

Evolução das Linguagens Faladas

IM-UFRJ 13/05/2009

Paulo Murilo Castro de Oliveira

pmco@if.uff.br

Instituto de Física, Universidade Federal Fluminense

Dois livros imperdíveis

Dois livros imperdíveis

Genes, Peuples et Langues

Luigi Luca Cavalli-Sforza

Odile Jacob, Paris (1996)

Dois livros imperdíveis

Genes, Peuples et Langues

Luigi Luca Cavalli-Sforza

Odile Jacob, Paris (1996)

Chi Siamo. La Storia della Diversità Umana

Luigi Luca e Francesco Cavalli-Sforza

Arnoldo Modadori, Milano (1993)

Extraído de um deles

Extraído de um deles

Durante a expansão do homem moderno, novas regiões e novos continentes foram ocupados aos poucos por grupos que, separando-se de sua comunidade original, se instalavam mais longe;

Extraído de um deles

Durante a expansão do homem moderno, novas regiões e novos continentes foram ocupados aos poucos por grupos que, separando-se de sua comunidade original, se instalavam mais longe; depois, outros grupos se desgarravam de novo, continuando em direção a novos locais.

Extraído de um deles

Durante a expansão do homem moderno, novas regiões e novos continentes foram ocupados aos poucos por grupos que, separando-se de sua comunidade original, se instalavam mais longe; depois, outros grupos se desgarravam de novo, continuando em direção a novos locais. Através destas divisões e deslocamentos em série, o homem moderno acabou em regiões tão distantes que manter relações com as origens se tornou difícil ou impossível.

Extraído de um deles

Durante a expansão do homem moderno, novas regiões e novos continentes foram ocupados aos poucos por grupos que, separando-se de sua comunidade original, se instalavam mais longe; depois, outros grupos se desgarravam de novo, continuando em direção a novos locais. Através destas divisões e deslocamentos em série, o homem moderno acabou em regiões tão distantes que manter relações com as origens se tornou difícil ou impossível. O isolamento de vários grupos assim formados gerou dois fenômenos: a formação de diferentes genéticas, e a formação de diferentes linguagens.

Extraído de um deles

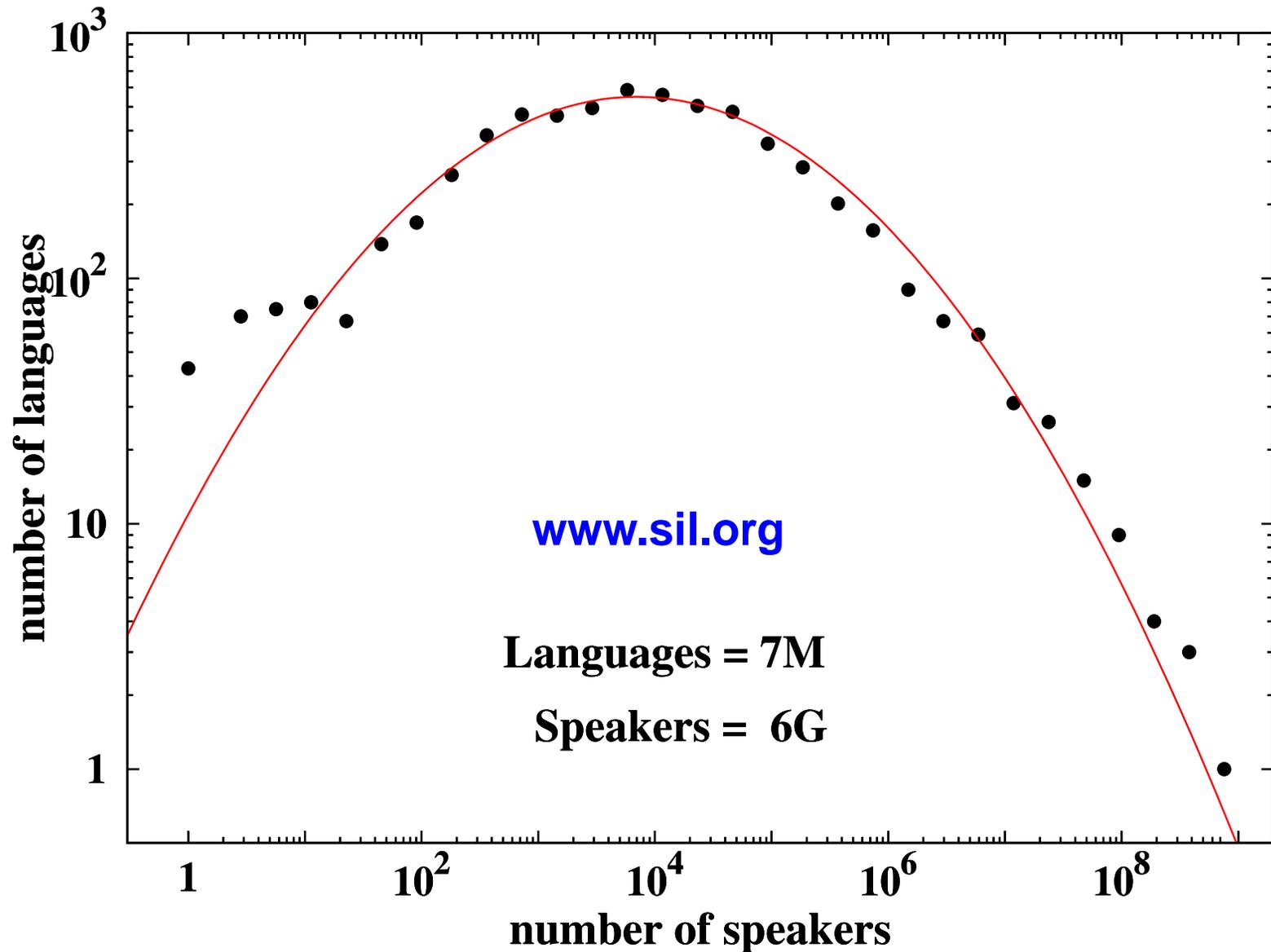
Durante a expansão do homem moderno, novas regiões e novos continentes foram ocupados aos poucos por grupos que, separando-se de sua comunidade original, se instalavam mais longe; depois, outros grupos se desgarravam de novo, continuando em direção a novos locais. Através destas divisões e deslocamentos em série, o homem moderno acabou em regiões tão distantes que manter relações com as origens se tornou difícil ou impossível. O isolamento de vários grupos assim formados gerou dois fenômenos: a formação de diferentes genéticas, e a formação de diferentes linguagens. Estes dois fenômenos seguiram cada um o seu caminho e dinâmica próprios;

Extraído de um deles

Durante a expansão do homem moderno, novas regiões e novos continentes foram ocupados aos poucos por grupos que, separando-se de sua comunidade original, se instalavam mais longe; depois, outros grupos se desgarravam de novo, continuando em direção a novos locais. Através destas divisões e deslocamentos em série, o homem moderno acabou em regiões tão distantes que manter relações com as origens se tornou difícil ou impossível. O isolamento de vários grupos assim formados gerou dois fenômenos: a formação de diferentes genéticas, e a formação de diferentes linguagens. Estes dois fenômenos seguiram cada um o seu caminho e dinâmica próprios; mas a história que se reconstitui a partir das línguas ou a partir da genética é a das separações, que são as causas da diferenciação, e esta história é a mesma.

Barbara Grimes, Ethnologue (2000)

Barbara Grimes, Ethnologue (2000)



Porque log-normal?

Porque log-normal?

Nos sistemas evolucionários, vários tipos de elementos co-evoluem, cada tipo se reproduz a partir de si mesmo

$$\frac{dN_i}{dt} = A N_i + \dots$$

Porque log-normal?

Nos sistemas evolucionários, vários tipos de elementos co-evoluem, cada tipo se reproduz a partir de si mesmo

$$\frac{dN_i}{dt} = A N_i + \dots$$

Mas A não é constante, porque o comportamento de i depende das quantidades N_j dos outros elementos.

Porque log-normal?

Nos sistemas evolucionários, vários tipos de elementos co-evoluem, cada tipo se reproduz a partir de si mesmo

$$\frac{dN_i}{dt} = A N_i + \dots$$

Mas A não é constante, porque o comportamento de i depende das quantidades N_j dos outros elementos.

Então, A flutua no tempo, $A = \bar{A} + \eta(t)$, ou

$$\frac{d \ln(N_i)}{dt} = \bar{A} + \eta(t) + \dots$$

Porque log-normal?

Nos sistemas evolucionários, vários tipos de elementos co-evoluem, cada tipo se reproduz a partir de si mesmo

$$\frac{dN_i}{dt} = A N_i + \dots$$

Mas A não é constante, porque o comportamento de i depende das quantidades N_j dos outros elementos.

Então, A flutua no tempo, $A = \bar{A} + \eta(t)$, ou

$$\frac{d \ln(N_i)}{dt} = \bar{A} + \eta(t) + \dots$$

Esta equação é do tipo Langevin, cujo resultado é uma distribuição Gaussiana para $\ln(N_i)$.

Famílias de línguas

Famílias de línguas

Os biólogos classificam os seres vivos em hierarquias:

Famílias de línguas

Os biólogos classificam os seres vivos em hierarquias:

- **Espécies** = grupos de indivíduos semelhantes (no caso sexuado, o critério consagrado é a possibilidade de reprodução de filhos férteis);

Famílias de línguas

Os biólogos classificam os seres vivos em hierarquias:

- **Espécies** = grupos de indivíduos semelhantes (no caso sexuado, o critério consagrado é a possibilidade de reprodução de filhos férteis);
- **Gêneros** = grupos de **espécies** semelhantes;

Famílias de línguas

Os biólogos classificam os seres vivos em hierarquias:

- **Espécies** = grupos de indivíduos semelhantes (no caso sexuado, o critério consagrado é a possibilidade de reprodução de filhos férteis);
- **Gêneros** = grupos de **espécies** semelhantes;
- **Famílias** = grupos de **gêneros** semelhantes;

Famílias de línguas

Os biólogos classificam os seres vivos em hierarquias:

- **Espécies** = grupos de indivíduos semelhantes (no caso sexuado, o critério consagrado é a possibilidade de reprodução de filhos férteis);
- **Gêneros** = grupos de **espécies** semelhantes;
- **Famílias** = grupos de **gêneros** semelhantes;
- Ordens Reinos (vegetal, animal, etc).

Famílias de línguas

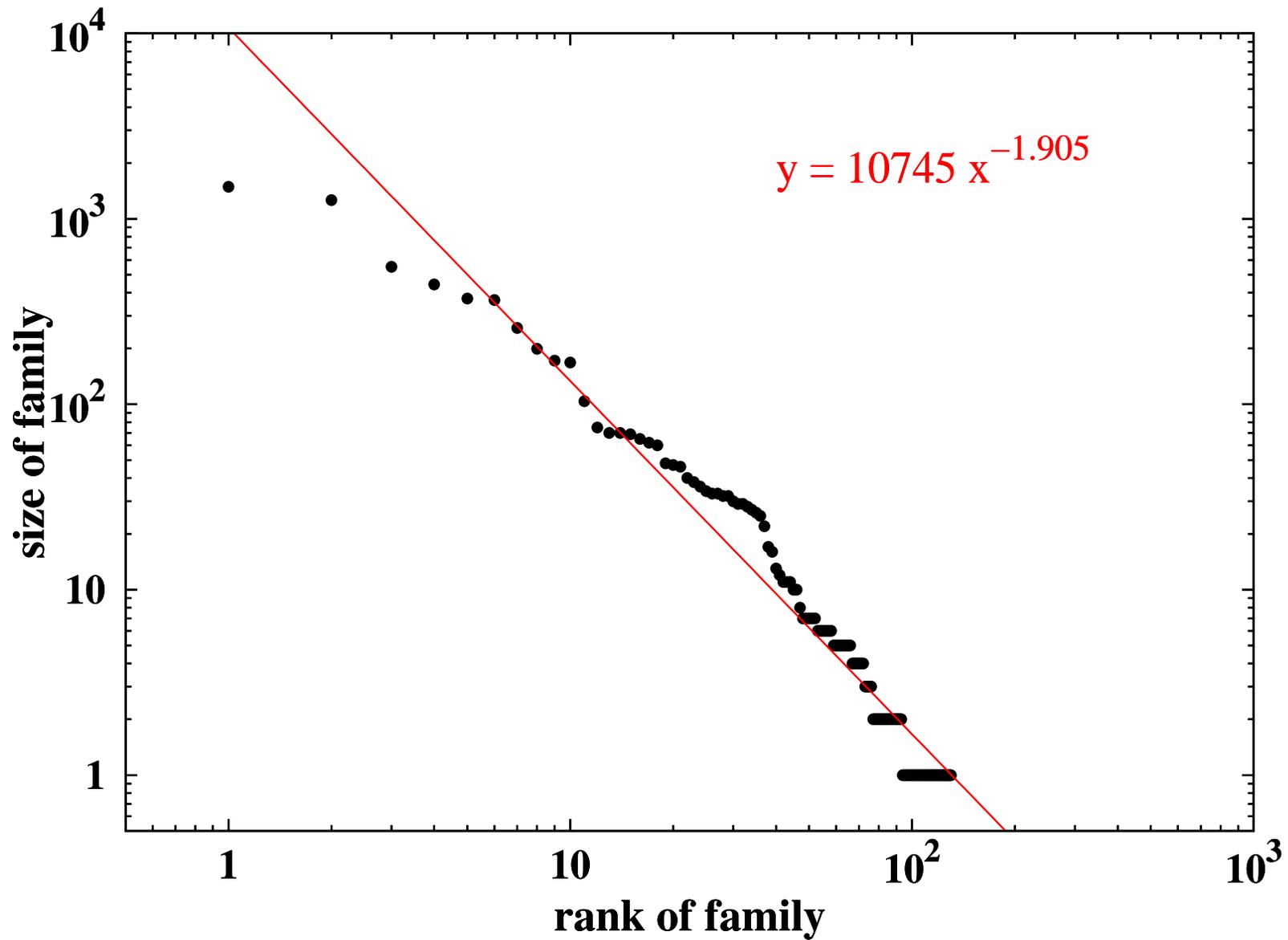
Os biólogos classificam os seres vivos em hierarquias:

- **Espécies** = grupos de indivíduos semelhantes (no caso sexuado, o critério consagrado é a possibilidade de reprodução de filhos férteis);
- **Gêneros** = grupos de **espécies** semelhantes;
- **Famílias** = grupos de **gêneros** semelhantes;
- Ordens Reinos (vegetal, animal, etc).

Meu mais novo colaborador, o linguista Søren Wichmann (Max Planck Inst. for Evolutionary Anthropology, Leipzig) usou apenas duas hierarquias,

línguas e **famílias de línguas**.

Søren Wichmann *J. Linguistics* **41**, 117 (2005)



Nosso problema

Nosso problema

- Modelar a evolução das línguas a partir da língua mãe, as demais são criadas dinamicamente.

Nosso problema

- Modelar a evolução das línguas a partir da língua mãe, as demais são criadas dinamicamente.
- Comparar a situação num dado instante da evolução com dois dados reais:

A distribuição por número de falantes (log-normal com desvio na região de pequenas populações);

A distribuição de famílias (lei de potência com expoente ≈ -1.5).

Nosso problema

- Modelar a evolução das línguas a partir da língua mãe, as demais são criadas dinamicamente.

- Comparar a situação num dado instante da evolução com dois dados reais:

A distribuição por número de falantes (log-normal com desvio na região de pequenas populações);

A distribuição de famílias (lei de potência com expoente ≈ -1.5).

- Os linguistas tratam o problema da frente para trás: comparam as línguas atuais, definem as famílias e inferem as distribuições. Não há dinâmica.

Nosso problema

- Modelar a evolução das línguas a partir da língua mãe, as demais são criadas dinamicamente.

- Comparar a situação num dado instante da evolução com dois dados reais:

A distribuição por número de falantes (log-normal com desvio na região de pequenas populações);

A distribuição de famílias (lei de potência com expoente ≈ -1.5).

- Os linguistas tratam o problema da frente para trás: comparam as línguas atuais, definem as famílias e inferem as distribuições. Não há dinâmica.
- Nós fazemos na ordem histórica: populações, línguas e famílias se definem ao longo da evolução dinâmica.

Publicações anteriores

Publicações anteriores

- V.M. de Oliveira, M.A.F. Gomes e I.R. Tsang, *Physica A361*, 361 (2006)

V.M. de Oliveira, P.R.A. Campos, M.A.F. Gomes e I.R. Tsang, *Physica A368*, 257 (2006).

Publicações anteriores

- V.M. de Oliveira, M.A.F. Gomes e I.R. Tsang, *Physica* **A361**, 361 (2006)

V.M. de Oliveira, P.R.A. Campos, M.A.F. Gomes e I.R. Tsang, *Physica* **A368**, 257 (2006).

- P.M.C. de Oliveira, D. Stauffer, F.W.S. Lima, A.O. Sousa, C. Schulze e S.M. Moss de Oliveira, *Physica* **A376**, 609 (2007) (www.arXiv.org PHYSICS/0608204)

P.M.C. de Oliveira, D. Stauffer, S. Wichmann e S. Moss de Oliveira, *J. of Linguistics* **44**, 659 (2008) (www.arXiv.org PHYSICS/0709.0868)

Publicações anteriores

- V.M. de Oliveira, M.A.F. Gomes e I.R. Tsang, *Physica A361*, 361 (2006)

V.M. de Oliveira, P.R.A. Campos, M.A.F. Gomes e I.R. Tsang, *Physica A368*, 257 (2006).

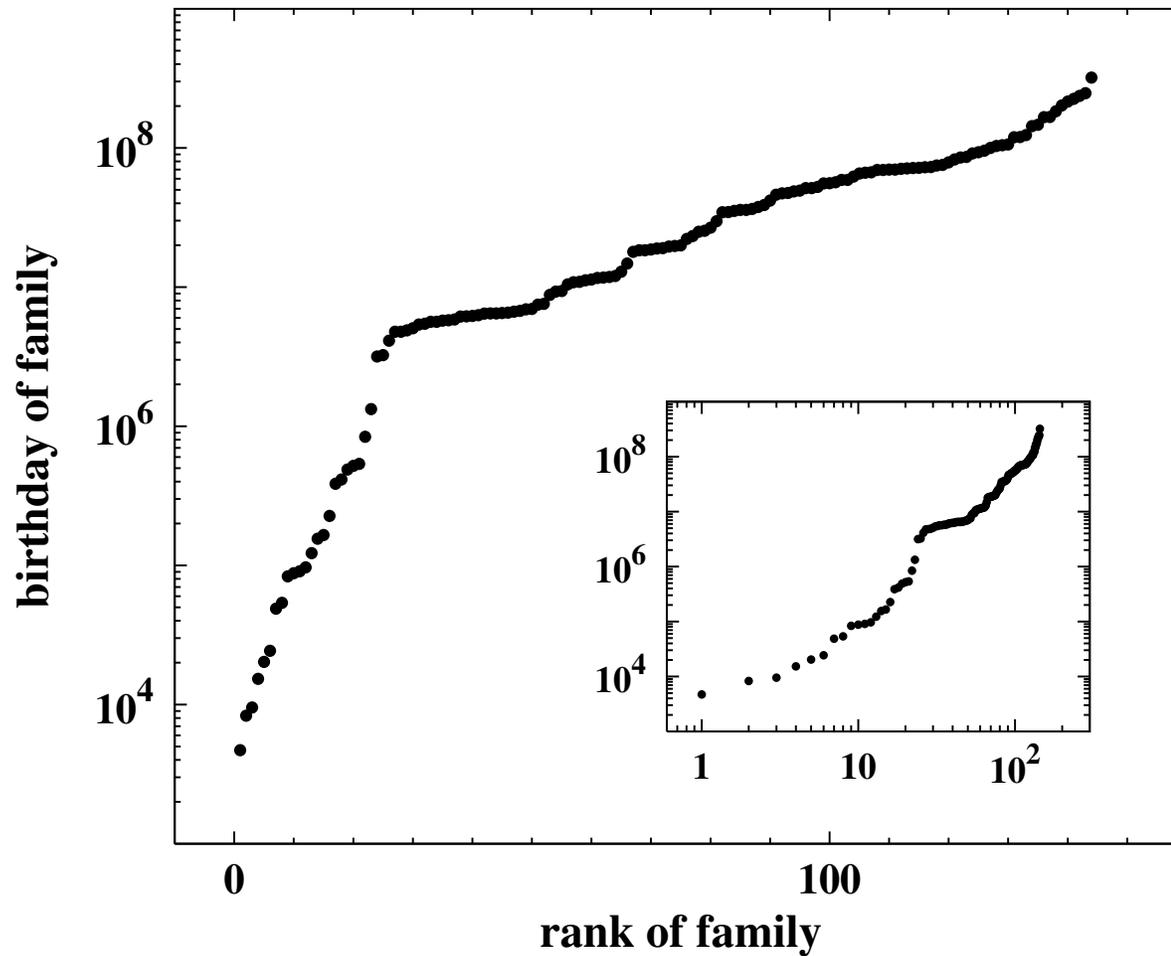
- P.M.C. de Oliveira, D. Stauffer, F.W.S. Lima, A.O. Sousa, C. Schulze e S.M. Moss de Oliveira, *Physica A376*, 609 (2007) (www.arXiv.org PHYSICS/0608204)

P.M.C. de Oliveira, D. Stauffer, S. Wichmann e S. Moss de Oliveira, *J. of Linguistics* 44, 659 (2008) (www.arXiv.org PHYSICS/0709.0868)

Todos baseados na ocupação geográfica gradativa de um território inicialmente vazio.

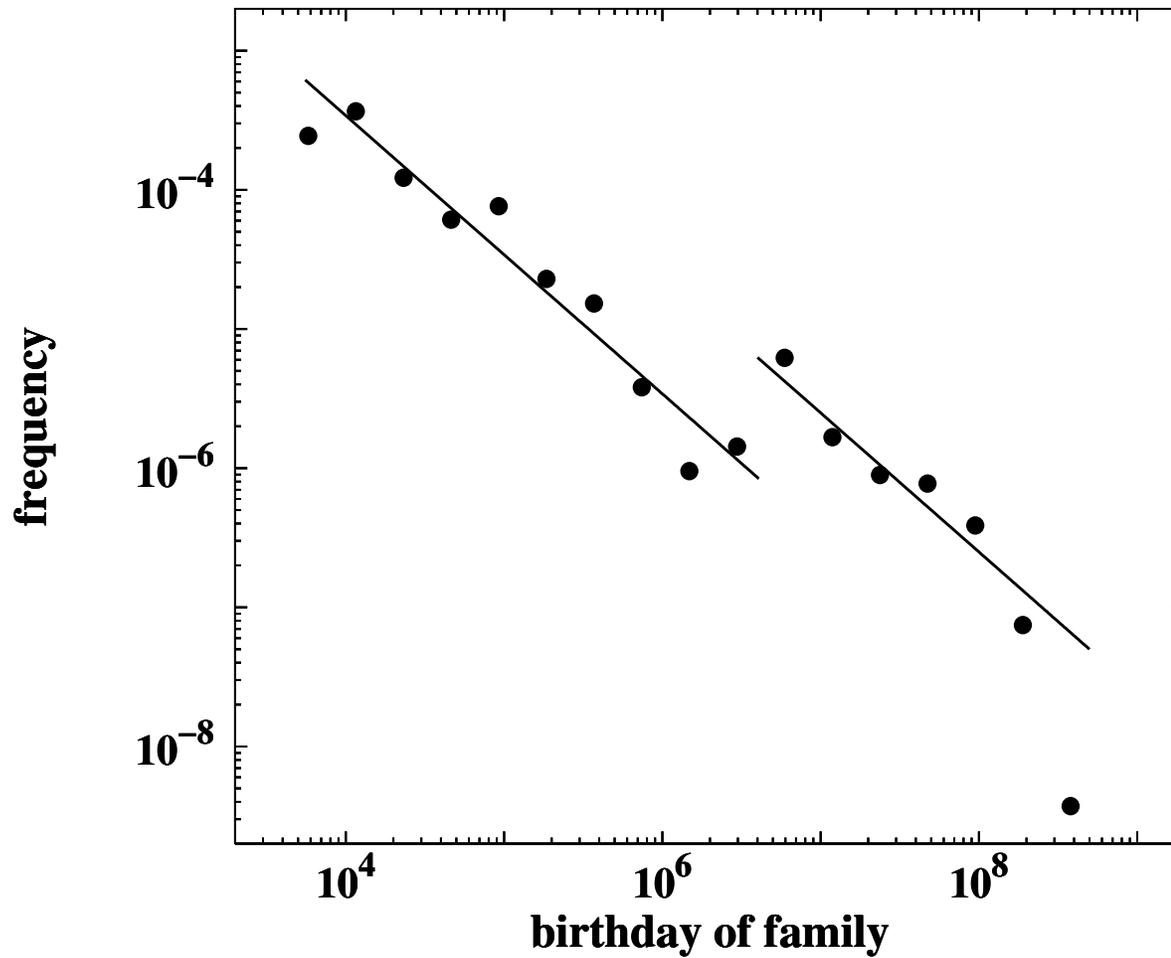
curiosidade: “elbow” nas famílias

curiosidade: “elbow” nas famílias



Distribuição de famílias

Distribuição de famílias



Trabalho atual

Trabalho atual

- V. Schwämmle e P.M.C. de Oliveira, *Physica A* (2009) (www.arXiv.org/abs/0811.2301)

S. Wichmann, P.M.C. de Oliveira, V. Vilupillai, A. Müller, D. Bakker e A. Grant, rejeitado na *Nature*, destino desconhecido.

Trabalho atual

- V. Schwämmle e P.M.C. de Oliveira, *Physica A* (2009) (www.arXiv.org/abs/0811.2301)
S. Wichmann, P.M.C. de Oliveira, V. Vilupillai, A. Müller, D. Bakker e A. Grant, rejeitado na *Nature*, destino desconhecido.
- Não há mais geografia, só a árvore histórica das bifurcações de línguas e famílias. As línguas se modificam a uma taxa constante. A cada passo uma língua se bifurca em duas com certa probabilidade.

Trabalho atual

- V. Schwämmle e P.M.C. de Oliveira, *Physica A* (2009) (www.arXiv.org/abs/0811.2301)
S. Wichmann, P.M.C. de Oliveira, V. Vilupillai, A. Müller, D. Bakker e A. Grant, rejeitado na *Nature*, destino desconhecido.
- Não há mais geografia, só a árvore histórica das bifurcações de línguas e famílias. As línguas se modificam a uma taxa constante. A cada passo uma língua se bifurca em duas com certa probabilidade.
- Apenas dois ingredientes são mantidos:
 - 1) redundância (estrutura da língua num espaço finito);
 - 2) definição de família (fundador).

explicação a seguir ...

Redundância

Redundância

- A estrutura de cada língua é resumida num conjunto de B características binárias (uma cadeia com B bits).

Redundância

- A estrutura de cada língua é resumida num conjunto de B características binárias (uma cadeia com B bits).
- A cada passo de tempo, um bit aleatório de cada língua é invertido, e sua população cresce por um fator constante (taxas constantes de modificação das línguas e de crescimento das populações).

Redundância

- A estrutura de cada língua é resumida num conjunto de B características binárias (uma cadeia com B bits).
- A cada passo de tempo, um bit aleatório de cada língua é invertido, e sua população cresce por um fator constante (taxas constantes de modificação das línguas e de crescimento das populações).
- A cada passo de tempo, com uma dada probabilidade, cada língua se bifurca em duas, a segunda difere da primeira por um bit aleatório invertido. Nesse caso, a população é dividida entre ambas aleatoriamente.

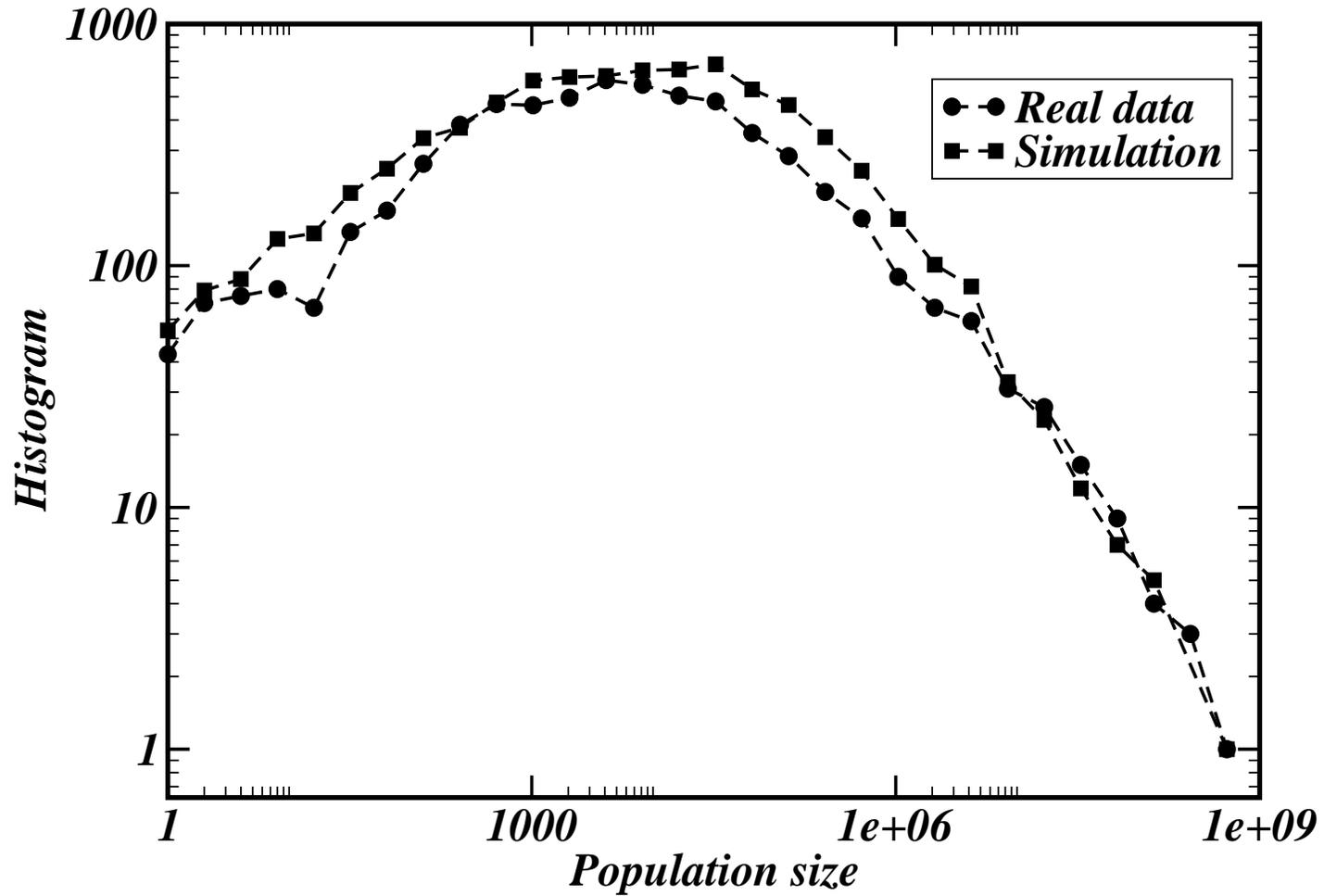
Redundância

- A estrutura de cada língua é resumida num conjunto de B características binárias (uma cadeia com B bits).
- A cada passo de tempo, um bit aleatório de cada língua é invertido, e sua população cresce por um fator constante (taxas constantes de modificação das línguas e de crescimento das populações).
- A cada passo de tempo, com uma dada probabilidade, cada língua se bifurca em duas, a segunda difere da primeira por um bit aleatório invertido. Nesse caso, a população é dividida entre ambas aleatoriamente.

Re-invenção de línguas por diferentes caminhos históricos. Reproduz a anomalia observada na curva log-normal das populações. Corresponde à resolução finita dos linguistas.

População

População



Fundação de famílias

Fundação de famílias

- A língua mãe funda a família original.

Fundação de famílias

- A língua mãe funda a família original.
- Qualquer língua pertencente à família original funda sua própria família com probabilidade $1/2$ (ou outro valor qualquer).

Fundação de famílias

- A língua mãe funda a família original.
- Qualquer língua pertencente à família original funda sua própria família com probabilidade $1/2$ (ou outro valor qualquer).
- Línguas pertencentes a outras famílias que não a original não podem fundar suas próprias famílias (seria gênero, e não família).

Fundação de famílias

- A língua mãe funda a família original.
- Qualquer língua pertencente à família original funda sua própria família com probabilidade $1/2$ (ou outro valor qualquer).
- Línguas pertencentes a outras famílias que não a original não podem fundar suas próprias famílias (seria gênero, e não família).

Respeita a hierarquia de formação de famílias. Na primeira geração a probabilidade de fundação é $1/2$. Na segunda, é $1/4$ em média, na terceira $1/8$ em média, etc.

Fundação de famílias

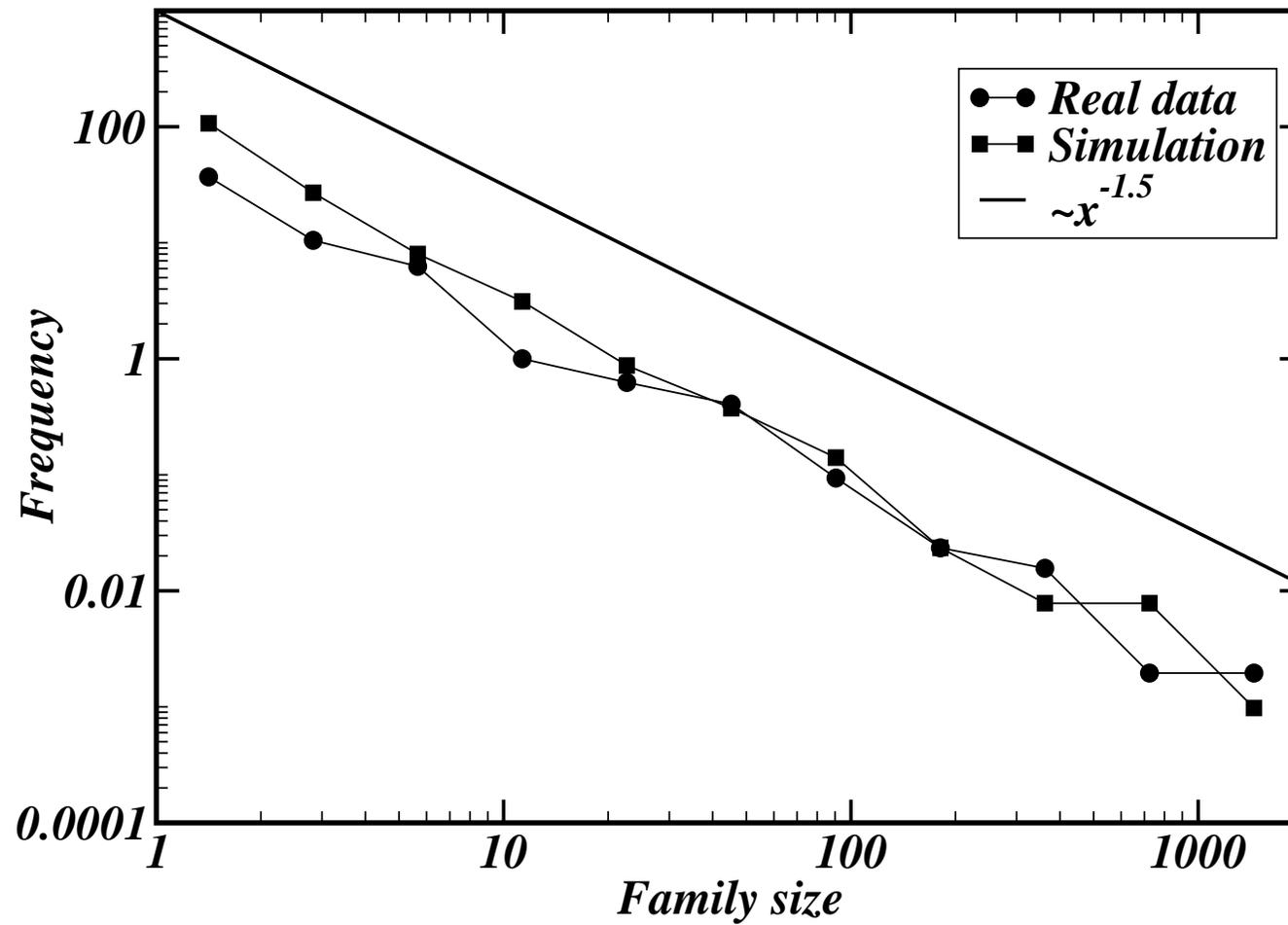
- A língua mãe funda a família original.
- Qualquer língua pertencente à família original funda sua própria família com probabilidade $1/2$ (ou outro valor qualquer).
- Línguas pertencentes a outras famílias que não a original não podem fundar suas próprias famílias (seria gênero, e não família).

Respeita a hierarquia de formação de famílias. Na primeira geração a probabilidade de fundação é $1/2$. Na segunda, é $1/4$ em média, na terceira $1/8$ em média, etc.

Introduz memória de longo prazo: para que uma língua possa fundar nova família é preciso conhecer todos os seus ancestrais até a língua mãe.

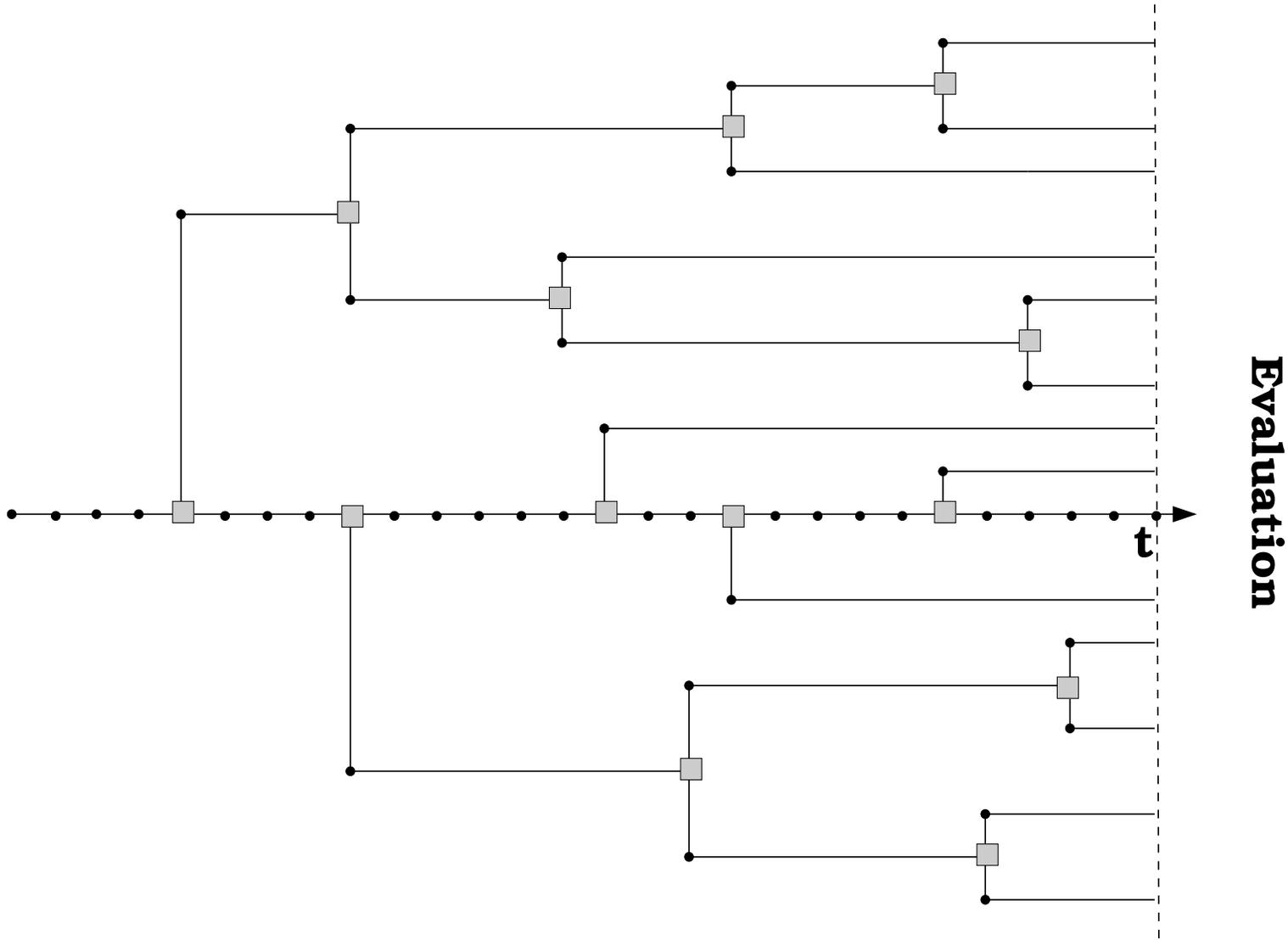
Famílias

Famílias



Árvore esquemática

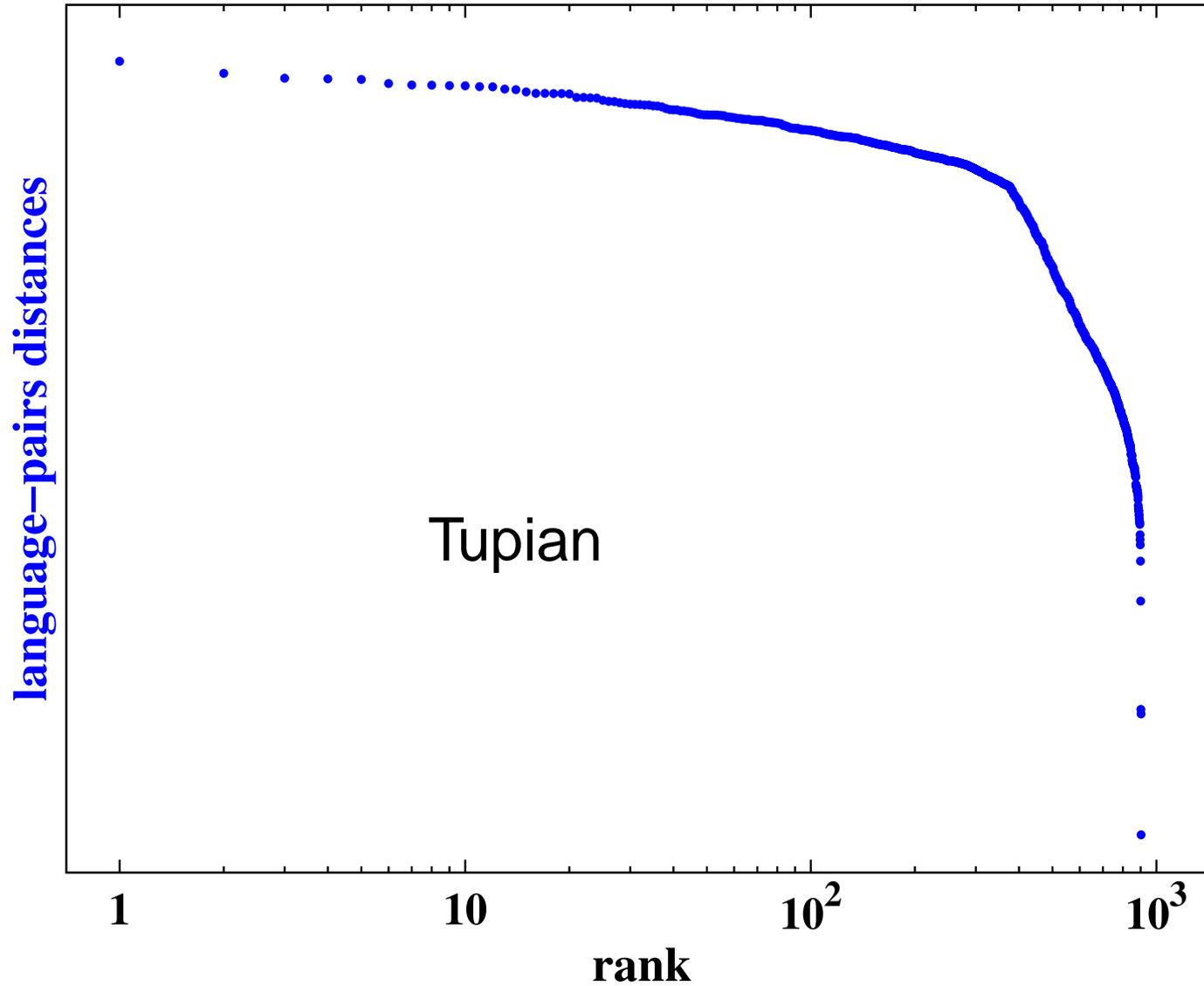
Árvore esquemática



Evaluation

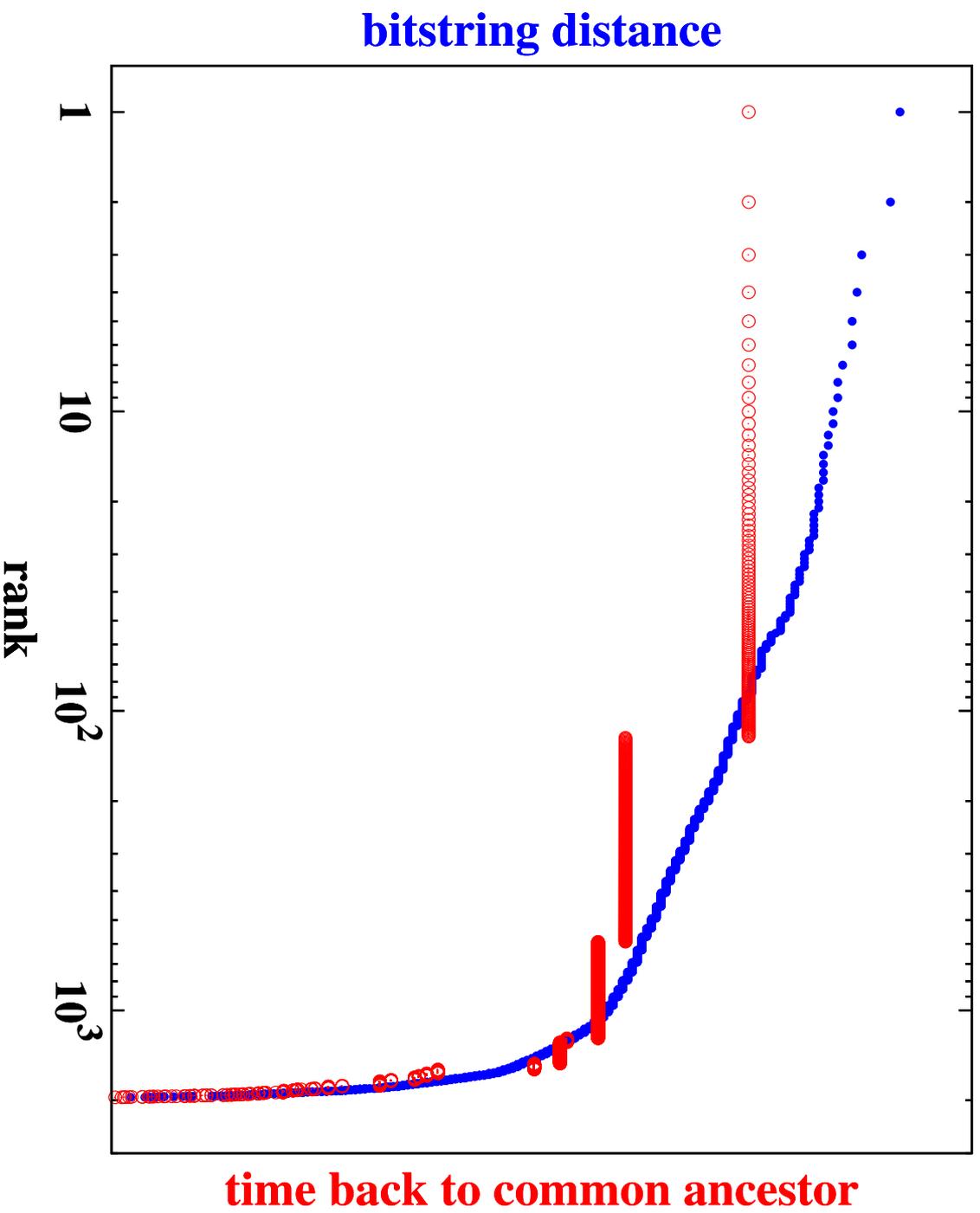
aparece o “elbow” (realidade)

aparece o “elbow” (realidade)



explicação: Gêneros (modelo)

explicação: Gêneros (modelo)



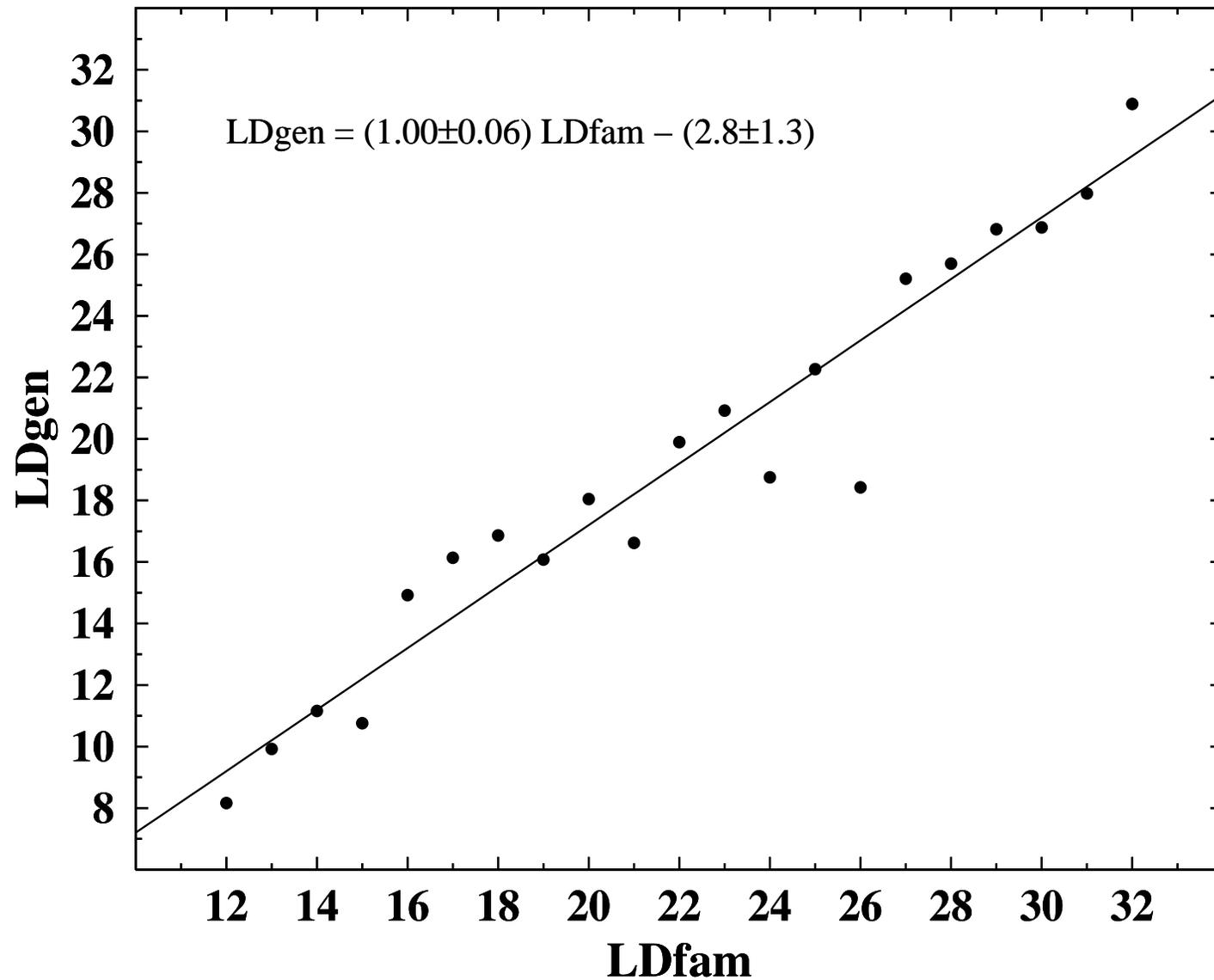
bitstring distance

time back to common ancestor

rank

Dados reais para 21 famílias

Dados reais para 21 famílias



Propaganda

Propaganda

Computing Boolean Statistical Models

P.M.C. de Oliveira

World Scientific, London New York Singapore, ISBN 981-02-0238-5 (1991)

Evolution, Money, War and Computers

Non-Traditional Applications of Computational Statistical Physics

S. Moss de Oliveira, P.M.C. de Oliveira and D. Stauffer

Teubner, Stuttgart Leipzig, ISBN 3-519-00279-5 (1999)

Biology, Sociology, Geology by Computational Physicists

D. Stauffer, S. Moss de Oliveira, P.M.C. de Oliveira and Jorge Sá Martins

Elsevier, Amsterdam, ISBN 0-444-52146-1 (2006)

O grupo da UFF

O grupo da UFF

- Suzana Moss
- Jorge Sá Martins
- Thadeu Penna
- Marcio Argollo
- Jürgen Stilck
- Dietrich Stauffer (penetra de Köln)
- PMCO

O grupo da UFF

- Suzana Moss
Jorge Sá Martins
Thadeu Penna
Marcio Argollo
Jürgen Stilck
Dietrich Stauffer (penetra de Köln)
PMCO
- Veit Schwämmle (Stuttgart)
Eduardo Aguilar
Klauko Mota
Cinthya Chianca
Nuno Crokidakis Peregrino
C. Eduardo Galhardo
Orahcio Felicio de Sousa

O grupo da UFF

- Suzana Moss
Jorge Sá Martins
Thadeu Penna
Marcio Argollo
Jürgen Stilck
Dietrich Stauffer (penetra de Köln)
PMCO
- Veit Schwämmle (Stuttgart)
Eduardo Aguilar
Klauko Mota
Cinthya Chianca
Nuno Crokidakis Peregrino
C. Eduardo Galhardo
Orahcio Felicio de Sousa