

Master Big Data y Business Analytics

Práctica Final

Análisis Redes Sociales



José Luis González Blázquez
30-12-2016

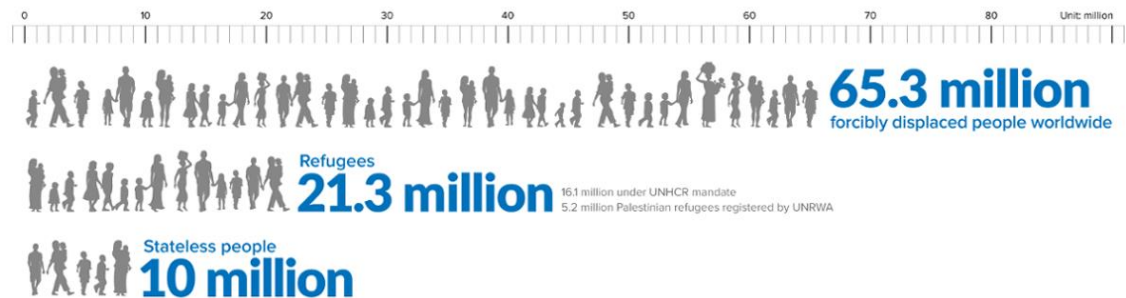
Contenido

1. Introducción	2
1.1. Contexto	2
1.2. Objetivo	2
1.3. Alcance	2
2. Preparación	3
2.1. Extracción de la información	3
2.2. Preparación de la información	3
2.3. Importación en Gephi	3
2.4. Configuración del Layout	6
3. Análisis	8
3.1. Análisis gráfico	8
1960 – 1979	8
1980 – 1999	9
2000 – 2009	10
2010 – 2015	12
3.2. Análisis estadístico	13
4. Conclusiones	16
5. Futuras mejoras	16

1. Introducción

1.1. Contexto

Durante los últimos años se está produciendo lo que se conoce en las noticias como “crisis de refugiados”. Con el aumento de las guerras en Oriente miles de familias intentan huir de los horrores de la guerra y encontrar un lugar donde poder sobrevivir. Según datos de UNHCR, solo en 2015 más de 3 millones de personas estaban en busca de asilo en algún país y 33.972 son forzadas diariamente a dejar sus hogares.



Sin embargo, aunque ahora se le dé más difusión, estas migraciones no son nuevas. En los 2000 fue la guerra de Irak, en los 90 la crisis de los Balcanes, en los 80 la guerra del Golfo y así sucesivamente.

Con este documento vamos a llevar a cabo un análisis de los diferentes movimientos migratorios a lo largo de distintas épocas para conocer cómo han evolucionado.

1.2. Objetivo

El objetivo de esta práctica es detallar los movimientos migratorios a lo largo del tiempo mediante grafos que indiquen las rutas seguidas.

Además, realizaremos análisis estadísticos que nos ayuden a llegar a conclusiones sobre estos movimientos de personas.

1.3. Alcance

El análisis se va a realizar con los datos recogidos de UNHCR (The UN Refugee Agency) desde 1960 hasta 2015.

Por otro lado, no se van a tener en cuenta aquellos valores en los que el origen de los refugiados sea desconocido.

2. Preparación

2.1. Extracción de la información

El dataset con el que vamos a trabajar ha sido extraído de la web Humanitarian Data Exchange de UNHCR: <http://popstats.unhcr.org/en/resettlement>.

La web permite varios filtros: año, país de asilo y país de origen. Para nuestro estudio vamos a seleccionar todos los años disponibles, así como todos los países de asilo y origen.

Por otro lado, tal y como nos avisa la web, tenemos que tener en cuenta lo siguiente:

This page presents information on resettlement arrivals of refugees, with or without UNHCR assistance. This dataset is based on Government statistics and, in principle, excludes humanitarian admissions. In the 2014 data, figures between 1 and 4 have been replaced with an asterisk (*). These represent situations where the figures are being kept confidential to protect the anonymity of individuals. Such figures are not included in any totals.

“Esta página presenta información de reasentamientos de refugiados, con o sin la ayuda de UNHCR. Este dataset está basado en estadísticas del gobierno y, en principio, excluye las admisiones humanitarias. En los datos de 2014, los valores entre 1 y 4 han sido reemplazados con un asterisco (). Esto representa situaciones donde los datos están siendo considerados confidenciales para proteger el anonimato de los individuos. Estos datos no están incluidas en ninguno de los totales.”*

2.2. Preparación de la información

Tal y como hemos indicado previamente, cuando el número de refugiados que cubren una ruta es menor o igual a 4 personas, se indica con un asterisco (*). Para poder representar estos datos, vamos a tomar cada asterisco como si fuesen 2 personas. Aunque introducimos un pequeño error en los datos, éste no es apreciable con la cantidad de datos que manejamos.

Por otro lado, también hay algunas rutas para las que el número de refugiados aparece vacío. Transformamos estos valores vacíos en un cero para poder trabajar con ellos convenientemente.

También observamos muchas líneas en las que el origen de los refugiados es indeterminado. Estas líneas no las tenemos en cuenta para nuestro estudio pues no tienen toda la información necesaria.

Por último, para trabajar con Gephi debemos indicarle tanto los valores que van a hacer de nodos como las aristas que tiene cada nodo (edges).

Para las aristas, creamos un nuevo campo “Type” donde indicamos que las relaciones son dirigidas y dos campos “starttime” y “endtime” para poder hacer la evolución temporal de las migraciones.

Esta fase de preparación de la información la hemos desarrollado en un notebook de Python.

2.3. Importación en Gephi

Importamos los ficheros resultantes del paso anterior en Gephi.

Fichero de nodos:

Importar hoja de cálculo

Pasos

1. **Opciones generales**
2. Parámetros de importación:

Opciones generales

Escoge un archivo CSV a importar:

D:\Máster\Redes Sociales\Ejercicio Final\nodes_refugees.csv ...

Separador: Tabla: Conjunto de caracteres:

Punto y coma Tabla de nodos UTF-8

Previsualización:

ID
Afghanistan
Albania
Algeria
Andorra
Angola
Antigua and Barbuda
Argentina
Armenia

< Atrás **Siguiente >** Terminar Cancelar Ayuda

Fichero de aristas:

Importar hoja de cálculo

Pasos

1. **Opciones generales**
2. Parámetros de importación:

Opciones generales

Escoge un archivo CSV a importar:

D:\Máster\Redes Sociales\Ejercicio Final\edges_refugees.csv ...

Separador: Tabla: Conjunto de caracteres:

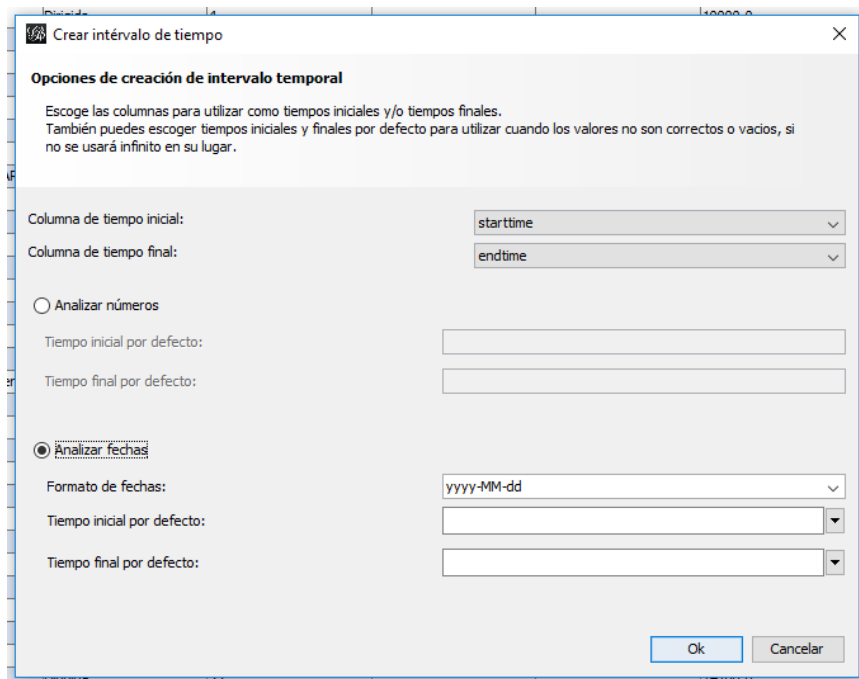
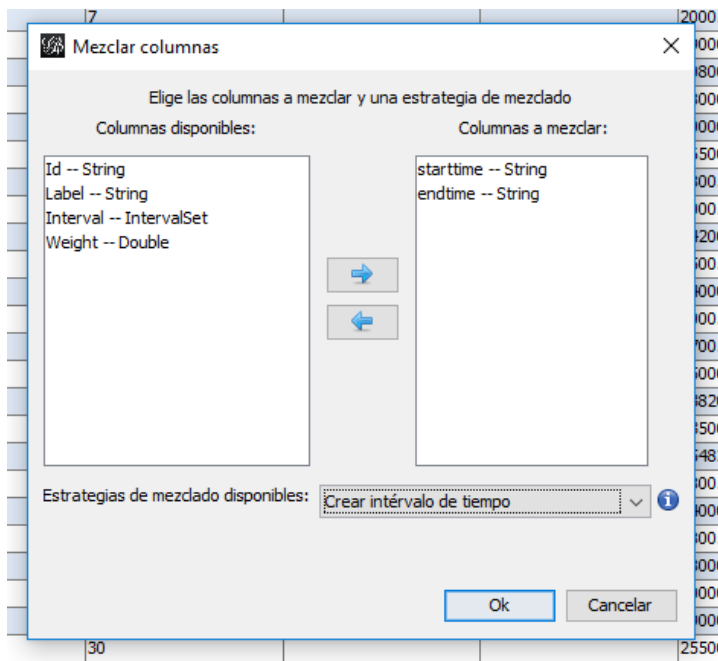
Punto y coma Tabla de aristas UTF-8

Previsualización:

Source	Target	Weight	starttime	endtime	Type
Angola	Dem. Re...	150000.0	1960-01...	1960-12...	Directed
Rwanda	Burundi	30000.0	1961-01...	1961-12...	Directed
Angola	Dem. Re...	150000.0	1961-01...	1961-12...	Directed
Rwanda	Dem. Re...	53000.0	1961-01...	1961-12...	Directed
Ghana	Togo	5000.0	1961-01...	1961-12...	Directed
Rwanda	United R...	12000.0	1961-01...	1961-12...	Directed
Rwanda	Uganda	52000.0	1961-01...	1961-12...	Directed
Rwanda	Burundi	30000.0	1962-01...	1962-12...	Directed

< Atrás **Siguiente >** Terminar Cancelar Ayuda

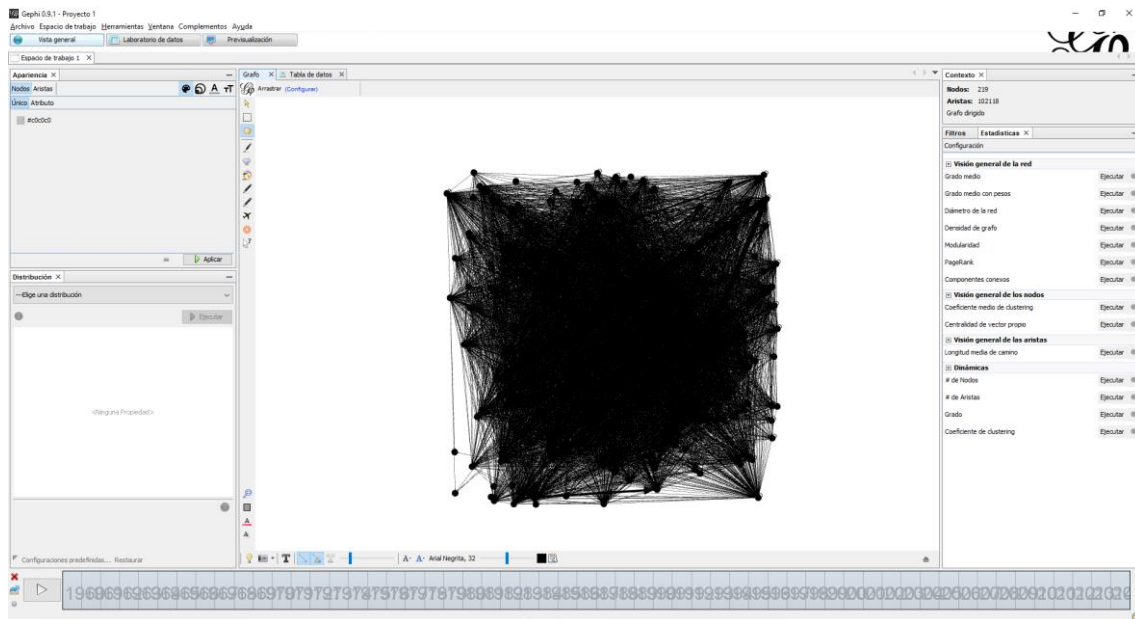
Creamos el campo con el intervalo de tiempo. Para ello, dentro de la pestaña “Tabla de datos” pulsamos en mezclar columnas:



De esta forma, activamos la línea temporal:



Como resultado, nos queda el siguiente grafo con 219 nodos y 102.118 aristas.



2.4. Configuración del Layout

Vamos a modificar la apariencia de los datos de forma que se puedan leer mejor visualmente.

Como algoritmo de espacialidad vamos a escoger un mapamundi que nos permita observar mejor los movimientos.

Algoritmo de espacialidad:

- GeoLayout
- Scale: 2000.
- Projection: Equirectangular.

A este algoritmo le sumamos la visualización del mapa:

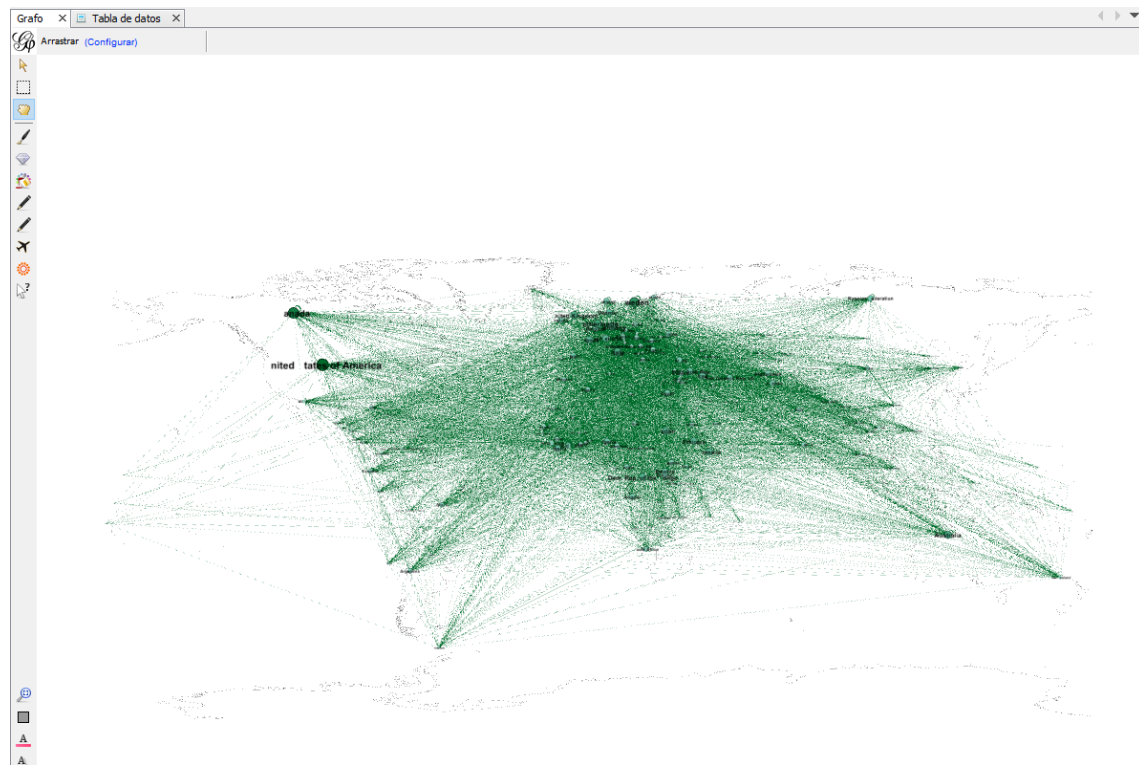
- Map of Countries.
- Scale: 2000.
- Projection: Equirectangular.
- Weight: 500.

Apariencia de nodos:

- Color verde en función del grado.
- Tamaño (20-50) en función del grado.
- Mostrar label en función del tamaño del nodo.

Apariencia de aristas:

- Color verde en función del peso.



3. Análisis

3.1. Análisis gráfico

Para hacer el análisis gráfico, vamos a centrarnos en los grados de entrada y de salida. Como estos datos por sí solos no nos indican la cantidad de personas que huyen de los países o que acogen otros, necesitamos ponderarlo con el peso de las aristas (número de personas que realizan un trayecto).

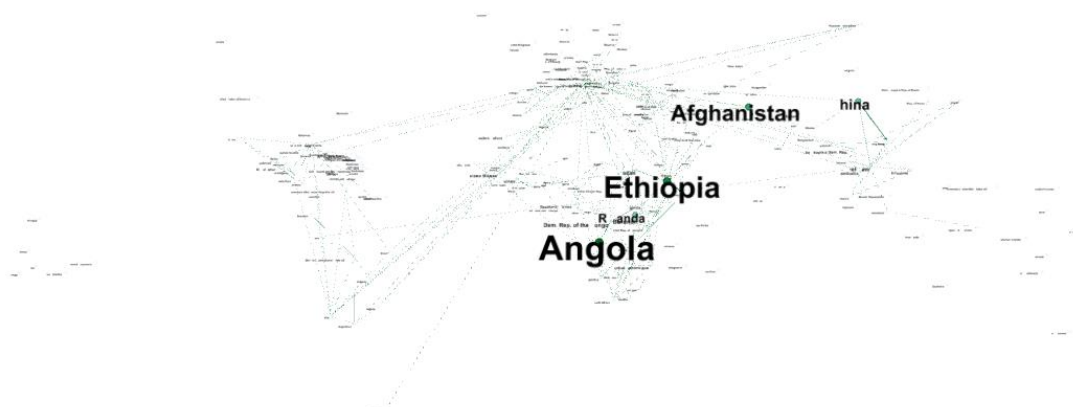
Para ello, ejecutamos *Estadísticas > Grado medio con pesos*.

- Grado de salida con pesos: El grado de salida nos indica cuáles son los países en los que los refugiados huían de manera más repartida. No es el número de personas que huyen del país, sino a cuántos países huyen esas personas. Al ponderarlo con los pesos de las aristas, sí que nos da una idea de los países de los que huyen el mayor número de personas.
- Grado de entrada con pesos: El grado de entrada indica los países que reciben refugiados del mayor número de países. Esto no significa que acojan a más refugiados, sino que los refugiados que acogen provienen de un mayor número de nacionalidades. Al ponderarlo con los pesos de las aristas, nos indica de manera más realista los países que acogen al mayor número de personas.

Para hacernos una idea de la evolución de los movimientos de los refugiados, vamos a analizar el gráfico con una ventana de dos décadas al principio y, a partir del 2000, con una ventana de una década.

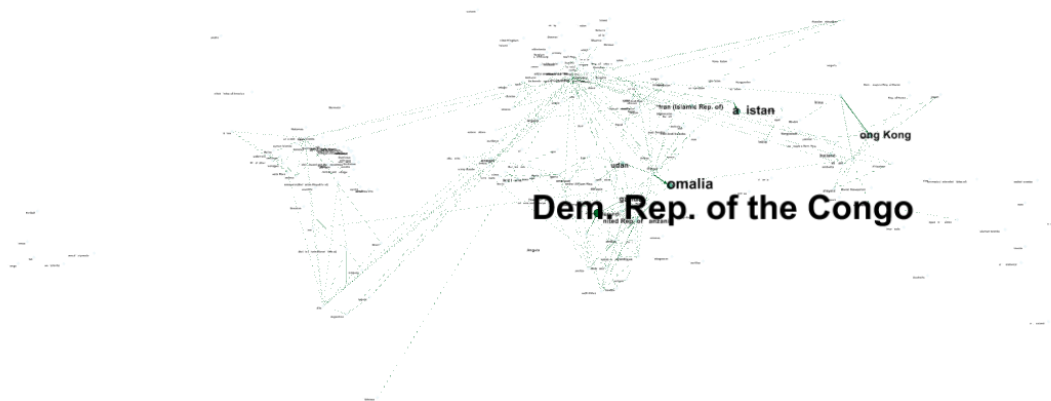
1960 – 1979

Grado de salida



Según vemos en el gráfico, durante estas dos décadas, los principales éxodos de personas se producían desde Angola, Etiopía y Afganistán. También en China se produjeron movimientos de personas reseñables.

Grado de entrada



Como vemos, el país que recogió más refugiados durante estas dos décadas fue el Congo. Esto se debe a las guerras que se sucedían alrededor de sus fronteras que hacía que las personas se refugiasen en este país.

También se ve que Pakistán y Hong Kong recibieron bastantes personas durante esta época.

Hechos relevantes

Durante los años 60, en Angola se produjo la “Guerra colonial portuguesa” que acabó con la independencia del país.

Por otro lado, a finales de los años 70 se produjo la invasión de Afganistán por parte de la Unión Soviética, que produjo un éxodo masivo hacia países limítrofes como Pakistán. Además, en Etiopía se libraba una guerra civil que dejó alrededor de 1.4 millones de muertos.

1980 – 1999

Grado de salida



Como vemos, en estos años destaca sobremanera la salida de personas que huían de Afganistán.

Además, observamos que en Europa del Este se aglutina mucho movimiento de entrada y salida de personas.

Lo primero que nos llama la atención de estos años es que vemos que se ha incrementado el número de personas que se han movido con estatus de refugiado a lo largo del mundo. Además de por el aumento de las guerras durante esta época, también se puede deber a que durante los años anteriores la agencia ACNUR tenía menos datos acerca de estos movimientos de refugiados.

Grado de entrada



Como hemos visto en el gráfico anterior, el principal éxodo de personas se produjo desde Afganistán, por lo que es lógico que los países que acogían al mayor número de personas de este éxodo sean Irán y Pakistán. Además, observamos que Irán también recibe muchos refugiados desde Irak.

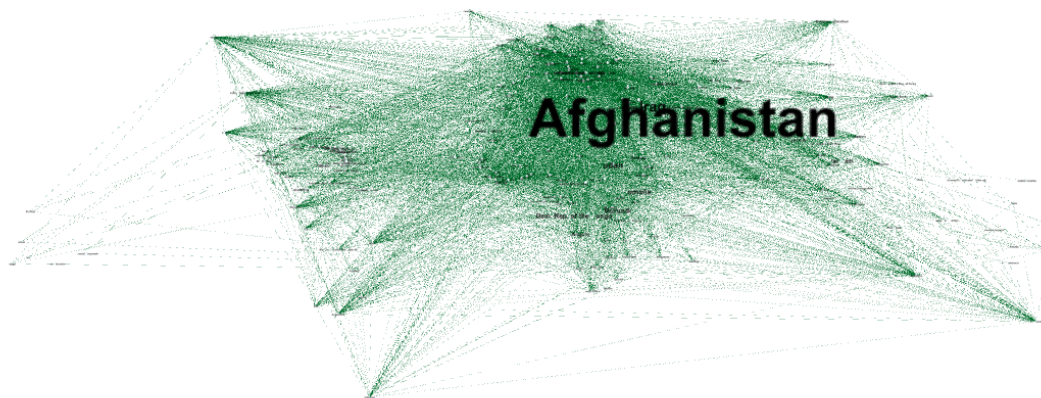
Hechos relevantes

El éxodo de personas de Afganistán se puede explicar debido a la invasión por parte de la Unión Soviética que había comenzado la década anterior. Además, al juntarse con la guerra del golfo en Irak, personas de ambos países se refugiaron en Irán, la cual se había convertido en un país radical islámico durante la década anterior.

Por otro lado, en Europa se dieron momentos importantes como la caída del muro de Berlín y la guerra de los Balcanes, que hicieron que el flujo de personas entre países aumentase considerablemente.

2000 – 2009

Grado de salida

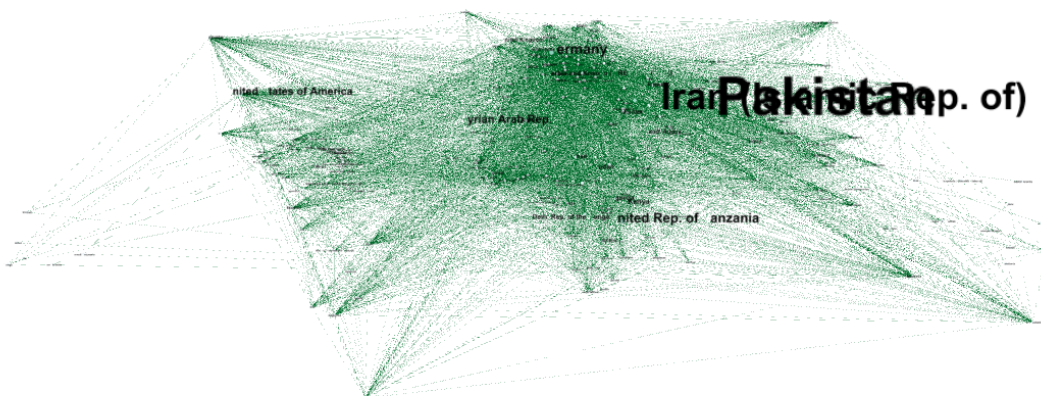


Durante la primera década de los 2000 vemos que Afganistán sigue siendo el país que mayor número de refugiados exporta.

Muy por detrás de Afganistán, pero aun así de una manera reseñable, observamos que de Irak también huyen un gran número de personas.

En Europa podemos destacar Croacia, Serbia y Bosnia como los países de los que más personas salen como refugiados.

Grado de entrada



Observamos que, como en las décadas anteriores, los países hacia los que más huyen las personas que salen de Afganistán son Irán y Pakistán.

No obstante, descubrimos otros movimientos interesantes. Vemos cómo Siria es uno de los países que más personas acoge huyendo de las guerras de Irak y Afganistán, así como del extremismo religioso de Irán.

En Europa, Alemania empieza a vislumbrarse como el país que más refugiados acoge provenientes tanto de la guerra de los Balcanes como de personas que han atravesado los países de Europa del Este hasta llegar a sus fronteras.

Como dato interesante, observamos que Tanzania está acogiendo a muchos refugiados provenientes del resto de países de África así como de Oriente Próximo.

Estados Unidos también aparece por primera vez como punto relevante de llegada de personas, aunque no llega a los números de Alemania, por ejemplo.

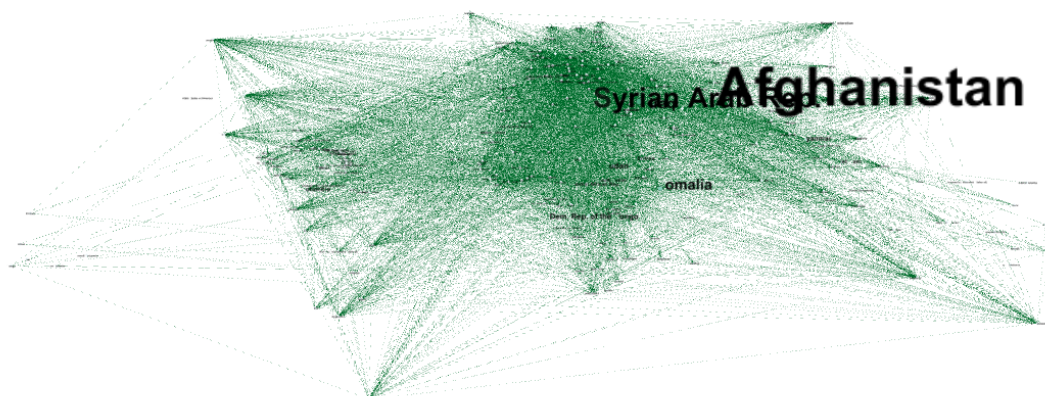
Hechos relevantes

En 2001 comienza la guerra de Afganistán tras el ataque de los talibanes a las torres gemelas en Estados Unidos. Además, en 2003 también comienza la segunda guerra del Golfo en Irak en el contexto de la lucha contra el terrorismo internacional.

Por otro lado, en 2004 se produjo la mayor ampliación de la historia de la Unión Europea, anexionando 7 países del antiguo bloque soviético. Esto explica el mayor movimiento de personas en Europa, y especialmente hacia Alemania.

2010 – 2015

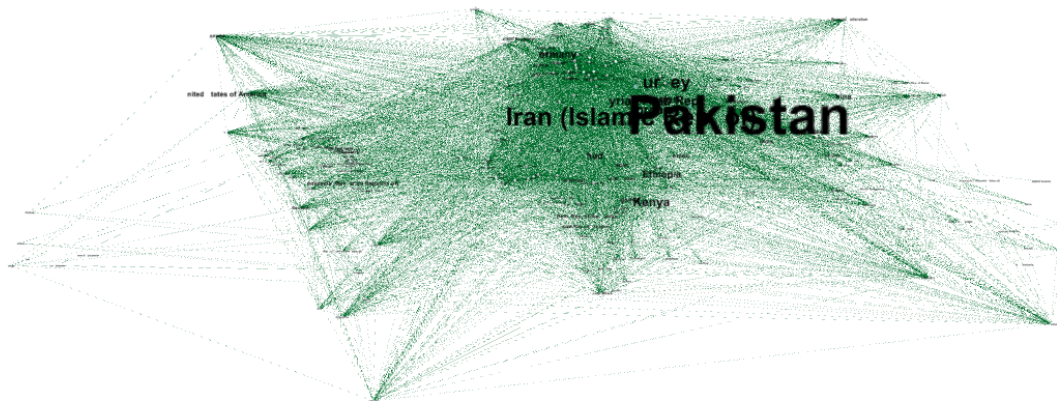
Grado de salida



Durante estos últimos 5 años, Afganistán sigue destacando como país del que huye el mayor número de personas. No obstante, se vislumbra cómo Siria también ha sufrido un gran éxodo de refugiados.

También observamos en África la salida de personas de Somalia hacia otros países.

Grado de entrada



Durante esta época Pakistán destaca notablemente sobre el resto de países en cuanto a acogida de personas. Asimismo, vemos que Irán, aunque sigue acogiendo a un gran número de personas, no es tan importante en este aspecto como lo fue anteriormente.

Por otro lado, podemos observar que Turquía ha crecido respecto a la acogida de refugiados en otras épocas. En Europa, también podemos seguir destacando a Alemania como un refugio de personas.

En África, observamos que Kenia y Etiopía han acogido a un gran número de personas.

Hechos relevantes

La guerra civil de Siria comenzó en 2011 como producto del terrorismo, así como de la inestabilidad que hemos observado anteriormente en los países vecinos.

Por otro lado, en 2009 comenzó en África la guerra de Somalia que dura actualmente. Esta guerra involucra a Etiopía y Somalia, aunque en territorio somalí. Esto explica el movimiento de personas en esa zona del planeta.

3.2. Análisis estadístico

El análisis estadístico lo vamos a hacer sobre la totalidad de los datos, sin distinguir entre épocas.

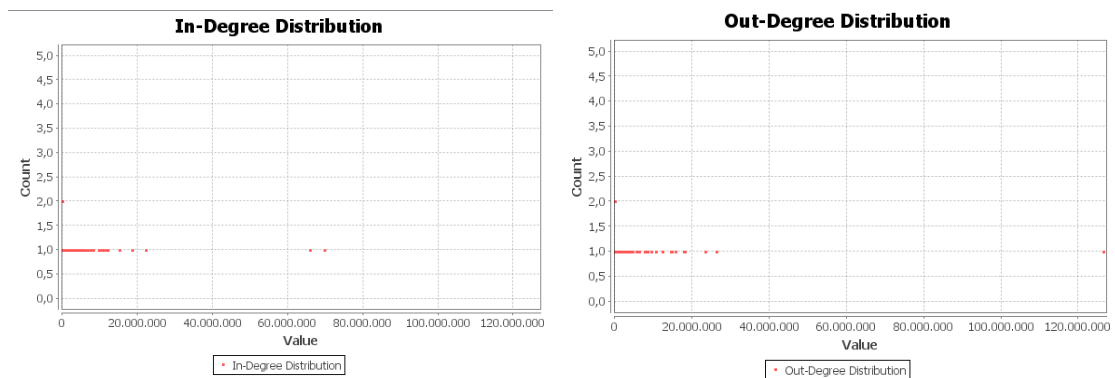
Datos que vamos a medir:

- Grado medio con pesos: lo mismo que hemos explicado en el punto anterior.
- Diámetro de la red: indica cuantos saltos de distancia hay entre los dos nodos más alejados de la red. Para ello utilizamos 3 métricas:
 - o Betweenness: Mide la frecuencia con la que un nodo aparece en el camino más corto entre nodos de la red.
 - o Closeness: La distancia media desde un nodo inicial a todos los demás nodos de la red.
 - o Eccentricity: La distancia desde un nodo al nodo más alejado de él en la red.
- Densidad del grafo: indica la relación entre el número de aristas actual y el número de aristas máximo que podría haber en un grafo con el mismo número de nodos.

- PageRank: La filosofía de PageRank es la siguiente: a más páginas apunten a mi página mayor será su reputación, y a más reputación tenga a su vez la página que me apunta, más importancia se le da a ese enlace. A su vez, y para evitar trampas, el algoritmo premia la exogamia, es decir, que a mi página la apunten desde diferentes comunidades poco relacionadas entre sí.

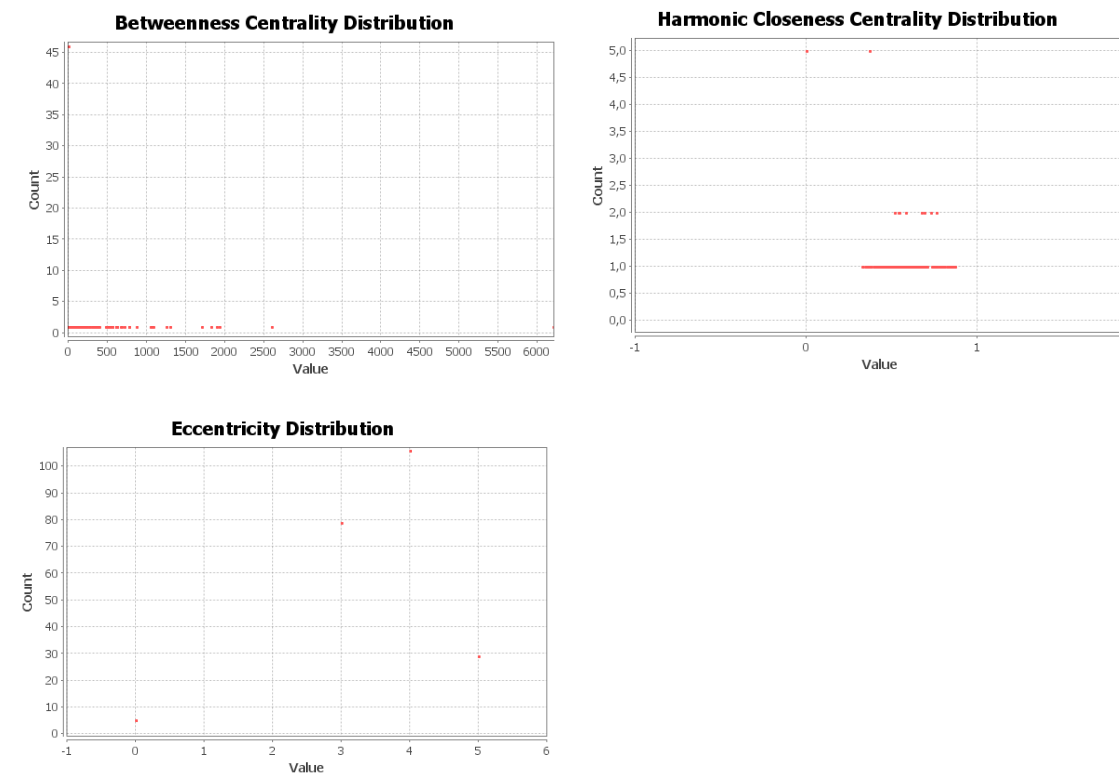
Grado medio con pesos

El grado medio con pesos tiene un valor de 1.956.252,283. La distribución de entradas y salidas queda de la siguiente forma:



Diámetro de la red

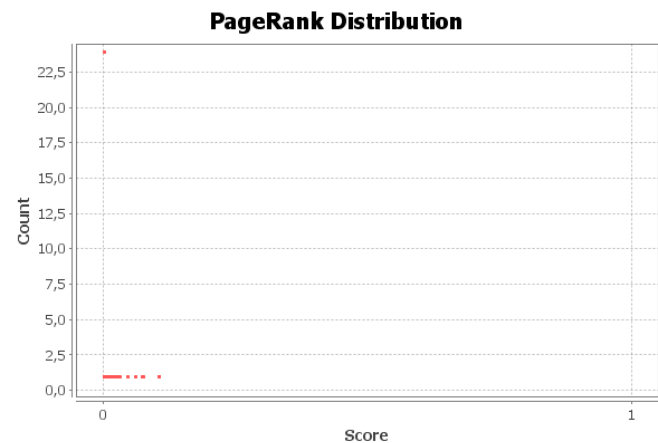
La red tiene un diámetro de 5 y una longitud media en cada paso de 2,089040762720109.



Densidad del grafo

Nuestro gráfico tiene una densidad de 2,139%.

PageRank



4. Conclusiones

Según hemos visto en los datos anteriores, las guerras son la principal causa por la que migra la gente de sus hogares.

Un dato interesante que hemos observado ha sido cómo la entrada de refugiados durante la primera década del 2000 ha sido un motivo más para que se desatase la guerra civil siria en la segunda década de este siglo. Esto ha provocado una nueva salida de refugiados de este territorio hacia Europa, Irán y Pakistán.

También hemos visto cómo Alemania es el país europeo que más refugiados acoge. En los últimos años, Francia está siendo el siguiente país en acogida de personas.

Por último, países como Reino Unido o Estados Unidos, que protestan en contra de la acogida y no quieren aceptar la inmigración no son los que más refugiados acogen en el mundo. De hecho, se encuentran muy alejados de los países que más acogen.

5. Futuras mejoras

Posibles mejoras que se pueden hacer en el futuro en este análisis:

- Se puede poner como peso del nodo la población del país. Esto ayudaría, tanto en salida como entrada de personas, a saber la relación con la población total del país.
- Estudio de polarización de la población respecto a este tema. En este sentido he intentado hacer un análisis de los enlaces de twitter que contengan el hashtag #refugiados. La idea es que 2 enlaces están conectados entre sí si un mismo usuario ha mencionado ambos (del estilo “los clientes que compran A también compran B”). Esto daría una idea de si leemos información parcial (sea a favor o en contra) o estamos abiertos a escuchar a la otra parte. No obstante, no he conseguido realizar este estudio debido al número limitado de tweets con los que contaba.