Information Management

Luis Seco
University of Toronto

Discussion #2

The financial sector in Georgia

FIM word cloud

Finance

Risk Management

Asset Management

Business Cases

Pricing

Blockchains

Credit Scoring

Financial products

Investment scoring

Trading

Hedging

VaR

Portfolio Optimization

Investment strategies

Data Science

Operations Research

Machine Learning

Market Simulation

Corporate finance

Stochastic Analysis

Model risk

Volatility modeling

Databases

Algorithms

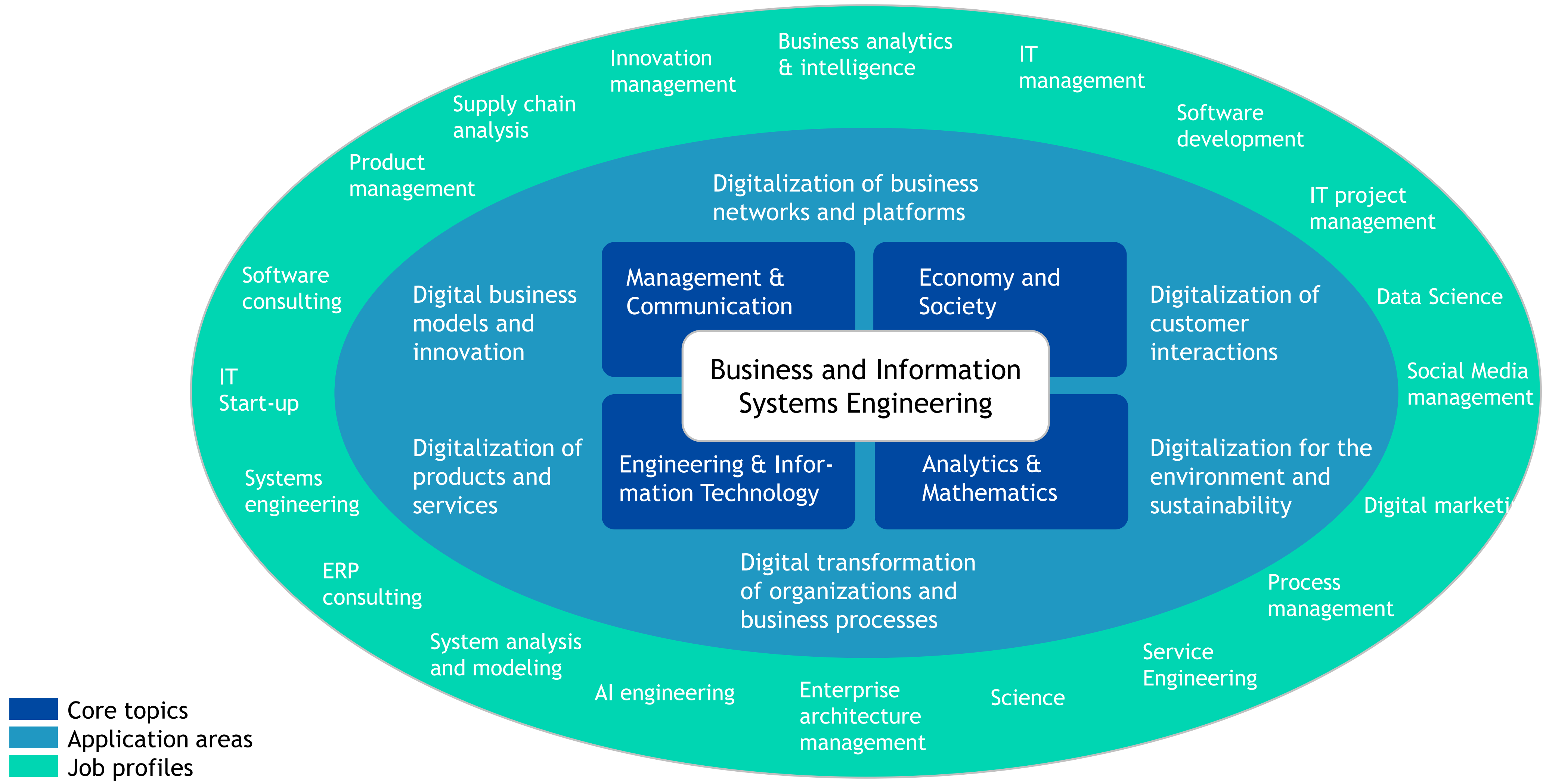
Data feeds

Numerical Methods

The science dealing with the design of

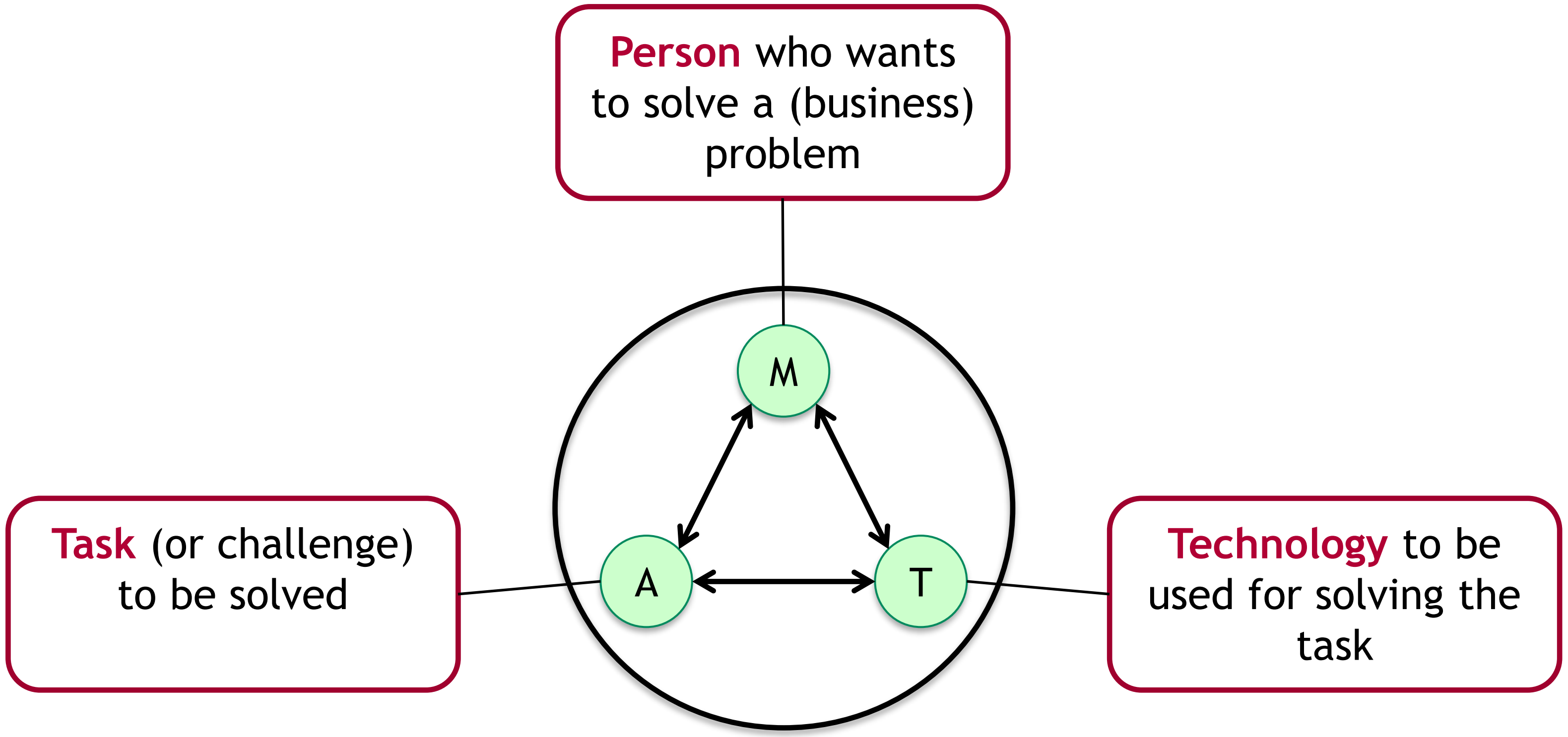
- **computer** information systems and their use in
    - business,
    - administration, and increasingly…
    - … in the immediate private environment.
- An independent, **interdisciplinary** subject essentially between
    - business administration and computer science.

Thematic map of Business and Information Systems Engineering



https://wirtschaftsinformatik.de/unsere-disziplin/themenlandkarte

Information systems = human + task + technology



Business and information systems engineering deals with human-task-technology systems.
Of particular interest is the design and use of information systems in organizations.

Simon (1994)

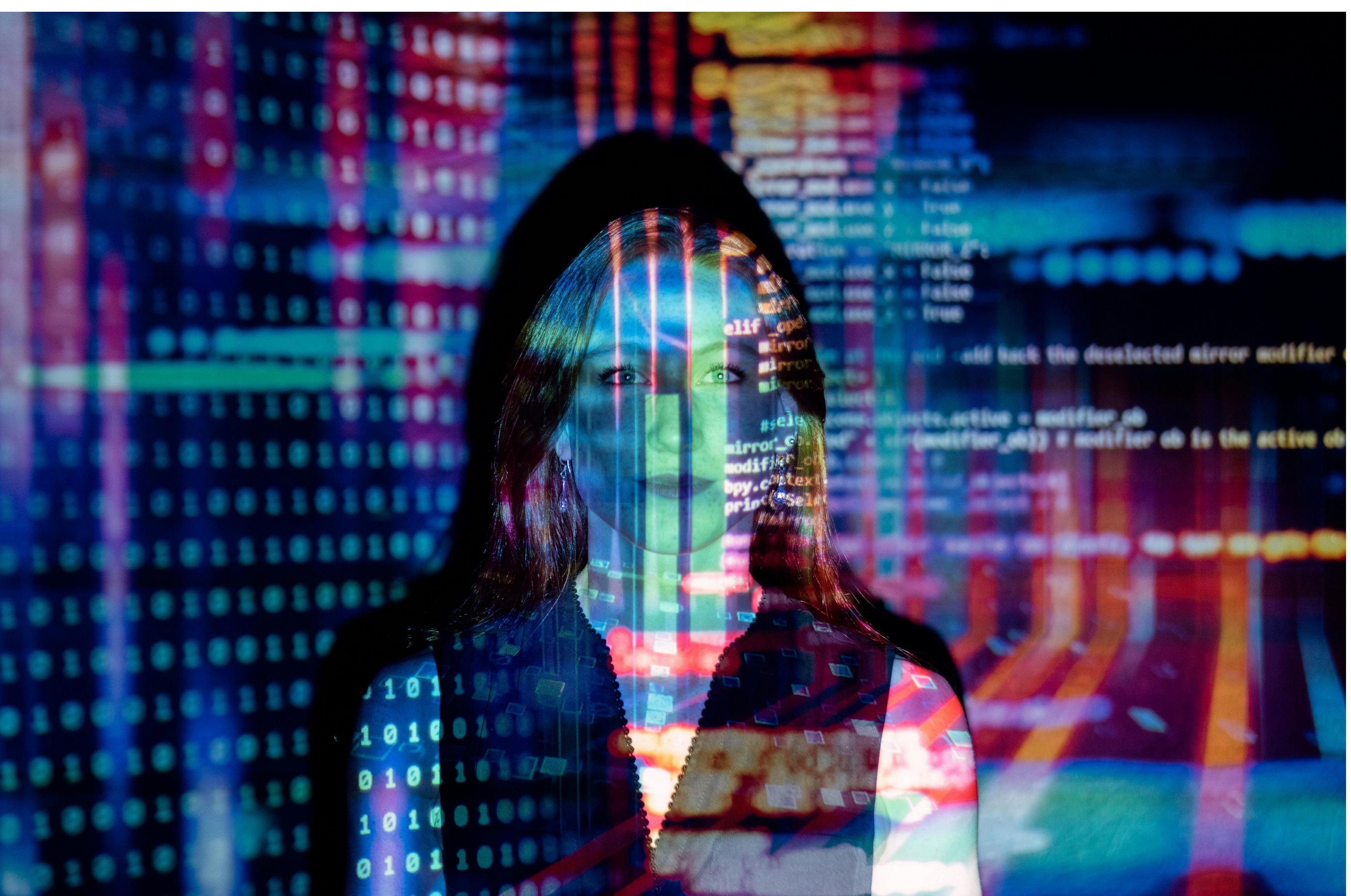
What is Digitalization?

The use of **digital technologies**

to **change** a business model

to provide **new value** opportunities

leading to **new revenue**



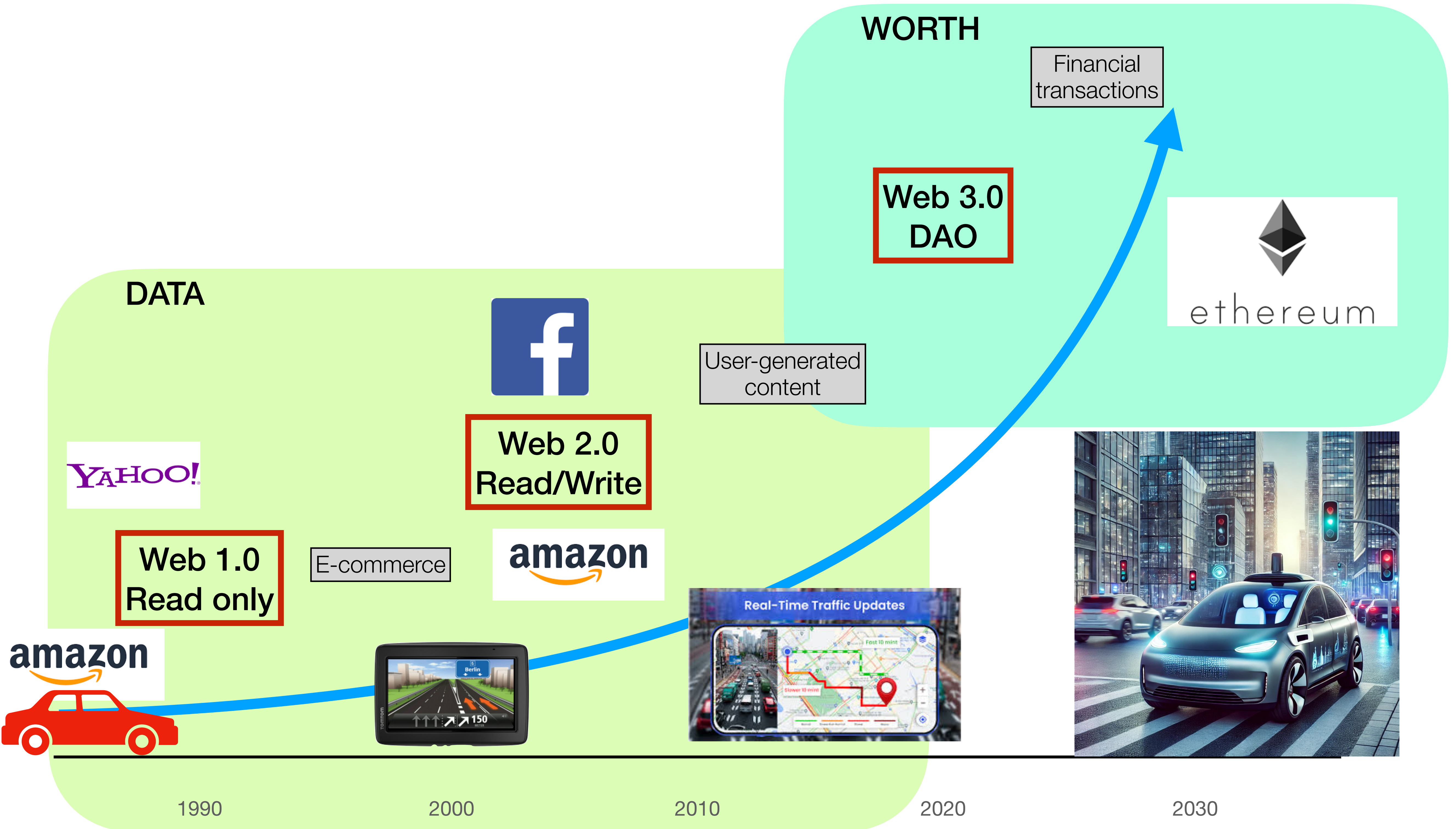
Largest taxi company:
owns no cars

Largest media company
owns no content

Most valuable retailer
owns no inventory

Largest accommodation
provider
owns no buildings

Digital Evolution



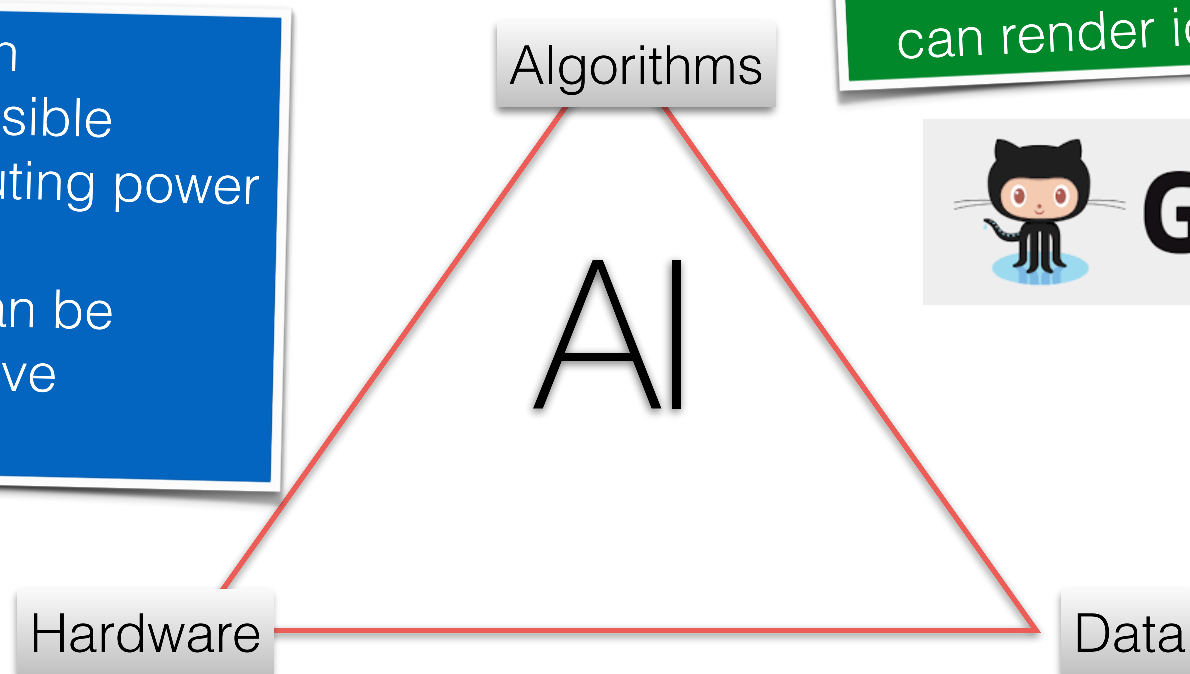
Moore's laws

- Calculation speed doubles every year
- Storage capacity doubles every 1.5 years
- Data speed doubles every two years

What IT is made of

The AI revolution
has been made possible
by the advances in computing power

Further enhances can be
especially disruptive
(crypto-threats)



Democratization of AI
has rendered Algorithms
a free commodity

Tensor flow
can render ideas free

Data monopolies
have the advantage

Data is king

Computational efficiency

- Mainstream algorithms are “free”
- Special purpose algorithms are valuable
- Good models/bad models make a difference

Computational efficiency over time

- 1997: 1 teraFLOP
- 2020: 400 petaFLOPs (400,000x)
- Moores' law: double per year

```
for x in range(10000)
        for y in range (10000)
                for z in range (10000)
                        for w in range (10000)

do something easy ...
```

Two business cases

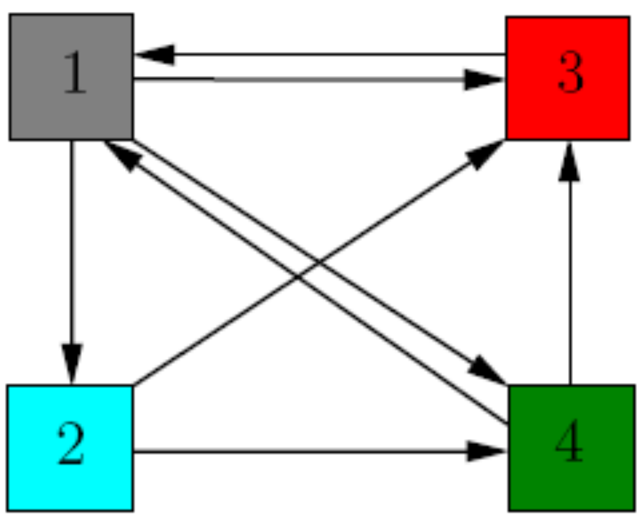
- Page rank & complexity
- Scotiabank business case & speed

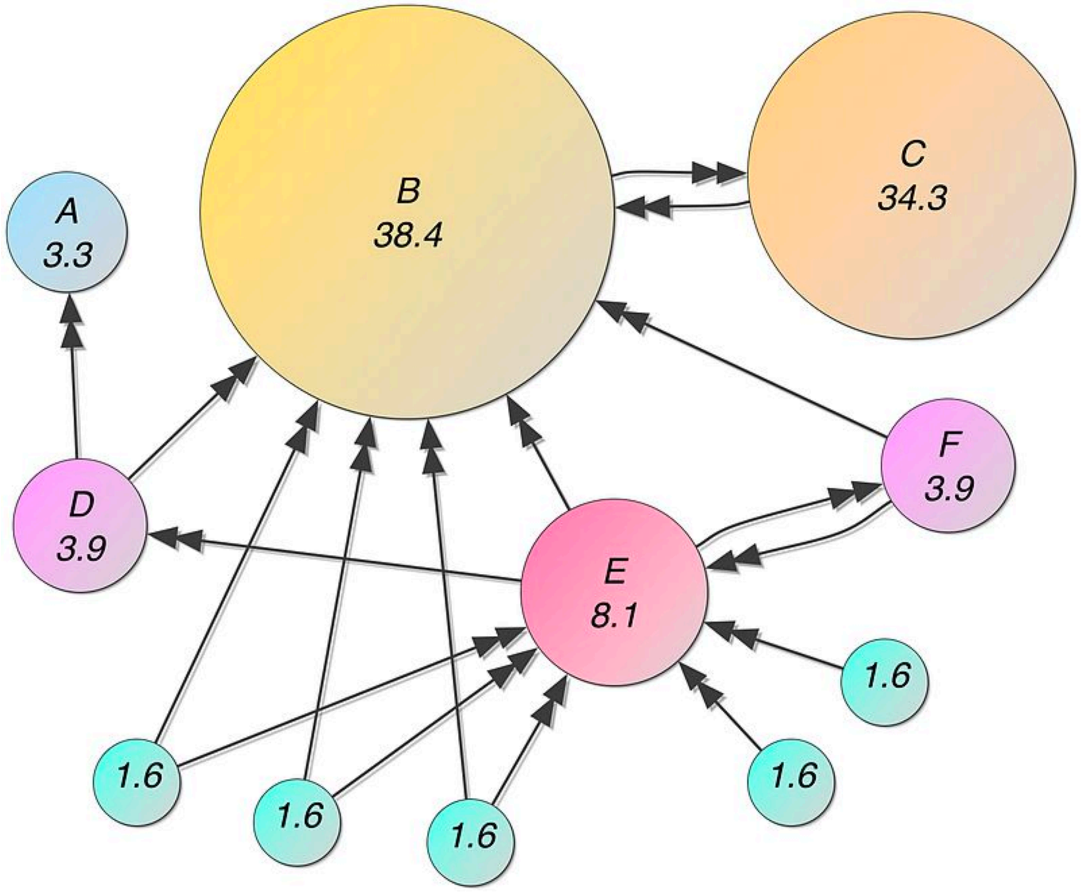
GOOGLE & Page Rank

- Larry Page and Sergey Brin developed PageRank at Stanford University in 1996.
- The name "PageRank" refers to the name of developer Larry Page, and to the concept of a web “page”.
- The word is a trademark of Google, but the PageRank process has been patented (U.S. Patent 6,285,999) by Stanford University (not by Google).
- Google has exclusive license rights on the patent. Stanford received 1.8 million shares of Google in exchange for use of the patent; Stanford sold the shares in 2005 for $336 million.


SEARCH: LINUS
linustechtips
HISTORY OF GOOGLE

Network graphs





Page Rank

- Consider websites w_i, $i = 1, \ldots$
- Consider the Matrix A such that if website j points to website i, then $a_{i,j} = 1$, and 0 otherwise.
- For a website i with n_i outgoing links to websites i_j, $j = 1, \ldots, n_i$

$$n_i = \sum_{j=1}^{\infty} a_{j,i}$$

- Page Rank of a website P_i is a function that satisfies the following properties:

$$P_i = \sum_{j=1}^{\infty} \frac{a_{i,j}}{n_j} P_j$$

$$P = M \cdot P$$

$$m_{i,j} = \frac{a_{i,j}}{n_j}$$

$$
\begin{pmatrix}
m_{1,1} & m_{1,2} & \cdots & m_{1,n} \\
m_{2,1} & m_{2,2} & \cdots & m_{2,n} \\
\cdots & & & \\
m_{n,1} & \cdots & & m_{n,n}
\end{pmatrix}
$$

$$
\begin{array}{lcccc}
+ & & & & \\
\hline
 & 1 & 1 & 1 & 1
\end{array}
$$

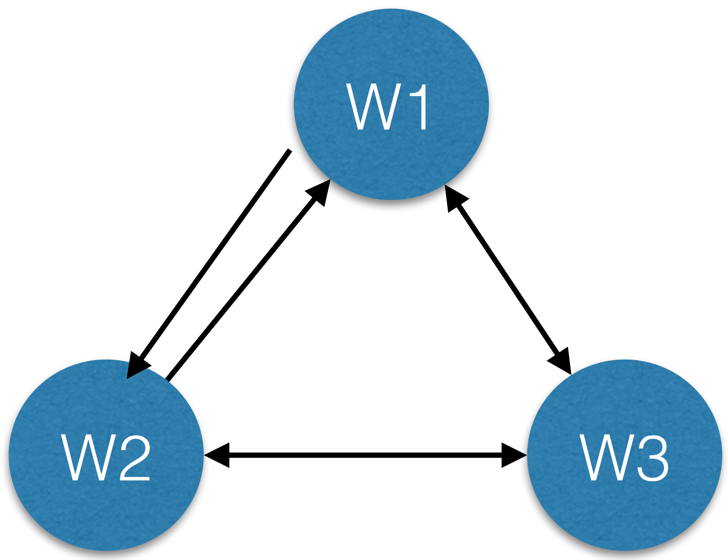
Properties of Markov matrices

If A is a Markov Matrix, then

- A has 1 as the largest eigenvalue
- All other eigenvalues are (in absolute value) less than 1.

Simple example of a network graph

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad n = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$



$$M = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1 \\ 1/2 & 0 & 0 \end{pmatrix}$$

- Consider the Matrix M such that if website j points to website i, then $m_{i,j} = \dfrac{1}{n_j}$, and 0 otherwise.

- The ranking has a vector $r = (r_1, \ldots)$ such that

$$M \cdot r = r$$

In other words, the ranking is the eigenvector with eigenvalue 1.

Page rank problem: Given a network graph determined by matrix M, determine the vector r.

Comparison of two approaches

Eigenvalue approach: Diagonalize the matrix M to find its eigenvector r with eigenvalue 1:

$$M \cdot r = r$$

Perron-Frobenius Approach: Take any ranking r_0, and form the sequence

$$r_{n+1} = M \cdot r_n$$

Then, $\lim_{n \to \infty} r_n = \lim_{n \to \infty} M^n r_0$ is the rank we are looking for.

Complexity of Page Rank

- The complexity of a system solver for

$$M \cdot r = r$$

is p^3 where p is the number of websites (today, $p > 10^9$)

- The complexity of each multiplication

$$r_{n+1} = M \cdot r_n$$

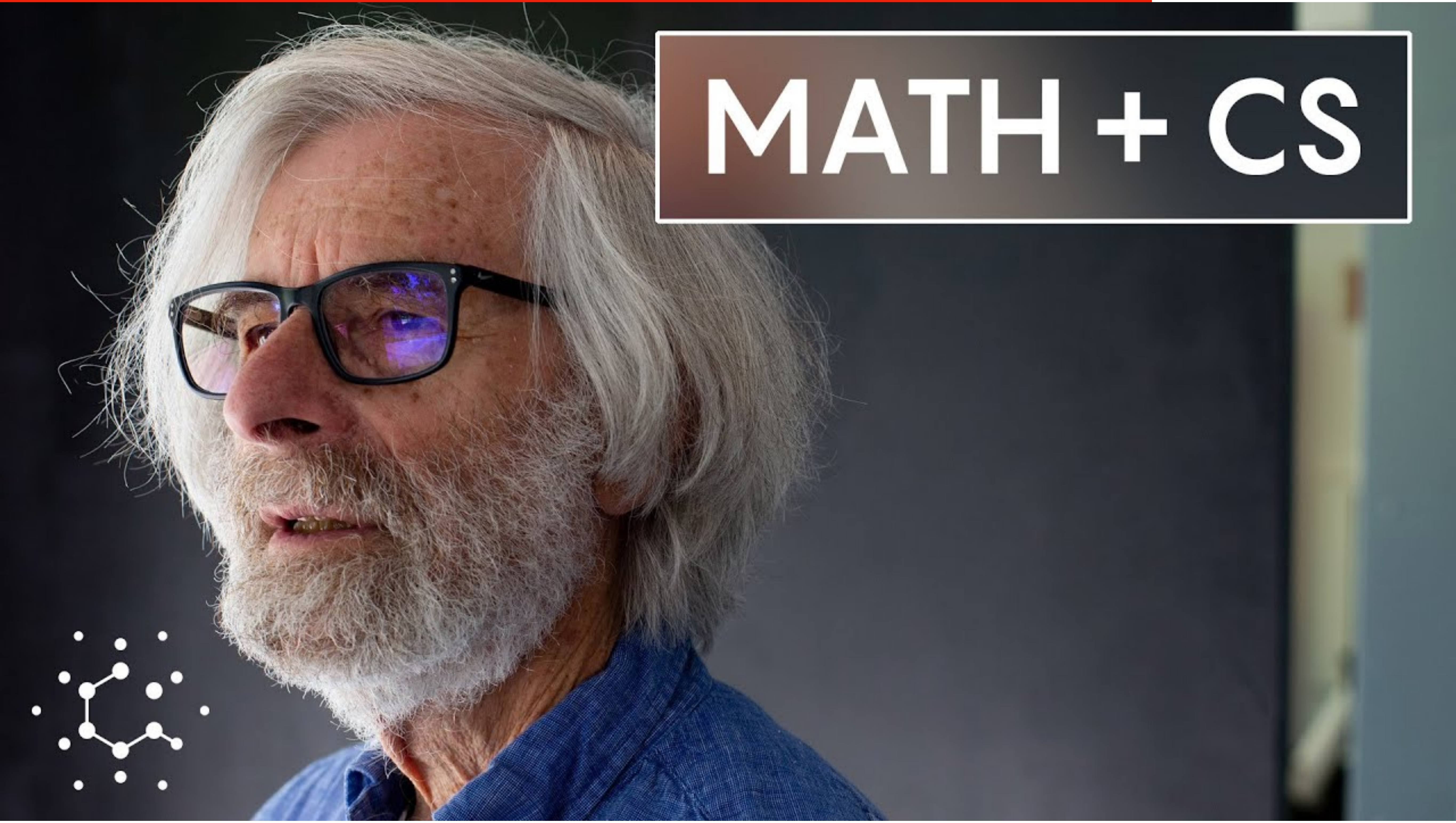
is L, the number of links between websites.

- Since $L \ll p$, the second method is generally more efficient if it converges quickly. Also, system solutions are often unstable due to poor condition numbers for the matrix.

Considerations

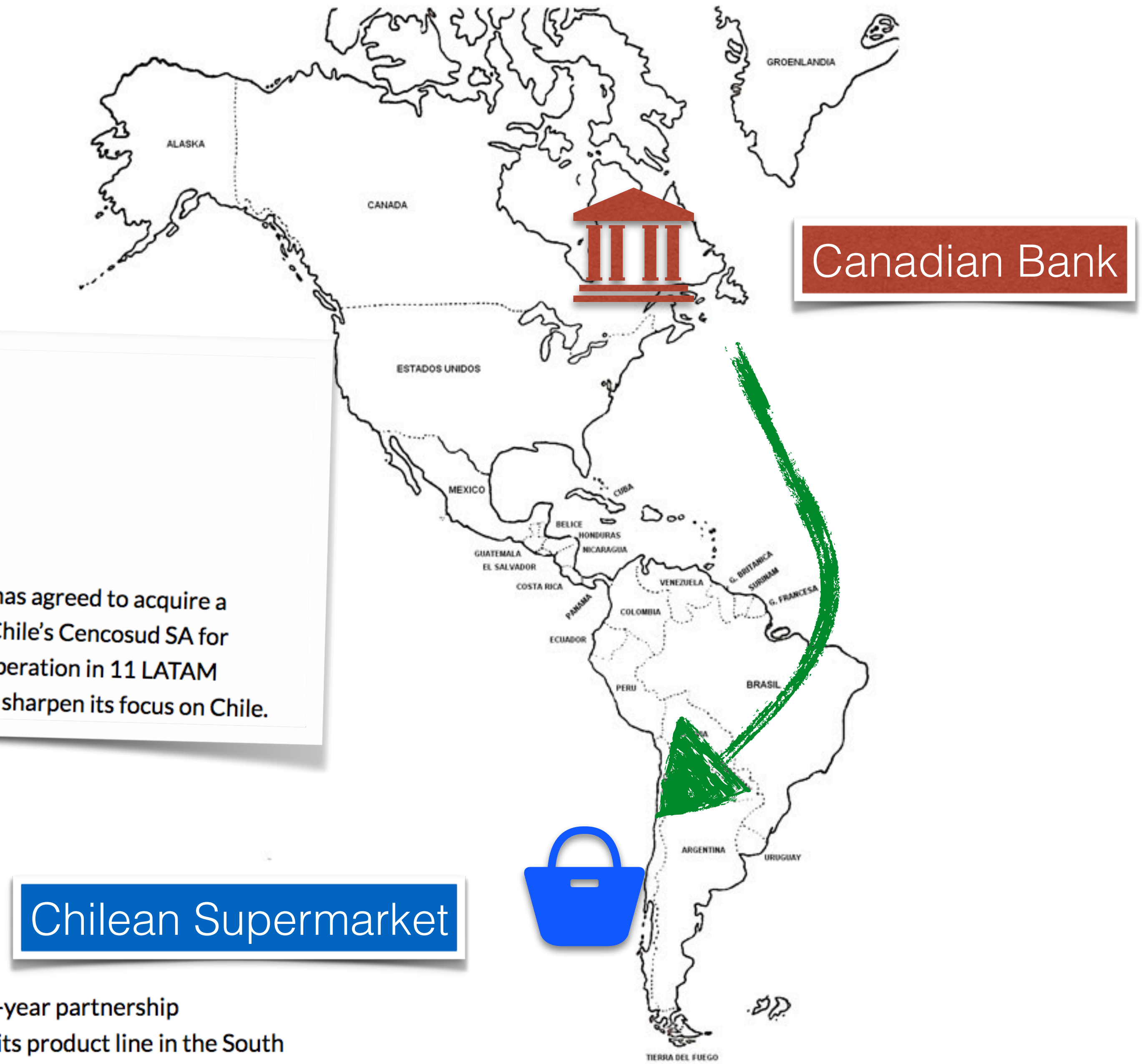
- The matrix A of web site links is very large, the vector of all web sites is very large too.
- Once a ranking system has been calculated, the matrix A of web site links may change as the individual web sites modify their content.
- When updates of A are calculated (from an old A to a new $\tilde{A}$), we simply update the rankings by repeating $r_{n+1} = \tilde{A} \cdot r_n$ starting with r_0 being equal to the old ranking.

Programming vs coding



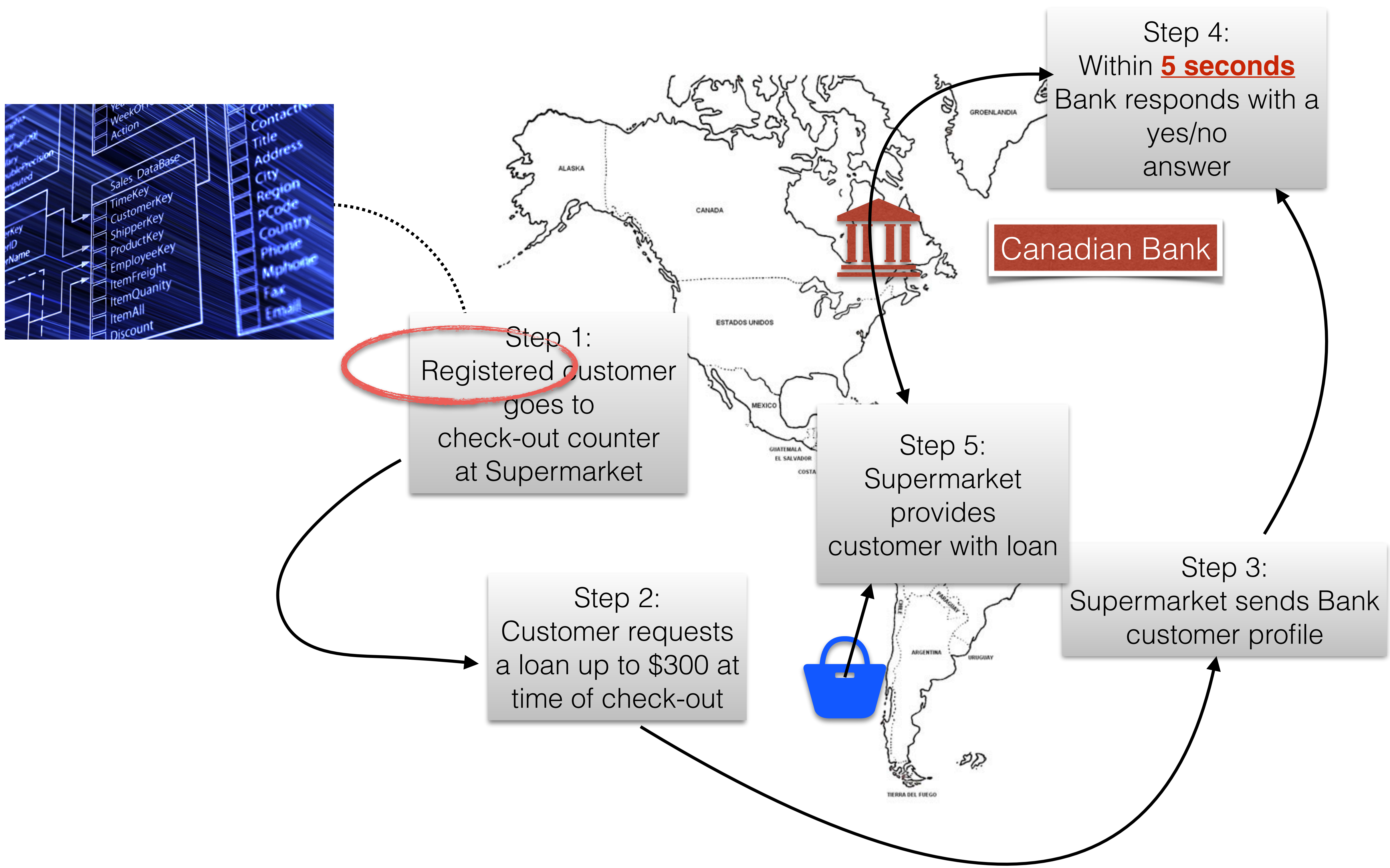
Scotiabank case

Canada's third biggest financial service provider Scotiabank has agreed to acquire a majority stake in the credit card and consumer loans unit of Chile's Cencosud SA for $280 million. Toranto-based bank – which has an extensive operation in 11 LATAM countries including Peru, Colombia, Mexico – says it wants to sharpen its focus on Chile.



In addition to the stake purchase, Scotiabank has sealed a 15-year partnership agreement with the Chilean bank that will allow it to expand its product line in the South American country. Scotiabank's business in Colombia, Banco Colpatria, has a similar partnership arrangement with Cencosud in that market.

Micro-loan fintech analysis



Metadata: the new data

- Cross-referencing of data coming from different business lines
  - Supermarket shoppers who are already banking clients
  - Relationships between commercial lending and retail lending
- Shopping patterns, family relationships, etc.
- There is a need to extract the meaning from large data sets and find a way to
  - enhance the different business lines
  - provide enhanced risk-based systems
    - The transition from **risk-based** to **rules-based** processes and systems will be increasingly harder and give rise to regulatory challenges unless regulators evolve accordingly
- As a consequence, the challenges of having to integrate many disparate data centres gives rise to the opportunity to harness data relationships: metadata.

Data issues

Customer name
Address
Profession
Income
Address
Shopping patterns by
- time
- goods purchased
- store locations

.... etc...

Customer Database

Subject to manipulation

Dynamic database: updates in real time

Decision and monitoring systems updated in real time

Not Bank-owned

Machine Friendly
Human un-friendly

Unstructured Data

Customer name vs Household name
Customer address vs. store address
Mining the customer network

Cybersecurity

- Data transmission between supermarket and bank subject to external interference
- How to detect a breach?

  Time frame: seconds

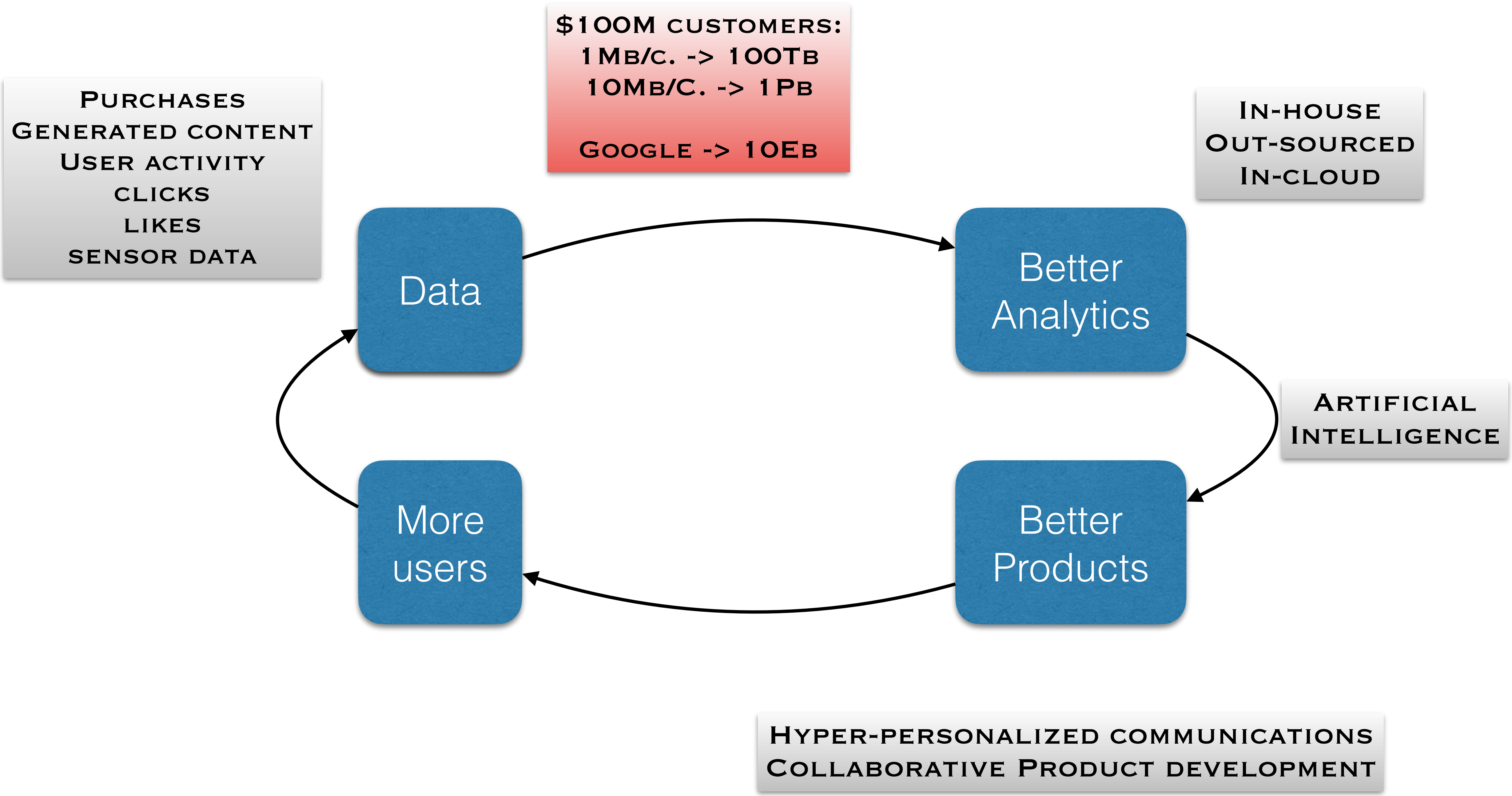
- Actions if breach occurs
  - Who gets notified?
    - Law enforcement?

      Time frame: minutes

    - Customers?

      Time frame: hours

The Data virtuous cycle



Conclusions

- It pays to be smart
- To simply have an algorithm is often not enough
- Mathematical structures save time
- Time is more important than money (?)

Market structures

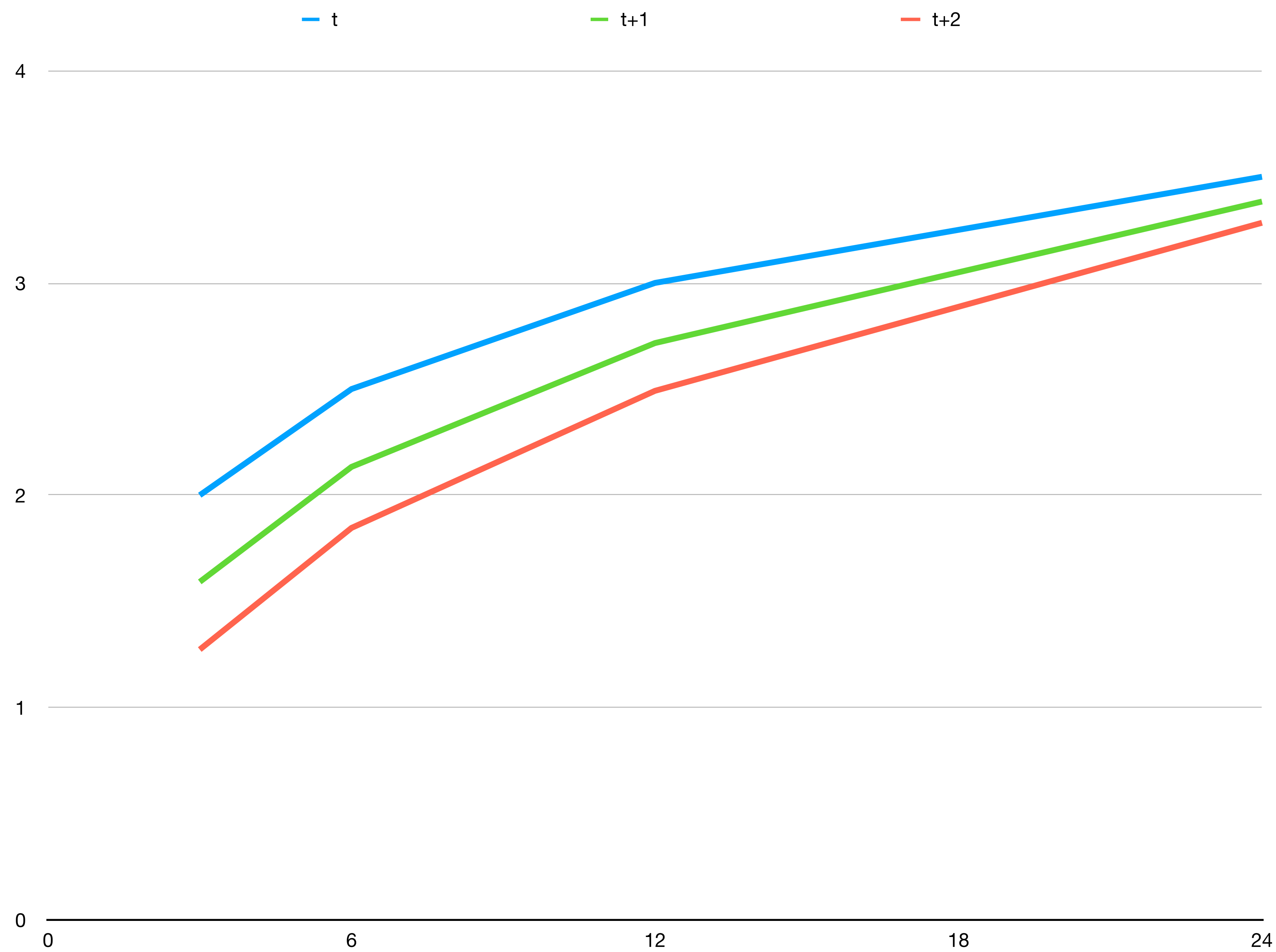
The Music of Financial Markets

Market factors

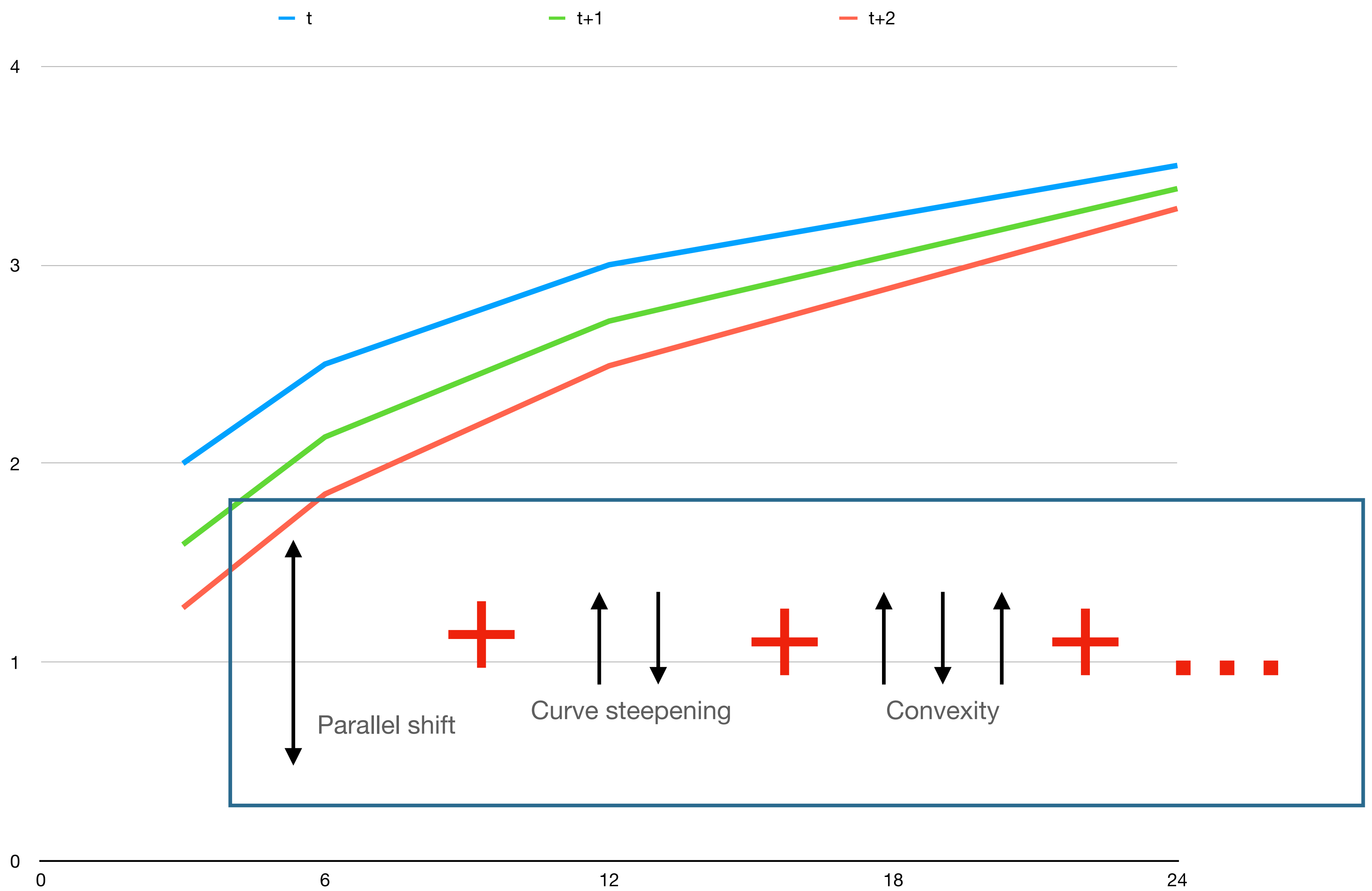
If you are a stock investor…

…why do you care if the stock index goes up or down?

Yield curve evolution



Yield risk factors



Fixed income factors

$$r_{t,T}, \quad t \geq 0, \quad T \geq 0$$

$$\frac{r_{t,T} - r_{t-1,T}}{r_{t-1,T}}, \quad t \geq 1, \quad T \geq 0$$

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & & \\ \vdots & \cdots & \ddots & \cdots \\ \sigma_{n,1} & \cdots & & \sigma_n^2 \end{bmatrix}$$

$$\begin{matrix} \lambda_1 & v_1 \\ \lambda_2 & v_2 \\ \lambda_3 & v_3 \\ \vdots & \cdots \end{matrix}$$

Stock markets

Significance

Eigenvalue ($\times 10^{-3}$)	2.027	0.336	0.269	0.216	0.192	0.155	0.137	0.127	0.103	0.077
% Explained Var.	55.7 %	9.2 %	7.4 %	5.9 %	5.3 %	4.3 %	3.8 %	3.5 %	2.8 %	2.1 %
	-0.320	0.079	-0.134	-0.024	0.004	-0.123	0.224	0.583	-0.517	0.448
	-0.359	0.190	0.098	-0.052	-0.011	-0.074	0.354	0.306	0.772	0.031
	-0.243	0.042	-0.273	0.047	0.864	0.144	-0.299	0.006	0.075	-0.001
	-0.316	0.094	-0.215	-0.050	-0.113	-0.047	0.061	0.193	-0.222	-0.859
Eigenvector	-0.274	0.178	-0.443	-0.028	-0.406	0.660	-0.190	-0.139	0.075	0.186
	-0.140	0.044	-0.240	0.121	0.149	-0.047	0.720	-0.584	-0.141	0.049
	-0.385	-0.903	0.025	-0.159	-0.059	0.032	-0.004	-0.053	0.054	0.026
	-0.457	0.296	0.590	-0.438	0.032	0.032	-0.141	-0.319	-0.197	0.041
	-0.247	0.102	-0.361	0.046	-0.210	-0.715	-0.379	-0.253	0.106	0.145
	-0.312	-0.005	0.344	0.870	-0.070	0.069	-0.111	-0.034	-0.057	-0.018

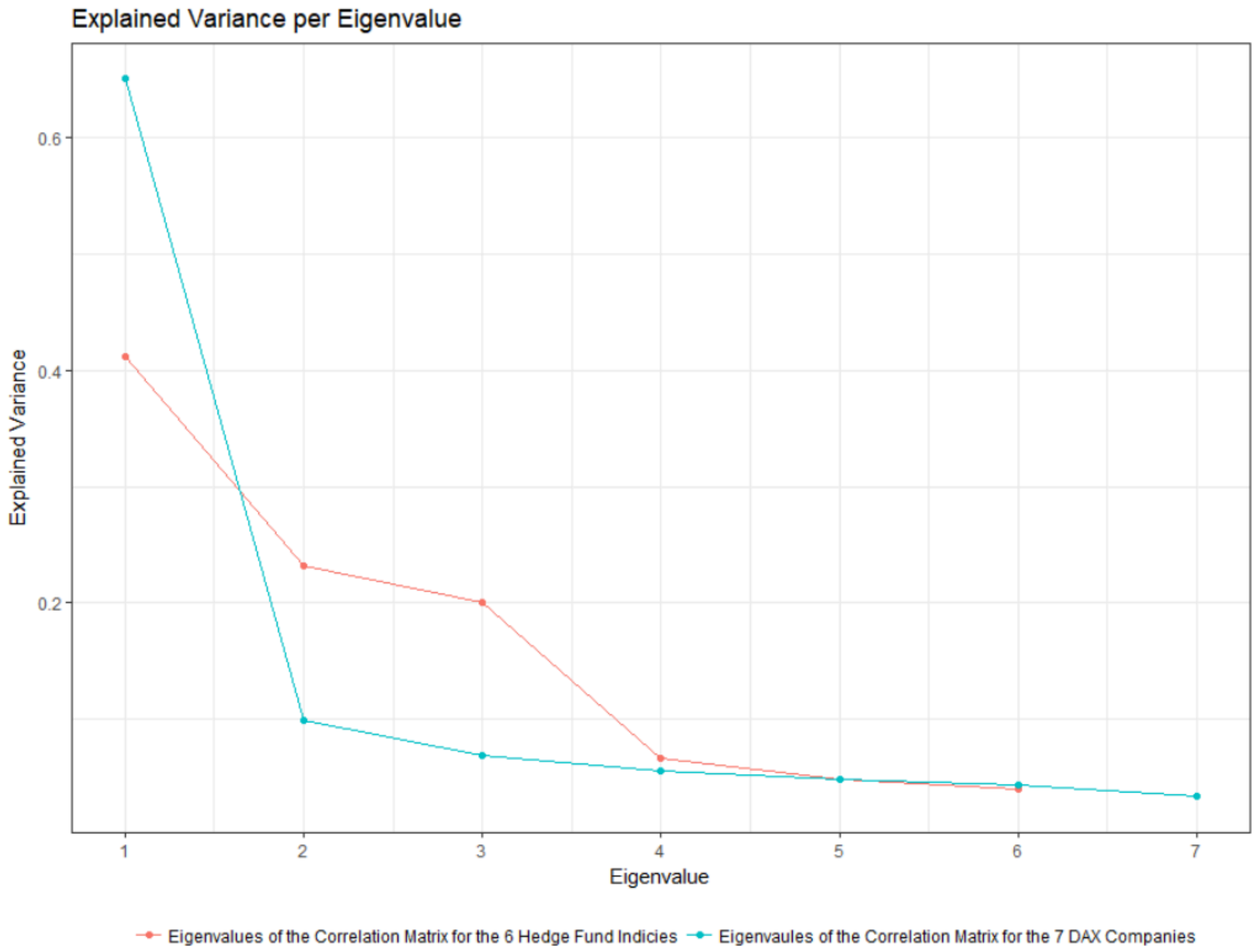
Principal components →

Market Factors

- Securities prices evolve in random ways but with a strong internal dependence structure
- That interdependence is an important market invariant
- It can be calculated using spectral theory
  - In their simplest form, with eigenvalues and eigenfunctions
  - In more complex terms, with neural embeddings, random forests and other

Music vs. noise

Stocks vs hedge funds

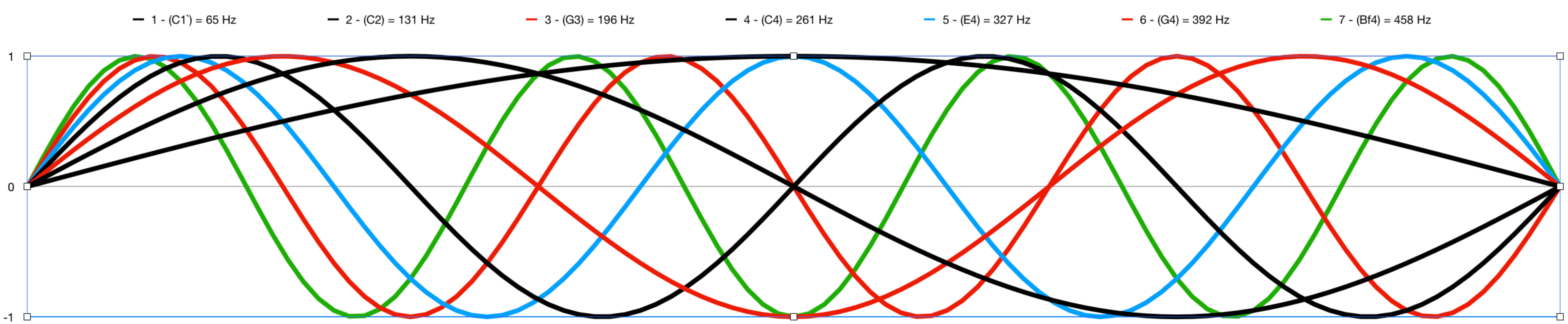


Music

Wave operator

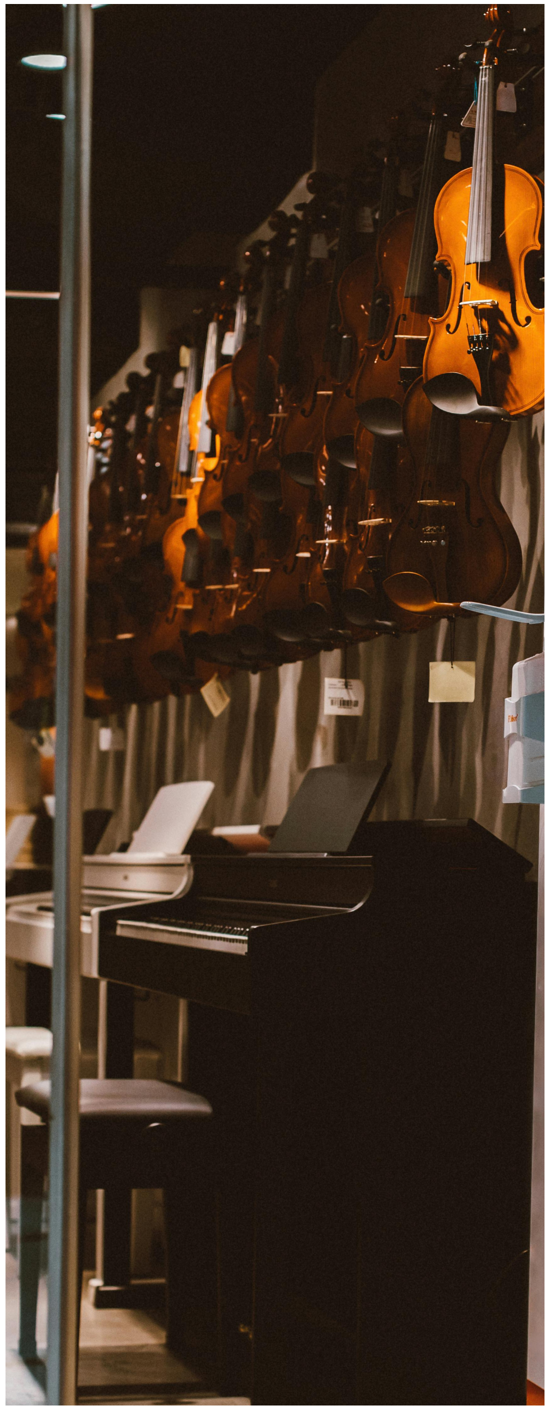
Fundamental Frequency

$$-\partial_x^2 \ \sin(n \cdot x) = n^2 \ \sin(n \cdot x)$$



Pythagoras’ piano

Pythagoras' piano



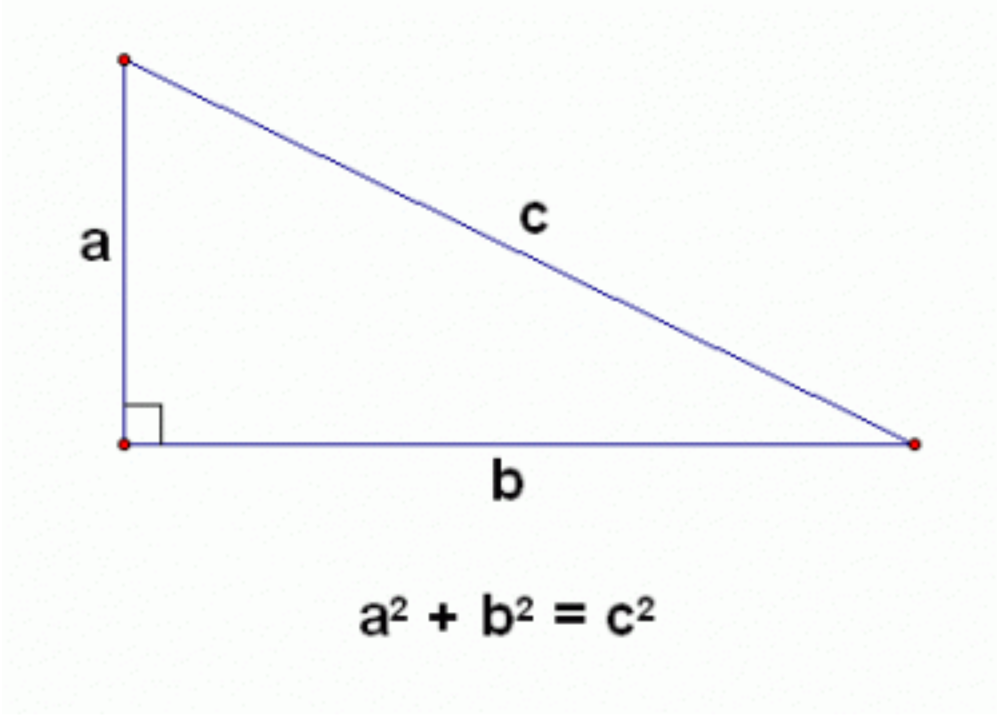
Western music

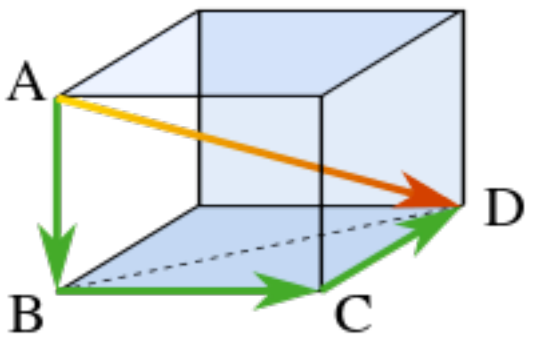
- There should be an infinite number of notes ?!
- In Western modern music, we have just 12 notes:

$$3^{12} \approx 2^{19}$$

- Take a note; build a progression of its dominants (“fifths"); you will, after 12 twelve iterations, get very close to the same note you started with, 19 octaves higher.
- "Well-tempered klavier” (J. S. Bach)

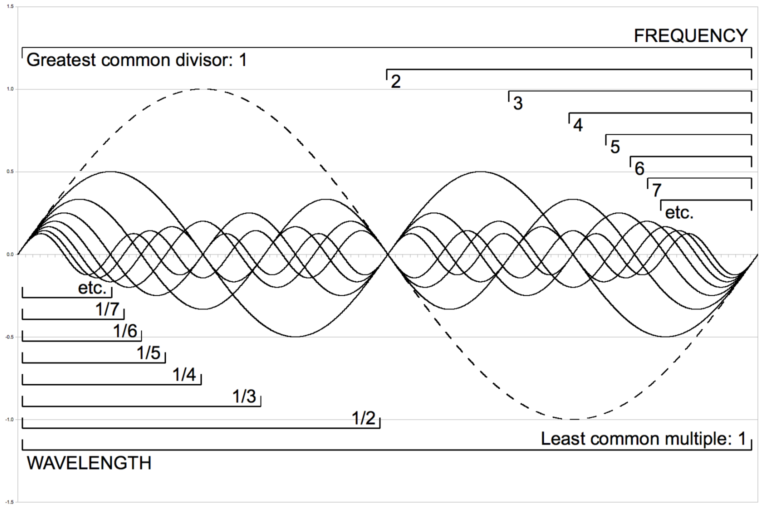
Pythagoras theorem



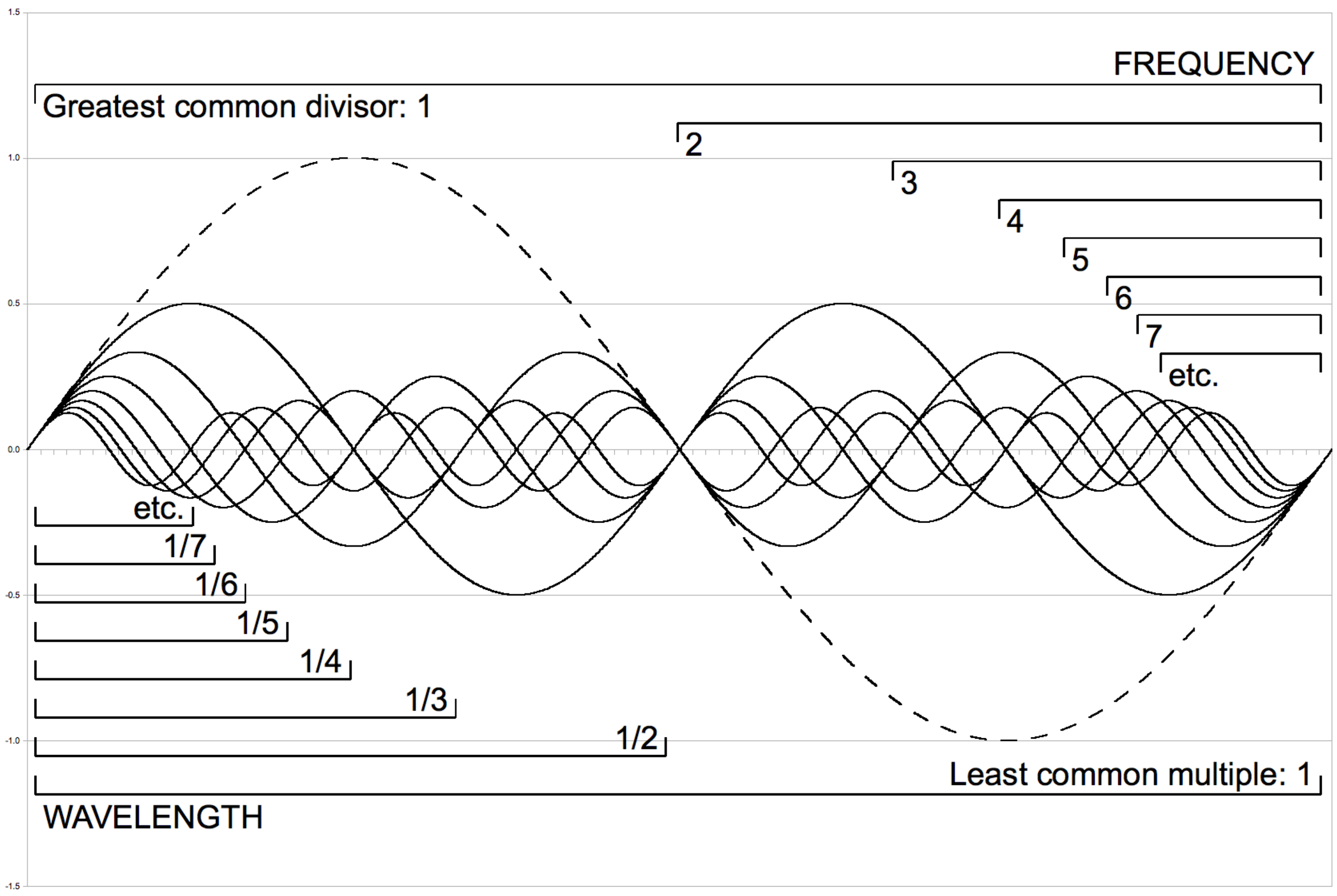


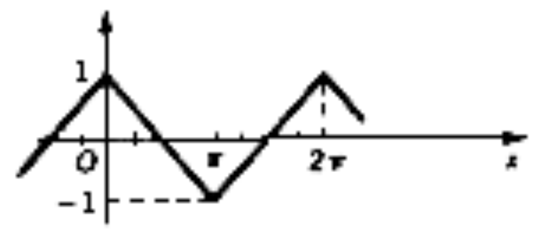
$$\int_0^1 |f(x)|^2\,dx = \sum_n |\hat{f}(n)|^2$$

$$\int_R |f(x)|^2\,dx = \int_R |\hat{f}(\xi)|^2\,d\xi$$



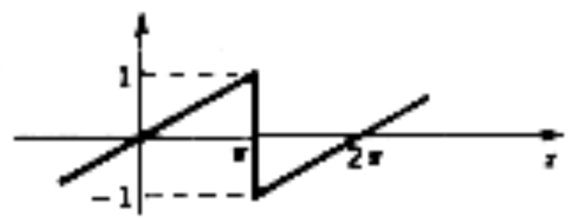
Music


1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
FREQUENCY
Greatest common divisor: 1
2
3
4
5
6
7
etc.
etc.
1/7
1/6
1/5
1/4
1/3
1/2
Least common multiple: 1
WAVELENGTH



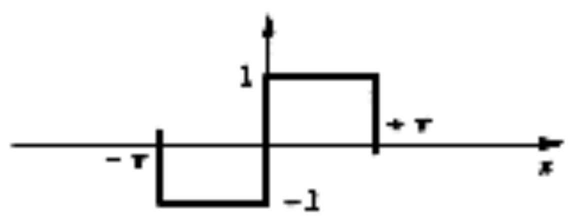
Triangular wave:

$$\frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}\cos(2n+1)x$$



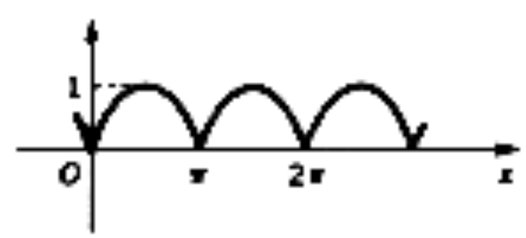
Rectangular sawtooth wave:

$$\frac{2}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}\sin nx$$



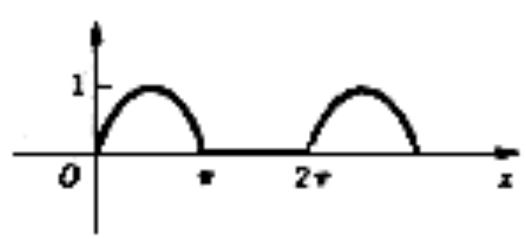
Square wave:

$$\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin(2n+1)x$$



Absolute value sine wave:

$$\frac{2}{\pi}-\frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{4n^2-1}\cos 2nx$$



Half sine wave:

$$\frac{1}{\pi}+\frac{1}{2}\sin x-\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{4n^2-1}\cos 2nx$$

Köszönöm

תודה

Спасибі

Thanks

ありがとう

Gracias

ευχαριστώ

감사해요

cảm ơn

Kösz

Dankon

Teşekkürler

Danke

متشکرم

Merci

Grazie

tack

rahmat

Takk

谢谢

ਧੰਨਵਾਦ

Gratias

dzięki

спасибо

धन्यवाद

ขอบใจ

شكرًا

asante

dankie

շնորհակալություն