LIFE DATA EPIDEMIOLOGY

lect. 1: Introduction to computational epidemiology
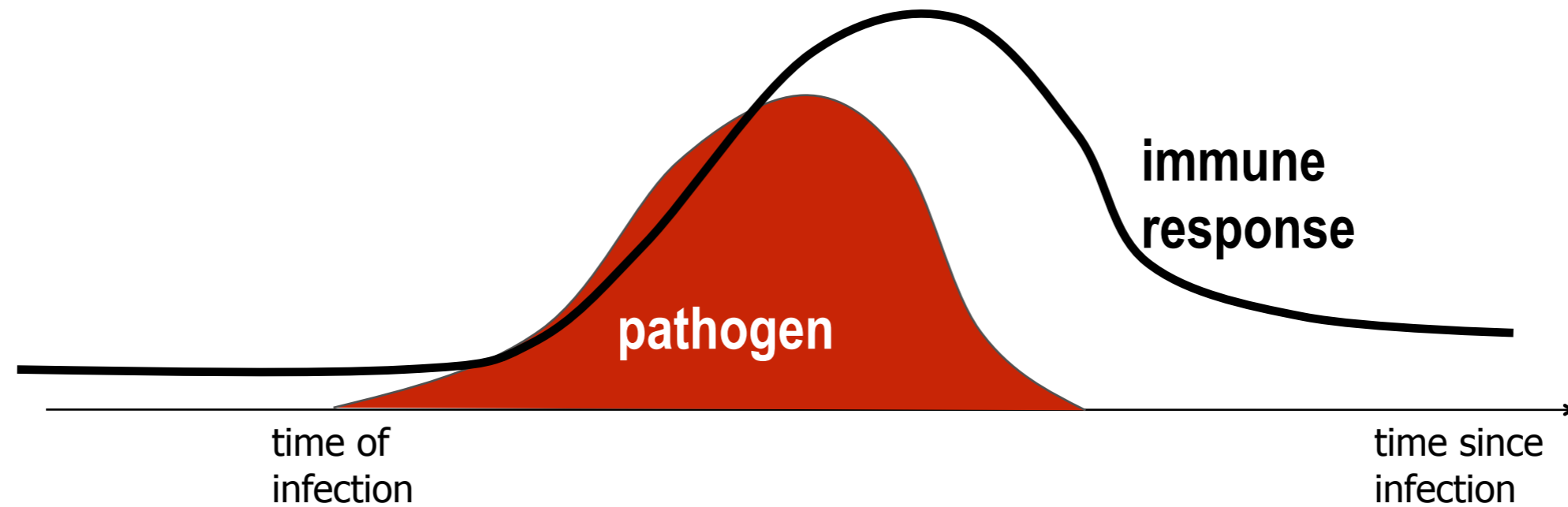
Chiara Poletto

polettoc@gmail.com

Life data epidemiology

- What is computational epidemiology?

- modelling epidemics in space - the meta population models

- Introduction to temporal network epidemiology

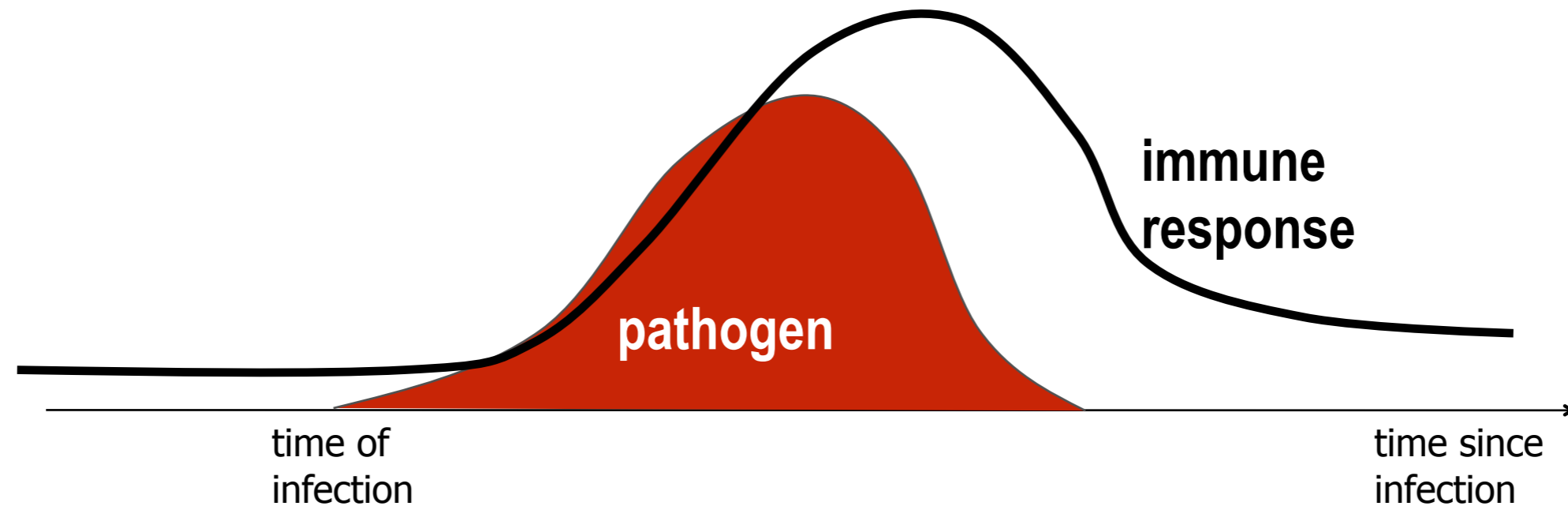- Confronting models with data introduction to model fitting

Epidemiology is interdisciplinary

- **Public health**: because of the emphasis on disease prevention

- **Clinical medicine**: because of the emphasis on disease classification and diagnosis (numerators)

- **Pathophysiology**: because of the need to understand basic biological mechanisms in disease (natural history)

- **Social sciences**: because of the need to understand the social context in which disease occurs and presents (social determinants of health phenomena)

- **Statistics** *[+ Mathematics + Physics + …]*: because of the need to quantify disease frequency and its relationships to antecedents
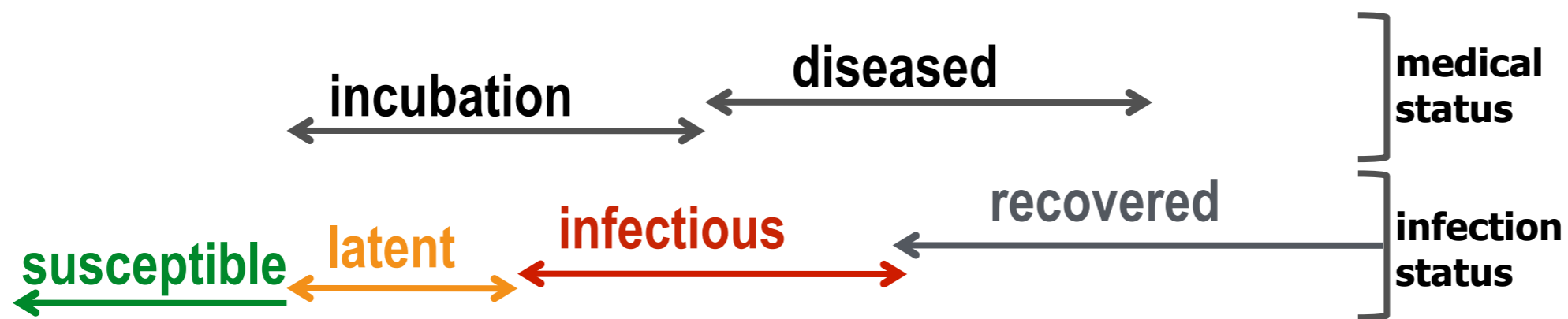
epidemiological vs. clinician point of view

[Keeling & Rohani, Modeling Infectious Diseases (2008)]

epidemiological vs. clinician point of view

pathogen

immune response

time of infection

time since infection

Coarse graining of the dynamics:

incubation

diseased

medical status

recovered

infection status

susceptible   latent   infectious

[Keeling & Rohani, Modeling Infectious Diseases (2008)]

Kermack & McKendrick model
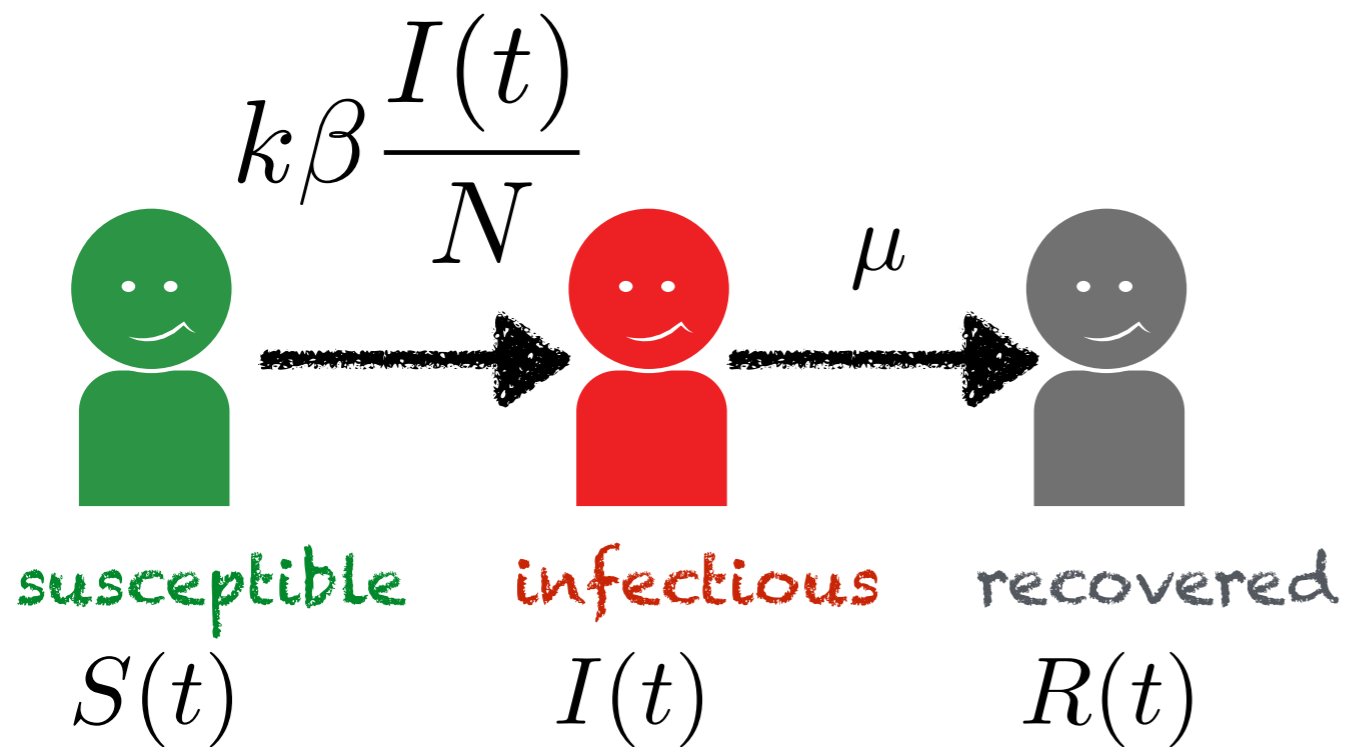
$$d_t S(t) = -k\beta \frac{I(t)}{N} S(t)$$

$$d_t I(t) = k\beta \frac{I(t)}{N} S(t) - \mu I(t)$$
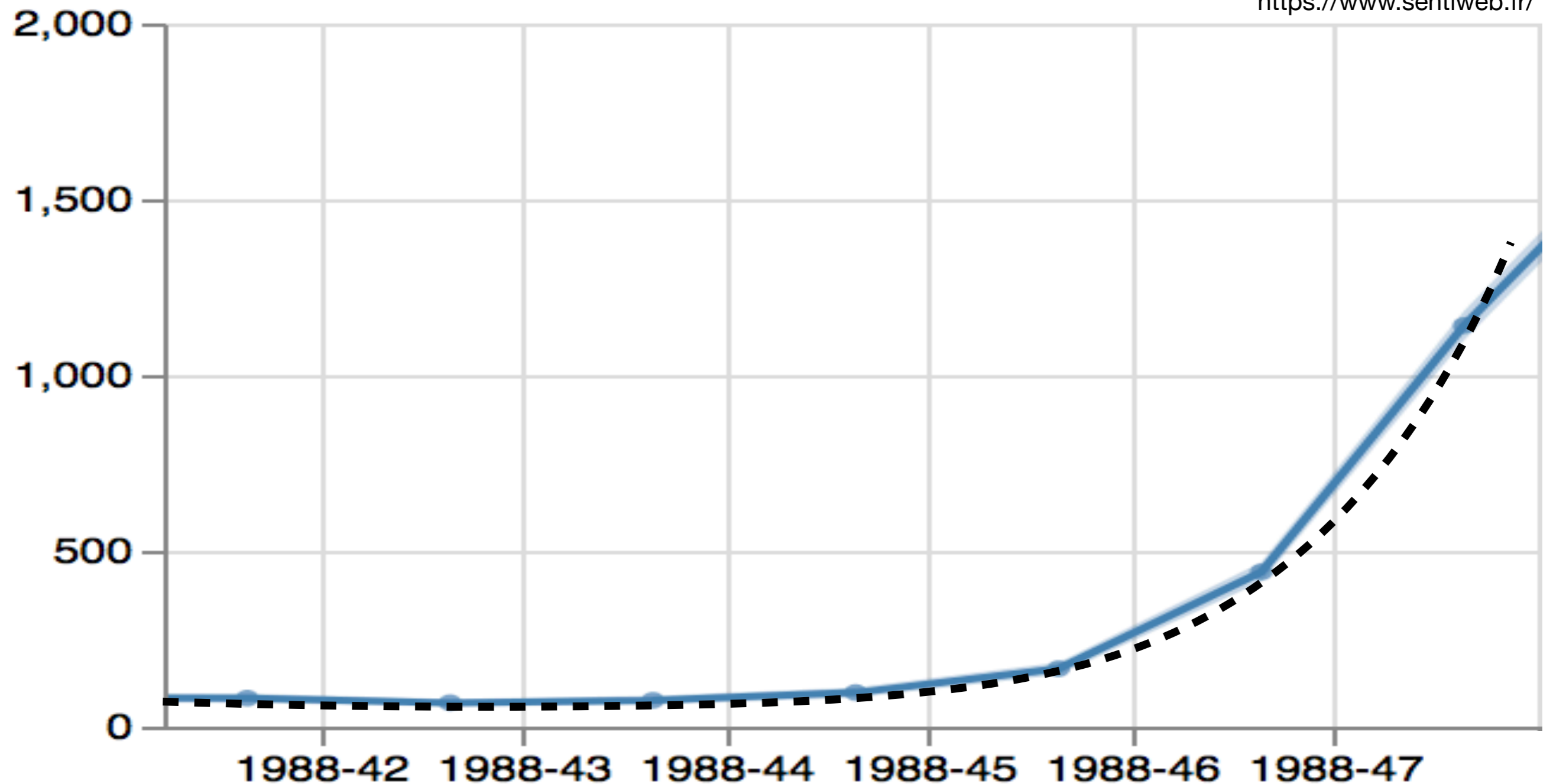
$$d_t R(t) = \mu I(t)$$

$$S(t) + I(t) + R(t) = N$$

[Kermack & McKendrick Proc Roy Soc A 1927, Keeling & Rohani, Modeling Infectious Diseases (2008)]

Kermack & McKendrick model

epidemiology = making sense of data
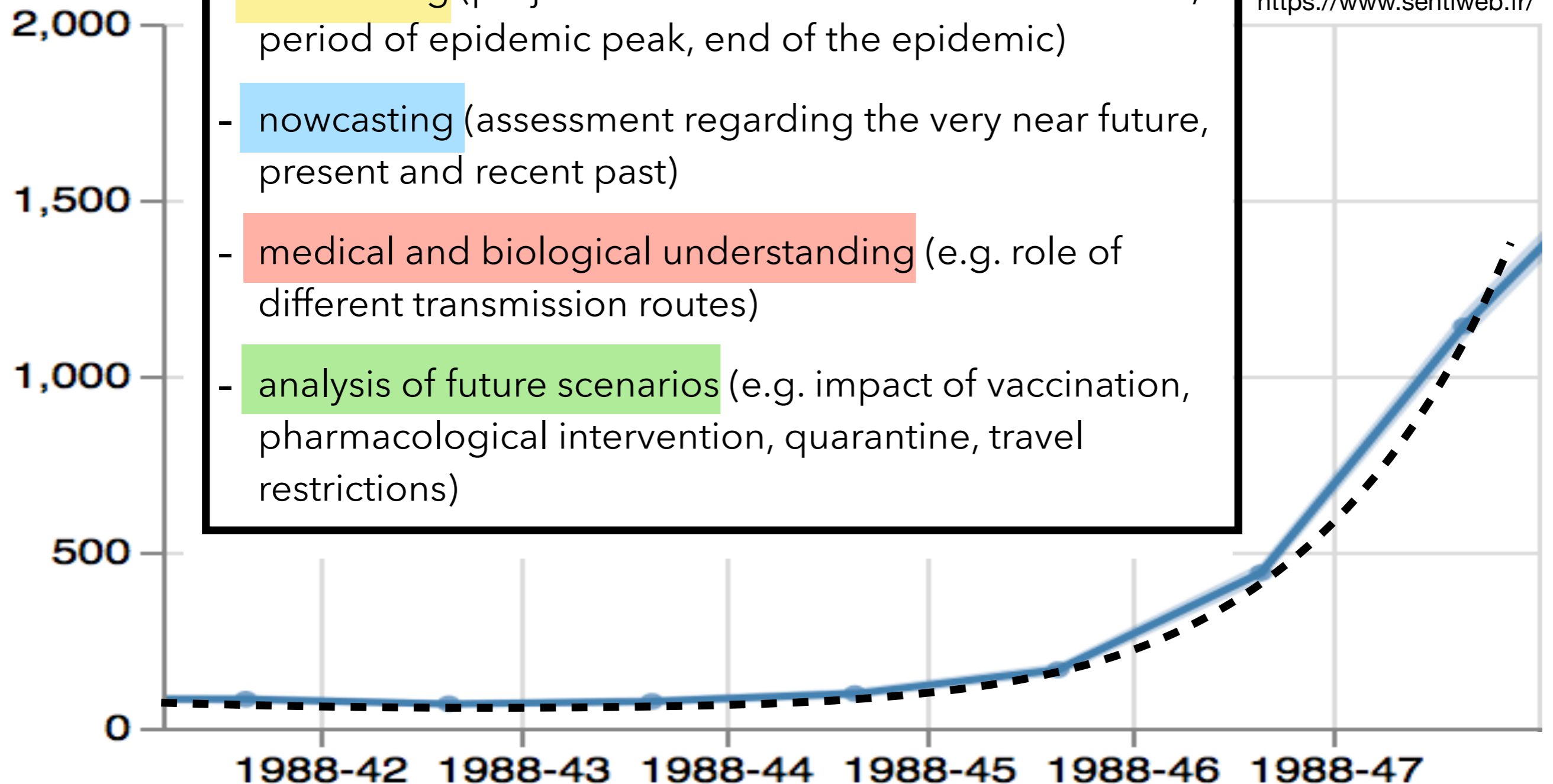
Kermack & McKendrick model

epidemiology = making sense of data

- forecasting (projections on the future number of cases, period of epidemic peak, end of the epidemic)

- nowcasting (assessment regarding the very near future, present and recent past)

- medical and biological understanding (e.g. role of different transmission routes)

- analysis of future scenarios (e.g. impact of vaccination, pharmacological intervention, quarantine, travel restrictions)

Sentinelles
Réseau Sentinelles
https://www.sentiweb.fr/

2,000

1,500

1,000

500

0

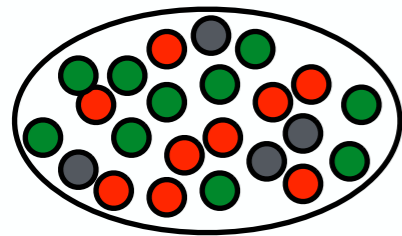1988-42   1988-43   1988-44   1988-45   1988-46   1988-47

pathogen polymorphism, evolution, interaction with other pathogens

COMPLEXITY

Host immunological & genetical profile, microbiome

COMPLEXITY

homogeneous mixing compartmental models
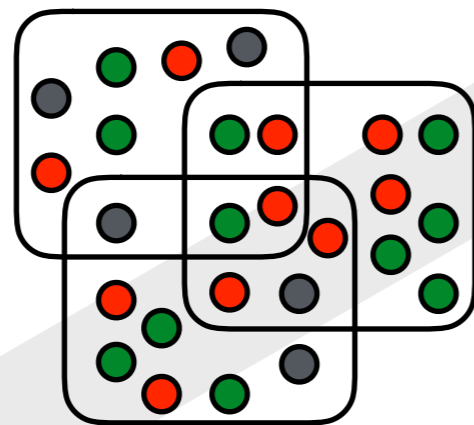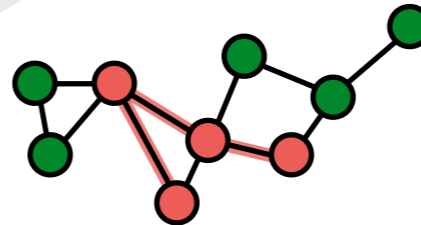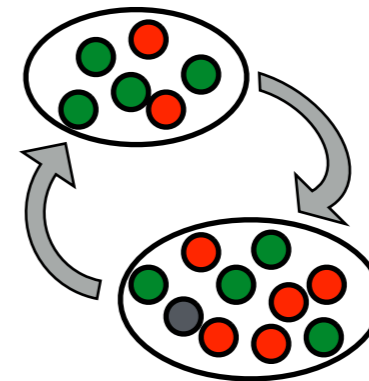
COMPLEXITY

Host demography - behaviour - contacts - spatial structure

pathogen polymorphism, evolution, interaction with other pathogens

COMPLEXITY

multi strain, evolution

Host immunological & genetical profile, microbiome

COMPLEXITY

population structure

network

metapopulation

agent based

hospital

school

home

work

homogeneous mixing compartmental models
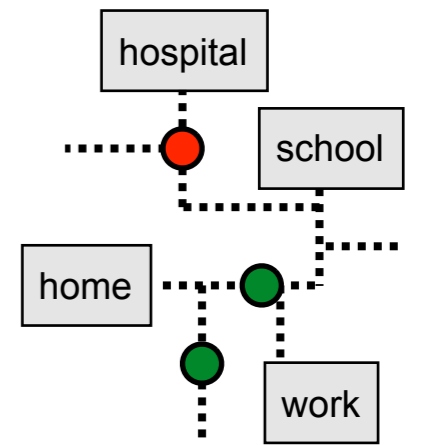
COMPLEXITY

Host demography - behaviour - contacts - spatial structure

COMPLEXITY
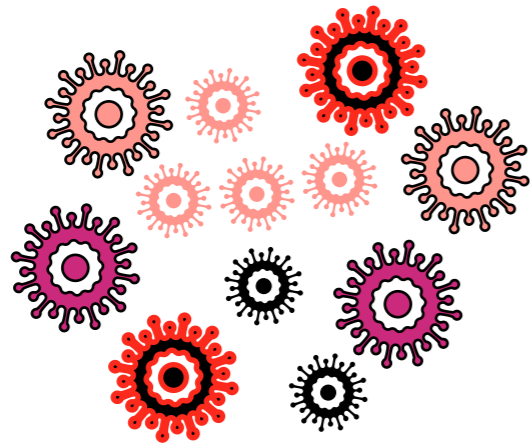
multi strain, evolution

pathogen sequences
phenotypic information
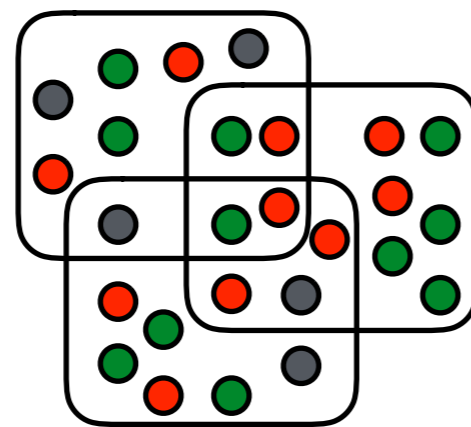
population structure
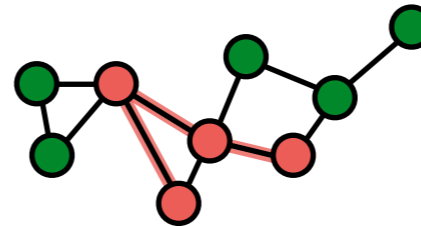
network
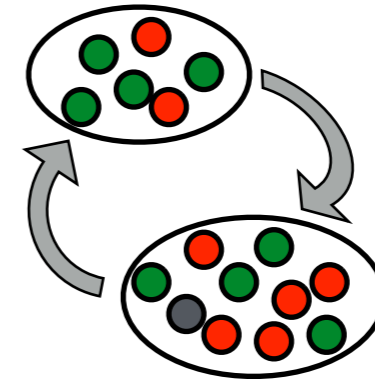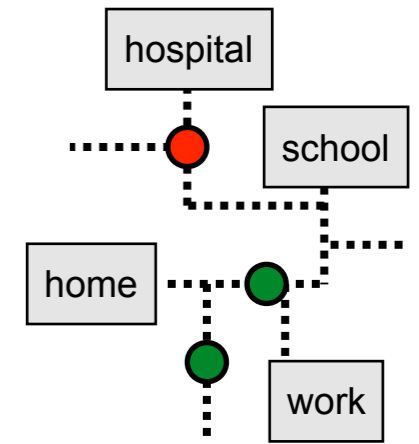
metapopulation

agent based

age profile, contact statistics risk classes

individual level contact dynamics

spatial distribution, mobility

detail statistics of the population (households, workplaces, schools)

population structure

network

metapopulation

agent based

hospital

school

home

work

EPIDEMIC MODEL

MOBILITY DATA
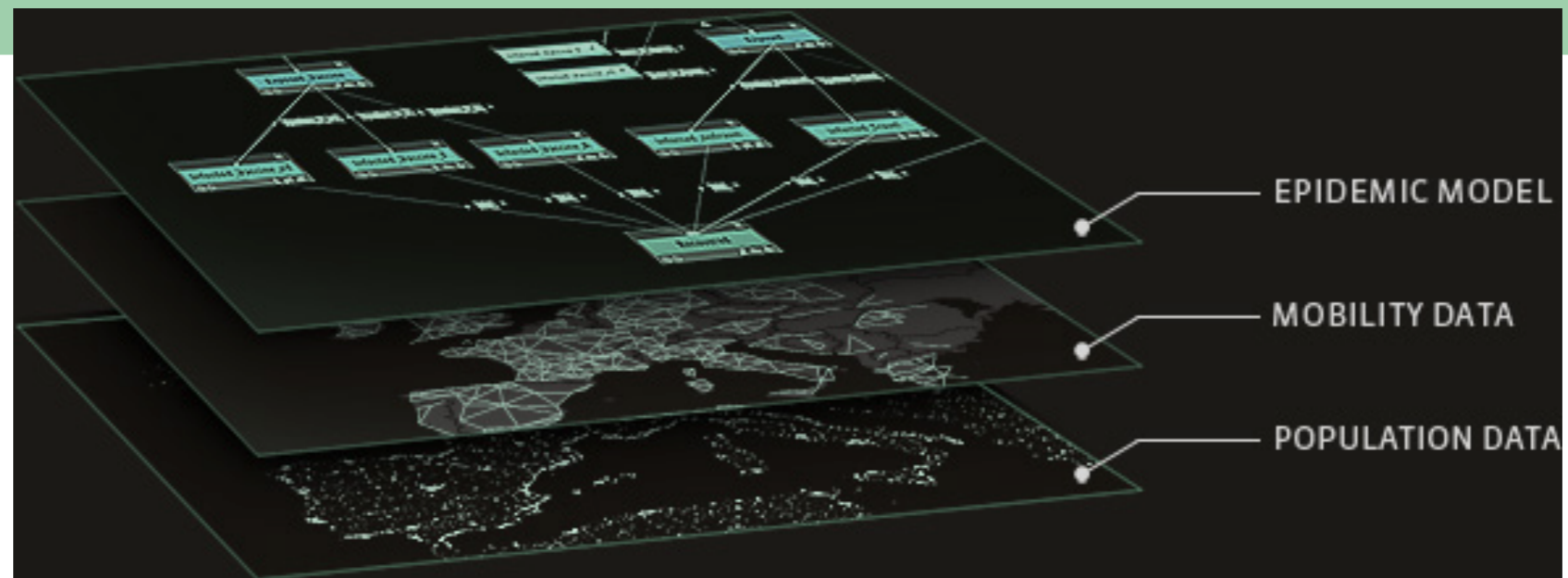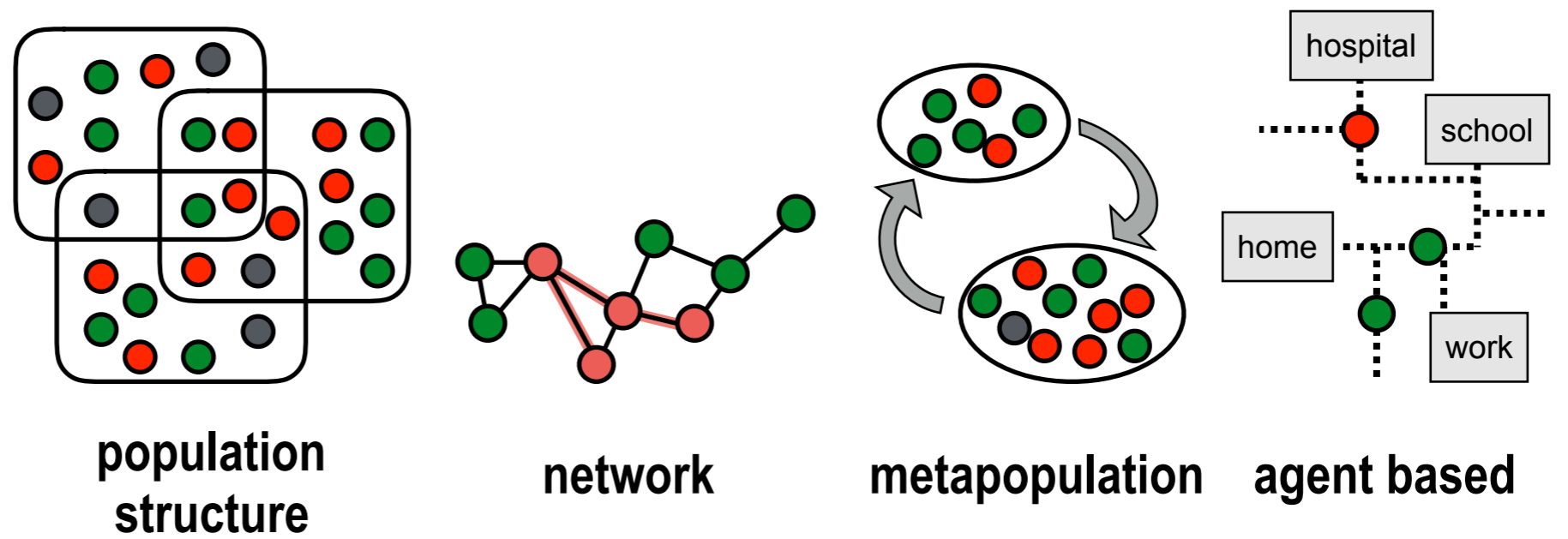
POPULATION DATA

big data!

epidemiological data - surveillance

incidence: flu

1988-1989 Influenza season in France

https://www.sentiweb.fr/
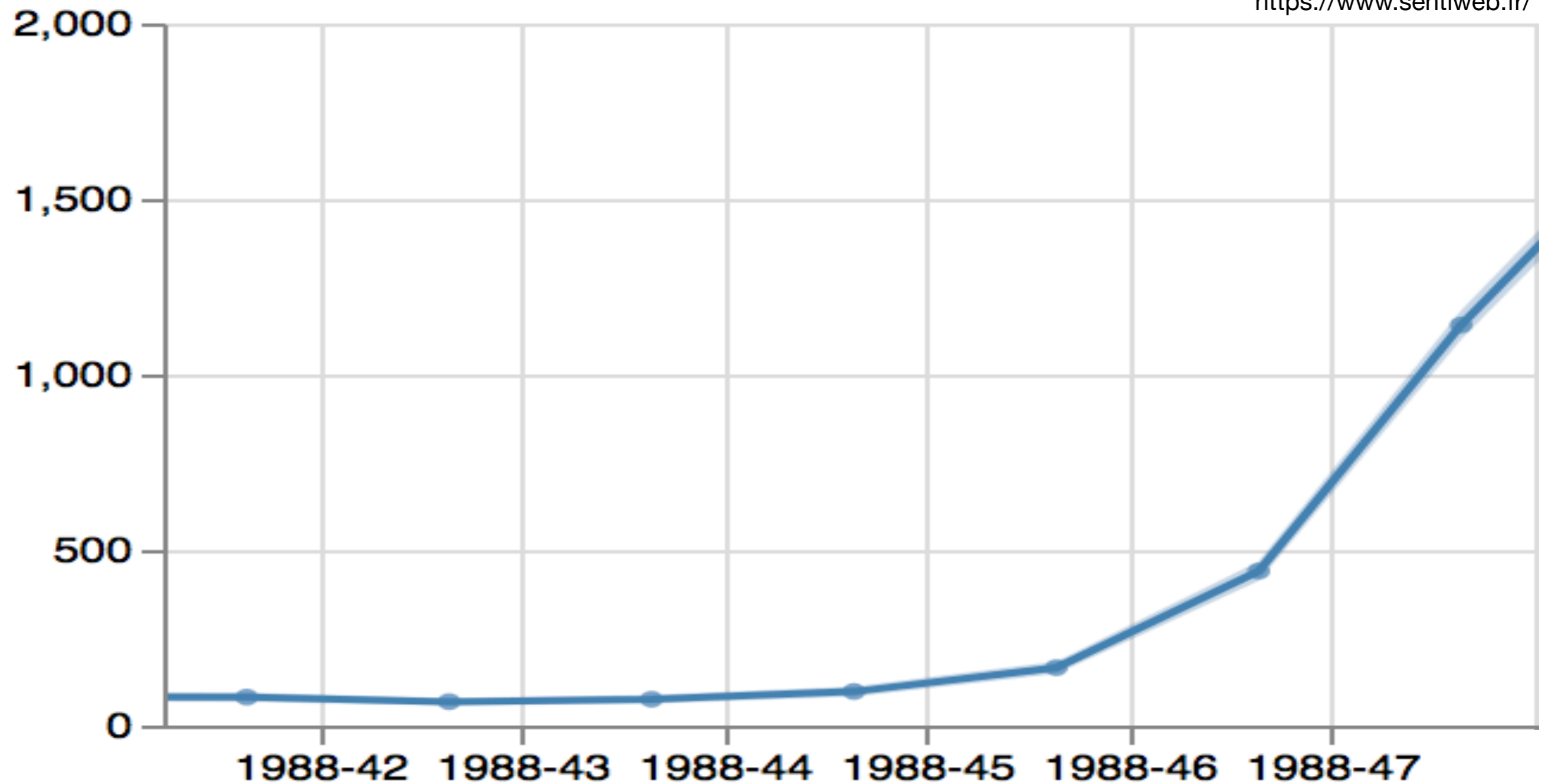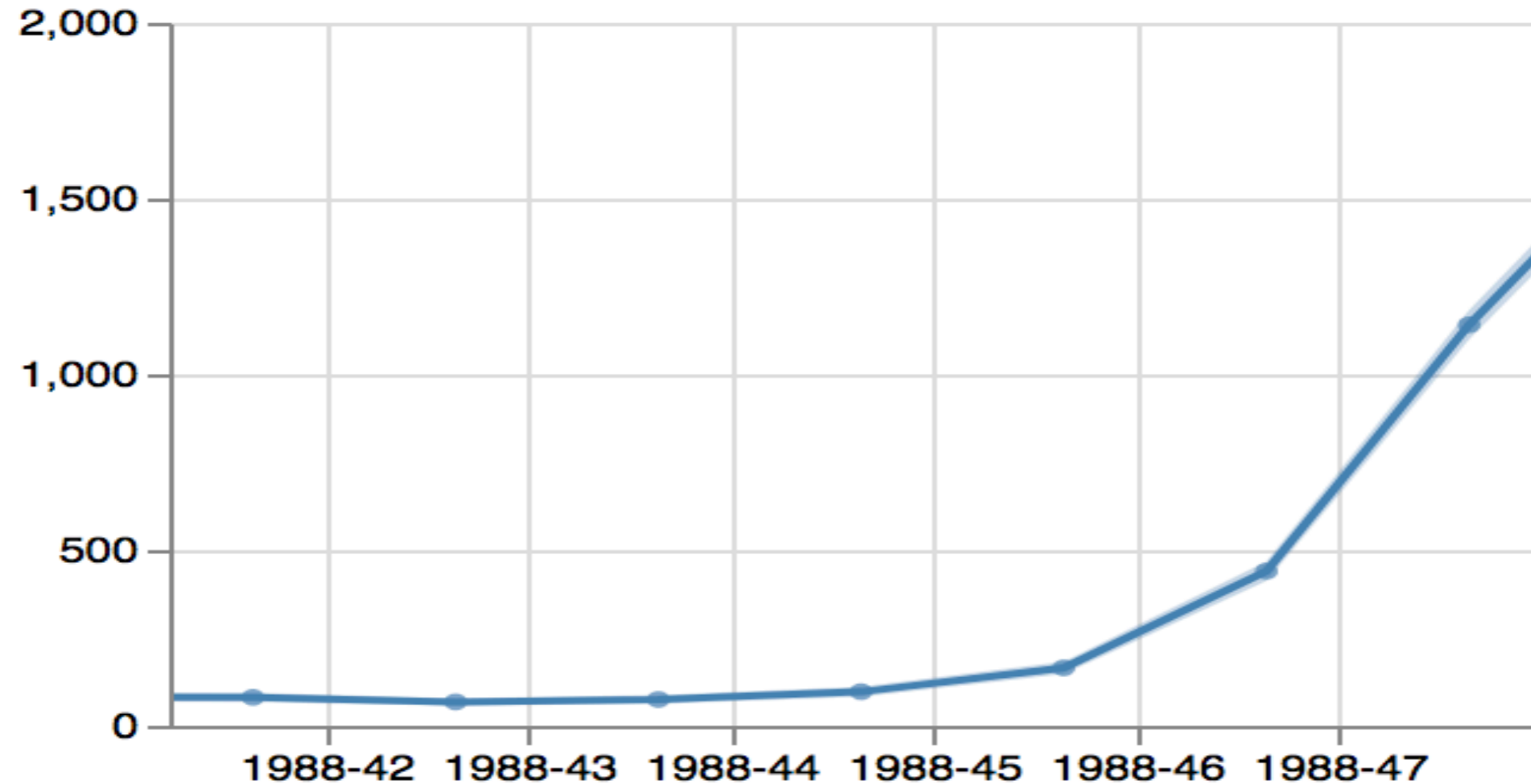
Incidence in a given area *a* at time *t* = fraction of the population catching flu in *a* at *t*

incidence: flu

1988-1989 Influenza season in France

Sentinelles
Réseau Sentinelles
https://www.sentiweb.fr/
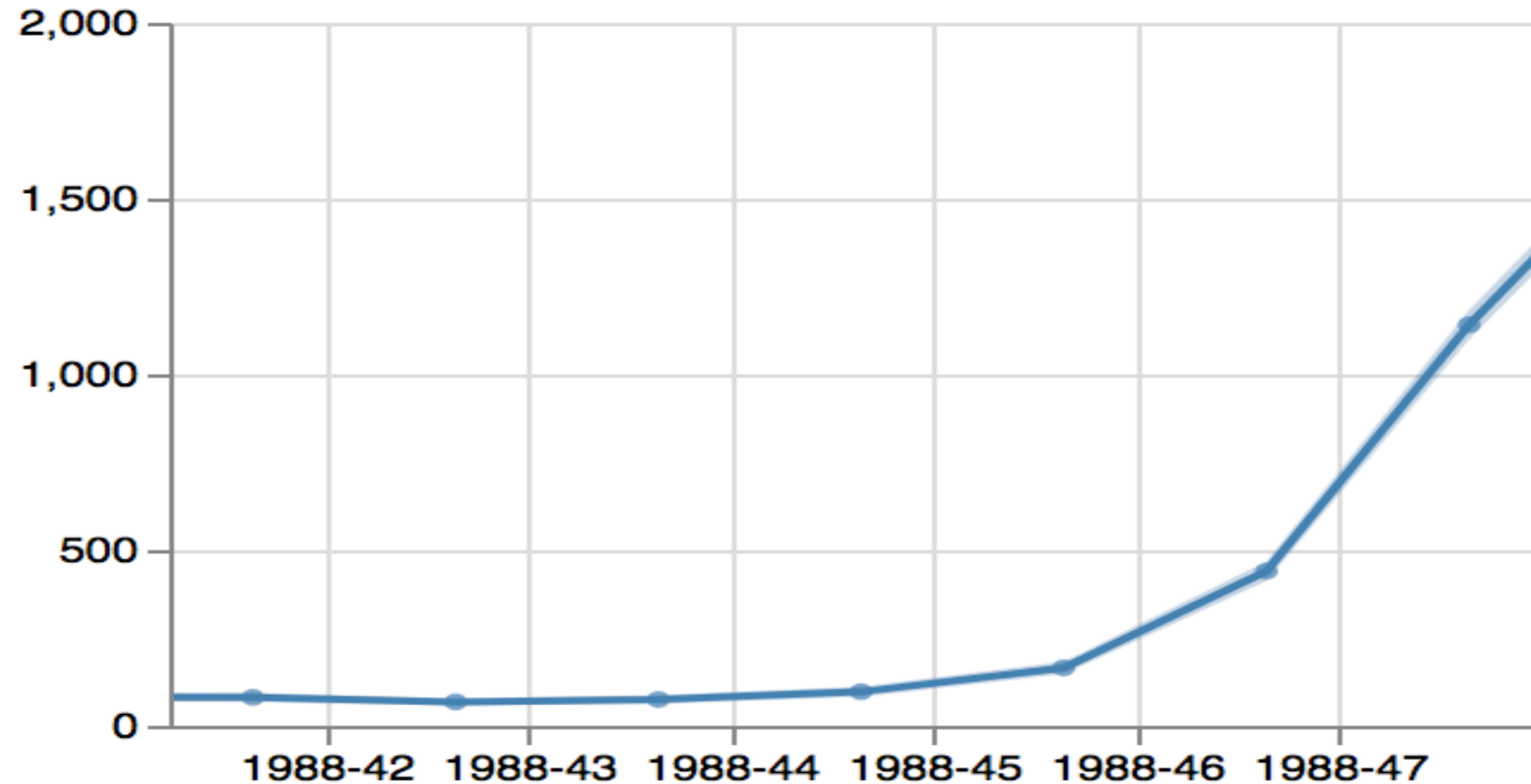
Incidence in a given area a= numerator$_a$/denominator$_a$

numerator = the number of people hit by flu

denominator = the population at risk

numerator: case definition

case definition = set of criteria used in making a decision as to whether an individual has a disease or health event of interest

possible criteria: clinical, laboratory characteristics, information regarding the person
cases can be classified: confirmed, probable, possible

case definition: sensitivity and specificity

A **sensitive** case definition will detect many cases but may also count as cases individuals who do not have the disease.

A **specific** case definition is more likely to include only persons who truly have the disease under investigation but also more likely to miss some cases

	Disease is truly present	Disease is truly absent	Total
complies to case definition	a	b	all cases
does not comply to case definition	c	d	all non-cases
	all 'diseased'	all 'non-diseased'	all people in the study sample

Sensitivity = [a / (a+c)]

Specificity = [d / (b+d)]

(https://wiki.ecdc.europa.eu/fem/w/wiki/sensitivity-and-specificity-of-a-case-definition)

numerator: flu

numerator for a area *a*= cases seen by General Partitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

symptoms of flu

- no symptoms (~30%)

- upper respiratory symptoms, e.g. nasal stuffiness, runny nose, sore throat, sneezing, hoarseness, ear pressure, or earache (~60%)

- lower respiratory symptoms, e.g. cough, breathing difficulty, and chest discomfort (~2%)

- fever (~35%)

(Carrat et al. Am J Epidemiol 2008)

numerator: flu

numerator for a area *a*= cases seen by General Partitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

symptoms of flu

- no symptoms (~30%)

- upper respiratory symptoms, e.g. nasal stuffiness, runny nose, sore throat, sneezing, ear pressure, or earache (~60%)

- lower respiratory symptoms, e.g. cough, breathing difficulty (~2%)

- fever (~35%)

- from early to peak symptoms ~1 day

clinical case definition

- fever > 39 °C AND myalgia

- sudden onset

- respiratory symptoms

Sentinelles
Réseau Sentinelles

(Carrat et al. Am J Epidemiol 2008)

numerator: flu

numerator for a area *a*= cases seen by General Partitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

symptoms of flu

- no symptoms (~30%)

- upper respiratory symptoms, e.g. nasal stuffiness, runny nose, sore throat, sneezing, hoarseness, ear pressure, or earache (~60%)

- lower respiratory symptoms, e.g. cough, breathing difficulty, and chest discomfort (~2%)

- fever (~35%)

(Carrat et al. Am J Epidemiol 2008)

clinical case definition

- fever OR malaise OR headache OR myalgia

- sudden onset

- cough OR sore throat OR shortness of breath

numerator: flu

numerator for a area *a=* cases seen by General Partitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

clinical case definition

Sentinelles
Réseau Sentinelles

- fever > 39 ℃ AND myalgia

- sudden onset

- respiratory symptoms

higher specificity

ecdc
EUROPEAN CENTRE FOR
DISEASE PREVENTION
AND CONTROL

- fever OR malaise OR headache OR myalgia

- sudden onset

- cough OR sore throat OR shortness of breath

higher sensitivity

numerator: flu

numerator for a area *a=* cases seen by General Partitioners defined based on *clinical criteria*. These are in fact possible cases. Laboratory confirmation available only for a small proportion of cases.

clinical case definition

Sentinelles
Réseau Sentinelles

- fever > 39 ℃ AND myalgia

- sudden onset

- respiratory symptoms

ecdc
EUROPEAN CENTRE FOR
DISEASE PREVENTION
AND CONTROL

- fever OR malaise OR headache OR myalgia

- sudden onset

- cough OR sore throat OR shortness of breath

higher specificity   **higher sensitivity**

many cases are missed (no symptoms for ~30%)

denominator: flu

denominator for the area *a*= catchment population, i.e. all the people living in the catchment area of the General Partitioner reporting the cases, who would usually seek healthcare at the site when they get sick
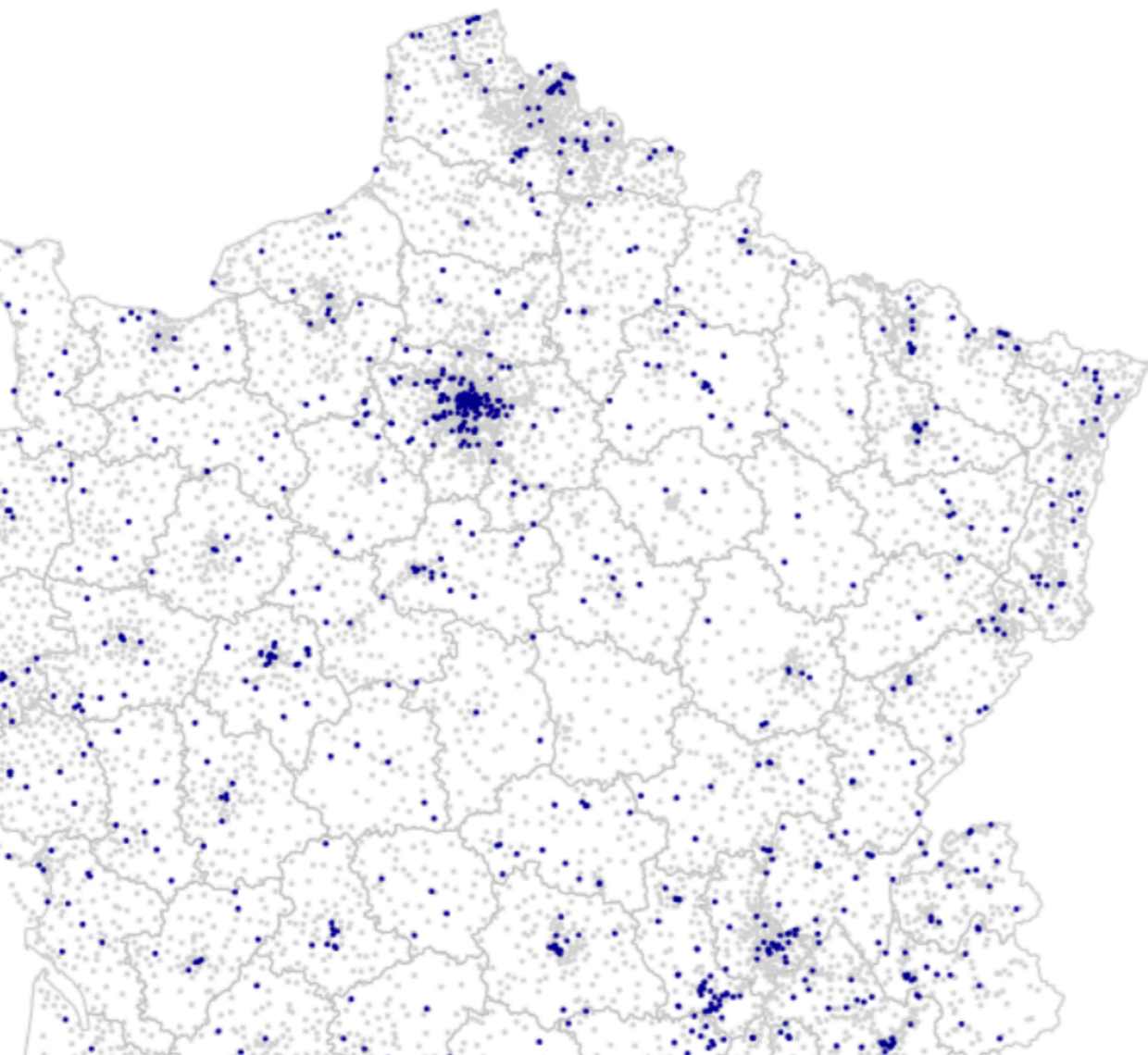
The Surveillance Networks (SN) is based on a *fraction of General Partitioner (~1%), who are* volunteers

denominator a= Population$_a$ GP$_{SN,a}$/GP$_a$

(Horvitz DG, Thompson DJ. A JASA. 1952;47:663–85)

incidence: flu

Incidence in a given area a= numerator$_a$/denominator$_a$

but … consultancy rate

- highly variable by age

- variable geographically (e.g. dependent on the GP density)

- variable according to the period of the year

- highly dependent on the health-care system (how expensive is going to the GP? Do you need a permit for staying at home from work?)

incidence: flu

Monitored population

C (with ILI symptoms)

A (infected)

B (detectable)

E (recorded as ILI)

F (tested)

D (GP)

(Baguelin et al. PLOS Med, 2013)
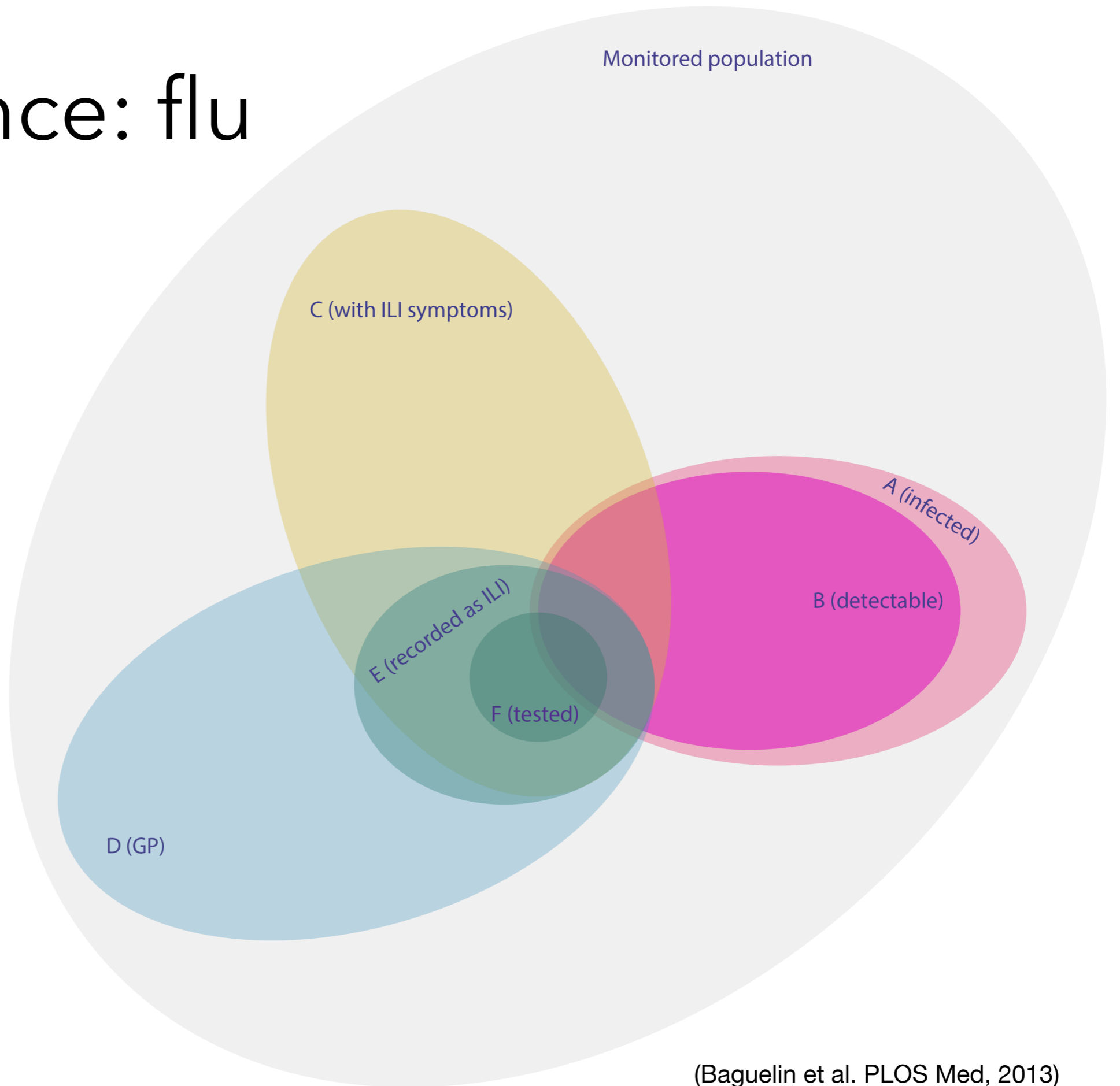
incidence: MERS

- emerging disease (first detected in 2012)

- severe respiratory infection (mortality rate ~35%)

- zoonotic origin (dromedary camels)

- limited human-to-human transmission

- zoonotic transmission (from camel to human)
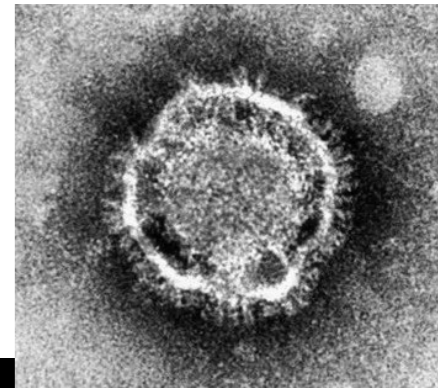
Middle East Respiratory Syndrome (MERS) Coronavirus

Al Jawf:2
Dammam:6
Dhahran:2   Khobar:8
Khafji:2
Taimah:3   Hail:1   Abqig:1
Dhebaa:1   Jubail:1
                Al Uyun:1   Buraidah:10
Badea Qseem:1   Al Zulfi:1
Alrass:2   Unayzah:2
            Shaqra:1   Hofuf:47
                Huraymeila:1   Riyadh:264
Madinah:8   Alfadliah:1
Ad Dawadimi:1
                            Alkharj:6
Al Quwayiyah:3

Al-Aflai:1

Al Khurma:1
Jeddah:16
            Turaba:2
Mecca:3   Taif:9

Number of cases
·   1
•   10
●   100

Province
Affected
Not affected

Qunfuthah:1   Al Namas:2

Abha:1
Khamis Mushayet:1   Najran:
Jazan:1

incidence: MERS

- case definition varying in time: range of symptoms unknown

- not wide spread: unclear the region where it is spreading, thus denominator hard to estimate

- quality and coverage of the health care systems unknown

- reporting rate highly variable in time: from passive to active surveillance, to saturation of the surveillance system

- there can be a retrospective correction fo the number of cases: further source of biases in case of a real time analysis

computational epidemiology

epidemiology = making sense of data

computational epidemiology
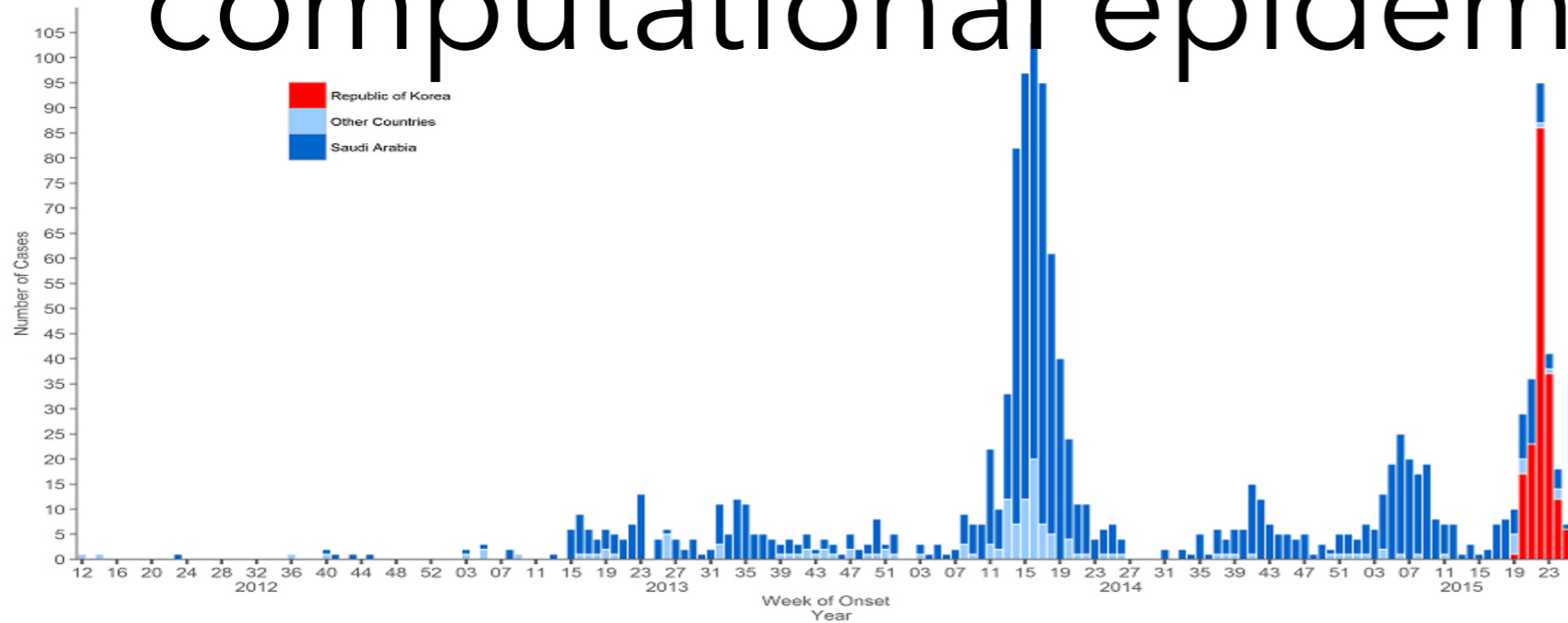
computational epidemiology = making sense of scarse data using big data
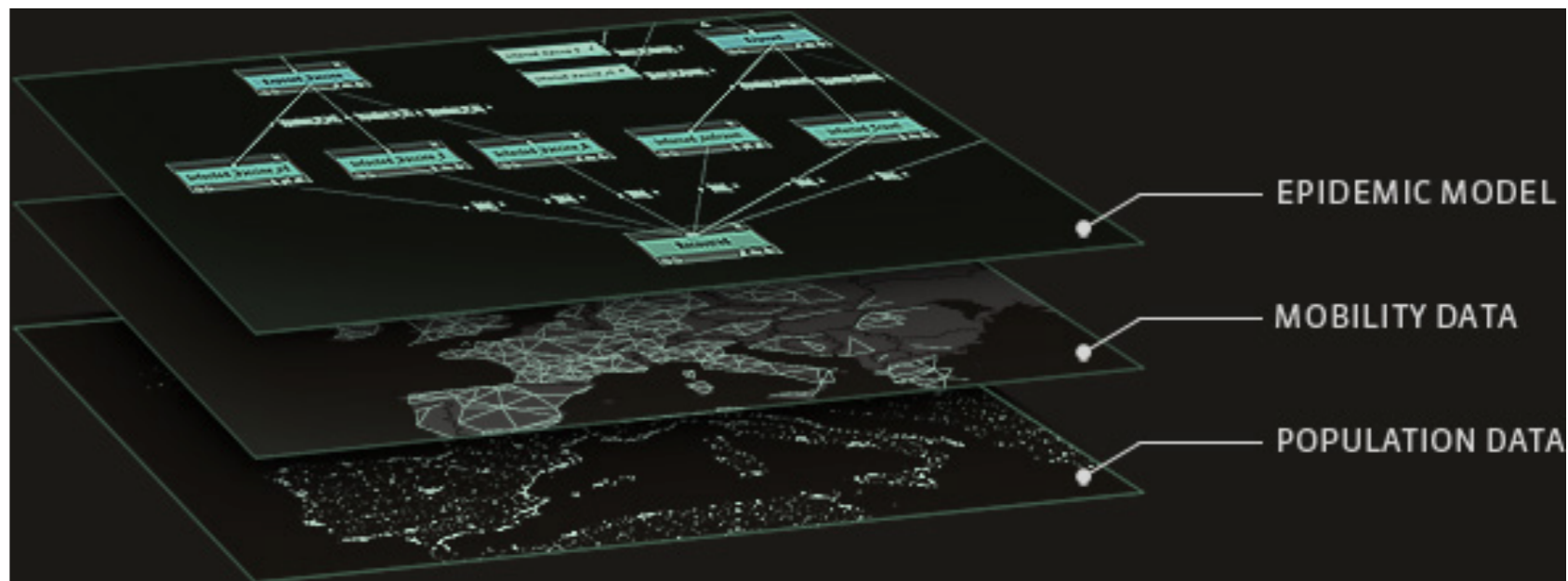
projects

- MERS outbreak in South Korea and role of super-spreading events

- 2014 Ebola outbreak

- Competition between strains of influenza virus

- Spread of sexual-transmitted-diseases (STDs) on a prostitution network

- Measles spread and vaccine hesitancy

- Critical community size and effect of spatial structure

project 1: MERS

Hospital outbreak of MERS in South Korea, 2015: 186 cases and 36 deaths

SCIENCE VISUALIZED

Male, 35 years old
May 27–29
A man who shared a hospital room with the first patient in the South Korean MERS outbreak infected the most people. On May 27, he went to Samsung Medical Center in Seoul, where he had to wait in the emergency room for a bed to become available. Over the next two and a half days, more than 80 people who had passed through the ER contracted the virus.

Ambulance June 6

Konkuk University Hospital June 6

Female, 75 years old
June 5–6
An elderly woman spread the virus to 11 other people in a fourth round of infection. Multiple rounds of infection are a concern because sustained transmission can lead to a pandemic.

Kyung Hee University Hospital June 5–6

Unknown exposure

Male, 68 years old
May 15–17
South Korean officials traced the MERS outbreak to a businessman who visited the Middle East in April and early May. Soon after returning, the man was admitted to St. Mary's Hospital in Pyeongtaek, where he infected about 30 people, mostly visitors and fellow patients.

St. Mary's Hospital May 15–17

Samsung Medical Center May 27–29

Good Morning Hospital May 25–27

Sacred Heart Hospital May 27–31

Daejeon Dae Cheong Hospital May 25–28

Konyang University Hospital May 28–30

Anatomy of a MERS outbreak
In 2015, South Korea experienced an outbreak of Middle East respiratory syndrome, or MERS. Between May and July, 186 people contracted the MERS virus; 38 eventually died. This diagram shows how quickly the pathogen spread within and between hospitals via a handful of "superspreaders." — *Tina Hesman Saey*

Male, 40 years old
May 25–30
A patient's visits to two hospitals led to 27 infections. Overall the virus reached dozens of health centers (not all labeled), in part because South Koreans tend to shop around for the best care.

Legend
- Patient
- Patient, deceased
- Superspreader
- Visitor
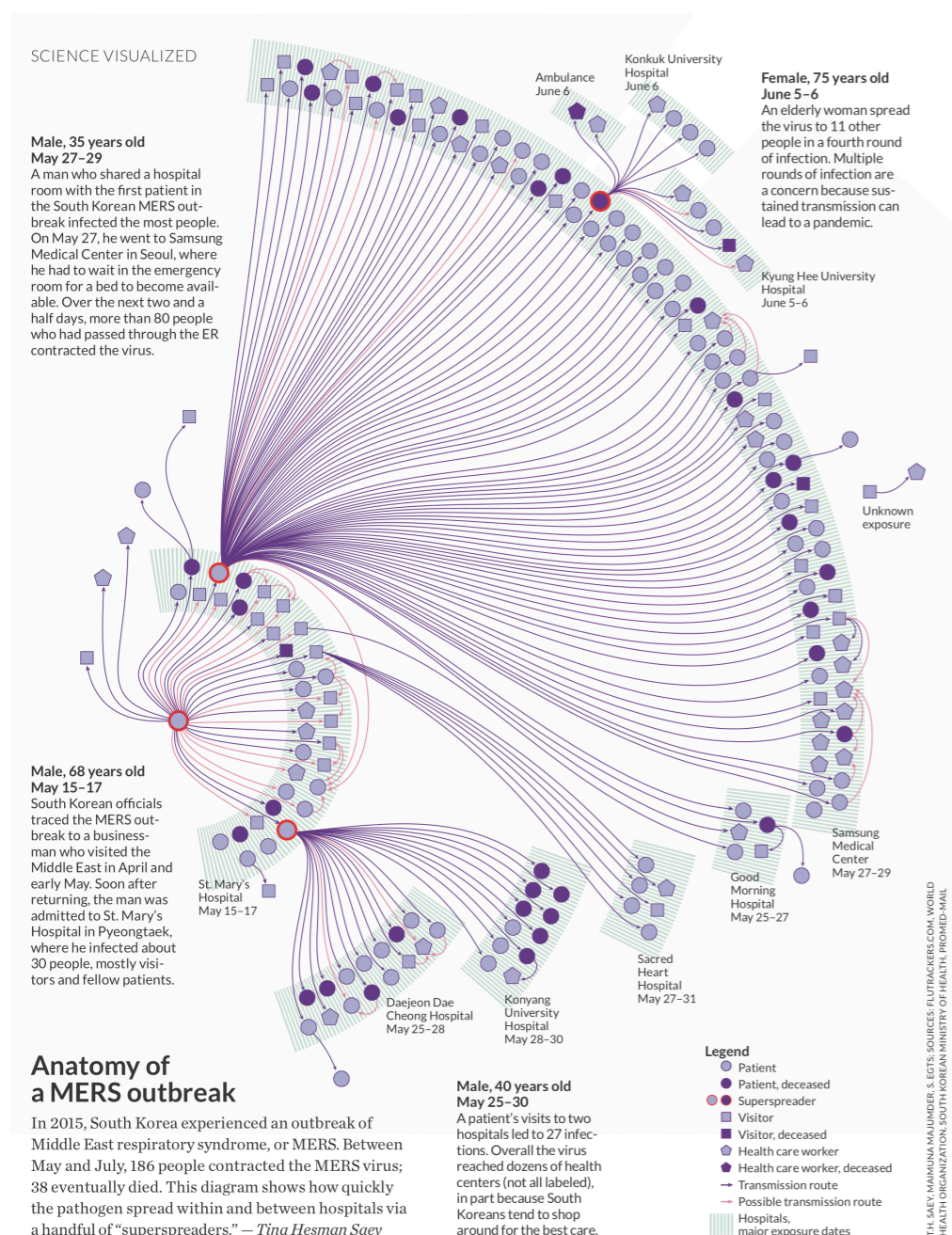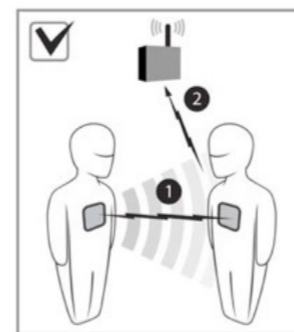- Visitor, deceased
- Health care worker
- Health care worker, deceased
- → Transmission route
- → Possible transmission route
- Hospitals, major exposure dates

T.H. SAEY, MAHNUNA MAJUMDER. S. EGTS; SOURCES: FLUTRACKERS.COM, WORLD HEALTH ORGANIZATION, SOUTH KOREAN MINISTRY OF HEALTH, PROMED-MAIL

[science news, Dec 26, 2015]

Heterogeneities in contacts can lead to super-spreading events

data on face-to-face contacts in a hospital from sociopattern.org

–Characterise the properties of the network from the data

–Simulate a MERS epidemic starting from the different nodes comparing different level of transmissibility. Is this different according to the node initially infected?

–What are the properties of the node that mostly affect the outbreak size? What are the category of people sensible to cause bigger outbreak?

project 2: Ebola

Largest Ebola outbreak in the history: 28,000 cases; 10,000 deaths; 10 countries affected

[Nature 514, October 2014]

Ebola data and Statistics
[http://apps.who.int/gho/data/node.ebola-sitrep]

Althaus CL. PLOS Currents Outbreaks. 2014. doi: 10.1371/currents.outbreaks. 91afb5e0f279e7f29e7056095255b288.

Legrand J, Grais RF, Boelle PY, Valleron AJ, Flahault A. Epidemiol Infect. 2007;135(4):610-21.
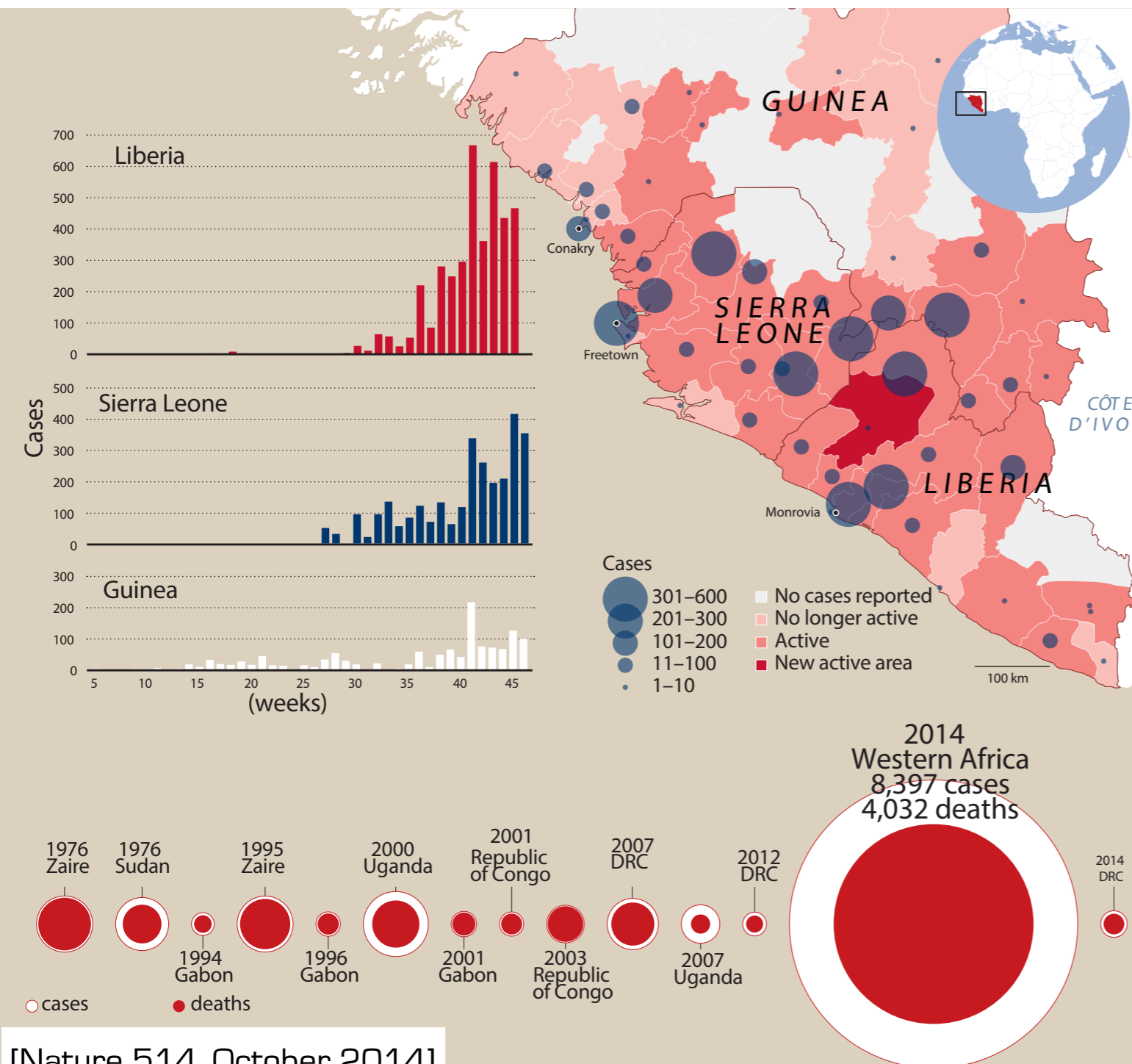
-Analyse the epidemic curves by fitting a compartmental model appropriate for Ebola and estimate R0.

-compare the three different countries, different time windows, different compartmental models

project 3: Influenza A

Pigs often a source of flu pandemics as they can be infected by bird, human & swine flus.

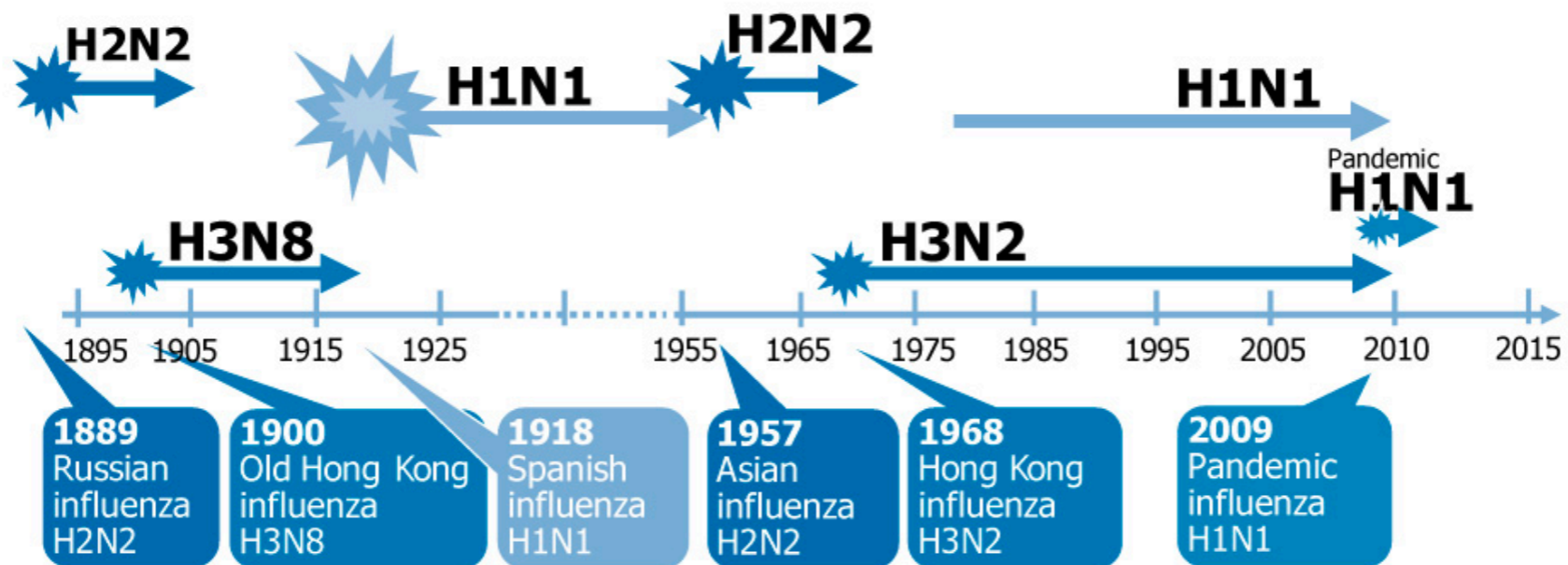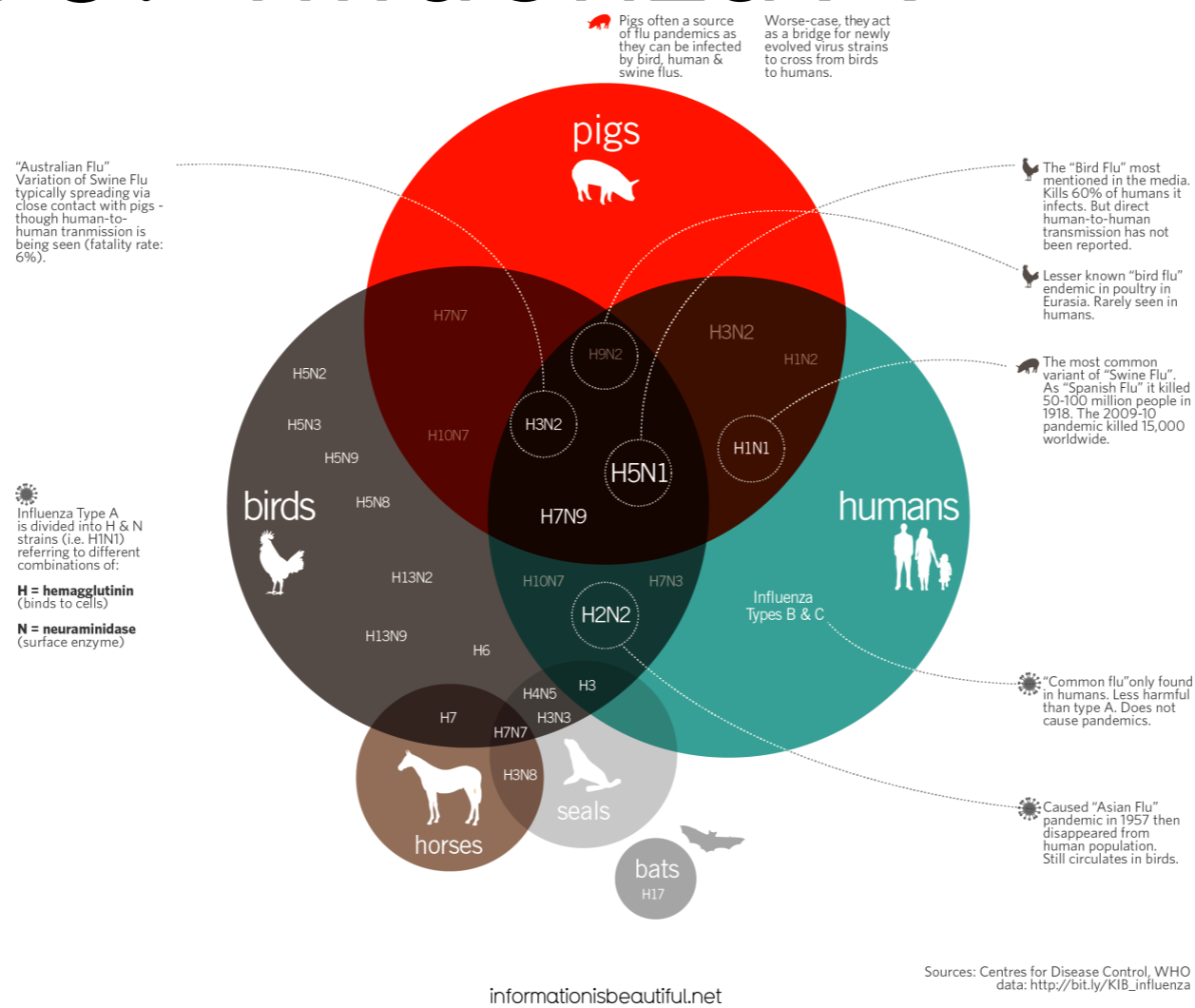Worse-case, they act as a bridge for newly evolved virus strains to cross from birds to humans.

"Australian Flu" Variation of Swine Flu typically spreading via close contact with pigs - though human-to-human tranmission is being seen (fatality rate: 6%).

The "Bird Flu" most mentioned in the media. Kills 60% of humans it infects. But direct human-to-human transmission has not been reported.

Lesser known "bird flu" endemic in poultry in Eurasia. Rarely seen in humans.

Influenza Type A is divided into H & N strains (i.e. H1N1) referring to different combinations of:

H = hemagglutinin (binds to cells)

N = neuraminidase (surface enzyme)

The most common variant of "Swine Flu". As "Spanish Flu" it killed 50-100 million people in 1918. The 2009-10 pandemic killed 15,000 worldwide.

Influenza Types B & C

"Common flu"only found in humans. Less harmful than type A. Does not cause pandemics.

Caused "Asian Flu" pandemic in 1957 then disappeared from human population. Still circulates in birds.

informationisbeautiful.net

Sources: Centres for Disease Control, WHO
data: http://bit.ly/KIB_influenza

Source: European Centre for Disease Prevention and Control (ECDC) 2009

project 3: Influenza A

2 types (A and B); 2 subtypes of type A (H3N2 and H1N1)

known facts about influenza A:

-Subtypes (strains) may have (slightly) different transmissibilities.

-Children and Adults have a different susceptibility to each strain: adults more immune than children; individuals more immune to strains that encountered early in life (antigenic seniority or antigenic sin)

-Social contacts structured by age: individuals enter in contact more with individuals of similar age; children have more contacts than adults.

-Subtypes interact with cross-immunity: after getting infected with a strain I become fully immune to that strain and partially immune to the other (cross-immunity lasts few months)

-Strain specific immunity wanes after some time (estimated to be between 1 to 7 years).

Study the competition between the two strains, considering two age-classes: children (< =15 years old), and adults (> 15 years old)

Identify the dominant strain as a function of the strain specific transmissibility and fraction of immune among the two age classes (also strain-specific)

Consider a single influenza season: account for full cross-immunity, and in a second moment partial cross immunity. Bonus: Account also for the waning of immunity

project 4: STDs

Internet mediated prostitution: sexual contacts between 6,624 escorts and 10,106 sex buyers extracted from an online community
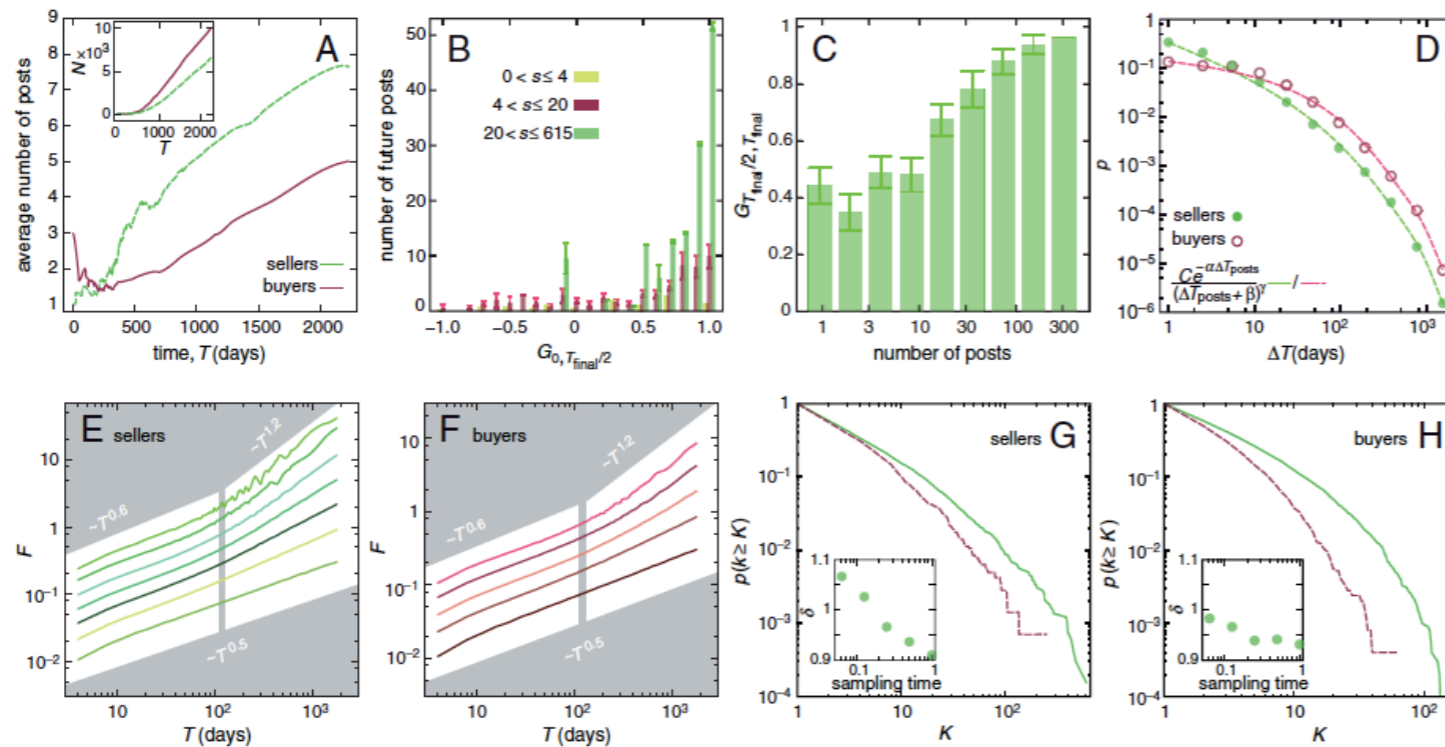
Fig. 1. Statistics of the dynamics of the community. (A) Time evolution of the average number of posts by sex buyers and about sex buyers. The *Inset* shows the growth in the number of sex sellers and sex buyers in the data. (B) The number of new posts according to the previous average grade at $T_{final}/2 = 1,116$ days for three different activity levels, or total number of posts, s. The R^2-values of these data are 0.19 ($0 < s \leq 4$), 0.29 ($4 < s \leq 20$), and 0.33 ($20 < s$). (C) The average future grade of sellers as a function of their number of contacts at half of the total sampling time (the data is logarithmically binned along the abscissa). (D) Shows the distribution of the time elapsed between two posts T_{posts} for buyers and sellers. Many posts were written during the same day, respectively, $p(T_{posts} = 0) = 0.495$ and $p(T_{posts} = 0) = 0.246$. The distributions are well fitted by $p(T_{posts}) = C \exp(-\alpha T_{posts}) = (T_{posts} + \beta)^\gamma$, with: $C = 2.9 \pm 0.5$ days$^\gamma$, $\alpha = 0.0023 \pm 0.0001$ days^{-1}, $\beta = 3.1 \pm 0.4$ days, and $\gamma = 1.49 \pm 0.04$ (for sellers); and $C = 12 \pm 8$ days$^\gamma$, $\alpha = 0.0021 \pm 0.0002$ days^{-1}, $\beta = 18 \pm 4$ days, and $\gamma = 1.5 \pm 0.1$ (for buyers). (E) and (F) shows statistics the DFA fluctuation function as a function of the time-scale ΔT for sellers and buyers, resp. The different curves correspond to different activity levels—from bottom to top they represent less than 3, 3–7, 8–20, 21–54, 55–148, 149–403, and more than 403 posts (about sellers or from buyers) resp. *Black Lines* are inserted for reference. $T^{1/2}$ corresponds to uncorrelated interaction. (G) and (H) show degree distributions for sex sellers (G) and buyers (H) cumulative degree distributions for the full sampling time (*Solid Line*) and a yearlong window (starting one year after the full dataset; *Dashed Line*) for sex sellers and -buyers, resp. The *Insets* show the exponent of preferential attachment (Eq. 1).

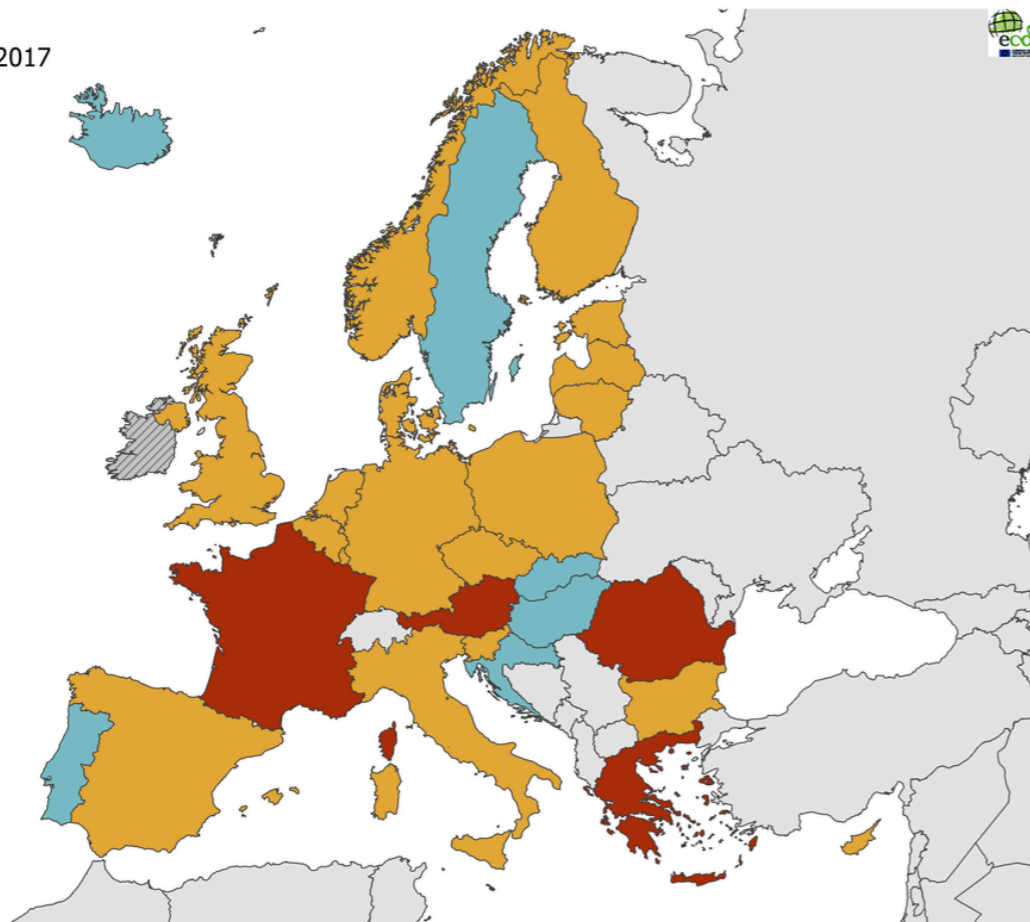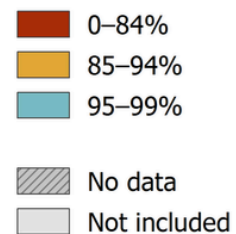Dataset made available
[LEC. Rocha, et al, PNAS 2009]

Study the impact of the structure of the network on the dynamics of an infection by comparing simulations on this network with simulations on randomized reference models

project 5: measles

Vaccine hesitancy listed among the 10 top threats to global health by WHO. Key facts about measles:

- $R_0 = [15\text{-}20]$ - vaccination rate to attain eradication 95%
- mortality rate 0.1-0.3%
- 1998: AJ Wakefield publishes on Lancet a paper that links MMR vaccine to autism. The study was later found to be a deliberate fraud, the paper was retracted, Wakefield was struck off the UK medical register
- "pockets" of unvaccinated individuals

Coverage of second dose of measles-containing vaccine, 2017

- 0–84%
- 85–94%
- 95–99%
- No data
- Not included

Luxembourg

Malta

Source: WHO/UNICEF Estimates of National Immunization Coverage   Produced 06 Feb 2019 using ECDC map maker: https://emma.ecdc.europa.eu

Study the simultaneous spread of measles and a fake news regarding the measles vaccine. Design a suitable model and simulate its dynamics.

project 6: measles & critical community size

In 1960, Bartlett defined the critical community size for measles as :

'the size for which measles is as likely as not to fade out after a major epidemic until reintroduced from outside, corresponding to a mean time to fade-out of about two years or about 30 in terms of average weekly notifications'.
[Bartlett MS Journal of the Royal Statistical Society Series A (General). 1960;123(1):37-44.]

Things are much more complex if we account for realistic ingredients, e.g. spatial structure

Consider the simple case of two cities, with a certain number of individuals traveling from one city to another. Model measles and analyse what are the conditions of persistence / extinction of the disease. Explore in depth the role of mobility and population size.