



UNIVERSITAT D'ALACANT

Facultat de Ciències

Facultad de Ciencias

**FACULTAD DE CIENCIAS**

**GRADO EN MATEMÁTICAS**

**TRABAJO FIN DE GRADO**

**CURSO ACADÉMICO [2023-2024]**

TÍTULO:

**TEORÍA DE REDES: FUNDAMENTOS Y APLICACIONES EN LAS  
CIENCIAS CLIMÁTICAS**

AUTOR:

**ANA GIL MOLINA**

## Resumen

*Este trabajo de fin de grado se centra en la exploración de los fundamentos matemáticos de la Teoría de Redes y su aplicación en la ciencia del clima. Dado que el clima es un sistema complejo con múltiples variables interrelacionadas, la Teoría de Redes representa una herramienta poderosa para entender su estructura y dinámica.*

*Se inicia con una introducción que aborda brevemente la historia de la Teoría de Redes, la definición de red y los objetivos del trabajo. A continuación, se profundiza en los conceptos generales de la Teoría de Grafos, incluyendo definiciones, ejemplos de grafos y las métricas fundamentales para caracterizar las estructuras de red.*

*Posteriormente, se examinan las redes aleatorias, estudiando diferentes modelos como el de Erdős-Rényi, el de Barabási-Albert o el de Watts-Strogatz, así como su implementación en Python. También se introducen los mundos pequeños y los coeficientes de agrupamiento de redes.*

*Finalmente, se aborda el estudio de la temperatura del aire a escala global utilizando la Teoría de Redes. Este último capítulo se dedica a aplicar los conocimientos adquiridos en la Teoría de Redes para analizar datos climáticos realistas provenientes de un modelo numérico. Este enfoque busca no solo proporcionar una mejor comprensión de la Teoría de Redes, sino también aplicar estos conocimientos en un contexto práctico y demostrar su potencial utilidad para abordar desafíos climáticos actuales, como la identificación de patrones climáticos y sus teleconexiones.*

**Palabras clave:** red, grafo, nodo, arco, clima

## Abstract

*This final project focuses on exploring the mathematical foundations of Network Theory and its application in climate science. Given that climate is a complex system with multiple interrelated variables, Network Theory represents a powerful tool to understand its structure and dynamics.*

*It begins with an introduction that briefly addresses the history of Network Theory, the definition of a network, and the objectives of the project. Following that, it delves into the general concepts of Graph Theory, including definitions, graph examples, and the fundamental metrics for characterizing network structures.*

*Subsequently, random networks are examined, studying different models such as Erdős-Rényi, Barabási-Albert, or Watts-Strogatz, as well as their implementation in Python. Small-world phenomenon and network clustering coefficients are also introduced.*

*Finally, the study of global air temperature using Network Theory is addressed. This final chapter is dedicated to applying the acquired knowledge in Network Theory to analyze realistic climate data from a numerical model. This approach aims not only to provide a better understanding of Network Theory, but also to apply this knowledge in a practical context and demonstrate its potential usefulness in addressing current climate challenges, such as the identification of climate patterns and their teleconnections.*

**Keywords:** network, graph, node, edge, climate

# Índice

<b>1. Introducción</b>	<b>5</b>
1.1. Historia de la Teoría de Redes . . . . .	5
1.2. Definición de Red: Estructura y Dinámica . . . . .	6
1.3. Objetivos del Trabajo de Fin de Grado . . . . .	7
<b>2. Conceptos generales de la Teoría de Grafos</b>	<b>8</b>
2.1. Definiciones de la Teoría de Grafos . . . . .	8
2.1.1. Caminos y circuitos . . . . .	10
2.1.2. Representación matricial de redes . . . . .	11
2.2. Ejemplos de grafos . . . . .	13
2.3. Métricas: Medidas de centralidad . . . . .	18
2.3.1. Centralidad de grado . . . . .	18
2.3.2. Centralidad de cercanía . . . . .	19
2.3.3. Centralidad de intermediacin . . . . .	21
2.3.4. Ejemplos de medidas de centralidad . . . . .	22
<b>3. Redes aleatorias</b>	<b>27</b>
3.1. El modelo de Erdős-Renyi de redes aleatorias . . . . .	27
3.1.1. Distribucin de grado . . . . .	30
3.2. El modelo de Barabsi-Albert de redes aleatorias . . . . .	31
3.3. Mundos pequeos . . . . .	32
3.4. Coeficientes de agrupamiento de redes . . . . .	33
3.4.1. El coeficiente de agrupamiento de Watts-Strogatz . . . . .	33
3.4.2. El coeficiente de agrupamiento de Newman . . . . .	34
3.5. El modelo de Watts-Strogatz . . . . .	34
3.6. Ejemplo: ´Arbol de Cayley . . . . .	36
3.7. Implementacin de redes aleatorias en Python . . . . .	40
3.7.1. Modelo de ´Erdos-Renyi en Python . . . . .	40
3.7.2. Modelo de Barabsi-Albert en Python . . . . .	41

---

3.7.3. Modelo de Watts-Strogatz en Python . . . . .	43
<b>4. Aplicación de la Teoría de Redes al clima: buscando patrones en la temperatura del aire</b>	<b>44</b>
4.1. Pasos en la construcción de la red de temperatura del aire . . . . .	45
4.2. Estudio de las propiedades de la red obtenida mediante representaciones gráficas . . . . .	47
<b>5. Conclusiones</b>	<b>54</b>
<b>Referencias</b>	<b>55</b>
<b>A. Detalles del procesamiento de datos de temperatura del aire para el cálculo de la matriz de adyacencia</b>	<b>56</b>
A.0.1. Conservación de los valores NaN . . . . .	56
A.0.2. Interpolación lineal para estimar los valores NaN . . . . .	57
<b>B. Detalles del cálculo de las correlaciones</b>	<b>60</b>
<b>C. Detalles del cálculo de la matriz de adyacencia</b>	<b>62</b>
<b>D. Detalles del desarrollo del trabajo</b>	<b>63</b>

## 1. Introducción

### 1.1. Historia de la Teoría de Redes

La ciencia de redes, originalmente fundamentada en la teoría de grafos, ha experimentado una expansión significativa al nutrirse de diversas disciplinas. En la actualidad, ha evolucionado al incorporar aportes del análisis de redes sociales y la teoría de control, entre otras. Sus diversas aplicaciones abarcan campos tan variados como las ciencias físicas, donde se utiliza, por ejemplo, en la dinámica del clima; las ciencias biológicas, donde se aplica en la jerarquización de redes tróficas; y las ciencias computacionales, donde se emplea en la teoría de la información, entre otros muchos campos.

Se pueden distinguir tres etapas fundamentales en el desarrollo de esta disciplina. En primer lugar, se ubica el periodo temprano pre-redes (1736-1966), en el cual la ciencia de redes se manifestaba principalmente como la matemática de los grafos. Posteriormente, se desarrolló el periodo meso-redes (1967-1998), marcado por la aparición de aplicaciones concretas de las redes. Por último, la etapa moderna (1998-presente), se caracteriza por el establecimiento de los fundamentos actuales de la ciencia de redes, además de la aplicación de la misma a diversos campos, como los arriba mencionados (Lewis, 2009).

El inicio de la teoría de grafos tuvo lugar en 1736, cuando Leonhard Euler empleó grafos para resolver un problema práctico: encontrar un recorrido para cruzar a pie toda la ciudad pasando solo una vez por cada uno de los puentes y regresando al mismo punto de inicio (problema de los puentes de Königsberg). Euler proporcionó una solución general para problemas de red, introduciendo conceptos como el de grafo, nodo, arco, grado de nodo y camino (Estrada and Knight, 2015).

En la década de 1950, el matemático húngaro Paul Erdős (1913-1996) revitalizó la teoría de grafos y creó la rama conocida como matemáticas discretas. El procedimiento generativo de Erdős-Rényi para construir grafos aleatorios marcó un segundo hito histórico en 1959-60.

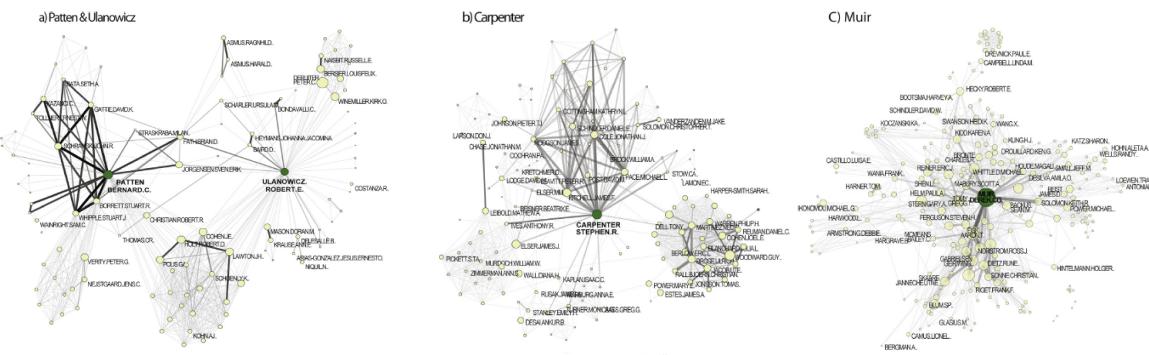
En las décadas de 1960 y 1970, la teoría de grafos fue empleada para modelar redes sociales y estudiar el comportamiento humano en grupos. Stanley Milgram introdujo la noción de red de mundo pequeño, generando interés en investigaciones sobre cómo la topología de las redes podría influir en el comportamiento humano (Milgram, 1967).

En la década de 1990, algunos científicos comenzaron a utilizar las redes como modelos de fenómenos físicos y biológicos. En particular, el trabajo de Duncan Watts, Steven Strogatz (Watts and Strogatz, 1998) y Albert-László Barabási (2016) generó un notable

interés en la aplicación del análisis matemático de redes a la física.

En la actualidad, la ciencia de redes se encuentra en permanente evolución. Se continúan realizando descubrimientos y desarrollos de forma mensual en distintas disciplinas (Lewis, 2009). Por ejemplo, algunos avances recientes incluyen el impacto de las Redes Neuronales de Grafos (GNNs) en la clasificación de nodos y la predicción de enlaces (Hamilton et al., 2017), la detección de comunidades, es decir, estructuras clave en redes, revelando patrones organizativos en sistemas complejos (Blondel et al., 2008), la adaptación de modelos de grafos a entornos dinámicos (Wang et al., 2018), entre otros.

En la Figura 1 se muestran como ejemplo diferentes redes ecológicas usadas para jerarquizar la relación entre especies dentro de un ecosistema.



**Figura 1:** Figura que presenta tres ejemplos de redes en el contexto de la ecología: (a) Red de Bernard C. Patten y R.E. Ulanowicz, (b) Red de S.R. Carpenter y (c) Red de D.C.G. Muir. La teoría de redes destaca en la ecología por sus contribuciones al análisis de redes alimenticias y ciclos en los ecosistemas, así como por el desarrollo del software Ecopath para análisis de redes (Borrett et al., 2014).

## 1.2. Definición de Red: Estructura y Dinámica

Los grafos están formados por nodos, arcos y una función de mapeo que define cómo se conectan los nodos. En los grafos estáticos, las propiedades como el número de nodos y arcos o la forma de la función de mapeo permanecen constantes en el tiempo. Sin embargo, en los grafos dinámicos, estas propiedades varían con el tiempo. Partiendo de la teoría de grafos se pueden definir las redes (Lewis, 2009).

En una red, se representan diversas estructuras en las que las entidades del sistema complejo son nodos, mientras que las relaciones o interacciones entre estas entidades se capturan mediante los arcos de la red (Estrada and Knight, 2015). Por tanto, la teoría de

redes es el estudio de la estructura de una colección de nodos y arcos que representan algún aspecto real, así como del comportamiento dinámico de la agrupación de dichos nodos y arcos. Es importante tener en cuenta que la teoría de redes estudia modelos aproximados de la realidad, pero no la realidad en sí misma.

La estructura de una red se puede definir gracias a la teoría de grafos. Un grafo se define como un conjunto  $G = \{V, E, f\}$ , donde  $V$  es un conjunto de nodos,  $E$  un conjunto de arcos, y  $f : E \rightarrow V \times V$  una función de mapeo que define la estructura de  $G$ , es decir, cómo los nodos están conectados entre sí a través de los arcos. La función  $f$ , llamada topología de  $G$ , permite dibujar el grafo usando puntos como nodos y líneas como arcos.

Por otro lado, para definir el comportamiento dinámico de la red es necesario describir el comportamiento de los nodos y los arcos en el tiempo. Por tanto, se define la red como  $G(t) = \{V(t), E(t), f(t) : J(t)\}$ , donde los conjuntos de nodos  $V$  y de arcos  $E$ , así como la función de mapeo  $f$ , son función del tiempo  $t$ , y en la que se añade  $J$  un algoritmo que describe el comportamiento de los nodos y los arcos frente al tiempo. De esta forma, una red es un grafo dinámico.

En conclusión, la teoría de redes es el estudio de los fundamentos teóricos de la estructura y dinámica de las redes, y la aplicación de las redes a diversos subcampos, entre los cuales se encuentran el análisis de redes sociales, los sistemas de ciencias físicas y de ciencias de la vida, las redes de colaboración, internet, entre otros (Lewis, 2009).

### 1.3. Objetivos del Trabajo de Fin de Grado

Los objetivos principales del trabajo son en primer lugar realizar una revisión de la literatura sobre la Teoría de Redes y sus fundamentos matemáticos. Esto implica comprender los conceptos básicos de la teoría de grafos y las métricas más relevantes para caracterizar las estructuras en forma de red.

Además, se exploran algunos de los algoritmos que se emplean para generar redes, destacando entre ellas las redes aleatorias.

Por último, se profundiza en los conceptos mediante una aplicación de la Teoría de Redes en el estudio del sistema climático. Se busca aplicar los algoritmos revisados en un contexto climático concreto: el estudio de la temperatura del aire a 2 metros sobre el nivel del mar entre los años 1940 y 2024 a escala global.

Estos objetivos buscan proporcionar una comprensión sólida de la Teoría de Redes y de sus posibilidades en un contexto realista, poniendo a prueba para ello algunas de las habilidades y conocimientos adquiridos a lo largo de los estudios de grado.

## 2. Conceptos generales de la Teoría de Grafos

### 2.1. Definiciones de la Teoría de Grafos

En esta sección se definen los conceptos generales de la teoría de grafos, que serán utilizados para el estudio de las redes.

Sea  $V$  un conjunto finito y sea  $E \subseteq V \otimes V$ , cuyos elementos no son necesariamente todos distintos. Se dice que:

- $E$  es **reflexivo** si  $(v, v) \in E, \forall v \in V$
- $E$  es **antirreflexivo** si  $(v, v) \notin E, \forall v \in V$
- $E$  es **simétrico** si  $(v_1, v_2) \in E \iff (v_2, v_1) \in E$

**Definición 2.1** *Un grafo o red,  $G$ , es un par  $(V, E)$ , tal que  $V$  es un conjunto cuyos elementos son los **nodos** o **vértices** de  $G$ , y  $E$  es un conjunto de **arcos** o **enlaces** entre los elementos de  $V$ . Según Lewis (2009), en un **grafo estático**, las propiedades se mantienen en el tiempo, mientras que en un **grafo dinámico**, ciertas propiedades como el número de nodos y arcos varían con el tiempo. En ocasiones, se habla de red para referirse a un grafo dinámico.*

**Definición 2.2** *Sea  $G = (V, E)$  un grafo o red. Se dice que:*

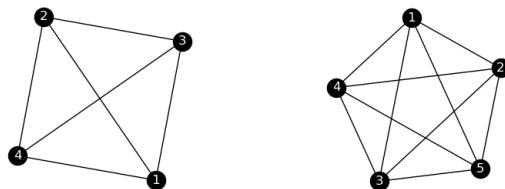
- $G$  es una **red no dirigida** si  $E$  es simétrico.
- $G$  es una **red simple** si  $E$  es simétrico y antirreflexivo y no contiene arcos duplicados.
- $G$  es un **multigrafo** si  $E$  contiene arcos duplicados, es decir, más de un arco relacionando los mismos nodos.
- $G$  es una **red dirigida o digrafo** si  $E$  es asimétrico.

**Definición 2.3** *Sea  $G = (V, E)$  un grafo. Se definen los siguientes conceptos elementales de la teoría de grafos:*

(I)  $G_s = (V_s, E_s)$  es un **subgrafo** de  $G$  si  $V_s \subseteq V$  y  $E_s \subseteq V_s \otimes V_s \cap E$ .

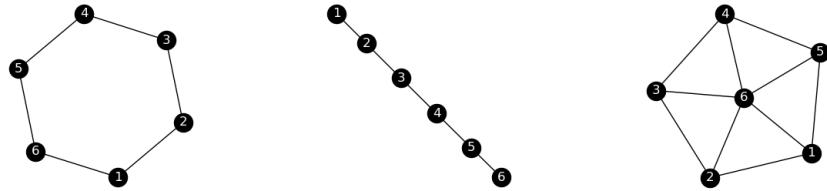
(II) Un **bucle** es un arco de la forma  $(v, v)$ . Las redes simples no tienen bucles.

- (III) Sea  $e = (v_1, v_2)$  un arco de  $G$ . Se dice que  $v_1$  es incidente a  $e$ ,  $v_2$  es incidente desde  $e$  y  $v_1$  es **adyacente** a  $v_2$ . Además, siguiendo la nomenclatura de Lewis (2009), se denomina **cola** a  $v_1$  y **cabeza** a  $v_2$ . Si  $G$  es no dirigido,  $v_1$  adyacente a  $v_2 \implies v_2$  adyacente a  $v_1$ .
- (IV)  $G_1 = (V_1, E_1)$  y  $G_2 = (V_2, E_2)$  son **isomórficos** si hay una correspondencia inyectiva  $f : V_1 \rightarrow V_2$  tal que para todos  $u, v \in V_1$ , el número de arcos entre  $u$  y  $v$  coincide con el número de arcos entre  $f(u)$  y  $f(v)$ .
- (V) Si  $G$  es simple, su **complementario**  $\overline{G} = (V, \overline{E})$  es la red simple con los mismos nodos que  $G$  tal que  $(u, v) \in E \iff (u, v) \notin \overline{E}$ .
- (VI) Sean  $G_1 = (V_1, E_1)$  y  $G_2 = (V_2, E_2)$  tal que  $V_1 \cap V_2 = \emptyset$ . La **unión**  $G_1 \cup G_2$  tiene conjunto de nodos  $V_1 \cup V_2$  y conjunto de arcos  $E_1 \cup E_2$ . Esta unión tendrá al menos dos componentes disjuntas. Además,  $G_1$  y  $G_2$  son subgrafos de  $G_1 \cup G_2$ .
- (VII) Sea  $F \subseteq E$ . Entonces  $H = (V, E - F)$  es el subgrafo obtenido al eliminar los arcos en  $F$  de  $G$ . Si  $\widehat{G} = (V, F)$ , entonces se escribe  $H = G - \widehat{G}$ .
- (VIII) El **grado de salida** de un nodo  $v \in V$ ,  $\text{out\_d}(v)$ , es el número de arcos a los que es incidente, y su **grado de entrada**,  $\text{in\_d}(v)$ , es el número de arcos desde los que es incidente. En redes no dirigidas, ambos coinciden y se les llama **grado** del nodo, denotado  $d(v)$ . Según Lewis (2009), se llama **hub** al nodo (o nodos) con mayor grado. Un nodo con grado 0 se llama **nodo aislado**, y uno con grado 1 se llama **nodo final**.
- (IX) Se llama **red completa** con  $n$  nodos, denotada  $K_n$ , a aquella en la que hay un arco entre cada nodo distinto. En este caso, el grado de cada nodo es  $n - 1$ .  
En la Figura 2 se muestran dos ejemplos de redes completas con 4 y 5 nodos.



**Figura 2:** Las redes completas  $K_4$  y  $K_5$ .

- (X) Una red con  $n$  nodos pero sin arcos se llama **grafo nulo**, y se denota  $N_n$ . En un grafo nulo, cada nodo es aislado.
- (XI) Una red es  **$k$ -regular** o **regular de grado  $k$**  si todos los nodos tienen el mismo grado  $k$ . Por ejemplo,  $N_n$  es 0-regular, y  $K_n$  es  $(n-1)$ -regular.
- (XII) En un **grafo ciclo** con  $n$  nodos,  $C_n$ , los nodos se pueden ordenar de manera que cada uno esté conectado a sus vecinos inmediatos. Si se quita un arco de  $C_n$  se obtiene un **grafo camino**,  $P_{n-1}$ . Si se añade un nodo a  $C_n$  y se conecta a todos los demás nodos se obtiene un **grafo rueda**,  $W_{n+1}$ .
- En la Figura 3 se representan un grafo ciclo, un grafo camino y un grafo rueda con 6 nodos cada uno.



**Figura 3:** Un ciclo  $C_6$ , un camino  $P_5$  y una rueda  $W_6$  con 6 nodos.

### 2.1.1. Caminos y circuitos

**Definición 2.4** Un **camino** es una secuencia de arcos (no necesariamente distintos):

$$(u_1, v_1), (u_2, v_2), \dots, (u_p, v_p)$$

tal que  $v_i = u_{i+1}$  ( $i = 1, 2, \dots, p - 1$ ). Si  $v_p = u_1$  el camino es **cerrado**.

Un **recorrido** es un camino en el que todos los arcos son distintos. Un **camino simple** es un recorrido en el que todos los  $u_i$  son distintos. Un camino simple cerrado se llama **ciclo** o **circuito**. Un grafo sin ciclos se llama **acíclico**. Según Lewis (2009), un ciclo de longitud 1 se llama **bucle**, y uno de longitud 3 se llama **triángulo**.

La **longitud** de un camino/recorrido/camino simple viene dada por  $p$ , y se mide en **saltos** (número de arcos a lo largo del camino).

**Observación 2.5** Un camino simple de longitud  $p$  induce un subgrafo  $P_p$ , y un ciclo de longitud  $p$  induce un subgrafo  $C_p$  (Estrada and Knight, 2015).

**Definición 2.6** El *camino más corto*, también llamado *camino directo* entre dos nodos  $u$  y  $v$ , es usado como el camino conectando ambos nodos. Su longitud se llama *distancia* entre  $u$  y  $v$ . La *longitud promedio de camino* de un grafo  $G$  es igual a la media de las longitudes de todos los caminos directos (Lewis, 2009).

**Teorema 2.7** El camino más corto entre dos nodos distintos es también el camino simple más corto entre ambos.

*Demostración.* Un camino de longitud 1 es también un camino simple.

Por inducción, para un camino de longitud mayor que 1 entre dos nodos  $u$  y  $v$ , se considera el camino sin el último arco, que conectará  $u$  con otro nodo  $w$ . Este será el camino más corto de  $u$  a  $w$ , por lo que por la hipótesis de inducción será un camino simple.

Este camino simple no contiene el nodo  $v$  ni el arco  $(w, v)$ , por lo tanto el camino de  $u$  a  $v$  debe ser también un camino simple. ■

**Observación 2.8** El camino cerrado más corto entre dos nodos no necesariamente es el ciclo más corto (Estrada and Knight, 2015).

**Definición 2.9** Se dice que existe un arco *duplicado* en un grafo  $G$  cuando más de un arco conecta el mismo par de nodos (Lewis, 2009).

**Definición 2.10** Una red está *conectada* si existe un camino conectando dos nodos cualesquiera.

**Observación 2.11** Si una red no está conectada, entonces se puede dividir en *componentes*, cada una de las cuales está conectada.

- En una red no dirigida, las componentes son disjuntas.
- En una red dirigida, puedes dejar una componente e ingresar en otra, pero no regresar. Una componente de una red dirigida de la que no puedes salir se llama *fuertemente conectada*.

**Definición 2.12** Un *árbol* es una red conectada sin ciclos. Un *bosque* es una unión de árboles (Estrada and Knight, 2015).

### 2.1.2. Representación matricial de redes

**Definición 2.13** Sea  $G = (V, E)$  un grafo, con  $V = \{1, 2, \dots, n\}$ .

Para  $1 \leq i, j \leq n$ , sea  $k$  el número de arcos que conectan  $i$  a  $j$ , y sea:

$$c_{ij} = \begin{cases} k, & (i, j) \in E \\ 0, & (i, j) \notin E \end{cases} \quad (1)$$

Entonces la matriz cuadrada  $C = (c_{ij})$  es la **matriz de conexión** de  $G$  (Lewis, 2009).

**Definición 2.14** Sea  $G = (V, E)$  un grafo, con  $V = \{1, 2, \dots, n\}$ .

Para  $1 \leq i, j \leq n$ , sea:

$$a_{ij} = \begin{cases} 1, & (i, j) \in E \\ 0, & (i, j) \notin E \end{cases} \quad (2)$$

Entonces la matriz cuadrada  $A = (a_{ij})$  es la **matriz de adyacencia** de  $G$  (Estrada and Knight, 2015).

**Observación 2.15** Si el grafo es no dirigido, entonces la matriz de conexión y la matriz de adyacencia son simétricas (Lewis, 2009).

**Observación 2.16** Si  $G$  es una red simple con matriz de adyacencia  $A$ , su complementario  $\bar{G}$  tiene matriz de adyacencia  $E - I - A$ , siendo  $E$  la matriz de unos e  $I$  la matriz identidad.

**Observación 2.17** En una red simple, los elementos de  $A^p$  indican el número de caminos de longitud  $p$  entre cada par de nodos. Así, la longitud del camino más corto entre dos nodos será el primer valor de  $p$  para el cual el elemento  $(i, j)$ -ésimo de  $A^p$  es distinto de cero. Esto se puede generalizar a cualquier red usando la matriz de conexión; así los elementos de  $C^p$  indicarán el número de caminos de longitud  $p$  entre cada par de nodos.

**Definición 2.18** Sea  $G = (V, E)$  un grafo donde  $V = \{1, 2, \dots, n\}$  y  $E = \{e_1, e_2, \dots, e_m\}$  con  $e_i = (u_i, v_i)$ . Para  $1 \leq i \leq m$  y  $1 \leq j \leq n$ , sea:

$$b_{ij} = \begin{cases} 1, & u_i = j \\ -1, & v_i = j \\ 0, & \text{en otro caso} \end{cases} \quad (3)$$

La matriz rectangular  $B = (b_{ij})$  es la **matriz de incidencia** de  $G$ .

Si  $e_i$  es un bucle, tomamos  $b_{iu_i} = 1$  y cualquier otro elemento de la fila  $i$  se toma 0.

Además, si  $G$  es simple y  $(u, v) \in E$ , entonces  $(v, u) \in E$ . En este caso, solo incluimos uno de estos pares en la matriz, que se llamará **matriz de incidencia orientada** (Estrada and Knight, 2015).

**Definición 2.19** Sea  $G = (V, E)$  un grafo, con  $V = \{1, 2, \dots, n\}$ .

Para  $1 \leq i, j \leq n$ , sea:

$$d_{ij} = \begin{cases} d(i), & j = i \\ 0, & \text{en otro caso} \end{cases} \quad (4)$$

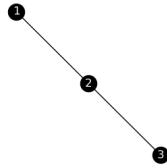
Entonces la matriz  $L = C - D$ , donde  $C$  es la matriz de conexión y  $D = (d_{ij})$  es la matriz de grados (diagonal), se llama la **matriz Laplaciana** de  $G$ .

**Definición 2.20** Sea  $G = (V, E)$  un grafo. La **matriz de camino**  $P$  almacena los saltos a lo largo del camino directo entre cada par de nodos, es decir, la longitud de los caminos más cortos. Si no existe ningún camino entre un par de nodos, el elemento correspondiente es cero. Además, los elementos diagonales son cero, pues no se tienen en cuenta los bucles ni los arcos duplicados (Lewis, 2009).

## 2.2. Ejemplos de grafos

En esta sección se presentan algunos ejemplos de grafos para poner en práctica los conceptos estudiados en la sección 2.1. Para representar los grafos, se usará la librería *NetworkX* de Python.

**Ejemplo 2.21** Sea  $G = (V, E)$  un grafo, donde  $V = \{1, 2, 3\}$  y  $E = \{(1, 2), (2, 1), (2, 3), (3, 2)\}$ , representado en la Figura 4.



**Figura 4:** Grafo no dirigido con 3 nodos.

- $E$  no es reflexivo, pues por ejemplo  $1 \in V$  pero  $(1, 1) \notin E$ .
- $E$  es antirreflexivo, pues  $(1, 1), (2, 2), (3, 3) \notin E$ .
- $E$  es simétrico, pues se cumple que  $(1, 2), (2, 1) \in E$  y  $(2, 3), (3, 2) \in E$ , luego  $(v_1, v_2) \in E \iff (v_2, v_1) \in E$ .

Por tanto,  $G$  es un grafo no dirigido por ser  $E$  simétrico, y además, como  $E$  también es antirreflexivo y no contiene arcos duplicados, se tiene que  $G$  es una red simple.

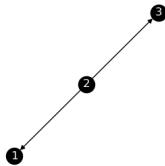
Los grados de los nodos de  $G$  son  $d(1) = 1$ ,  $d(2) = 2$  y  $d(3) = 1$ , por lo que el nodo 2 es el hub del grafo.

La longitud del camino del nodo 1 al 3 es de 2 saltos. Por otro lado, la longitud del camino de 1 a 2 y del de 2 a 3 es de 1 salto en ambos casos. Se observa además que estos caminos son todos directos. Por lo tanto, la longitud promedio de camino de  $G$  es de  $\frac{1+1+2}{3} = \frac{4}{3} = 1.33$  saltos.

Por otro lado, la matriz de conexión de  $G$  es simétrica por ser  $G$  no dirigido. Como  $E = \{(1, 2), (2, 1), (2, 3), (3, 2)\}$ , se tendrá que  $c_{12} = c_{21} = c_{23} = c_{32} = 1$  y  $c_{ij} = 0$  para  $(i, j) \notin E$ , por lo que:

$$C(G) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (5)$$

**Ejemplo 2.22** Sea  $G = (V, E)$  un grafo, donde  $V = \{1, 2, 3\}$  y  $E = \{(2, 1), (2, 3)\}$ , representado en la Figura 5.



**Figura 5:** Grafo dirigido con 2 arcos dirigidos.

- $E$  no es reflexivo, pues por ejemplo  $1 \in V$  pero  $(1, 1) \notin E$ .
- $E$  es antirreflexivo, pues  $(1, 1), (2, 2), (3, 3) \notin E$ .
- $E$  no es simétrico, pues por ejemplo  $(2, 1) \in E$  pero  $(1, 2) \notin E$ .

Entonces,  $G$  es un grafo dirigido por ser  $E$  asimétrico.

Los grados de salida de los nodos de  $G$  son  $\text{out\_d}(1) = 0$ ,  $\text{out\_d}(2) = 2$  y  $\text{out\_d}(3) = 0$ , y los grados de entrada son  $\text{in\_d}(1) = 1$ ,  $\text{in\_d}(2) = 0$ ,  $\text{in\_d}(3) = 1$ .

Se observa que no hay ningún camino del nodo 1 al 3, ya que los arcos son dirigidos. Sin embargo, sí hay un camino del nodo 2 al 1 y otro del nodo 2 al 3, ambos de longitud de 1 salto.

Por otro lado, la matriz de conexión de  $G$  no es simétrica por ser  $G$  dirigido. Como  $E = \{(2, 1), (2, 3)\}$ , se tendrá que  $c_{21} = c_{23} = 1$  y  $c_{ij} = 0$  para  $(i, j) \notin E$ , por lo que (Lewis, 2009):

$$C(G) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (6)$$

**Ejemplo 2.23** Sean  $G = (V, E)$  y  $G' = (V, E')$  grafos tales que  $V = \{1, 2, 3, 4\}$  y

$$E = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2), (3, 4), (4, 3)\}$$

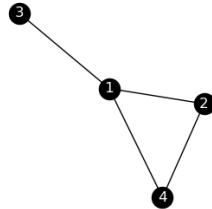
$$E' = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 3), (2, 3), (3, 1), (3, 2), (3, 2), (3, 4), (4, 3)\}$$

Ambos son no dirigidos. Además,  $G$  es un grafo simple, mientras que  $G'$  es un multigrafo. En la Figura 6 se muestran ambos grafos,  $G$  y  $G'$ .



**Figura 6:** Grafo simple  $G$  y multigrafo  $G'$ .

Ni el etiquetado ni la representación de una red son únicos. Por ejemplo, se puede representar  $G = (V, E'')$ , donde  $E'' = \{(1, 2), (1, 3), (1, 4), (2, 1), (2, 4), (3, 1), (4, 1), (4, 2)\}$ . En la Figura 7 se muestra esta representación equivalente de  $G$ .



**Figura 7:** Una representación equivalente de  $G$ . Aquí,  $G = (V, E'')$ .

Observamos lo siguiente:

- (I)  $G$  es un subgrafo de  $G'$ .
- (II)  $(V, E)$  y  $(V, E'')$  son isomorfos mediante  $f : V \rightarrow V$ , donde  $f(1) = 3$ ,  $f(2) = 2$ ,  $f(3) = 4$ ,  $f(4) = 1$ .

(III)  $\overline{G} = (V, \overline{E})$ , donde  $\overline{E} = \{(1, 4), (4, 1), (2, 4), (4, 2)\}$ .

En la figura 8 se representa  $\overline{G}$ , el complementario de  $G$ .



**Figura 8:** El complementario,  $\overline{G}$ , del grafo  $G$ .

(IV) Sea  $F = \{(1, 1), (2, 3), (2, 3), (3, 2), (3, 2)\}$ , entonces  $(V, E) = (V, E') - (V, F)$ .

(V) Los grados de los nodos en  $G$  son  $d(1) = 2$ ,  $d(2) = 2$ ,  $d(3) = 3$  y  $d(4) = 1$ , y de los nodos en  $G'$  son  $d(1) = 3$ ,  $d(2) = 4$ ,  $d(3) = 5$  y  $d(4) = 1$ .

(VI) El nodo 3 de  $\overline{G}$  es un nodo aislado, y los nodos 1 y 2 son nodos finales.

(VII) En  $G$ , se tiene el triángulo  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ . Además, el camino  $4 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 3$  es un recorrido, mientras que  $4 \rightarrow 3 \rightarrow 2 \rightarrow 1$  es un camino simple.

Ahora se calculan algunas matrices para  $G$  y  $G'$ . En primer lugar, para calcular las matrices de adyacencia, se tiene en cuenta que:

- $E = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2), (3, 4), (4, 3)\}$ , luego  $a_{12} = a_{13} = a_{21} = a_{23} = a_{31} = a_{32} = a_{34} = a_{43} = 1$  y  $a_{ij} = 0$  para  $(i, j) \notin E$ , donde  $A(G) = (a_{ij})$ .
- $E' = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 3), (2, 3), (2, 3), (3, 1), (3, 2), (3, 2), (3, 2), (3, 4), (4, 3)\}$ , luego  $a'_{11} = a'_{12} = a'_{13} = a'_{21} = a'_{23} = a'_{31} = a'_{32} = a'_{34} = a'_{43} = 1$  y  $a'_{ij} = 0$  para  $(i, j) \notin E'$ , donde  $A(G') = (a'_{ij})$ .

Así, las matrices de adyacencia son:

$$A(G) = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad A(G') = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (7)$$

Además, la matriz de conexión de  $G$  coincide con su matriz de adyacencia por ser  $G$  un grafo simple, luego no contiene arcos duplicados. Sin embargo, la matriz de conexión de  $G'$  es diferente de su matriz de adyacencia ya que hay arcos que se repiten (el arco  $(2, 3)$

y el  $(3, 2)$  se repiten 3 veces cada uno, luego  $c_{23} = 3$  y  $c_{32} = 3$ ). Así, las matrices de conexión son:

$$C(G) = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad C(G') = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (8)$$

Se pueden obtener algunas potencias de  $A = A(G)$ :

$$A^2 = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \quad A^3 = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 3 & 2 & 4 & 1 \\ 4 & 4 & 2 & 3 \\ 1 & 1 & 3 & 0 \end{bmatrix} \quad A^4 = \begin{bmatrix} 7 & 6 & 6 & 4 \\ 6 & 7 & 6 & 4 \\ 6 & 6 & 11 & 2 \\ 4 & 4 & 2 & 3 \end{bmatrix} \quad (9)$$

Se observa que los elementos de la diagonal de  $A^2$  son los grados de los nodos, y coinciden con los que se han calculado anteriormente.

Ahora, para obtener las matrices de incidencia se observa que:

- Como  $G$  es simple, luego si  $(u, v) \in E$  entonces  $(v, u) \in E$ , se toma solo uno de estos pares para construir la matriz de incidencia orientada. Así, los arcos de  $G$  son  $e_1 = (1, 2), e_2 = (1, 3), e_3 = (2, 3), e_4 = (3, 4) \in E$ . Si  $e_i = (u_i, v_i)$ , entonces  $B(G) = (b_{ij})$  donde para  $1 \leq i \leq 4$  y  $1 \leq j \leq 4$ :

$$b_{ij} = \begin{cases} 1, & u_i = j \\ -1, & v_i = j \\ 0, & \text{en otro caso} \end{cases} \quad (10)$$

Por ejemplo,  $b_{11} = 1$  ya que  $u_1 = 1$ , mientras que  $b_{12} = -1$  porque  $v_1 = 2$ , y  $b_{13} = 0$  ya que  $u_1 = 1 \neq 3$  y  $v_1 = 2 \neq 3$ .

- De forma análoga para  $G'$ , teniendo en cuenta que en este caso  $G'$  no es simple, luego se deben considerar todos los arcos. Así, se obtiene  $e'_1 = (1, 1), e'_2 = (1, 2), e'_3 = (1, 3), e'_4 = (2, 1), e'_5 = (2, 3), e'_6 = (2, 3), e'_7 = (2, 3), e'_8 = (3, 1), e'_9 = (3, 2), e'_{10} = (3, 2), e'_{11} = (3, 2), e'_{12} = (3, 4), e'_{13} = (4, 3) \in E'$ . Se observa que como  $e'_1 = (u'_1, v'_1) = (1, 1)$  es un bucle, se tiene que  $b_{1u'_1} = b_{11} = 1$  y  $b_{12} = b_{13} = b_{14} = 0$ .

Así, las matrices de incidencia de  $G$  y  $G'$  son (Estrada and Knight, 2015):

$$B(G) = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad B(G') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad (11)$$

## 2.3. Métricas: Medidas de centralidad

La centralidad de un nodo es un concepto usado en la determinación de los nodos más importantes en una red, en base a diversas características. Así, se llega a diferentes medidas de centralidad, que se estudian a continuación.

### 2.3.1. Centralidad de grado

La centralidad de grado mide la capacidad de un nodo de comunicarse directamente con otros. Para ello, se denota el grado de un nodo  $i$  como  $k_i$ , es decir,  $k_i = d(i)$ . De forma análoga, se denota por  $k_i^{in} = in\_d(i)$  al grado de entrada y por  $k_i^{out} = out\_d(i)$  al grado de salida de un nodo  $i$ .

**Proposición 2.24** *Sea  $G$  una red y sea  $i$  un nodo de  $G$ . Sea  $A$  la matriz de adyacencia de  $G$  y sea  $e$  el vector de unos. Se tiene que:*

- *Si  $G$  es una red simple, el grado de  $i$  se puede obtener como:*

$$k_i = \sum_{j=1}^n a_{ij} = (e^T A)_i = (Ae)_i \quad (12)$$

- *Si  $G$  es una red dirigida, el grado de entrada de  $i$  se obtiene como:*

$$k_i^{in} = \sum_{j=1}^n a_{ji} = (e^T A)_i \quad (13)$$

- Si  $G$  es una red dirigida, el grado de salida de  $i$  se obtiene como:

$$k_i^{out} = \sum_{j=1}^n a_{ij} = (Ae)_i \quad (14)$$

**Definición 2.25** Sea  $G$  una red simple, y sean  $i$  y  $j$  nodos de  $G$ . Con la centralidad de grado, se dice que  $i$  es más **central** que  $j$  si  $k_i > k_j$

**Observación 2.26** Si  $G$  es una red dirigida entonces, de forma análoga al caso de la red simple, se pueden definir dos medidas de centralidad de grado usando  $k_i^{in}$  y  $k_i^{out}$ .

**Proposición 2.27** Con la notación anterior, se cumple que:

$$(I) \quad k_i = (A^2)_{ii}$$

$$(II) \quad \sum_{i=1}^n k_i = 2m, \text{ donde } m \text{ es el número de arcos.}$$

$$(III) \quad \sum_{i=1}^n k_i^{in} = \sum_{i=1}^n k_i^{out} = m, \text{ donde } m \text{ es el número de arcos.}$$

**Definición 2.28** En una red dirigida  $G$ , se dice que un nodo es una **fuente** si su grado de entrada es 0 y su grado de salida es distinto de 0, un **sumidero** si su grado de salida es 0 y su grado de entrada es distinto de 0, y un **nodo aislado** si tanto su grado de salida como el de entrada son ambos 0.

### 2.3.2. Centralidad de cercanía

La centralidad de cercanía mide qué tan cerca está un nodo del resto de nodos, y se mide en términos de la distancia del camino más corto.

**Definición 2.29** La **cercanía** del nodo  $i$  en una red no dirigida  $G$  se define como:

$$CC(i) = \frac{n - 1}{s(i)} \quad (15)$$

donde la suma de distancias  $s(i)$  se calcula a partir de las distancias de camino más corto  $d(i, j)$  como:

$$s(i) = \sum_{j \in V(G)} d(i, j) \quad (16)$$

**Observación 2.30** Si  $G$  es una red dirigida, entonces de forma análoga se pueden definir la centralidad de cercanía de entrada, que indica qué tan cerca está un nodo de los nodos de los que recibe información, y la centralidad de cercanía de salida, que corresponde a qué tan cerca está un nodo de aquellos a los que envía información. En este caso, el camino más corto es una pseudodistancia (puede ser cero entre dos nodos distintos) debido a una posible falta de simetría.

**Proposición 2.31** Sea  $CC(i)$  la centralidad de cercanía del  $i$ -ésimo nodo de la red camino  $P_{n-1}$  denotada por  $1 - 2 - 3 - 4 - \cdots - (n-1) - n$ . Entonces:

$$(a) CC(i) = \frac{2(n-1)}{(i-1)i + (n-i)(n-i+1)}$$

(b) Sea  $i$  el nodo (o nodos) en el centro de  $P_{n-1}$ . Si  $n$  es impar, entonces  $i = \frac{n+1}{2}$  y  $CC\left(\frac{n+1}{2}\right) = \frac{4}{(n+1)}$ . Si  $n$  es par, entonces  $i = \frac{n}{2}$  e  $i = \frac{n}{2} + 1$  son los nodos centrales y  $CC\left(\frac{n}{2}\right) = \frac{4(n-1)}{n^2}$ ,  $CC\left(\frac{n}{2} + 1\right) = \frac{4(n-1)}{n^2}$ .

(c) La centralidad de cercanía del nodo (o nodos) central es la más grande en  $P_{n-1}$ .

*Demostración.*

(a) Se considera la suma de todas las distancias de un nodo al resto de nodos en el camino:

$$\sum_{j \neq 1} d(1, j) = 1 + 2 + \cdots + n - 1$$

$$\sum_{j \neq 2} d(2, j) = 1 + 1 + 2 + \cdots + n - 2$$

$$\sum_{j \neq 3} d(3, j) = 2 + 1 + 1 + 2 + \cdots + n - 3$$

⋮

$$\sum_{j \neq i} d(i, j) = (i-1) + (i-2) + \cdots + 2 + 1 + 1 + 2 + \cdots + n - i \quad (17)$$

Para cada nodo  $i$ , la suma (17) corresponde a la suma  $1 + 2 + \cdots + n - i = \frac{(n-i)(n-i+1)}{2}$  de las distancias de los nodos a la derecha de  $i$ , más la suma  $(i-1) + (i-2) + \cdots + 2 + 1 = \frac{(i-1)i}{2}$  de las distancias de los nodos a la izquierda de  $i$ . Así, se obtiene:

$$CC(i) = \frac{n-1}{s(i)} = \frac{n-1}{\frac{(i-1)i}{2} + \frac{(n-i)(n-i+1)}{2}} = \frac{2(n-1)}{(i-1)i + (n-i)(n-i+1)} \quad (18)$$

(b) Si  $n$  es impar el nodo central es  $i = \frac{n+1}{2}$ . Sustituyendo en (18):

$$\begin{aligned} CC\left(\frac{n+1}{2}\right) &= \frac{2(n-1)}{\left(\frac{n+1}{2} - 1\right) \frac{n+1}{2} + \left(n - \frac{n+1}{2}\right) \left(n - \frac{n+1}{2} + 1\right)} = \\ &= \frac{2(n-1)}{\left(\frac{n-1}{2}\right) \frac{n+1}{2} + \left(\frac{n-1}{2}\right) \left(\frac{n+1}{2}\right)} = \frac{2(n-1)}{2 \left(\frac{n-1}{2}\right) \left(\frac{n+1}{2}\right)} = \frac{4}{(n+1)} \end{aligned} \quad (19)$$

Si  $n$  es par los nodos centrales son  $i = \frac{n}{2}$  e  $i = \frac{n}{2} + 1$ . Sustituyendo en (18):

$$\begin{aligned} CC\left(\frac{n}{2}\right) &= \frac{2(n-1)}{\left(\frac{n}{2}-1\right)\frac{n}{2} + \left(n-\frac{n}{2}\right)\left(n-\frac{n}{2}+1\right)} = \\ &= \frac{2(n-1)}{\left(\frac{n}{2}-1\right)\frac{n}{2} + \frac{n}{2}\left(\frac{n}{2}+1\right)} = \frac{2(n-1)}{\frac{n^2}{4} - \frac{n}{2} + \frac{n^2}{4} + \frac{n}{2}} = \frac{2(n-1)}{\frac{n^2}{2}} = \frac{4(n-1)}{n^2} \end{aligned} \quad (20)$$

$$\begin{aligned} CC\left(\frac{n}{2}+1\right) &= \frac{2(n-1)}{\left(\frac{n}{2}+1-1\right)\left(\frac{n}{2}+1\right) + \left(n-\frac{n}{2}-1\right)\left(n-\frac{n}{2}-1+1\right)} = \\ &= \frac{2(n-1)}{\frac{n}{2}\left(\frac{n}{2}+1\right) + \left(\frac{n}{2}-1\right)\frac{n}{2}} = \frac{2(n-1)}{\frac{n^2}{4} + \frac{n}{2} + \frac{n^2}{4} - \frac{n}{2}} = \frac{2(n-1)}{\frac{n^2}{2}} = \frac{4(n-1)}{n^2} \end{aligned} \quad (21)$$

(c) Por (18) se tiene:

$$\begin{aligned} CC(i+1) - CC(i) &= \frac{2(n-1)}{i(i+1) + (n-i-1)(n-i)} - \frac{2(n-1)}{(i-1)i + (n-i)(n-i+1)} = \\ &= \frac{2n-2}{i^2 + i + n^2 - ni - ni + i^2 - n + i} - \frac{2n-2}{i^2 - i + n^2 - ni + n - ni + i^2 - i} = \\ &= \frac{2n-2}{-2 + 2i + n^2 - 2ni - n} - \frac{2n-2}{-2 - 2i + n^2 - 2ni + n} = \\ &= \frac{(-2 - 2i + n^2 - 2ni + n)(2n-2) - (-2 + 2i + n^2 - 2ni - n)(2n-2)}{(-2 + 2i + n^2 - 2ni - n)(-2 - 2i + n^2 - 2ni + n)} = \\ &= \frac{(2n-2)(-2 - 2i + n^2 - 2ni + n - (-2 + 2i + n^2 - 2ni - n))}{(-2 + n^2 - 2ni)^2 - (2i - n)^2} = \\ &= \frac{(2n-2)(-4i + 2n)}{(-2 + n^2 - 2ni)^2 - (2i - n)^2} = \frac{4(n-1)(n-2i)}{(-2 + n^2 - 2ni)^2 - (2i - n)^2} \end{aligned} \quad (22)$$

Como  $4(n-1)(n-2i) > 0$  si  $i < n/2$  y  $4(n-1)(n-2i) < 0$  si  $i > n/2$ , entonces por (22) se tiene que  $CC(i+1) > CC(i)$  si  $i < n/2$  y  $CC(i+1) < CC(i)$  si  $i > n/2$ , por lo que  $CC(i)$  alcanza su valor máximo en el centro del camino.

■

### 2.3.3. Centralidad de intermediación

La centralidad de intermediación caracteriza la importancia de un nodo en la comunicación entre otros pares de nodos, es decir, representa la proporción de información que pasa a través de un nodo en comunicaciones entre otros pares de nodos. La intermediación asume que la información viaja a través de los caminos más cortos.

**Definición 2.32** La *intermediación* de un nodo  $i$  en una red no dirigida  $G$  se define como:

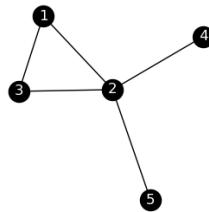
$$BC(i) = \sum_j \sum_k \frac{\rho(j, i, k)}{\rho(j, k)}, \quad i \neq j \neq k, \quad (23)$$

donde  $\rho(j, k)$  es el número de caminos más cortos conectando el nodo  $j$  al  $k$ , y  $\rho(j, i, k)$  es el número de estos caminos más cortos que pasan a través del nodo  $i$ .

**Observación 2.33** Si la red es dirigida, la intermediación de  $i$  se define de forma análoga, pero considerando que  $\rho(j, i, k)$  es el número de caminos simples dirigidos de  $j$  a  $k$  que pasan a través de  $i$ , y  $\rho(j, k)$  es el número total de caminos simples dirigidos de  $j$  a  $k$ .

#### 2.3.4. Ejemplos de medidas de centralidad

**Ejemplo 2.34** Se considera la red representada en la Figura 9.



**Figura 9:** Una red simple.

Su matriz de adyacencia es:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (24)$$

Por lo tanto, el vector con los grados de los nodos es:

$$k = Ae = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 2 \\ 1 \\ 1 \end{bmatrix} \quad (25)$$

Así, los grados de los nodos son:  $k_1 = k_3 = 2$ ,  $k_2 = 4$ ,  $k_4 = k_5 = 1$ . Por ello, el nodo más central es el 2, por ser su grado el mayor.

**Ejemplo 2.35** Se considera la red representada en la Figura 10.

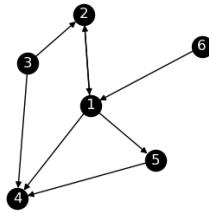


Figura 10: Una red dirigida.

Su matriz de adyacencia es:

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (26)$$

Los vectores con los grados de entrada y de salida se obtienen:

$$k^{in} = (e^T A)^T = A^T e = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 3 \\ 1 \\ 0 \end{bmatrix} \quad (27)$$

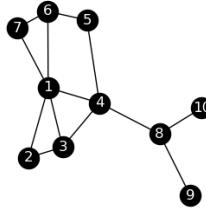
$$k^{out} = Ae = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad (28)$$

Los nodos 3 y 6 son fuentes ya que  $k_3^{in} = 0$  y  $k_6^{in} = 0$ , pero  $k_3^{out} = 2 \neq 0$  y  $k_6^{out} = 1 \neq 0$ .

El nodo 4 es un sumidero porque  $k_4^{out} = 0$  pero  $k_4^{in} = 3 \neq 0$ .

El nodo más central en el envío de información a sus vecinos más cercanos es el nodo 1, por ser  $k_1^{out} = 3$  el mayor grado de salida, y en la recepción de información es el nodo 4, ya que  $k_4^{in} = 3$  es el mayor grado de entrada.

**Ejemplo 2.36** Se considera la red representada en la Figura 11.



**Figura 11:** Una red donde la centralidad de cercanía no coincide con la centralidad de grado.

Se calcula la suma de distancias  $s(i)$  para cada nodo  $i \in V = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ , y a partir de ella se calcula la centralidad de cercanía de cada nodo:

$$s(1) = \sum_{j \in V} d(1, j) = 0 + 1 + 1 + 1 + 2 + 1 + 1 + 2 + 3 + 3 = 15 \implies CC(1) = \frac{9}{15} = 0.6$$

$$s(2) = \sum_{j \in V} d(2, j) = 1 + 0 + 1 + 2 + 3 + 2 + 2 + 3 + 4 + 4 = 22 \implies CC(2) = \frac{9}{22} = 0.409$$

$$s(3) = \sum_{j \in V} d(3, j) = 1 + 1 + 0 + 1 + 2 + 2 + 2 + 2 + 3 + 3 = 17 \implies CC(3) = \frac{9}{17} = 0.529$$

$$s(4) = \sum_{j \in V} d(4, j) = 1 + 2 + 1 + 0 + 1 + 2 + 2 + 1 + 2 + 2 = 14 \implies CC(4) = \frac{9}{14} = 0.643$$

$$s(5) = \sum_{j \in V} d(5, j) = 2 + 3 + 2 + 1 + 0 + 1 + 2 + 2 + 3 + 3 = 19 \implies CC(5) = \frac{9}{19} = 0.474$$

$$s(6) = \sum_{j \in V} d(6, j) = 1 + 2 + 2 + 2 + 1 + 0 + 1 + 3 + 4 + 4 = 20 \implies CC(6) = \frac{9}{20} = 0.45$$

$$s(7) = \sum_{j \in V} d(7, j) = 1 + 2 + 2 + 2 + 2 + 1 + 0 + 3 + 4 + 4 = 21 \implies CC(7) = \frac{9}{21} = 0.429$$

$$s(8) = \sum_{j \in V} d(8, j) = 2 + 3 + 2 + 1 + 2 + 3 + 3 + 0 + 1 + 1 = 18 \implies CC(8) = \frac{9}{18} = 0.5$$

$$s(9) = \sum_{j \in V} d(9, j) = 3 + 4 + 3 + 2 + 3 + 4 + 4 + 1 + 0 + 2 = 26 \implies CC(9) = \frac{9}{26} = 0.346$$

$$s(10) = \sum_{j \in V} d(10, j) = 3 + 4 + 3 + 2 + 3 + 4 + 4 + 1 + 2 + 0 = 26 \implies CC(10) = \frac{9}{26} = 0.346$$

Así, el nodo más central es el nodo 4, pues  $CC(4) = 0.643$  es la cercanía mayor. Sin embargo, si se usa la centralidad de grado, entonces se obtiene que el nodo más central es el nodo 1, como se ve a continuación:

$$k = Ae = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 3 \\ 4 \\ 2 \\ 3 \\ 2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \quad (29)$$

Ahora veamos cómo obtener la centralidad de intermediación para el nodo 1. Para ello, se obtiene el número de caminos más cortos desde cada par de nodos que pasan a través del nodo 1,  $\rho(j, 1, k)$ , así como el número total de caminos más cortos desde cada par de nodos,  $\rho(j, k)$ :

$$\rho(2, 1, 4) = 1 \qquad \rho(2, 4) = 2$$

$$\rho(2, 1, 5) = 2 \qquad \rho(2, 5) = 3$$

$$\rho(2, 1, 6) = 1 \qquad \rho(2, 6) = 1$$

$$\begin{array}{ll}
\rho(2, 1, 7) = 1 & \rho(2, 7) = 1 \\
\rho(2, 1, 8) = 1 & \rho(2, 8) = 2 \\
\rho(2, 1, 9) = 1 & \rho(2, 9) = 2 \\
\rho(2, 1, 10) = 1 & \rho(2, 10) = 2 \\
\rho(3, 1, 6) = 1 & \rho(3, 6) = 1 \\
\rho(3, 1, 7) = 1 & \rho(3, 7) = 1 \\
\rho(4, 1, 6) = 1 & \rho(4, 6) = 2 \\
\rho(4, 1, 7) = 1 & \rho(4, 7) = 1 \\
\rho(6, 1, 8) = 1 & \rho(6, 8) = 2 \\
\rho(6, 1, 9) = 1 & \rho(6, 9) = 2 \\
\rho(6, 1, 10) = 1 & \rho(6, 10) = 2 \\
\rho(7, 1, 8) = 1 & \rho(7, 8) = 1 \\
\rho(7, 1, 9) = 1 & \rho(7, 9) = 1 \\
\rho(7, 1, 10) = 1 & \rho(7, 10) = 1
\end{array}$$

Observar que para el resto de pares de nodos  $j, k$ , se tiene que  $\rho(j, 1, k) = 0$ . Así, la centralidad de intermediación del nodo 1 es:

$$BC(1) = \sum_j \sum_k \frac{\rho(j, 1, k)}{\rho(j, k)} = \frac{1}{2} + \frac{2}{3} + 1 + 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + 1 + 1 + \frac{1}{2} + 1 + \frac{1}{2} + \frac{1}{2} + 1 + 1 + 1 = 12.667$$

De forma análoga, se obtiene la centralidad de intermediación para cada nodo:

$$BC = [12.667, 0, 2.333, 20.167, 2, 1.833, 0, 15, 0, 0]$$

Así, el nodo 4 es el más central, es decir, es el más importante en la comunicación entre otros pares de nodos (Estrada and Knight, 2015).

### 3. Redes aleatorias

Las redes aleatorias fueron algunas de las primeras redes estudiadas, remontándose a la década de 1950. Sin embargo, no son representativas de la naturaleza. De hecho, la mayoría de sistemas físicos y biológicos no están dispuestos al azar, sino que contienen estructura. Sin embargo, el análisis de las redes aleatorias sirve como una base para contrastar con redes que tienen una estructura más específica o planificada. Esto ayuda a identificar patrones, propiedades o características de las redes más estructuradas y proporciona información valiosa sobre la función y organización de distintos tipos de redes en diversos campos (Lewis, 2009).

Para la generación de redes aleatorias, se puede pensar en un modelo en el cual los pares de nodos estén conectados con cierta probabilidad. Es decir, si se empieza con  $n$  nodos, para cada uno de los  $\frac{n(n-1)}{2}$  enlaces posibles, se conectarán un par de nodos  $u, v$  con probabilidad  $p_{u,v}$ . Así, si se fijan una serie de parámetros de la red y se crean los arcos mediante un proceso aleatorio, se podrán crear modelos que nos permitan entender la influencia de estos parámetros en la estructura de las redes (Estrada and Knight, 2015).

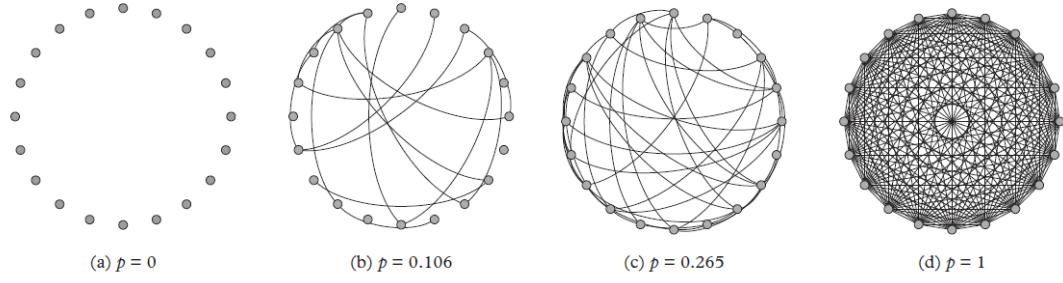
#### 3.1. El modelo de Erdös-Rényi de redes aleatorias

Según este modelo, llevado a cabo por Pál Erdös y Alfréd Rényi en 1959, para construir una red aleatoria se siguen los siguientes pasos:

- 1) Se empieza con  $n$  nodos aislados, y se fija un  $p > 0$ .
- 2) Se selecciona un par de nodos y se genera uniformemente un número aleatorio  $r \in [0, 1]$ . Si  $p > r$  se conecta el par de nodos seleccionado con un arco, y en otro caso se dejan desconectados.
- 3) Se repite el paso 2) para cada uno de los  $\frac{n(n-1)}{2}$  pares de nodos (Barabási, 2016).

**Notación 3.1** La red aleatoria de Erdös-Rényi (ER) se denota por  $G_{ER}(n, m)$  o  $G_{ER}(n, p)$ , donde  $n$  es el número de nodos,  $m$  es el número de arcos y  $p > 0$  es la probabilidad de enlace.

En la Figura 12 se muestran varios ejemplos de redes aleatorias generadas por el modelo de Erdös-Rényi usando diferentes probabilidades de enlace.



**Figura 12:** Ejemplos de redes aleatorias de Erdős-Rényi con 20 nodos y diferentes probabilidades de enlace. Para  $p = 0$  la red permanece totalmente desconectada, y para  $p = 1$  se obtiene una red completa (Estrada and Knight, 2015).

**Proposición 3.2** Algunas propiedades de las redes aleatorias generadas por el modelo de Erdős-Rényi son las siguientes:

- a) El número esperado de arcos es  $\overline{m} = \frac{n(n-1)p}{2}$ .
- b) El grado esperado de cada nodo es  $\overline{k} = (n - 1)p$ .
- c) La longitud promedio de camino es  $\overline{l}(G) = \frac{\ln n - \gamma}{\ln(pn)} + \frac{1}{2}$ , donde  $\gamma \approx 0.577$  es la constante de Euler-Mascheroni (Estrada and Knight, 2015).

*Demostración.* Se demostrarán los puntos a) y b). La probabilidad de que una red aleatoria tenga exactamente  $m$  arcos es el producto de 3 términos:

- La probabilidad de que  $m$  de los intentos de conectar los  $\frac{n(n-1)}{2}$  pares de nodos den como resultado un arco, que es  $p^m$ .
- La probabilidad de que los  $\frac{n(n-1)}{2} - m$  intentos restantes no resulten en un nodo, que es  $(1 - p)^{\frac{n(n-1)}{2} - m}$ .
- El número de formas diferentes en que podemos colocar  $m$  arcos entre  $\frac{n(n-1)}{2}$  pares de nodos, que es  $\binom{\frac{n(n-1)}{2}}{m}$ .

Entonces se escribe la probabilidad de que una red aleatoria tenga exactamente  $m$  arcos como:

$$p_m = \binom{\frac{n(n-1)}{2}}{m} p^m (1 - p)^{\frac{n(n-1)}{2} - m} \quad (30)$$

De esta forma, como (30) es una distribución binomial  $\text{Bin}\left(\frac{n(n-1)}{2}, m\right)$ , calculando su esperanza se obtiene el número esperado de arcos en una red aleatoria, que será:

$$\bar{m} = \sum_{m=0}^{\frac{n(n-1)}{2}} mp_m = \frac{n(n-1)p}{2} \quad (31)$$

Así,  $\bar{m}$  es el producto de la probabilidad  $p$  de que dos nodos estén conectados y del número de pares que intentamos conectar, es decir,  $\frac{n(n-1)}{2}$ .

Ahora, a partir de (31) se puede obtener el grado esperado de cada nodo como:

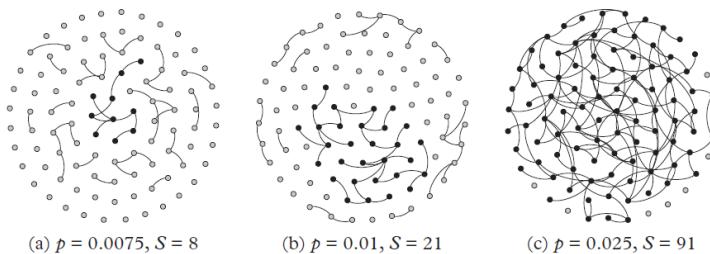
$$\bar{k} = \frac{2\bar{m}}{n} = (n-1)p \quad (32)$$

De esta forma,  $\bar{k}$  es el producto de la probabilidad  $p$  de que dos nodos estén conectados y de  $(n-1)$ , que es el número máximo de arcos que un nodo puede tener en una red con  $n$  nodos (Barabási, 2016). ■

**Observación 3.3** La estructura de  $G_{ER}(n, p)$  cambia en función de  $p = \frac{\bar{k}}{n-1}$ , dando lugar a 3 posibles etapas:

- (I) **Subcrítica** ( $\bar{k} < 1$ ): todas las componentes son simples y muy pequeñas. El tamaño de la componente más grande es  $S = O(\ln n)$ .
- (II) **Crítica** ( $\bar{k} = 1$ ): el tamaño de la componente más grande es  $S = O(n^{2/3})$ .
- (III) **Supercrítica** ( $\bar{k} > 1$ ):  $\lim_{n \rightarrow \infty} P((f - \varepsilon)n < S < (f + \varepsilon)n) = 1$ , con  $\varepsilon > 0$  y donde  $f = f(\bar{k})$  es la solución positiva de  $e^{-\bar{k}f} = 1 - f$ . El resto de componentes son muy pequeñas, y el tamaño de la segunda componente más grande es  $O(\ln n)$ .

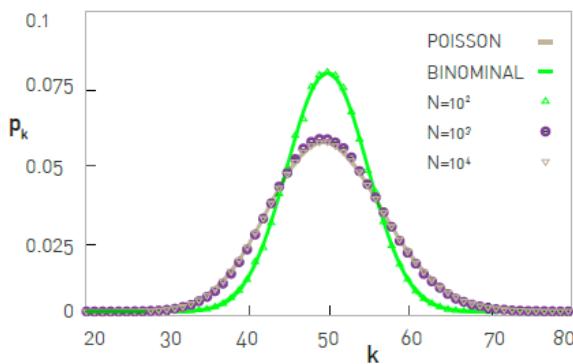
En la Figura 13 se ilustra este comportamiento para varias redes aleatorias ER con 100 nodos y diferentes probabilidades de enlace.



**Figura 13:** Redes aleatorias ER con 100 nodos y diferentes probabilidades de enlace. Los nodos en la componente conectada más grande están coloreados de un tono más oscuro (Estrada and Knight, 2015).

### 3.1.1. Distribución de grado

Sea  $p(k) = \frac{n(k)}{n}$ , donde  $n(k)$  es el número de nodos de grado  $k$  en una red de tamaño  $n$ . Entonces  $p(k)$  representa la probabilidad de que un nodo seleccionado uniformemente de manera aleatoria tenga grado  $k$ . Trazando  $p(k)$  frente a  $k$  se puede representar la distribución de grado de una red (Estrada and Knight, 2015), tal como se muestra en la Figura 14 para varias redes aleatorias ER.



**Figura 14:** Distribución de los grados en redes aleatorias ER con  $\bar{k} = 50$  y  $n = 10^2, 10^3, 10^4$ . Para redes pequeñas ( $n = 10^2$ ) la distribución de grado se desvía significativamente de la forma de Poisson, ya que la condición para la aproximación de Poisson,  $n \gg \bar{k}$ , no se satisface. Por ello, se usa la forma binomial. Sin embargo, para redes más grandes ( $n = 10^3, 10^4$ ), la distribución de grado se vuelve indistinguible de la distribución de Poisson. Así, para  $n$  grande, la distribución de grado es independiente del tamaño de la red  $n$  (Barabási, 2016).

En una red aleatoria de Erdös-Rényi, la probabilidad de que un nodo dado tenga exactamente  $k$  enlaces es el producto de tres términos:

- La probabilidad de que  $k$  de sus arcos estén presentes, que es  $p^k$ .
- La probabilidad de que los  $(n - 1 - k)$  posibles arcos restantes no estén presentes, es decir,  $(1 - p)^{n-1-k}$ .
- El número de formas diferentes en que se pueden seleccionar  $k$  arcos entre los  $n - 1$  arcos potenciales que un nodo puede tener, que es  $\binom{n-1}{k}$ .

Consecuentemente, la distribución de grado de una red aleatoria ER sigue una **distribución binomial**  $\text{Bin}(n - 1, k)$ , pues:

$$p(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \quad (33)$$

Según (33), dicha distribución de grado depende del número de nodos  $n$ , y de la probabilidad  $p$ . No obstante, en la mayoría de redes reales,  $\bar{k} \ll n$ . En este límite, la distribución de grado se aproxima correctamente por la **distribución de Poisson** que, según (34), depende únicamente del grado esperado de cada nodo,  $\bar{k}$ , (Barabási, 2016).

$$p(k) = \frac{e^{-\bar{k}} \bar{k}^k}{k!} \quad (34)$$

### 3.2. El modelo de Barabási-Albert de redes aleatorias

Se ha observado empíricamente que en muchas redes del mundo real, los grados siguen una distribución de cola gruesa (que presenta una gran asimetría o curtosis). Esto contrasta notablemente con la distribución de Poisson que se observa en las redes aleatorias de Erdös-Rényi. De esta forma, en 1999, Barabási y Albert propusieron un modelo para generar redes en el que la probabilidad de encontrar un nodo de grado  $k$  decae como una ley potencial respecto al grado.

Se comienza con una red pequeña con  $m_0$  nodos. En cada paso, se añade un nuevo nodo  $u$  a la red y se conecta a  $m \leq m_0$  de los nodos existentes  $v \in V$ . La probabilidad de unir el nodo  $u$  al nodo  $v$  es proporcional al grado de  $v$ , siendo más probable conectar nuevos nodos a aquellos nodos existentes con alto grado. Este proceso se conoce como conexión preferencial.

Se asume que la red aleatoria inicial está conectada y es del tipo ER con  $m_0$  nodos,  $G_{ER} = (V, E)$ . El algoritmo de Barabási-Albert (BA) es un proceso mediante el cual pequeñas heterogeneidades en la distribución de grado de la red ER crecen con el tiempo.

En las redes generadas por el modelo de Barabási-Albert, la probabilidad de que un nodo tenga grado  $k \geq d$  viene dada por (35).

$$p(k) = \frac{2d(d-1)}{k(k+1)(k+2)} \approx k^{-3} \quad (35)$$

Además, la longitud de camino promedio es:

$$\bar{l} = \frac{\ln n - \ln(d/2) - 1 - \gamma}{\ln \ln n + \ln(d/2)} + \frac{3}{2} \quad (36)$$

donde  $\gamma \approx 0.577$  es la constante de Euler-Mascheroni (Estrada and Knight, 2015).

### 3.3. Mundos pequeños

El **fenómeno de mundo pequeño**, también conocido como **seis grados de separación**, establece que dos personas, en cualquier parte del mundo, se pueden conectar a través de una cadena de seis conocidos o menos. En el contexto de la Teoría de Redes, el fenómeno de mundo pequeño establece que la distancia entre dos nodos escogidos aleatoriamente en una red es corta.

Se considera una red aleatoria con grado esperado  $\bar{k}$ . Un nodo en esta red tiene en promedio  $\bar{k}^d$  nodos a distancia  $d$ . Por ejemplo, si  $\bar{k} \approx 1000$ , que es el número estimado de conocidos que tiene una persona, se esperan  $10^6$  individuos a distancia 2, y mil millones a distancia 3.

El número esperado de nodos a una distancia menor o igual que  $d$  desde nuestro nodo inicial es:

$$N(d) = 1 + \bar{k} + \bar{k}^2 + \cdots + \bar{k}^d = \frac{\bar{k}^{d+1} - 1}{\bar{k} - 1} \quad (37)$$

$N(d)$  no debe exceder el número total de nodos,  $n$ . Por tanto, las distancias no pueden tomar valores arbitrarios. Se identificará la distancia máxima o **diámetro** de la red,  $d_{max}$ , al establecer:

$$N(d_{max}) \approx n \quad (38)$$

Asumiendo que  $\bar{k} \gg 1$ , se puede ignorar el término  $-1$  en el numerador y el denominador de (37), obteniendo:

$$\bar{k}^{d_{max}} \approx n \quad (39)$$

Por lo tanto, el diámetro de una red aleatoria cumple:

$$d_{max} \approx \frac{\ln n}{\ln \bar{k}} \quad (40)$$

que representa la formulación matemática del fenómeno de mundo pequeño. Veamos su interpretación:

- (40) predice el diámetro de la red,  $d_{max}$ , usando el número de nodos  $n$ . Sin embargo, para la mayoría de las redes, (40) ofrece una mejor aproximación a la distancia promedio entre dos nodos elegidos al azar,  $\bar{d}$ , que a  $d_{max}$ . Esto se debe a que  $d_{max}$  está dominado por algunos caminos extremos, mientras que  $\bar{d}$  se promedia sobre todos los pares de nodos. Por ello, la propiedad de mundo pequeño típicamente se define por:

$$\bar{d} \approx \frac{\ln n}{\ln \bar{k}} \quad (41)$$

que describe la dependencia de la distancia promedio de una red con  $n$  y con  $\bar{k}$ .

- En general  $\ln n \ll n$ , por lo que la dependencia de  $\bar{d}$  en  $\ln n$  implica que las distancias en una red aleatoria son órdenes de magnitud más pequeñas que el tamaño de la red. Así, en el fenómeno de mundo pequeño, por *pequeño* nos referimos a que la longitud de camino promedio o el diámetro dependen de manera logarítmica del tamaño de la red. *Pequeño* significa que  $\bar{d}$  es proporcional a  $\ln n$ , en lugar de a  $n$  o a alguna potencia de  $n$ .
- El término  $\ln \bar{k}$  en el denominador de (41) implica que cuanto más densa sea la red (más arcos tenga), más pequeña es la distancia entre nodos.

Veamos las implicaciones de (41) en las redes sociales. Usando  $n \approx 8 \cdot 10^9$  como la población mundial y  $\bar{k} \approx 10^3$  como el número promedio de conocidos de una persona, se obtiene:

$$\bar{d} \approx \frac{\ln(8 \cdot 10^9)}{\ln(10^3)} = 3.30 \quad (42)$$

Por lo tanto, todas las personas en la Tierra se deberían poder conectar a través de tres o cuatro conocidos, según el resultado de Sola Pool and Kochen (1978). La estimación (42) probablemente sea más cercana al valor real que los frecuentemente citados seis grados (Barabási, 2016).

### 3.4. Coeficientes de agrupamiento de redes

Muchas redes en el mundo real se caracterizan por la presencia de un gran número de triángulos, como consecuencia de la alta transitividad. Nuestra medida relativa está entre la proporción de triángulos existentes en una red y el número potencial de triángulos que puede soportar dados los grados de sus nodos. Para cuantificar esta propiedad se usan los coeficientes de agrupamiento de la red.

#### 3.4.1. El coeficiente de agrupamiento de Watts-Strogatz

Se supone que el agrupamiento de un nodo  $i$  es proporcional a:

$$C_i = \frac{\text{número de relaciones transitivas del nodo } i}{\text{número total de relaciones transitivas posibles del nodo } i} \quad (43)$$

y que  $t_i$  denota el número de triángulos unidos al nodo  $i$  de grado  $k_i$ , entonces:

$$C_i = \frac{t_i}{k_i(k_i - 1)/2} = \frac{2t_i}{k_i(k_i - 1)} \quad (44)$$

En consecuencia, el coeficiente de agrupamiento promedio de la red es:

$$\bar{C} = \frac{1}{n} \sum_i C_i \quad (45)$$

**Ejemplo 3.4** ■ Para las redes aleatorias generadas por el modelo de Erdős-Rényi, el coeficiente de agrupamiento promedio es  $\bar{C} = p$ .

■ Para las redes aleatorias generadas por el modelo de Barabási-Albert, el coeficiente de agrupamiento promedio,  $\bar{C}$ , se aproxima a  $\frac{d-1}{8} \frac{\log^2 n}{n}$  cuando  $n \rightarrow \infty$ .

### 3.4.2. El coeficiente de agrupamiento de Newman

También conocido como índice de transitividad. Sea  $t = |C_3|$  el número total de triángulos, y sea  $|P_2|$  el número de caminos simples de longitud 2 en la red. Entonces (Estrada and Knight, 2015):

$$C = \frac{3t}{|P_2|} = \frac{3|C_3|}{|P_2|} \quad (46)$$

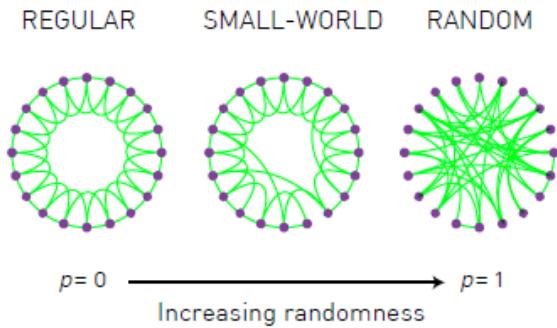
## 3.5. El modelo de Watts-Strogatz

Ducan Watts y Steven Strogatz propusieron una extensión del modelo de Erdős-Rényi motivados por dos observaciones:

- (a) **Propiedad de Mundo Pequeño:** en las redes reales la distancia promedio entre dos nodos depende logarítmicamente de  $n$ , en lugar de seguir un polinomio esperado para grafos regulares.
- (b) **Alto agrupamiento:** el coeficiente de agrupamiento promedio de las redes reales es mucho mayor del esperado para una red aleatoria ER con similar número de nodos y de arcos.

El modelo de Watts-Strogatz (también llamado el modelo de mundo pequeño) interpola entre un grafo regular (donde cada nodo tiene el mismo grado), que tiene un alto agrupamiento pero carece del fenómeno de mundo pequeño, y una red aleatoria ER, que tiene un bajo agrupamiento pero muestra la propiedad de mundo pequeño. Así, la longitud promedio de camino del modelo es baja pero el coeficiente de agrupamiento es alto, reproduciendo así la coexistencia de un alto agrupamiento y fenómenos de mundo pequeño.

En la Figura 15 se muestra este proceso de reconexión usado en el modelo de Watts-Strogatz.



**Figura 15:** Representación del proceso de reconexión de Watts-Strogatz. (Barabási, 2016).

El modelo de Watts-Strogatz se resume de la siguiente manera:

- 1) Se comienza con un anillo con  $n$  nodos, donde cada nodo está conectado a  $k$  vecinos inmediatos y siguientes.
- 2) Se recreenetan algunos arcos: cada uno de los arcos originales tiene una probabilidad  $p$  de que uno de sus extremos se mueva a un nuevo nodo elegido al azar. Para valores pequeños de  $p$ , la red mantiene un alto agrupamiento, aunque los enlaces aleatorios pueden disminuir drásticamente las distancias entre los nodos.
- 3) Para  $p = 1$  todos los arcos se han reconectado, de forma que la red se vuelve una red aleatoria, es decir, nos acercamos al modelo ER (Barabási, 2016).

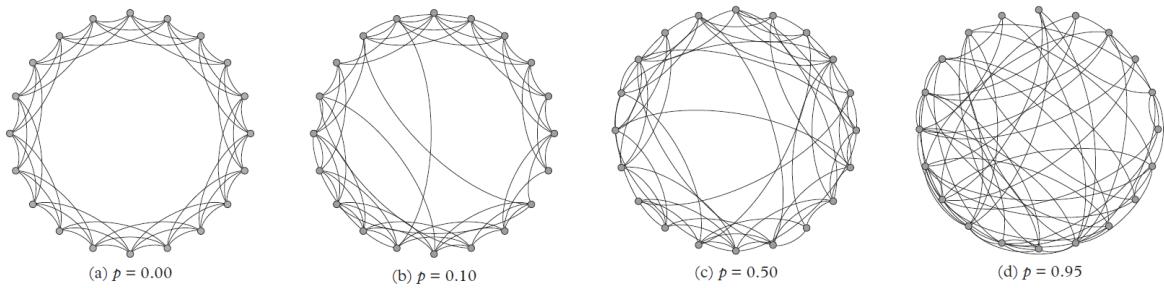
**Proposición 3.5** *Las redes generadas por el modelo de Watts-Strogatz cumplen las siguientes propiedades:*

- a) *El coeficiente de agrupamiento promedio es  $\overline{C} = \frac{3(k-2)}{4(k-1)}$ .*
- b) *La longitud de camino promedio decae muy rápido desde la de un grafo circulante,*

$$\bar{l} = \frac{(n-1)(n+k-1)}{2kn} \quad (47)$$

*para acercarse a la de una red aleatoria (Estrada and Knight, 2015).*

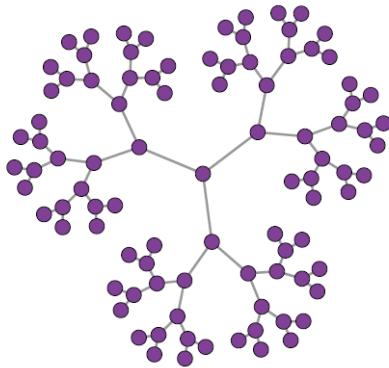
En la Figura 16 se muestran varios ejemplos de redes aleatorias de Watts-Strogatz para diferentes probabilidades de reconexión.



**Figura 16:** Ejemplos de redes aleatorias de Watts–Strogatz con  $n = 20$ ,  $k = 6$  y con diferentes probabilidades de reconexión  $p$  (Estrada and Knight, 2015).

### 3.6. Ejemplo: Árbol de Cayley

Un árbol de Cayley es un árbol simétrico, construido empezando desde un nodo central de grado  $k$ . Cada nodo a distancia  $d$  del nodo central tiene grado  $k$ , hasta que se alcanzan los nodos a distancia  $P$ , que tienen grado 1 y se llaman hojas (Barabási, 2016). En la Figura 17 se muestra un ejemplo de un árbol de Cayley.



**Figura 17:** Un árbol de Cayley con  $k = 3$  y  $P = 5$  (Barabási, 2016).

- a) Calculemos el número de nodos alcanzables en  $t$  pasos desde el nodo central.

Se observa que el nodo central, que digamos que es el nodo 0, está conectado con otros  $k$  nodos, digamos los nodos  $1, 2, \dots, k$ . Entonces, en  $t = 1$  pasos se pueden alcanzar  $k$  nodos desde el nodo central.

A su vez, cada uno de estos  $k$  nodos están conectados con el nodo 0 y con otros  $k - 1$  nodos. Es decir, que si se toma un nodo  $i_1 \in \{1, 2, \dots, k\}$ , existirá un arco desde este nodo  $i_1$  hasta el 0, y también existirán otros  $k - 1$  arcos desde  $i_1$  hasta  $k - 1$  nodos que

llamaremos  $i_11, i_12, \dots, i_1(k-1)$ . De esta forma, en  $t = 2$  pasos se podrán alcanzar  $k(k-1)$  nodos desde el nodo central.

Ahora, sea  $i_2 \in \{1, 2, \dots, k-1\}$ , entonces el arco  $i_1i_2$  estará conectado a su vez con el arco  $i_1$  y con  $k-1$  arcos que llamaremos  $i_1i_21, i_1i_22, \dots, i_1i_2(k-1)$ . Por tanto, en  $t = 3$  pasos se podrán alcanzar  $k(k-1)(k-1) = k(k-1)^2$  nodos desde el nodo central. Y así sucesivamente hasta alcanzar los nodos a distancia  $P$ , que llamaremos  $i_1i_2\dots i_{P-1}i_P$ , para  $i_1 \in \{1, 2, \dots, k\}, i_2, \dots, i_P \in \{1, 2, \dots, k-1\}$ , de forma que cada uno de ellos únicamente estará conectado al nodo  $i_1i_2\dots i_{P-1}$ .

Sea  $\hat{N}(t)$  el número de nodos a distancia  $t$  del nodo 0. Entonces:

$$\hat{N}(t) = \begin{cases} 1, & \text{si } t = 0 \\ k(k-1)^{t-1}, & \text{si } t \in \{1, 2, \dots, P\} \\ 0, & \text{si } t > P \end{cases} \quad (48)$$

A partir de (48) se puede calcular el número de nodos alcanzables en  $t$  pasos (o menos). Para ello, primero se considera que  $t < P$ , y se obtiene:

$$\begin{aligned} N(t) &= \sum_{i=0}^t \hat{N}(i) = 1 + \sum_{i=1}^t k(k-1)^{i-1} = 1 + k \left( \frac{1 - (k-1)^t}{1 - (k-1)} \right) = \\ &= \frac{(k-2) + k((k-1)^t - 1)}{k-2} = \frac{k(k-1)^t - 2}{k-2} \end{aligned} \quad (49)$$

Por otro lado, si  $t \geq P$ , entonces todos los nodos de la red son alcanzables en  $t$  (o menos) pasos, es decir:

$$\begin{aligned} N(t) &= \sum_{i=0}^P \hat{N}(i) = 1 + \sum_{i=1}^P k(k-1)^{i-1} = 1 + k \left( \frac{1 - (k-1)^P}{1 - (k-1)} \right) = \\ &= \frac{(k-2) + k((k-1)^P - 1)}{k-2} = \frac{k(k-1)^P - 2}{k-2} \end{aligned} \quad (50)$$

En conclusión, juntando (49) y (50), se llega a que el número de nodos alcanzables en  $t$  pasos (o menos) será:

$$N(t) = \begin{cases} \frac{k(k-1)^t - 2}{k-2}, & \text{si } t < P \\ \frac{k(k-1)^P - 2}{k-2}, & \text{si } t \geq P \end{cases} \quad (51)$$

b) Calculemos la distribución de grado de la red.

La distribución de grado  $p(m)$  de una red proporciona la probabilidad de que un nodo cualquiera de la red tenga exactamente  $m$  enlaces. Es decir, que se debe calcular la fracción de nodos en la red con grado  $m$ , para  $m \geq 0$ . En el caso del árbol de Cayley, todos los arcos tienen grado  $k$ , a excepción de los nodos que se encuentran a distancia  $P$  del nodo central, que tienen grado 1.

Por ello, se necesita saber cuántos nodos tienen grado  $k$  y cuántos tienen grado 1. Para ello, se puede usar el apartado a). Los nodos con grado  $k$  serán el nodo central más los nodos a distancia  $t$  del nodo central, con  $t \in \{1, 2, \dots, P-1\}$ . Es decir, que si se denota por  $n(m)$  el número de nodos con grado  $m$ , usando (48) se tendrá:

$$\begin{aligned} n(k) &= \sum_{t=0}^{P-1} \hat{N}(t) = 1 + \sum_{t=1}^{P-1} k(k-1)^{t-1} = 1 + k \left( \frac{1 - (k-1)^{P-1}}{1 - (k-1)} \right) = \\ &= \frac{(k-2) + k((k-1)^{P-1} - 1)}{k-2} = \frac{k(k-1)^{P-1} - 2}{k-2} \end{aligned} \quad (52)$$

Por otro lado, los nodos con grado 1 serán aquellos a distancia  $P$  del nodo central, es decir:

$$n(1) = \hat{N}(P) = k(k-1)^{P-1} \quad (53)$$

Ahora se puede calcular la distribución de grado como  $p(m) = \frac{n(m)}{n}$ , donde  $n$  es el número total de nodos de la red, es decir:

$$\begin{aligned} n &= \sum_{t=0}^P \hat{N}(t) = 1 + \sum_{t=1}^P k(k-1)^{t-1} = 1 + k \left( \frac{1 - (k-1)^P}{1 - (k-1)} \right) = \\ &= \frac{(k-2) + k((k-1)^P - 1)}{k-2} = \frac{k(k-1)^P - 2}{k-2} \end{aligned} \quad (54)$$

Entonces, usando (53) y (54), la probabilidad de que un nodo cualquiera tenga un enlace será:

$$p(1) = \frac{n(1)}{n} = \frac{k(k-1)^{P-1}}{\frac{k(k-1)^P - 2}{k-2}} = \frac{k(k-1)^{P-1}(k-2)}{k(k-1)^P - 2} \quad (55)$$

Además, por (52) y (54), la probabilidad de que un nodo tenga  $k$  enlaces (grado  $k$ ) será:

$$p(k) = \frac{n(k)}{n} = \frac{\frac{k(k-1)^{P-1}-2}{k-2}}{\frac{k(k-1)^P-2}{k-2}} = \frac{k(k-1)^{P-1}-2}{k(k-1)^P-2} \quad (56)$$

Por último, la probabilidad de que un nodo tenga grado  $m$ , con  $m \neq 1, k$  será 0. Así, juntando (55) y (56), la distribución de grado de un árbol de Cayley será:

$$p(m) = \begin{cases} \frac{k(k-1)^{P-1}(k-2)}{k(k-1)^{P-2}}, & \text{si } m = 1 \\ \frac{k(k-1)^{P-1}-2}{k(k-1)^{P-2}}, & \text{si } m = k \\ 0, & \text{en otro caso} \end{cases} \quad (57)$$

- c) Calculemos el diámetro  $d_{max}$ .

Para ello, se debe calcular la distancia máxima que existe entre pares de nodos (considerando todos los pares de nodos posibles), usando la longitud del camino más corto entre dichos pares de nodos.

Se observa que los pares de nodos a mayor distancia entre sí son aquellos que se encuentran a distancia  $P$  del nodo central (son hojas) y que además se encuentran en ramas distintas del árbol. Más específicamente, usando la notación del apartado a), son aquellos de la forma  $i_1 i_2 \dots i_{P-1} i_P$  e  $i'_1 i'_2 \dots i'_{P-1} i'_P$ , para  $i_1, i'_1 \in \{1, 2, \dots, k\}$ ,  $i_2, \dots, i_P, i'_2, \dots, i'_P \in \{1, 2, \dots, k-1\}$ , con  $i_1 \neq i'_1$ .

Por lo tanto, se debe calcular la distancia entre estos pares de nodos, y esta será el diámetro de la red. Pero como estos nodos están a distancia  $P$  del nodo central, y cualquier camino que los una debe pasar por el nodo central, al ser  $i_1 \neq i'_1$ , entonces la distancia entre estos nodos será 2 veces la distancia de cada uno de ellos al nodo central, es decir,  $2P$ . Así:

$$d_{max} = 2P \quad (58)$$

- d) Obtengamos una expresión para el diámetro  $d_{max}$  en función del número total de nodos  $n$  de la red.

En el apartado b) se ha obtenido por (54) que el número total de nodos es:

$$n = \frac{k(k-1)^P - 2}{k-2} \quad (59)$$

Se puede despejar  $P$  de (59), con lo que:

$$\begin{aligned} \frac{n(k-2)+2}{k} &= (k-1)^P \\ P &= \log_{k-1} \left( \frac{n(k-2)+2}{k} \right) = \frac{\ln \left( \frac{n(k-2)+2}{k} \right)}{\ln(k-1)} \end{aligned} \quad (60)$$

Por lo tanto, sustituyendo (60) en (58) el diámetro será:

$$d_{max} = 2P = 2 \cdot \frac{\ln\left(\frac{n(k-2)+2}{k}\right)}{\ln(k-1)} \quad (61)$$

- e) Ahora, a partir del apartado anterior se observa que el árbol de Cayley cumple la propiedad de mundo pequeño, pues según (61) el diámetro depende logarítmicamente de  $n$ .

### 3.7. Implementación de redes aleatorias en Python

En este apartado se estudia cómo implementar las redes aleatorias vistas a lo largo de las secciones 3.1, 3.2 y 3.5 usando la librería *NetworkX* de Python.

#### 3.7.1. Modelo de Érdos-Rényi en Python

En la librería *NetworkX* existe una función que permite construir redes aleatorias mediante el modelo de Érdos-Rényi:

`erdos_renyi_graph(n, p, seed = None, directed = False)`

que devuelve una red  $G_{ER}(n, p)$ , cuyos parámetros son:

- **n: int**  
Número de nodos.
- **p: float**  
Probabilidad de enlace.
- **seed: integer, random\_state, o None (por defecto)**  
Indicador del estado de generación de números aleatorios.
- **directed: bool, optional (por defecto=False)**  
Sí es True, la función devuelve una red dirigida.

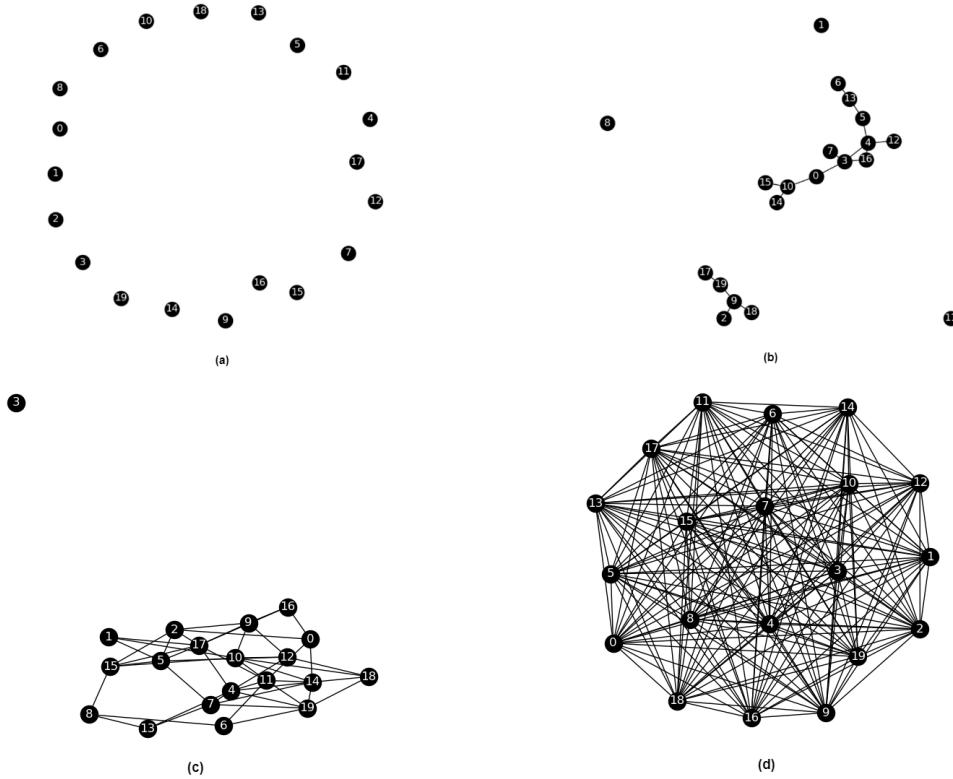
Usando esta función se pueden construir las redes vistas en la Figura 12. Por ejemplo para crear y representar la red (a), se puede usar el siguiente código en Python para construir una red aleatoria de Érdos-Rényi con  $n = 20$  nodos y probabilidad de enlace  $p = 0$ :

```
import networkx as nx
import matplotlib.pyplot as plt
```

```
GERa = nx.erdos_renyi_graph(n=20,p=0)

fig , ax = plt.subplots(figsize=(4, 4))
nx.draw(GERa, with_labels=True, ax=ax, node_color='black', font_color='white')
```

Análogamente se pueden obtener las redes (b) usando  $p = 0.106$ , (c) usando  $p = 0.265$  y (d) usando  $p = 1$ , obteniendo las representaciones de la Figura 18.



**Figura 18:** Redes aleatorias de Erdős-Rényi generadas en Python con  $n = 20$  nodos y probabilidades de enlace (a)  $p = 0$ , (b)  $p = 0.106$ , (c)  $p = 0.265$ , (d)  $p = 1$ .

### 3.7.2. Modelo de Barabási-Albert en Python

En la librería *NetworkX* también existe una función que permite generar redes aleatorias mediante el modelo de Barabási-Albert:

```
barabasi_albert_graph(n, m, seed = None, initial_graph = None)
```

cuyos parámetros son:

- **n: int**

Número de nodos.

- **m: int**

Número de arcos que conectan un nodo nuevo a nodos existentes.

- **seed: integer, random\_state, o None (por defecto)**

Indicador del estado de generación de números aleatorios.

- **initial\_graph: Graph o None (por defecto)**

Red inicial para el algoritmo de Barabási–Albert. Debe ser un grafo conectado para la mayoría de los casos. Se utiliza una copia de *initial\_graph*. Si es *None*, comienza desde un grafo de estrella con  $(m + 1)$  nodos.

Se debe cumplir  $1 \leq m < n$  y  $m \leq m_0 \leq n$ , donde  $m_0$  es el número de nodos de la red inicial.

Por ejemplo, tomando como red inicial la red (c) de la figura 18, generada mediante el modelo de Érdos-Rényi, con  $m_0 = 20$ , se puede generar una nueva red mediante el modelo de Barabási–Albert. El siguiente código de Python construye una red aleatoria mediante el método de Barabási-Albert con  $n = 25$  nodos y  $m = 10$ , partiendo de una red aleatoria generada mediante el método de Érdos-Rényi con  $m_0 = 20$  nodos y una probabilidad de enlace de  $p = 0.265$ :

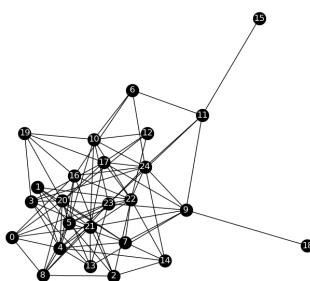
```
import networkx as nx
import matplotlib.pyplot as plt

GERc = nx.erdos_renyi_graph(n=20, p=0.265)

GBA = nx.barabasi_albert_graph(n=25, m=10, initial_graph = GERc)

fig, ax = plt.subplots(figsize=(4, 4))
nx.draw(GBA, with_labels=True, ax=ax, node_color='black', font_color='white')
```

Se obtiene la siguiente red representada en la Figura 19.



**Figura 19:** Red aleatoria generada en Python mediante el método de Barabási-Albert con  $n = 25$  nodos y  $m = 10$ , a partir de una red aleatoria obtenida mediante el método de Érdos-Rényi con  $m_0 = 20$  nodos y probabilidad de enlace  $p = 0.265$ .

### 3.7.3. Modelo de Watts-Strogatz en Python

Por otro lado, en la librería *NetworkX* también se encuentra una función para generar redes aleatorias mediante el modelo de Watts-Strogatz:

`watts_strogatz_graph(n, k, p, seed = None)`

cuyos parámetros son:

- ***n: int***

Número de nodos.

- ***k: int***

Cada nodo está conectado a sus *k* vecinos más cercanos.

- ***p: float***

Probabilidad de reconectar cada arco.

- ***seed: integer, random\_state, o None (por defecto)***

Indicador del estado de generación de números aleatorios.

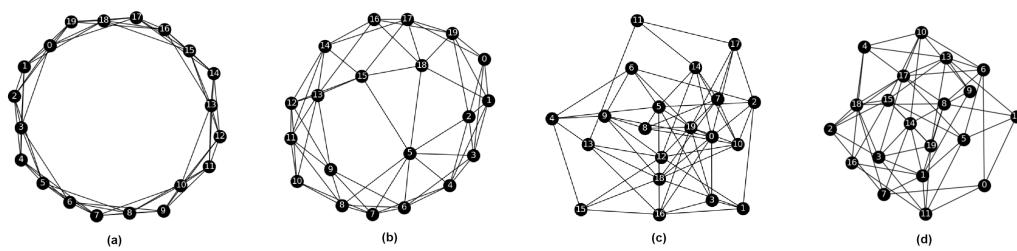
Probemos a construir las redes vistas en la figura 16. Por ejemplo, para crear y representar la red (a) se puede usar el siguiente código en Python que contruye una red aleatoria mediante el método de Watts-Strogatz con *n* = 20 nodos y *k* = 6, con probabilidad de reconexión *p* = 0:

```
import networkx as nx
import matplotlib.pyplot as plt

GWSa = nx.watts_strogatz_graph(n=20, k=6, p=0)

fig, ax = plt.subplots(figsize=(4, 4))
nx.draw(GWSa, with_labels=True, ax=ax, node_color='black', font_color='white')
```

Análogamente se pueden obtener las redes (b) usando *p* = 0.1, (c) usando *p* = 0.5 y (d) usando *p* = 0.95, con lo cual se obtienen las representaciones de la Figura 20.



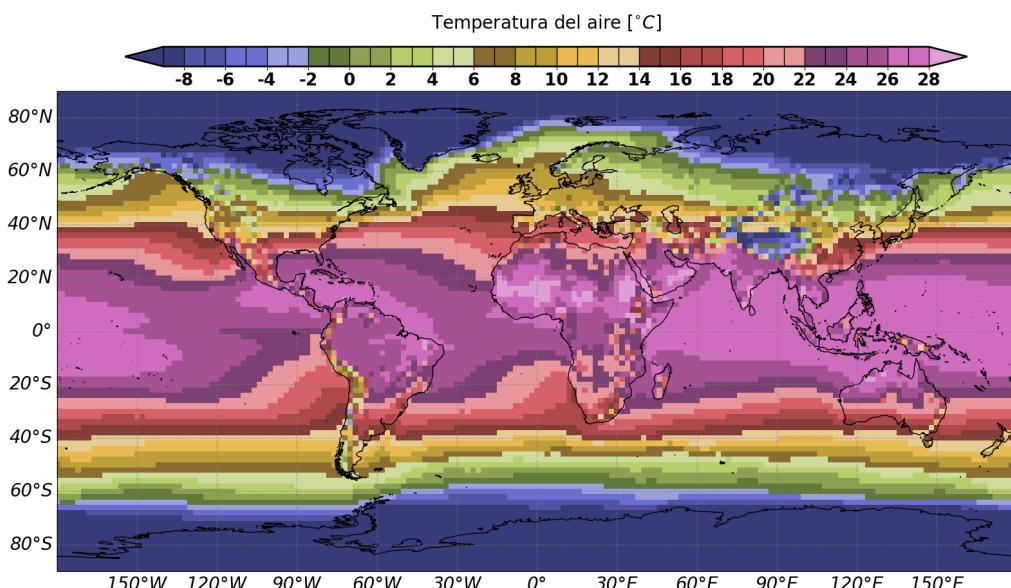
**Figura 20:** Redes aleatorias de Watts-Strogatz generadas en Python con *n* = 20 nodos, *k* = 6, con probabilidades de reconexión (a) *p* = 0, (b) *p* = 0.1, (c) *p* = 0.5, (d) *p* = 0.95.

## 4. Aplicación de la Teoría de Redes al clima: buscando patrones en la temperatura del aire

Las redes climáticas se construyen mediante el análisis estadístico de las interdependencias entre pares de series temporales que describen variables climatológicas en diferentes regiones geográficas de la Tierra, las cuales, a su vez, constituirían los nodos de la red (Dijkstra et al., 2019). Estas interdependencias se pueden estimar con diferentes técnicas, siendo la más sencilla de ellas el cálculo de correlaciones.

En esta sección, se aplica la Teoría de Redes a una base de datos de temperatura del aire a 2 metros sobre el nivel del mar ( $t2m$  a partir de ahora). Los datos de  $t2m$  provienen de una simulación climática de alta resolución, ERA 5, consistente en un reanálisis de cobertura global con una resolución nativa de  $0.25^\circ \times 0.25^\circ$  de resolución espacial (es decir, cada cuarto de grado en latitud y longitud se dispone de una serie de tiempo). Los datos de ERA 5 se encuentran alojados en la Copernicus Climate Data Store, y son accesibles de manera gratuita en <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>.

En la Figura 21 se representa el mapa promedio de  $t2m$  para el periodo 01/1940–03/2024.



**Figura 21:** Mapa de temperatura promedio del aire a una altura de 2 metros sobre el nivel del mar ( $t2m$ ) desde enero de 1940 hasta marzo de 2024.

#### 4.1. Pasos en la construcción de la red de temperatura del aire

Para construir las redes de temperatura, es necesario obtener la matriz de adyacencia entre los diferentes nodos del sistema, en este caso las celdas del modelo numérico. Se van a seguir los siguientes pasos resumidos en la Figura 22 (Dijkstra et al., 2019):

- **Reducción del tamaño de los datos:** Aunque originalmente los datos son horarios, por limitaciones computacionales se han empleado promedios mensuales desde enero de 1940 hasta marzo de 2024. Por la misma razón, se ha tomado un dato de cada 8, de manera que la resolución espacial es de  $2^\circ \times 2^\circ$ . En total se dispone de 16380 ( $180 \times 91$ ) localizaciones, y para cada una de ellas se tiene una serie de tiempo de 1011 valores de temperatura del aire.
- **Eliminación de la tendencia lineal y del ciclo estacional:** Para estudiar la relación de las series de temperatura entre diferentes localizaciones a escala interanual, es conveniente eliminar la tendencia lineal y el ciclo estacional. La presencia de tendencias a largo plazo (causadas por ejemplo por el calentamiento global antropogénico, o cambios en parámetros astronómicos como la nutación o la precesión) y variaciones estacionales (a causa del movimiento de traslación terrestre), opacan la identificación de otras señales, puesto que suelen ser dominantes. Por otra parte, las señales climáticas con variabilidad interanual son muy diversas, y su grado de importancia depende de la zona de estudio. Entre las más importantes cabe citar El Niño Oscilación del Sur, la Oscilación del Atlántico Norte, o la Oscilación Ártica, entre muchas otras. Los detalles sobre la eliminación de la tendencia lineal y el ciclo estacional se encuentran en A. Detalles del procesamiento de datos de temperatura del aire para el cálculo de la matriz de adyacencia.
- **Cálculo de las correlaciones:** Para medir el grado de interdependencia entre las series de anomalía de  $t2m$  para todas las celdas del modelo, se procedió a calcular el coeficiente de correlación de un nodo con respecto a todo el resto de nodos del sistema. Para tener en cuenta tanto las relaciones lineales como las no lineales, y ver como el grado de no linealidad puede alterar la estructura de la red, se han empleado dos coeficientes de correlación: el de Pearson y el de Spearman.

**Definición 4.1** *En el análisis de series temporales, el coeficiente de correlación de Pearson es una medida de similaridad entre dos series temporales, X e Y, sincrónicamente muestreadas, en función de un desplazamiento temporal (o lag),  $\tau$ , de una*

con respecto a la otra. Se define como:

$$\rho_{X,Y}(\tau) = \frac{\text{cov}(X(t), Y(t + \tau))}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{E[(X(t) - \mu_X)(Y(t + \tau) - \mu_Y)]}{\sigma_X \sigma_Y} \quad (62)$$

donde  $\text{cov}(X, Y)$  es la covarianza de las dos series,  $\mu_X$  y  $\mu_Y$  son sus medias, y  $\sigma_X^2$  y  $\sigma_Y^2$  son sus varianzas. Una medida alternativa de correlación entre dos variables es la correlación de Spearman, que es igual a la correlación de Pearson entre los valores de rango de esas dos variables (es decir, las etiquetas de posición relativas).

Los detalles sobre el cálculo de las correlaciones se encuentran en B. Detalles del cálculo de las correlaciones.

- **Obtención de las matrices de adyacencia:** Una vez se han obtenido todas las correlaciones, queda establecer si existe una conexión o no entre dos nodos cualquiera. Esto requiere estudiar la dependencia entre series temporales en diferentes ubicaciones geográficas. Para ello es habitual definir un límite del coeficiente de correlación, a partir del cual considerar que sí existe una interdependencia.

Supongamos ahora que  $S_{ij}$  es uno de los coeficientes de correlación anteriormente definidos, midiendo la interdependencia entre pares de series temporales. Se considera que los enlaces significativos son aquellos que tienen valores  $S_{ij}$  mayores que un cierto umbral  $W$ . En este caso, se ha seleccionado un umbral para el coeficiente de correlación de  $W = 0.5$  (en valor absoluto). Entonces la matriz de adyacencia se determina como:

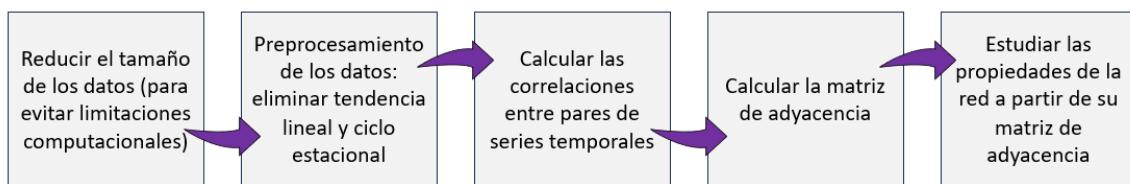
$$A_{ij} = H(S_{ij} - W) \quad (63)$$

donde  $H$  es la función de Heaviside:

$$H(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1, & \text{si } x \geq 0 \end{cases} \quad (64)$$

El 1 en la definición (64) se entiende como la existencia de una conexión (un arco) entre los nodos (las dos ubicaciones geográficas). Los detalles sobre cómo implementar el cálculo de la matriz de adyacencia en Python se encuentran en C. Detalles del cálculo de la matriz de adyacencia.

## PASOS EN LA CONSTRUCCIÓN Y ESTUDIO DE LA RED DE TEMPERATURA DEL AIRE



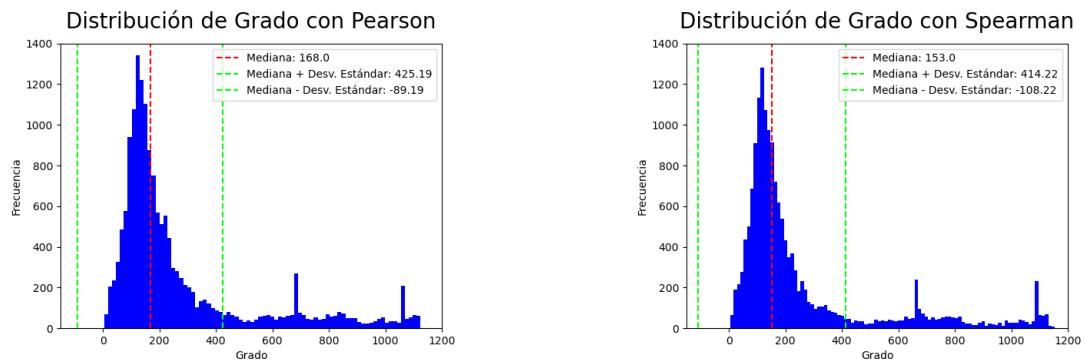
**Figura 22:** Esquema con los pasos para la construcción y el análisis de la red de temperatura del aire.

### 4.2. Estudio de las propiedades de la red obtenida mediante representaciones gráficas

En este último apartado se presentan los resultados del análisis de las redes obtenidas, basadas en las matrices de adyacencia calculadas en la sección C. El estudio se centra en las principales propiedades estructurales de la red, como el número de nodos, el grado de los nodos, el diámetro y diversas medidas de centralidad. Estas propiedades son fundamentales para comprender la dinámica y la estructura interna de la red.

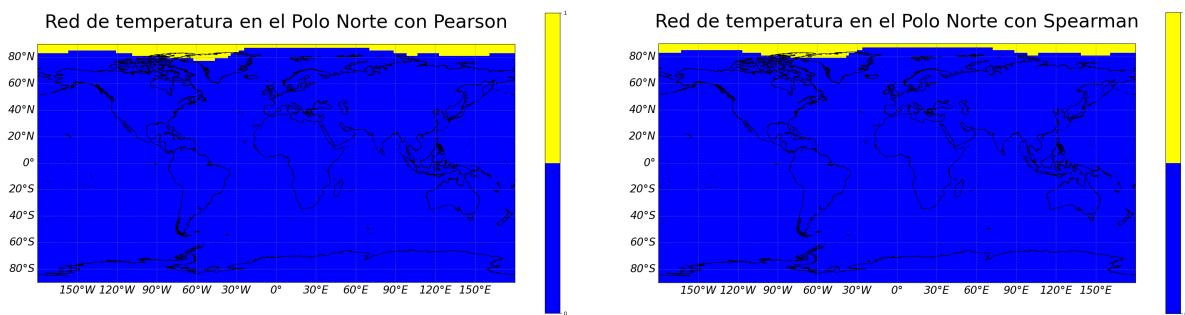
Además, dichas propiedades se implementan a través de representaciones gráficas, que permiten identificar relaciones importantes entre los nodos, así como destacar los nodos más centrales dentro de la estructura de la red. Para cada propiedad de la red, se realizarán dos gráficos: uno utilizando la matriz de adyacencia calculada con el coeficiente de correlación de Pearson, y el otro empleando la matriz calculada con el coeficiente de correlación de Spearman. Esto permite comparar cómo varían las propiedades de la red en función del coeficiente de correlación utilizado para construir la matriz de adyacencia.

Para empezar, se obtiene la distribución de grado. Para ello, se calcula el grado de cada nodo de la red, y se crea un histograma a partir de dichos grados. En la Figura 23 se muestran los histogramas del grado de las redes asociadas a cada nodo, para cada coeficiente de correlación estudiado. Como se observa, tanto la mediana como la desviación estándar de las distribuciones son muy similares, lo que indica que en las series de tiempo domina la linealidad. Esto no es sorprendente puesto que se está trabajando con promedios mensuales, por lo que los efectos no lineales se deben haber filtrado durante el preprocesado de los datos.



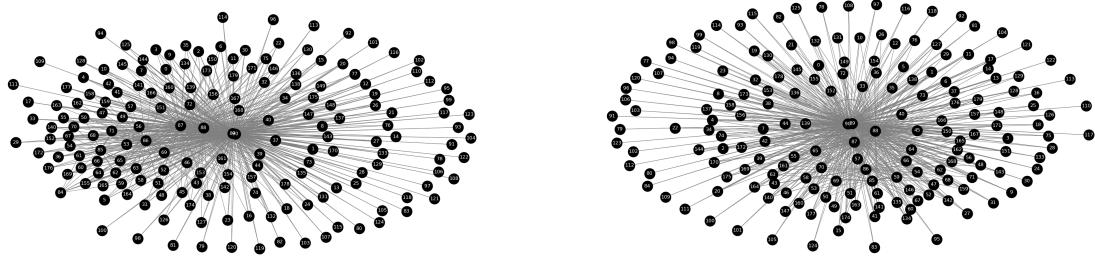
**Figura 23:** Distribución de grado de las redes de temperatura obtenidas usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

A continuación, se van a seleccionar varios nodos que puedan ser interesantes y, para cada uno de ellos, se representan en un mapa las localizaciones cuyos datos de temperatura estén relacionados con los del nodo seleccionado. Es decir, se muestran los nodos que están enlazados con el nodo seleccionado según la matriz de adyacencia. Además, se hará un dibujo de la red obtenida para el nodo seleccionado. Se empieza eligiendo un nodo en el polo Norte, con latitud  $90^{\circ}$  y longitud  $0^{\circ}$  (Figura 24).



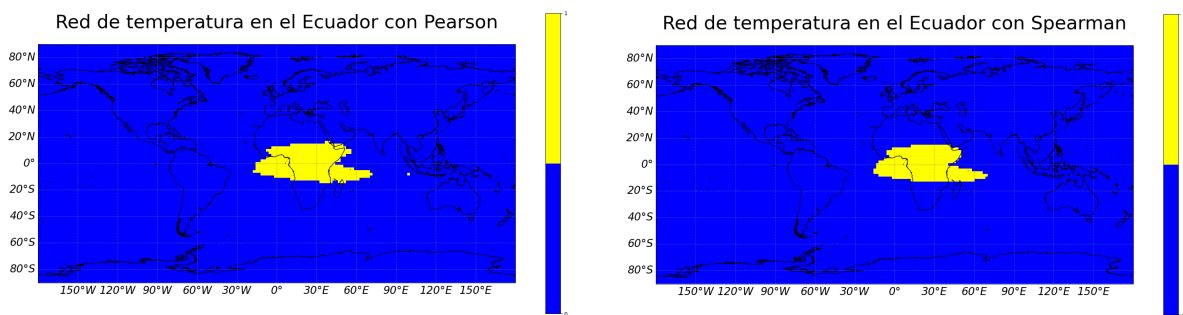
**Figura 24:** Mapas de las redes de temperatura en el Polo Norte usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

Si se grafica la red asociada al nodo seleccionado para cada coeficiente de correlación se obtiene la Figura 25. Como se observa la estructura de la red es similar, aunque en el caso del coeficiente de Pearson (panel izquierdo) hay más nodos concentrados hacia la izquierda, mientras que en el caso de Spearman (panel derecho) están más concentrados hacia abajo y la derecha.



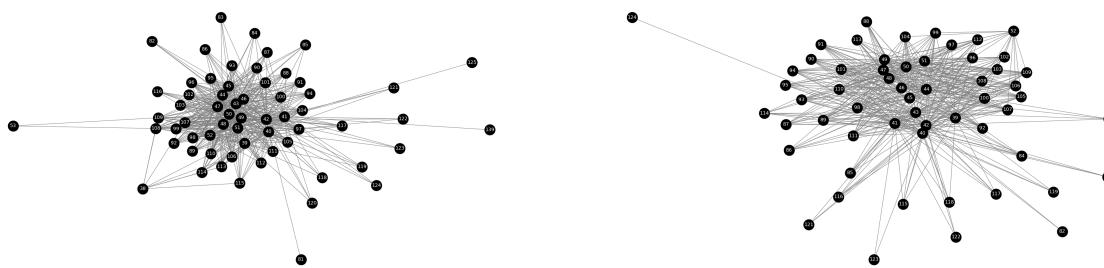
**Figura 25:** Redes de temperatura en el Polo Norte usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

El siguiente punto está ubicado en el ecuador, con latitud  $0^\circ$  y longitud  $0^\circ$  (Figura 26).



**Figura 26:** Mapas de las redes de temperatura en el Ecuador usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

A diferencia del punto ubicado en el norte, las redes asociadas al punto del ecuador están mucho más concentradas (Figura 27).

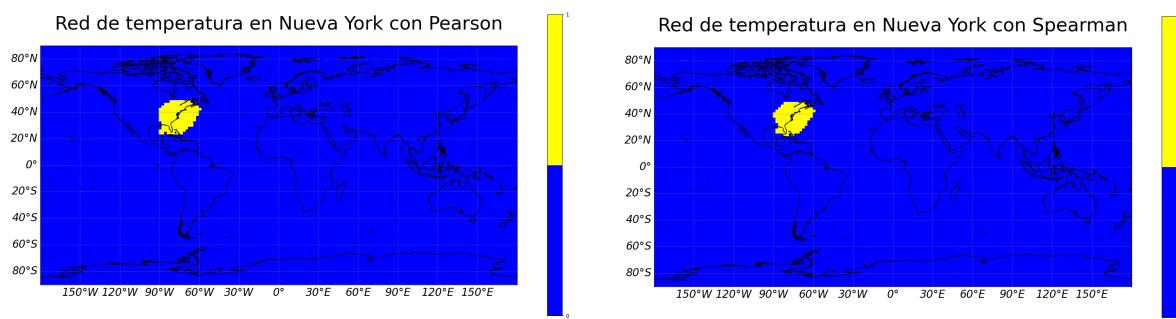


**Figura 27:** Redes de temperatura en el Ecuador usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

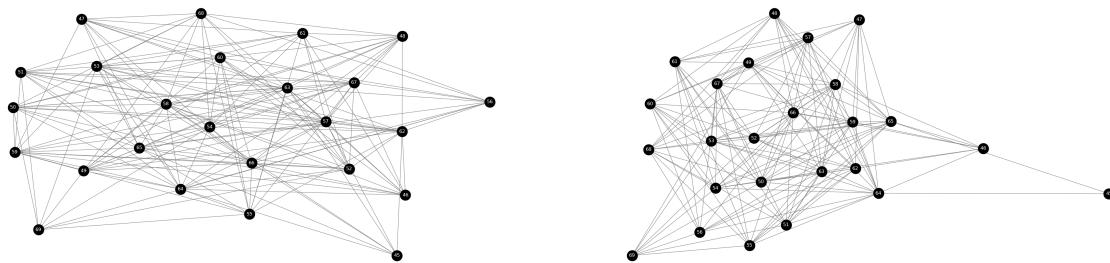
El hecho de que las redes en el polo Norte estén más dispersas se puede explicar en parte por la mayor cobertura geográfica y su mayor variabilidad. Cabe destacar que las

zonas polares son especialmente sensibles al calentamiento global antropogénico, además de experimentar fenómenos como el deshielo estacional y la congelación, que pueden afectar grandes extensiones de terreno simultáneamente. Por otra parte, la red asociada al punto ubicado en el ecuador es probable que esté más concentrada a causa de su mayor homogeneidad térmica en las primeras capas de la troposfera.

Por último, se han seleccionado puntos ubicados en latitudes medias. En primer lugar se ha elegido el más cercano a la ciudad de Nueva York (Figura 28), con latitud  $40.7128^{\circ}$  y longitud  $-74.0060^{\circ}$ .

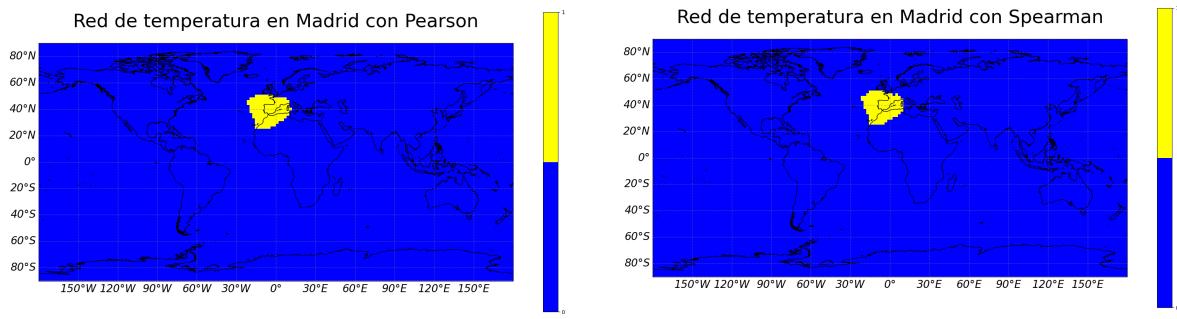


**Figura 28:** Mapas de las redes de temperatura en Nueva York usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

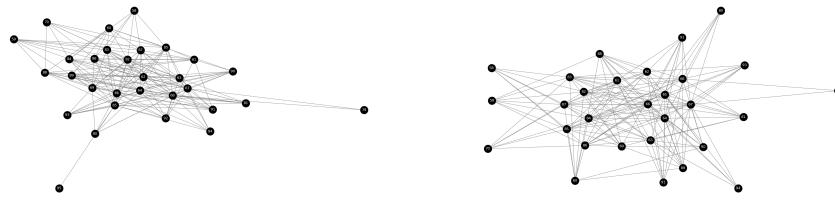


**Figura 29:** Redes de temperatura en Nueva York usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

En este caso, hay un menor número de nodos con correlaciones por encima del umbral, encontrándose bastante dispersos entre ellos. Sin embargo esta dispersión es menor en el caso del coeficiente de correlación de Spearman en comparación con el de Pearson (comparar paneles en la Figura 29). Si se repite el análisis para el nodo más cercano a Madrid, con latitud  $40.4168^{\circ}$  y longitud  $-3.7038^{\circ}$  (Figura 30), los resultados son similares, aunque ligeramente las estructuras presentan una dispersión menor (Figura 31).

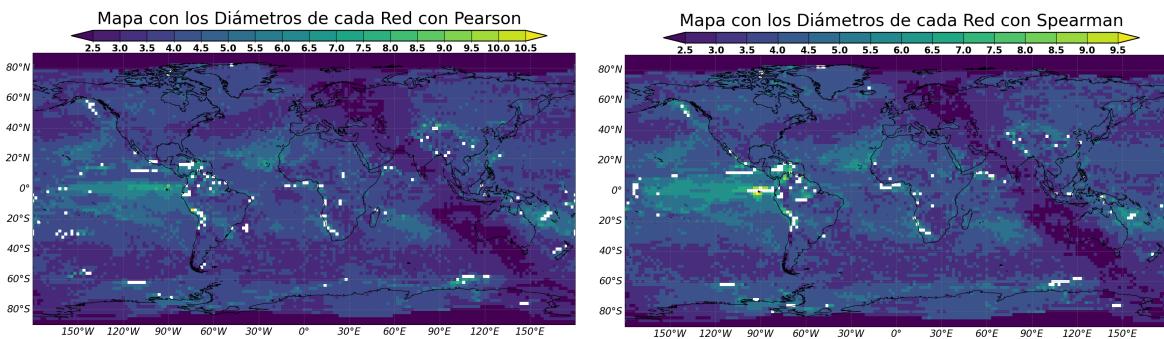


**Figura 30:** Mapas de las redes de temperatura en Madrid usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).



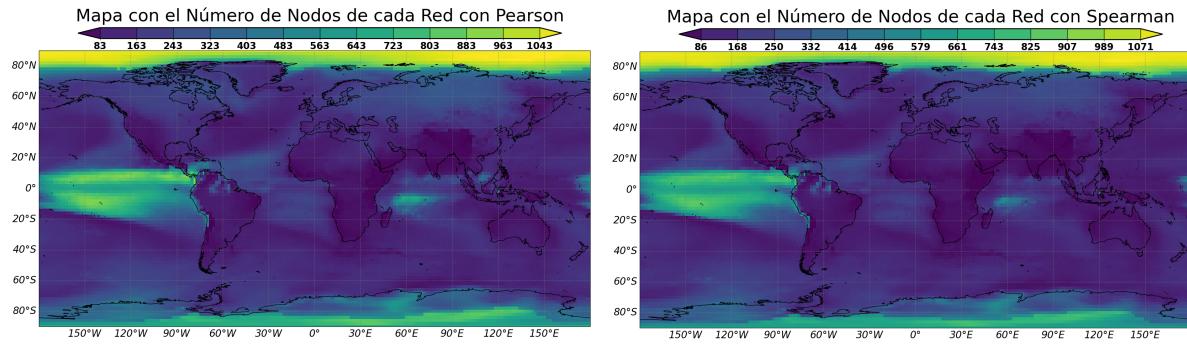
**Figura 31:** Redes de temperatura en Madrid usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

Como es imposible mostrar las más de 16000 redes, a continuación se muestran algunas propiedades de las redes asociadas a cada uno de los nodos sobre un mapa, tal que el valor en cada nodo debe entenderse como el valor asociado a la red de ese nodo. Para empezar, se incluirá un mapa en el que se indique el diámetro de la red correspondiente a cada nodo. Se omitirán aquellos nodos cuya red asociada no esté conectada, identificados por el color blanco (Figura 32). El diámetro toma valores entre 2 (latitudes medias preferentemente) y 11 (zonas tropicales).



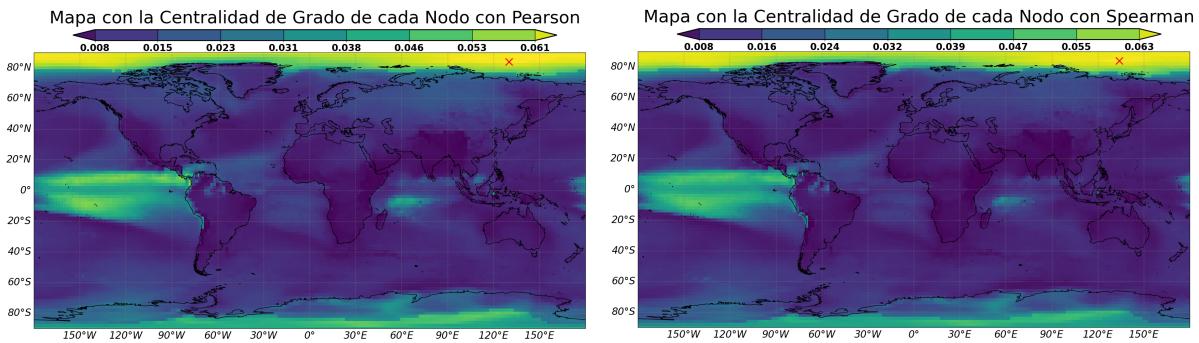
**Figura 32:** Mapa con el diámetro de las redes para cada nodo, usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

En los mapas siguientes se representa el número de nodos conectados a cada localización, es decir, el número de nodos de la red correspondiente a la localización seleccionada. Tal como se ha indicado anteriormente, se puede observar que las redes con mayor número de nodos se encuentran en los polos y en el ecuador, oscilando el número entre 3 y alrededor de 1100 (Figura 33).



**Figura 33:** Mapa con el número de nodos de las redes asociadas a cada nodo, usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

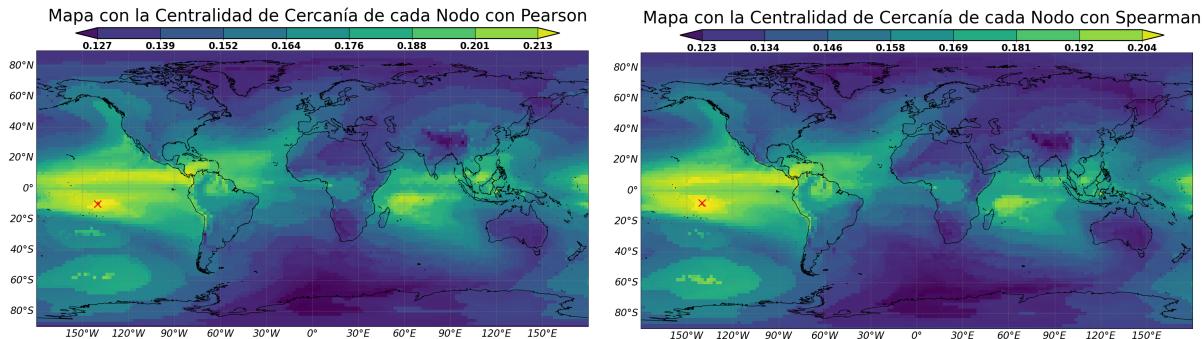
Para finalizar, vamos a analizar medidas de centralidad a través de los mapas. En los siguientes mapas, se representa la centralidad de grado de cada nodo de la red, marcando con una cruz roja el nodo más central según la centralidad de grado (los valores de la centralidad de grado están normalizados dividiéndolos por el grado máximo posible,  $n - 1$ , donde  $n$  es el número de nodos en la red). Es notable destacar que las redes cuyos nodos más relevantes están en el centro de la red se encuentran en las regiones polares y en el área donde se produce el fenómeno de El Niño (Pacífico oriental, ver Figura 34)



**Figura 34:** Mapa con la centralidad de grado de cada nodo, marcando con una cruz roja el nodo más central de la red, usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

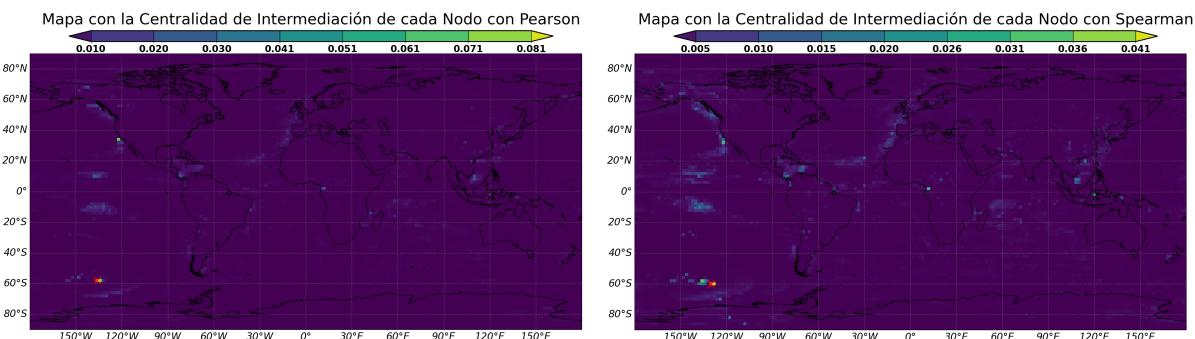
Esto sugiere que las regiones polares, particularmente el polo Norte, juegan un papel crucial en la conectividad global de la red de temperatura.

En la Figura 35 se muestra la centralidad de cercanía (normalizada) para cada uno de los nodos de la red, siendo marcado con una cruz roja el nodo con una mayor centralidad de cercanía. Se observa que en este caso los nodos con mayor centralidad se encuentran mayoritariamente cercanos al ecuador.



**Figura 35:** Mapa con la centralidad de cercanía de cada nodo, marcando con una cruz roja el nodo más central de la red, usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

En la Figura 36 se representa la centralidad de intermediación de los nodos de la red, marcando al nodo con mayor centralidad de intermediación con una cruz roja. Es destacable que las regiones con mayor centralidad de intermediación se encuentran preferentemente sobre los océanos.



**Figura 36:** Mapa con la centralidad de intermediación de cada nodo, marcando con una cruz roja el nodo más central de la red, usando el coeficiente de correlación de Pearson (a la izquierda) y de Spearman (a la derecha).

## 5. Conclusiones

En este trabajo final de grado se han revisado los fundamentos matemáticos de la Teoría de Redes, tanto desde el punto de vista teórico como computacional, y su aplicación en las ciencias del clima.

En primer lugar, se ha comenzado con la exploración de los principales conceptos de la Teoría de Grafos, iniciando con las definiciones básicas, y adentrándonos en la manera de representar las redes en forma matricial, algo de gran utilidad a la hora de analizar sistemas complejos como el clima. Además, por medio de las medidas de centralidad, se han examinado diversas métricas asociadas a las redes.

A continuación, se han estudiado diferentes modelos de redes aleatorias que, a pesar de no ser representativas de la naturaleza, pues la mayoría de sistemas reales tienen una estructura definida, nos han proporcionado una comprensión más profunda de los principios fundamentales que hay detrás de la estructura y la dinámica de las redes en general.

Finalmente, se ha profundizado en una aplicación específica de la Teoría de Redes consistente en el estudio del sistema climático mediante el análisis de las interdependencias de series de temperatura del aire a escala global. El estudio de datos climáticos a través de la matriz de adyacencia ha permitido identificar patrones y tendencias, potencialmente útiles para comprender la distribución y la evolución del clima a lo largo del tiempo.

El estudio de las redes de temperatura del aire a partir de la matriz de adyacencia ha permitido ejemplificar la complementariedad de este tipo de técnicas con respecto a los métodos de análisis de datos más tradicionales. En el contexto del clima, estas técnicas aún no están muy extendidas, y su combinación con técnicas de aprendizaje automático/profundo ofrecen nuevas oportunidades para monitorizar cambios en variables de especial interés, así como para acercar algo de luz a problemas en los que es urgente extender el horizonte de predicción a causa de sus consecuencias dramáticas, como por ejemplo la intensidad de los huracanes, las olas de calor o el fenómeno de El Niño.

## Referencias

- Barabási, A.-L. (2016). *Network Science*. Cambridge University Press, Cambridge, UK.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.*
- Borrett, S. R., Moody, J., and Edelmann, A. (2014). The rise of network ecology: Maps of the topic diversity and scientific collaboration. *Ecological Modelling*, 293:111–127.
- de Sola Pool, I. and Kochen, M. (1978). Contacts and influence. *Social Networks*, 1(1):5–51.
- Dijkstra, H. A., Hernández-García, E., Masoller, C., and Barreiro, M. (2019). *Networks in Climate*. Cambridge University Press, Cambridge, UK.
- Estrada, E. and Knight, P. (2015). *A First Course in Network Theory*. Oxford University Press, Oxford, UK.
- Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Inductive representation learning on large graphs. *NIPS 2017*.
- Ibáñez, R. M. (2017a). *Estadística básica I. Introducción a la estadística*. Publicaciones de la Universidad de Alicante, Alicante, España.
- Ibáñez, R. M. (2017b). *Estadística básica: II. Probabilidad: variables aleatorias*. Publicaciones de la Universidad de Alicante, Alicante, España.
- Lewis, T. G. (2009). *Network Science: Theory and Applications*. John Wiley Sons, Hoboken, New Jersey.
- Milgram, S. (1967). The small world problem. *Psychology Today*.
- Tortosa, C. G. (2020). *Introducción al cálculo numérico*. Publicaciones de la Universidad de Alicante, Alicante, España.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2018). Dynamic graph cnn for learning on point clouds. *arXiv*.
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.

## A. Detalles del procesamiento de datos de temperatura del aire para el cálculo de la matriz de adyacencia

Dado que la variable  $t2m$  contiene valores faltantes (NaN), estos pueden ser tratados de distintas formas. A continuación se muestran dos posibles aproximaciones:

### A.0.1. Conservación de los valores NaN

En este primer enfoque se eliminan la tendencia lineal y la estacionalidad de los datos sin modificar los valores faltantes NaN.

Para ello, en Python se puede usar un bucle doble para iterar sobre cada punto de latitud y longitud en los datos de temperatura. A continuación, se crea un array booleano *mask* que contendrá *True* si el valor correspondiente en la serie temporal no es NaN y *False* en caso contrario.

Se selecciona la serie temporal de temperatura y se ajusta un modelo de regresión lineal utilizando únicamente los valores no faltantes mediante la función *linregress* de *scipy.stats*. Esta función devuelve, entre otros, la pendiente y la intersección de la regresión lineal. Se usa la pendiente para eliminar la tendencia lineal de los datos de temperatura, multiplicando la pendiente por el tiempo y restando el resultado de los datos originales.

```
import numpy as np
import scipy.stats as scst

# Variable para almacenar los datos de temperatura sin tendencia lineal
t2m_detrended = t2m.copy()

# Para cada punto de latitud y longitud, ajustar el modelo y eliminar la tendencia lineal
for i in range(len(latitude)):
    for j in range(len(longitude)):

        # Array booleano para seleccionar los valores válidos (no NaN)
        mask = ~np.isnan(t2m[:, i, j])

        # Obtener los datos de temperatura para la latitud y longitud específicas
        serie = t2m[:, i, j]

        # Regresión lineal
        slope, intercept, r_value, p_value, std_err = scst.linregress(time[mask], serie[mask])

        # Eliminar la tendencia lineal
        t2m_detrended[:, i, j] = t2m[:, i, j] - (slope*time)
```

Una vez eliminada la tendencia lineal, se debe eliminar el ciclo estacional. Para ello, se

usa un bucle triple para iterar sobre cada punto de latitud y longitud, así como sobre los 12 meses del año. Ahora, se selecciona dicha serie y se calcula la media de los valores no faltantes para el mes correspondiente. Por último, se elimina el ciclo estacional restando la media del mes correspondiente de los datos de temperatura. A continuación se muestra el código en Python:

```
import numpy as np

# Variable para almacenar los datos de temperatura sin tendencia lineal ni ciclo
# estacional
t2m_det_des = t2m_detrended.copy()

# Matriz para guardar los ciclos estacionales de todos los puntos
seasonal_cycle = np.zeros((12, len(latitude), len(longitude)))

# Para cada punto de latitud y longitud, y para cada mes del año, calcular la media
# y eliminar el ciclo estacional
for i in range(len(latitude)):
    for j in range(len(longitude)):
        for k in range(0,12):

            # Calcular la media de los valores no faltantes
            seasonal_cycle[k,i,j] = np.nanmean(t2m_detrended[k::12,i,j])

            # Eliminar el ciclo estacional
            t2m_det_des[k::12,i,j] = t2m_detrended[k::12,i,j] - seasonal_cycle[k,i,j]
```

Después de aplicar los procesos de eliminación de tendencia lineal y ciclos estacionales, el resultado final es un conjunto de datos de temperatura, *t2m\_det\_des* que ha sido preprocesado para eliminar tanto la tendencia lineal como los ciclos estacionales. En esta serie se siguen conservando los valores NaN. Esto puede ser relevante, pues dichos valores indican la ausencia de datos, lo que puede ser información relevante en sí misma. Por ejemplo, pueden indicar la falta de observaciones en ciertos períodos de tiempo o ubicaciones geográficas. Conservando estos valores mantenemos la completitud y la estructura de los datos originales.

### A.0.2. Interpolación lineal para estimar los valores NaN

Este segundo enfoque aborda el problema de los valores NaN en los datos climáticos mediante la interpolación, lo que permite rellenar los huecos disponibles para el análisis. Además, al ser la temperatura una variable continua y con variaciones normalmente suaves, sobretodo si hablamos de promedios mensuales, puede ser razonable interpolar para estimar los valores faltantes.

Para ello, usando Python, se itera sobre todas las posiciones de latitud y longitud, y

se obtienen los datos de temperatura para cada una de estas ubicaciones. De esta forma, se crea una serie temporal de temperaturas para cada ubicación específica.

A continuación, se realiza una interpolación para estimar los valores NaN en la serie temporal de temperaturas, usando la función *interp* de la librería *NumPy*, utilizando los índices de tiempo para los cuales no faltan datos, es decir,  $\sim np.isnan(temperatura)$ .

Por último, una vez obtenidos los valores estimados para los puntos faltantes, se reemplazan los valores NaN de la serie temporal con los valores interpolados. A continuación se detalla el código para realizar este proceso:

```
import numpy as np

for i in range(len(latitude)):
    for j in range(len(longitude)):

        # Obtener los datos de temperatura para la latitud y longitud específicas
        temperatura = t2m[:, i, j]

        # Realizar interpolación lineal
        temperatura_interp = np.interp(np.arange(len(temperatura)),
                                         np.arange(len(temperatura))[~np.isnan(temperatura)],
                                         temperatura[~np.isnan(temperatura)])

        # Reemplazar los valores NaN con la interpolación lineal
        t2m[:, i, j] = np.where(np.isnan(t2m[:, i, j]), temperatura_interp, t2m[:, i, j])
```

Una vez estimados los valores NaN mediante interpolación, se procede a eliminar la tendencia lineal de los datos, es decir, ajustar los cambios en la temperatura a lo largo del tiempo. Para ello, se usará regresión lineal. Para crear un modelo de regresión lineal, se puede utilizar la clase *LinearRegression* del módulo *sklearn.linear\_model*.

De nuevo, se itera sobre todas las posiciones de latitud y longitud, y para cada ubicación se obtienen los datos de temperatura y se ajusta el modelo de regresión lineal. En dicho modelo, la variable dependiente es la temperatura correspondiente, y la variable independiente es el tiempo, *time*.

Una vez ajustado el modelo, se utiliza el método *predict* para obtener la tendencia lineal, y a continuación se resta esta tendencia de los datos originales de temperatura, lo que elimina la tendencia lineal de los datos. En el siguiente código se desarrolla este proceso:

```
from sklearn.linear_model import LinearRegression

# Crear un modelo de regresión lineal
lr = LinearRegression()

# Variable dependiente y: temperatura (t2m)
```

```
# Variable independiente X: tiempo (time)

# Las variables independientes deben estar en una matriz en la que cada fila representa
# una muestra, y cada columna una característica
X = time.reshape(-1, 1)

# Para cada punto de latitud y longitud, ajustar el modelo y eliminar la tendencia lineal
for i in range(len(latitude)):
    for j in range(len(longitude)):

        # Obtener los datos de temperatura para la latitud y longitud específicas
        y = t2m[:, i, j]

        # Ajustar el modelo de regresión lineal
        lr.fit(X, y)

        # Eliminar la tendencia lineal
        tendencia_lineal = lr.predict(X)
        t2m[:, i, j] = t2m[:, i, j] - tendencia_lineal
```

Para finalizar, veamos cómo eliminar el ciclo estacional de los datos. Para ello, de nuevo se recorren todos los puntos de latitud y longitud en los datos de temperatura. Para cada ubicación específica, se realiza una descomposición estacional de la serie temporal de temperatura usando *seasonal\_decompose* de *statsmodels*. Mediante el parámetro *period* = 12 se indica que la serie tiene un ciclo estacional de 12 períodos (meses en un año).

La descomposición estacional produce tres componentes: tendencia, estacionalidad y residuo. Aquí, estamos interesados en la componente de estacionalidad. Luego, el ciclo estacional se elimina de los datos originales de temperatura restando la estacionalidad estimada.

```
import statsmodels.api as sm

# Para cada punto de latitud y longitud, hacer la descomposición estacional y eliminar
# el ciclo estacional
for i in range(len(latitude)):
    for j in range(len(longitude)):

        # Descomposición estacional con período de 12
        descomposicion = sm.tsa.seasonal_decompose(t2m[:, i, j], period=12)

        # Eliminar el ciclo estacional
        estacionalidad = descomposicion.seasonal
        t2m[:, i, j] = t2m[:, i, j] - estacionalidad
```

## B. Detalles del cálculo de las correlaciones

En este apartado, se presenta cómo se ha realizado el cálculo de las correlaciones entre pares de series temporales. Para ello, se pueden usar las funciones *pearsonr* y *spearmanr* de la biblioteca *scipy.stats* de Python, que servirán para calcular los coeficientes de correlación de Pearson y de Spearman respectivamente.

Para reducir el coste computacional, se realizará un reshape de los datos de temperatura almacenados en *t2m\_det\_des*, que tiene 3 dimensiones (tiempo, latitud y longitud), transformándolo en una matriz de 2 dimensiones donde la primera dimensión es el tiempo, y la segunda es el producto de la latitud por la longitud. Además, los coeficientes de las correlaciones se almacenarán en matrices de 2 dimensiones, donde cada dimensión será el producto de las dimensiones espaciales, que serán finalmente expandidas a matrices de 4 dimensiones.

Para calcular los coeficientes, se irá iterando sobre los pares de puntos. Así, para cada par de puntos ( $k, j$ ):

- Se extraen las series temporales de los puntos  $k$  y  $j$ .
- Se excluyen los valores faltantes NaN (en el caso de que se hayan conservado en los pasos anteriores).
- Se calcula el coeficiente de correlación correspondiente y el p-valor asociado.
- Se almacenan los resultados en la matriz de correlaciones y de p-valores respectivamente.

Veamos cómo implementar este procedimiento en Python. Se muestra el caso con el coeficiente de correlación de Pearson, pero con el coeficiente de Spearman se hace de forma análoga usando *spearmanr* en lugar de *pearsonr*.

```
import numpy as np
from scipy.stats import pearsonr

nx = len(longitude)
ny = len(latitude)

# Matrices para almacenar las correlaciones y los p-valores
S_pearson = np.zeros((ny*nx, ny*nx))
p_values_pearson = np.zeros((ny*nx, ny*nx))

# Reshape de los datos de temperatura de 3D a 2D
t2m_reshape = np.reshape(t2m_det_des, [len(time), nx*ny])
```

```
# Se recorren los pares de puntos de la red
for k in range(0,ny*nx):
    for j in range(0,ny*nx):

        # Extraer las 2 series temporales
        serie1 = t2m_reshape[:, k]
        serie2 = t2m_reshape[:, j]

        # Máscara para excluir los valores NaN
        msk = (~np.isnan(serie1) & ~np.isnan(serie2))

        # Cálculo del coeficiente de correlación y el p-valor
        coef_pearson, p_value = pearsonr(serie1[msk], serie2[msk])

        # Se almacenan los resultados en matrices
        S_pearson[k, j] = coef_pearson
        p_values_pearson[k,j] = p_value

# Expandir a 4D
original_shape = (ny, nx, ny, nx)
expanded_S_pearson = np.reshape(S_pearson, original_shape)
expanded_p_values_pearson = np.reshape(p_values_pearson, original_shape)
```

## C. Detalles del cálculo de la matriz de adyacencia

En este apartado, nos centraremos en cómo calcular la matriz de adyacencia, una representación esencial para construir redes climáticas a partir de datos de temperatura. Tal y como se define en este caso, la matriz de adyacencia permite identificar la existencia de una relación estadística significativa entre series de tiempo de  $t2m$  de ubicaciones geográficas diferentes, es decir, ayuda a detectar aquellos lugares con una variabilidad interanual similar en las series de temperatura.

Como se ha indicado en la sección 4.1, esta relación se estima mediante los coeficientes de correlación de Pearson y Spearman. A continuación, se muestra cómo implementar el cálculo de la matriz de adyacencia en Python, usando el coeficiente de correlación de Pearson. De manera análoga, se podría realizar con el coeficiente de correlación de Spearman.

```
import numpy as np

# Fijamos un umbral
W = 0.5

# Función de Heaviside
def H(x):
    if x < 0:
        return 0
    elif x >= 0:
        return 1

# Inicializar la matriz de adyacencia
A_pearson_abs = np.zeros_like(expanded_S_pearson)

# Rellenar la matriz de adyacencia
for i in range(ny):
    for j in range(nx):
        for k in range(ny):
            for l in range(nx):
                A_pearson_abs[i, j, k, l] = H(abs(expanded_S_pearson[i, j, k, l]) - W)
```

Se observa que la matriz de adyacencia obtenida tiene 4 dimensiones. Sin embargo, es posible extraer submatrices de 2 dimensiones para cada nodo de la red. De esta forma, para cada nodo que escojamos, se puede construir una red con aquellos nodos que estén enlazados con el nodo seleccionado. Esto es de utilidad a la hora de representar para algunos nodos seleccionados su red correspondiente, es decir, cómo un nodo cualquiera está conectado con todo el resto.

## D. Detalles del desarrollo del trabajo

Tarea	Tiempo (horas)
Recopilación de materiales	5
Estudio de bibliografía	45
Elaboración de resultados gráficos/numéricos	45
Redacción de la memoria	55
Total	150

**Tabla 1:** Distribución aproximada del tiempo dedicado al trabajo final de grado.

Asignatura	Páginas	Descripción
Álgebra lineal I y II	11-13	Representación matricial de las redes.
Álgebra lineal I y II	46	Matriz de adyacencia en el estudio de la temperatura.
Algoritmia	40-43	Implementación de redes aleatorias en Python.
Algoritmia	48-53	Algoritmos en Python para estudiar la temperatura.
Análisis de datos I y II	56-62	Procesamiento de datos de temperatura en Python.
Cálculo numérico I	57-58	Interpolación: se explica en el libro Tortosa (2020).
Estadística	56-59	Regresión lineal: se explica en el libro Ibáñez (2017a).
Probabilidad	27-31	Distribución binomial y Poisson: están en Ibáñez (2017b).
Probabilidad	27-36	Redes aleatorias: probabilidades de enlace entre nodos.
Series temporales	44-53	Estudio de la temperatura del aire a lo largo del tiempo.
Topología básica	6-11	Los grafos son estructuras topológicas.
Topología básica	32-33	Fenómeno de mundo pequeño, es propiedad topológica.
Topología básica	33-34	Coeficientes de agrupamiento, son medidas topológicas.

**Tabla 2:** Asignaturas relacionadas con el trabajo.