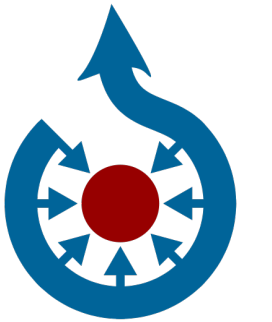
Análisis de Comunidades Virtuales con R

Felipe Ortega
GSyC/Libresoft
Universidad Rey Juan Carlos

twitter/identi.ca
@jfelipe

Comunidades virtuales

- Grupos de personas que interactúan y colaboran usando sistemas de información y comunicación (normalmente basados en Internet).
  - Blogs.
  - Microblogs.
  - Wikis.
  - Redes sociales.
  - Software libre.
  - Foros...



¿Por qué R?

- Software libre.
- Comunidad amplia y muy activa.
- Extensa documentación.
- Bibliotecas disponibles.
- Adecuado para automatizar.
- Buen rendimiento.
- Soporte para reproducibilidad.

Caso de estudio: Wikipedia

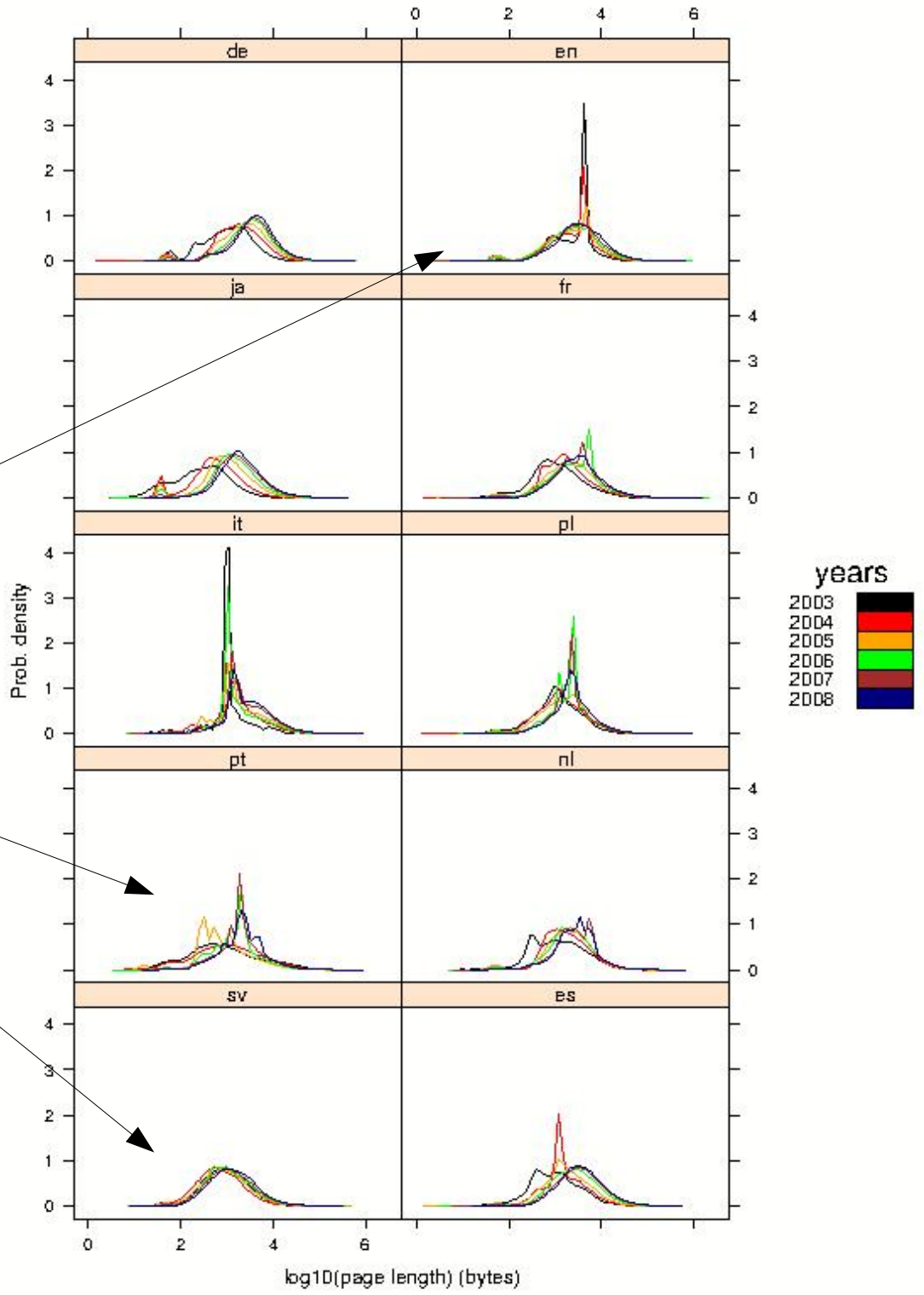
- Análisis cuantitativo macroscópico.
  - "*Wikipedia: A quantitative analysis*" (2009).
  - Comparativa de las 10 mayores Wikipedias.
  - http://felipeortega.net/sites/default/files/thesis-jfelipe.pdf
- Gran cantidad de datos para analizar.
  - +300 millones de cambios, sólo para la Wikipedia en inglés.
  - Casi 10 millones de artículos enciclopédicos.
- WikiXRay: Python + MySQL + R.
  - http://meta.wikimedia.org/wiki/WikiXRay

Condiciones de estudio

- Similares a muchos otros proyectos *open data*.
- Análisis del conjunto completo de datos disponibles.
  - No se realiza muestreo.
- Investigación reproducible
  - Y replicable.
- Automatización.

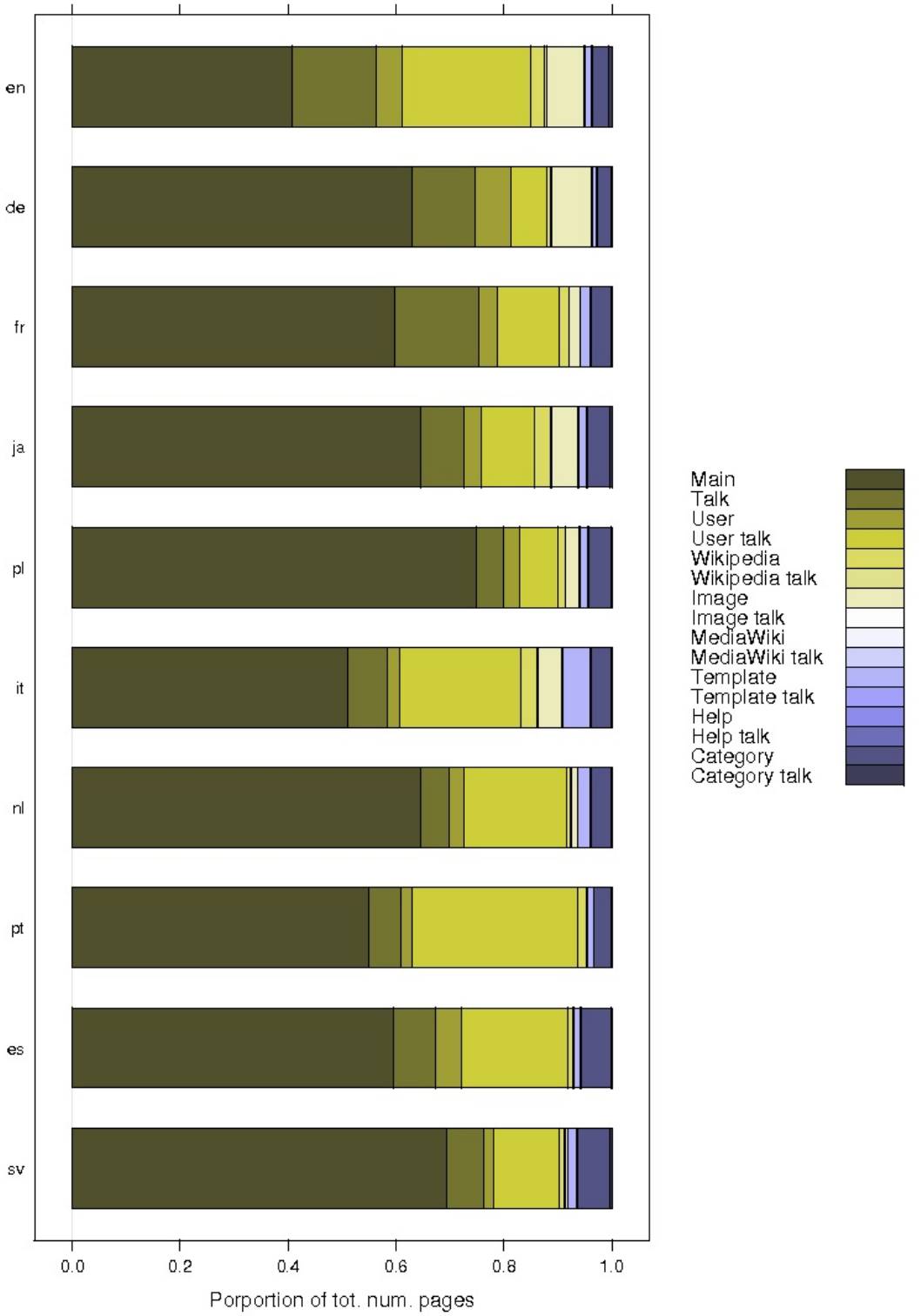
Contenidos

Diferentes
Patrones
evolutivos



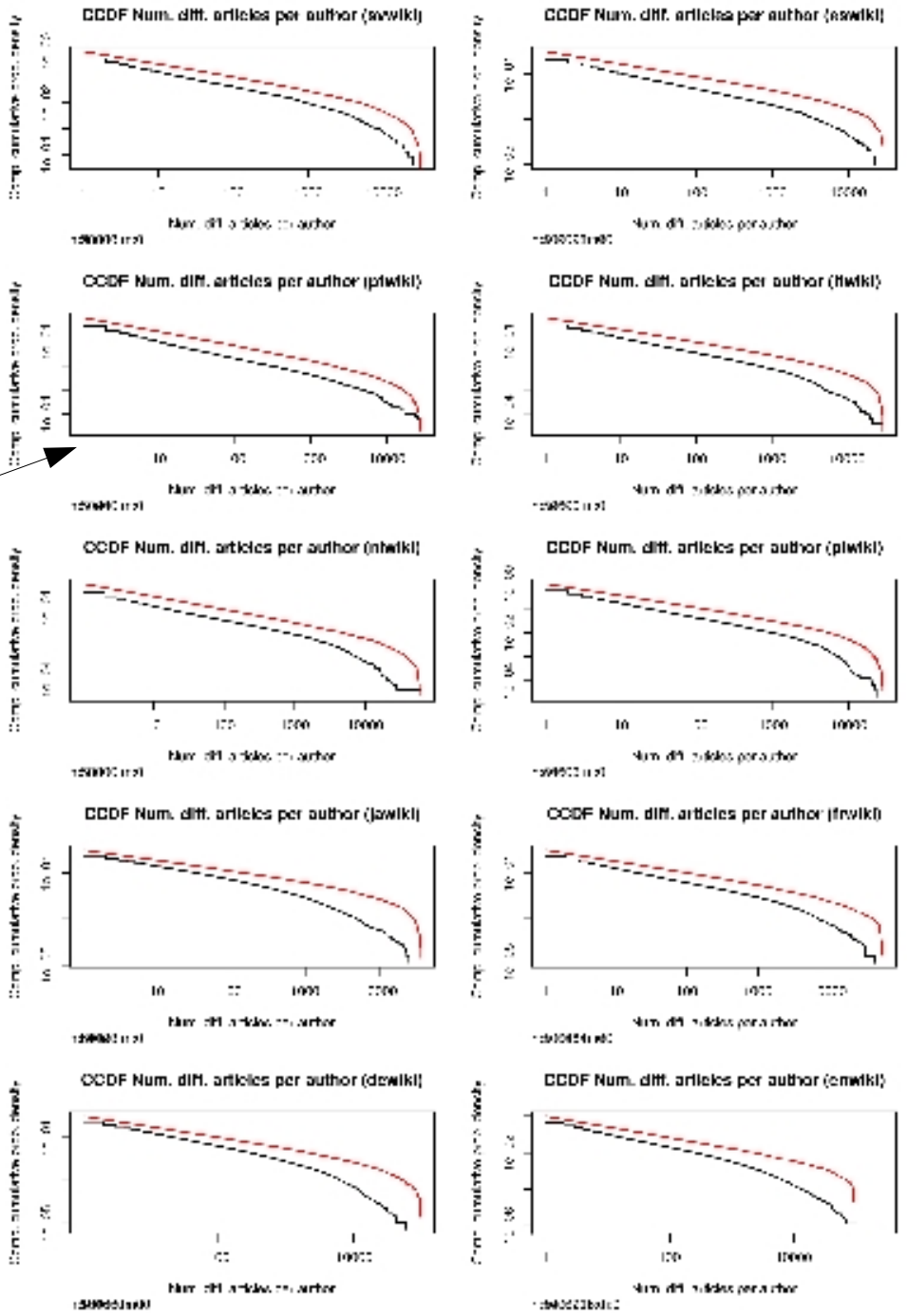
Interacción

Distribución del trabajo
e interés en diferentes
vertientes de la colaboración



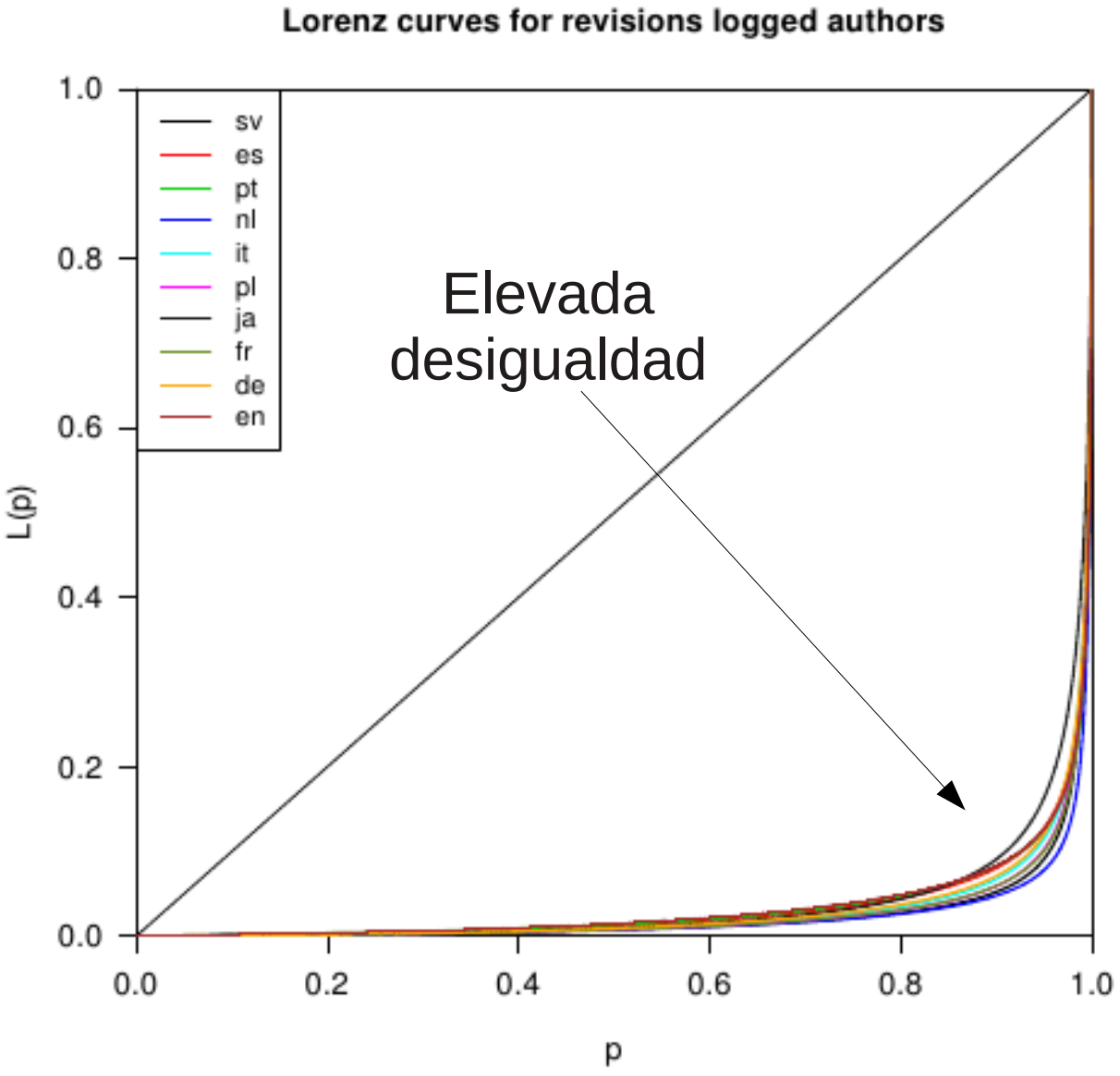
Editores

Distribución de Pareto truncada

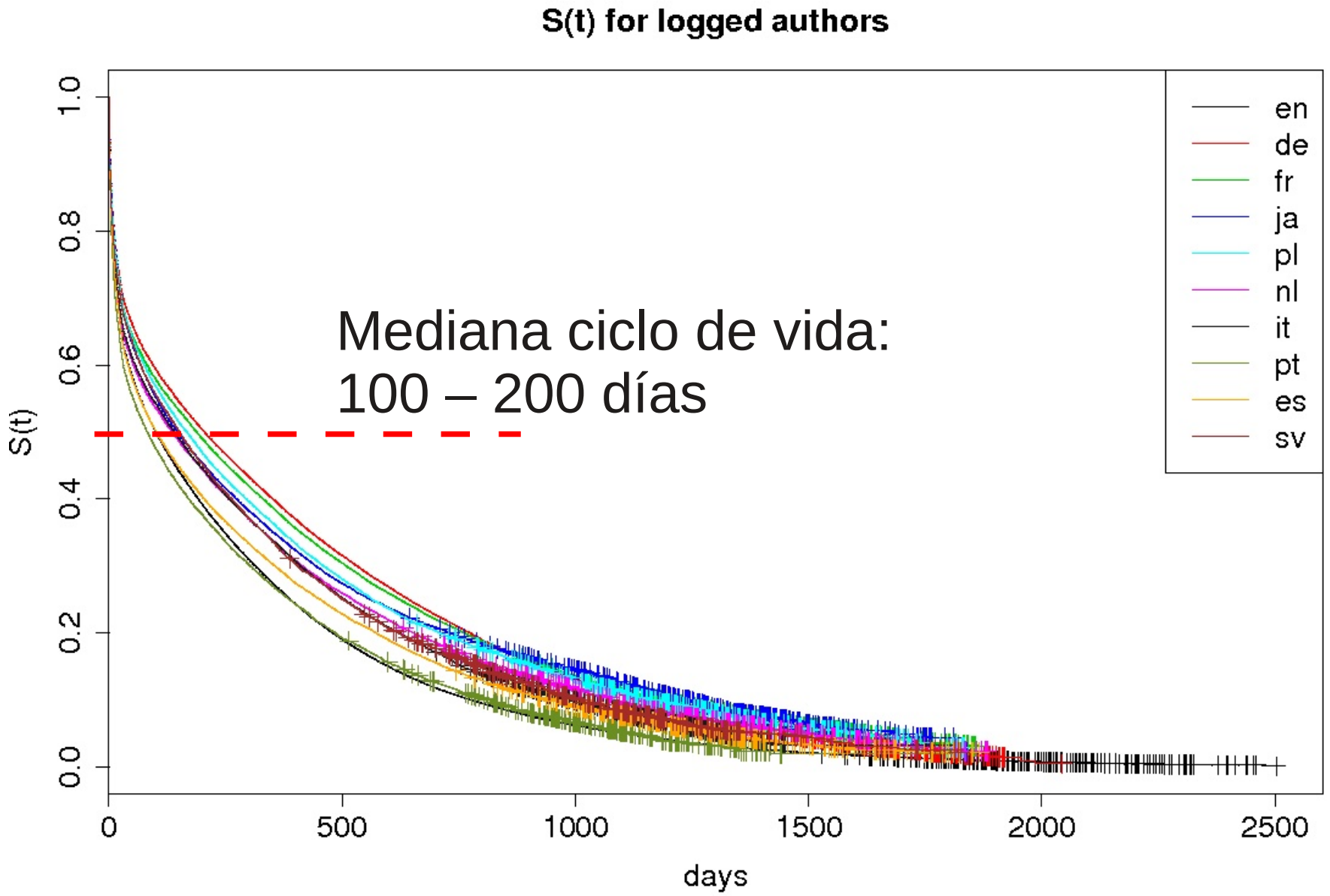


Reparto de esfuerzo

Rango de Coeficientes de Gini:
92% - 95%



Ciclo de vida



Conclusiones

- Motivos para utilizar R.
  - Potencia, flexibilidad, rendimiento.
  - Fácil automatización del proceso de análisis.
  - Favorecer reproducibilidad de estudios.
  - Software libre: incorporación de mejoras y aportaciones.
- Próximos pasos.
  - Análisis de redes sociales.
  - Análisis longitudinal: predicción y *clustering*.

Créditos

- Portada:
  - Water drops on spider web, por Tttrung [Public domain], via Wikimedia Commons.
- Página 3:
  - Twitter-logo, via Wikimedia Commons.
  - Wikimedia Commons logo, (c) Wikimedia Foundationi, via Wikimedia Commons.
  - Wikipedia-logo-es, por Nohat (c) Wikimedia Foundationi, via Wikimedia Commons.
  - Flickr-wordmark, por Brands of the Worldi, via Wikimedia Commons.
  - Debian open logo, por Debiani, via Wikimedia Commons.
  - KDE logo, por KDEi, via Wikimedia Commons.
  - Gnome logo, por David Vignoni, via Wikimedia Commons.
- Página 4:
  - R logo, por R Foundation, via Wikimedia Commons.
- Página 6:
  - R-square-logo, por Rsquaremca (Own work) [Dominio Público], via Wikimedia Commons (modificada).