Networks (of Everything)

Herman Y Sutarto

Research Director

PT Pusat Riset Energi (PRE)

ECF Extension Course Filsafat
Universitas Katolik Parahyangan

Outline

- Networks
- Data

Networks

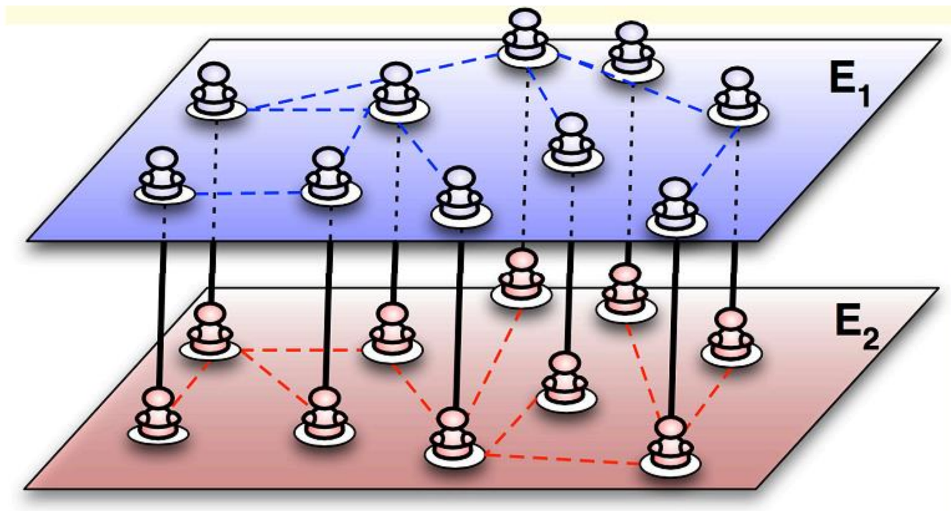
- Networks are everywhere, from the internet, to social networks, traffic, brain, financial, the genetic networks that determine our biological existence.

Patterns

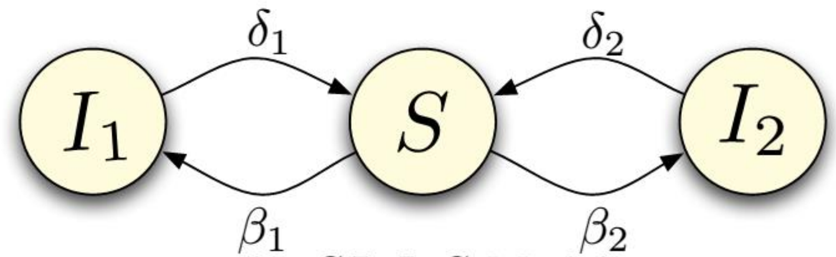
synchronisation

- Novel behavior
- Properties of the whole
- **Cannot be predicted** from properties of the components that make up the system

Model

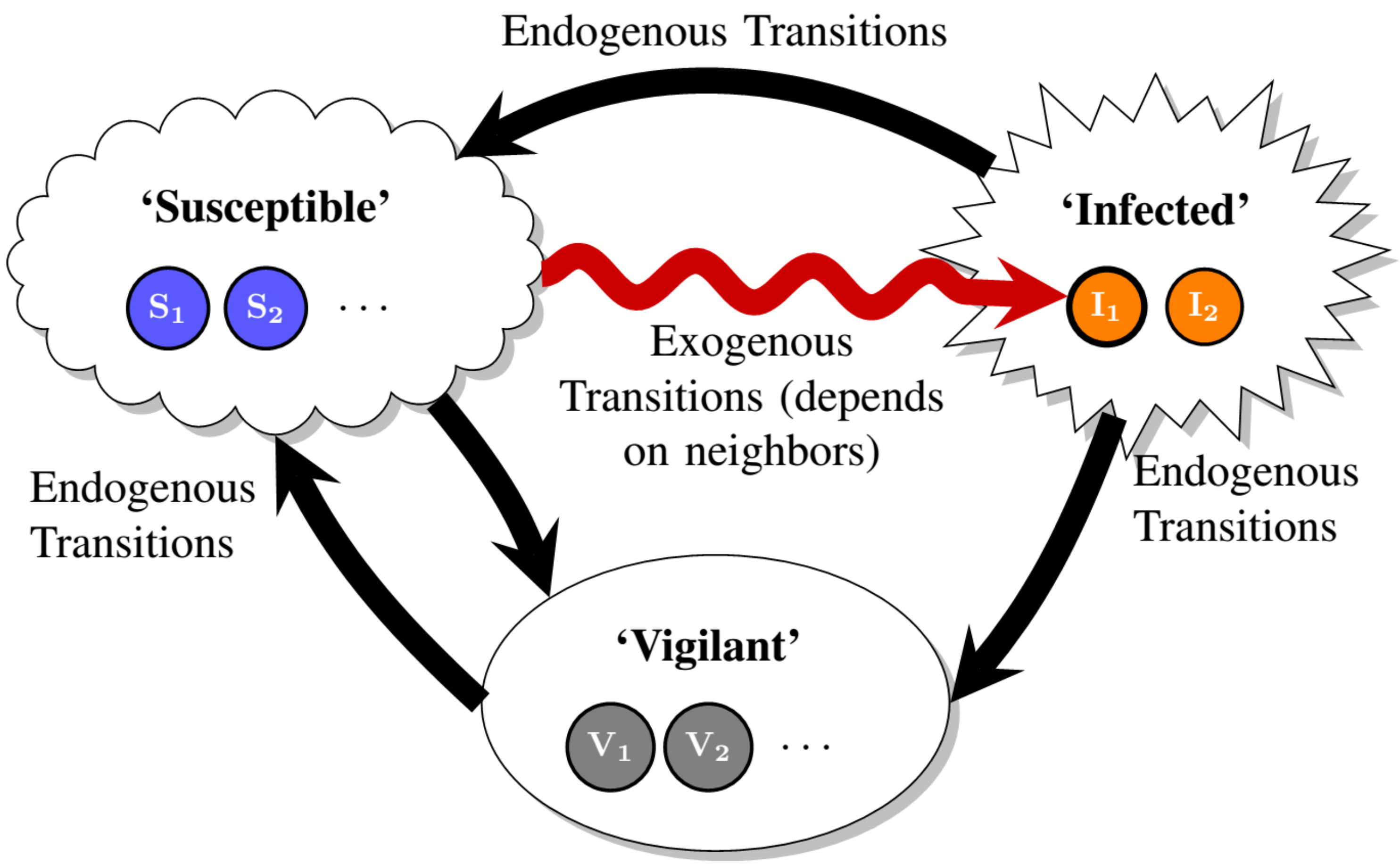


(a) Example Composite Network



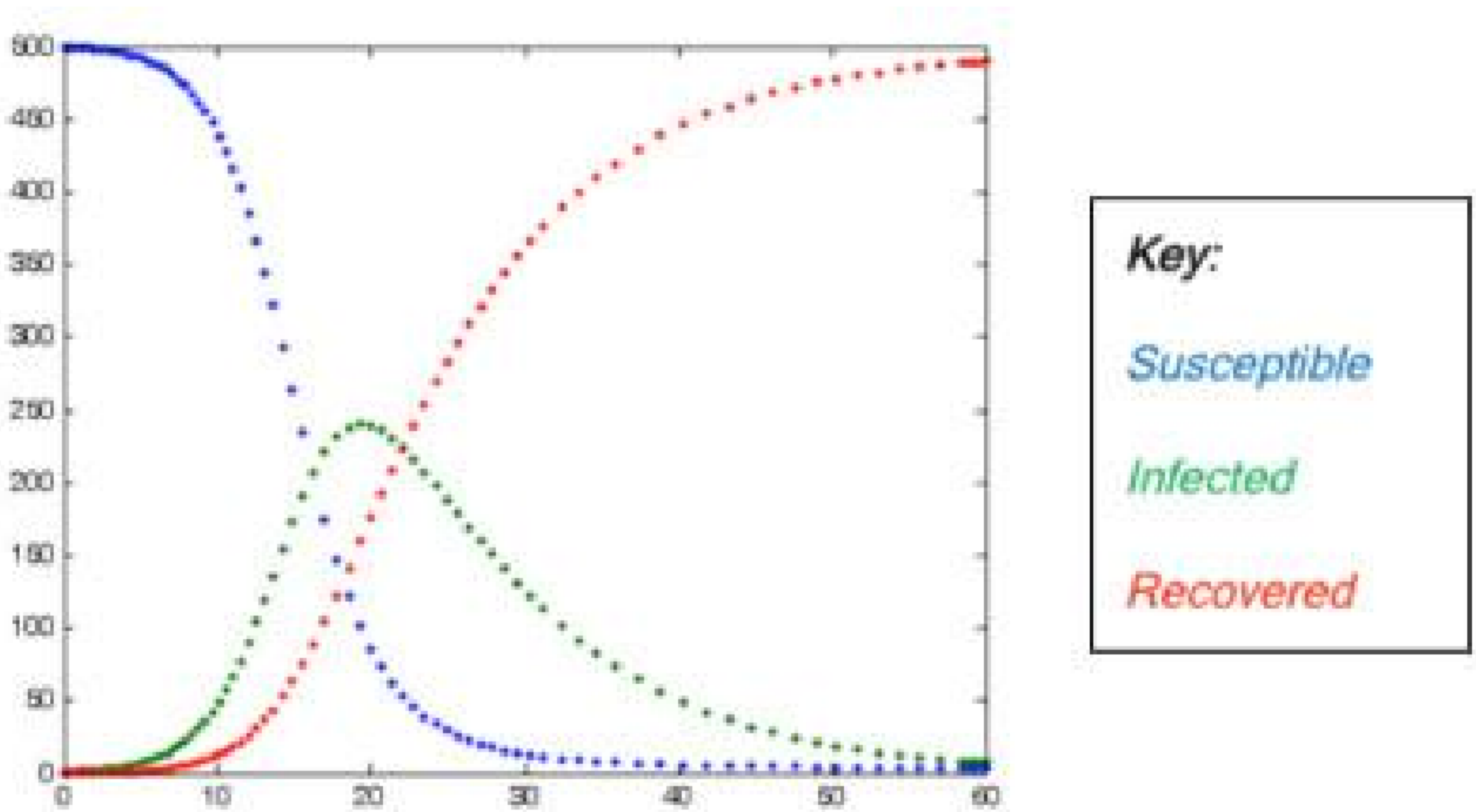
(b) SI_1I_2S Model

Model



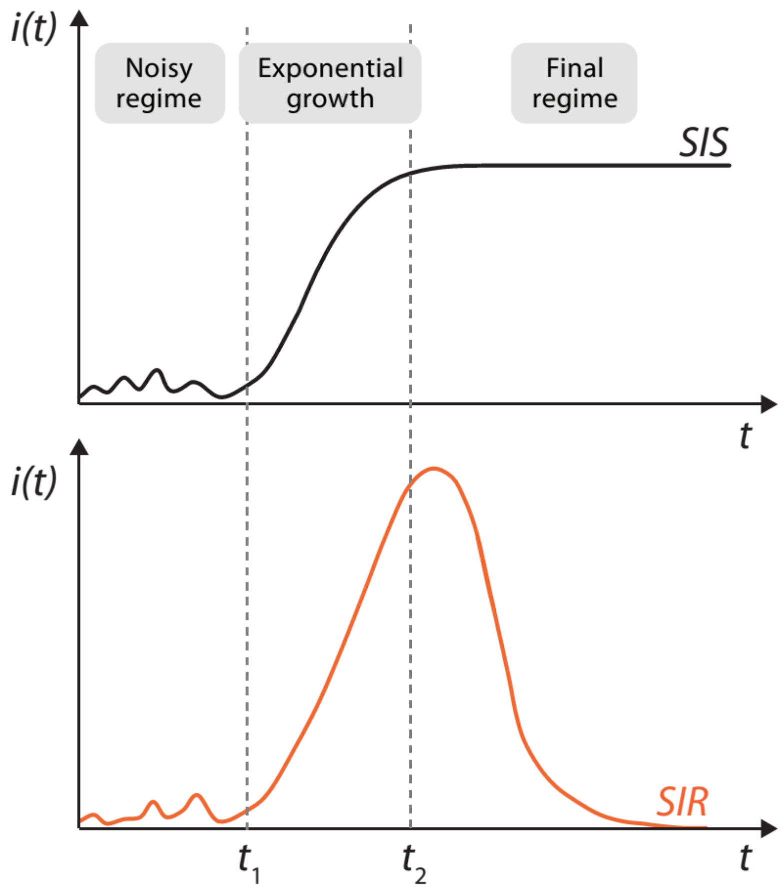
Endogenous Transitions
'Susceptible'
S_1 S_2 . . .
'Infected'
I_1 I_2
Exogenous Transitions (depends on neighbors)
Endogenous Transitions
Endogenous Transitions
'Vigilant'
V_1 V_2 . . .

Rumor Model



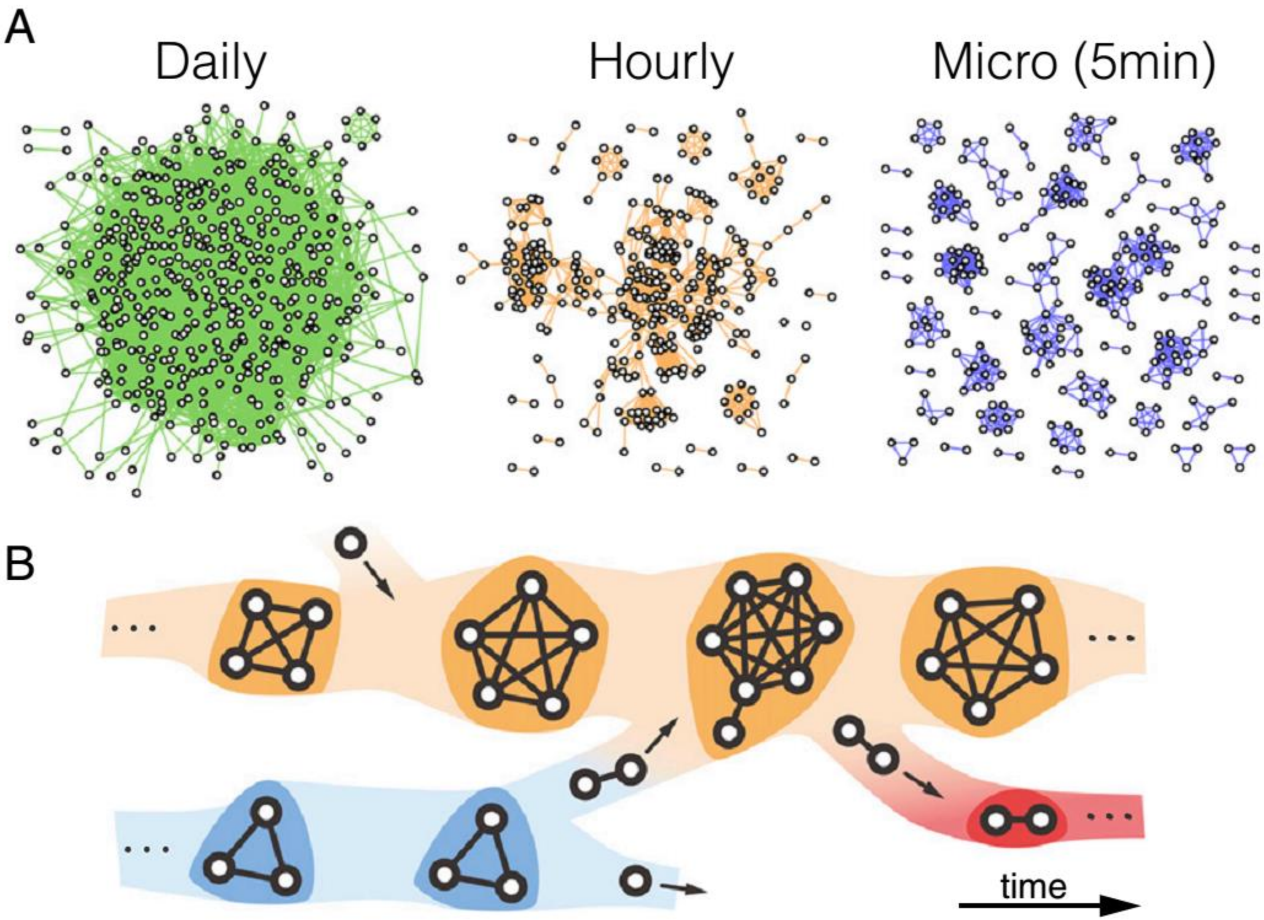
A SIR model at work. The vertical axis shows the number of people in each category as time goes by.

Rumor Model

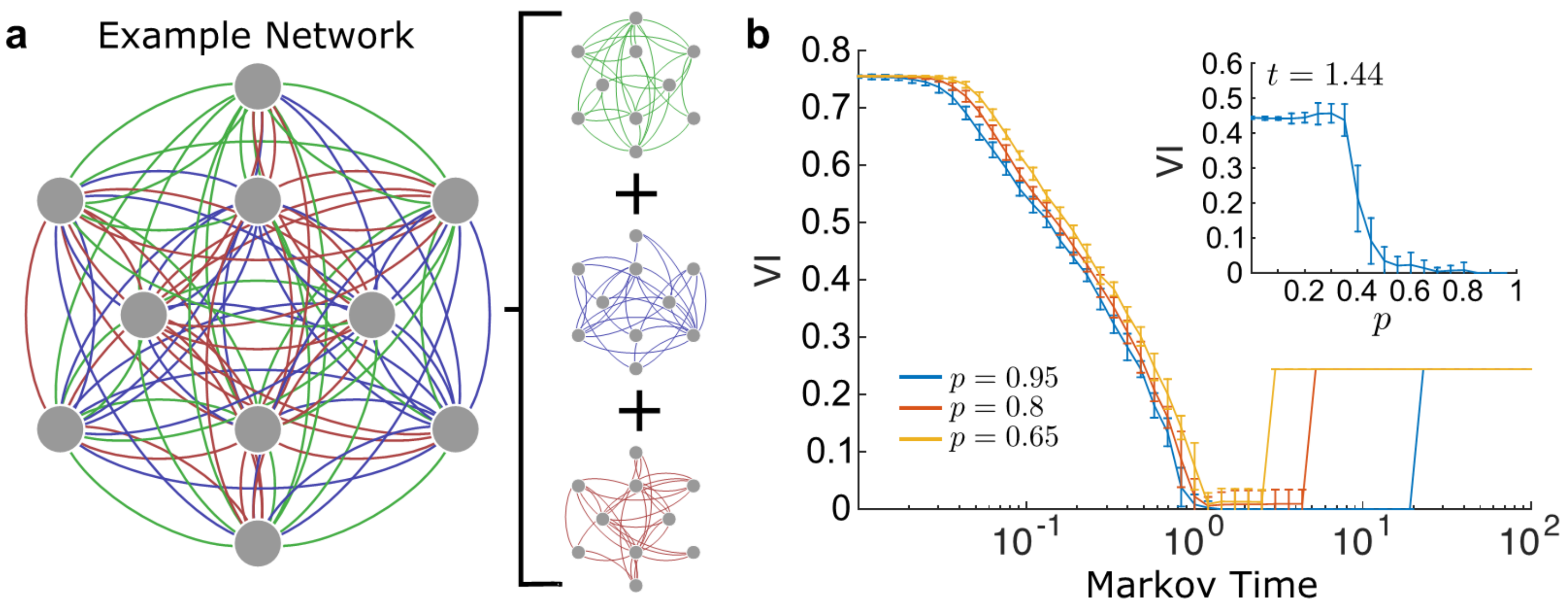


i(t)
Noisy regime
Exponential growth
Final regime
SIS
t
i(t)
SIR
t_1
t_2
t

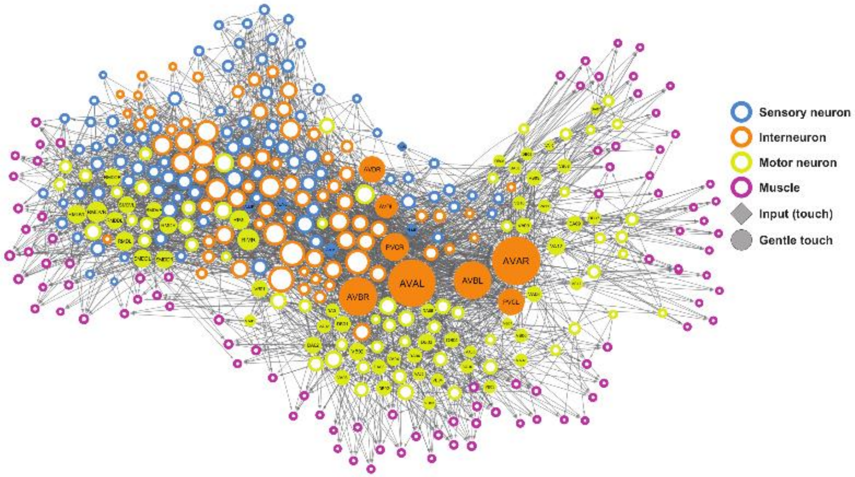
Temporal



Model


a
Example Network
b
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
VI
10^{-1}
10^{0}
10^{1}
10^{2}
Markov Time
$p = 0.95$
$p = 0.8$
$p = 0.65$
$t = 1.44$
0.6
0.5
0.4
0.3
0.2
0.1
0
VI
0.2 0.4 0.6 0.8 1
p

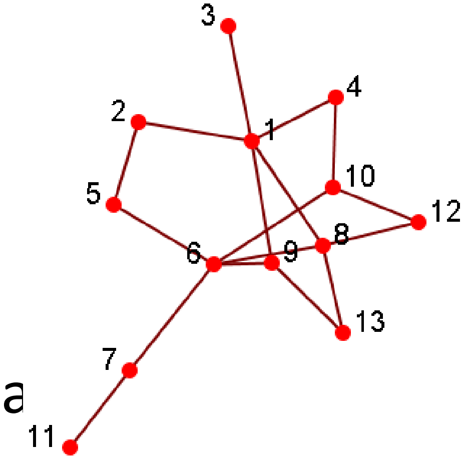
Brain Networks



The brain project utilizes concepts and tools from network science to understand the structural principles of and functional implications for connectomes across species, from the nervous system of the model organism Caenorhabditis elegans, to the mouse, to the human. The brain is inherently multiscale in nature and may be conceptualized as a network at each level; from that of individual neurons and synapses to the integration of macroscopic brain regions.

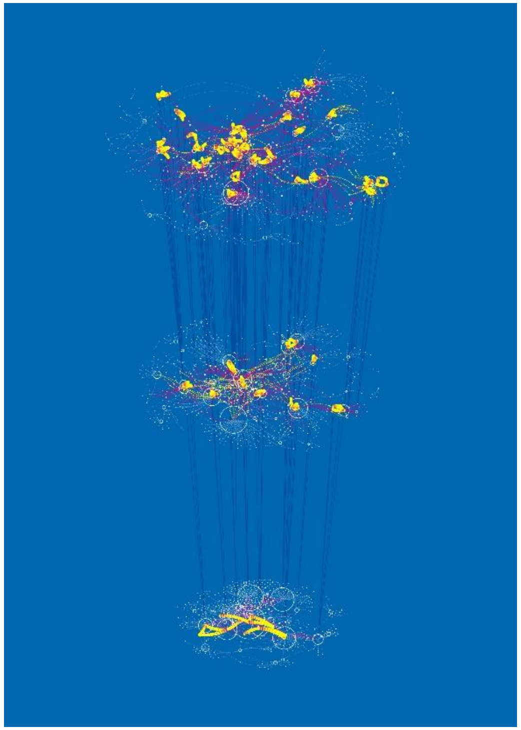
Social Networks

- A social structure made of nodes (individuals or organizations) that are related to each other by various interdependencies like friendship, kinship, etc.
- Graphical representation
  - Nodes = members
  - Edges = relationships
- Various realizations
  - Friendship networks (facebook, WA,myspa
  - Blogosphere
  - Media Sharing (Flickr, Youtube)
  - Folksonomies



Social Network

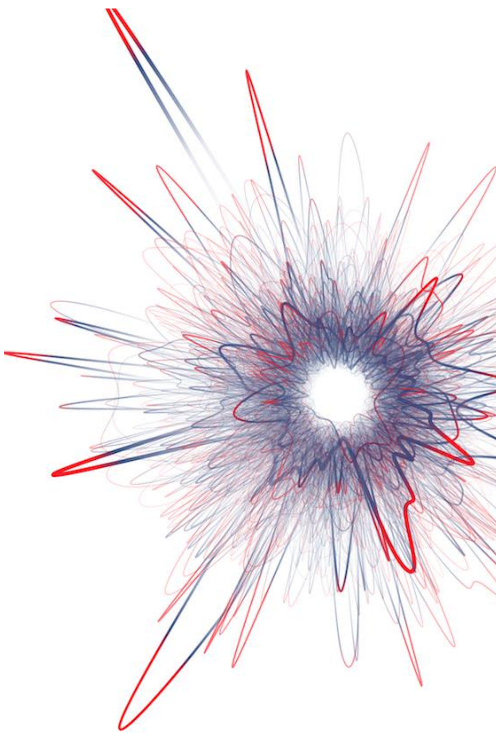
Network Dynamics and Control



We are working on a number of studies that develop mathematical and theoretical models for understanding internal control mechanisms for complex self-organized systems. One can control the behavior of a large network by taking control actions on a comparatively small number of nodes because the network structure broadcasts the influence of these "driver nodes" to distant parts of the network. These findings have tremendous implications for designing, disrupting, or facilitating system capabilities, including physical systems (e.g., climate change and resilience of habitats), technological systems, and biological systems.

Urban Traffic

Science of Success



The goal of the Science of Success project is to develop measures, models and predictions that offer actionable information towards a quantitative evaluation of success in a diverse range of competitive settings, from science to sports and software development. The work is driven by the hypothesis that success can become predictable to a substantial extent if we see it not as an individual phenomenon, but rather as a collective one. For a scientific finding, an athlete, or **a software product to be successful**, **it is not enough to be novel, fundamental or high performing - the community must agree that it is worthy of praise and follow-up.** The aim is to understand the fundamental patterns that govern community impact by analyzing the evolution of career paths, of individual and team performances, and the dynamics of impact, using large-scale data sets that provide quantitative information on performance and success.

Discovery

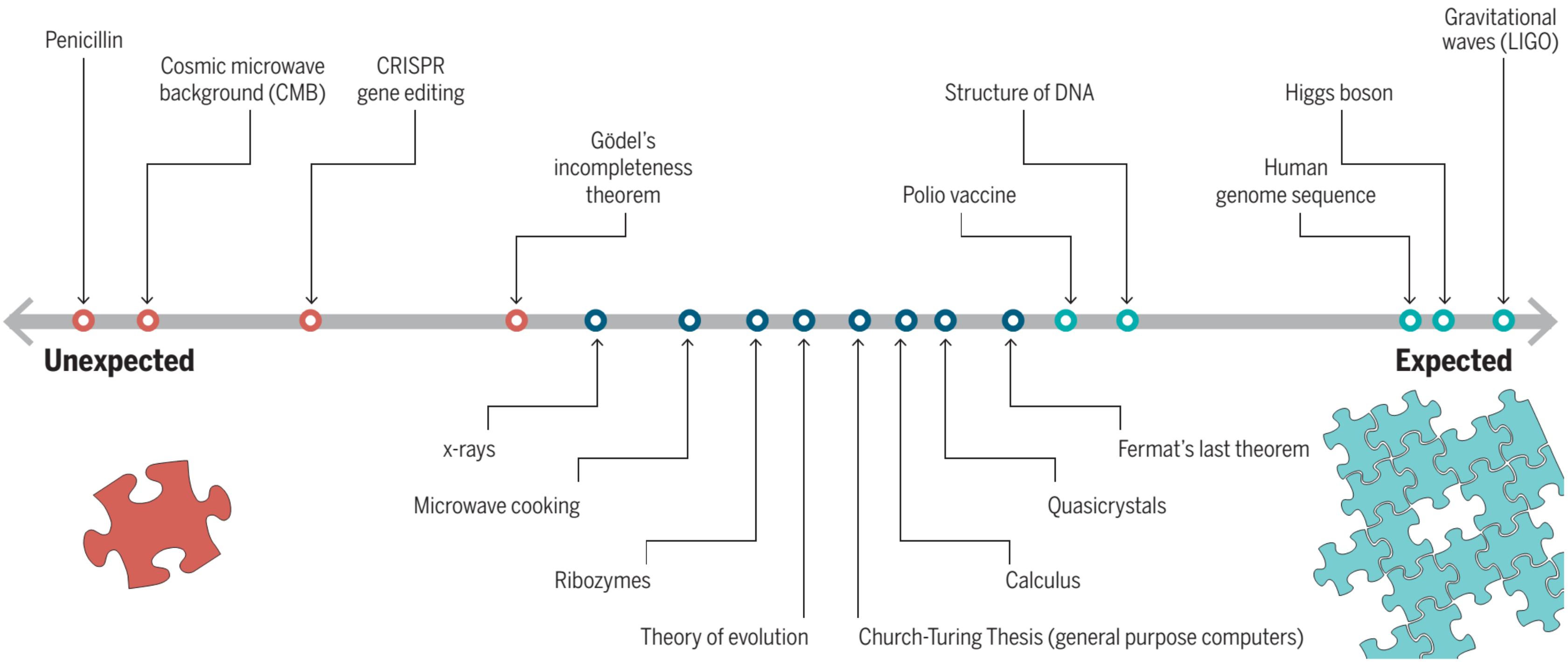


Fig. 1. How unexpected is a discovery? Scientific discoveries vary in how unexpected they were relative to existing knowledge. To illustrate this perspective, 17 examples of major scientific discoveries are arranged from the unanticipated (like antibiotics, programmable gene editing, and cosmic microwave background radiation) to expected discoveries (like the observation of gravitational waves, the structure of DNA, or the decoding of the human genome).

Who will publish the next breakthrough?

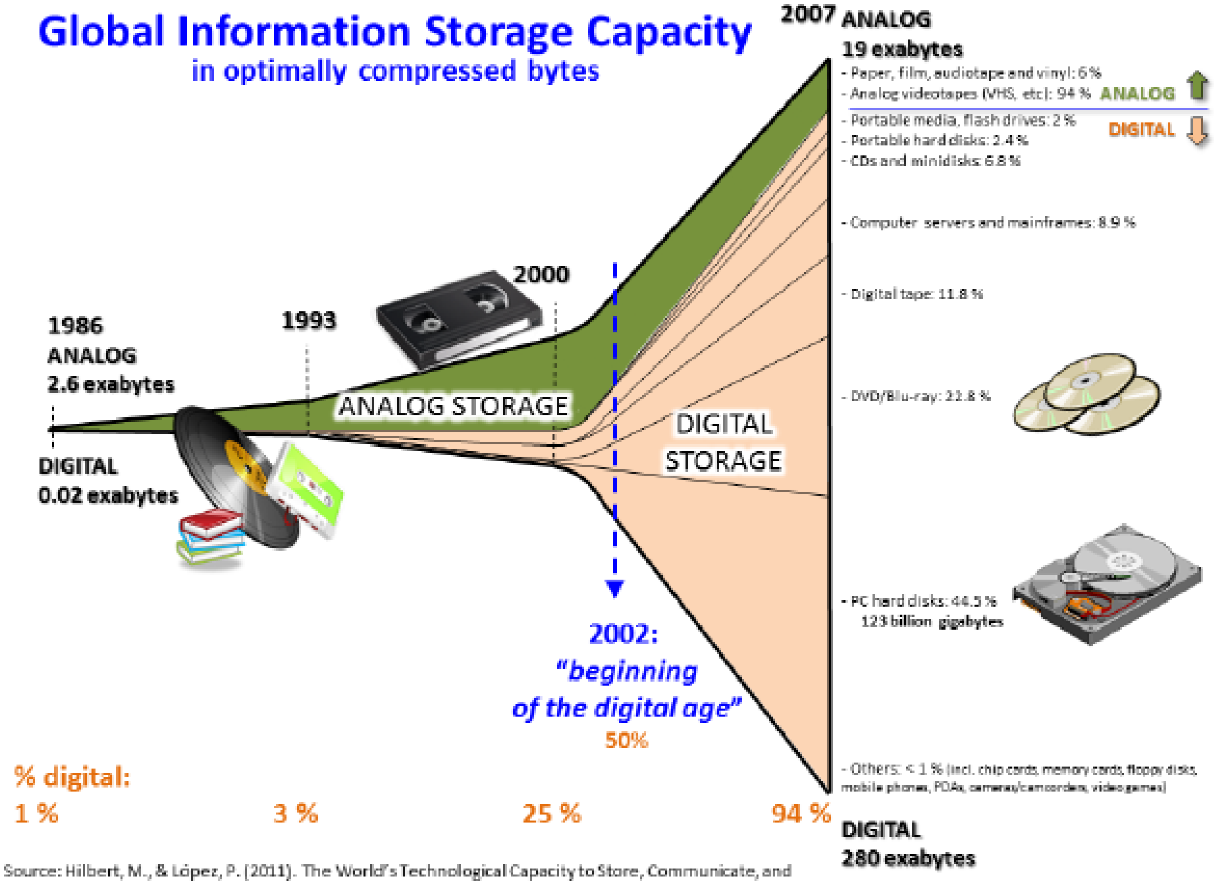
Who will publish the next breakthrough? Who will get grants? Who will get tenure?

Data Science

The (Long) History of Data Processing

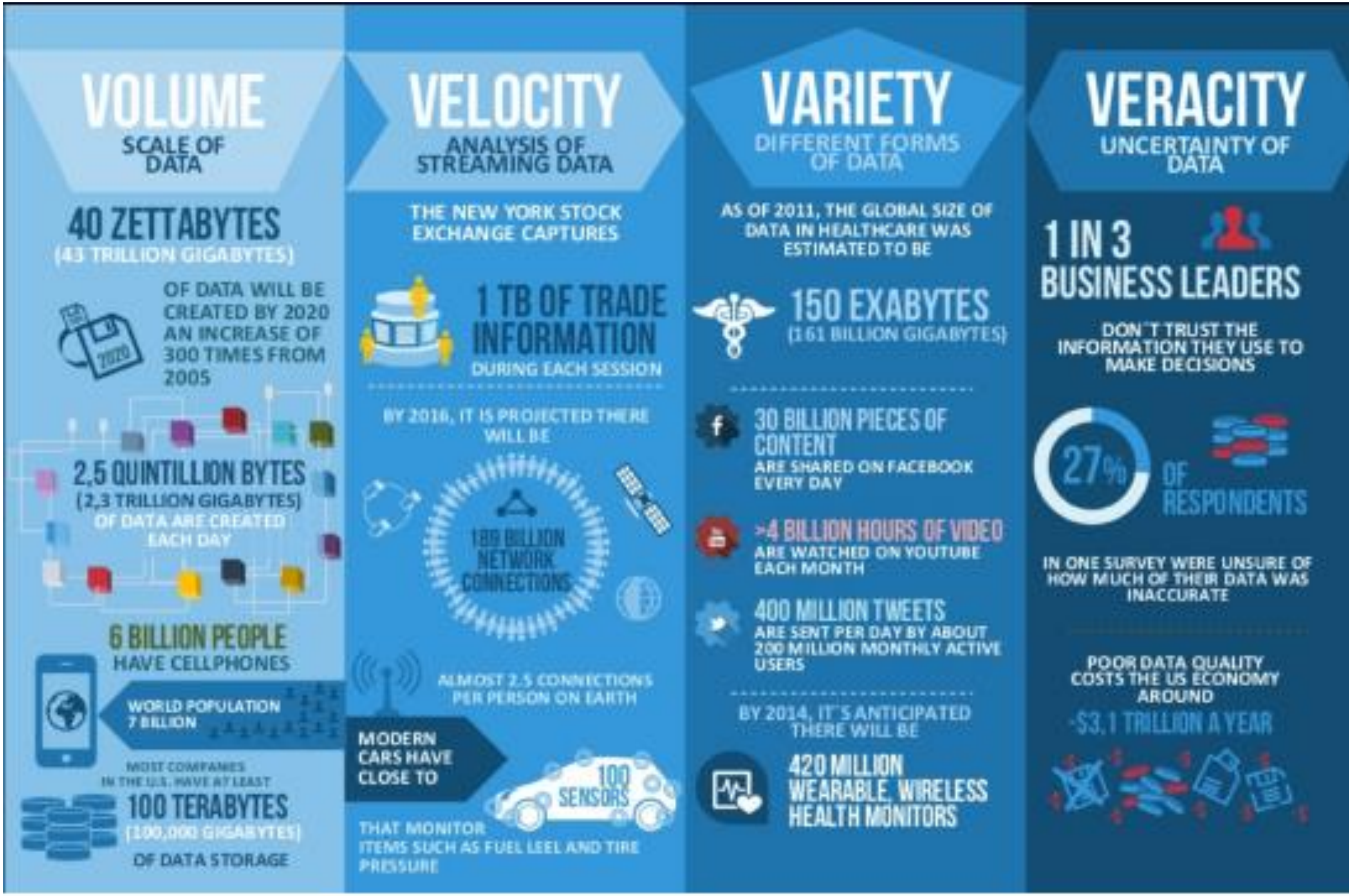
- (…) Manual Data Processing
- (1832) Punch cards
- (1936) Turing Machine
- (1944) First IBM Computers (SSEC)
- (1945) Von Neumann Architecture
- (1950s) Sort & Search Algorithms
- (1950s) Heuristics Methods
- (1936) Pattern Recognition
- (1951) The First Neural Network Machine
- (1955) Concerns on data explosion by Fremont Rider, Wesleyan University Librarian
- (1960s) DBMS
- (1960s) Data Analysis Methods (Bayesian, Time Series, Stochastic, …)
- (1968) Knuth – The Art of Computer Programming
- (1970s) Relational DBMS
- (1974) SQL
- (1975) First PC (MITS Altair 8800)
- (1990s) Data Mining / KDD
- (1990s) Complex-Event Processing
- (1990s) Data Stream Processing
- (1990s) Social Network Analysis
- (1998) The Term "Big Data" was first coined by John Mashey
- (1999) Internet of Things
- (2001) Volume, Velocity, Variety by Doug Laney
- (2001) "Data Science" by William Cleveland
- (2004) MapReduce
- (2009) No-SQL
- (2011) Global Information Storage Capacity grows at 25% annually by Martin Hilbert, Priscila López

How Big is Big?



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

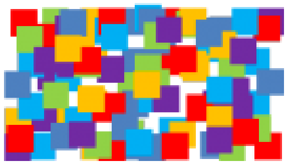
As of 2012, every day 2.5 exabytes (2.5×10^{18}) of data are generated
[IBM, "What is Big Data"]



https://www.slideshare.net/PresentationLoad/big-data-51890676

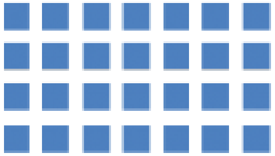
Data Structures

Unstructured Data

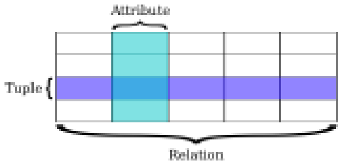


Semi-structured

Structured Data

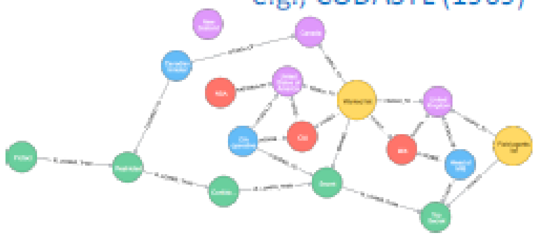


Relational

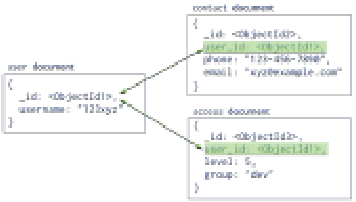


Around 80-90% of all potentially usable business information may originate in unstructured form [Merill Lynch, 1998]

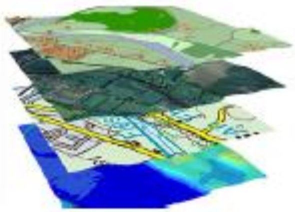
Graph DB → Network Model DB e.g., CODASYL (1969)



Document DB → Hierarchical Model DB e.g., IBM IMS (1969)



Other Data Structures



Spatial / Geospatial Data

Spatio-temporal Data

→ e.g., Moving Objects

Biological Data

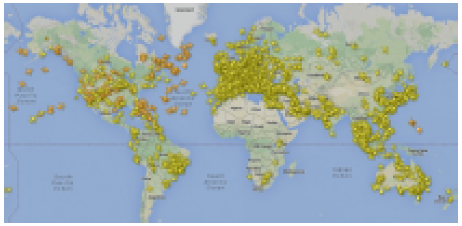
1000 Genomes Project → >200 Terabytes
https://aws.amazon.com/1000genomes/

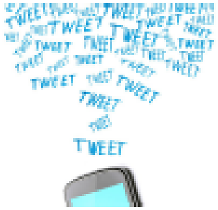
Million Human Genomes project → ???

Data Stream Characteristics

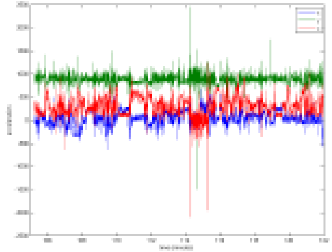
- Continuous flow of data
- Infinite length
  - not just BIG but "UNLIMITED"
  - impractical to store all data
- Examples:

Call Detail Records (CDR)





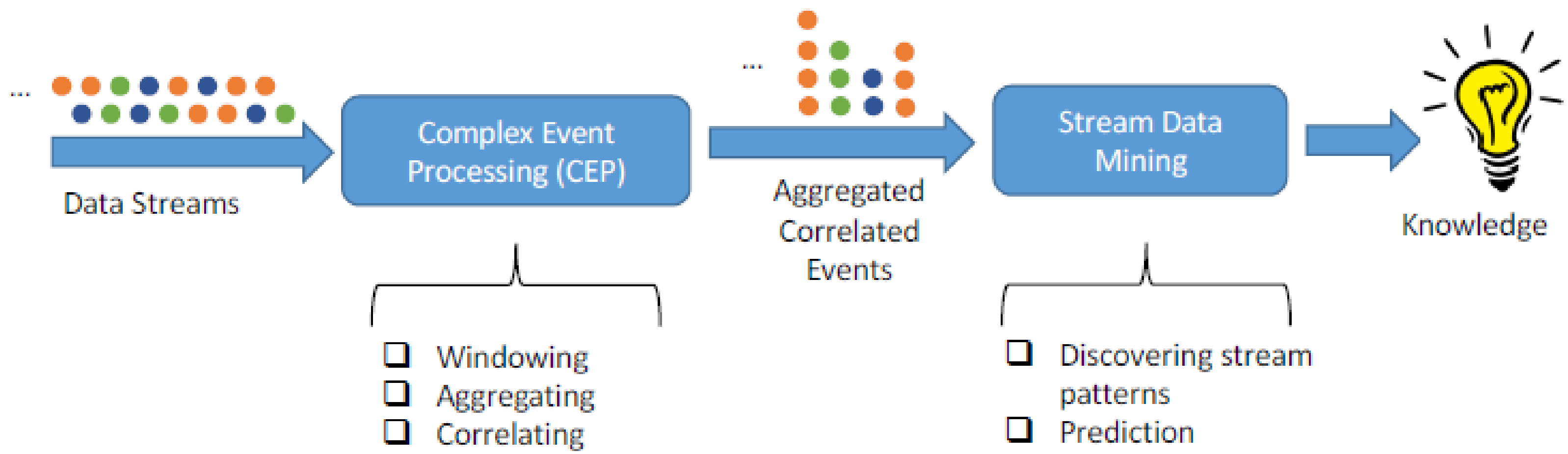
Tweets

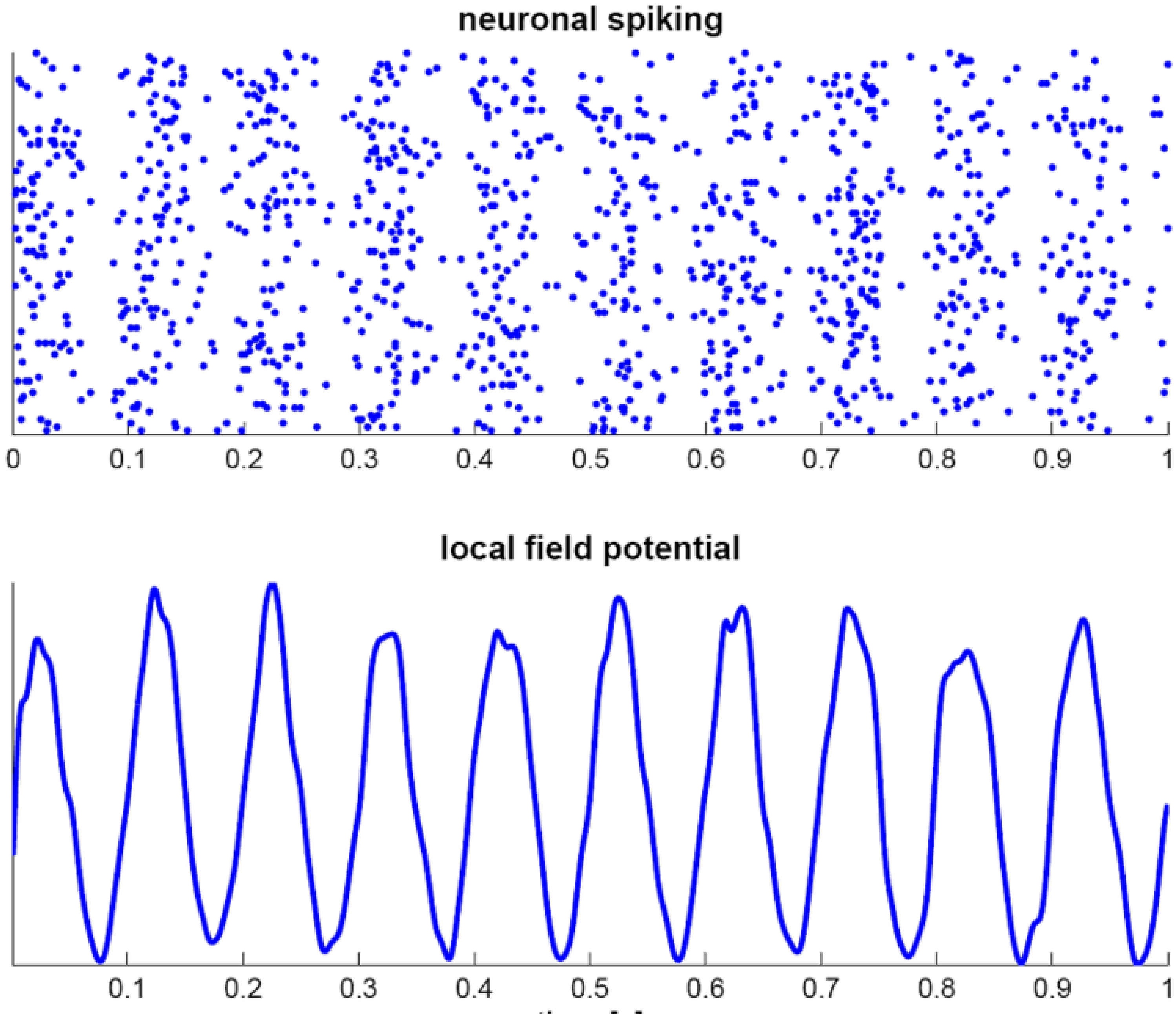


Sensor Data

Data Stream Processing

An example of a typical data stream processing flow




neuronal spiking
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
local field potential
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
time [s]