

Los números de Wikipedia

MediaLab Prado, Madrid

2 de Febrero de 2010.



Felipe Ortega, Libresoft (URJC)

jfelipe@libresoft.es

Contenidos

1. Introducción.
2. Artículos.
3. Comunidad
4. Continuidad.



CC-BY : steffenz (Flickr)

1. INTRODUCCIÓN

Por qué Wikipedia?

1. Introducción

“ El problema con Wikipedia es que sólo funciona en la práctica. En teoría, nunca podría funcionar”

Mikka Ryokas, citado en artículo del NY Times

Trivial de Wikipedia

1. Initial problem and goals

1. ¿Cuántas personas han escrito en las 10 mayores versiones de Wikipedia?
 - **3,203,546** editores registrados.
2. ¿Número total de ediciones en la versión inglesa de Wikipedia?
 - **296.387.800** (10 de octubre de 2009).
3. ¿Tiempo de los editores en Wikipedia?
 - **346.9 días** (media); **141.6 días** (mediana)

Principales cuestiones

1. Introducción

- Estudios comparativos.
- Perspectiva global.
 - Contenido y comunidad
- Focos de interés.
 - Calidad.
 - Evolución.
 - Continuidad del proyecto.



Respuestas

1. Introducción

- Tesis doctoral sobre Wikipedia (abril, 2009).
 - *“Wikipedia: A quantitative Analysis”*.



CC-BY : William Hoiles

Algunos términos básicos

1. Introducción

- **Página.**
 - Cualquier página wiki en una versión de Wikipedia.
- **Revisión (a.k.a. Contribución, edición).**
 - Un cambio en una página de Wikipedia.
 - **Revert:** Deshacer revisión anterior.
- **Editor.**
 - Persona que cambia una página.
 - Dos tipos.
 - **Registrados vs. anónimos.**



Organización y contenidos

1. Introducción

- **Diferentes tipos de páginas.**
 - Artículos enciclopédicos.
 - Páginas de discusión asociadas a artículos.
 - Páginas de usuario.
- **Proyecto de evaluación de calidad.**
 - Artículos destacados (*Featured Articles*)
 - Revisados en detalle por la comunidad.
 - Promocionados tras un proceso de votación

WikiXRay

2. Methodology and tools

- Herramienta de análisis para el estudio de cualquier versión de Wikipedia



WikiXRay on-line

1. Introducción

Try Beta [Log in / create account](#)

[content page](#) [discussion](#) [edit](#) [history](#)

Please [comment](#) on the proposal for [global sysops](#).

WikiXRay

From Meta, a Wikimedia project coordination wiki

The main goal of this project is to develop a robust and extensible software tool for an in-depth quantitative analysis of the whole Wikipedia project. This project is currently developed by José Felipe Ortega ([GlimmerPhoenix](#)) at the [Librosoft Group](#), at [Universidad Rey Juan Carlos](#).

Currently, this tool includes a set of [Python](#) and [GNU R](#) scripts to obtain statistics, graphics and quantitative results for any Wikipedia language version. Current functionality includes:

- Downloading the 7zip database dump of the target language version.
- Construction and decompression of the database dump in a local storage media.
- Creating additional database tables with useful statistics and quantitative information.
- Generating graphics and data files with quantitative results, adequately organized in a per-language directory substructure.

Some of these capabilities still require manual insertion of parameters in a common configuration file, though also some of them work automatically.

The source code is publicly available under the [GNU GPL license](#), and could be found in the official WikiXRay project page [WikiXRay Project Page](#) at [BerliOS](#). The home page is currently under development, so please wait a few moments for the project summary page to load.

Please note that this software is still in a very early development stage (pre-alpha level). Any useful contributions will be of course welcomed (first contact with the project administrators).

Following, we summarize some of the most relevant results we have obtained so far. These results comes from a quantitative analysis focused on the top 10 Wikipedias (attending to their total number of articles).

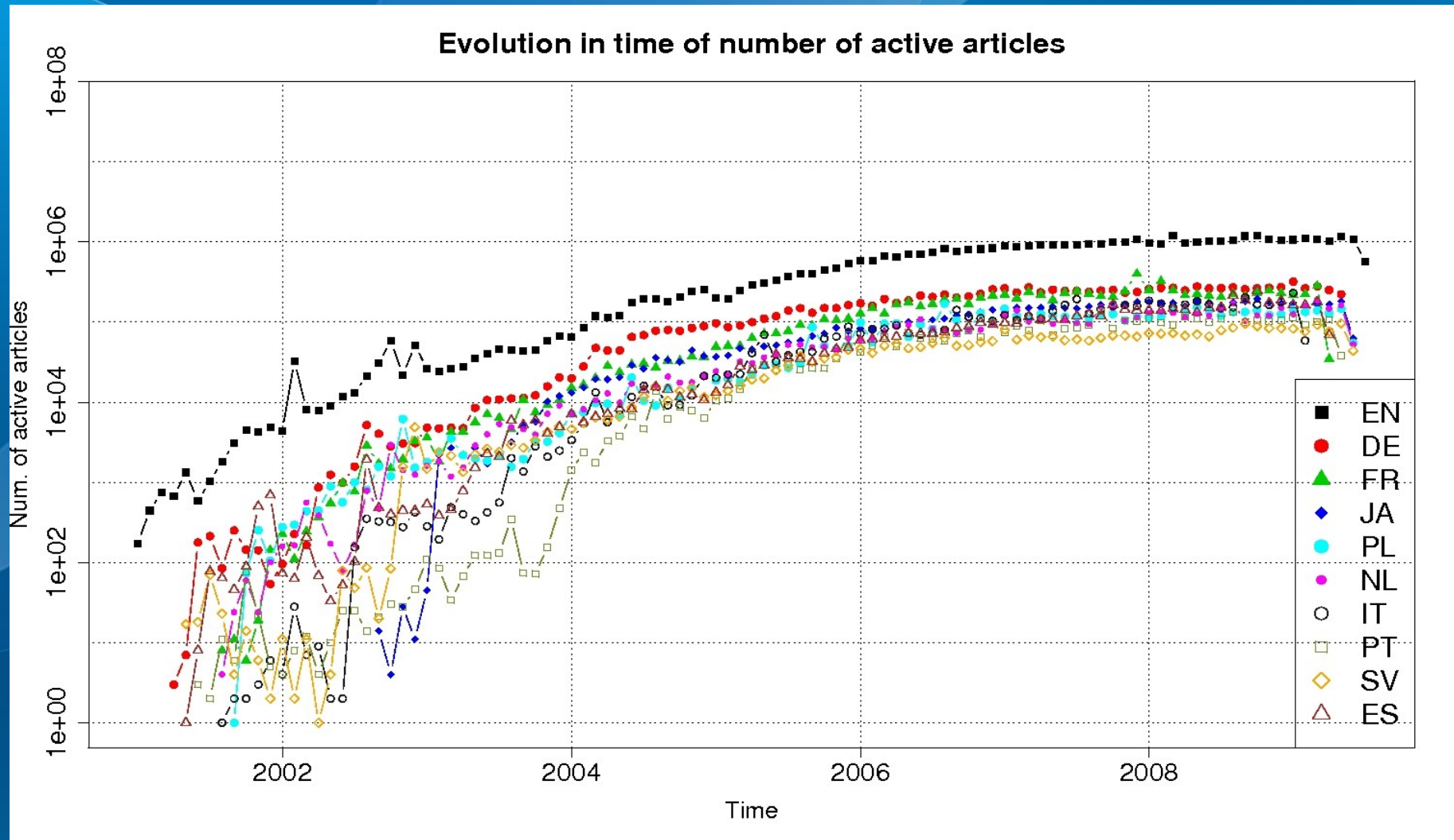
Contents [\[hide\]](#)

- 1 Python Parser
- 2 Article size distribution (HISTOGRAMS)
- 3 Share of bots edits
- 4 Evolution in time
- 5 Correlation between authors and...

2. ARTÍCULOS

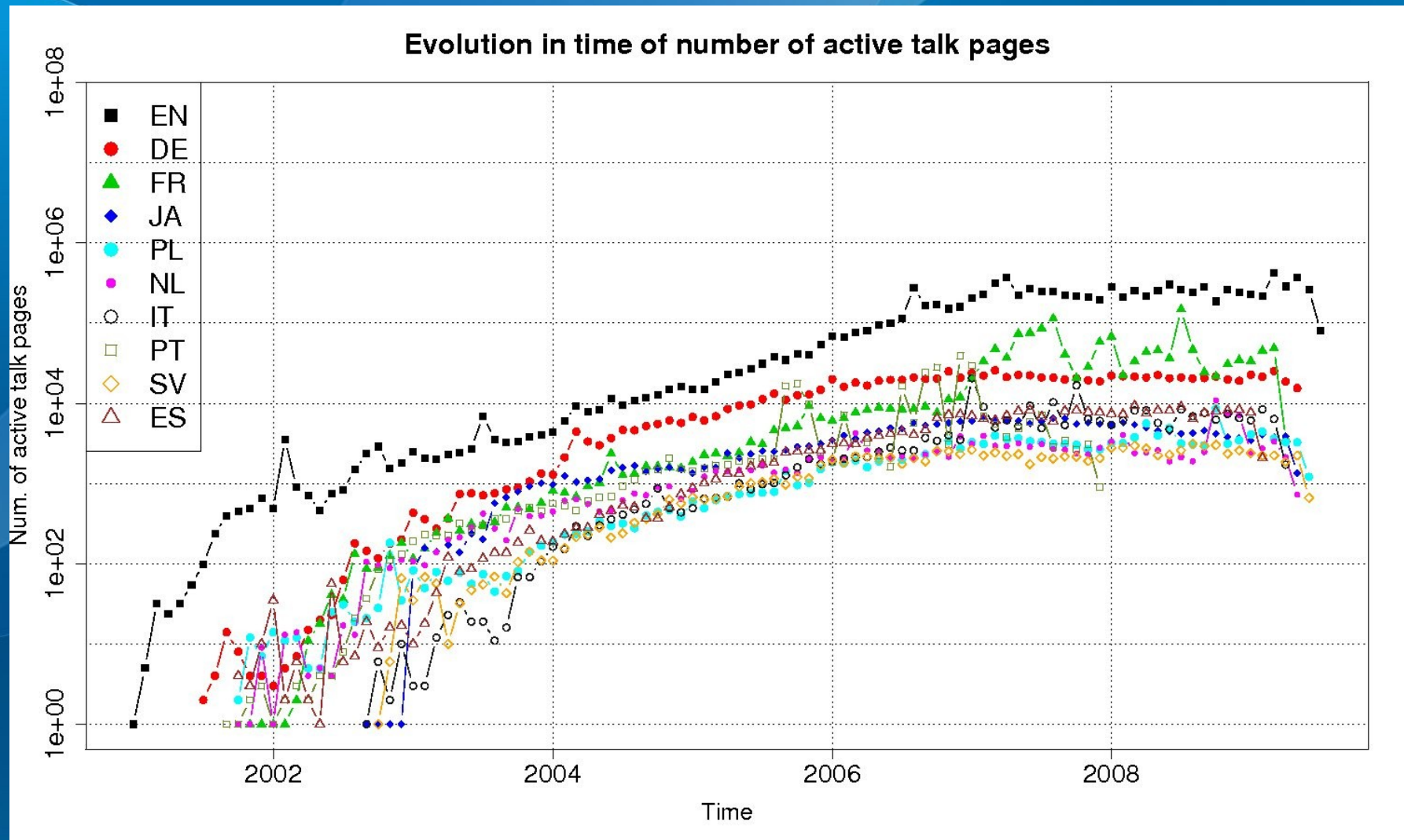
Evolución de artículos activos

3. Artículos



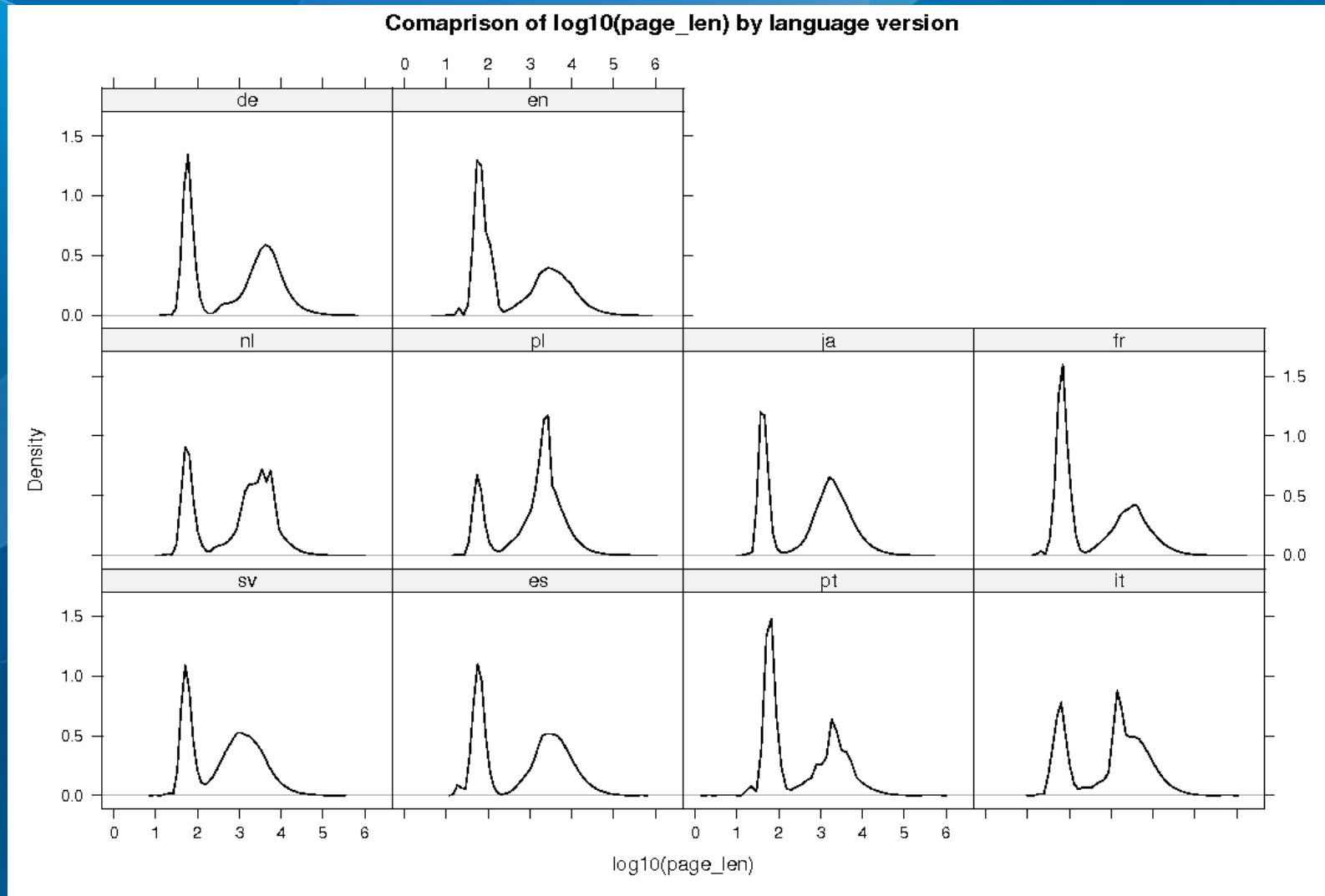
Evol. páginas discusión activas

3. Artículos



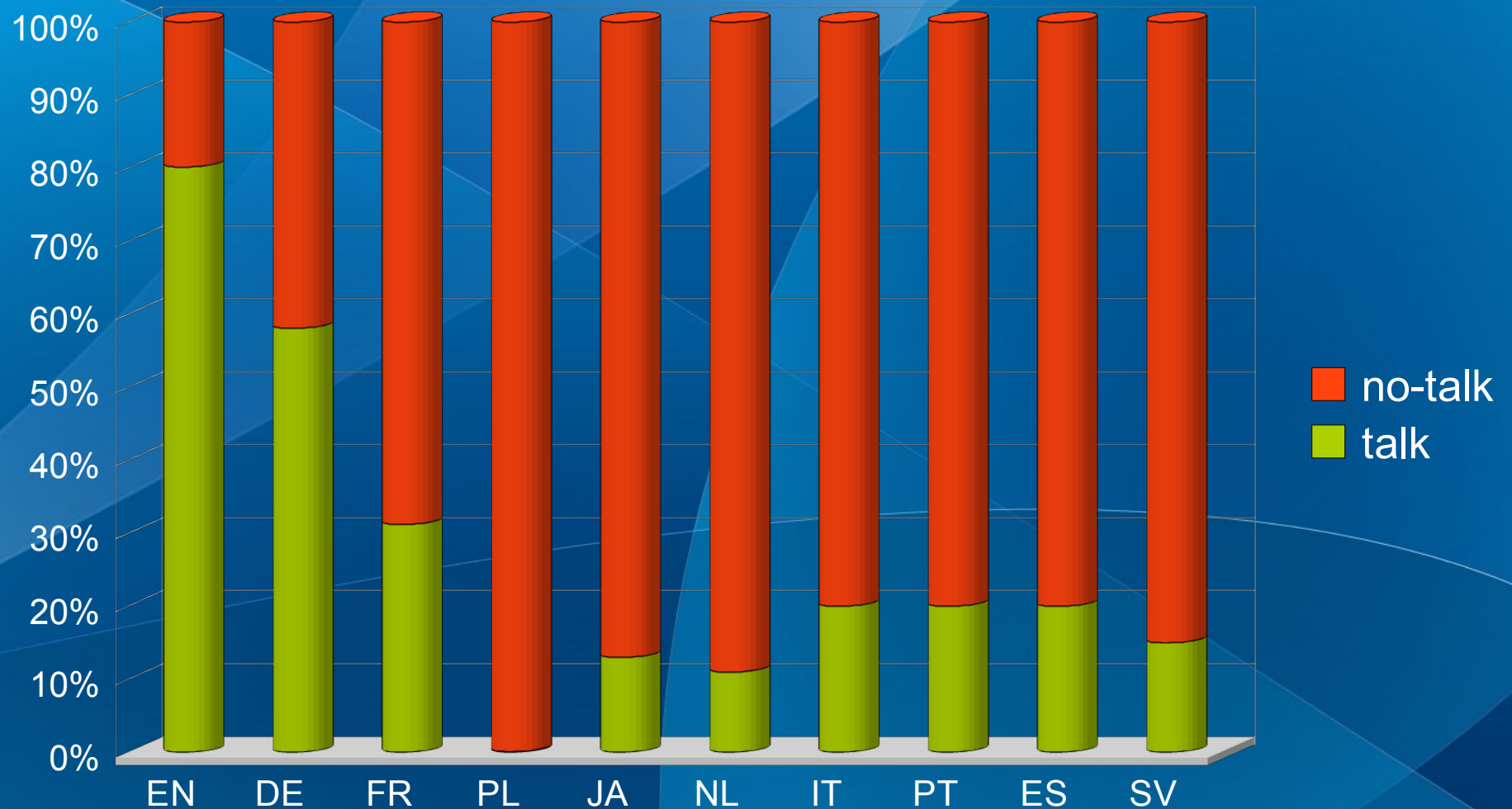
Poblaciones de artículos

3. Artículos



Importancia páginas de discusión

3. Artículos



3. COMUNIDAD

Una multitud... ¿todos iguales?

4. Comunidad



CC-BY-SA : José Felipe Ortega

Pero...¿quién escribe Wikipedia?

4. Comunidad

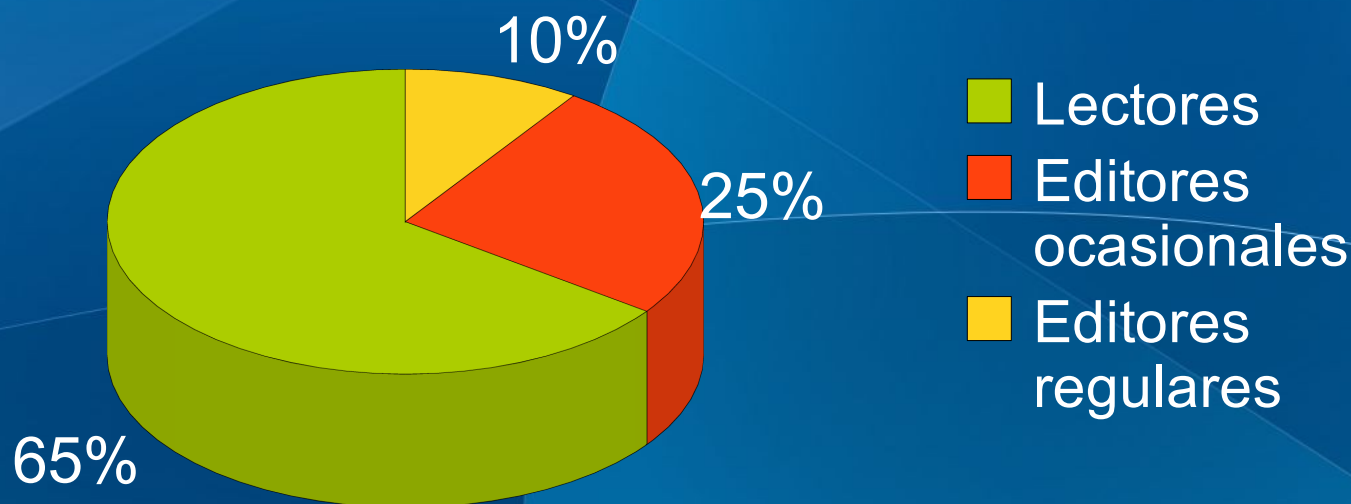


© Wulffmorgenthaler comic strip (8/11/2007)

Estudio general de Wikipedia

4. Comunidad

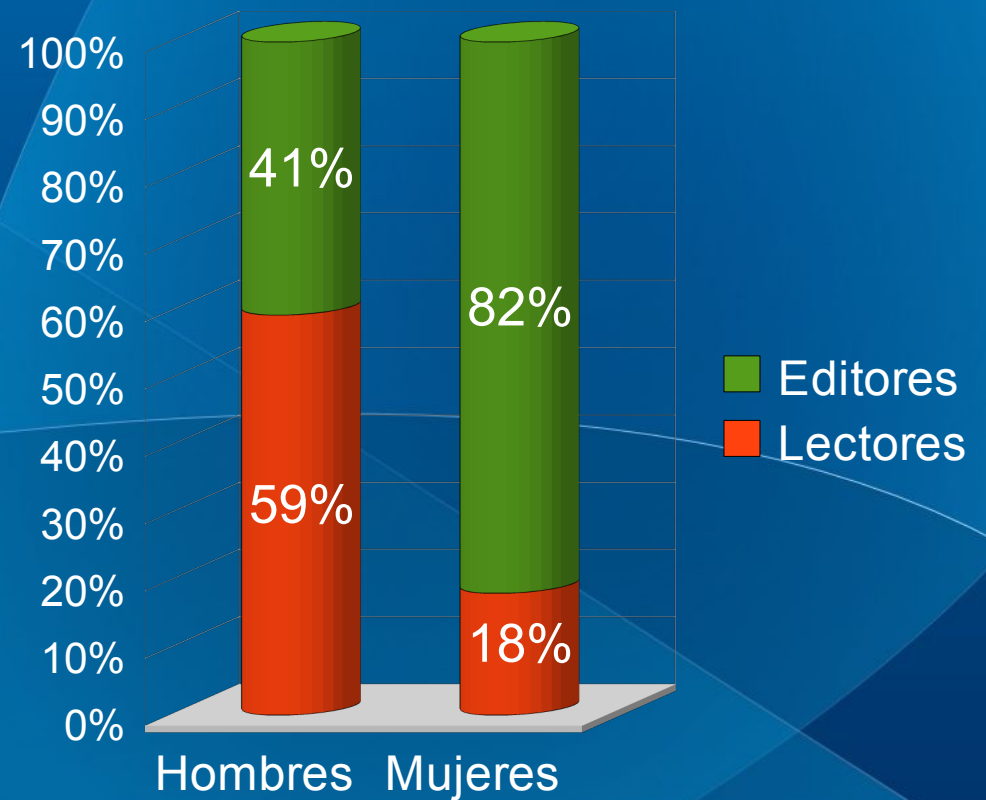
- Realizado por UNU-MERIT.
- Basado en 130,576 casos, hasta abril 2009.
- **Actividad.**



Wikipedia general survey

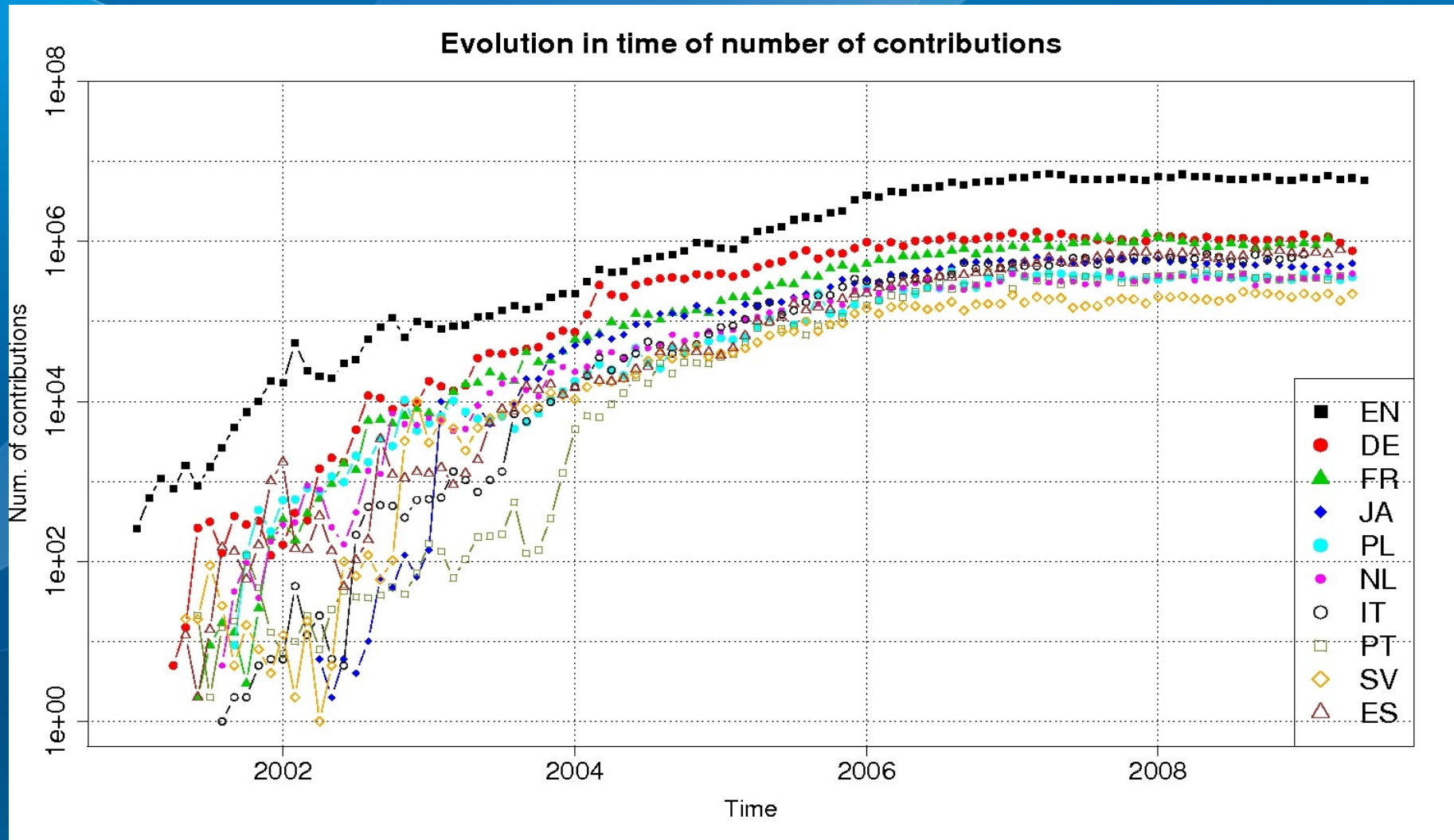
4. Comunidad

• Género



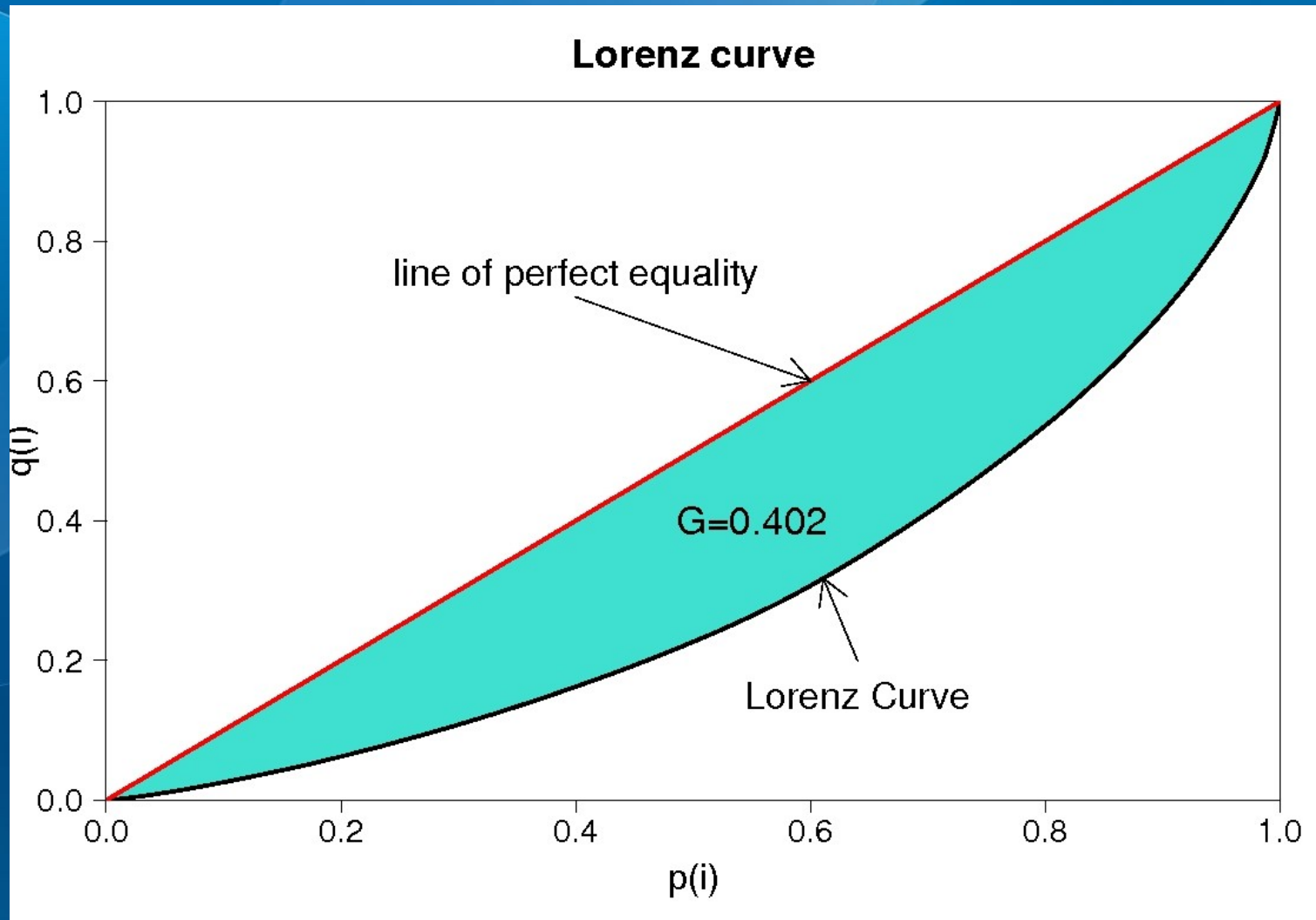
Evolución de ediciones

4. Comunidad



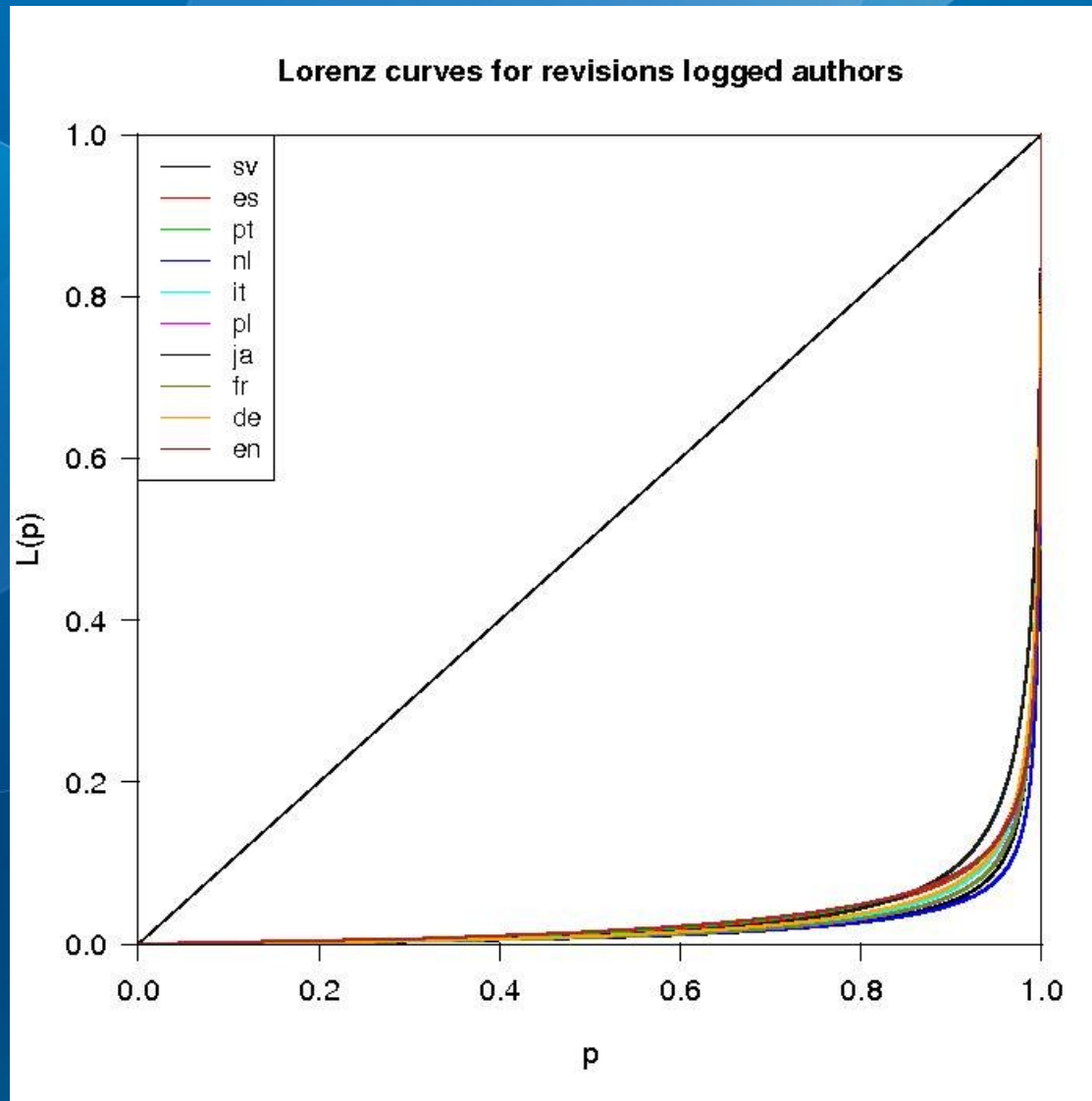
Curva de Lorenz y coef. Gini

4. Comunidad



Lorenz curves in Wikipedia

4. Comunidad



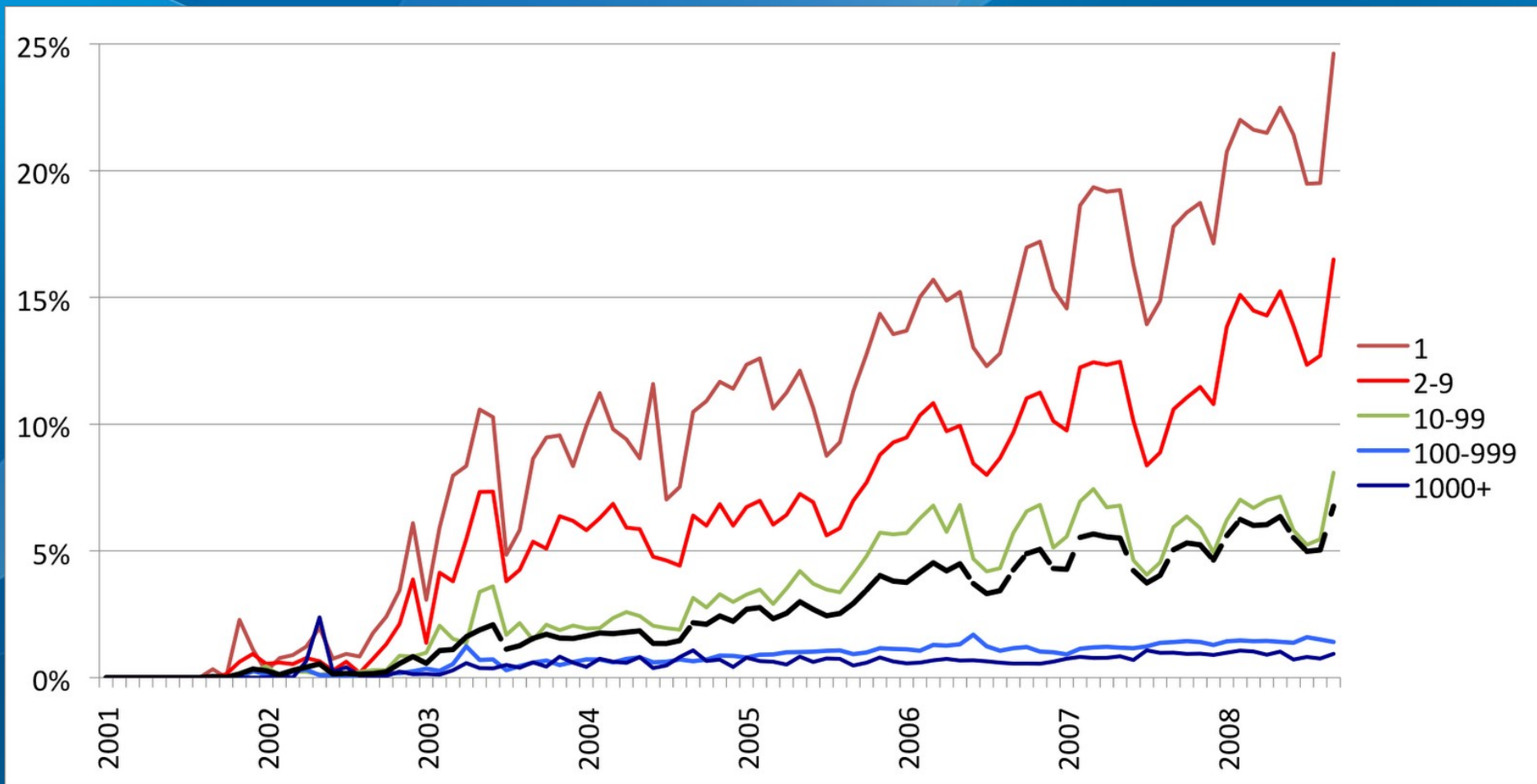
Desigualdad en ediciones

4. Comunidad

- Hay un núcleo de editores muy activos.
 - Valida la percepción de Jimmy Wales.
 - Coeficientes de Gini del 93% al 95%.
 - Alrededor de un 5% de editores efectúa más del 90% del total de revisiones.
- Desigualdad en artículos menos acusada.
 - Coeficientes Gini del 56% al 69%.

Aumento de los *reverts*

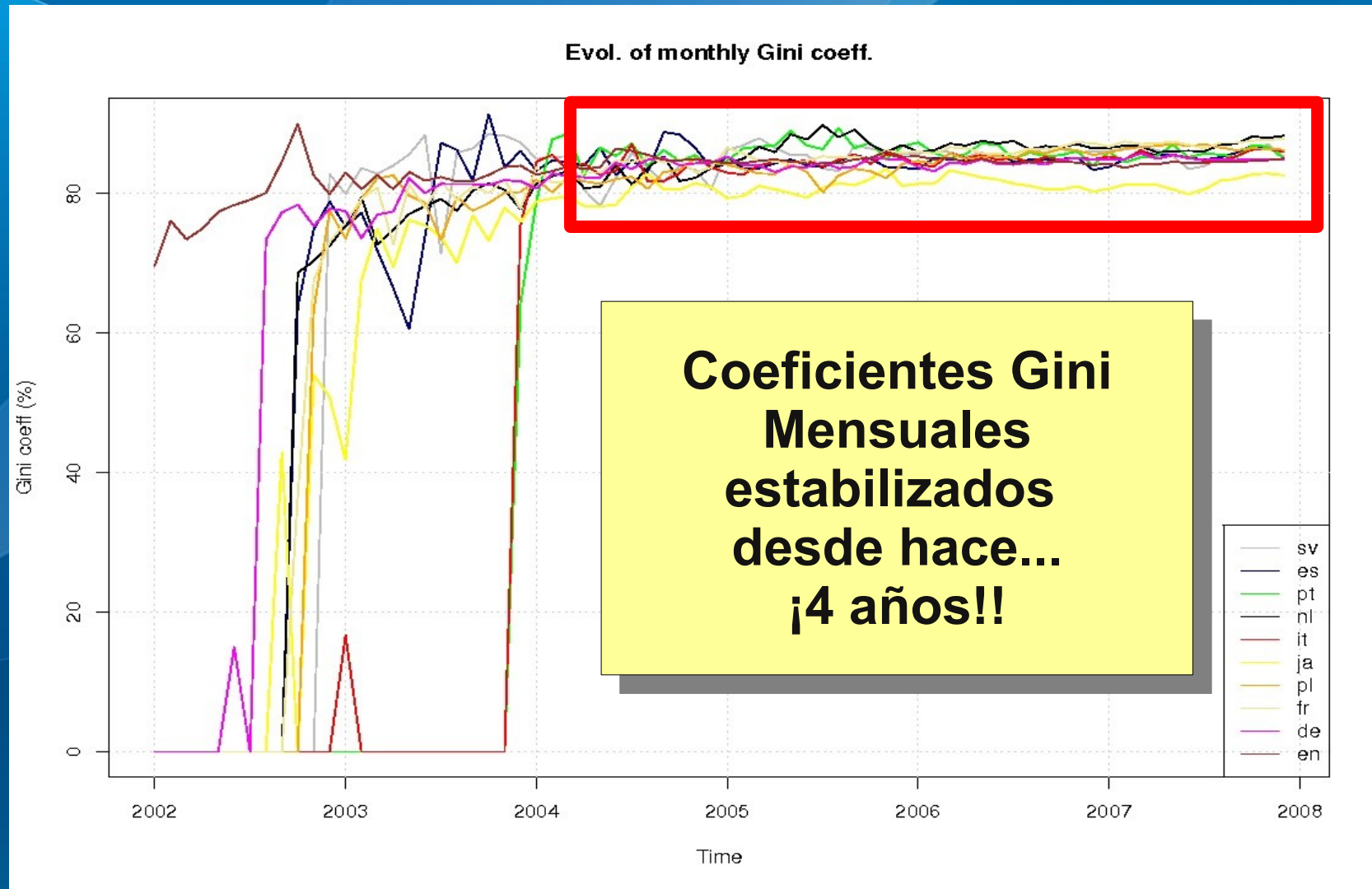
4. Comunidad



Augmented Social Cognition group, PARC (EE.UU.)

Evolution of inequality

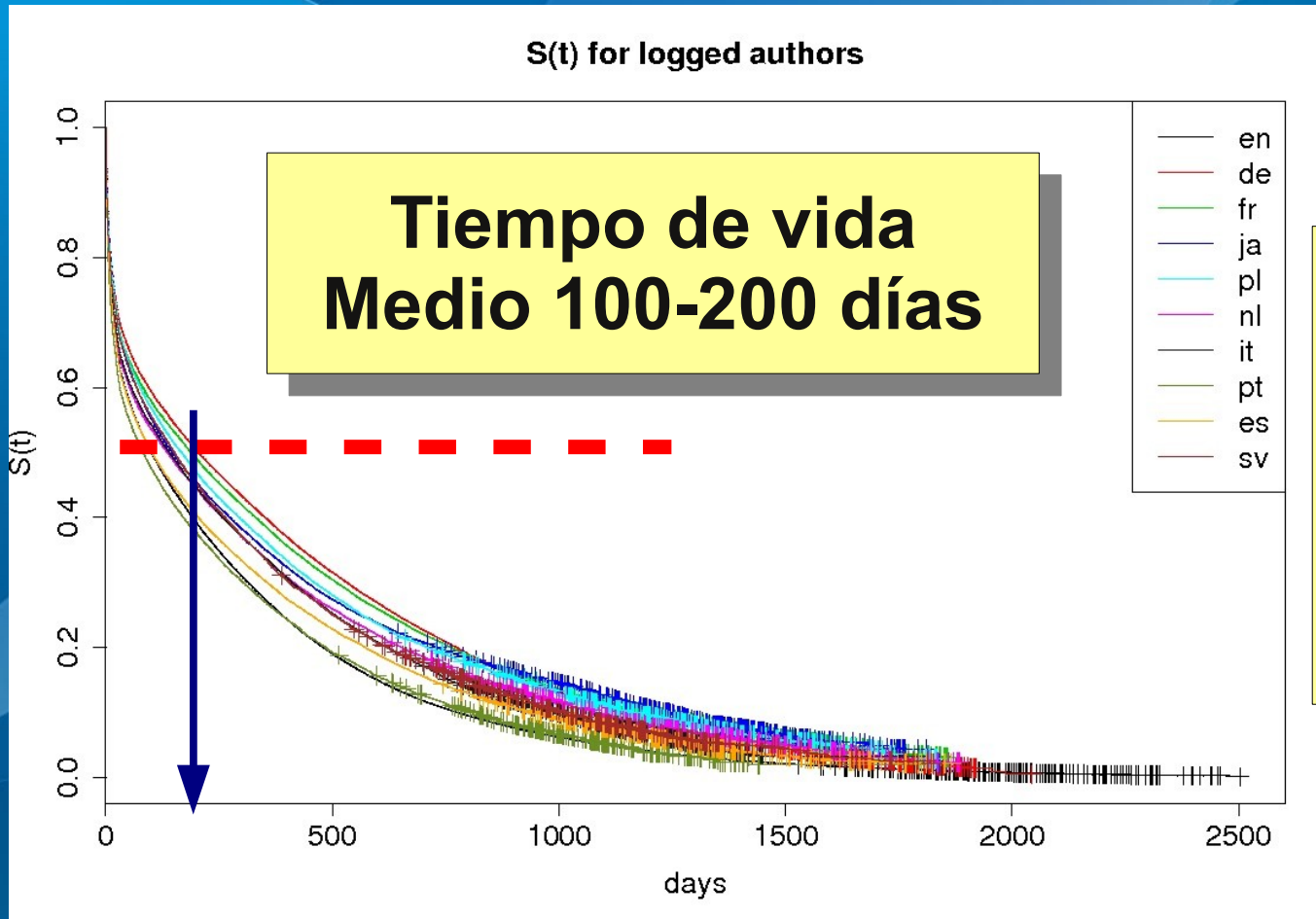
4. Community



4. CONTINUIDAD

Análisis de supervivencia

5.Continuidad



50% llegan al núcleo antes de 120 días

50% dejan el núcleo tras 100 días.

Aspectos de calidad

5.Continuidad

- 1.000 días para que un artículo sea nominado como FA.
- 10 veces más editores en FAs que en artículos normales (casi **200 veces** en la versión en inglés).
- FAs revisados por editores con **más experiencia** (cerca de 1.000 días en el proyecto).
- Especialmente importante mantener a estos editores para garantizar la calidad del contenido.

¿El ocaso de Wikipedia?

5.Continuidad

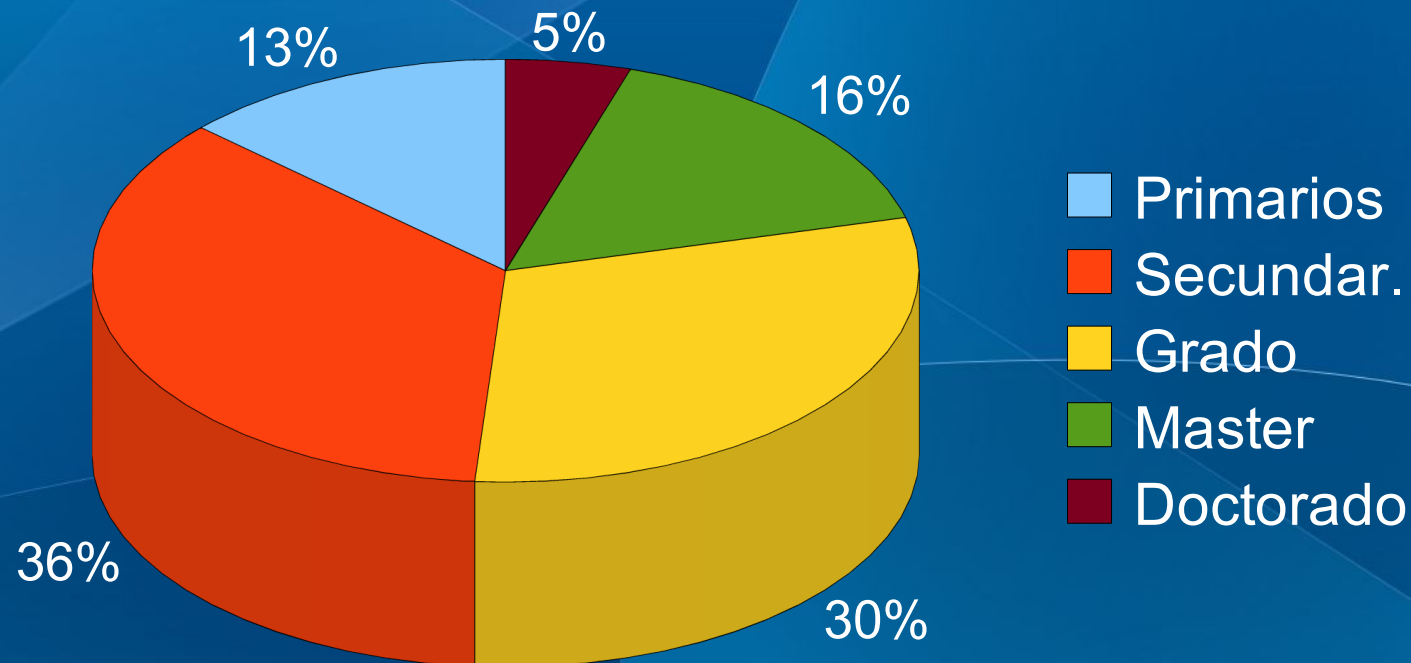


CC-BY-SA : José Felipe Ortega

Lecciones del estudio de WKP

5.Continuidad

- **Desglose de los editores según su nivel de estudios.**



Lecciones del estudio de WKP

5.Continuidad

- Razones para contribuir.
 - “Me gusta la idea de compartir conocimiento y quiero contribuir”.
 - “Quería corregir un error”.
 - “Por razones profesionales”.
 - “La información debería estar disponible para cualquiera que quiera consultarla”.

Lecciones del estudio de WKP

5.Continuidad

- **Razones para no contribuir**
 - **(25%)** “No sé cómo contribuir”.
 - **(42,2%)** “Sería más sencillo contribuir si supiese qué áreas precisan mi aportación”.
- **Razones para no hacer una donación.**
 - **(42%)** “No sé cómo puedo hacerla”.

Estrategias de mejora

5. Continuidad

- Wikipedia y educación.
 - Universidades, académicos.
 - Institutos, centros formación.
- Wikipedia: comunidad.
 - Mejorar experiencia de los usuarios.
 - Información extra sobre el proceso editorial
 - Información extra sobre los editores.
 - Añadir enlaces/soporte para redes sociales.



CC-BY-SA: Bios