

Análisis mediante teoría de redes de interacciones en sistemas complejos reales

Efecto del Brexit en la red de proyectos europeos

Trabajo Fin de Máster
Máster de Física y Tecnologías Físicas

Junio 2018

Realizado por:
Leticia Pérez Sienes

Directores:
David Iñiguez Dieste
Jesús Gómez Gardeñes



Índice general

1. Introducción	1
1.1. Redes Complejas	3
1.1.1. Tipos de redes	3
1.1.2. Medidas sobre la estructura y características de las redes	4
1.2. Red aleatoria y libre de escala	10
2. Programa Horizon 2020	12
2.1. Obtención y filtrado de los datos	13
3. Efecto del Brexit en la red total de proyectos	15
3.1. Análisis por comunidades	15
3.2. Comunidades sin Reino Unido	17
4. Efecto del Brexit por programas	20
4.1. Excellent Science	21
4.2. Industrial Leadership	23
4.3. Societal Challenges	25
5. Conclusiones	27

Capítulo 1

Introducción

La teoría de grafos es la base matemática sobre la que se construye el estudio de las redes complejas [1][2][3]. Un grafo es una representación gráfica de un conjunto de puntos (también nodos, vértices) unidos por líneas (enlaces, aristas), que representan elementos e interacciones entre los mismos. Una correcta representación del mismo supone una gran simplificación del modelo a tratar. Resulta sorprendente la cantidad de modelos que se encuentran en la naturaleza que pueden ser estudiados mediante teorías de grafos. Sin duda, es este hecho el que en las últimas décadas ha contribuido a la proliferación de numerosas investigaciones en este ámbito, llevándolo a ser uno de los campos científicos en auge. Hasta hace relativamente poco tiempo campos como la biología, física, sociología y economía se nos aparecían como compartimentos estancos, resultaba difícil imaginar que pudiera existir una conexión entre ellos y por tanto, descubrimientos en cualquiera de estas áreas pasaban desapercibidos en las otras. La principal característica de las redes complejas es su multidisciplinariedad. Esto supone que el avance en el estudio y análisis de redes complejas implica la colaboración de expertos en distintas materias que aporten sus conocimientos para obtener una perspectiva transversal y holística de las mismas.

A pesar de que el estudio de las redes complejas es muy reciente (hablamos de algo más de un par de décadas) la teoría de grafos sobre la que se sustenta comenzó a desarrollarse ya en el siglo XVIII, de la mano del matemático alemán Leonhard Euler en un artículo de 1736 conocido como *el problema de los puentes de Königsberg* [4], la actual ciudad rusa de Kaliningrado. La teoría de grafos surgió como el intento de buscar una solución al siguiente problema:

“Dado el mapa de Königsberg, con el río Pregel dividiendo la ciudad en cuatro regiones distintas que se encuentran unidas a través de siete puentes, ¿es posible dar un paseo comenzando desde cualquier región de manera que sólo se atravesase una vez cada puente y se regrese al mismo punto?”

Para resolver el problema Euler recurrió a una representación abstracta del mapa que sirviera para otros ejemplos de semejantes características: las distintas regiones fueron sustituidas por puntos y los puentes que las conectaban por enlaces.

Euler concluyó que, debido a que los 4 puntos estaban conectados con un número impar de líneas, resultaba imposible encontrar un camino que satisficiera los requisitos anteriormente mencionados.

Desde entonces, en la teoría de grafos existe el concepto *ciclo euleriano*, que consiste en

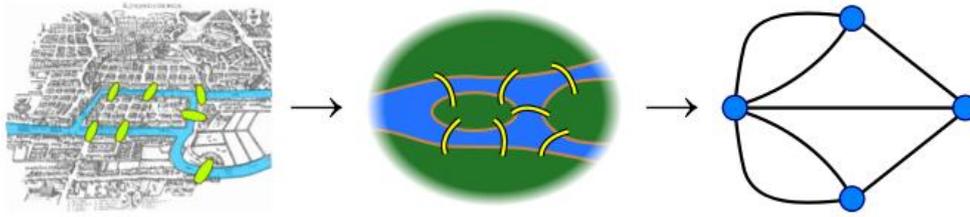


Figura 1.1: Esquema del problema de los puentes de Königsberg [5].

encontrar un camino que recorra todos los nodos una sola vez y regresando al nodo original. Euler consideró que en este tipo de problemas las medidas de los objetos son irrelevantes, lo que importa son sus propiedades estructurales. Esto dio lugar al término que hoy se traduce como *topología* en geometría.

A partir de entonces aparece de nuevo la teoría de grafos, esta vez como herramienta de análisis de redes eléctricas, que llevó a Gustav Kirchoff plantear las leyes de los circuitos que permiten calcular corrientes y voltajes. Se considera que fue la primera aplicación de las redes a un problema de ingeniería.

En 1852 Francis Guthrie planteó el conocido problema de “¿cuántos colores son necesarios para dibujar un mapa político, con la condición de que dos países vecinos no puedan tener el mismo color?”. En 1976 Kenneth Appel y Wolfgang Haken resolvieron el problema matemáticamente [6], definiendo los conceptos teóricos y términos fundamentales de la incipiente teoría de grafos.

La consolidación de las redes complejas aparece en la última década del siglo pasado con dos artículos: el de Watts & Strogatz sobre *redes de mundo pequeño* [7], y el de Barabasi & Albert sobre la *aparición de redes complejas invariantes de escala* [8]. En estos trabajos surge la idea de que en sistemas complejos como Internet, redes sociales, red de carreteras, redes tróficas, redes neuronales, etc., subyace una estructura que describe el complejo conjunto de interacciones entre las unidades elementales que constituyen el sistema.

Red	Nodos	Enlaces
Internet	Routers	Líneas físicas
Red de aeropuertos	Aeropuertos	Vuelos entre aeropuertos
Red de colaboraciones	Científicos	Coautores de un artículo
Red social	Personas	Amistades
Red lingüística	Palabras	Sinonimia
Red trófica	Especie	Alimento
Red económica	Bancos	Transacciones
Red neuronal	Neuronas	Sinapsis

Tabla 1.1: Ejemplos de redes complejas.

La red compleja difiere del grafo en que este último no tiene por qué representar ningún sistema complejo real, mientras que la red compleja sí lo hace. De este modo, un mismo grafo puede representar distintos sistemas complejos que nada tengan que ver entre sí, pero que

tengan una estructura común. Este hecho sugiere que la naturaleza se organiza siguiendo ciertos patrones que son comunes desde la evolución biológica hasta sistemas que se originan a partir de la actividad humana, como en la red de Internet, de tráfico aéreo, etc. En la tabla 1.1 se recogen algunos ejemplos de redes complejas.

1.1. Redes Complejas

Una red compleja está constituida por un conjunto finito de nodos $N = \{n_1, \dots, n_N\}$ y enlaces $L = \{l_1, \dots, l_L\}$ que los unen. Como hemos visto en la tabla 1.1, los nodos pueden representar todo tipo de objetos o individuos, y los enlaces las interacciones entre los mismos. Cada nodo n_i viene determinado por su orden i en el conjunto, y cada enlace que une dos nodos de la red viene denotado por $l_k = \{i, j\} = \{n_i, n_j\} = l_{ij}$ si une el nodo n_i con el n_j . Se dice entonces que n_i y n_j son *vecinos*.

1.1.1. Tipos de redes

Dependiendo de las características de la red éstas pueden ser (no todas son excluyentes):

- I. *Red no dirigida*: el orden de los índices en las uniones no altera la configuración, $l_{ij} = l_{ji}$.
- II. *Red dirigida*: el orden de los índices es relevante, de modo que $l_{ij} \neq l_{ji}$.
- III. *Red multicapa*: existen autoenlaces l_{ii} o varios enlaces entre dos nodos.
- IV. *Red pesada*: los enlaces tienen asignados un valor numérico que indica la intensidad de la unión. Estos pesos pueden ser todos del mismo signo (*unsigned*) o de diferente (*signed*).
- V. *Redes dinámicas*: el número de nodos, las conexiones cambian a lo largo del tiempo, por lo que el sistema no está representado siempre por una misma red.

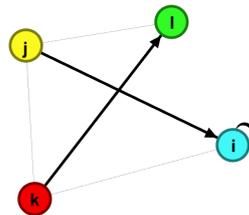


Figura 1.2: Ejemplo de una red mixta (enlaces dirigidos y no dirigidos) y pesada.

Una red puede ser dirigida o no dirigida, presentar autointeracciones que a su vez pueden ser pesadas, y enlaces múltiples entre pares de nodos. Las posibilidades son muchas, pudiéndose dar el caso en el que una misma red pueda analizarse de distintas maneras según convenga tener en cuenta unas propiedades u otras.

1.1.2. Medidas sobre la estructura y características de las redes

Una estructura tan compleja queda determinada por su matriz de adyacencia, esto es, una matriz A donde las entradas de las filas y las columnas hacen referencia a los nodos existentes en la red y las posiciones a_{ij} son números que indican el valor del enlace entre el nodo n_i y n_j . Las características de la red se ven reflejadas en la matriz A , por ejemplo: una red sin autointeracciones tendrá ceros en la diagonal, una red no pesada simple tendrá términos a_{ij} binarios (0 si no hay enlace, 1 si lo hay), una red no dirigida presenta una matriz de A simétrica, etc.

A continuación se presentan algunas de las medidas más habituales sobre redes. Si no se especifica, se considera que se tiene un grafo simple y no dirigido $G = (V, E)$ con N vértices y M enlaces:

Grado. El grado de un nodo es el número de primeros vecinos del mismo, es decir, el número de enlaces que presenta. En el caso de una red dirigida se distingue el grado de entrada y el grado de salida, en función de los enlaces que llegan al nodo o parten de él. Al grado del nodo i se le denota por k_i o $k(i)$. En caso de ser dirigida se tiene además el grado de entrada $k_{in}(i)$ y el de salida $k_{out}(i)$ (la suma de ambos es el grado total). Considerando los términos a_{ij} de la matriz de adyacencia, el grado de un nodo i será:

$$k_i = \sum_{j=1}^N a_{ij} \quad (1.1)$$

Y el grado medio de la red se obtiene al sumar todas las entradas de A y dividir las para el número de nodos:

$$\bar{k} = \frac{1}{N} \sum_{i=1}^N k_i \quad (1.2)$$

Distribución de grado. La distribución de grados $P(k)$ aporta información importante sobre la estructura de la red. $P(k)$ da la probabilidad de que un determinado nodo elegido al azar tenga grado k , es decir, esté conectado a exactamente k vecinos. Según sea la forma de la distribución de grado las redes se clasifican en diversas topologías, las más conocidas son:

- *Topología de Poisson.* Las redes conocidas como *aleatorias* o de *Érdos-Renyi*, en las que cada nodo se conecta a otro de manera independiente con probabilidad p , presentan la siguiente distribución binomial de grado:

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \quad (1.3)$$

donde N es el número total de nodos, k el grado y p la probabilidad de que un par de nodos esté conectado.

Sin embargo, las redes reales suelen ser dispersas, esto es, cuando el grado medio es mucho menor que N , y por ello la distribución binomial se aproxima a la de Poisson:

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (1.4)$$

donde λ es el grado medio.

- *Topología exponencial.* La distribución exponencial aparece en el contexto de crecimiento de redes que se construyen a partir de lo que se conoce como *enlace igualitario*. Esto es la situación en la que un nodo nuevo puede alcanzar cualquiera de los nodos de la red con idéntica probabilidad:

$$P(k) = Ce^{-\alpha k} \tag{1.5}$$

- *Topología ley de potencia.* Las redes que con mayor frecuencia aparecen en la naturaleza siguen esta distribución:

$$P(k) = Ck^{-\gamma} \tag{1.6}$$

siendo γ una constante.

Mientras que en las redes aleatorias la mayoría de los nodos tiene similar grado, en aquellas cuya distribución de grado sigue una exponencial o la ley de potencia los nodos presentan gran heterogeneidad.

Coefficiente de asortatividad. La asortatividad representa la tendencia de los nodos para establecer enlaces con otros similares a ellos [9]. Normalmente esta tendencia viene determinada por el grado, es decir, se consideran dos nodos similares si tienen parecido grado. Dependiendo del tipo de red los nodos tienen ciertos atributos que podrían ser determinantes para el cálculo de este coeficiente (en una red social podría ser la etnia, religión, ideología política, etc.).

Para medir la correlación entre los grados de los pares de nodos de una red calculamos el coeficiente de asortatividad r , que toma valores entre -1 y +1. Este coeficiente de correlación de Pearson evalúa si los nodos de alto grado tienden a conectarse con nodos de también alto grado y, a su vez, si los nodos de bajo grado lo hacen con nodos de bajo grado; entonces $r > 0$ (red asortativa). Si sucede todo lo contrario, es decir, los nodos de alto grado tienden a conectarse con nodos de menor grado y viceversa el coeficiente es $r < 0$ (red disortativa). En el caso en el que los nodos se unen con los demás nodos independientemente de su grado $r = 0$. El cálculo de r viene dado por la siguiente expresión [9]:

$$r_k = \frac{\frac{1}{M} \sum_{i,j} k_i k_j - \left\{ \frac{1}{M} \sum_{i,j} \frac{1}{2} (k_i + k_j) \right\}^2}{\frac{1}{M} \sum_{i,j} \frac{1}{2} (k_i^2 + k_j^2) - \left\{ \frac{1}{M} \sum_{i,j} \frac{1}{2} (k_i + k_j) \right\}} \tag{1.7}$$

siendo k_i y k_j los grados de los nodos de ambos extremos del enlace $(i, j) \in E$, y M el número de enlaces totales.

Coefficiente de clustering. El coeficiente de clustering cuantifica la tendencia de agrupamiento de los nodos. Si un vértice está conectado con sus vecinos y todos ellos lo están entre sí, entonces el coeficiente es máximo; mientras que si sus vecinos apenas interaccionan entonces el valor será bajo (ver ejemplo de la figura 1.3). En las redes reales los nodos tienden a establecer relaciones en grupos con una densidad de enlaces alta.

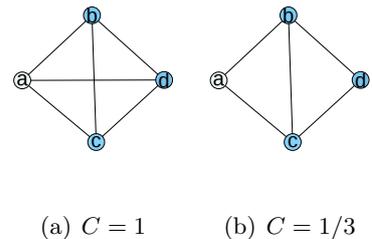


Figura 1.3: Ejemplos de coeficiente de clustering.

Dado un nodo v_i de una red no dirigida con grado k_i , entonces el máximo número de enlaces que pueden existir entre sus primeros vecinos son $\frac{k_i(k_i-1)}{2}$, de manera que su coeficiente de clustering vendrá dado por:

$$C_i = \frac{2T(i)}{k_i(k_i - 1)} \quad (1.8)$$

donde $T(i)$ son los enlaces de los nodos v_j con v_k de la vecindad de v_i , o lo que es lo mismo, número de *triángulos* que pasan por v_i . También puede calcularse la media de clustering de todos los nodos de la red para tener una idea general del agrupamiento local de los nodos:

$$\bar{C} = \frac{1}{N} \sum_i C_i. \quad (1.9)$$

Transitividad. El concepto de transitividad está estrechamente ligado al de clustering. La transitividad mide el agrupamiento global de la red, y viene dado por la expresión:

$$T = \frac{3 \times \text{número de triángulos de la red}}{\text{número de conjuntos de tres vértices}} \quad (1.10)$$

donde T toma valores entre 0 y 1, siendo $T = 0$ el caso en el que no existe ningún camino cerrado en la red, y $T = 1$ el caso en el que todos los nodos están conectados con todos los demás.

Comunidades. El concepto de comunidad se introdujo a raíz del estudio de redes sociales, en el que se observa una tendencia de los individuos de dividirse en grupos dentro de los cuáles existen numerosas conexiones y, en cambio, escasas conexiones entre los grupos [10].

En una misma comunidad cada nodo puede alcanzar cualquier nodo de su misma comunidad, de manera que la conectividad y el coeficiente de clustering son altos. En ocasiones, se ha visto como las comunidades representan unidades funcionales (ciclos en rutas metabólicas, etc.) con propiedades diferentes a las que pueda tener la red completa. Por tanto, limitarse al estudio de las propiedades y estructuras de la red sin reparar en las comunidades puede llevar a ignorar información importante del sistema. De este modo, en los últimos años se ha visto un notable crecimiento en el interés por las técnicas de detección de comunidades.

A pesar de que la noción de comunidad se nos aparece sencilla, resulta difícil establecer un definición formal y, en consecuencia, diseñar un algoritmo eficiente para dar con una partición optimizada de los nodos de la red. No existe una única manera de hacer una partición (ver figura 1.4), por ello la detección de comunidades es un tema bajo continuo estudio [11]. Se considera que una partición $\mathcal{P} = C_1, \dots, C_k$ del grafo $G = (V, N)$ ($\forall i, C_i \subseteq V$) es buena cuando la proporción de enlaces dentro de C_i es muy alta en comparación con la proporción de enlaces con el resto de C_j ($i \neq j$). Un modo de evaluar la calidad de la partición es atendiendo a la *modularidad*, una medida de la estructura interna de la red [12].

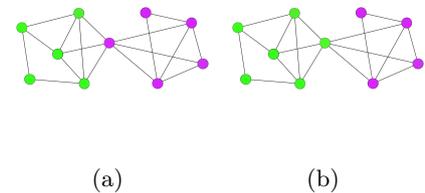


Figura 1.4: Ejemplo de dos particiones distintas de una misma red: en 1.4(a) las comunidades tienen 6 y 4 vértices, y en 1.4(b) ambas tienen 5 vértices.

En el presente trabajo las comunidades se han obtenido mediante el algoritmo de Newman [10] y de ‘walktrap’ [13].

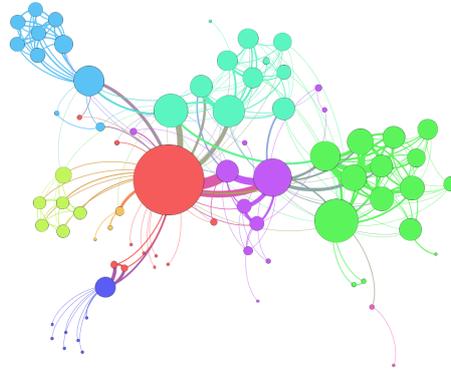


Figura 1.5: Distintas comunidades de una red representadas con diferente color [14].

Modularidad. Como hemos mencionado, la modularidad mide la calidad de la división de la red en componentes. Una partición del grafo que de lugar a subgrafos muy densamente conexos pero poco conectados entre ellos presentará una modularidad alta; si por el contrario la red no presenta apenas subgrafos densos o están muy bien conectados se obtendrá un valor de modularidad bajo. Por tanto, la modularidad mide el número mínimo de elementos o enlaces que se necesitan eliminar para desconectar la red. Esta propiedad permite la detección de comunidades grandes pero resulta insuficiente para detectar comunidades pequeñas.

Conectividad. La conectividad entre dos nodos es el número de caminos distintos que hay entre ellos. Si se quiere ser más explícito pueden diferenciarse dos tipos de conectividades: conectividad de nodo y conectividad de enlace. La conectividad de nodo cuenta los posibles caminos diferentes que atraviesan nodos distintos para llegar de v_i a v_j , mientras que la conectividad de enlace cuenta los posibles caminos distintos, pudiendo visitarse un mismo nodo. En la figura 1.6 los nodos a y b están conectados por un único nodo c pero por dos caminos independientes: (a, b, c, f, g) y (a, d, c, e, g) (también podrían escogerse (a, b, c, e, g) y (a, d, c, f, g)). Su conectividad es muy sensible a la eliminación de nodos: basta suprimir c para que queden desconectados. También lo es a la eliminación de aristas aunque en menor medida: es necesario eliminar dos enlaces, los que conectan a con c o c con g , para que a y g estén incomunicados.

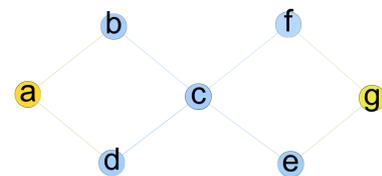


Figura 1.6: Conectividad entre dos nodos a y g . Existen dos caminos independientes pero ambos comparten el nodo c .

Por tanto, la conectividad de un par de vértices es en-

tonces una medida que cuantifica cómo de fuerte es la conexión entre los nodos. Dos nodos que estén conectados únicamente por un camino presentarán conectividad más débil que dos nodos con varios caminos que los conecten. Esta propiedad resulta importante debido a la información que aporta sobre la resiliencia de la red. Por ejemplo, una red que siga una ley de potencia será muy robusta frente a la eliminación de nodos escogidos aleatoriamente, puesto que la gran mayoría presentan pocos enlaces; en cambio, es muy frágil si se seleccionan a conciencia, puesto que son unos pocos los de alto grado y desconectarían gran parte de la red.

Las redes pueden ser conexas o no conexas. Una red conexa es aquella en la que desde cualquier nodo i se puede llegar a cualquier otro nodo j , con más o menos nodos intermedios. En cambio, si existe algún nodo i desde el que no se puede llegar a un nodo j , es no conexa. Esto significa que la red está constituida por varias *componentes*. De entre ellas la de mayor relevancia es la de mayor tamaño, la *componente gigante*.

Distancias. Si consideramos una red simple (aquella cuyos enlaces son no dirigidos y de valor 1 de manera que los $a_{ij} = 0, 1$) la distancia o *camino geodésico* d_{ij} entre los nodos n_i y n_j es el número de enlaces mínimo que deben recorrerse para llegar del nodo n_i al nodo n_j . Si no existe camino que los una se dice que la distancia es infinita. La longitud geodésica media de la red es entonces:

$$l_G = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} d_{ij} \tag{1.11}$$

donde d_{ij} es la distancia del camino más corto que une los nodos i y j .

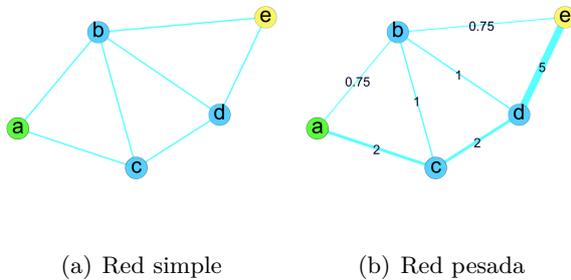


Figura 1.7: En la imagen *a* el camino más corto del nodo *a* al nodo *e* es el que pasa por *b*, mientras que en la imagen *b* el camino más corto recorre los nodos *c* y *d*.

Mientras que en una red simple el camino geodésico es el que recorre menos nodos intermedios, en una red cuyos enlaces sean pesados las distancias ya no son proporcionales al número de intermedios, sino que depende de los pesos de los enlaces recorridos. Por ejemplo, en la imagen 1.7(a) tenemos una red simple y las distancias entre un nodo y sus primeros vecinos es en todos ellos 1. Sin embargo, en la imagen 1.7(b) los enlaces son pesados, indicando qué nodos presentan conexiones más fuertes. El enlace $a-b$ es más débil que los enlaces $a-c$, $b-c$ o $d-e$. Cabe esperar que la información fluya con mayor facilidad y eficiencia entre los nodos en los que los pesos de los enlaces sean mayores, y en consecuencia, la distancia (entendida como accesibilidad

entre los nodos) sea más corta.

En la literatura existen distintos algoritmos que calculan la distancia más corta entre un par de nodos, entre ellos están: Bellman Ford, Dijkstra y Floyd-Warshall [15]. Dependiendo de si la red es pesada y del tamaño de la misma conviene utilizar uno u otro. El segundo es un algoritmo más optimizado que el primero y calcula la menor de las distancias de un nodo con cada nodo de la red, y el tercero es un método que ahorra en tiempo, ya que calcula de una vez todos los caminos más cortos.

Otra medida que aporta información sobre las dimensiones de la red es el *diámetro*, es decir, la mayor de las distancias geodésicas entre todos los pares de nodos. Resulta sorprendente comprobar que por mucho que la red aumente su tamaño incorporando una gran cantidad de nodos el diámetro apenas crece, se mantiene pequeño. Este efecto se conoce como red de *mundo pequeño*. Un ejemplo muy conocido de este fenómeno es el de *seis grados de separación*¹, el cuál postula que dos personas están separadas en un máximo de 6 intermediarios.

Comunicabilidad. Esta magnitud fue introducida por Ernesto Estrada y Naomichi Hatano quienes presentaron el concepto por primera vez [17]. Exponían cómo medir la eficiencia espacial en una red sobre las bases de comunicación entre los nodos. La comunicabilidad entre dos nodos i y j es:

$$C(i, j) = \sum_{j \neq i}^N e^A = (I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots)_{ij} \quad (1.12)$$

con A la matriz de adyacencia. $C(i, j)$ representa todos los caminos posibles que conectan i con j . La potencia A^n representa la cantidad de caminos de longitud n (suponiendo que cada enlace mida la unidad) que une ambos nodos. Otorgaron una relevancia decreciente a los caminos cada vez más largos, el factorial fue una opción y con ella el cálculo se limita a resolver la exponencial de la matriz. Si se trata de una matriz no dirigida entonces A será simétrica, y se puede encontrar una matriz diagonal que simplifique el cálculo: $A = Q^T \Lambda Q$, con Q la matriz cuyas columnas son los autovectores de A y Λ la matriz diagonal cuyas entradas son los autovalores de A .

Si extendemos el cálculo de $C(i, j)$ a todos los pares de nodos entonces tenemos la comunicabilidad total de la red:

$$C(A) = \sum_i \sum_{j \neq i}^N e^A = \sum_i \sum_{j \neq i}^N v_i(k) v_k(j) e^{-\lambda_k} \quad (1.13)$$

con $v_i(j)$ la componente j -ésima del i -ésimo autovector y λ_k el k -ésimo autovalor.

Esta medida tiene un alto coste computacional cuando se estudian redes muy grandes, por lo que puede ser interesante calcular en su lugar la comunicabilidad de centralidad, esto es, el coeficiente $C(i, i)$ para todos los nodos. Este coeficiente representa la cantidad de caminos cerrados que vuelven al nodo de partida i .

Por otro lado, el índice de Estrada $EE(A)$ supone el límite inferior de la comunicabilidad total de la red, y es:

$$EE(A) = \sum_i e^{\lambda_i} \quad (1.14)$$

mientras que el límite superior es ne^{λ_1} , con λ_1 el mayor de los autovalores. Así, $EE(A) < C(A) < ne^{\lambda_1}$. Es fácil ver que en el caso en el que se tenga una red vacía, esto es n vértices y ningún enlace, A será nula y $C(A) = n$. Por el contrario, una red completa tendría todos los enlaces posibles y esto llevaría a un $C(A) = n^2 e^{n-1} - n$.

Eficiencia. La eficiencia de una red es una medida sobre la capacidad que tiene de intercambiar información [18]. El concepto puede aplicarse a escala global y local: la eficiencia global

¹La teoría fue inicialmente propuesta en 1930 por el escritor húngaro F. Karinthy, sin embargo fue en 1950 cuando se hizo famosa gracias al experimento propuesto por el psicólogo estadounidense S- Milgram [16].

mide la capacidad de la red de enviar información a través de toda ella, mientras que la eficiencia local mide la resistencia de la red de fallar a pequeña escala. La eficiencia del nodo i mide cómo es el intercambio de información entre sus vecinos cuando se elimina i . Es interesante conocer la eficiencia media, definida así:

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d(i,j)}, \quad (1.15)$$

con $d(i, j)$ la distancia geodésica entre i y j .

Medidas de centralidad. Numerosas investigaciones en redes se han dedicado a definir el concepto de centralidad de un grafo. La pregunta que motiva tal hecho es buscar cuáles son los vértices más importantes o principales. Hay diversas definiciones de centralidad, sin embargo, una de las más simples y conocidas es la centralidad de grado.

- **Centralidad de grado.** Un nodo se considera central si presenta muchas conexiones con otros nodos. En redes sociales esto supone tener acceso a gran cantidad de información.
- **Centralidad de autovector o de Bonacich.** La importancia de un nodo ya no sólo se mide en función de la cantidad de conexiones que tiene sino de a dónde apuntan sus enlaces [19]. Así, un nodo con grado pequeño puede tener fácil acceso a la información si está conectado con nodos de alto grado.

1.2. Red aleatoria y libre de escala

Una propiedad muy importante que caracteriza la estructura de una red compleja es la distribución del grado de los nodos, que viene dada por $P(k)$ y, que como ya hemos visto en 1.1.2, nos da la probabilidad de que un nodo escogido al azar tenga k enlaces o primeros vecinos. Los tres tipos de distribuciones más importantes son: de Poisson, exponencial y libre de escala.

Las redes que siguen la distribución de Poisson son las que se conocen como redes aleatorias o redes Erdős-Renyi, y son muy importantes sobretudo por razones históricas, ya que fueron las primeras que se estudiaron. No obstante, son poco realistas puesto que apenas se encuentran ejemplos de este modelo en la naturaleza.

Una red cuya distribución sigue la ley de potencia $P(k) = Ck^{-\gamma}$, con C una constante y γ tomando valores típicamente entre 2 y 3 (valores encontrados en la gran mayoría de las redes libres de escala reales) se conoce como *red libre de escala*. La importancia de este tipo de redes es que son las que aparecen con mayor frecuencia en la naturaleza, lejos de la idea original que se tenía de que lo esperado eran las redes aleatorias. La principal característica de estas redes es que existen muchos nodos con poca conectividad y pocos nodos con muy alta conectividad (véase la figura 1.9). Las diferencias principales entre una red aleatoria y otra libre de escala se encuentran en la cola de la distribución.

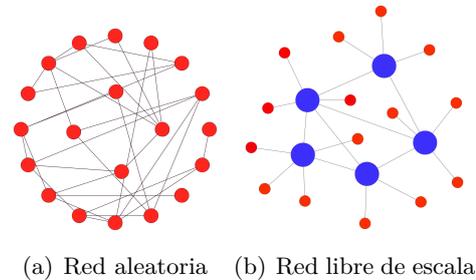


Figura 1.8: Ejemplo de red aleatoria y red libre de escala con el mismo número de nodos.

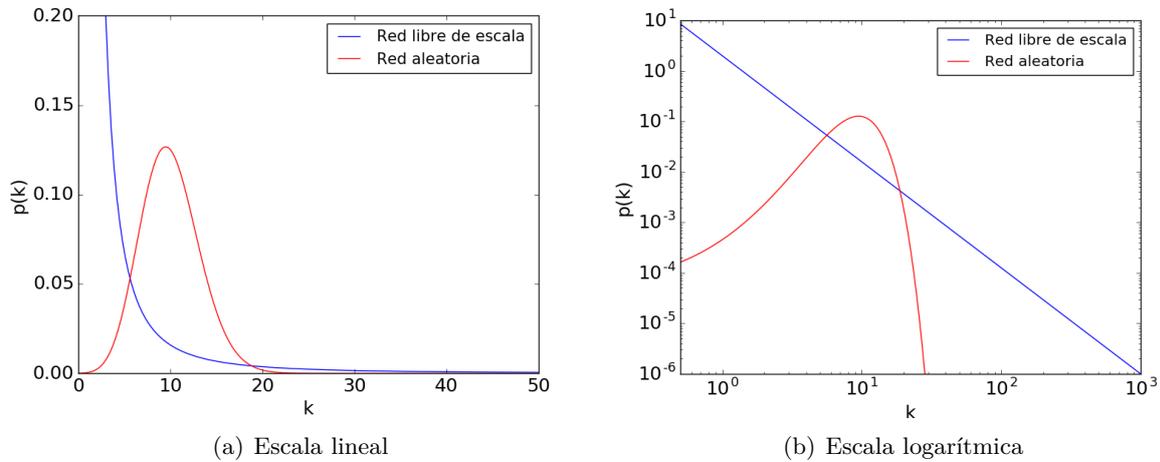


Figura 1.9: Distribución de grado de una red aleatoria y una libre de escala en diferentes escalas.

- Para valores pequeños de grado la distribución de la ley de potencia se encuentra por encima de la de Poisson, puesto que tiene muchos nodos con grado pequeño.
- Para valores del grado cercanos a la media la distribución de Poisson se encuentra por encima de la distribución de la ley de potencia, ya que existen muchos nodos con grado similar a la media.
- Para valores altos de grado la distribución que sigue la ley de potencia está de nuevo por encima de la de Poisson. Esto implica que existe una probabilidad no despreciable de encontrar nodos con grado mucho más alto que la media.

A los nodos que están altamente conectados se les conoce comúnmente como “*hubs*” o conectores, puesto que contribuyen a conectar gran parte de la red. Es por ello que las redes libres de escala son muy robustas en cuanto a la eliminación de un nodo al azar, pero frágiles si se elimina de manera selectiva alguno de sus conectores. El nombre “libre de escala” hace referencia a la característica de que independientemente de la escala en la que observemos la red su comportamiento es similar; es decir, se mantienen sus propiedades.

En general las redes no son estáticas, sino que crecen (y decrecen). Cuando un nodo nuevo se incorpora a la red no establece enlaces de manera aleatoria, sino que existe una probabilidad de conectarse con un nodo y no con otro, lo que se conoce como enlace preferencial.

En las gráficas de la red de proyectos europeos del capítulo 3 se verá que presentan distribuciones de grado que se ajustan a la distribución libre de escala. Esta propiedad resulta relevante a la hora de eliminar nodos de alto grado, ya que contribuyen a la desconexión de la red.

Capítulo 2

Programa Horizon 2020

Horizon 2020 es un programa fundado y financiado por la Comisión Europea que da cobertura a una gran variedad de proyectos de investigación e innovación. En ellos confluyen numerosas instituciones repartidas por toda la geografía europea, que colaboran en equipos multidisciplinares y transnacionales. Horizon 2020 ha destinado 80 mil millones de euros para proyectos durante el periodo 2014-2020, con el objetivo de promover avances en diversas áreas del conocimiento, liderar descubrimientos y dar apoyo a los laboratorios y centros de investigación para llevar sus ideas al mercado. Horizon 2020 pretende otorgar notoriedad y consolidar a Europa como un líder firme en investigación, innovación y desarrollo.

Este programa financia proyectos en los que participan instituciones de varios países, de manera que el resultado es una red de colaboración estrecha entre universidades y centros de investigación de toda Europa. Dentro de la red, no todos los países están igualmente implicados ni tampoco lo están en todas las 22 áreas (agricultura, energía, alimentación, seguridad, ciencias sociales y espacio, entre otras). Sin embargo, se reconoce una gran participación en muchos de ellos de Reino Unido, Francia, Alemania, Italia y España. La reciente voluntad de Reino Unido de salir de la Unión Europea afectará indudablemente a la red de proyectos europeos, más aún siendo éste uno de los países líderes en I+D. En el presente trabajo se estudia el impacto de la salida de Reino Unido de la U.E. en el contexto de la red de proyectos europeos. Para una mejor comprensión de las consecuencias, además se compara con el efecto de la salida de instituciones europeas escogidas al azar con independencia de su localización.

Los distintos programas de H2020 son:

1. Excellent Science
2. Industrial Leadership
3. Societal Challenges
4. Spreading Excellence and Widening Participation
5. Science with and for Society
6. Euratom

De todos los programas los que involucran más centros e instituciones son los tres primeros, para los cuáles construiremos las redes correspondientes y mediremos sus propiedades.

2.1. Obtención y filtrado de los datos

La página web de la comisión europea permite el acceso a las bases de datos de los proyectos de H2020 que se han realizado desde el 2014 hasta ahora [20]. De los archivos que pueden descargarse uno de ellos es una base de datos que contiene información de cada uno de los centros que han participado alguna vez en algún proyecto: nombre de la institución, ciudad, país, número identificativo, página web, etc. Otro documento es una base de datos que contiene los proyectos organizados por programa en el que participan, y un tercero constituye otra base de datos con información sobre cada proyecto en particular, entre los que se especifica el número de identificación del proyecto, centro coordinador, centros participantes, cuantía económica recibida, fecha de realización, etc.

El proceso de filtrado y construcción de la red se ha hecho utilizando el lenguaje de programación Python. Se ha considerado una óptima herramienta debido a la variedad de librerías que proporciona con funciones ya programadas. Entre ellas se incluye la librería *Networkx*, que incluye funciones para la generación de redes y cálculo de medidas.

El objetivo de este trabajo es analizar el impacto del Brexit, y para ello resulta conveniente analizar la red de proyectos de cada uno de los ámbitos de I+D. En consecuencia, el primer paso fue elaborar un programa que clasificara los proyectos según la clase de programa al que pertenecían. De este modo, para cada uno de ellos se disponía de la lista de todos los proyectos.

A continuación, hubo que elaborar otro programa que tomara de otro archivo los datos del coordinador, participante, sus números identificativos, sus respectivas asignaciones económicas en el proyecto y países a los que pertenecen. Este paso lo hacía dos veces, una en la que todos los nodos se incluían (red completa de proyectos) y otra en la que si el nodo era británico pero participante el proyecto se mantenía sin dicho nodo, y si el coordinador era británico se eliminaba todo el proyecto (red sin Reino Unido). Se encontraron centros cuyo papel es *host institution* o *beneficiary*, pero representaban un porcentaje muy bajo del total y por ello se consideró tratarlos como coordinador y participante respectivamente. Por otro lado, algún proyecto presentaba más de un coordinador (generalmente no más de dos), por lo que para simplificar el problema se decidió tomar al azar uno de ellos como único coordinador del proyecto, el otro se añadió al conjunto de participantes del mismo. De esta manera se generan las redes deseadas: una con todos los países europeos junto a sus colaboradores, y otra sin los nodos de Reino Unido, que en consecuencia se llevan consigo también a centros que únicamente han colaborado con centros de Reino Unido. Una vez hecho esto para cada programa el siguiente paso fue construir la red de nodos y enlaces. Para ello, se decidió establecer únicamente los enlaces que parten del coordinador y que llegan a cada uno de los participantes con un peso igual a su contribución económica relativa en el proyecto. De esta manera, dos nodos están conectados si al menos una vez han colaborado juntos, siendo uno de ellos la institución coordinadora. De otra forma la red sería demasiado densa en cuanto a enlaces, y se considera que el centro coordinador es el alma mater del proyecto, sin cuya participación el proyecto sería prácticamente inviable.

En un principio se tuvieron en cuenta las contribuciones del propio coordinador, que consiste en conectar el nodo consigo mismo. Más tarde se descartó esta opción, puesto que para el cálculo de las propiedades de la estructura de la red no es relevante y, lo que es más, puede inducir a errores del tipo caer en un bucle infinito a la hora de medir distancias.

El siguiente paso fue elaborar un programa que tomase los datos de los nodos y enlaces de las redes de proyectos de cada una de las diferentes clases de programas de Horizon 2020. Para

cada programa se generó una red con todos los centros, y otra red con los centros que sobreviven al eliminar los de Reino Unido, esto es, la red que resulta de eliminar los centros de Reino Unido y aquellos centros que sólo colaboran con ellos y ningún otro centro de la comunidad europea. Además, para un entendimiento mejor construyeron 100 redes en las que se eliminaron la misma cantidad de nodos pertenecientes a Reino Unido, pero de manera aleatoria (es decir, se eliminan tantos nodos como centros de Reino Unido existían en la red completa, pero pudiendo pertenecer ahora a cualquiera de los países), de las cuáles se tomaron las mismas medidas que en los otros casos y se promediaron. Esto permite la comparación de los efectos de escoger los nodos de manera selectiva o al azar. La finalidad es observar qué medidas se ven más afectadas en cada programa para determinar el papel que juegan los centros británicos en la red. Inevitablemente la eliminación de una considerable cantidad de nodos conlleva una variación en los valores de las medidas, precisamente por ello se compara con una media estadística de redes de tamaño similar. Se han considerado de especial interés para la comparación de propiedades de las redes las siguientes medidas: grado, asortatividad, transitividad, clustering, eficiencia, camino geodésico medio. De esta manera, podrá determinarse si el impacto se debe a la cantidad de centros que abandonan la red o si depende de su pertenencia a Reino Unido.

Paralelamente se utilizó la herramienta de *Kampal* para el proyecto *Brexit* [21], que trabaja con la misma base de datos de Horizon 2020. *Kampal*, entre otras cosas, permite visualizar y obtener información acerca de las comunidades que se forman, también permite seleccionar muestras de los conjuntos de la red total (por países, por proyectos) y calcula propiedades de la red muestral, comparándolos con los de la red total. De *Kampal* se obtuvieron bases de datos de los países integrantes de cada comunidad de la red total de proyectos tanto con Reino Unido como sin él, y de sus respectivas contribuciones dentro de la comunidad.

Además, la herramienta de visualización de redes *Gephi* [14] permitió obtener gráficos y algunas medidas que se calcularon con el código.

Capítulo 3

Efecto del Brexit en la red total de proyectos

3.1. Análisis por comunidades

Para hacernos una idea general de la situación actual de la red de proyectos europeos hemos agrupado todos los centros participantes de Horizon H2020 por país al que pertenecen. De este modo, se detectan 3 grandes comunidades cuyos líderes son Alemania con 52 países integrantes, Holanda con 46 y España con 35. En la gráfica 3.1 se presenta el porcentaje de los fondos destinados a cada una de las tres comunidades.

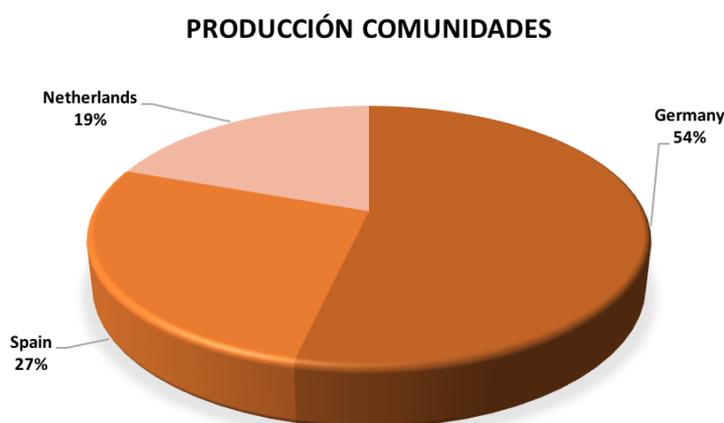


Figura 3.1: Porcentaje de la producción de las comunidades en Horizon 2020.

En las gráficas 3.2, 3.3 y 3.4 se presentan los 10 países que más colaboran dentro de sus comunidades. Los porcentajes hacen referencia a la asignación de fondos destinados a los proyectos de cada uno de los países en esa comunidad, que consideramos aproximadamente proporcional a la producción interna de dichos países en la misma.

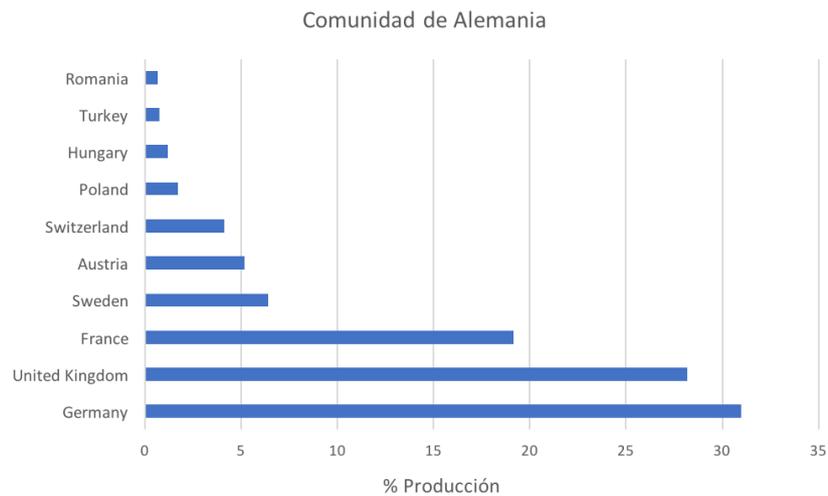


Figura 3.2: Países con más fondos en la comunidad liderada por Alemania.

En la comunidad de Alemania se incluyen Reino Unido y Francia, países de gran población que además se encuentran a la cabeza en la inversión en I+D. En cambio, como vemos en la gráfica 3.3, en la comunidad de Holanda los países colaboradores son los de menor tamaño en términos de población: Bélgica, Dinamarca, Noruega, etc.

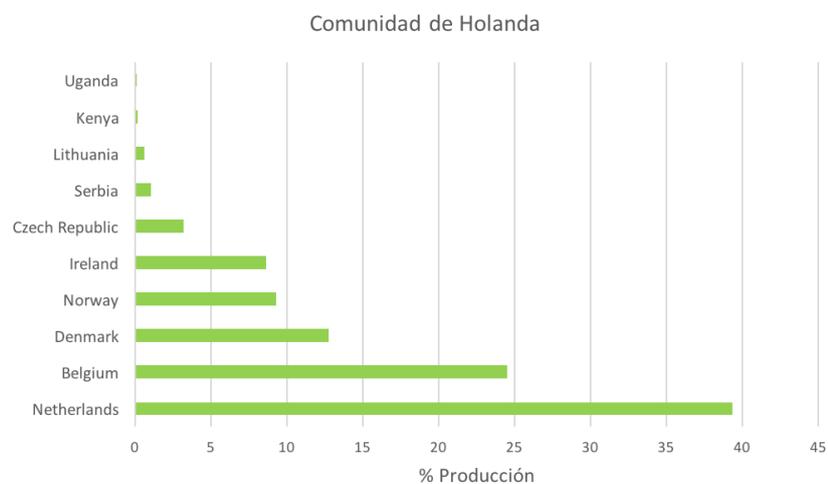


Figura 3.3: Países con más fondos en la comunidad liderada por Holanda.

Cabe destacar que el liderazgo de la comunidad recae en Alemania, pero Reino Unido juega un papel fundamental puesto que es el siguiente país con mayor asignación de fondos en su comunidad, por lo que se esperaría que este grupo sea el que más impacto sufra tras el Brexit.

Por último, en la gráfica 3.4, vemos que los colaboradores de España son principalmente los países mediterráneos: Italia, Grecia, Portugal, Chipre, etc.



Figura 3.4: Países con más fondos en la comunidad liderada por España.

3.2. Comunidades sin Reino Unido

Debido a cómo hemos concebido la red cuando eliminamos los nodos de Reino Unido con ellos se van centros que sólo habían colaborado en proyectos donde éstos fueran coordinadores. Es interesante ver la proporción de estos centros en la red sin Reino Unido y en las redes generadas al eliminar centros al azar, puesto que si decrece considerablemente el tamaño de la red indicará que Reino Unido lidera muchos proyectos y tiene cierta relevancia en Horizon 2020.

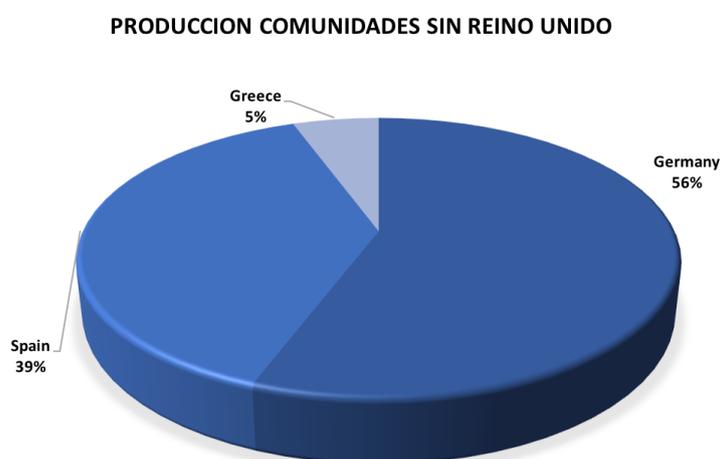


Figura 3.5: Porcentaje de la producción de las comunidades sin Reino Unido.

Con esta nueva configuración las comunidades se ven alteradas. Ahora España lidera la segunda comunidad que más produce en la red total y aparece Grecia como líder de la tercera comunidad. También es interesante comprobar los integrantes de estas nuevas comunidades. La comunidad liderada por Alemania se mantiene como principal productora, sin embargo hay que destacar el auge de la nueva comunidad liderada por España, pasando de producir un 27% a

un 39%. Veremos que una parte de la comunidad de Holanda es absorbida por la liderada por España.

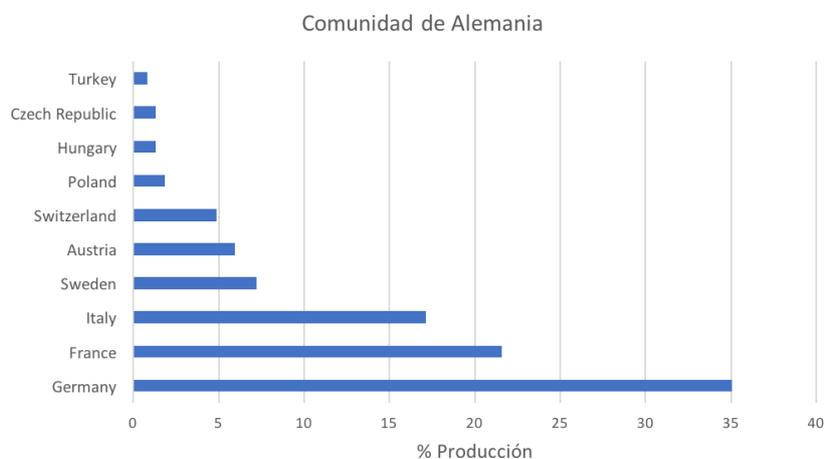


Figura 3.6: Países con más fondos en la comunidad liderada por Alemania sin Reino Unido.

Observamos en la gráfica 3.6 que la comunidad liderada por Alemania parece mantenerse inalterada, pese a la incorporación de Italia, principal colaborador de España en el caso en el que Reino Unido estaba en la red. En la gráfica 3.2 vemos cómo países integrantes de la comunidad

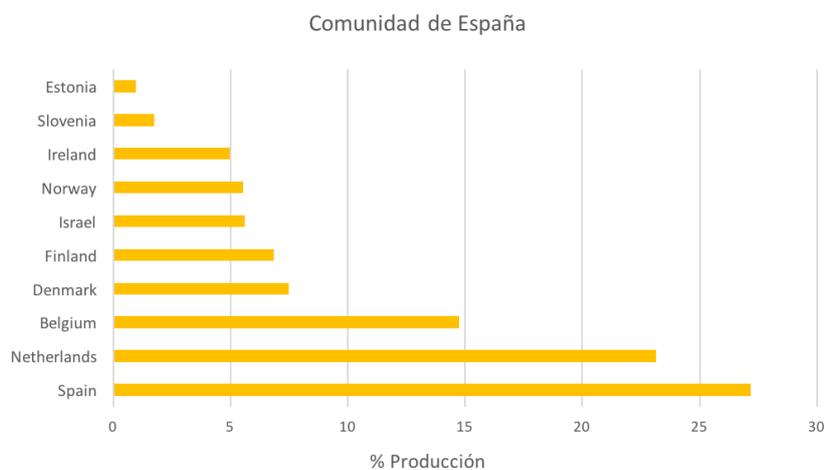


Figura 3.7: Países con más fondos en la comunidad liderada por España sin Reino Unido.

de Holanda ahora pasan a formar parte de la comunidad liderada por España (Bélgica, Noruega, Irlanda), incluida Holanda.

La comunidad liderada por Grecia es la más pequeña, los principales cambios son la inclusión de Portugal como máximo participante y de Chipre (antes integrado en la comunidad liderada por España), y el número de países extracomunitarios involucrados en H2020 (Ucrania, Bosnia y Herzegovina, China). Hay que tener presente que aunque se sitúen entre los 10 países que más contribuyen en esta comunidad el peso de ésta es pequeño, recordamos que tan solo representa



Figura 3.8: Países con más fondos en la comunidad liderada por Grecia sin Reino Unido.

el 5% de la red total de Horizon 2020.

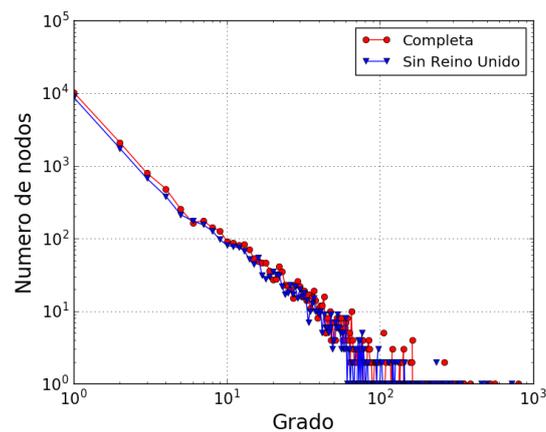


Figura 3.9: Distribución de grado es escala logarítmica para la red total de proyectos de todos los programas de Horizon 2020.

En la gráfica 3.9 se representa la distribución de grado para toda la red de Horizon 2020 antes y después del Brexit; constatamos que se ajustan muy bien a redes libres de escala, comentadas en la sección 1.2 del capítulo introductorio.

Capítulo 4

Efecto del Brexit por programas

Es interesante estudiar el impacto de la salida de Reino Unido en cada uno de los diferentes programas de Horizon 2020. De los mencionados en el capítulo 2 analizaremos las redes de aquellos que más proyectos tienen: *Excellent Science*, *Industrial Leadership* y *Societal Challenges*.

De las redes de los programas se ha tomado la componente gigante (la componente conexas de mayor tamaño) de la red total y de la red sin los nodos de Reino Unido. El motivo es que los cálculos se simplifican considerablemente mientras que los valores de las medidas permanecen prácticamente iguales. Como los valores por sí solos no pueden aportarnos demasiada información, se ha hecho estadística sobre 100 redes resultantes de eliminar aleatoriamente la misma cantidad de nodos que centros de Reino Unido existen, de manera que la comparación con estas medidas nos aportarán información valiosa sobre el papel que Reino Unido juega en los proyectos de I+D europeos.

En cada programa veremos cuáles son los países más implicados, se presentará la gráfica de la distribución de grado para el caso de la red completa y de la red sin Reino Unido, además de una tabla con las medidas recogidas para ambos casos y el de las medidas promedio de las redes aleatorias.

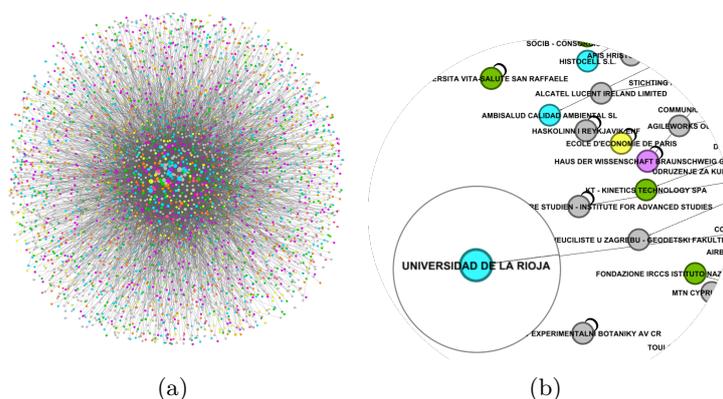


Figura 4.1: Red de la componente gigante del programa Excellent Science (a) y aumento de una zona de la misma donde se visualizan los nombres de los centros (b).

4.1. Excellent Science

En este programa se incluyen proyectos que investigan en el campos de las ciencias, educación, nuevas tecnologías, etc. De los tres programas que aquí se analizan éste es el que menos nodos tiene en la componente gigante, unos 3380.



Figura 4.2: Comunidades de colaboración entre países en Excellent Science, identificadas por el país con mayor asignación económica en cada una de ellas.

Fijándonos en las gráficas de la figura 4.2 Reino Unido lidera la comunidad con más fondos disponibles a día de hoy, por lo que es de esperar que el efecto de su salida de la Unión Europea tenga cierta repercusión. Los países se reparten de la siguiente manera (mostramos sólo los que más participan y en orden de mayor a menor implicación):

- La comunidad liderada por **Reino Unido** está formada por 42 países, entre los cuáles se encuentran: Alemania, Francia, Holanda, España, Italia y Portugal.
- La comunidad liderada por **Suiza** está formada por 17 países, entre los cuáles se encuentran: Suecia, Finlandia, Hungría, Turquía y Eslovenia.
- La comunidad liderada por **Bélgica** está formada por 8 países, entre los cuáles se encuentran: Dinamarca, Irlanda, Luxemburgo, Rumanía, Estonia y Ucrania.
- La comunidad liderada por **Israel** está formada por 8 países, entre los cuáles se encuentran: Austria, Grecia, Chipre, Croacia, Eslovaquia, y Serbia.

Los fondos destinados a Reino Unido son de unos 1960 millones de euros, que supone entorno al 25 % del total del capital destinado a Excellent Science [21]. Alemania, con una asignación absoluta de unos 1830 millones de euros, es el segundo país con más fondos en la comunidad tras Reino Unido, y es sobre quien recaería el liderazgo de la misma. La comunidad liderada por Suiza queda absorbida casi al completo por la liderada por Alemania. Este hecho sucede porque aunque segmentemos en comunidades los países en la realidad existen numerosas colaboraciones entre gran parte de ellos, la salida de Reino Unido conlleva la supresión de muchos proyectos (enlaces en nuestra red) pero, a cambio, otras relaciones cobran mayor importancia (enlaces que pasan a tener un peso mayor sobre un total que se ha visto disminuido).

La tabla 4.1 muestra el valor absoluto de las medidas sobre la red completa del programa Excellent Science, los valores en porcentaje de la componente gigante de la red resultante de eliminar los nodos de Reino Unido, y el valor promedio de las redes resultantes de eliminar

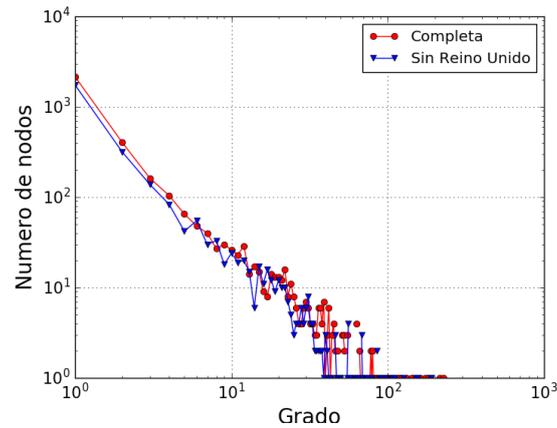


Figura 4.3: Distribución de grado en escala logarítmica del programa Excellent Science.

Medida	Red completa	Red sin Reino Unido (%)	Red aleatoria (%)
Nodos	3380	80	83 ± 2
Enlaces	8460	72	80 ± 3
Grado medio	5	90	96 ± 2
Grado medio pesado	0.533	87	92 ± 2
Asortatividad	-0.122	101	102 ± 5
Transitividad	0.071	91	99 ± 5
Clustering	0.078	94	97 ± 6
Eficiencia	0.274	97	100 ± 1
Camino geodésico medio	3.88	104	99 ± 1

Tabla 4.1: Medidas en la red de proyectos Excellent Science.

nodos de manera aleatoria. Los centros de Reino Unido que participan en este programa son casi 400, sin embargo, con ellos se van otros nodos que únicamente participaban en proyectos donde centros británicos son el centro coordinador. Comprobamos que la red se reduce a un 80 %, un 3 % menos del resultado de eliminar 400 nodos al azar, aunque teniendo en cuenta el margen de ± 2 podemos afirmar que no es una variación muy significativa. En cambio, se aprecia que el número de enlaces es inferior: la red se vuelve menos densa al eliminar nodos de Reino Unido que de manera aleatoria y, en consecuencia, el número medio de colaboradores por centro también desciende.

El Brexit tiene un efecto perceptible en la transitividad, eficiencia y camino geodésico de la red Excellent Science. La ausencia de numerosos enlaces conlleva que la conexión global de la red se vea diluida. Esto se explica porque tras el Brexit algunos centros que tenían en común centros de Reino Unido con sus vecinos pierden ese enlace. También se ve reflejado en la eficiencia, que se ve resentida ya que sin los nodos británicos la comunicación entre otros centros es peor, a la vez que las distancias entre los mismos aumenta. Por último, observamos que la salida de Reino Unido apenas afecta a la asortatividad.

4.2. Industrial Leadership

Este programa destina fondos a los proyectos que invierten en la investigación en el ámbito de las tecnologías de las comunicaciones, nanotecnología, biotecnología, manufactura y tecnologías espaciales entre otras. De todos los programas éste es el segundo en el que más instituciones participan en proyectos, entorno a 6100 centros.



Figura 4.4: Comunidades de colaboración entre países en Industrial Leadership, identificadas por el país con mayor presencia en cada una de ellas.

Observando las gráficas de la figura 4.4 vemos que las comunidades apenas se ven alteradas. Antes de eliminar los nodos de Reino Unido las configuraciones de las comunidades eran las siguientes (de nuevo mencionaremos sólo los países más representativos en orden de mayor a menor participación):

- La comunidad liderada por **Alemania** está formada por 55 países entre los cuáles destacan: España, Reino Unido, Francia, Italia, Suiza y Grecia.
- La comunidad liderada por **Holanda** está formada por 8 países entre los que destacan: Bélgica, República Checa y Serbia.
- La comunidad liderada por **Irlanda** está formada por 24 países entre los que destacan: Israel, Eslovenia, Rumanía y Letonia.

Tras la salida de Reino Unido los países permanecen en las mismas comunidades, no se ve alterada esta estructura pese a que la asignación de fondos a Reino Unido, en torno a unos 640 millones de euros, representa el 10% del total del capital destinado a este programa, una cantidad nada despreciable. Esto nos lleva a pensar que Reino Unido no es coordinador de tantos proyectos como en el caso anterior (hablando siempre en porcentajes, pues esta red es el doble de grande que la de Excellent Science), o que sus colaboradores sobreviven a su salida porque tienen muchas relaciones con otros países de la Unión Europea.

Observando la tabla 4.2 corroboramos lo anteriormente planteado: en este caso la salida de Reino Unido no “arrastra” tantos centros consigo como sucedía en el programa anterior. Los centros de Reino Unido que participan en este programa son entorno a 940, y con ellos desaparecen unos 340 centros, un porcentaje alto pero similar al de nodos que se van cuando los centros que se eliminan se escogen al azar. La cantidad de enlaces disminuye sustancialmente con respecto a la red completa; sin embargo, no se aprecian diferencias entre la red sin Reino Unido y las redes aleatorias. Fijándonos en el valor del número de enlaces del caso anterior,

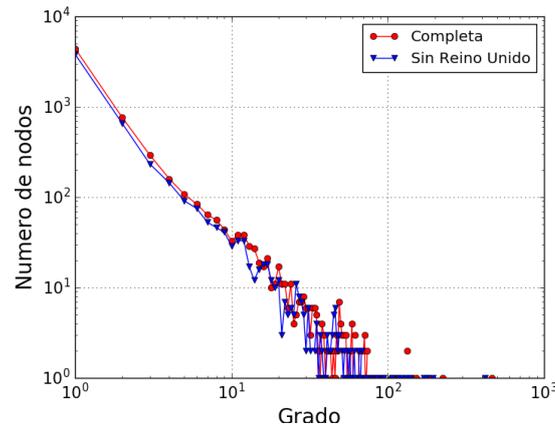


Figura 4.5: Distribución de grado en escala logarítmica del programa Industrial Leadership.

Medida	Red completa	Red sin Reino Unido (%)	Red aleatoria (%)
Nodos	6100	85	84 ± 2
Enlaces	10930	82	81 ± 2
Grado medio	4	97	97 ± 1
Grado medio pesado	0.372	96	87 ± 1
Asortatividad	-0.101	107	103 ± 8
Transitividad	0.027	103	100 ± 4
Clustering	0.052	106	95 ± 5
Eficiencia	0.255	100	98 ± 1
Camino geodésico medio	4.24	100	100 ± 2

Tabla 4.2: Medidas en la red de proyectos Industrial Leadership.

una disminución hasta el 80%, vemos que el dato es en proporción prácticamente igual al de las redes aleatorias, lo que constata que Reino Unido no juega un papel tan importante en Industrial Leadership como lo hace en Excellent Science. El grado medio de los nodos de los nodos se mantiene constante, lo que significa que se han ido nodos con grado heterogéneo, hecho que podemos confirmar con la figura 4.5 donde vemos que la línea azul se mantiene ligeramente por debajo de la roja en todo el tramo de $k = 1$ hasta algo más de $k = 10$.

Por el contrario, el grado medio pesado de la red sin Reino Unido se mantiene alto en comparación con la red aleatoria, indicando que los centros de Reino Unido no suponen las instituciones que más fondos tienen asignados en los proyectos, más bien al contrario, puesto que la elección de centros al azar para su eliminación ha llevado a una caída importante en esta magnitud. La asortatividad se ve afectada en ambos casos, resulta superior en el caso de Reino Unido pero no muy lejos del caso de redes aleatorias. El valor de la transitividad muestra que en ambos casos las redes mantienen la cohesión global. El clustering crece tras el Brexit, mostrando una estructura de agrupamiento de centros, mientras que en el caso de las redes aleatorias la estructura no presenta regiones tan densamente conexas en comparación con la estructura global. Por último, ni la eficiencia ni el camino geodésico medio se ven modificados.

4.3. Societal Challenges

En este programa se incluyen los proyectos destinados a la investigación en temas de salud, cambios demográficos y bienestar, seguridad alimentaria, agricultura sostenible, energías eficientes y cuidado del entorno natural entre otros. Este es con diferencia el programa en el que más instituciones participan: más de 12000 centros que resulta aproximadamente el doble que en Industrial Leadership y el triple que en Excellent Science.



Figura 4.6: Países líderes de las comunidades que más producen en Societal Challenges.

Atendiendo a las gráficas de la figura 4.6 vemos que las comunidades aparecen notablemente alteradas. Antes de eliminar los nodos de Reino Unido las configuraciones de las comunidades eran las siguientes (de nuevo mencionaremos sólo los países más representativos en orden decreciente según su participación):

- La comunidad liderada por **Alemania** está formada por 59 países entre los cuáles destacan: Reino Unido, Francia, Holanda, Bélgica, Suecia y Austria.
- La comunidad liderada por **España** está formada por 43 países entre los que destacan: Italia, Grecia, Finlandia, Portugal, Eslovenia e Israel.
- La comunidad liderada por **Dinamarca** está formada por 20 países entre los que destacan: Noruega, Hungría, Islandia, Croacia y Suiza.

La asignación de fondos a Reino Unido en el programa de Societal Challenge es alrededor de 1180 millones de euros, que supone casi un 14% del total. Es el segundo país por detrás de Alemania, quien percibe un 15%.

El efecto del Brexit reestructura las comunidades: la comunidad de Alemania no sólo pierde a Reino Unido, sino que su comunidad se ve reducida a 42 países. Holanda y Bélgica, países que más colaboraban en esta comunidad, pasan a integrarse en la comunidad liderada por España, que también ve reducido su número de países integrantes a 34. La comunidad más “beneficiada” es la que aparece liderada por Italia, con un total de 42 países integrantes entre los que se encuentran Grecia, Portugal e Irlanda.

La tabla 4.3 nos muestra que las medidas sobre la red sin los nodos de Reino Unido comparada con la media de las medidas sobre redes aleatorias apenas cambian. Los centros de Reino Unido que participan en este programa son unos 1200, y son 730 los que desaparecen con ellos tras el Brexit. El número de instituciones y colaboraciones que se lleva consigo es similar al caso

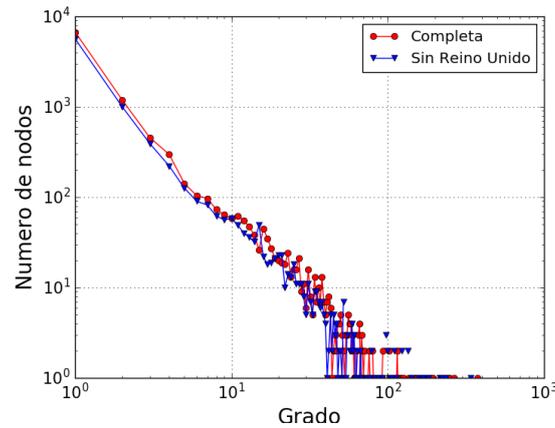


Figura 4.7: Distribución de grado en escala logarítmica del programa Industrial Leadership.

Medida	Red completa	Red sin Reino Unido (%)	Red aleatoria (%)
Nodos	9780	83	84 ± 1
Enlaces	19390	79	82 ± 2
Grado medio	4	95	97 ± 1
Grado medio pesado	0.307	93	92 ± 1
Asortatividad	-0.127	104	101 ± 4
Transitividad	0.035	99	100 ± 4
Clustering	0.063	102	97 ± 3
Eficiencia	0.247	100	99 ± 1
Camino geodésico medio	4.32	100	101 ± 1

Tabla 4.3: Medidas en la red de proyectos Societal Challenges.

de las redes aleatorias y, en consecuencia, el grado medio se mantiene parecido. Atendiendo a la gráfica 4.7 vemos que los grados de los centros que desaparecen son muy heterogéneos: de nuevo la línea azul se aprecia bien que se mantiene por debajo de la roja en el tramo de 1 a 30. En el caso de la red sin Reino Unido la asortatividad es superior a la de la red completa, mientras que en la red aleatoria se mantiene muy parecida. La transitividad de la red se mantiene, por lo que a nivel global la densidad de enlaces entre distintas regiones no se ve afectada, sin embargo el valor del coeficiente de clustering es mayor tras el Brexit que en el caso de las redes aleatorias. Este hecho nos lleva a pensar que los centros de Reino Unido participan en numerosos proyectos con una gran variedad de países conectando así pequeñas subcomunidades, y tras el Brexit esos enlaces se pierden y dejan entrever una estructura con un nivel mayor de agrupamientos.

Capítulo 5

Conclusiones

Después de haber analizado las redes de los programas Excellent Science, Industrial Leadership y Societal Challenges, podemos asegurar que la salida de Reino Unido de la Unión Europea va a tener impacto en ellos, pero este va a depender del programa en cuestión.

Así en el caso de Excellent Science el capital asignado a Reino Unido supone un 25 % del total de fondos destinados al programa; la comunidad liderada por Suiza queda englobada en la comunidad liderada por Alemania, donde anteriormente Reino Unido era el país con mayor presencia. La red sin Reino Unido pierde muchos proyectos y, en consecuencia, colaboraciones; la cohesión general de la red disminuye, aumentando el agrupamiento local de centros.

En Industrial Leadership Reino Unido recibe fondos en menor proporción que en el programa anterior, un 10 %. Ahora las colaboraciones que se pierden son similares al caso de la salida del mismo número de centros elegidos al azar; la cantidad media de colaboraciones aumenta, por lo que intuimos que desaparecen muchos centros que tienen pocos enlaces. En este programa, al contrario que en Excellent Science, tanto la cohesión de toda la red como la clusterización de los centros es mayor tras la eliminación de los nodos de Reino Unido que en el caso de la salida aleatoria de centros; es decir, la red está más conectada a escala global y a escala local.

En Societal Challenges (recordamos que es el mayor de los programas en cuanto a número de participación de centros) la cantidad de centros y proyectos que desaparecen es parecido al caso de la red aleatoria. En este programa Reino Unido dispone del 14 % de los fondos destinados a este tipo de proyectos. Su ausencia no afecta a la cohesión global de la red; sin embargo sí afecta al coeficiente de clustering. Esta medida aumenta cuando eliminamos nodos de Reino Unido, indicando que aparecen más grupos locales con mayor densidad de colaboraciones cuando eliminamos que cuando desaparecen centros al azar.

En resumen, podemos concluir que:

1. Se produce un cambio radical en la estructura de comunidades. La remoción de Reino Unido conlleva un reagrupamiento de las comunidades y la polarización en dos comunidades predominantes.
2. En cuanto a programas, el programa que más ve afectada su estructura de colaboraciones tras el Brexit es el de Excellent Science, ya que conlleva una disminución significativa de la cohesión de la red resultante.

El análisis del impacto del Brexit sobre la red de proyectos europeos presentado en este trabajo ofrece una visión general del escenario en el que la Unión Europea se verá en un futuro próximo. Para un estudio más profundo de esta situación es conveniente paralelizar los algoritmos del código que miden las propiedades de la red, con objeto de agilizar los cálculos para: calcular otras medidas como la comunicabilidad y el diámetro que ofrezcan información sobre las dimensiones de la red pero tienen un alto coste computacional, poder incluir más nodos analizados en cada red permitiendo así una mayor precisión en los valores estimados, generar un mayor número de redes aleatorias sobre las que hacer estadística. Además, convendría analizar el resto de programas (Spreading Excellence and Widening Participation, Science with and for Society y Euratom) tanto por separado como los seis conjuntamente, con la finalidad de obtener medidas de toda la red de proyectos europeos de Horizon 2020.

Agradecimientos

Quiero agradecer a mis directores David Iñiguez y Jesús Gómez por sus consejos y recomendaciones, que me han ayudado a ser crítica y perseguir el rigor científico en el desarrollo del trabajo y los resultados obtenidos, competencias que me llevaré conmigo en mi vida profesional.

Además, quiero agradecer a Alfonso Tarancón y Gonzalo Ruiz por su disposición y ayuda a la hora de trabajar con las bases de datos de Horizon 2020, con el manejo de la herramienta Kampal, así como con el desarrollo mismo del trabajo elaborado.

Bibliografía

- [1] Ernesto Estrada, Philip A Knight, and Philip Knight. *A first course in network theory*. Oxford University Press, USA, 2015.
- [2] Vito Latora, Vincenzo Nicosia, and Giovanni Russo. *Complex networks: principles, methods and applications*. Cambridge University Press, 2017.
- [3] Newman Mej. *Networks: an introduction*, 2010.
- [4] Leonhard Euler. Leonhard euler and the königsberg bridges. *Scientific American*, 189(1):66–72, 1953.
- [5] Problema de los puentes de konigsberg. <https://es.wikipedia.org/>.
- [6] Kenneth Appel and Wolfgang Haken. Every planar map is four colorable. *Bulletin of the American mathematical Society*, 82(5):711–712, 1976.
- [7] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440, 1998.
- [8] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [9] Alfonso Allen-Perkins, Juan Manuel Pastor, and Ernesto Estrada. Two-walks degree assortativity in graphs and networks. *Applied Mathematics and Computation*, 311:262–271, 2017.
- [10] Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.
- [11] Martin Rosvall, Jean-Charles Delvenne, Michael T Schaub, and Renaud Lambiotte. Different approaches to community detection. *arXiv preprint arXiv:1712.06468*, 2017.
- [12] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- [13] Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. In *International symposium on computer and information sciences*, pages 284–293. Springer, 2005.
- [14] Gephi. <https://gephi.org>.

- [15] Amgad Madkour, Walid G Aref, Faizan Ur Rehman, Mohamed Abdur Rahman, and Saleh Basalamah. A survey of shortest-path algorithms. *arXiv preprint arXiv:1705.02044*, 2017.
- [16] Mark Newman, Albert-Laszlo Barabasi, and Duncan J Watts. *The structure and dynamics of networks*, volume 19. Princeton University Press, 2011.
- [17] Ernesto Estrada and Naomichi Hatano. Communicability in complex networks. *Physical Review E*, 77(3):036111, 2008.
- [18] Paolo Crucitti, Vito Latora, Massimo Marchiori, and Andrea Rapisarda. Efficiency of scale-free networks: error and attack tolerance. *Physica A: Statistical Mechanics and its Applications*, 320:622–642, 2003.
- [19] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182, 1987.
- [20] Portal de datos abiertos de la u.e. <https://data.europa.eu>.
- [21] Kampal research. <http://research.kampal.com>.