Downloaded from http://rsif.royalsocietypublishing.org/ on May 27, 2015

INTERFACE

rsif.royalsocietypublishing.org

Research

Punctuated equilibrium in the large-scale evolution of programming languages†

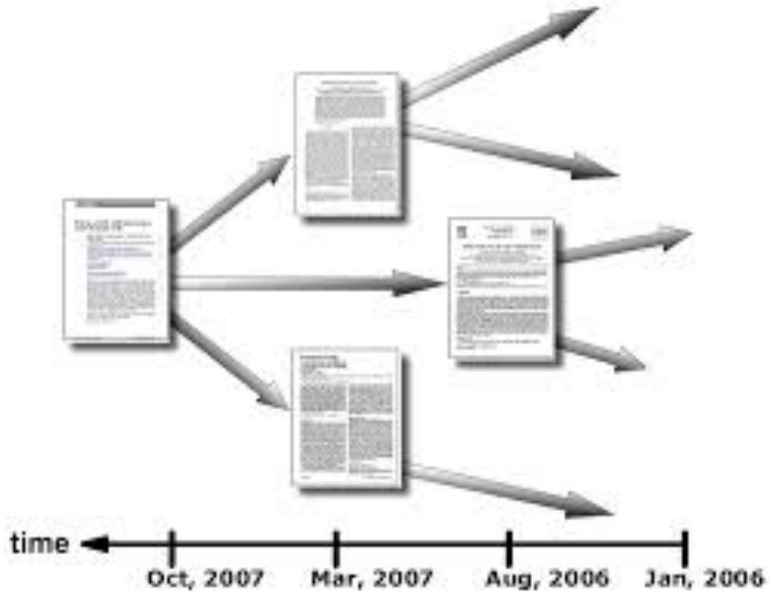
Sergi Valverde[1,2] and Ricard V. Solé[1,2,3]

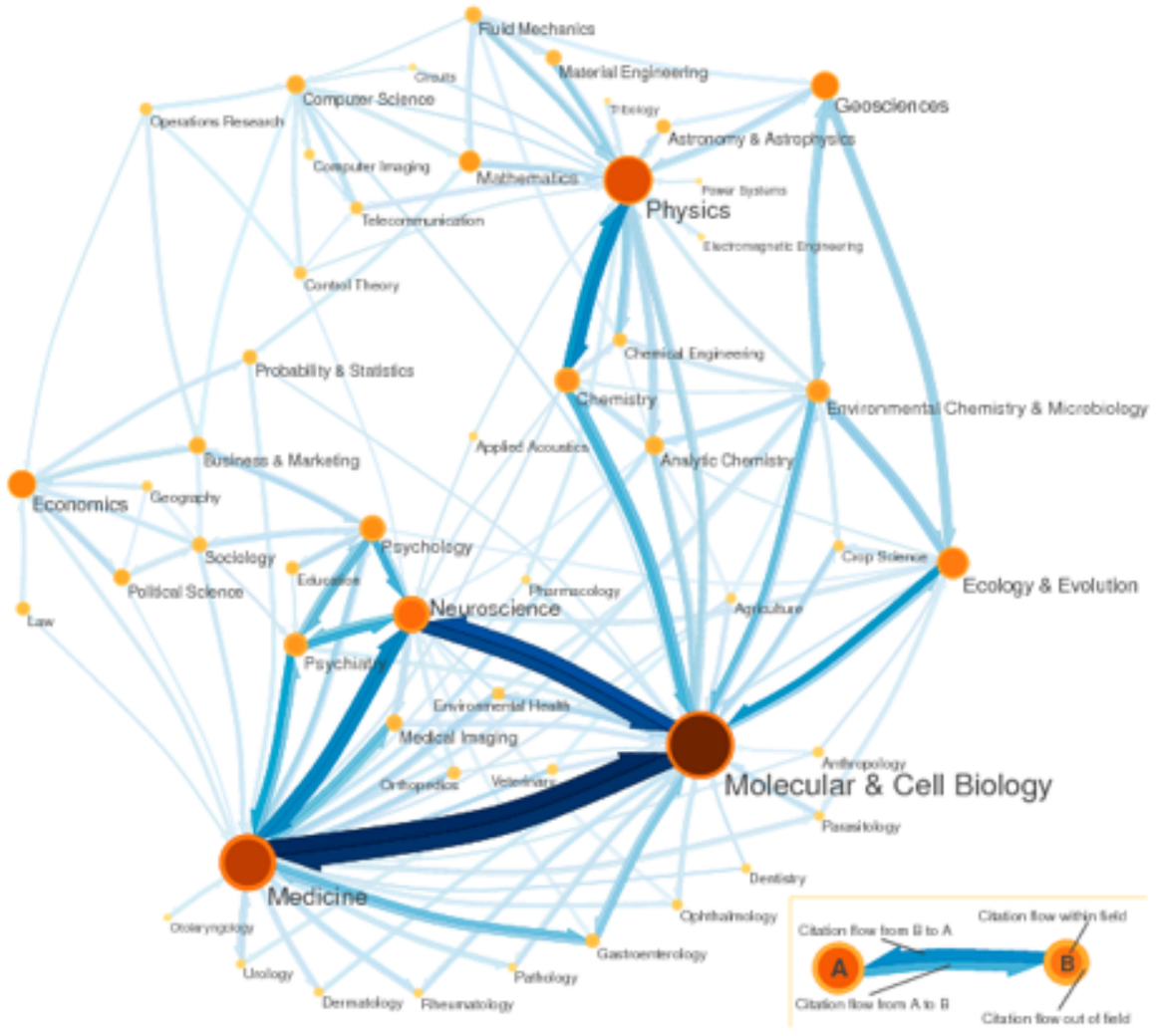
[1]ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Dr Aiguader 80, 08003 Barcelona, Spain
[2]Institut de Biologia Evolutiva, CSIC-UPF, Pg Maritim de la Barceloneta 37, 08003 Barcelona, Spain
[3]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Journal Club (KKY)
9/17/2015

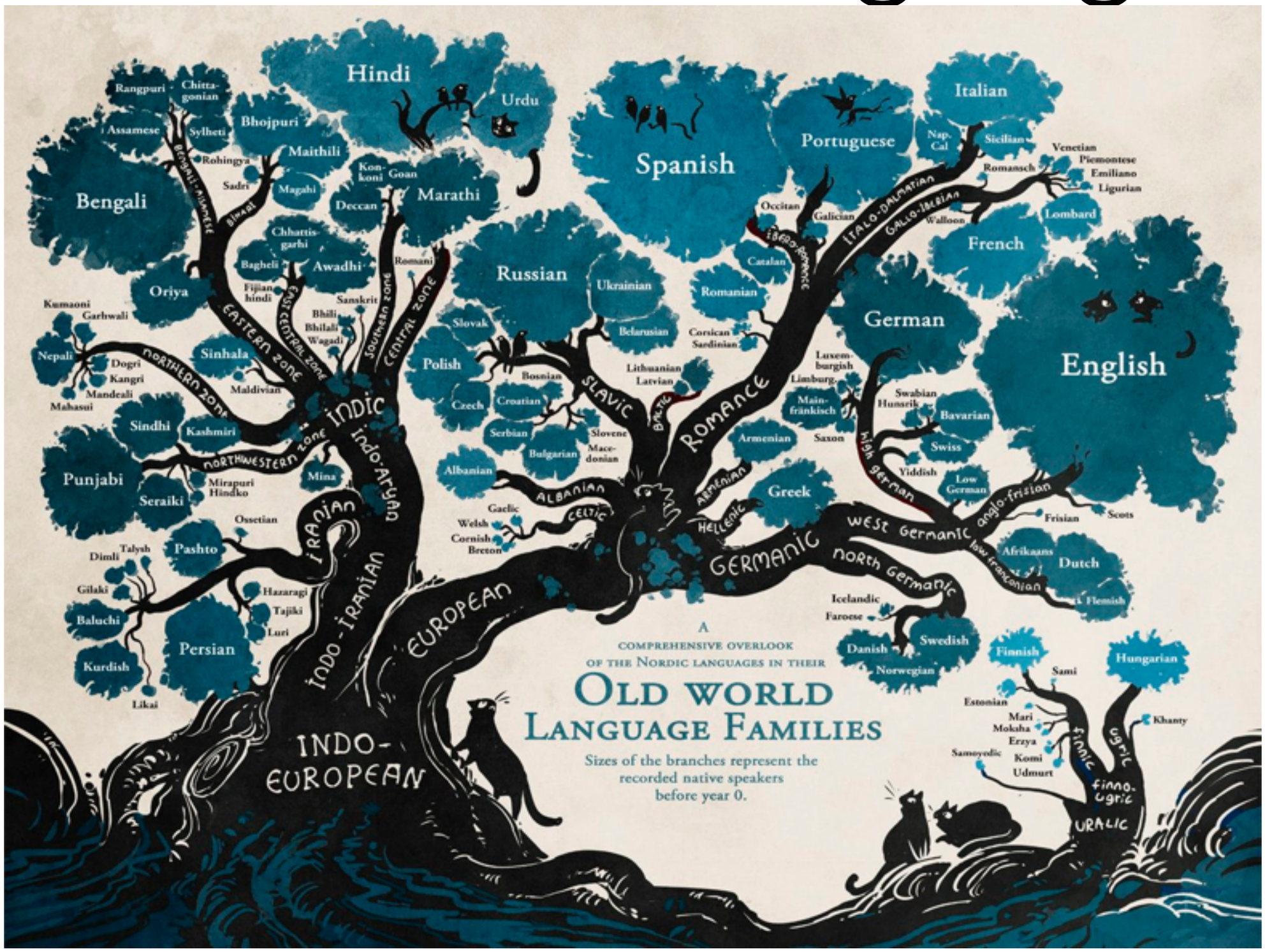
Evolution of technology resembles biological evolution

the same for the accumulation/ evolution of knowledge





also human languages



Evolution of Programming Languages (PLs)

- Compared to human languages, it is harder to quantify the divergence in terms of vocabulary and syntax.

- Programming language is kind of technology

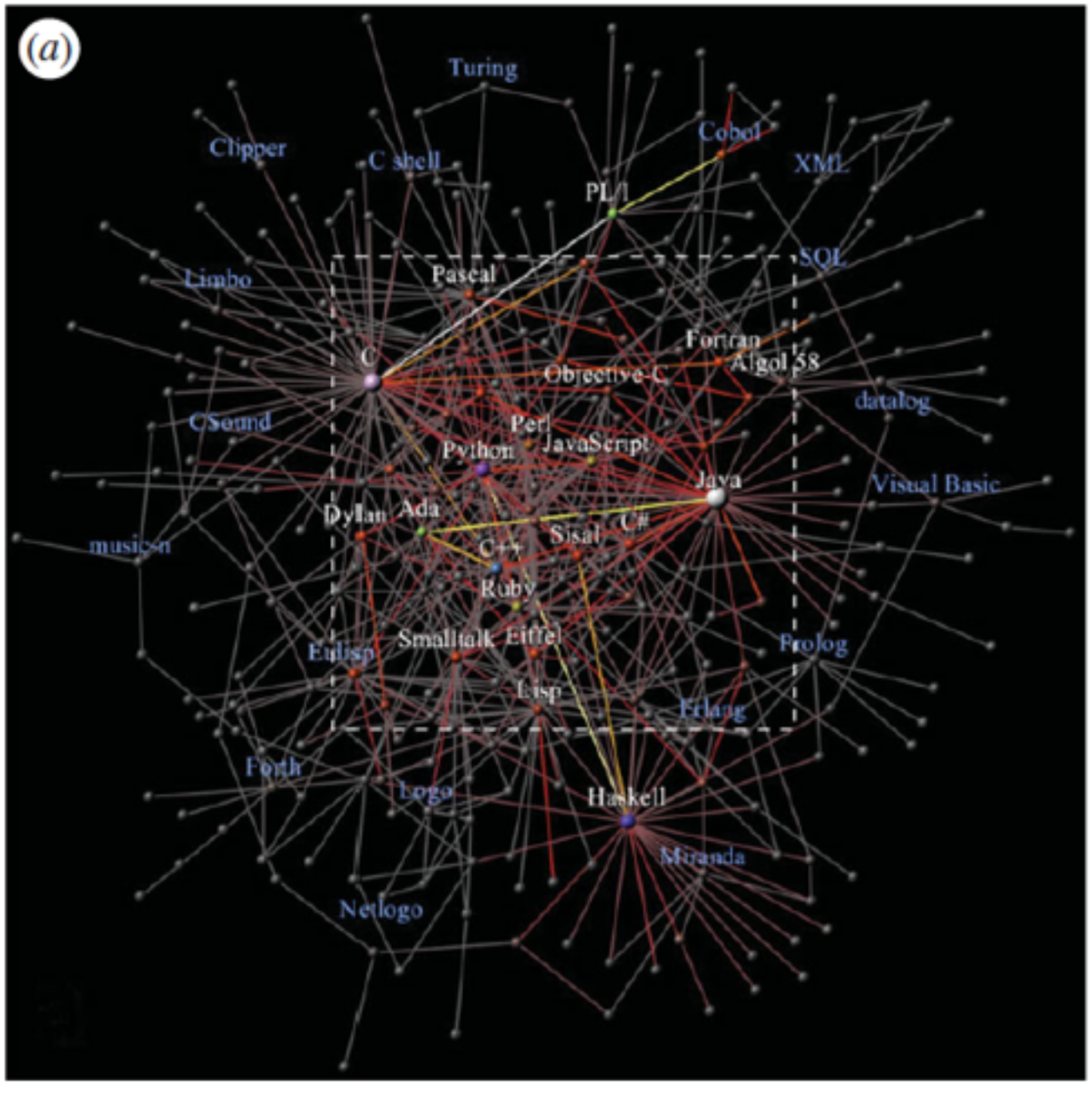
- As the decouple of hardware and software, it is a kind of knowledge evolution

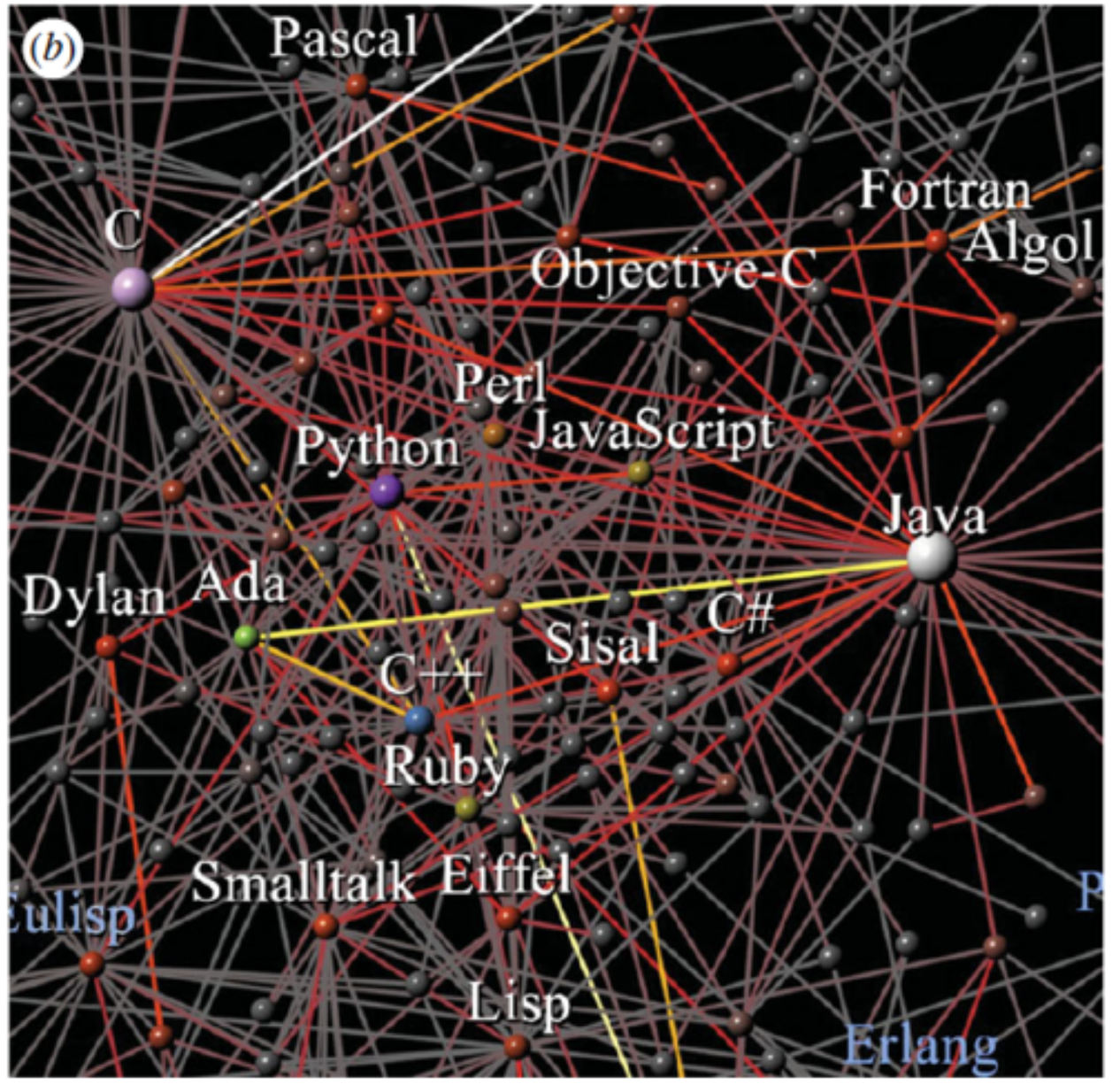
- Like all scientific advances depend on previous works, new programming languages depend on existing languages in certain fuzzy ways

Dataset of programming languages

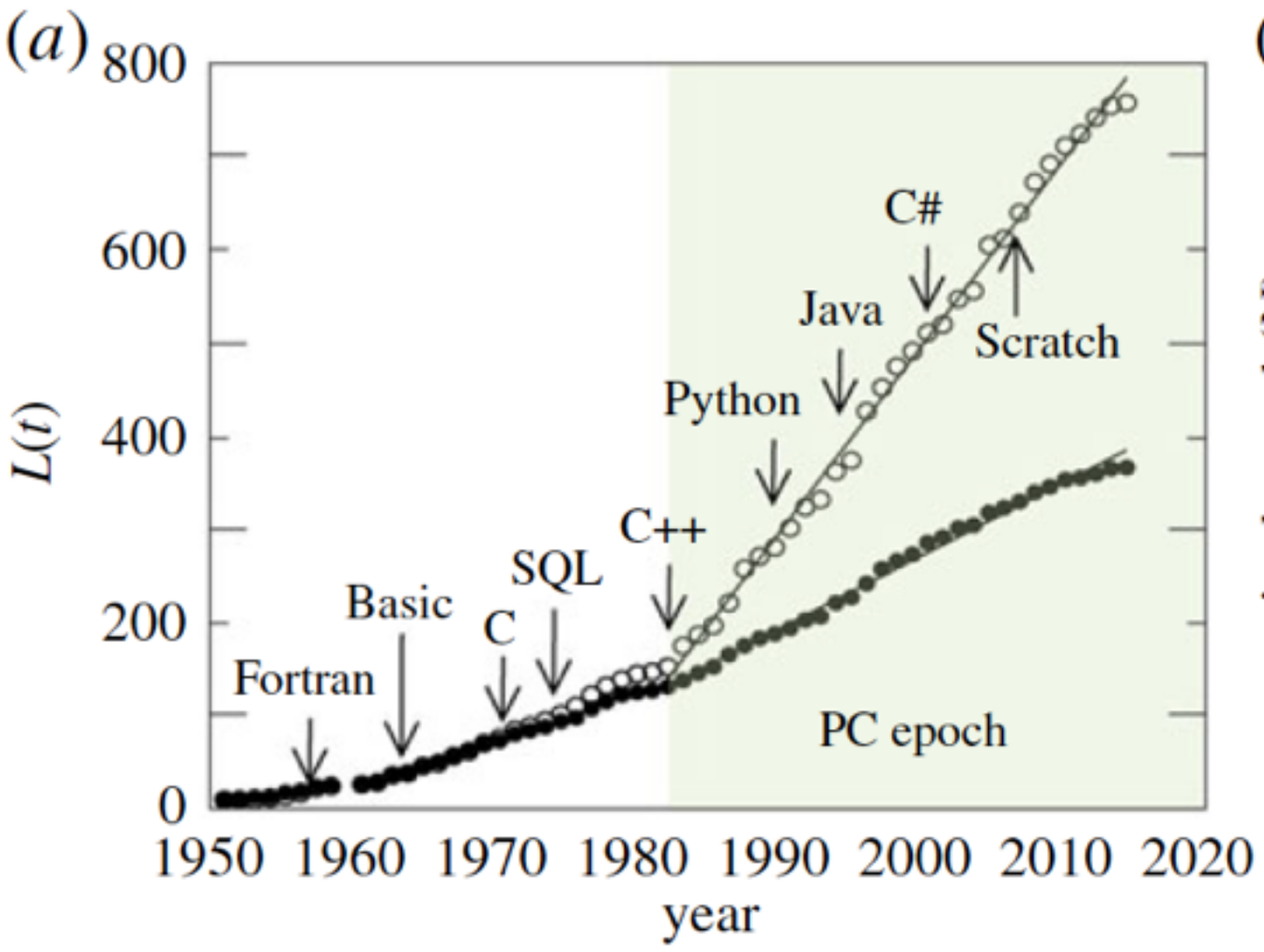
- Downloaded and parsed all the pages describing programming languages in wikipedia.
- 347 programming languages, from 1952 to 2010.
- Connections among languages are defined in terms of influence (not usage).
- A directed graph representing how one influence another (not phylogenetic map)

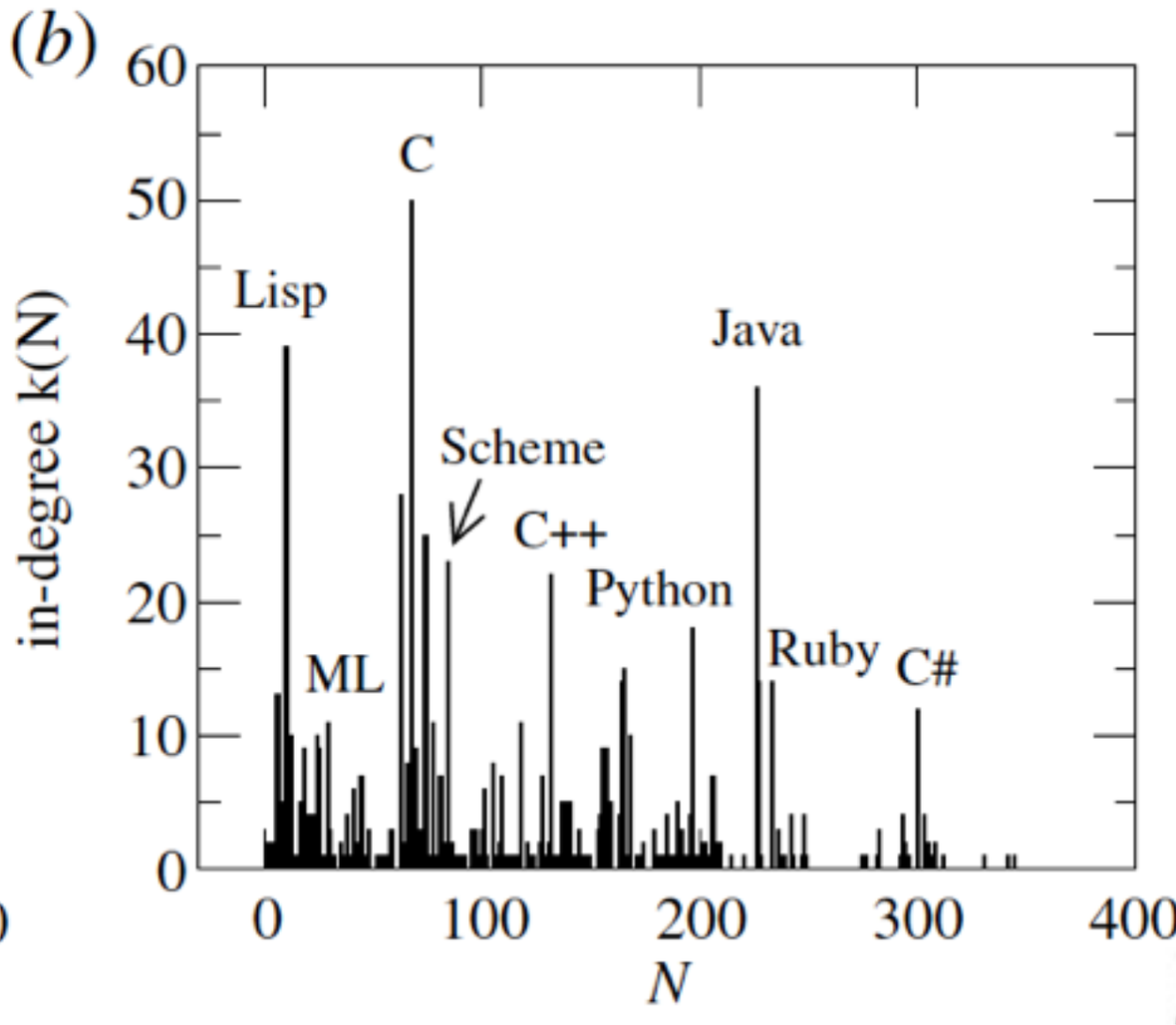
Programming languages form a dependency network



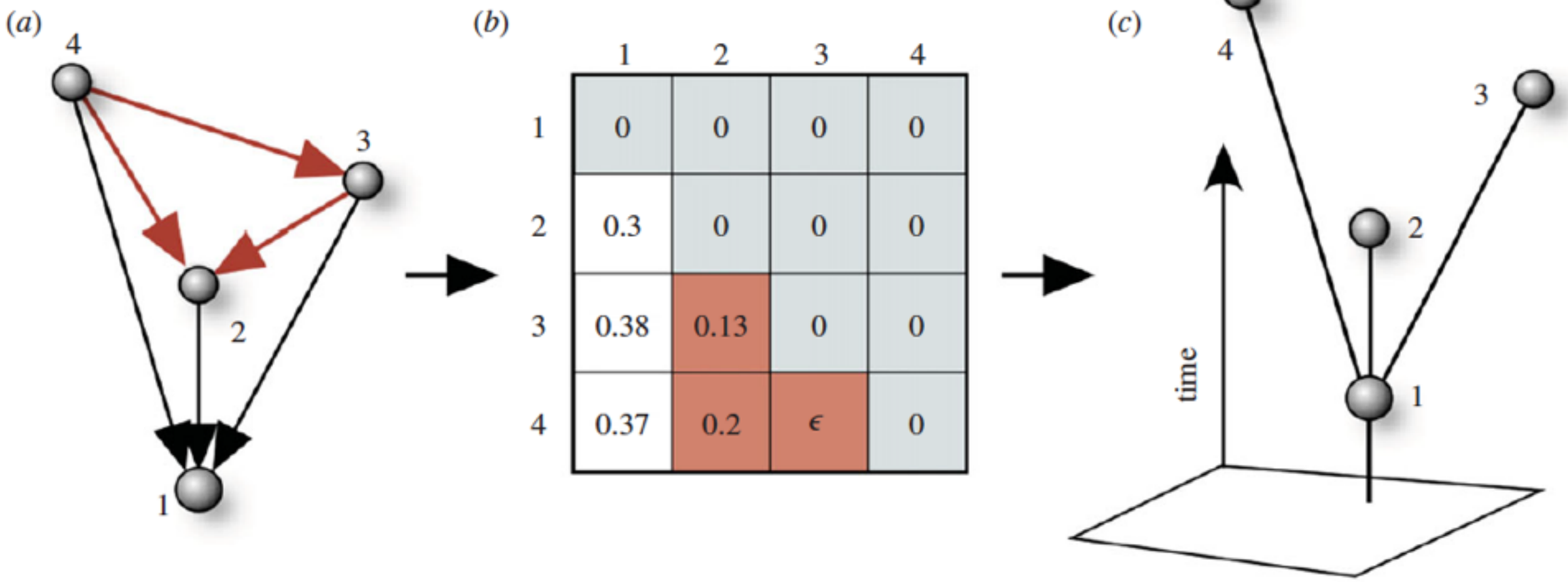


Statistics of the network

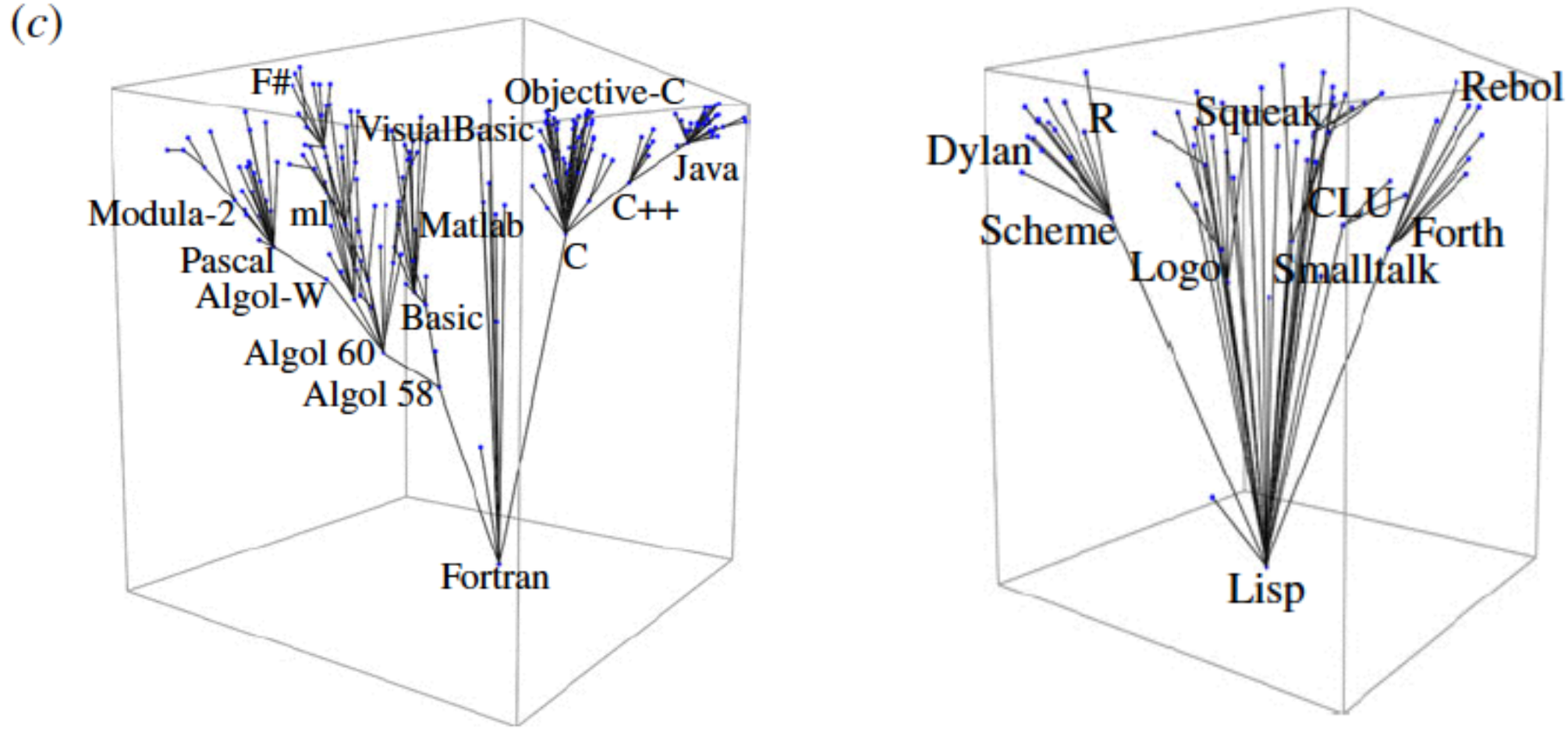




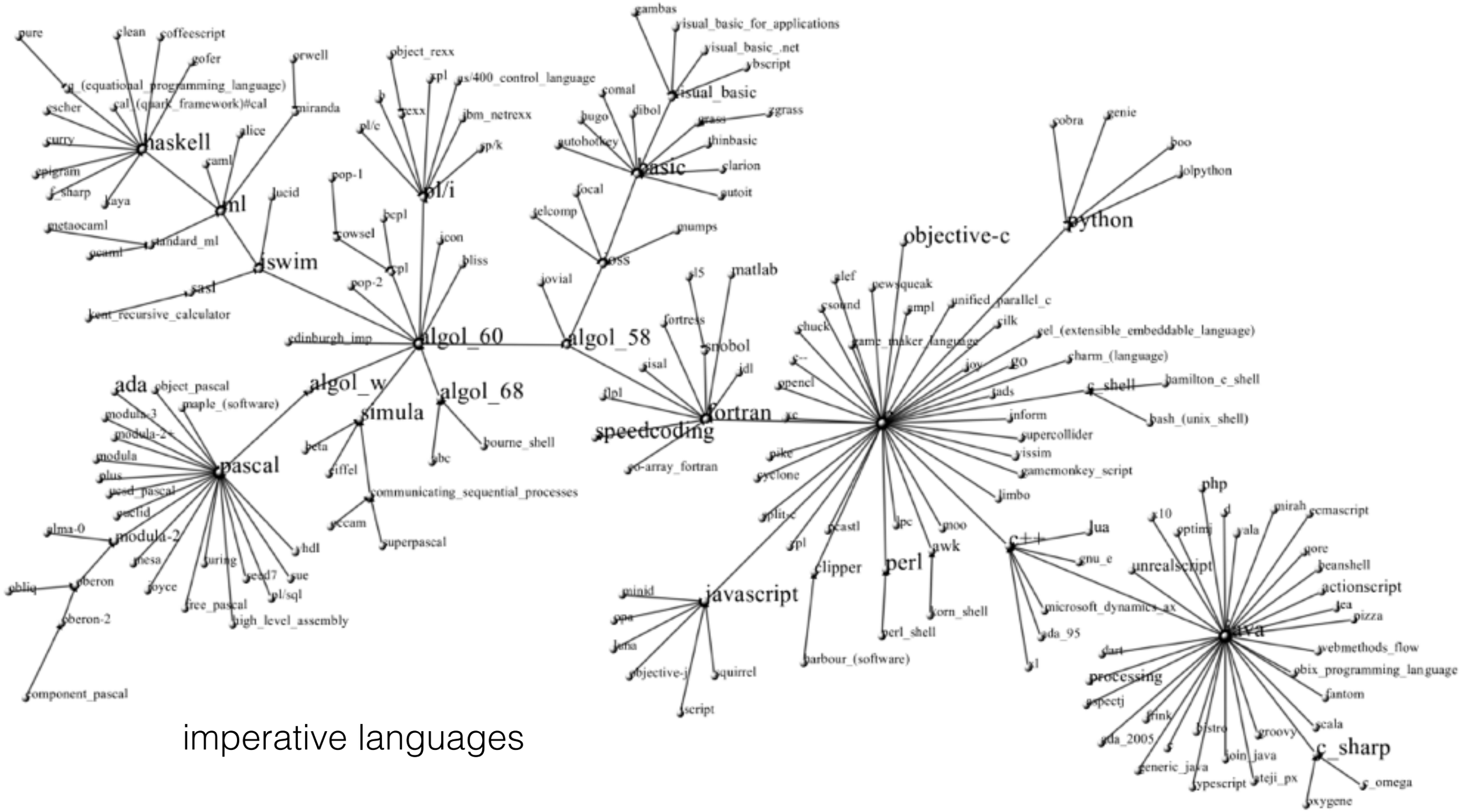
Transform dependency network to a phylogenetic tree



Network reveals clades of Programming languages



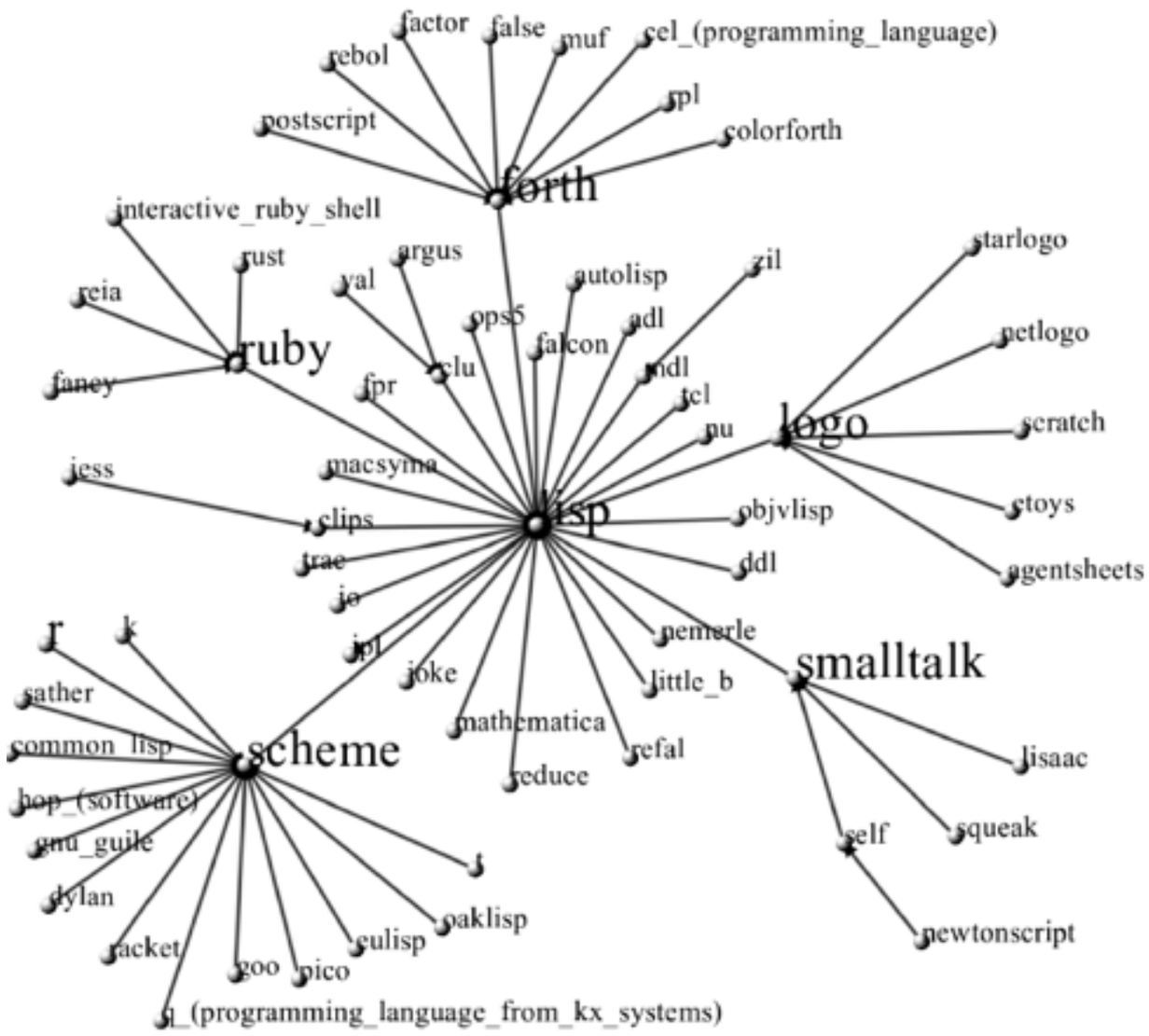
Network reveals clades of Programming languages



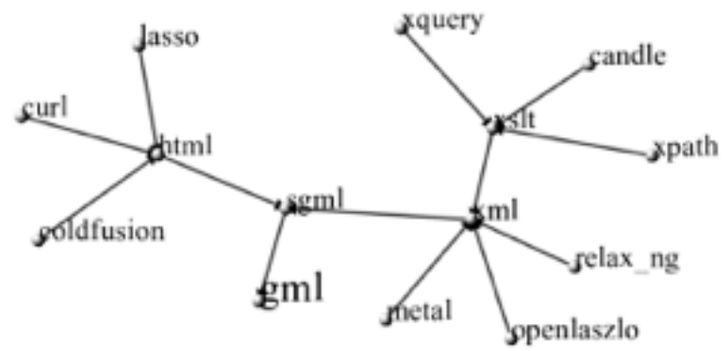
imperative languages

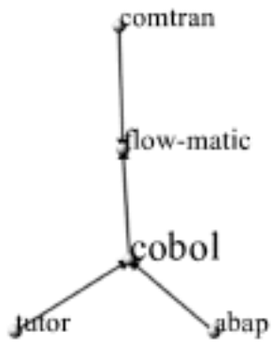
Network reveals clades of Programming languages

functional languages

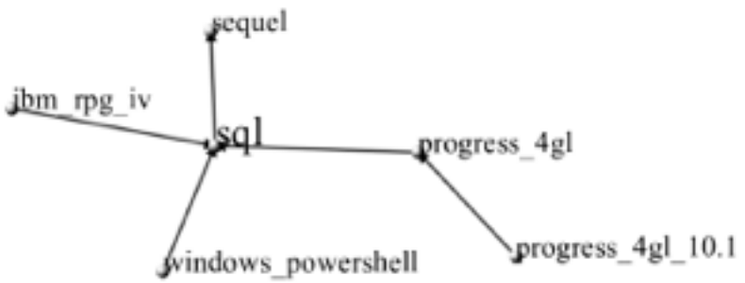


data definition languages

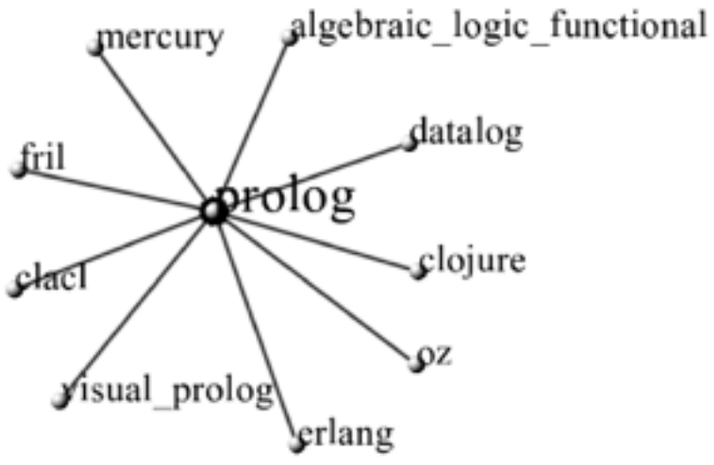


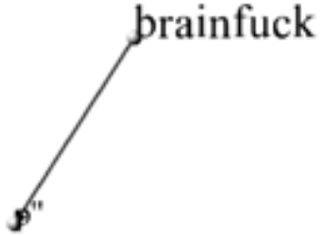


and many more..

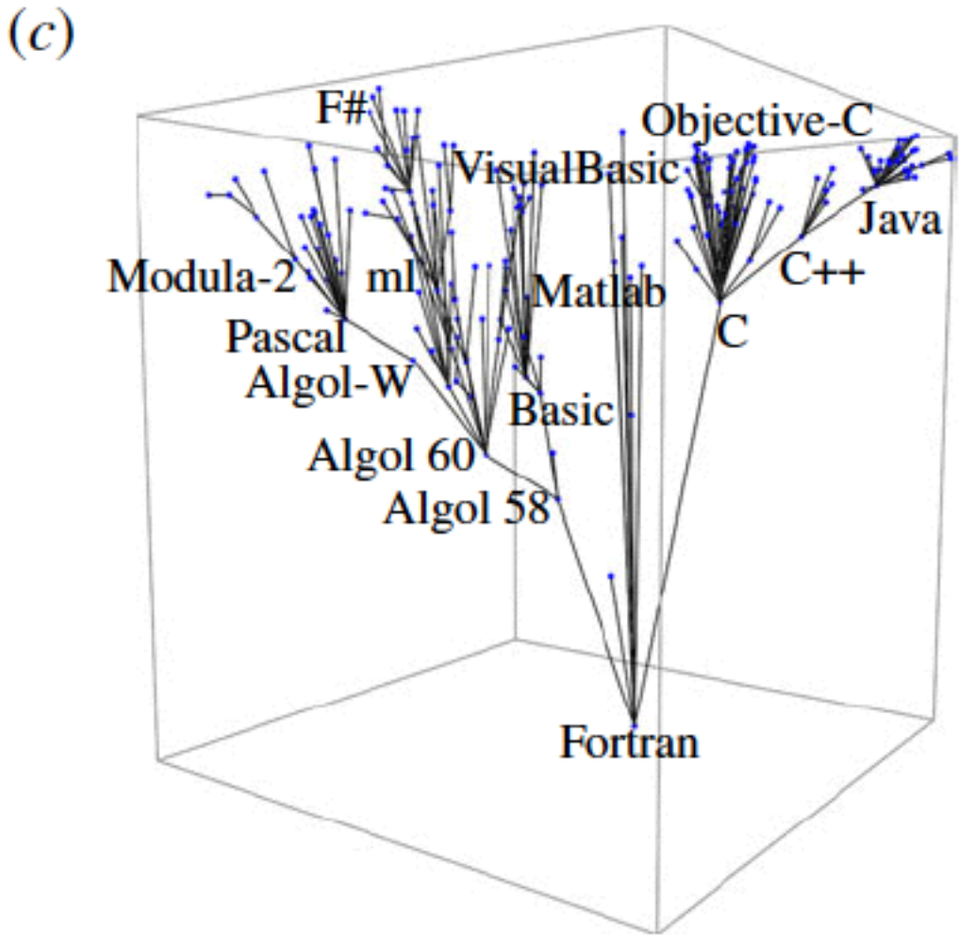


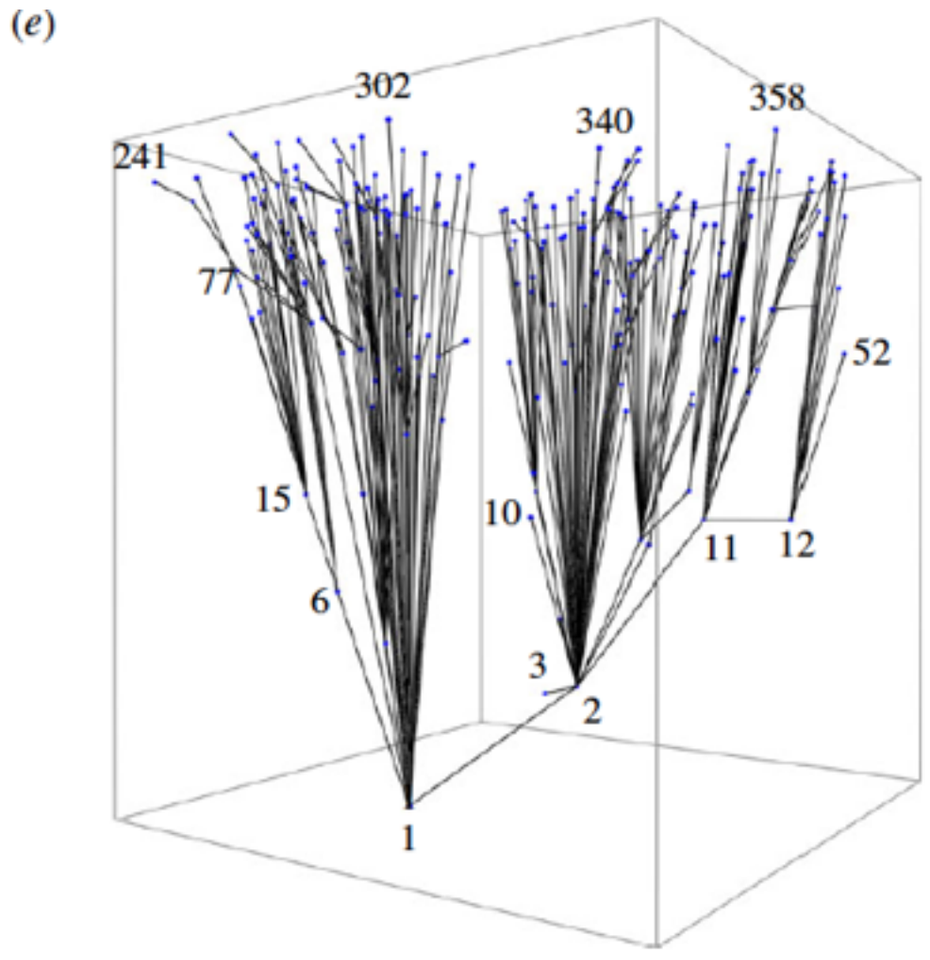
logic languages





Tree imbalance





$$\langle d \rangle = \frac{1}{N} \sum_i d(r, i),$$

Punctuated equilibrium

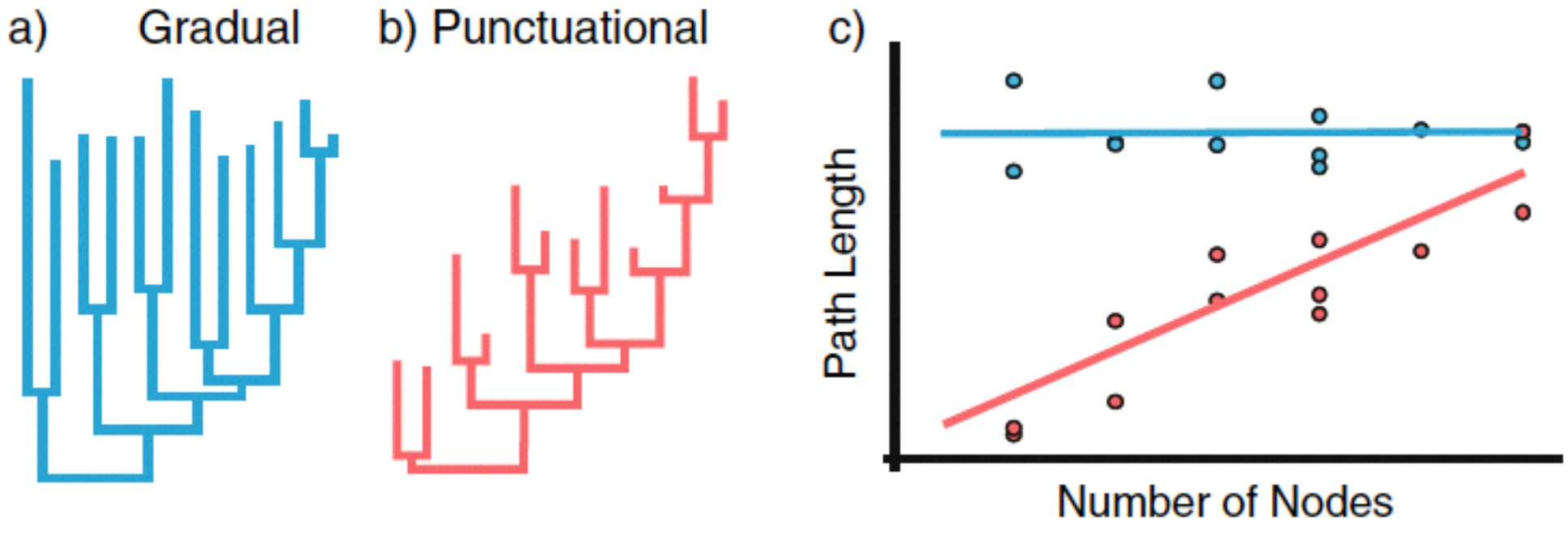
- A longstanding debate in evolutionary biology concerns whether species diverge gradually through time (Darwin's gradualistic view) or by rapid punctuated bursts at the time of speciation.

- The theory of punctuated equilibrium (Stephen Jay Gould et al.) states that evolutionary change is characterized by short periods of rapid evolution followed by longer periods of stasis in which no change occurs. The majority of evolutionary change occurs at or around the time of speciation.

Punctuated equilibrium

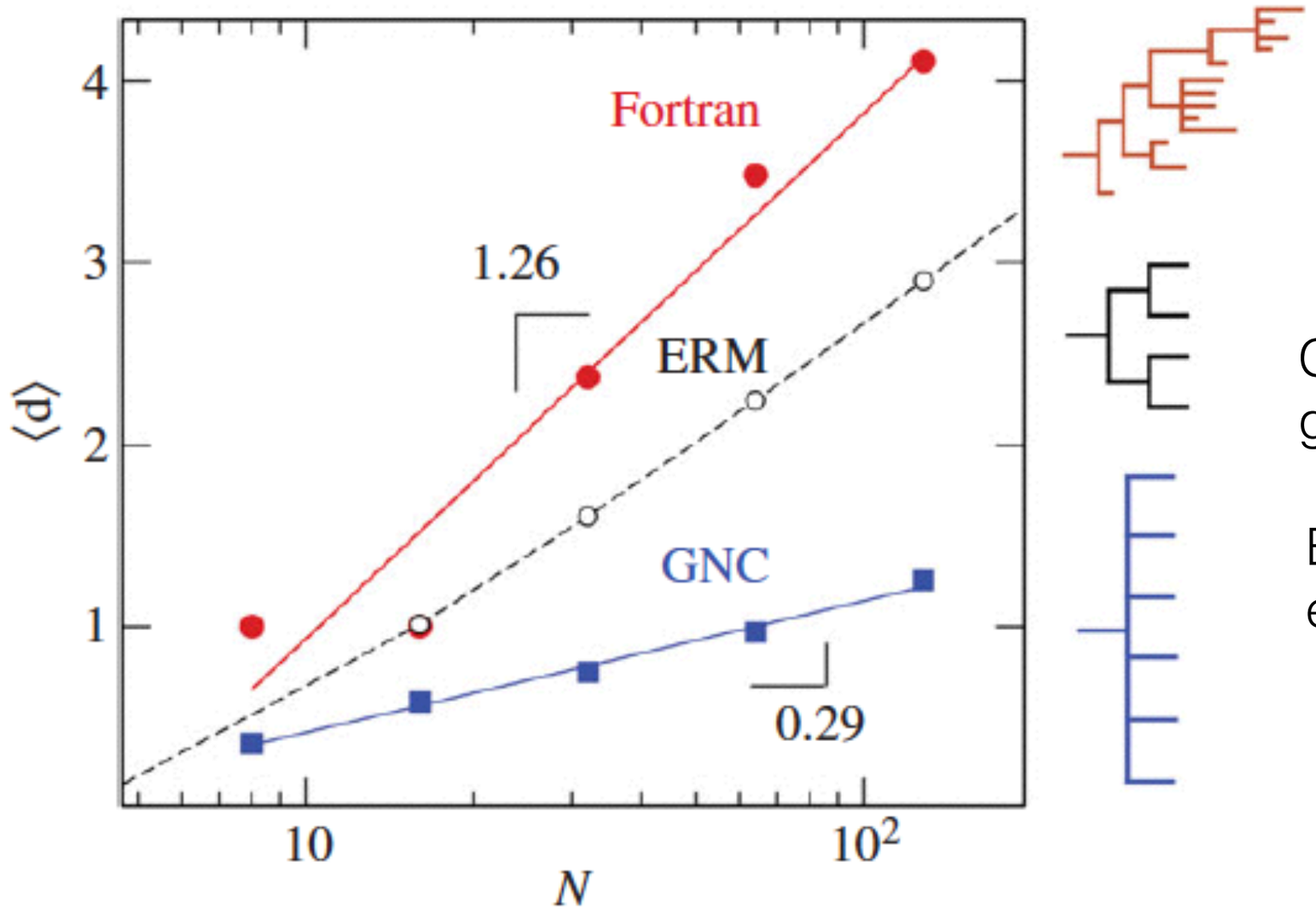
- suggested evidences were first found in fossil record. many species show rapid bursts of change that are often followed by long periods in which little or no change occurs (e.g. Cambrian explosion).

- recently, it has been proposed that punctuated equilibrium could be examined by phylogenetic trees.



Venditti and Pagel, 2008

Tree imbalance



GNC:
growth network by copying

ERM:
equal rate Markov

- The trees suggest the evolution of programming languages undergoes similar bursty patterns.

Discussion

- In biology, we have DNA sequences for phylogenetic analysis. In many other contexts, using dependency maps instead of phylogenetic trees presents certain advantages

- human-driven evolution and biological evolution exhibit similar bursty patterns. Apart from programming languages, punctuated equilibrium seems to be the case in the evolution of science, and human languages

brainfuck

This is the code for printing hello world.

```
+++++ +++++
[
    > +++++ ++
    > +++++ +++++
    > +++
    > +
<<<< -
]
> ++ .
> + .
+++++ ++ .
.
+++ .
> ++ .
<< +++++ +++++ +++++ .
> .
+++ .
----- - .
----- --- .
> + .
> .
```

```
There are eight commands:
+ : Increments the value at the current cell by one.
- : Decrements the value at the current cell by one.
> : Moves the data pointer to the next cell (cell on the right).
< : Moves the data pointer to the previous cell (cell on the left).
. : Prints the ASCII value at the current cell (i.e. 65 = 'A').
, : Reads a single input character into the current cell.
[ : If the value at the current cell is zero, skips to the corresponding ] .
    Otherwise, move to the next instruction.
] : If the value at the current cell is zero, move to the next instruction.
    Otherwise, move backwards in the instructions to the corresponding [ .

[ and ] form a while loop. Obviously, they must be balanced.
```