

# De Fisher até o Big Data: continuidades e descontinuidades

---

Renato Assunção  
DCC - UFMG

# Mensagem principal

---

- Estamos vivendo o início de uma revolução na história humana.

# Mensagem principal

---

- Estamos vivendo o início de uma revolução na história humana.
- Vocês tiraram a sorte grande!
- Estão no lugar certo, na hora certa.

# Mensagem principal

---

- Estamos vivendo o início de uma revolução na história humana.
- Vocês tiraram a sorte grande!
- Estão no lugar certo, na hora certa.
- O mundo inteiro está vivendo uma revolução.

# Mensagem principal

---

- Estamos vivendo o início de uma revolução na história humana.
- Vocês tiraram a sorte grande!
- Estão no lugar certo, na hora certa.
- O mundo inteiro está vivendo uma revolução.
- E estatística é um dos principais motores desta revolução.

# Mensagem principal

---

- Estatística também vive uma mudança fundamental, radical.

# Mensagem principal

---

- Estatística vive também uma mudança fundamental, radical.
- Aproveitem esta oportunidade única, que raramente aparece.

# Mensagem principal

- Estatística vive também uma mudança fundamental, radical.
- Aproveitem esta oportunidade única, que raramente aparece.
- Posicione-se para os próximos 30 anos.
- Você terá oportunidades incríveis.

E tudo começou em...



# E tudo começou em...

- Em 1500...



Livro Sapiens,  
de Yuval Harari

# E tudo começou em...

- Em 1500.
- A descoberta da ignorância.



# E tudo começou em...

---

- Em 1500.
- A descoberta da ignorância.
- Até 1500, religião ou filósofos antigos eram capazes de responder a todas as perguntas.

# Era só perguntar

---

- **Cristãos, judeus, chineses, muçulmanos...**
- **Tudo que precisava ser conhecido, já era conhecido.**

# Era só perguntar

---

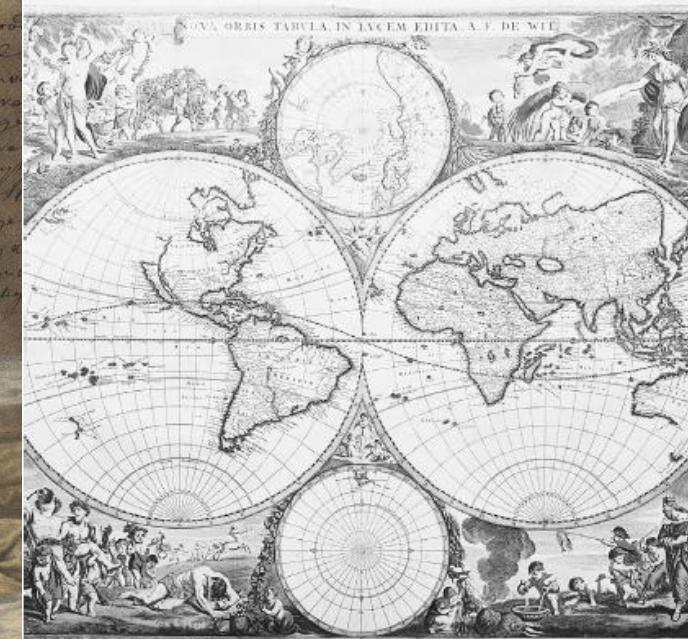
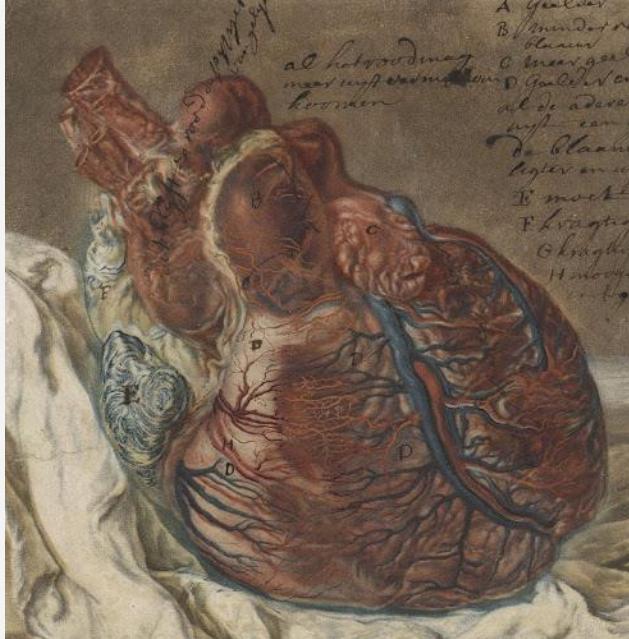
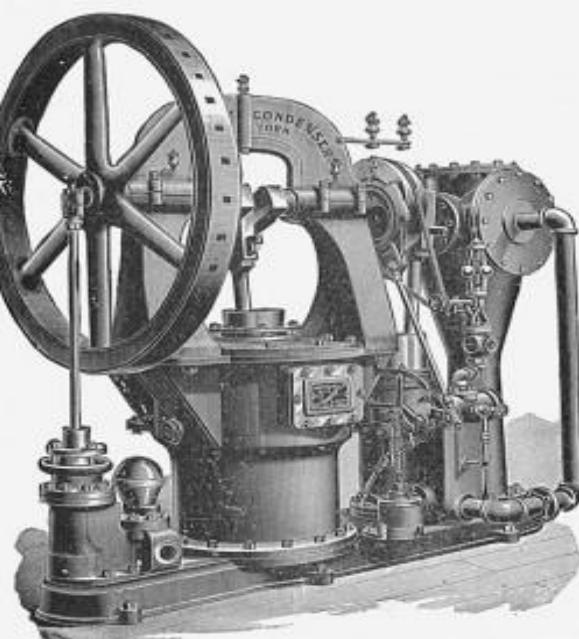
- Cristãos, judeus, chineses, muçulmanos...
- Tudo que precisava ser conhecido, já era conhecido.
- Se alguém não soubesse algo?
  - pergunte ao sábio de plantão. A resposta estaria nas escrituras sagradas.

# Era só perguntar

- Cristãos, judeus, chineses, muçulmanos: tudo que precisava ser conhecido, já era conhecido.
- Se alguém não soubesse algo?
  - pergunte ao sábio de plantão. A resposta estaria nas escrituras sagradas.
  - Se ninguém soubesse a resposta para a pergunta (como aparece o arco íris?), então ela é irrelevante.

# Ciência moderna

- O nascimento da ciência moderna significa um momento único.



# Ciência moderna

- O nascimento da ciência moderna significa um momento único.
- É quando descobrimos nossa ignorância coletiva.

# Ciência moderna

- O nascimento da ciência moderna significa um momento único.
- É quando descobrimos nossa ignorância coletiva.
- Admitimos abertamente que não sabemos.
- Admitimos também que, quando sabemos algo, não é a resposta final.

# Ciência moderna

---

- E também que não temos respostas ainda para questões fundamentais.

# Ciência moderna

---

- E também que não temos respostas ainda para questões fundamentais.
- Até hoje ainda não sabemos, por exemplo, como o cérebro produz consciência.

# Ciência moderna

- E também que não temos respostas ainda para questões fundamentais.
- Ainda não sabemos, por exemplo, como o cérebro produz consciência.
- E não ficamos parados: procuramosativamente respostas baseados em:
  - fatos empíricos,
  - previsões verificáveis
  - construídas com matemática.

# Ciência moderna e capitalismo

---

- Casamento muito bem sucedido.
- Riqueza, saúde, educação...

# Ciência moderna e capitalismo

---

- Casamento muito bem sucedido.
- Riqueza, saúde, educação...
- Compare um camponês europeu de 1500 e um trabalhador rural do Brasil em 2017.

# Comparando



# Ciência moderna e capitalismo

- Compare um rei francês em 1500 e um trabalhador rural do Brasil em 2017.



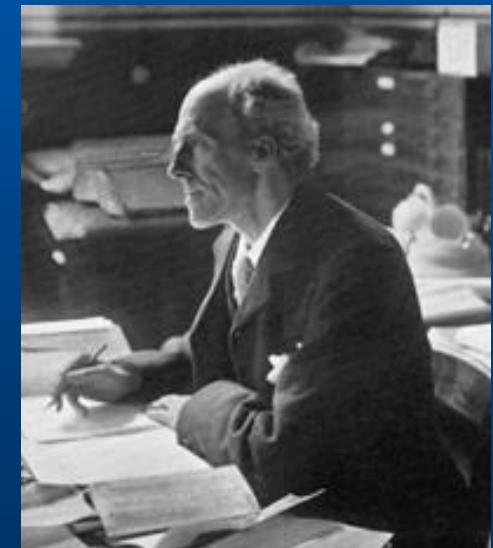
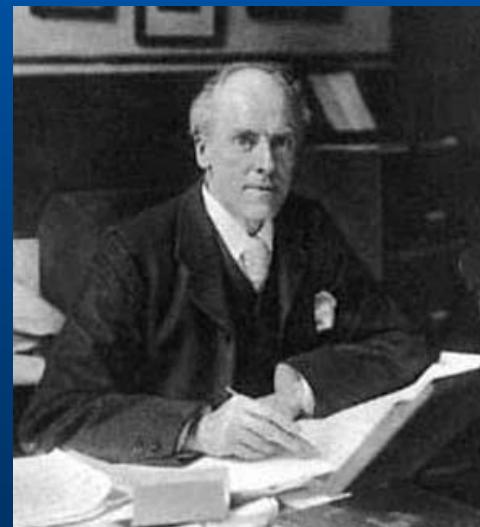
# Um salto na história

---

# E tudo começa em ...

---

- 1900
- Karl Pearson



# Pearson foi um gigante

---

- Movimento eugênico.
- Correlation coefficient.
- Method of moments.
- Pearson's system of continuous curves.
- Chi Squared test.
- P-value.
- Principal component analysis.

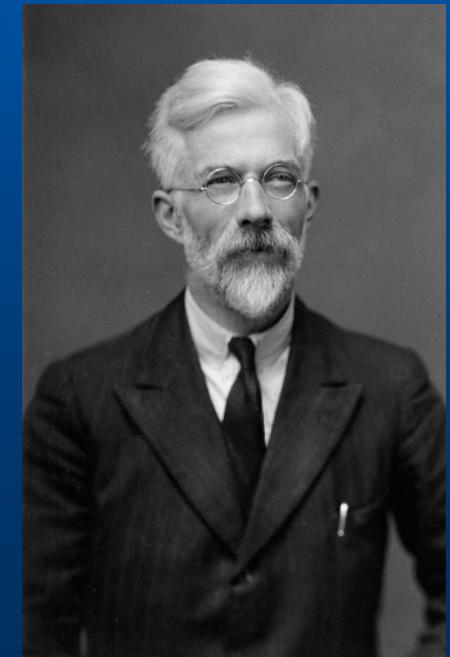
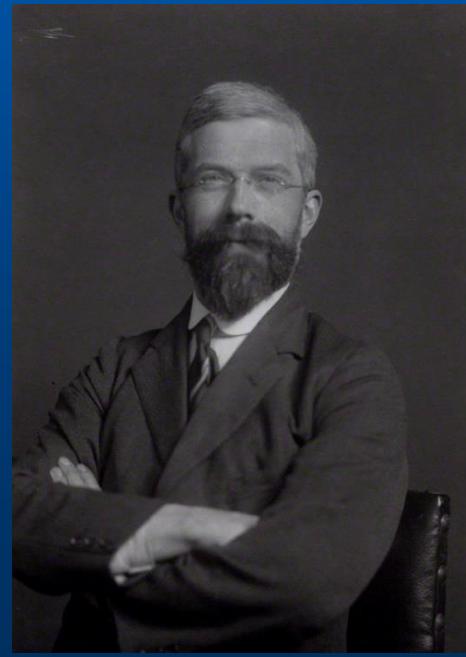
# Pearson e a estatística

---

- Fundou o 1º departamento de estatística do mundo em 1911: Department of Applied Statistics, University of London, como apoio da Drapers' Company
- Fundou a Biometrika em 1901, revista científica de estatística, viva e importante até hoje.

# E a luz se fez...

- Ronald A. Fisher
- Dois artigos, em 1922 e 1925 (aos 32-35 anos de idade)



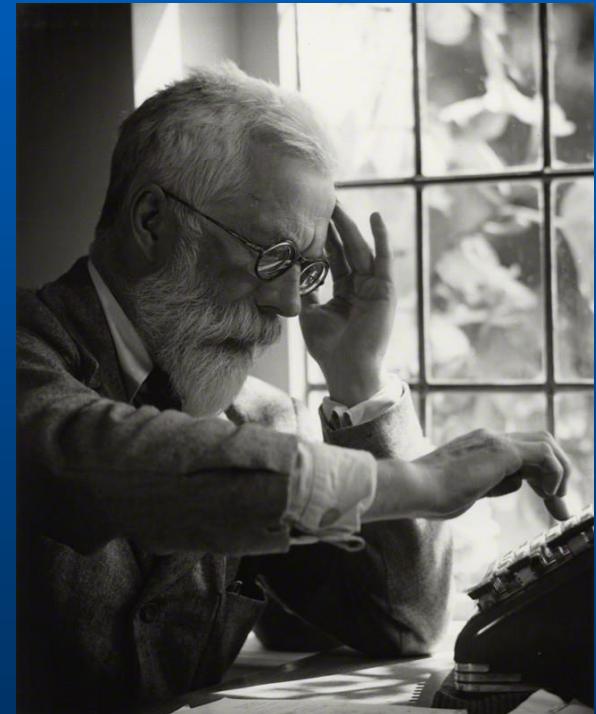
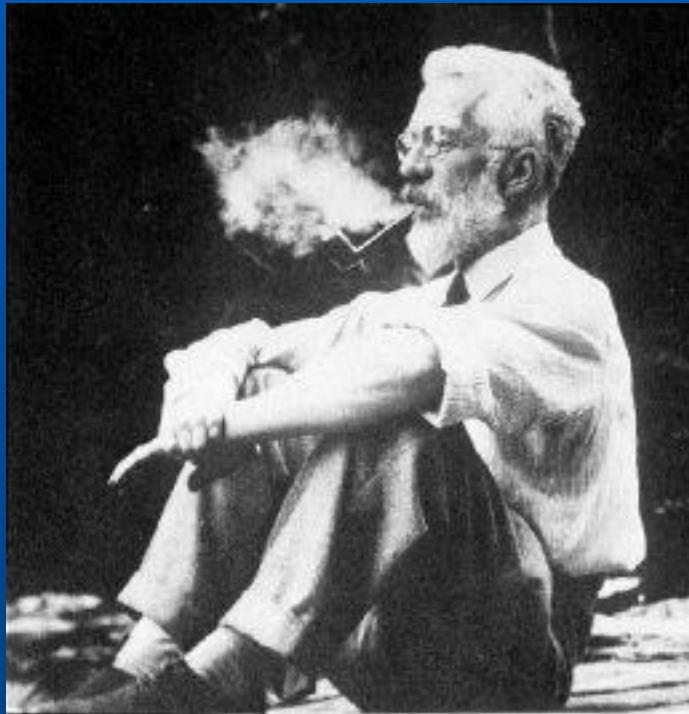
# Quem foi Fisher?

---

- Anders Hald: "um gênio que criou praticamente sozinho as fundações para a moderna ciência estatística"
- Richard Dawkins: "o maior dos sucessores de Darwin"

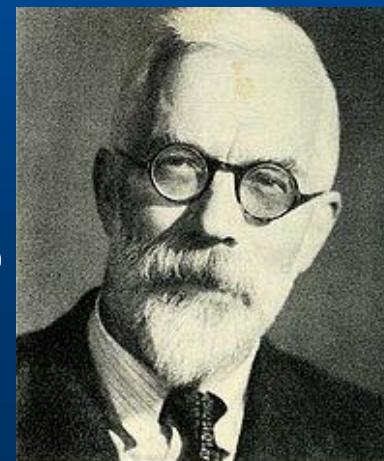
# Mais dois momentos Fisher

---



# Artigo de 1922 e de 1925

- Fisher, R. A. (1922). "On the Mathematical Foundations of Theoretical Statistics" *Philosophical Transactions of the Royal Society A*. **222** (594–604): 309–368
- Fisher, R. A. (1925). "Theory of Statistical Estimation". *Proceedings of the Cambridge Philosophical Society*. **22** (5): 700–725.



# Principia Statistica

## IX. On the Mathematical Foundations of Theoretical Statistics.

By R. A. FISHER, M.A., Fellow of Gonville and Caius College, Cambridge, Chief Statistician, Rothamsted Experimental Station, Harpenden.

Communicated by DR. E. J. RUSSELL, F.R.S.

Received June 25,--Read November 17, 1921.

### CONTENTS.

Section		Page
1.	The Neglect of Theoretical Statistics	310
2.	The Purpose of Statistical Methods	311
3.	The Problems of Statistics	313
4.	Criteria of Estimation	316
5.	Examples of the Use of Criterion of Consistency	317
6.	Formal Solution of Problems of Estimation	323
7.	Satisfaction of the Criterion of Sufficiency	330
8.	The Efficiency of the Method of Moments in Fitting Curves of the Pearsonian Type III	332
9.	Location and Scaling of Frequency Curves in general	338
10.	The Efficiency of the Method of Moments in Fitting Pearsonian Curves	342
11.	The Reason for the Efficiency of the Method of Moments in a Small Region surrounding the Normal Curve	355
12.	Discontinuous Distributions	356
	(1) The Poisson Series	359
	(2) Grouped Normal Data	359
	(3) Distribution of Observations in a Dilution Series	363
13.	Summary	366

### DEFINITIONS.

*Centre of Location.*—That abscissa of a frequency curve for which the sampling errors of optimum location are uncorrelated with those of optimum scaling. (9.)

*Consistency.*—A statistic satisfies the criterion of consistency, if, when it is calculated from the whole population, it is equal to the required parameter. (4.)

*Distribution.*—Problems of distribution are those in which it is required to calculate the distribution of one, or the simultaneous distribution of a number, of functions of quantities distributed in a known manner. (3.)

*Efficiency.*—The efficiency of a statistic is the ratio (usually expressed as a percentage) which its intrinsic accuracy bears to that of the most efficient statistic possible. It

700 *Mr Fisher, Theory of statistical estimation*

*Theory of Statistical Estimation.* By Mr R. A. FISHER, Gonville and Caius College.

[Received 17 March, read 4 May, 1925.]

### PREFATORY NOTE.

It has been pointed out to me that some of the statistical ideas employed in the following investigation have never received a strictly logical definition and analysis. The idea of a frequency curve, for example, evidently implies an infinite hypothetical population distributed in a definite manner; but equally evidently the idea of an infinite hypothetical population requires a more precise logical specification than is contained in that phrase. The same may be said of the intimately connected idea of random sampling. These ideas have grown up in the minds of practical statisticians and lie at the basis especially of recent work; there can be no question of their pragmatic value. It was not part of my original intention to deal with the logical bases of these ideas, but some comments which Dr Burnside has kindly made have convinced me that it may be desirable to set out for criticism the manner in which I believe the logical foundations of these ideas may be established.

The idea of an infinite hypothetical population is, I believe, implicit in all statements involving mathematical probability. If, in a Mendelian experiment, we say that the probability is one half that a mouse born of a certain mating shall be white, we must conceive of our mouse as one of an infinite population of mice which might have been produced by that mating. The population must be infinite for in sampling from a finite population the fact of one mouse being white would affect the probability of others being white, and this is not the hypothesis which we wish to consider; moreover, the probability may not always be a rational number. Being infinite the population is clearly hypothetical, for not only must the actual number produced by any parents be finite, but we might wish to consider the possibility that the probability should depend on the age of the parents, or their nutritional conditions. We can, however, imagine an unlimited number of mice produced upon the conditions of our experiment, that is, by similar parents, of the same age, in the same environment. The proportion of white mice in this imaginary population appears to be the actual meaning to be assigned to our statement of probability. Briefly, the hypothetical population is the conceptual resultant of the conditions which we are studying. The probability, like other statistical parameters, is a numerical characteristic of that population.

We only need the conception of an infinite hypothetical population, in connection with random sampling. The ultimate logical elucidation of the one idea implies that of the other. Also, the word infinite is to be taken in its proper mathematical sense as denoting the limiting conditions approached by increasing a finite number indefinitely. I imagine that an exact meaning can be given to all the ideas required by some process such as the following.

Imagine a population of  $N$  individuals belonging to  $s$  classes, the number in class  $k$  being  $p_k N$ . This population can be arranged in order in  $N!$  ways. Let it be so arranged and let us call the first  $n$  individuals in each arrangement a sample of  $n$ . Neglecting the order within the sample, these samples can be classified into the several possible types of sample according to the number of individuals of each class which appear. Let this be done, and denote the proportion of samples which belong to type  $j$  by  $g_j$ , the number of types being  $t$ . Consider the following proposition.

Given any series of proper fractions  $P_1, P_2, \dots, P_s$ , such that  $S(P_k)=1$ ,

# Fisher (1922) e Fisher (1925)

- Fisher tinha 32 e 35 anos de idade.
- Nestes dois artigos:
  - Informação de Fisher: Menor variância possível...
  - Máxima verossimilhança
  - Propriedades de MLE: atinge o ótimo
  - Suficiência
- Criou os conceitos que nos guiam até hoje

# Tomando decisões

---

- 1940-1950
- Abraham Wald



# Contribuições

---

- Teoria da decisão
  - Estados da natureza
  - Dados
  - Ações
  - Função utilidade
- Máxima verossimilhança: propriedades assintóticas e testes
- Análise sequencial

# Um trabalho durante a guerra

## PART II

### MAXIMUM VALUE OF THE PROBABILITY THAT A PLANE WILL BE DOWNED BY A GIVEN NUMBER OF HITS<sup>1</sup>

The symbols defined and the results obtained in part I will be used here without further explanation. The purpose of this

memorandum is to derive the least upper bound of  $x_i = \sum_{j=1}^i x_j$  and

that of  $P_i$  ( $i = 1, \dots, n$ ) under the restriction that

$q_1 \geq q_2 \geq \dots \geq q_n$ .

First, we shall show that  $x_i$  is a strictly increasing function of  $p_j$  for  $j \leq i$ . Let us replace  $p_j$  by  $p_j + \delta$  ( $\delta > 0$ ) and let us study the effect of this change on  $x_1, \dots, x_i$ . Denote the changes in  $x_1, \dots, x_i$  by  $\delta_1, \dots, \delta_i$ , respectively. Clearly,

$\delta_1 = \dots = \delta_{j-1} = 0$ . It follows easily from equation 9 that

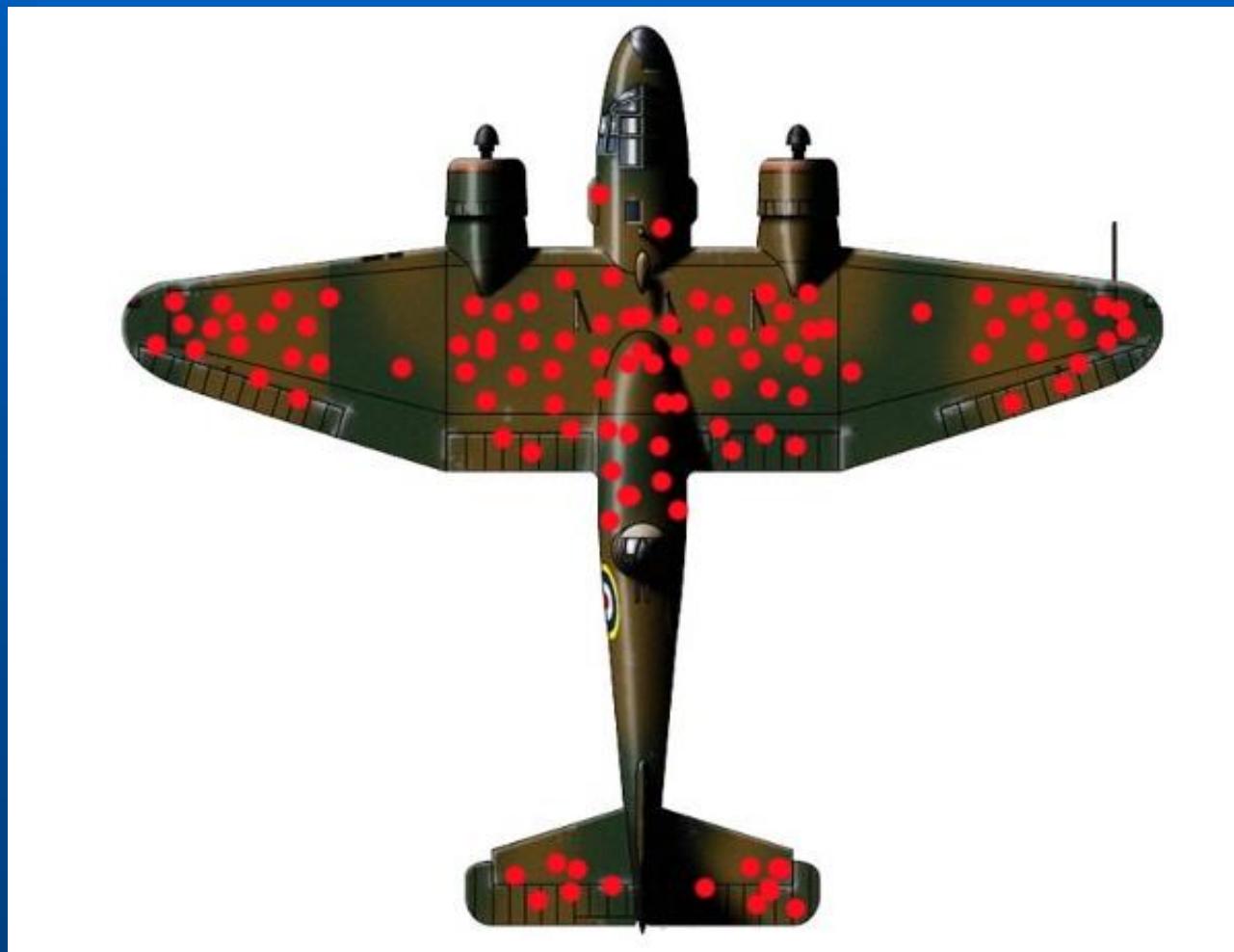
$\delta_j > 0$  and

$$\delta_{j+1} = -p_{j+1} \delta_j$$

Hence,

$$\delta_j + \delta_{j+1} = (1 - p_{j+1}) \delta_j > 0.$$

# Um trabalho durante a guerra



# Matematizando a Estatística

- 50's e 60's: Matematização da estatística
  - Berkeley: Estimadores UMVU, Testes UMPs, teoria da decisão
  - Resultados ótimos limitados a situações simples (1-2 parâmetros) ou assintóticas.
  - Pacotes estatísticos começam a espalhar-se (SAS, SPSS,...)
  - No fim do período: Necessidade de atacar problemas maiores e mais complexos (mais parâmetros, distribuições não normais...)

# Anos 70

- E nascem os modelos de análise de dados
- Modelos Box-Jenkins ou ARIMA (1970)
- Modelo linear generalizado (1972)
- Riscos proporcionais de Cox (1972)
- Diagnóstico em regressão múltipla:
  - D de Cook (1977), DFFITS (B+K+W, 1980)
- Tukey: EDA e robustez (1977)

# O tempo escorre em bits e bytes...

- 80's: Métodos algorítmicos (e não apenas numéricos)
  - Efron - Bootstrap (1982)
  - CART (1984), Loess (1988), ACE(1985), Projection Pursuit (1985)
  - Kernel smoothing
- Pacotes estatísticos tradicionais mostram suas limitações:
  - precisamos de ambientes de análise de dados
  - e não de uma coleção de procedimentos

# Parâmetros vão ao infinito

- 90's: modelos pouco parcimoniosos, com muitos parâmetros ou não paramétricos, modelos mais realistas, alta capacidade preditiva.
  - GAM, Redes Neurais, Support Vector Machine, random forests, Bagging...
  - MCMC: Reverendo Bayes invade a praia
  - Inferência sobre estruturas de dados mais complexas: dados espaciais, grafos, árvores de ramificação...
  - Modelos de efeitos mistos – aleatórios
- S, Splus, R, Ox, BUGS
- E mais C++, Java...

# Novo milênio: 2000 a 2015

- Modelos espaciais e espaço-temporais
- Modelos para os IMENSOS grafos aleatórios (web, redes sociais)
- Processamento natural de linguagem
- Recomendação de itens a usuários
- Robótica: taxi sem motorista
- Deep Learning
- LDA: Organização automática de textos

# Sensores em todos os lugares

---

- **Internet das coisas**
- **Nano-bio sensores**
- **Exoesqueleto**
- **Robôs cuidando de idosos e doentes**
- **Drones carregando encomendas e atendimento de urgência**
- **Carros autonômos**

# Atualizando Fisher

- "Suficiência" evoluiu para "compressão de dados"
- "eficiência" → computação e escala.
- "parâmetros" e "máxima verossimilhança" → modelos semi e não paramétricos, centenas de parâmetros, com robustez.
- Bayes → tornou-se possível, na prática.
- "bootstrap" e "aprendizado de máquina" → algoritmos probabilísticos.

# Alguns exemplos pessoais

---

# Infection Hot Spot Mining from Social Media Trajectories

Roberto Souza, Renato Assunção, Derick Oliveira, Denise Brito  
and Wagner Meira Jr.

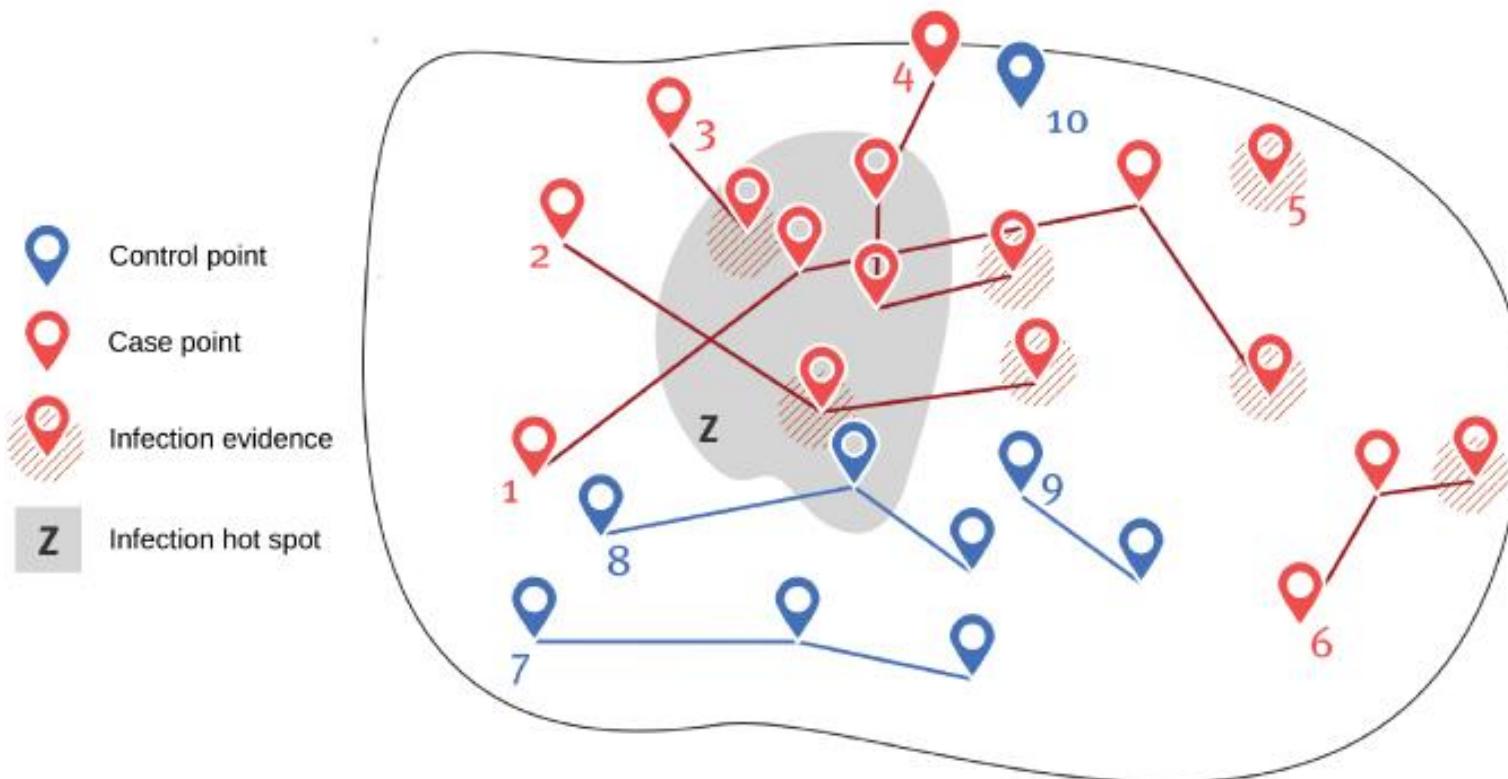
Department of Computer Science  
Universidade Federal de Minas Gerais – Brazil



2016 RIVA DEL GARDA

September 20, 2016

# Problem Definition



# Dois modelos

---

- **Modelo de visita: the most likely zone that a diseased person visits**
- **Modelo de infecção:the most likely zone a person gets infected while visiting**
- **Mesmos dados → Diferentes verossimilhanças**

# Contrastando os dois modelos

## Visit Model:

$P(\text{Tweets from } Z \mid \text{Is a case})$

## Infection Model:

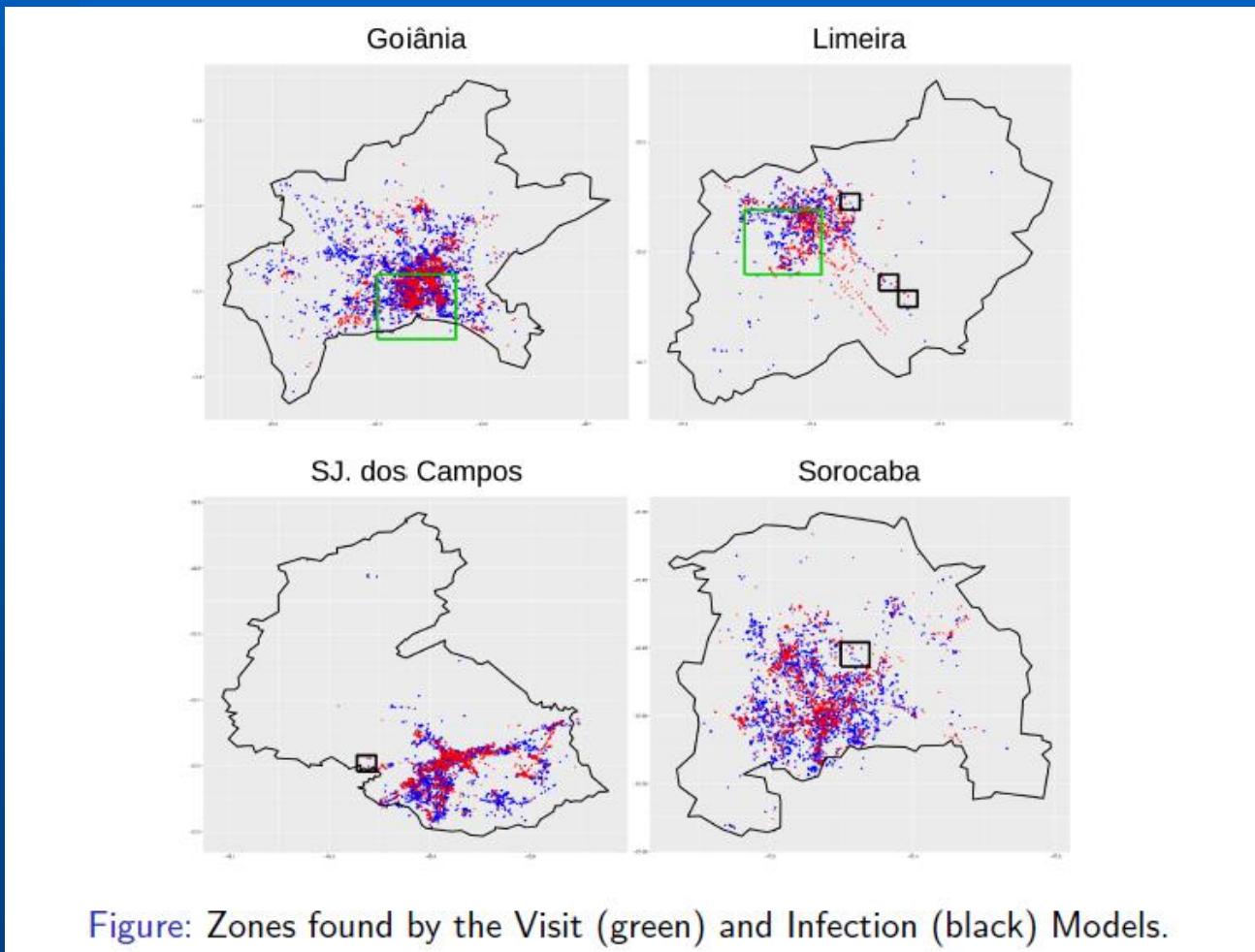
$P(\text{Is a case} \mid K \text{ Tweets from } Z)$

# Dados



- We collected the data from Twitter.
- Geo-located tweets from Brazil.
- Period: Jan 1st, 2015 – Dec 31th, 2015.
- Total of 106,784,441 tweets.
- City-level analysis.
- Selected 11 municipalities, including cities facing strong surges.

# Alguns resultados



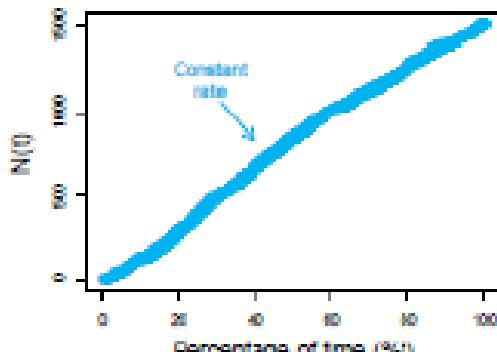


*BuSe: um modelo de mistura de processos pontuais estocásticos para tempos entre eventos de serviços na Web*

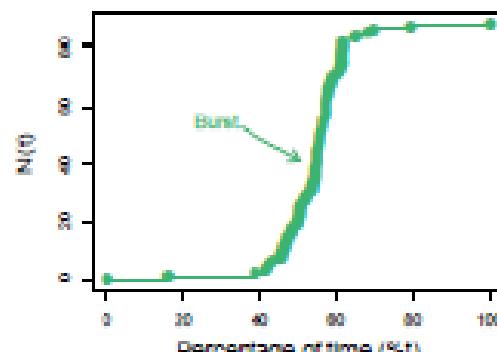
Rodrigo Alves, Renato Assunção e Pedro Vaz-de-Melo

DCC- UFMG

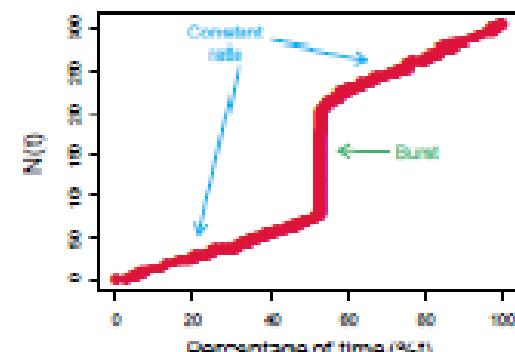
# Exemplos de dados do Twitter



(a) #wheretheydothatat

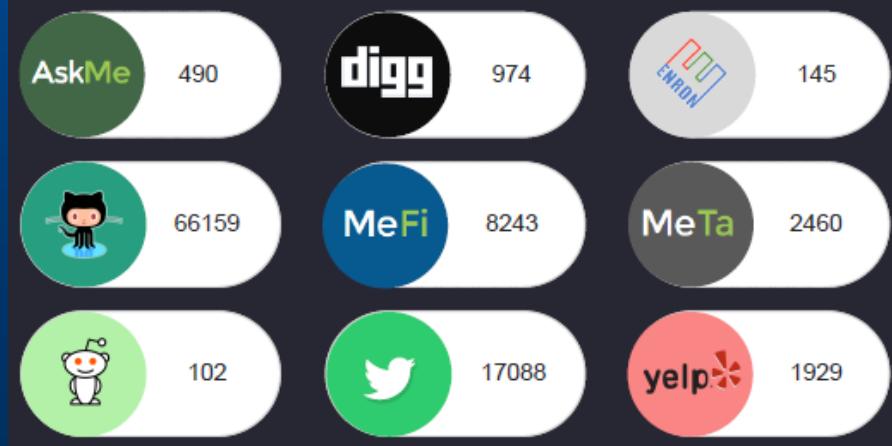


(b) #cotto

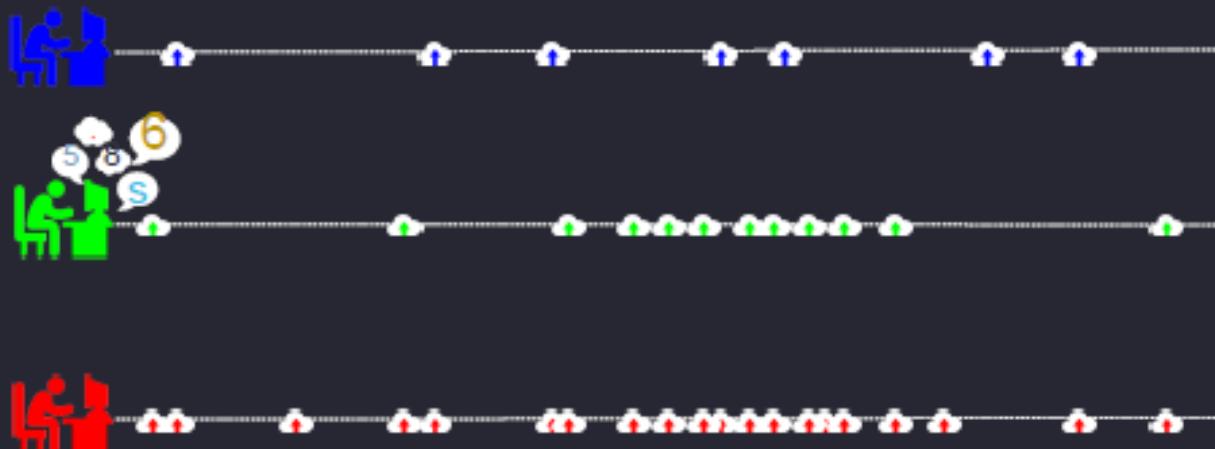


(c) #ta

## Bases de dados



# Modelo de Mistura



## Intensidade dos Processos

PP

Processo de Poisson

$$\lambda(t|H_t) = \lambda_{pp}$$

SFP

Self-Feeding Process

$$\lambda_s(t|\mathcal{H}_t) = \frac{1}{\mu/e + \Delta t_i}$$

where  $\Delta t_i = t_i - t_{i-1}$

and  $t_i = \max_k \{t_k : t_k \leq t\}$

# Função de verossimilhança da mistura

$$\lambda_{SFP}(t|H_t) = \begin{cases} \frac{1}{\mu/\epsilon} & \text{se } t \leq t_1 \\ \frac{1}{\mu/\epsilon + t_1} & \text{se } t_1 < t \leq t_2 \\ \frac{1}{\mu/\epsilon + \Delta t} & \text{se } t > t_2 \end{cases}$$

$$Pr\{t_1 \in SFP\} = 1 \text{ e } Pr\{t_2 \in SFP\} = 1$$

$$Pr\{t_i \in SFP\} = \frac{\lambda_{SFP}(t_i|H_{t_i})}{\lambda_{MISTURA}(t_i|H_{t_i})} = \frac{\lambda_{SFP}(t_i|H_{t_i})}{\lambda_{SFP}(t_i|H_{t_i}) + \lambda_{PP}}$$

Por consequência:

$$Pr\{t_i \in PP\} = \frac{\lambda_{PP}}{\lambda_{SFP}(t_i|H_{t_i}) + \lambda_{PP}} = 1 - Pr\{t_i \in SFP\}$$

$$\begin{aligned} E[\lambda_{SFP}(t_i|H_{t_i})] &= Pr\{t_{i-1} \in PP\} \times E[\lambda_{SFP}(t_{i-1}|H_{t_{i-1}})] + Pr\{t_{i-1} \in SFP\} \\ &\quad \times \sum_{k=2}^{i-2} \left\{ Pr\{t_k \in SFP\} \times [\prod_{j=k+1}^{i-2} Pr\{t_j \in PP\}] \times E[\lambda_{SFP}(t_i|H_{(t_k, t_{i-1})})] \right\} \end{aligned}$$

# A geometria das redes sociais

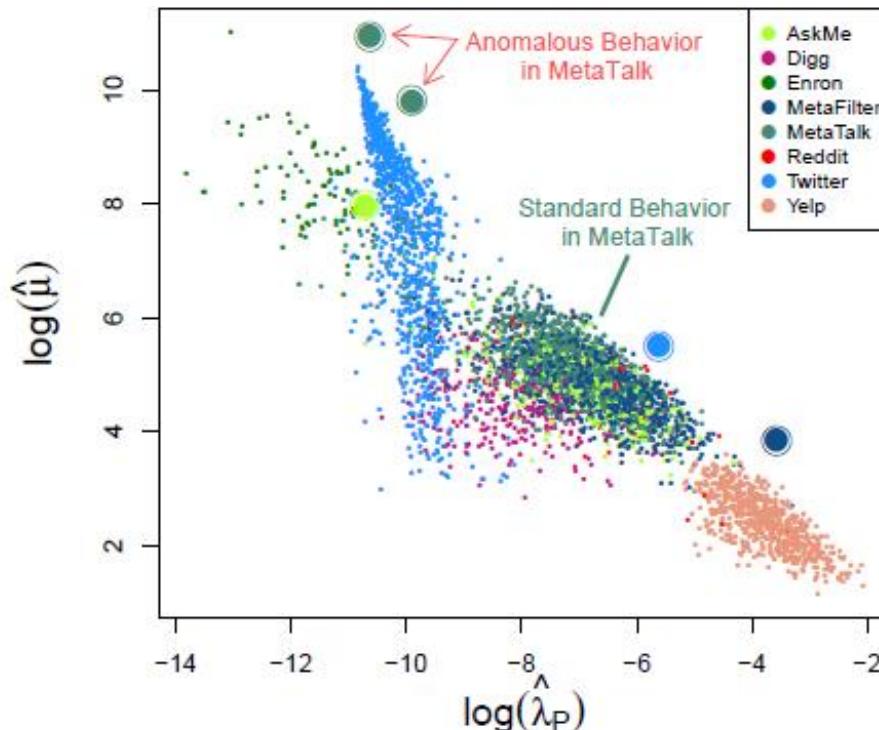
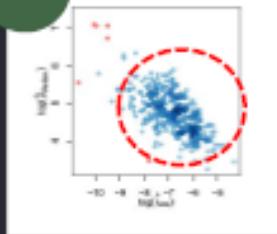


Figure 11: Estimates  $(\hat{\lambda}_p, \hat{\mu})$  for all events streams from the eight databases (logarithmic scale). A few anomalous time series are highlighted as large dots.

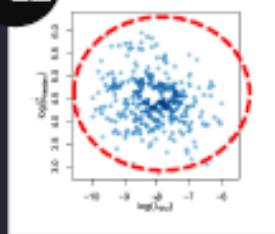
# Detecção de anomalias

## Detecção de Anomalias

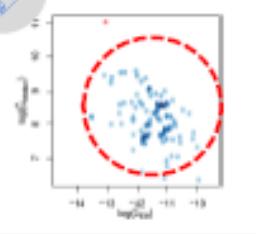
AskMe



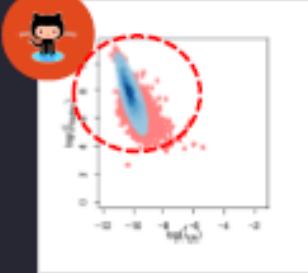
digg



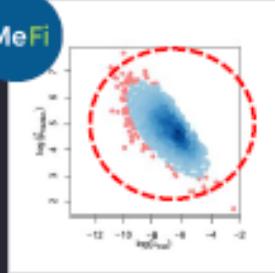
Techno



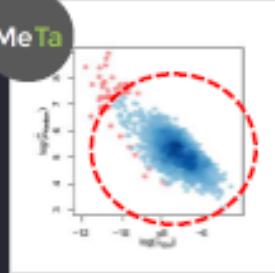
GitHub



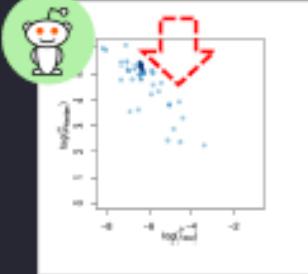
MeFi



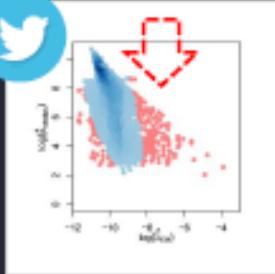
MeTa



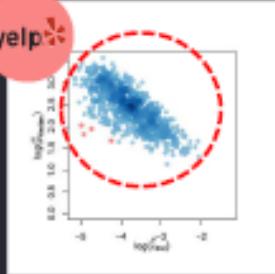
Robot



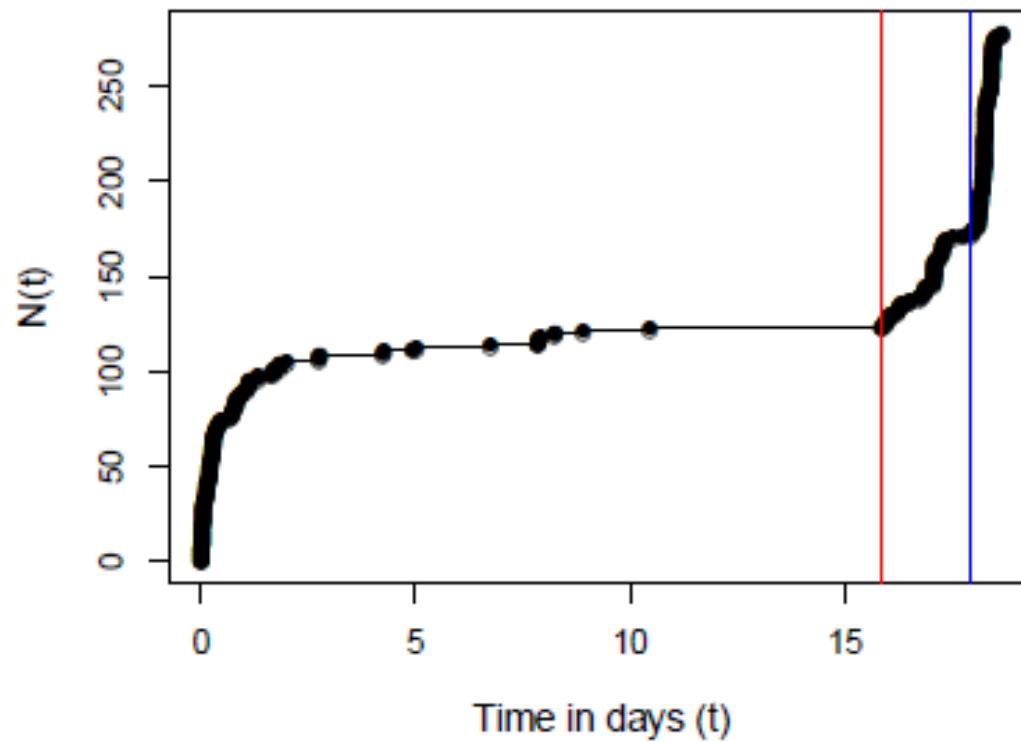
Twitter



yelp



# Onde está o Totó?



(a) Askme: #219940

# Experiências em robótica

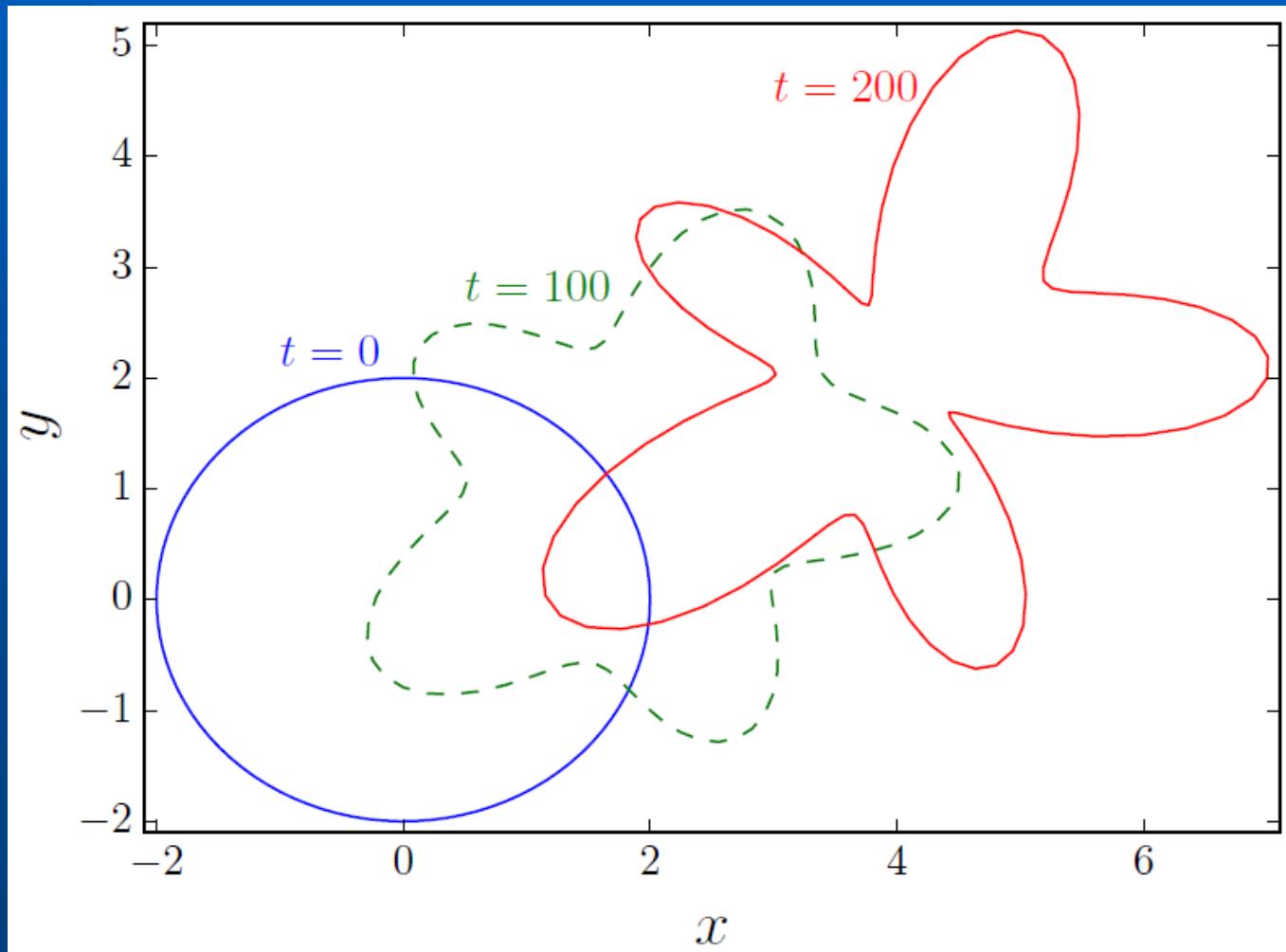
IEEE ROBOTICS AND AUTOMATION LETTERS. PREPRINT VERSION. ACCEPTED JANUARY, 2016

1

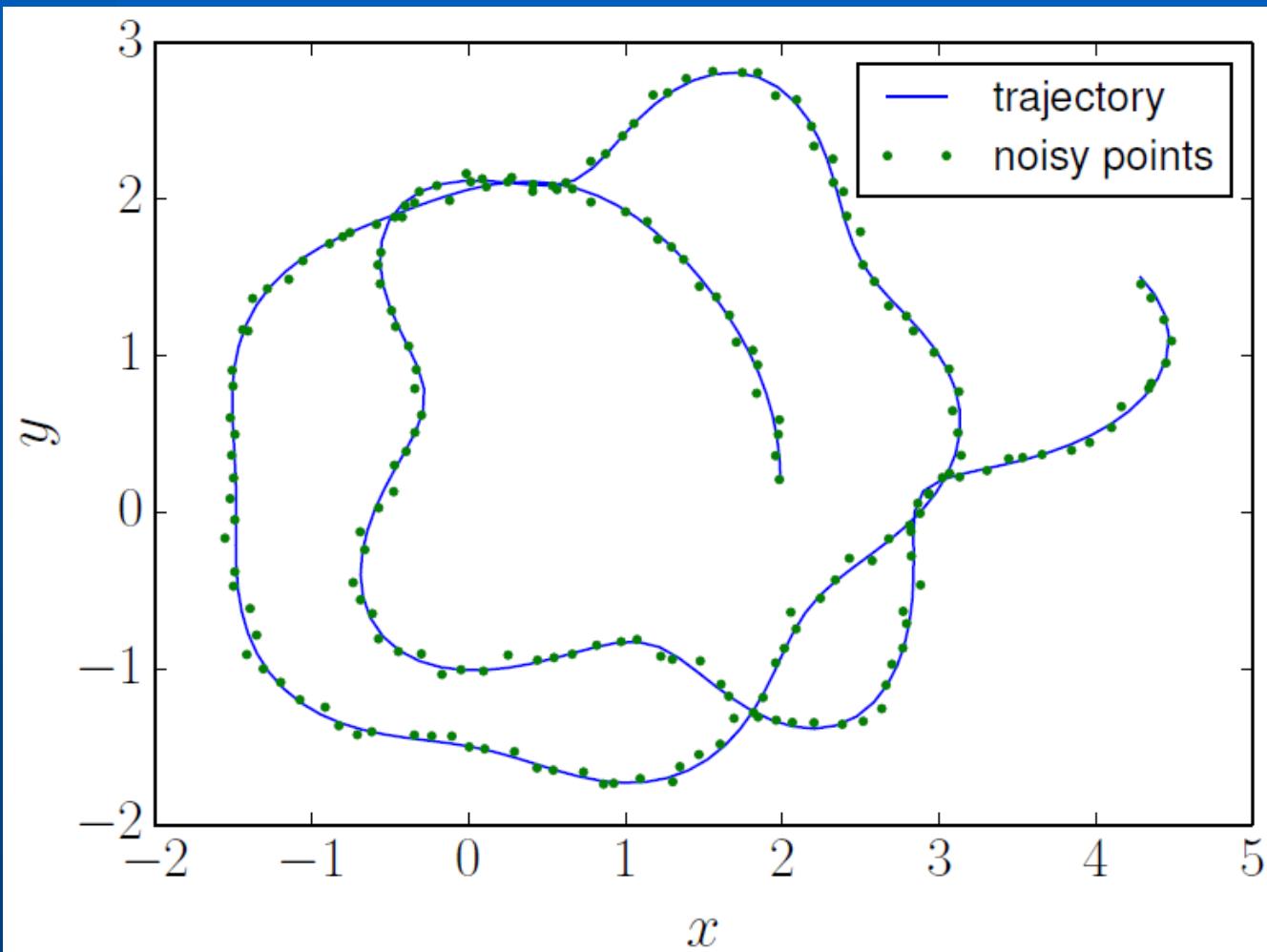
## Predicting Environmental Boundary Behaviors with a Mobile Robot

David Saldaña<sup>1</sup>, Renato Assunção<sup>1</sup>, and Mario F. M. Campos<sup>1</sup>

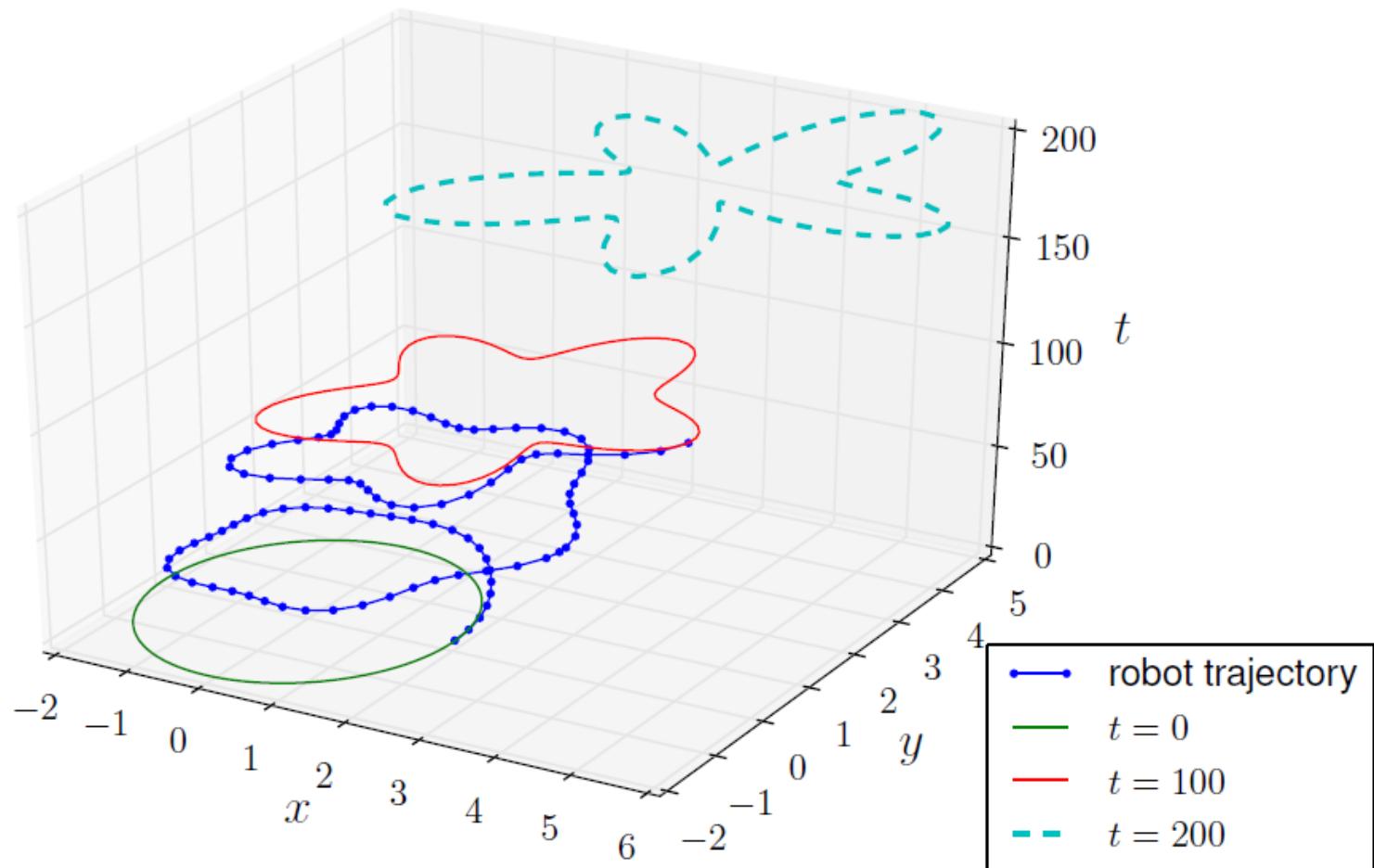
# Anomalia dinâmica



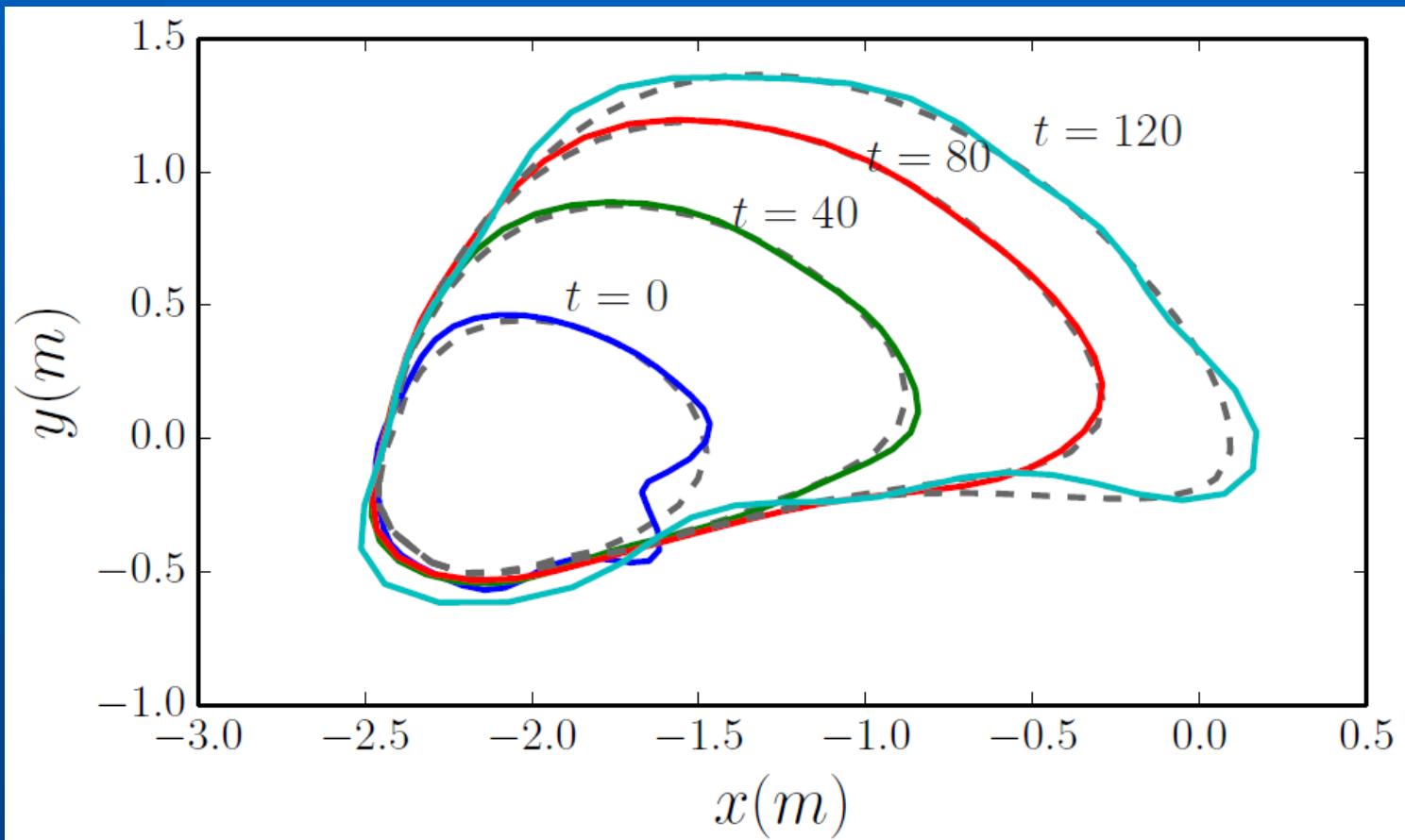
# Incerteza também na fronteira



# Visão 3-dim



# Resultados



[Video](#)

# Nuvens passageiras

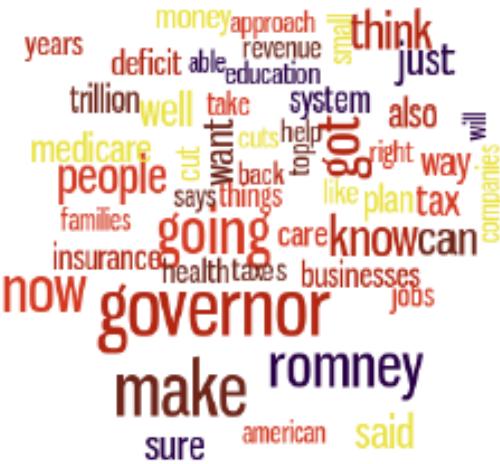
# Mais de uma ao mesmo tempo



A word cloud visualization of Governor Romney's speech. The words are colored in shades of orange, yellow, and red, with larger words representing more frequent terms. The most prominent words include 'Romney' (large, red), 'make' (large, orange), 'got' (large, orange), 'plan' (large, orange), 'small' (medium, orange), 'businesses' (medium, orange), 'now' (medium, orange), 'taxes' (medium, orange), 'cut' (medium, orange), 'jobs' (medium, orange), 'help' (medium, orange), 'sure' (medium, orange), 'just' (medium, orange), 'also' (medium, orange), 'getting' (medium, orange), 'insurance' (medium, orange), 'care' (medium, orange), 'people' (medium, orange), 'know' (medium, orange), 'way' (medium, orange), 'think' (medium, orange), 'right' (medium, orange), 'Now' (medium, orange), 'tax' (medium, orange), 'years' (medium, orange), 'reason' (medium, orange), 'trillion' (medium, orange), 'tax' (medium, orange), 'health' (medium, orange), 'Medicare' (medium, orange), 'education' (medium, orange), 'care' (medium, orange), 'system' (medium, orange), 'cuts' (medium, orange), 'revenue' (medium, orange), 'back' (medium, orange), 'American' (medium, orange), 'opportunity' (medium, orange), and 'things' (medium, orange). The word 'Governor' is displayed in large, bold, yellow letters at the bottom left of the cloud.

able sanctions people now making work nuclear also Israel kind Iraq Americans tale done way jobs years need back just sure make world troops China military Iran Governor region want come Afghanistan America countries help got know

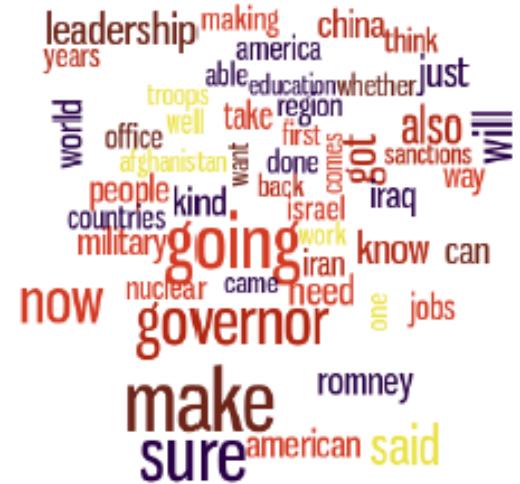
# Nossa proposta: COWORDS



A word cloud visualization of political speech content. The words are colored in a rainbow gradient. The most prominent words include 'years', 'think', 'just', 'got', 'gov', 'make', 'romney', 'sure', 'american', and 'said'. Other visible words include 'money', 'approach', 'deficit', 'revenue', 'education', 'trillion', 'well', 'take', 'system', 'also', 'medicare', 'cut', 'want', 'cuts', 'help', 'right', 'way', 'people', 'cut', 'say', 'things', 'back', 'top', 'like', 'plant', 'tax', 'families', 'going', 'care', 'know', 'can', 'insurance', 'health', 'taxes', 'businesses', 'now', 'governor', 'make', 'romney', 'sure', 'american', and 'said'.



A word cloud visualization of political speech content. The words are colored in a rainbow gradient. The most prominent words include 'years', 'think', 'just', 'got', 'gov', 'make', 'romney', 'sure', 'american', and 'said'. Other visible words include 'making', 'everybody', 'candy', 'percent', 'education', 'new', 'young', 'well', 'actually', 'say', 'gas', 'done', 'believe', 'like', 'oil', 'president', 'tax', 'families', 'country', 'going', 'governor', 'make', 'romney', 'sure', 'american', and 'said'.



A word cloud visualization of political speech content. The words are colored in a rainbow gradient. The most prominent words include 'leadership', 'making', 'china', 'think', 'just', 'got', 'also', 'will', 'world', 'office', 'well', 'take', 'region', 'first', 'sanctions', 'will', 'people', 'kind', 'back', 'iran', 'iraq', 'countries', 'military', 'going', 'governor', 'make', 'romney', 'sure', 'american', and 'said'. Other visible words include 'years', 'trrops', 'afghanistan', 'want', 'done', 'comes', 'go', 'israel', 'work', 'know', 'can', 'nuclear', 'came', 'need', 'one', 'jobs', 'gov', 'make', 'romney', 'sure', 'american', and 'said'.

$$\pi(\mathbf{W}) \propto \exp \left\{ - \sum_{t=1}^T \sum_{i \sim j} \alpha_{ij}^2 \right\} \quad \prod_{t=1}^T \prod_{i \neq j} \mathbb{1}_{[S_{ij}]} \quad \text{Equation 1}$$

# Gerando playlists

## **The Fast and Winding Roads that Lead to The Doors: Generating Heterogeneous Music Playlists**

Marcos A. de Almeida

Department of Computer Science  
UFMG

Belo Horizonte, Brazil

[marcos.almeida@dcc.ufmg.br](mailto:marcos.almeida@dcc.ufmg.br)

Pedro O.S. Vaz de Melo

Department of Computer Science  
UFMG

Belo Horizonte, Brazil

[olmo@dcc.ufmg.br](mailto:olmo@dcc.ufmg.br)

Carolina C. Vieira

Department of Computer Science  
UFMG

Belo Horizonte, Brazil

[carolcoimbra@dcc.ufmg.br](mailto:carolcoimbra@dcc.ufmg.br)

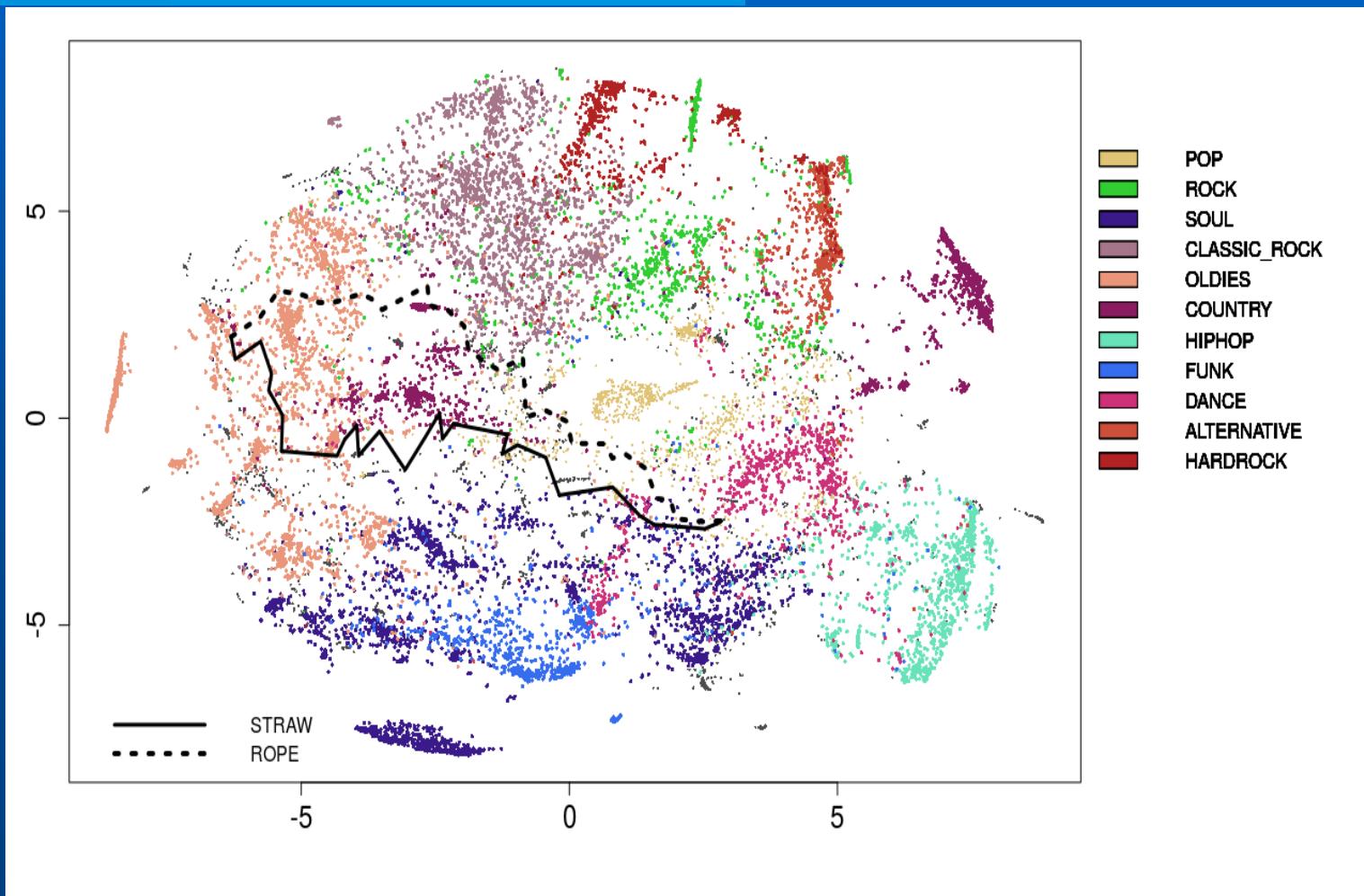
Renato Assunção

Department of Computer Science  
UFMG

Belo Horizonte, Brazil

[assuncao@dcc.ufmg.br](mailto:assuncao@dcc.ufmg.br)

# Uma playlist bem heterogênea



# Gerador de playlists

---

[Gerador de playlists](#) no YouTube

# Outros trabalhos

---

- **Raquel Aoki:**
  - Luck is hard to beat: The difficulty of sports prediction
- **Ramon Lopes:**
  - Efficient Bayesian Methods for Graph-based Recommendation
- **Bruno Barbarioli:**
  - Anomaly Detection Under Cost Constraint
- **Guilherme Oliveira**
  - Subestimação de mortalidade infantil

# Outros trabalhos

- Márcia Barbian
  - Spatial subsemble estimator for large geostatistical data
- Denise Britto
  - Análise de clusters de processos pontuais espaciais do Twitter
- Bráulio Veloso
  - Comportamento estocástico do leitor de jornais online
- Pablo Marcondes:
  - Galo x Raposa: a geografia das torcidas

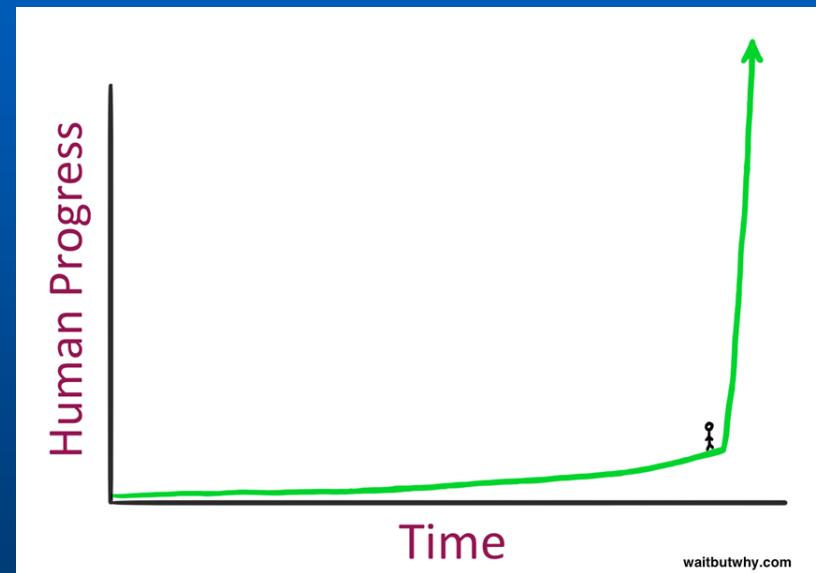
# KDD 2016

---

- 2800 participantes
- 1 milhão de dólares em patrocínio.
- Empresas: facebook, snapchat, AirBnb, Uber, Didi, Baidou, Amazon, etc, etc, etc
- Todas contratando no evento!
- Festas para recrutamento!!!

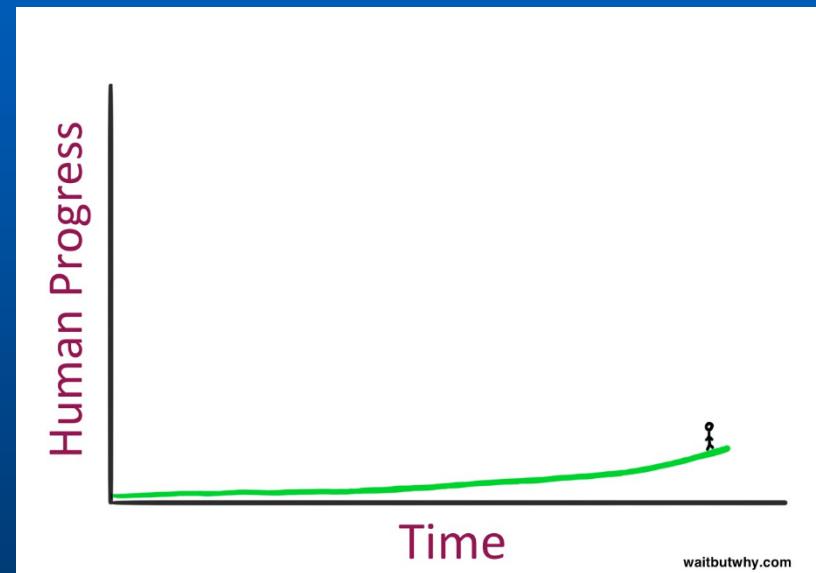
# A beira de uma revolução

- Qual a sensação de estar aqui?
- Parece ser um momento de mudança radical na nossa vida.
- Mas...



# A beira de uma revolução

- Aqui está o que realmente podemos sentir.
- Parece ser um momento como outro qualquer.



# O melhor está por vir

