Social Physics and Complexity (SPAC)

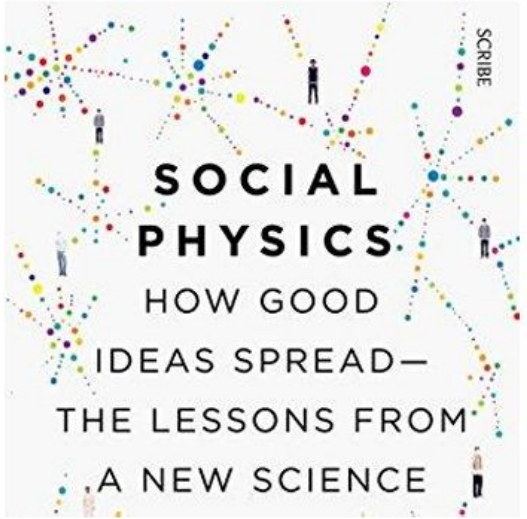
Lília Perfeito
lperfeito@lip.pt

09/07/2024

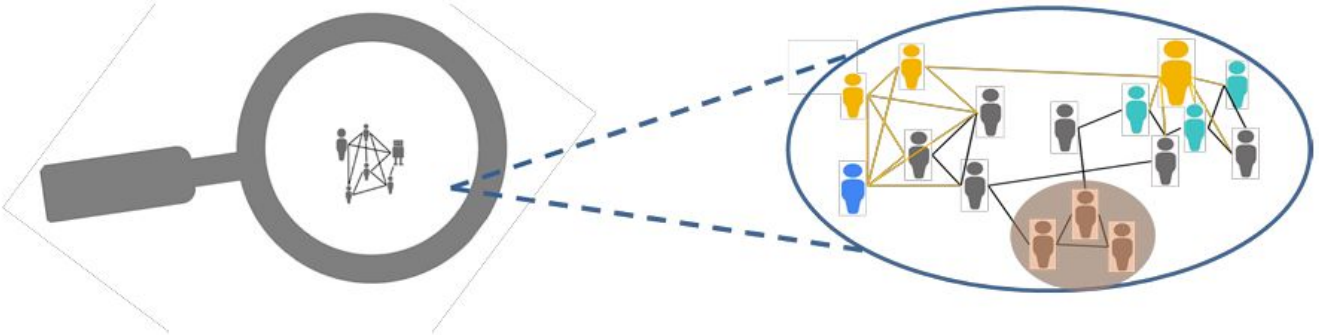
These problems—and a wide range of similar problems in the biological, medical, psychological, economic, and political sciences—are just too complicated to yield to the old nineteenth-century techniques which were so dramatically successful on two-, three-, or four-variable problems of simplicity. These new problems, moreover, cannot be handled with the statistical techniques so effective in describing average behavior in problems of disorganized complexity.

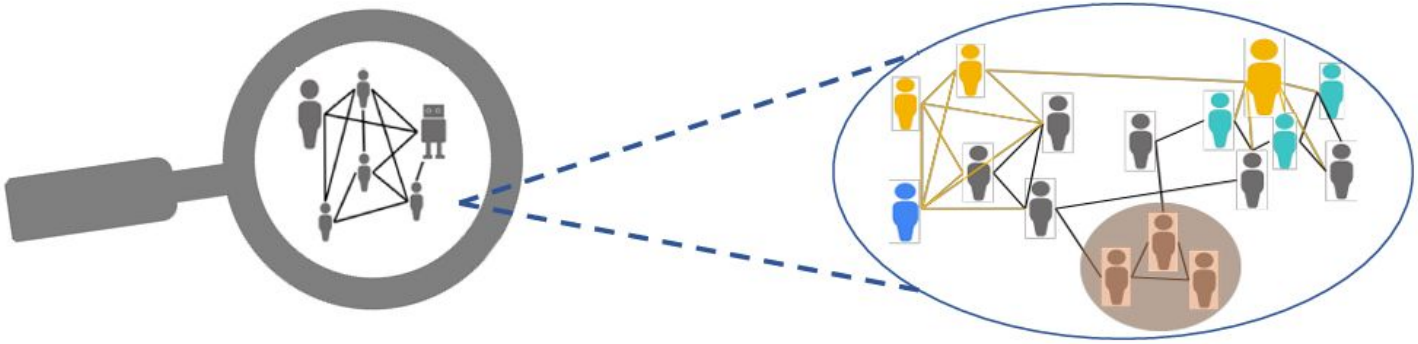
These new problems, and the future of the world depends on many of them, requires science to make a third great advance, an advance that must be even greater than the nineteenth-century conquest of problems of simplicity or the twentieth-century victory over problems of disorganized complexity. Science must, over the next 50 years, learn to deal with these problems of organized complexity.

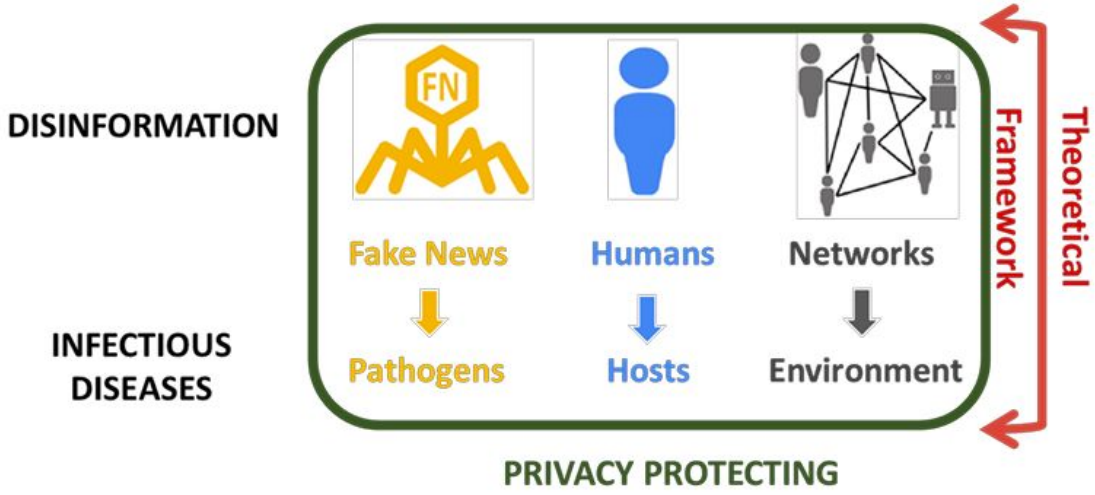
Warren Weaver, 1947

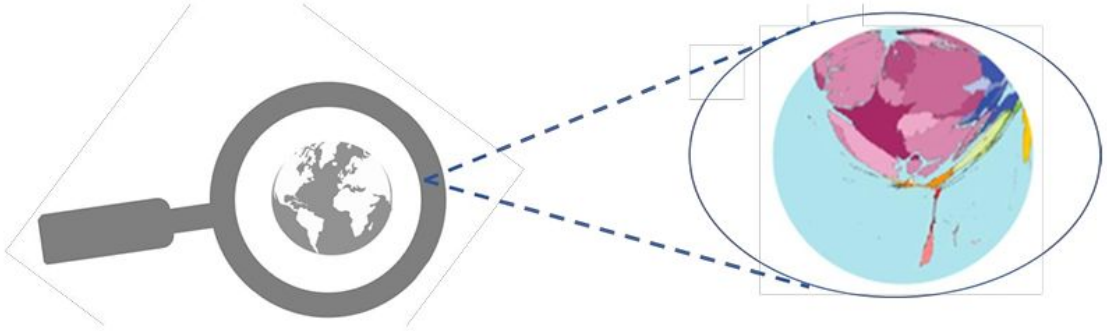


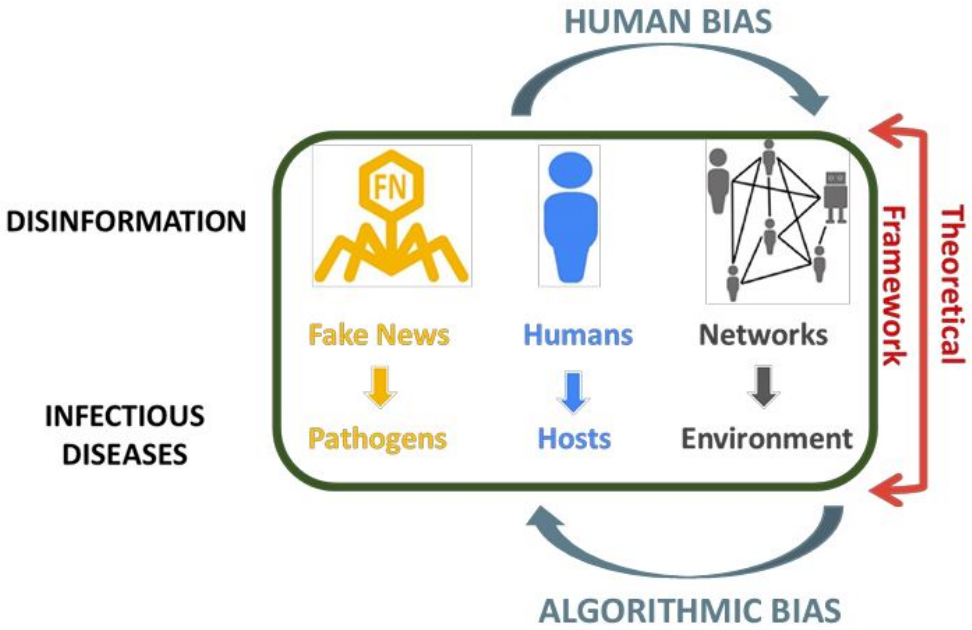
Alex Pentland, 2014

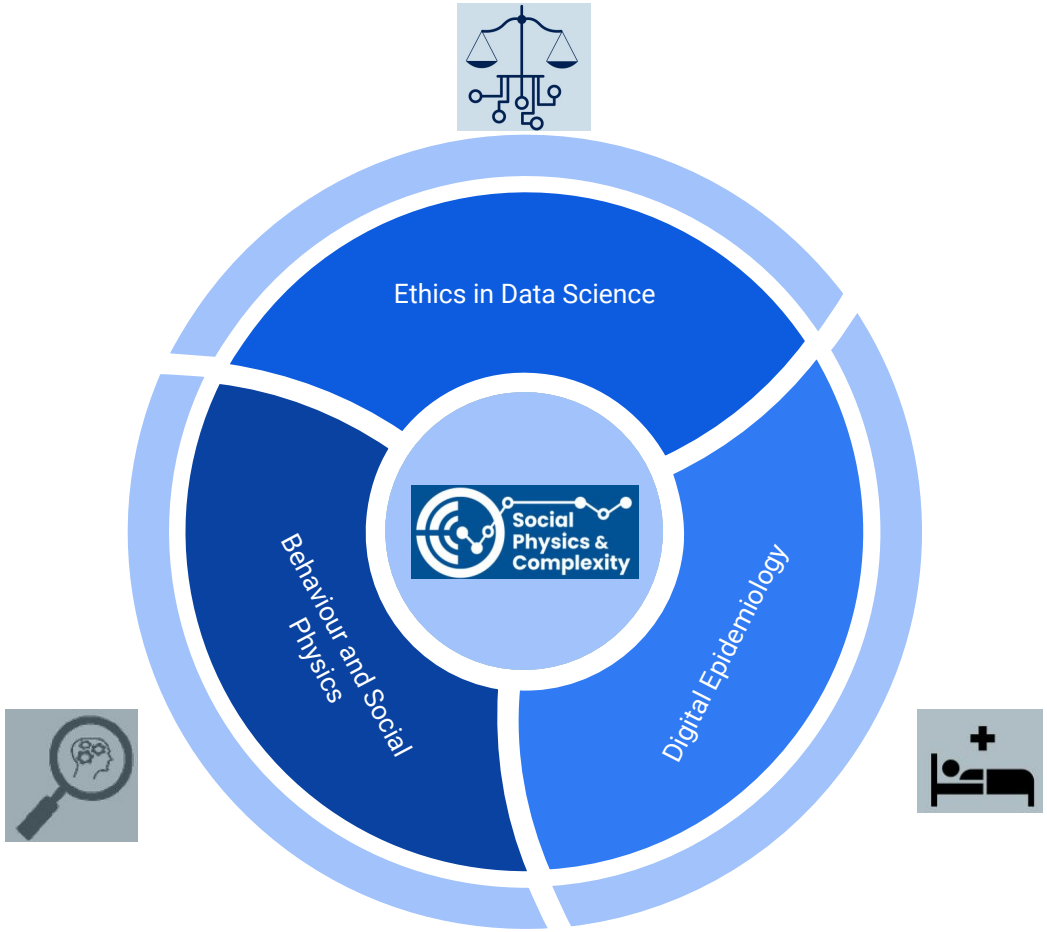




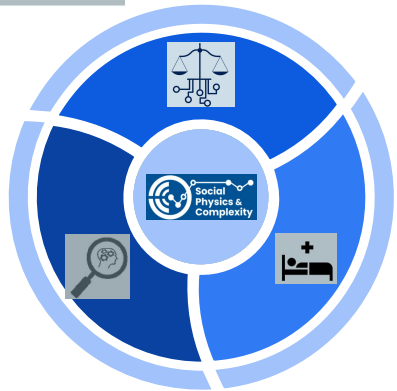

DISINFORMATION
INFECTIOUS DISEASES
FN
Fake News
Humans
Networks
Pathogens
Hosts
Environment
Framework
Theoretical
PRIVACY PROTECTING



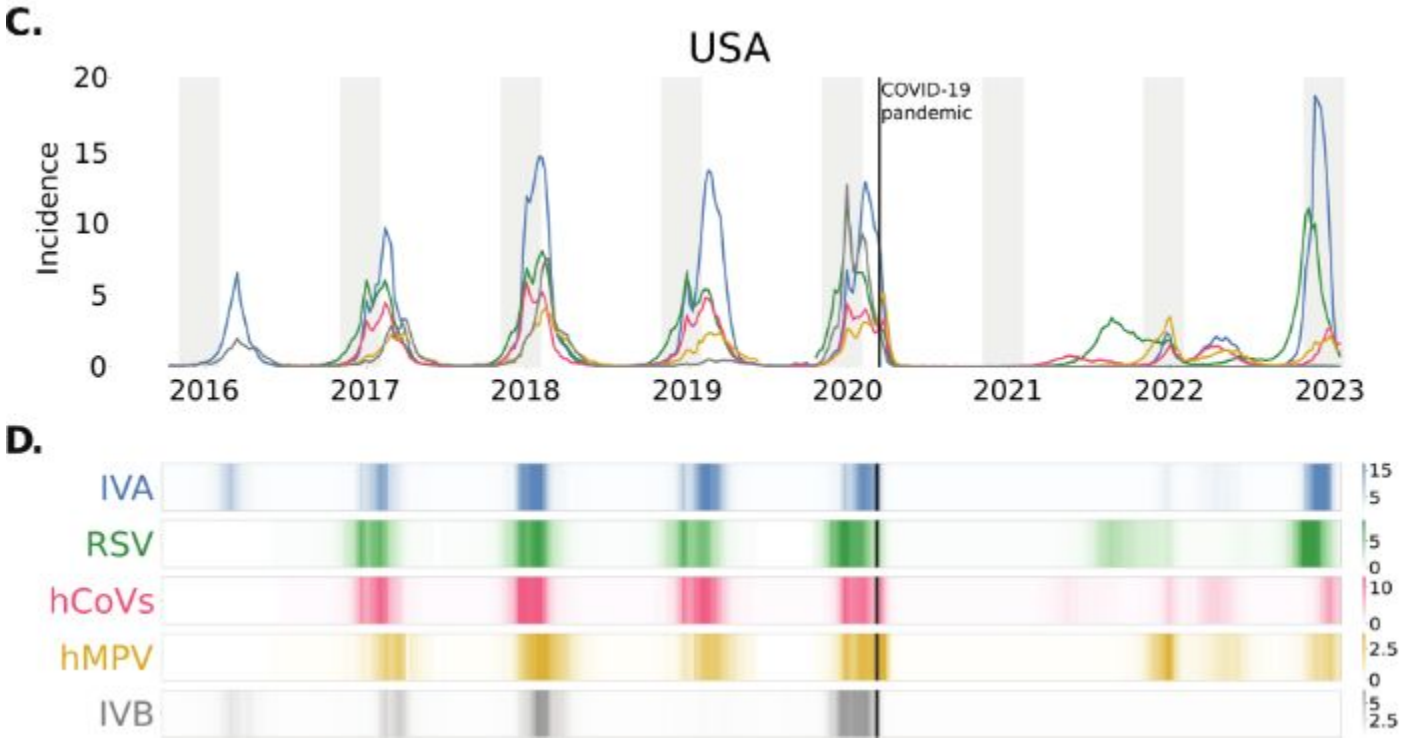

HUMAN BIAS
DISINFORMATION
INFECTIOUS DISEASES
FN
Fake News
Humans
Networks
Pathogens
Hosts
Environment
Framework
Theoretical
ALGORITHMIC BIAS


Ethics in Data Science
Social Physics & Complexity
Behaviour and Social Physics
Digital Epidemiology

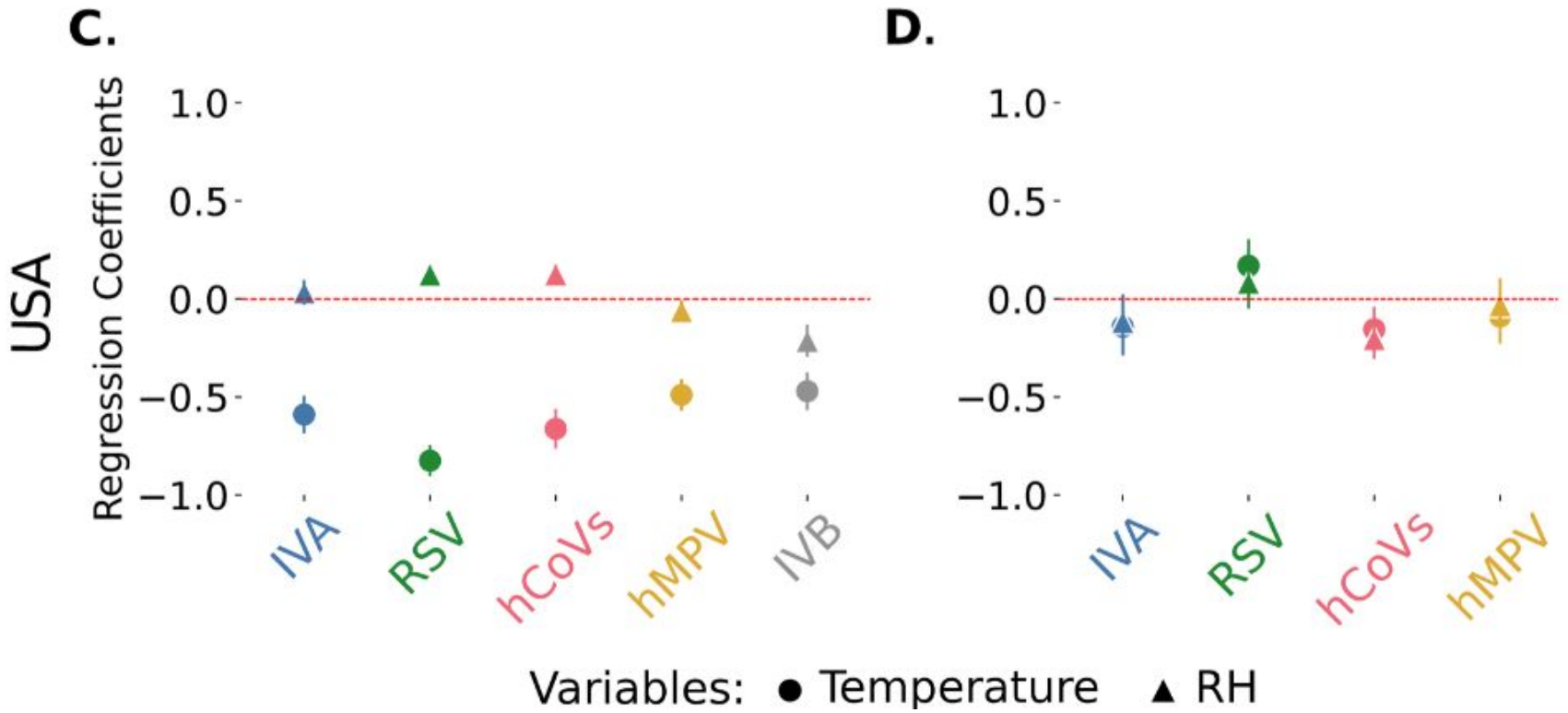
	QUESTIONS	DATA	TOOLS
HEALTH	Emergency Now-casting Antibiotic Over-prescription Infectious Disease Dynamics **From prescription to diagnosis**	**Google Trends** SNS24 Twitter ER acceptance /times **SPMS e-prescriptions**	Math Modelling ML Epidemiology
BEHAVIOUR	Cognitive Biases Attitudes Towards Science Privacy Protecting Analysis	**Large scale surveys** **Behavioral experiments** **Twitter**	Networks Math Modelling Psychology Information



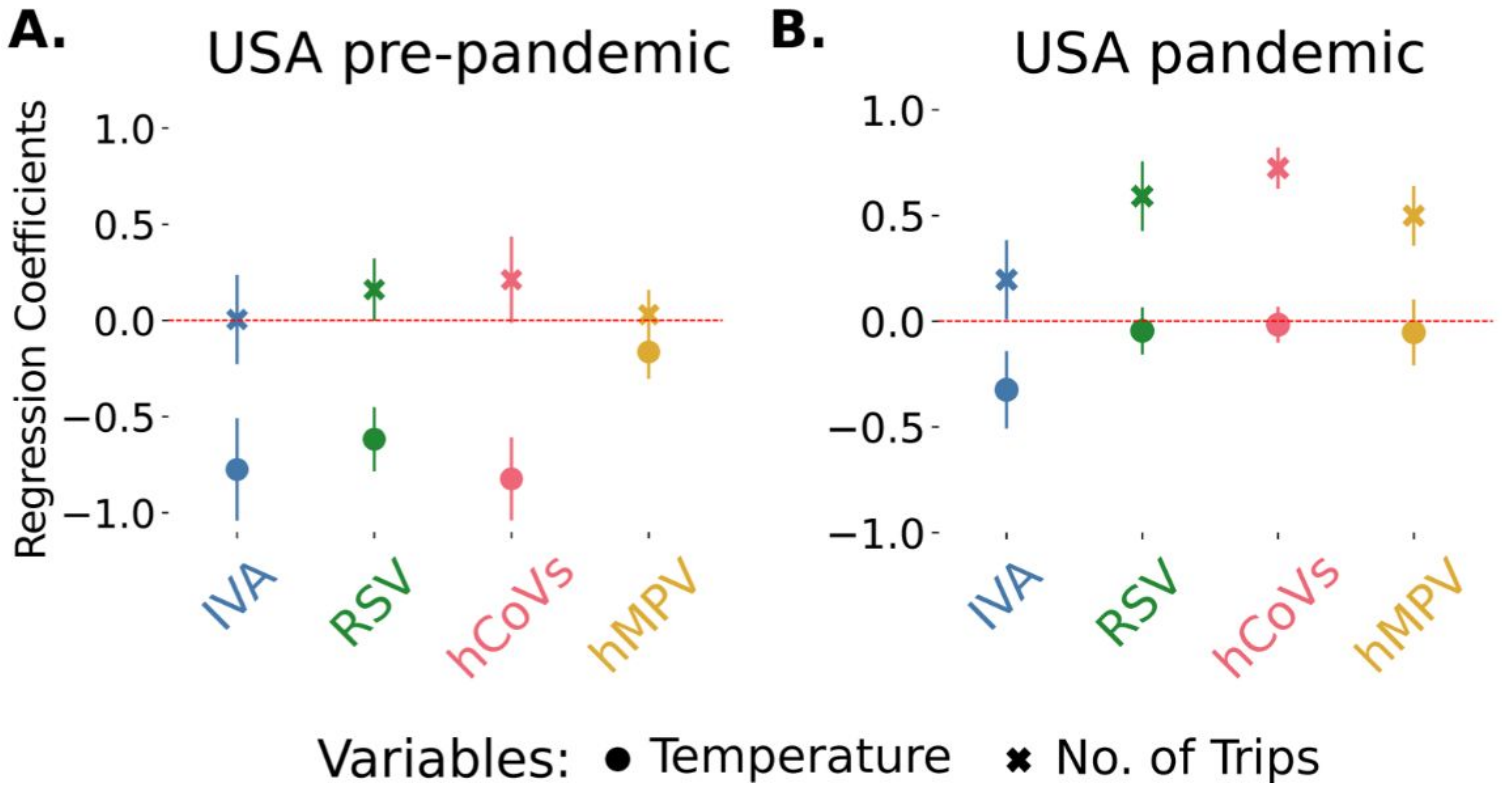
Infectious disease dynamics



The influence of weather changed during the pandemic



What might be driving infections then?

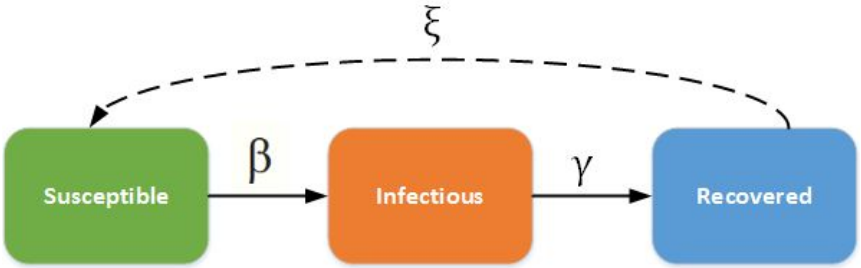


Varela-Lasheras, I., Perfeito, L., Mesquita, S., & Gonçalves-Sá, J. (2023). The effects of weather and mobility on respiratory viruses dynamics before and during the COVID-19 pandemic in the USA and Canada. *PLOS Digital Health*, *2*(12), e0000405.

What did we learn?

Seasonality in respiratory disease is not driven by weather alone

During the Covid-19 pandemic (masks, confinements, changes in behaviour, etc.), the movement of people may have been the limiting factor



$$\beta_{v_i}(t) = f(\text{weather, behavioural factors})$$

(Re)using electronic health records to gain insights into public health

Designed for multiple purposes:
Financial, medical care, human resources, etc.

How is disease incidence estimated?

Notifiable diseases like the flu, measles, yellow fever, HIV- All diagnosis are reported and stored centrally at the state or country level

Non-notifiable diseases like diabetes, asthma, depression - Incidence is estimated from surveys, which can be costly, incomplete, biased

Electronic prescriptions

- **Designed to**
  - Inform the patients of their therapy - what, when, how;
  - Control the sale of medical substances;
  - Keep track of costs with medication.

- **Useful to**
  - Describe the prescribing habits of MDs;
  - Assess the health status of the population.

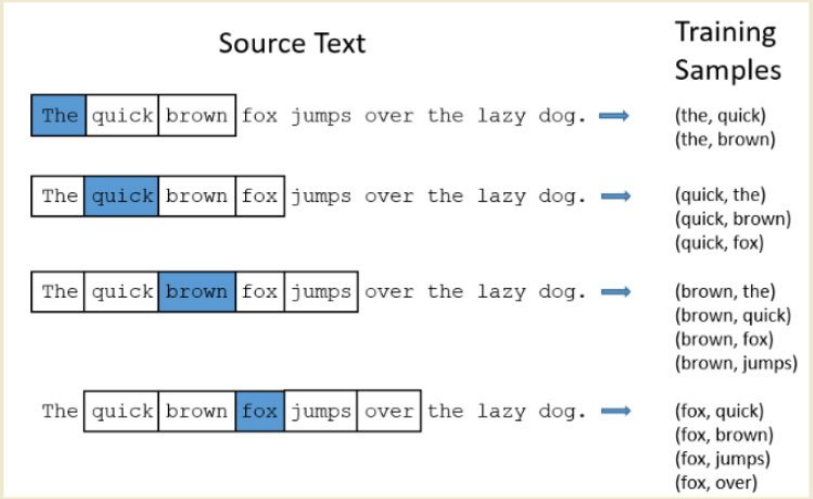
In Portugal

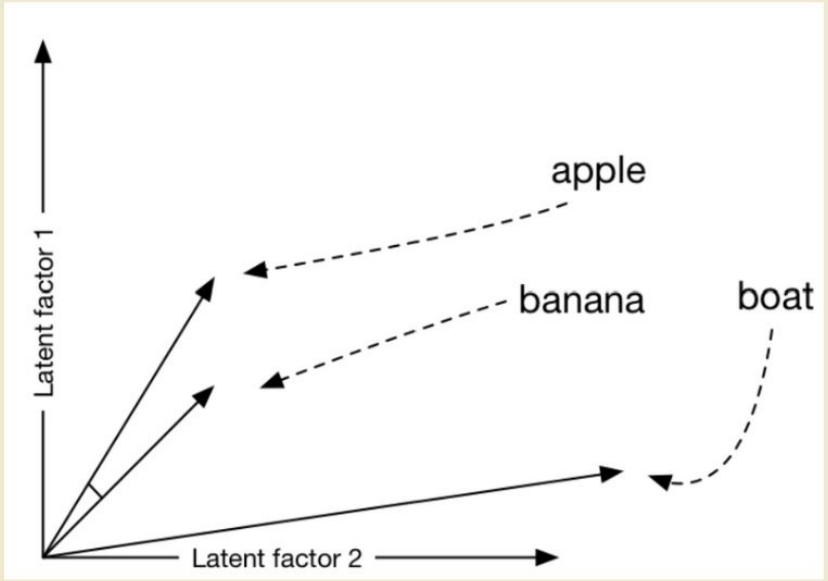
- Electronic prescriptions introduced in 2011;
- By 2017, 97% of all prescriptions were electronic;
- PEM is managed by SPMS and includes all medical prescriptions of regulated drugs, including the private sector

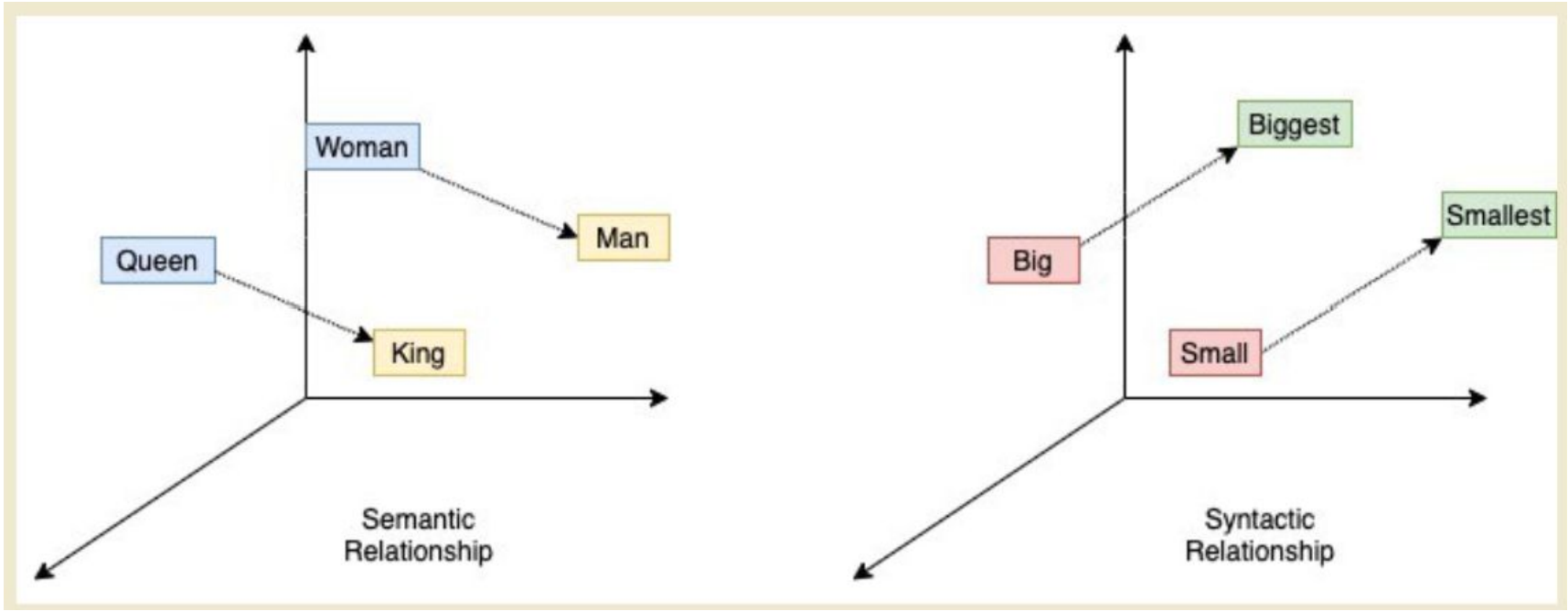
Data and Methods

- The Data
  - Pseudo-Anonymized medical prescriptions 2017-2019
  - Only substance name, grouped by patient
- The method
  - Embeddings model - Word2vec - to infer distance between substances
  - Reciprocal clustering algorithm

Word2vec

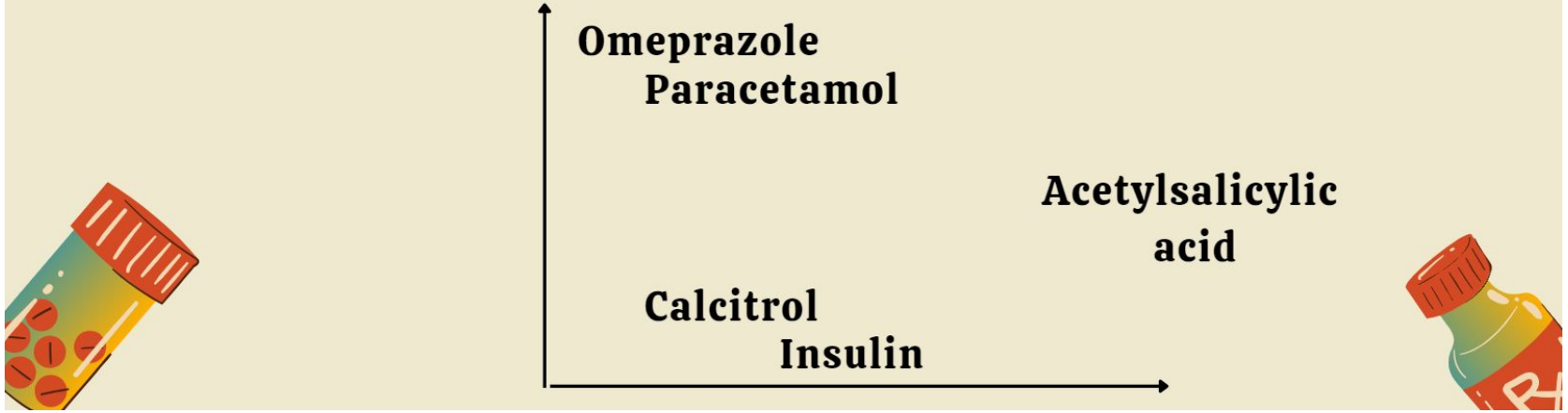

Source Text
Training Samples
The quick brown fox jumps over the lazy dog.
(the, quick)
(the, brown)
The quick brown fox jumps over the lazy dog.
(quick, the)
(quick, brown)
(quick, fox)
The quick brown fox jumps over the lazy dog.
(brown, the)
(brown, quick)
(brown, fox)
(brown, jumps)
The quick brown fox jumps over the lazy dog.
(fox, quick)
(fox, brown)
(fox, jumps)
(fox, over)

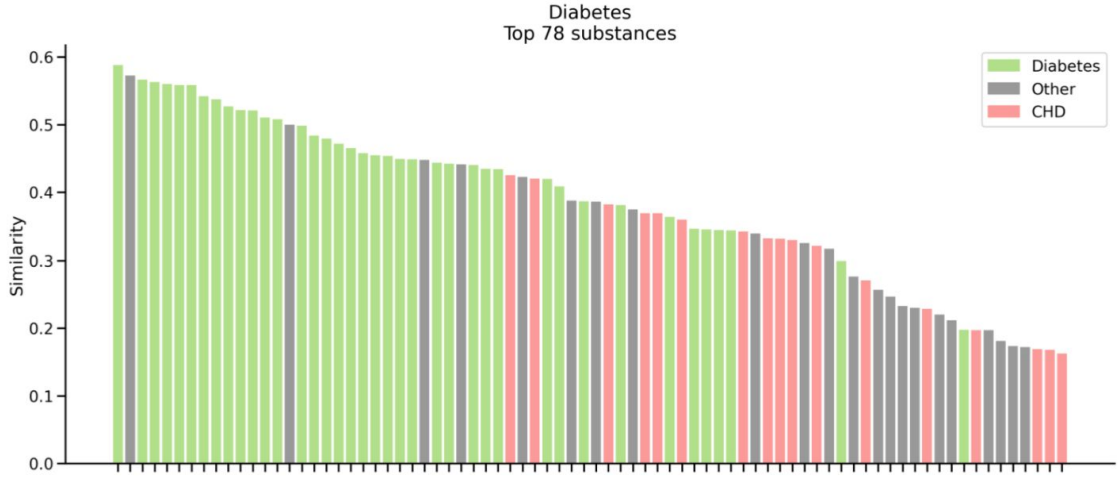

Latent factor 1
Latent factor 2
apple
banana
boat


Woman
Man
Queen
King
Semantic
Relationship
Biggest
Smallest
Big
Small
Syntactic
Relationship

Word2vec applied to prescribed drugs

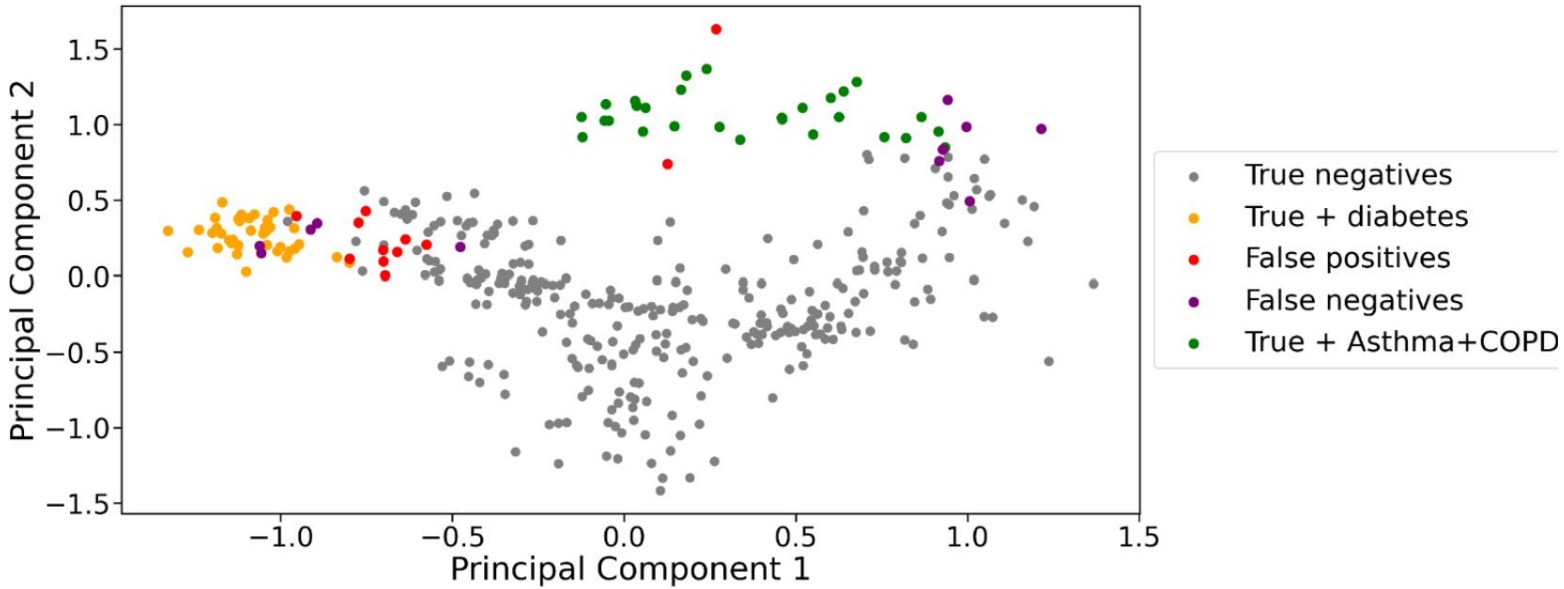
Patient 1 - Omeprazole, Paracetamol
Patient 2 - Acetylsalicylic acid
Patient 3 - Calcitriol, Human insulin
....





drug	sim	disease
Pioglitazona	0.588210	Diabetes
Fenofibrato	0.572582	Nan
Dapagliflozina	0.566599	Diabetes
Nateglinida	0.563022	Diabetes
Empagliflozina	0.560064	Diabetes
Metformina	0.558796	Diabetes
Gliclazida	0.558517	Diabetes
Glimepirida	0.541760	Diabetes
Exenatido	0.537630	Diabetes
Metformina + Dapagliflozina	0.527170	Diabetes

Prescription space



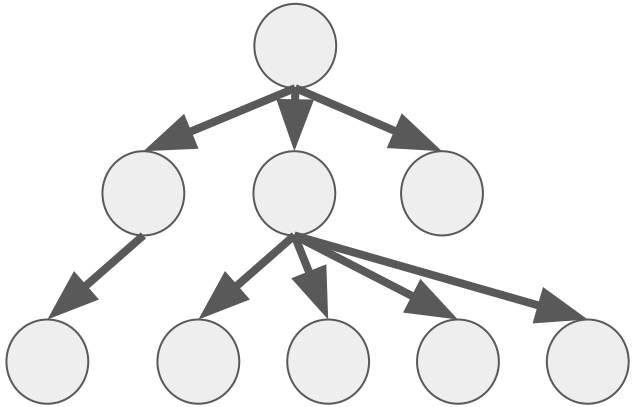
We developed a method able to identify co-prescribed drugs, which are associated with specific diseases

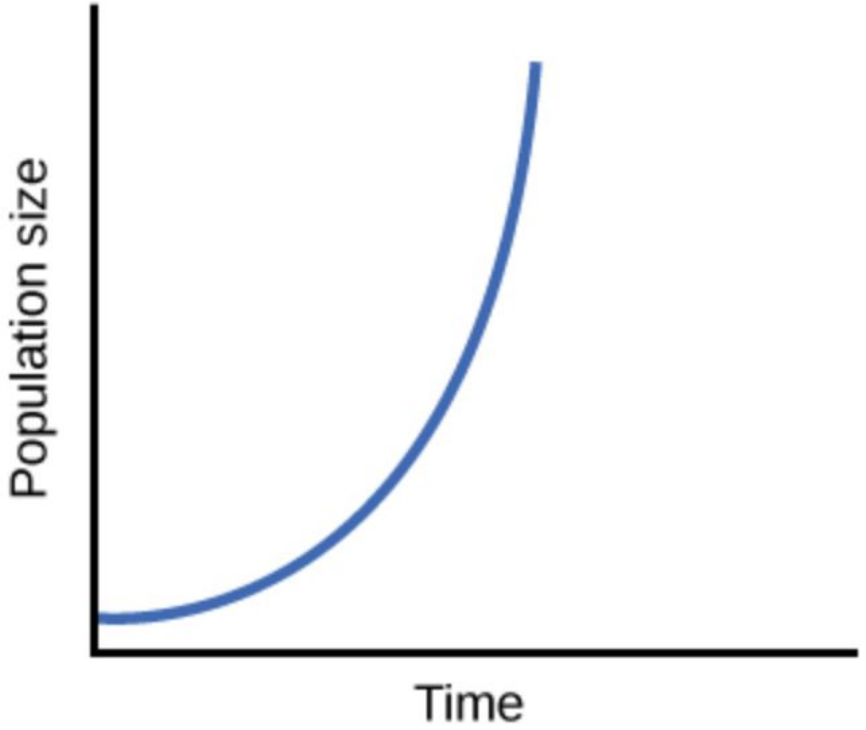
Next step is to enrich it with textual descriptions of the drugs

The objective is to use these clusters to infer the prevalence of non-notifiable diseases

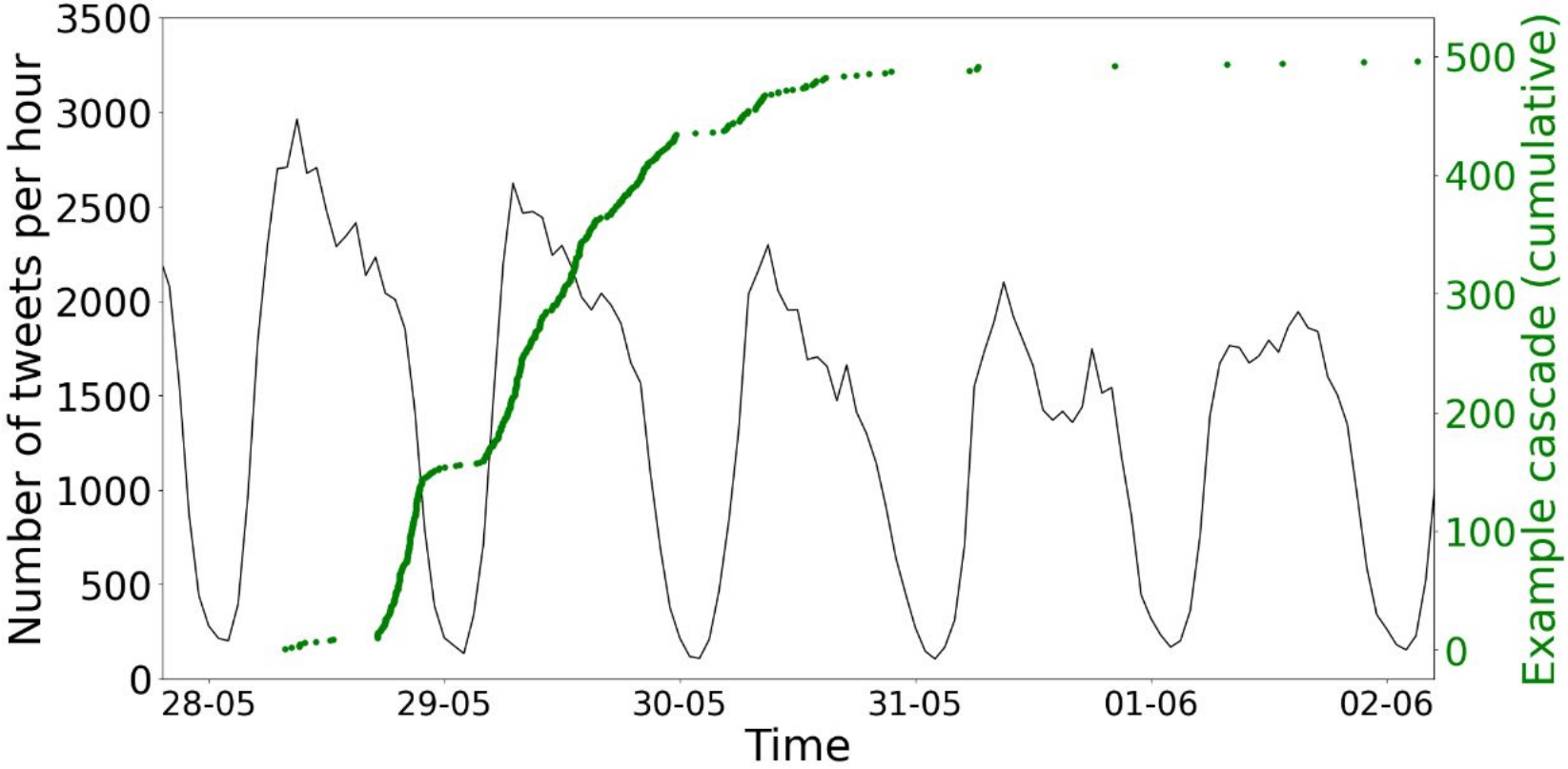
MSc thesis of Tiago Miranda

The diffusion of information on social media

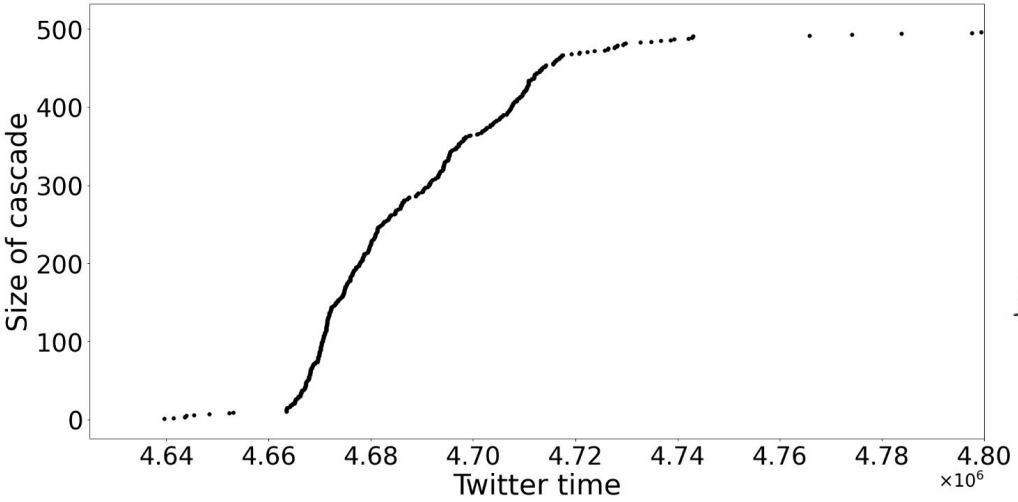




Time is not linear in social media



Model of information cascade growth

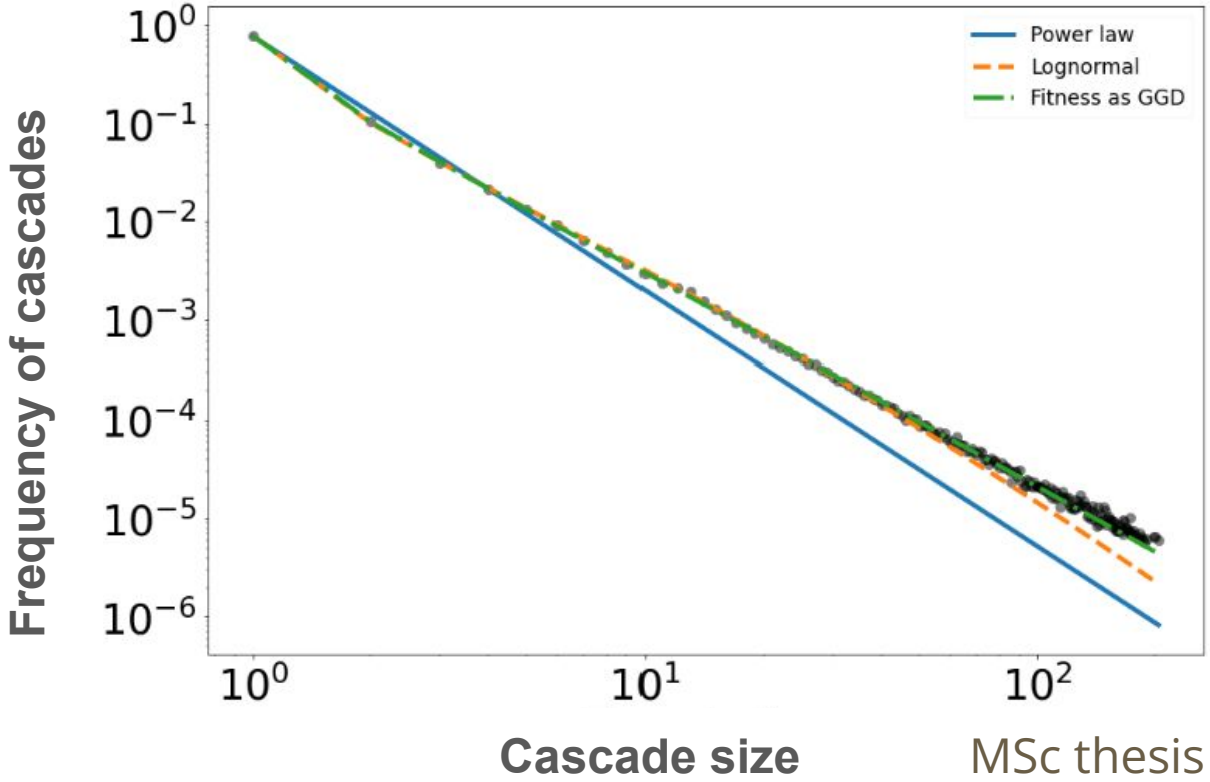


$$\frac{dN}{dt} = N \cdot a \cdot e^{-g \cdot t}$$

$$N(t) = N(0) \cdot e^{a/g} e^{-a \cdot e^{-g \cdot t}/g}$$

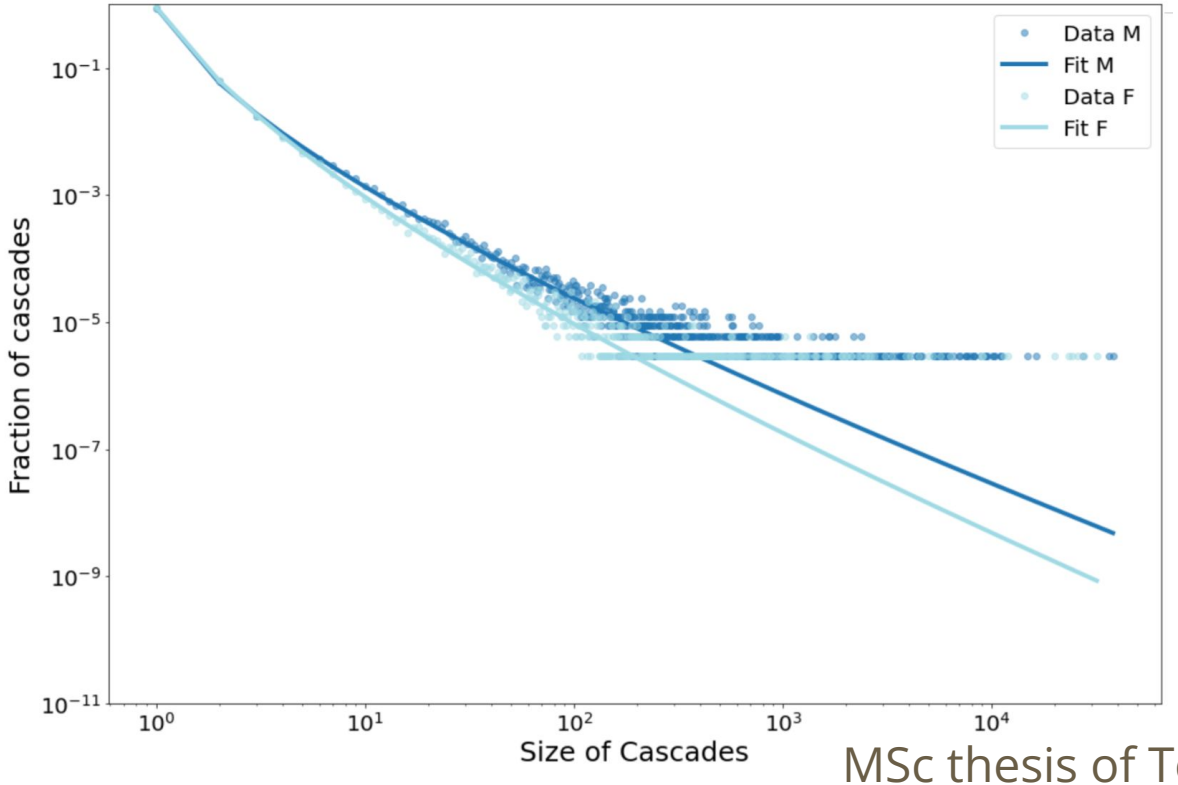
$$\lim_{t \to +\infty} N(t) = e^{\frac{a}{g}}$$

Cascade growth model gives rise to the observed size distribution



MSc thesis of Pedro Duarte (2023)

Information spreads differently depending on the gender of the original poster



MSc thesis of Tomás Silva

Time should be measured in attention rather than minutes in social media

A simple growth model (without the network) can explain the observed distribution of cascade sizes

Preliminary data point to socio-demographic biases in the success of tweets

How (un)biased are search engines?

- In a representative survey in the United States, three-quarters of respondents said they **trusted the information they found on search engines**: 28% do so for all or almost all, 45% for most information → Purcell K, Brenner J, Rainie L. Search engine use 2012. *Washington, DC*, https://www.eff.org/files/pew_2012_0.pdf (2012, accessed 11 January 2021).

- More than three-quarters (78%) of European Internet and online platform users **trust that the results displayed in search engines are the most relevant**. → European Commission. Special Eurobarometer 447: online platforms. Report, European Commission, Belgium, 2016.

- **Search engine trustworthiness is comparable to traditional news media,** as shown by a representative study of Internet users from 28 markets, including the United States, China and Germany. → Edelman. Edelman trust barometer2020, https://www.edelman.De/research/edelman-trust-barometer-2020 (2020, accessed 12 August 2020).

Iris Damião, Paulo Almeida

why am i so

- why am i so **tired**
- why am i so **ugly**
- why am i so **gassy**
- why am i so **thirsty**
- why am i so **angry**
- why am i so **itchy**
- why am i so **sad**
- why am i so **hungry**
- why am i so **emotional**
- why am i so **bloated**

como posso ser

- como posso ser **amigo de alguem**
- como posso ser **feliz**
- como posso ser **inteligente**
- como posso ser **uma pessoa melhor**
- como posso ser **salvo**
- como posso ser **rico**
- como posso ser **feliz sozinho**
- como posso ser **um hacker**
- como posso ser **popular no facebook**
- como posso ser **cantora**

pourquoi je suis

- pourquoi je suis **moche**
- pourquoi je suis **triste**
- pourquoi je suis **toujours fatigué**
- pourquoi je suis **célibataire**
- pourquoi je suis **toujours célibataire**
- pourquoi je suis **devenu rebelle pdf**
- pourquoi je suis **seule**
- pourquoi je suis **toujours fatiguée**
- pourquoi je suis **jalouse**
- pourquoi je suis **triste sans raison**

how to

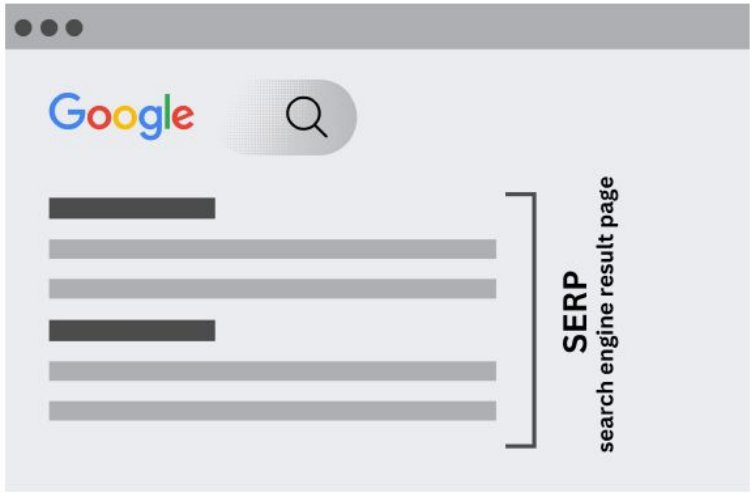
- how to **make slime**
- how to **tie a tie**
- how to **buy bitcoin**
- how to **lose weight**
- how to **draw**
- how to **buy ripple**
- how to **kiss**
- how to **make pancakes**
- how to **mine bitcoin**
- how to **train your dragon**

como é que se

- como é que se **beija**
- como é que se **diz eu te amo**
- como é que se **beija de lingua**
- como é que se **engravida**
- como é que se **beija na boca**
- como é que se **escreve**
- como é que se **beija pela primeira vez**
- como é que se **faz um facebook**
- como é que se **faz um relatório**
- como é que se **faz panquecas**

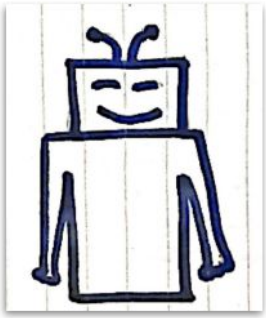
comment faire

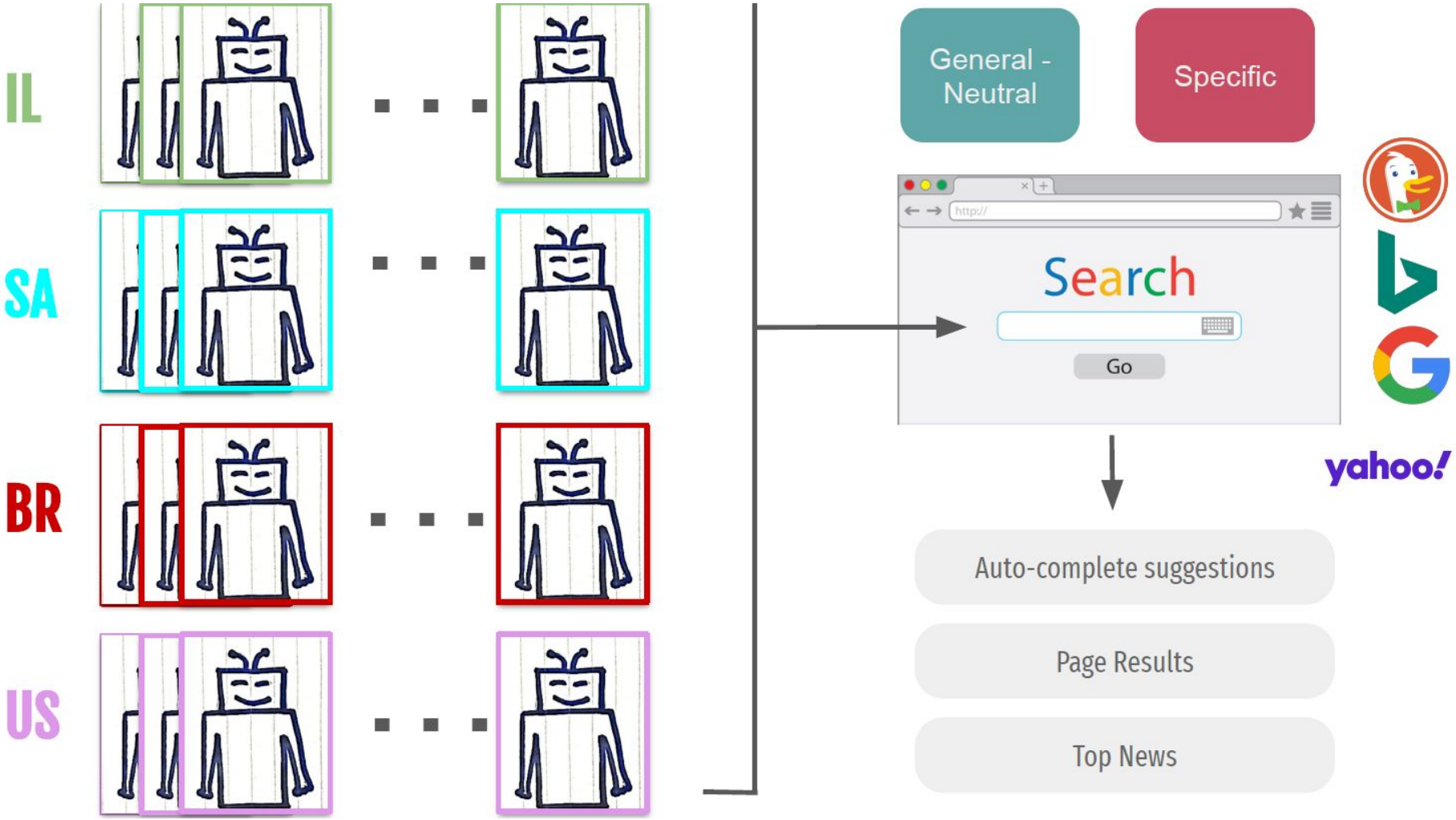
- comment faire **du slime**
- comment faire **un cv**
- comment faire **des crepes**
- comment faire **une dissertation**
- comment faire **une capture d'écran**
- comment faire **une bibliographie**
- comment faire **un gateau**
- comment faire **du caramel**
- comment faire **de la glue**
- comment faire **du pain**


Google
SERP
search engine result page

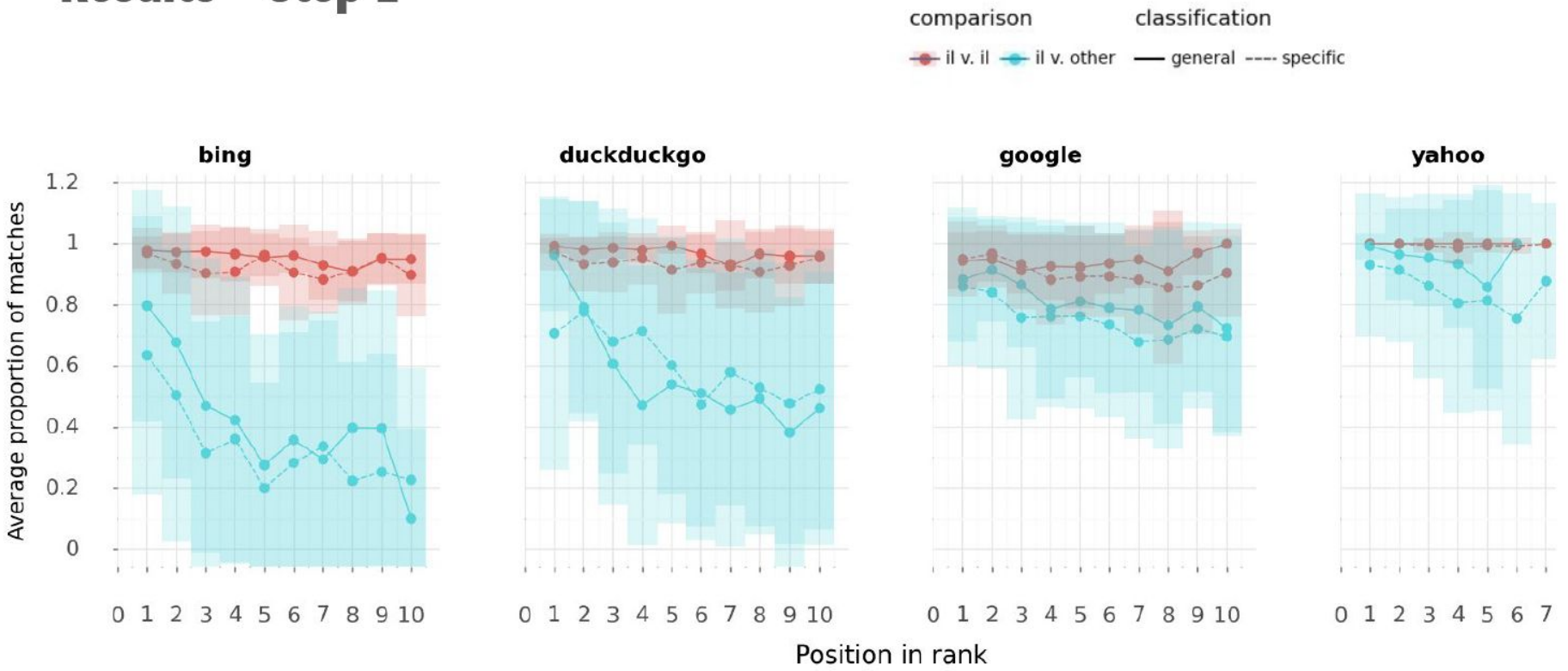
Web - Crawlers

1. Locations
2. Browsing definitions
3. Browsing histories




IL
SA
BR
US
General - Neutral
Specific
http://
Search
Go
yahoo!
Auto-complete suggestions
Page Results
Top News

Results – Step 1



On going conclusions

We observe substantially different results depending on the “location” of the bot

We also observe differences depending on the profile of the bot

Ongoing work to identify how meaningful these differences are

The SPAC group

PhD Students
Sara Mesquita, Public Health
Íris Damião, Computer Science

Postdocs
Cristina Mendonça, Psychology
Ana Vranic, Physics
José Reis, Law

MSc Students
Tiago Miranda, Data Science
Miguel Félix, Physics
Tomás Silva, Physics

Project Manager
Rita Saraiva

Senior Researcher
Lília Perfeito, Biology

Researchers
Hugo Cachitas, Programmer
Paulo Almeida, Lead Programmer, DPO
Hamid Shahzad, Part-time programmer

PI
Joana Gonçalves-Sá, Physics, Systems Biology