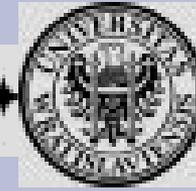Smoluchowski Insitute of Physics, Jagiellonian University (Cracow)
UNESCO Chair of Interdisciplinary Studies, Wroclaw University

Contact networks and the spread of MRSA in hospitals

Andrzej Jarynowski, **Frederik Liljeros**

12.06.2012 Wroclaw

Outline

- The MRSA outbreak in Stockholm 1999-2005

- Statistical Analyze

- Stationary Case

- Dynamic Case

- Artificial network and results

Outline

- The MRSA outbreak in Stockholm 1999-2005

- Statistical Analyze (classic-frequencial)

- Stationary Case

- Dynamic Case (Bayesian)

- Artificial network and results

Outline

- The MRSA outbreak in Stockholm 1999-2005

- Statistical Analyze (classic-frequencial)

- Stationary Case

  Ising model (physics)

- Dynamic Case (Bayesian)

  complex networks (physics)

- Artificial network and results

MRSA - Background and Social Network

The bacterium Meticillin Resistant *Staphylococcus Aureus* **(MRSA)** is resistant against more than half of all antibiotics and is known to alone be the largest care related the infection problem.

For such a infectious diseases, where a close contact is needed for a transmission to occur, the individual's position in the contact network is important for the person's risk to get infected. The awareness of the importance of contact network has brought methods from sociological studies of social networks into the area of preventive infectious disease protection work.
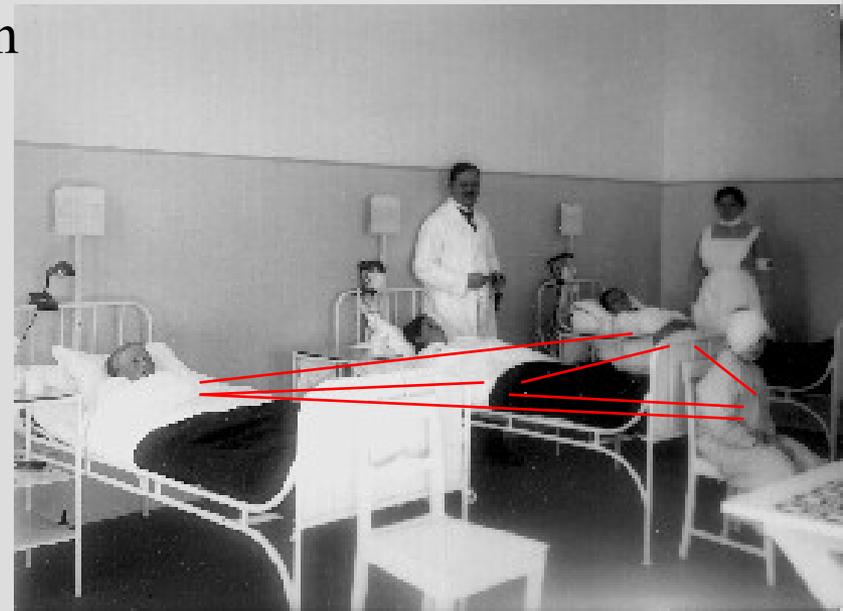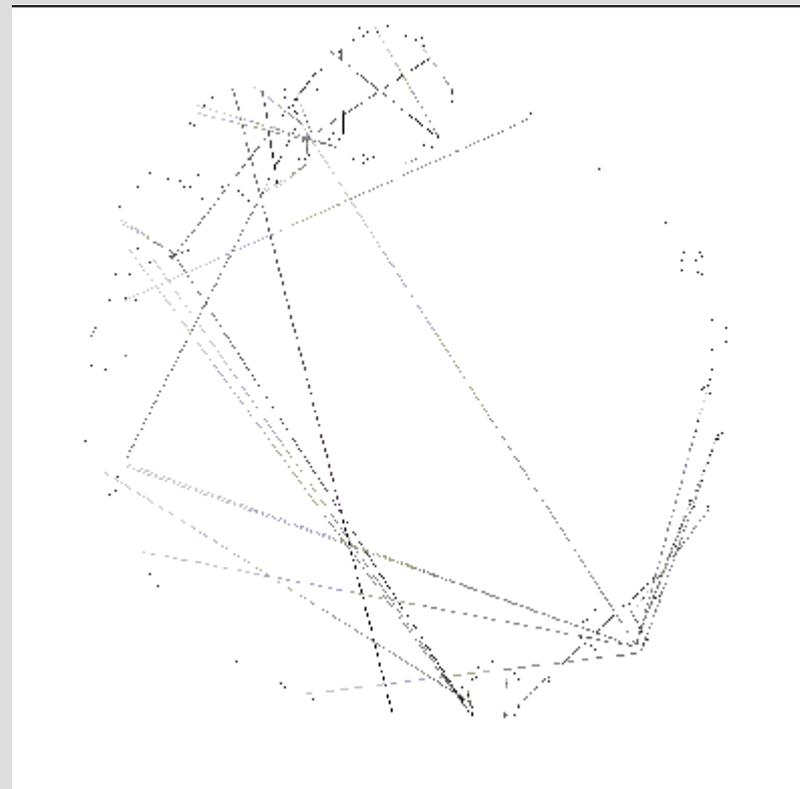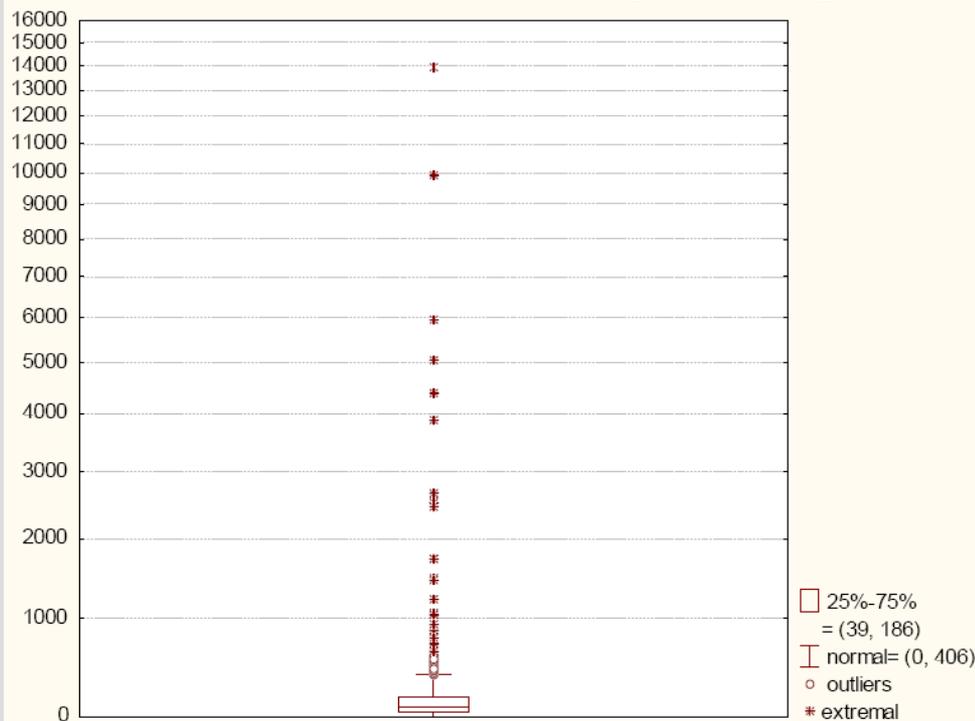
The MRSA outbreak in Stockholm 1999-2005

- 1337 Cases
- Population 2 314 517 Patients
- 210 different types of MRSA
- UK-E15 is the most frequent one
- The outbreak is now under control

  Dataset contains information about all in- and outpatient visits within Stockholm County during the period outbreak and a registry over diagnosed MRSA cases

MRSA – Some statistical properties

	binomial test values(ill-1 or health-0)			
Efect	Evaluation	standard error	Wald Stat.	p
free element	-4,01480	0,025059	25669,39	0,000000
days	-0,00027	0,000486	0,31	0,579326

□ 25%-75% = (39, 186)
⊥ normal= (0, 406)
○ outliers
* extremal

Box plot of number of contact ill persons with others. Most cases are between 39 and 186, but there are few outliers as well.

Logistic regression as a joining function with binomial depending variable (being infected or not). The null hypothesis was that the days spent with infected person have significant influence on being infected. P-Values is above 0,1 so there is no significant dependence between explaining and dependent variables.

MRSA – Method of realization

We study matrix of disease transition in hospitals population. This matrix **P** is our first goal. In rows are Infected and in Columns people, who could sent infection. Elements of matrix are probabilities, what Infection was sent by indicated person. Diagonal elements are probabilities of being infected by someone out of hospital, but they are in first approximation zeros. Unfortunately ¼ of all infected are patients, who had no contact with no other infected person.

B

t ol

A

Probabilities calculation is based on time of contact (time of sharing the same ward)

Matrix of probabilities (*P*)

	1 Inf by	2 Inf by	3 Inf by	4 Inf by		1 Inf by	2 Inf by	3 Inf by	4 Inf by
1	0	0	0,41	0	1	0	0	2,2	0
2	0	0	0	0	2	0	0	0	0
3	0	0	0	0	3	0	0	0	0
4	0,38	0	0,62	0	4	8,8	0	14,3	0
5	0	0	0	0	5	0	0	0	0
6	0	0	0	0	6	0	0	0	0
7	0	0	0	0	7	0	0	0	0
8	0	0	0	0	8	0	0	0	0
9	0	0	0	0	9	0	0	0	0
10	0	0	0	0	10	0	0	0	0
11	0	0	0	0	11	0	0	0	0
12	0	0	0	0	12	0	0	0	0
13	0	0	0	1	13	0	0	0	5,4
14	0	0	0	0	14	0	0	0	0

Part of contact's matrixes for type UK-E15. In rows are Infected (291) and in Columns people, who could sent infection (291 who also were infected). Elements of left matrix are probabilities $p_{A<-B}$ opposite to shifted overlapping times t_{Ol} on the right hand side.

MRSA – Improved matrix of p

A-individual infected
B-individual infected
$p_{B<-A} < p_{A<-B}$ if B was detected earlier

Stationary case...what next?

At this point we should discuss limitation of data. Advantage of closed community come into disadvantage because of physician's interventions. Hospital staff reacting on diagnosed MRSA case by quarantine and other medical tools. Clinical trials of testing on MRSA are not systematically provide on randomized cohorts, but *in vivo*, based on medical intuition. Date of positive test on having MRSA gives bias if we want to correspond it to date of disease transmission. Isolation and decolonisation action also take place.

Matrix Q corresponds to second order neighborhood on the level of clinic.
Matrix P' corresponds to contacts with all patients.

MRSA – Dynamic case
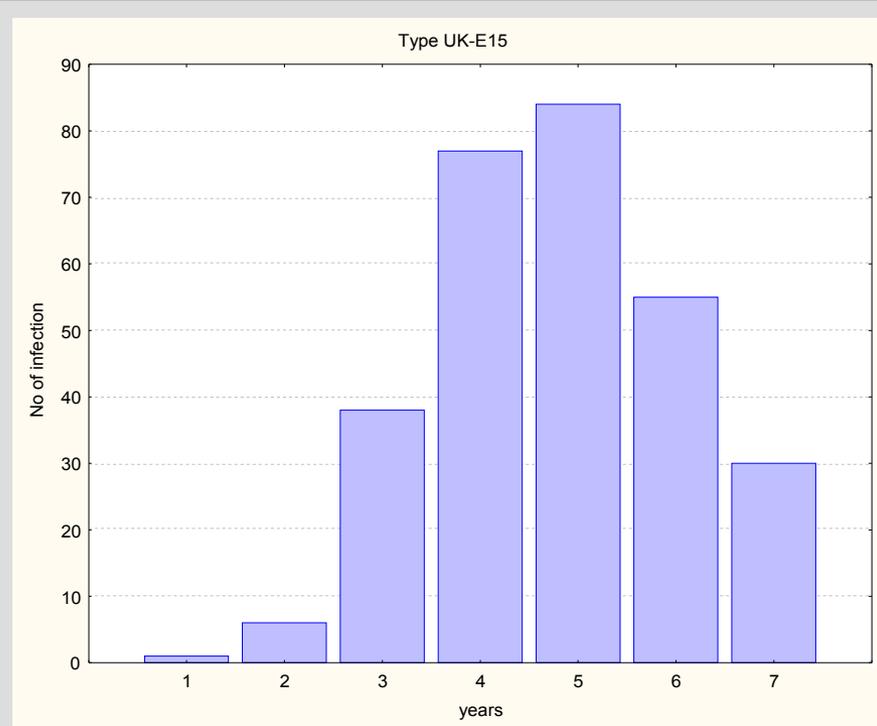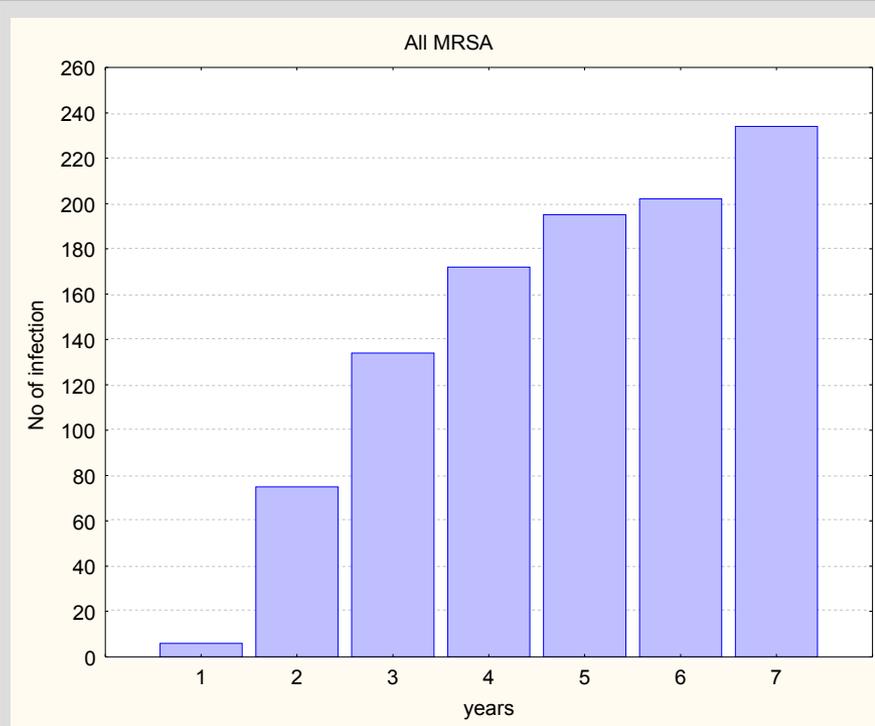
There is a difference between shape of histograms (numbers of infections in time) for all MRSA cases (on left) and for type UK-E15 (on right) which looks more like epidemic curve.

MRSA – Dynamic case

Let build matrixes of contacts in smaller times intervals (*dt*). Let starts from time t_0 (1999) and matrix P_0. That can divided The Stockholm Outbreak 1999-2005 into smaller periods, but last P will show situation during period (2005-*dt*). From that data we will also need matrix matrixes P' which will not be normalized and tell us some characteristics of numbers of contacts for patients in that time intervals (with contacts with all patients).

The best *dt* is whole year because of periodicity of visits to hospitals. That mean, that there will be 7 matrixes (P_t, Q_t and P'_t) in which we are considering contact networks in that time interval.

Histogram of entries
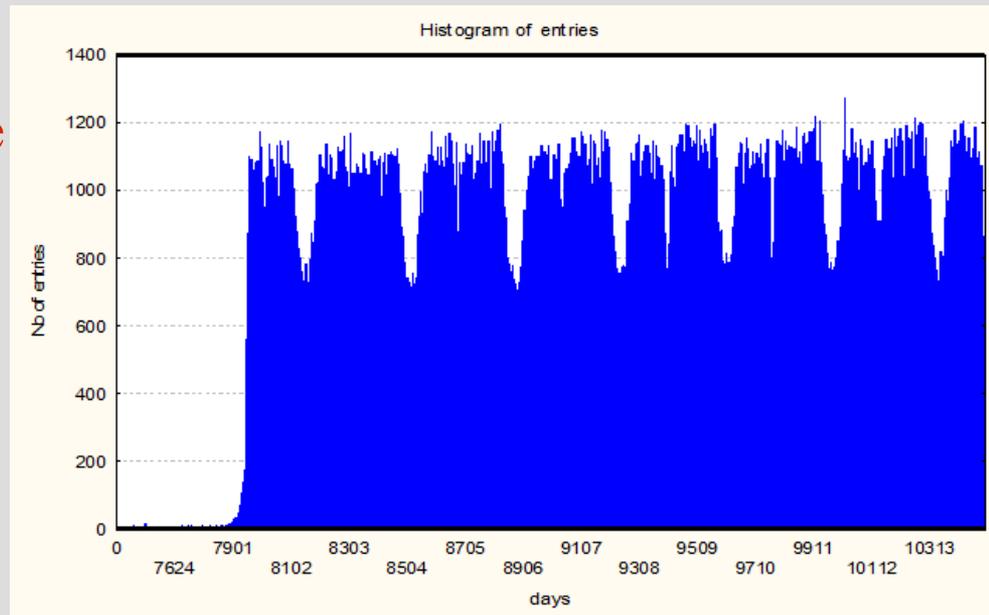
q-MCMC simulation

We can look at how MRSA is spreading in time and tried to simulate similar scenarios using q-MCMC (quasi- Markov Chain Monte Carlo) class of algorithms, which are based constructing a Markov chain that has the desired distribution as its equilibrium distribution.

- We assume that vector of our population (0-health, 1-ill) evolve in time.
- We want to find mechanism of change.

0 0 ... 0 1 ... 0

0 1 ... 0 1 ... 0 change after time interval

- All individuals can have specific p_t (probability transition). We can estimate that probabilities using historical data and try to run MCMC to predict future states.

q-MCMC simulation

To get the most likely probabilities let use Metropolis Algorithm. One (tested positive) has more influence on the sending infection, whereas the second one (not tested) is also included influence. We studied, via computer simulations (like Ising model with Metropolis Monte Carlo algorithm), the interplay between states in depends on such factors as: fraction of contacted infected, and the possibility of contact between all patient. That model has it roots in the domain of magnetism, but the meaning of magnetic spins has been changed into health states (0-health or ill-1).

Linear reaction on shifted t_{ol} values combine with heavy-tailed distribution of contacts seems to be to strong so square root suits better to describe phenomenon.

Metropolis Monte Carlo algorithm

In a first step we try to start with a vector of 291 patient, who have UK-E15. At beginning only one patient was ill (was tested positive during 1 year). So first state has 290-zeros and only one-1.There are two reasons for this initial condition:

1) this it typical way of thinking about epidemiological models in Sweden. All first cases of disease come from abroad. For example Swed had accident in Poland and was infected in Polish hospital. After moving to Sweden and being treated in home hospitals he/she infected some other patients and started outbreak.

2) it was in data – on first year only one patient was tested positive.

Our goal is to get at the end of simulation also similar number of infected (this is condition for calibrating parameters of model) so this model can be understand as SI because MRSA is usually incurable.

Metropolis Monte Carlo algorithm

We have matrixes $\boldsymbol{P_t}$, $\boldsymbol{Q_t}$ and $\boldsymbol{P'_t}$ for each year but there are not normalized. Formula for individual change state in one time step can be written as:

$$p_i(t) = 1 - \exp\left(-\sum_j \sqrt{\boldsymbol{P_t}(i,j)} \cdot s - \sum_j \sqrt{\boldsymbol{Q_t}(i,j)} \cdot k - \sum_{j'} \sqrt{\boldsymbol{P'_t}(i,j')} \cdot m\right)$$

Using characteristics of exponent we can rewrite formula above in multiplicative form, because our goal is to know who could infected patient i.

$$p_i(t) = 1 - \Pi_j(P_{i<-j}(t)) \cdot \Pi_j(Q_{i<-j}(t)) \cdot P'_{i<-indirect}(t)$$

Where:

$P_{i<-j}(t)$ represents that j send infection to i in year t on ward connections ($Q_{i<-j}(t)$ equivalent on clinic connections)

$P'_{i<-indirect}(t)$ is the representation of being infected indirectly in year t.

We need 3 parameters:

Metropolis Monte Carlo algorithm

We need 3 parameters:

- s- normalization on influence of contacts with infected (there is important to treat all of possible senders individually to find path of transition)

- k- normalization on influence of second order contacts with infected (there is also important to treat all of possible senders individually)

- m- normalization on influence of contacts with all patients (there is no need to treat it individually so the best measure is person · day)

The third parameter m- as we showed before can be calculate to reproduce ¼ of all infections which cannot be explain at all by a social network of infected. Both parameters k and s come from spatial relation so why do not combine them.

Metropolis Monte Carlo algorithm

Proposition of combining *s* and *k* is to have the same infectivity level both for ward and clinic neighborhood and it can be done by transformation: $k=s \cdot N_W/N_C$, where N_W is number connections on level of ward and N_C is number connections on level of clinic during whole outbreak. The *a priori* selected configurations of parameters bring at least only one parameter to fit: *s*.

One parameter left: let allow complex system work!

Metropolis Monte Carlo algorithm

Proposition of combining *s* and *k* is to have the same infectivity level both for ward and clinic neighborhood and it can be done by transformation: $k=s \cdot N_W/N_C$, where N_W is number connections on level of ward and N_C is number connections on level of clinic during whole outbreak. The *a priori* selected configurations of parameters bring at least only one parameter to fit: *s*.

One parameter left: let allow complex system work!

* Parameter free model - Great!

Metropolis Monte Carlo algorithm

Proposition of combining *s* and *k* is to have the same infectivity level both for ward and clinic neighborhood and it can be done by transformation: $k=s \cdot N_W/N_C$, where N_W is number connections on level of ward and N_C is number connections on level of clinic during whole outbreak. The *a priori* selected configurations of parameters bring at least only one parameter to fit: *s*.

One parameter left: let allow complex system work!

* Parameter free model - Great!

* One parameter - It can be

Metropolis Monte Carlo algorithm

Proposition of combining *s* and *k* is to have the same infectivity level both for ward and clinic neighborhood and it can be done by transformation: $k=s \cdot N_W/N_C$, where N_W is number connections on level of ward and N_C is number connections on level of clinic during whole outbreak. The *a priori* selected configurations of parameters bring at least only one parameter to fit: *s*.

One parameter left: let allow complex system work!

* Parameter free model - Great!

* One parameter - It can be

* 2 or more – It's like infinity...no sense to analyze

Artificial *vs* Real network

The simplest method to stay with the same statistics as original network is permutation of matrixes P_t, Q_t. Ones it is done, new network is ready.
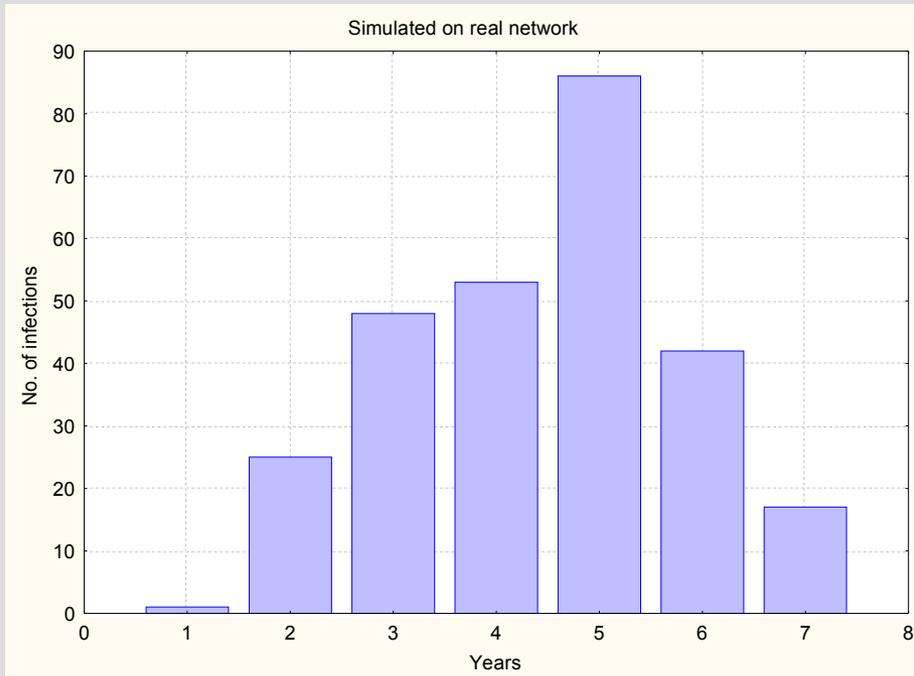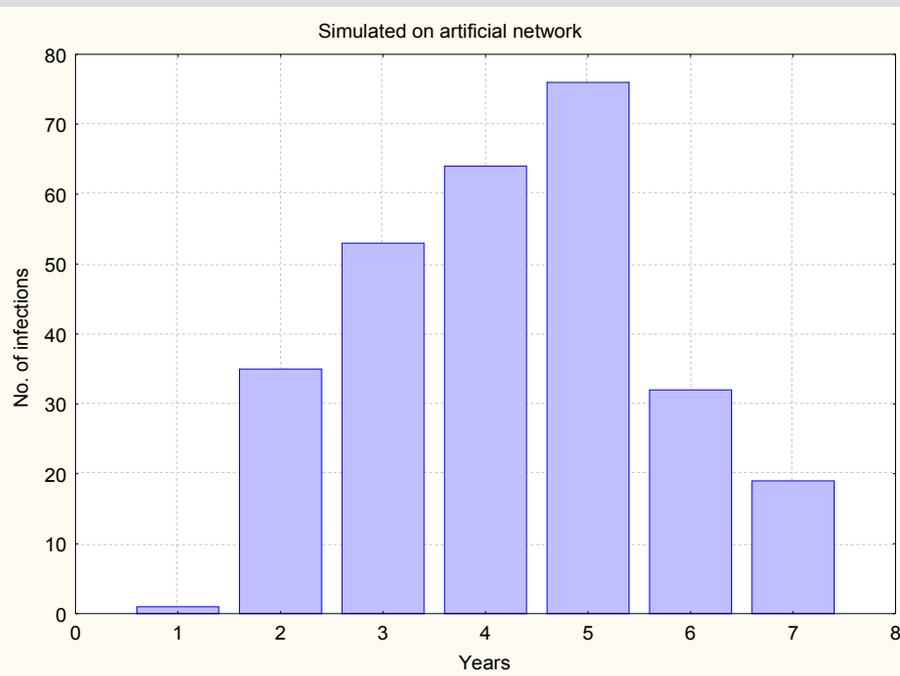
When parameter *s* is fixed after calibration, artificial oubreak is constructed.

Let repeat observed outbreak:
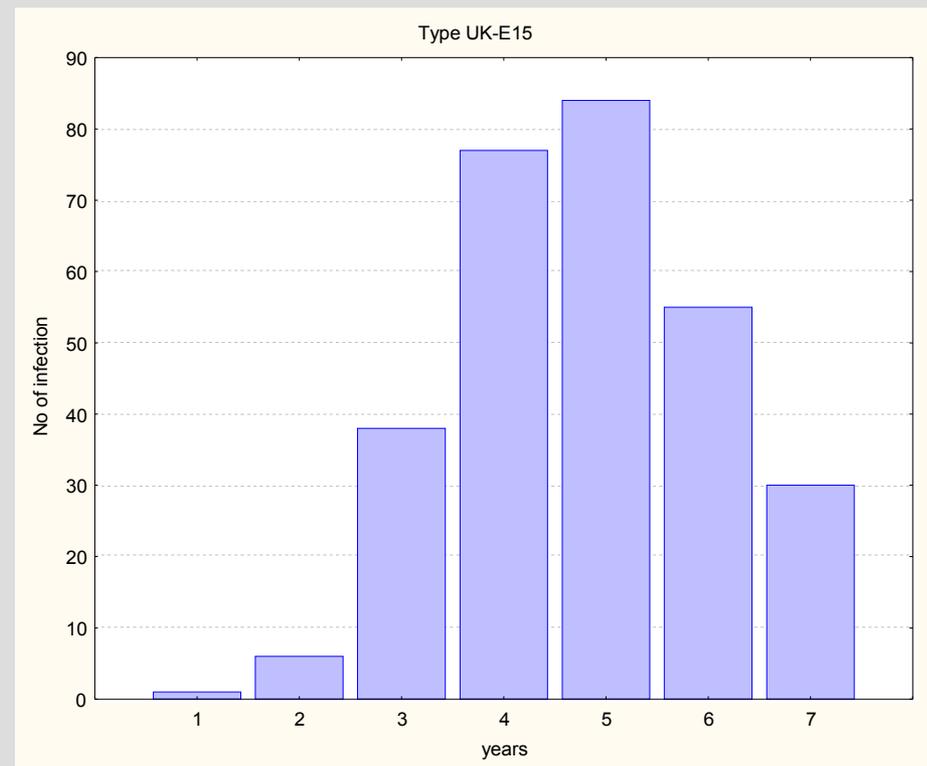
Results

Selected realizations: histograms (numbers of infections) for simulation on artificial (left) and real (right) network. Compare with real situation.

Results

	Error for cumulative No. of infections					
	1 Year	2 Year	3 Year	4 Year	5 Year	6 Year
artificial	408%	159%	55%	18%	4%	5%
real	382%	71%	7%	4%	6%	9%

Errors after each year of simulation (real number of infections is the pattern)

First evaluation partly failed (error around 400% after first year and 70% after second for real network and even worse for artificial). Let consider reaction of physicians who are watching start of outbreak – they are testing bigger number of potential infected. This fact explain that number of infections in first years is underestimated in registered data. Moreover the highest pick can contain patients who were infected earlier but tested positive during period of massive testing.

Summary

One step to get most likely paths of transmission
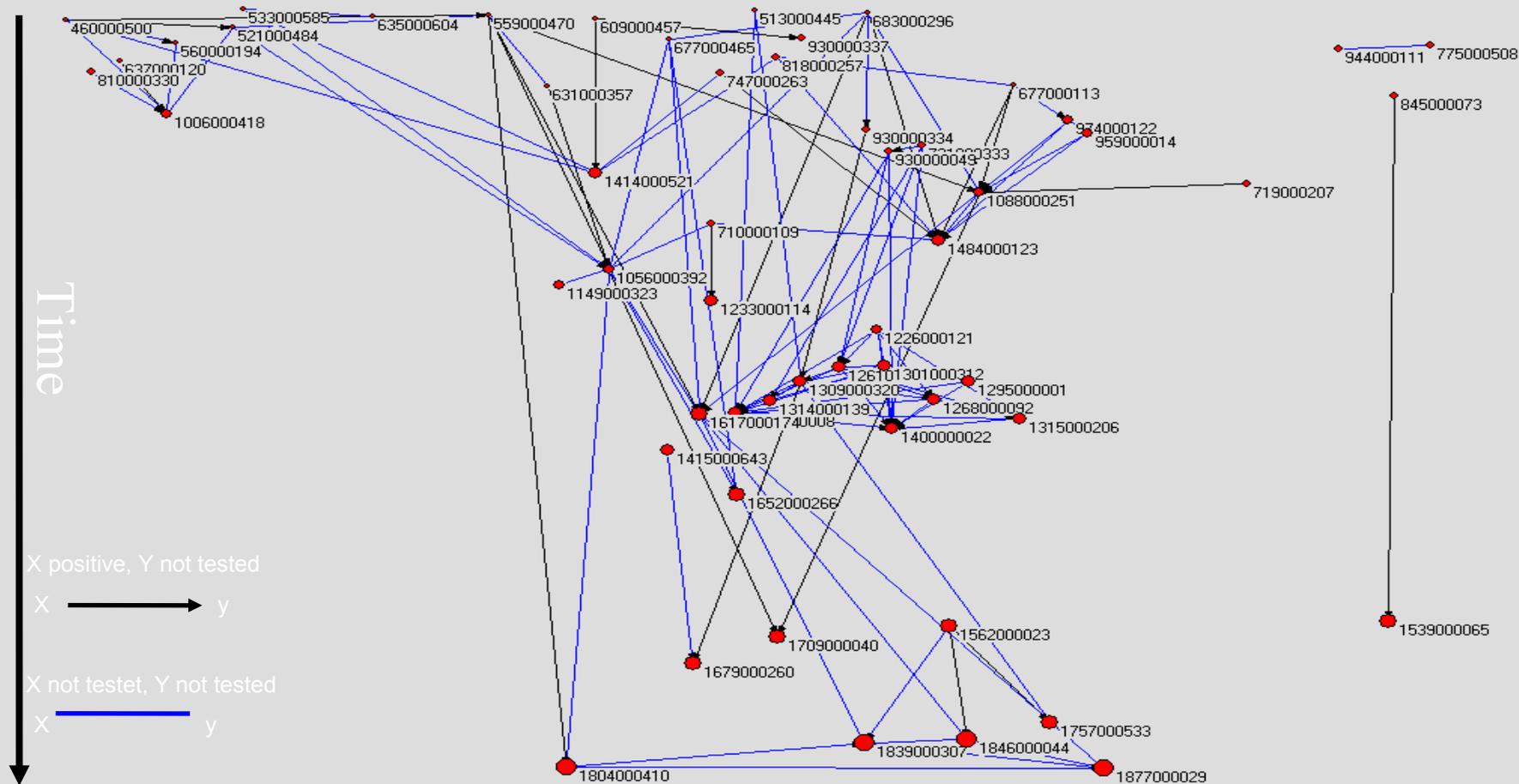
Limits of model

- Is the best path (in meaning of Hamming distance) most likely one?

- **Maximal length of tree (6)**

- **Validation of model on controlled outbreak**

- **Problems of operational research**

Acknowledgent

- Ph.D. **Lisa Brouwers** (Swedish Institute for Infectious Disease Control) for inviting me to DSV department in Stockholm;

- Ph.D. **Frederik Lilieros** and M.Sc. **Lu Xin** (Department of Sociology, University of Stockholm) for releasing unique dataset of MRSA patient and for giving sociological advises in coupling with that data.

Acknowledgent

Thank You for Your attention!