## Desequilibrios en la representatividad de los datos abiertos pertenecientes a la plataforma IMDb

Carlos Aguilar Paredes

Universitat de Barcelona

ORCID: 0000-0002-2165-4279

Javier Sanz Aznar.

Universitat de Barcelona

ORCID: 0000-0002-9837-761X

## Resumen

La base de datos abierta IMDb tiene más de 250 millones de visitantes mensuales e indexa más de 500 millones de interacciones de sus usuarios registrados en forma de votaciones y 3,9 millones en forma de comentarios (IMDb Stats, 2018). Su popularidad, la gran cantidad de metadatos que vincula alrededor de las obras indexadas (más de 5 millones entre audiovisuales cinematográficos y televisivos) y la constante interacción de sus usuarios la convierten en una gran herramienta para la investigación.

Sin embargo, la gran cantidad de datos aportados de forma libre por los usuarios puede generar desequilibrios en la representatividad de los grupos poblacionales votantes. Puede suceder que los usuarios al expresar su opinión o voto respecto a una obra concreta, gran cantidad de los votantes perteneciesen a un rango concreto de edad, un género o una nacionalidad concreta y esa situación produjese que las votaciones resultantes estuviesen condicionadas por las preferencias de un grupo social y silenciasen el resto de los grupos poblacionales. Esta situación implicaría en que las investigaciones que toman IMDb como una fuente neutra de información puedan obtener resultados sesgados poblacionalmente, lo que llevaría a unas conclusiones que no podrían extenderse más allá del reflejo del sesgo dominante.

Para realizar nuestra investigación seleccionar tomamos todos los films indexados en IMDb con más de 10.000 votos, lo que nos da un total de 7.569 películas. Con el universo seleccionado, para obtener un nivel de confianza del 99% con un margen de error del 5% debemos seleccionar un mínimo de 664 películas, por lo que consideramos la muestra representativa. De estos films extraemos los metadatos relacionados referentes a los datos poblacionales a los que IMDb nos da acceso, que son sexo y rango de edad.

Tras un análisis de las 7.569 más votadas detectamos que el 80,72% de los votos corresponden a población masculina y el 85,5% a los votantes que tienen una edad entre 18 y 45 años. Además, estos porcentajes de participación varían según el género cinematográfico del film, país de producción y año de estreno del film.

Los resultados obtenidos definen claramente como la votación de los films está determinada por la población masculina, mostrando un claro desequilibrio de género en la puntuación promediada asociada a cada film. De la misma forma que la población comprendida entre 18 y 45 años es la que tiene una mayor representatividad en el voto e influyen en mayor medida en la puntuación final asignada, dando poca representatividad a la opinión de votantes menores de 18 años o mayores de 45. Conforme la película es de producción más reciente se reducen las diferencias entre los porcentajes de votos

| pertenecientes a hombres y representatividad equilibrada | mujeres, a<br>entre géneros | pesar de | estar | todavía | muy | lejos | de | una |
|----------------------------------------------------------|-----------------------------|----------|-------|---------|-----|-------|----|-----|
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |
|                                                          |                             |          |       |         |     |       |    |     |