

PROGRAMMING LANGUAGE ADOPTION AS AN EPIDEMIOLOGICAL PHENOMENON

This briefing reports scientific evidence on the epidemic nature of programming language adoption by developers of open source software.

FINDINGS

- By using concepts of epidemiology, powered by mathematical models of epidemics, we are able to understand, describe and forecast the adoption of programming languages in SE.
- The Richards function was used to model the phenomenon:

$$I(t) = \frac{K}{1 + e^{-r(t-t_i)}}$$

K : total case number of the infection
 r : per capita growth of the infecte population
 a : exponent of deviation the standard logistic curve
 t_i : time of the curve's inflection point

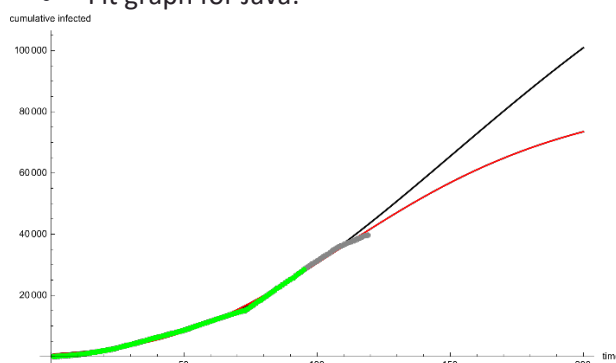
- Fit parameters found:

| Language | K | r | t_m | a |
|-------------|--------|--------|----------|-----------------------|
| C (100%) | 38244 | 0.017 | -748.71 | 7.56×10^{-7} |
| C (80%) | 29085 | 0.022 | -592.45 | 8.96×10^{-7} |
| C++ (100%) | 78057 | 0.0133 | -905.46 | 1.33×10^{-6} |
| C++ (80%) | 61705 | 0.015 | -816.75 | 1.21×10^{-6} |
| Java (100%) | 86338 | 0.02 | 32.14 | 0.225 |
| Java (80%) | 179248 | 0.011 | -1129.89 | 1.43×10^{-6} |
| PHP (100%) | 63607 | 0.015 | -157.17 | 0.017 |
| PHP (80%) | 89569 | 0.013 | -934.20 | 1.25×10^{-6} |

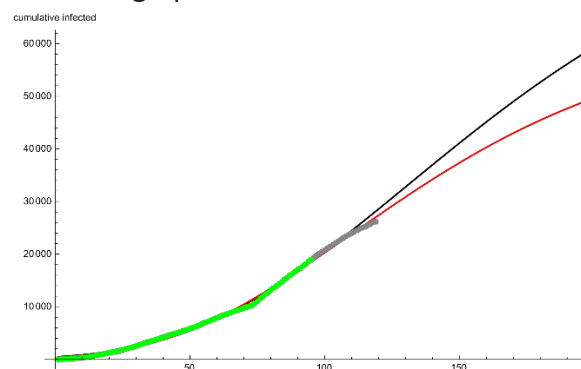
- Fit was considered satisfactory according to the goodness of fit analysis:

| Language | R^2 | Székely | p-value |
|-------------|--------|---------|---------|
| C (100%) | 0.9988 | 1297.63 | 0.987 |
| C (80%) | 0.9980 | 3967.76 | 0.706 |
| C++ (100%) | 0.9989 | 1433.59 | 0.996 |
| C++ (80%) | 0.9977 | 2194.39 | 0.961 |
| Java (100%) | 0.9988 | 1793.37 | 0.993 |
| Java (80%) | 0.9979 | 2040.49 | 0.987 |
| PHP (100%) | 0.9990 | 942.95 | 0.996 |
| PHP (80%) | 0.9981 | 1286.83 | 0.994 |

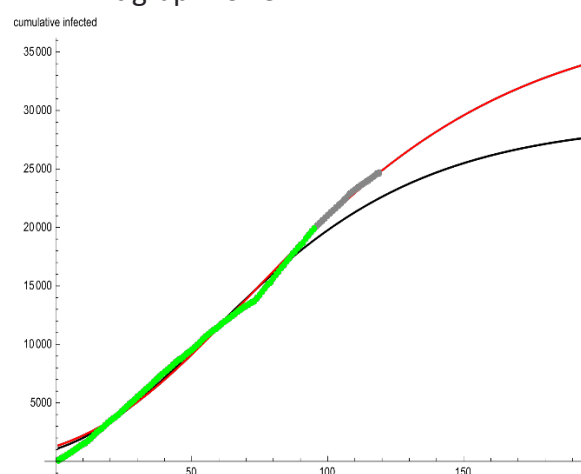
- Fit graph for Java:



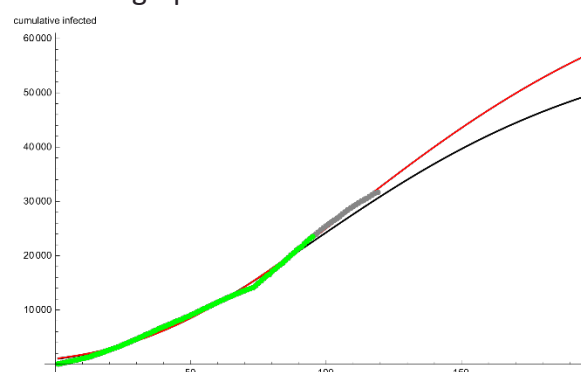
- Fit graph for PHP:



- Fit graph for C:



- Fit graph for C++:



- Even after using only 80% of the dataset, the resulting fitted model is still able to describe 100% of the data.
- It is possible, from the current findings, to predict satisfactorily 25% of the size of the used dataset into the future.
-

Who is this briefing for?

Practitioners interested in industry trends/standards related to the adoption of programming languages. **Researchers** interested in the dynamics of adoption of programming languages. **Students** seeking for suggestions of which programming language to study.

Where the findings come from?

All findings of this briefing were extracted from Sourceforge data, made available by Meyerovich et al. (Empirical Analysis of Programming Language Adoption, OOPSLA'13)

What is included in this briefing?

Data from the adoption of C, C++, Java and PHP from 2000 to 2009.

For additional information about ESEG:

<https://sites.google.com/site/eseportal>

ORIGINAL RESEARCH REFERENCE

Emanoel Barreiros et al. *Programming Language Adoption as an Epidemiological Phenomenon*. Proceedings of the 31st Brazilian Symposium on Software Engineering (SBES'17), pages 255-260, ISBN: 978-1-4503-5326-7, 2017. <https://doi.org/10.1145/3131151.3131188>