

# LDA Latent Dirichlet Allocation(LDA)

```
In [57]: # Topic modeling is a method for unsupervised classification of documents, similar to clustering on numeric data, whi  
# of items (topics) even when we're not sure what we're looking for. (https://towardsdatascience.com/Latent-dirichlet
```

```
In [5]: import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from sklearn.feature_extraction import text  
import nltk  
from nltk.corpus import stopwords  
from nltk.classify import SklearnClassifier  
  
import re  
from nltk.tokenize import wordpunct_tokenize  
  
from nltk.tokenize import RegexpTokenizer  
from nltk.stem.snowball import SnowballStemmer  
from nltk.stem.porter import PorterStemmer
```

```
C:\Users\farav\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn\rcmod.py:82: DeprecationWarning: di  
stutils Version classes are deprecated. Use packaging.version instead.  
    if LooseVersion(mpl.__version__) >= "3.0":  
C:\Users\farav\AppData\Local\Programs\Python\Python310\lib\site-packages\setuptools\_distutils\version.py:346: Depre  
cationWarning: distutils Version classes are deprecated. Use packaging.version instead.  
    other = LooseVersion(other)
```

```
In [1]: !pip install pyLDAvis
```

```

Requirement already satisfied: pyLDAvis in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (3.3.1)
Requirement already satisfied: gensim in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (4.3.0)
Requirement already satisfied: funcy in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (1.17)
Requirement already satisfied: jinja2 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (3.1.2)
Requirement already satisfied: pandas>=1.2.0 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (1.4.3)
Requirement already satisfied: sklearn in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (0.0.post2)
Requirement already satisfied: future in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (0.18.3)
Requirement already satisfied: numpy>=1.20.0 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (1.23.2)
Requirement already satisfied: scipy in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (1.9.0)
Requirement already satisfied: scikit-learn in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (1.1.2)
Requirement already satisfied: joblib in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (1.1.0)
Requirement already satisfied: setuptools in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (65.0.2)
Requirement already satisfied: numexpr in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyLDAvis) (2.8.4)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pandas>=1.2.0->pyLDAvis) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pandas>=1.2.0->pyLDAvis) (2022.2.1)
Requirement already satisfied: FuzzyTM>=0.4.0 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from gensim->pyLDAvis) (2.0.5)
Requirement already satisfied: Cython==0.29.32 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from gensim->pyLDAvis) (0.29.32)
Requirement already satisfied: smart-open>=1.8.1 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from gensim->pyLDAvis) (5.2.1)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from jinja2->pyLDAvis) (2.1.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from scikit-learn->pyLDAvis) (3.1.0)
Requirement already satisfied: pyfume in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from FuzzyTM>=0.4.0->gensim->pyLDAvis) (0.2.25)
Requirement already satisfied: six>=1.5 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from python-dateutil>=2.8.1->pandas>=1.2.0->pyLDAvis) (1.16.0)
Requirement already satisfied: simpful in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (2.9.0)
Requirement already satisfied: fst-pso in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (1.8.1)
Requirement already satisfied: miniful in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from fst-pso->pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (0.0.6)
Requirement already satisfied: requests in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from simpful->pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (2.28.1)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (2022.6.15)
Requirement already satisfied: idna<4,>=2.5 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (1.26.11)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0->gensim->pyLDAvis) (2.1.0)

```

In [3]: `from collections.abc import Mapping`

In [4]: `print(Mapping)`

```
<class 'collections.abc.Mapping'>
```

In [6]: `!pip install gensim`

```
Requirement already satisfied: gensim in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (3.8.1)
Requirement already satisfied: numpy>=1.11.3 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from gensim) (1.23.2)
Requirement already satisfied: scipy>=0.18.1 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from gensim) (1.9.0)
Requirement already satisfied: six>=1.5.0 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from gensim) (1.16.0)
Requirement already satisfied: smart_open>=1.8.1 in c:\users\farav\appdata\local\programs\python\python310\lib\site-packages (from gensim) (5.2.1)
```

```
In [6]: import gensim
from gensim import corpora
import pyLDAvis.gensim_models
```

```
In [7]: nltk.download('stopwords')
nltk.download('punkt')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\farav\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\farav\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

```
Out[7]: True
```

```
In [8]: base = pd.read_csv('C:/Users/farav/OneDrive/Documentos/Python Scripts/base_tit_senti_fecha.csv', encoding='iso8859_2')
```

```
In [9]: base_elobse = base[base['diario'] == 'elobse']
```

```
In [10]: base_elobse_fecha = base_elobse[base_elobse['fecha'] == '2021-07-23']
```

```
In [11]: stop_words = list(stopwords.words('spanish'))
espanol = pd.read_csv('spanish.txt', header=None)
espanol.columns = ['palabra']
list_espanol = list(espanol['palabra'])
stop_words.extend(list_espanol)

unique_stop_words = []
[unique_stop_words.append(word) for word in stop_words if word not in unique_stop_words]
len(unique_stop_words)
```

```
Out[11]: 615
```

```
In [12]: def limpiar_tokenizar(texto):
    porter_stemmer = PorterStemmer()
    nuevo_texto = texto.lower()
    nuevo_texto = re.sub("\d+", ' ', nuevo_texto)
    nuevo_texto = wordpunct_tokenize(nuevo_texto) # esto toma espacios, símbolos, comas, etc. y los tokeniza junto con los espacios
    nuevo_texto = [token for token in nuevo_texto if len(token) > 3]
    nuevo_texto = [token for token in nuevo_texto if token not in unique_stop_words]
    stemmers = [porter_stemmer.stem(word) for word in nuevo_texto]
    nuevo_texto = [stem for stem in stemmers if stem.isalpha() and len(stem) > 1]

    return(nuevo_texto)
```

```
<ipython>:4: DeprecationWarning: invalid escape sequence '\d'
<ipython>:4: DeprecationWarning: invalid escape sequence '\d'
C:\Users\farav\AppData\Local\Temp\ipykernel_4668\2669027614.py:4: DeprecationWarning: invalid escape sequence '\d'
    nuevo_texto = re.sub("\d+", ' ', nuevo_texto)
```

```
In [13]: base_elobse_fecha['tit_tokenizado'] = base_elobse_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))
tokenizado = list(base_elobse_fecha['tit_tokenizado'])
```

```
C:\Users\farav\AppData\Local\Temp\ipykernel_4668\3183837091.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
base_elobse_fecha['tit_tokenizado'] = base_elobse_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))
```

```
In [14]: dictionary = corpora.Dictionary(tokenizado)
dictionary.filter_extremes(no_below=3, keep_n=300)

In [15]: print(dictionary)
Dictionary<12 unique tokens: ['inauguración', 'juego', 'olímpico', 'tokio', 'cancha']...>

In [16]: corpus = [dictionary.doc2bow(text) for text in tokenizado]

In [17]: corpus[0][:10]

Out[17]: [(0, 1), (1, 1), (2, 1), (3, 1)]

In [18]: # Define the LDA model
ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics=5, id2word=dictionary, passes=10)# revisar los argumentos

# Save the topics and top 5 words
topics = ldamodel.print_topics(num_words=5)

# Print the results
for topic in topics:
    print(topic)

(0, '0.295*"juego" + 0.295*"olímpico" + 0.213*"tokio" + 0.131*"inauguración" + 0.009*"vacuna"')
(1, '0.444*"montevideo" + 0.338*"vacuna" + 0.022*"olímpico" + 0.022*"inauguración"')
(2, '0.280*"cancha" + 0.280*"peñarol" + 0.280*"clásico" + 0.018*"montevideo" + 0.018*"uruguay"')
(3, '0.380*"dosi" + 0.378*"pfizer" + 0.025*"olímpico" + 0.025*"juego" + 0.025*"peñarol"')
(4, '0.688*"uruguay" + 0.032*"tokio" + 0.031*"olímpico" + 0.031*"juego" + 0.029*"pfizer"')

In [21]: lda_display_elobse = pyLDAvis.gensim_models.prepare(ldamodel, corpus, dictionary, sort_topics=False)

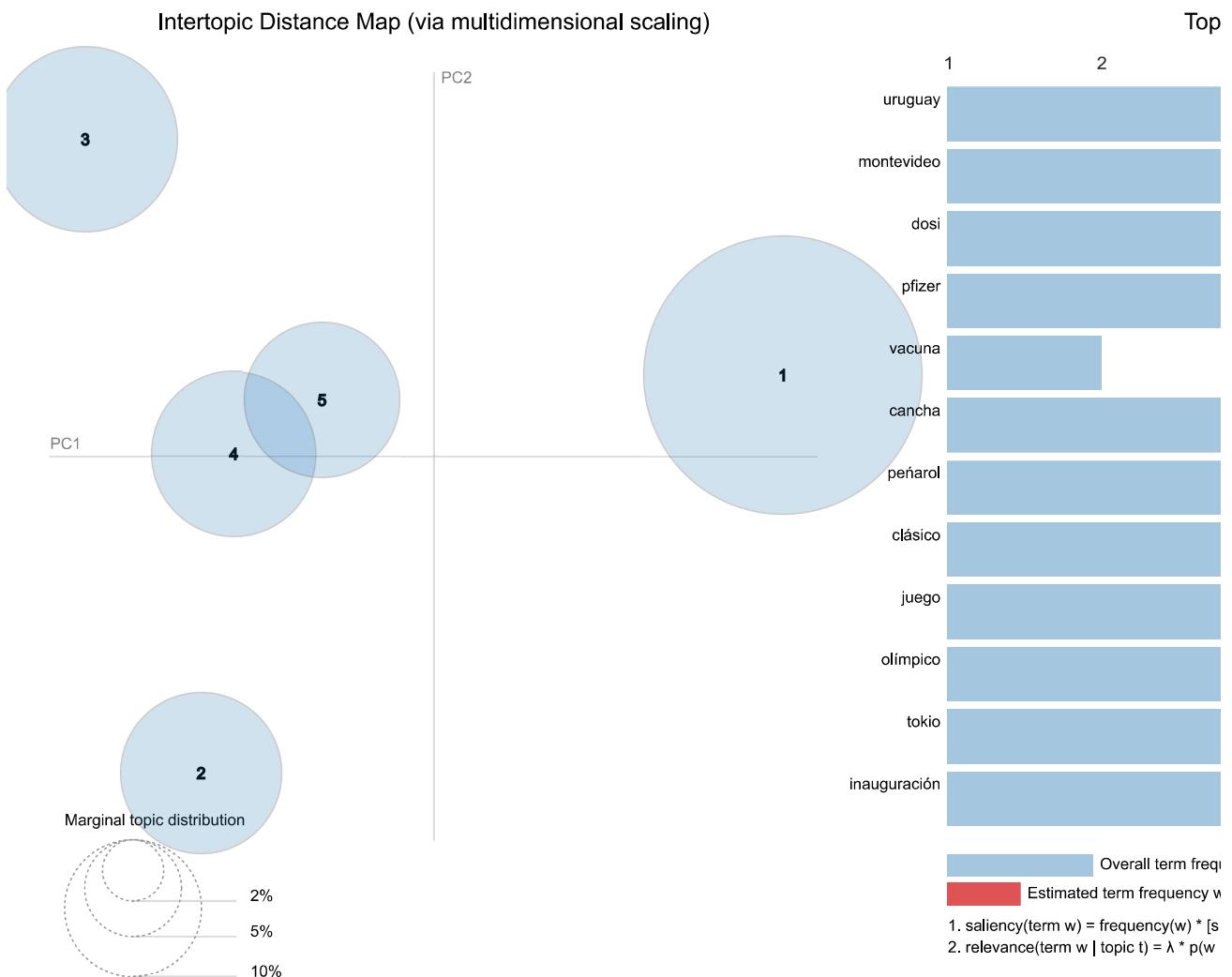
C:\Users\farav\AppData\Local\Programs\Python\Python310\lib\site-packages\pyLDAvis\_prepare.py:247: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only.
    by='saliency', ascending=False).head(R).drop('saliency', 1)

In [22]: pyLDAvis.display(lda_display_elobse)
```

Out[22]:

Selected Topic: 0    [Previous Topic](#) [Next Topic](#) [Clear Topic](#)

Slide to adjust relevance metric (2)

 $\lambda = 1$ 

In [23]:

```
base_ladiar = base[base['diario'] == 'ladiar']
base_ladiar_fecha = base_ladiar[base_ladiar['fecha'] == '2021-07-23']
```

In [24]:

```
base_ladiar_fecha['tit_tokenizado_1'] = base_ladiar_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))
tokenizado_1 = list(base_ladiar_fecha['tit_tokenizado_1'])
```

C:\Users\farav\AppData\Local\Temp\ipykernel\_4668\3857919374.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
base_ladiar_fecha['tit_tokenizado_1'] = base_ladiar_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))
```

In [27]:

```
dictionary_1 = corpora.Dictionary(tokenizado_1)
dictionary_1.filter_extremes(no_below=3, keep_n=300)
```

In [28]:

```
print(dictionary_1)
```

Dictionary<20 unique tokens: ['canelon', 'juego', 'olímpico', 'tokio', 'estafa']...>

In [29]:

```
corpus_1 = [dictionary_1.doc2bow(text) for text in tokenizado_1]
```

```
In [30]: # Define the LDA model
ldamodel_1 = gensim.models.ldamodel.LdaModel(corpus_1, num_topics=5, id2word=dictionary_1, passes=10)

# Save the topics and top 5 words
topics = ldamodel_1.print_topics(num_words=5)

# Print the results
for topic in topics:
    print(topic)

(0, '0.270*"uruguay" + 0.269*"pandemia" + 0.182*"tiempo" + 0.095*"apunt" + 0.052*"uruguayo"')
(1, '0.248*"juego" + 0.208*"olímpico" + 0.128*"tokio" + 0.128*"estafa" + 0.127*"coronaviru"')
(2, '0.184*"gobierno" + 0.184*"mujer" + 0.100*"coronaviru" + 0.100*"uruguayo" + 0.100*"tokio"')
(3, '0.247*"gigant" + 0.200*"número" + 0.152*"abierto" + 0.152*"cabildo" + 0.057*"apunt"')
(4, '0.169*"falta" + 0.169*"canelon" + 0.169*"política" + 0.168*"niño" + 0.063*"apunt")
```

```
In [31]: lda_display_ladiar = pyLDAvis.gensim_models.prepare(ldamodel_1, corpus_1, dictionary_1, sort_topics=False)
```

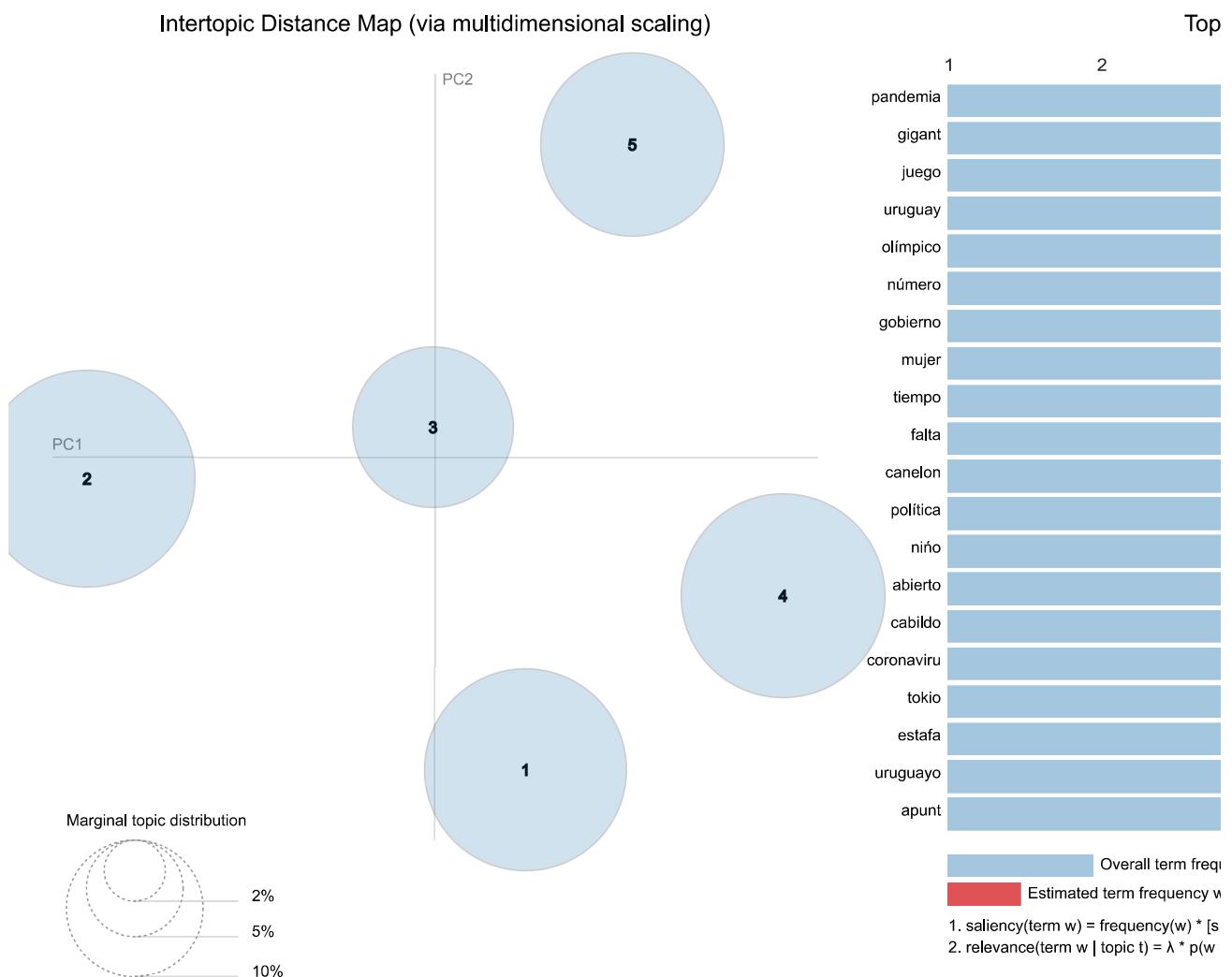
```
C:\Users\farav\AppData\Local\Programs\Python\Python310\lib\site-packages\pyLDAvis\_prepare.py:247: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only.
  by='saliency', ascending=False).head(R).drop('saliency', 1)
```

```
In [32]: pyLDAvis.display(lda_display_ladiar)
```

Out[32]:

Selected Topic: 0    [Previous Topic](#) [Next Topic](#) [Clear Topic](#)

Slide to adjust relevance metric (2)

 $\lambda = 1$ 

In [33]:

```
base_elpais = base[base['diario'] == 'elpais']
base_elpais_fecha = base_elpais[base_elpais['fecha'] == '2021-07-23']
```

In [34]:

```
base_elpais_fecha['tit_tokenizado_2'] = base_elpais_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))
tokenizado_2 = list(base_elpais_fecha['tit_tokenizado_2'])
```

C:\Users\farav\AppData\Local\Temp\ipykernel\_4668\494036624.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)  
`base_elpais_fecha['tit_tokenizado_2'] = base_elpais_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))`

In [35]:

```
dictionary_2 = corpora.Dictionary(tokenizado_2)
dictionary_2.filter_extremes(no_below=3, keep_n=300)
```

In [36]:

```
print(dictionary_2)
```

Dictionary<10 unique tokens: ['peñarol', 'juego', 'olímpico', 'uruguay', 'netflix']...>

In [37]:

```
corpus_2 = [dictionary_2.doc2bow(text) for text in tokenizado_2]
```

```
In [38]: # Define the LDA model
ldamodel_2 = gensim.models.ldamodel.LdaModel(corpus_2, num_topics=5, id2word=dictionary_2, passes=10)

# Save the topics and top 5 words
topics = ldamodel_2.print_topics(num_words=5)

# Print the results
for topic in topics:
    print(topic)

(0, '0.398*"marca" + 0.398*"vuelta" + 0.029*"uruguay" + 0.026*"videojuego" + 0.025*"baja"')
(1, '0.266*"netflix" + 0.266*"peñarol" + 0.265*"baja" + 0.099*"videojuego" + 0.020*"clasificación"')
(2, '0.527*"clasificación" + 0.200*"vuelta" + 0.036*"uruguay" + 0.035*"baja" + 0.034*"peñarol"')
(3, '0.317*"olímpico" + 0.317*"juego" + 0.218*"videojuego" + 0.030*"uruguay" + 0.020*"baja"')
(4, '0.736*"uruguay" + 0.030*"vuelta" + 0.029*"juego" + 0.029*"olímpico" + 0.029*"baja"')

In [39]: lda_display_elpais = pyLDAvis.gensim_models.prepare(ldamodel_2, corpus_2, dictionary_2, sort_topics=False)

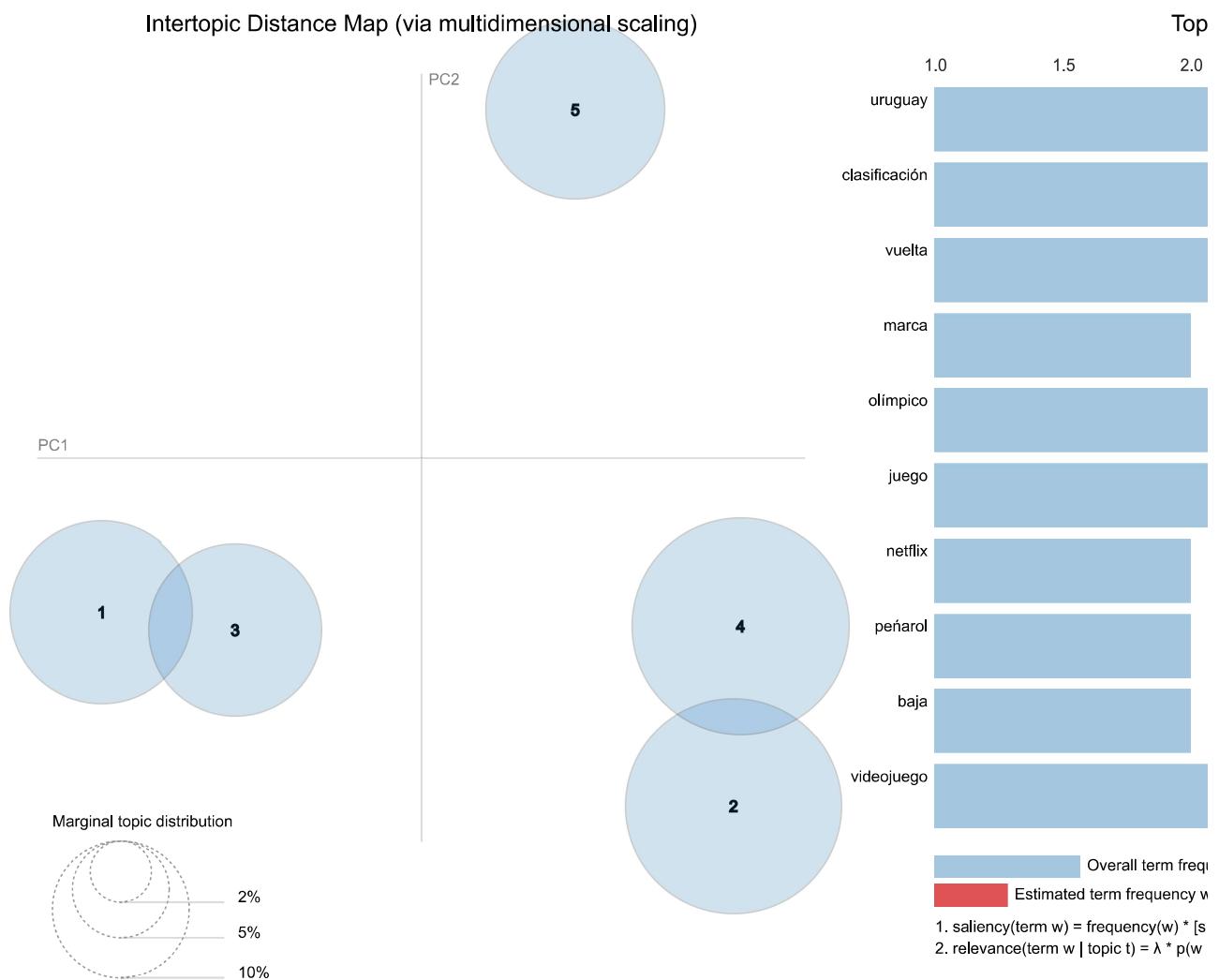
C:\Users\farav\AppData\Local\Programs\Python\Python310\lib\site-packages\pyLDAvis\_prepare.py:247: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only.
  by='salience', ascending=False).head(R).drop('salience', 1)

In [40]: pyLDAvis.display(lda_display_elpais)
```

Out[40]: Selected Topic: 0    Previous Topic    Next Topic    Clear Topic

Slide to adjust relevance metric (2)

$\lambda = 1$



```
In [41]: base_montev = base[base['diario'] == 'montev']
base_montev_fecha = base_montev[base_montev['fecha'] == '2021-07-23']

In [42]: base_montev_fecha['tit_tokenizado_3'] = base_montev_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))
tokenizado_3 = list(base_montev_fecha['tit_tokenizado_3'])

C:\Users\farav\AppData\Local\Temp\ipykernel_4668\1743992655.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    base_montev_fecha['tit_tokenizado_3'] = base_montev_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))

In [43]: dictionary_3 = corpora.Dictionary(tokenizado_3)
dictionary_3.filter_extremes(no_below=3, keep_n=300)

In [44]: print(dictionary_3)

Dictionary<10 unique tokens: ['uruguay', 'clásico', 'nacion', 'peñarol', 'juego']...>

In [45]: corpus_3 = [dictionary_3.doc2bow(text) for text in tokenizado_3]
```

```
In [46]: # Define the LDA model
ldamodel_3 = gensim.models.ldamodel.LdaModel(corpus_3, num_topics=5, id2word=dictionary_3, passes=10)

# Save the topics and top 5 words
topics = ldamodel_3.print_topics(num_words=5)

# Print the results
for topic in topics:
    print(topic)

(0, '0.607*"delta" + 0.056*"peñarol" + 0.055*"sudamericana" + 0.042*"juego" + 0.041*"uruguay"')
(1, '0.501*"sudamericana" + 0.213*"nacion" + 0.039*"delta" + 0.036*"peñarol" + 0.035*"hombr"')
(2, '0.617*"juego" + 0.062*"opinión" + 0.047*"cuba" + 0.042*"nacion" + 0.039*"hombr"')
(3, '0.323*"uruguay" + 0.314*"opinión" + 0.243*"cuba" + 0.025*"delta" + 0.016*"nacion"')
(4, '0.310*"nacion" + 0.221*"peñarol" + 0.181*"clásico" + 0.138*"hombr" + 0.107*"sudamericana")
```

```
In [47]: lda_display_montev = pyLDAvis.gensim_models.prepare(ldamodel_3, corpus_3, dictionary_3, sort_topics=False)
```

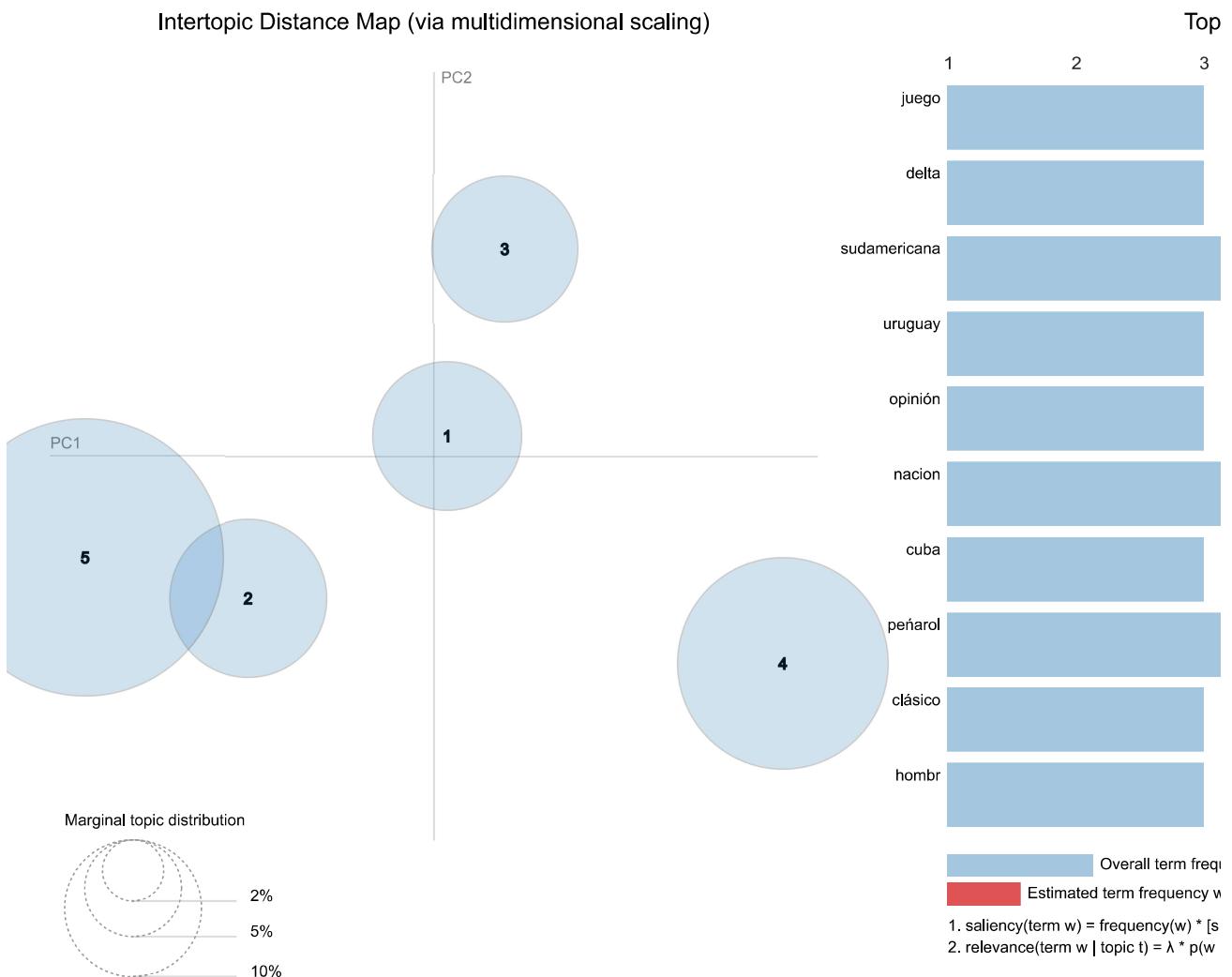
C:\Users\farav\AppData\Local\Programs\Python\Python310\lib\site-packages\pyLDAvis\\_prepare.py:247: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only.  
by='saliency', ascending=False).head(R).drop('saliency', 1)

```
In [48]: pyLDAvis.display(lda_display_montev)
```

Out[48]: Selected Topic: 0    Previous Topic    Next Topic    Clear Topic

Slide to adjust relevance metric (2)

$\lambda = 1$



```
In [49]: base_republ = base[base['diario'] == 'republ']
base_republ_fecha = base_republ[base_republ['fecha'] == '2021-07-23']

In [50]: base_republ_fecha['tit_tokenizado_4'] = base_republ_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))
tokenizado_4 = list(base_republ_fecha['tit_tokenizado_4'])

C:\Users\farav\AppData\Local\Temp\ipykernel_4668\881438736.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    base_republ_fecha['tit_tokenizado_4'] = base_republ_fecha['titular'].apply(lambda x: limpiar_tokenizar(x))

In [51]: dictionary_4 = corpora.Dictionary(tokenizado_4)
dictionary_4.filter_extremes(no_below=3, keep_n=300)

In [52]: print(dictionary_4)
Dictionary<3 unique tokens: ['juego', 'olímpico', 'mujer']>

In [53]: corpus_4 = [dictionary_4.doc2bow(text) for text in tokenizado_4]
```

```
In [54]: # Define the LDA model
ldamodel_4 = gensim.models.ldamodel.LdaModel(corpus_4, num_topics=5, id2word=dictionary_4, passes=10)

# Save the topics and top 5 words
topics = ldamodel_4.print_topics(num_words=5)

# Print the results
for topic in topics:
    print(topic)

(0, '0.334*"mujer" + 0.333*"juego" + 0.333*"olímpico"')
(1, '0.334*"mujer" + 0.333*"juego" + 0.333*"olímpico"')
(2, '0.884*"mujer" + 0.058*"juego" + 0.058*"olímpico"')
(3, '0.345*"olímpico" + 0.345*"juego" + 0.311*"mujer"')
(4, '0.485*"olímpico" + 0.485*"juego" + 0.031*"mujer")
```

```
In [55]: lda_display_republ = pyLDAvis.gensim_models.prepare(ldamodel_4, corpus_4, dictionary_4, sort_topics=False)
```

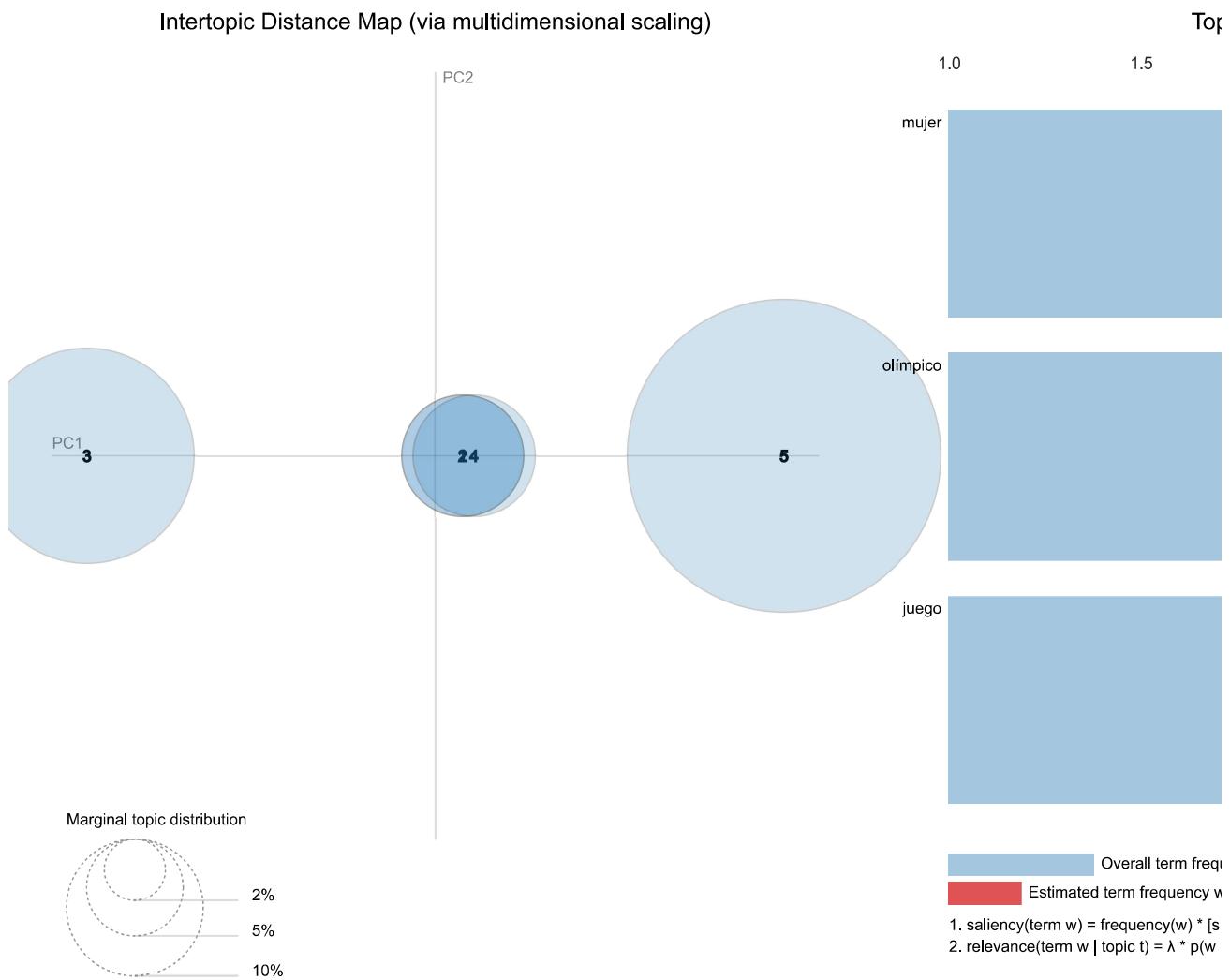
```
C:\Users\farav\AppData\Local\Programs\Python\Python310\lib\site-packages\pyLDAvis\_prepare.py:247: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only.
    by='saliency', ascending=False).head(R).drop('saliency', 1)
```

```
In [56]: pyLDAvis.display(lda_display_republ)
```

Out[56]: Selected Topic: 0    [Previous Topic](#) [Next Topic](#) [Clear Topic](#)

Slide to adjust relevance metric (2)

$\lambda = 1$



In [ ]: