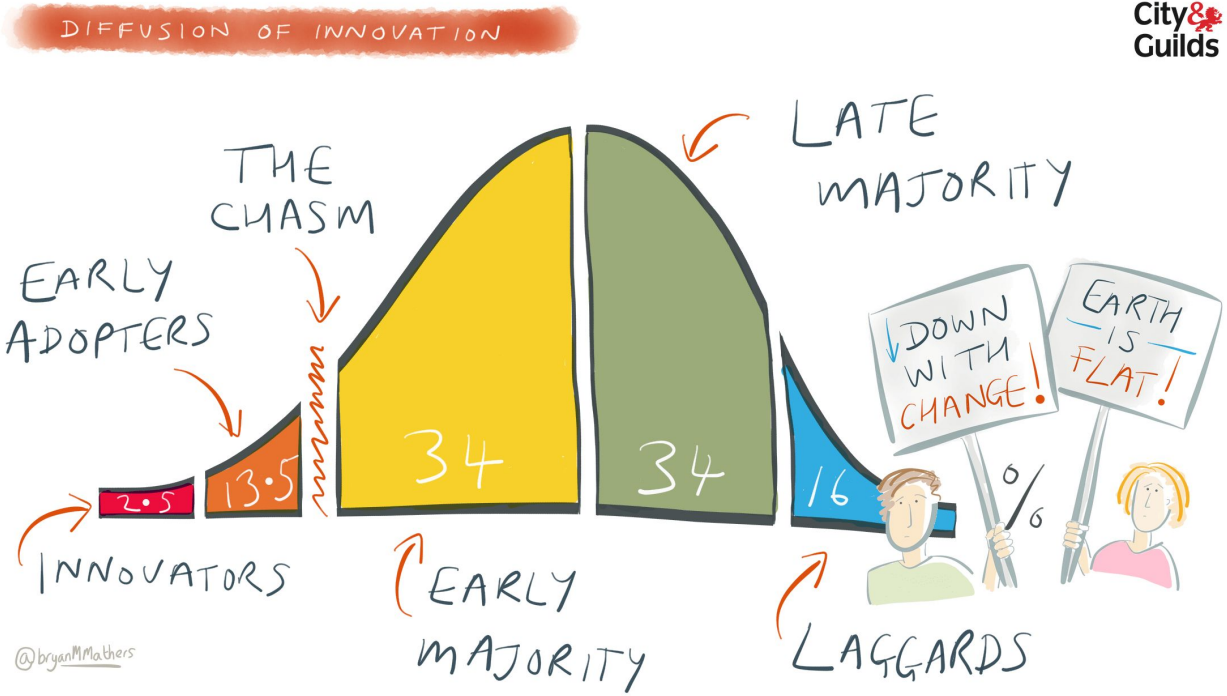
Innovation Diffusion: a (Big) Data-driven approach to the study of the geographic spreading of IT trends

Candidato: Enrico Palumbo, Data Scientist, Istituto Superiore Mario Boella

Relatore: Prof. Giancarlo Ruffo, UniTo, Dipartimento di Informatica

Innovation Diffusion

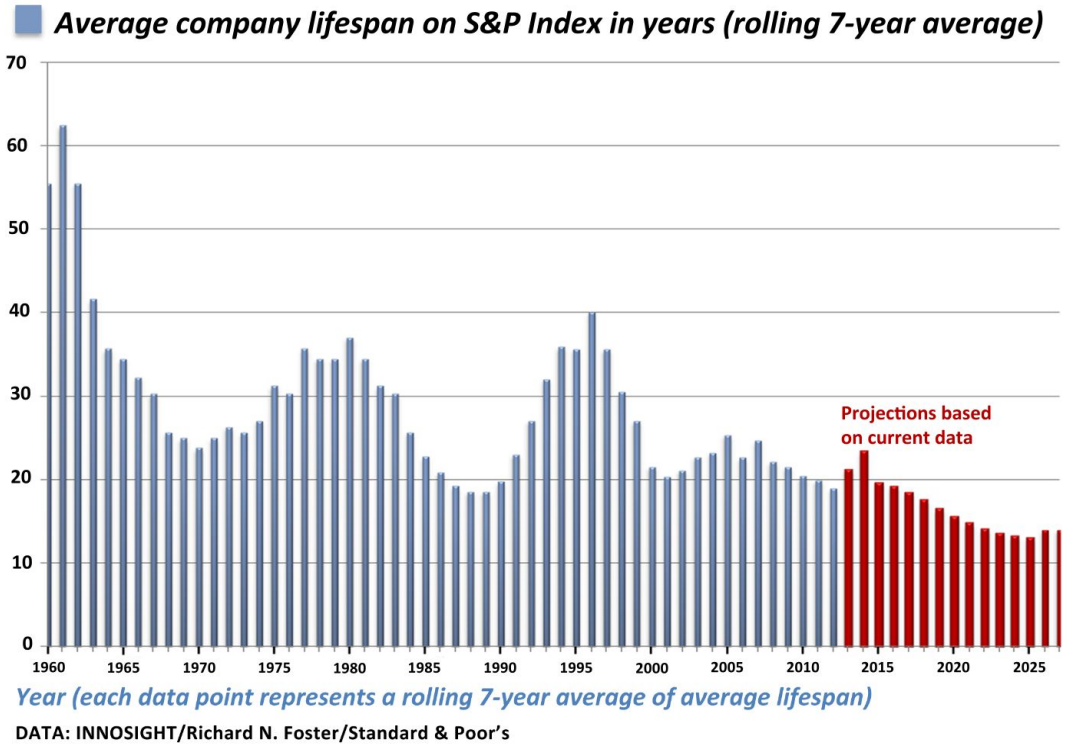
Pioneering work of Everett Rogers in 1962



http://bryanmmathers.com/

Average Lifespan in S&P 500

Average lifespan in S&P 500 has shrunk from 61 years in 1958 to 18 years in 2012



Gartner's Hype Cycle 2016

Predicting evolution of IT trends is paramount

Research questions

1. Are IT trends synchronous worldwide?
2. Is a good time for investing in the USA necessarily a good time for investing in Italy?
3. Is there a geographic diffusion process of IT trends?
4. Is there a delay between IT trends?

A (Big) Data-driven approach

Datafication: digitalization is turning everything into data

Google queries: enormous quantity of data (Big Data) reflecting people's interests

Google Trend: measure of Google queries in a specific location and in a specific temporal interval

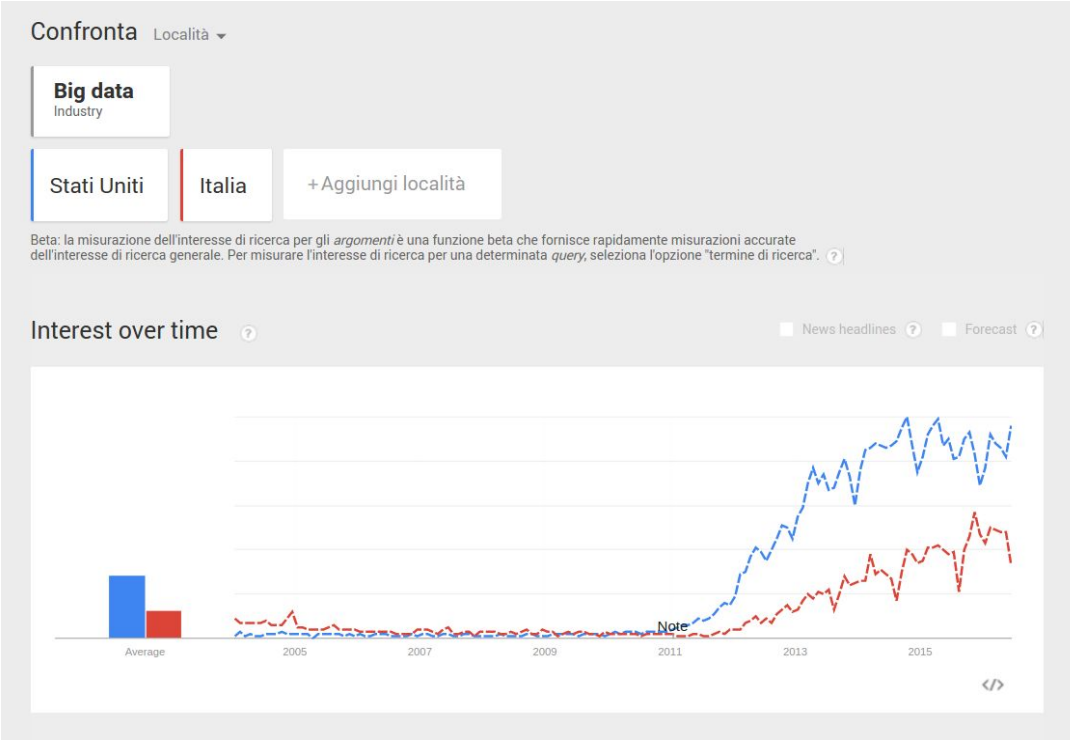
Study the geographic diffusion of IT trends through the analysis of Google Trend data

Google Trend Data

Query term: **Big Data**

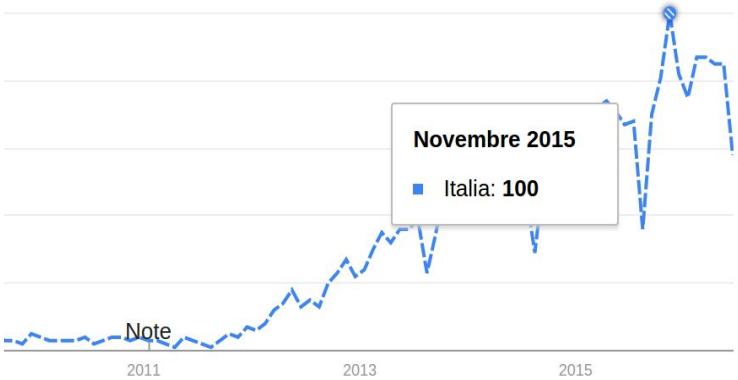
Countries: **USA, GBR, ITA, FRA, DEU, CAN, CHI, BRA, IND, RUS, JAP**

Time interval: **2008-2015**



From Google queries to Google Trend

1. Queries with semantically related keywords are aggregated in topic i
2. Number of queries **q(i,j,t)** on topic i, in place j and at time t
3. Frequency of queries **f(i,j,t)** on topic i, in place j and at time t
4. Rescaling: **max f(i,j,t) = 100**



Data Analysis

1. Raw Data

Smoothing

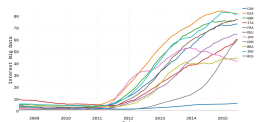
2. Trend

Plot

Max

Fitting

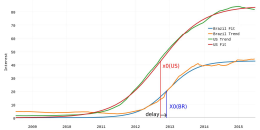
2a. Representation



2b. Peaks

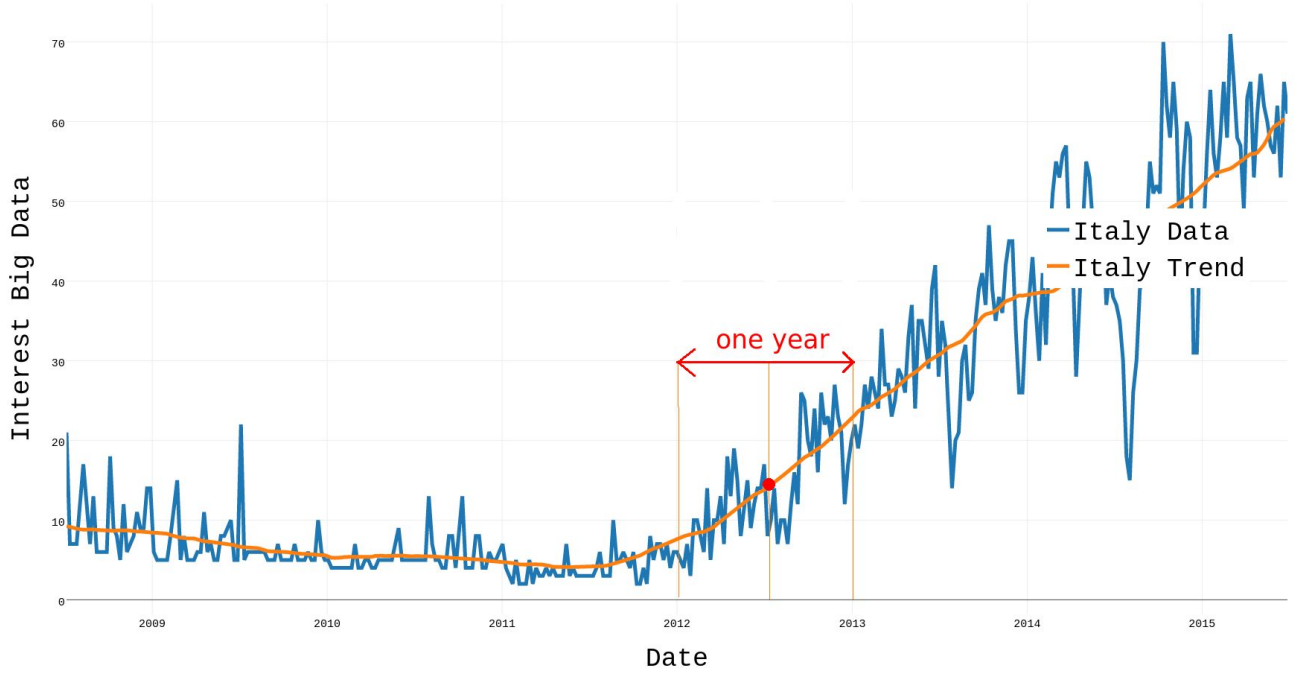
Country	Time of peak
Japan	15-12-2013

2d. Delays



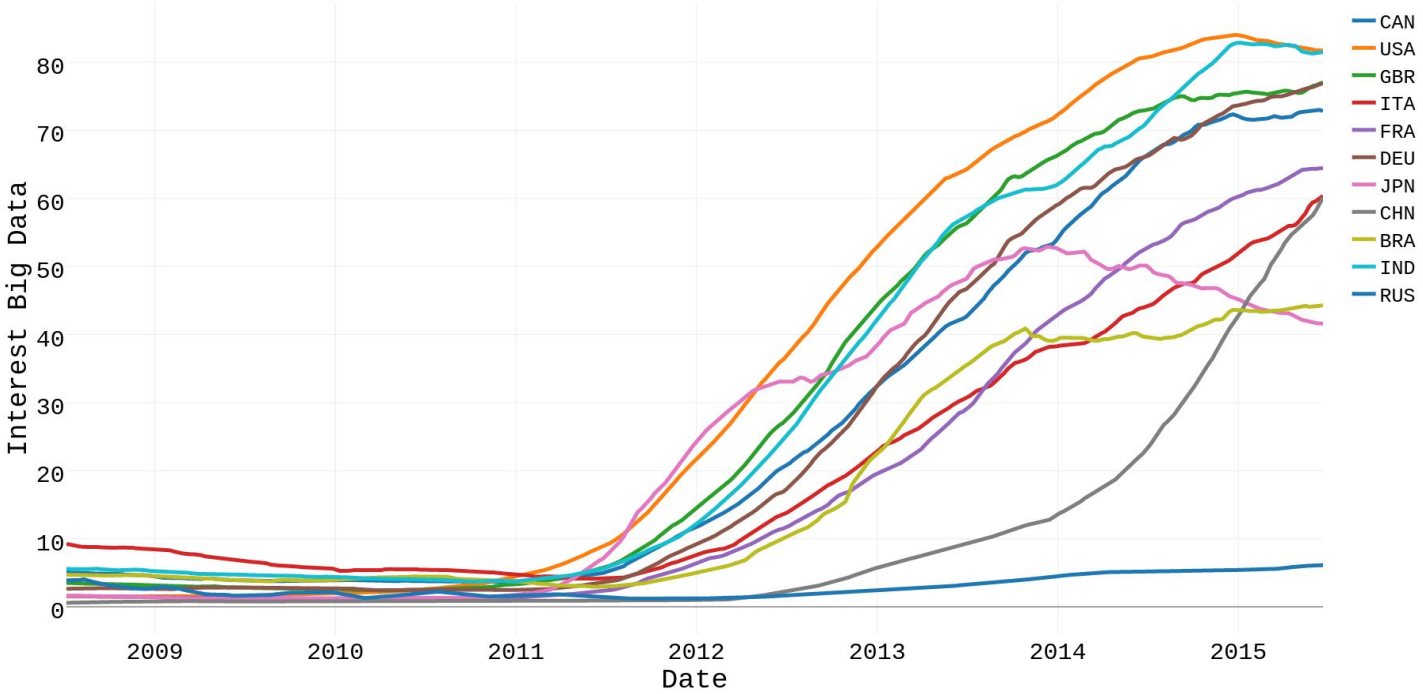
Smoothing

Moving Average with annual resolution, to reduce random and seasonal effects



Representation

Depiction of the Big Data trend from 2008 to 2015 for the eleven countries



Peaks

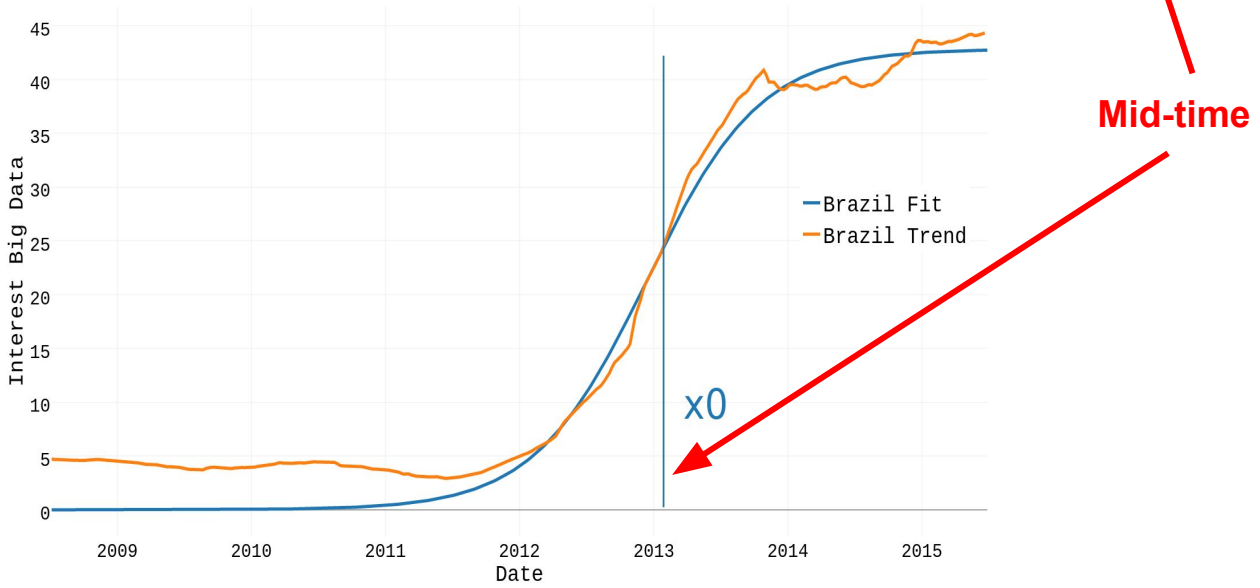
Country	Time of peak	Delay w.r.t. JAP [weeks]
Japan	15-12-2013	0
United States	21-12-2014	53
India	04-01-2015	55
Canada	14-06-2015	78
Great Britain	21-06-2015	79
Brazil	21-06-2015	79
Germany	21-06-2015	79
France	21-06-2015	79
Italy	21-06-2015	79
China	21-06-2015	79

Fitting

Method: non-linear least squares

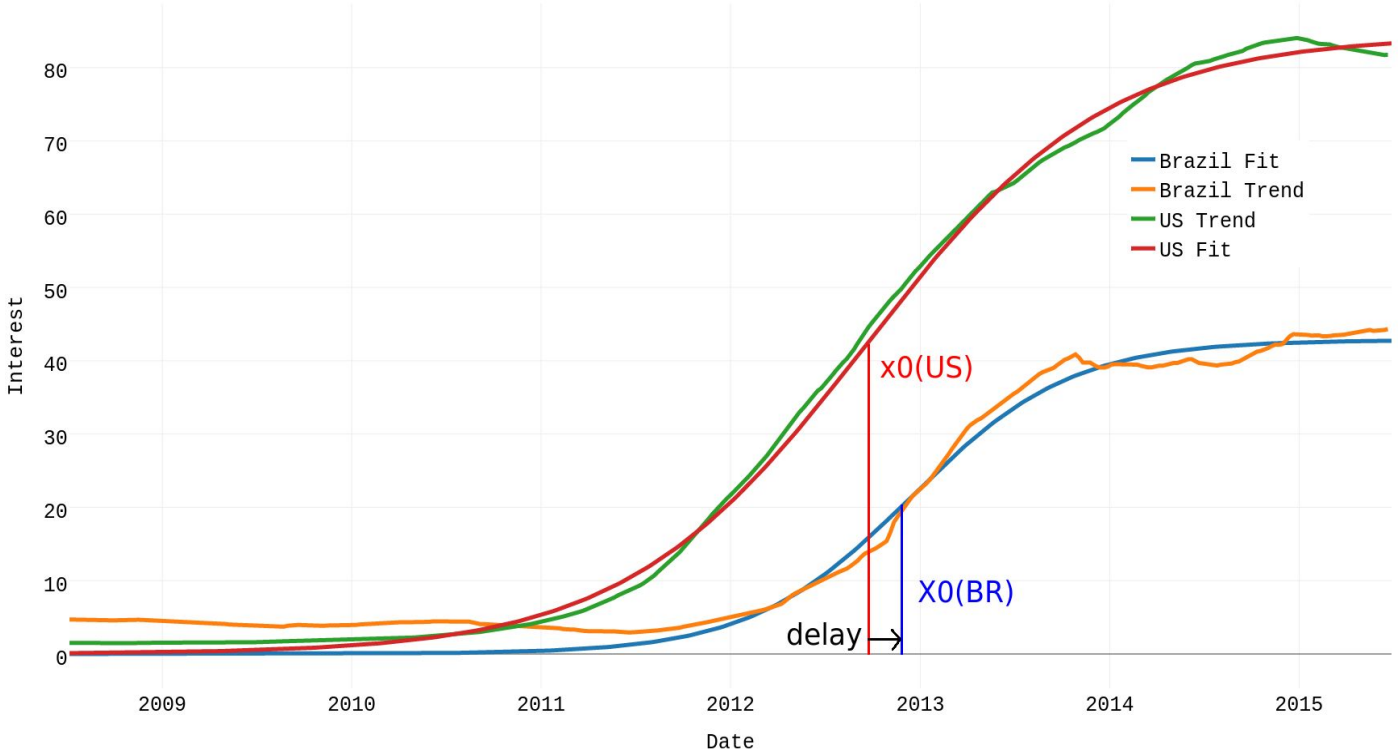
Model: logistic function

$$P(x) = \frac{K}{1 + e^{-r(x - x_0)}}$$



Delays

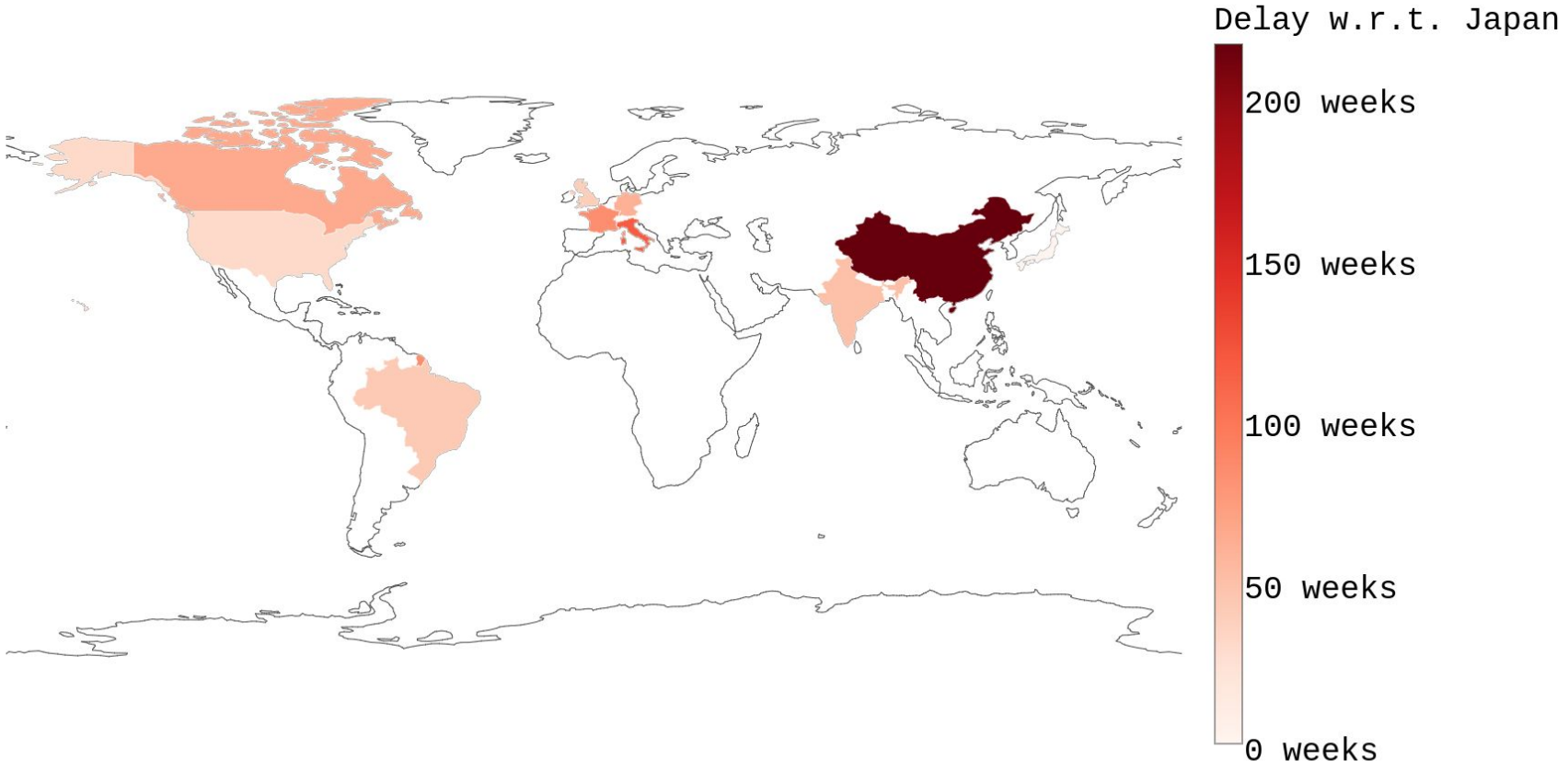
Delay: x0(Country 1) - x0(Country 2)



Delay with respect to Japan

Country	Delay w.r.t. JAP [weeks]
Japan	0
United States	30.4 ± 0.8
Great Britain	39.1 ± 0.9
Brazil	42.9 ± 1.2
India	50.1 ± 1.3
Germany	59.5 ± 0.9
Canada	64.9 ± 1.3
France	83.3 ± 0.9
Italy	116.3 ± 7.7
China	216.2 ± 5.9

Map of delay with respect to Japan



Conclusions

1. Google Trend Data is a useful tool to study the geographic diffusion of IT trends
2. Geography matters in studying the diffusion of trends, **delays can be up to 4 years** (Japan - China)
3. Big Data is still a growing trend in most countries under analysis
4. English speaking countries and Japan appear to be more responsive to the Big Data trend

Future work

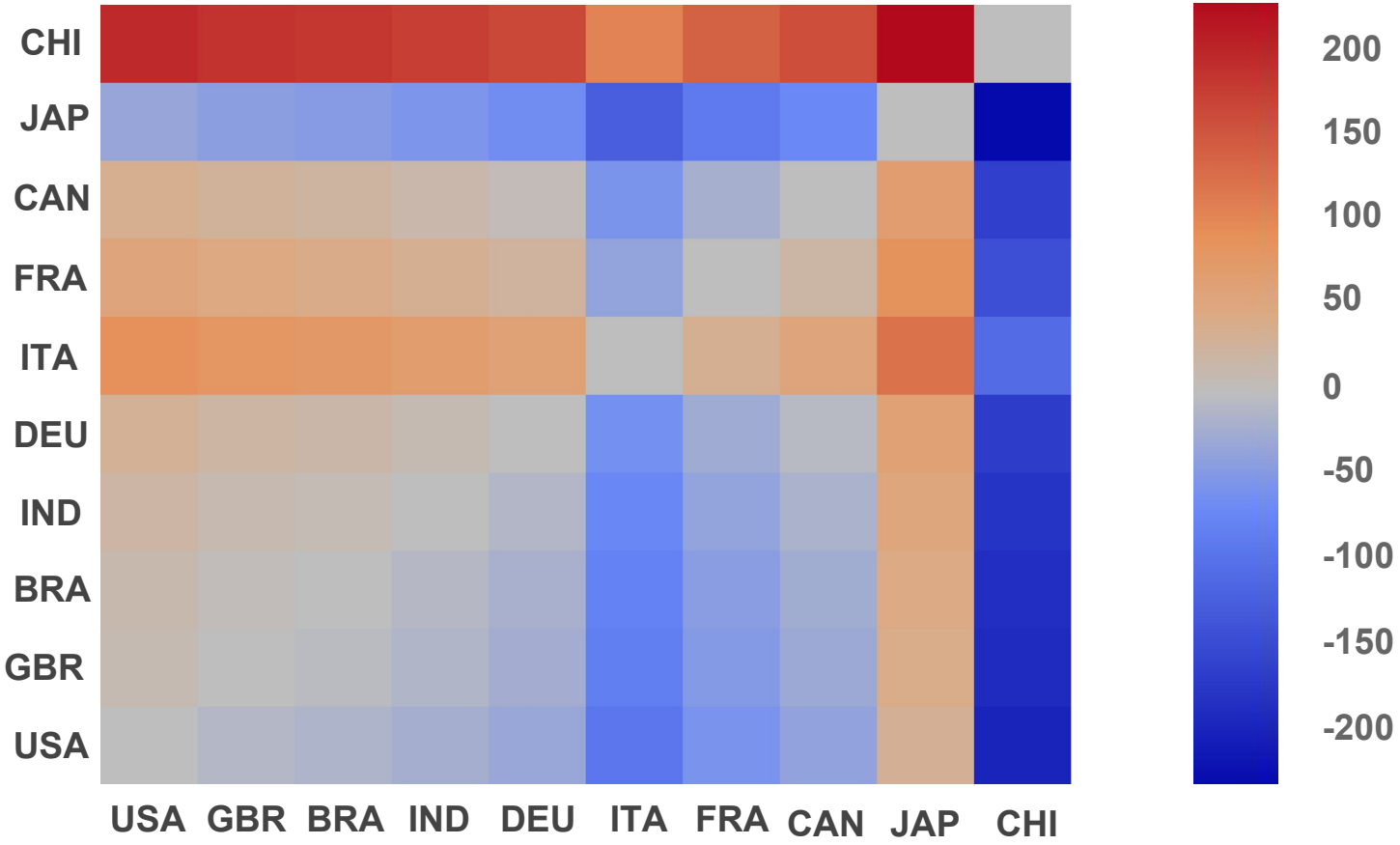
1. Further validation extending the study to other trends and other countries
2. Use Google Trend Data to predict economic indicators (e.g. market share of the technology)
3. Model the diffusion process (e.g. complex network model)

Thank you!

Enrico Palumbo, palumbo@ismb.it, www.ismb.it, @enricopalumbo91

Innovation Development area, Istituto Superiore Mario Boella (ISMB)

Heatmap delay



Data Selection

How to measure the evolution of the interest for Big Data in time across different countries?

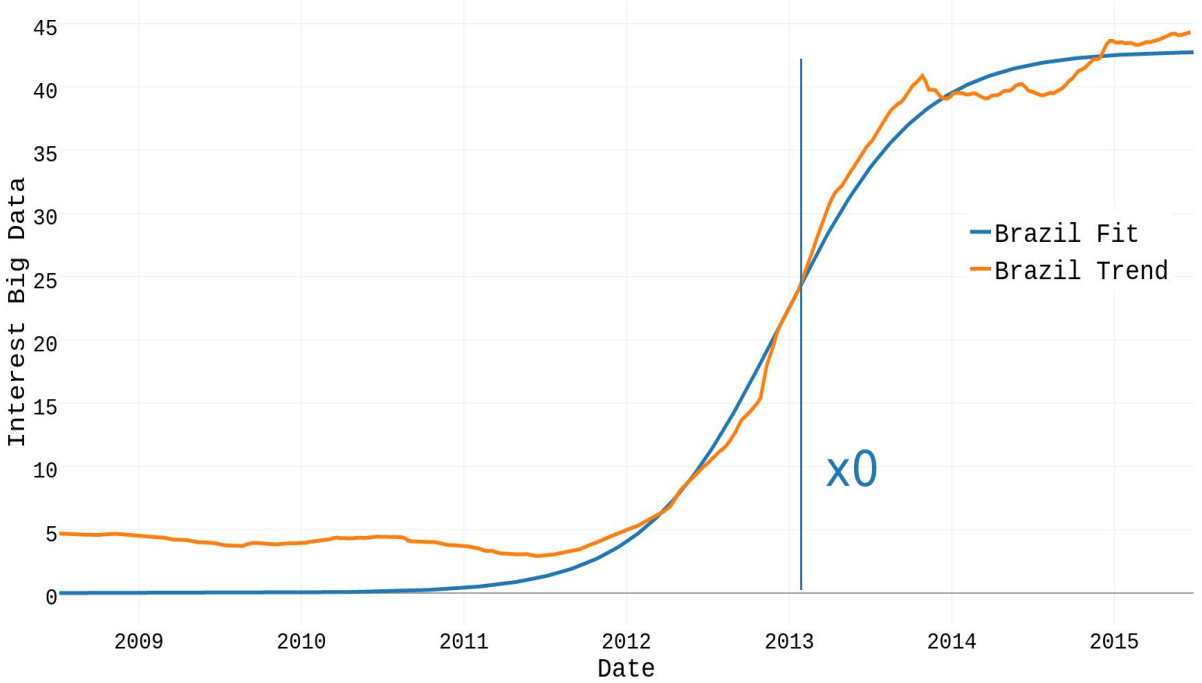
	Geographic	Temporal	Access
Job posts	yes	limited	limited
Scientific papers	no	yes	limited
Reports	yes	limited	limited
Google Trend	**yes**	**yes**	**yes**

Data Analysis

1. **Smoothing:** reducing the noise and extracting the trend
2. **Trend Representation:** plotting the curves representing the trends
3. **Peak of Interest:** computation of the time of peak of interest
4. **Clustering:** grouping similar curves together
5. **Fitting and delay estimation:** fitting the curves and estimating parameters to measure delay between them

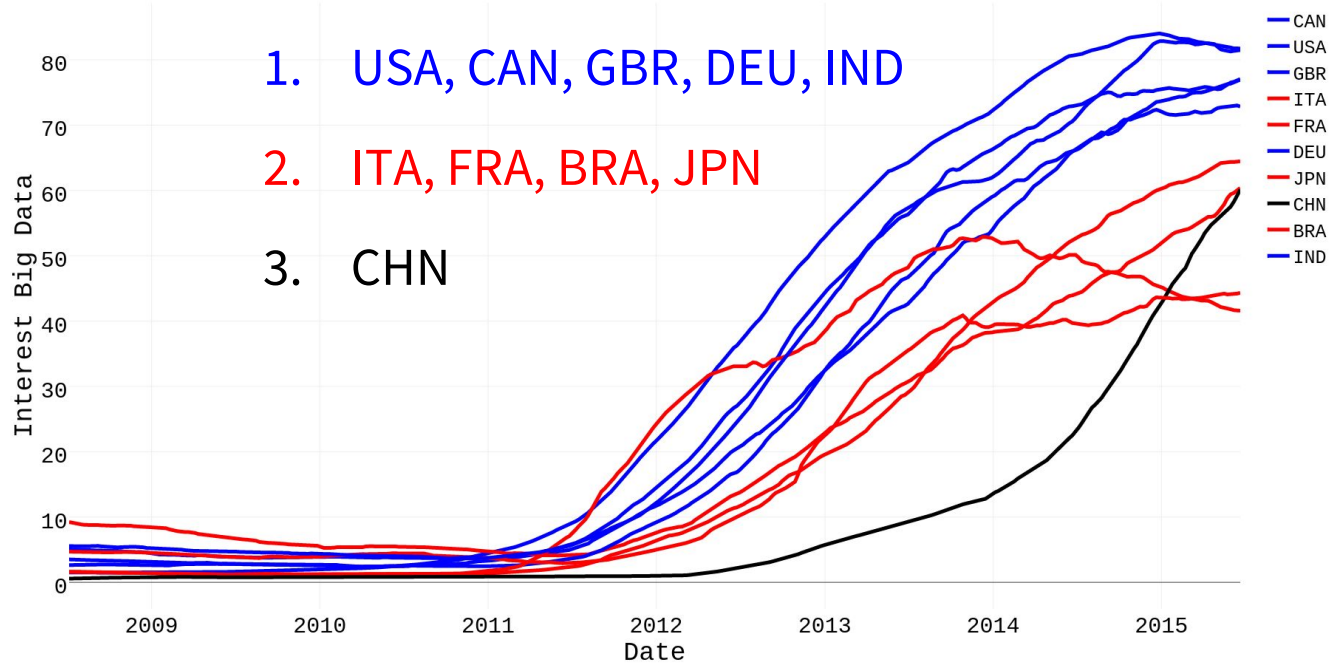
Fitting

Model: logistic curve **Parameters:** K, r, **mid-time x0** **Method:** non-linear least squares



Clusters

Grouping similar curves together, K-means algorithm: 3 groups



Other parameters

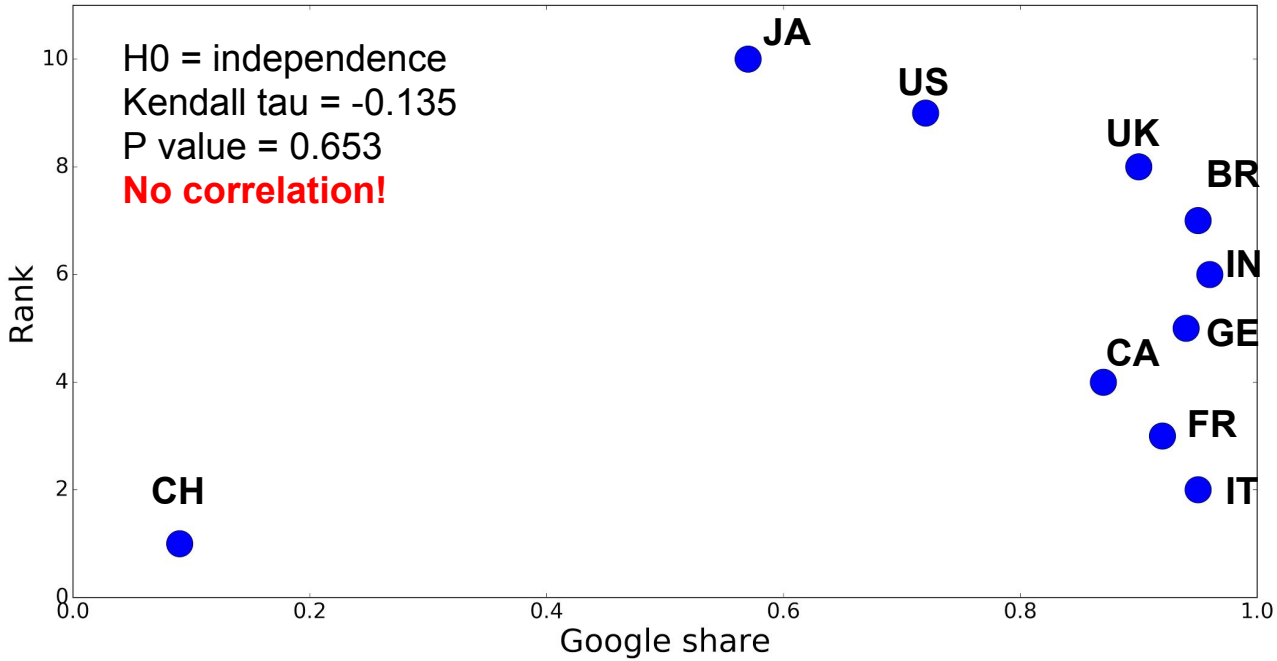
Country	x0	K	1/r
Japan	188.9 ± 0.8	48.0 ± 0.3	22.8 ± 0.7
United States	219.3 ± 0.3	84.4 ± 0.2	33.1 ± 0.2
Great Britain	227.9 ± 0.4	77.4 ± 0.3	31.4 ± 0.3
Brazil	231.7 ± 0.9	42.8 ± 0.4	22.0 ± 0.8
India	239 ± 1	84.1 ± 0.7	36.5 ± 0.7
Germany	248.3 ± 0.5	77.7 ± 0.4	33.5 ± 0.4
Canada	254 ± 1	80.4 ± 0.7	41.7 ± 0.6
France	272.1 ± 0.4	70.5 ± 0.3	38.0 ± 0.3
Italy	305 ± 8	84 ± 4	66 ± 3
China	405 ± 6	221 ± 20	44.3 ± 0.7

Google search engine market share 2015

Country	Google market share
Japan	57%
United States	72%
Great Britain	90%
Brazil	95%
India	96%
Germany	94%
Canada	87%
France	92%
Italy	95%
China	9%

1: http://returnonnow.com/internet-marketing-resources/2015-search-engine-market-share-by-country/
2: http://www.chinainternetwatch.com/17415/search-engine-2012-2018e/

Correlation?



S&P 500 turnover

Sample companies that have entered and exited since 2002

Summary: interest in Big Data

1. Peak: JAP, USA, IND, CAN
2. Mid-time: JAP, USA, GBR, BRA, IND

USA, India, Canada, Great Britain, Japan