

---

# Análisis de listas de correo en el software libre: un caso de estudio

*Israel Herraiz, Juanjo Amor, Álvaro Navarro*  
{herraiz, jjamor, anavarro}@gsyc.escet.urjc.es  
{israel.herraiz, juanjo.amor,  
alvaro.navarro}@hispalinux.es



*Córdoba, 5 de noviembre de 2005*

---

(cc) 2005 Israel Herraiz, Juanjo Amor, Álvaro Navarro  
Some rights reserved. This work is licensed under the Creative  
Commons Attribution-ShareAlike License.

To view a copy of this license, visit  
<http://creativecommons.org/licenses/by-sa/2.0/>  
or send a letter to Creative Commons, 559 Nathan Abbot Way,  
Stanford, California 94305, USA

## Objetivo

- Estudiar la actividad y la participación en una comunidad de software libre.
- Listas de correo electrónico
  - Medio de comunicación principal
  - Discusiones acerca de la organización, intercambio de información, decisiones relativas a la comunidad
- Hipótesis
- Actividad y participación en listas de correo indican actividad y participación en la comunidad

## Caso de Estudio

### Hispalinux

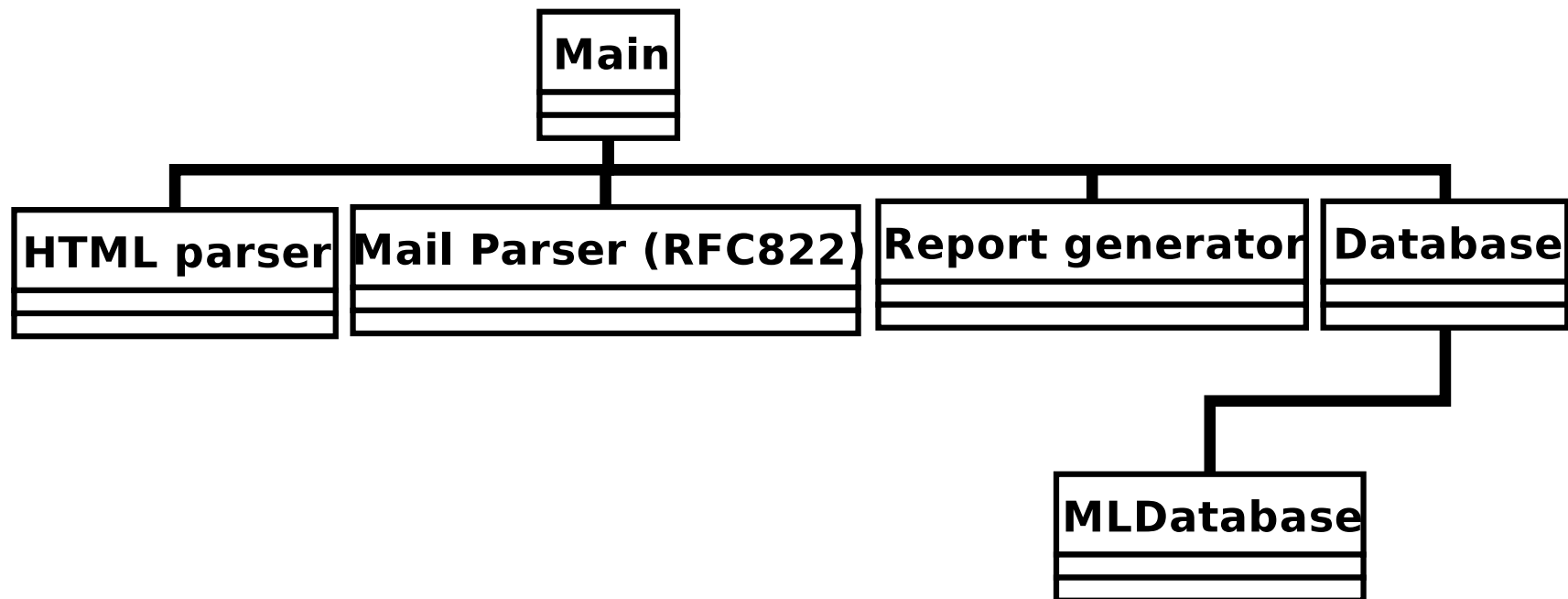
- Fundada en 1997.
- El mayor grupo de usuarios de software libre de España (7000 socios).
- Hazte socio en <http://www.hispalinux.es>
- Tensiones internas.
- Varias listas de correo. Grupos de trabajo. Socios.

# MLStats

## Herramienta para analizar el archivo de una lista de correo

1. Descargar página web de descripción del archivo de la lista (HTTP, FTP)
2. *Parser* del HTML. Búsqueda de enlaces a ficheros *mbox*.
3. Descarga de cada fichero *mbox*.
4. Descomprimir si es necesario.
5. *Parser* de cada mensaje, siguiendo el estándar RFC-822.
6. Volcado de cabeceras y/o contenido del mensaje a una base de datos.
7. Generación de gráficas y estadísticas a partir de la base de datos.

## Arquitectura



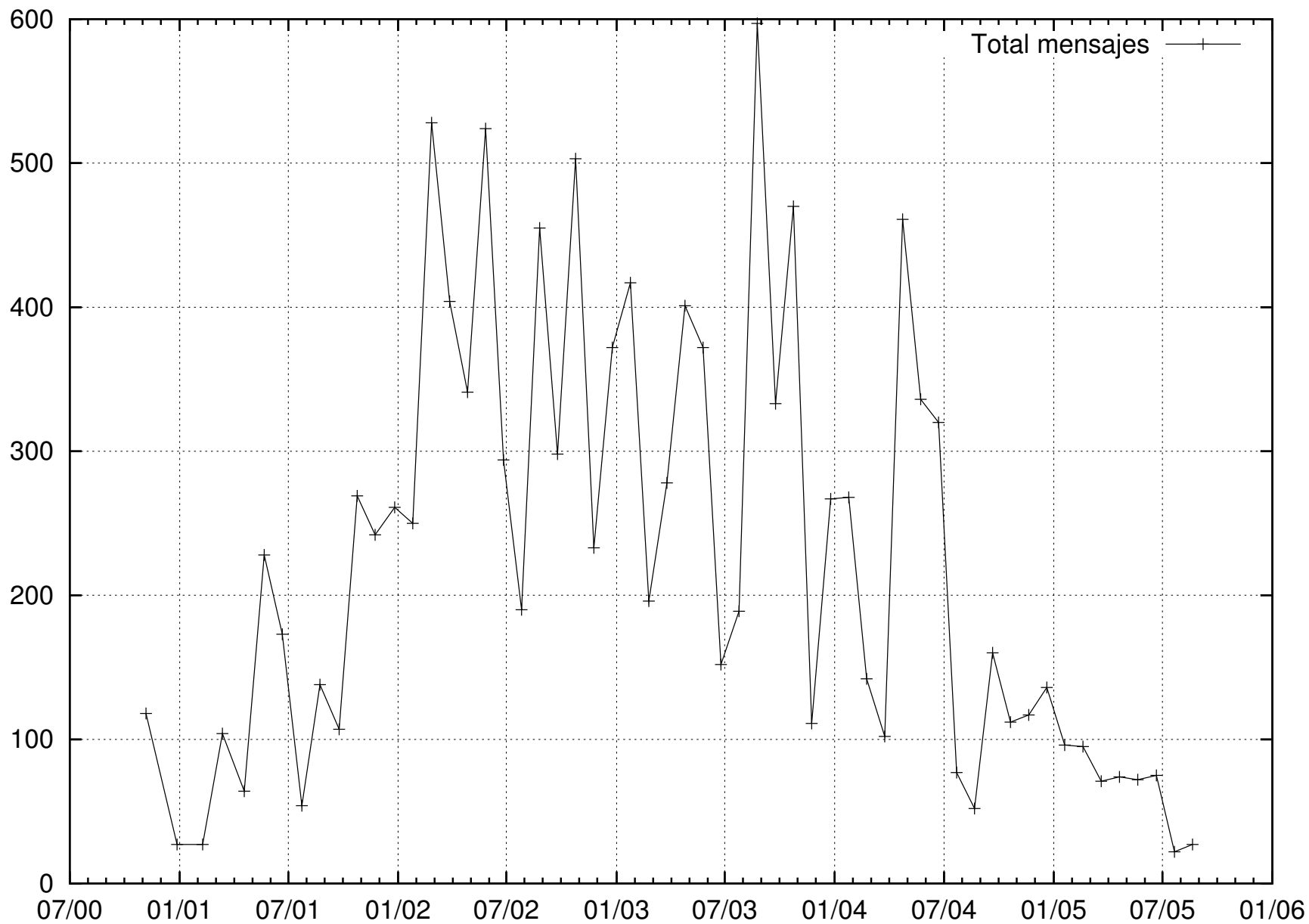
## Resultados

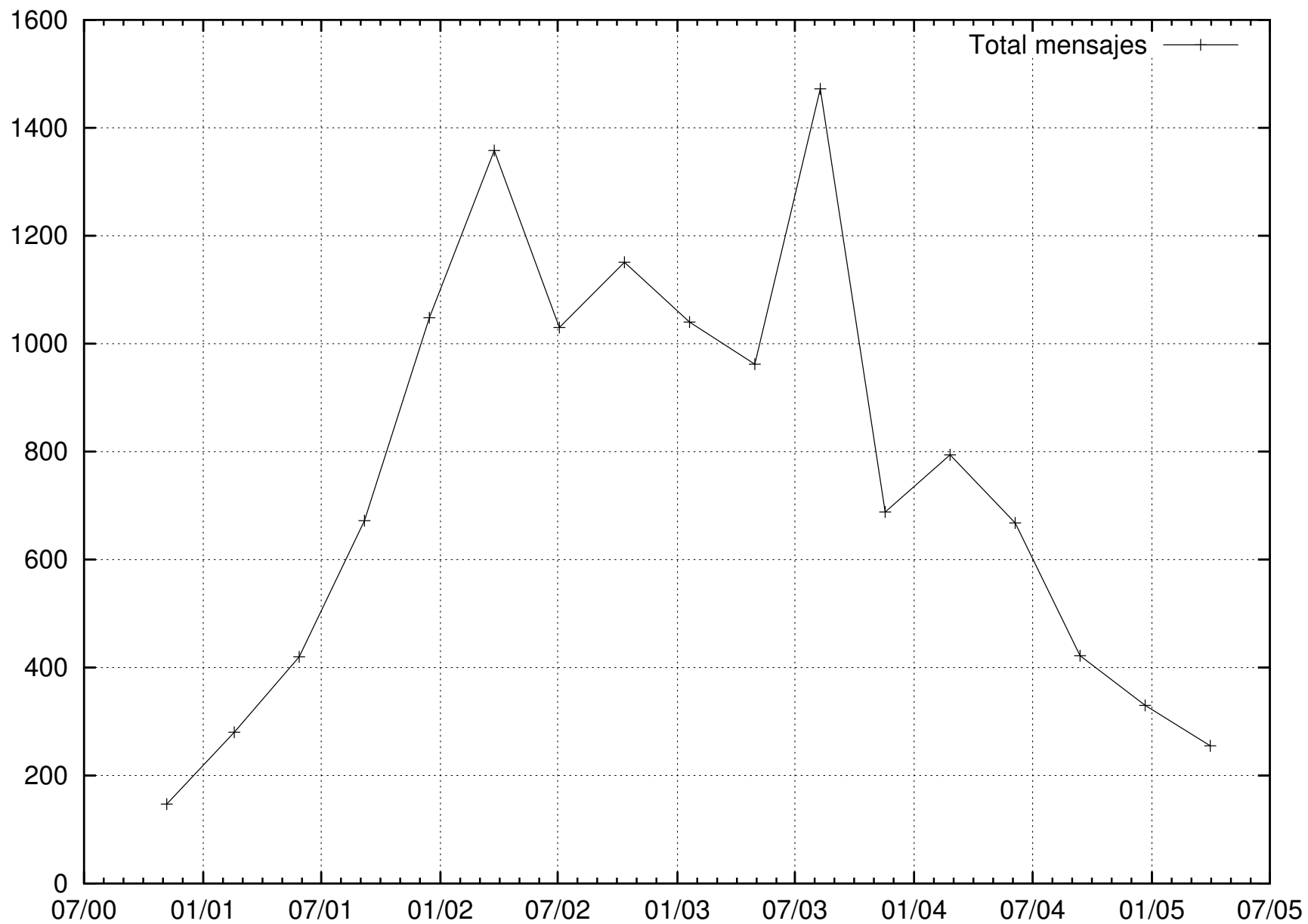
- Tamaño del archivo: 65 megabytes
- 12820 mensajes
- 1 mensaje no cumplía RFC-822 y fue eliminado de la muestra
- Primer mensaje: 6 de noviembre de 2000
- Último mensaje: 27 de septiembre de 2005
- Mes con más mensajes: Septiembre de 2003
- Mes con menos mensajes: Agosto de 2005

## Número de participantes en la lista

- 1134 direcciones de correo electrónico
- 1197 nombres
- 1696 mensajes con diferente From
- Mes con más participantes: noviembre de 2002
- Mes con menos participantes: agosto de 2005
- Concentración de la participación:
  - 24 direcciones de e-mail han hecho el 50 % de los envíos
  - 32 nombres han hecho el 50 % de los envíos
- Correlación lineal mensajes - direcciones de email:  $r = 0,9666$
- Correlación lineal mensajes - nombres:  $r = 0,9702$

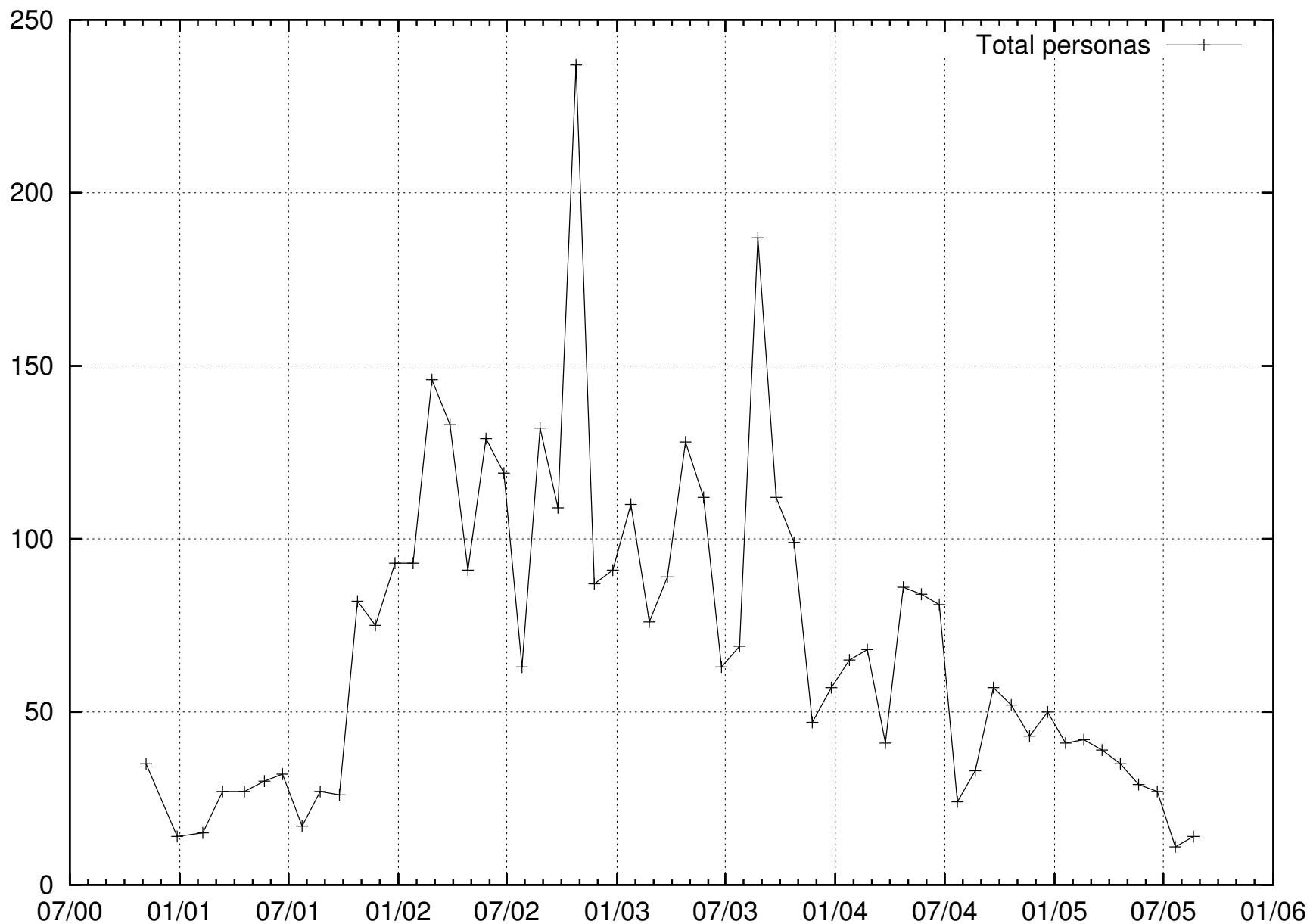
## **Evolución de la actividad**

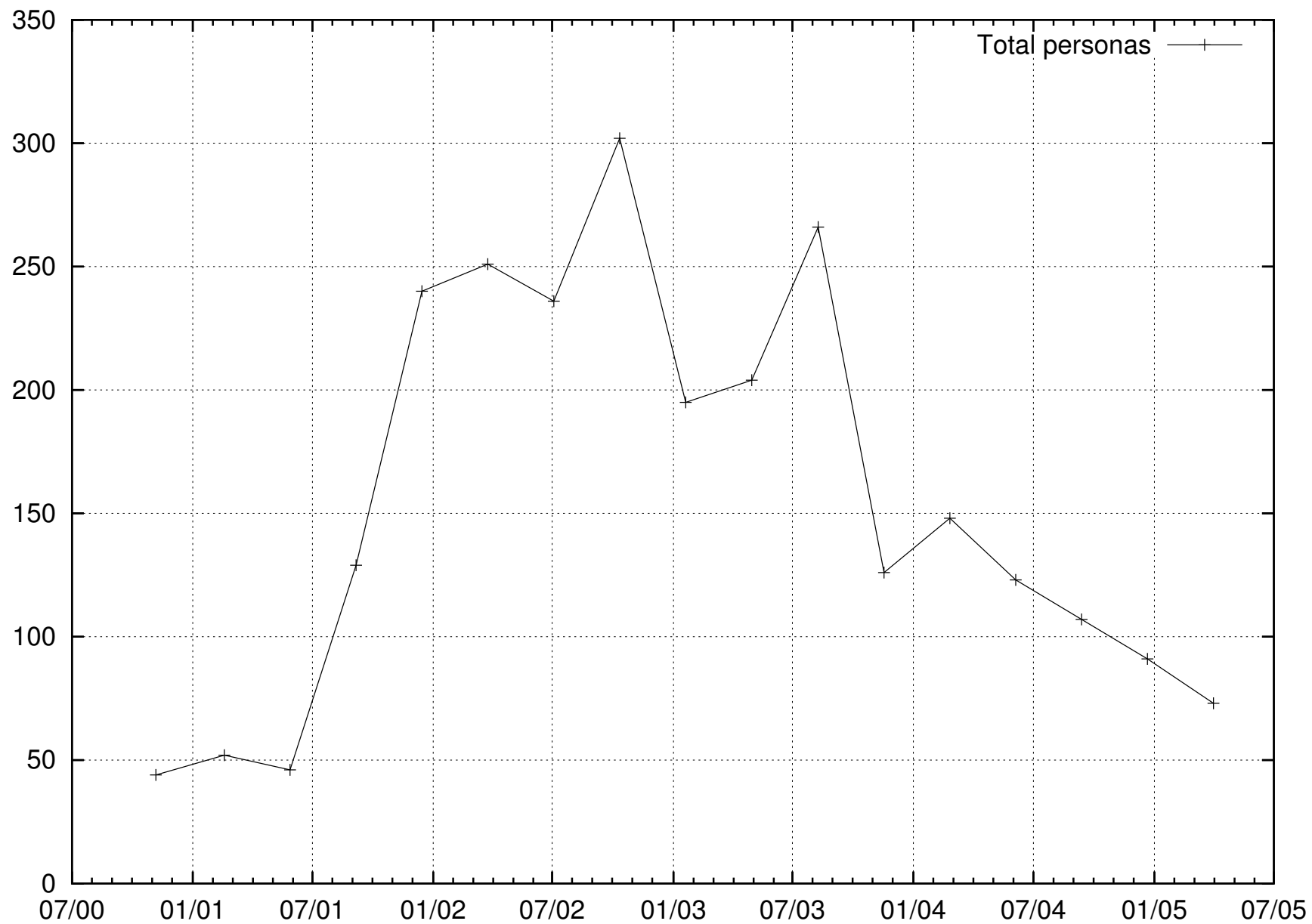




# **Evolución de la participación**

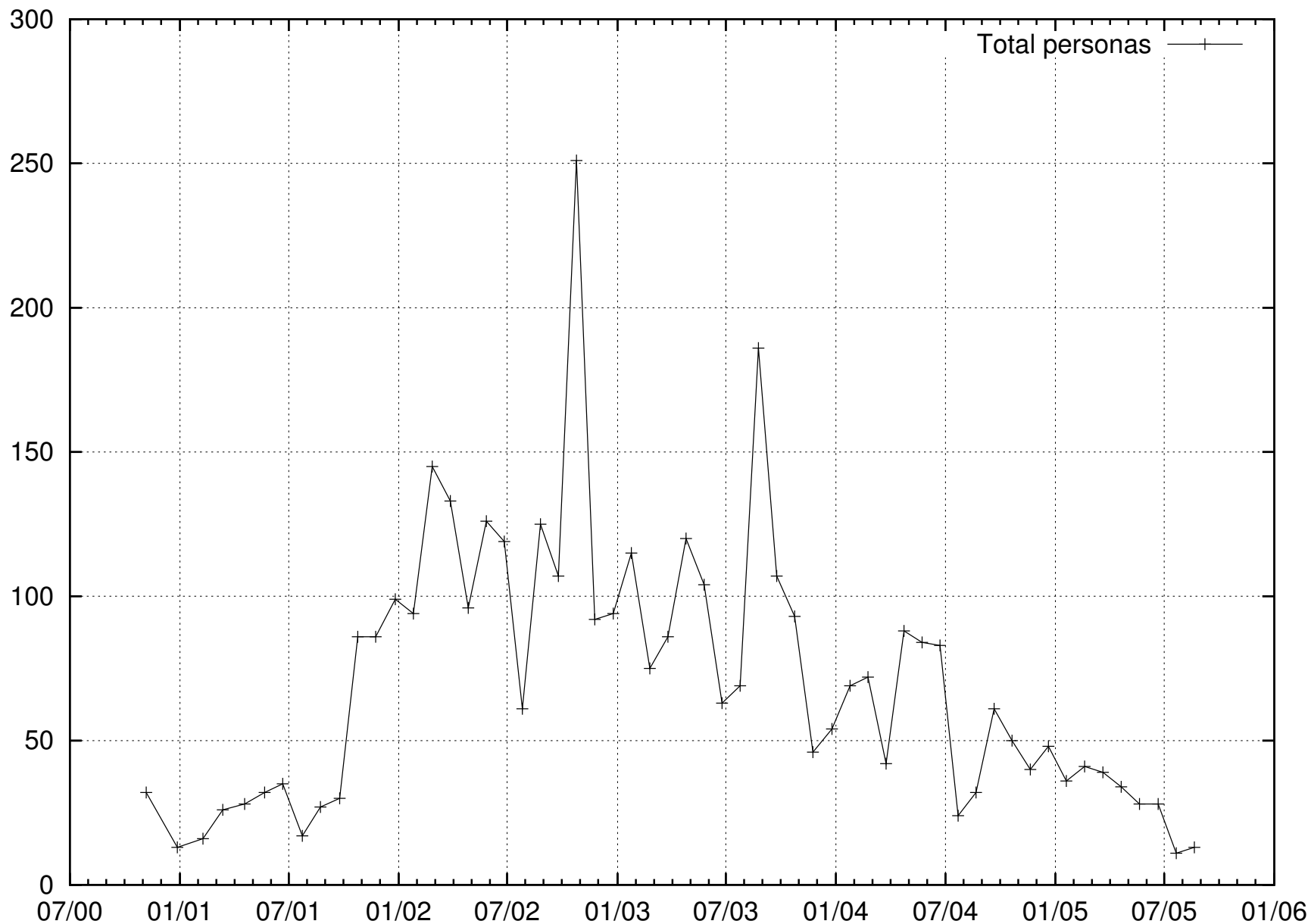
**Diferentes nombres**

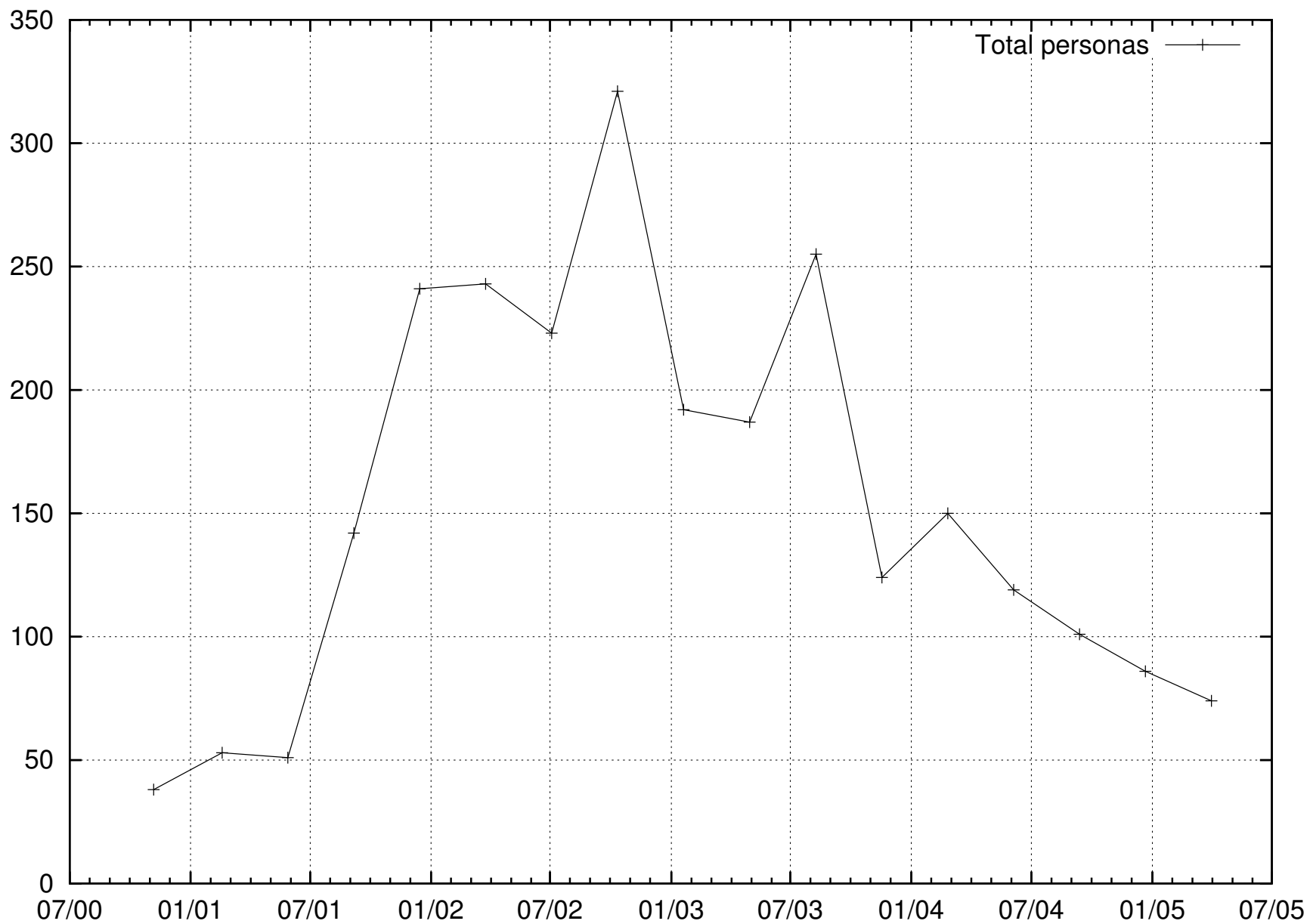




# **Evolución de la participación**

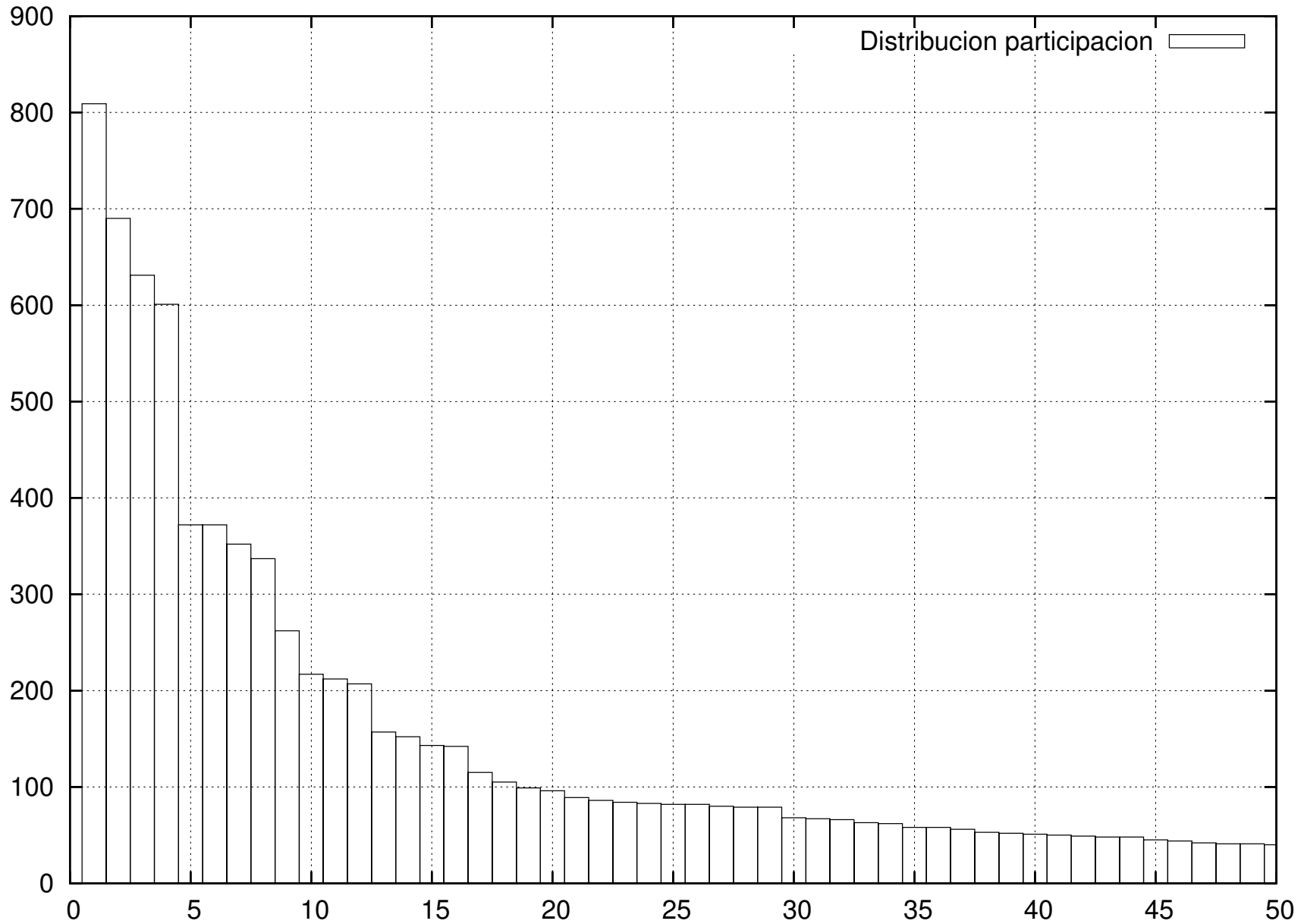
**Diferentes emails**

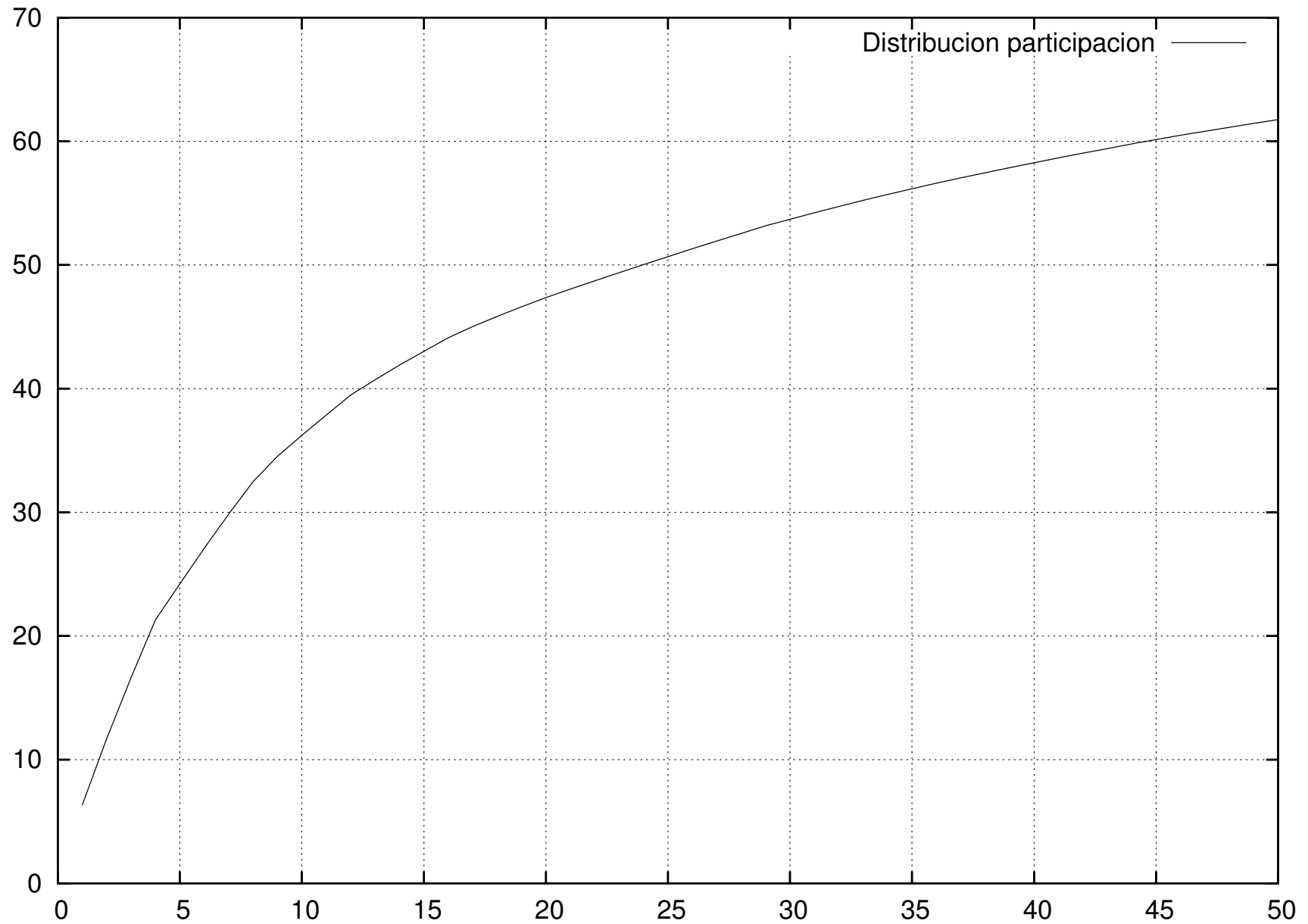




# Concentración de la participación

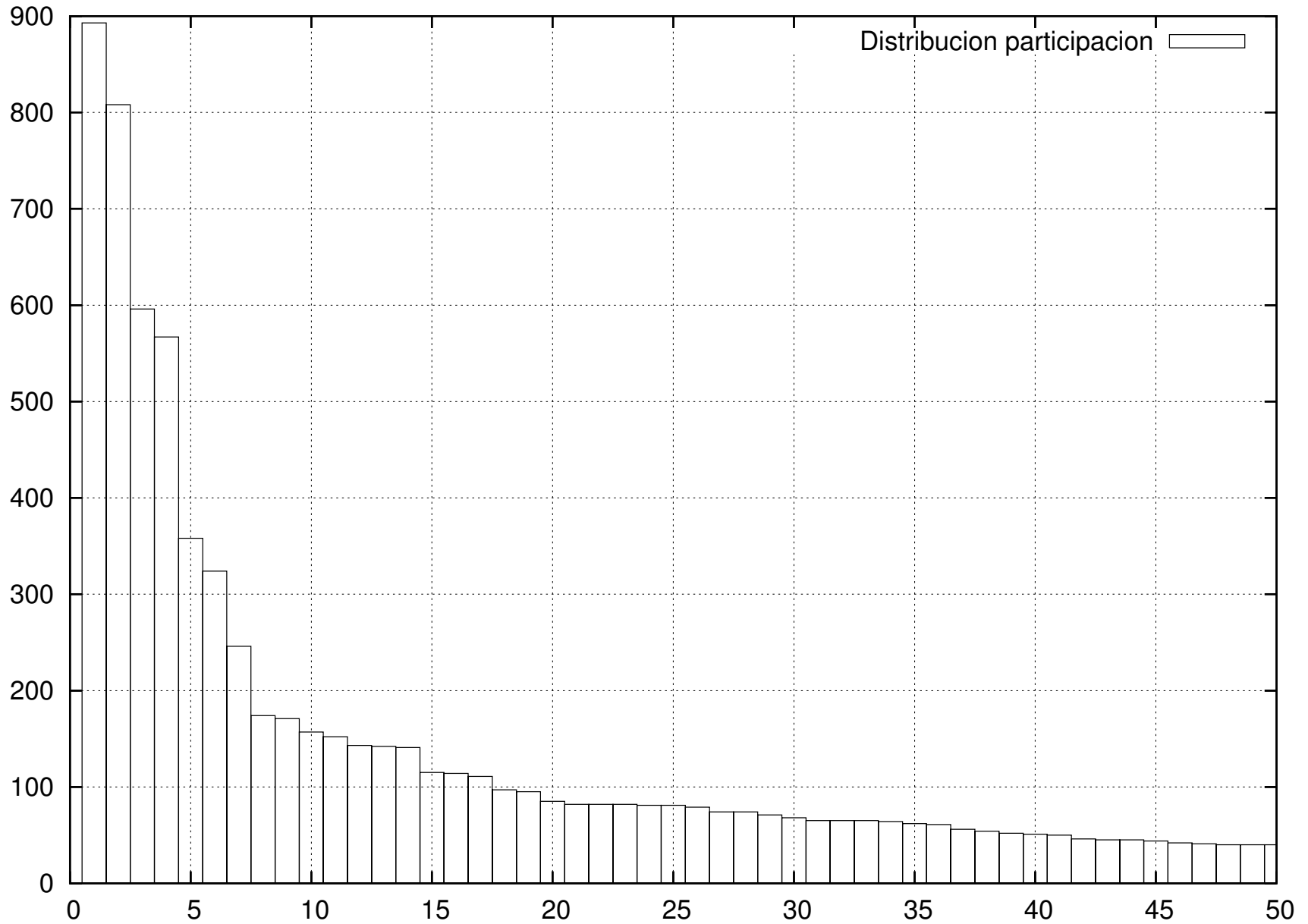
**Diferentes nombres**

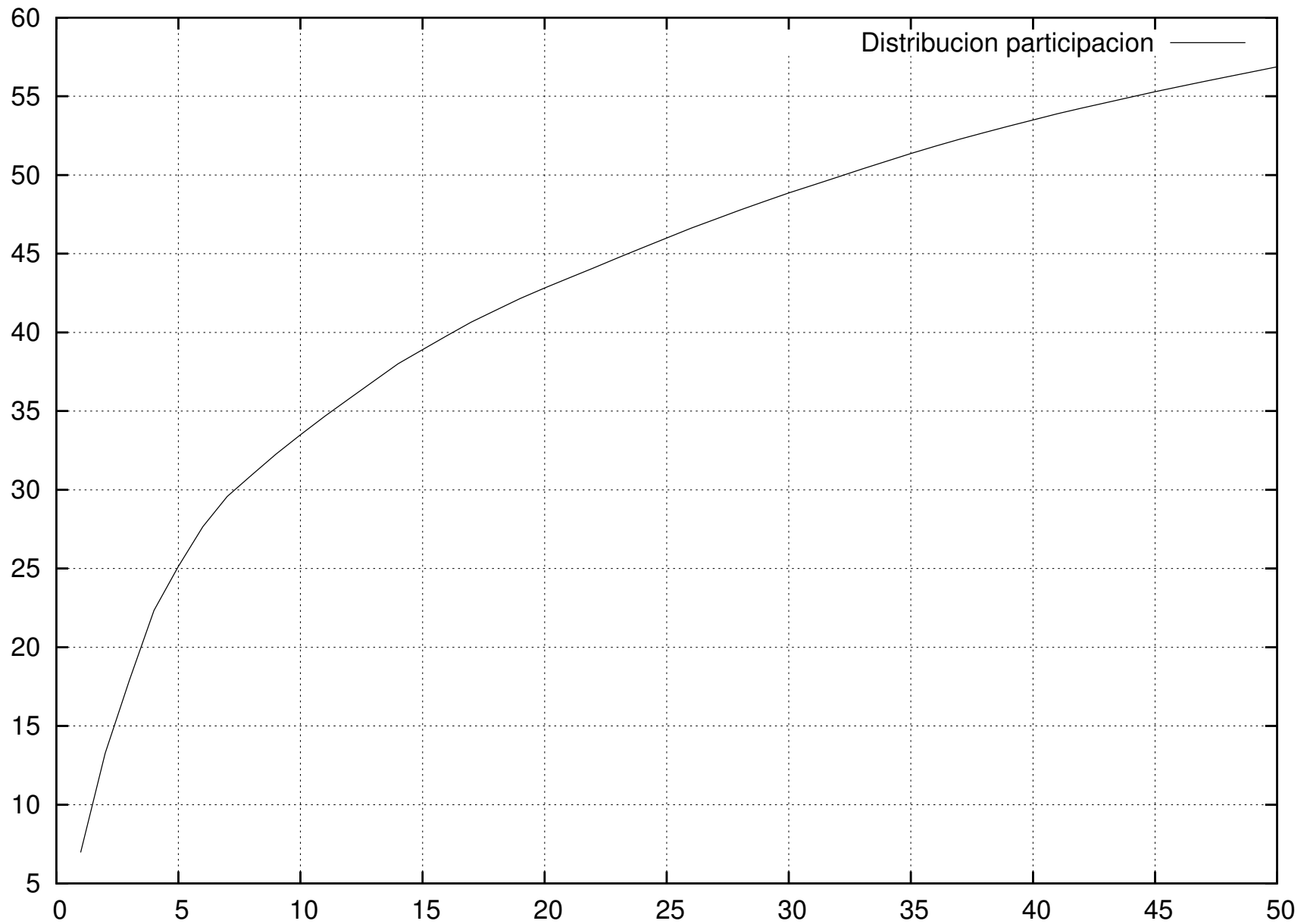




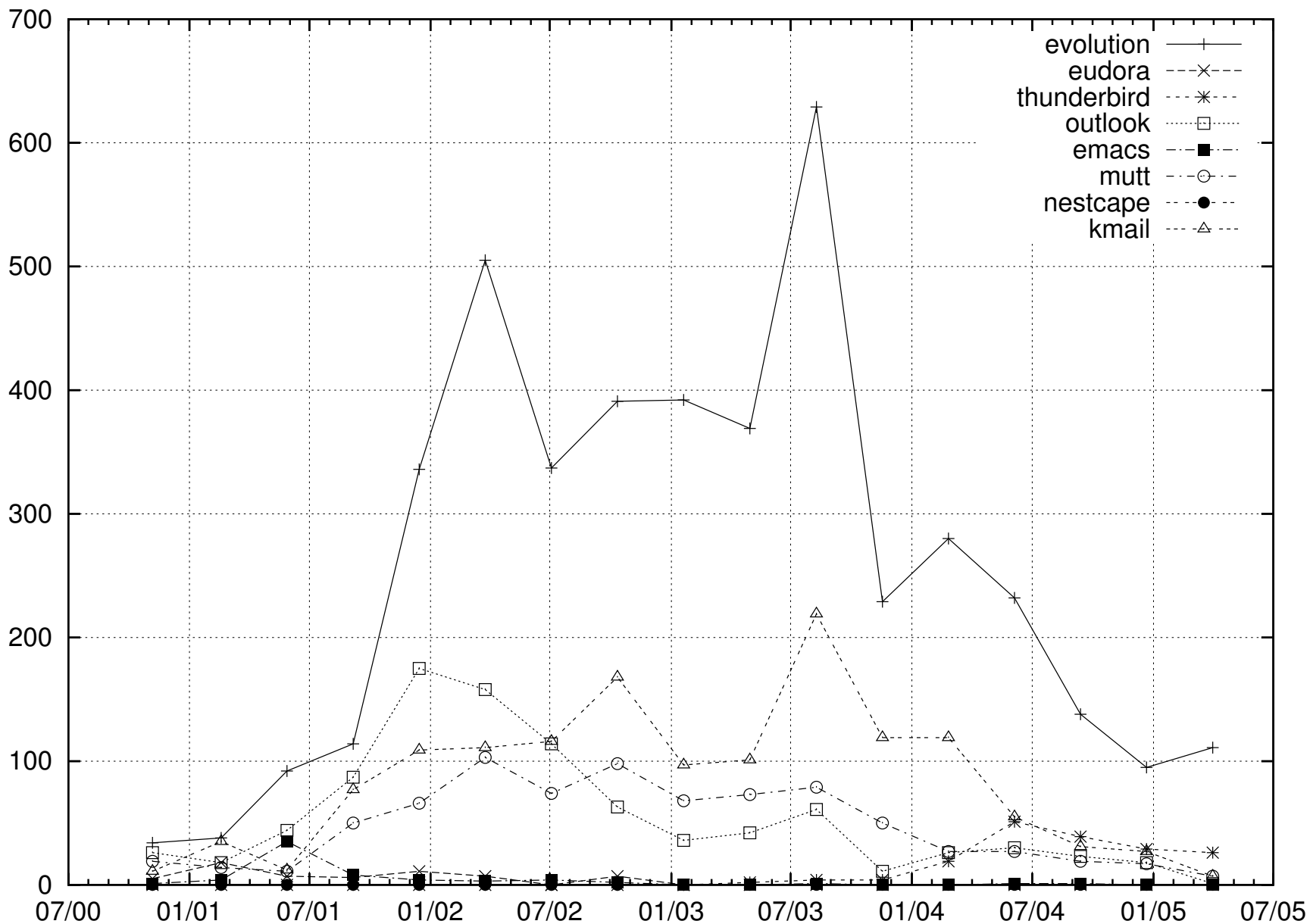
# Concentración de la participación

**Diferentes emails**

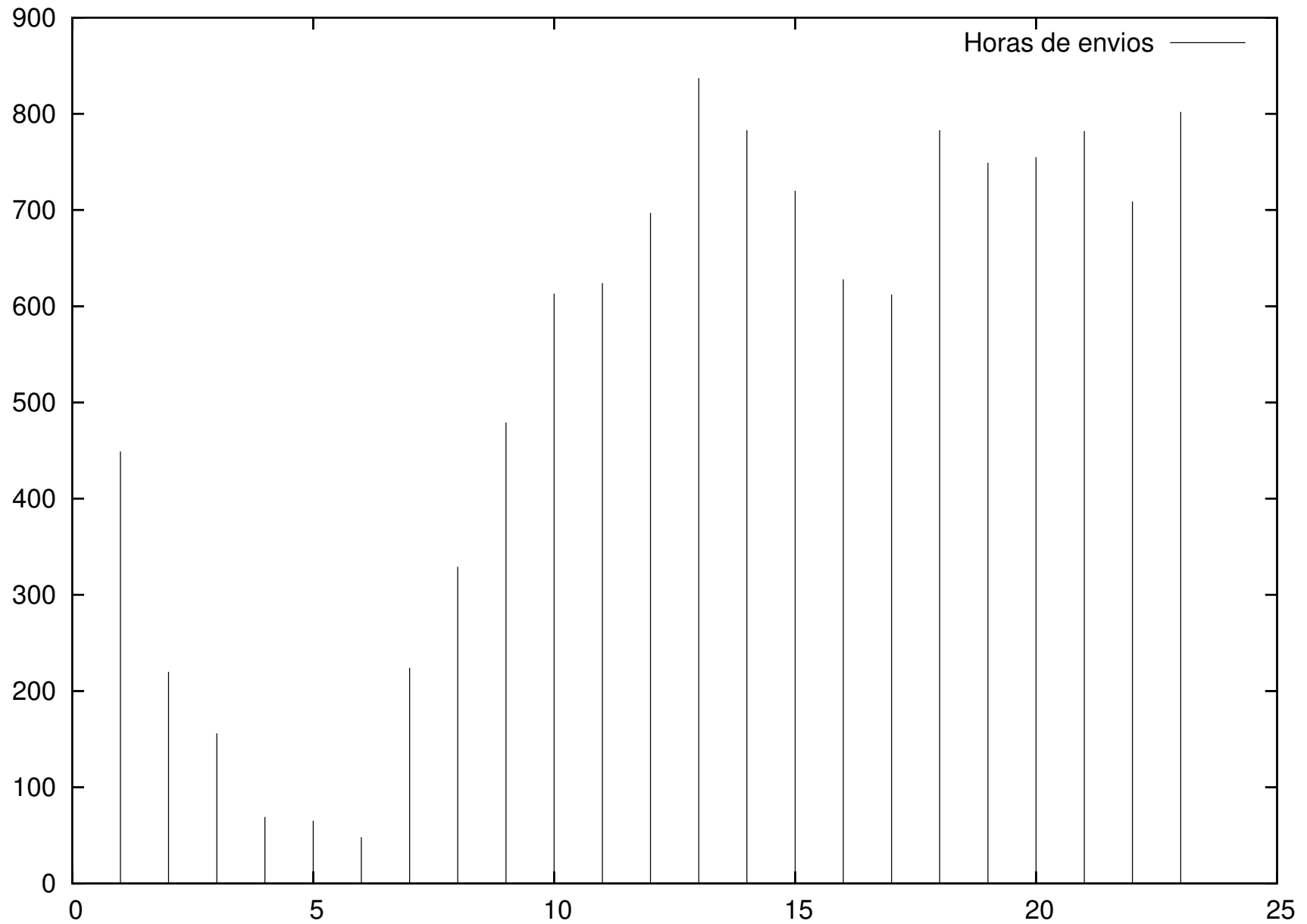




# **Evolución de los clientes de correo electrónico usados**



## Horas de envío de mensajes



## Conclusiones

- Lista de socios. Elemento de comunicación entre los miembros de Hispalinux.
- Lejos de su apogeo. Actividad y participación disminuyendo.
- No hay individuos imprescindibles. Actividad bien distribuida entre todos los participantes.
- Mayoría de mensajes en horario habitual de España.
- ¿Por qué ha ocurrido esto?
  - Probablemente debido a un conjunto de causas.
  - En nuestra opinión, debido a la deslocalización. Auge de los grupos locales (162 grupos en España).
  - Transferencia de participación desde un grupo centralizado a muchos grupos distribuidos.

**Gracias por vuestra atención.**

**¿Preguntas?**

**No os olvidéis de visitar  
<http://libresoft.urjc.es>**