



# Proyecto final

BÚSQUEDA Y ANÁLISIS DE LA  
INFORMACIÓN (BAIN)

Alina Rojas Reynoso  
3º MAIS





# Episodios Nacionales: Primera Serie

Benito Pérez Galdós

# Objeto de estudio

1. **Trafalgar** (1873)
2. **La corte de Carlos IV** (1873)
3. **El 19 de marzo y el 2 de mayo** (1873)
4. **Bailén** (1873)
5. **Napoleón en Chamartín** (1874)
6. **Zaragoza** (1874)
7. **Gerona** (1874)
8. **Cádiz** (1874)
9. **Juan Martín el Empecinado** (1874)
10. **La batalla de los Arapiles** (1875)

The logo for Project Gutenberg, featuring the words "Project" and "Gutenberg" in a stylized, black, gothic-style font. The "P" in "Project" and the "G" in "Gutenberg" are larger and more ornate, with a red outline. The text is set against a light beige background.

# Planteamiento

## 1. **Técnicas de Text Mining**

1. Frecuencia de palabras con Wordcloud
2. TF-IDF
3. Ley de Zipf
  1. Ley de potencia
4. Topic Modeling
5. Análisis de Sentimiento

## 2. **SNA**

# Estructura de los datos

	text	book	line_number	chapter
1	Se me permitirá que antes de referir el gran suceso de que f...	Trafalgar	7	1
2	diga algunas palabras sobre mi infancia, explicando por qué...	Trafalgar	8	1
3	manera me llevaron los azares de la vida a presenciar la terri...	Trafalgar	9	1
4	catástrofe de nuestra marina.	Trafalgar	10	1
5	Al hablar de mi nacimiento, no imitaré a la mayor parte de l...	Trafalgar	11	1
6	cuentan hechos de su propia vida, quienes empiezan nombr...	Trafalgar	12	1
7	parentela, las más veces noble, siempre hidalga por lo men...	Trafalgar	13	1
8	dicen descendientes del mismo Emperador de Trapisonda. Y...	Trafalgar	14	1
9	parte, no puedo adornar mi libro con sonoros apellidos; y fu...	Trafalgar	15	1
10	madre, a quien conocí por poco tiempo, no tengo noticia d...	Trafalgar	16	1
11	mis ascendientes, si no es de Adán, cuyo parentesco me par...	Trafalgar	17	1
12	indiscutible. Doy principio, pues, a mi historia como Pablos, ...	Trafalgar	18	1
13	de Segovia: afortunadamente Dios ha querido que en esto s...	Trafalgar	19	1
14	parezcamos.	Trafalgar	20	1

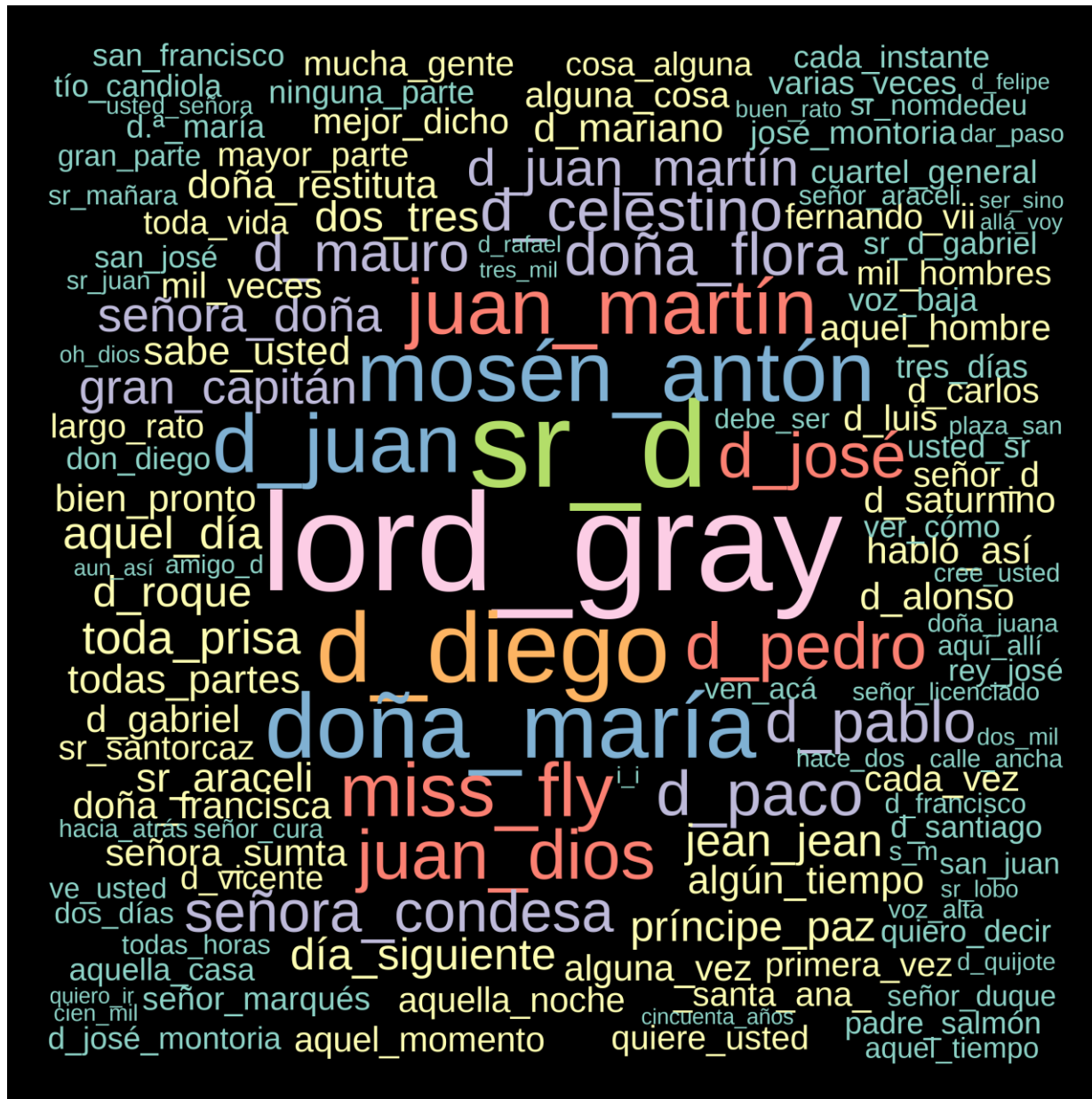
# Muestra

```
> summary(books_metadata_filtered)
```

text	book	line_number	chapter
Length:63617	Length:63617	Min. : 5	Min. : 1.00
Class :character	Class :character	1st Qu.:1656	1st Qu.: 8.00
Mode :character	Mode :character	Median :3254	Median :15.00
		Mean :3340	Mean :16.27
		3rd Qu.:4870	3rd Qu.:24.00
		Max. :8768	Max. :43.00

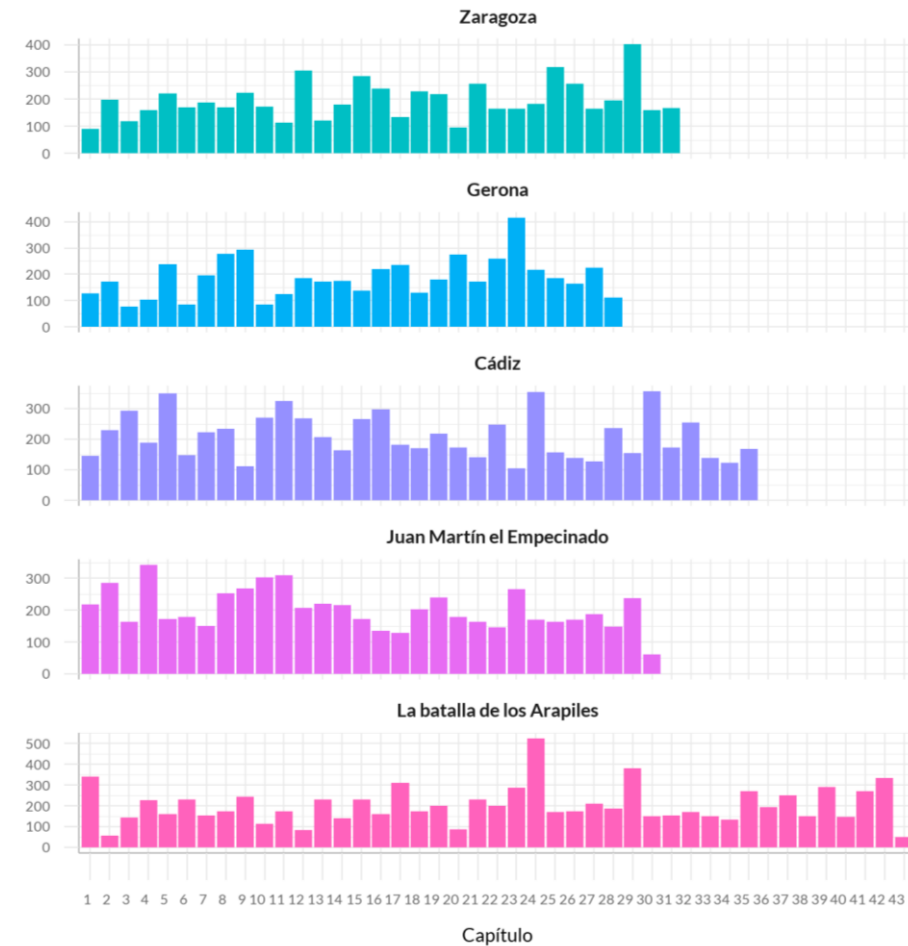
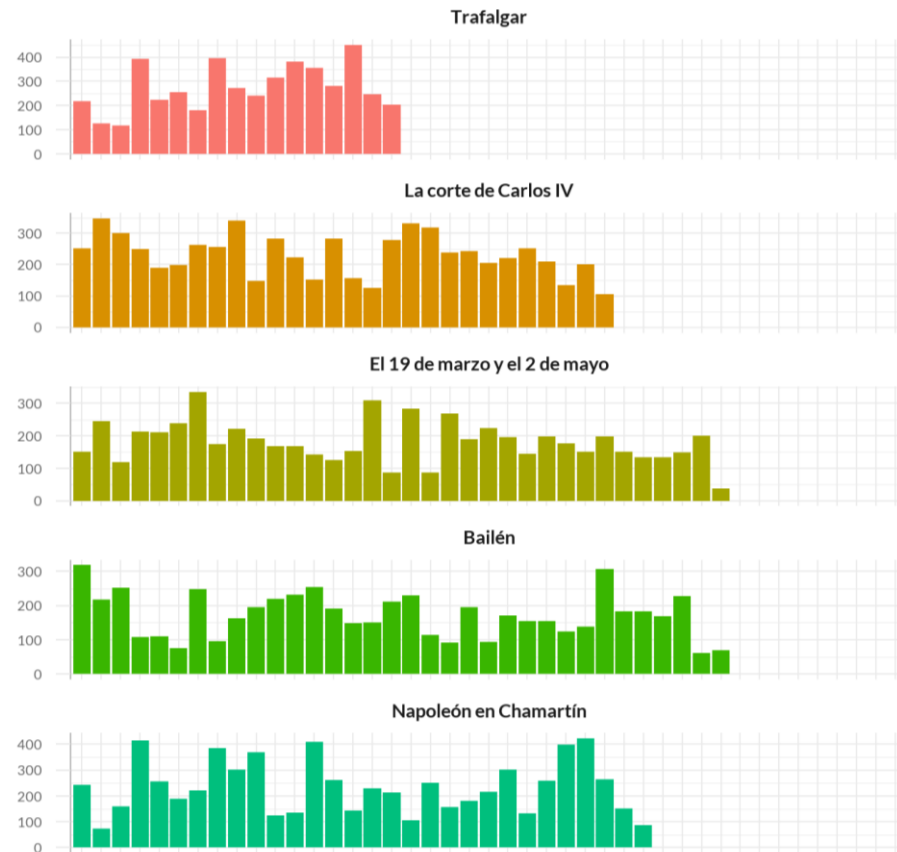
# Análisis de texto





Términos  
frecuentes

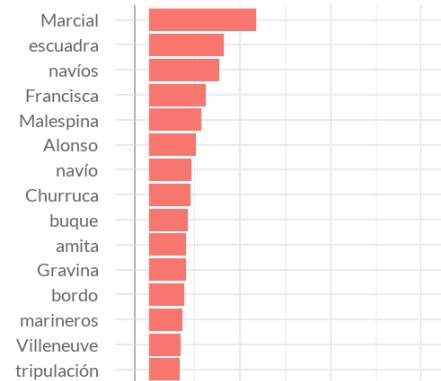




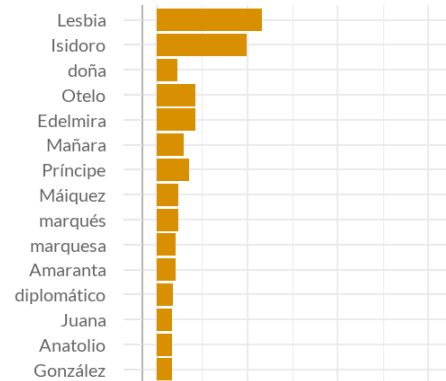
Distribución del número de líneas por capítulo y libro

# Palabras clave por libro según TF-IDF

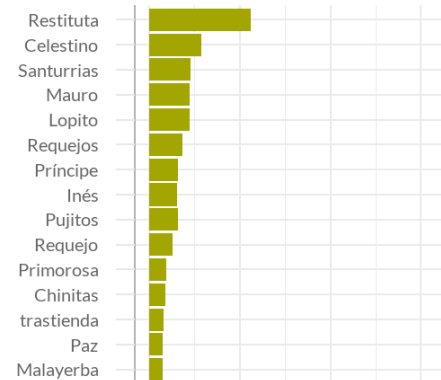
Trafalgar



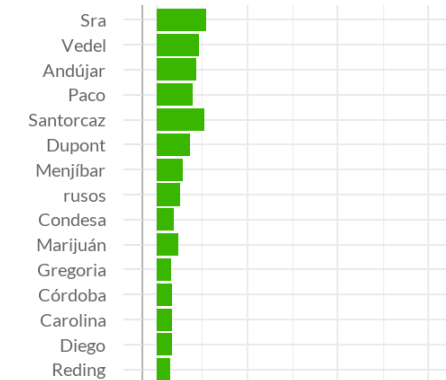
La corte de Carlos IV



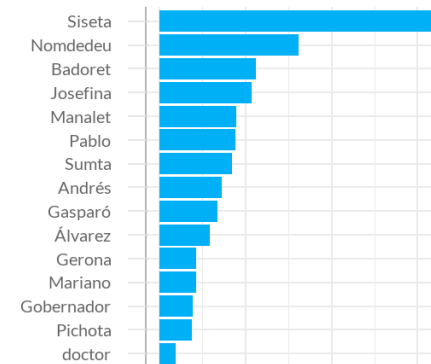
El 19 de marzo y el 2 de mayo



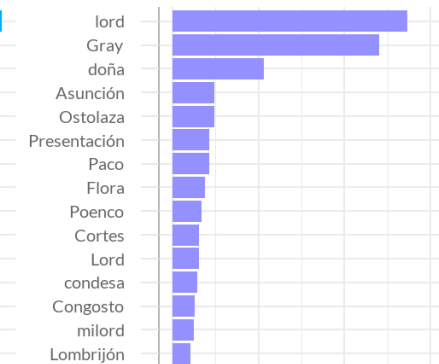
Bailén



Gerona



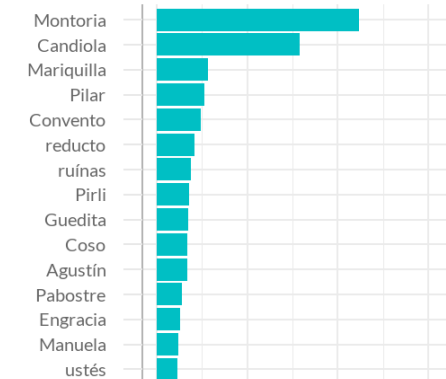
Cádiz



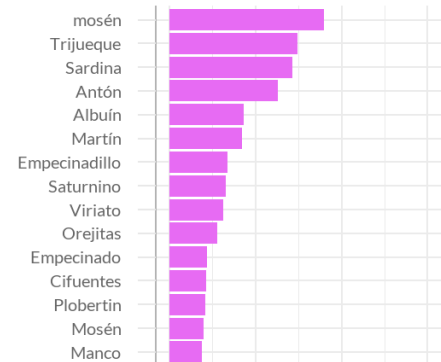
Napoleón en Chamartín



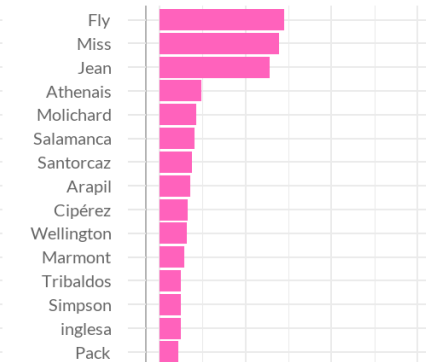
Zaragoza



Juan Martín el Empecinado



La batalla de los Arapiles



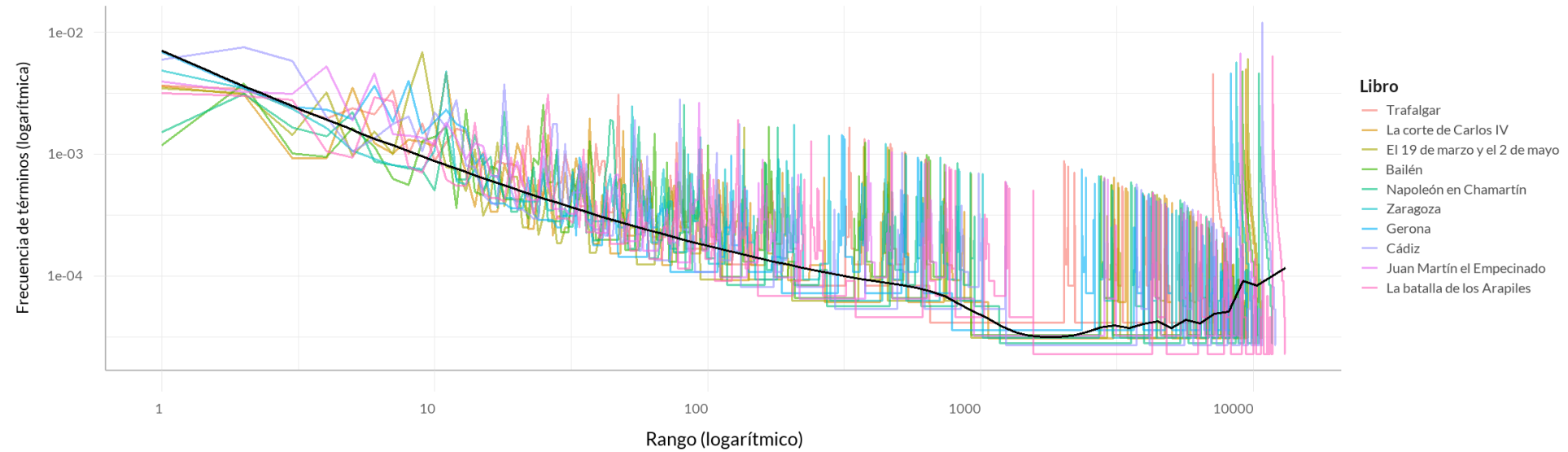
0.000 0.005 0.010 0.015

0.000 0.005 0.010 0.015

TF-IDF

## Distribución de Zipf por Libro

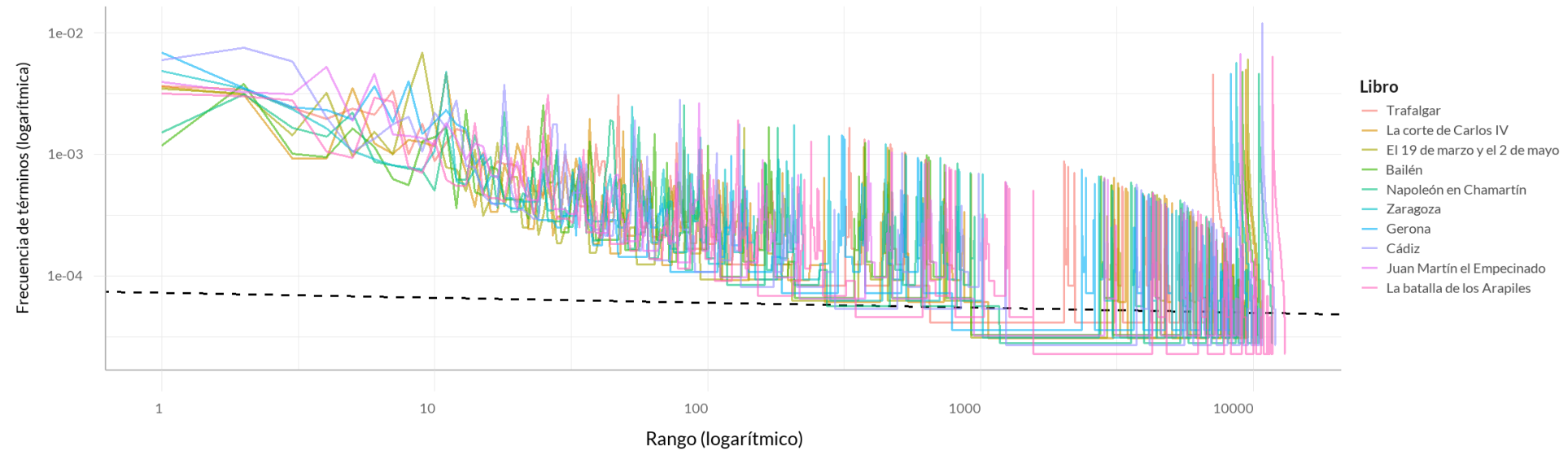
Análisis de los Episodios Nacionales: Primera Serie de Galdós



# Ley de Zipf

## Ley de potencia

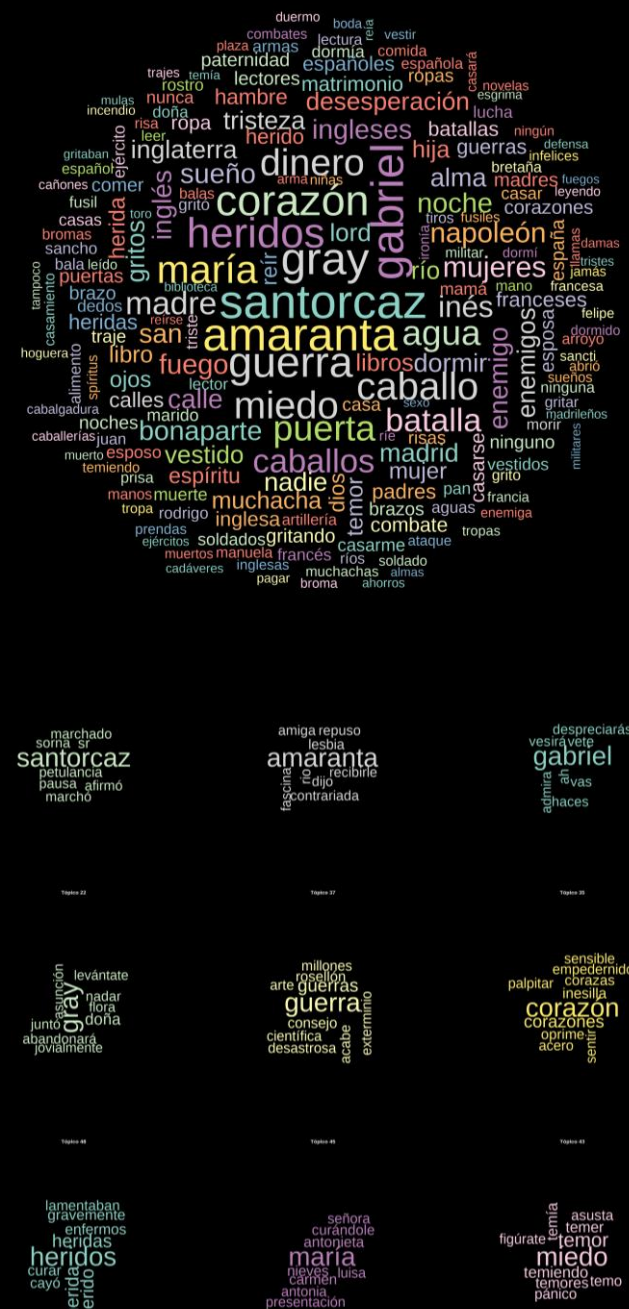
Análisis de los Episodios Nacionales: Primera Serie de Galdós



# Ley de Zipf

# Topic Models

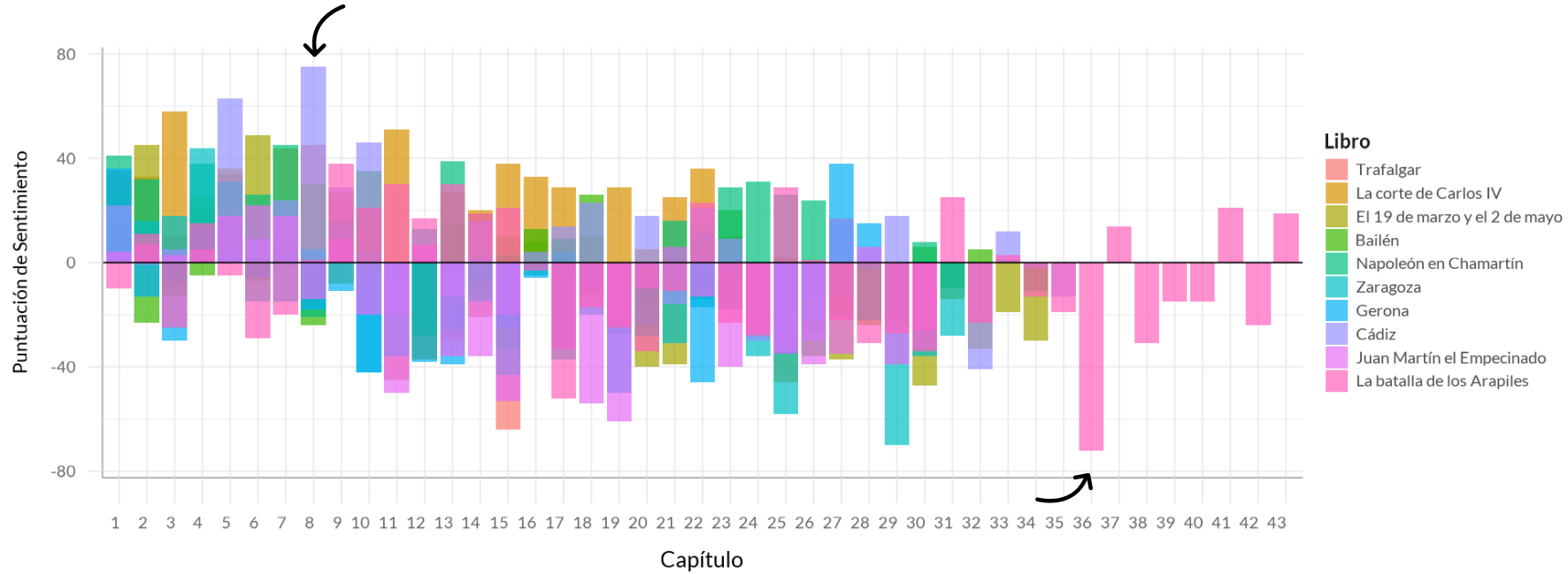
- Presencia de personajes principales
- Temática bélica
- Temática cristiana





## Análisis de Sentimiento a lo largo de los capítulos por cada libro

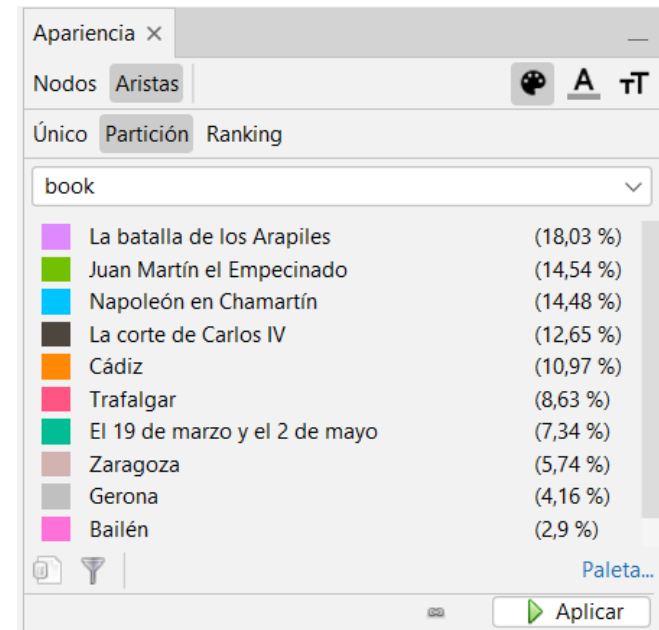
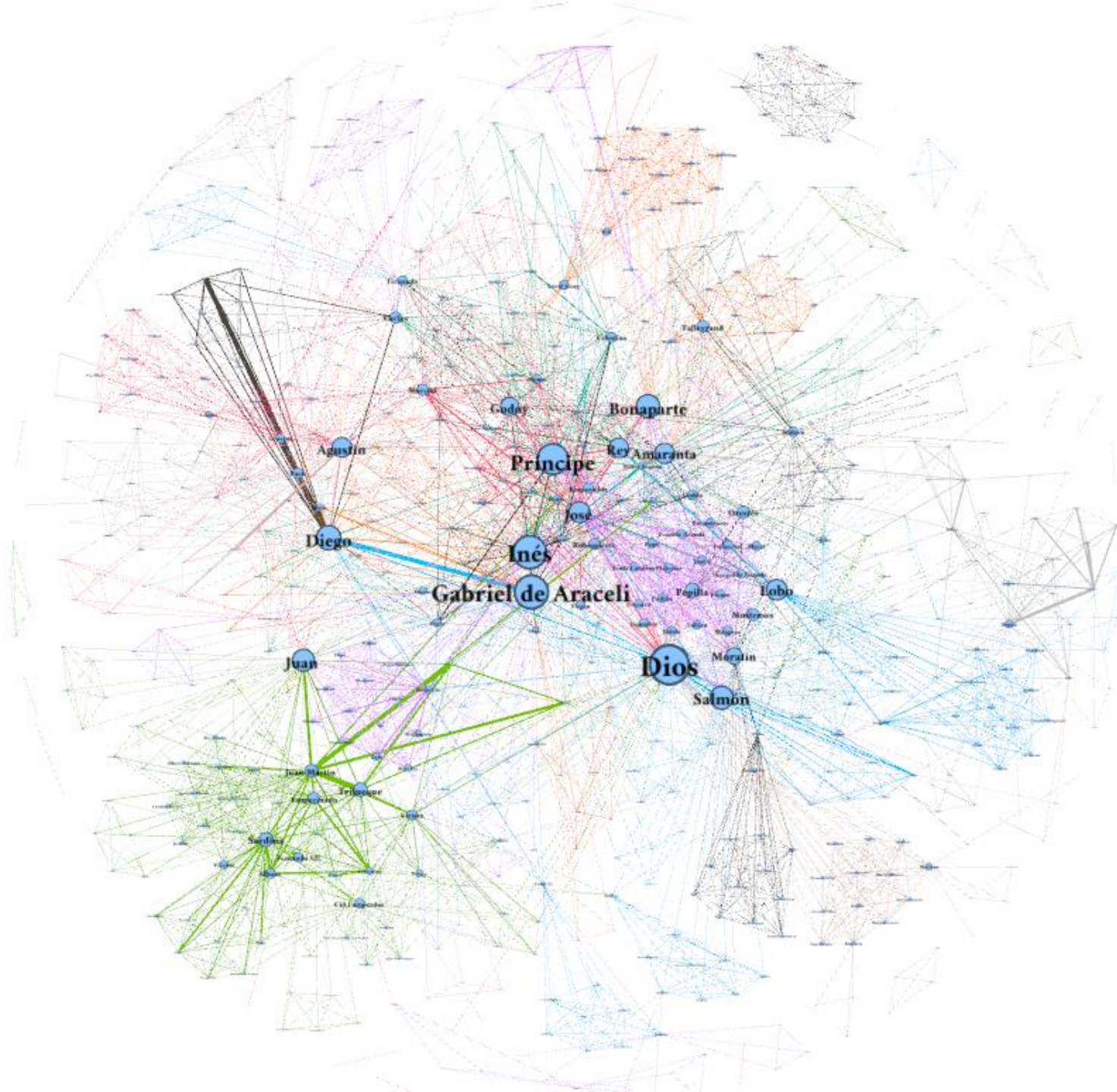
Análisis de los Episodios Nacionales: Primera Serie de Galdós



# Análisis de Sentimiento

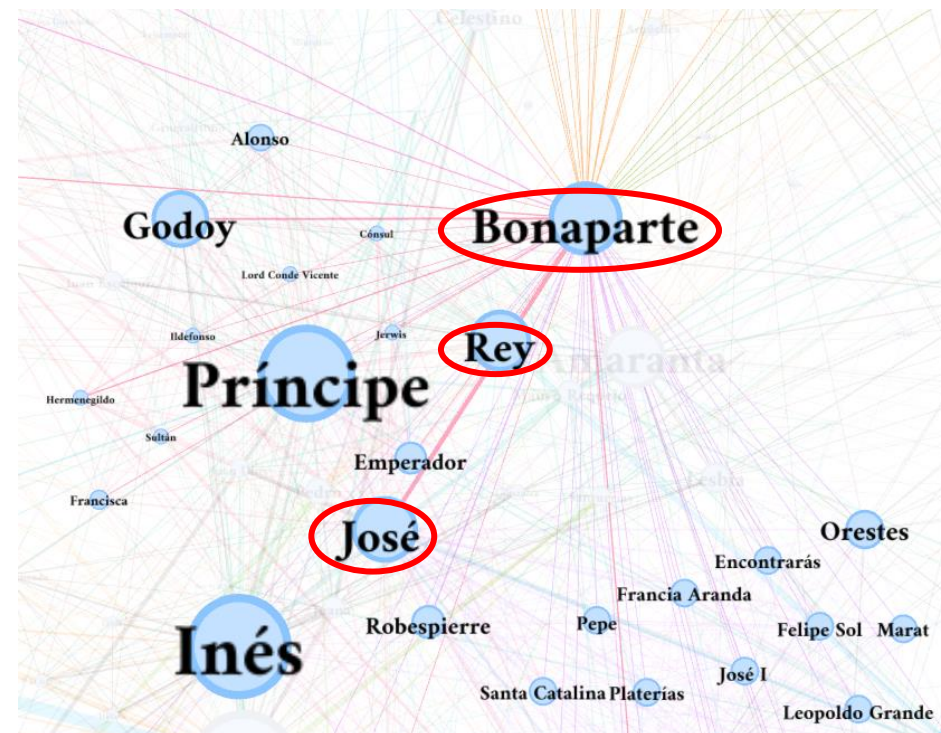
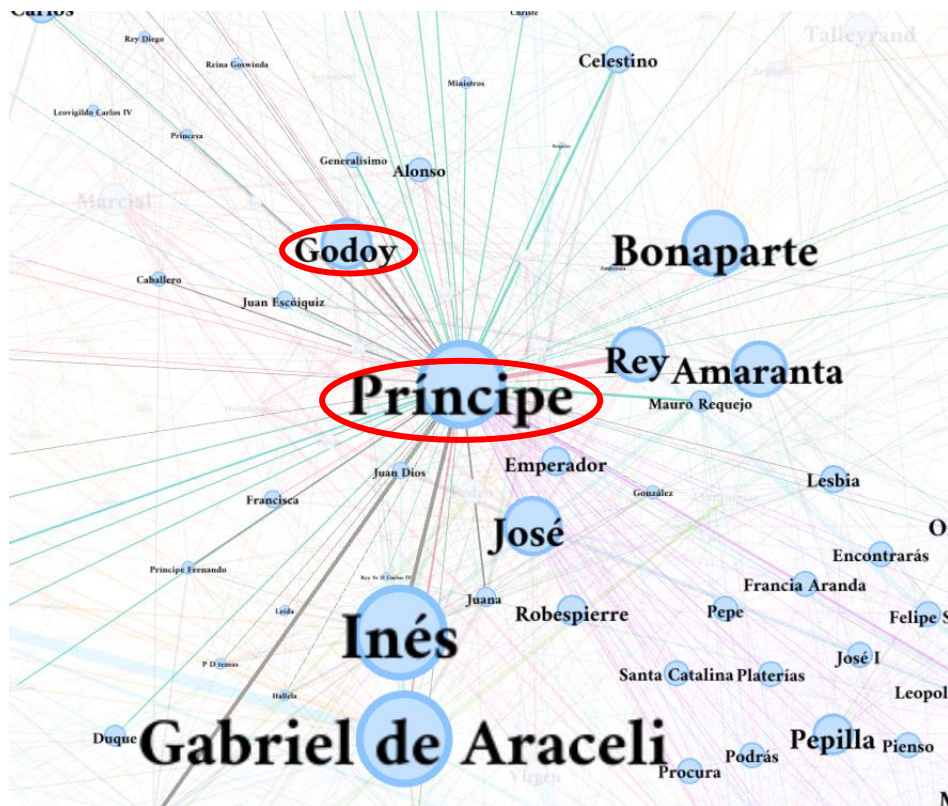
# Análisis de redes sociales

# Grafo





# Detalles del grafo





*Muchas  
gracias*