Big Data for Public Health – Public Data for Big Health

Stefan Thurner

www.complex-systems.meduniwien.ac.at www.santafe.edu

wien may 24 2017

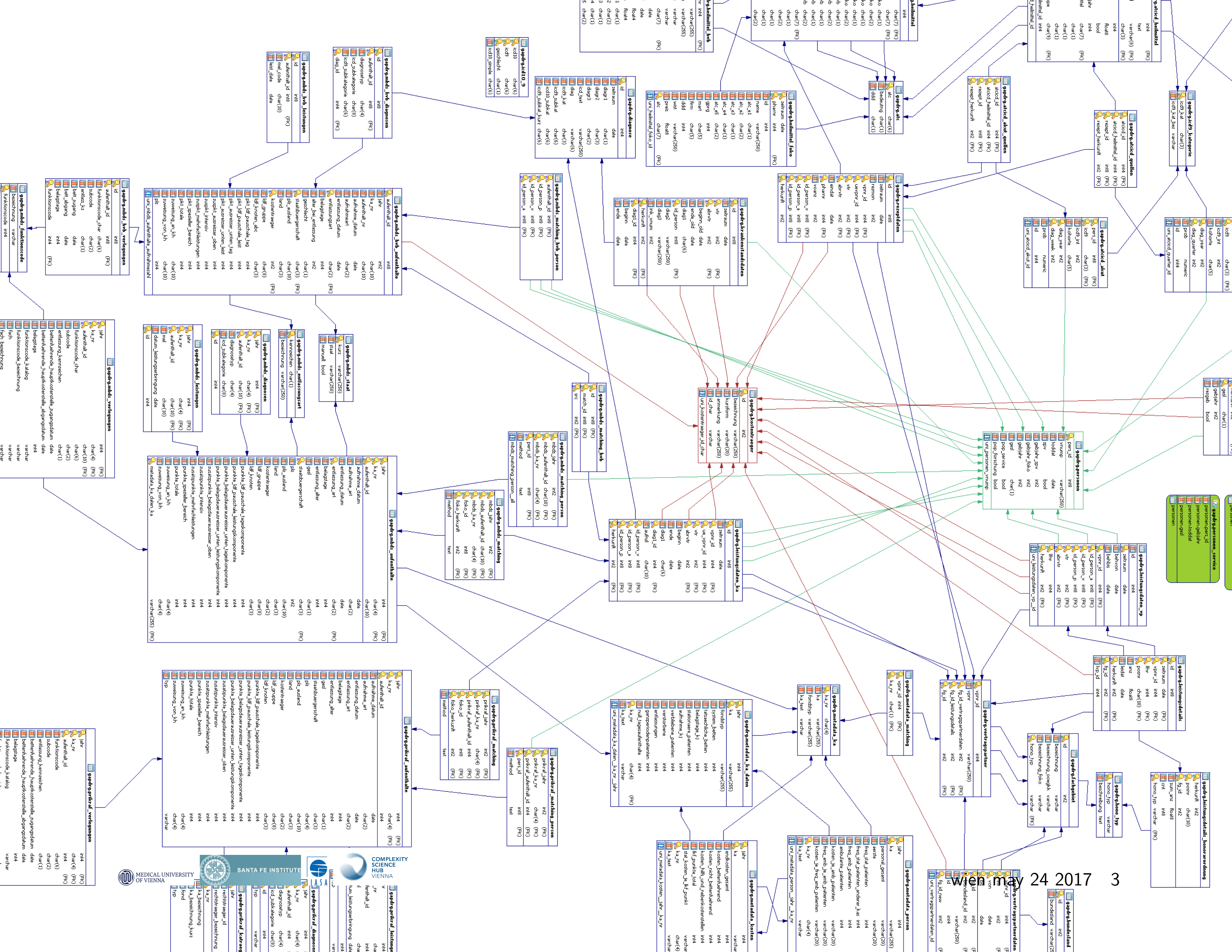
with

Peter Klimek, Silke Aichberger, Anna Chmiel

supported by 

EC FP7 projects LASAGNE and MULTIPLEX

date — patient ID — HCP ID — location — diagnosis — side diagnoses — prescription — price of generic drug/treatment — pharmacy ID — price of drug — date of purchase



MEDICAL UNIVERSITY OF VIENNA

SANTA FE INSTITUTE

IIASA

COMPLEXITY SCIENCE HUB VIENNA

100.000.000 lines per year

Data set

$\rightarrow$ medical claims data

for **every** payed **healthcare service** there is one data line

Data set

- 8,000,000 patients
- 100,000,000 patient visits per year
- 2,000,000 hospitalisations per year
- 12,000 health care providers
- 6,102 diseases (ICD10 code)
- 1,171 drugs (ATC code)
- 255 hospitals
- 1,238 pharmacies

Network medicine

Co-morbidity networks

starting point: usually patients have more than one disease

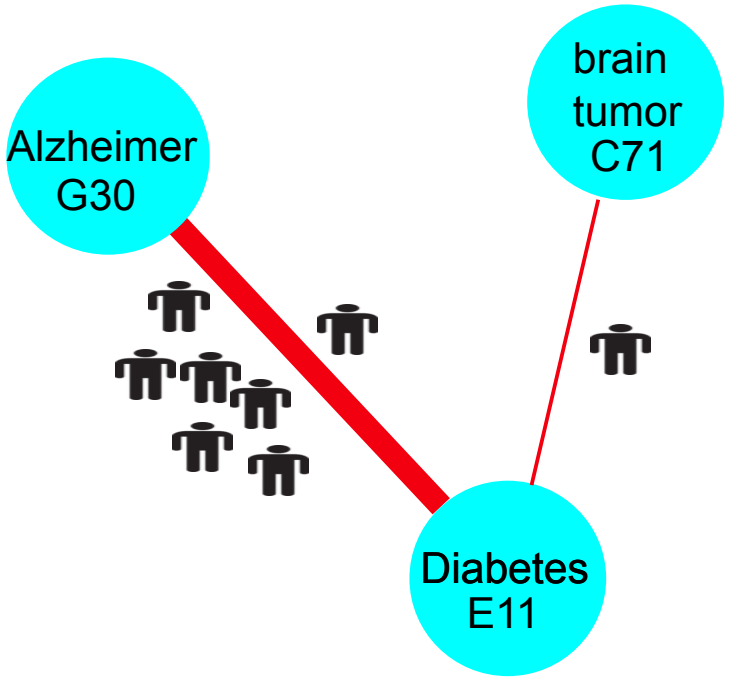
which diseases occur together? $\rightarrow$ co-morbidity networks

co-morbidity networks = health state of population (phenotype)

what can we learn from **co-morbidity networks**?

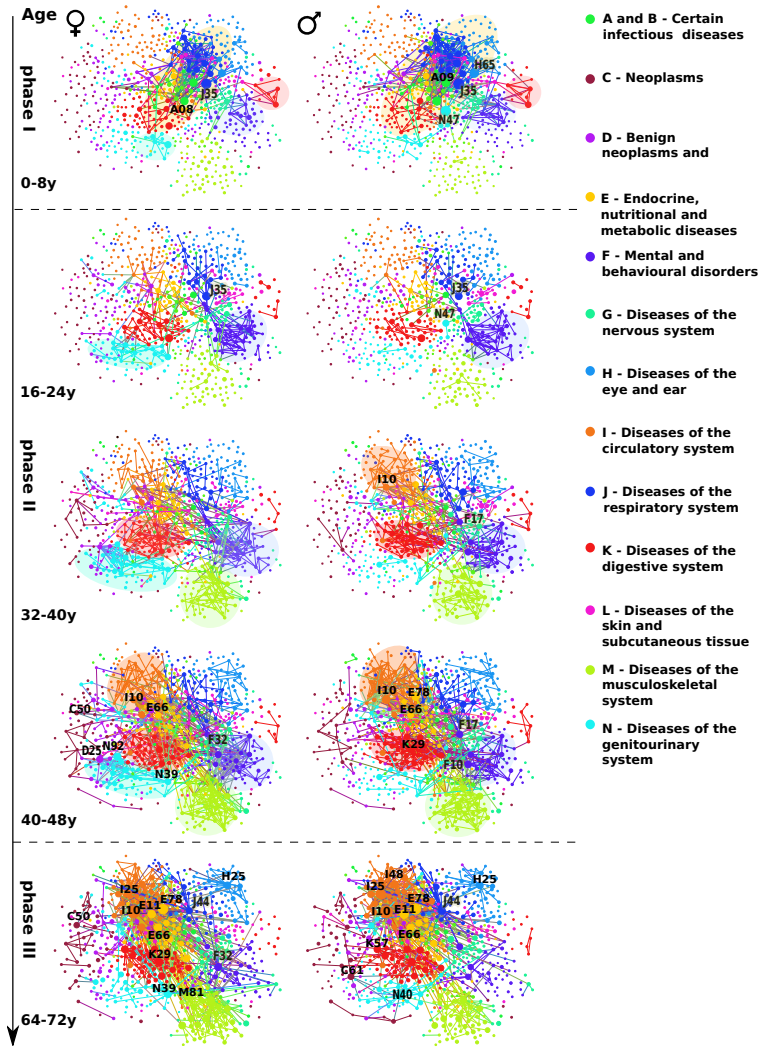
What is a co-morbidity network?

co-occurence of diseases in population

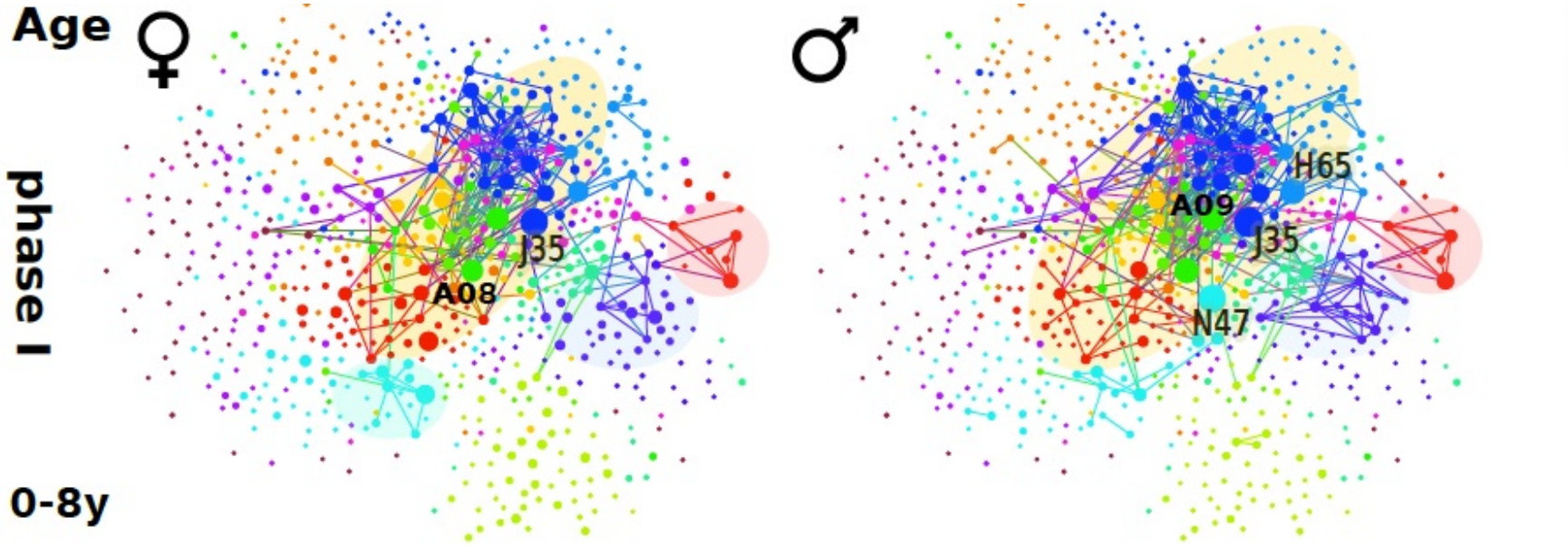


diseases ‘linked’ if many patients have both at same time

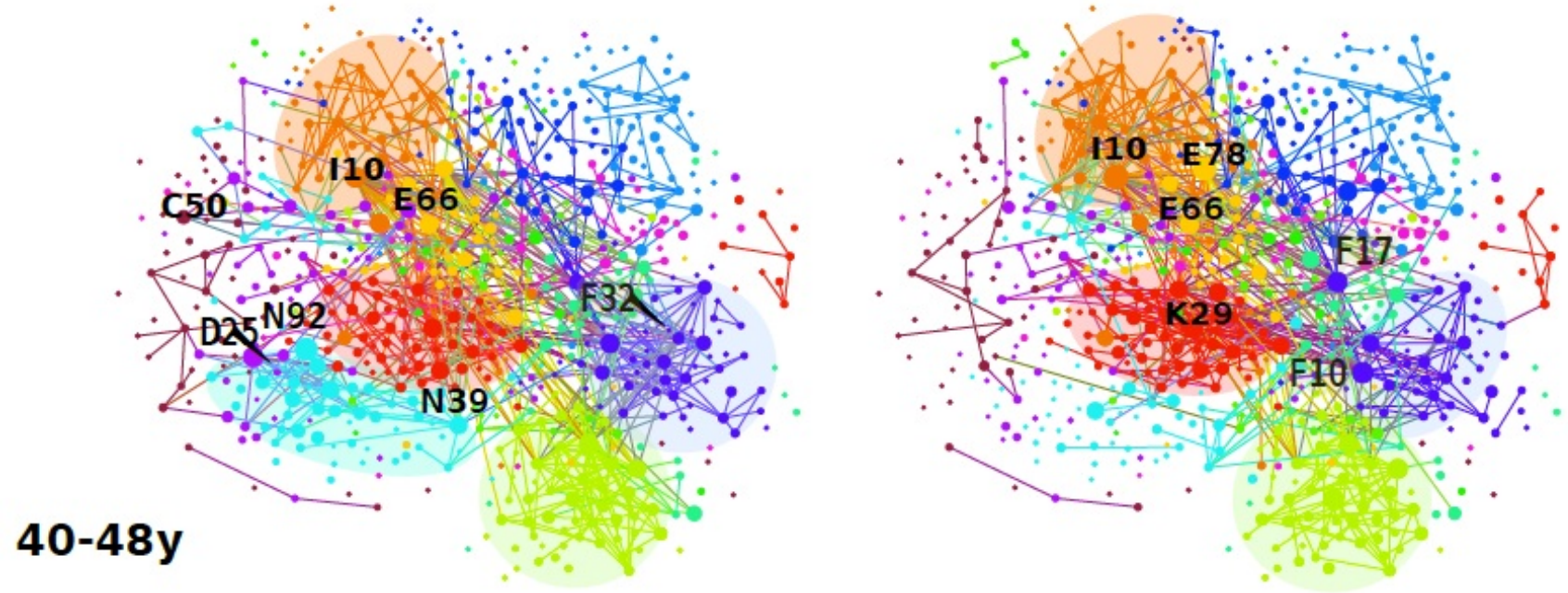
Co-morbidity network of Austria



The co-morbidity network: children



The co-morbidity network: adults age 40-48



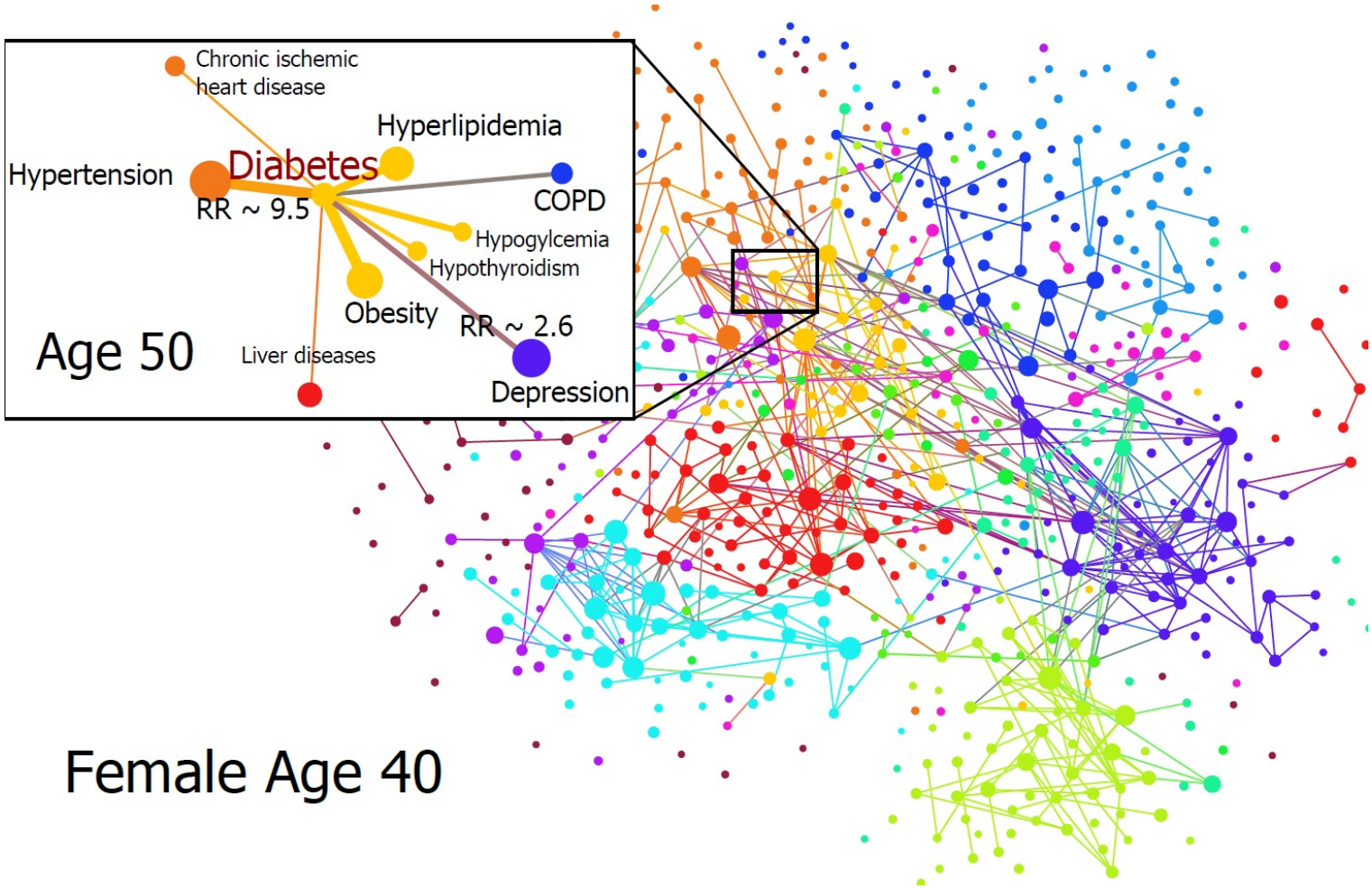
Empirical finding

the way individual diseases occur = diffusion on these networks

A Chmiel, P Klimek, S Thurner, New Journal of Physics 16, 115013, (2014)

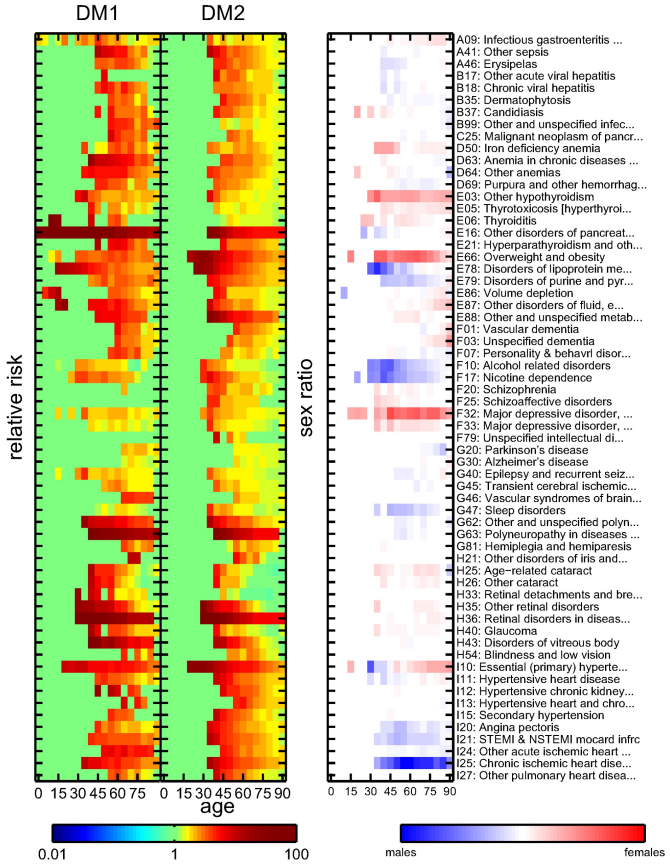
Prediction of health trajectories

Co-morbidity networks allow predictions



Co-morbidity networks allow predictions: DM

if have diabetes what is odds ratio to have any other disease?



P Klimek, A Kautzky, A Chmiel, S Thurner, PLoS Comput Biol 11(4): e1004125 (2015)

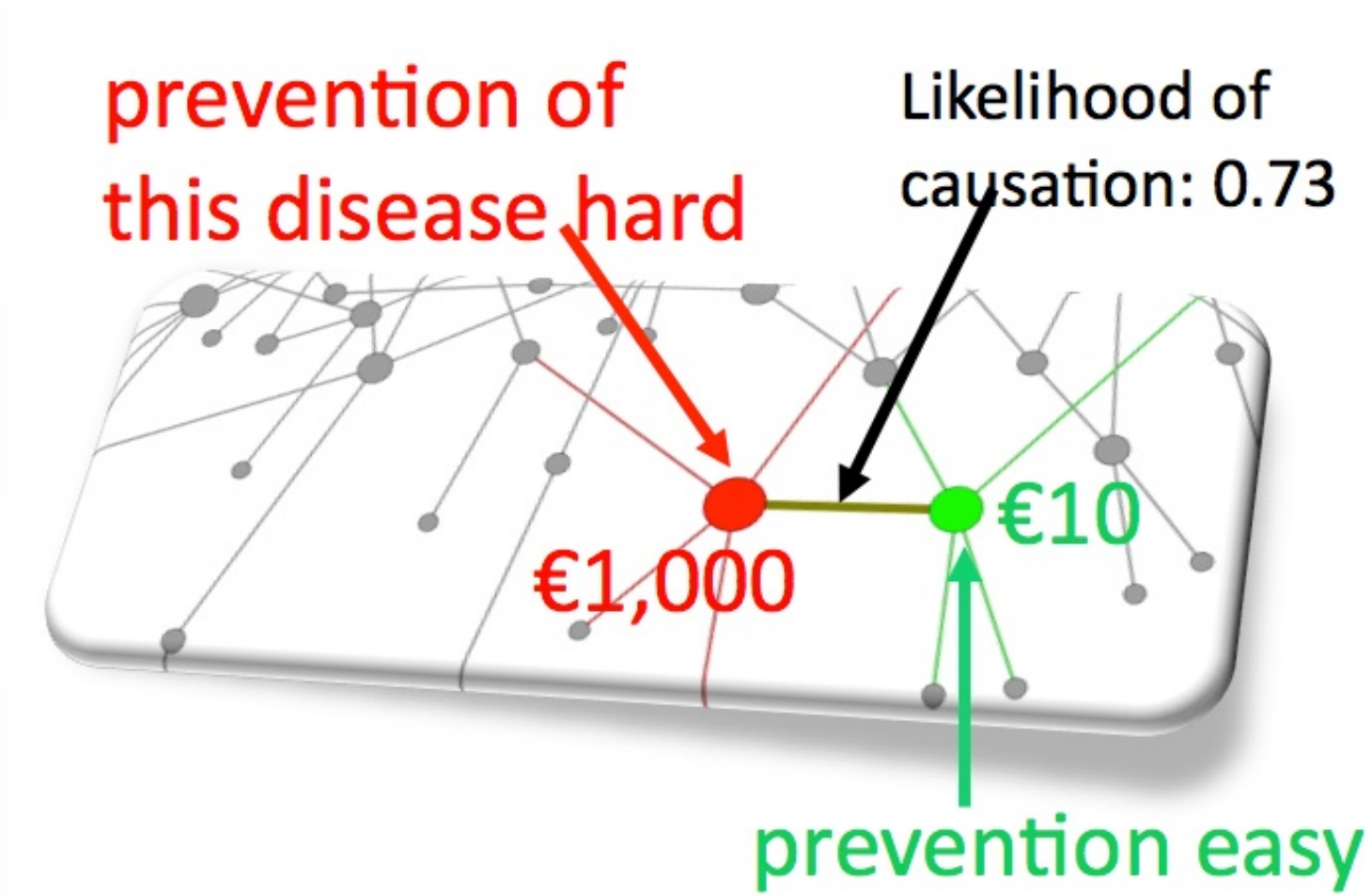
Co-morbidity network of diabetes explains ...

- analysis equivalent to 40,000 individual epidem. studies
- check which co-morbidity is causal
- confirm controversial relation of increased risk for Parkinson
- gender differences in progression of congestive heart failure
- females lower risk of hypertension during fertile age
- type 1 diabetes **leads** detection of depressions
- schizo-affective disorders **lead** type 2 diabetes, suggesting similar pathogenic or medication-related mechanisms

P Klimek, A Kautzky, A Chmiel, S Thurner, PLoS Comput Biol 11(4): e1004125(2015)

Efficacy of prevention

Co-morbidity networks and prevention



identify co-morbidities – check 'causality' – treat cause

New classification of diseases

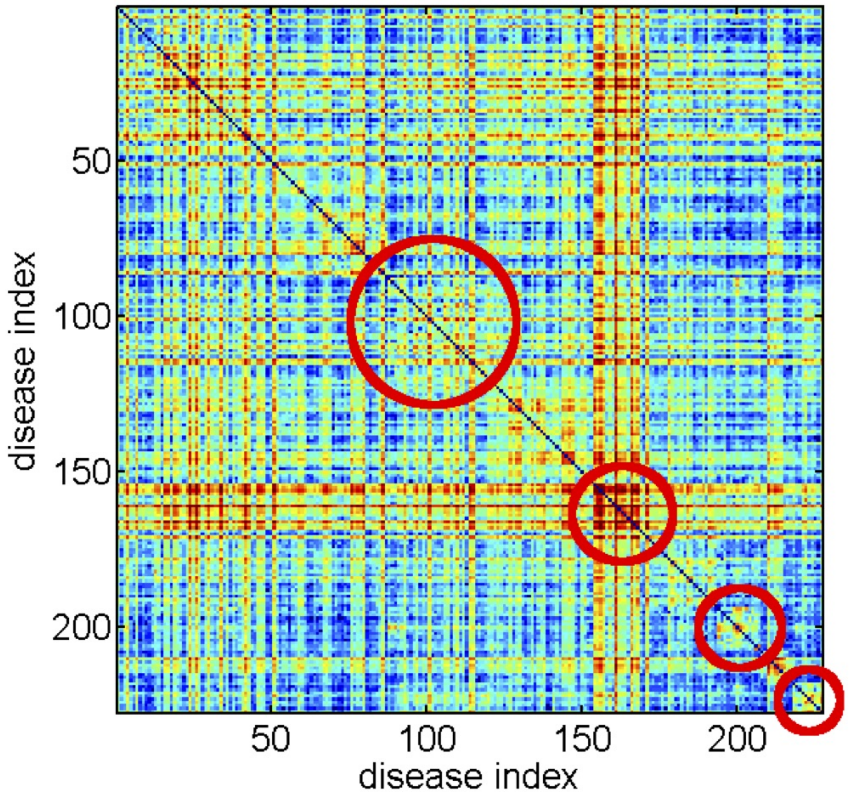
What is Diabetes?

Observe:

diabetes co-occurs with other diseases in robust patterns

$\rightarrow$ **allows us to classify diabetes differently**

New “types” of diabetes – defined by co-morbidity



define new phenomenological types of DM through co-morbidity

Which drug / therapy works?

- take a disease for which 2 therapies exist A and B
- compute all co-morbidities following therapy A
- compute all co-morbidities following therapy B
- compare: follow up costs, hospitalization time, co-morbitities

How genetic is your disease?

What explains a disease?

- genetic factors
- metabolic factors
- environmental / toxicogenetic factors
- epigenetic factors

How genetic is diabetes?

Genes associated with diabetes type 2

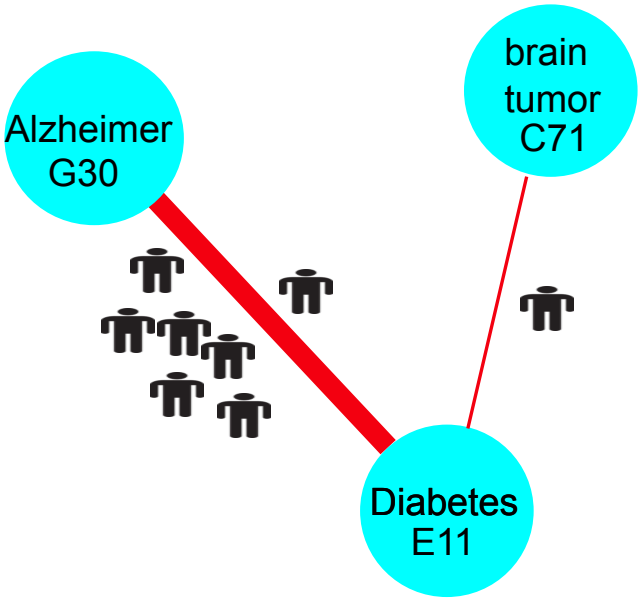
• HHEX/IDE/KIF11 • TCF7L2 • KCNJ11 • MTNR1B • HNF1A • FTO • GCKR • PPARG • ADCY5 • CDKAL1 • SLC30A8 • CRY2 • FADS1

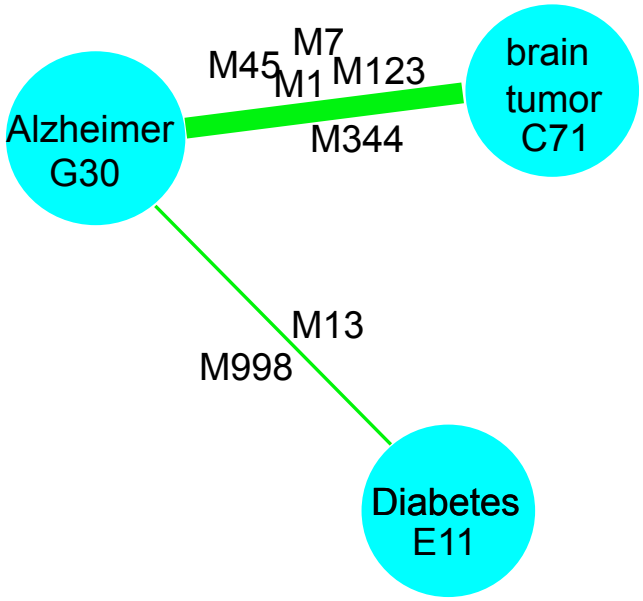
type 2 diabetes 25% hereditary

5-10% of variance explained by gene variants*

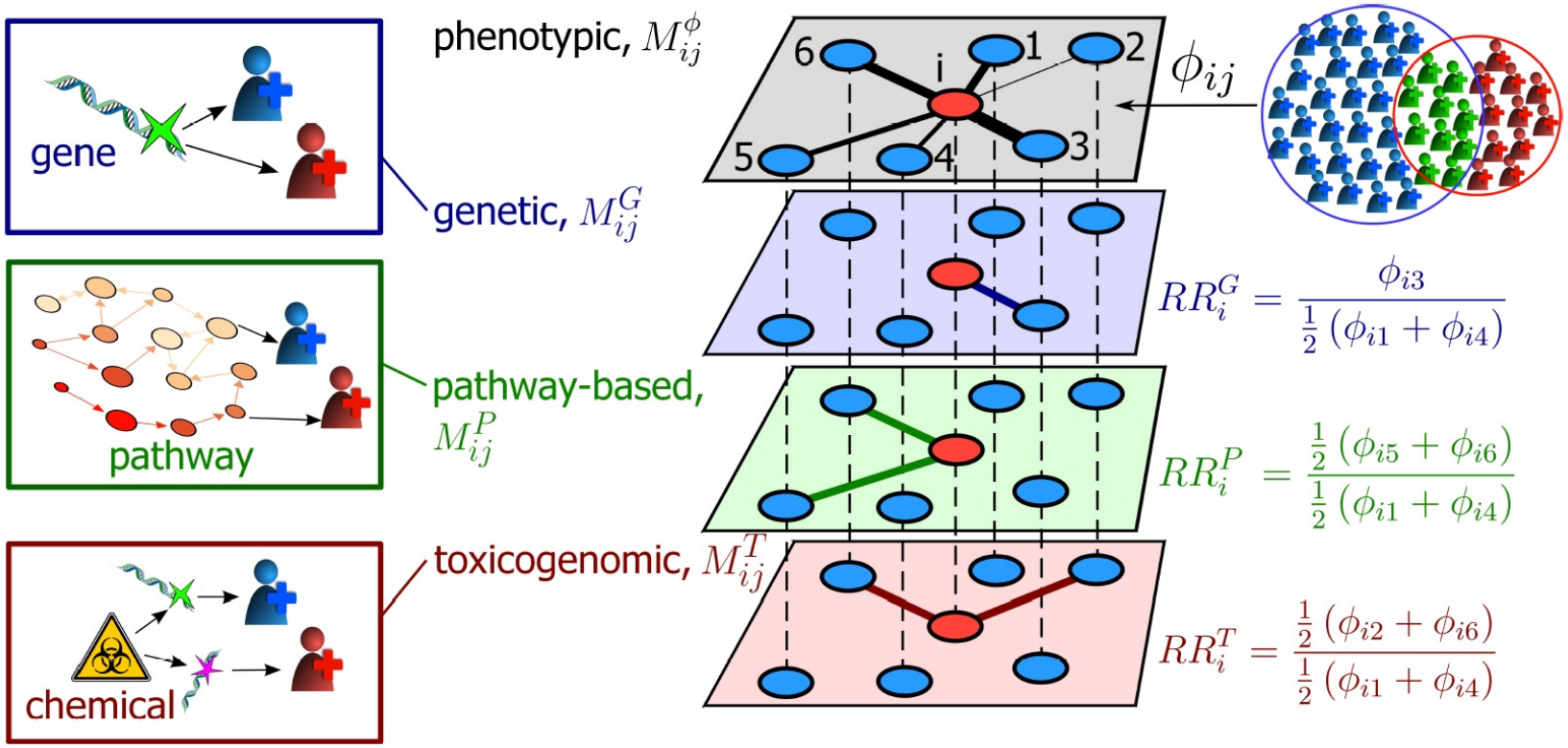
→ **hard to tell!**

* ME Travers MI McCarthy, Human Genetics 130 41-58 (2011)


Alzheimer
G30
brain
tumor
C71
Diabetes
E11


M7
M45
M1
M123
M344
Alzheimer
G30
brain
tumor
C71
M13
M998
Diabetes
E11

Compare co-morbidity and genotype networks



if co-morbidity network is "similar" to genetic network $\rightarrow$ yes it is genetic

Ranking of likelihood of genetic cause in multi-factoral diseases

- compute "similarity" between phenotypic and genetic network
- **hypothesis:** the more similar – the "more" genetic influence
- rank multi-factoral diseases wrt similarity in pheno-geno networks

only take cases that are unlikely to be of statistical origin $p < 0.00001$

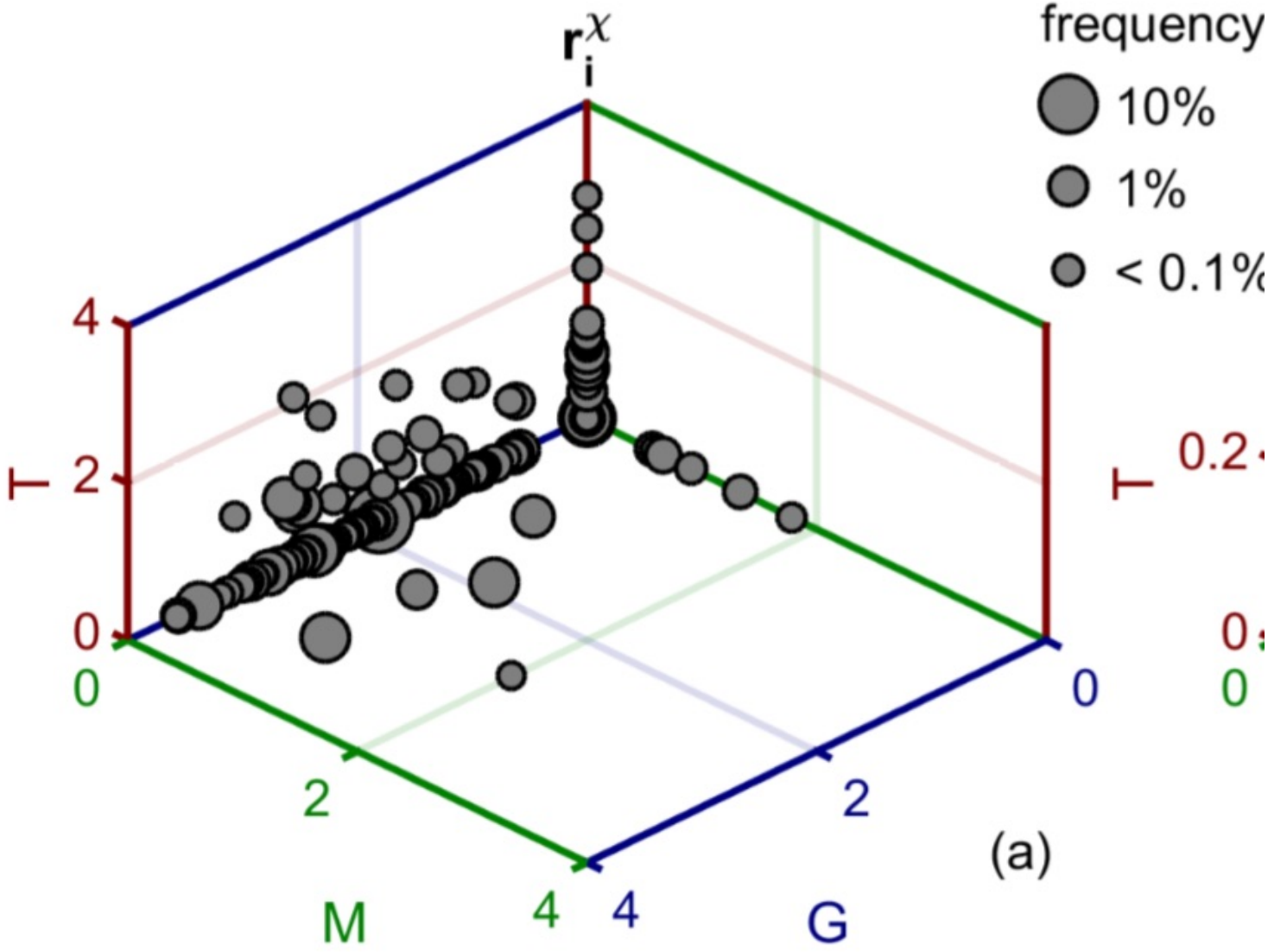
Type 1 diabetes mellitus	E10	0,50
Transient cerebral ischemic attacks and related syndromes	G45	0,50
Benign neoplasm of colon, rectum, anus and anal canal	D12	0,33
Eating disorders	F50	0,33
Polycythemia vera	D45	0,25
Other diseases of intestine	K63	0,25
Other cerebrovascular diseases	I67	0,21
Other and unsp diseases of blood and blood-forming organs	D75	0,21
Other congenital malformations of heart	Q24	0,20
Malignant neoplasm of heart, mediastinum and pleura	C38	0,16
Mesothelioma	C45	0,16
Specific personality disorders	F60	0,16
Overweight and obesity	E66	0,13
Other cardiac arrhythmias	I49	0,13
Cerebral infarction	I63	0,13
Type 2 diabetes mellitus	E11	0,11
Secondary parkinsonism	G21	0,11
Other and unspecified myopathies	G72	0,11
Congenital malformations of cardiac chambers and connections	Q20	0,10
Other congenital malformations of eye	Q15	0,09
Congenital malformations of aortic and mitral valves	Q23	0,09
Parkinsons disease	G20	0,08
[illegible]	G71	0,07
Essential (primary) hypertension	I10	0,07
Anoph	Q11	0,06

Classify multi-factoral diseases

- do the same with metabolic, environmental, pathway networks

$\longrightarrow$ **every disease gets assigned 3 numbers:**

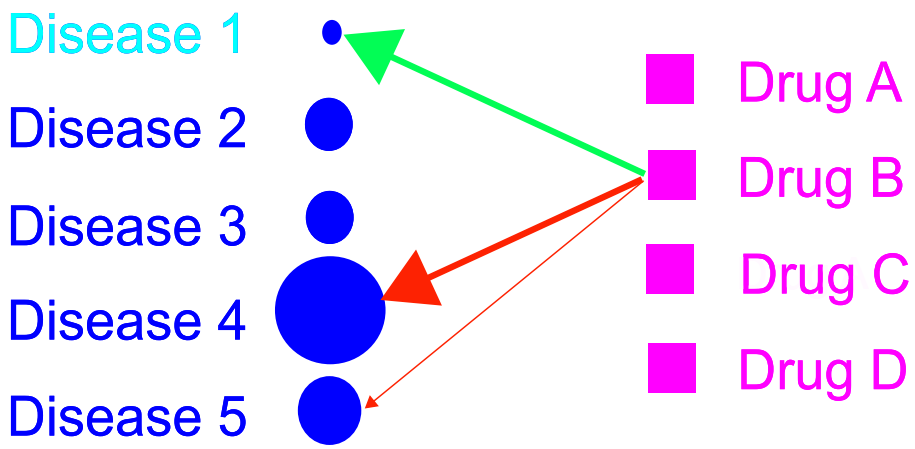
- genetic rank
- toxicogenetic rank
- pathway importance


r_i^{χ}
frequency
10%
1%
< 0.1%
T
4
2
0
0
2
4
M
4
2
0
G
T
0.2
0
0
(a)

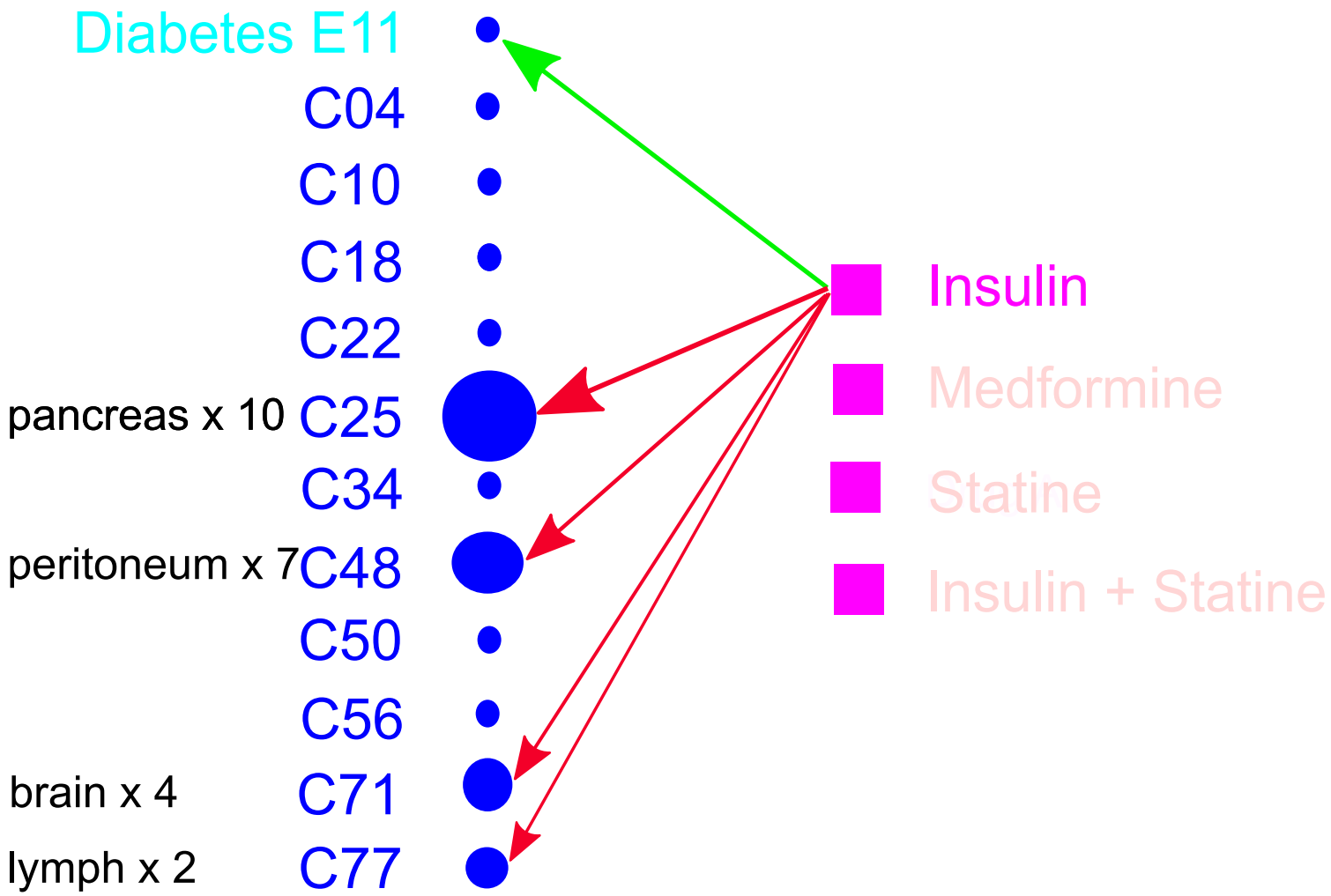
Side effects – personalized

Side effect networks

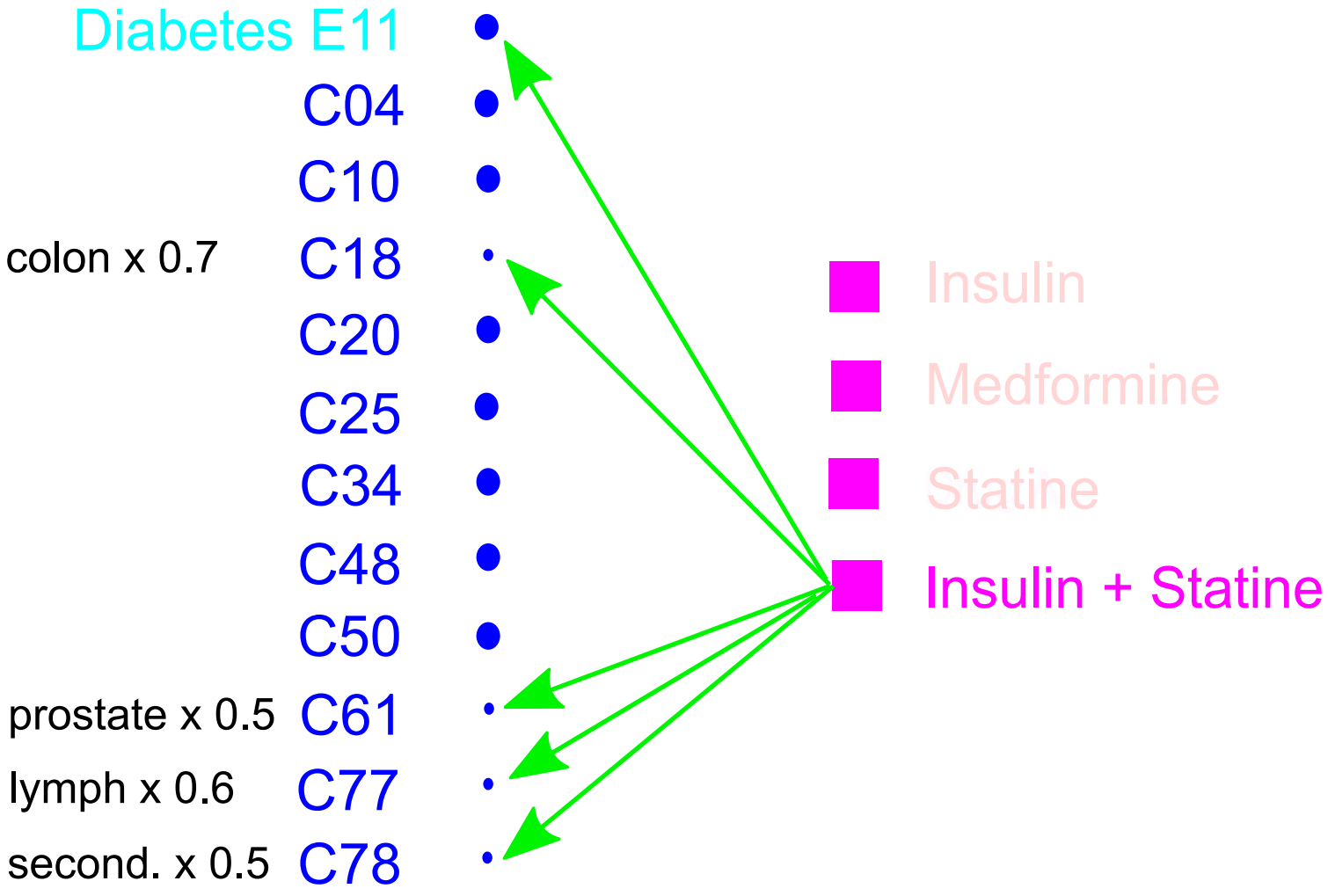
you have disease $x \rightarrow$ get medication $y \rightarrow y$ causes disease z



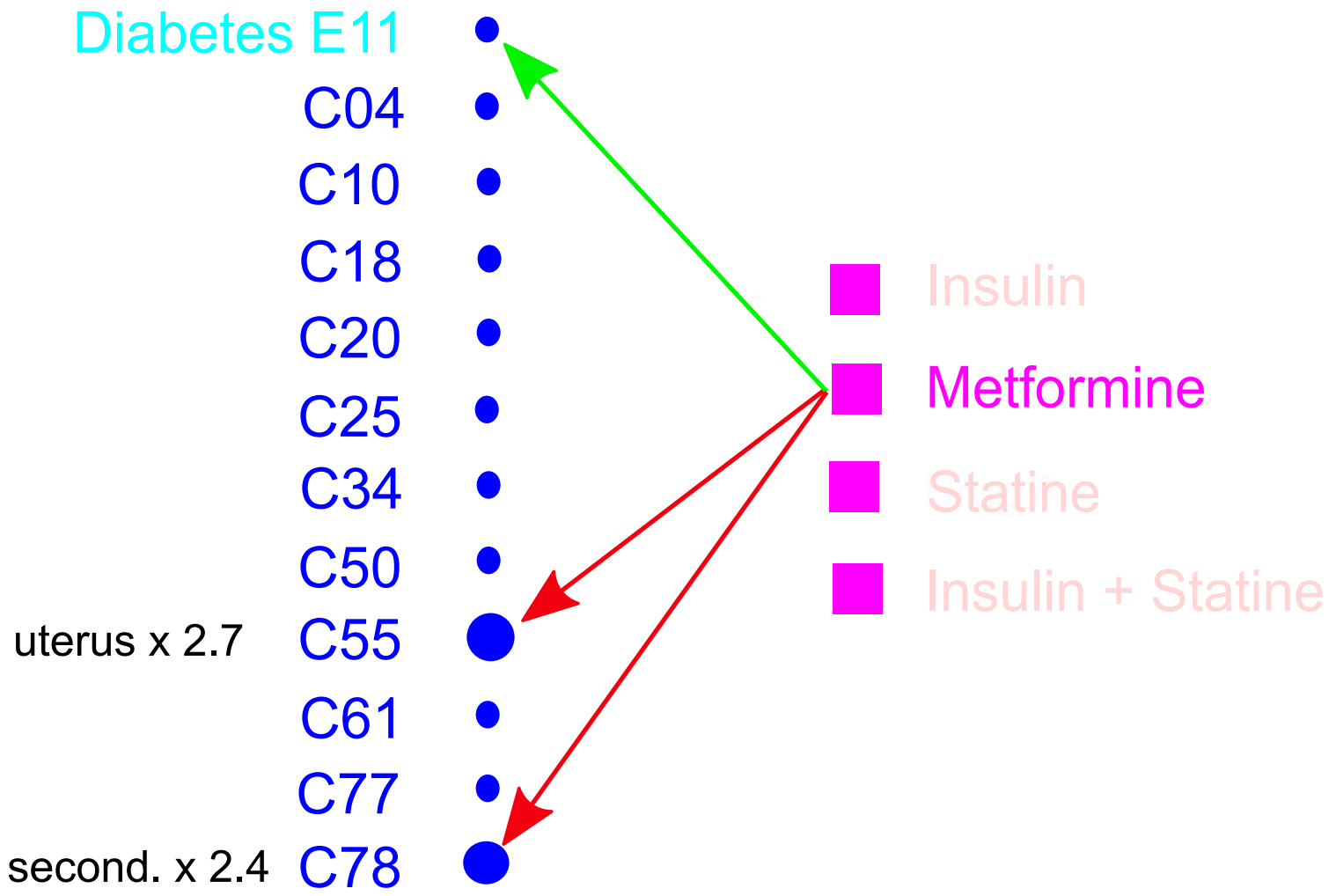
Side effects for diabetes treatments: Insulin



Side effects for treatment: Insulin+Statines



What are the side effects of Metformin?



Disclaimer

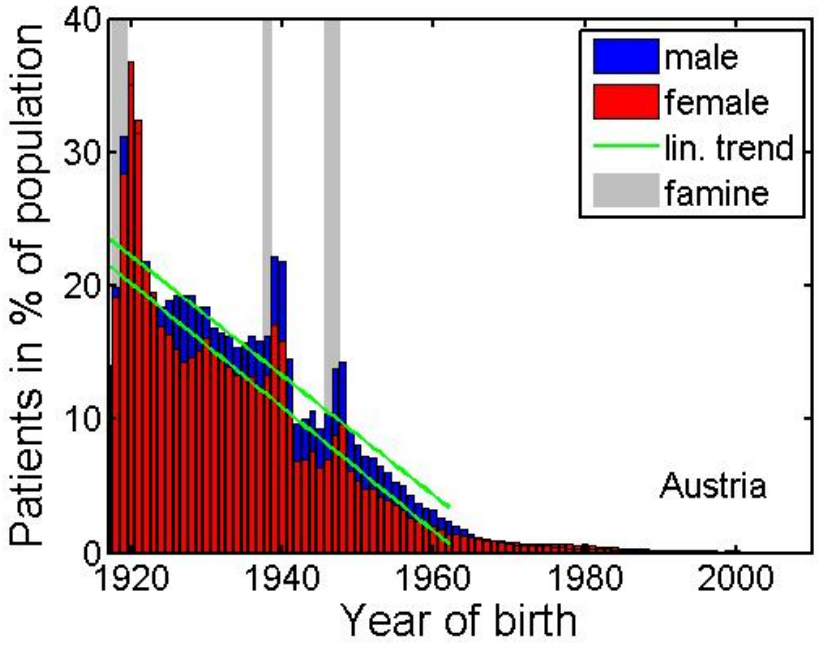
NO medical statement are made here!

- This reflects the *status quo* in the population only
- No understanding why
- No mechanism clarified
- No medical understanding
- Need experts for this

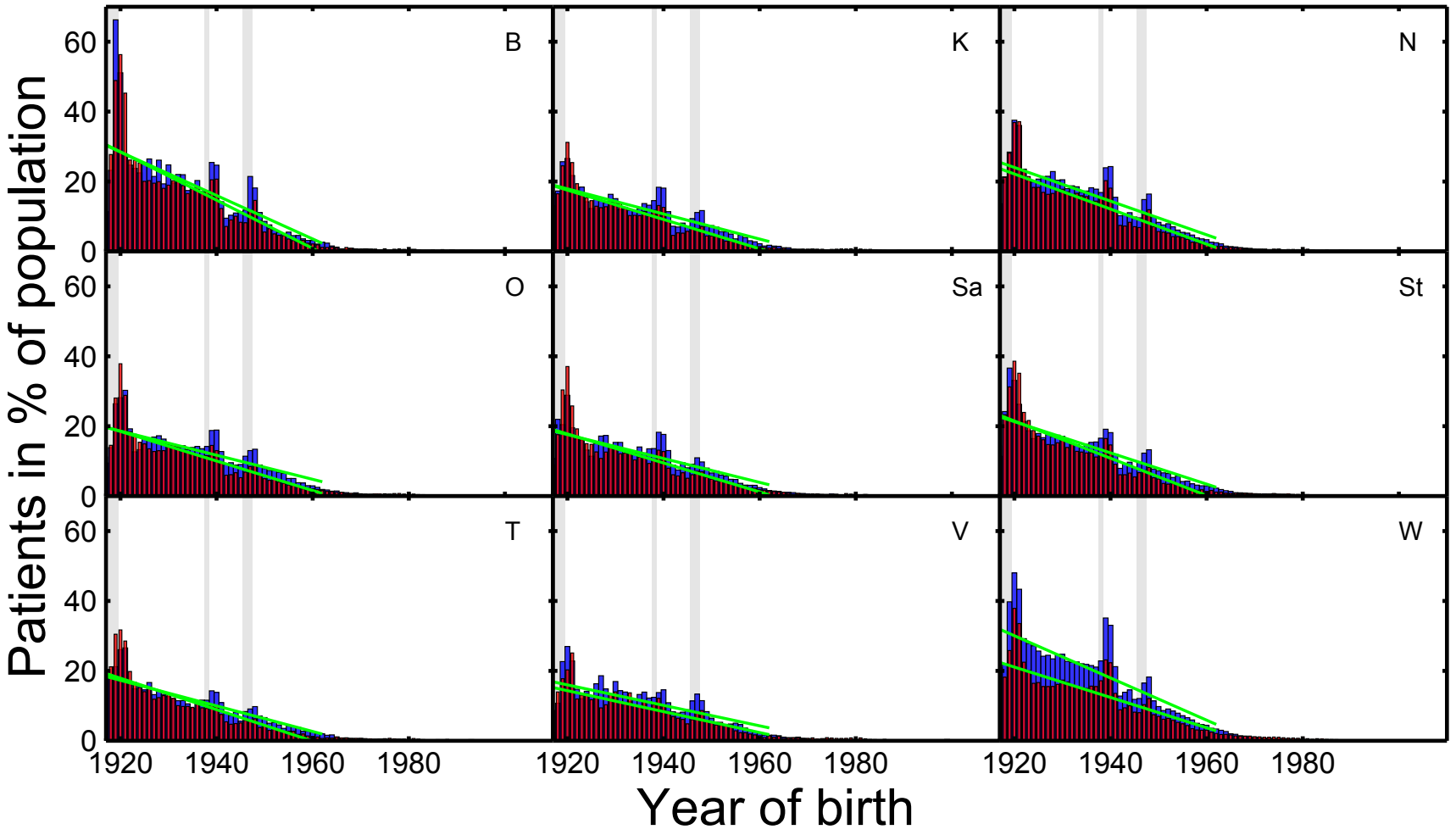
A telescope into the past

Unexpected causes for diabetes?

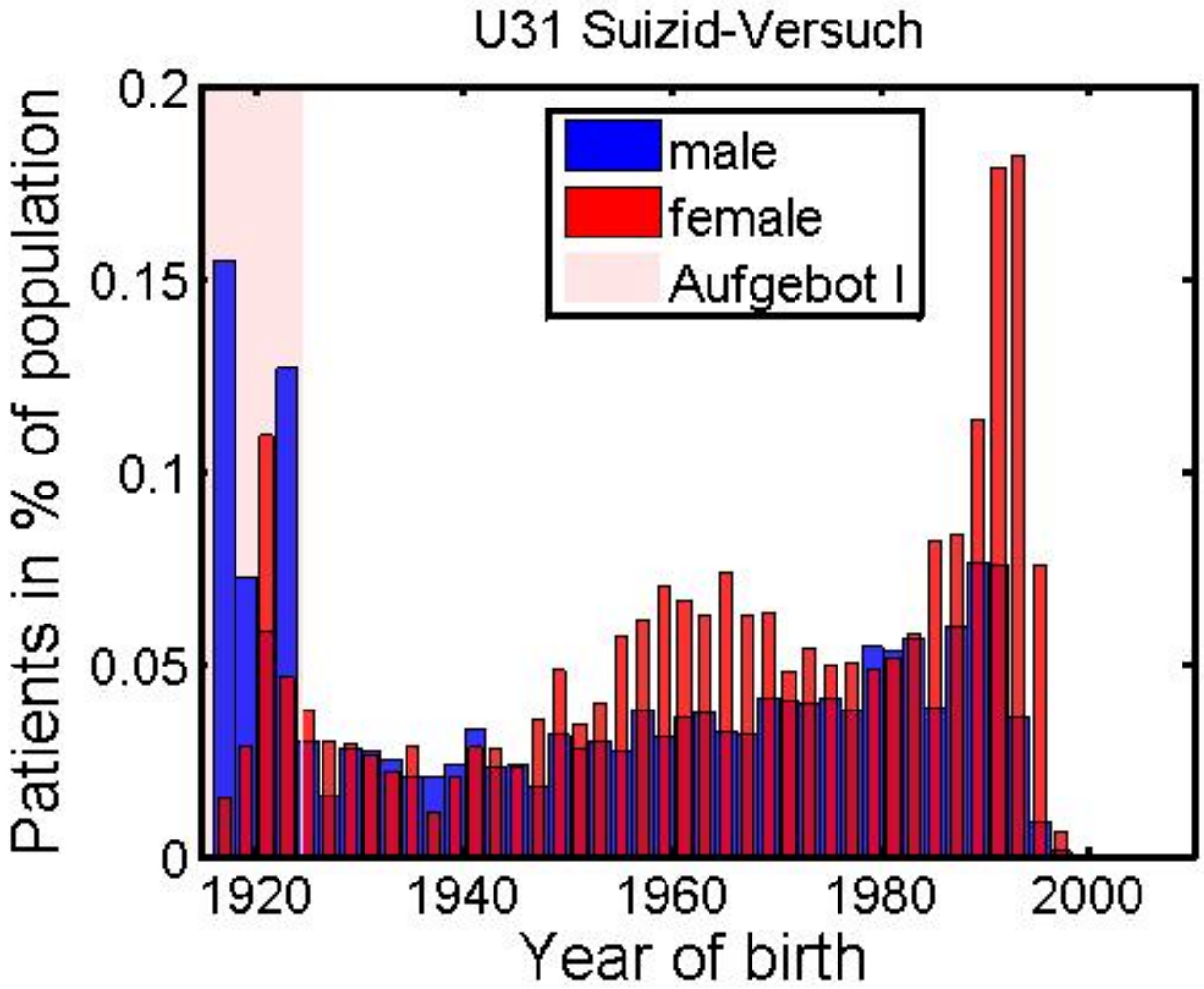
take all ∼300.000 diabetes patients. Fraction of patients in population given birth date? → **famines in Austria**



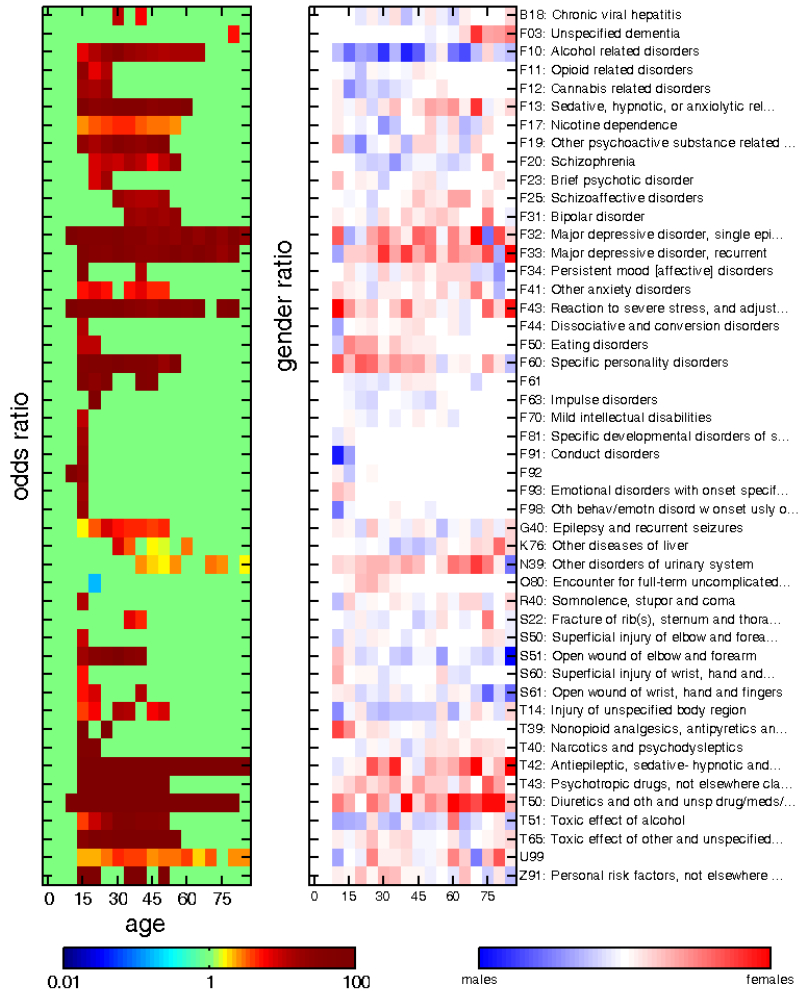
Message: mother suffers hunger in pregnancy → baby develops diabetes in later life S Thurner et al. PNAS 110, 4703-4707, (2013)

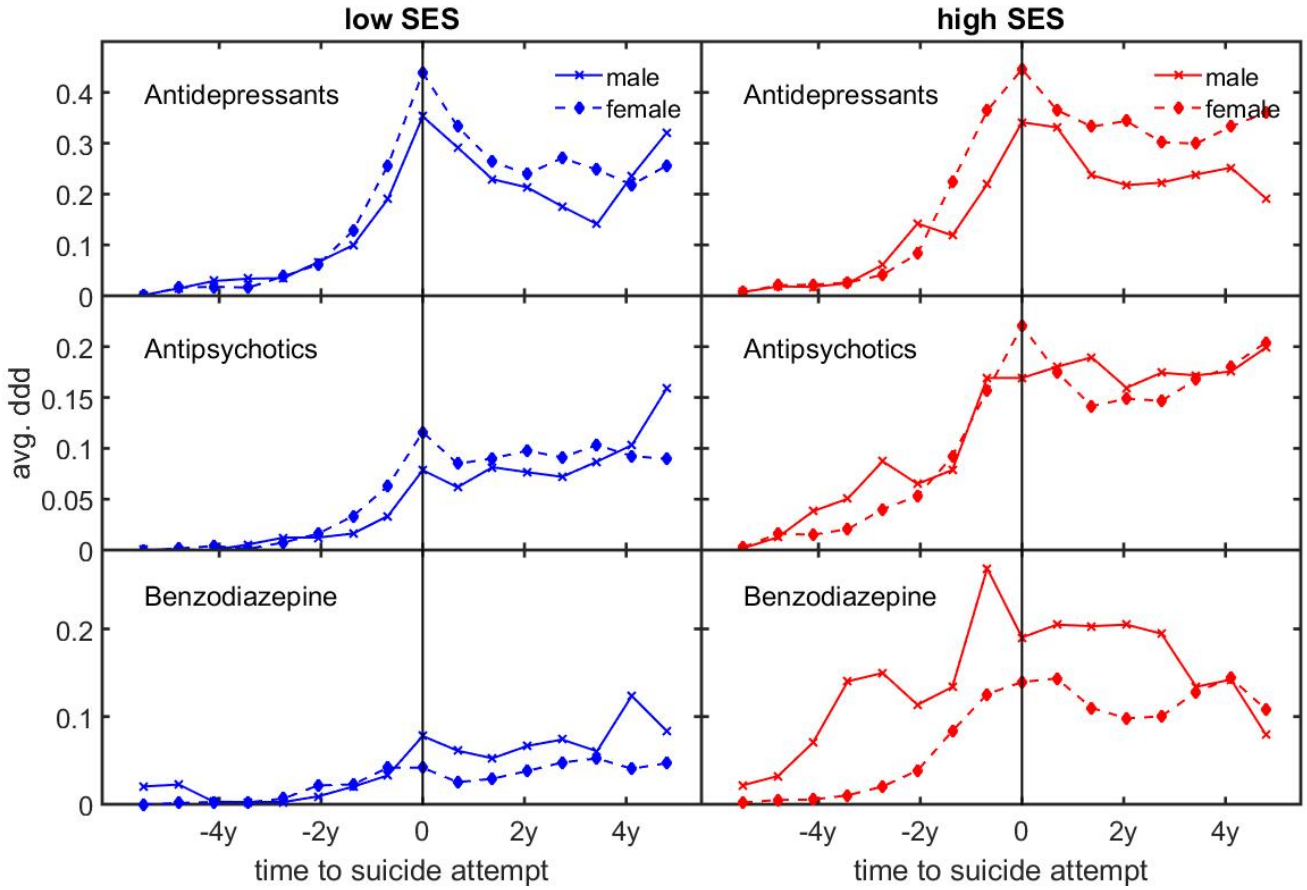

Patients in % of population
Year of birth
B
K
N
O
Sa
St
T
V
W
0
20
40
60
1920
1940
1960
1980

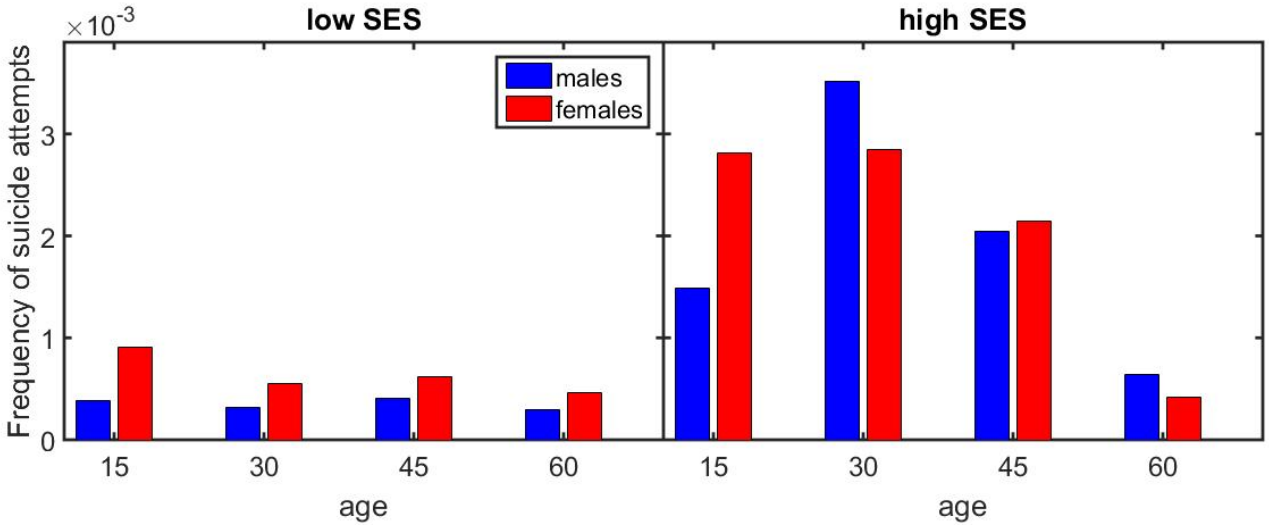
A window into the past II



A window into the past II




low SES
high SES
Antidepressants
Antipsychotics
Benzodiazepine
male
female
avg. ddd
0.4
0.3
0.2
0.1
0
0.2
0.15
0.1
0.05
0
0.2
0.1
0
-4y
-2y
0
2y
4y
time to suicide attempt


low SES
high SES
$\times 10^{-3}$
Frequency of suicide attempts
males
females
0
1
2
3
15
30
45
60
age

Vizualize healthcare system

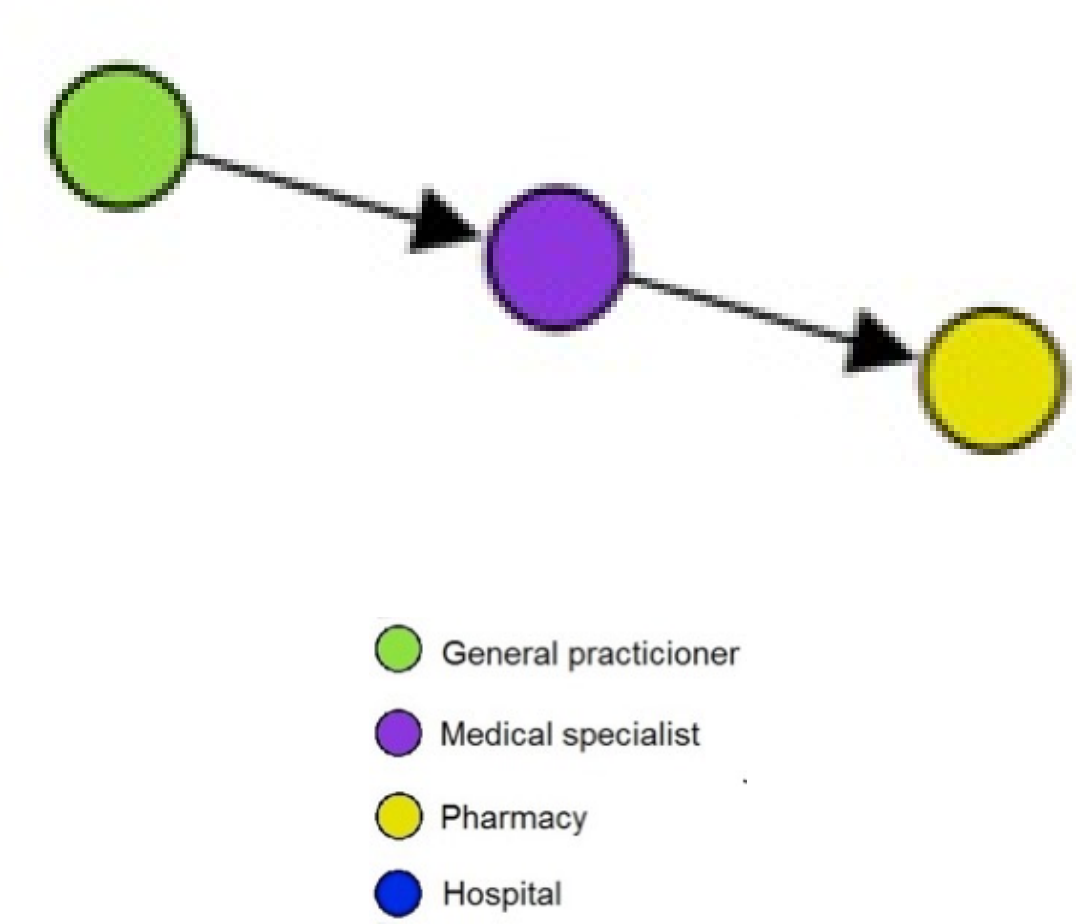
The healthcare system is ...

- network of patient flows
- network of information flows
- network of cash flows
- it is a co-evolving multi-layer network !

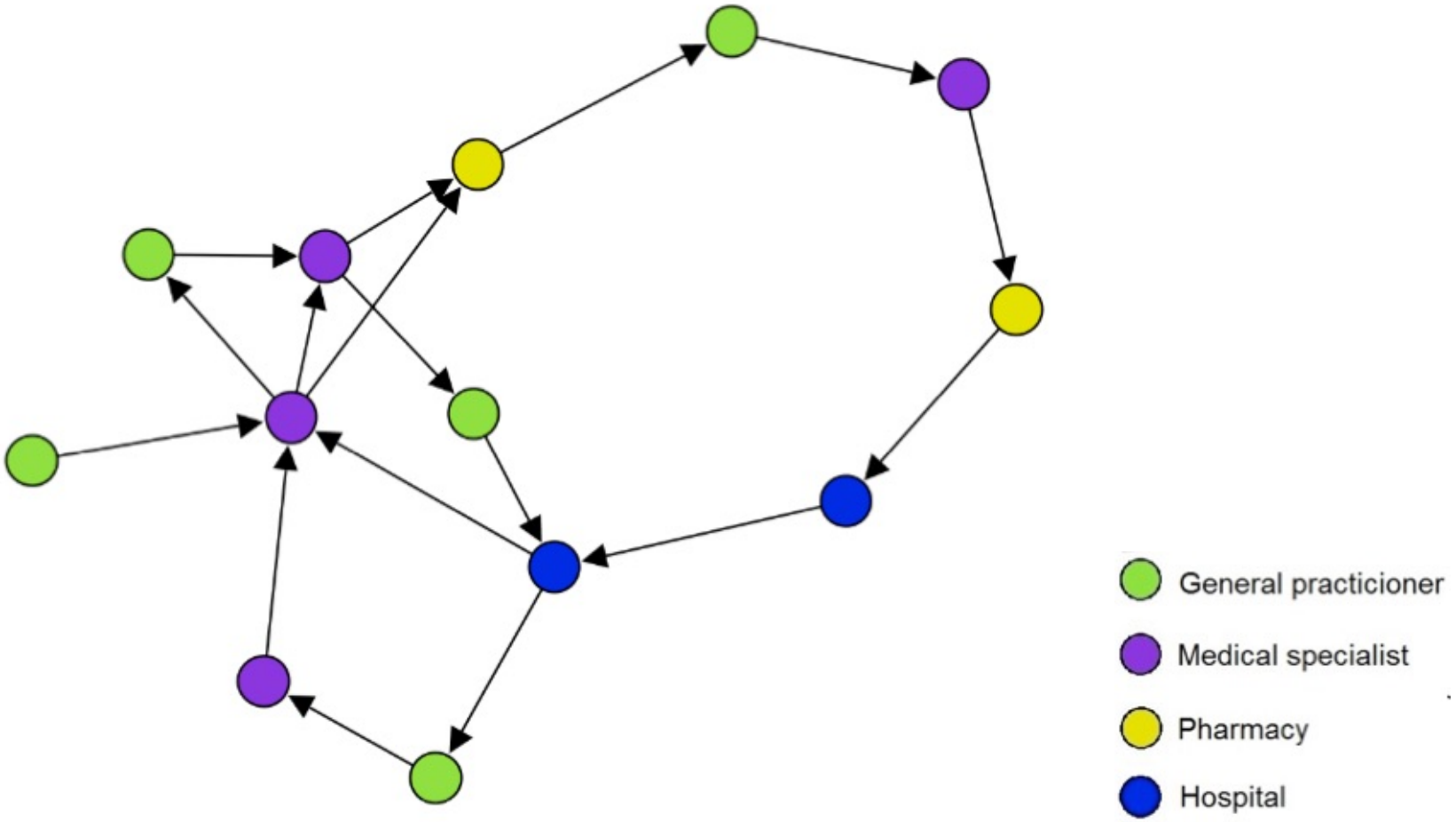
Patient-flow network

- many patient flows are medically reasonable – many are not
- health care costs can be completely transparent – if wanted
- patient flows + comorbidity across age $\rightarrow$ future costs

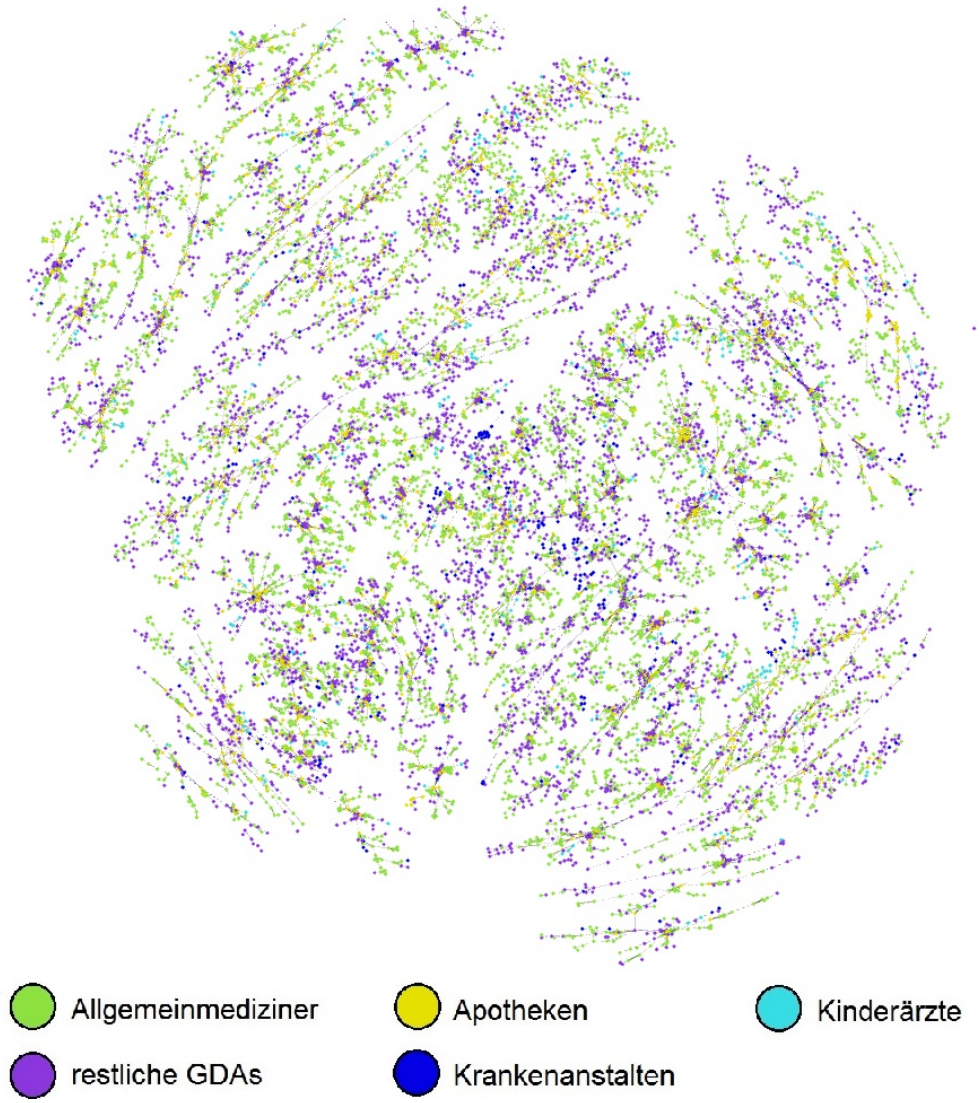
Patient-flow network as we like it



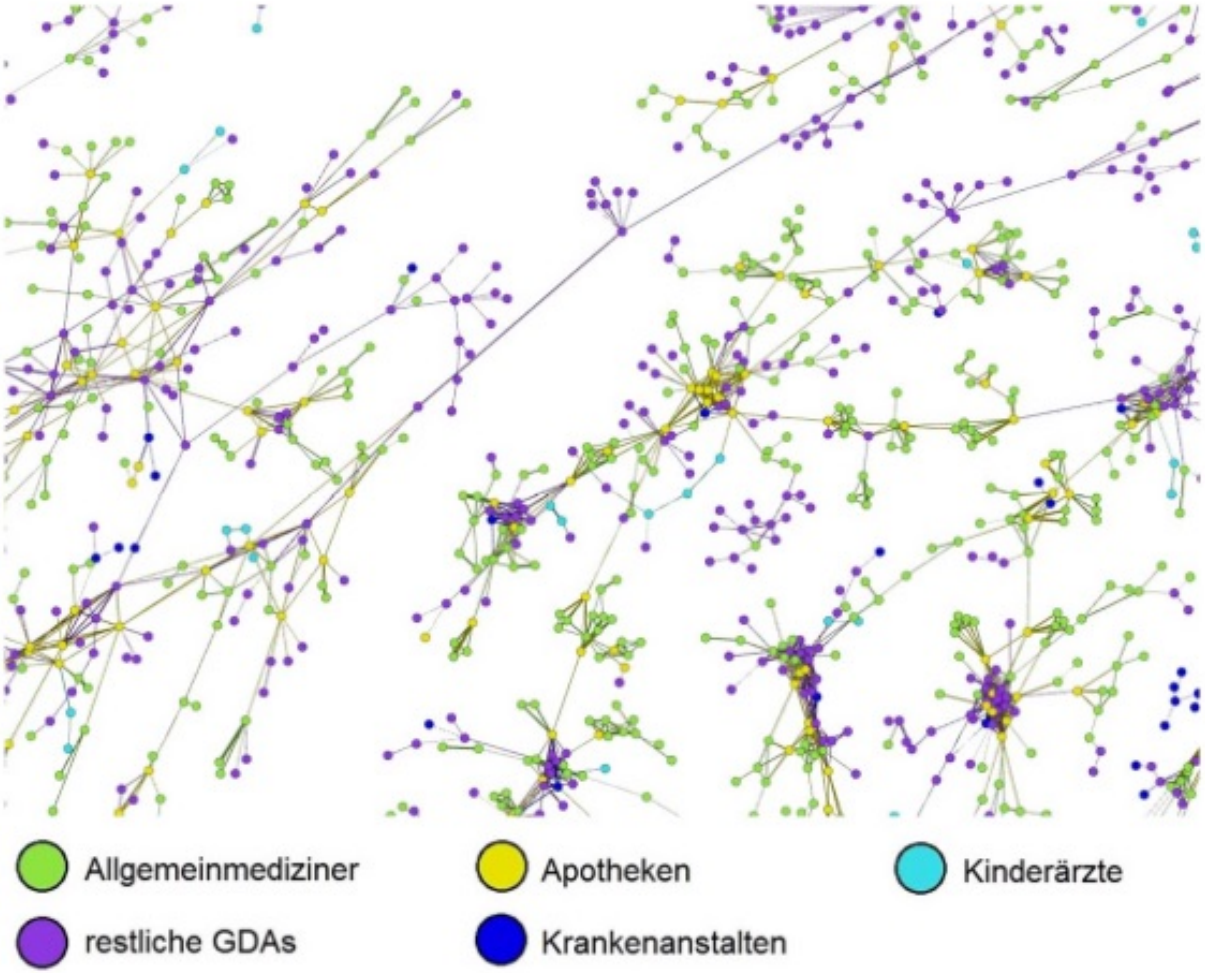
Patient-flow network



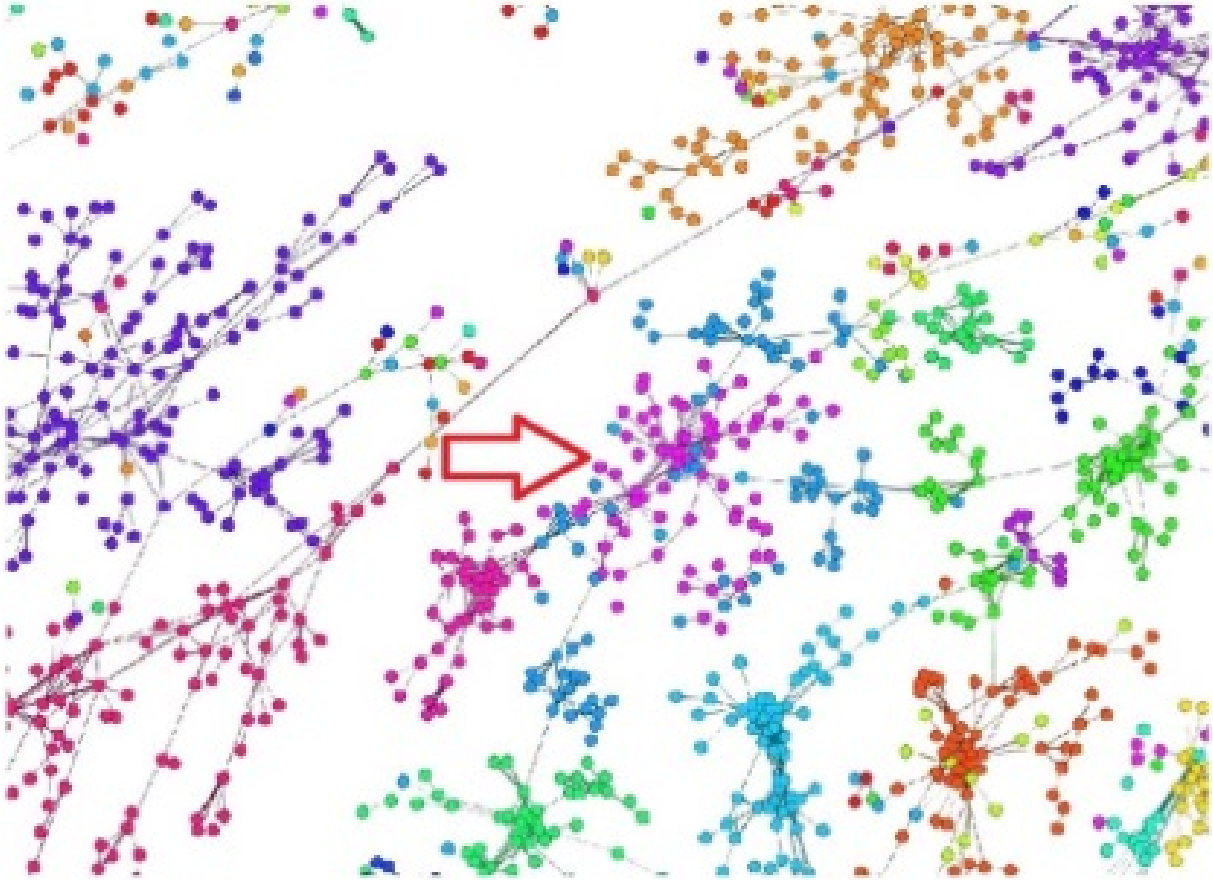
Patient-flow network of Austria 2006



Patient-flow network: patterns



Dorda W, et al. Analyse von Behandlungsnetzwerken (HVB, 2013)

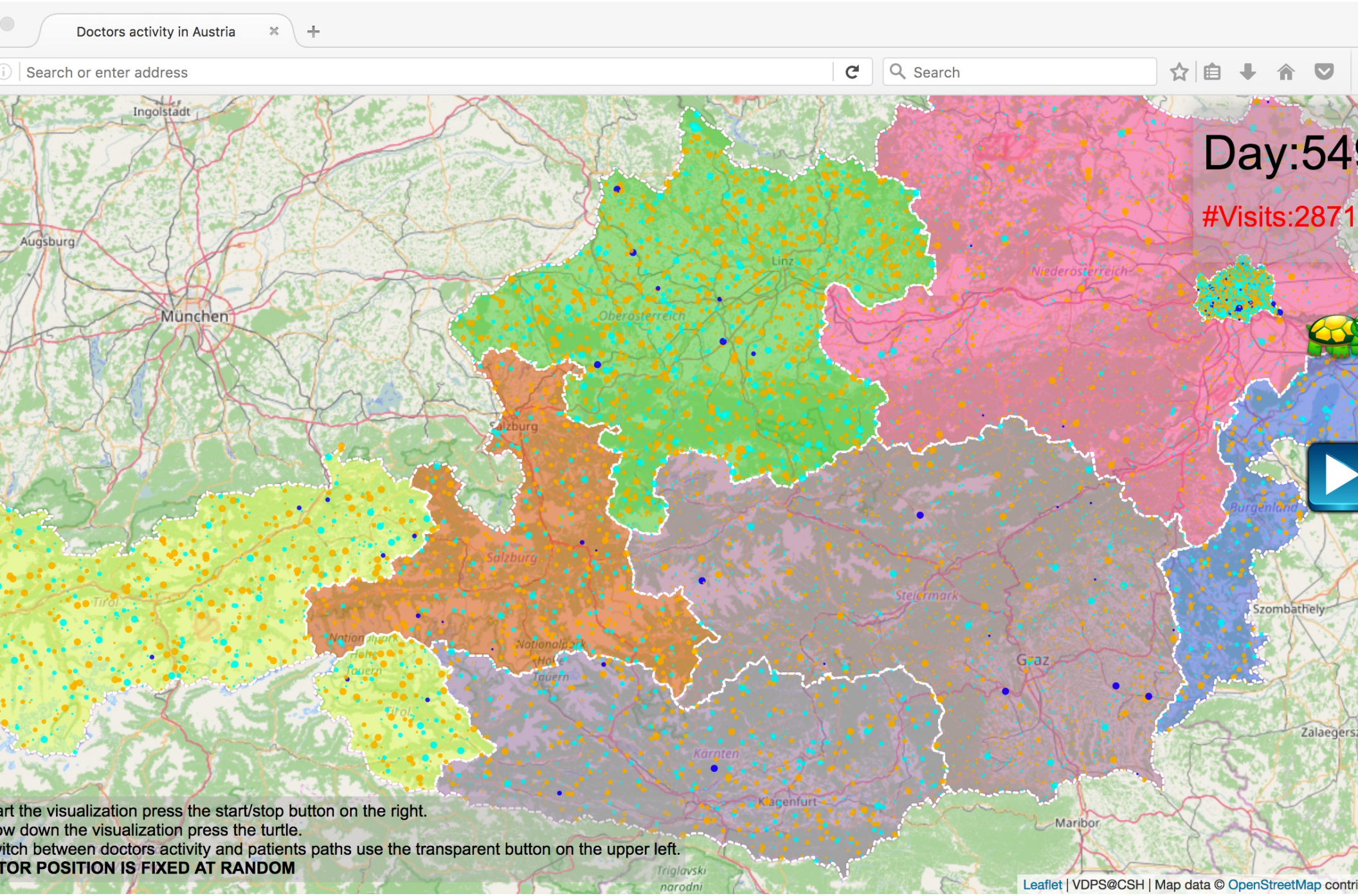


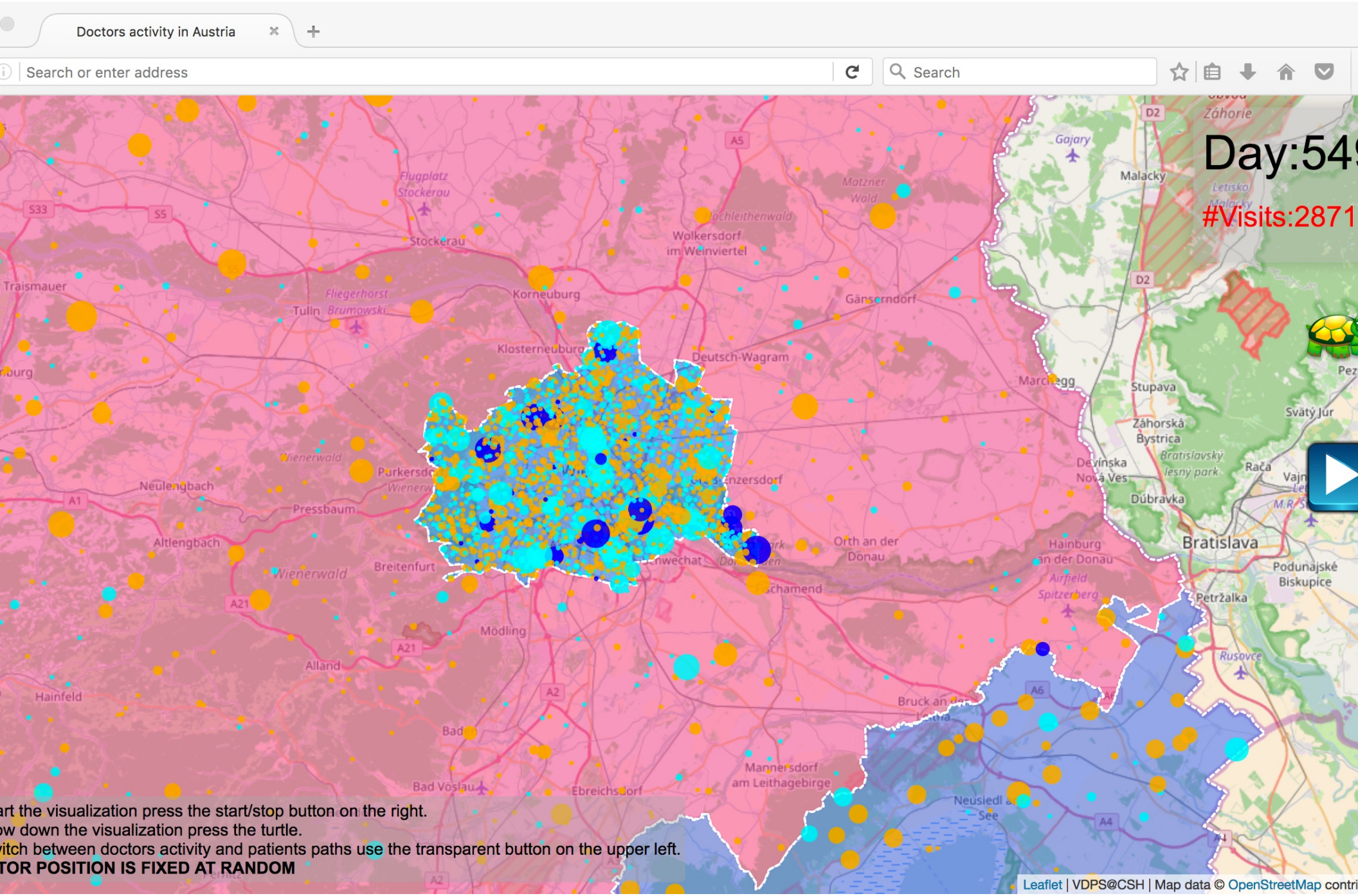
With this information one can monitor ...

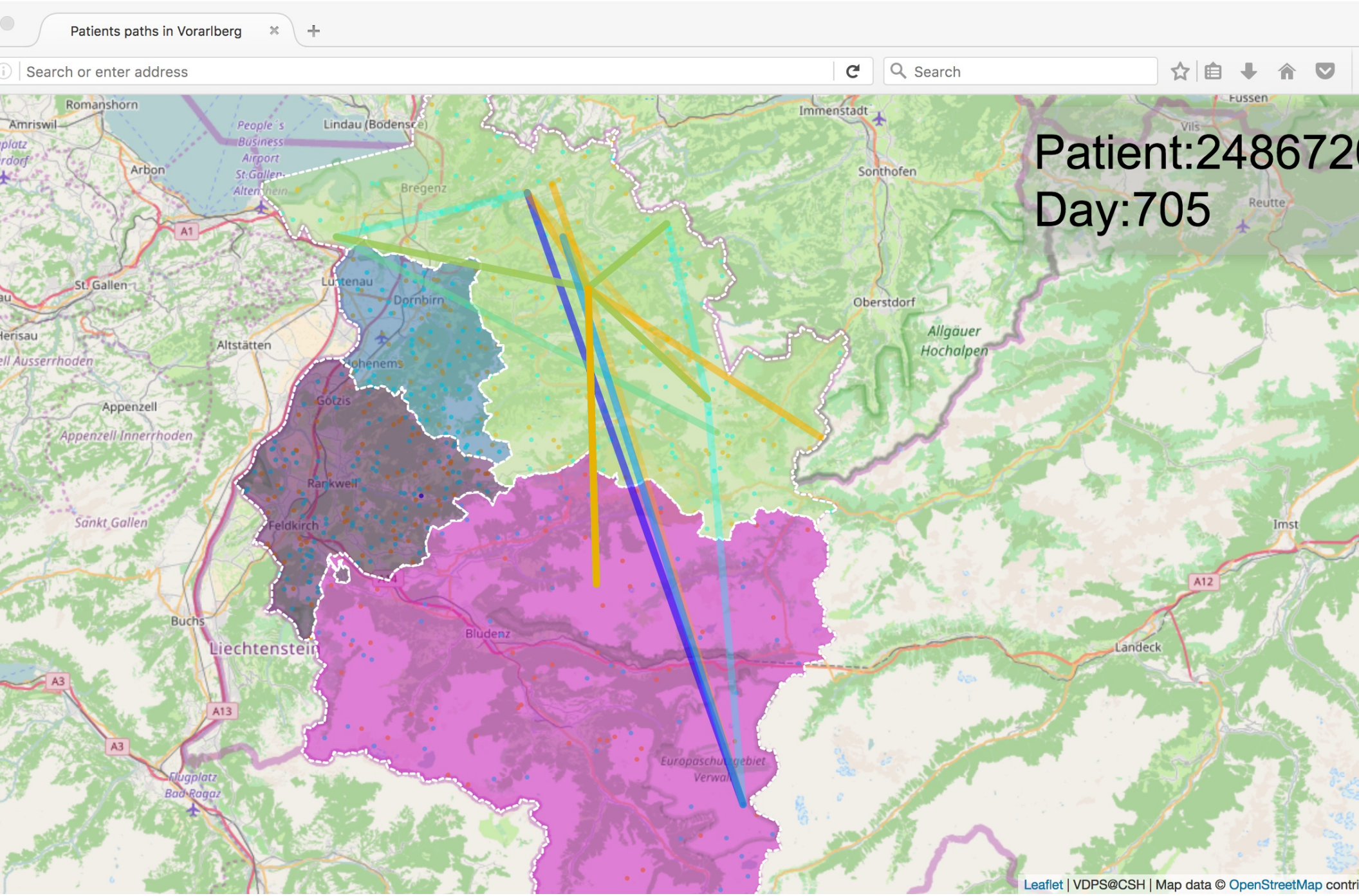
- regional differences in quality of care / prescriptions / success rates / diagnoses / costs / transparency /
- 'cyclic flows': which ones are medically necessary?
- success of prevention schemes (medical & economic)
- nation-wide behaviour of patients: drug use, vaccination rates
- visiting frequency as function of accessibility of HCP
- optimal health care coverage densities

Vision

- 1:1 agent based model of the Austrian health care system
- use data to fully calibrate the model
- estimate how individual patients take decisions
- estimate how HCP take decisions
- make policy experiments: insurers and politicians


Doctors activity in Austria
Search or enter address
Search
Day:54
#Visits:2871
Ingolstadt
Augsburg
München
Linz
Oberösterreich
Niederösterreich
Salzburg
Tirol
Steiermark
Burgenland
Szombathely
Graz
Kärnten
Klagenfurt
Maribor
art the visualization press the start/stop button on the right.
ow down the visualization press the turtle.
witch between doctors activity and patients paths use the transparent button on the upper left.
TOR POSITION IS FIXED AT RANDOM
Leaflet | VDPS@CSH | Map data © OpenStreetMap contri


Doctors activity in Austria
Search or enter address
Search
Day:54
#Visits:2871
art the visualization press the start/stop button on the right.
ow down the visualization press the turtle.
vitch between doctors activity and patients paths use the transparent button on the upper left.
TOR POSITION IS FIXED AT RANDOM
Leaflet | VDPS@CSH | Map data © OpenStreetMap contri


Patients paths in Vorarlberg
Search or enter address
Search
Patient:248672
Day:705
Romanshorn
Amriswil
Arbon
Lindau (Bodensee)
Bregenz
Immenstadt
Sonthofen
Oberstdorf
Allgäuer Hochalpen
Reutte
Füssen
Vils
St. Gallen
Dornbirn
Altstätten
Appenzell
Appenzell Innerrhoden
Götzis
Rankweil
Feldkirch
Sankt Gallen
Buchs
Liechtenstein
Bludenz
Landeck
Imst
A1
A3
A13
A12
Flugplatz Bad Ragaz
Leaflet | VDPS@CSH | Map data © OpenStreetMap contri

Summary

- begin to predict health trajectories from co-morbidity networks
- see which medication works
- make gender differences visible
- compute personalized side effects
- new classification of diseases in terms of co-morbidity
- compute medical and economical value of prevention schemes
- quantify resilience, robustness, sustainability of health care system

Collaboration partners – experts

Alexandra Kautzky-Willer, MUW
Gottfried Endel, Hauptverband
Miriam Leitner, MUW
Irmgard Schiller-Frühwirth, Hauptverband
Herwig Ostermann, Gesundheit Österreich
Klaus Kratochwill, MUW

Statistics

- Phenotype NW: ϕ is correlation coefficient (binary), Kramers coefficient

- Relative Risk:

$$\frac{a/(a+c)}{b/(b+d)}$$

- Odds Ratio:

$$\frac{P(A)(1-P(B))}{P(B)(1-P(A))}$$

where $P(A) = a/(a+c)$ and $P(B) = b/(b+d)$