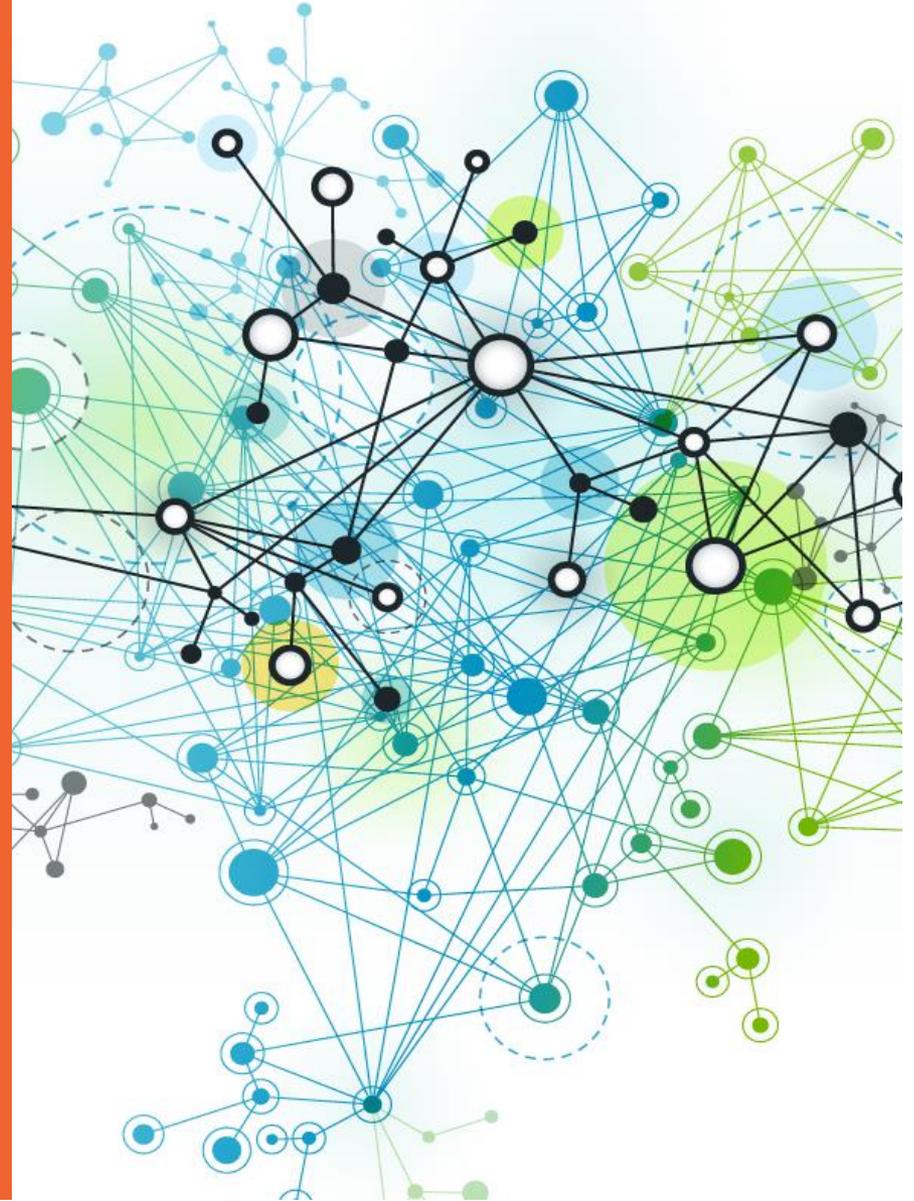


The logo for Nébih, featuring the word "nébih" in a lowercase, sans-serif font. The letters "n" and "i" are black, while "é" and "h" are green. There are two small green circles above the "é" and one above the "h".

nébih

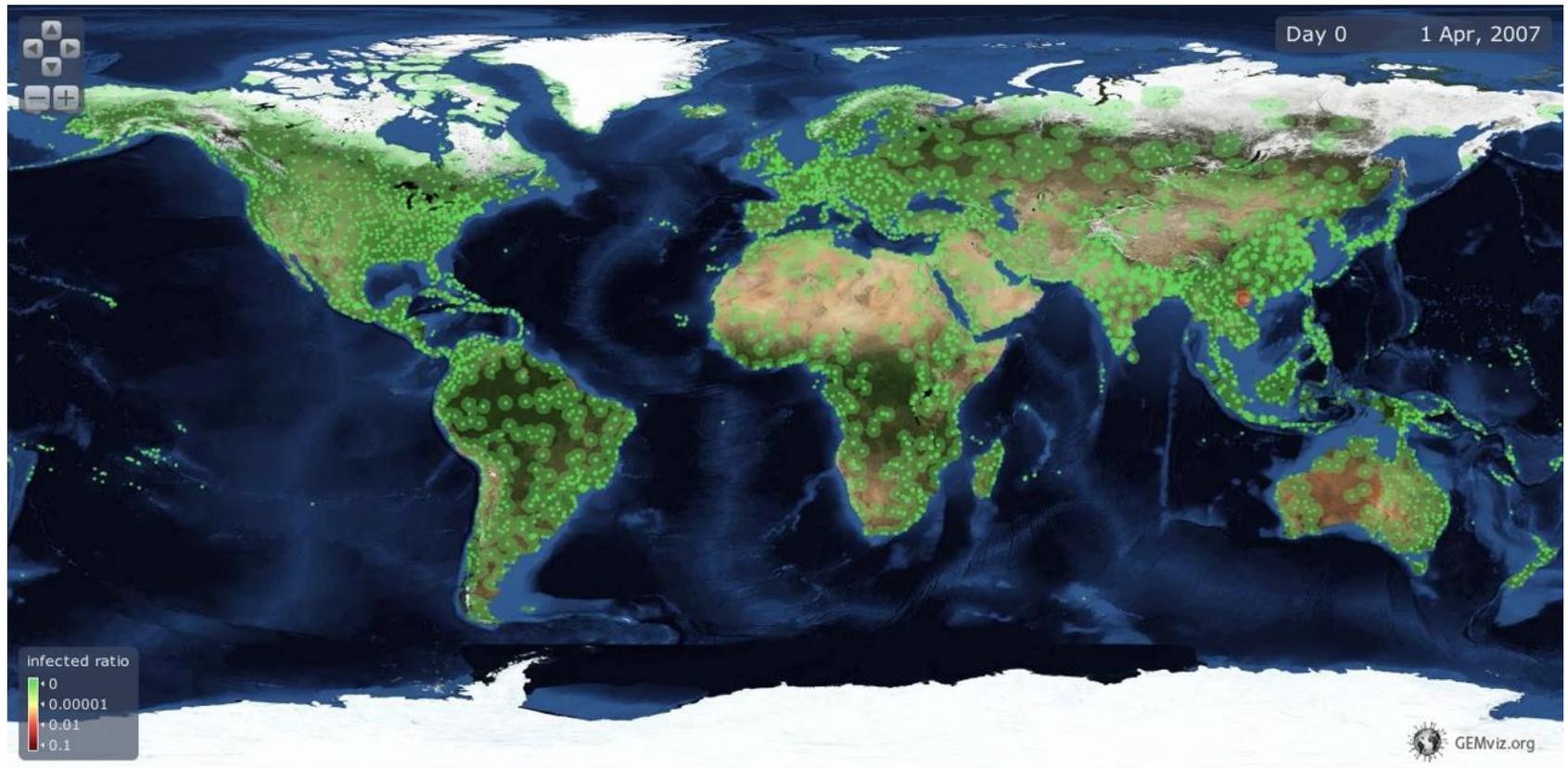
Role of **Data Science** in Food Chain Safety Decision Making

**Ákos Józwiak**  
NÉBIH, Hungary



# Food chain...

...is a **complex embedded network** of different entities and/or processes

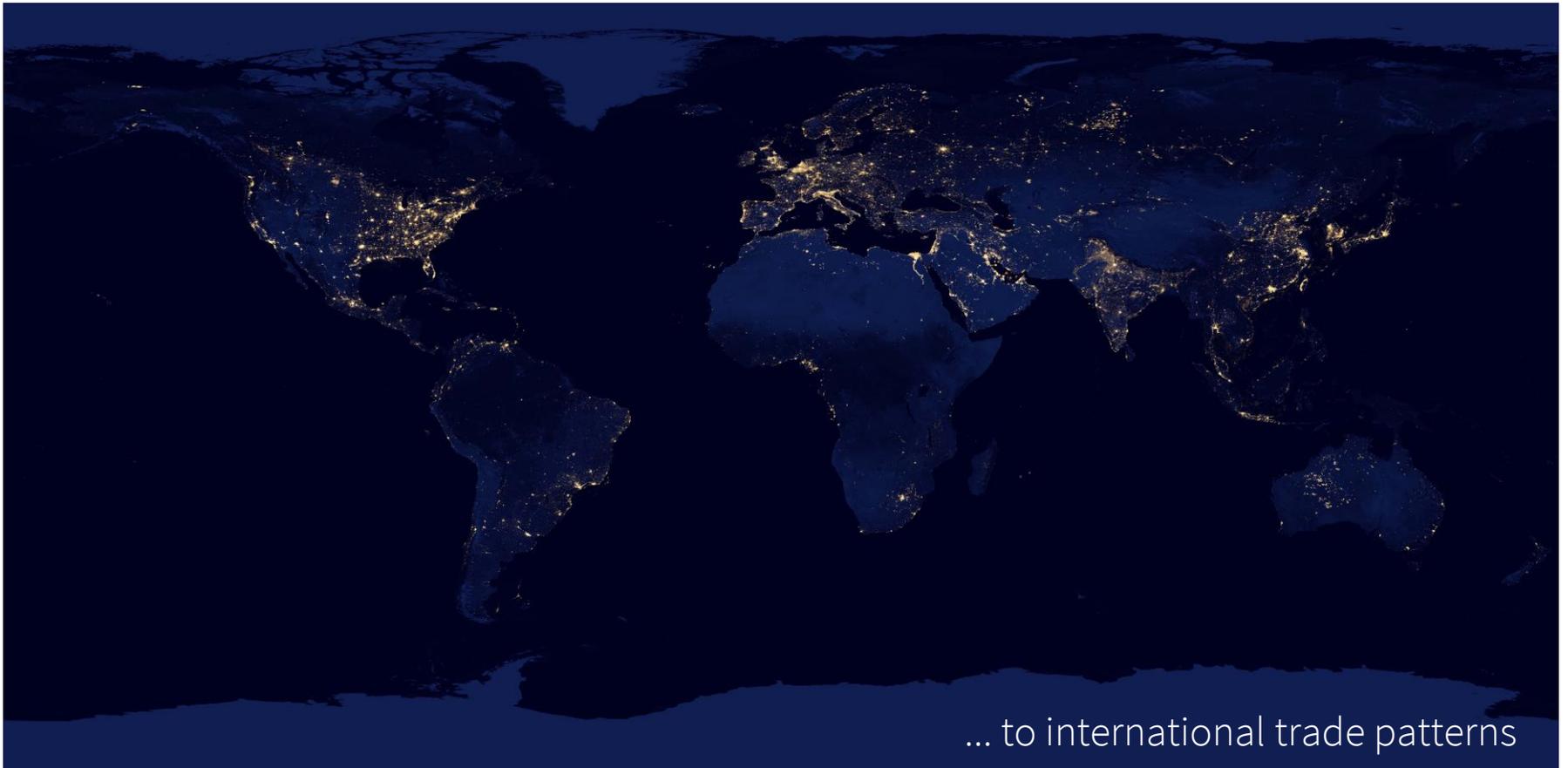


Source: [www.gleamviz.org](http://www.gleamviz.org)

# Food chain...

...overarches many orders of magnitude in size:

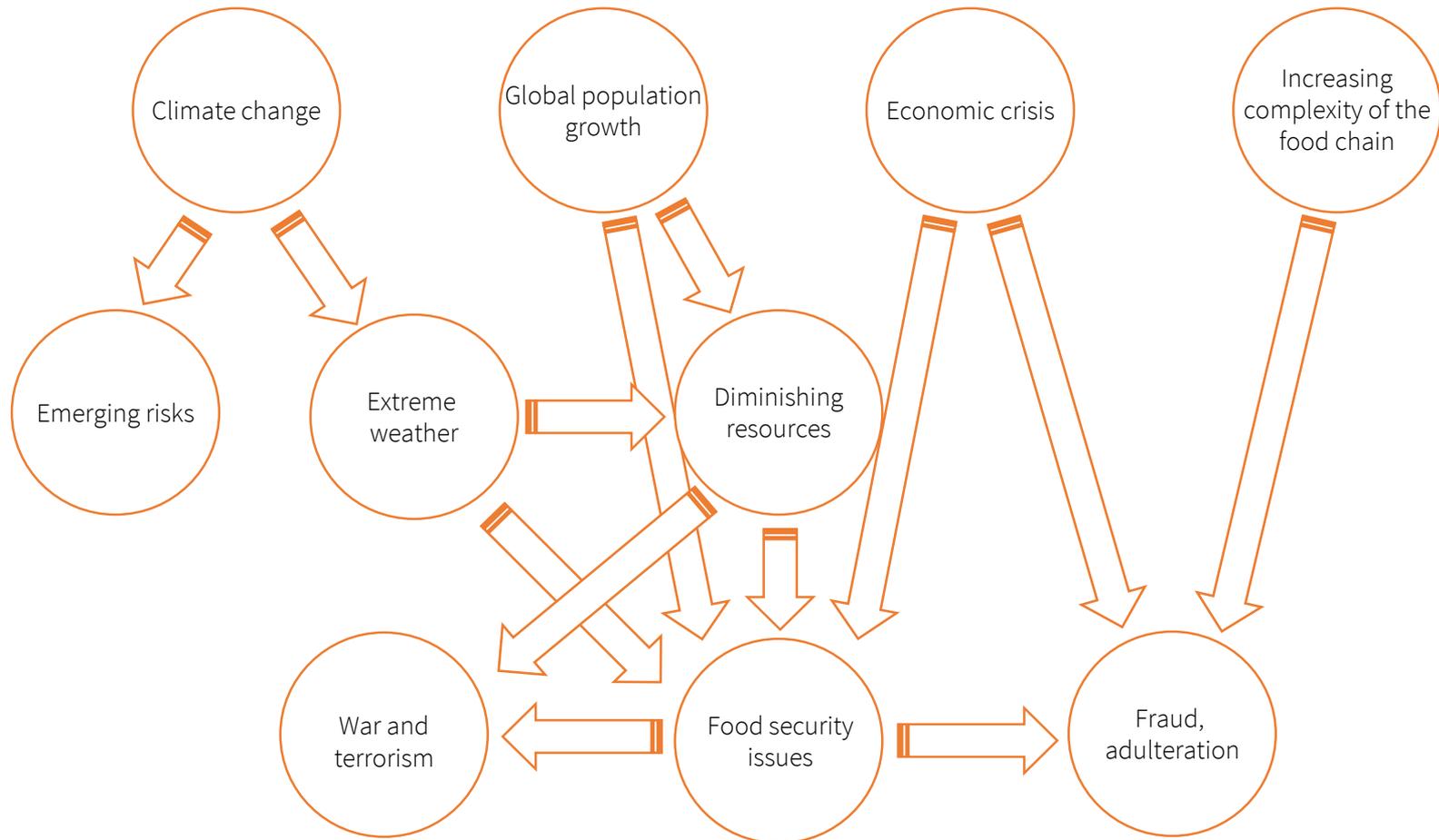
From nano-particles...



... to international trade patterns

# Food chain...

...is influenced by many **complex drivers**

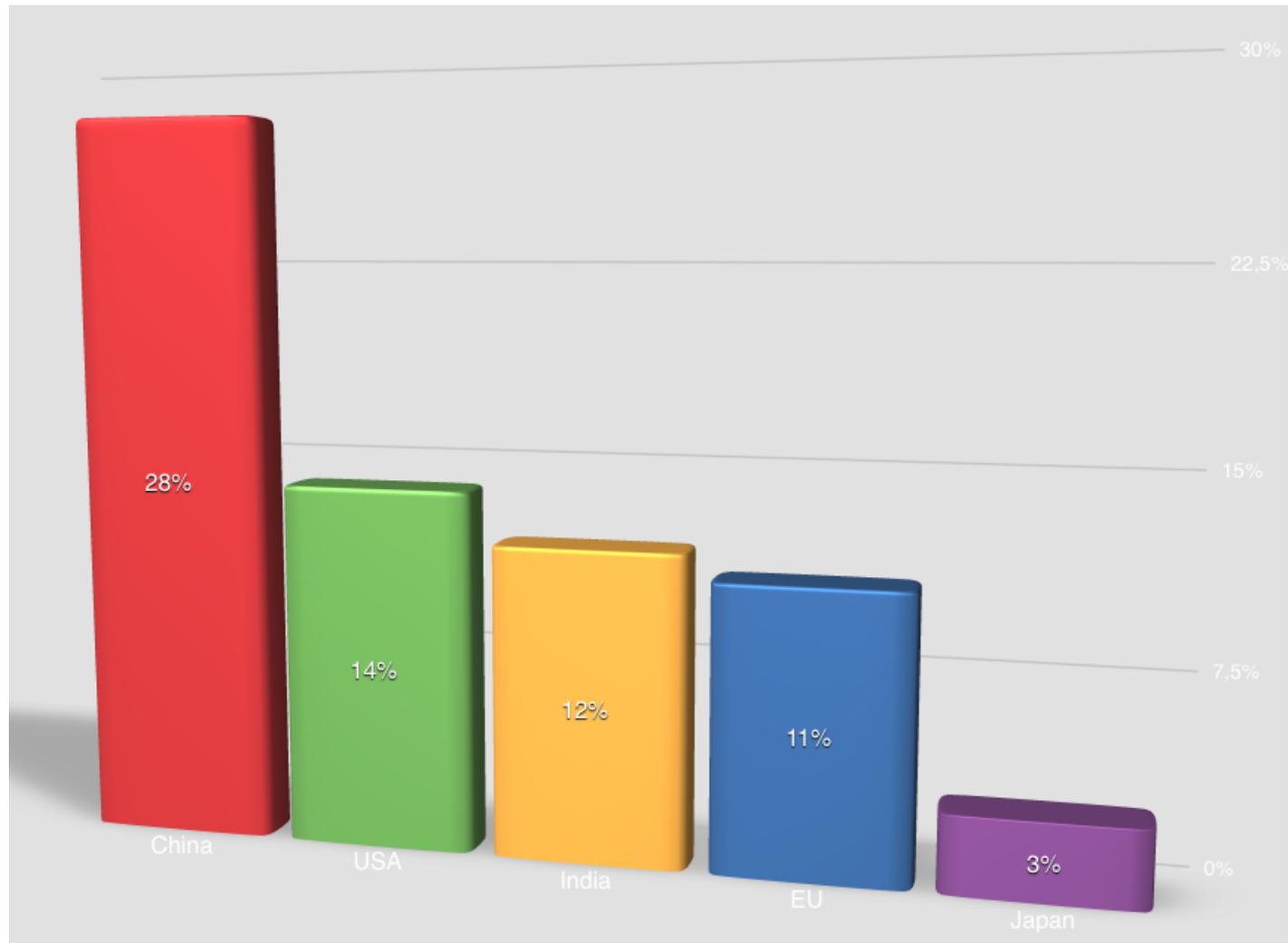




Increasing  
**volume & complexity**  
of the food chain

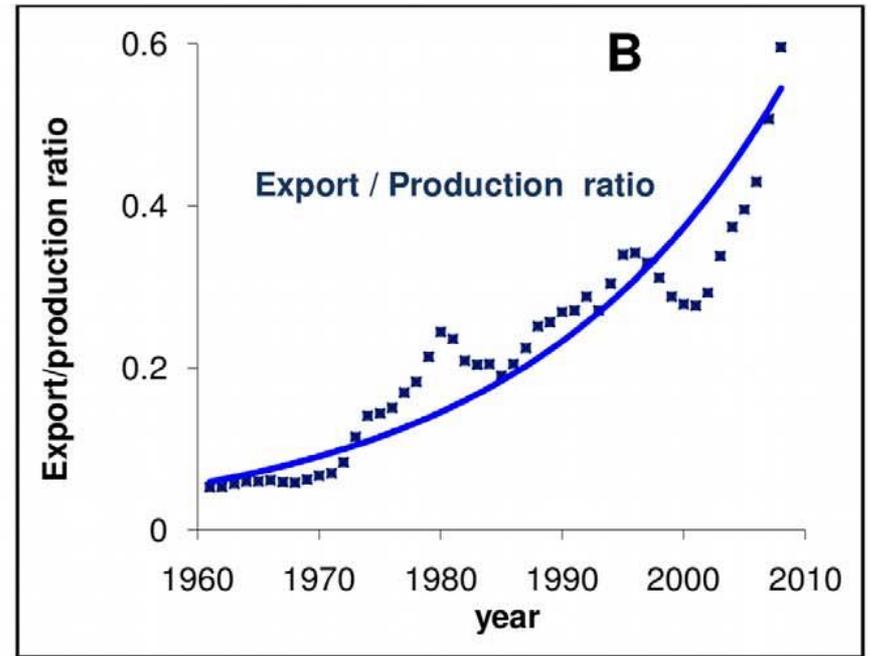
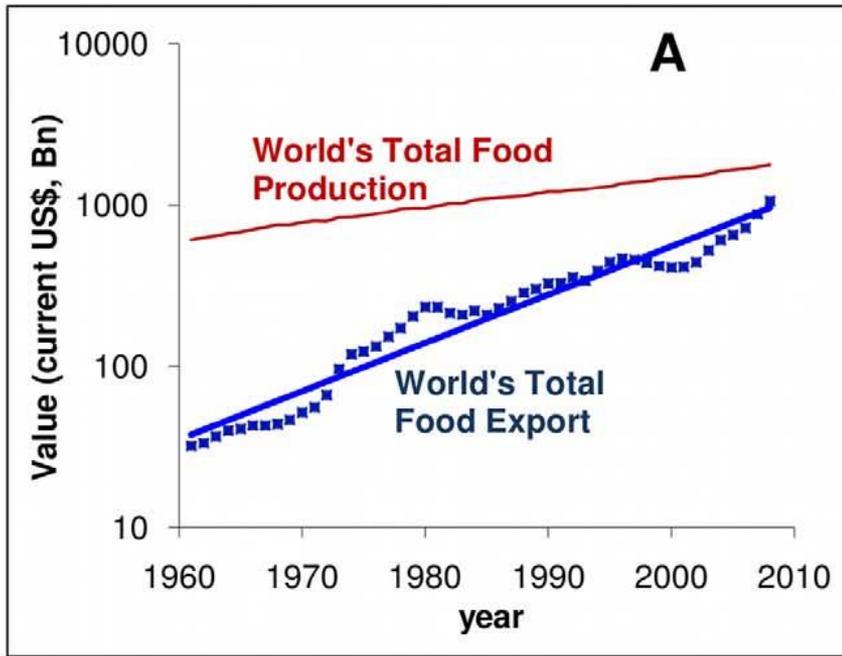


# Predicted share of global economy by 2050

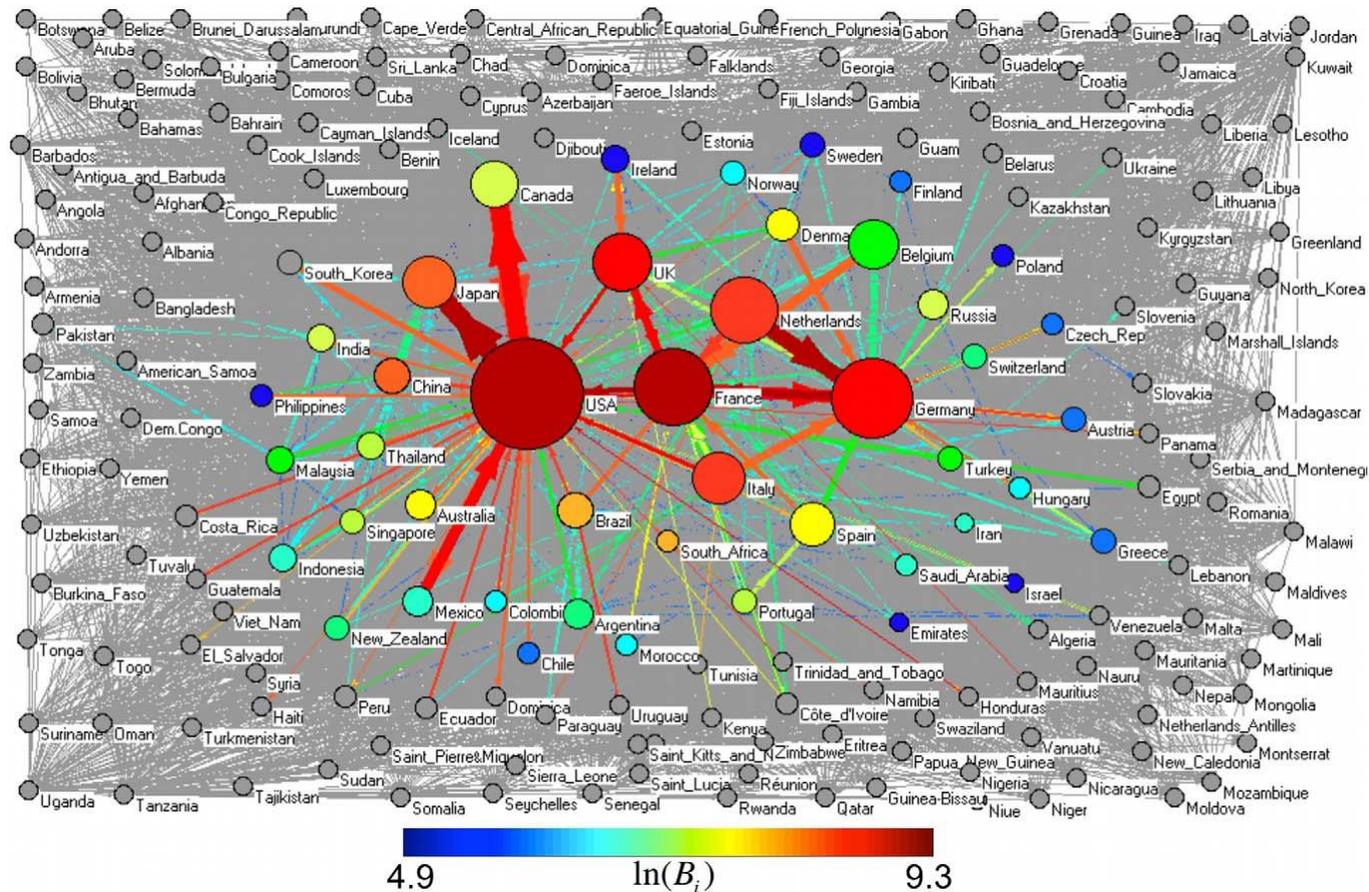


Source: Fouré, Jean, Agnès Bénassy-Quéré and Lionel Fontagné: The World Economy in 2050: a Tentative Picture, Paris, 2010 .

# The world's food trade grows faster than the food production



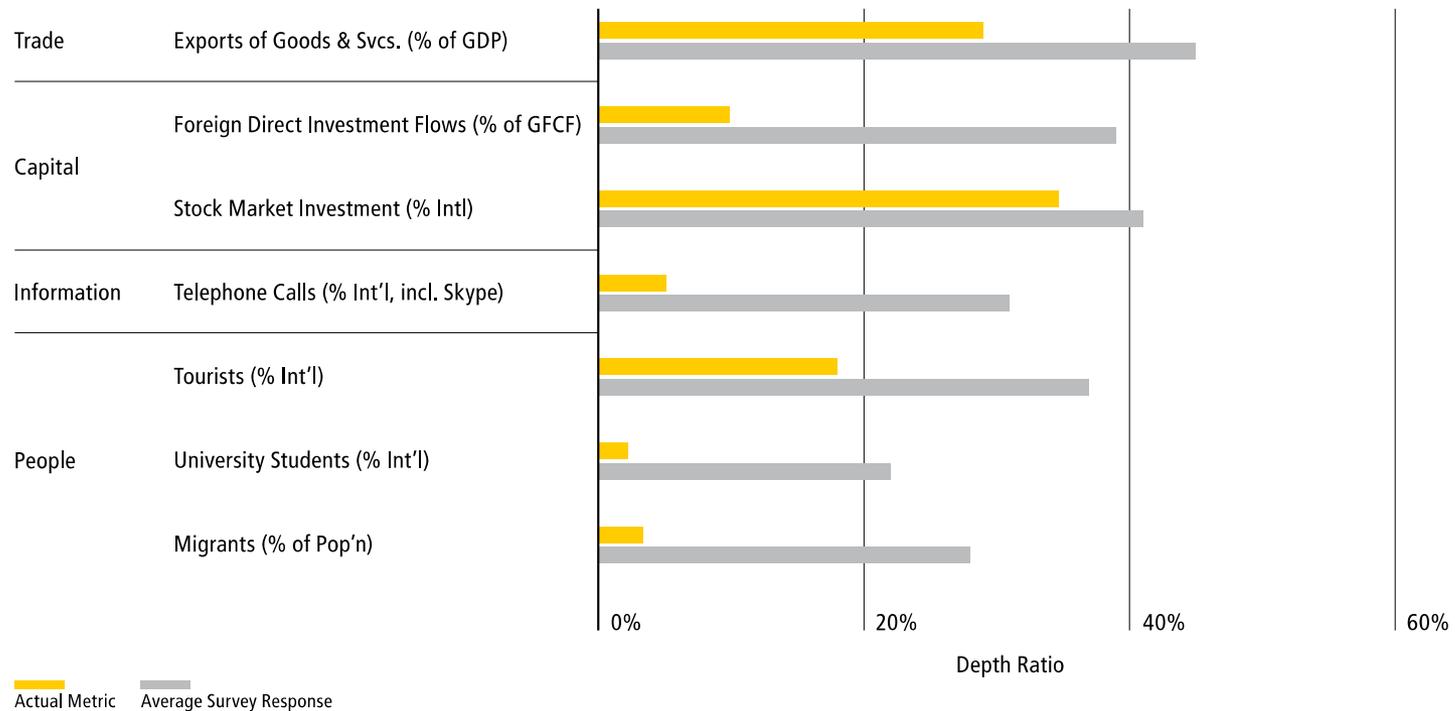
7 countries form the core of the agri-food trade network, each trading with over 77% of all the countries in the world



Source: Ercsey-Ravasz M, Toroczka Z, Lakner Z, Baranyi J (2012) Complexity of the International Agro-Food Trade Network and Its Impact on Food Safety. PLoS ONE 7(5): e37810. doi:10.1371/journal.pone.0037810

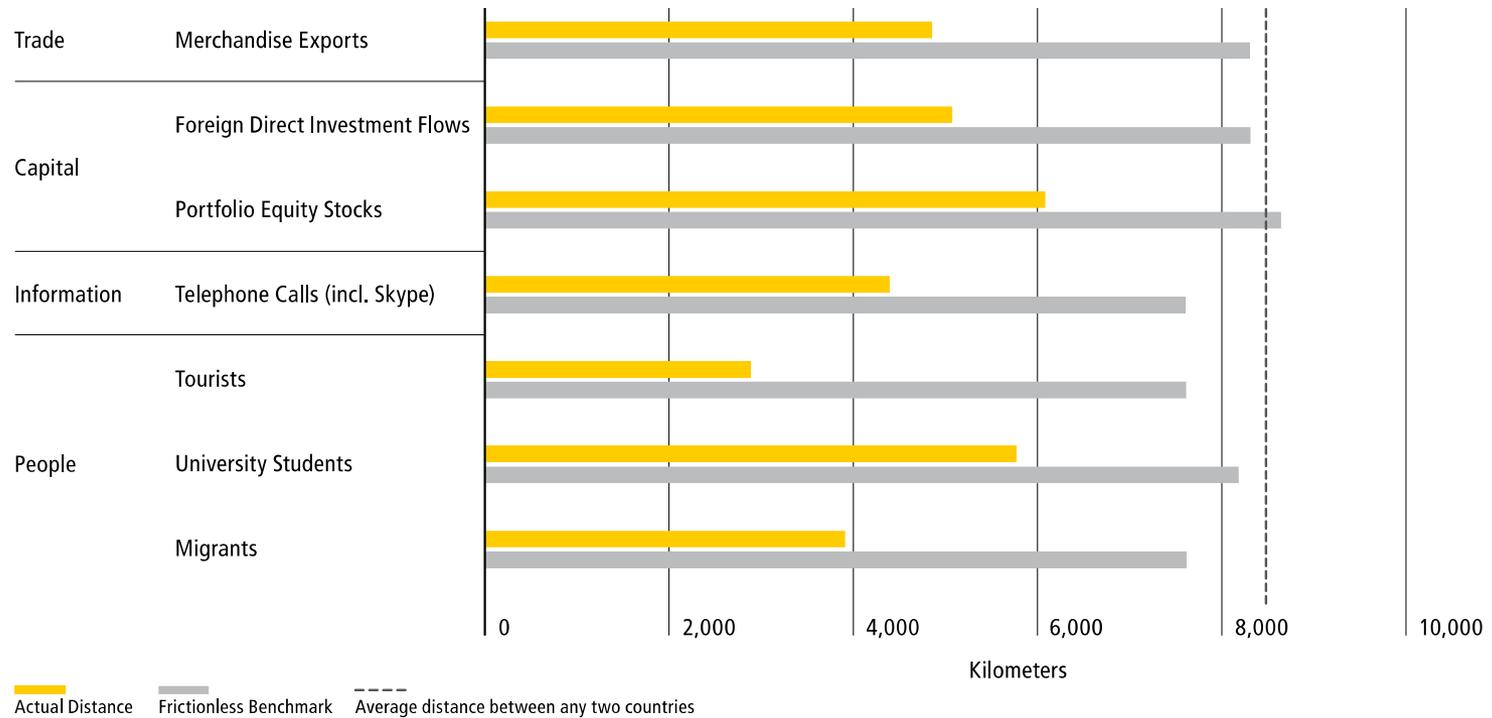
# The world we know is more local than global

## GLOBAL DEPTH MEASURES VERSUS US SURVEY ESTIMATES



# The world we know is more local than global

## GLOBAL AVERAGE DISTANCE VERSUS BENCHMARKS WITHOUT DISTANCE EFFECTS





Increasing  
**volume & complexity**  
of the food chain



Growth of the amount of  
**data available**  
for analysis





# Information boom

- From the dawn of the civilization to 2003 humans produced **5 exabytes** of data in total
- IBM has estimated in 2016 that 2.5 exabytes (2.5 million terabytes) of data are produced every day
- Now it is around 5 exabytes **daily**



# Post-normal science

- Policy-related scientific problems:
  - uncertain facts
  - disputes over ethics and values
  - urgent decisions needed
  - that may have far-reaching consequences
- Policy makers are required to make difficult and firm decisions based on data characterized by high levels of uncertainty.

# Age of **post-truth** politics

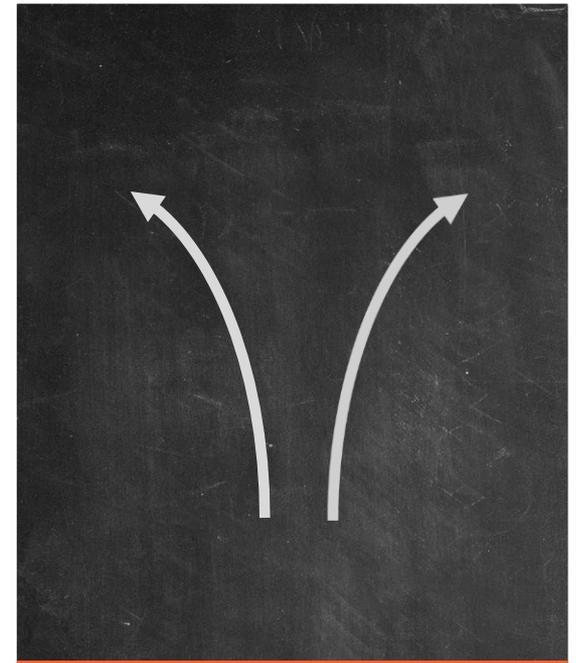
- Oversupply of facts in the 21st century
  - too many sources
  - too many methods
  - with varying levels of credibility
- We are in the middle of a **transition from a society of facts to a society of data.**
- Numbers are being generated much faster than we have any specific use for.
- But the data can be mined to get a sense of how people are behaving and what they are thinking
  - Analysis of reactions and feelings rather than dealing with facts
  - Numbers are viewed more as indicators of current sentiment, rather than as statements about reality



Increasing  
**volume & complexity**  
of the food chain



Growth of the amount of  
**data available**  
for analysis



Better evidence-based  
**decision making?**





# Computational science as a solution

Computational science:

- Able to detect patterns which can not be detected by a smaller set of data
- Those **emerging patterns** can be surprising & counter-intuitive
- 'more is different'



---

# Computational science as a solution

- Big data
- Data mining
- Network science
- Predictive modelling
- (Business) intelligence
- Artificial intelligence
- ...

## Example

Identifying trending topics in news based on text mining and network analysis



# The **idea**

- Data source: Europe Media Monitor (EMM) MediSys
- Words co-occurring in the same text (or sentence or paragraph) are connected to each other
- Same words in different news texts imply their relative importance
- Words connected to each other form a network
  - node size and colour: occurrence frequency of the word (weighted degree)
  - edge thickness: co-occurrence frequency
  - edge colour: occurring in the same text
- This network could be analyzed and visualized



## Example

Identifying emerging areas  
based on network cluster analysis

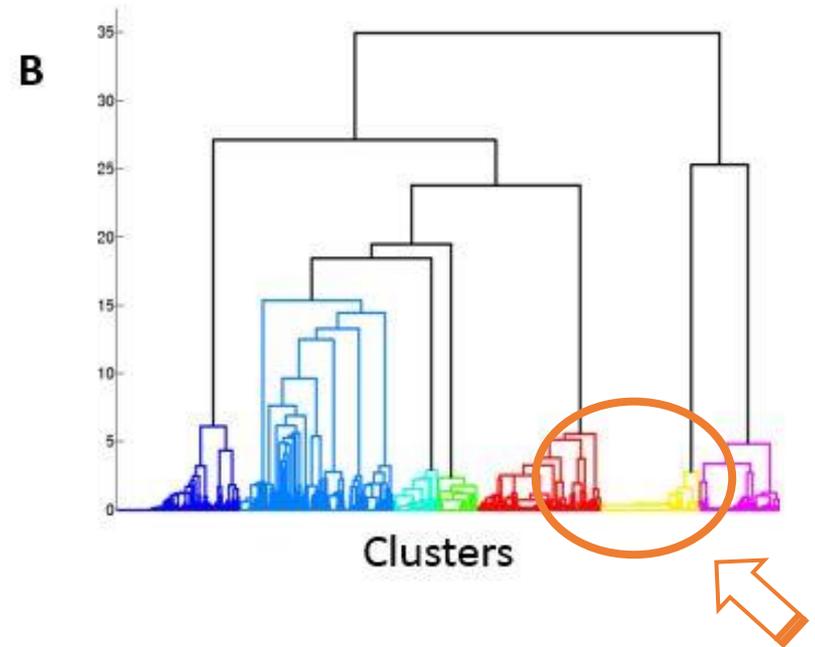
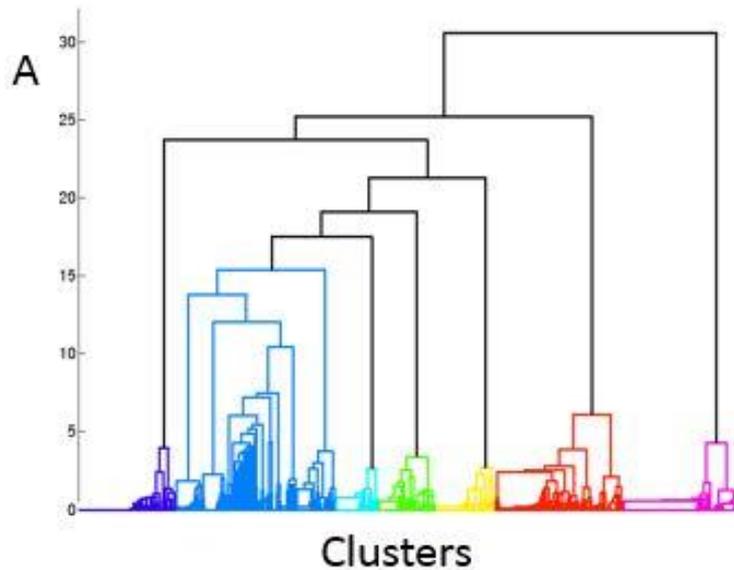
# Patent network analysis

- Based on: *Péter Érdi, Kinga Makovi, Zoltán Somogyvári, Katherine Strandburg, Jan Tobochnik, Péter Volf, László Zalányi: Prediction of Emerging Technologies Based on Analysis of the U.S. Patent Citation Network*
- Connecting patents based on citation
- Patents receiving citations from different patent categories could be grouped based on similarity of this citation pattern
- New groups emerge with time (as a consequence of emerging technological-innovation activities)

# Changes in the structure of clusters reflect **technological evolution**

- Temporal changes in the cluster structure of the patent system can be visualized on dendrograms
- The dendrogram structure is presented in different times
- Comparing the hierarchical structure in different years, we can observe both quantitative changes, when only the height of the branching point (branch separation distance) changed, and qualitative changes, when a new branching point has appeared.

# Temporal changes in the cluster structure of the patent system



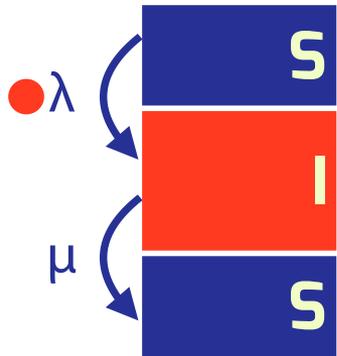
Dendrograms representing the results of the hierarchical Ward clustering of patents in SC11, based on their citation vector similarity on Jan. 1, 1994 (graph A) and Dec. 31, 1999 (graph B).



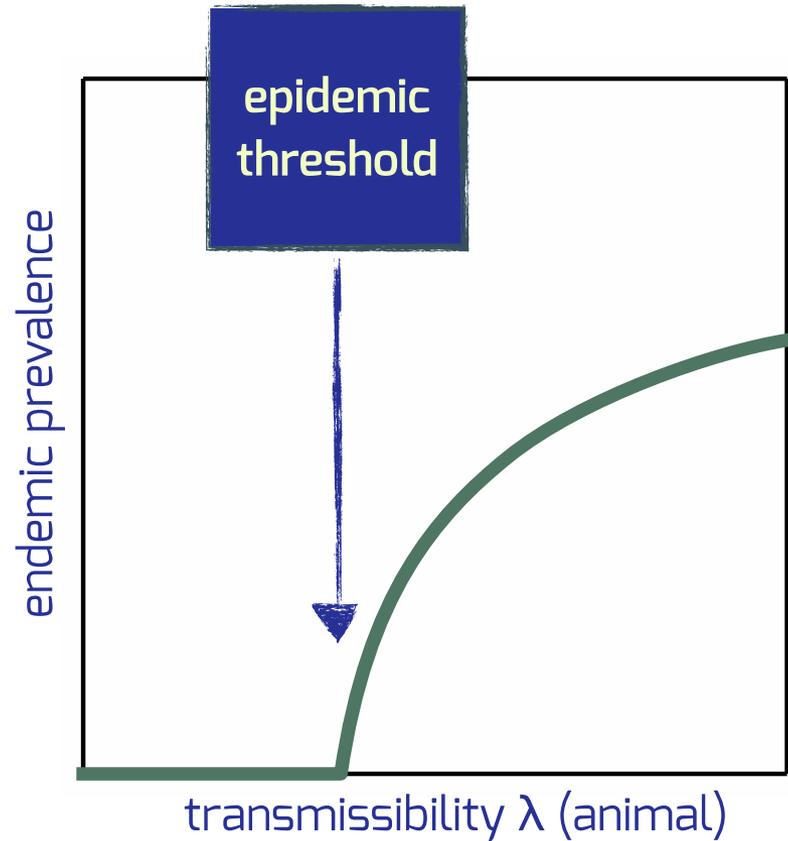
## Example

Cattle trade movements in Europe  
– a multicountry study

# vulnerability epi threshold



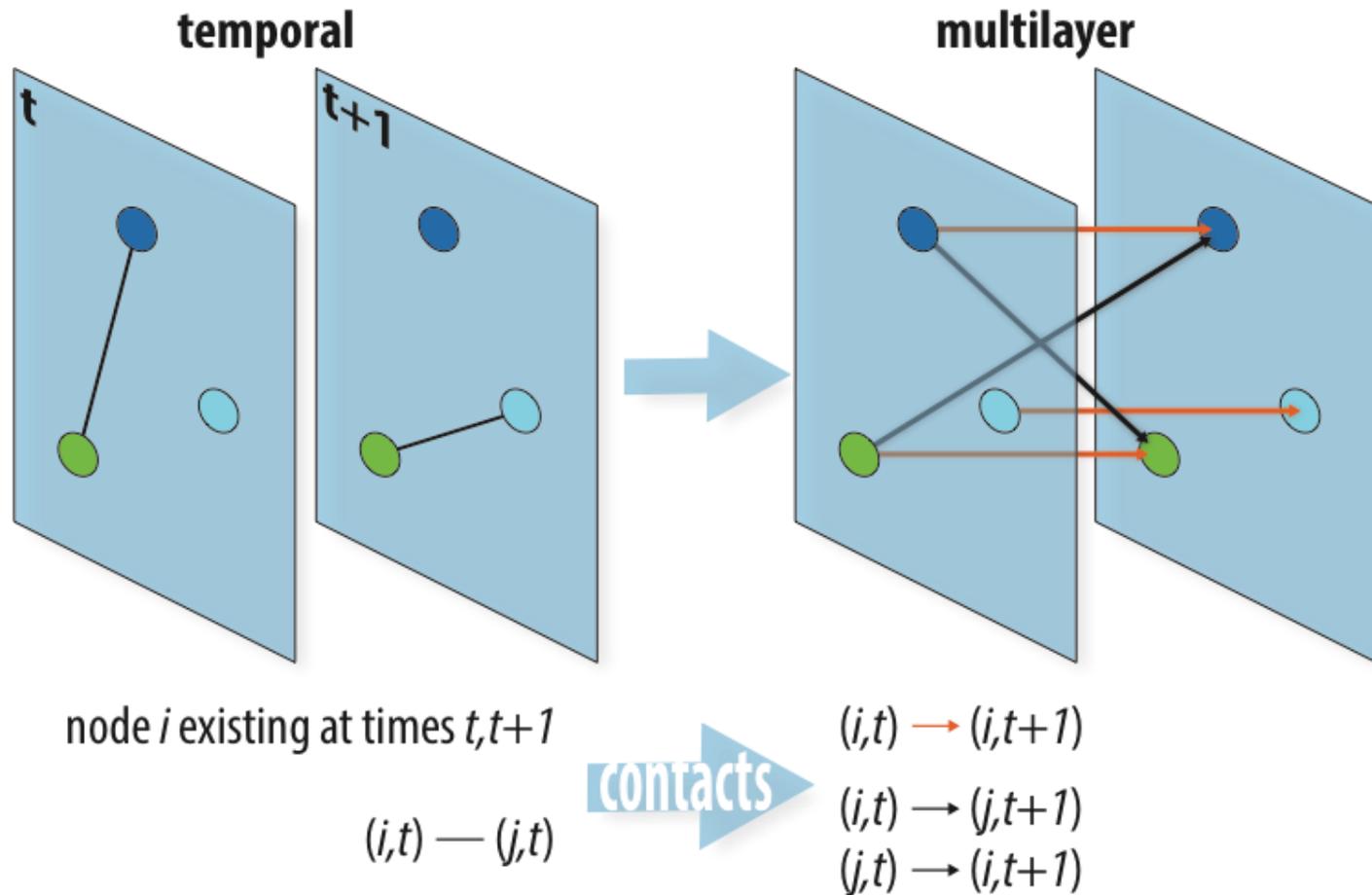
$\mu = 1/(\text{avg outbreak duration})$



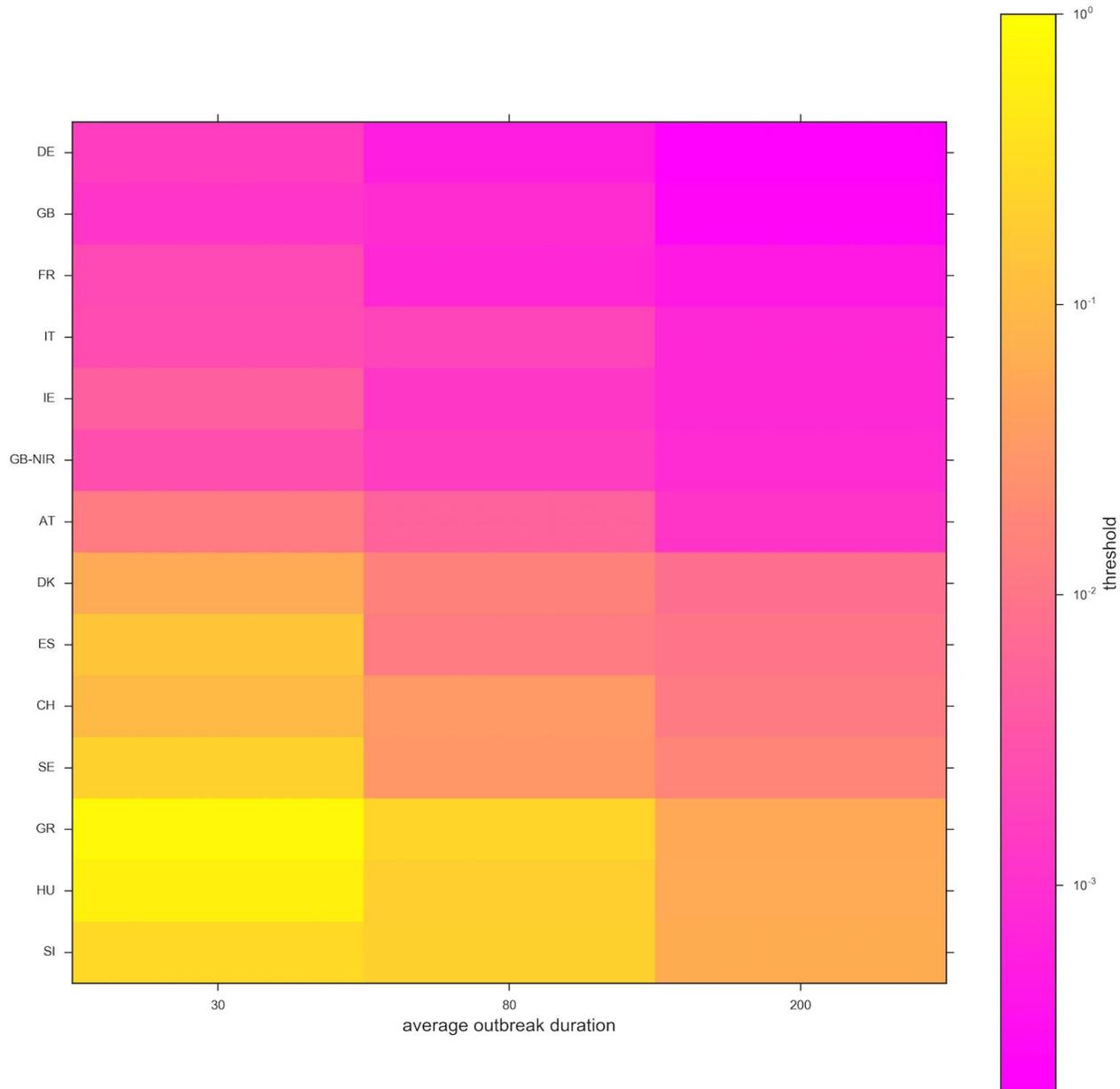
low threshold  
high vulnerability

high threshold  
low vulnerability

# Calculation of epidemic threshold on temporal networks: multilayer approach



# vulnerability results



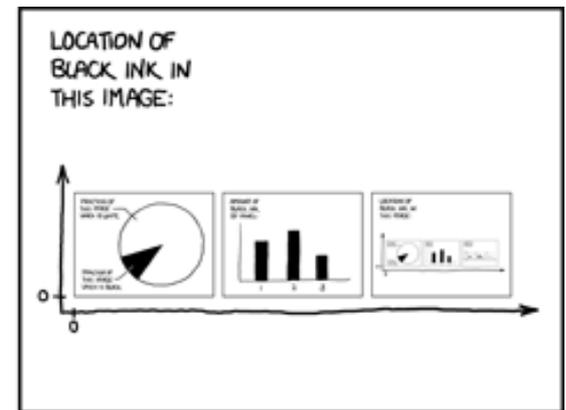
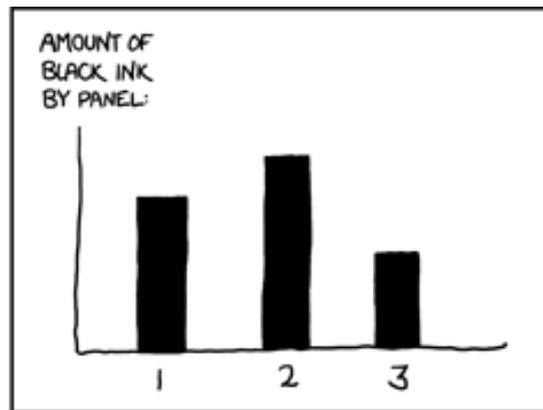
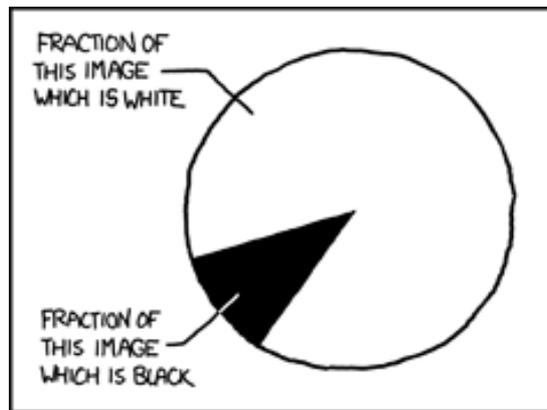
# Meaning of the epidemic threshold

- The lower the threshold, the more vulnerable the country is
  - The network behind the epidemics is more suitable for spreading the disease fast
- It doesn't tell anything about the interventions, the biosecurity nor the real level of risk
- Epidemic threshold is an analysis based purely on network topology.
- The more vulnerable countries should pay more attention to the biosecurity, traceability, control measures, etc. to prevent disease spreading

 Takeaway

# Important is...

- ... the story behind the numbers!



Source: <https://xkcd.com/688/>

- Creation and development of (big) databases is not only an IT problem
- The ability of analysis and evaluation of input data and results: high-level knowledge of food chain science is needed enabling interpretation and validation.

The background of the slide is a complex network diagram. It features numerous nodes of varying sizes and colors (blue, green, black, white) connected by thin lines. Some nodes are highlighted with larger, semi-transparent circles in blue or green. The overall aesthetic is clean and modern, suggesting a digital or scientific network.

nébih

**Thank you!**

**Contact**

Ákos Józwiak

NÉBIH, Hungary

[jozwiaka@nebih.gov.hu](mailto:jozwiaka@nebih.gov.hu)

 [@jozwi](https://twitter.com/jozwi)