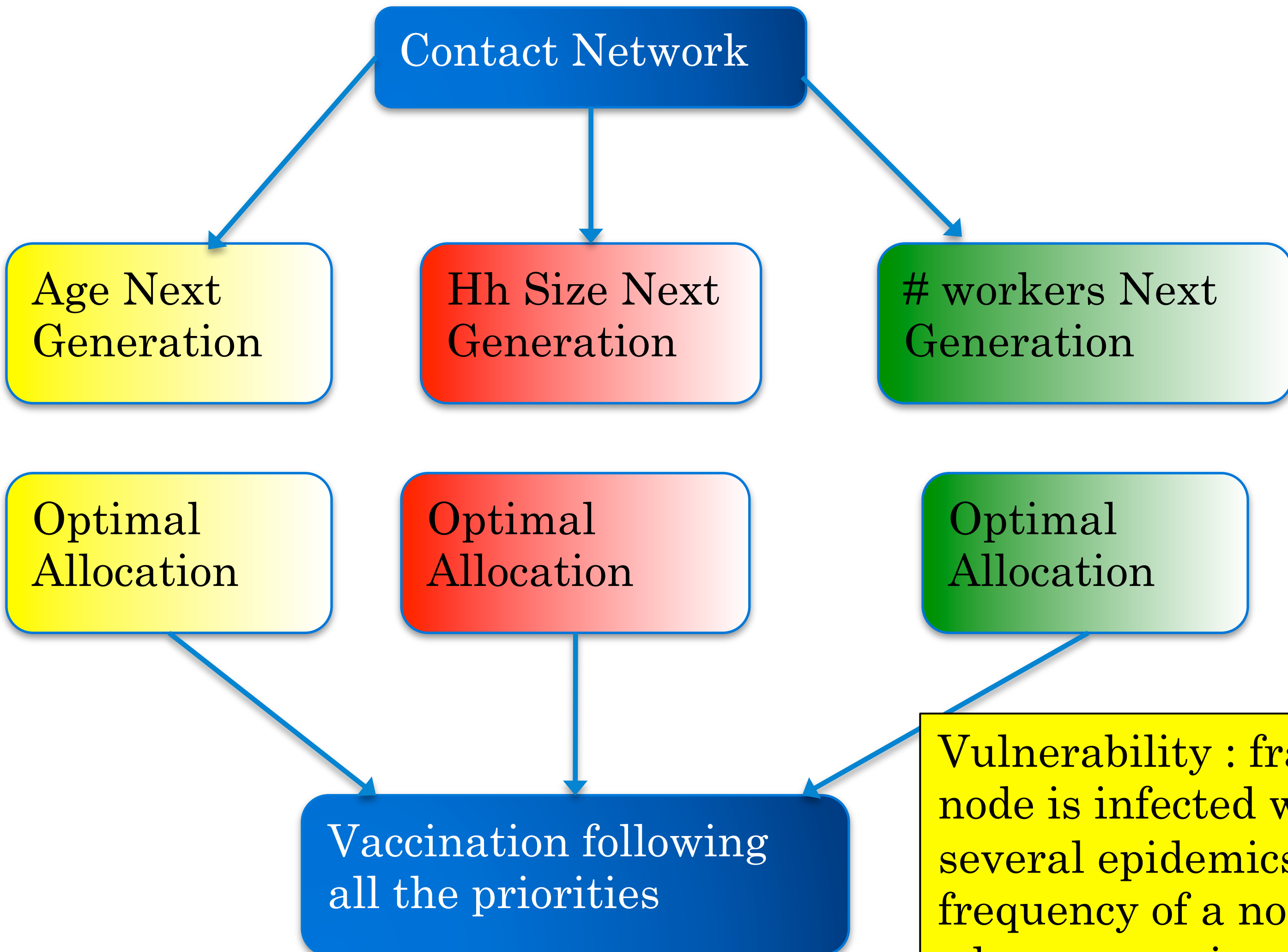
Correlation 7-8%

Vulnerables and sociable represent a specific subset of the network. Depend on the way network has been constructed and can not be traced back

Correlation 75%

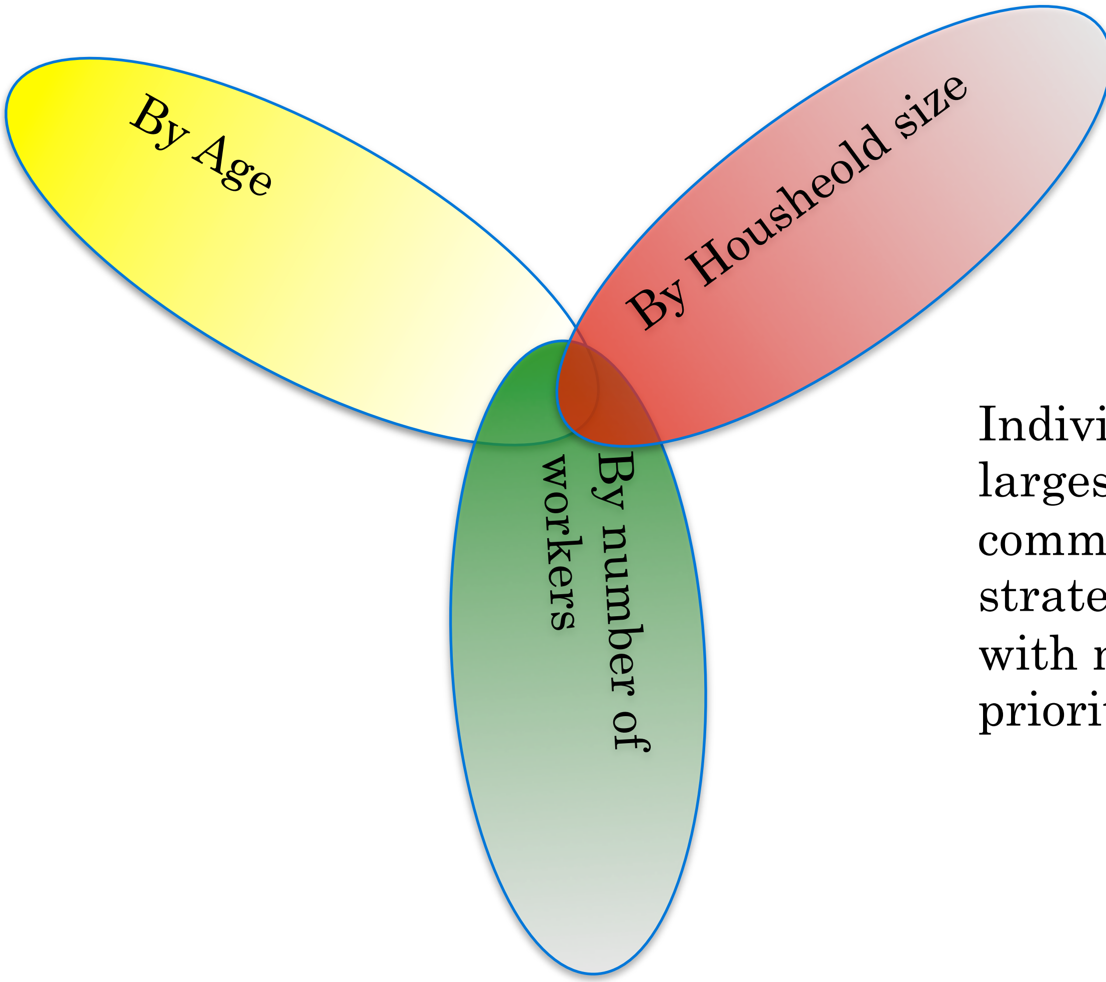
Although appealing this vaccination scheme can not be implemented as policy.

TOWARDS VULNERABILITY AND BEYOND



Vulnerability : fraction of times a node is infected when we consider several epidemics outbreaks; frequency of a node gets infected when many simulations are run

TOWARDS VULNERABILITY AND BEYOND

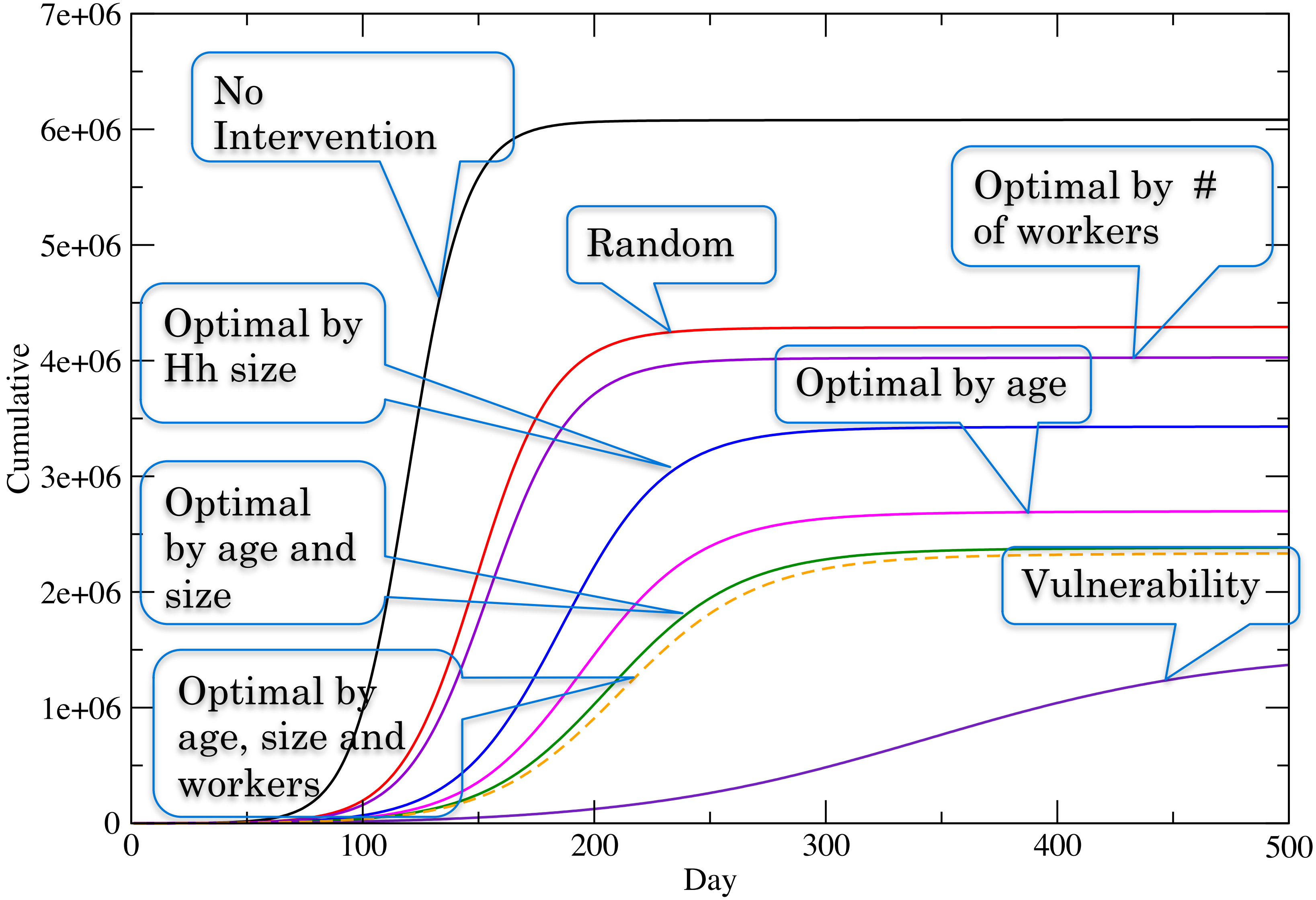


Individuate the largest subgroup common to the three strategies. Vaccinate with maximum priority

Towards Vulnerability And Beyond

- Age Groups (0-5,6-18,19-50,>50): priority to school children (they are the only one receiving)
- Size Groups (1,2-3,4,>4): priority to household with more than 4 members
- Number of workers infamily(0,1,2,3): priorities to household with more than 1 worker
- We vaccinate school children in large families: vaccinating almost all the middle and the rich and just few poor.
- Comparison to a vulnerability related: individuals with larger contact duration are vaccinated

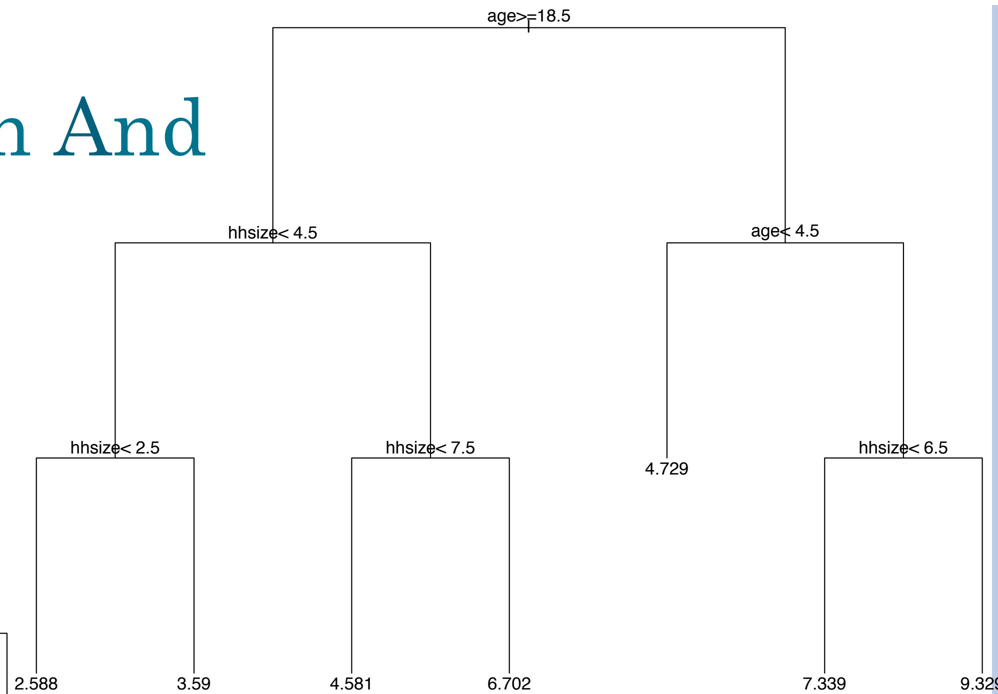
Towards Vulnerability And Beyond



IMPLEMENTING

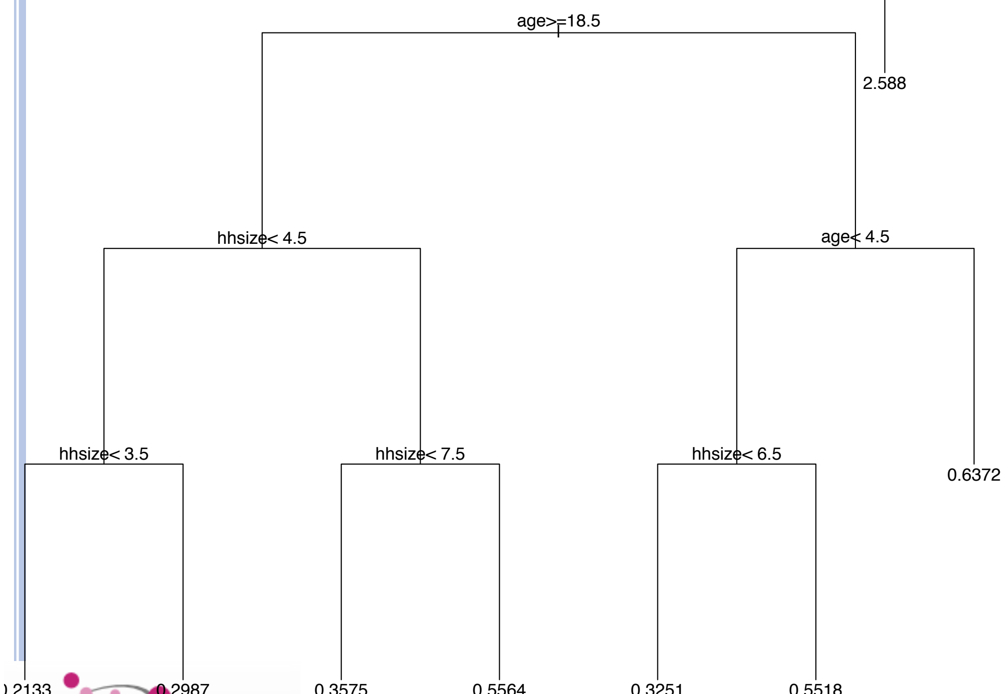
- Method above can not work: we can consider all the possible combination of demographic data; method depends also on groups' sizes….
- Using vulnerability (and contact time) as guiding parameters: we regress demographic variables related to vulnerability; this provides in one shot "important" variables, size groups.
- We use optimal vaccine allocation procedure

C.A.R.T. (Classification And Regression Technique)



Contact time

Vulnerability



	Vulnerability age	size	Duration age	size
group 1	0-4	1-5	0-4	
group 2	0-4	≥ 6	5-18	0-6
group 3	5-18		5-18	≥ 7
group 4	≥ 19	0-3	≥ 19	0-2
group 5	≥ 19	4-5	≥ 19	3-4
group 6	≥ 19	6-7	≥ 19	5-7
group 7	≥ 19	≥ 8	≥ 19	≥ 8

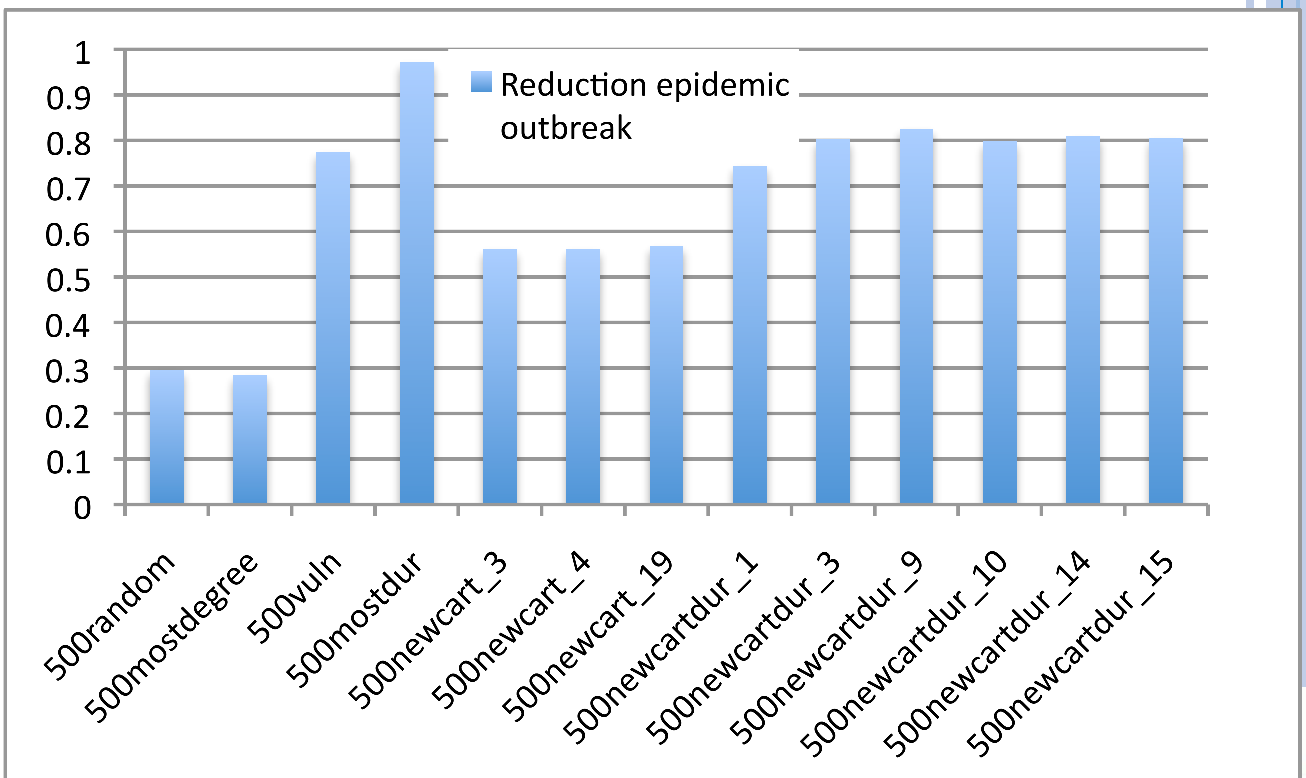
Vaccine policies and results

	new cart
cart 3	0.179 group 2; 0.41 group 3
cart 4	0.316 group 2; 0.393 group 3
cart 19	0.398 group 3; 0.18 group 7

	new dur
cart 1	0.238 group 1; 0.686 group 3
cart 3	0.891 group 2; 0.715 group 3
cart 9	0.723 group 3; 0.166 group 4
cart 10	0.935 group 2; 0.69 group 3
cart 14	0.761 group 3; 0.157 group 5
cart 15	0.857 group 2; 0.735 group 3

Some of the vaccine policies are comparable with results from vaccinating only most vulnerables

Notice that group 3 (the most sociable, consisting of school children in large families) is always vaccinated with other groups

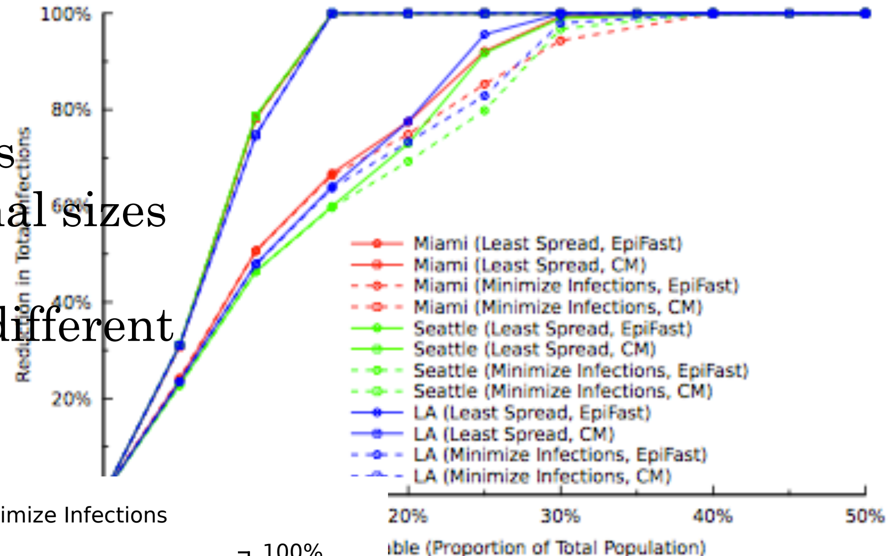


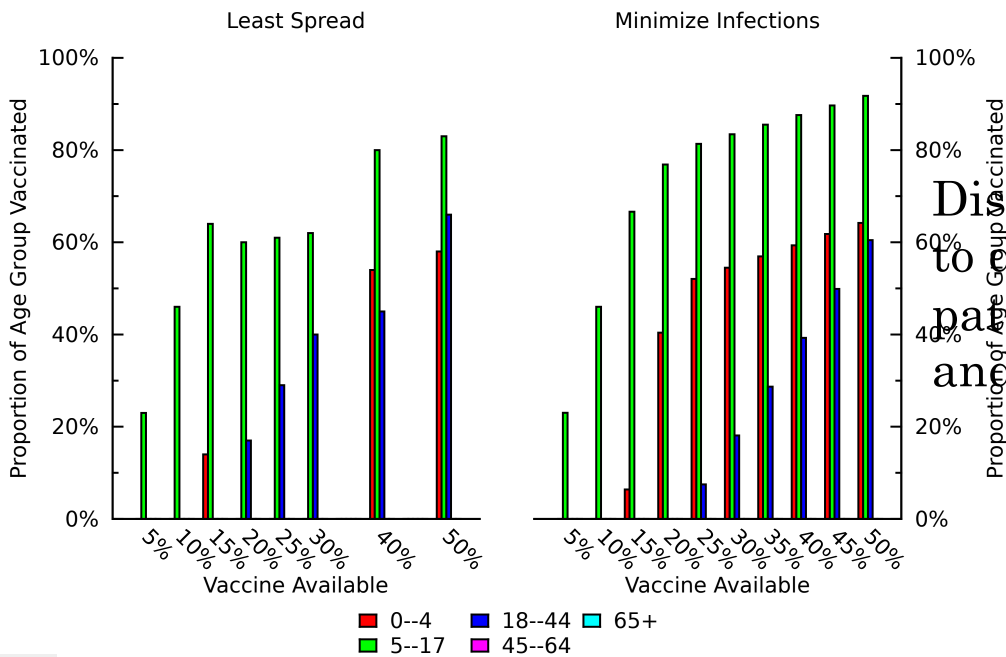
Comparison with Galvani Medlock method

Compartmental model based on optimizing certain function: number of infected; mortality etc…

Considered 3 cities:

- Provide the same contact matrices
- Evaluate parameters such that final sizes are comparable
- Evaluate optimal partition using different simulation engines



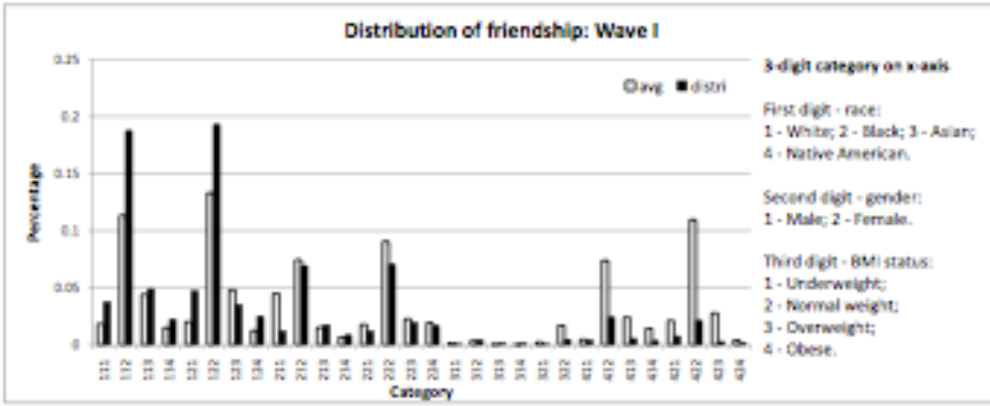


Distribution varies due to different contact pattern of age groups, and heterogeneities

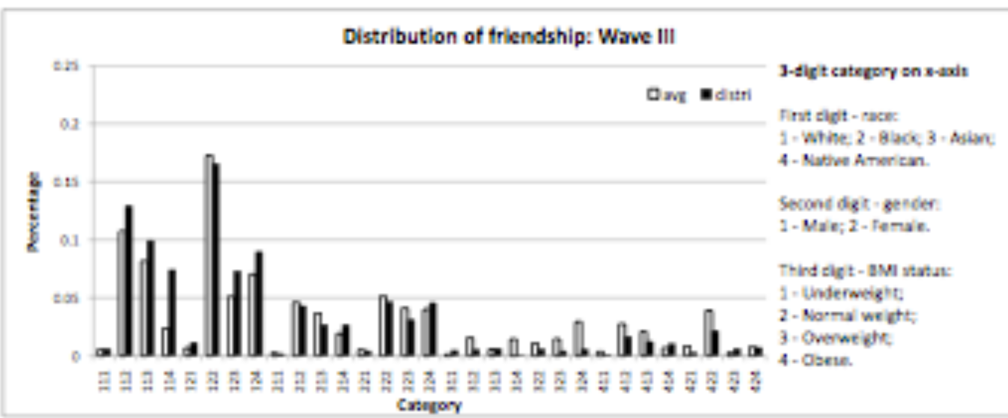
Conclusion And future experiment

- Method predicts an optimal allocation procedure compared to the random ones.
- Method strongly depends on partition criteria and selected strata
- Questions: does an optimal partition exist? What are characteristic vulnerable nodes?
- Vulnerability and contact duration provide parameters for guiding partition.
- C.A.R.T. analysis + Optimal vaccine allocation provide results comparable with vaccinating only vulnerables
- Comparison with other model and outperformance

Obesity in Adolescent



- Obesity epidemics conundrium of different social and health factors
- We use Add Health data to study popularity of obese people during 3 waves of data collection 1994-2002
- Estimates based on mutual relationship



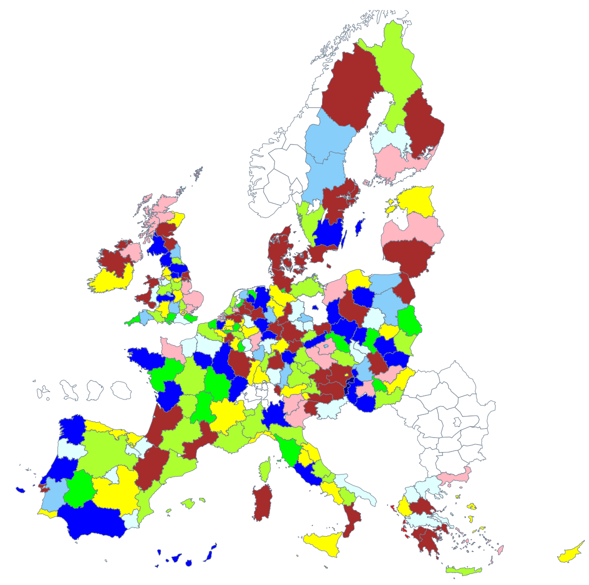
- Overweight and obese has fewer friends
- Differences according to race groups: White obese are less popular; Native American obese have higher popularity
- Obesity causes social marginalization not viceversa

Institut des systèmes complexes
Complex Systems Institute
Rhône-Alpes
IXXI

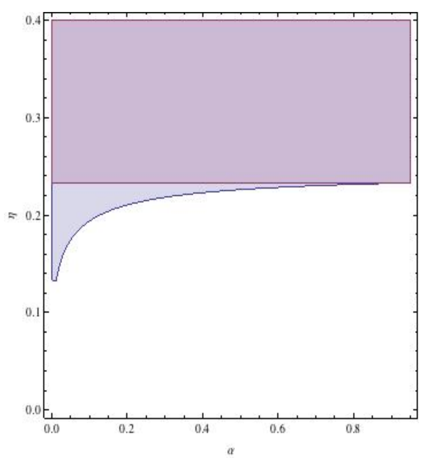
ECOLE NORMALE SUPERIEURE DE LYON

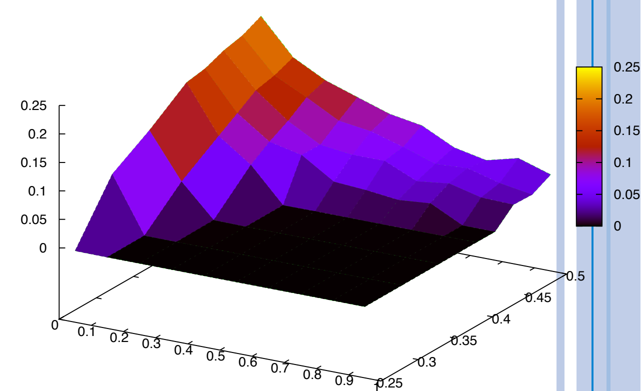
Works in Progress

- Scientometric: Analysis of citation and relation to collaboration distance



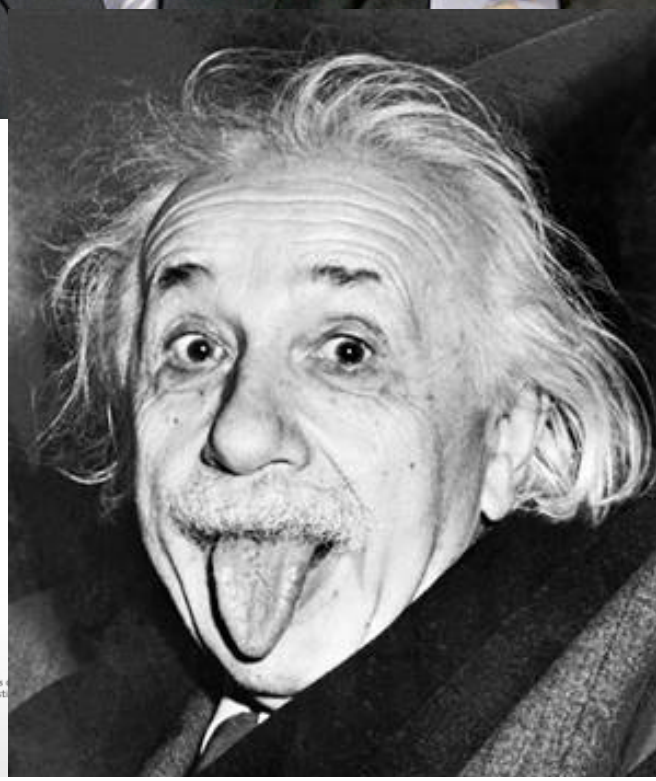
- Meta-population and structured cities. Effect of finite contact range at global level

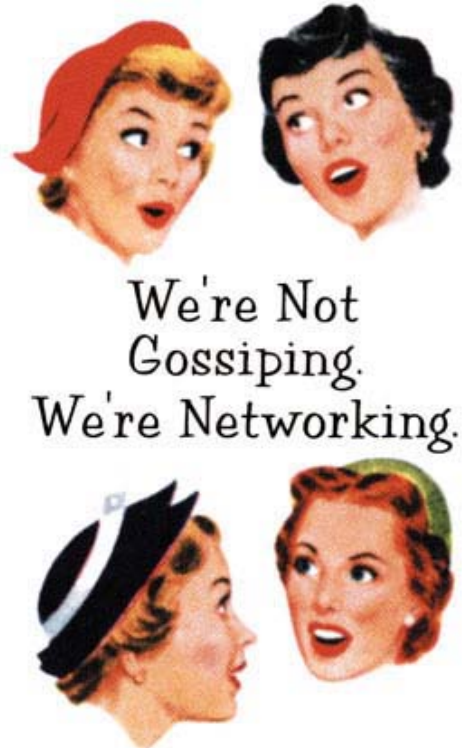




Courtesy ISI Turin

SOCIO PHYSICS AND RUMORS SPREADING





A. A., K. Channashekawa, L. Durbeck, M. Khan, C. Kuhlman, B. Lewis, S. Swarup (N.D.S.S.L.)

SOCIO-PHYSICS
(older than particle one)

FROM INTERACTIONS BETWEEN INDIVIDUALS (social psychology)…

FROM MICROSCOPIC TO MACROSCOPIC BEHAVIOUR: large systems to avoid single individual choices, interaction leads to consensus

NO CHARACTERISTIC SCALE: Critical exponents, finite size scaling, universality

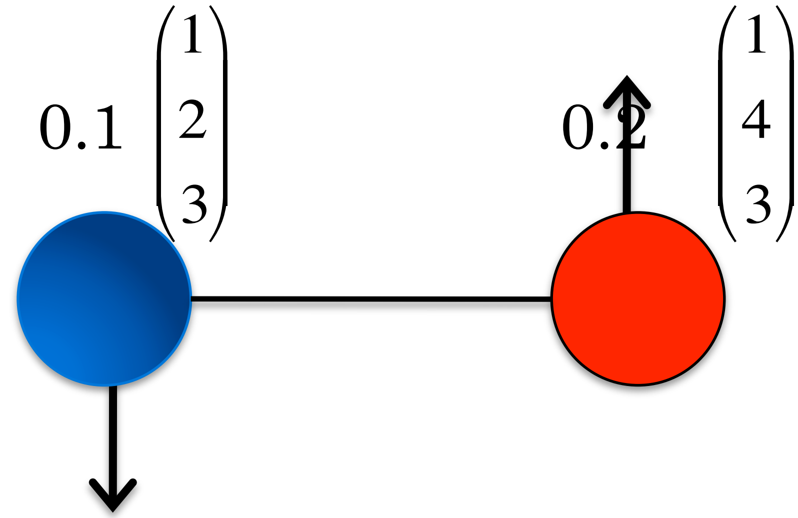
DIFFERENT BEHAVIOURS: Phase transition

…TO COLLECTIVE BEHAVIOUR (markets, fashion, politics…)

KISS: Keep It Simple, Stupid

Individuals have social characteristic (opinion) and live in a network. Each time 2 agents are chosen

Voter model: opinion is a spin variable; The agent copies other agent's opinion.
If and When consensus is reached



Deffuant model: opinion is a continuous variable. Agents interacts if not distant *(bounded confidence)* and in the case they converge average opinions. Interest in opinion areas

Axelrod: Opinion is a vector.
Interaction if enough similarity.
Change a dissimilar attribute
Probability of switching given common element.

Strength: give qualitative behavior; Numerical support to social science

Critics: No time evolution; Theoretical model of network; Interaction is not individualized (the same for everybody); restricted to small set of parameter

Motivation

- Importance at different level: gossip; political opinion shaping; financial behavior; awareness or panic during epidemics; viral marketing.
- Resemblance with epidemics processes

Precedent works

- M. Nekovee and Y. Moreno and G. Bianconi and M. Marsili, *Theory of rumor spreading in complex social networks*, Physica A, vol.374,(2007), pp. 457-470.

Institut des systèmes complexes
Complex Systems Institute
Rhône-Alpes
IXXI

ECOLE NORMALE SUPERIEURE DE LYON

COMMUNICATION BY WORD OF MOUTH

- Importance at different level:
  - gossip;
  - political opinion shaping (Galam, *The 9/11 in France*);
  - financial behavior (Kosfeld, *Crashing of market*);
  - awareness or panic during epidemics (Funk, *Spread of awareness*);
  - viral marketing (Domingos et al. *Mining Knowledge*
- Controlling and spreading: where is the truth? (Agliari et al. *Efficiency of information spreading*)
- Resemblance with epidemics processes (Nekovee et Al. *Theory of rumor spreading in complex social networks)*
  - It spreads by word of mouth, needs a local (social) network
  - People who heard rumor are likely to spread
  - It could self-sustain after a certain threshold

Assumptions

Network :static, with theoretical degree distribution; evolving according to particular rules

Transmission probability: constant and average

Our Contribution

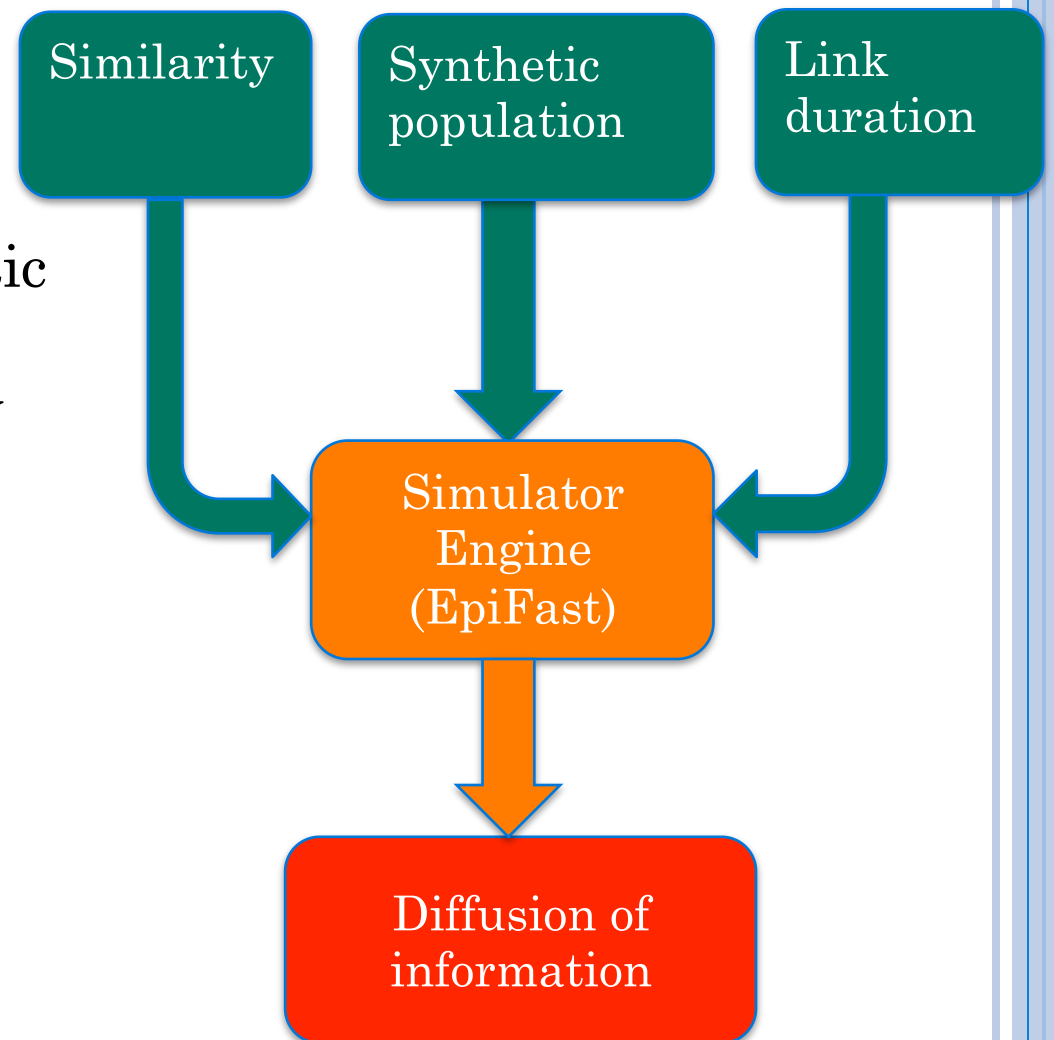
Realistic Network
Individuals follow a certain routine, meet individuals at a specific time of the day and in particular places

Individuals dynamics:
Communications between conversants with some relation

Conversation time:
Importance of the message is related to the time we need to reach the topic

Our contribution

- We consider a diffusion of information by words of mouth in a realistic social network, built using synthetic population.
- Transmission can occur only if:
  - there is enough similarity between conversants
  - the contact duration time is larger than a certain threshold (indicated as *ndt* and expressed in terms of time units=10 min)



SIMILARITY

Homophily Principle: people interact with similar and try to become more similar

Status: race, ethnicity, sex age, income, household size

Value: internal states that shape behaviors such as, beliefs, values and attitude

Relation status and values

No information about individual values

Consider only status similarity (age, income, household size)

Probability of spreading is given by fraction of common elements

LINK DURATION

- Discussions need time
- Importance of the topic
- Different time scale implies pruning network: parameter *ndt*: 1,10,20 t.u.(=10 min)

0 Second:

Gossip
Sport news
Triviality

10 minute:

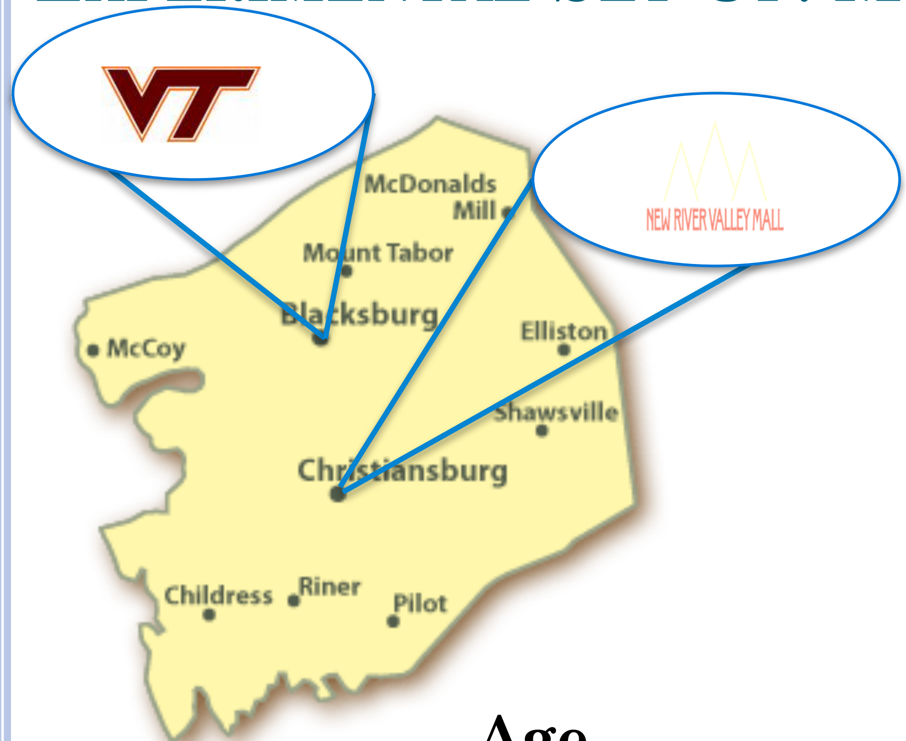
Marketing
communication

1 hour:

Politics
Personal affairs

>2 hours:

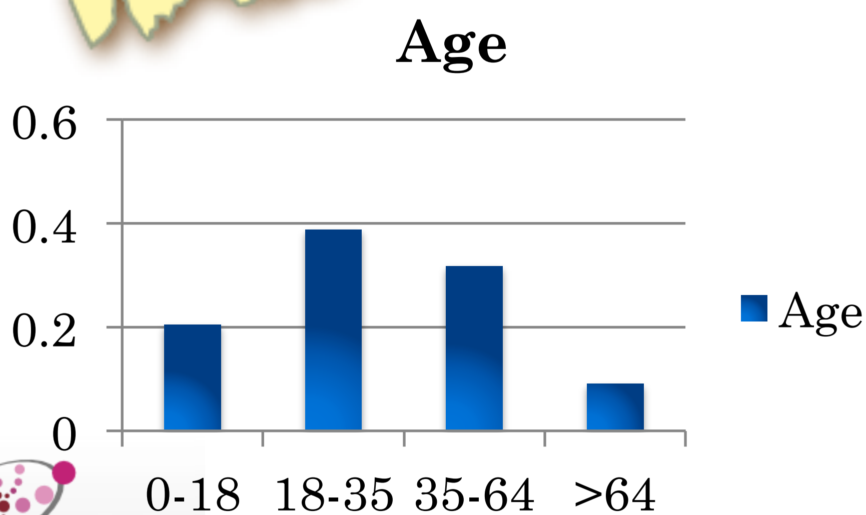
EXPERIMENTAL SET UP: MONTGOMERY COUNTY (VA)

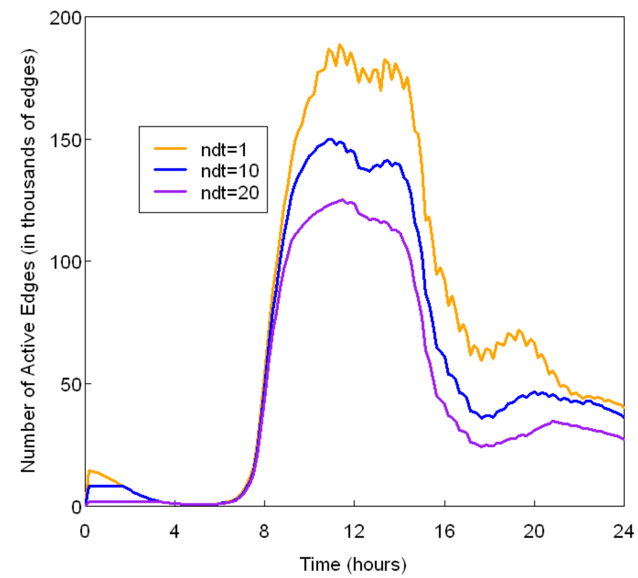


- Population:74.360
- Links created~1.800.000
- Similarity based on:

Age ([0;18],[18;35][35;64],>64);
Hh Size (1,2-3,>4);
Hh Income k$ ([0;25],[25;50],[50;75],>75);
48 possible groups

- Links duration threshold (*ndt*):1, 10, 20 t.u.;
- 1 Initially infected chosen randomly or among kids
- Experiment lasts for 3 days and is repeated 100 times

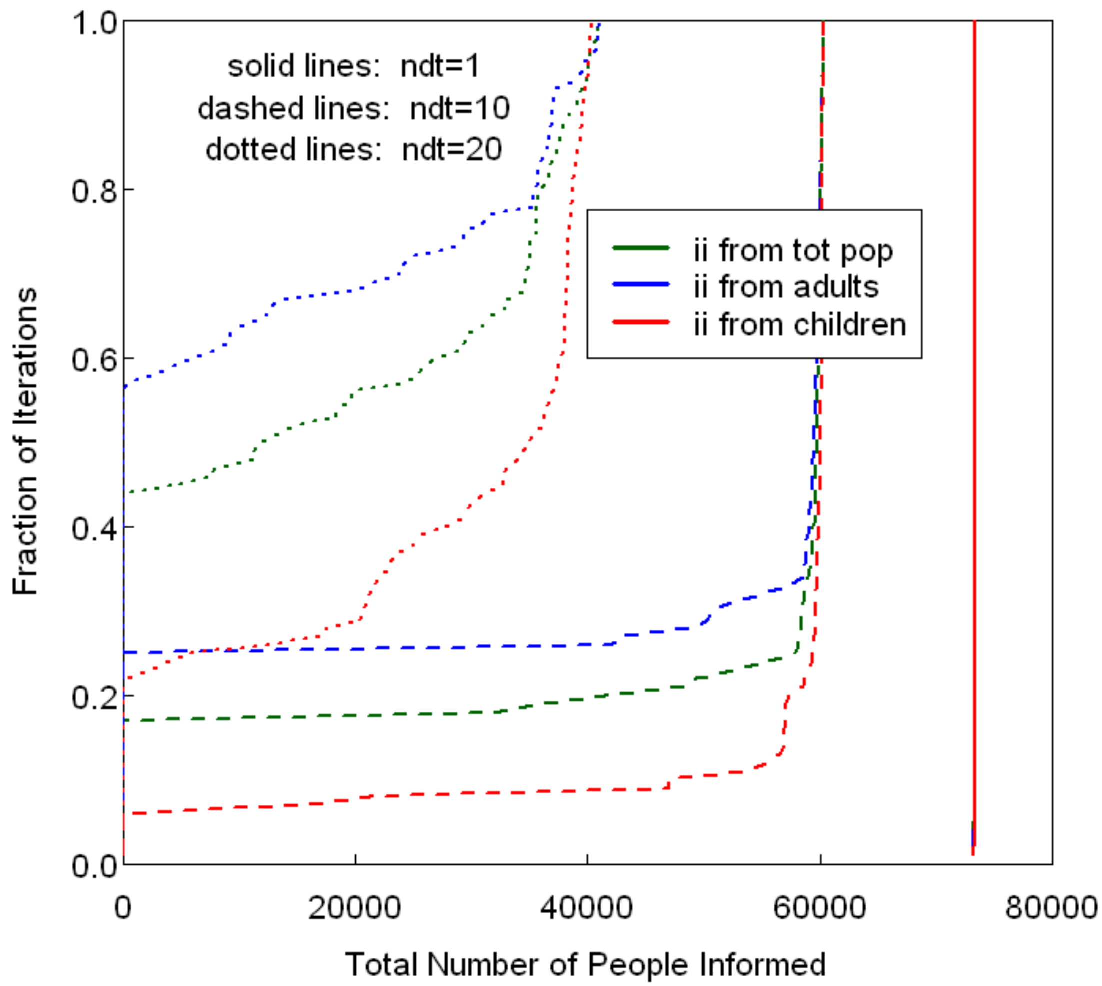




INFLUENCING FACTORS

- Diffusion depends on the number and type of source
- Diffusion strongly depends on link duration

Kids are more communicative

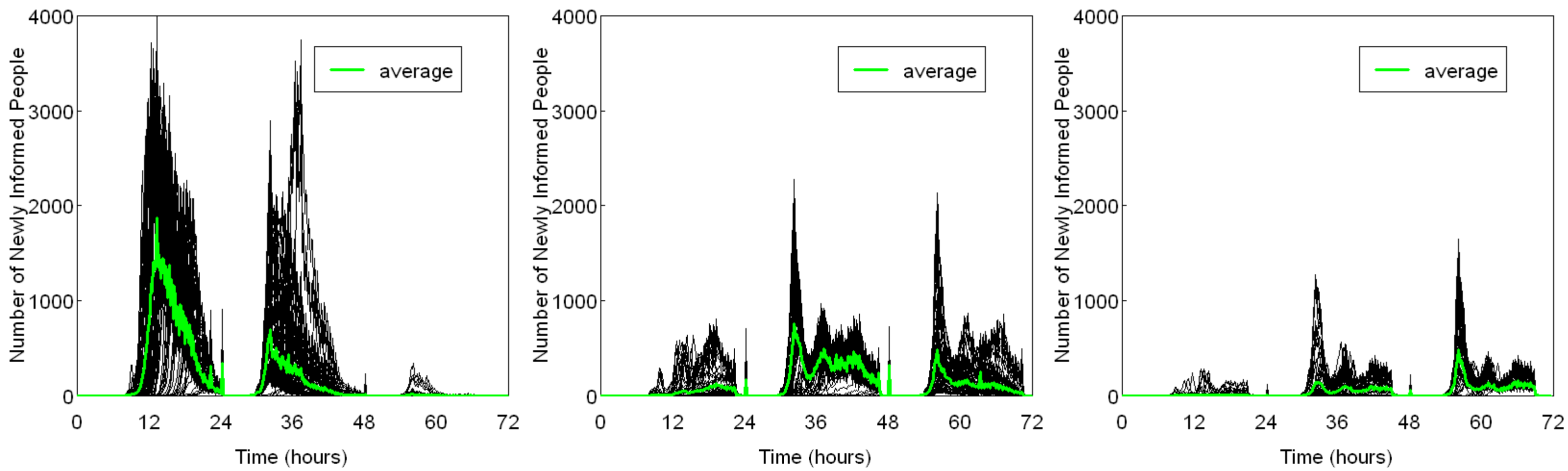


What we are interested in

Effect of link duration, type of source on:

- Distribution in the groups: What are the most informed groups?
- Communication channel: How does communication flow among age group?
- When and where could communication occur?

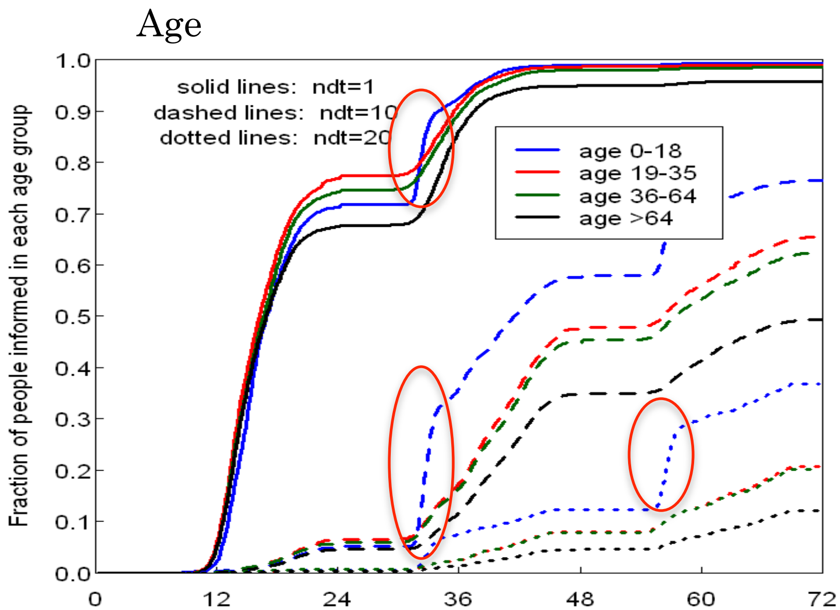
TEMPORAL EVOLUTION

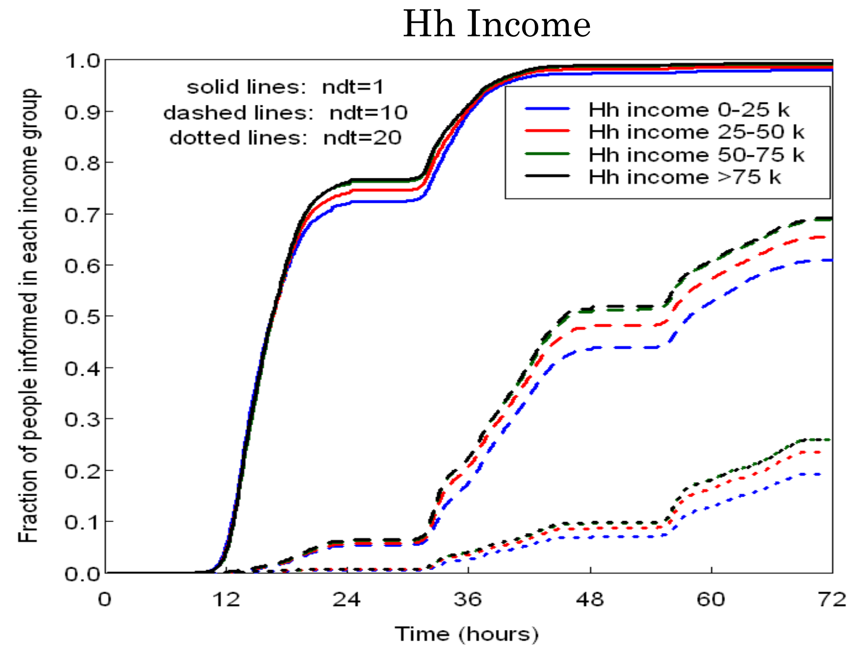


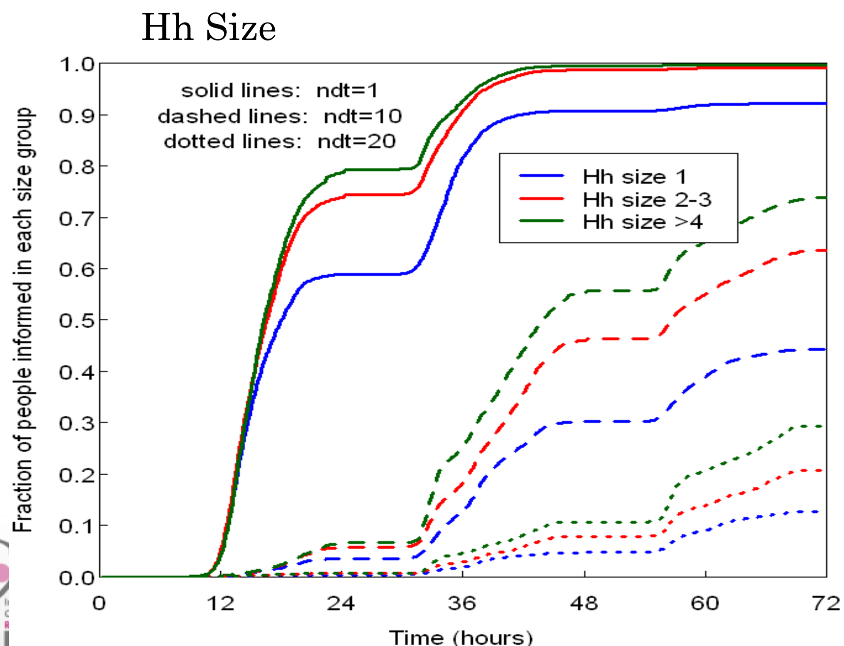
R
a
n
d
o
m

- Increasing link threshold, peak is shifted on the future
- Total number of informed people decreases

How are they distributed?





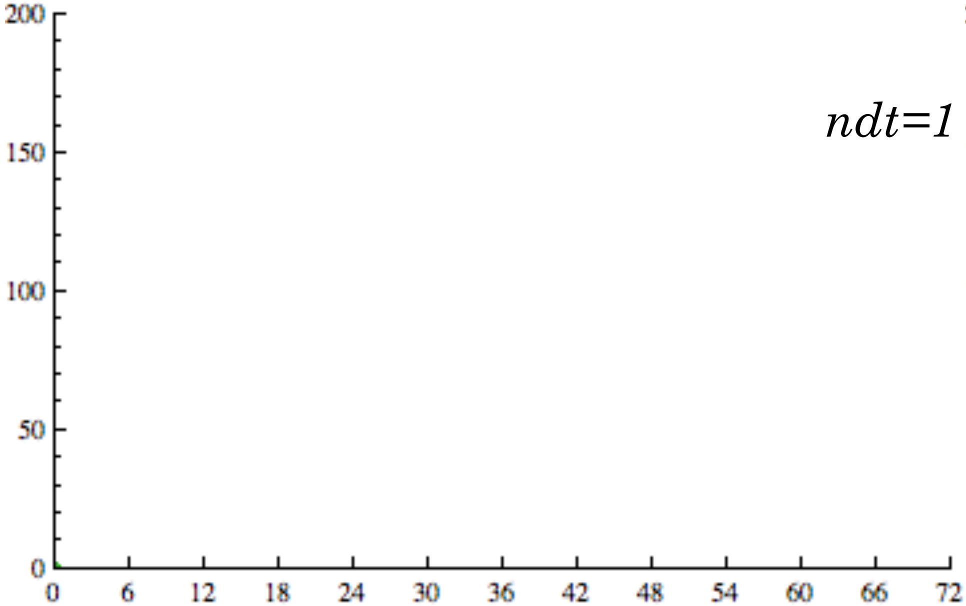


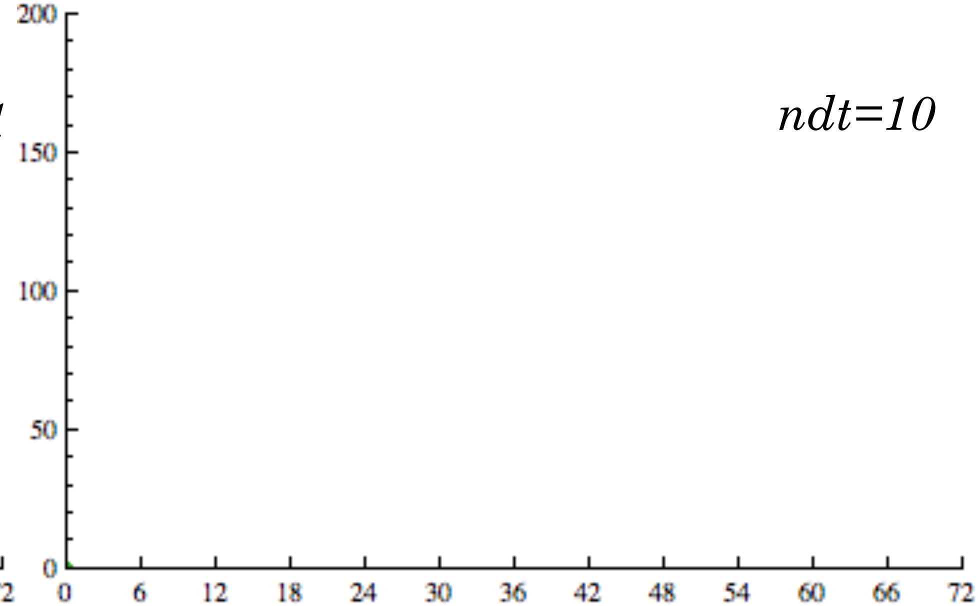
- Importance age groups and activity changes with link duration
- Income has not differentiating role
- Differences Hh size>1 increase with increasing link duration. Singles less informed

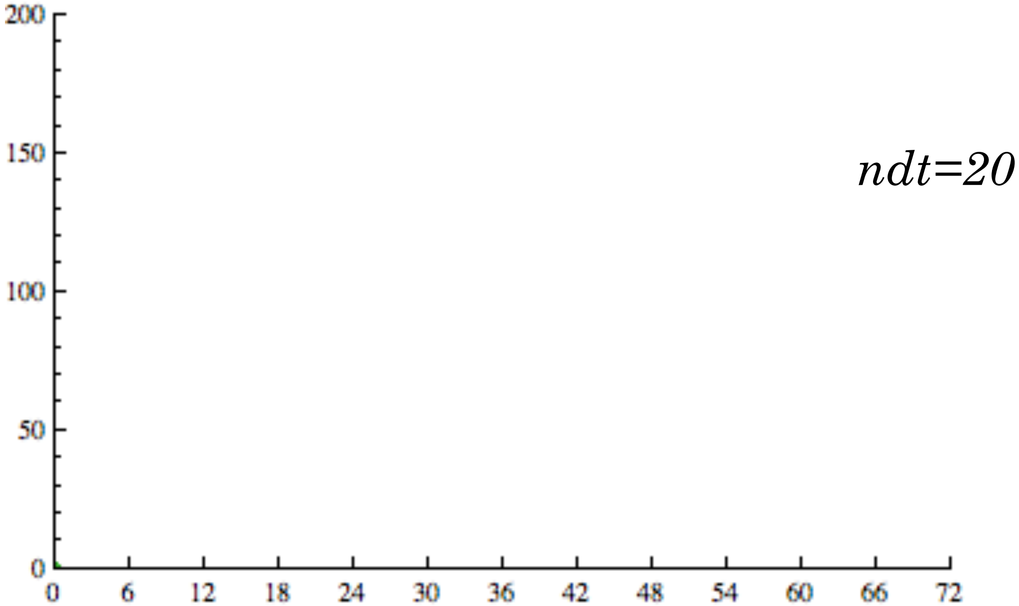
ECOLE NORMALE SUPERIEURE DE LYON

WHO INFORMS WHO

● Child => child
● Adult=>adult
● Child=>adult
● Adult=>child



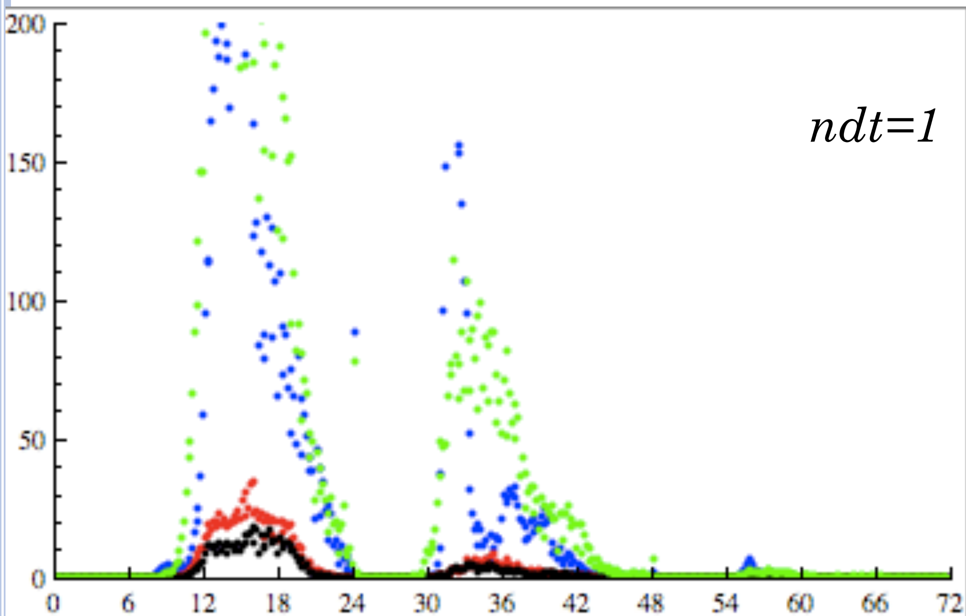


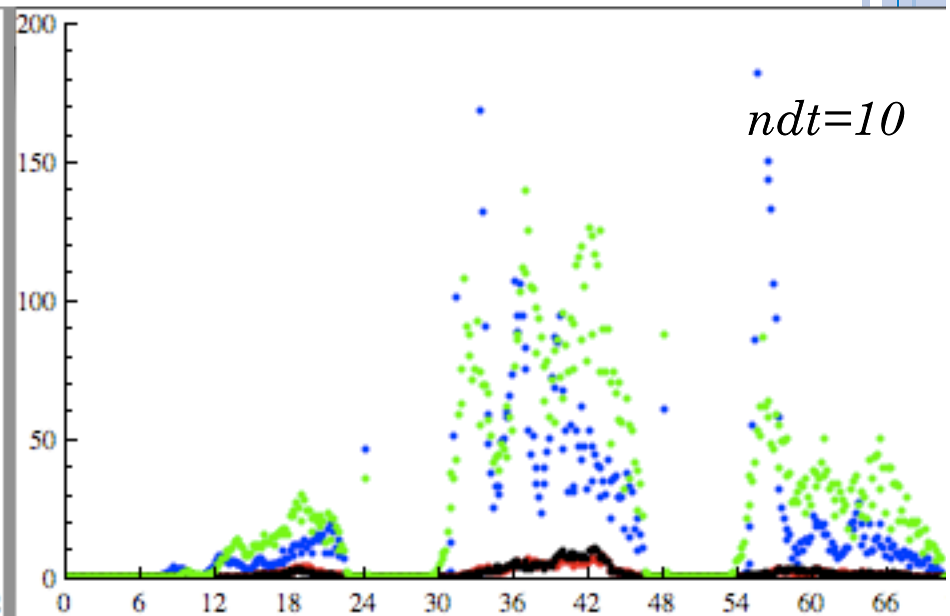


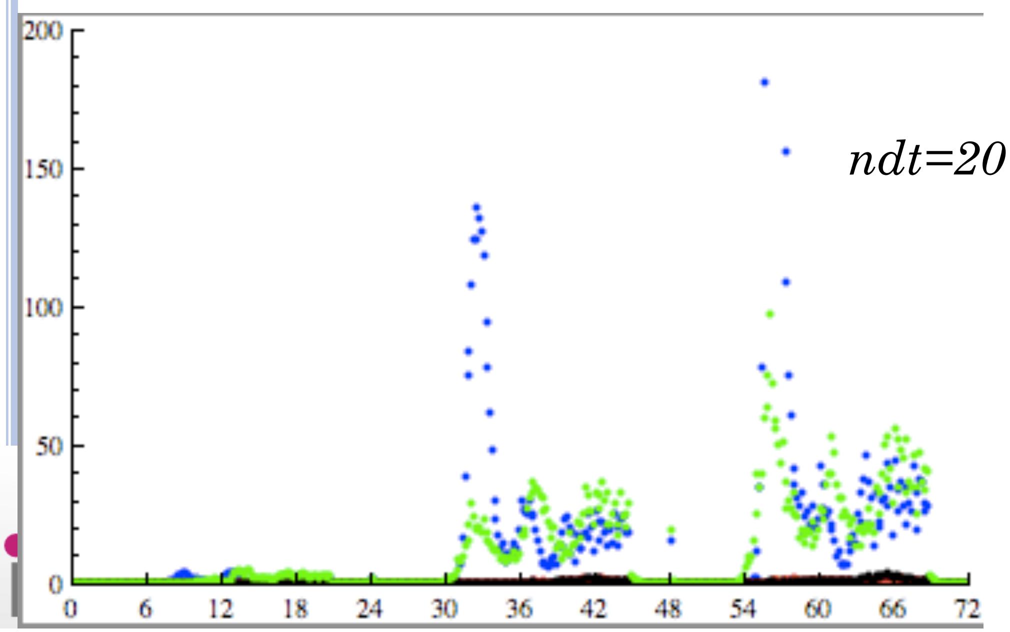
Communication is enhanced in the same age group.
Increasing *ndt*, shift and reduction of informed. Kids become more important
Kids listen to adults more than adults listen to kids

WHO INFORMS WHO

- Child => child
- Adult=>adult
- Child=>adult
- Adult=>child

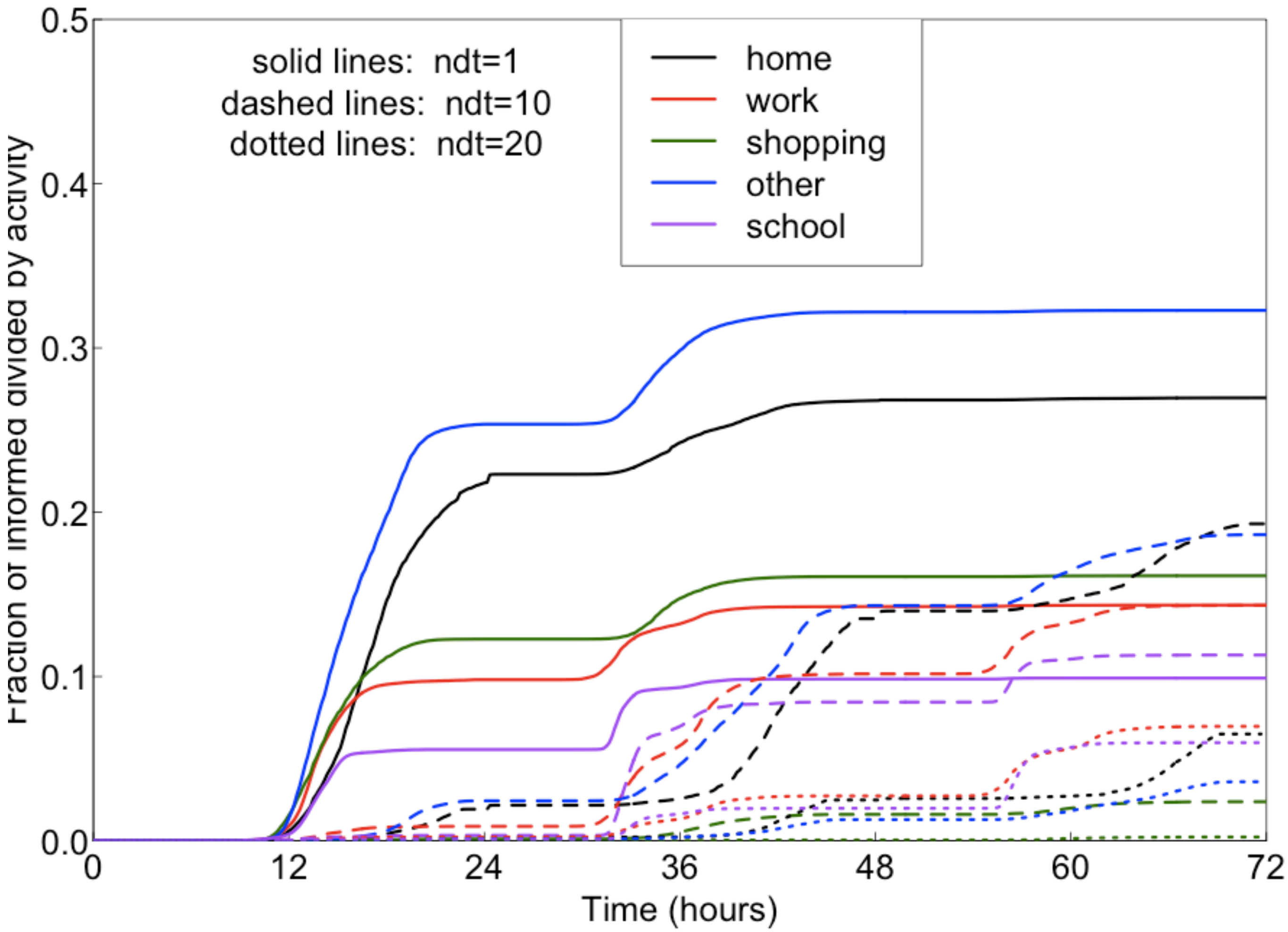






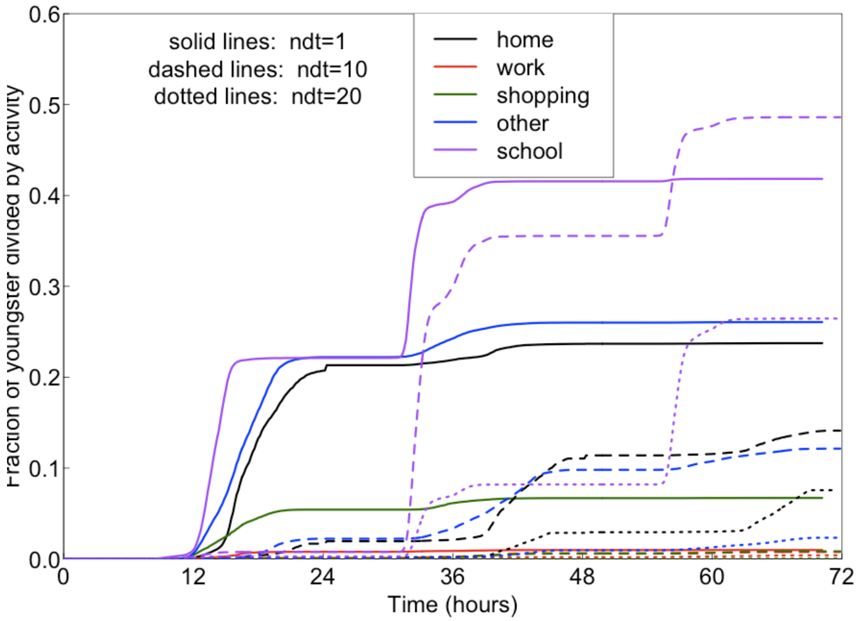
Communication is enhanced in the same age group.
Increasing *ndt*, shift and reduction of informed. Kids become more important
Kids listen to adults more than adults listen to kids

WHAT ARE THEY DOING?

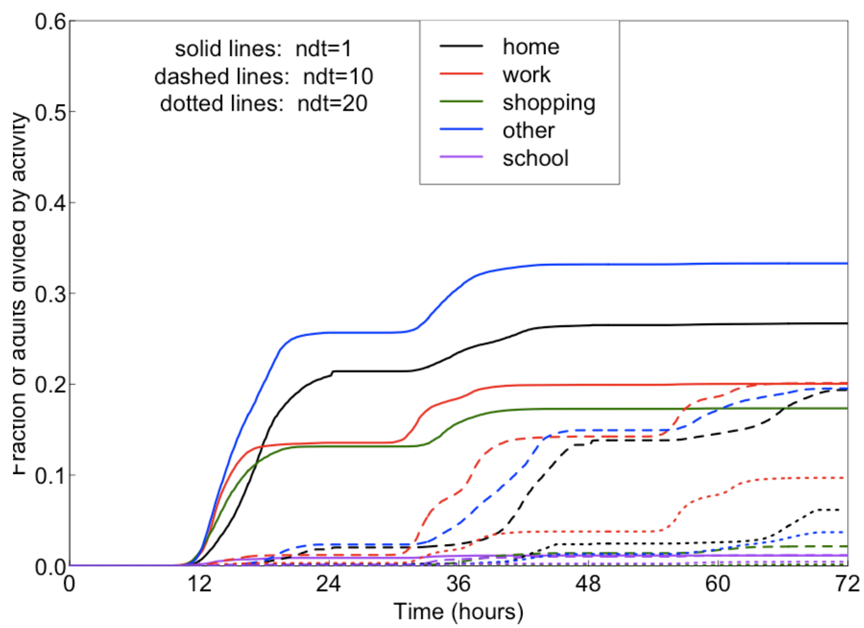


- For small *ndt* shopping recreational and transport main *foci* of communication: more individuals gather together
- Increasing *ndt* school work and home become important: this is due to length of contact

YOUNGSTER (0-18) AND ADULT(35-64)



School is the most important place for communication for children and increases the importance with *ndt*



At the beginning recreational activities and home are the places where communication happens more often. Increasing *ndt* work becomes predominant

Conclusions

- Tools for social scientist: synthetic populations and simulator engines
- Information flows depend on demographics and activity patterns
- Information flows depend on initial condition
- Link duration : prune the social network, delay and reduce diffusion (confirm Granovetter theory)
- Link duration : decrease communication, increase communication in-group.
- Link duration: change the importance of communication *foci*

Conclusions

- New tool for studying complex phenomena evolving in time. Theoretical analysis is really hard
- The model used is highly detailed and realistic
- Can be applied in different fields (epidemics, economics, power demand, social phenomena)
- Can be used for orienting policies but also theoretical analysis
- Requires large computing capabilities and data acquisition