

# Análisis de las menciones a países en las noticias de prensa chilena sobre inmigración: Un enfoque basado en grafos para visualizar las relaciones y cambios a lo largo del tiempo

Cristian Cofré Sepúlveda

26 de junio de 2023

## 1. Introducción

En la era de la información y la comunicación, los medios de prensa desempeñan un papel fundamental en la difusión de noticias y la formación de opiniones en la sociedad. El estudio de los medios de prensa y su influencia se ha vuelto cada vez más relevante en diversos campos, como la sociología, la comunicación y la ciencia política. En particular, el análisis de grafos ha demostrado ser una herramienta poderosa para comprender y visualizar las complejas redes de relaciones presentes en el mundo de los medios de prensa.

En este contexto, el presente trabajo se centra en el estudio de las noticias de prensa chilena que abordan el tema de la inmigración, utilizando técnicas de análisis de grafos. La inmigración es un fenómeno global que ha generado debates y desafíos en diversas sociedades, y su cobertura mediática desempeña un papel crucial en la formación de percepciones y actitudes hacia los inmigrantes.

El objetivo principal de este estudio es analizar la evolución de las menciones a países en las noticias de prensa chilena sobre inmigración y visualizar las relaciones entre ellos mediante la construcción de un grafo. Los grafos son estructuras que representan las interconexiones entre elementos, y su aplicación en el análisis de medios de prensa permite identificar patrones, agrupamientos y cambios en las relaciones entre países a lo largo del tiempo. Además, el análisis de grafos nos permite identificar agrupamientos geográficos, destacando las conexiones entre países vecinos o de una misma región.

## 2. Hipótesis

Se plantea la hipótesis de que existirán subgrafos dentro del grafo principal que reflejen agrupamientos geográficos de los países mencionados en las noticias. Además, se espera que estos subgrafos evolucionen con el tiempo, mostrando cambios en las conexiones entre los países y la incorporación o exclusión de algunos de ellos.

### 2.1. Objetivos del estudio

- Analizar la evolución de las menciones a países en noticias de prensa chilena relacionadas con el tema de la inmigración.
- Identificar y visualizar agrupamientos geográficos de los países mencionados en estas noticias.
- Investigar los cambios en las conexiones entre países y la incorporación o exclusión de países en los agrupamientos a lo largo del tiempo.

### 2.2. Preguntas de investigación

- ¿Qué países han sido mencionados en las noticias de prensa chilena que tratan sobre inmigración desde 2016 hasta 2021?
- ¿Existen agrupamientos geográficos de países en el grafo de menciones?
- ¿Cómo han evolucionado estos agrupamientos a lo largo del tiempo?
- ¿Cuáles son los países que se han agregado o excluido de estos agrupamientos durante el periodo de estudio?

## 3. Descripción del Dataset

En este anexo se proporciona una descripción del dataset utilizado en este estudio, así como los métodos empleados para su construcción y limpieza. Además, se discuten las limitaciones asociadas a los datos y las consideraciones importantes a tener en cuenta al interpretar los resultados.

El dataset utilizado en este estudio consiste en una recopilación de noticias de prensa Chilena en que se menciona la palabra "inmigración", abarcando el período comprendido entre 2015 y 2023. Cada noticia se recopiló a partir de medios de prensa Chilenos. Se registraron diversos atributos para cada noticia, como la fecha de publicación, el título, el contenido y el medio de comunicación.

Se llevó a cabo un proceso de limpieza y preprocesamiento de los datos. Esto implicó la eliminación de duplicados, la eliminación de valores nulos y de espacios en blancos innecesarios.

Es importante tener en cuenta algunas limitaciones asociadas al dataset utilizado en este estudio. En primer lugar, debido a la naturaleza de la recopilación

de noticias, es posible que exista un sesgo inherente en la selección de las fuentes de información y la palabra clave utilizada ("inmigración "). Esto podría influir en la representatividad del conjunto de datos y en la interpretación de los resultados.

Se realizó un procesamiento para determinar la cantidad de noticias publicadas por cada medio de prensa en el dataset, lo que se representa en la Figura 1 y de forma similar se calculó la cantidad de noticias publicadas por año, la cual se ilustra en la Figura 2. Es importante señalar que todas las noticias contienen la palabra "inmigración".

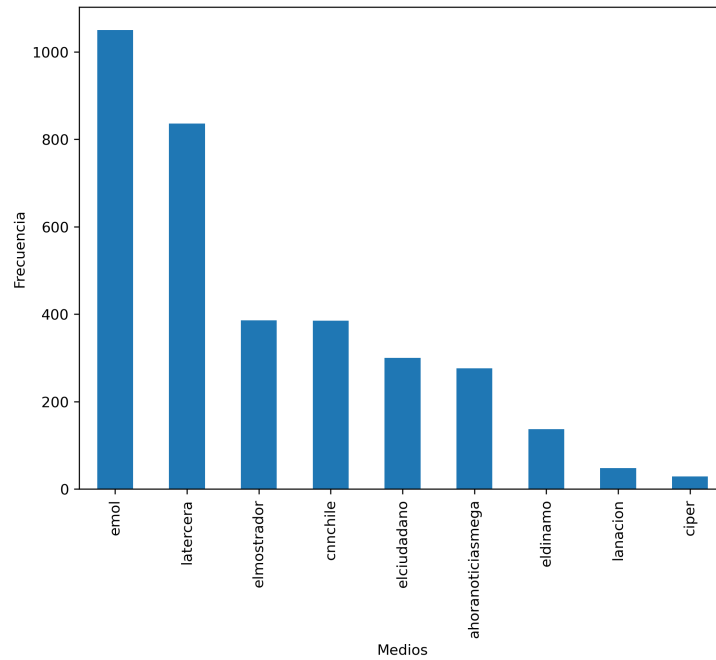


Figura 1: Cantidad de noticias publicadas por cada medio de prensa

## 4. Metodología

En esta sección, se explica el proceso para obtener los países mencionados en cada noticia prensa utilizando una lista de países y sus respectivos gentilicios en singular y plural. Este procedimiento se presenta de manera secuencial, siguiendo una serie de pasos. También se abordará el algoritmo de Louvain Communities utilizado en el estudio.

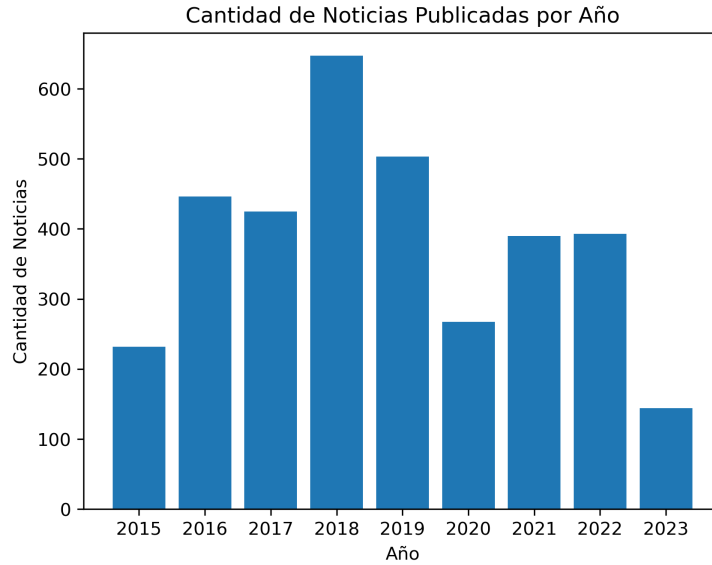


Figura 2: Cantidad de noticias publicadas por año

#### 4.1. Extracción de países mencionados en las noticias

- Se recopiló una lista de países, de los gentilicios en singular de cada país y de sus gentilicios en plural. Esto se utilizará como referencia para identificar los países mencionados en las noticias.
- Se crearon dos diccionarios que contienen los gentilicios en singular y plural de los países con el respectivo nombre del país. Esto permitirá realizar transformaciones necesarias en las menciones de los países para asegurar una consistencia en los datos.
- Para cada noticia en el dataset, se realizó un procesamiento de texto para identificar las menciones de países. Si se encontró una coincidencia entre el texto de la noticia y algún país de la lista, se registró el país mencionado en una columna adicional junto a la noticia.
- De manera similar al paso anterior se realizó un procesamiento de texto y si se encontró una mención en forma de gentilicio en singular o plural, se utilizó el diccionario previamente creado para transformar la mención al nombre del país correspondiente.

#### 4.2. Algoritmo de Louvain Communities

El algoritmo de Louvain Communities es una técnica de detección de comunidades en grafos. Su objetivo es identificar grupos o agrupamientos de nodos que estén fuertemente interconectados entre sí y que tengan conexiones más

débiles con nodos externos al grupo. Este algoritmo utiliza una estrategia de optimización iterativa para maximizar la modularidad del grafo, es decir, la calidad de la partición en comunidades. En cada iteración, los nodos se reasignan a comunidades de manera que se maximice la ganancia en modularidad. Este proceso se repite hasta que se alcanza un estado de estabilidad en el que no hay mejoras significativas. Como resultado, se obtienen comunidades o subgrafos que representan agrupamientos de nodos con fuertes conexiones internas.

## 5. Resultados y Discusión

Al aplicar la el método de extracción de países, se obtuvo un nuevo dataset donde una columna muestra el identificador de la noticia y otra columna contiene los países mencionados en cada noticia. Con este nuevo dataset fue posible obtener los países que tienen recurrencia en las noticias de prensa, esto se visualiza en la Figura 3.

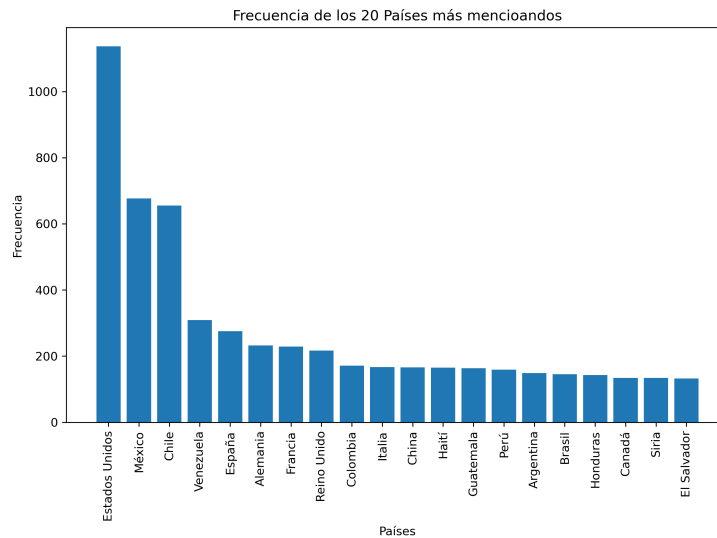


Figura 3: 20 países mas mencionados en noticias de prensa

En base al nuevo dataset obtenido, se construyó un grafo en el que los nodos representan los países y las conexiones entre ellos se establecen en función de su aparición conjunta en las noticias. Es decir, si dos países son mencionados en la misma noticia, se crea una conexión entre sus respectivos nodos en el grafo.

El "peso" de la conexión entre los nodos depende de la frecuencia con la que los países son mencionados juntos en las noticias. Si dos países son mencionados varias veces en diferentes noticias, el peso de la conexión entre sus nodos aumenta. El grafo resultante se muestra en la Figura 4.

A continuación, se procedió a identificar los pares de países que se encuentran

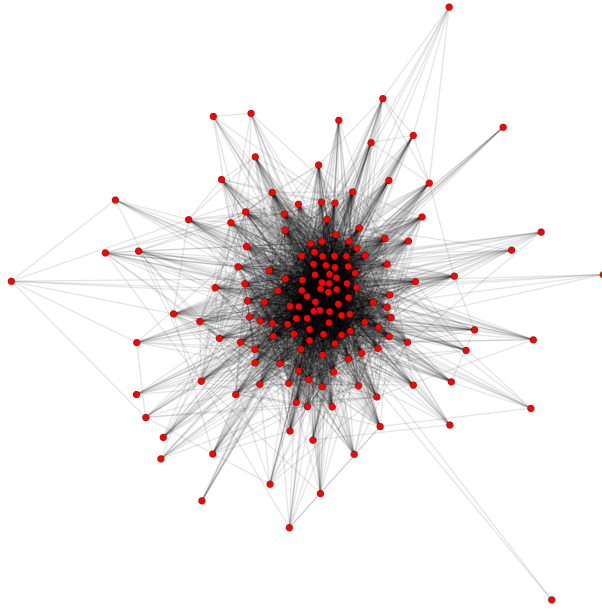


Figura 4: Grafo resultante que muestra cómo los países están interconectados en función de sus menciones en las noticias.

a una distancia menor a 5000 kilómetros y cuya conexión en el grafo original tenía un peso mayor a 5. Estos pares de países cumplen con dos criterios: proximidad geográfica y una relación más significativa basada en la frecuencia de sus menciones en las noticias.

Después de identificar los pares de países, se procedió a crear un grafo adicional. En este nuevo grafo, cada país se representa como un nodo, y las conexiones entre los países se establecen en base a su proximidad geográfica y la intensidad de su relación, la cual se determina por el peso de la conexión en el grafo original. De esta manera, se obtiene una representación visual de la estructura y las interacciones entre los países seleccionados. Para lograr visualizar el nuevo grafo generado, se obtuvieron las coordenadas de los países involucrados para posteriormente marcarlas en un mapa del planeta y también se mantuvieron las conexiones entre los países. Lo anterior se ilustra en la Figura 5.

Utilizando el algoritmo de detección de la comunidad de Louvain, se llevó a cabo un análisis más profundo del grafo anterior. Al aplicar la función el algoritmo al grafo, se logró separar el conjunto de países en diferentes comunidades. Esto permitió observar la formación de subgrafos que agrupan países con relaciones más fuertes entre sí en términos de menciones conjuntas en las noticias

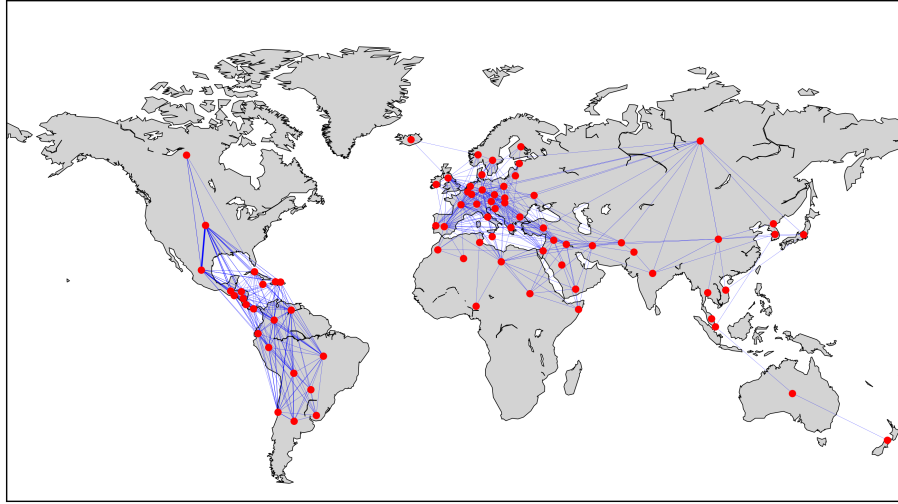


Figura 5: Grafo creado entre países con una distancia menor a 5000 km y con un peso mayor a 5

de prensa chilena relacionadas con la inmigración.

En particular, se identificaron subgrafos que reflejan agrupamientos geográficos. Por ejemplo, se formó un subgrafo que agrupa países de América del Norte y América Central (color negro), lo cual indica una mayor co-aparición en las noticias de países pertenecientes a esta región geográfica. De manera similar, se encontró un subgrafo para Sudamérica (color púrpura), otro para Europa, África y algunos países de Asia (color rojo), y finalmente un subgrafo que incluye países de Asia y Oceanía (color blanco). Los subgrafos se pueden ver coloreados en la Figura 6.

### 5.1. Agrupamientos a lo largo del tiempo

Después de identificar los subgrafos que surgieron a partir del análisis de la comunidad de Louvain, se llevó a cabo un estudio de su evolución a lo largo del período comprendido entre 2016 y 2021. El objetivo fue comprender cómo cambian estos subgrafos con el tiempo y qué países se agregan o se excluyen de cada uno de ellos. Se seleccionaron dos subgrafos principales para este estudio, centrándonos en regiones geográficas específicas. El primer subgrafo se enfocó en América del Norte y América Central, el segundo en América del sur.

Para ilustrar la presencia de los países en cada subgrafo a lo largo de los años, se creó una matriz en la que las columnas representan los años y las filas representan los países, ver Figura 7. En esta matriz, se utilizó el color negro para indicar la presencia de un país en el subgrafo en un año determinado, mientras que el color blanco representó la ausencia de ese país en el subgrafo durante ese período.

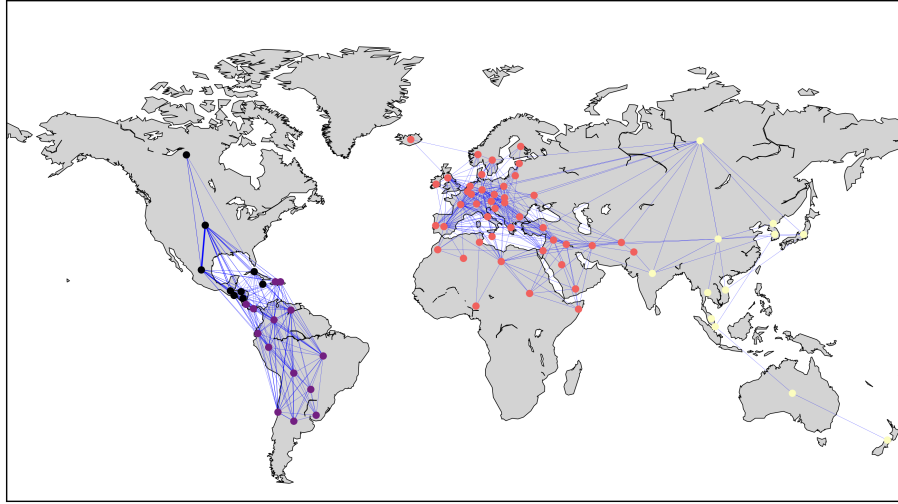
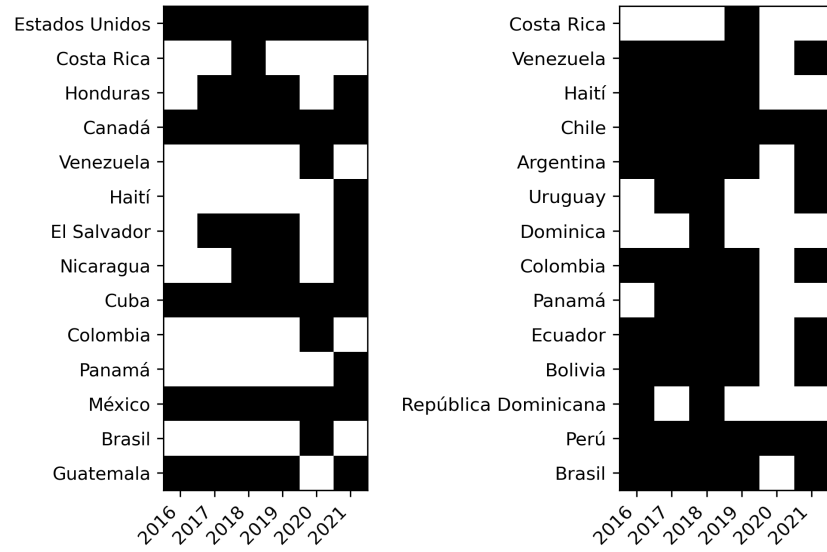


Figura 6: Subgrafos (marcados con colores) generados a partir del grafo principal



(a) Subgrafo de América del Norte y (b) Subgrafo de América del Sur en el tiempo Central en el tiempo

Figura 7: Presencia (negro) y ausencia (blanco) de países en los diferentes subgrafos

Basándose en los países que muestran mayor fluctuación entre las compo-



nentes, como Colombia, Venezuela, Haití, Brasil, Costa Rica y Panamá, y considerando el algoritmo Louvain para la detección de comunidades en el grafo, podemos inferir un análisis.

La fluctuación entre las componentes en estos países puede deberse a varios factores. En primer lugar, los cambios en las menciones de países en las noticias pueden reflejar cambios en los flujos migratorios o eventos relacionados con la inmigración en diferentes regiones. Por ejemplo, en un año determinado, un país puede recibir mayor atención mediática debido a problemas migratorios específicos, lo que puede llevar a su inclusión en una componente determinada. Sin embargo, en años subsiguientes, la situación puede cambiar y los flujos migratorios pueden redirigirse hacia otras regiones o los problemas migratorios pueden disminuir, lo que resulta en la aparición del país en una componente diferente. Además, los cambios en la composición de las componentes también pueden estar influenciados por factores políticos, económicos o sociales que afectan los patrones migratorios. Políticas migratorias más restrictivas o acuerdos bilaterales entre países pueden influir en los flujos migratorios y, por lo tanto, en las menciones en noticias relacionadas.

Es importante destacar que la fluctuación en las componentes no implica necesariamente un cambio significativo en la posición geopolítica de un país, sino más bien una variación en la atención mediática y en los flujos migratorios que se reflejan en los datos analizados.

## 6. Conclusión

En conclusión, la teoría de grafos ha sido una herramienta fundamental en este estudio, permitiéndonos analizar las relaciones y conexiones entre los países mencionados en las noticias de prensa relacionadas con la inmigración en Chile. A través del algoritmo de Louvain Communities, hemos identificado subgrafos que reflejan agrupamientos geográficos y su evolución a lo largo del tiempo.

El estudio realizado proporciona evidencia que respalda las hipótesis planteadas. Se encontraron subgrafos dentro del grafo principal que reflejan agrupamientos geográficos de los países mencionados en las noticias de prensa chilena relacionadas con la inmigración. Estos subgrafos evolucionaron a lo largo del tiempo, mostrando cambios en las conexiones entre los países y la incorporación o exclusión de algunos de ellos. Se identificaron agrupamientos geográficos significativos, como América del Norte y América Central, y América del Sur, que reflejan patrones coherentes con las expectativas iniciales. Los países más influyentes en estas noticias varían según la región geográfica, destacando México y Estados Unidos en América del Norte y América Central (Subgrafo 1), y Chile y Perú en América del Sur (Subgrafo 2) (Ver Anexo donde se encuentran gráficos matrices de centralidad del vector propio). Además, se observaron cambios en los agrupamientos a medida que algunos países fueron agregados o excluidos a lo largo del periodo de estudio. Como trabajo futuro, sería interesante realizar un análisis detallado de las componentes correspondientes a los otros continentes, como Europa, Asia, África y Oceanía.

## 7. Anexo

Deseo destacar que todo el proyecto, incluyendo los códigos, los datos utilizados, los gráficos generados y más información relevante, se encuentra disponible en un repositorio de GitHub: . Este repositorio proporciona un recurso completo y accesible para aquellos interesados en explorar y replicar el estudio realizado. Los códigos permiten la reproducción de los análisis y visualizaciones, lo que brinda la oportunidad de profundizar en los detalles metodológicos y realizar investigaciones adicionales.