



Universidad de Valladolid

FACULTAD DE CIENCIAS

Grado en Matemáticas

Análisis de redes complejas

Alumno: Patricia Baz Domínguez

Tutor: José Ignacio Farrán Martín

Fecha: 18 de septiembre de 2024

Resumen

Las redes sociales se modelan mediante grafos, cuyas aristas representan las relaciones entre los miembros de la red. Este tipo de grafos, llamados redes complejas por muchos autores, modelan también otro tipo de situaciones, como los enlaces y conexiones entre páginas *web*, grafos de colaboración, o redes de comunicaciones, entre otros. Estas redes complejas pueden estudiarse desde muchos puntos de vista: por ejemplo, establecer medidas de centralidad, que en el caso de redes sociales se corresponde con la localización de *influencers*; detectar comunidades de usuarios con alta densidad de interconexiones; estudiar la evolución de la red a lo largo del tiempo; estudiar la distribución de probabilidad del número de conexiones en los nodos, entre muchos otros. Este trabajo se centra en algunos de estos problemas, desde un punto de vista matemático.

Palabras claves: grafos, redes sociales, redes complejas

Abstract

Social networks are modeled using graphs, where the edges represent the relationships between the members of the network. These types of graphs, often referred to as complex networks by many authors, also model other types of situations, such as links and connections between web pages, collaboration graphs, or communication networks, among others. These complex networks can be studied from various perspectives: for example, determining centrality measures, which, in the case of social networks, corresponds to identifying influencers; detecting communities of users with high interconnection density; studying the evolution of the network over time; examining the probability distribution of the number of connections at the nodes, among many others. This work focuses on some of these problems from a mathematical standpoint.

Keywords: graphs, social networks, complex networks

Índice general

Lista de figuras	VII
Lista de tablas	VII
Índice de ecuaciones	XII
1 Introducción	1
1.1 Objetivos del trabajo	2
1.2 Estructura de la memoria	2
2 Conceptos previos	5
2.1 Teoría de grafos	5
2.1.1 Tipos	6
2.1.2 Terminología	9
2.1.3 Representación	12
2.1.4 Operaciones	14
2.1.5 Conectividad	14
2.2 Aproximaciones asintóticas y series infinitas	17
2.2.1 Factorial	17
2.2.2 Función exponencial	18
2.2.3 Función $(1 - \frac{x}{n})^{-k}$	18
2.3 Teoría de la probabilidad	18
2.3.1 Combinatoria	19
2.3.2 Modelos de variables aleatorias	20
3 Grafos y redes sociales	23
3.1 Introducción	23
3.2 Excentricidad	24
3.3 Densidad	25
3.4 Centralidad	27
3.5 Algoritmo de Floyd-Warshall	32
3.6 Detección de comunidades	34
3.6.1 Medida	34
3.6.2 Algoritmos de división	38
4 Redes complejas	41
4.1 Introducción	41

4.2	Modelos	42
4.2.1	Redes sociales	42
4.2.2	Redes de comunicación	43
4.2.3	Redes de información	43
4.2.4	Diseño de <i>software</i>	43
4.2.5	Redes de transporte	43
4.2.6	Redes biológicas	44
4.2.7	Torneos	44
4.3	Estudio	44
4.3.1	Estructura	44
4.3.2	Dinámica	45
4.4	Distribución de vecinos	45
4.4.1	Topología de Poisson	47
4.4.2	Crecimiento de redes	50
4.4.3	Topología exponencial (modelo de enlace igualitario)	53
4.4.4	Topología libre de escala (modelo de enlace preferencial)	54
4.5	Envejecimiento de nodos	58
4.5.1	Modelo de enlace preferencial con adaptabilidad	61
4.6	Islas gigantes	61
4.6.1	Distribución de Poisson	62
4.6.2	Distribución libre de escala	62
5	Ejemplos de casos reales	65
5.1	Grafo dirigido	66
5.1.1	Descripción del <i>dataset</i>	66
5.1.2	Importación de los datos	67
5.1.3	Contexto	67
5.1.4	Estadísticas	68
5.1.5	Resultados	70
5.1.6	Discusión	94
5.1.7	Conclusiones	96
5.2	Grafo no dirigido	96
5.2.1	Descripción del <i>dataset</i>	96
5.2.2	Importación de los datos	98
5.2.3	Contexto	98
5.2.4	Estadísticas	98
5.2.5	Resultados	98
5.2.6	Discusión	110
5.2.7	Conclusiones	111
I	Apéndices	113
A	Acrónimos	115
B	Métodos de crecimiento de redes	117
B.1	Método de la ecuación maestra	117

B.2	Método continuo	118
B.3	Método cinético	119
C	Transiciones de fase en redes complejas con topología de Poisson	121
C.1	Transición de segundo orden	121
C.2	Transición de primer orden	121
C.3	Conclusión	121
D	Código	123
E	Cálculos	125
E.1	Esquema	125
E.2	Desarrollo	128
	Bibliografía	143

Índice de figuras

2.1	Ejemplo de grafo.	5
2.2	Tipos de grafos.	7
2.3	Tipos especiales de grafos.	9
2.4	Cubo 3D en forma de grafo.	9
3.1	Grafos de ejemplo para centralidad y detección de comunidades.	28
4.1	Crecimiento de redes.	51
4.2	Gráfica log-log de la distribución de conectividades $P(k)$ dada en la Ecuación 4.48. Tomada de [3].	57
4.3	(A) Gráfica log-log de la distribución de conectividades $P(k)$ para una red WWW consistente en 260000 páginas. (B) Conectividad k de los nodos de la red como función de su edad. Tomadas de [1].	60
4.4	Probabilidad p de pertenecer a la isla gigante como función de la conectividad promedio z de la red, pare el caso de la topología de Poisson. Tomada de [3].	63
5.1	Ejemplo 1 (DE): Grado medio.	71
5.2	Ejemplo 1 (DE): Grado medio con pesos.	71
5.3	Ejemplo 1 (DE): Diámetro.	72
5.4	Ejemplo 1 (DE): HITS.	72
5.5	Ejemplo 1 (DE): PageRank.	73
5.6	Ejemplo 1 (DE): Modularidad.	73
5.7	Ejemplo 1 (DE): Inferencia estadística.	73
5.8	Ejemplo 1 (DE): Centralidad de vector propio.	74
5.9	Ejemplo 1 (ENGB): Grado medio.	75
5.10	Ejemplo 1 (ENGB): Grado medio con pesos.	75
5.11	Ejemplo 1 (ENGB): Diámetro.	76
5.12	Ejemplo 1 (ENGB): HITS.	76
5.13	Ejemplo 1 (ENGB): PageRank.	77
5.14	Ejemplo 1 (ENGB): Modularidad.	77
5.15	Ejemplo 1 (ENGB): Inferencia estadística.	77
5.16	Ejemplo 1 (ENGB): Centralidad de vector propio.	78
5.17	Ejemplo 1 (ES): Grado medio.	79
5.18	Ejemplo 1 (ES): Grado medio con pesos.	79
5.19	Ejemplo 1 (ES): Diámetro.	80
5.20	Ejemplo 1 (ES): HITS.	80
5.21	Ejemplo 1 (ES): PageRank.	81

5.22	Ejemplo 1 (ES): Modularidad.	81
5.23	Ejemplo 1 (ES): Inferencia estadística.	81
5.24	Ejemplo 1 (ES): Centralidad de vector propio.	82
5.25	Ejemplo 1 (FR): Grado medio.	83
5.26	Ejemplo 1 (FR): Grado medio con pesos.	83
5.27	Ejemplo 1 (FR): Diámetro.	84
5.28	Ejemplo 1 (FR): HITS.	84
5.29	Ejemplo 1 (FR): PageRank.	85
5.30	Ejemplo 1 (FR): Modularidad.	85
5.31	Ejemplo 1 (FR): Inferencia estadística.	85
5.32	Ejemplo 1 (FR): Centralidad de vector propio.	86
5.33	Ejemplo 1 (PTBR): Grado medio.	87
5.34	Ejemplo 1 (PTBR): Grado medio con pesos.	87
5.35	Ejemplo 1 (PTBR): Diámetro.	88
5.36	Ejemplo 1 (PTBR): HITS.	88
5.37	Ejemplo 1 (PTBR): PageRank.	89
5.38	Ejemplo 1 (PTBR): Modularidad.	89
5.39	Ejemplo 1 (PTBR): Inferencia estadística.	89
5.40	Ejemplo 1 (PTBR): Centralidad de vector propio.	90
5.41	Ejemplo 1 (RU): Grado medio.	91
5.42	Ejemplo 1 (RU): Grado medio con pesos.	91
5.43	Ejemplo 1 (RU): Diámetro.	92
5.44	Ejemplo 1 (RU): HITS.	92
5.45	Ejemplo 1 (RU): PageRank.	93
5.46	Ejemplo 1 (RU): Modularidad.	93
5.47	Ejemplo 1 (RU): Inferencia estadística.	93
5.48	Ejemplo 1 (RU): Centralidad de vector propio.	94
5.49	Ejemplo 2: Grado medio.	99
5.50	Ejemplo 2: Diámetro sin normalizar centralidades en el rango $[0, 1]$	99
5.51	Ejemplo 2: Diámetro normalizando centralidades en el rango $[0, 1]$	100
5.52	Ejemplo 2: HITS.	101
5.53	Ejemplo 2: PageRank.	102
5.54	Ejemplo 2: Modularidad con aleatoriedad y uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).	103
5.55	Ejemplo 2: Modularidad con aleatoriedad sin uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).	104
5.56	Ejemplo 2: Modularidad sin aleatoriedad con uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).	106
5.57	Ejemplo 2: Modularidad sin aleatoriedad ni uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).	107
5.58	Ejemplo 2: Inferencia estadística (primera ejecución).	108
5.59	Ejemplo 2: Inferencia estadística (segunda ejecución).	108
5.60	Ejemplo 2: Coeficiente medio de <i>clustering</i>	108
5.61	Ejemplo 2: Centralidad de vector propio con resolución 10.	109
5.62	Ejemplo 2: Centralidad de vector propio con resolución 100.	109
5.63	Ejemplo 2: Centralidad de vector propio con resolución 1000.	110

Índice de tablas

2.1	Resumen de las características de los tipos de grafos estudiados.	7
2.2	Lista de adyacencia del grafo de la Figura 2.4.	13
4.1	Estudio de las redes complejas.	44
4.2	Algunas de las redes libres de escala que se encuentran en la naturaleza. Sólo se muestra el exponente de entrada para las redes dirigidas. Los cuadros con líneas “—” indican que no se tiene el dato correspondiente al momento de escribir estas notas.	47
5.1	Comparativa de características de la red entre países.	95

Índice de ecuaciones

2.1	Teorema 1	11
2.2	Demostración del Corolario 1	11
2.3	Aplicación del Teorema 1	11
2.4	Teorema 2	12
2.5	Definición de los elementos de la matriz adyacente	13
2.6	Definición de los elementos de la matriz incidente	14
2.7	Fórmula de Stirling	17
2.8	Definición de la función exponencial real como serie de potencias	18
2.9	Definición de la función exponencial real como límite	18
2.10	Aproximación asintótica de la función $(1 - \frac{x}{n})^{-k}$	18
2.11	Expansión en serie de Taylor de $(1 - \frac{x}{n})^{-k}$ alrededor de $\frac{x}{n} = 0$	18
2.12	Variaciones sin repetición	19
2.13	Variaciones con repetición	19
2.14	Permutaciones sin repetición	19
2.15	Permutaciones con repetición	19
2.16	Combinaciones	19
2.17	Función de probabilidad de una variable aleatoria de Bernoulli	20
2.18	Función de probabilidad de una variable aleatoria binomial	20
2.19	Función de probabilidad de una variable aleatoria de Poisson	21
2.20	Función de probabilidad de $X_n \sim \text{Bi}(n, p_n)$	21
2.21	Límite infinito de la función de probabilidad de $X_n \sim \text{Bi}(n, p_n)$	21
2.22	Aproximación en el infinito del primer término de la Ecuación 2.20	21
2.23	Definición del coeficiente binomial	21
2.24	Definición explícita del coeficiente binomial	22
2.25	Reescritura de la definición explícita del coeficiente binomial	22
2.26	Simplificación del segundo término de la Ecuación 2.25	22
2.27	Reescritura de los términos del producto de la Ecuación 2.26	22
2.28	Sustitución de la Ecuación 2.27 en la Ecuación 2.26	22
2.29	Aproximación en el infinito del producto de la Ecuación 2.28	22
2.30	Aproximación en el infinito del tercer término de la Ecuación 2.20	22
2.31	Sustitución de la Ecuación 2.22 y de la Ecuación 2.30 en la Ecuación 2.20	22
3.1	Grado medio de un grafo	23
3.2	Camino característico del grafo de la Figura 2.4	25
3.3	Número total de aristas posibles en un grafo	25
3.4	Densidad de un grafo	26
3.5	Densidad de un digrafo	26

3.6	Número total posible de triángulos en un subgrafo con un vértice concreto	26
3.7	Coficiente de agrupamiento de un grafo	27
3.8	Inverso de la medida de centralidad de un nodo por cercanía	27
3.9	Medida de centralidad de un nodo por cercanía	27
3.10	Medida de centralidad de un nodo por intermediación	27
3.11	Probabilidad de que las geodésicas entre dos nodos concretos pasen por un tercer nodo determinado	28
3.12	Sistema lineal de ecuaciones cuya solución es la medida de centralidad de un nodo por vector propio	28
3.13	Medida de centralidad del nodo v_1 por cercanía	29
3.14	Medida de centralidad del nodo v'_1 por cercanía	29
3.15	Medida de centralidad del nodo v''_1 por cercanía	29
3.16	Medida de centralidad del nodo v_1 por intermediación	29
3.17	Medida de centralidad del nodo v'_1 por intermediación	30
3.18	Medida de centralidad del nodo v''_1 por intermediación	30
3.19	Sistema lineal de ecuaciones cuya solución es la medida de centralidad de un nodo de G por vector propio	30
3.20	Sistema lineal de ecuaciones cuya solución es la medida de centralidad de un nodo de G' por vector propio	31
3.21	Sistema lineal de ecuaciones cuya solución es la medida de centralidad de un nodo de G'' por vector propio	32
3.22	Definición de los elementos de la matriz de pesos D del algoritmo de Floyd-Warshall	33
3.23	Matriz de pesos D del ejemplo del algoritmo de Floyd-Warshall	33
3.24	Matriz inicial C del ejemplo del algoritmo de Floyd-Warshall	33
3.25	Matriz de distancias D del ejemplo del algoritmo de Floyd-Warshall	33
3.26	Matriz de nodos C del ejemplo del algoritmo de Floyd-Warshall	33
3.27	Modularidad de una partición de un grafo	34
3.28	Índice Surprise de una partición de un grafo	35
3.29	Modularidad de G - Partición 1	36
3.30	Modularidad de G' - Partición 1	36
3.31	Modularidad de G'' - Partición 1	37
3.32	Índice Surprise de G - Partición 4	38
3.33	Índice Surprise de G' - Partición 4	38
4.1	Función de distribución de una red con topología de Poisson	46
4.2	Función de distribución de una red con topología exponencial	46
4.3	Función de distribución de una red con topología libre de escala	46
4.4	Número total de parejas que se pueden formar en un conjunto de N nodos	48
4.5	Probabilidad de que una pareja arbitraria seleccionada al azar esté enlazada en un conjunto de N nodos y M enlaces	48
4.6	Probabilidad de que un nodo concreto esté contenido en k parejas de las M posibles (distribución binomial)	48
4.7	Probabilidad de que un nodo concreto esté contenido en k parejas de las M posibles (distribución de Poisson)	48
4.8	Reescritura del coeficiente binomial $\binom{N-1}{k}$ que aparece en la Ecuación 4.6	49
4.9	Reescritura de la Ecuación 4.6	49
4.10	Reescritura de la Ecuación 4.9	49

4.11	Arreglo de la multiplicación del numerador del primer término de la Ecuación 4.10 y el denominador del tercero	49
4.12	Arreglo de la multiplicación de los términos segundo y cuarto de la Ecuación 4.10	49
4.13	Reescritura de la Ecuación 4.10 a partir de los arreglos de la Ecuación 4.11 y la Ecuación 4.12	50
4.14	Aproximación en el infinito de la probabilidad de que un nodo concreto esté contenido en k parejas	50
4.15	Probabilidad de que un nodo arbitrario tenga k conexiones al tiempo t	51
4.16	Probabilidad de que un nodo arbitrario tenga k conexiones al tiempo t (despeje)	51
4.17	Ecuación maestra que determina la evolución temporal de $P(n, k, t)$	52
4.18	Condición inicial para resolver la Ecuación 4.17	52
4.19	Sumatorio de la Ecuación 4.17	52
4.20	Reescritura de la Ecuación 4.19 con $N(t) = t + 1$	52
4.21	Sumatorio hasta $t + 1$ de la Ecuación 4.20	53
4.22	Reescritura de la Ecuación 4.21	53
4.23	Sustitución de la Ecuación 4.22 en la Ecuación 4.20	53
4.24	Probabilidad de enlace igualitario	53
4.25	Sustitución de la Ecuación 4.24 en la Ecuación 4.23	53
4.26	Ecuación 4.25 en un estado estacionario	53
4.27	Solución de la Ecuación 4.26 de recurrencia	54
4.28	Solución de la Ecuación 4.26 para el caso $k = 1$	54
4.29	Solución de la Ecuación 4.26 para el caso $k > 1$	54
4.30	Solución de la Ecuación 4.26 para el caso $k = 2$	54
4.31	Solución de la Ecuación 4.26 para el caso $k = 3$	54
4.32	Solución de la Ecuación 4.26 $\forall k$	54
4.33	Probabilidad de enlace preferencial	55
4.34	Sumatorio de las conectividades de los nodos de un grafo al tiempo t	55
4.35	Sustitución de la Ecuación 4.34 en la Ecuación 4.33	55
4.36	Sustitución de la Ecuación 4.35 en la Ecuación 4.23	55
4.37	Ecuación 4.36 en un estado estacionario	55
4.38	Despeje de $\delta_{k,1}$ en la Ecuación 4.37	55
4.39	Ecuación 4.38 tomando límites en la Ecuación 4.36	56
4.40	Solución de la Ecuación 4.39 de recurrencia	56
4.41	Solución de la Ecuación 4.39 para el caso $k = 1$ (sin despejar)	56
4.42	Solución de la Ecuación 4.39 para el caso $k = 1$	56
4.43	Solución de la Ecuación 4.39 para el caso $k > 1$	56
4.44	Solución de la Ecuación 4.39 para el caso $k = 2$ (sin despejar)	56
4.45	Solución de la Ecuación 4.39 para el caso $k = 2$	56
4.46	Solución de la Ecuación 4.39 para el caso $k = 3$ (sin despejar)	57
4.47	Solución de la Ecuación 4.39 para el caso $k = 3$	57
4.48	Solución de la Ecuación 4.39 $\forall k$	57
4.49	Probabilidad de enlace no lineal	58
4.50	Cambio de la conectividad de un nodo concreto proporcional a la probabilidad de enlace	58
4.51	Sustitución de la Ecuación 4.35 en la Ecuación 4.50	58
4.52	Solución de la Ecuación 4.51 diferencial ordinaria de primer orden	59

4.53	Separación de variables en la Ecuación 4.51	59
4.54	Integración de la Ecuación 4.53	59
4.55	Resolución de las integrales de la Ecuación 4.53	59
4.56	Exponenciación de la Ecuación 4.55	59
4.57	Solución general de la Ecuación 4.51	59
4.58	Solución de la Ecuación 4.51 para $k_n(n) = 1$	59
4.59	Constante de integración	59
4.60	Solución final de la Ecuación 4.51	60
4.61	Desigualdad que implica que los nodos más viejos tienen, en promedio, conectividades mayores	60
4.62	Probabilidad de enlace preferencial con adaptabilidad	61
4.63	Probabilidad de que un nodo escogido aleatoriamente no pertenezca a la isla gigante	62
4.64	Probabilidad de que un nodo escogido aleatoriamente no pertenezca a la isla gigante (distribución de Poisson)	62
4.65	Probabilidad de que un nodo escogido aleatoriamente pertenezca a la isla gigante (distribución de Poisson)	62
4.66	Probabilidad de enlace (distribución libre de escala)	63
4.67	Probabilidad de que un nodo escogido aleatoriamente no pertenezca a la isla gigante (distribución libre de escala)	63
4.68	Probabilidad de que un nodo escogido aleatoriamente pertenezca a la isla gigante (distribución libre de escala)	63
B.1	Probabilidad de que un sistema se encuentre en un estado concreto en un instante de tiempo determinado	118
B.2	Tasa de cambio de la conectividad de un nodo	118
B.3	Integración de la Ecuación B.3 diferencial	118
B.4	Solución general de la Ecuación B.2	118
B.5	Exponenciación de la Ecuación B.4	119
B.6	Solución de la Ecuación B.2	119
B.7	Aproximación del número de nodos con grado, al menos, k	119
B.8	Evolución de la probabilidad de que un nodo alcance un grado concreto en un tiempo determinado	119
B.9	Ecuación B.8 en un estado estacionario	119
E.1	Modularidad de G - Partición 1	128
E.2	Modularidad de G - Partición 2	129
E.3	Modularidad de G - Partición 3	130
E.4	Modularidad de G - Partición 4	131
E.5	Modularidad de G - Partición 5	132
E.6	Modularidad de G - Partición 6	133
E.7	Modularidad de G - Partición 7	134
E.8	Modularidad de G - Partición 8	135
E.9	Modularidad de G - Partición 9	136
E.10	Modularidad de G - Partición 10	137
E.11	Modularidad de G - Partición 11	138
E.12	Modularidad de G - Partición 12	139
E.13	Modularidad de G - Partición 13	140
E.14	Modularidad de G - Partición 14	141

Capítulo 1

Introducción

El análisis y estudio de grafos emerge como una de las ramas más dinámicas de la matemática aplicada en los últimos años, abarcando un amplio espectro de aplicaciones que van desde la informática teórica hasta la optimización de redes sociales y la Inteligencia Artificial (IA). Un grafo puede considerarse como una estructura compuesta por nodos, que representan entidades individuales, y aristas, que indican la relación o interacción entre estas entidades. Debido a su simplicidad y flexibilidad, los grafos son adoptados como herramientas clave para modelar una vasta variedad de sistemas reales complejos, incluyendo sistemas biológicos, infraestructuras de transporte, redes de comunicación y, de forma destacada, las redes sociales.

El vertiginoso desarrollo de las plataformas sociales y la creciente interconexión entre individuos han dado lugar a nuevas y fascinantes problemáticas, que van desde la identificación de comunidades hasta la detección de patrones en las interacciones. Este proyecto tiene como objetivo analizar, modelar y visualizar estas redes sociales a través de grafos, haciendo uso de herramientas avanzadas como el *software* Gephi. Dicho *software* permite generar representaciones visuales de redes complejas y aplicar algoritmos que descomponen dichas redes en estructuras manejables para su análisis, facilitando así la comprensión de la dinámica interna de las mismas.

El uso de redes sociales no solo revoluciona la forma en que las personas se comunican, sino que también ofrece a los investigadores la oportunidad de explorar cómo emergen los comportamientos colectivos y cómo las estructuras de las redes influyen en la difusión de la información y el capital social. Esto convierte el análisis de redes sociales en una tarea interdisciplinaria que involucra áreas como la teoría de grafos, la probabilidad, la sociología y la ciencia de datos.

Este trabajo tiene como punto de partida los fundamentos teóricos de la teoría de grafos, una rama de la matemática discreta que desarrolla de manera significativa en las últimas décadas. Tras abordar las definiciones matemáticas y conceptuales que subyacen a la teoría de redes, se realiza un análisis práctico sobre redes sociales reales, aplicando herramientas de *software* que permiten representar y estudiar estos sistemas.

En concreto, se examinan las redes sociales Twitch y Facebook, ambas con características estructurales y dinámicas propias que requieren de distintos enfoques analíticos. Para Twitch, se trabaja con un grafo dirigido, en el cual las relaciones entre nodos (usuarios) tienen una dirección clara, mientras que en el caso de Facebook se emplea un grafo no dirigido, donde las relaciones entre usuarios son bidireccionales y simétricas. A través de estos estudios de caso, se busca no solo visualizar y modelar las redes, sino también comprender mejor los patrones de interacción y las posibles aplicaciones en áreas como el marketing digital, la política y el análisis de datos masivos.

En resumen, este trabajo tiene como propósito principal ofrecer una visión clara de cómo la teoría de grafos y las herramientas computacionales actuales pueden emplearse para estudiar y comprender las redes sociales modernas, proporcionando ejemplos prácticos que facilitan su aplicabilidad en diversos contextos.

1.1. Objetivos del trabajo

El presente proyecto se enmarca en la problemática asociada al estudio de redes complejas y la detección de comunidades en estas redes. En esta línea de trabajo se identifican varios objetivos:

- **OBJ-01:** Revisar el estado del arte de los algoritmos de detección de comunidades en redes complejas.
- **OBJ-02:** Clasificar las redes complejas según la probabilidad de que un nodo al azar tenga un determinado número de conexiones o aristas.
- **OBJ-03:** Estudiar cómo evolucionan estas probabilidades a medida que las redes crecen.
- **OBJ-04:** Aplicar este estudio probabilístico al caso de las redes sociales.

Cabe destacar que el alcance y la duración del proyecto están limitados por la carga de trabajo que especifica la asignatura Trabajo Fin de Grado (TFG), que son 12 Sistema Europeo de Transferencia y Acumulación de Créditos o *European Credit Transfer System* (ECTS). Es por esta restricción que los objetivos descritos, aunque ambiciosos, no ahondan más en el tema propuesto.

1.2. Estructura de la memoria

El presente documento está dividido en cinco capítulos que abordan los distintos objetivos del proyecto.

1. **Capítulo 1 (Introducción):** Introduce al lector en el dominio del proyecto. En él se especifican los objetivos del mismo (Sección 1.1) y la estructura de la memoria (Sección 1.2). Es el capítulo actual.
2. **Capítulo 2 (Conceptos previos):** Revisa las definiciones matemáticas generales necesarias para que el lector comprenda los capítulos posteriores. Profundiza en la teoría de grafos (Sección 2.1), aproximaciones asintóticas y series infinitas (Sección 2.2), y conceptos clave de la teoría de la probabilidad (Sección 2.3).
3. **Capítulo 3 (Grafos y redes sociales):** Contextualiza el proyecto en el ámbito de las redes sociales (Sección 3.1) y ahonda en métricas clave como la excentricidad (Sección 3.2), la densidad (Sección 3.3), y la centralidad (Sección 3.4). Además, explica algoritmos esenciales como son el de Floyd-Warshall (Sección 3.5) y algunos de detección de comunidades (Sección 3.6).

4. **Capítulo 4 (Redes complejas)**: Aborda los modelos clásicos de redes complejas (Sección 4.2), la clasificación del estudio (Sección 4.3) y de las distribuciones de vecinos de una red (Sección 4.4). Explora también fenómenos el envejecimiento de nodos (Sección 4.5) y la formación de islas gigantes (Sección 4.6).
5. **Capítulo 5 (Ejemplos de casos reales)**: Muestra dos ejemplos reales de redes sociales estudiadas con la herramienta *software* Gephi: Twitch como grafo dirigido (Sección 5.1) y Facebook como grafo no dirigido (Sección 5.2).

También se incluyen cinco apéndices o anexos, que amplían y complementan lo explicado en los capítulos.

1. **Anexo A (Acrónimos)**: Presenta una lista de los acrónimos utilizados a lo largo del trabajo.
2. **Anexo B (Métodos de crecimiento de redes)**: Explica los métodos empleados para modelar el crecimiento de las redes.
3. **Anexo C (Transiciones de fase en redes complejas)**: Aborda las transiciones de fase en redes complejas con topología de Poisson, tanto de primer como de segundo orden.
4. **Anexo D (Código)**: Contiene el código fuente utilizado para realizar los análisis de redes complejas, en concreto, el Ejemplo 37 del Capítulo 3.
5. **Anexo E (Cálculos)**: Detalla los cálculos realizados en los ejemplos de redes, en concreto, el Ejemplo 42 del Capítulo 3.

Se advierte que, durante toda la memoria, los nombres de archivos se muestran con una fuente monoespaciada, mientras que los extranjerismos y términos técnicos se escriben en cursiva.

Capítulo 2

Conceptos previos

El estudio de redes complejas requiere de una sólida comprensión de varios conceptos matemáticos fundamentales. En este capítulo se introducen las nociones básicas de la teoría de grafos, incluyendo la clasificación de diferentes tipos de grafos y la terminología asociada a ellos [13] [5]. Además, se exploran aproximaciones asintóticas y las bases de la teoría de la probabilidad, las cuales juegan un papel crucial en la modelización y el análisis de redes complejas. Estos conceptos previos sientan las bases para comprender los algoritmos y métodos que se aplican en capítulos posteriores.

2.1. Teoría de grafos

Los grafos son estructuras discretas compuestas por vértices, también llamados nodos y representados por puntos, y aristas, también llamadas arcos y representadas por segmentos de línea o flechas. Cada arista tiene asociado uno o dos vértices, llamados extremos o puntos finales de la arista. Con posibles características opcionales (nombres, etiquetas, pesos, colores, etc.), los grafos sirven para representar muchas situaciones prácticas, en muy diversos contextos, donde hay un número finito de elementos y relaciones entre ellos (Definición 7).

En la Figura 2.1 se puede observar el ejemplo de un grafo con 4 vértices etiquetados: $\{v_1, v_2, v_3, v_4\}$; 12 aristas: (v_1) , (v_1, v_2) , (v_1, v_4) , (v_2, v_1) , (v_2, v_3) , (v_2, v_4) , (v_3, v_2) , (v_3, v_4) , (v_4, v_1) , (v_4, v_3) , (v_4) ; y sus correspondientes pesos.

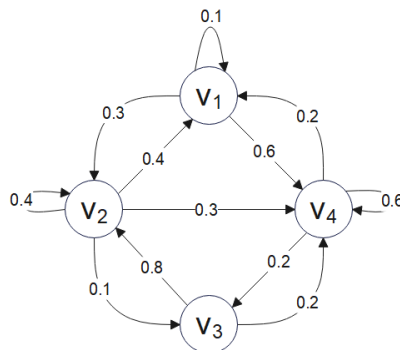


Figura 2.1: Ejemplo de grafo.

Nota 1. Las aristas se representan entre paréntesis por ser dirigidas; de lo contrario, se usan llaves.

2.1.1. Tipos

Existen seis tipos de grafos diferentes según sus aristas.

Definición 1. Un **grafo simple** es un par $G = (V, E)$, donde $V \neq \emptyset$ es un conjunto no vacío de vértices y E es un conjunto de aristas o pares no ordenados de vértices distintos de V . Así pues, las aristas no están orientadas, es decir, no existen bucles o lazos ¹ ni aristas paralelas o múltiples ².

En la Figura 2.2 se puede observar el ejemplo de un grafo simple con 4 vértices etiquetados: $\{v_1, v_2, v_3, v_4\}$; y 5 aristas: $\{v_1, v_2\}$, $\{v_1, v_4\}$, $\{v_2, v_3\}$, $\{v_2, v_4\}$, $\{v_3, v_4\}$.

Definición 2. Un **multigrafo** es un grafo con un multi-conjunto de aristas. En este caso particular de grafo puede haber aristas paralelas.

En la Figura 2.2 se puede observar el ejemplo de un multigrafo con 4 vértices etiquetados: $\{v_1, v_2, v_3, v_4\}$; y 8 aristas: 3 aristas paralelas $\{v_1, v_2\}$, $\{v_1, v_4\}$, $\{v_2, v_3\}$, $\{v_2, v_4\}$, 2 aristas paralelas $\{v_3, v_4\}$.

Definición 3. Un **pseudografo** es un multigrafo con bucles o lazos. En este caso particular de grafo puede haber varios bucles sobre un mismo vértice.

En la Figura 2.2 se puede observar el ejemplo de un pseudografo con 4 vértices etiquetados: $\{v_1, v_2, v_3, v_4\}$; y 10 aristas: $\{v_1\}$, 3 aristas paralelas $\{v_1, v_2\}$, $\{v_1, v_4\}$, $\{v_2, v_3\}$, $\{v_2, v_4\}$, $\{v_3\}$, 2 aristas paralelas $\{v_3, v_4\}$.

Definición 4. Un **digrafo** (o grafo dirigido) es un grafo donde las aristas son representadas mediante flechas, ya que se trata de un conjunto de pares ordenados de vértices (distintos o no). En este caso particular de grafo puede haber dos aristas paralelas para cada par de vértices (un arco en cada sentido), así como bucles.

En la Figura 2.2 se puede observar el ejemplo de un digrafo con 4 vértices etiquetados: $\{v_1, v_2, v_3, v_4\}$; y 9 aristas: (v_1, v_2) , (v_1, v_4) , (v_2, v_1) , (v_2) , (v_2, v_4) , (v_3, v_2) , (v_3, v_4) , (v_4, v_3) , (v_4) .

Definición 5. Un **multidigrafo** (o multigrafo dirigido) es un digrafo donde se permiten aristas múltiples, es decir, varias aristas paralelas en el mismo sentido. También se puede definir como un multigrafo donde todas sus aristas son dirigidas.

En la Figura 2.2 se puede observar el ejemplo de un multidigrafo con 4 vértices etiquetados: $\{v_1, v_2, v_3, v_4\}$; y 11 aristas: 3 aristas paralelas (v_1, v_2) , (v_2) , (v_2, v_3) , (v_2, v_4) , 2 aristas paralelas (v_3, v_4) , (v_4, v_1) , (v_4, v_3) , (v_4) .

Definición 6. Un **grafo mixto** es un grafo con aristas dirigidas y no dirigidas.

En la Figura 2.2 se puede observar el ejemplo de un grafo mixto con 4 vértices etiquetados: $\{v_1, v_2, v_3, v_4\}$; y 11 aristas: 3 aristas paralelas (v_1, v_2) , $\{v_1, v_4\}$, (v_2) , $\{v_2, v_3\}$, (v_2, v_4) , 2 aristas paralelas (v_3, v_4) , (v_4, v_3) , (v_4) .

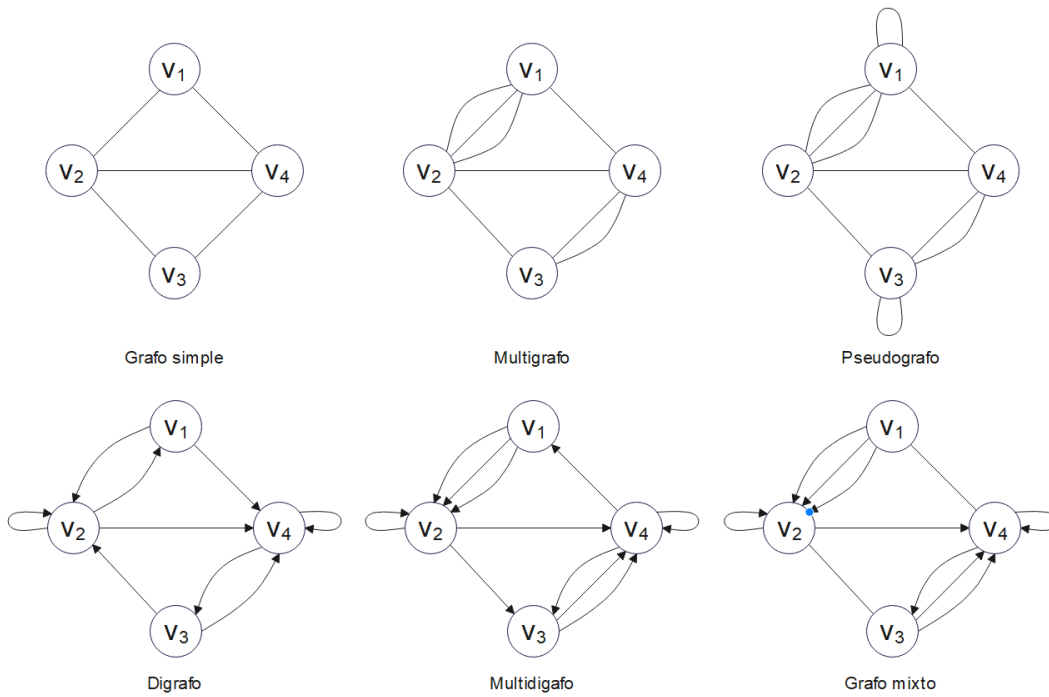


Figura 2.2: Tipos de grafos.

Tipo	Aristas orientadas	Aristas múltiples	Bucles
Grafo simple	No	No	No
Multigrafo	No	Sí	No
Pseudografo	No	Sí	Sí
Digrafo	Sí	No	Sí
Multidigrafo	Sí	Sí	Sí
Grafo mixto	Sí y No	Sí	Sí

Tabla 2.1: Resumen de las características de los tipos de grafos estudiados.

En la Tabla 2.1 se puede observar el resumen de las características de los tipos de grafos estudiados.

Definición 7. Un *grafo infinito* es un grafo con un conjunto de vértices o un número de aristas infinito. En este trabajo, se utilizan únicamente los grafos finitos.

Tipos especiales

Se definen ahora nueve grafos simples especiales.

¹Arista que une un vértice consigo mismo, es decir, una arista del tipo v donde v es un vértice. En un bucle no importa el sentido de recorrido.

²Aristas que unen el mismo par de vértices.

Definición 8. Una *cadena* L_n consta de $n \geq 1$ nodos: $\{v_1, \dots, v_n\}$; y $n - 1$ aristas: $\{v_i, v_{i+1}\}$ para $i = 1, \dots, n - 1$.

En la Figura 2.3 se puede observar el ejemplo de una cadena L_n .

Definición 9. Un *ciclo* C_n consta de $n \geq 3$: $\{v_1, \dots, v_n\}$; y n aristas: $\{v_i, v_{i+1}\}$ para $i = 1, \dots, n - 1$, junto con la arista $\{v_1, v_n\}$.

En la Figura 2.3 se puede observar el ejemplo de un ciclo C_n .

Definición 10. Una *estrella* S_n consta de $n + 1 \geq 4$ nodos: $\{v_0, v_1, \dots, v_n\}$; y n aristas: $\{v_0, v_i\}$ para $i = 1, \dots, n$.

En la Figura 2.3 se puede observar el ejemplo de una estrella S_n .

Definición 11. Una *rueda* W_n consta de $n + 1 \geq 4$ nodos: $\{v_0, v_1, \dots, v_n\}$; y $2n$ aristas: $\{v_0, v_i\}$ para $i = 1, \dots, n$, $\{v_i, v_{i+1}\}$ para $i = 1, \dots, n - 1$, y $\{v_1, v_n\}$.

En la Figura 2.3 se puede observar el ejemplo de una rueda W_n .

Definición 12. Un *completo* o *clique* K_n consta de $n \geq 1$ nodos: $\{v_1, \dots, v_n\}$; y todas las $\binom{n}{2}$ aristas posibles.

En la Figura 2.3 se puede observar el ejemplo de un completo K_6 .

Definición 13. Una *malla* G_{n_1, n_2} consta de $n_1 \times n_2$ nodos dispuestos en n_1 filas y n_2 columnas, y todas las posibles aristas horizontales y verticales.

En la Figura 2.3 se puede observar el ejemplo de una malla $G_{2,4}$.

Definición 14. Un *hipercubo* Q_n (de dimensión n , para $n \geq 0$) se define de manera inductiva como sigue:

- Q_0 es un punto (sin aristas).
- Una vez definido Q_n , se hace una copia exacta Q'_n y se unen, uno a uno, los nodos equivalentes de ambos grafos.

En la Figura 2.3 se puede observar el ejemplo de un hipercubo Q_3 .

Definición 15. Un *grafo bipartito* B_{n_1, n_2} es aquel en el que $V = V_1 \cup V_2$ con $V_1 \cap V_2 = \emptyset$ (donde $|V_1| = n_1$, $|V_2| = n_2$ y $|V| = n_1 + n_2$), de manera que todas sus aristas conectan un nodo de V_1 con otro de V_2 .

En la Figura 2.3 se puede observar el ejemplo de un bipartito $B_{3,2}$.

Nota 2. Los grafos bipartitos se usan en problemas de asignación y de emparejamiento, entre otros.

Definición 16. Un *grafo bipartito completo* K_{n_1, n_2} es un grafo bipartito en el que se toman todas las aristas posibles, es decir, $|E| = n_1 \cdot n_2$.

En la Figura 2.3 se puede observar el ejemplo de un bipartito completo $K_{3,2}$.

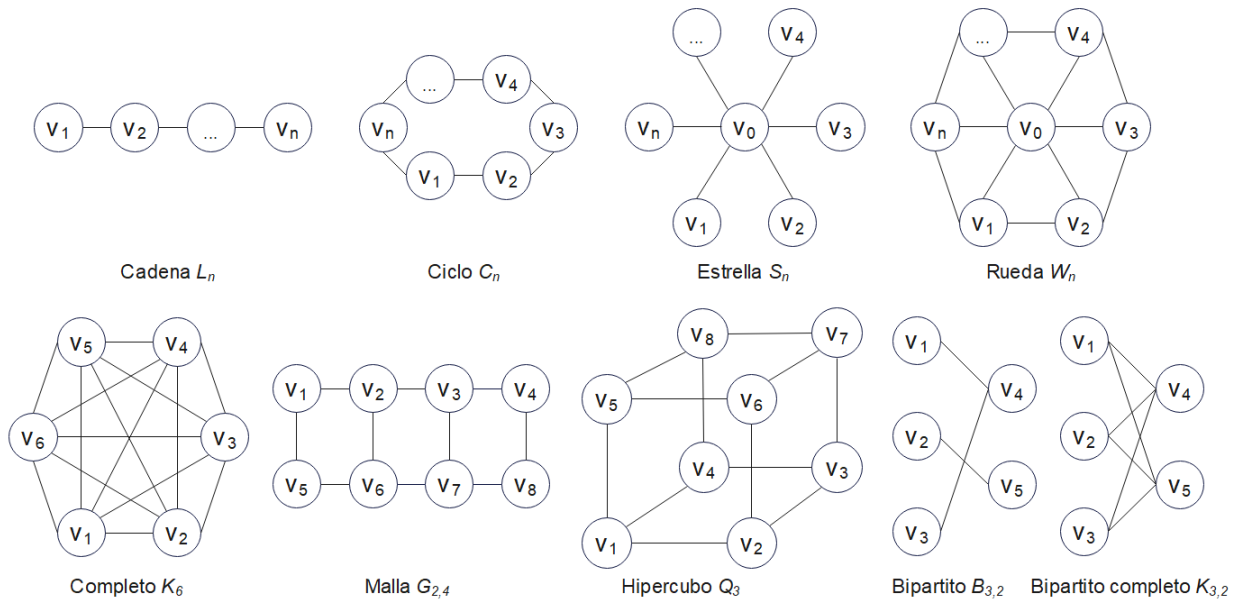


Figura 2.3: Tipos especiales de grafos.

2.1.2. Terminología

En esta subsección, se definen una serie de términos fundamentales a la hora de trabajar con un grafo $G = (V, E)$.

Definición 17. El *orden* de G es el número de vértices del mismo, es decir, $|V|$. Suele denotarse por la letra n .

Ejemplo 1. En la Figura 2.4 se representa el cubo 3 Dimensiones o 3 Dimensions (3D) en forma de grafo. El orden del grafo de la Figura 2.4 es 8 y coincide con el número de vértices del

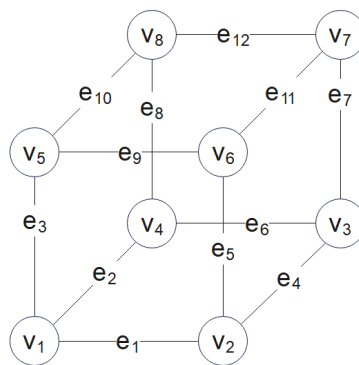


Figura 2.4: Cubo 3D en forma de grafo.

cubo.

Ejemplo 2. Se sabe que la red social Facebook es un grafo no dirigido, donde los nodos son los perfiles de usuario, y las aristas, las relaciones de amistad entre ellos. El orden de este grafo es

el número de usuarios de la red. En julio de 2024, el número de usuarios activos mensuales de Facebook a nivel mundial es de 3070 millones [20].

Ejemplo 3. Se sabe que la red social Instagram es un grafo dirigido, donde los nodos son los perfiles de usuario, y las aristas, las relaciones de seguimiento entre ellos. El orden de este grafo es el número de usuarios de la red. En julio de 2024, el número de usuarios activos mensuales de Instagram a nivel mundial es de 2000 millones [20].

Definición 18. El **tamaño** de G es el número de aristas del mismo, es decir, $|E|$. Suele denotarse por la letra m .

Ejemplo 4. El tamaño del grafo de la Figura 2.4 es 12 y coincide con el número de aristas del cubo.

Ejemplo 5. El tamaño de Facebook es el número de relaciones de amistad de la red.

Ejemplo 6. El tamaño de Instagram es el número de relaciones de seguimiento de la red.

Existen varias definiciones remarcables alrededor de los grafos no dirigidos.

Definición 19. Dos **vértices** $u, v \in V$ son **adyacentes** (o **vecinos**) si existe una arista $e \in E$ que los une, es decir, $e = \{u, v\} \in E$. En tal caso, se dice que “ e es incidente con u (y con v)”, que “ e conecta u y v ”, o que “ u y v son los extremos de e ”.

Ejemplo 7. Los vértices v_1 y v_2 del grafo de la Figura 2.4 son adyacentes.

Ejemplo 8. Dos usuarios de Facebook son adyacentes o vecinos si son amigos en dicha red.

Definición 20. Dos **vértices** son **independientes** si no existe una arista que los une.

Ejemplo 9. Los vértices v_1 y v_6 del grafo de la Figura 2.4 son independientes.

Ejemplo 10. Dos usuarios de Facebook son independientes si no son amigos en dicha red.

Definición 21. Dos **aristas** son **independientes** si no tienen vértices en común.

Ejemplo 11. Las aristas e_1 y e_6 del grafo de la Figura 2.4 son independientes.

Ejemplo 12. Dos relaciones de amistad de Facebook son independientes si no tienen usuarios en común.

Definición 22. El **vecindario** de $u \in V$, $N(u)$, es el conjunto de todos los vértices de G adyacentes a u .

Si A es un subconjunto de V ($A \subseteq V$), se denota por $N(A)$ al conjunto de todos los vértices de G que son adyacentes a, al menos, un vértice de A , entonces $N(A) = \cup_{v \in A} N(v)$.

Ejemplo 13. El vecindario del vértice v_1 del grafo de la Figura 2.4 son los vértices v_2 , v_4 y v_5 .

Ejemplo 14. El vecindario de un usuario de Facebook es el conjunto de todos sus amigos en dicha red.

Definición 23. El **grado** de $u \in V$, $\delta(u)$, es el número de aristas de G que son incidentes con u (excepto los bucles, que se cuentan dos veces).

Ejemplo 15. El grado del vértice v_1 del grafo de la Figura 2.4 es 3. Coincide con el tamaño de su vecindario. De hecho, todos los vértices del cubo tienen el mismo grado.

Ejemplo 16. El grado de un usuario de Facebook es el número de amigos que tiene en dicha red. Coincide con el tamaño de su vecindario.

Cada arista contribuye en dos unidades a la suma de los grados de los vértices, porque una arista es incidente con exactamente dos vértices, pudiendo tratarse del mismo en el caso de los bucles. Así pues, cuando se suman los grados de todos los vértices de un grafo, se obtiene el Teorema 1, denominado el teorema del apretón de manos (*handshaking*).

Teorema 1. Sea $G = (V, E)$ un grafo no dirigido con $m = |E|$ aristas, entonces se cumple la Ecuación 2.1.

$$2 \cdot m = \sum_{v \in V} \delta(v) \quad (2.1)$$

Nota 3. La fórmula del Teorema 1 (Ecuación 2.1) es válida incluso con bucles y aristas paralelas.

El Teorema 1 muestra que la suma de los grados de los vértices de un grafo no dirigido es par. Este hecho tiene muchas consecuencias, una de las cuales se expresa en el Corolario 1.

Corolario 1. Siempre hay un número par de vértices de grado impar.

Demostración. Sea $G = (V, E)$ un grafo no dirigido con $m = |E|$ aristas. Sean $V_1, V_2 \subseteq V$ el conjunto de vértices de grado par y el conjunto de vértices de grado impar de G , respectivamente. Entonces, se cumple la Ecuación 2.2.

$$2 \cdot m = \sum_{v \in V} \delta(v) = \sum_{v \in V_1} \delta(v) + \sum_{v \in V_2} \delta(v) \quad (2.2)$$

La primera igualdad de la Ecuación 2.2 es el Teorema 1. La segunda igualdad de la Ecuación 2.2 procede de las dos siguientes: $V_1 \cup V_2 = V$ y $V_1 \cap V_2 = \emptyset$.

Dado que $\delta(v)$ es par $\forall v \in V_1$, el primer término en el lado derecho de la última igualdad, $\sum_{v \in V_1} \delta(v)$, es par. Además, la suma de los dos términos en el lado derecho de la última igualdad es par, ya que es $2 \cdot m$. Por eso, el segundo término en el lado derecho de la última igualdad, $\sum_{v \in V_2} \delta(v)$, es también par. Puesto que todos los términos de $\sum_{v \in V_2} \delta(v)$ son impares, debe haber un número par de tales términos. Así, $|V_2|$ es par, es decir, hay un número par de vértices de grado impar. \square

Nota 4. Una aplicación del Teorema 1 es poder contar el número de aristas sumando los grados de los vértices y dividiendo el resultado por 2, como se observa en la Ecuación 2.4.

$$m = \frac{\sum_{v \in V} \delta(v)}{2} \quad (2.3)$$

La terminología de los grafos dirigidos refleja el hecho de que sus aristas tienen dirección.

Definición 24. Si $(u, v) \in E$ (es decir, hay una flecha desde u hasta v), se dice que u es *adyacente a* v , y que v es *adyacente desde* u .

Ejemplo 17. Si un usuario de Instagram sigue a otro se dice que el primero es adyacente al segundo y que el segundo es adyacente desde el primero, es decir, un seguidor es adyacente al seguido y un seguido es adyacente desde su seguidor.

En este caso, hay que distinguir entre el grado entrante y el grado saliente de un vértice.

Definición 25. El **grado entrante** de $u \in V$, $\delta^-(u)$ es el número de aristas que tienen a u como vértice final o extremo.

Definición 26. El **grado saliente** de $u \in V$, $\delta^+(u)$ es el número de aristas que tienen a u como vértice inicial u origen.

Ejemplo 18. El grado entrante de un usuario de Instagram es el número de seguidores que tiene en dicha red. El grado saliente, por su parte, es el número de usuarios seguidos por dicho usuario.

Nota 5. Un bucle contribuye en una unidad en ambos casos.

Para el caso de los digrafos, existe un teorema análogo al Teorema 1. Como cada arista tiene un vértice inicial y un vértice terminal, la suma de los grados de entrada y de los grados de salida de todos los vértices es la misma, y coincide con el número de aristas del grafo. Esta es la explicación del Teorema 2.

Teorema 2. Sea $G = (V, E)$ un grafo dirigido con $m = |E|$ aristas, entonces se cumple la Ecuación 2.4.

$$\sum_{v \in V} \delta^-(v) = \sum_{v \in V} \delta^+(v) = m \quad (2.4)$$

2.1.3. Representación

Sea $G = (V, E)$ un grafo simple donde $V = \{v_1, v_2, \dots, v_n\}$, $|V| = n$, $E = \{e_1, e_2, \dots, e_m\}$, $|E| = m$.

Lista de adyacencia

Definición 27. La **lista de adyacencia** de G consiste en enumerar, para cada vértice $v \in V$, sus nodos adyacentes.

Nota 6. En las listas de adyacencia de los multigrafos se pueden repetir nodos. Por su parte, en el caso de los digrafos, las listas contienen los extremos de las aristas salientes en cada vértice, es decir, para cada vértice se anota el extremo final de las aristas que lo tienen como extremo origen.

Ejemplo 19. En la Tabla 2.2 se puede observar la lista de adyacencia del grafo de la Figura 2.4.

Matriz de adyacencia

Definición 28. La **matriz adyacente** A de G , A_G , es la matriz estándar de dimensión $n \times n$ con (i, j) -ésima entrada unitaria cuando v_i y v_j son adyacentes, y nula en caso contrario, es

Nodos	Adyacentes
v_1	v_2, v_4, v_5
v_2	v_1, v_3, v_6
v_3	v_2, v_4, v_7
v_4	v_1, v_3, v_8
v_5	v_1, v_6, v_8
v_6	v_2, v_5, v_7
v_7	v_3, v_6, v_8
v_8	v_4, v_5, v_7

Tabla 2.2: Lista de adyacencia del grafo de la Figura 2.4.

decir, cuando son independientes. En otras palabras, si esta matriz adyacente es $A_G = [a_{ij}]$, entonces a_{ij} es una función definida a trozos como indica la Ecuación 2.5.

$$a_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \text{ es una arista de } G, \text{ es decir, } (v_i, v_j) \in E \\ 0 & \text{en otro caso, es decir, } (v_i, v_j) \notin E \end{cases} \quad (2.5)$$

Nota 7. En las matrices de adyacencia de los multigrafos, se sustituye la unidad por el número de aristas incidentes a los nodos correspondientes.

Nota 8. Las matrices de adyacencia de los digrafos, en general, no son simétricas.

Ejemplo 20. A continuación, se puede observar la matriz de adyacencia del grafo de la Figura 2.4.

$$\begin{pmatrix} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 \\ v_1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ v_2 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ v_3 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ v_4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ v_5 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ v_6 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ v_7 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ v_8 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Matriz de incidencia

Definición 29. La **matriz incidente** B de G , B_G , es la matriz estándar de dimensión $n \times m$ con (i, j) -ésima entrada unitaria cuando la arista e_j es incidente con el nodo v_i y nula cuando no se cumple tal condición. En otras palabras, si esta matriz incidente es $B_G = [b_{ij}]$, entonces b_{ij} es una función definida a trozos como indica la Ecuación 2.6.

$$b_{ij} = \begin{cases} 1 & \text{cuando la arista } e_j \text{ es incidente con } v_i \\ 0 & \text{en otro caso} \end{cases} \quad (2.6)$$

Nota 9. En las matrices de incidencia de los multigrafos, pueden existir columnas repetidas, que se corresponden con aristas paralelas; y columnas con un solo 1, que se corresponden con bucles.

Nota 10. En las matrices de incidencia de los digrafos, se anota 1 en el extremo origen y -1 en el extremo final.

Ejemplo 21. A continuación, se puede observar la matriz de incidencia del grafo de la Figura 2.4.

$$\begin{pmatrix} & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \\ v_1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_2 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_3 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ v_4 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ v_5 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ v_6 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ v_7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ v_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

2.1.4. Operaciones

Sea $G = (V, E)$ un grafo simple.

Definición 30. Se dice que el grafo $G' = (V', E')$ es un **subgrafo** de G si $V' \subseteq V$, $E' \subseteq E$ y las aristas de E' solo son incidentes con vértices de V' .

Definición 31. Dados los grafos $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$, se define el **grafo unión** de G_1 y G_2 por $G = (V, E)$, donde $V = V_1 \cup V_2$ y $E = E_1 \cup E_2$.

Definición 32. Una **subdivisión elemental** $G' = (V', E')$ de G se obtiene añadiendo un nuevo vértice $w \notin V$, eliminando una arista $\{u, v\} \in E$ y añadiendo dos nuevas aristas $\{u, w\}$ y $\{w, v\}$.

Definición 33. Dos grafos $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$ son **isomorfos** si existe un isomorfismo entre V_1 y V_2 , es decir, si existe una biyección (permutación de nodos) $\phi : V_1 \rightarrow V_2$ tal que $u, v \in V_1$ son adyacentes ($\{u, v\} \in E_1$) si y solo si $\phi(u), \phi(v) \in V_2$ son adyacentes ($\{\phi(u), \phi(v)\} \in E_2$).

Nota 11. Si G_1 y G_2 son isomorfos, entonces tienen el mismo número de nodos y de aristas, los mismos grados, los mismos ciclos, etc.

Definición 34. Dos grafos $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$ son **homeomorfos** si existen sucesiones de subdivisiones elementales que dan lugar a dos grafos $G'_1 = (V'_1, E'_1)$ y $G'_2 = (V'_2, E'_2)$ isomorfos.

Nota 12. En grafos homeomorfos se pueden contraer cadenas de aristas en una sola arista y viceversa, así como permutar vértices.

2.1.5. Conectividad

Sea $G = (V, E)$ un grafo simple.

Caminos

Definición 35. Sean $u, w \in V$ y n un entero no negativo, un **camino** de longitud n de u (origen) a w (destino) es una sucesión $P = \{v_0, e_1, v_1, e_2, v_2, \dots, v_{n-1}, e_n, v_n\}$, donde $v_i \in V \forall i \in [0, n]$, $e_j \in E \forall j \in [1, n]$, $v_0 = u$, $v_n = w$, y cada arista $e_k = \{v_{k-1}, v_k\}$ tiene extremos v_{k-1} y $v_k \forall k \in [1, n]$.

Nota 13. Si G es un grafo simple, basta dar la sucesión de vértices por los que pasa, ya que las aristas a recorrer están unívocamente determinadas. Por tanto, a pesar de que en esta subsección se trata con digrafos, se puede generalizar fácilmente a grafos simples.

Definición 36. La **longitud** de un camino es simplemente el número de aristas del mismo.

Sea P el camino de la Definición 35.

Definición 37. P es **cerrado (circuito)** si $v_0 = v_n$ y $n \geq 3$.

Definición 38. P es **simple** si no repite ninguna arista, es decir, si $e_i \neq e_j \forall i, j \in [1, n]$, $i \neq j$.

Definición 39. P es **elemental** si no repite ningún vértice, es decir, $v_i \neq v_j \forall i, j \in [0, n]$, $i \neq j$.

Definición 40. Un **ciclo** es un camino elemental cerrado, es decir, un camino en el que solo se repiten los extremos.

Proposición 1. Todo camino elemental es simple, pero el recíproco no es cierto.

Demostración.

\Rightarrow Todo camino elemental es simple.

Se supone que P es un camino elemental. Por definición, esto significa que P no repite ningún vértice. Para que P repitiera una arista, tendría que pasar por los mismos dos vértices dos veces. Pero si P no repite ningún vértice, entonces no puede pasar dos veces por la misma arista, porque hacerlo implicaría pasar dos veces por los vértices que definen esa arista. Por lo tanto, si P no repite ningún vértice, tampoco puede repetir ninguna arista. Entonces, queda demostrado que todo camino elemental es simple.

\Leftarrow Un camino simple no necesariamente es elemental.

Para demostrar que no todo camino simple es elemental, se proporciona un contraejemplo.

Se considera el camino $P = \{v_0, e_1, v_1, e_2, v_2, e_3, v_3, e_4, v_4, e_5, v_2\}$. P es un camino que recorre las aristas $e_1 = \{v_0, v_1\}$, $e_2 = \{v_1, v_2\}$, $e_3 = \{v_2, v_3\}$, $e_4 = \{v_3, v_4\}$ y $e_5 = \{v_2, v_4\}$. Este camino es simple porque no repite ninguna arista. Sin embargo, P no es elemental porque repite el vértice v_2 . Por lo tanto, existe, al menos, un camino que es simple y no es elemental.

□

Nota 14. En el caso de un digrafo, en la Definición 35, las aristas deben ser orientadas, es decir, $e_k = (x_{k-1}, x_k)$. En otras palabras, no se puede recorrer una arista en sentido contrario a su orientación fijada.

Nota 15. En el caso de un digrafo, la Proposición 1 se cumple en ambos sentidos, es decir, un camino es elemental si y solo si es simple.

Ejemplo 22. El modelo gráfico no orientado para calcular el camino más corto entre dos ciudades es un grafo sobre un mapa de carreteras dado (con longitudes de aristas asignadas, pesos según el consumo o el tiempo estimado, etc.) en el que las ciudades son nodos y, las carreteras, aristas.

Ejemplo 23. El modelo gráfico orientado para hallar el camino más corto entre dos puntos de una ciudad viajando en coche es un grafo sobre el plano de la ciudad en el que las aristas son calles con su sentido de circulación y, los nodos, sus intersecciones.

Conexión

Definición 41. G es **conexo** si entre dos vértices $u, v \in V$ cualesquiera de G existe siempre, al menos, un camino de u a v .

Proposición 2. Si G es conexo, siempre existe un camino elemental que une dos vértices cualesquiera, pues siempre que se repite un nodo $v_i = v_j$ se pueden suprimir todas las aristas intermedias e_{i+1}, \dots, e_j .

Demostración. Dado que G es conexo, por definición, existe, al menos, un camino entre cualquier par de vértices arbitrarios $u, w \in V$. Sea $P = \{v_0, e_1, v_1, e_2, v_2, \dots, v_{n-1}, e_n, v_n\}$ un camino cualquiera que conecta u y w (donde $v_0 = u$ y $v_n = w$). Si P es elemental, entonces ya se ha encontrado el camino elemental deseado entre u y w . Si P no es elemental, al menos un vértice se repite en el camino P . Se supone que $v_i = v_j$ para $i < j$. En este caso, se puede “cortar” el ciclo formado por los vértices entre v_i y v_j para obtener un nuevo camino que sigue siendo válido y que es más corto. Se forma un nuevo camino P' eliminando los vértices y aristas desde v_{i+1} hasta v_j . El nuevo camino es $P' = \{v_0, e_1, v_1, \dots, v_{i-1}, e_i, v_i, e'_{j+1}, v_{j+1}, \dots, v_{n-1}, e_n, v_n\}$. Este proceso de eliminación de ciclos puede repetirse cada vez que haya un vértice repetido. Al continuar eliminando ciclos de la manera descrita anteriormente, finalmente se obtiene un camino que no tiene vértices repetidos, es decir, un camino elemental. Como cada eliminación de un ciclo reduce la longitud del camino, y dado que el número de vértices en G es finito, este proceso necesariamente termina. El resultado final es un camino elemental que conecta u y w . \square

Definición 42. Un subgrafo G' de G es **maximal** con respecto a una propiedad P si:

- G' cumple la propiedad P .
- No existe un subgrafo $G'' = (V'', E'')$ de G tal que G' es un subgrafo propio de G'' ($V' \subsetneq V''$ y $E' \subsetneq E''$) y G'' también cumple la propiedad P .

En otras palabras, G' es maximal con respecto a P si cualquier intento de agregar un vértice o una arista de G a G' resulta en un subgrafo de G que no cumple P .

Definición 43. Un subgrafo conexo maximal de G es una **componente conexa** de G . Esto significa que es un subgrafo conexo al que no se le puede agregar ningún otro vértice o arista de G sin que deje de ser conexo. Esto es, un subgrafo conexo maximal de G es un subgrafo de G que es maximal con respecto a la propiedad de conexión.

Proposición 3. Si G no es conexo, se puede dividir en componentes conexas, que son subgrafos conexos maximales.

Demostración. Dado que G no es conexo, entonces existe, al menos, un par de vértices $u, w \in V$ tal que no hay camino entre ellos. Para construir una componente conexa, se escoge un vértice $v \in V$ cualquiera y se considera el conjunto de vértices $N(v)$, que incluye todos los vértices que son alcanzables desde v mediante algún camino. Se construye el subgrafo inducido por $N(v)$, G_v . Por construcción, G_v es conexo porque todos los vértices en $N(v)$ son alcanzables entre sí. Para demostrar la maximalidad de la componente conexa G_v , se supone que no es maximal. Esto implica que existe otro vértice de G , $z \notin N(v)$, tal que $G_v \cup \{z\}$ es conexo. Sin embargo, por la definición de $N(v)$, z no es alcanzable desde ningún vértice en $N(v)$, lo que contradice la hipótesis de que $G_v \cup \{z\}$ es conexo. Por lo tanto, G_v es un subgrafo conexo maximal. La partición en componentes conexas se consigue repitiendo el proceso anterior para todos los vértices que aún no son incluidos en alguna componente conexa. Cada vez que se selecciona un nuevo vértice no incluido en las componentes previas, se construye una nueva componente conexa. Este proceso finalmente cubre todos los vértices del grafo G , porque cada vértice pertenece a exactamente una de estas componentes conexas. Así, queda demostrado que el grafo G puede ser particionado en subgrafos conexas maximales, los cuales son llamados componentes conexas. \square

En el caso de aristas dirigidas, se suele distinguir entre digrafos fuertemente conexos y débilmente conexos.

Definición 44. *Cuando se puede ir de cualquier nodo u a cualquier nodo v respetando las orientaciones de las aristas se trata de un digrafo **fuertemente conexo**.*

Definición 45. *Cuando no siempre se pueden respetar dichas orientaciones, pero el grafo subyacente eliminando orientaciones es conexo se trata de un digrafo **débilmente conexo**.*

Ejemplo 24. *El digrafo de un plano de una ciudad, considerando los sentidos de circulación de las calles, es un ejemplo de grafo fuertemente conexo.*

2.2. Aproximaciones asintóticas y series infinitas

En matemáticas, las aproximaciones asintóticas y las series infinitas son herramientas fundamentales para describir el comportamiento de funciones y secuencias en límites específicos, como cuando una variable tiende a infinito. Estas técnicas permiten obtener expresiones simplificadas y manejar casos donde las fórmulas exactas son difíciles de calcular o no están disponibles. Esta aproximación es útil en diversas aplicaciones, particularmente en probabilidades y estadísticas, donde se encuentran con frecuencia términos de este tipo.

En las subsecciones siguientes, se exploran ejemplos clave de estas herramientas, incluyendo la aproximación asintótica de factoriales mediante la fórmula de Stirling, y la representación de la función exponencial como serie de potencias y como límite. Por último, se aproxima $(1 - \frac{x}{n})^{-k}$ como función de x .

2.2.1. Factorial

La fórmula de Stirling es una aproximación para los factoriales de números grandes y se observa en la Ecuación 2.7.

$$n! \approx e^{-n} \cdot n^n \text{ para } n \text{ grande} \quad (2.7)$$

2.2.2. Función exponencial

La función exponencial real, $\exp : \mathbb{R} \rightarrow \mathbb{R}$, se puede caracterizar de varias maneras equivalentes. Más comúnmente, se define por la serie de potencias de la Ecuación 2.8. Como el radio de convergencia de esta serie de potencias es infinito, esta definición es, de hecho, aplicable a todos los números complejos $z \in \mathbb{C}$.

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (2.8)$$

Por medio del teorema del binomio y la definición de la serie de potencias de la Ecuación 2.8, la función exponencial también se puede definir como el límite de la Ecuación 2.9.

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \quad (2.9)$$

2.2.3. Función $\left(1 - \frac{x}{n}\right)^{-k}$

Otra importante aproximación asintótica aparece en expresiones de la forma $\left(1 - \frac{x}{n}\right)^{-k}$, que son comunes en el análisis de series y distribuciones probabilísticas. Cuando n es muy grande y x y k son números fijos, el término $\frac{x}{n}$ y sus múltiplos son muy pequeños.

En estas condiciones, la expresión se puede aproximar como se observa en la Ecuación 2.10.

$$\left(1 - \frac{x}{n}\right)^{-k} \approx 1 \quad (2.10)$$

Esta aproximación se basa en el hecho de que, para valores pequeños de $\frac{x}{n}$, la expansión en serie de Taylor de $\left(1 - \frac{x}{n}\right)^{-k}$ alrededor de $\frac{x}{n} = 0$ es la que se observa en la Ecuación 2.11.

$$\left(1 - \frac{x}{n}\right)^{-k} = 1 - k \cdot \frac{x}{n} + \frac{k(k+1)}{2} \cdot \left(\frac{x}{n}\right)^2 + \dots \quad (2.11)$$

Como n es grande, $\frac{x}{n}$ es muy pequeño, y los términos de orden superior a $k \cdot \frac{x}{n}$ son insignificantes, la aproximación de la Ecuación 2.10 queda justificada.

2.3. Teoría de la probabilidad

La teoría de la probabilidad es una rama fundamental de las matemáticas que se ocupa del estudio de fenómenos aleatorios. Su importancia radica en su capacidad para modelar y analizar situaciones en las que el resultado de un experimento no puede predecirse con certeza, sino que se caracteriza por la incertidumbre. En este contexto, la probabilidad proporciona las herramientas necesarias para cuantificar y gestionar esta incertidumbre, permitiendo hacer inferencias y tomar decisiones basadas en datos.

En esta sección, se abordan los conceptos básicos de la teoría de la probabilidad, comenzando con los principios de la combinatoria, que sirven como base para el conteo y la organización de eventos en espacios muestrales. Después, se exploran los diferentes modelos de variables aleatorias

que son de gran relevancia en aplicaciones como el análisis de grafos y la modelización de eventos raros.

2.3.1. Combinatoria

La combinatoria estudia las diversas formas de realizar agrupaciones con los elementos de un conjunto, formándolas y calculando su número. Existen distintas formas de realizar estas agrupaciones, según se repitan los elementos o no, según se puedan tomar todos los elementos o no y dependiendo de si influye o no el orden de colocación de los elementos.

Para diferenciar variaciones, permutaciones y combinaciones lo primero es identificar los m elementos disponibles y los n elementos que se toman. En segundo lugar, hay que preguntarse si el orden de colocación de los elementos importa: en caso afirmativo, se trata de variaciones o permutaciones; de lo contrario, son combinaciones. Si se toman todos los elementos disponibles se trata de permutaciones, con $m = n$. Si sólo se toman algunos, son variaciones. Por último, si se pueden repetir los elementos se trata de variaciones o permutaciones con repetición. En caso negativo, son variaciones o permutaciones sin repetición. Este es el esquema que se sigue a continuación, donde se listan las fórmulas combinatorias [21].

1. Variaciones

a) Sin repetición. Se puede observar en la Ecuación 2.12.

$$V_m^n = m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - n + 1) = \frac{m!}{(m - n)!} \quad (2.12)$$

b) Con repetición. Se puede observar en la Ecuación 2.13.

$$VR_m^n = m^n \quad (2.13)$$

2. Permutaciones

a) Sin repetición. Se puede observar en la Ecuación 2.14.

$$P_n = V_m^m = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 3 \cdot 2 \cdot 1 = n! \quad (2.14)$$

b) Con repetición. Se puede observar en la Ecuación 2.15.

$$PR_n^{a,b,c} = \frac{n!}{a! \cdot b! \cdot c!} \quad (2.15)$$

3. Combinaciones. Se puede observar en la Ecuación 2.16.

$$C_m^n = \frac{\text{Variaciones}}{\text{Permutaciones}} = \frac{V_m^n}{P_n} = \binom{m}{n} = \frac{m!}{n! \cdot (m - n)!} \quad (2.16)$$

2.3.2. Modelos de variables aleatorias

En el estudio de la teoría de la probabilidad, es esencial distinguir entre dos tipos fundamentales de variables aleatorias: las variables discretas y las variables continuas. Las variables aleatorias discretas toman un conjunto contable de valores, y su comportamiento se describe mediante una función de probabilidad que asigna una probabilidad a cada uno de estos valores.

Por otro lado, las variables aleatorias continuas pueden tomar un número infinito de valores dentro de un intervalo determinado. Su comportamiento se describe mediante una función de densidad de probabilidad, que permite calcular la probabilidad de que la variable aleatoria caiga dentro de un intervalo específico.

Esta subsección se enfoca en tres tipos de variables aleatorias discretas que son relevantes para el estudio de grafos: la variable aleatoria de Bernoulli, la variable aleatoria binomial y la variable aleatoria de Poisson [8]. Aunque la distribución de Poisson es la más utilizada en el análisis de grafos debido a su capacidad para modelar eventos raros y distribuciones de grado, su comprensión se basa en la binomial, la cual a su vez se construye a partir de la distribución de Bernoulli. Por esta razón, se tratan primero las distribuciones de Bernoulli y binomial, para luego introducir la distribución de Poisson en un contexto adecuado.

Variable aleatoria de Bernoulli La variable aleatoria de Bernoulli se utiliza para modelar un experimento con dos posibles resultados, denominados *éxito* y *fracaso*. Usualmente se asigna el valor 1 para el éxito y el valor 0 para el fracaso. Se supone que el éxito aparece con probabilidad $p \in (0, 1)$ y, por tanto, el fracaso con probabilidad $q = 1 - p$. Este tipo de experimentos se denominan experimentos de Bernoulli.

Más formalmente, una variable aleatoria X sigue una distribución de Bernoulli de parámetro $p \in (0, 1)$, si su función de probabilidad es la que se observa en la Ecuación 2.17.

$$\mathbb{P}(X = 0) = 1 - p \text{ y } \mathbb{P}(X = 1) = p \quad (2.17)$$

Se denota $X \sim \text{Be}(p)$.

Variable aleatoria binomial Se supone un experimento de Bernoulli que es repetido n veces. Además, los experimentos son (mutuamente) independientes y la probabilidad de éxito, p , se mantiene constante en las n repeticiones.

Más formalmente, una variable aleatoria $X =$ número de éxitos en los n ensayos Bernoulli, es una variable aleatoria binomial de parámetros n y p , si su función de probabilidad es la que se observa en la Ecuación 2.18.

$$\mathbb{P}(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}, k = 0, \dots, n \text{ con } q = 1 - p \quad (2.18)$$

Se denota $X \sim \text{Bi}(n, p)$.

Obviamente, si $n = 1$ la variable binomial coincide con la distribución de Bernoulli, es decir, $X \sim \text{Bi}(1, p) = \text{Be}(p)$.

Variable aleatoria de Poisson Una variable aleatoria de Poisson de parámetro $\lambda > 0$ es una variable aleatoria X que toma valores en los enteros no negativos y que tiene una función de probabilidad como la que se observa en la Ecuación 2.19.

$$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, k = 0, 1, \dots \quad (2.19)$$

Se denota $X \sim \text{Po}(\lambda)$.

Proposición 4. *La distribución de Poisson es un caso límite de la distribución binomial.*

Demostración. Sean $X_n \sim \text{Bi}(n, p_n)$ con $p_n = \frac{\lambda}{n}$, con $\lambda > 0$. La función de probabilidad de X_n es la que se observa en la Ecuación 2.20.

$$\mathbb{P}(X_n = k) = \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k}, k = 0, \dots, n \quad (2.20)$$

Si se hace tender n (número de ensayos de Bernoulli) a infinito en la Ecuación 2.20, se obtiene la Ecuación 2.21 (Proposición 5).

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \forall k \geq 0 \text{ entero} \quad (2.21)$$

Por lo tanto, la distribución de Poisson puede entenderse como el límite de probabilidades binomiales cuando el número de ensayos tiende a infinito y la probabilidad de éxito tiende a cero, manteniéndose constante el valor esperado $\lambda = n \cdot p_n$. \square

Proposición 5. *Si se hace tender n (número de ensayos de Bernoulli) a infinito en la Ecuación 2.20, se obtiene la Ecuación 2.21.*

Demostración. Para explicar el paso de la Ecuación 2.20 a la Ecuación 2.21 se aproxima cada uno de sus términos.

Para valores grandes de n ($n \rightarrow \infty$), el primer término de la Ecuación 2.20 se puede aproximar como se observa en la Ecuación 2.22.

$$\binom{n}{k} \approx \frac{n^k}{k!} \quad (2.22)$$

Para demostrar la aproximación de la Ecuación 2.22, primero se recuerda la definición del coeficiente binomial que se observa en la Ecuación 2.23.

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} \quad (2.23)$$

Cuando n es muy grande en comparación con k , se pueden hacer algunas simplificaciones en la Ecuación 2.23. Primero, se puede escribir en una forma más explícita, como se observa en la Ecuación 2.24.

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{k!} \quad (2.24)$$

La Ecuación 2.24 se puede reescribir como se observa en la Ecuación 2.25.

$$\binom{n}{k} = \frac{n^k}{k!} \cdot \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{n^k} \quad (2.25)$$

El segundo término de la Ecuación 2.25 se puede simplificar aún más como se observa en la Ecuación 2.26.

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{n^k} = \frac{n}{n} \cdot \frac{(n-1)}{n} \cdot \frac{(n-2)}{n} \cdot \dots \cdot \frac{(n-k+1)}{n} \quad (2.26)$$

Cada uno de los términos del producto de la Ecuación 2.26 se puede reescribir como en la Ecuación 2.27.

$$\frac{n-i}{n} = 1 - \frac{i}{n} \quad (2.27)$$

Así, la Ecuación 2.26 queda como se observa en la Ecuación 2.28.

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{n^k} = 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) \quad (2.28)$$

Ahora, cuando $n \rightarrow \infty$, cada término de la Ecuación 2.28 se aproxima a 1 y, por tanto, el producto de todos estos términos también se aproxima a 1 (Ecuación 2.29).

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{n^k} \approx 1 \quad (2.29)$$

Para valores grandes de n ($n \rightarrow \infty$), el tercer término de la Ecuación 2.20 se puede aproximar como se observa en la Ecuación 2.30.

$$\left(1 - \frac{\lambda}{n}\right)^{n-k} = \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \approx e^{-\lambda} \quad (2.30)$$

Para ello, se utiliza el límite exponencial de la Ecuación 2.9 y la aproximación asintótica de la Ecuación 2.10, sustituyendo x por λ .

Sustituyendo las aproximaciones de la Ecuación 2.22 y de la Ecuación 2.30 en la fórmula original de la Ecuación 2.20 como se observa en la Ecuación 2.31, se obtiene la Ecuación 2.21.

$$\mathbb{P}(X_n = k) = \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \approx \frac{n^k}{k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot e^{-\lambda} = \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} \cdot e^{-\lambda} \quad (2.31)$$

□

Capítulo 3

Grafos y redes sociales

Las redes sociales pueden ser modeladas matemáticamente mediante grafos [12]. Este capítulo analiza cómo las propiedades de los grafos, como la centralidad, la densidad y la excentricidad, son esenciales para entender la estructura y el comportamiento de las redes sociales. Asimismo, se introducen algoritmos clave como el de Floyd-Warshall y métodos de detección de comunidades, que permiten identificar influencias y patrones en la red. Todo esto proporciona las herramientas necesarias para analizar redes sociales desde una perspectiva cuantitativa.

3.1. Introducción

Sea $G = (V, E)$ un grafo que representa una red social con $n = |V|$ nodos y $m = |E|$ aristas. En principio, G es un grafo simple, aunque muchas de las definiciones pueden adaptarse a digrafos. Se supone, por simplicidad, que el grafo es conexo.

Nota 16. *En el contexto del análisis de redes, es común denotar el número de nodos con la letra N y el número de aristas con la letra M . Sin embargo, al considerar que estos valores pueden variar debido al crecimiento y la evolución de la red, resulta más apropiado utilizar minúsculas, es decir, $n = N(t)$ y $m = M(t)$, para representar estas cantidades. Este cambio refleja que, en realidad, no son constantes, sino variables que pueden aumentar o disminuir en función del desarrollo de la red. De este modo, se evita la connotación de inmutabilidad que podría asociarse con las letras en mayúscula, facilitando un entendimiento más preciso y adaptable al contexto dinámico del análisis de redes.*

En busca de los nodos más influyentes (en inglés, *influencers*), lo primero que se puede mirar es su grado. En particular, interesa el grado máximo, así como localizar los nodos con grados más altos.

El **grado medio**, es decir, la media de los grados, mide el grado de cohesión de la red. Aplicando el Teorema 1, el grado medio es como se indica en la Ecuación 3.1.

$$\frac{1}{n} \cdot \sum_{v \in V} \delta(v) = \frac{\delta(v_1) + \delta(v_2) + \dots + \delta(v_n)}{n} = \frac{2 \cdot m}{n} \quad (3.1)$$

Ejemplo 25. *El grado medio del grafo de la Figura 2.4 es $\frac{2 \cdot 12}{8} = \frac{24}{8} = 3$.*

Nota 17. En el caso de digrafos, estos conceptos se aplican a las aristas salientes y entrantes. En particular, de cara a estudiar influencers, interesa más el grado entrante de un nodo (seguidores o, en inglés, followers) que el saliente (seguidos o, en inglés, following).

3.2. Excentricidad

Definición 46. Se denomina camino geodésico o **geodésica** a todo camino (Definición 35) de longitud mínima entre dos nodos.

Nota 18. Entre dos nodos puede haber más de un camino geodésico, pero todos tienen la misma longitud.

Ejemplo 26. Un camino geodésico entre los vértices v_1 y v_6 del grafo de la Figura 2.4 es $\{v_1, e_1, v_2, e_5, v_6\}$. Otro camino geodésico entre los vértices v_1 y v_6 es $\{v_1, e_3, v_5, e_9, v_6\}$. No hay más caminos de longitud 2 o menor entre ambos vértices.

Definición 47. Se denomina **distancia** entre dos nodos $u, v \in V$, y se denota por $d(u, v) \in \mathbb{N}$, a la longitud de cualquiera de las geodésicas entre u y v .

Ejemplo 27. La distancia entre los vértices v_1 y v_6 del grafo de la Figura 2.4 es $d(v_1, v_6) = 2$.

Definición 48. Se denomina **excentricidad** de un nodo $u \in V$ a la mayor distancia entre u y otro nodo cualquiera de G .

Ejemplo 28. La excentricidad del vértice v_1 del grafo de la Figura 2.4 es 3.

Definición 49. Se denomina **diámetro** a la mayor excentricidad de G . En otras palabras, es la longitud del camino geodésico más largo posible en G .

Ejemplo 29. El diámetro del grafo de la Figura 2.4 es 3.

Los nodos con mayor excentricidad se consideran en la **periferia** del grafo, mientras que los de menor excentricidad se consideran el centro o **núcleo** del mismo. Los nodos del núcleo son los que en menos pasos pueden llegar a la totalidad de los nodos del grafo; por ejemplo, para difundir una información.

Definición 50. Se denomina **camino característico** (en inglés, average path length) de G a la longitud media de un camino geodésico.

Nota 19. Las medias se calculan sobre el número total de parejas del grafo. Si el grafo no es conexo, la distancia entre dos nodos que no se pueden conectar con un camino es infinita, y esta media se hace sobre todas las parejas a distancia finita.

Ejemplo 30. El camino característico del grafo de la Figura 2.4 se calcula en la Ecuación 3.2. El número total de parejas del grafo es $\binom{n}{2} = \binom{8}{2} = \frac{8!}{2!6!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1) \cdot (6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} = \frac{56}{2} = 28$.

$$\begin{aligned}
 & \frac{1}{28} \cdot [d(v_1, v_2) + d(v_1, v_3) + d(v_1, v_4) + d(v_1, v_5) + d(v_1, v_6) + d(v_1, v_7) + d(v_1, v_8) + \\
 & \quad d(v_2, v_3) + d(v_2, v_4) + d(v_2, v_5) + d(v_2, v_6) + d(v_2, v_7) + d(v_2, v_8) + \\
 & \quad d(v_3, v_4) + d(v_3, v_5) + d(v_3, v_6) + d(v_3, v_7) + d(v_3, v_8) + \\
 & \quad d(v_4, v_5) + d(v_4, v_6) + d(v_4, v_7) + d(v_4, v_8) + \\
 & \quad d(v_5, v_6) + d(v_5, v_7) + d(v_5, v_8) + \\
 & \quad d(v_6, v_7) + d(v_6, v_8) + \\
 & \quad d(v_7, v_8)] = \\
 & \quad \frac{1}{28} \cdot (1 + 2 + 1 + 1 + 2 + 3 + 2 + \\
 & \quad 1 + 2 + 2 + 1 + 2 + 3 + \\
 & \quad 1 + 3 + 2 + 1 + 2 + \\
 & \quad 2 + 3 + 2 + 1 + \\
 & \quad 1 + 2 + 1 + \\
 & \quad 1 + 2 + \\
 & \quad 1) = \\
 & \quad \frac{1}{28} \cdot (12 + 11 + 9 + 8 + 4 + 3 + 1) = \\
 & \quad \frac{1}{28} \cdot (48) = \frac{48}{28} = \frac{12}{7} = 1,714
 \end{aligned} \tag{3.2}$$

La distancia entre dos vértices depende de si están en la misma arista, en la misma cara, o en caras distintas.

Nota 20. En el caso de digrafos, los caminos que interesan son los caminos entrantes en cada nodo. Por tanto, se debe respetar el sentido de las aristas en los caminos.

3.3. Densidad

Se supone que los grafos con más cantidad de aristas son los más cohesionados. Esto tiene interés en la detección de comunidades, que se ve en la Sección 3.6.

En primer lugar, el número total de aristas posibles en G es el número combinatorio de la Ecuación 3.3, es decir, el número de parejas posibles con n elementos o combinaciones sin repetición (Sección 2.3.1).

$$\binom{n}{2} = \frac{n \cdot (n - 1)}{2} \tag{3.3}$$

Nota 21. Si G tiene todas las aristas posibles se trata de un grafo completo o clique (Definición 12).

Definición 51. La *densidad* de G , $\rho(G)$, es la ratio entre el número total de aristas y el número total de aristas posibles. Se calcula según la Ecuación 3.4.

$$\rho(G) = \frac{m}{\frac{n \cdot (n-1)}{2}} = \frac{2 \cdot m}{n \cdot (n-1)} \quad (3.4)$$

Ejemplo 31. La densidad del grafo de la Figura 2.4 es $\rho(Q_3) = \frac{2 \cdot 12}{8 \cdot (8-1)} = \frac{24}{56} = \frac{3}{7} = 0,429$.

Definición 52. Se dice que G es *disperso* (en inglés, *sparse*) si existe una constante $c \leq 10$ tal que $m \leq c \cdot n$. En caso contrario, G es *denso*, más o menos según su densidad sea o no próxima a 1.

Ejemplo 32. Dado que $12 \leq 2 \cdot 8$, el grafo de la Figura 2.4 es disperso. Se puede intuir también de su densidad, 0,429, más cercana a 0 que a 1.

Este concepto de densidad es global; por ello, se aplica directamente a los digrafos, con la única diferencia de que el número total posible de aristas orientadas es $n \cdot (n-1)$. Por tanto, se tiene que la densidad se calcula según la Ecuación 3.5.

$$\rho(G) = \frac{m}{n \cdot (n-1)} \quad (3.5)$$

Esta definición puede aplicarse también a un subgrafo S de G . Típicamente, estos subgrafos son o bien el conjunto de vecinos de un nodo dado, o una comunidad de usuarios.

Definición 53. Dado un nodo $u \in V$ de G , se considera el subgrafo S formado por u y sus nodos adyacentes (vecinos). Entonces, se denomina *coeficiente de agrupamiento* (en inglés, *clustering coefficient*) **de u** a la densidad del subgrafo S , es decir, $\rho(S)$. Se denota por $C(u)$, y mide cuan cohesionada está la red en el entorno de este nodo.

Ejemplo 33. El coeficiente de agrupamiento del vértice v_1 del grafo de la Figura 2.4 es $C(v_1) = \rho(S_{v_1}) = \frac{3}{3 \cdot (3-1)} = \frac{3}{3 \cdot 2} = \frac{1}{2} = 0,5$

Se puede demostrar que, en el caso de grafos simples, el coeficiente de agrupamiento de u coincide con la densidad de los triángulos en S , es decir, la ratio entre el número de triángulos existentes en S con u en uno de los vértices y el número total posible de triángulos en S con u en uno de los vértices. Esta última cantidad se expresa en la Ecuación 3.6, donde $N(u)$ es el conjunto de vecinos de u y $\delta(u)$ es el número de vecinos de u , que coincide con la cardinalidad de $N(u)$, es decir, $\delta(u) = |N(u)|$, y con el grado de u . Para calcularla, hay que observar que los triángulos que incluyen a u se forman a partir de los nodos en su vecindario $N(u)$. Para que exista un triángulo, se necesitan tres nodos: el nodo u y dos de sus vecinos. Por tanto, el número total de formas de seleccionar dos vecinos de u es el número de combinaciones de 2 elementos de un conjunto de tamaño $|N(u)|$.

$$\binom{\delta(u)}{2} = \frac{\delta(u) \cdot (\delta(u) - 1)}{2} \quad (3.6)$$

Haciendo la media de los coeficientes de agrupamiento de todos los nodos de G como muestra la Ecuación 3.7, se obtiene el **coeficiente de agrupamiento de G** , $C(G)$.

$$C(G) = \frac{1}{n} \cdot \sum_{v \in V} C(v) \quad (3.7)$$

Ejemplo 34. El coeficiente de agrupamiento del grafo de la Figura 2.4 es $C(Q_3) = \frac{1}{8} \cdot [C(v_1) + \dots + C(v_8)] = \frac{1}{8} \cdot (0,5 + \dots + 0,5) = \frac{1}{8} \cdot 8 \cdot 0,5 = 0,5$

Nota 22. Hay un fenómeno conocido como “fenómeno del pequeño mundo” (en inglés, *small world*), que ocurre cuando entre dos nodos cualesquiera existe siempre (o casi siempre) un camino corto. Esto suele ser consecuencia de una combinación entre alta densidad, camino característico corto y alto coeficiente de agrupamiento.

Ejemplo 35. Un estudio (referencia [19]) sobre 700 millones de usuarios de Facebook muestra que la distancia media entre dos usuarios es 4,74, es decir, entre ellos solo hay 4 intermediarios.

Ejemplo 36. Se conjetura que entre dos personas cualesquiera en un país existen, como mucho, 6 grados de separación, es decir, que se puede establecer una cadena de “conocidos” de manera que en un número reducido de pasos ambas personas están conectadas. Esta conjetura surge en 1967 tras un experimento sociológico que se realiza en Estados Unidos.

3.4. Centralidad

Hay tres medidas que determinan cómo de central es un nodo $u_k \in V$; se listan a continuación.

- Por cercanía (en inglés, *closeness*). La Ecuación 3.8 muestra la medida de centralidad de un nodo por cercanía, $C_C(u_k)$.

$$\frac{1}{C_C(u_k)} = \sum_{i=1}^n d(u_i, u_k) \quad (3.8)$$

Es decir, esta centralidad es el inverso de la suma de las distancias geodésicas al resto de nodos (Ecuación 3.9).

$$C_C(u_k) = \frac{1}{\sum_{i=1}^n d(u_i, u_k)} \quad (3.9)$$

- Por intermediación (en inglés, *betweenness*). La Ecuación 3.10 muestra la medida de centralidad de un nodo por intermediación, $C_B(u_k)$.

$$C_B(u_k) = \sum_{i < j}^n I_{i,j}(u_k) \quad (3.10)$$

A continuación, se define cada uno de los términos utilizados en la Ecuación 3.10.

- $g_{i,j}(u_k)$ es el número de caminos geodésicos entre u_i y u_j que pasan por u_k .
- $g_{i,j}$ es el número total de geodésicas entre u_i y u_j .

- $I_{i,j}(u_k)$ se calcula a partir de $g_{i,j}(u_k)$ y $g_{i,j}$ según la Ecuación 3.11.

$$I_{i,j}(u_k) = \frac{g_{i,j}(u_k)}{g_{i,j}} \quad (3.11)$$

En otras palabras, esta centralidad mide cuan probable es que las geodésicas pasen por el nodo u_k .

- Por vector propio (en inglés, *eigenvector*). El vector de centralidades $\mathbf{p} = (C_V(u_1), \dots, C_V(u_n))$ muestra la medida de centralidad de un nodo por vector propio y se calcula como una solución del sistema lineal de ecuaciones de la Ecuación 3.12, donde $A' = (A_G)^T$ es la traspuesta de la matriz de adyacencia de G e I_n es la matriz identidad de orden n . La centralidad del nodo u_k se refleja en el valor específico $p_k = C_V(u_k)$, que es la k -ésima entrada del vector propio \mathbf{p} .

$$(A' - I_n) \cdot \mathbf{p} = 0 \quad (3.12)$$

En términos de Álgebra Lineal, \mathbf{p} es un vector propio de A' asociado al valor propio $\lambda = 1$.

Ejemplo 37. En la Figura 3.1 se representan tres grafos de ejemplo, formados por dos triángulos con una arista que los une: e_7 , e'_7 y e''_7 , respectivamente. De izquierda a derecha se denotan como G , G' y G'' .

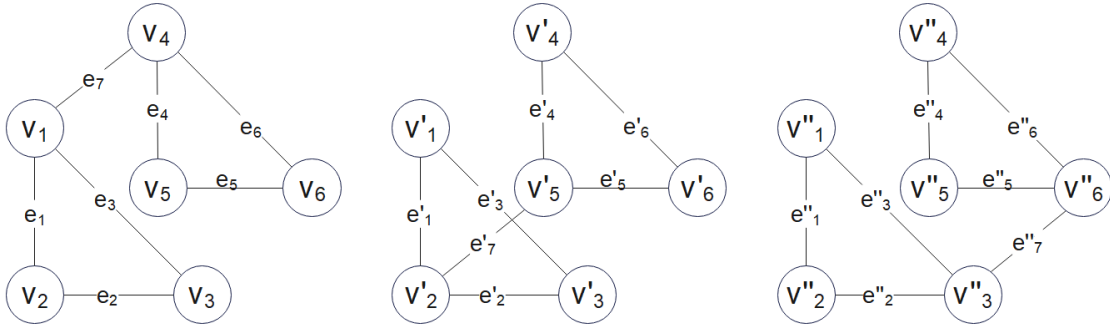


Figura 3.1: Grafos de ejemplo para centralidad y detección de comunidades.

- La Ecuación 3.13 muestra la medida de centralidad del nodo v_1 por cercanía, $C_C(v_1)$.

$$\begin{aligned} C_C(v_1) &= \frac{1}{\sum_{i=1}^6 d(v_i, v_1)} \\ &= \frac{1}{[d(v_1, v_1) + d(v_2, v_1) + d(v_3, v_1) + d(v_4, v_1) + d(v_5, v_1) + d(v_6, v_1)]} \\ &= \frac{1}{(0 + 1 + 1 + 1 + 2 + 2)} = \frac{1}{7} = 0,149 \end{aligned} \quad (3.13)$$

La Ecuación 3.14 muestra la medida de centralidad del nodo v'_1 por cercanía, $C_C(v'_1)$.

$$\begin{aligned}
C_C(v'_1) &= \frac{1}{\sum_{i=1}^6 d(v'_i, v'_1)} \\
&= \frac{1}{[d(v'_1, v'_1) + d(v'_2, v'_1) + d(v'_3, v'_1) + d(v'_4, v'_1) + d(v'_5, v'_1) + d(v'_6, v'_1)]} \\
&= \frac{1}{(0 + 1 + 1 + 3 + 2 + 3)} = \frac{1}{10} = 0,1
\end{aligned} \tag{3.14}$$

La Ecuación 3.15 muestra la medida de centralidad del nodo v''_1 por cercanía, $C_C(v''_1)$.

$$\begin{aligned}
C_C(v''_1) &= \frac{1}{\sum_{i=1}^6 d(v''_i, v''_1)} \\
&= \frac{1}{[d(v''_1, v''_1) + d(v''_2, v''_1) + d(v''_3, v''_1) + d(v''_4, v''_1) + d(v''_5, v''_1) + d(v''_6, v''_1)]} \\
&= \frac{1}{(0 + 1 + 1 + 3 + 3 + 2)} = \frac{1}{10} = 0,1
\end{aligned} \tag{3.15}$$

- La Ecuación 3.16 muestra la medida de centralidad del nodo v_1 por intermediación, $C_B(v_1)$.

$$\begin{aligned}
C_B(v_1) &= \sum_{i < j}^6 I_{i,j}(v_1) \\
&= [I_{1,2}(v_1) + I_{1,3}(v_1) + I_{1,4}(v_1) + I_{1,5}(v_1) + [I_{1,6}(v_1)]] \\
&+ [I_{2,3}(v_1) + I_{2,4}(v_1) + I_{2,5}(v_1) + I_{2,6}(v_1)] \\
&+ [I_{3,4}(v_1) + I_{3,5}(v_1) + I_{3,6}(v_1)] \\
&+ [I_{4,5}(v_1) + I_{4,6}(v_1)] \\
&+ [I_{5,6}(v_1)] \\
&= \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1}\right) + \left(\frac{0}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1}\right) + \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1}\right) + \left(\frac{0}{1} + \frac{0}{1}\right) + \left(\frac{0}{1}\right) \\
&= (1 + 1 + 1 + 1 + 1) + (0 + 1 + 1 + 1) + (1 + 1 + 1) + (0 + 0) + (0) = 11
\end{aligned} \tag{3.16}$$

La Ecuación 3.17 muestra la medida de centralidad del nodo v'_1 por intermediación,

$C_B(v'_1)$.

$$\begin{aligned}
 C_B(v'_1) &= \sum_{i < j}^6 I_{i,j}(v'_1) \\
 &= [I_{1,2}(v'_1) + I_{1,3}(v'_1) + I_{1,4}(v'_1) + I_{1,5}(v'_1) + I_{1,6}(v'_1)] \\
 &\quad + [I_{2,3}(v'_1) + I_{2,4}(v'_1) + I_{2,5}(v'_1) + I_{2,6}(v'_1)] \\
 &\quad + [I_{3,4}(v'_1) + I_{3,5}(v'_1) + I_{3,6}(v'_1)] \\
 &\quad + [I_{4,5}(v'_1) + I_{4,6}(v'_1)] \\
 &\quad + [I_{5,6}(v'_1)] \\
 &= \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1}\right) + \left(\frac{0}{1} + \frac{0}{1} + \frac{0}{1} + \frac{0}{1}\right) + \left(\frac{0}{1} + \frac{0}{1} + \frac{0}{1}\right) + \left(\frac{0}{1} + \frac{0}{1}\right) + \left(\frac{0}{1}\right) \\
 &= (1 + 1 + 1 + 1 + 1) + (0 + 0 + 0 + 0) + (0 + 0 + 0) + (0 + 0) + (0) = 5
 \end{aligned} \tag{3.17}$$

La Ecuación 3.18 muestra la medida de centralidad del nodo v''_1 por intermediación, $C_B(v''_1)$.

$$\begin{aligned}
 C_B(v''_1) &= \sum_{i < j}^6 I_{i,j}(v''_1) \\
 &= [I_{1,2}(v''_1) + I_{1,3}(v''_1) + I_{1,4}(v''_1) + I_{1,5}(v''_1) + I_{1,6}(v''_1)] \\
 &\quad + [I_{2,3}(v''_1) + I_{2,4}(v''_1) + I_{2,5}(v''_1) + I_{2,6}(v''_1)] \\
 &\quad + [I_{3,4}(v''_1) + I_{3,5}(v''_1) + I_{3,6}(v''_1)] \\
 &\quad + [I_{4,5}(v''_1) + I_{4,6}(v''_1)] \\
 &\quad + [I_{5,6}(v''_1)] \\
 &= \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1}\right) + \left(\frac{0}{1} + \frac{0}{1} + \frac{0}{1} + \frac{0}{1}\right) + \left(\frac{0}{1} + \frac{0}{1} + \frac{0}{1}\right) + \left(\frac{0}{1} + \frac{0}{1}\right) + \left(\frac{0}{1}\right) \\
 &= (1 + 1 + 1 + 1 + 1) + (0 + 0 + 0 + 0) + (0 + 0 + 0) + (0 + 0) + (0) = 5
 \end{aligned} \tag{3.18}$$

- El vector de centralidades $\mathbf{p} = (C_V(v_1), \dots, C_V(v_6))$ muestra la medida de centralidad de un nodo por vector propio y se calcula como una solución del sistema lineal de ecuaciones de la Ecuación 3.19. En términos de Álgebra Lineal, \mathbf{p} es un vector propio de A_1 asociado al valor propio $\lambda = 1$.

$$(A_1 - I_6) \cdot \mathbf{p} = 0 \tag{3.19}$$

La traspuesta de la matriz de adyacencia de G , $A_1 = (A_G)^T$, es la misma que A_G , ya que es simétrica.

$$\begin{pmatrix}
 & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\
 v_1 & 0 & 1 & 1 & 1 & 0 & 0 \\
 v_2 & 1 & 0 & 1 & 0 & 0 & 0 \\
 v_3 & 1 & 1 & 0 & 0 & 0 & 0 \\
 v_4 & 1 & 0 & 0 & 0 & 1 & 1 \\
 v_5 & 0 & 0 & 0 & 1 & 0 & 1 \\
 v_6 & 0 & 0 & 0 & 1 & 1 & 0
 \end{pmatrix}$$

Luego, se calcula $(A_1 - I_6)$, donde I_6 es la matriz identidad de orden 6.

$$\begin{pmatrix} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ v_1 & -1 & 1 & 1 & 1 & 0 & 0 \\ v_2 & 1 & -1 & 1 & 0 & 0 & 0 \\ v_3 & 1 & 1 & -1 & 0 & 0 & 0 \\ v_4 & 1 & 0 & 0 & -1 & 1 & 1 \\ v_5 & 0 & 0 & 0 & 1 & -1 & 1 \\ v_6 & 0 & 0 & 0 & 1 & 1 & -1 \end{pmatrix}$$

El cálculo de la Ecuación 3.19 se puede realizar utilizando un método numérico como descomposición de valores propios (Anexo D). La centralidad del nodo v_1 se refleja en el valor específico $p_1 = C_V(v_1)$, que es la primera entrada del vector propio \mathbf{p} , es decir, $p_1 = C_V(v_1) = -0,628$.

$$\mathbf{p} = [-0,62796303, 0,22985042, 0,22985042, 0,62796303, -0,22985042, -0,22985042]$$

El vector de centralidades $\mathbf{p}' = (C_V(v'_1), \dots, C_V(v'_6))$ muestra la medida de centralidad de un nodo por vector propio y se calcula como una solución del sistema lineal de ecuaciones de la Ecuación 3.20. En términos de Álgebra Lineal, \mathbf{p}' es un vector propio de A_2 asociado al valor propio $\lambda = 1$.

$$(A_2 - I_6) \cdot \mathbf{p}' = 0 \quad (3.20)$$

La traspuesta de la matriz de adyacencia de G' , $A_2 = (A_{G'})^T$, es la misma que $A_{G'}$, ya que es simétrica.

$$\begin{pmatrix} & v'_1 & v'_2 & v'_3 & v'_4 & v'_5 & v'_6 \\ v'_1 & 0 & 1 & 1 & 0 & 0 & 0 \\ v'_2 & 1 & 0 & 1 & 0 & 1 & 0 \\ v'_3 & 1 & 1 & 0 & 0 & 0 & 0 \\ v'_4 & 0 & 0 & 0 & 0 & 1 & 1 \\ v'_5 & 0 & 1 & 0 & 1 & 0 & 1 \\ v'_6 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Luego, se calcula $(A_2 - I_6)$.

$$\begin{pmatrix} & v'_1 & v'_2 & v'_3 & v'_4 & v'_5 & v'_6 \\ v'_1 & -1 & 1 & 1 & 0 & 0 & 0 \\ v'_2 & 1 & -1 & 1 & 0 & 1 & 0 \\ v'_3 & 1 & 1 & -1 & 0 & 0 & 0 \\ v'_4 & 0 & 0 & 0 & -1 & 1 & 1 \\ v'_5 & 0 & 1 & 0 & 1 & -1 & 1 \\ v'_6 & 0 & 0 & 0 & 1 & 1 & -1 \end{pmatrix}$$

Nuevamente, se resuelve el sistema de la Ecuación 3.20 (Anexo D). La centralidad del nodo v'_1 se refleja en el valor específico $p'_1 = C_V(v'_1)$, que es la primera entrada del vector propio \mathbf{p}' , es decir, $p'_1 = C_V(v'_1) = -0,354$.

$$\mathbf{p}' = [-0,35355339, -0,5, -0,35355339, -0,35355339, -0,5, -0,35355339]$$

El vector de centralidades $\mathbf{p}'' = (C_V(v_1''), \dots, C_V(v_6''))$ muestra la medida de centralidad de un nodo por vector propio y se calcula como una solución del sistema lineal de ecuaciones de la Ecuación 3.21. En términos de Álgebra Lineal, \mathbf{p}'' es un vector propio de A_3 asociado al valor propio $\lambda = 1$.

$$(A_3 - I_6) \cdot \mathbf{p}'' = 0 \quad (3.21)$$

La traspuesta de la matriz de adyacencia de G'' , $A_3 = (A_{G''})^T$, es la misma que $A_{G''}$, ya que es simétrica.

$$\begin{pmatrix} v_1'' & v_1'' & v_2'' & v_3'' & v_4'' & v_5'' & v_6'' \\ v_1'' & 0 & 1 & 1 & 0 & 0 & 0 \\ v_2'' & 1 & 0 & 1 & 0 & 0 & 0 \\ v_3'' & 1 & 1 & 0 & 0 & 0 & 1 \\ v_4'' & 0 & 0 & 0 & 0 & 1 & 1 \\ v_5'' & 0 & 0 & 0 & 1 & 0 & 1 \\ v_6'' & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Luego, se calcula $(A_3 - I_6)$.

$$\begin{pmatrix} v_1'' & v_1'' & v_2'' & v_3'' & v_4'' & v_5'' & v_6'' \\ v_1'' & -1 & 1 & 1 & 0 & 0 & 0 \\ v_2'' & 1 & -1 & 1 & 0 & 0 & 0 \\ v_3'' & 1 & 1 & -1 & 0 & 0 & 1 \\ v_4'' & 0 & 0 & 0 & -1 & 1 & 1 \\ v_5'' & 0 & 0 & 0 & 1 & -1 & 1 \\ v_6'' & 0 & 0 & 1 & 1 & 1 & -1 \end{pmatrix}$$

Al igual que en los casos anteriores, se resuelve el sistema de la Ecuación 3.21 (Anexo D). La centralidad del nodo v_1'' se refleja en el valor específico $p_1'' = C_V(v_1'')$, que es la primera entrada del vector propio \mathbf{p}'' , es decir, $p_1'' = C_V(v_1'') = -0,354$.

$$\mathbf{p}'' = [-0,35355339, -0,35355339, -0,5, -0,35355339, -0,35355339, -0,5]$$

Nota 23. En el caso de digrafos, son interesantes los caminos entrantes de cada nodo para la centralidad por cercanía. Además, por simplicidad, se supone que el digrafo es fuertemente conexo, pues, en caso contrario, solo deben considerarse las parejas de nodos a distancia finita.

Para la centralidad por intermediación, hay que hacer la suma para todo par de índices (i, j) , sin la restricción $i < j$.

3.5. Algoritmo de Floyd-Warshall

Este algoritmo se construye a partir de la matriz de pesos D , que es la matriz estándar de dimensión $n \times n$ con diagonal nula, (i, j) -ésima entrada unitaria cuando v_i y v_j , $i \neq j$, son adyacentes, e infinita en caso contrario, es decir, cuando son independientes. En otras palabras, si esta matriz de pesos es $D = [d_{ij}]$, entonces d_{ij} es una función definida a trozos como indica la Ecuación 3.22.

$$d_{ij} = \begin{cases} 0 & \text{cuando } i = j \\ 1 & \text{cuando los nodos } v_i, v_j, \text{ con } i \neq j, \text{ son adyacentes} \\ \infty & \text{en otro caso} \end{cases} \quad (3.22)$$

En otros contextos, se puede realizar con la matriz D en la que d_{ij} es el peso de la arista correspondiente, o bien 1 si no hay arista.

El objetivo del algoritmo es encontrar las distancias geodésicas entre todos los pares de nodos, junto con una matriz C que permite hallar uno de los caminos geodésicos correspondientes.

Aparte de la matriz D , se inicializa otra matriz C con $c_{ij} = i$. Cuando se termina el algoritmo, se tiene el penúltimo nodo de un camino óptimo que une los nodos v_i y v_j , salvo $c_{ii} = i$. La matriz C final permite reconstruir de forma recursiva un camino óptimo (geodésico) para ir de un nodo cualquiera a otro.

Los pasos del algoritmo de Floyd son dos para cada valor de $1 \leq k \leq n$, $1 \leq i \leq n$ y $1 \leq j \leq n$, y se listan a continuación.

- Si $D_{ik} + D_{kj} < D_{ij}$, entonces $D_{ij} = D_{ik} + D_{kj}$ y $C_{ij} = C_{kj}$.
- En caso contrario, las matrices se dejan como están.

Ejemplo 38. Se considera la matriz de pesos de la Ecuación 3.23 y la matriz inicial de la Ecuación 3.24.

$$D = \begin{pmatrix} & v_0 & v_1 & v_2 & v_3 \\ v_0 & 0 & 1 & \infty & \infty \\ v_1 & 1 & 0 & 1 & \infty \\ v_2 & \infty & 1 & 0 & 1 \\ v_3 & \infty & \infty & 1 & 0 \end{pmatrix} \quad (3.23)$$

$$C = \begin{pmatrix} & v_0 & v_1 & v_2 & v_3 \\ v_0 & 0 & 0 & 0 & 0 \\ v_1 & 1 & 1 & 1 & 1 \\ v_2 & 2 & 2 & 2 & 2 \\ v_3 & 3 & 3 & 3 & 3 \end{pmatrix} \quad (3.24)$$

Al finalizar el algoritmo, se obtienen la matriz de distancias de la Ecuación 3.25 y la matriz de nodos intermedios de la Ecuación 3.26.

$$D = \begin{pmatrix} & v_0 & v_1 & v_2 & v_3 \\ v_0 & 0 & 1 & 2 & 3 \\ v_1 & 1 & 0 & 1 & 2 \\ v_2 & 2 & 1 & 0 & 1 \\ v_3 & 3 & 2 & 1 & 0 \end{pmatrix} \quad (3.25)$$

$$C = \begin{pmatrix} & v_0 & v_1 & v_2 & v_3 \\ v_0 & 0 & 0 & 1 & 2 \\ v_1 & 1 & 1 & 1 & 2 \\ v_2 & 1 & 2 & 2 & 2 \\ v_3 & 1 & 2 & 3 & 3 \end{pmatrix} \quad (3.26)$$

Por ejemplo, si se quiere ir del nodo v_0 al v_3 , en la matriz de distancias se observa que la distancia es 3, y que se llega al nodo v_3 desde el nodo v_2 . Siguiendo el hilo, del v_0 al v_2 se llega desde el v_1 , y del v_0 al v_1 se llega directamente mediante una arista desde el v_0 ; esto se observa en la matriz de nodos intermedios. Se trata de un grafo dirigido, pues la matriz C es asimétrica.

3.6. Detección de comunidades

De una manera imprecisa, una **comunidad** dentro de una red se define como un conjunto de nodos cuya densidad es alta. Dependiendo del umbral utilizado para considerar esta densidad “alta”, se obtiene un mayor o un menor número de comunidades.

Este tema es complejo y, además, los autores no se ponen de acuerdo en si las comunidades son disjuntas o pueden tener **solapamientos**, es decir, si algunos nodos pueden pertenecer simultáneamente a varias comunidades. En este trabajo, y por simplicidad, se consideran particiones de redes en comunidades disjuntas. También para simplificar la nomenclatura, a un subgrafo completo se le llama *clique*.

Antes de avanzar en el tema, es necesario conocer una clasificación de las aristas que se produce cuando el grafo se encuentra particionado en comunidades.

Definición 54. En una partición de un grafo, una arista **intracomunitaria** es una arista que conecta dos nodos o vértices que pertenecen a la misma comunidad o grupo dentro de esa partición.

Nota 24. Estas aristas son importantes para medir la cohesión dentro de una comunidad, ya que representan las conexiones internas.

Definición 55. Las aristas que conectan nodos de diferentes comunidades se llaman aristas **intercomunitarias**.

3.6.1. Medida

Para una partición dada, hay dos números que miden lo buena que es la partición; se listan a continuación.

- Modularidad Q . Se calcula según la Ecuación 3.27, donde $A_G = [a_{ij}]$ es la matriz de adyacencia, k_i representa el grado del nodo $u_i \in V$ (es decir, $k_i = \delta(u_i)$), n es el número de nodos de G ($n = |V|$), m es el número de aristas de G ($m = |E|$), y la función δ vale 1 si ambos nodos están en la misma comunidad, y 0 en caso contrario.

$$Q = \frac{1}{2 \cdot m} \cdot \sum_{i=1}^n \sum_{j=1}^n \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot m} \right) \cdot \delta(u_i, u_j) \quad (3.27)$$

- Índice H (Surprise). Se calcula según la Ecuación 3.28, donde F es el número máximo posible de aristas en G , es decir, $\binom{n}{2}$, m es, como en la Ecuación 3.27, el número real de aristas de G , M es el número máximo posible de aristas “intracomunitarias” para esa partición, y p es el número real de aristas intracomunitarias.

$$H = -\log \sum_{j=p}^{\min(M,m)} \frac{\binom{M}{j} \cdot \binom{F-M}{m-j}}{\binom{F}{m}} \quad (3.28)$$

Debido al signo menos del logaritmo, cuanto mayor es H mejor es la partición de la red. Por defecto, el logaritmo es neperiano, pero esto solo influye en una constante de proporcionalidad.

Ejemplo 39. *Se supone una partición en la que cada nodo forma una comunidad distinta (Partición 1 del Anexo E). Esto significa que la función $\delta(v_i, v_i) = 1 \forall i \in [1, 6]$ y nula en el resto, por lo que los sumandos que no coinciden con la diagonal se anulan. A continuación se calcula la modularidad para cada uno de los grafos.*

- *Modularidad Q . Se calcula según la Ecuación 3.29, donde $A_1 = A_G = [a_{ij}]$ es la matriz de adyacencia, k_i representa el grado del nodo $v_i \in V$ (es decir, $k_i = \delta(v_i)$).*

$$\begin{aligned}
 Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
 &= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) \right. \\
 &\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
 &\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
 &\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
 &\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
 &\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
 &= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
 &= \frac{1}{14} \cdot \left[\left(\frac{-9}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-9}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-4}{14} \right) \right] \\
 &= \frac{1}{14} \cdot \left[\left(2 \cdot \frac{-9}{14} \right) + \left(4 \cdot \frac{-4}{14} \right) \right] \\
 &= \frac{1}{14} \cdot \left[\left(\frac{-9}{7} \right) + \left(\frac{-8}{7} \right) \right] = \frac{1}{14} \cdot \left(\frac{-17}{7} \right) = \frac{-17}{14 \cdot 7} = \frac{-17}{98} = -0,173
 \end{aligned} \tag{3.29}$$

- *Modularidad Q' . Se calcula según la Ecuación 3.30, donde $A_2 = A'_G = [a'_{ij}]$ es la matriz de adyacencia, k'_i representa el grado del nodo $v'_i \in V$ (es decir, $k'_i = \delta(v'_i)$).*

$$\begin{aligned}
 Q' &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a'_{ij} - \frac{k'_i \cdot k'_j}{2 \cdot 7} \right) \cdot \delta(v'_i, v'_j) \\
 &= \frac{1}{14} \cdot \left[\left(a'_{11} - \frac{k'_1 \cdot k'_1}{14} \right) \cdot \delta(v'_1, v'_1) \right. \\
 &\quad + \left(a'_{22} - \frac{k'_2 \cdot k'_2}{14} \right) \cdot \delta(v'_2, v'_2) \\
 &\quad + \left(a'_{33} - \frac{k'_3 \cdot k'_3}{14} \right) \cdot \delta(v'_3, v'_3) \\
 &\quad + \left(a'_{44} - \frac{k'_4 \cdot k'_4}{14} \right) \cdot \delta(v'_4, v'_4) \\
 &\quad + \left(a'_{55} - \frac{k'_5 \cdot k'_5}{14} \right) \cdot \delta(v'_5, v'_5) \\
 &\quad \left. + \left(a'_{66} - \frac{k'_6 \cdot k'_6}{14} \right) \cdot \delta(v'_6, v'_6) \right] \\
 &= \frac{1}{14} \cdot \left[\left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right. \\
 &\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
 &\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
 &= \frac{1}{14} \cdot \left[\left(\frac{-4}{14} \right) + \left(\frac{-9}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-9}{14} \right) + \left(\frac{-4}{14} \right) \right] \\
 &= \frac{1}{14} \cdot \left[\left(2 \cdot \frac{-9}{14} \right) + \left(4 \cdot \frac{-4}{14} \right) \right] \\
 &= \frac{1}{14} \cdot \left[\left(\frac{-9}{7} \right) + \left(\frac{-8}{7} \right) \right] = \frac{1}{14} \cdot \left(\frac{-17}{7} \right) = \frac{-17}{14 \cdot 7} = \frac{-17}{98} = -0,173
 \end{aligned} \tag{3.30}$$

- *Modularidad Q'' . Se calcula según la Ecuación 3.31, donde $A_3 = A''_G = [a''_{ij}]$ es la matriz*

de adyacencia, k_i'' representa el grado del nodo $v_i'' \in V$ (es decir, $k_i'' = \delta(v_i'')$).

$$\begin{aligned}
Q'' &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij}'' - \frac{k_i'' \cdot k_j''}{2 \cdot 7} \right) \cdot \delta(v_i'', v_j'') \\
&= \frac{1}{14} \cdot \left[\left(a_{11}'' - \frac{k_1'' \cdot k_1''}{14} \right) \cdot \delta(v_1'', v_1'') \right. \\
&\quad + \left(a_{22}'' - \frac{k_2'' \cdot k_2''}{14} \right) \cdot \delta(v_2'', v_2'') \\
&\quad + \left(a_{33}'' - \frac{k_3'' \cdot k_3''}{14} \right) \cdot \delta(v_3'', v_3'') \\
&\quad + \left(a_{44}'' - \frac{k_4'' \cdot k_4''}{14} \right) \cdot \delta(v_4'', v_4'') \\
&\quad + \left(a_{55}'' - \frac{k_5'' \cdot k_5''}{14} \right) \cdot \delta(v_5'', v_5'') \\
&\quad \left. + \left(a_{66}'' - \frac{k_6'' \cdot k_6''}{14} \right) \cdot \delta(v_6'', v_6'') \right] \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right. \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right] \\
&= \frac{1}{14} \cdot \left[\left(\frac{-4}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-9}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-4}{14} \right) + \left(\frac{-9}{14} \right) \right] \\
&= \frac{1}{14} \cdot \left[\left(2 \cdot \frac{-9}{14} \right) + \left(4 \cdot \frac{-4}{14} \right) \right] \\
&= \frac{1}{14} \cdot \left[\left(\frac{-9}{7} \right) + \left(\frac{-8}{7} \right) \right] = \frac{1}{14} \cdot \left(\frac{-17}{7} \right) = \frac{-17}{14 \cdot 7} = \frac{-17}{98} = -0,173
\end{aligned} \tag{3.31}$$

Se supone ahora la Partición 4 del Anexo E. A continuación se calcula el índice Surprise para cada uno de los grafos.

- **Índice H.** Se calcula según la Ecuación 3.32, donde F es el número máximo posible de aristas en G , es decir, $\binom{6}{2} = \frac{6!}{2! \cdot 4!} = \frac{6 \cdot 5}{2} = 15$, $M = 1$ es el número máximo posible de aristas intracomunitarias para esa partición, y $p = 1$ es el número real de aristas intraco-

munitarias.

$$\begin{aligned}
 H &= -\ln \sum_{j=1}^{\min(1,7)} \frac{\binom{1}{j} \cdot \binom{15-1}{7-j}}{\binom{15}{7}} = -\ln \sum_{j=1}^1 \frac{\binom{1}{j} \cdot \binom{15-1}{7-j}}{\binom{15}{7}} = -\ln \frac{\binom{1}{1} \cdot \binom{15-1}{7-1}}{\binom{15}{7}} \\
 &= -\ln \frac{\binom{14}{6}}{\binom{15}{7}} = -\ln \frac{14!}{6! \cdot 8!} = -\ln \frac{14! \cdot 7! \cdot 8!}{15! \cdot 6! \cdot 8!} = -\ln \frac{7}{15} = -(-0,762) = 0,762
 \end{aligned} \tag{3.32}$$

- *Índice H' . Se calcula según la Ecuación 3.33, donde $F' = 15$ es el número máximo posible de aristas en G' , $M' = 1$ es el número máximo posible de aristas intracomunitarias para esa partición, y $p' = 0$ es el número real de aristas intracomunitarias. En la igualdad de la Ecuación 3.33 se usa fórmula recursiva aditiva del coeficiente binomial.*

$$\begin{aligned}
 H &= -\ln \sum_{j=0}^{\min(1,7)} \frac{\binom{1}{j} \cdot \binom{15-1}{7-j}}{\binom{15}{7}} = -\ln \sum_{j=0}^1 \frac{\binom{1}{j} \cdot \binom{15-1}{7-j}}{\binom{15}{7}} = -\ln \left[\frac{\binom{1}{0} \cdot \binom{15-1}{7-0}}{\binom{15}{7}} + \frac{\binom{1}{1} \cdot \binom{15-1}{7-1}}{\binom{15}{7}} \right] \\
 &= -\ln \left[\frac{\binom{14}{7}}{\binom{15}{7}} + \frac{\binom{14}{6}}{\binom{15}{7}} \right] = -\ln \left[\frac{\binom{14}{7} + \binom{14}{6}}{\binom{15}{7}} \right] = -\ln \left[\frac{\binom{15}{7}}{\binom{15}{7}} \right] = -\ln(1) = 0
 \end{aligned} \tag{3.33}$$

- *Índice H'' . Dado que el valor de las variables de la ecuación que define esta medida coincide en G' y G'' para la Partición 4, se calcula según la Ecuación 3.33.*

3.6.2. Algoritmos de división

Existen muchos algoritmos para dividir una red en comunidades (o *clusters*); los más sencillos se listan a continuación.

UVCluster

Se selecciona un nodo al azar, y se elige un nodo adyacente; si no existe, este nodo es aislado, forma una comunidad y se retira de la red, para proseguir de la misma forma con el resto de nodos. Si existe, el nodo y su adyacente forman un *clique*; se busca ahora un nodo que sea adyacente a ambos nodos, y se añade, buscando entonces un nodo que sea adyacente a estos tres, y así sucesivamente hasta tener que parar; en ese caso, se tiene otra comunidad, que se retira de la red junto con las aristas que sean incidentes a alguno de los nodos eliminados, y se continúa con el resto de nodos.

El algoritmo termina cuando la red queda vacía, al haber retirado todas las comunidades posibles.

Nótese que para este algoritmo todas las comunidades son *cliques*. Se puede cambiar la condición de *clique* con la de tener una densidad por encima de un cierto umbral que se fija al principio del algoritmo.

Ejemplo 40. *Se aplica el algoritmo UVCluster sobre los grafos de la Figura 3.1.*

1. Se selecciona el nodo v_1 y se elige uno de sus adyacentes, v_2 . Después, se añade v_3 , que es adyacente a v_1 y v_2 . Estos 3 nodos forman un clique, y no existe un nodo en G que sea adyacente a todos a la vez. Por tanto, la primera comunidad encontrada es $\{v_1, v_2, v_3\}$, que se retira de la red junto con las aristas incidentes a alguno de los nodos de la misma, es decir, $\{e_1, e_2, e_3, e_7\}$. Por llevar un orden, se selecciona el nodo v_4 , y se elige uno de sus adyacentes, v_5 . Después, se añade v_6 , que es adyacente a v_4 y v_5 . Estos 3 nodos forman un clique, y no existen más nodos sin comunidad en G . Por tanto, la segunda comunidad encontrada es $\{v_4, v_5, v_6\}$, que se retira de la red junto con las aristas incidentes a alguno de los nodos de la misma, es decir, $\{e_4, e_5, e_6, e_7\}$.
2. Se selecciona el nodo v'_1 y se elige uno de sus adyacentes, v'_2 . Después, se añade v'_3 , que es adyacente a v'_1 y v'_2 . Estos 3 nodos forman un clique, y no existe un nodo en G' que sea adyacente a todos a la vez. Por tanto, la primera comunidad encontrada es $\{v'_1, v'_2, v'_3\}$, que se retira de la red junto con las aristas incidentes a alguno de los nodos de la misma, es decir, $\{e'_1, e'_2, e'_3, e'_7\}$. Por llevar un orden, se selecciona el nodo v'_4 , y se elige uno de sus adyacentes, v'_5 . Después, se añade v'_6 , que es adyacente a v'_4 y v'_5 . Estos 3 nodos forman un clique, y no existen más nodos sin comunidad en G' . Por tanto, la segunda comunidad encontrada es $\{v'_4, v'_5, v'_6\}$, que se retira de la red junto con las aristas incidentes a alguno de los nodos de la misma, es decir, $\{e'_4, e'_5, e'_6, e'_7\}$.
3. Se selecciona el nodo v''_1 y se elige uno de sus adyacentes, v''_2 . Después, se añade v''_3 , que es adyacente a v''_1 y v''_2 . Estos 3 nodos forman un clique, y no existe un nodo en G'' que sea adyacente a todos a la vez. Por tanto, la primera comunidad encontrada es $\{v''_1, v''_2, v''_3\}$, que se retira de la red junto con las aristas incidentes a alguno de los nodos de la misma, es decir, $\{e''_1, e''_2, e''_3, e''_7\}$. Por llevar un orden, se selecciona el nodo v''_4 , y se elige uno de sus adyacentes, v''_5 . Después, se añade v''_6 , que es adyacente a v''_4 y v''_5 . Estos 3 nodos forman un clique, y no existen más nodos sin comunidad en G'' . Por tanto, la segunda comunidad encontrada es $\{v''_4, v''_5, v''_6\}$, que se retira de la red junto con las aristas incidentes a alguno de los nodos de la misma, es decir, $\{e''_4, e''_5, e''_6, e''_7\}$.

SCluster

Se selecciona un nodo al azar y se forma un *cluster* con este nodo y todos sus adyacentes, retirándose de la red y procediendo de igual forma con el resto de nodos de la red.

En este caso, los *clusters* no necesariamente son *cliques* y el algoritmo es más rápido.

En lugar de elegir los nodos al azar, se pueden elegir mediante algún criterio; por ejemplo, los de grado mayor, que se supone que tienen más influencia.

Ejemplo 41. Se aplica el algoritmo SCluster sobre los grafos de la Figura 3.1.

1. Dado que todos los nodos de G tienen grado 2 excepto v_1 y v_4 que tienen grado 3, se selecciona uno de estos al azar: el nodo v_1 . Por tanto, la primera comunidad está formada por v_1 y todos sus adyacentes, es decir, $\{v_1, v_2, v_3, v_4\}$, que se retiran de la red. Se selecciona el nodo v_5 y todos sus adyacentes que aún no pertenecen a ninguna comunidad, es decir, v_6 . Por tanto, la segunda comunidad es $\{v_5, v_6\}$.
2. Dado que todos los nodos de G' tienen grado 2 excepto v'_2 y v'_5 que tienen grado 3, se selecciona uno de estos al azar: el nodo v'_2 . Por tanto, la primera comunidad está formada

por v'_2 y todos sus adyacentes, es decir, $\{v'_1, v'_2, v'_3, v'_5\}$, que se retiran de la red. Se selecciona el nodo v'_4 y todos sus adyacentes que aún no pertenecen a ninguna comunidad, es decir, v'_6 . Por tanto, la segunda comunidad es $\{v'_4, v'_6\}$.

3. Dado que todos los nodos de G'' tienen grado 2 excepto v''_3 y v''_6 que tienen grado 3, se selecciona uno de estos al azar: el nodo v''_3 . Por tanto, la primera comunidad está formada por v''_3 y todos sus adyacentes, es decir, $\{v''_1, v''_2, v''_3, v''_6\}$, que se retiran de la red. Se selecciona el nodo v''_4 y todos sus adyacentes que aún no pertenecen a ninguna comunidad, es decir, v''_5 . Por tanto, la segunda comunidad es $\{v''_4, v''_5\}$.

Algoritmo de Blondel

Este algoritmo tiene como objetivo maximizar la modularidad Q . Análogamente se podría intentar optimizar el índice H .

Inicialmente, cada nodo constituye una comunidad y, en cada iteración, se recorren (en principio, al azar) todos los nodos, comprobándose si al cambiar la comunidad del nodo correspondiente por la comunidad de alguno de sus vecinos, la modularidad de la partición aumenta; si es así, el nodo se mueve a la comunidad en la que la modularidad aumenta más y, en caso contrario, se deja en la comunidad que está.

El algoritmo termina si, en una iteración, ningún nodo cambia de comunidad.

Ejemplo 42. Se aplica el algoritmo de Blondel sobre el primer grafo de la Figura 3.1. Se parte de los datos de modularidad del Ejemplo 39. Cuando cada nodo constituye una comunidad, el valor de la modularidad Q de G es $-0,173$. Finalmente, se obtiene que la mejor partición es la intuitiva, la Partición 14, es decir, la formada por las dos comunidades que se listan a continuación.

- Comunidad 1: $\{v_1, v_2, v_3\}$
- Comunidad 2: $\{v_4, v_5, v_6\}$

El valor de la modularidad Q de G en este caso es $0,357$. El paso a paso y los detalles se pueden consultar en el Anexo E.

Nota 25. Estos algoritmos se pueden iterar varias veces, ya que tienen una componente aleatoria, y quedarse con la partición que obtenga una mejor puntuación, respecto de Q o H . Incluso podrían ejecutarse en paralelo y quedarse con el algoritmo que obtenga una mejor puntuación.

Capítulo 4

Redes complejas

Las redes complejas abarcan una amplia variedad de sistemas que van desde redes sociales hasta redes biológicas y de transporte. En este capítulo se describen diferentes modelos de redes complejas y sus propiedades estructurales y dinámicas. Se hace hincapié en la importancia de los distintos tipos de topologías, como la de Poisson y la libre de escala, y su impacto en la conectividad de la red. Asimismo, se abordan fenómenos como el envejecimiento de nodos y la formación de islas gigantes, que son esenciales para entender la evolución y el comportamiento global de estas redes [3].

4.1. Introducción

El interés actual en las redes complejas radica en que dichas redes abundan en la naturaleza, son parte de la vida diaria de las personas y se presentan a diferentes niveles de organización. Las redes complejas son ubicuas, están por todos lados. Por ejemplo, algunas redes biológicas que se encuentran en el nivel microscópico son las redes de regulación genética, las redes de proteínas, las redes neuronales y las redes metabólicas. Por otro lado, a un nivel de organización mucho mayor, se encuentran las redes de comunicación e informáticas (la red de internet, la red Red de Alcance Mundial o *World Wide Web* (WWW), las redes telefónicas, etc.), las redes sociales (redes de amistades, contactos sexuales, colaboradores científicos, propagación de enfermedades, etc.) y las redes ecológicas (redes de interacciones tróficas en un ecosistema). Incluso se estudia la red de super héroes en el Universo de Marvel, donde dos super héroes están conectados si han aparecido por lo menos una vez en el mismo cómic, siendo el hombre araña el super héroe más popular, es decir, con la mayor conectividad [6].

Es un hecho sobresaliente el que todas estas redes, tan diferentes en naturaleza y en tamaño, tengan muchas propiedades estructurales similares. Este hecho, tan simple como sorprendente, hace posible que se puedan formular modelos matemáticos para entender y explicar las propiedades estructurales (y, en algunos casos, también las propiedades dinámicas) de las redes complejas.

Definición 56. *Un sistema complejo se caracteriza por estar compuesto de muchas partes que interactúan entre sí. En él, cada parte tiene su propia estructura interna y está encargada de llevar a cabo una función específica, de modo que lo que ocurra a una parte del sistema afecta de manera altamente no lineal a todo el sistema. Además, presentan comportamientos emergentes, de tal manera que el todo no es la simple suma de sus partes.*

Ejemplo 43. *Un ejemplo típico de sistema complejo es la célula. Evidentemente, está compuesta de muchas partes (ribosomas, mitocondrias, núcleo, membrana, retículo endoplasmático, Ácido Desoxirribonucleico (ADN), Ácido Ribonucleico (ARN), etc.), y cada una de ellas se encarga de realizar una función específica dentro de la célula. Las partes de la célula responden de forma no lineal ante perturbaciones externas. Por ejemplo, algunas veces una mutación en el ADN no tiene ningún efecto en la célula, mientras que otras veces una sólo mutación puede ser fatal¹. Además, la célula presenta comportamientos emergentes que no pueden explicarse en términos de las propiedades de sus partes individuales. Así, se puede hablar de una célula enferma, pero no se puede decir que un ribosoma o una proteína estén enfermos. La enfermedad es una propiedad que emerge como resultado de la organización colectiva de todos los constituyentes de la célula.*

4.2. Modelos

Los grafos son ampliamente utilizados para modelar diferentes tipos de redes y estructuras en diversos campos. En esta subsección, se describen ejemplos específicos de cómo los grafos se aplican en contextos sociales, de comunicación, de información, de diseño de *software*, de transporte, biológicos y en torneos deportivos [5] [13].

4.2.1. Redes sociales

En las redes sociales (Capítulo 3), los grafos modelan las relaciones entre individuos u organizaciones. Los vértices representan a las personas o entidades, y las aristas representan las relaciones entre ellos.

A continuación, se listan ejemplos específicos.

- *Grafos simples de relación y amistad.* Conectan a personas que se conocen.
- *Digrafos de influencia.* Modelan la influencia de unas personas sobre otras, donde una flecha de A a B significa que A influye en B .
- *Grafos simples de colaboración.* Conectan a personas que trabajan juntas en un proyecto.
- *Grafos de relaciones familiares.* Conectan a personas que son familiares cercanos.
- *Grafos de contagio de enfermedades.* Dos personas están conectadas si una contagia a la otra.
- *Grafos de relaciones sexuales.* Conectan a personas que tienen una relación sexual.
- *Grafos de coautoría científica.* Conectan a científicos que son coautores de un artículo.
- *Grafos de actores.* Conectan a actores que aparecen en la misma película.

¹La anemia falciforme, por ejemplo, es una enfermedad de los glóbulos rojos de la sangre que se origina por una sólo mutación en uno de los más de 600 aminoácidos que conforman a la proteína β -globina encargada de capturar oxígeno.

4.2.2. Redes de comunicación

En las redes de comunicación, los grafos modelan las conexiones entre dispositivos. Los vértices representan los dispositivos y las aristas los enlaces de comunicación.

Un ejemplo son los *multidigrafos de llamadas*, que representan redes telefónicas, donde cada número es un vértice y cada llamada una arista dirigida.

4.2.3. Redes de información

Los grafos modelan redes que vinculan diferentes tipos de información.

A continuación, se listan ejemplos específicos.

- *Grafos simples de redes de ordenadores*. Conectan máquinas con una arista si hay una interconexión entre ellas.
- *Digrafo de la Web (WWW)*. Modela la web, donde una arista va de la página A a la página B si hay un enlace en A que apunta a B.
- *Digrafos de citas*. Representan citaciones en documentos, con una arista desde un documento a otro si el primero cita al segundo.
- *Grafos de palabras*. Dos palabras están conectadas si en el diccionario una aparece en la definición de la otra.
- *Grafos de sinónimos*. Dos palabras están conectadas si son sinónimos.
- *Grafos de Internet*. Dos computadoras están conectadas si hay un cable que las conecta.

4.2.4. Diseño de *software*

En el diseño de *software*, los grafos modelan las relaciones entre módulos de un programa. Es crucial para el diseño, pruebas y mantenimiento del software.

Un ejemplo es el *digrafo de precedencia de tareas en un proyecto*, donde los nodos representan las tareas y las flechas indican qué tareas preceden a otras.

4.2.5. Redes de transporte

Los grafos también son útiles en el modelado de redes de transporte.

A continuación, se listan ejemplos específicos.

- *Multidigrafos de rutas aéreas*. Modelan los vuelos de una aerolínea, considerando múltiples vuelos entre los mismos aeropuertos en un día.
- *Digrafos de tráfico*. Representan redes viales, donde las flechas indican el sentido de circulación en las calles.
- *Grafos simples de carreteras*. Permiten calcular el camino más corto entre dos ciudades.

4.2.6. Redes biológicas

En biología, los grafos modelan redes complejas.

A continuación, se listan ejemplos específicos.

- *Grafos simples de interacción de proteínas.* Representan la interacción entre proteínas en una célula viva, esenciales para funciones biológicas.
- *Grafos de redes genéticas.* Conectan genes si uno regula la expresión del otro.
- *Grafos de redes neuronales.* Conectan neuronas si existe una conexión sináptica entre ellas.
- *Grafos de redes ecológicas.* Conectan especies si una se alimenta de la otra.

4.2.7. Torneos

Los grafos modelan torneos deportivos, donde los vértices representan equipos y las aristas los enfrentamientos.

A continuación, se listan ejemplos específicos.

- *Grafos simples completos.* Representan torneos de "todos contra todos", donde cada equipo juega contra todos los demás una vez.
- *Grafos de torneos de eliminación.* Modelan torneos donde los equipos se eliminan progresivamente.

4.3. Estudio

El estudio general de las redes complejas puede dividirse en dos campos diferentes y complementarios, que se observan en la Tabla 4.1 y son los que dan título a las subsecciones de la sección actual.

Estructura	Dinámica
Distribución de conexiones	Sincronización
Coefficiente de agrupamiento	Transiciones de fase
Longitud promedio	Aprendizaje
Componente gigante	Procesos difusivos

Tabla 4.1: Estudio de las redes complejas.

4.3.1. Estructura

En el primer campo de estudio interesa determinar las propiedades estructurales (o topológicas) de la red, es decir, las propiedades que expresan cómo están conectados los nodos entre sí. Algunas de las propiedades más importantes que determinan la estructura (o topología) de una red se definen a continuación.

Definición 57. La *distribución de conexiones* (o vecinos) la red, $\mathbf{P}(\mathbf{k})$, es la probabilidad de que un nodo escogido al azar tenga k conexiones (o vecinos).

Definición 58. El *coeficiente de agregación* de la red, \mathbf{C} , es la probabilidad de que dos nodos conectados directamente a un tercer nodo, estén conectados entre sí.

Definición 59. La *longitud mínima* entre dos nodos v_i y v_j , \mathbf{L}_{ij} , es el número mínimo de aristas que separan dos nodos v_i y v_j de la red. Coincide con la Definición 47 del Capítulo 2, $d(v_i, v_j)$.

Definición 60. La *longitud promedio* de la red, \mathbf{L} , es el promedio de las longitudes mínimas L_{ij} entre todas las posibles parejas de nodos (v_i, v_j) de la red. Coincide con la Definición 50 del Capítulo 2.

Definición 61. La *distribución de tamaños de las islas*, $\mathbf{P}(\mathbf{s})$, es la probabilidad de que una isla esté compuesta por s nodos.

Definición 62. El *tamaño de la isla más grande* de la red se denota por \mathbf{S}_∞ .

4.3.2. Dinámica

Además de estar conectados, los nodos de una red también interactúan, y las interacciones pueden dar lugar a fenómenos dinámicos muy interesantes. Por lo tanto, además de estudiar las propiedades estructurales de una red, también es importante estudiar sus propiedades dinámicas una vez que se sabe de qué manera interactúan los nodos.

Ejemplo 44. Las enfermedades en una sociedad no son estáticas, sino que se propagan por toda la población dando lugar a epidemias.

Ejemplo 45. Las neuronas en el cerebro están conectadas físicamente unas con otras por medio de las uniones entre dendritas y axones, llamadas sinapsis. A través de dichas uniones, las neuronas se transmiten señales eléctricas que se propagan por todo el cerebro dando lugar a una serie de fenómenos dinámicos muy interesantes; entre ellos destacan el reconocimiento de imágenes y sonido, la motricidad de los músculos, el lenguaje, el pensamiento y, finalmente, la consciencia.

Ejemplo 46. Otros ejemplos son la propagación de virus informáticos en la red Internet, o la comunicación entre los peces que da lugar a bancos moviéndose todos en la misma dirección.

Existen tantos fenómenos dinámicos en redes complejas como interacciones físicas, químicas, informáticas, o sociales se puedan imaginar, y cada día aparecen más artículos en la literatura científica donde se estudian nuevos procesos dinámicos sobre redes complejas.

Este trabajo se enfoca más en el estudio de las propiedades estructurales de las redes complejas, dejando un poco de lado sus propiedades dinámicas.

4.4. Distribución de vecinos

Existen tres tipos de distribuciones $P(k)$ importantes, cada una de las cuales determina una estructura o topología² diferente. Se listan a continuación.

²En el contexto de las redes complejas, la palabra “topología” no significa lo mismo que en Análisis Funcional, sino que es sinónimo de “estructura” o “arquitectura”.

- Topología de Poisson. La distribución de conexiones de la red sigue la función de la Ecuación 4.1.

$$P(k) = e^{-z} \cdot \frac{z^k}{k!} \quad (4.1)$$

- Topología exponencial. La distribución de conexiones de la red sigue la función de la Ecuación 4.2.

$$P(k) = C \cdot e^{-\alpha \cdot k} \quad (4.2)$$

- Topología libre de escala ³. La distribución de conexiones de la red sigue la función de la Ecuación 4.3.

$$P(k) = C \cdot k^{-\gamma} \quad (4.3)$$

Las redes con topología de Poisson son importantes principalmente por razones históricas, ya que son las primeras que se analizan matemáticamente. Este análisis lo llevan a cabo los matemáticos húngaros Paul Erdős y Alfréd Rényi en la década de 1950. Ellos también reportan más adelante la primera transición de fase topológica observada en redes con topología de Poisson (Anexo C). Por lo tanto, a estas redes también se les conoce como redes tipo Erdős-Rényi. Sin embargo, a pesar de su importancia histórica, las redes con topología de Poisson están lejos de ser una representación realista de las redes reales observadas en la naturaleza. No es sino hasta 1998 que se comienza el estudio sistemático de las propiedades topológicas de las redes complejas reales. En este estudio participan, principalmente y de forma independiente, Albert-László Barabási, Ricard Solé y Mark J. Newman, quienes encuentran que la topología exponencial aparece algunas veces en las redes reales. Pero el resultado más sorprendente de sus estudios es la ubicuidad de las redes con topología libre de escala. En la Tabla 4.2 se listan algunas de las redes con topología libre de escala encontradas en los últimos 10 años.

De la topología libre de escala sorprenden dos cosas. Por un lado, es fascinante que se encuentre en redes tan diferentes y de tan gran variedad como las listadas en la Tabla 4.2. Esta omnipresencia sugiere que la formación de redes tan diferentes puede estar gobernada por la misma ley fundamental. Por otro lado, su aparición es inesperada. En la naturaleza existen muchos procesos aleatorios que generan distribuciones de Poisson o distribuciones exponenciales, pero existen muy pocos procesos conocidos que generan distribuciones libres de escala como la dada en la Ecuación 4.3. De hecho, el trabajo de Erdős y Rényi demuestra que las redes que se construyen añadiendo nuevos nodos y conexiones al azar presentan topologías de Poisson o exponenciales. Y ésta es precisamente la paradoja, que redes que se forman aleatoriamente añadiendo nuevos nodos y nuevas conexiones a lo largo del tiempo, como las de la Tabla 4.2, no presenten topologías de Poisson o exponenciales como predice la pareja de húngaros.

La red de Poisson, que se parece mucho a las redes con topología exponencial, se ve más aleatoria y más homogénea que la red libre de escala. En las redes de Poisson todos los nodos tienen más o menos el mismo número de conexiones. Algunos nodos están más conectados que otros, pero, en promedio, todos tienen la misma conectividad; es decir, las conexiones en una red de Poisson están distribuidas homogéneamente entre sus nodos. Por el contrario, la característica

³A las distribuciones libres de escala también se les llama distribuciones de potencia.

Red	Número de nodos	Número de conexiones	γ_i
dominio www.nd.edu	325729	1469680	2,1
Páginas de WWW encontradas por Altavista	2711×10^9	2130×10^9	2,1
Dominios en la WWW	$2,60 \times 10^5$	-	1,9
Nivel de inter-dominio de la Internet	4389	8256	2,2
Sistemas autónomos de la Internet	6374	13641	2,2
Nivel de ruteador de la Internet	150000	200000	2,3
Citas en la base de datos ISI	783339	6716198	3,0
Citatas de la revista Phys. Rev. D	24296	351872	2,3
Red de colaboraciones de actores de Hollywood	212250	61085555	2,3
Red de colaboradores en las revistas <i>Medline</i>	1388989	1028×10^7	2,5
Colaboradores en las revistas de matemáticas	70975	132×10^5	2,1
Colaboradores en las revistas de neurociencias	209293	121×10^6	2,4
Red de interacciones metabólicas en (<i>E. coli</i>)	778	$\sim 1500 - 3000$	2,2
Red de interacciones protéicas en levadura	1870	2240	2,5
Co-ocurrencia de palabras	470000	17000000	2,7
Red de palabras sinónimas	22311	—	2,8
Circuitos digitales	2×10^4	4×10^4	3,0
Llamadas telefónicas	47×10^6	8×10^7	2,1
Red de interacciones sexuales en humanos	2810	—	3,4
Redes alimenticias (interacciones tróficas)	154	405	1,0

Tabla 4.2: Algunas de las redes libres de escala que se encuentran en la naturaleza. Sólo se muestra el exponente de entrada para las redes dirigidas. Los cuadros con líneas “—” indican que no se tiene el dato correspondiente al momento de escribir estas notas.

más importante de las redes libres de escala es su alta heterogeneidad, ya que existen nodos con muy pocas conexiones, nodos medianamente conectados y nodos extremadamente conectados. Los nodos altamente conectados se denominan los núcleos o centros de la red ⁴. En una red libre de escala no se puede decir que todos los nodos tienen aproximadamente la misma conectividad.

4.4.1. Topología de Poisson

Sea un conjunto de N nodos distribuidos aleatoriamente e inicialmente desconectados. Al tiempo $t = 0$ se escoge aleatoriamente una pareja de nodos y se unen con una arista. Después de enlazar esta pareja, se escoge aleatoriamente otra para unir. Se pueden escoger nodos que están conectados con otros nodos, pero no es válido enlazar más de una vez la misma pareja de nodos. Se repite este proceso sucesivamente M veces, escogiendo aleatoriamente una pareja de

⁴En inglés, se les llama *hubs*.

nodos cada vez. Al final del proceso se establecen M enlaces entre M parejas diferentes de nodos, generando así una red.

Intuitivamente, es claro que si M (el número total de enlaces) es pequeño comparado con N (el número total de nodos), entonces la red resultante está desmembrada en varias islas pequeñas. Dentro de cada isla los nodos están enlazados entre sí, pero están desconectados de las otras islas. Sin embargo, si M es grande comparado con N , casi la totalidad de los nodos termina conectada. Probablemente hay islas muy pequeñas desconectadas de la red principal, pero seguramente la gran mayoría de nodos forma parte de una isla principal: la isla gigante (Sección 4.6).

Nota 26. *Este proceso no permite la formación de bucles en la red resultante.*

Teorema 3. *La red que resulta de unir M parejas en un conjunto total de N nodos tiene una distribución de conexiones $P(k)$ de Poisson (Ecuación 4.1).*

Demostración. Para comenzar, el número total N_p de parejas (grupos de 2) que se pueden formar en un conjunto de N nodos es $C_{N,2}$ (Sección 2.3.1) y se observa en la Ecuación 4.4.

$$N_p = C_{N,2} = \binom{N}{2} = \frac{N!}{2! \cdot (N-2)!} = \frac{N \cdot (N-1) \cdot \cancel{(N-2)} \cdot \cancel{(N-3)} \cdot \dots \cdot \cancel{1}}{2 \cdot \cancel{(N-2)} \cdot \cancel{(N-3)} \cdot \dots \cdot \cancel{1}} = \frac{N \cdot (N-1)}{2} \quad (4.4)$$

Como se enlazan M parejas de nodos, la **probabilidad** p_e de que una pareja arbitraria seleccionada al azar esté **enlazada** se observa en la Ecuación 4.5.

$$p_e = \frac{M}{N_p} = \frac{M}{\frac{N \cdot (N-1)}{2}} = \frac{2 \cdot M}{N \cdot (N-1)} \quad (4.5)$$

Ahora se pone el foco sobre un nodo particular v_j de la red, escogido al azar. El número total de parejas que pueden contener a v_j es $N-1$, ya que v_j se puede enlazar con los $N-1$ nodos restantes de la red. Sin embargo, en los M enlaces que se llevan a cabo, no necesariamente se escoge al nodo v_j todas las veces posibles. Se supone entonces que, de las M parejas que se escogen, el nodo v_j está solamente en k de ellas. La probabilidad de que v_j esté contenido en k parejas de las M posibles se observa en la Ecuación 4.6.

$$P(k) = \binom{N-1}{k} \cdot (p_e)^k \cdot (1-p_e)^{N-1} \quad (4.6)$$

Esta es una distribución binomial (Sección 2.3.2) para N y M finitas. Pero si se considera que la red es muy grande y se toma el límite $N \rightarrow \infty$ y $M \rightarrow \infty$ de tal forma que la cantidad $z = \frac{2 \cdot M}{N}$ permanece finita, entonces la distribución se transforma en la distribución de Poisson (Sección 2.3.2) con promedio z de la Ecuación 4.7.

$$P(k) = e^{-z} \cdot \frac{z^k}{k!} \quad (4.7)$$

□

Proposición 6. *La distribución binomial de la Ecuación 4.6 se transforma en la distribución de Poisson con promedio z de la Ecuación 4.7 cuando se toma el límite $N \rightarrow \infty$ y $M \rightarrow \infty$ de tal forma que la cantidad $z = \frac{2 \cdot M}{N}$ permanece finita.*

Demostración. Utilizando la fórmula de Stirling (Ecuación 2.7) sobre $(N-1)!$ y $(N-1-k)!$ y simplificando, el coeficiente binomial $\binom{N-1}{k}$ que aparece en la Ecuación 4.6 puede escribirse, para N grande, como se observa en la Ecuación 4.8.

$$\begin{aligned} \binom{N-1}{k} &= \frac{(N-1)!}{k! \cdot (N-1-k)!} \\ &\approx \frac{e^{-(N-1)} \cdot (N-1)^{N-1}}{k! \cdot e^{-(N-1-k)} \cdot (N-1-k)^{N-1-k}} \\ &= \frac{e^{-k} \cdot (N-1-k)^k}{k!} \cdot \left(\frac{N-1}{N-1-k}\right)^{N-1} \end{aligned} \quad (4.8)$$

Por lo tanto, la distribución binomial de la Ecuación 4.6 queda, para N grande, como se observa en la Ecuación 4.9.

$$P(k) = \frac{e^{-k} \cdot (N-1-k)^k}{k!} \cdot \left(\frac{N-1}{N-1-k}\right)^{N-1} \cdot (p_e)^k \cdot (1-p_e)^{N-1} \quad (4.9)$$

Para continuar, la cantidad $z = \frac{2 \cdot M}{N}$ es la **conectividad promedio** de cada nodo en la red. En términos de esta cantidad, la probabilidad de la Ecuación 4.5 puede escribirse como $p_e = \frac{z}{N-1}$, con lo cual la Ecuación 4.9 queda como se observa en la Ecuación 4.10.

$$P(k) = \frac{e^{-k} \cdot (N-1-k)^k}{k!} \cdot \left(\frac{N-1}{N-1-k}\right)^{N-1} \cdot \left(\frac{z}{N-1}\right)^k \cdot \left(1 - \frac{z}{N-1}\right)^{N-1} \quad (4.10)$$

Después, se arregla la multiplicación del numerador del primer término y el denominador del tercero como se observa en la Ecuación 4.11.

$$\left(\frac{N-1-k}{N-1}\right)^k = \left(1 - \frac{k}{N-1}\right)^k \quad (4.11)$$

Se arregla también la multiplicación de los términos segundo y cuarto, utilizando en última instancia la Ecuación 4.11, como se observa en la Ecuación 4.12.

$$\begin{aligned} \left(\frac{N-1}{N-1-k}\right)^{N-1} \cdot \left(1 - \frac{z}{N-1}\right)^{N-1} &= \left(\frac{N-1}{N-1-k} \cdot \left(1 - \frac{z}{N-1}\right)\right)^{N-1} \\ &= \left(\frac{1 - \frac{z}{N-1}}{\frac{N-1-k}{N-1}}\right)^{N-1} = \left(\frac{1 - \frac{z}{N-1}}{1 - \frac{k}{N-1}}\right)^{N-1} \end{aligned} \quad (4.12)$$

A partir de los arreglos de la Ecuación 4.11 y la Ecuación 4.12, la Ecuación 4.10 puede escribirse como se observa en la Ecuación 4.13.

$$P(k) = \frac{e^{-k}}{k!} \cdot z^k \cdot \left(1 - \frac{k}{N-1}\right)^k \cdot \left(\frac{1 - \frac{z}{N-1}}{1 - \frac{k}{N-1}}\right)^{N-1} \quad (4.13)$$

Tomando el límite $N \rightarrow \infty$ y $M \rightarrow \infty$ de tal forma que la conectividad promedio permanece constante, y utilizando la definición de e^x (Ecuación 2.9) como se observa en la Ecuación 4.14, se obtiene finalmente la Ecuación 4.7 deseada.

$$P(k) \approx \frac{e^{-k}}{k!} \cdot z^k \cdot 1^k \cdot \frac{e^{-z}}{e^{-k}} = e^{-z} \cdot \frac{z^k}{k!} \quad (4.14)$$

□

Por lo tanto, para una red muy grande, el proceso de unir parejas de nodos escogidas al azar genera una distribución de conexiones de Poisson.

4.4.2. Crecimiento de redes

En la Sección 4.4.1 se supone una población fija de N nodos y un número fijo M de conexiones añadidas aleatoriamente. Sin embargo, en la realidad esto no ocurre, las redes no están fijas; por el contrario, evolucionan y crecen en el tiempo. Por lo tanto, es importante que los modelos de formación de redes incorporen el hecho de que se pueden añadir nuevos nodos y nuevas conexiones a la red, así como el hecho de que pueden eliminarse nodos y conexiones ya existentes.

En el modelo más simple de crecimiento de redes se añade un nuevo nodo en cada paso de tiempo, que puede conectarse con alguno de los nodos ya existentes. Siendo k_i la conectividad al tiempo t del i -ésimo nodo ya existente, éste puede ser seleccionado para la conexión con una probabilidad $\pi(k_i, t)$.

Los pasos de este modelo se listan a continuación y, los 4 primeros, se representan en la Figura 4.1.

- El proceso comienza al tiempo $t = 0$ con un único nodo inicial v_0 .
- Al tiempo $t = 1$ se añade un nuevo nodo v_1 , que se conecta con el único nodo ya existente v_0 con probabilidad $\pi(k_0, t) = 1$.
- Al tiempo $t = 2$ se añade el nodo v_2 , que se conecta con alguno de los nodos ya existentes v_0 y v_1 con la misma probabilidad $\pi(k_0, t) = \pi(k_1, t) = \frac{1}{2}$. En este momento, los nodos v_0 , v_1 y v_2 ya no tienen todos la misma conectividad: alguno de ellos tiene dos conexiones mientras que los otros dos nodos sólo una.
- Al tiempo $t = 3$ se añade el nodo v_3 , que se conecta con alguno de los nodos ya existentes v_0 , v_1 o v_2 con una probabilidad $\pi(k_0, t) = \pi(k_1, t) = \pi(k_2, t)$ que es función de sus conectividades k_0 , k_1 y k_2 , respectivamente.
- Continuando con este proceso, al tiempo $t + 1$ se añade el nodo v_{t+1} , que se conecta con alguno de los nodos ya existentes v_0, v_1, \dots, v_t , el cual es seleccionado con una probabilidad $\pi(k_i, t)$, siendo k_i la conectividad al tiempo t del nodo v_i .

Es claro que la probabilidad de que al tiempo t un nodo arbitrario de la red tenga k conexiones, denotada por $P(k, t)$, depende del tiempo. Sin embargo, si se continúa añadiendo nodos por un tiempo suficientemente largo, se espera que la función $P(k, t)$ alcance un estado estacionario independiente del tiempo. Esto no significa que la red alcance un estado estacionario; la red sigue

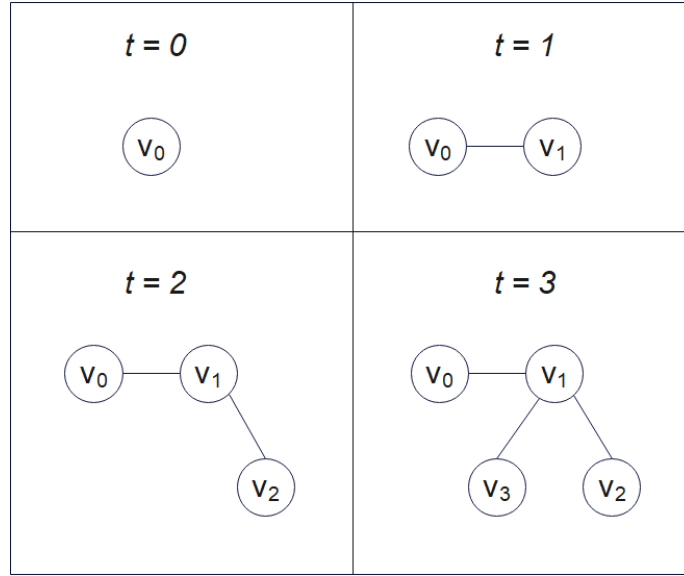


Figura 4.1: Crecimiento de redes.

creciendo mientras se continúa añadiendo nodos, es únicamente la distribución de conectividades $P(k, t)$ la que llega a un estado estacionario en el cual $P(k, t + 1) = P(k, t) = P(k)$.

Método de la ecuación maestra

En la Sección B.1 del Anexo B se da una breve introducción al planteamiento de la ecuación maestra, enfocada a aquellos que no conocen su formalismo. Se trata de una ecuación muy fácil de escribir pero, la mayoría de las veces, muy difícil de resolver.

La probabilidad de que el nodo v_n tenga k conexiones al tiempo t , denotada por $P(n, k, t)$, está asociada al nodo específico v_n . Sin embargo, se puede obtener la probabilidad $P(k, t)$ de que un nodo arbitrario tenga k conexiones al tiempo t a través de la media que se observa en la Ecuación 4.15, donde $N(t)$ es el número total de nodos de la red al tiempo t .

$$P(k, t) = \frac{1}{N(t)} \cdot \sum_{n=0}^t P(n, k, t) \Rightarrow \sum_{n=0}^t P(n, k, t) = P(k, t) \cdot N(t) \quad (4.15)$$

Despejando en la Ecuación 4.15 se obtiene la Ecuación 4.16, más cómoda.

$$\sum_{n=0}^t P(n, k, t) = P(k, t) \cdot N(t) \quad (4.16)$$

La ecuación que determina la evolución temporal de $P(n, k, t)$ se obtiene notando que, en cada paso de tiempo, hay dos contribuciones a dicha probabilidad. Ambas posibilidades se listan a continuación.

1. Al tiempo t el nodo v_n tiene $k_n = k - 1$ conexiones y es seleccionado para conectarse con el nuevo nodo añadido a la red con probabilidad $\pi(k - 1, t)$. Por lo tanto, al tiempo $t + 1$ el nodo v_n tiene $k_n = k$ conexiones.

2. Al tiempo t el nodo v_n ya tiene $k_n = k$ conexiones y no es seleccionado para conectarse con el nuevo nodo añadido a la red, lo cual ocurre con probabilidad $1 - \pi(k, t)$. Por lo tanto, al tiempo $t + 1$ el nodo v_n sigue teniendo $k_n = k$ conexiones.

Teniendo en cuenta estas dos contribuciones, la ecuación maestra que determina la evolución temporal de $P(n, k, t)$ es la Ecuación 4.17.

$$\begin{aligned}
 P(n, k, t + 1) = & P(n, k - 1, t) \quad k - 1 \text{ conexiones al tiempo } t \\
 & \cdot \pi(k - 1, t) \quad v_n \text{ es seleccionado} \\
 + & \\
 & P(n, k, t) \quad k \text{ conexiones al tiempo } t \\
 & \cdot (1 - \pi(k, t)) \quad v_n \text{ no es seleccionado}
 \end{aligned} \tag{4.17}$$

Como el nodo v_n “nace” al tiempo $t = n$ con una sólo conexión, la condición inicial para resolver la Ecuación 4.17 viene dada por la función *delta de Dirac*, definida en la Ecuación 4.18.

$$P(n, k, t)|_{t=n} = P(n, k, n) = \delta_{k,1} = \begin{cases} 0 & \text{si } k \neq 1 \\ 1 & \text{si } k = 1 \end{cases} \tag{4.18}$$

Sumando sobre n en la Ecuación 4.17 desde $n = 0$ hasta $n = t$ y tomando en consideración la Ecuación 4.16, se obtiene la Ecuación 4.19.

$$\begin{aligned}
 \sum_{n=0}^t P(n, k, t + 1) &= \sum_{n=0}^t P(n, k - 1, t) \cdot \pi(k - 1, t) + P(n, k, t) \cdot (1 - \pi(k, t)) \\
 &= \sum_{n=0}^t P(n, k - 1, t) \cdot \pi(k - 1, t) + \sum_{n=0}^t P(n, k, t) \cdot (1 - \pi(k, t)) \\
 &= \pi(k - 1, t) \cdot \sum_{n=0}^t P(n, k - 1, t) + (1 - \pi(k, t)) \cdot \sum_{n=0}^t P(n, k, t) \\
 &= \pi(k - 1, t) \cdot P(k - 1, t) \cdot N(t) + (1 - \pi(k, t)) \cdot P(k, t) \cdot N(t) \\
 &= N(t) \cdot (P(k - 1, t) \cdot \pi(k - 1, t) + P(k, t) \cdot (1 - \pi(k, t)))
 \end{aligned} \tag{4.19}$$

Ahora bien, como en cada paso de tiempo se añade un nuevo nodo (comenzando con el nodo v_0 al tiempo $t = 0$), entonces $N(t) = t + 1$. Por lo tanto, la Ecuación 4.19 puede escribirse como la Ecuación 4.20.

$$\sum_{n=0}^t P(n, k, t + 1) = (t + 1) \cdot (P(k - 1, t) \cdot \pi(k - 1, t) + P(k, t) \cdot (1 - \pi(k, t))) \tag{4.20}$$

En el lado izquierdo de la Ecuación 4.20 se puede completar la suma hasta $t + 1$ como en la Ecuación 4.21, donde se utiliza el hecho de que $P(t + 1, k, t + 1) = \delta_{k,1}$, tal y como lo establece la Ecuación 4.18.

$$\begin{aligned}
 \sum_{n=0}^t P(n, k, t+1) &= \sum_{n=0}^{t+1} P(n, k, t+1) - P(t+1, k, t+1) \\
 &= \sum_{n=0}^{t+1} P(n, k, t+1) - \delta_{k,1}
 \end{aligned} \tag{4.21}$$

También se utiliza la Ecuación 4.16 (evaluada en $t+1$) y el hecho de que $N(t+1) = t+2$ como se observa en la Ecuación 4.22.

$$\sum_{n=0}^t P(n, k, t+1) = N(t+1) \cdot P(k, t+1) - \delta_{k,1} = (t+2) \cdot P(k, t+1) - \delta_{k,1} \tag{4.22}$$

Sustituyendo el resultado de la Ecuación 4.22 en la Ecuación 4.20 se obtiene la Ecuación 4.23.

$$(t+2) \cdot P(k, t+1) - \delta_{k,1} = (t+1) \cdot (P(k-1, t) \cdot \pi(k-1, t) + P(k, t) \cdot (1 - \pi(k, t))) \tag{4.23}$$

Para poder resolver la Ecuación 4.23 se necesita conocer explícitamente la función $\pi(k, t)$.

Como se explica en las siguientes secciones, diferentes formas de la probabilidad $\pi(k, t)$ conducen a topologías diferentes.

4.4.3. Topología exponencial (modelo de enlace igualitario)

Si la probabilidad para conectarse con el nuevo nodo añadido es la misma para cualquiera de los nodos ya existentes en la red, entonces $\pi(k, t)$ es independiente de k y queda dada por la Ecuación 4.24.

$$\pi(k, t) = \frac{1}{N(t)} = \frac{1}{t+1} \tag{4.24}$$

Sustituyendo esta forma de $\pi(k)$ en la Ecuación 4.23 se obtiene la Ecuación 4.25.

$$(t+2) \cdot P(k, t+1) - \delta_{k,1} = P(k-1, t) + t \cdot P(k, t) \tag{4.25}$$

Después de un tiempo muy largo, la función $P(k, t)$ alcanza un estado estacionario en el que $P(k, t+1) = P(k, t) = P(k)$ y $P(k-1, t+1) = P(k-1, t) = P(k-1)$. En dicho estado estacionario, la Ecuación 4.25 se transforma en la Ecuación 4.26.

$$\begin{aligned}
 (t+2) \cdot P(k) - \delta_{k,1} &= P(k-1) + t \cdot P(k) \\
 &\Rightarrow 2 \cdot P(k) = P(k-1) + \delta_{k,1} \\
 &\Rightarrow P(k) = \frac{1}{2} \cdot (P(k-1) + \delta_{k,1})
 \end{aligned} \tag{4.26}$$

Como todos los nodos de la red tienen, al menos, una conexión, es claro que $P(0) = 0$.

Proposición 7. *Con esta condición inicial, la Ecuación 4.26 de recurrencia tiene su solución en la Ecuación 4.27, que no es más que la distribución exponencial.*

$$P(k) = 2^{-k} \quad (4.27)$$

Demostración. Para explicar el paso de la Ecuación 4.26 a la Ecuación 4.27, hay que advertir que la ecuación tiene un término diferente dependiendo de si $k = 1$ o no.

Resolver la Ecuación 4.26 para el caso $k = 1$ sabiendo que $P(0) = 0$ y $\delta_{1,1} = 1$ es sencillo y se observa en la Ecuación 4.28.

$$P(1) = \frac{1}{2} \cdot (P(0) + \delta_{1,1}) = \frac{1}{2} \cdot (0 + 1) = \frac{1}{2} \quad (4.28)$$

Para el caso $k > 1$, se tiene que $\delta_{k,1} = 0$ y la Ecuación 4.26 se simplifica a la Ecuación 4.29.

$$P(k) = \frac{1}{2} \cdot P(k-1) \quad (4.29)$$

La Ecuación 4.29 significa que cada valor de $P(k)$ es la mitad del valor anterior, $P(k1)$.

En la Ecuación 4.30 y la Ecuación 4.31 se siguen calculando los primeros valores para ver el patrón de la serie.

$$P(2) = \frac{1}{2} \cdot P(1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad (4.30)$$

$$P(3) = \frac{1}{2} \cdot P(2) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} \quad (4.31)$$

Y, así, sucesivamente. La generalización se observa en la Ecuación 4.32.

$$P(k) = \frac{1}{2^k} = 2^{-k} \quad (4.32)$$

□

Aunque se comenta al comienzo de la sección actual, es importante señalar que esta distribución aparece en el contexto del crecimiento de redes como el resultado de lo que se llama **enlace igualitario**, es decir, en una situación en la cual cada nodo nuevo que se añade a la red se puede enlazar a cualquiera de los nodos ya existentes con la misma probabilidad. En este sentido, el nuevo enlace que se forma es igualitario, porque no discrimina entre los nodos ya existentes.

4.4.4. Topología libre de escala (modelo de enlace preferencial)

En la vida real, las conexiones entre diferentes nodos no se dan de manera igualitaria.

Por esta razón, Barabási inventa el concepto de **enlace preferencial**, en el cual los nuevos nodos que se añaden a la red se conectan preferentemente con los nodos ya existentes que tienen el mayor número de conexiones. Intuitivamente, se puede pensar que el enlace preferencial consiste

en que uno siempre trata de estar conectado con los nodos más “populares”, es decir, con los nodos de mayor conectividad.

Para incorporar este comportamiento, Barabási sugiere que la probabilidad de enlace $\pi(k, t)$ debe tomar la forma de la Ecuación 4.33, donde k_n es la conectividad del n -ésimo nodo ya existente al tiempo t . El factor $\left(\sum_{n=0}^{N(t)} k_n\right)^{-1}$ es simplemente para garantizar que la probabilidad $\pi(k, t)$ está normalizada. De esta forma, $\pi(k, t)$ es proporcional a k y se tiene enlace preferencial, ya que cuanto más grande es la conectividad k de un nodo, mayor es la probabilidad de conectarse con él.

$$\pi(k, t) = \left(\sum_{n=0}^{N(t)} k_n \right)^{-1} \cdot k \quad (4.33)$$

Como en cada paso de tiempo se añade un nuevo nodo con una conexión, comenzando con 0 conexiones al tiempo $t = 0$, entonces para cualquier tiempo $t > 0$ se tiene la Ecuación 4.34.

$$\sum_{n=0}^{N(t)} k_n = 2 \cdot t \quad (4.34)$$

Por lo tanto, sustituyendo la igualdad de la Ecuación 4.34 en la Ecuación 4.33, se obtiene la Ecuación 4.35.

$$\pi(k, t) = \frac{k}{2 \cdot t} \quad (4.35)$$

Sustituyendo el resultado de la Ecuación 4.35 en la Ecuación 4.23 se obtiene la Ecuación 4.36.

$$(t+2) \cdot P(k, t+1) - \delta_{k,1} = \frac{t+1}{2 \cdot t} \cdot ((k-1) \cdot P(k-1, t) + (2 \cdot t - k) \cdot P(k, t)) \quad (4.36)$$

De la Ecuación 4.36 a la Ecuación 4.37 se pasa teniendo en cuenta el estado estacionario que el sistema alcanza cuando $t \rightarrow \infty$.

$$(t+2) \cdot P(k) - \delta_{k,1} = \frac{t+1}{2 \cdot t} \cdot ((k-1) \cdot P(k-1) + (2 \cdot t - k) \cdot P(k)) \quad (4.37)$$

Despejando en la Ecuación 4.37, se obtiene la Ecuación 4.38.

$$\begin{aligned} \delta_{k,1} &= (t+2) \cdot P(k) - \frac{t+1}{2 \cdot t} \cdot ((k-1) \cdot P(k-1) + (2 \cdot t - k) \cdot P(k)) \\ &= (t+2) \cdot P(k) - (t+1) \cdot P(k) - \frac{t+1}{2 \cdot t} \cdot ((k-1) \cdot P(k-1) - k \cdot P(k)) \\ &= (t+2 - t - 1) \cdot P(k) - \frac{t+1}{2 \cdot t} \cdot ((k-1) \cdot P(k-1) - k \cdot P(k)) \end{aligned} \quad (4.38)$$

También se tiene $\lim_{t \rightarrow \infty} \frac{t+1}{2 \cdot t} = \frac{1}{2}$. Por lo tanto, tomando el $\lim_{t \rightarrow \infty}$ en la Ecuación 4.36 se obtiene la distribución de conexiones estacionaria $P(k)$, lo que conduce a la Ecuación 4.39.

$$P(k) + \frac{1}{2} \cdot (k \cdot P(k) - (k-1) \cdot P(k-1)) = \delta_{k,1} \quad (4.39)$$

Como todos los nodos de la red tienen, al menos, una conexión, es claro que $P(0) = 0$.

Proposición 8. *Con esta condición inicial, la Ecuación 4.39 de recurrencia tiene su solución en la Ecuación 4.40.*

$$P(k) = \frac{4}{k \cdot (k+1) \cdot (k+2)} \quad (4.40)$$

Demostración. Para explicar el paso de la Ecuación 4.39 a la Ecuación 4.40, hay que advertir que la ecuación tiene un término diferente dependiendo de si $k = 1$ o no.

Resolver la Ecuación 4.39 para el caso $k = 1$ sabiendo que $P(0) = 0$ y $\delta_{1,1} = 1$ es sencillo y se observa en la Ecuación 4.49.

$$P(1) + \frac{1}{2} \cdot (1 \cdot P(1) - 0 \cdot P(0)) = P(1) + \frac{1}{2} \cdot P(1) = \frac{3}{2} \cdot P(1) = \delta_{1,1} = 1 \quad (4.41)$$

Despejando $P(1)$ en la Ecuación 4.49 se obtiene su valor, que se observa en la Ecuación 4.42.

$$P(1) = \frac{2}{3} \quad (4.42)$$

Para el caso $k > 1$, se tiene que $\delta_{k,1} = 0$ y la Ecuación 4.39 se simplifica a la Ecuación 4.43.

$$P(k) + \frac{1}{2} \cdot (k \cdot P(k) - (k-1) \cdot P(k-1)) = 0 \quad (4.43)$$

En la Ecuación 4.45 y la Ecuación 4.47 se siguen calculando los primeros valores para ver el patrón de la serie.

$$\begin{aligned} P(2) + \frac{1}{2} \cdot (2 \cdot P(2) - 1 \cdot P(1)) &= P(2) + \frac{1}{2} \cdot (2 \cdot P(2) - P(1)) \\ &= P(2) + \frac{1}{2} \cdot (2 \cdot P(2) - \frac{2}{3}) \\ &= 2 \cdot P(2) - \frac{2}{3} = 0 \\ \Rightarrow 2 \cdot P(2) &= \frac{2}{3} \end{aligned} \quad (4.44)$$

$$P(2) = \frac{1}{3} \quad (4.45)$$

$$\begin{aligned}
P(3) + \frac{1}{2} \cdot (3 \cdot P(3) - 2 \cdot P(2)) &= P(3) + \frac{1}{2} \cdot (3 \cdot P(3) - \frac{2}{3}) \\
&= \frac{5}{2} \cdot P(3) - \frac{1}{3} = 0 \\
\Rightarrow \frac{5}{2} \cdot P(3) &= \frac{1}{3}
\end{aligned} \tag{4.46}$$

$$P(3) = \frac{2}{15} \tag{4.47}$$

Y, así, sucesivamente. La generalización se observa en la Ecuación 4.48.

$$P(k) = \frac{4}{k \cdot (k+1) \cdot (k+2)} \tag{4.48}$$

□

Nota 27. Una gráfica log-log es un tipo de representación gráfica en la cual tanto el eje x como el eje y están escalados logarítmicamente. Esto significa que en lugar de marcarse en escala lineal, los valores en ambos ejes se representan usando su logaritmo. Este tipo de gráfica es particularmente útil para visualizar relaciones que siguen leyes de potencia o exponentes.

Aun cuando no es una ley de potencias “perfecta”, para valores grandes de k la distribución $P(k)$ se comporta como $P(k) \sim k^{-3}$. La línea punteada roja de la Figura 4.2 es la gráfica de la ley de potencias “perfecta” $f(k) = 4k^{-3}$. Es preciso remarcar que $P(k) \approx 4k^{-3}$ para valores grandes de k .

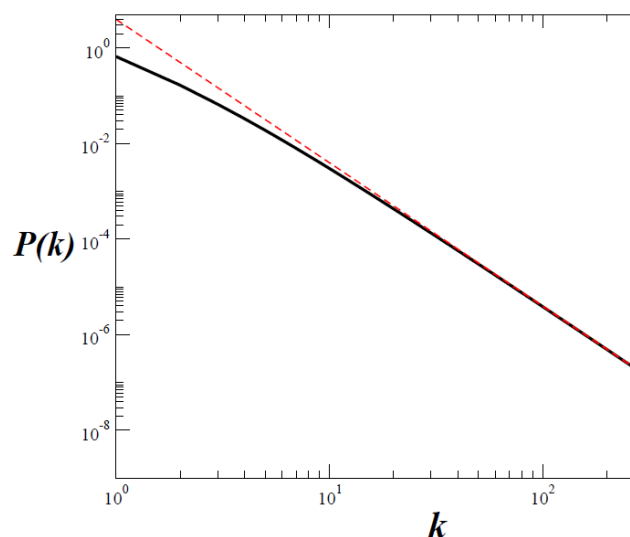


Figura 4.2: Gráfica log-log de la distribución de conectividades $P(k)$ dada en la Ecuación 4.48. Tomada de [3].

Por un lado, este es un resultado muy interesante: el enlace preferencial genera distribuciones de conectividades con colas libres de escala, y son estas colas largas las responsables de que existan elementos altamente conectados.

Por otro lado, este resultado también es desalentador ya que el proceso de enlace preferencial *siempre* da el exponente $\gamma = 3$, el cual no se encuentra frecuentemente en la naturaleza, como se observa en la Tabla 4.2.

En otras palabras, aun cuando el método de enlace preferencial da una ley de potencias, es incapaz de reproducir la gran variedad de exponentes encontrados en la naturaleza. Por esta razón, se han propuesto diferentes formas de la función $\pi(k, t)$ que corresponden a diferentes tipos de enlace preferencial.

Por ejemplo, Krapivsky, Redner y Leyvraz (referencia [11]) proponen la función de enlace $\pi(k, t)$ no lineal de la Ecuación 4.49, donde α es un exponente arbitrario.

$$\pi(k, t) = \left(\sum_{n=0}^{N(t)} k_n^\alpha \right)^{-1} \cdot k^\alpha \quad (4.49)$$

Desafortunadamente, sólo para $\alpha = 1$ se tiene que la distribución $P(k)$ tiene una cola libre de escala. El trabajo [11] muestra que la naturaleza libre de escala de la red queda destruida cuando el enlace preferencial obedece una regla no lineal como la dada en la Ecuación 4.49 cuando $\alpha \neq 1$. Es importante mencionar que todavía no existen modelos de crecimiento de redes que generen todos los exponentes listados en la Tabla 4.2.

4.5. Envejecimiento de nodos

Otro problema con el modelo de enlace preferencial propuesto por Barabási es que predice que los nodos más viejos, es decir, los que se añaden primero a la red, son los que finalmente adquieren el mayor número de conexiones. En otras palabras, en una red libre de escala se puede identificar fácilmente a los nodos más viejos: son aquellos altamente conectados. Y cuanto más viejo es un nodo, mayor es su conectividad.

Para ver que esto efectivamente es un resultado del modelo de enlace preferencial, se considera un nodo particular v_n de la red. Su conectividad k_n cambia a una tasa que es proporcional a la probabilidad $\pi(k_n, t)$ de que este nodo adquiera más conexiones, como se observa en la Ecuación 4.50.

$$\frac{dk_n}{dt} = \pi(k_n, t) \quad (4.50)$$

Para el caso particular en el que $\pi(k_n, t)$ está dada como en la Ecuación 4.35, se tiene que la Ecuación 4.50 se transforma en la Ecuación 4.51.

$$\frac{dk_n}{dt} = \frac{k_n}{2 \cdot t} \quad (4.51)$$

Como el nodo v_n nace al tiempo $t = n$ con una conexión, la condición inicial para la Ecuación 4.51 es $k_n(n) = 1$.

Proposición 9. *Con esta condición inicial, la Ecuación 4.51 tiene su solución en la Ecuación 4.52.*

$$k_n(t) = \left(\frac{t}{n}\right)^{\frac{1}{2}} \quad (4.52)$$

Demostración. Para explicar el paso de la Ecuación 4.51 a la Ecuación 4.52 hay que advertir que se trata de una ecuación diferencial ordinaria de primer orden. Una forma estándar de resolverla es separando variables en la Ecuación 4.51; se reorganiza para separar k_n y t en diferentes lados de la igualdad, como se observa en la Ecuación 4.53.

$$\frac{dk_n}{k_n} = \frac{dt}{2 \cdot t} \quad (4.53)$$

Ahora, se integran ambos lados de la Ecuación 4.53 como se observa en la Ecuación 4.54.

$$\int \frac{1}{k_n} dk_n = \int \frac{1}{2 \cdot t} dt \quad (4.54)$$

La resolución de las integrales se observa en la Ecuación 4.55, donde C es la constante de integración.

$$\ln |k_n| = \frac{1}{2} \cdot \ln |t| + C = \ln |t|^{\frac{1}{2}} + C \quad (4.55)$$

Al aplicar la exponencial a ambos lados de la igualdad en la Ecuación 4.55 se obtiene la Ecuación 4.56. Se tiene en cuenta que tanto k_n como t son números enteros positivos.

$$e^{\ln |k_n|} = |k_n| = k_n = e^{\ln |t|^{\frac{1}{2}} + C} = e^{\ln |t|^{\frac{1}{2}}} \cdot e^C = |t|^{\frac{1}{2}} \cdot e^C = e^C \cdot t^{\frac{1}{2}} \quad (4.56)$$

Renombrando la constante e^C como C' , la solución general es la Ecuación 4.57.

$$k_n(t) = C' \cdot t^{\frac{1}{2}} \quad (4.57)$$

Para encontrar C' se utiliza $k_n(n) = 1$, lo cual conduce a la Ecuación 4.58.

$$k_n(n) = C' \cdot n^{\frac{1}{2}} = 1 \quad (4.58)$$

Resolviendo para C' , se obtiene la Ecuación 4.59.

$$C' = \frac{1}{n^{\frac{1}{2}}} \quad (4.59)$$

Finalmente, sustituyendo C' en la solución general (Ecuación 4.57), se obtiene la solución final de la ecuación diferencial dada la condición inicial, que se observa en la Ecuación 4.60.

$$k_n(t) = n^{-\frac{1}{2}} \cdot t^{\frac{1}{2}} = \left(\frac{t}{n}\right)^{\frac{1}{2}} \quad (4.60)$$

□

De la Ecuación 4.60 se deduce claramente que, en cualquier instante t , se cumple la desigualdad de la Ecuación 4.61. Esto implica que los nodos que aparecen primero tienen, en promedio, conectividades mayores que aquellos que llegan posteriormente.

$$k_1(t) > k_2(t) > k_3(t) > \dots > k_n(t) \quad (4.61)$$

Sin embargo, este comportamiento predicho por el modelo de enlace preferencial, en el que los nodos más viejos son los más conectados, no siempre se observa en la naturaleza.

Ejemplo 47. *Por ejemplo, Lada A. Adamic y Bernardo Huberman estudian una red que consiste en 260000 páginas WWW [1], considerando que dos páginas están conectadas si una contiene un hipervínculo a la otra. Adamic y Huberman encuentran que, aún cuando esta red tiene topología libre de escala con un exponente $\gamma \sim 2$ (Figura 4.3 (A)), no existe correlación entre la conectividad de los nodos y su edad. La Figura 4.3 (B) muestra la gráfica de la conectividad de los nodos como función de su edad. Como puede observarse, estas dos cantidades no parecen estar relacionadas.*

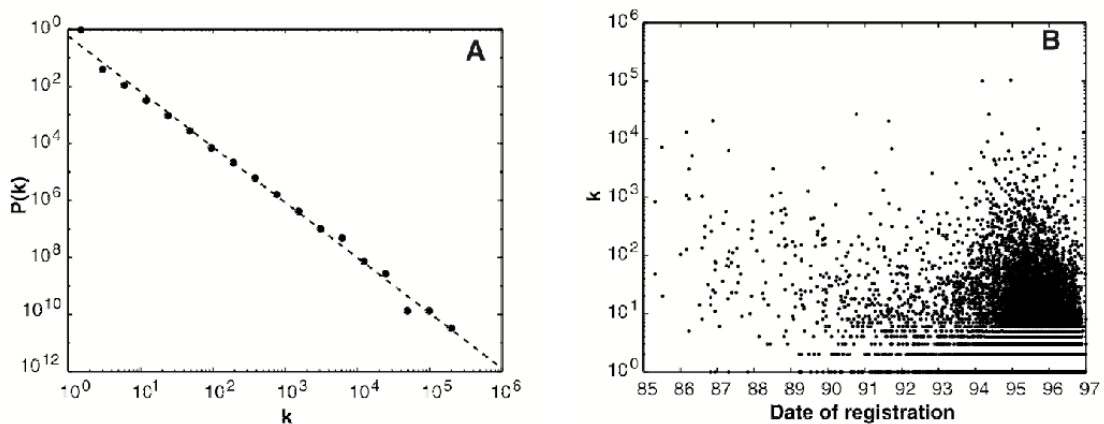


Figura 4.3: (A) Gráfica log-log de la distribución de conectividades $P(k)$ para una red WWW consistente en 260000 páginas. (B) Conectividad k de los nodos de la red como función de su edad. Tomadas de [1].

Ejemplo 48. *Probablemente el ejemplo más contundente que contradice los resultados de la Ecuación 4.61 es Google, el robot buscador de la red WWW. Altavista y Yahoo ya existían y estaban bien establecidos en 1998, año en que apareció Google; éste tomó la delantera rápidamente, convirtiéndose en el buscador más popular en el mundo informático. Este éxito repentino se debió probablemente a que estaba mejor diseñado, era más rápido y más eficiente que Altavista y Yahoo. En otras palabras, Google nace con características que lo hacen un robot “mejor adaptado” que sus competidores.*

4.5.1. Modelo de enlace preferencial con adaptabilidad

Esta observación hace que Barabási proponga el concepto de adaptabilidad en el contexto de las redes complejas. Así, cada uno de los nodos v_n , además de tener una conectividad k_n , también tiene un parámetro de **adaptabilidad** asociado w_n . Este parámetro mide cuan bien está adaptado el nodo v_n a su entorno: cuanto más grande es el valor de w_n , mayor es el grado de adaptabilidad de v_n . Específicamente, lo que Barabási hace es proponer una **función de enlace preferencial** $\pi(k, w, t)$ que, además de ser proporcional a la conectividad k de los nodos ya existentes, también es proporcional a su adaptabilidad w (referencia [4]). De esta forma, la probabilidad de que un nuevo nodo añadido a la red se conecte con el nodo v_n ya existente, cuya conectividad es k_n y cuya adaptabilidad es w_n , queda dada por la Ecuación 4.62, donde C es una constante de normalización.

$$\pi(k_n, w_n, t) = C \cdot w_n \cdot k_n \quad (4.62)$$

Las adaptabilidades w_n asociadas a los nodos de la red son variables aleatorias que se escogen de una distribución $A(w)$. Cada nodo nace con su propia adaptabilidad, la cual no cambia en el tiempo; la adaptabilidad de un nodo es innata.

Ejemplo 49. *Un resultado sorprendente del modelo de enlace preferencial con adaptabilidad es que puede ser mapeado exactamente a un gas de Bose-Einstein. En este mapeo cada nodo de la red representa un nivel energético del gas, mientras que las conectividades de los nodos representan los números de ocupación de los respectivos niveles de energía. Más que una simple metáfora, es un mapeo matemático preciso. Por eso, bajo ciertas condiciones, la red en crecimiento puede presentar “condensación de Bose-Einstein”, lo cual significa que, repentinamente, un nodo puede adquirir la mayoría de las conexiones de la red, independientemente de su edad (como sucede con Google) [4].*

4.6. Islas gigantes

En esta sección se regresa al modelo descrito en la Sección 4.4.1. Inicialmente, todos los nodos están desconectados, pero conforme se van enlazando parejas de nodos, se comienzan a formar islas de nodos conectados. Al principio, las islas son pequeñas, pero al ir añadiendo conexiones entre los nodos, las islas crecen y se conectan entre sí. Eventualmente, se forma una isla gigante, es decir, una isla que es mucho más grande que todas las demás.

Ejemplo 50. *El tamaño de la isla gigante es importante en el estudio de la propagación de epidemias en una sociedad, donde se tienen personas que se contagian unas a otras. Las islas consisten en los grupos de personas que están infectadas, y la enfermedad se convierte en una epidemia cuando se forma una isla gigante que abarca a la mayoría de la sociedad.*

Dado un conjunto de elementos, es interesante estudiar cuántos enlaces (contagios) tienen que establecerse para que se forme la isla gigante. En este sentido, Erdős y Rényi son los primeros en mostrar la existencia de una **transición de fase** en la teoría de redes. Dicha transición de fase consiste, precisamente, en la formación de la isla gigante. Aunque el trabajo original lo llevan a cabo para redes con topología de Poisson, es fácil generalizarlo a redes con topologías arbitrarias.

Para calcular el tamaño de la isla gigante se utiliza un razonamiento de consistencia.

Proposición 10. *Sea q la probabilidad de que un nodo v_n escogido aleatoriamente no pertenezca a la isla gigante. Supongamos que v_n tiene k vecinos. Claramente, v_n no pertenece a la isla gigante si y sólo si ninguno de sus k vecinos pertenece a la isla gigante. Por lo tanto, la probabilidad q debe satisfacer la Ecuación 4.63 de consistencia.*

$$q = \sum_k P(k) \cdot q^k \quad (4.63)$$

Demostración. El lado izquierdo de la Ecuación 4.63 es simplemente la probabilidad q de que v_n no pertenezca a la isla gigante. El lado derecho es la probabilidad $P(k)$ de que v_n tenga k vecinos, multiplicada por la probabilidad q^k de que ninguno de estos k vecinos pertenezca a la isla gigante. La suma sobre k tiene en cuenta todos los posibles vecinos que v_n puede llegar a tener. \square

4.6.1. Distribución de Poisson

En el caso en que $P(k)$ es una distribución de Poisson, la Ecuación 4.63 se transforma en la Ecuación 4.64. En la primera igualdad se utiliza la Ecuación 4.1 y, en la tercera, la Ecuación 2.8.

$$q = \sum_{k=0}^{\infty} e^{-z} \cdot \frac{(z \cdot q)^k}{k!} = e^{-z} \cdot \sum_{k=0}^{\infty} \frac{(z \cdot q)^k}{k!} = e^{-z} \cdot e^{z \cdot q} = e^{z \cdot (q-1)} \quad (4.64)$$

Denotando como $p = 1 - q$ a la **probabilidad de que un nodo arbitrario sí pertenezca a la isla gigante**, se obtiene la Ecuación 4.65.

$$p = 1 - e^{-z \cdot p} \quad (4.65)$$

La Ecuación 4.65 es trascendental y no se puede resolver de forma analítica. Sin embargo, es fácil resolverla de forma numérica y encontrar p para cada valor de z . La Figura 4.4 muestra p como función de z .

Para $z < 1$ la probabilidad de pertenecer a la isla gigante es nula, en otras palabras, no hay isla gigante para $z < 1$. Justo en $z = 1$ aparece y, conforme aumenta el valor de z , lo cual es equivalente a añadir más conexiones a la red, la isla gigante comienza a formarse. Este es un resultado interesante: se necesita, en promedio, sólo una conexión por nodo para que aparezca la isla gigante. Si se sigue aumentando el valor de z , el tamaño de la isla gigante aumenta más y más, como se puede observar en la Figura 4.4, la cual muestra el comportamiento característico de una transición de fase de segundo orden que ocurre en $z = 1$ (Anexo C).

4.6.2. Distribución libre de escala

Si se considera ahora el caso de las **redes con topología libre de escala**, la distribución $P(k)$ toma la forma de la Ecuación 4.66, donde $\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$ es la función ζ de Riemman, que simplemente juega el papel de la constante de normalización.

$$P(k) = \frac{1}{\zeta(\gamma)} \cdot k^{-\gamma} \quad (4.66)$$

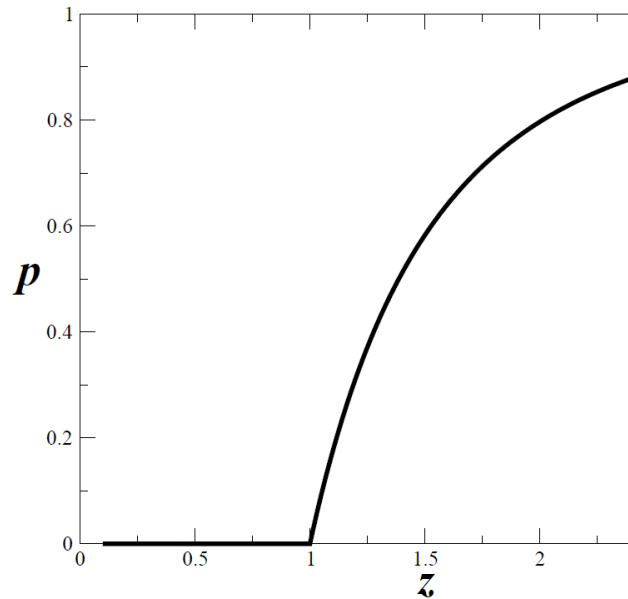


Figura 4.4: Probabilidad p de pertenecer a la isla gigante como función de la conectividad promedio z de la red, para el caso de la topología de Poisson. Tomada de [3].

Para esta distribución la Ecuación 4.64 de consistencia se convierte en la Ecuación 4.67, donde $Li_\gamma(q) = \sum_{k=1}^{\infty} k^{-\gamma} \cdot q^k$ es la función poli-logarítmica de orden γ .

$$q = \frac{1}{\zeta(\gamma)} \cdot \sum_{k=1}^{\infty} \frac{q^k}{k^\gamma} = \frac{1}{\zeta(\gamma)} \cdot Li_\gamma(q) \quad (4.67)$$

En términos de la probabilidad $p = 1 - q$, la Ecuación 4.65 queda como la Ecuación 4.68.

$$p = 1 - \frac{1}{\zeta(\gamma)} \cdot Li_\gamma(1 - p) \quad (4.68)$$

Como $Li_\gamma(0) = 0$, es claro que $p = 1$ siempre es solución de la Ecuación 4.68. Por lo tanto, en una red libre de escala la componente gigante siempre existe y la probabilidad de pertenecer a ella es $p = 1$. Este interesante resultado es una consecuencia de la existencia de los elementos altamente conectados en redes libres de escala. Dichos elementos mantienen a toda la red conectada, impidiendo que se fracture en pequeñas islas.

Ejemplo 51. Los resultados presentados en esta sección muestran que, en redes con **topología de Poisson**, siempre se puede vacunar a un número suficiente de personas para garantizar que la conectividad promedio z de la red de individuos infectados se mantenga inferior a 1. En tal caso, se puede detener la epidemia, ya que no hay isla gigante, sólo pequeñas islas de personas infectadas. Sin embargo, las redes sociales no tienen topología de Poisson, sino que exhiben **topologías libres de escala**. Para tales topologías siempre existe una isla gigante, y la probabilidad de pertenecer a ella es $p = 1$. Por lo tanto, en redes libres de escala las enfermedades tarde o temprano se propagan a todas las personas que no son vacunadas. Es decir, la única forma de detener

la propagación de una epidemia en redes libres de escala es vacunando a todas las personas de la sociedad. Los efectos aterradores del resultado anterior pueden observarse claramente en África, donde comunidades enteras han desaparecido a causa del Síndrome de Inmunodeficiencia Adquirida (SIDA). Más información sobre la propagación de epidemias en redes sociales puede encontrarse en las referencias [18] [16].

Capítulo 5

Ejemplos de casos reales

Gephi es un *software* de código abierto para visualizar y analizar gráficos de grandes redes. Gephi utiliza un motor de renderizado 3D para mostrar gráficos en tiempo real y acelerar la exploración. Se puede usar para explorar, analizar, espacializar, filtrar, ordenar, manipular y exportar todo tipo de gráficos [10].

Gephi ofrece diferentes estrategias para combinar las aristas paralelas y calcular el peso de la arista resultante; se listan a continuación.

- Suma (*Sum*): El peso de la arista combinada es la suma de los pesos de todas las aristas individuales entre los dos nodos. Esta estrategia es útil para que la arista combinada refleje la “intensidad total” de la relación entre los nodos.
- Promedio (*Average*): El peso de la arista combinada es el promedio de los pesos de las aristas originales. Esta estrategia es útil para suavizar las diferencias y obtener un valor representativo.
- Máximo (*Maximum*): El peso de la arista combinada es el máximo peso de las aristas originales. Esta estrategia es útil para que la arista combinada refleje la interacción más fuerte entre los nodos.
- Mínimo (*Minimum*): El peso de la arista combinada es el mínimo peso de las aristas originales. Esta estrategia es útil si interesa que la interacción sea menos intensa.
- Mediana (*Median*): El peso de la arista combinada es la mediana de los pesos de las aristas originales. Esta estrategia es útil para evitar el efecto de valores extremos.

Elegir la estrategia adecuada depende del contexto del análisis y del significado de los pesos en la red. En la mayoría de los análisis de redes sociales, usar “Suma (*Sum*)” es generalmente la opción más coherente y útil, ya que proporciona una medida clara y acumulativa de las conexiones. Una de las razones es que captura la intensidad de la relación, pues permite reflejar la intensidad total de las interacciones entre dos usuarios. Si un usuario interactúa con otro varias veces, sumando los valores se cuantifica esta relación más fuertemente. Además, en las redes sociales, la suma de interacciones es un buen indicador de la fuerza de la relación entre usuarios, lo que es clave para identificar comunidades o nodos influyentes. Otra razón de peso es que se trata de una simplificación realista, ya que al sumar las aristas, se simplifica la red sin perder la información esencial sobre la cantidad de interacciones.

5.1. Grafo dirigido

En esta sección se expone el primer ejemplo de análisis de redes con Gephi. Se trata de un grafo dirigido [9].

5.1.1. Descripción del *dataset*

El *dataset* “Twitch Social Networks” en Kaggle contiene redes sociales de usuarios de Twitch divididas por países. Cada subarchivo (`edges.csv`) para un país específico describe las conexiones entre usuarios mediante dos columnas: *from* (usuario de origen o seguidor) y *to* (usuario de destino o seguido). Este *dataset* es ideal para analizar las estructuras de red, identificar comunidades, y calcular métricas como la centralidad de los nodos. Es un recurso excelente para estudiar la dinámica social en plataformas de *streaming*.

Nota 28. *Cada nodo representa un usuario de Twitch, pero un enlace entre dos usuarios puede representar varias interacciones posibles en la vida real: raid o usuario que redirige su audiencia a otro usuario; follow o usuario que sigue a otro; host o usuario que comparte la transmisión de otro en su propio canal; menciones y colaboraciones, que son interacciones directas que generan visibilidad para otro usuario. Este ejemplo se centra únicamente en los follows.*

La jerarquía de archivos del *dataset* se lista a continuación; hay un directorio por cada país estudiado.

- Alemania o *Denmark* (DE)
 - DE_edges.csv
 - DE_target.csv
 - DE.json
- Reino Unido o *English - Great Britain* (ENGB)
 - ENGB_edges.csv
 - ENGB_target.csv
 - ENGB.json
- España o *Spain* (ES)
 - ES_edges.csv
 - ES_target.csv
 - ES.json
- Francia o *France* (FR)
 - FR_edges.csv
 - FR_target.csv
 - FR.json
- Brasil o *Portuguese - Brazil* (PTBR)

- PTBR_edges.csv
- PTBR_target.csv
- PTBR.json
- Rusia o *Russia* (RU)
 - RU_edges.csv
 - RU_target.csv
 - RU.json

Se puede explorar y descargar el *dataset* en la siguiente Localizador Uniforme de Recursos o *Uniform Resource Locator* (URL): <https://www.kaggle.com/datasets/andreagarritano/twitch-social-networks>.

5.1.2. Importación de los datos

En primer lugar, se cambia el nombre de las columnas de cada uno de los archivos `edges.csv` para facilitar su comprensión por parte de la herramienta *software*. En concreto, *from* pasa a ser *source* y *to* se convierte en *target*.

Se abre la aplicación de escritorio Gephi. Se selecciona “Archivo” y “Nuevo Proyecto”. De nuevo, se selecciona “Archivo”, “Importar Hoja de Cálculo” y se elige el archivo `edges.csv` del país correspondiente. En el asistente de importación, se selecciona “Tabla de aristas” y se asignan las columnas *source* y *target* como nodos fuente y destino, respectivamente. Finalmente, se hace clic en “Finalizar” y, luego, en “Aceptar”.

En el caso de Twitch, donde es importante entender la intensidad global de las relaciones entre usuarios, la estrategia elegida es “Suma (*Sum*)”. Se considera la más adecuada porque permite ver cuáles son los vínculos más fuertes en términos de cantidad total de interacciones, lo que es crucial en redes sociales donde la frecuencia y el número de interacciones pueden ser indicadores importantes de influencia o conexión. Sin embargo, dado que las interacciones son de naturaleza diversa (Observación 28) y preocupa que muchas interacciones menores puedan sesgar la red, se puede considerar el “Promedio (*Average*)” para obtener un valor representativo que no esté dominado por nodos con muchas aristas.

5.1.3. Contexto

En la pestaña “Contexto” aparece un recuento de nodos y aristas.

Alemania

- Nodos: 9417
- Aristas: 153055

Reino Unido

- Nodos: 7015
- Aristas: 94597

España

- Nodos: 4594
- Aristas: 59334

Francia

- Nodos: 6486
- Aristas: 112621

Brasil

- Nodos: 1887
- Aristas: 31249

Rusia

- Nodos: 4301
- Aristas: 37270

5.1.4. Estadísticas

En la pestaña “Estadísticas”, se hace clic en “Ejecutar” para cada una de las medidas y parámetros que definen un grafo.

A continuación, se reproduce la definición de algunos de estos conceptos por parte de Gephi, aunque ya se exponen anteriormente en este trabajo.

Diámetro

Definición 63. *El **diámetro** es la distancia de grafo más larga entre dos nodos cualquiera de la red, es decir, cómo de lejos están los dos nodos más alejados. Coincide con la Definición 49.*

Nota 29. *La **distancia media** de grafo (Definición 47) se calcula entre todos los pares de nodos. Los nodos conectados tienen distancia 1.*

Nota 30. *Al calcular la estadística de la Definición 63, se puede elegir normalizar centralidades en el rango $[0, 1]$.*

Definición 64. *La **intermediación** mide la frecuencia con la que un nodo aparece en el camino más corto entre nodos de la red.*

Definición 65. *La **cercanía** es la distancia media desde un nodo inicial a todos los demás nodos de la red.*

Definición 66. *La **excentricidad** es la distancia desde un nodo al nodo más alejado de él en la red. Coincide con la Definición 48.*

Definición 67. *El **radio** de un grafo es la distancia mínima máxima de un nodo a todos los demás nodos, es decir, es la menor distancia máxima desde cualquier nodo del grafo a cualquier otro.*

Densidad

Definición 68. La *densidad* mide cómo de cerca está el grafo de ser completo. Un grafo completo tiene todas las aristas posibles y una densidad igual a 1. Coincide con la Definición 51.

HITS

Definición 69. *Búsqueda de temas inducida por hipervínculos o Hyperlink-Induced Topic Search (HITS)* computa dos valores separados para cada nodo. El primer valor, llamado *Authority*, mide cómo de valiosa es la información almacenada en ese nodo. El segundo valor, llamado *Hub*, mide la calidad de los enlaces de ese nodo.

Nota 31. En la estadística de la Definición 69, existe un parámetro. ϵ , cuyo valor por defecto es 0,0001, es el criterio de parada; cuanto menor es este valor, más tiempo toma la convergencia.

Pagerank

Definición 70. El algoritmo *PageRank* clasifica las “páginas” de los nodos de acuerdo a la frecuencia con la que un usuario, siguiendo enlaces, llega a la “página” del nodo de forma no aleatoria.

Nota 32. En la estadística de la Definición 70, existen dos parámetros. Por defecto, se usa una probabilidad $p = 0,85$, para simular aleatoriamente que el usuario reinicia la navegación web, y un criterio de parada $\epsilon = 0,001$; cuanto menor es este valor, más tiempo toma la convergencia. También se ofrece la opción de utilizar el peso de las aristas.

Ejemplo 52. Un usuario de Twitch con un PageRank alto podría ser un streamer popular que recibe muchos raids de otros streamers, tiene muchos seguidores, y es frecuentemente mencionado en otras transmisiones. Esto sugiere que otros usuarios de Twitch son más propensos a encontrar y ver sus transmisiones, lo que lo convierte en un nodo central dentro de la red de Twitch.

Componentes conexos

Nota 33. Se determina el número de **componentes conexos** en la red. Si el grafo es dirigido, detecta componentes fuertemente y débilmente conectados. Para el caso no dirigido, detecta solo componentes débilmente conectados.

Modularidad

Definición 71. La *modularidad* (Ecuación 3.27) es un algoritmo de detección de comunidades.

Nota 34. En la estadística de la Definición 71, existen cuatro parámetros. En el caso de utilizar aleatoriedad se produce una mejor descomposición, pero aumenta el tiempo de cómputo. Se puede utilizar o no el peso de las aristas. La resolución toma valores menores para obtener más comunidades (más pequeñas) y mayores que 1,0 para obtener menos comunidades (más grandes). La etiqueta de las clases empieza en 0 por defecto.

Coefficiente de *clustering*

Nota 35. El *coeficiente de clustering*, junto con el valor promedio del camino más corto, puede indicar un efecto “small world”. Indica cómo los nodos están incrustados entre sus nodos vecinos. El coeficiente medio de clustering es el valor medio de los coeficientes individuales y da una indicación general del clustering en la red.

Centralidad de vector propio

Definición 72. La *centralidad de vector propio* (Ecuación 3.12) es una medida de la importancia de un nodo en la red basada en sus conexiones.

Nota 36. En la estadística de la Definición 72, el número de iteraciones es 100 por defecto.

Definición 73. El *sum change* representa la diferencia acumulada entre los valores de centralidad de los nodos en dos iteraciones sucesivas del algoritmo que calcula la centralidad. En cada iteración, el algoritmo ajusta los valores de centralidad de vector propio de los nodos, y el sum change indica cuánto cambian esos valores entre iteraciones.

Nota 37. Dado que el sum change mide el cambio total en los valores de centralidad de todos los nodos entre una iteración y la siguiente, un valor más cercano a 0 indica que el algoritmo se estabiliza o converge, es decir, los valores de centralidad ya no cambian significativamente.

5.1.5. Resultados

Se muestran los resultados obtenidos por cada uno de los países para las estadísticas de la Sección 5.1.4. En este ejemplo se mantienen los valores por defecto para cada uno de los parámetros, debido a que lo interesante es comparar los resultados entre países.

Alemania

En el caso de Alemania se obtienen los valores que se observan a continuación.

1. Visión general de la red

- Grado medio: 16,253

En la Figura 5.1 se observa la distribución de los grados entrante y saliente, respectivamente.

- Grado medio con pesos: 16,262

En la Figura 5.2 se observa la distribución de los grados entrante y saliente con pesos, respectivamente.

- Diámetro: 14

Radio: 0

- ✓ Parámetro:

- Normalizar centralidades en el rango $[0, 1]$: No

- Resultado: Figura 5.3

- Densidad: 0,002

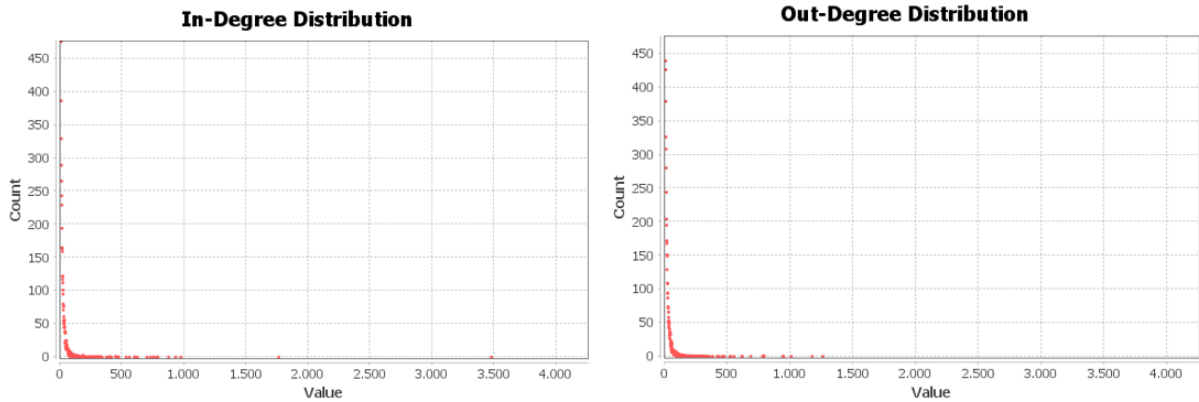


Figura 5.1: Ejemplo 1 (DE): Grado medio.

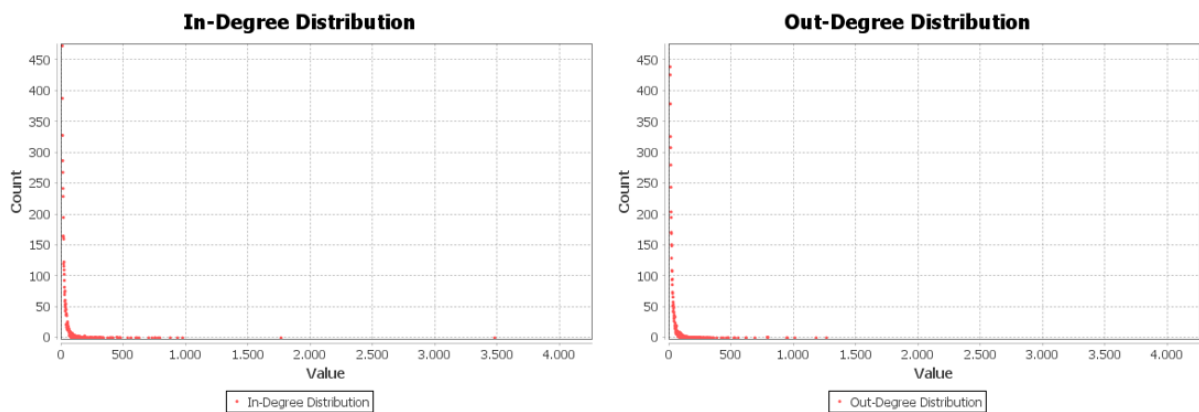


Figura 5.2: Ejemplo 1 (DE): Grado medio con pesos.

- HITS:
 - ✓ Parámetro:
 - ϵ : 0,0001
 - Resultado: Figura 5.4
- PageRank:
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - Resultado: Figura 5.5
- Componentes conexos: 1

2. Detección de comunidades

- Modularidad:
 - ✓ Parámetros:
 - Aleatorio: Sí

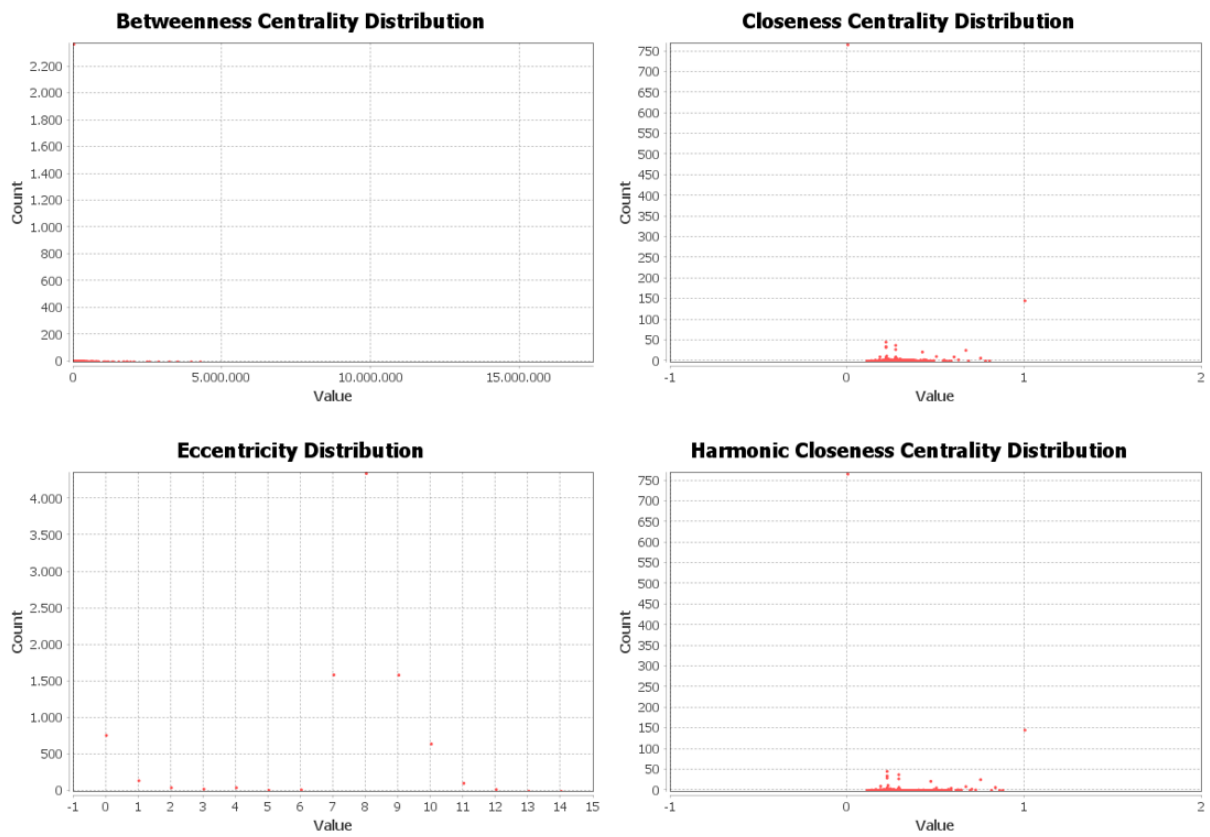


Figura 5.3: Ejemplo 1 (DE): Diámetro.

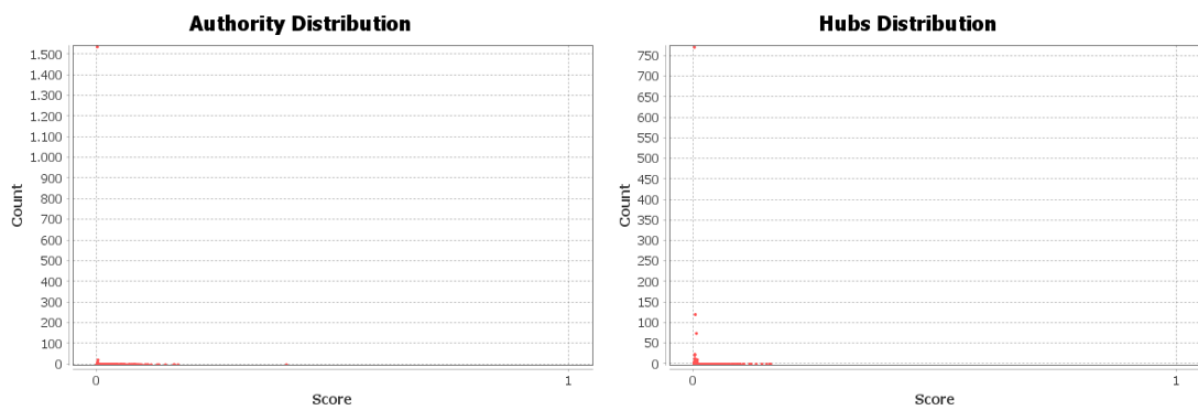


Figura 5.4: Ejemplo 1 (DE): HITS.

- Utilizar pesos: Sí
- Resolución: 1,0
- Resultados: Figura 5.6
 - Modularidad: 0,276
 - Modularidad con resolución: 0,276
 - Número de comunidades: 6

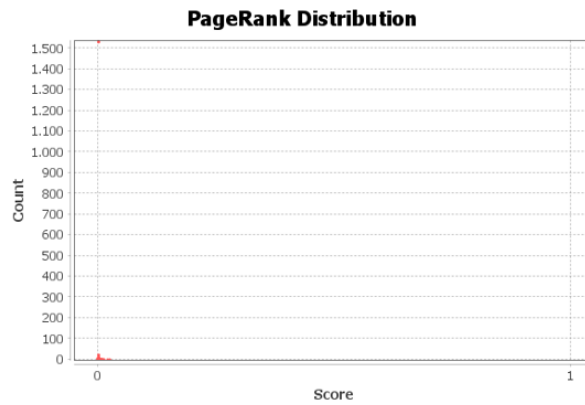


Figura 5.5: Ejemplo 1 (DE): PageRank.

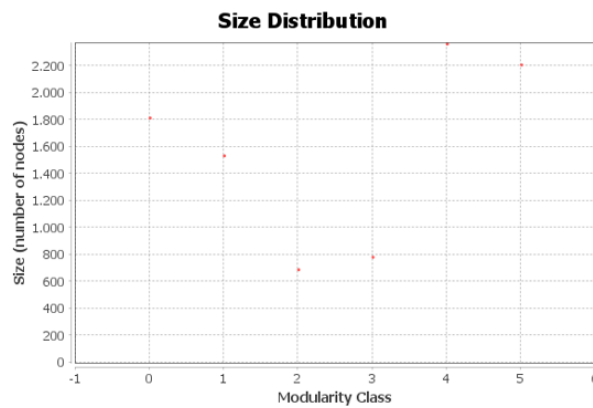


Figura 5.6: Ejemplo 1 (DE): Modularidad.

- Inferencia estadística:
 - Resultados: Figura 5.7
 - Longitud: 803109,355
 - Número de comunidades: 765

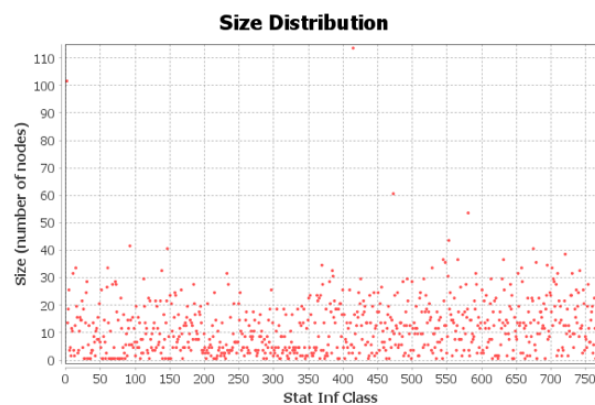


Figura 5.7: Ejemplo 1 (DE): Inferencia estadística.

3. Visión general de los nodos

- Coeficiente medio de *clustering*: 0,096
- Centralidad de vector propio:
 - ✓ Parámetros:
 - Número de iteraciones: 100
 - *Sum change*: 0,2516626930720811 \approx 0,252
 - Resultado: Figura 5.8.

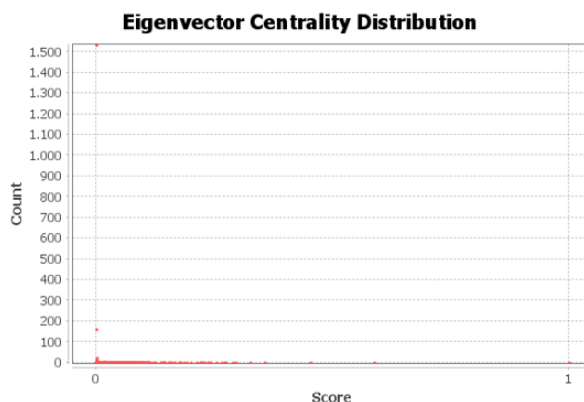


Figura 5.8: Ejemplo 1 (DE): Centralidad de vector propio.

4. Visión general de las aristas

- Longitud media de camino: 3,741846899244344 \approx 3,742

Reino Unido

En el caso de Reino Unido se obtienen los valores que se observan a continuación.

1. Visión general de la red

- Grado medio: 13,485
En la Figura 5.9 se observa la distribución de los grados entrante y saliente, respectivamente.
- Grado medio con pesos: 13,5
En la Figura 5.10 se observa la distribución de los grados entrante y saliente con pesos, respectivamente.
- Diámetro: 12
Radio: 0
 - ✓ Parámetro:
 - Normalizar centralidades en el rango [0, 1]: No
 - Resultado: Figura 5.11
- Densidad: 0,002

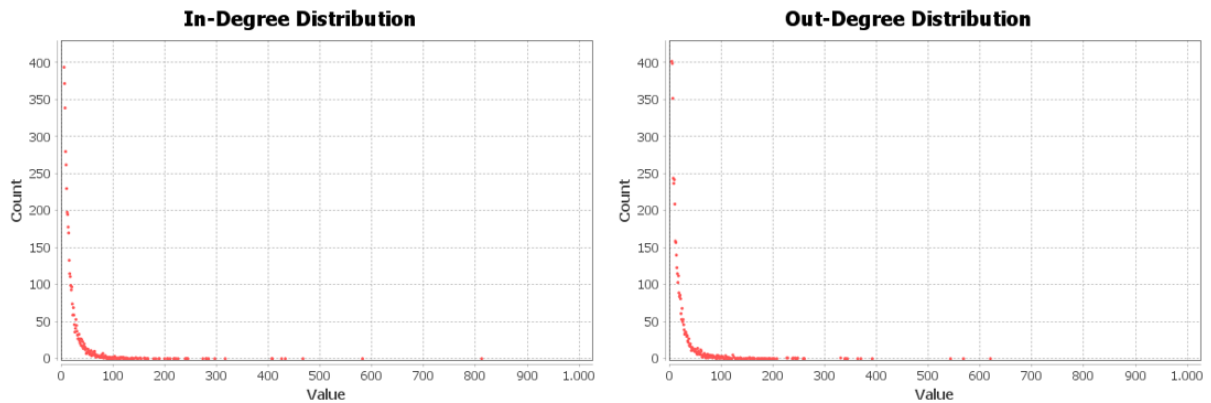


Figura 5.9: Ejemplo 1 (ENGB): Grado medio.

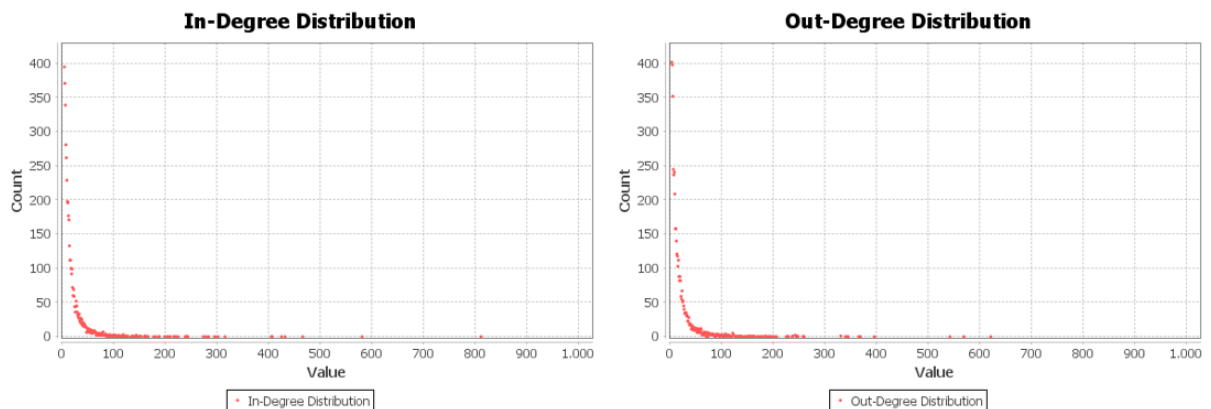


Figura 5.10: Ejemplo 1 (ENGB): Grado medio con pesos.

- HITS:
 - ✓ Parámetro:
 - ϵ : 0,0001
 - Resultado: Figura 5.12
- PageRank:
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - Resultado: Figura 5.13
- Componentes conexos: 1

2. Detección de comunidades

- Modularidad:
 - ✓ Parámetros:
 - Aleatorio: Sí

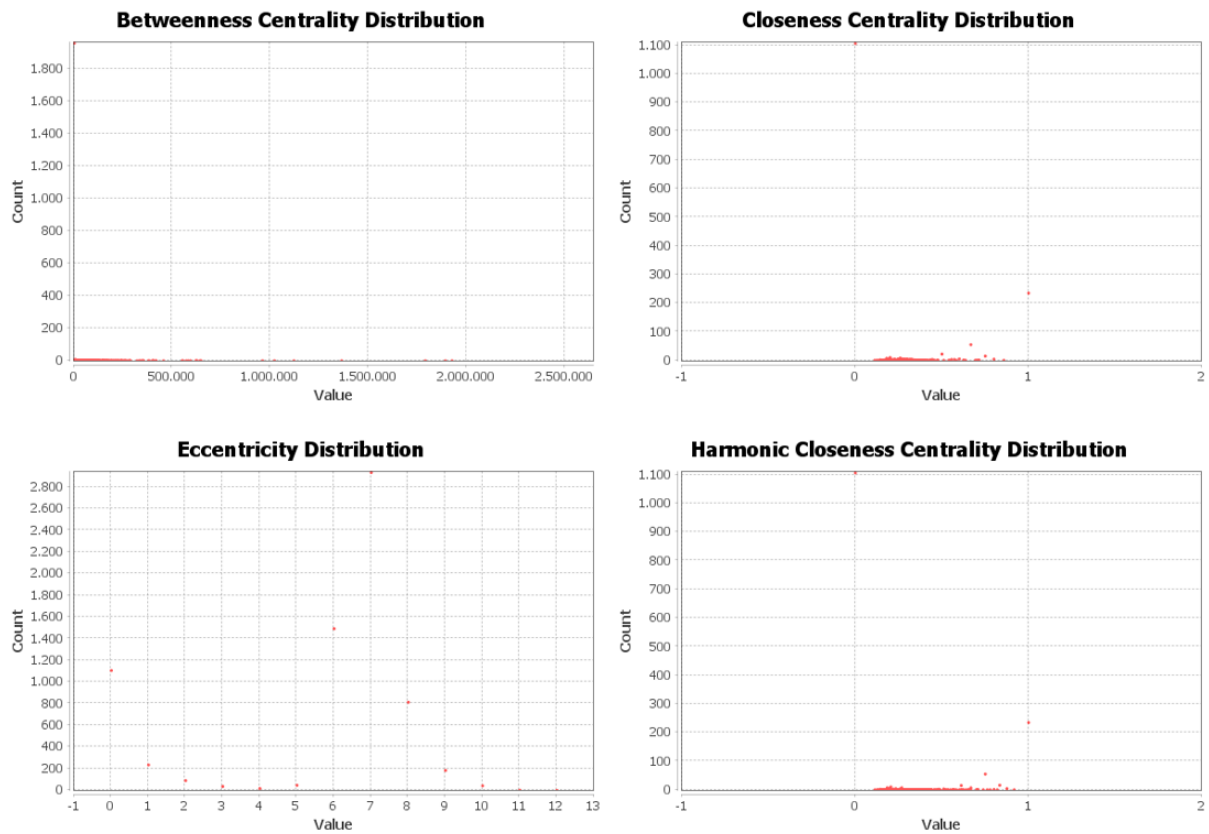


Figura 5.11: Ejemplo 1 (ENGB): Diámetro.

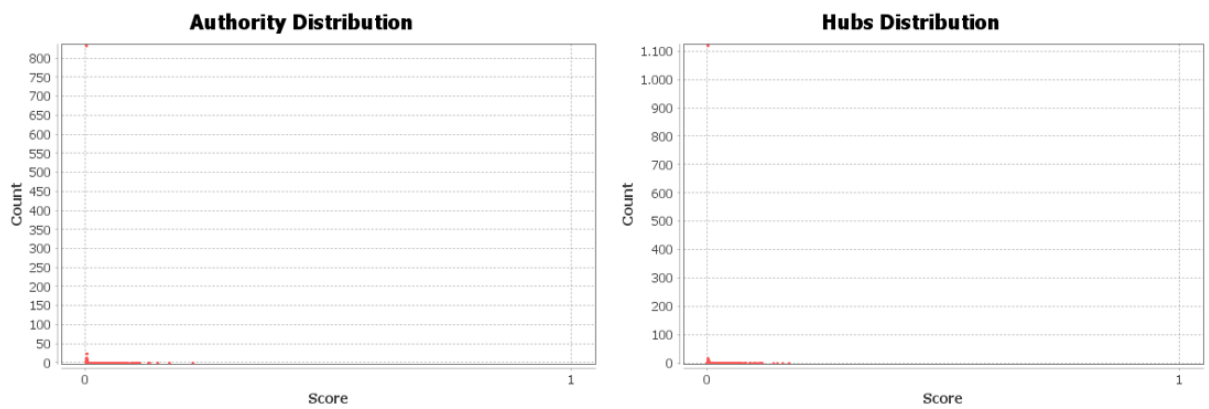


Figura 5.12: Ejemplo 1 (ENGB): HITS.

- Utilizar pesos: Sí
- Resolución: 1,0
- Resultados: Figura 5.14
 - Modularidad: 0,364
 - Modularidad con resolución: 0,364
 - Número de comunidades: 10

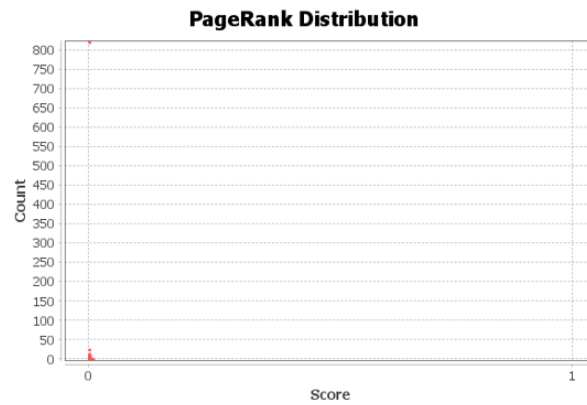


Figura 5.13: Ejemplo 1 (ENGB): PageRank.

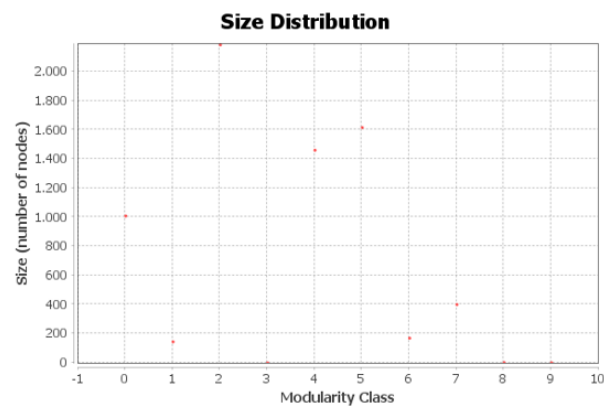


Figura 5.14: Ejemplo 1 (ENGB): Modularidad.

- Inferencia estadística:
 - Resultados: Figura 5.15
 - Longitud: 803109,355
 - Número de comunidades: 765

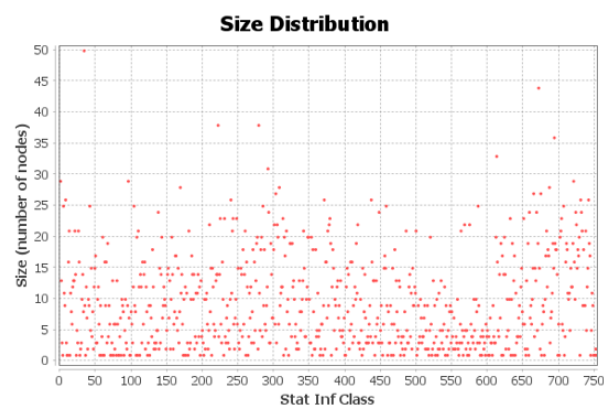


Figura 5.15: Ejemplo 1 (ENGB): Inferencia estadística.

3. Visión general de los nodos

- Coeficiente medio de *clustering*: 0,066
- Centralidad de vector propio:
 - ✓ Parámetros:
 - Número de iteraciones: 100
 - *Sum change*: 0,12156024858549262 \approx 0,122
 - Resultado: Figura 5.16.

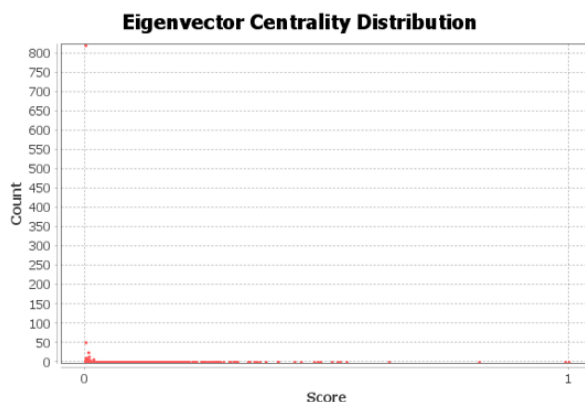


Figura 5.16: Ejemplo 1 (ENGB): Centralidad de vector propio.

4. Visión general de las aristas

- Longitud media de camino: 3,7858868224358133 \approx 3,786

España

En el caso de España se obtienen los valores que se observan a continuación.

1. Visión general de la red

- Grado medio: 12,916
En la Figura 5.17 se observa la distribución de los grados entrante y saliente, respectivamente.
- Grado medio con pesos: 12,926
En la Figura 5.18 se observa la distribución de los grados entrante y saliente con pesos, respectivamente.
- Diámetro: 12
Radio: 0
 - ✓ Parámetro:
 - Normalizar centralidades en el rango $[0, 1]$: No
 - Resultado: Figura 5.19
- Densidad: 0,003

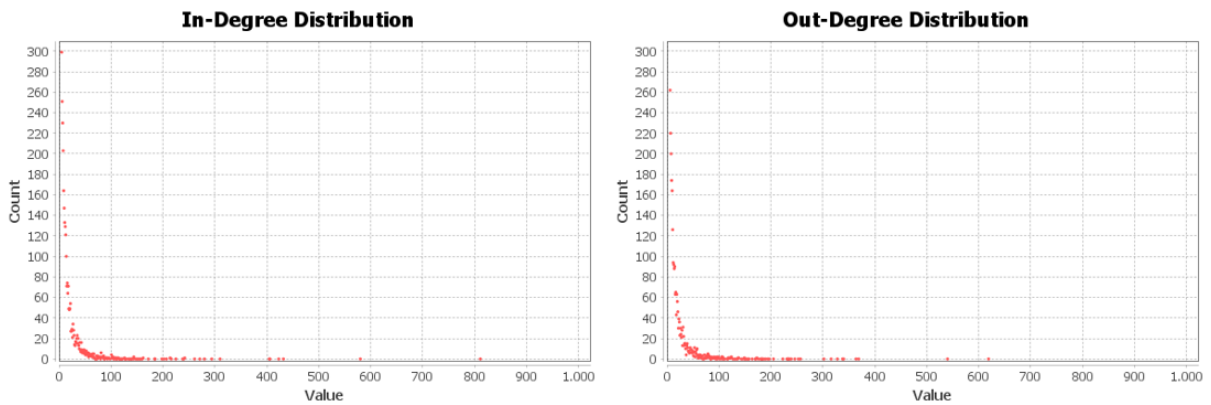


Figura 5.17: Ejemplo 1 (ES): Grado medio.

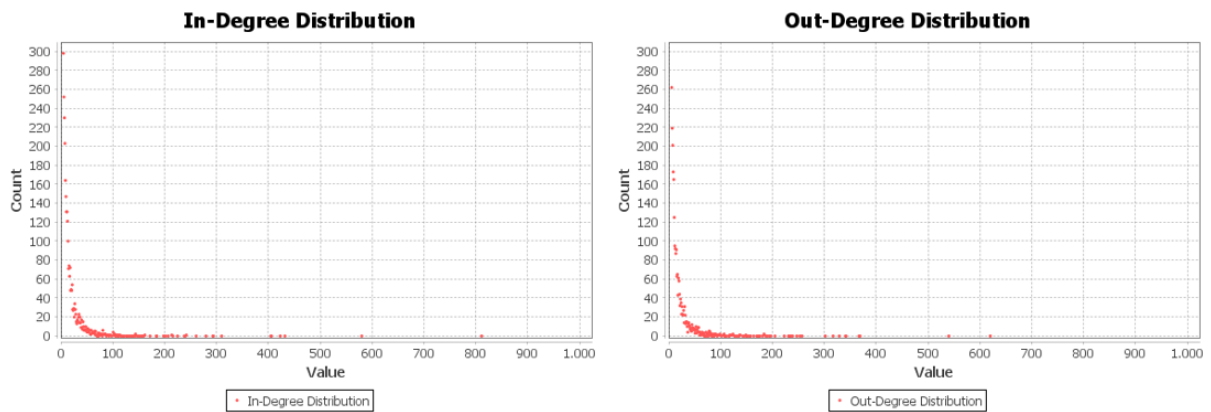


Figura 5.18: Ejemplo 1 (ES): Grado medio con pesos.

- HITS:
 - ✓ Parámetro:
 - ϵ : 0,0001
 - Resultado: Figura 5.20
- PageRank:
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - Resultado: Figura 5.21
- Componentes conexos: 1

2. Detección de comunidades

- Modularidad:
 - ✓ Parámetros:
 - Aleatorio: Sí

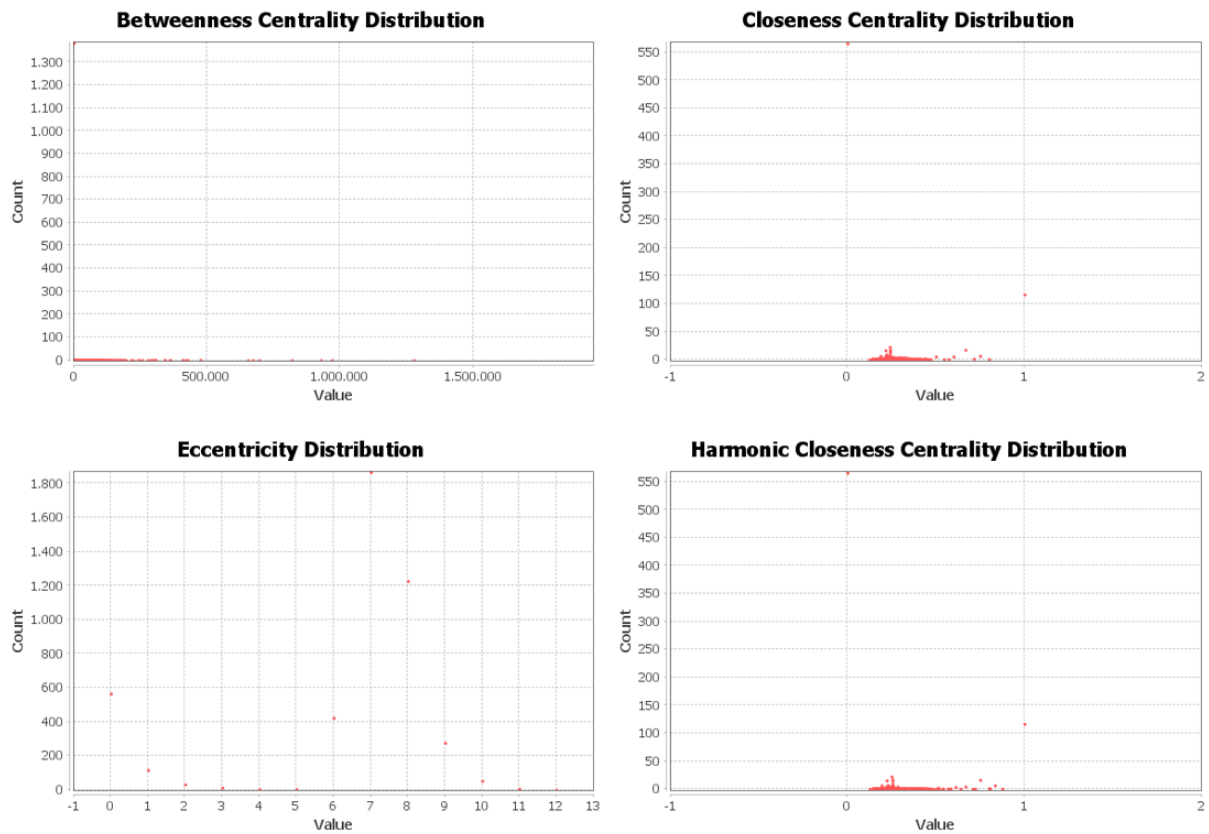


Figura 5.19: Ejemplo 1 (ES): Diámetro.

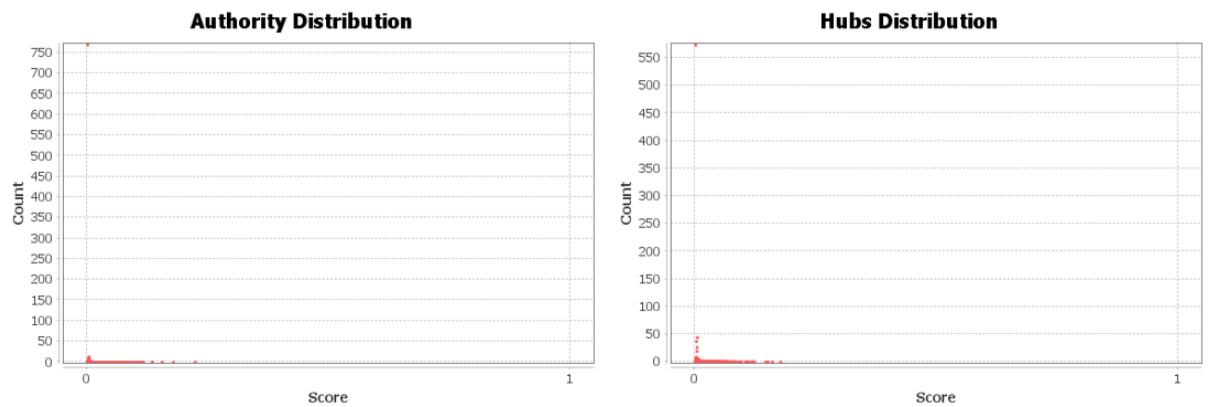


Figura 5.20: Ejemplo 1 (ES): HITS.

- Utilizar pesos: Sí
- Resolución: 1,0
- Resultados: Figura 5.22
 - Modularidad: 0,387
 - Modularidad con resolución: 0,387
 - Número de comunidades: 7

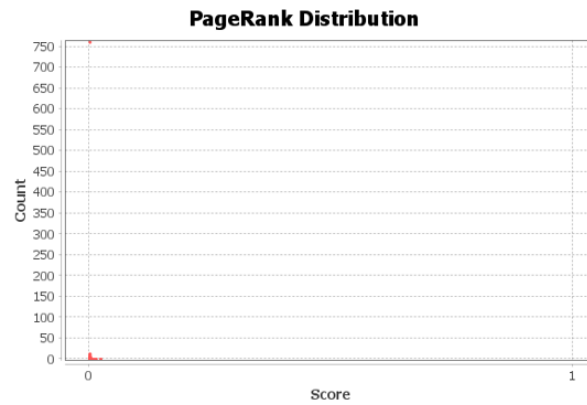


Figura 5.21: Ejemplo 1 (ES): PageRank.

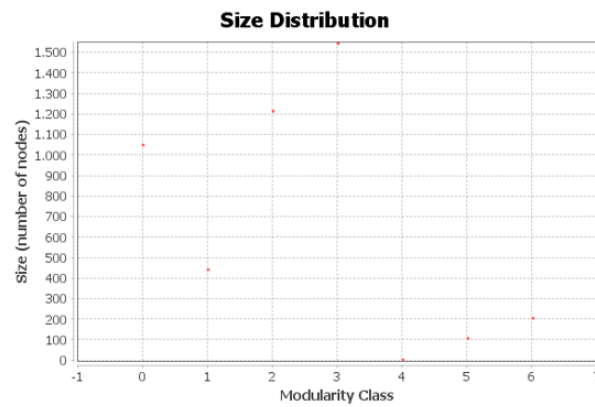


Figura 5.22: Ejemplo 1 (ES): Modularidad.

- Inferencia estadística:
 - Resultados: Figura 5.23
 - Longitud: 803109,355
 - Número de comunidades: 765

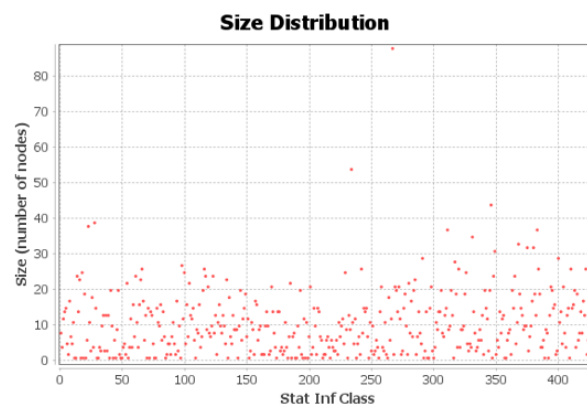


Figura 5.23: Ejemplo 1 (ES): Inferencia estadística.

3. Visión general de los nodos

- Coeficiente medio de *clustering*: 0,105
- Centralidad de vector propio:
 - ✓ Parámetros:
 - Número de iteraciones: 100
 - *Sum change*: 0,10078772023051352 \approx 0,101
 - Resultado: Figura 5.24.

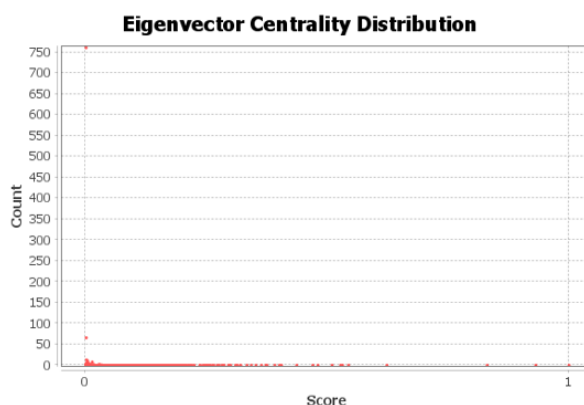


Figura 5.24: Ejemplo 1 (ES): Centralidad de vector propio.

4. Visión general de las aristas

- Longitud media de camino: 3,8621776477216763 \approx 3,862

Francia

En el caso de Francia se obtienen los valores que se observan a continuación.

1. Visión general de la red

- Grado medio: 17,364
En la Figura 5.25 se observa la distribución de los grados entrante y saliente, respectivamente.
- Grado medio con pesos: 17,371
En la Figura 5.26 se observa la distribución de los grados entrante y saliente con pesos, respectivamente.
- Diámetro: 14
Radio: 0
 - ✓ Parámetro:
 - Normalizar centralidades en el rango $[0, 1]$: No
 - Resultado: Figura 5.27
- Densidad: 0,003

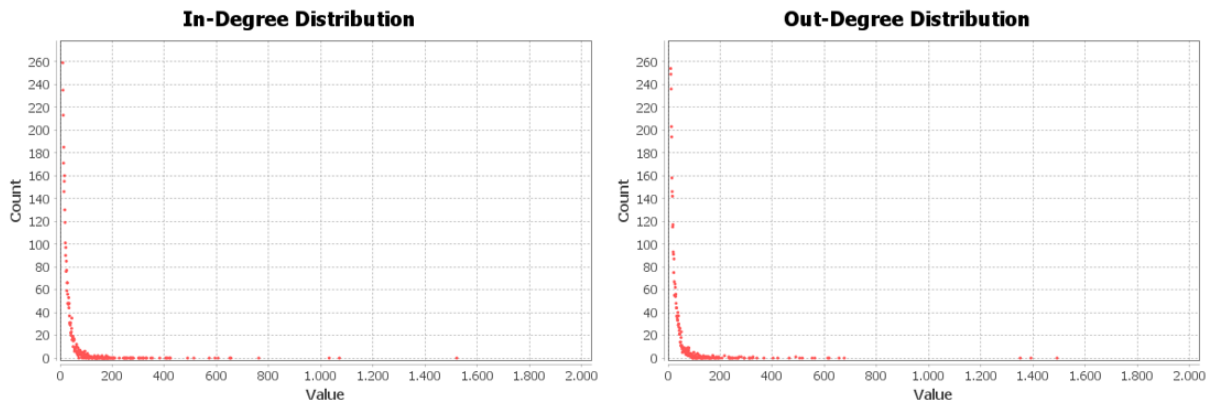


Figura 5.25: Ejemplo 1 (FR): Grado medio.

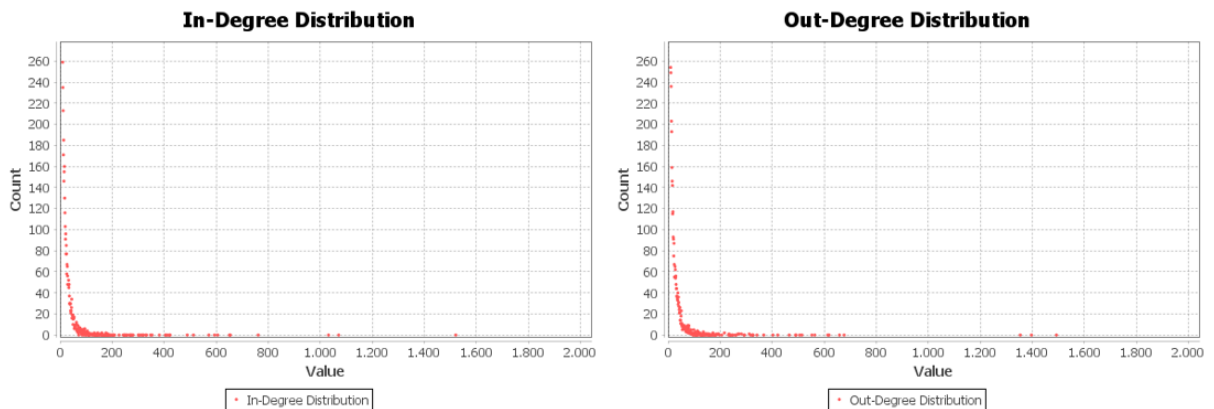


Figura 5.26: Ejemplo 1 (FR): Grado medio con pesos.

- HITS:
 - ✓ Parámetro:
 - ϵ : 0,0001
 - Resultado: Figura 5.28
- PageRank:
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - Resultado: Figura 5.29
- Componentes conexos: 2

2. Detección de comunidades

- Modularidad:
 - ✓ Parámetros:
 - Aleatorio: Sí

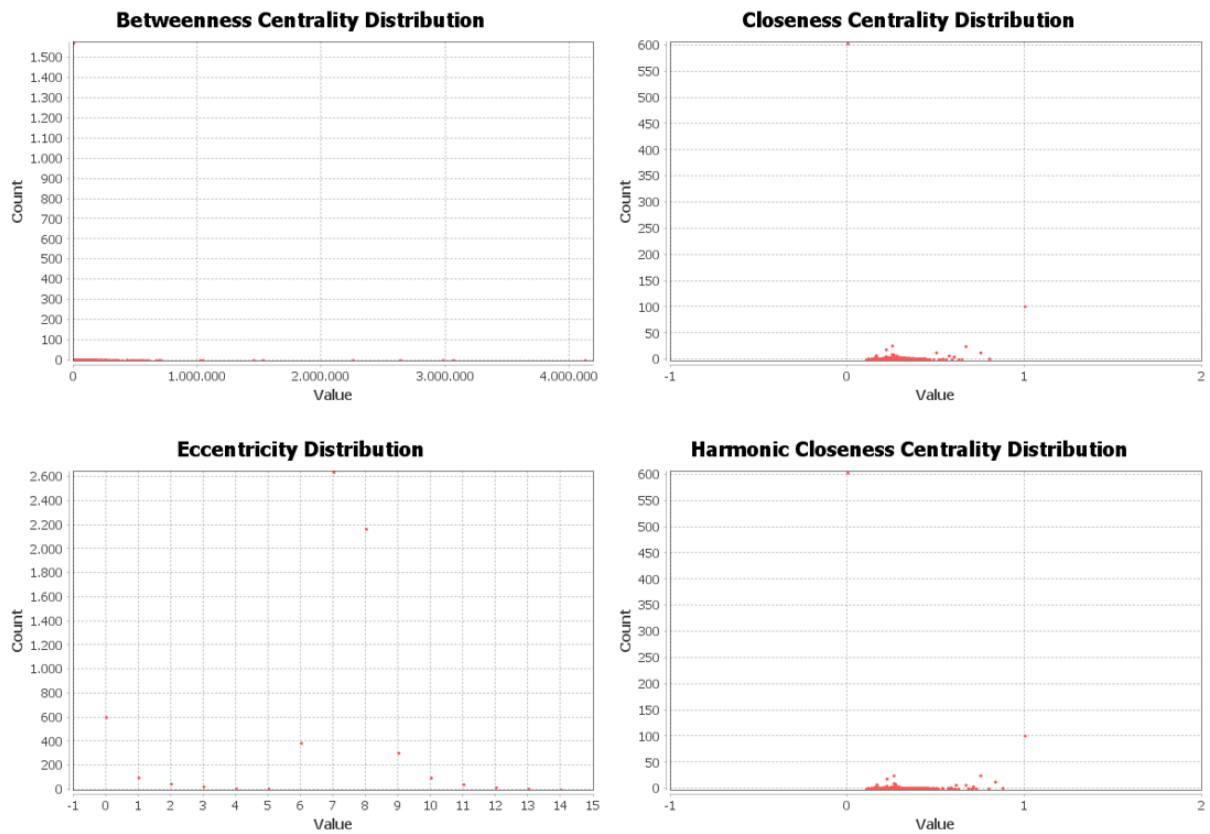


Figura 5.27: Ejemplo 1 (FR): Diámetro.

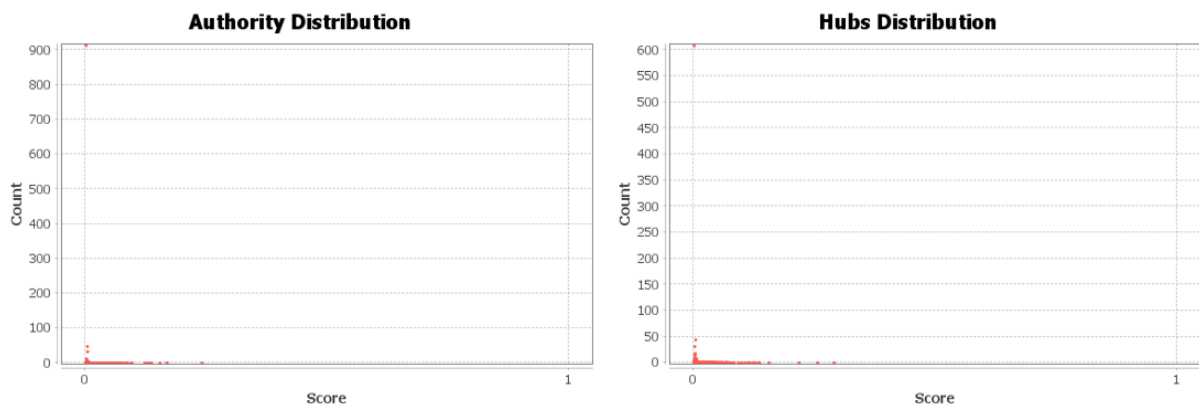


Figura 5.28: Ejemplo 1 (FR): HITS.

- Utilizar pesos: Sí
- Resolución: 1,0
- Resultados: Figura 5.30
 - Modularidad: 0,322
 - Modularidad con resolución: 0,322
 - Número de comunidades: 9

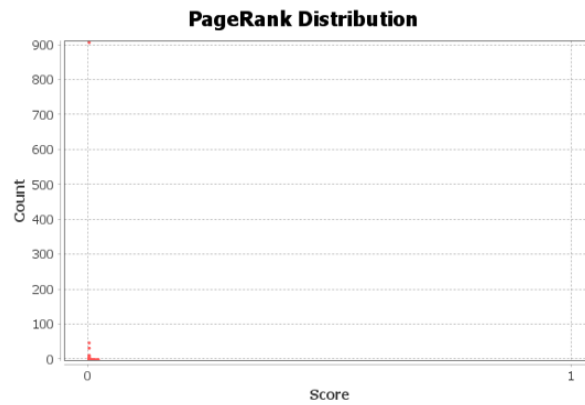


Figura 5.29: Ejemplo 1 (FR): PageRank.

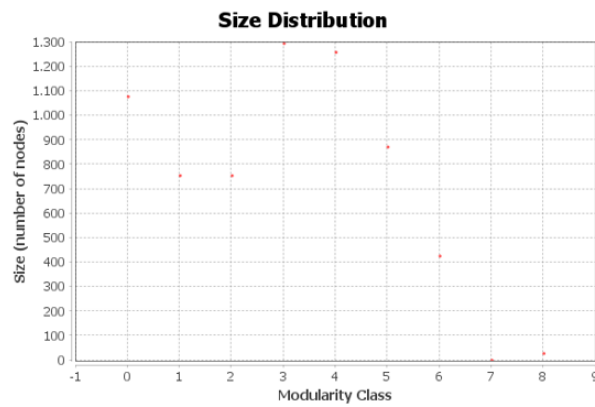


Figura 5.30: Ejemplo 1 (FR): Modularidad.

- Inferencia estadística:
 - Resultados: Figura 5.31
 - Longitud: 803109,355
 - Número de comunidades: 765

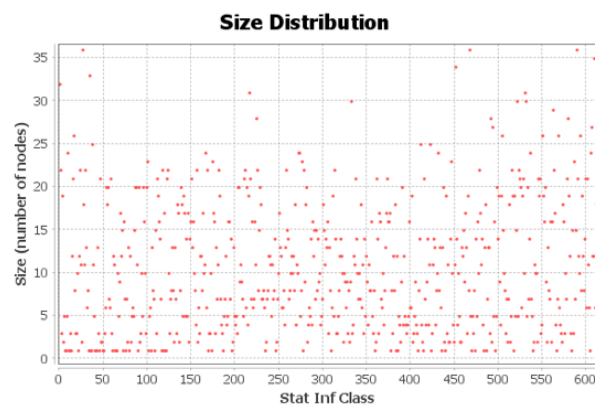


Figura 5.31: Ejemplo 1 (FR): Inferencia estadística.

3. Visión general de los nodos

- Coeficiente medio de *clustering*: 0,098
- Centralidad de vector propio:
 - ✓ Parámetros:
 - Número de iteraciones: 100
 - *Sum change*: 0,0949244707491212 \approx 0,095
 - Resultado: Figura 5.32.

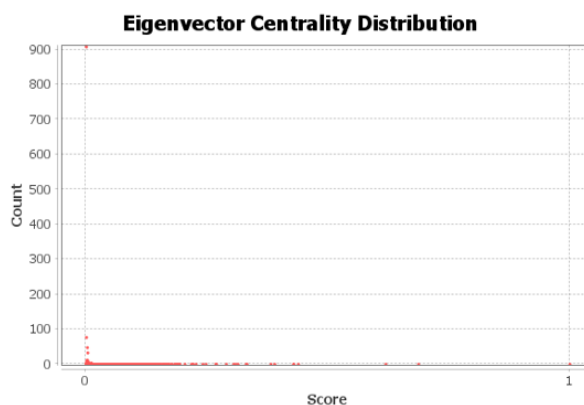


Figura 5.32: Ejemplo 1 (FR): Centralidad de vector propio.

4. Visión general de las aristas

- Longitud media de camino: 3,6002098018863884 \approx 3,6

Brasil

En el caso de Brasil se obtienen los valores que se observan a continuación.

1. Visión general de la red

- Grado medio: 16,56
En la Figura 5.33 se observa la distribución de los grados entrante y saliente, respectivamente.
- Grado medio con pesos: 16,587
En la Figura 5.34 se observa la distribución de los grados entrante y saliente con pesos, respectivamente.
- Diámetro: 10
Radio: 0
 - ✓ Parámetro:
 - Normalizar centralidades en el rango $[0, 1]$: No
 - Resultado: Figura 5.35
- Densidad: 0,009

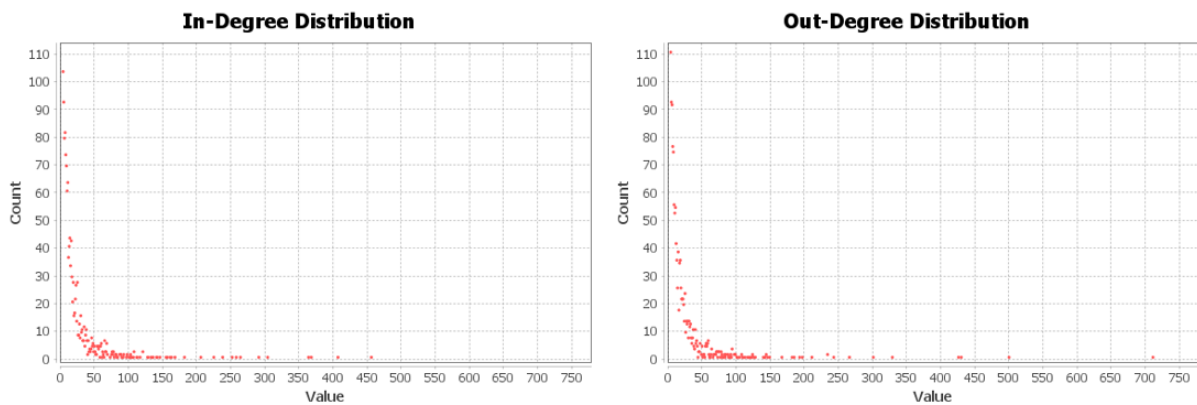


Figura 5.33: Ejemplo 1 (PTBR): Grado medio.

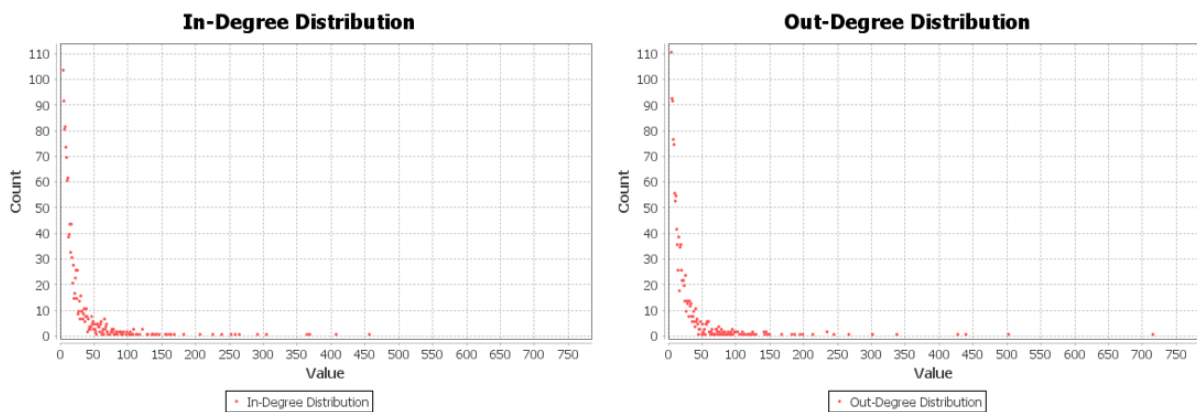


Figura 5.34: Ejemplo 1 (PTBR): Grado medio con pesos.

- HITS:
 - ✓ Parámetro:
 - ϵ : 0,0001
 - Resultado: Figura 5.36
- PageRank:
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - Resultado: Figura 5.37
- Componentes conexos: 1

2. Detección de comunidades

- Modularidad:
 - ✓ Parámetros:
 - Aleatorio: Sí

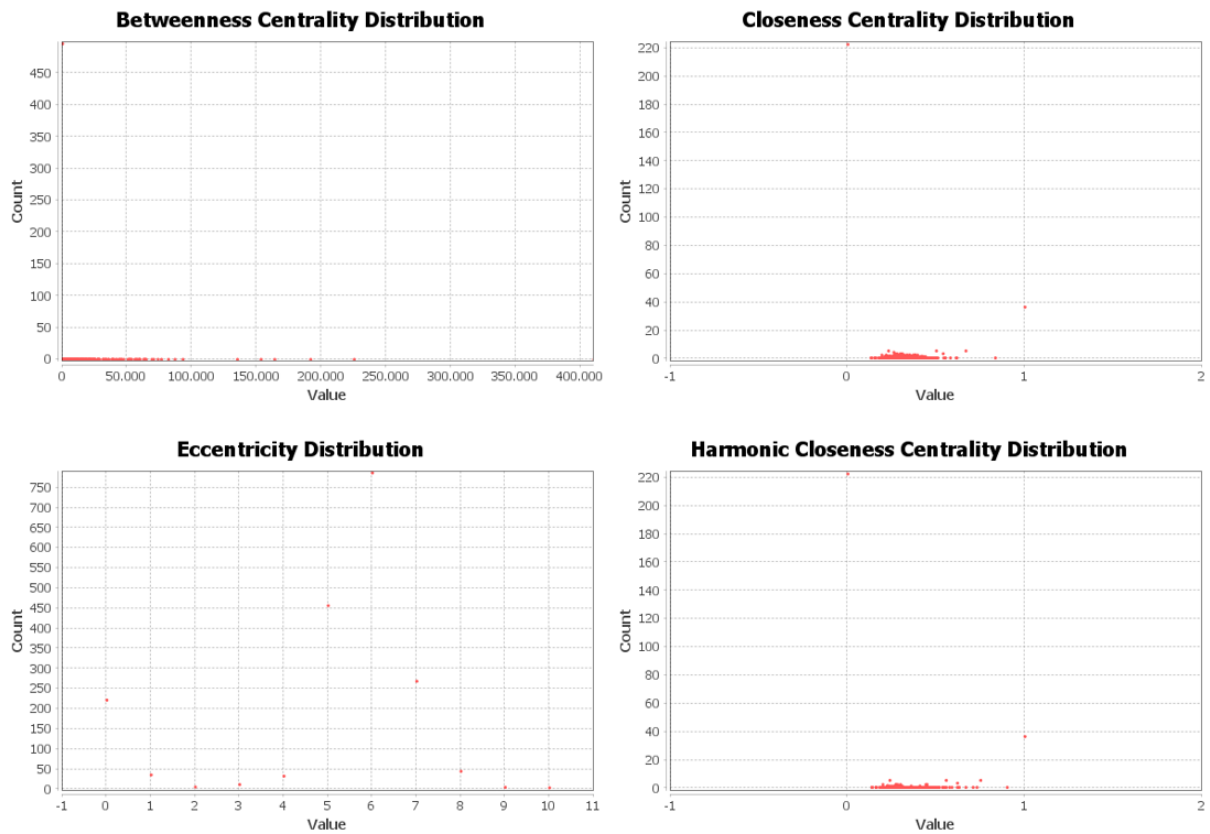


Figura 5.35: Ejemplo 1 (PTBR): Diámetro.

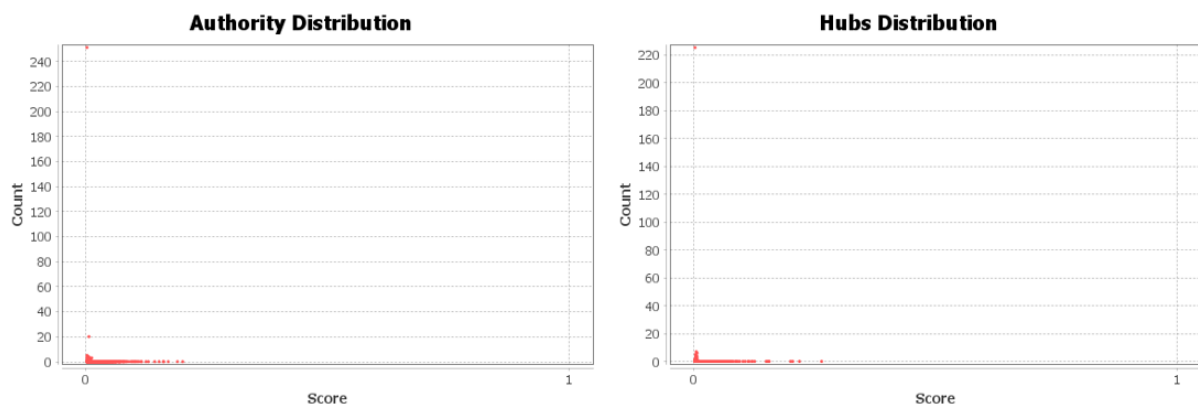


Figura 5.36: Ejemplo 1 (PTBR): HITS.

- Utilizar pesos: Sí
- Resolución: 1,0
- Resultados: Figura 5.38
 - Modularidad: 0,274
 - Modularidad con resolución: 0,274
 - Número de comunidades: 6

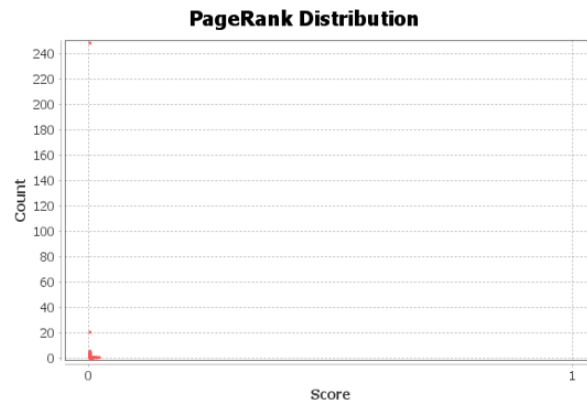


Figura 5.37: Ejemplo 1 (PTBR): PageRank.

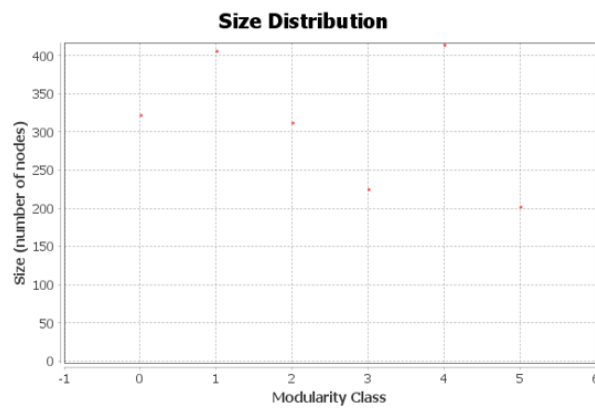


Figura 5.38: Ejemplo 1 (PTBR): Modularidad.

- Inferencia estadística:
 - Resultados: Figura 5.39
 - Longitud: 803109,355
 - Número de comunidades: 765

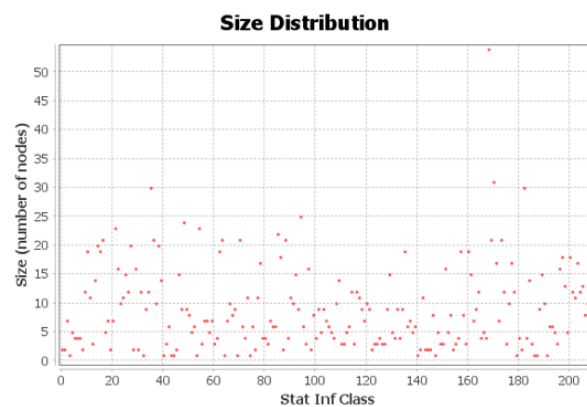


Figura 5.39: Ejemplo 1 (PTBR): Inferencia estadística.

3. Visión general de los nodos

- Coeficiente medio de *clustering*: 0,151
- Centralidad de vector propio:
 - ✓ Parámetros:
 - Número de iteraciones: 100
 - *Sum change*: 0,06587515848879211 \approx 0,066
 - Resultado: Figura 5.40.

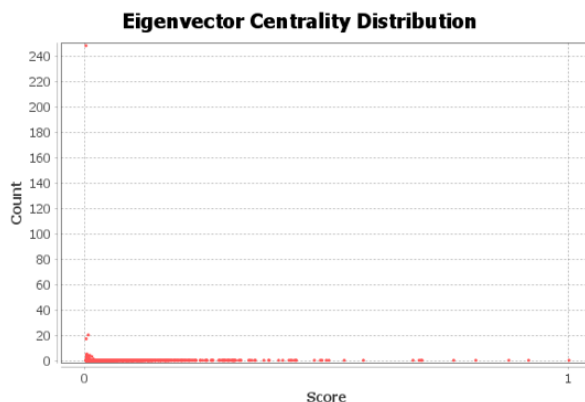


Figura 5.40: Ejemplo 1 (PTBR): Centralidad de vector propio.

4. Visión general de las aristas

- Longitud media de camino: 3,4101713026863685 \approx 3,41

Rusia

En el caso de Rusia se obtienen los valores que se observan a continuación.

1. Visión general de la red

- Grado medio: 8,665
En la Figura 5.41 se observa la distribución de los grados entrante y saliente, respectivamente.
- Grado medio con pesos: 8,673
En la Figura 5.42 se observa la distribución de los grados entrante y saliente con pesos, respectivamente.
- Diámetro: 13
Radio: 0
 - ✓ Parámetro:
 - Normalizar centralidades en el rango [0, 1]: No
 - Resultado: Figura 5.43
- Densidad: 0,002

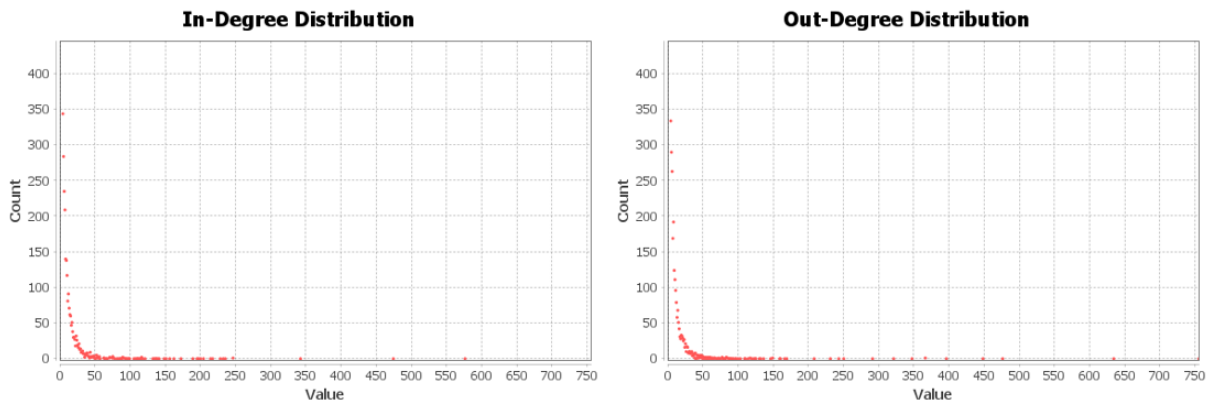


Figura 5.41: Ejemplo 1 (RU): Grado medio.

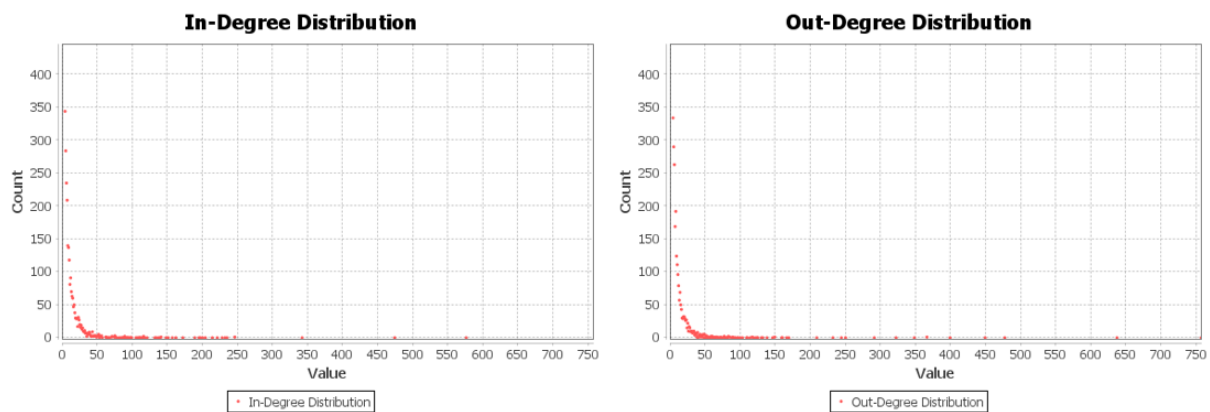


Figura 5.42: Ejemplo 1 (RU): Grado medio con pesos.

- HITS:
 - ✓ Parámetro:
 - ϵ : 0,0001
 - Resultado: Figura 5.44
- PageRank:
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - Resultado: Figura 5.45
- Componentes conexos: 2

2. Detección de comunidades

- Modularidad:
 - ✓ Parámetros:
 - Aleatorio: Sí

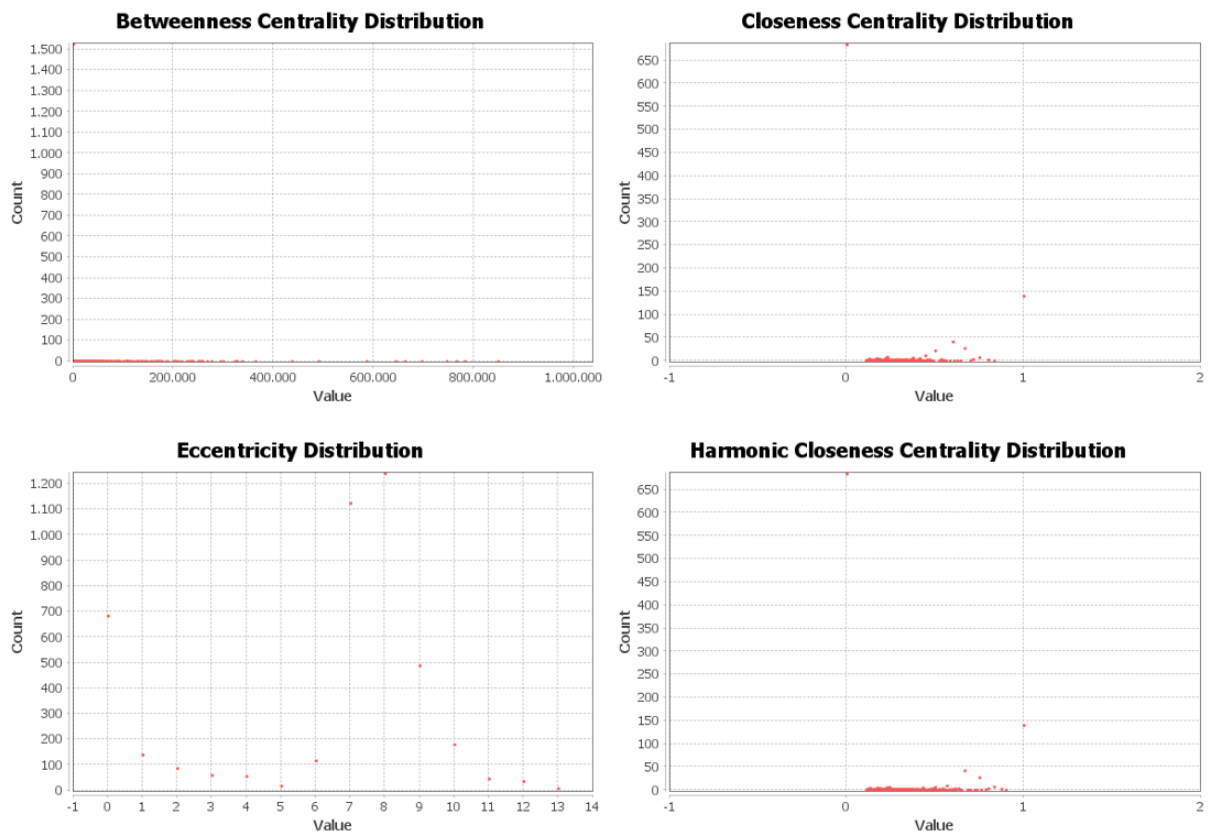


Figura 5.43: Ejemplo 1 (RU): Diámetro.

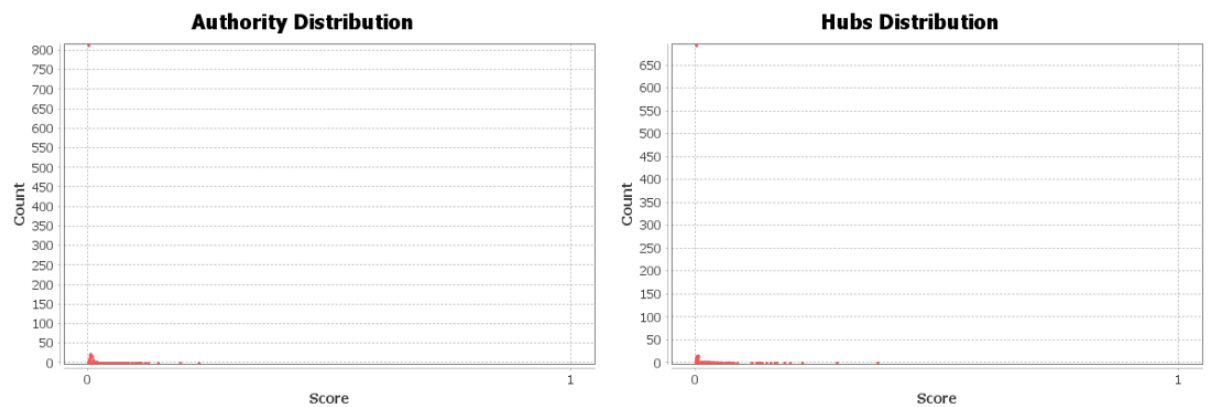


Figura 5.44: Ejemplo 1 (RU): HITS.

- Utilizar pesos: Sí
- Resolución: 1,0
- Resultados: Figura 5.46
 - Modularidad: 0,325
 - Modularidad con resolución: 0,325
 - Número de comunidades: 14

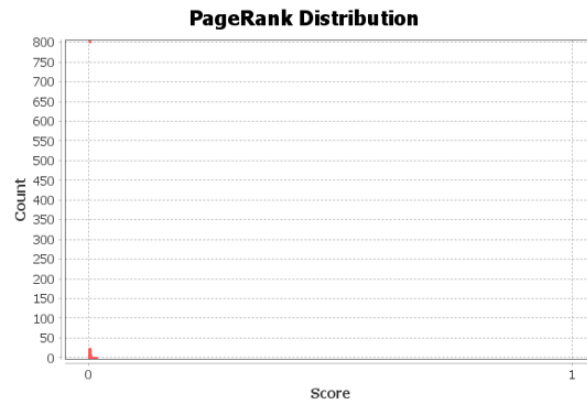


Figura 5.45: Ejemplo 1 (RU): PageRank.

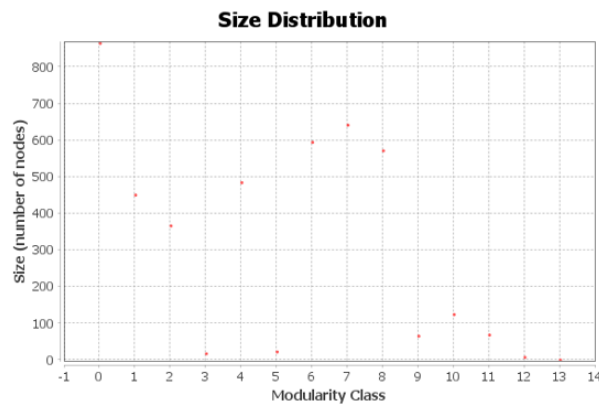


Figura 5.46: Ejemplo 1 (RU): Modularidad.

- Inferencia estadística:
 - Resultados: Figura 5.47
 - Longitud: 803109,355
 - Número de comunidades: 765

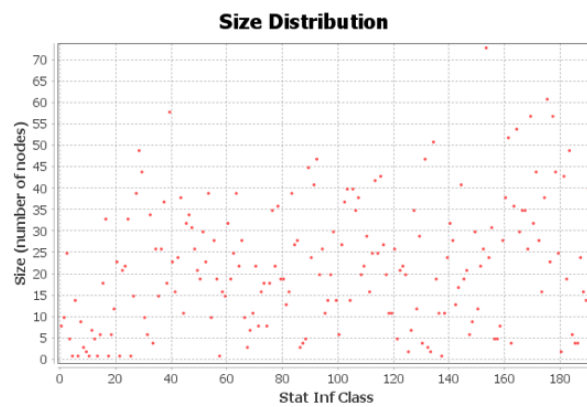


Figura 5.47: Ejemplo 1 (RU): Inferencia estadística.

3. Visión general de los nodos

- Coeficiente medio de *clustering*: 0,073
- Centralidad de vector propio:
 - ✓ Parámetros:
 - Número de iteraciones: 100
 - *Sum change*: 0,07903694791305273 \approx 0,079
 - Resultado: Figura 5.48.

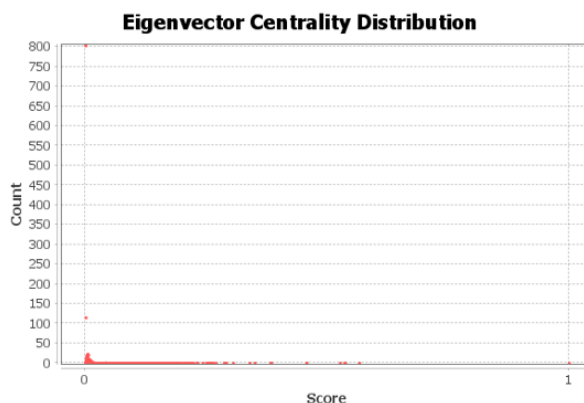


Figura 5.48: Ejemplo 1 (RU): Centralidad de vector propio.

4. Visión general de las aristas

- Longitud media de camino: 4,221816165650243 \approx 4,222

5.1.6. Discusión

El análisis de las redes sociales en Twitch, segmentadas por países, revela patrones interesantes en cuanto a la estructura y características de las redes. Para evaluar estas estructuras, se consideran varias métricas clave: el grado medio, el diámetro, la densidad, la modularidad, el coeficiente de *clustering* y la longitud media del camino. En la Tabla 5.1 se presenta una discusión detallada de los resultados obtenidos.

A continuación, se interpretan los resultados dependiendo de la estadística.

Grado medio

Francia presenta el mayor grado medio 17,364, mientras que Rusia tiene el menor 8,665.

Un grado medio elevado sugiere una red con alta conectividad entre los usuarios. Esto podría indicar una mayor interacción entre los usuarios de Twitch en Francia, lo cual es relevante para detectar nodos con alta influencia (potenciales *influencers*). En contraste, un grado medio más bajo en Rusia sugiere que la red es más dispersa, con menos conexiones entre usuarios.

País	Grado medio	Diámetro	Longitud media de camino	Densidad	Modularidad	Coef. clustering
Alemania	16,253	14	3,742	0,002	0,276	0,096
R. Unido	13,485	12	3,786	0,002	0,364	0,066
España	12,916	12	3,862	0,003	0,387	0,105
Francia	17,364	14	3,6	0,003	0,322	0,098
Brasil	16,560	10	3,41	0,009	0,274	0,151
Rusia	8,665	13	4,222	0,002	0,325	0,073

Tabla 5.1: Comparativa de características de la red entre países.

Diámetro y longitud media del camino

El diámetro más pequeño se observa en Brasil $D = 10$, mientras que el más grande se encuentra en Alemania y Francia $D = 14$. La longitud media del camino varía, con Rusia mostrando el valor más alto $L = 4,222$ y Brasil el más bajo $L = 3,41$.

Un diámetro grande y una longitud media del camino alta sugieren que la red es extensa y que los usuarios pueden estar más aislados, lo que aumenta la dificultad de propagación de la información. Por otro lado, un diámetro pequeño con una longitud media del camino baja (Brasil) indica una red más compacta, donde la información se puede propagar más rápidamente entre los nodos.

Densidad

Brasil tiene la densidad más alta $\rho = 0,009$, mientras que Alemania, Reino Unido y Rusia presentan las densidades más bajas $\rho = 0,002$. Aún así, la densidad de Brasil dista mucho de 1.

Una densidad baja (inferior a 0,5) indica una red esparcida, lo que podría reflejar una menor interacción entre usuarios en esos países, posiblemente debido a barreras culturales, lingüísticas o de otro tipo. Una densidad alta (superior a 0,5), sugiere una comunidad más cohesionada, donde los usuarios están más conectados entre sí.

Modularidad y número de comunidades

La modularidad más alta se observa en España $Q = 0,387$ y la más baja en Brasil $Q = 0,274$.

Una alta modularidad indica una estructura de comunidad fuerte, con grupos bien definidos y con alta densidad de conexiones internas en relación con las conexiones entre grupos. Esto es común en redes sociales donde los usuarios forman comunidades alrededor de intereses comunes o grupos sociales. En contraste, una modularidad baja sugiere una estructura menos definida de comunidades, lo que podría indicar una red más homogénea sin una división clara en subgrupos.

Coefficiente de *clustering*

Brasil muestra el coeficiente de *clustering* más alto $C = 0,151$, mientras que el Reino Unido tiene el más bajo $C = 0,066$.

Un alto coeficiente de *clustering* indica una red donde los amigos de un usuario tienden a ser amigos entre sí, lo que podría reflejar una red con fuerte cohesión social o nichos de interacción intensiva. Un coeficiente bajo sugiere menos agrupamientos y, posiblemente, una red más orientada a conexiones más dispersas y menos clanes.

5.1.7. Conclusiones

A partir del análisis anterior, se puede concluir que la red social Twitch muestra variaciones significativas en su estructura dependiendo del país, lo que refleja diferentes dinámicas de interacción entre los usuarios. Francia y Brasil destacan por sus altas conectividades y coeficientes de *clustering*, lo que sugiere redes más cohesionadas con mayores probabilidades de difusión rápida de información. Por otro lado, Rusia y el Reino Unido muestran redes más dispersas, con menor densidad y menor grado medio, lo que puede reflejar comunidades menos integradas y una propagación de la información más lenta.

Este análisis permite inferir que las características culturales y sociales de cada país influyen en la forma en que se estructuran las redes sociales en línea. La formalización de estos patrones mediante métricas matemáticas proporciona una base sólida para comprender y optimizar la estructura de redes en contextos específicos, lo cual es esencial para aplicaciones que van desde la identificación de *influencers* hasta la optimización de la difusión de contenido en plataformas como Twitch.

5.2. Grafo no dirigido

En esta sección se expone el segundo ejemplo de análisis de redes con Gephi. Se trata de un grafo no dirigido [14].

5.2.1. Descripción del *dataset*

El *dataset* “Círculos sociales: Facebook” consiste en “círculos” (o “listas de amigos”) de Facebook. Los datos son recolectados a través de una aplicación de encuesta de Facebook. El conjunto de datos incluye características de los nodos (perfiles), círculos y redes ego ¹.

Los datos de Facebook son anonimizados reemplazando los identificadores internos de Facebook por nuevos valores. Aunque se proporcionan vectores de características, la interpretación de esas características se oscurece. Por ejemplo, donde el conjunto de datos original contiene una característica “política = Partido Demócrata”, los nuevos datos simplemente contienen “política = característica anonimizada 1”. Por lo tanto, con los datos anonimizados es posible determinar si dos usuarios tienen las mismas afiliaciones políticas, pero no qué representan sus afiliaciones políticas individuales.

Nota 38. *También hay datos disponibles de Google+ y Twitter.*

A continuación, se listan una serie de estadísticas compiladas combinando las redes ego, incluyendo los nodos ego (junto con una arista hacia cada uno de sus amigos) [14].

¹“Ego” es un nodo “focal” individual. Una red tiene tantos egos como nodos. Los egos pueden ser personas, grupos, organizaciones o sociedades enteras. “Vecindario” es la colección del ego y de todos los nodos con los que el ego tiene una conexión a cierta longitud de camino.

- Nodos: 4039
- Aristas: 88234
- Nodos en la Componente Conexo Débil o *Weakly Connected Component* (WCC) más grande: 4039 miles
- Aristas en la WCC más grande: 88234 miles
- Nodos en la Componente Fuertemente Conexo o *Strongly Connected Component* (SCC) más grande: 4039 miles
- Aristas en la SCC más grande: 88234 miles
- Coeficiente de agrupamiento promedio: 0,606
- Número de triángulos: 1612010
- Fracción de triángulos cerrados: 0,265
- Diámetro (camino más corto más largo): 8
- Diámetro efectivo del 90 %: 4,7

La jerarquía de archivos del *dataset* se lista a continuación; hay un directorio por cada `nodeId` estudiado.

- `facebook.tar.gz`: Datos de Facebook (10 redes, anonimizadas).
 - `nodeId`: Toma los valores 0, 107, 348, 414, 686, 698, 1684, 1912, 3437, 3980.
 - `nodeId.edges`: Los bordes en la red ego para el nodo “`nodeId`”. Los bordes son no dirigidos para Facebook y dirigidos para Twitter y Google+ (Nota 38). El nodo “ego” no aparece, pero se asume que sigue a cada “`nodeId`” que aparece en este archivo.
 - `nodeId.circles`: El conjunto de círculos para el nodo ego. Cada línea contiene un círculo, que consiste en una serie de ids de nodos. La primera entrada en cada línea es el nombre del círculo.
 - `nodeId.feats`: Las características de cada uno de los nodos que aparecen en el archivo de bordes.
 - `nodeId.ego_feats`: Las características para el usuario ego.
 - `nodeId.feats_names`: Los nombres de cada una de las dimensiones de las características. Las características son 1 si el usuario tiene esta propiedad en su perfil, y 0 en caso contrario.
- `facebook_combined.txt.gz`: Aristas de todas las redes ego combinadas.
- `readme-Ego.txt`: Descripción de los archivos.

Se puede explorar y descargar el *dataset* en la siguiente URL: <https://snap.stanford.edu/data/ego-Facebook.html>.

5.2.2. Importación de los datos

En primer lugar, se cambia el nombre de las columnas del archivo `facebook_combined.txt` para facilitar su comprensión por parte de la herramienta *software*. En concreto, la primera pasa a ser *source* y la segunda se convierte en *target*. Antes de acudir a la herramienta, el archivo se convierte en Valores Separados por Comas o *Comma Separated Values* (CSV).

Se abre la aplicación de escritorio Gephi. Se selecciona “Archivo” y “Nuevo Proyecto”. De nuevo, se selecciona “Archivo”, “Importar Hoja de Cálculo” y se elige el archivo `facebook_combined.csv`. En el asistente de importación, se selecciona “Tabla de aristas” y se asignan las columnas *source* y *target* como nodos fuente y destino, respectivamente. Finalmente, se hace clic en “Finalizar” y, luego, en “Aceptar”.

5.2.3. Contexto

En la pestaña “Contexto” aparece un recuento de nodos y aristas.

- Nodos: 4039
- Aristas: 88234

5.2.4. Estadísticas

En la pestaña “Estadísticas”, se hace clic en “Ejecutar” para cada una de las medidas y parámetros que definen un grafo.

5.2.5. Resultados

Se muestran los resultados obtenidos para las estadísticas de la Sección 5.2.4. En este ejemplo se prueba con distintos valores para cada uno de los parámetros para comprobar cómo varían los resultados en cada caso.

Se obtienen los valores que se observan a continuación.

1. Visión general de la red

- Grado medio: 43,691

En la Figura 5.49 se observa la distribución del grado.

- Grado medio con pesos: 43,691

La figura de la distribución del grado con pesos en este caso coincide con la Figura 5.49, al igual que el valor de ambos grados medios.

- Diámetro: 8

Radio: 4

- ✓ Parámetro:

- Normalizar centralidades en el rango $[0, 1]$: No

- Resultado: Figura 5.50

- ✓ Parámetro:

- Normalizar centralidades en el rango $[0, 1]$: Sí

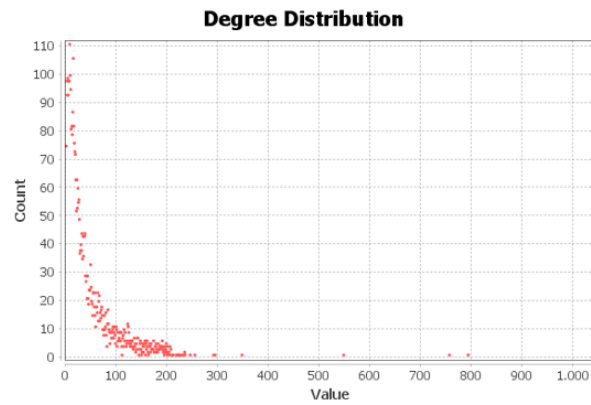


Figura 5.49: Ejemplo 2: Grado medio.

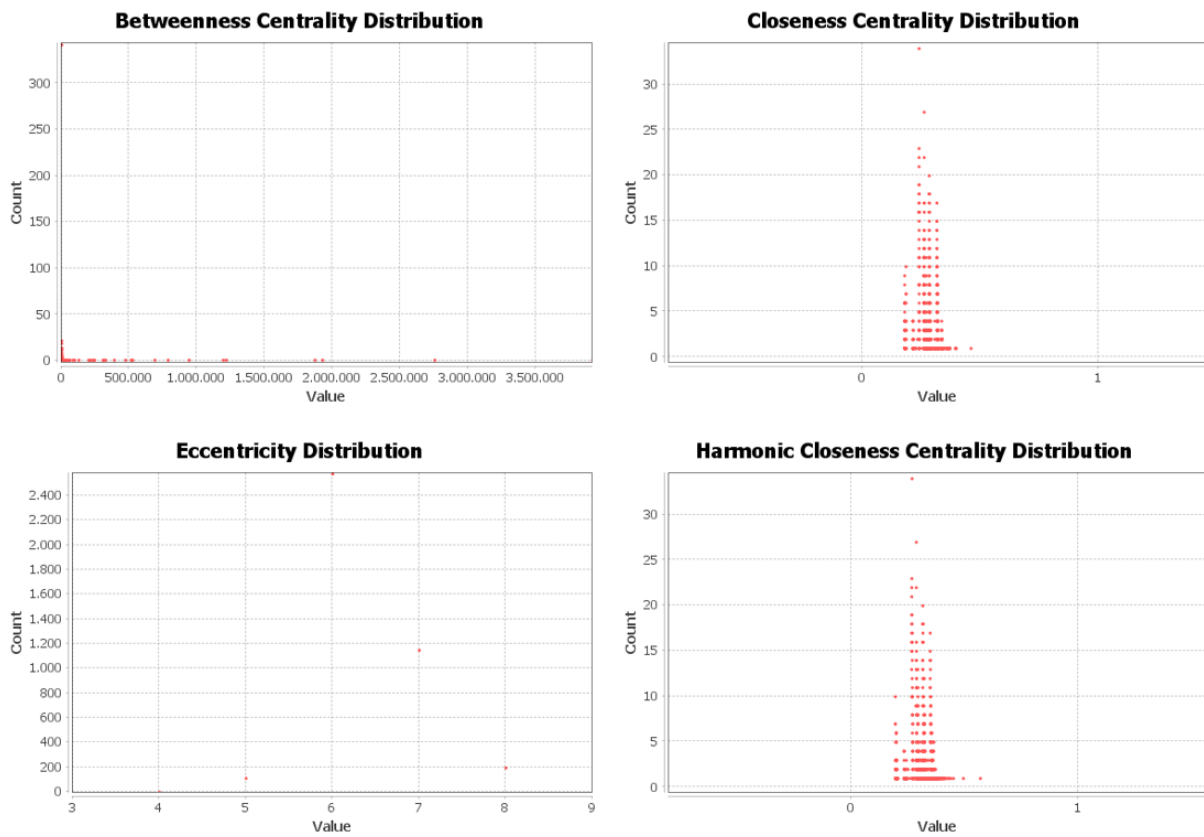


Figura 5.50: Ejemplo 2: Diámetro sin normalizar centralidades en el rango $[0, 1]$.

- Resultado: Figura 5.51
- Densidad: 0,011
- HITS:
 - ✓ Parámetro:
 - ϵ : 0,001
 - ✓ Parámetro:

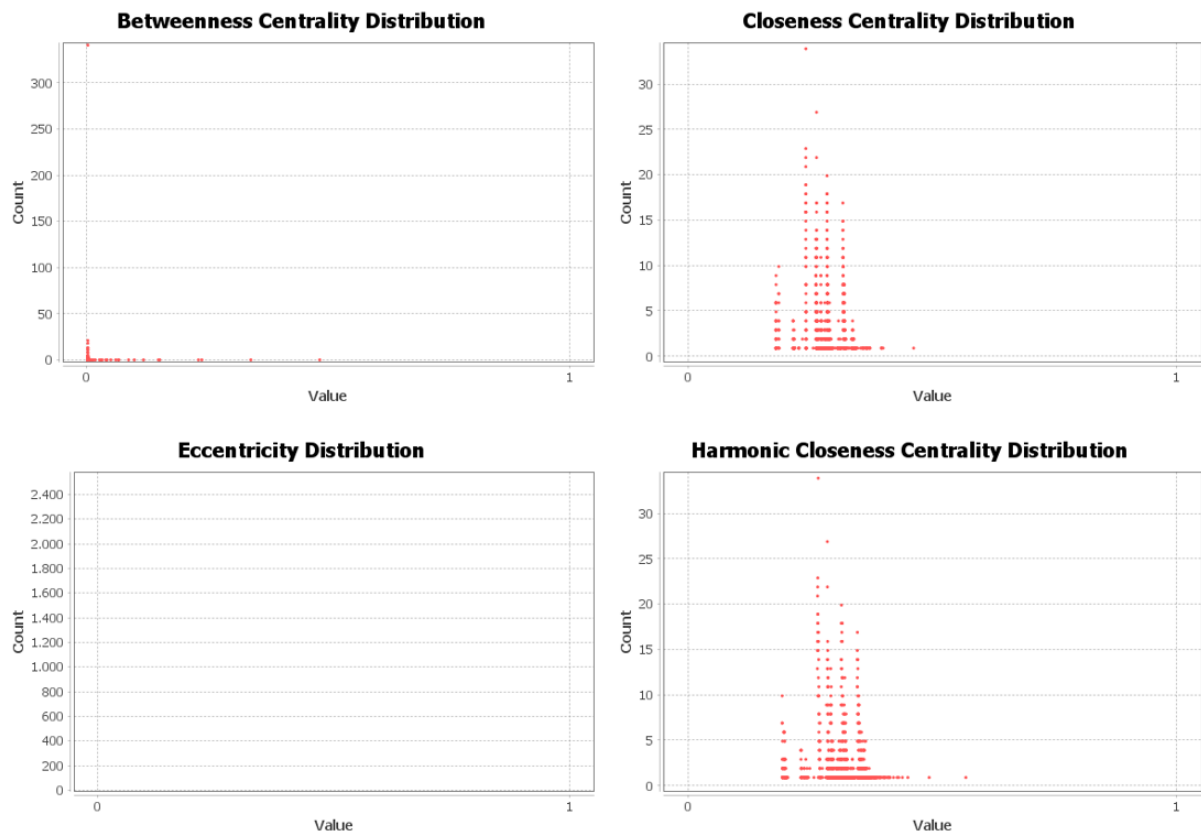


Figura 5.51: Ejemplo 2: Diámetro normalizando centralidades en el rango $[0, 1]$.

- ϵ : 0,0001
 - ✓ Parámetro:
 - ϵ : 0,00001
 - ✓ Parámetro:
 - ϵ : 0,000001
 - Resultado: Figura 5.52
- PageRank:
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,001
 - Utilizar peso de aristas: Sí
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,0001

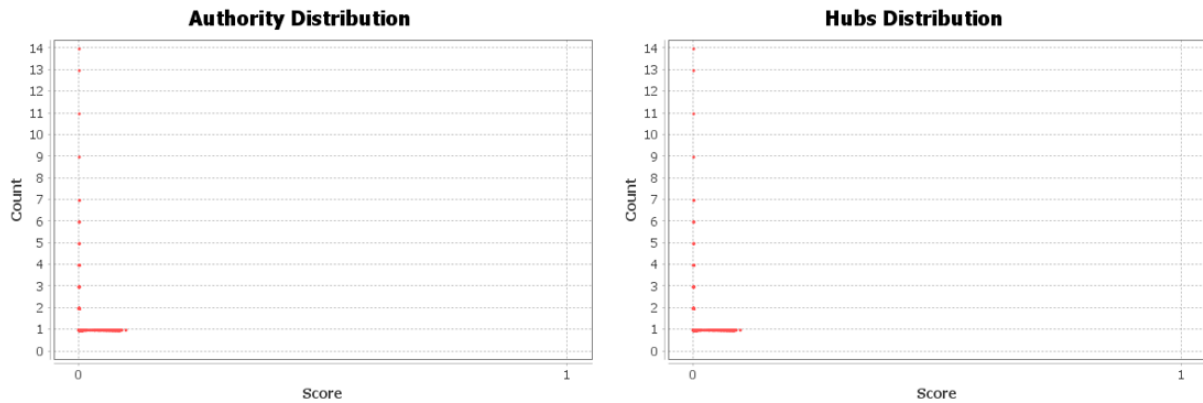


Figura 5.52: Ejemplo 2: HITS.

- Utilizar peso de aristas: No
 - ✓ Parámetros:
 - Probabilidad (p): 0,85
 - ϵ : 0,0001
 - Utilizar peso de aristas: Sí
 - ✓ Parámetros:
 - Probabilidad (p): 0,5
 - ϵ : 0,001
 - Utilizar peso de aristas: No
 - ✓ Parámetros:
 - Probabilidad (p): 0,5
 - ϵ : 0,001
 - Utilizar peso de aristas: Sí
 - ✓ Parámetros:
 - Probabilidad (p): 0,5
 - ϵ : 0,0001
 - Utilizar peso de aristas: No
 - ✓ Parámetros:
 - Probabilidad (p): 0,5
 - ϵ : 0,0001
 - Utilizar peso de aristas: Sí
 - Resultado: Figura 5.53
 - Componentes conexos: 1
2. Detección de comunidades
- Modularidad:
 - ✓ Parámetros:
 - Aleatorio: Sí



Figura 5.53: Ejemplo 2: PageRank.

- Utilizar pesos: Sí
- Resolución: 0,1
- Resultados: Figura 5.54 (izquierda y arriba)
 - Modularidad: 0,460
 - Modularidad con resolución: 0,030
 - Número de comunidades: 147
- ✓ Parámetros:
 - Aleatorio: Sí
 - Utilizar pesos: Sí
 - Resolución: 0,5
- Resultados: Figura 5.54 (derecha y arriba)
 - Modularidad: 0,832
 - Modularidad con resolución: 0,359
 - Número de comunidades: 22
- ✓ Parámetros:
 - Aleatorio: Sí
 - Utilizar pesos: Sí
 - Resolución: 1,0
- Resultados: Figura 5.54 (izquierda y abajo)
 - Modularidad: 0,834
 - Modularidad con resolución: 0,834
 - Número de comunidades: 15
- ✓ Parámetros:
 - Aleatorio: Sí
 - Utilizar pesos: Sí
 - Resolución: 2,0
- Resultados: Figura 5.54 (derecha y abajo)
 - Modularidad: 0,827

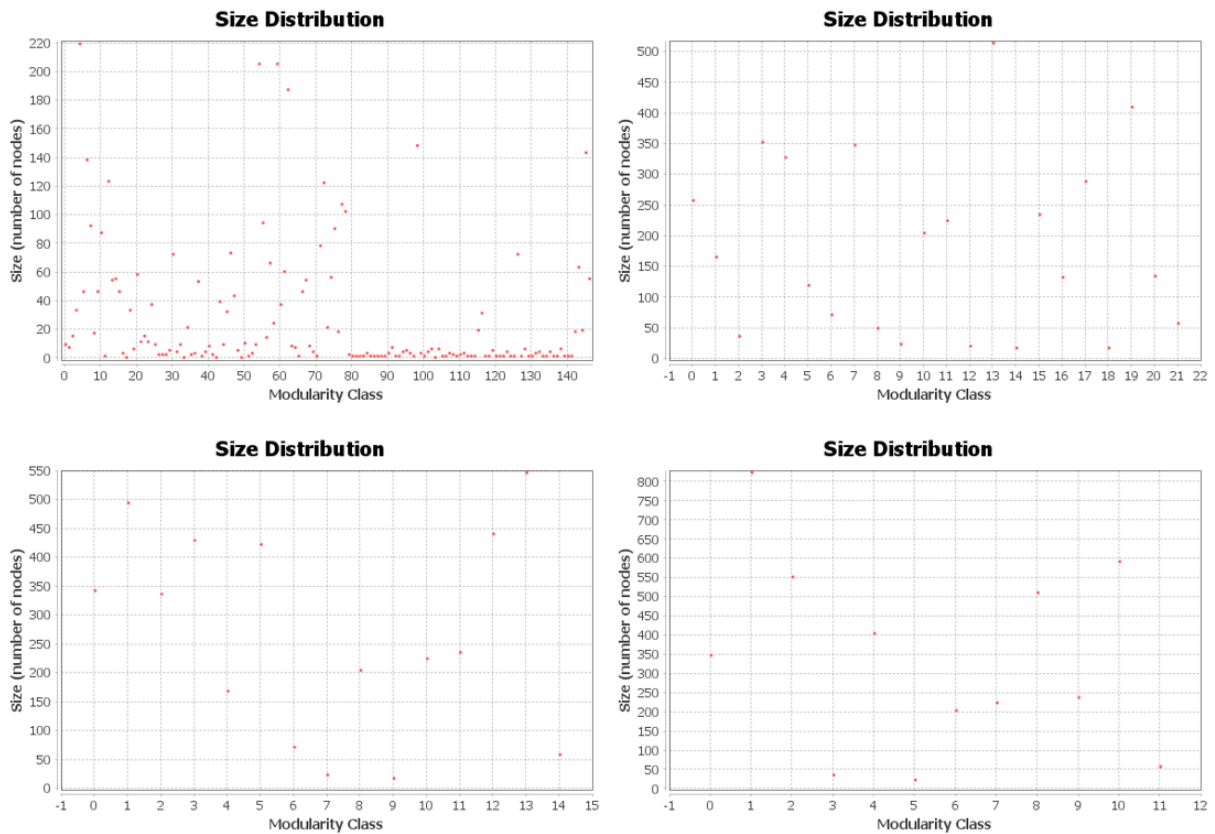


Figura 5.54: Ejemplo 2: Modularidad con aleatoriedad y uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).

- Modularidad con resolución: 1,798
- Número de comunidades: 12
- ✓ Parámetros:
 - Aleatorio: Sí
 - Utilizar pesos: No
 - Resolución: 0,1
- Resultados: Figura 5.55 (izquierda y arriba)
 - Modularidad: 0,456
 - Modularidad con resolución: 0,030
 - Número de comunidades: 145
- ✓ Parámetros:
 - Aleatorio: Sí
 - Utilizar pesos: No
 - Resolución: 0,5
- Resultados: Figura 5.55 (derecha y arriba)
 - Modularidad: 0,828
 - Modularidad con resolución: 0,359
 - Número de comunidades: 20

- ✓ Parámetros:
 - Aleatorio: Sí
 - Utilizar pesos: No
 - Resolución: 1,0
- Resultados: Figura 5.55 (izquierda y abajo)
 - Modularidad: 0,834
 - Modularidad con resolución: 0,834
 - Número de comunidades: 15
- ✓ Parámetros:
 - Aleatorio: Sí
 - Utilizar pesos: No
 - Resolución: 2,0
- Resultados: Figura 5.55 (derecha y abajo)
 - Modularidad: 0,827
 - Modularidad con resolución: 1,798
 - Número de comunidades: 12

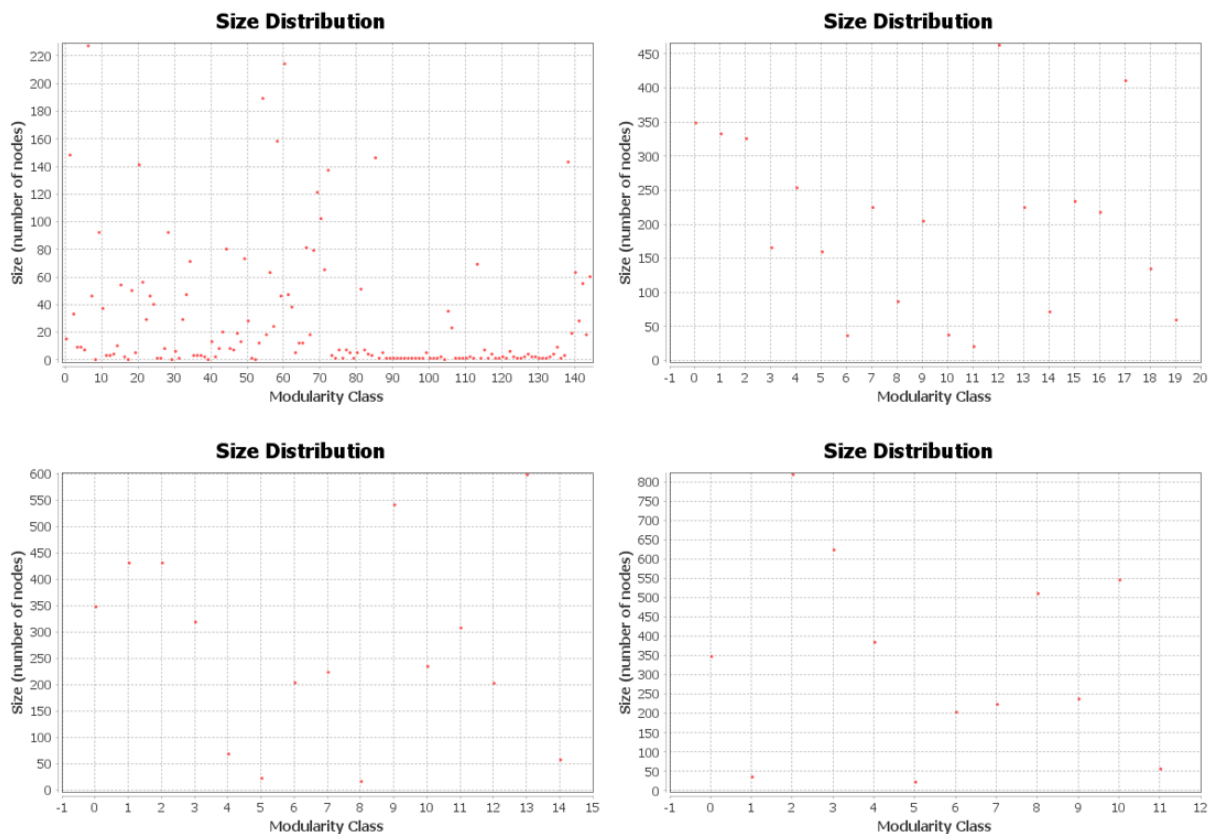


Figura 5.55: Ejemplo 2: Modularidad con aleatoriedad sin uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).

- ✓ Parámetros:

- Aleatorio: No
- Utilizar pesos: Sí
- Resolución: 0,1
- Resultados: Figura 5.56 (izquierda y arriba)
 - Modularidad: 0,456
 - Modularidad con resolución: 0,030
 - Número de comunidades: 147
- ✓ Parámetros:
 - Aleatorio: No
 - Utilizar pesos: Sí
 - Resolución: 0,5
- Resultados: Figura 5.56 (derecha y arriba)
 - Modularidad: 0,830
 - Modularidad con resolución: 0,359
 - Número de comunidades: 23
- ✓ Parámetros:
 - Aleatorio: No
 - Utilizar pesos: Sí
 - Resolución: 1,0
- Resultados: Figura 5.56 (izquierda y abajo)
 - Modularidad: 0,835
 - Modularidad con resolución: 0,835
 - Número de comunidades: 17
- ✓ Parámetros:
 - Aleatorio: No
 - Utilizar pesos: Sí
 - Resolución: 2,0
- Resultados: Figura 5.56 (derecha y abajo)
 - Modularidad: 0,827
 - Modularidad con resolución: 1,798
 - Número de comunidades: 12
- ✓ Parámetros:
 - Aleatorio: No
 - Utilizar pesos: No
 - Resolución: 0,1
- Resultados: Figura 5.57 (izquierda y arriba)
 - Modularidad: 0,456
 - Modularidad con resolución: 0,030
 - Número de comunidades: 150
- ✓ Parámetros:

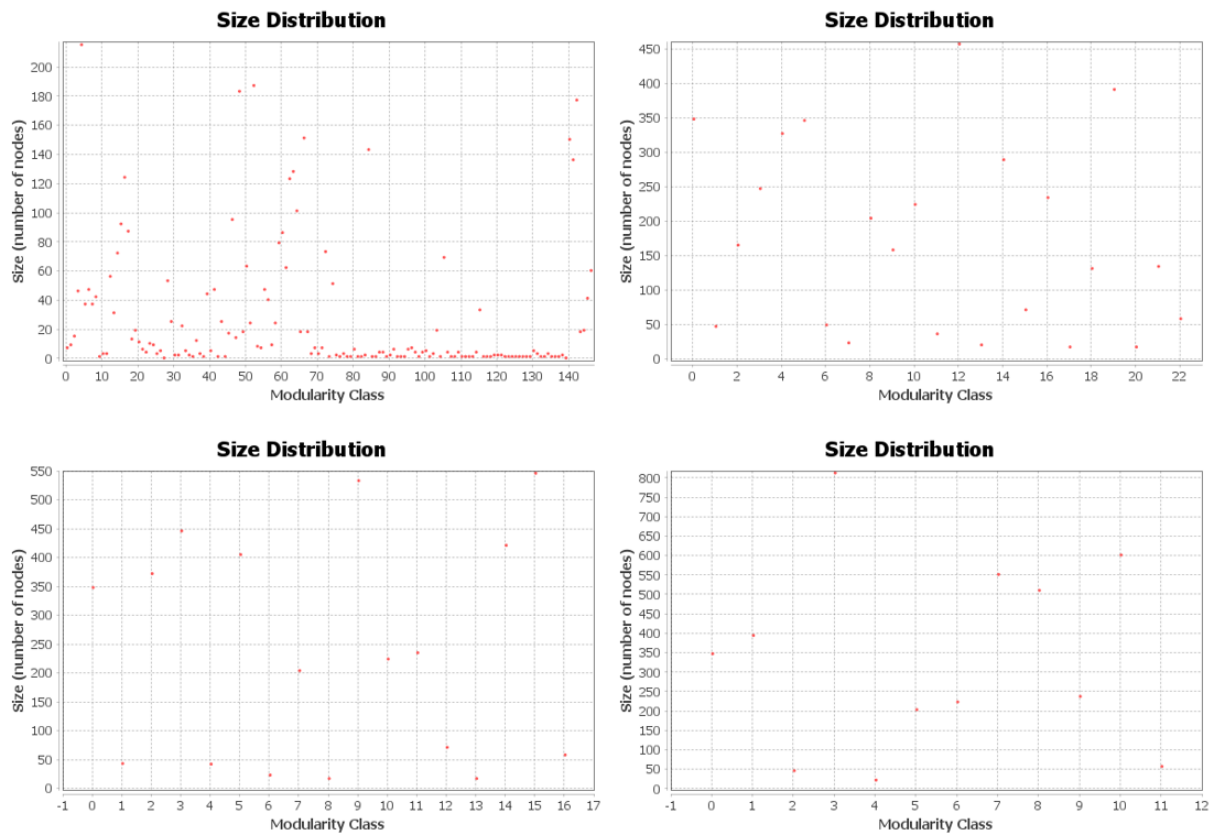


Figura 5.56: Ejemplo 2: Modularidad sin aleatoriedad con uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).

- Aleatorio: No
- Utilizar pesos: No
- Resolución: 0,5
- Resultados: Figura 5.57 (derecha y arriba)
 - Modularidad: 0,831
 - Modularidad con resolución: 0,359
 - Número de comunidades: 23
- ✓ Parámetros:
 - Aleatorio: No
 - Utilizar pesos: No
 - Resolución: 1,0
- Resultados: Figura 5.57 (izquierda y abajo)
 - Modularidad: 0,835
 - Modularidad con resolución: 0,835
 - Número de comunidades: 16
- ✓ Parámetros:
 - Aleatorio: No
 - Utilizar pesos: No

- Resolución: 2,0
- Resultados: Figura 5.57 (derecha y abajo)
 - Modularidad: 0,828
 - Modularidad con resolución: 1,798
 - Número de comunidades: 12

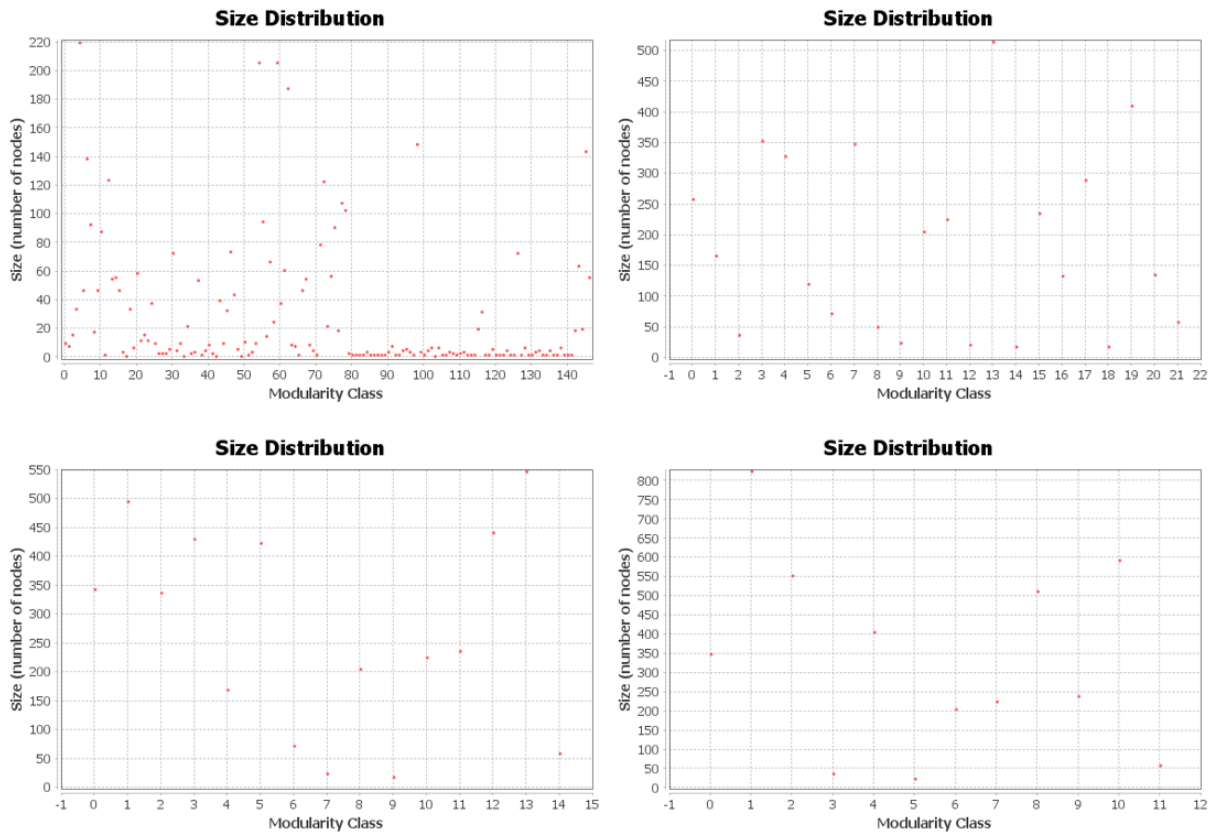


Figura 5.57: Ejemplo 2: Modularidad sin aleatoriedad ni uso de pesos (resoluciones 0,1; 0,5; 1 y 2, resp.).

- Inferencia estadística:
 - Resultados (primera ejecución): Figura 5.58
 - Longitud: 285245,807
 - Número de comunidades: 138
 - Resultados (segunda ejecución): Figura 5.59
 - Longitud: 285489,514
 - Número de comunidades: 148

3. Visión general de los nodos

- Coeficiente medio de *clustering*:
 - Resultados: Figura 5.60
 - Coeficiente medio de *clustering*: 0,617

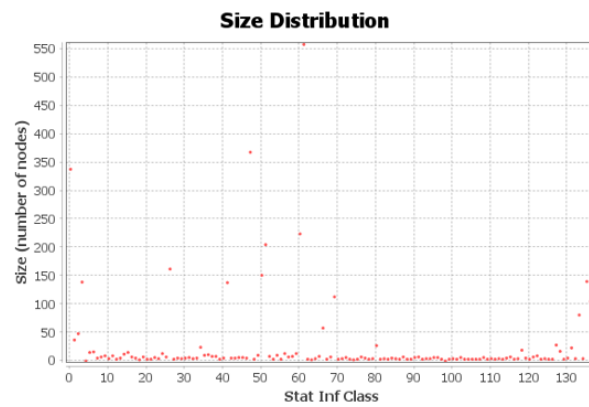


Figura 5.58: Ejemplo 2: Inferencia estadística (primera ejecución).

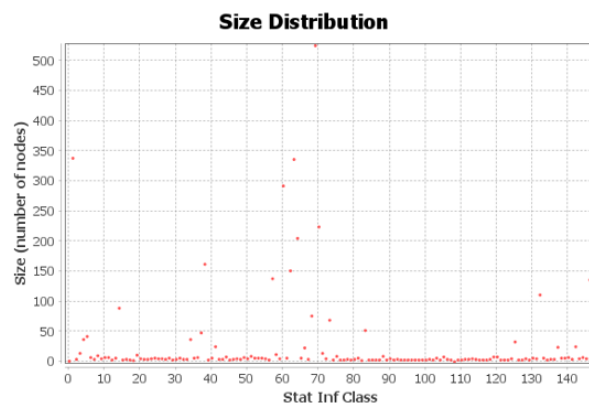


Figura 5.59: Ejemplo 2: Inferencia estadística (segunda ejecución).



Figura 5.60: Ejemplo 2: Coeficiente medio de *clustering*.

- Triángulos totales: 1612010
- Centralidad de vector propio:
 - ✓ Parámetros:
 - Número de iteraciones: 10

- *Sum change*: 6,448195362362457 \approx 6,448
- Resultado: Figura 5.61

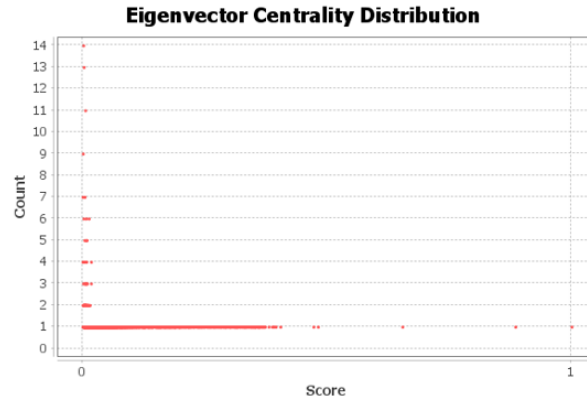


Figura 5.61: Ejemplo 2: Centralidad de vector propio con resolución 10.

- ✓ Parámetros:
 - Número de iteraciones: 100
 - *Sum change*: 0,4279073669804368 \approx 0,428
- Resultado: Figura 5.62

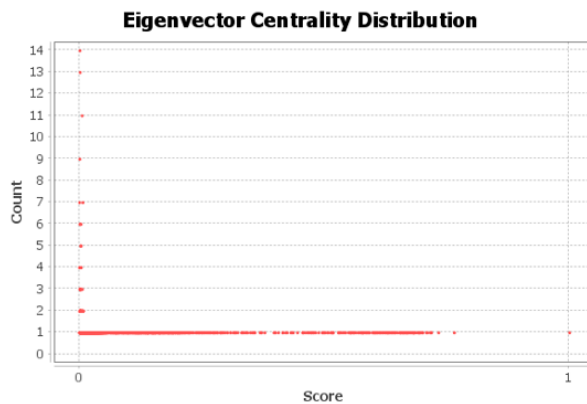


Figura 5.62: Ejemplo 2: Centralidad de vector propio con resolución 100.

- ✓ Parámetros:
 - Número de iteraciones: 1000
 - *Sum change*: 0,018387721960502874 \approx 0,018
- Resultado: Figura 5.63

4. Visión general de las aristas

- Longitud media de camino: 3,6925068496963913 \approx 3,693

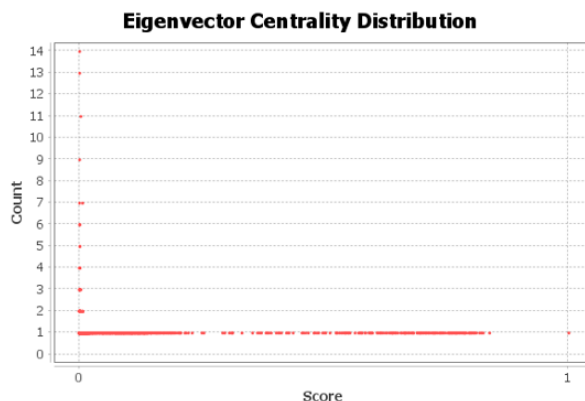


Figura 5.63: Ejemplo 2: Centralidad de vector propio con resolución 1000.

5.2.6. Discusión

El análisis del grafo no dirigido de Facebook revela una serie de características interesantes sobre la estructura de la red. A continuación, se interpretan los resultados dependiendo de la estadística.

Grado medio

El grado medio 43,691 es considerablemente alto, lo que indica una gran conectividad entre los nodos. Esto sugiere que, en promedio, cada usuario de la red está directamente conectado con muchos otros usuarios, lo que es característico de redes sociales grandes y densamente conectadas.

Diámetro y longitud media del camino

El diámetro del grafo $D = 8$ y la longitud media del camino $L = 3,693$ sugieren que la red es eficiente en términos de transmisión de información. Un diámetro relativamente bajo en comparación con el tamaño de la red es indicativo de un *small world* (Nota 22), donde cualquier nodo puede ser alcanzado desde otro a través de un pequeño número de pasos.

Densidad

La densidad de la red $\rho = 0,011$ es baja. Esto es típico en redes grandes donde, a pesar de la alta conectividad, el número de conexiones posibles supera ampliamente el número de conexiones existentes. Esta baja densidad, sin embargo, no impide la cohesión general de la red debido al alto grado medio y a la eficiente transmisión de información.

Modularidad y número de comunidades

Las distintas resoluciones utilizadas para calcular la modularidad arrojan valores entre $Q = 0,456$ y $Q = 0,835$, identificando entre 12 y 150 comunidades. Sin embargo, a partir de valores mayores de 0,5 de resolución, la modularidad oscila solo entre $Q = 0,827$ y $Q = 0,835$, y las comunidades entre 12 y 23. Una alta modularidad, sugiere que la red tiene una estructura de comunidades bien definida, lo que podría reflejar grupos de amigos o círculos sociales en Facebook que están densamente conectados entre sí, pero relativamente separados de otros grupos.

Coefficiente de *clustering*

El coeficiente medio de *clustering* $C = 0,617$ es alto, lo que indica que los nodos tienden a formar grupos cerrados de amigos. Es un comportamiento típico en redes sociales, donde si dos personas tienen un amigo en común, es probable que ellas también sean amigas.

5.2.7. Conclusiones

El análisis muestra que la red de Facebook es un ejemplo típico de red social con características de *small world* (Nota 22) y alto *clustering*. Estas propiedades sugieren que la red es muy eficiente para la propagación de información y permite la formación de comunidades locales fuertes. Este tipo de estructura es ideal para el intercambio rápido de información, manteniendo al mismo tiempo subgrupos cohesionados, que podrían representar diferentes círculos sociales o intereses comunes dentro de la red social.

Parte I

Apéndices

Apéndice A

Acrónimos

ADN	Ácido Desoxirribonucleico
ARN	Ácido Ribonucleico
CSV	Valores Separados por Comas o <i>Comma Separated Values</i>
3D	3 Dimensiones o 3 <i>Dimensions</i>
DE	Alemania o <i>Denmark</i>
ECTS	Sistema Europeo de Transferencia y Acumulación de Créditos o <i>European Credit Transfer System</i>
ENGB	Reino Unido o <i>English - Great Britain</i>
ES	España o <i>Spain</i>
FR	Francia o <i>France</i>
HITS	Búsqueda de temas inducida por hipervínculos o <i>Hyperlink-Induced Topic Search</i>
IA	Inteligencia Artificial
PTBR	Brasil o <i>Portuguese - Brazil</i>
RU	Rusia o <i>Russia</i>
SCC	Componente Fuertemente Conexo o <i>Strongly Connected Component</i>
SIDA	Síndrome de Inmunodeficiencia Adquirida
TFG	Trabajo Fin de Grado
URL	Localizador Uniforme de Recursos o <i>Uniform Resource Locator</i>
WCC	Componente Conexo Débil o <i>Weakly Connected Component</i>
WWW	Red de Alcance Mundial o <i>World Wide Web</i>

Apéndice B

Métodos de crecimiento de redes

Existen varios métodos para calcular la distribución de conectividades $P(k, t)$. Dado que en este trabajo ya se estudia el método de la ecuación maestra aplicado a redes complejas en la Sección 4.4.2, en este anexo se explica de una forma más general. Sin embargo, existen, por lo menos, otros dos métodos distintos: el método continuo (Sección B.2), inventado por Barabási, y el método cinético (Sección B.3), introducido por Krapivsky, Redner, y Leyvraz [Aldana2], que sí se explican aquí aplicados a redes complejas. Los tres métodos dan resultados equivalentes.

El objetivo es determinar cómo la probabilidad $P(k, t)$ de que un nodo tenga k enlaces en el tiempo t cambia debido a la adición de nuevos nodos y enlaces en la red. Se considera que la red crece añadiendo un nodo a la vez, cada uno con m enlaces.

Nota 39. En la Sección B.1 se considera $m = 1$.

B.1. Método de la ecuación maestra

Se supone un sistema que puede estar en cualquiera de los estados E_1, E_2, \dots, E_N . Sea $W_{m \rightarrow n}(t)$ la probabilidad condicional de que el sistema “salte” al estado E_n sabiendo que parte del estado E_m al tiempo t . Las probabilidades $W_{m \rightarrow n}(t)$ se denominan probabilidades de transición.

La probabilidad de que el sistema se encuentre en el estado E_n al tiempo t , $P(n, t)$, puede cambiar en el tiempo por los dos factores que se listan a continuación.

- El sistema está en el estado E_n al tiempo t y salta a otro estado E_m , lo cual claramente disminuye la probabilidad $P(n, t + 1)$ de encontrar al sistema en el estado E_n al tiempo $t + 1$.
- El sistema está en algún estado E_m al tiempo t y salta al estado E_n , lo cual aumenta la probabilidad $P(n, t + 1)$ de encontrar al sistema en el estado E_n al tiempo $t + 1$.

Teniendo en cuenta estos dos factores, la probabilidad de que el sistema se encuentre en el estado E_n al tiempo $t + 1$ está dada por la Ecuación B.1, denominada ecuación maestra del sistema.

$$P(n, t + 1) = \sum_{m=1}^N P(m, t) \cdot W_{m \rightarrow n}(t) - \sum_{m=1}^N P(n, t) \cdot W_{n \rightarrow m}(t) \quad (\text{B.1})$$

El primer término del lado derecho indica que, al tiempo t , el sistema puede estar en el estado E_m y saltar al estado E_n , aumentando así la probabilidad de estar en E_n al tiempo $t + 1$. El segundo término tiene en cuenta el hecho de que, al tiempo t , el sistema puede estar en el estado E_n y salirse de allí, saltando al estado E_m y disminuyendo así la probabilidad de estar en E_n al tiempo $t + 1$. Las sumas sobre m recogen todos los posibles estados hacia o desde los que puede saltar el sistema.

El objetivo es resolver la Ecuación B.1 para encontrar $P(n, t)$. Si las probabilidades de transición $W_{m \rightarrow n}$ son independientes del tiempo, la ecuación se puede resolver fácilmente; de lo contrario, es decir, si las probabilidades de transición dependen del tiempo, la resolución se complica enormemente. Afortunadamente, las ecuaciones maestras que aparecen en la teoría de crecimiento de redes tratada en este trabajo son relativamente fáciles de resolver.

B.2. Método continuo

El método continuo es introducido por Albert-László Barabási y Réka Albert en 1999 en su famoso modelo de redes de escala libre. El método continuo considera la evolución de la red como un proceso en el que la conectividad de los nodos, $k_i(t)$, es tratada como una variable continua. En lugar de analizar la probabilidad discreta de que un nodo tenga un número específico de conexiones en cada paso de tiempo, se estudia cómo crece el grado de un nodo en función del tiempo a través de una ecuación diferencial. Este método es especialmente útil para analizar redes en las que se observa un enlace preferencial, es decir, donde los nodos con más conexiones tienden a obtener nuevas conexiones a un ritmo mayor.

A partir de estos supuestos, se busca formular una ecuación diferencial que describa cómo cambia el grado $k_i(t)$ de un nodo v_i en función del tiempo t . Considerando que en cada paso temporal se añade un nodo con m enlaces, y que la probabilidad de que un nuevo nodo se conecte a un nodo v_i es $\pi(k_i) = \frac{k_i}{2 \cdot m \cdot t}$, se puede escribir la tasa de cambio del grado de v_i como se observa en la Ecuación B.2.

$$\frac{dk_i(t)}{dt} = \pi \cdot \frac{k_i(t)}{2 \cdot \pi \cdot t} = \frac{k_i(t)}{2 \cdot t} \quad (\text{B.2})$$

La Ecuación B.2 indica que el grado de un nodo aumenta lentamente con el tiempo, y depende directamente de su grado actual, capturando el efecto de “los ricos se hacen más ricos” característico de redes libres de escala.

Se integra la Ecuación B.3 diferencial.

$$\frac{dk_i}{k_i} = \frac{dt}{2 \cdot t} \quad (\text{B.3})$$

Y se obtiene la Ecuación B.4.

$$\ln k_i = \frac{1}{2} \cdot \ln t + C \quad (\text{B.4})$$

Donde C es una constante determinada por la condición inicial. Exponenciando para resolver k_i se obtiene la Ecuación B.5.

$$k_i(t) = C' \cdot t^{\frac{1}{2}} \quad (\text{B.5})$$

Dado que cuando el nodo v_i se añade en el tiempo t_i , su grado es m . Se obtiene así C' y se sustituye su valor para obtener la Ecuación B.6.

$$k_i(t) = m \cdot \left(\frac{t}{t_i} \right)^{\frac{1}{2}} \quad (\text{B.6})$$

Esto implica que el grado de un nodo crece como una raíz cuadrada del tiempo desde que es añadido a la red.

Para determinar la distribución de conectividades $P(k)$, se considera el número de nodos con grado al menos k , que, como se observa en la Ecuación B.7, es inversamente proporcional a t_i .

$$N(k, t) \approx \left(\frac{k}{m} \right)^{-2} \quad (\text{B.7})$$

Por lo tanto, la distribución de grados sigue una ley de potencias $P(k) \sim k^{-3}$. Este resultado es fundamental para explicar las propiedades de redes reales como Internet, donde la mayoría de los nodos tiene pocos enlaces, pero unos pocos nodos tienen muchos, formando *hubs*.

B.3. Método cinético

El método cinético, desarrollado por Krapivsky, Redner y Leyvraz en 2000, se basa en la formulación de ecuaciones de tasa que describen cómo cambia la fracción de nodos con un determinado número de conexiones a lo largo del tiempo. Este método es particularmente útil cuando se busca entender no solo el estado final de la red, sino también su evolución dinámica.

La evolución de $P(k, t)$ se describe por la Ecuación B.8 de balance que toma en cuenta las posibles maneras en que un nodo puede alcanzar un grado k en el tiempo t .

$$\frac{\partial P(k, t)}{\partial t} = \pi(k-1, t) \cdot P(k-1, t) - \pi(k, t) \cdot P(k, t) \quad (\text{B.8})$$

En el régimen estacionario, donde $\frac{\partial P(k, t)}{\partial t} = 0$ se obtiene la Ecuación B.9.

$$\frac{\partial P(k, t)}{\partial t} = \pi(k-1) \cdot P(k-1) = \pi(k) \cdot P(k) \quad (\text{B.9})$$

Si se asume que $\pi(k)$ es lineal en k , como en el caso de enlace preferencial $\pi(k) \sim k$, entonces la solución para $P(k)$ sigue la ley de potencias $P(k) \sim k^{-\gamma}$, donde γ es el exponente de la distribución, que puede variar según los parámetros específicos del sistema.

Apéndice C

Transiciones de fase en redes complejas con topología de Poisson

En el contexto de transiciones de fase en redes complejas con topología de Poisson, como las redes aleatorias de Erdős-Rényi, se observan principalmente dos órdenes distintos de transiciones de fase: de primer orden y de segundo orden. Sin embargo, la transición de fase topológica en estas redes, como la transición de percolación, típicamente se asocia con una transición de segundo orden.

C.1. Transición de segundo orden

En una red de Erdős-Rényi, que sigue una distribución de grado de Poisson, la transición de percolación es un ejemplo clásico de una transición de fase topológica de segundo orden. Aquí, el “orden” se refiere al hecho de que la transición ocurre de manera continua a medida que un parámetro crítico es cruzado, sin saltos bruscos.

Al aumentar el número de enlaces en la red, se alcanza un punto crítico donde se forma un componente gigante que conecta una fracción significativa de la red. Este cambio ocurre de manera continua y se caracteriza por la aparición de correlaciones a gran escala en la red.

C.2. Transición de primer orden

En redes con otras topologías o bajo ciertas condiciones, se pueden observar transiciones de fase de primer orden, donde la propiedad global de la red cambia de manera abrupta. Sin embargo, esto no es típico en redes con topología de Poisson sin modificaciones adicionales o bajo dinámicas específicas.

C.3. Conclusión

En resumen, en redes con topología de Poisson, como las redes de Erdős-Rényi, la transición de fase topológica es generalmente de segundo orden. En este caso específico, no se observan transiciones de primer orden de manera natural, a menos que se introduzcan variaciones en la dinámica o en la estructura de la red.

Apéndice D

Código

En este anexo se adjunta el código Python utilizado para resolver los cálculos del Ejemplo 37 (Ecuación 3.19, Ecuación 3.20 y Ecuación 3.21), utilizando un método numérico como descomposición de valores propios.

```
1 # Definir las matrices de adyacencia dadas por el usuario
2 # Grafo 1 proporcionado
3 A1_user = np.array([
4     [0, 1, 1, 1, 0, 0],
5     [1, 0, 1, 0, 0, 0],
6     [1, 1, 0, 0, 0, 0],
7     [1, 0, 0, 0, 1, 1],
8     [0, 0, 0, 1, 0, 1],
9     [0, 0, 0, 1, 1, 0]
10 ])
11 # Grafo 2 proporcionado
12 A2_user = np.array([
13     [0, 1, 1, 0, 0, 0],
14     [1, 0, 1, 0, 1, 0],
15     [1, 1, 0, 0, 0, 0],
16     [0, 0, 0, 0, 1, 1],
17     [0, 1, 0, 1, 0, 1],
18     [0, 0, 0, 1, 1, 0]
19 ])
20 # Grafo 3 proporcionado
21 A3_user = np.array([
22     [0, 1, 1, 0, 0, 0],
23     [1, 0, 1, 0, 0, 0],
24     [1, 1, 0, 0, 0, 1],
25     [0, 0, 0, 0, 1, 1],
26     [0, 0, 0, 1, 0, 1],
27     [0, 0, 1, 1, 1, 0]
28 ])
29
30 # Calcular  $A' - I6$  para cada grafo dado por el usuario
31 A1_user_prime_minus_I6 = A1_user.T - I6
32 A2_user_prime_minus_I6 = A2_user.T - I6
33 A3_user_prime_minus_I6 = A3_user.T - I6
34
35 # Resolver  $(A' - I6) * p = 0$  para cada grafo
36 _, p1_user = np.linalg.eig(A1_user_prime_minus_I6)
```

```
37 _, p2_user = np.linalg.eig(A2_user_prime_minus_I6)
38 _, p3_user = np.linalg.eig(A3_user_prime_minus_I6)
39
40 # Normalizar los vectores propios
41 p1_user_normalized = p1_user[:, 0] / np.linalg.norm(p1_user[:, 0])
42 p2_user_normalized = p2_user[:, 0] / np.linalg.norm(p2_user[:, 0])
43 p3_user_normalized = p3_user[:, 0] / np.linalg.norm(p3_user[:, 0])
44
45 p1_user_normalized, p2_user_normalized, p3_user_normalized
```

Listing D.1: Código en Python

Resultado

```
(array([-0.62796303,  0.22985042,  0.22985042,  0.62796303, -0.22985042,
        -0.22985042]),
 array([-0.35355339, -0.5          , -0.35355339, -0.35355339, -0.5          ,
        -0.35355339]),
 array([-0.35355339, -0.35355339, -0.5          , -0.35355339, -0.35355339,
        -0.5          ]))
```

Listing D.2: Output del código

Apéndice E

Cálculos

En este anexo se adjuntan los cálculos del Ejemplo 42.

E.1. Esquema

A continuación, se exponen los pasos realizados de una forma esquematizada. Además, se añade una serie de comentarios explicativos acerca de los resultados obtenidos.

1. $Q = -0,173$

- Comunidad 1: $\{v_1\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_3\}$
- Comunidad 4: $\{v_4\}$
- Comunidad 5: $\{v_5\}$
- Comunidad 6: $\{v_6\}$

2. $Q = -0,092$

- Comunidad 1: $\{v_1, v_2\}$
- Comunidad 2: $\{v_3\}$
- Comunidad 3: $\{v_4\}$
- Comunidad 4: $\{v_5\}$
- Comunidad 5: $\{v_6\}$

3. $Q = -0,092$

- Comunidad 1: $\{v_1, v_3\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_4\}$
- Comunidad 4: $\{v_5\}$
- Comunidad 5: $\{v_6\}$

Hasta este paso se observa que la modularidad crece si se cambia v_1 a la comunidad de un nodo adyacente a él, y que el crecimiento es igual para todos aquellos que tienen el mismo grado ($\delta(v_2) = \delta(v_3) = 2$).

4. $Q = -0,122$

- Comunidad 1: $\{v_1, v_4\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_3\}$
- Comunidad 4: $\{v_5\}$
- Comunidad 5: $\{v_6\}$

En este paso se observa que la modularidad crece si se cambia v_1 a la comunidad del nodo con el mayor grado y adyacente a él ($\delta(v_4) = 3 > 2 = \delta(v_2) = \delta(v_3)$).

5. $Q = -0,092$

- Comunidad 1: $\{v_1\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_3\}$
- Comunidad 4: $\{v_4, v_5\}$
- Comunidad 5: $\{v_6\}$

6. $Q = -0,092$

- Comunidad 1: $\{v_1\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_3\}$
- Comunidad 4: $\{v_4, v_6\}$
- Comunidad 5: $\{v_5\}$

Hasta este paso se observa que la modularidad crece si se cambia v_4 a la comunidad de un nodo adyacente a él, y que el crecimiento es igual para todos aquellos que tienen el mismo grado ($\delta(v_5) = \delta(v_6) = 2$) e igual que para el nodo v_1 con sus nodos adyacentes de grado 2.

7. $Q = -0,071$

- Comunidad 1: $\{v_1\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_3\}$
- Comunidad 4: $\{v_4\}$
- Comunidad 5: $\{v_5, v_6\}$

Ahora se observa que la modularidad crece si se cambia v_5 a la comunidad de un nodo adyacente a él, y que el crecimiento es mayor para v_6 ($\delta(v_6) = 2$) que para v_4 ($\delta(v_4) = 3$).

Nota 40. *Esto es, la modularidad aumenta cuando el grado de los nodos de la misma comunidad coincide, más aún que juntándolos con nodos de grado mayor.*

8. $Q = 0,02$

- Comunidad 1: $\{v_1, v_2, v_3\}$
- Comunidad 2: $\{v_4\}$
- Comunidad 3: $\{v_5\}$
- Comunidad 4: $\{v_6\}$

9. $Q = -0,087$

- Comunidad 1: $\{v_1, v_2, v_4\}$
- Comunidad 2: $\{v_3\}$
- Comunidad 3: $\{v_5\}$
- Comunidad 4: $\{v_6\}$

10. $Q = 0,092$

- Comunidad 1: $\{v_1\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_3\}$
- Comunidad 4: $\{v_4, v_5, v_6\}$

Estos tres pasos confirman que la modularidad crece cuando todos los nodos de la misma comunidad son adyacentes.

11. $Q = 0,01$

- Comunidad 1: $\{v_1, v_2, v_3, v_4\}$
- Comunidad 2: $\{v_5\}$
- Comunidad 3: $\{v_6\}$

12. $Q = 0,02$

- Comunidad 1: $\{v_1, v_4, v_5, v_6\}$
- Comunidad 2: $\{v_2\}$
- Comunidad 3: $\{v_3\}$

13. $Q = 0,087$

- Comunidad 1: $\{v_1, v_2, v_3\}$
- Comunidad 2: $\{v_4, v_5\}$
- Comunidad 3: $\{v_6\}$

14. $Q = 0,357$

- Comunidad 1: $\{v_1, v_2, v_3\}$
- Comunidad 2: $\{v_4, v_5, v_6\}$

La modularidad máxima se alcanza al dividir la red en vecinarios disjuntos.

15. $Q = 0$

- Comunidad 1: $\{v_1, v_2, v_3, v_4, v_5, v_6\}$

E.2. Desarrollo

Cambiando la partición de forma que v_1 y v_2 forman una única comunidad, la modularidad Q de G es $-0,092$, y su cálculo se observa en la Ecuación E.1.

$$\begin{aligned}
 Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
 &= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{12} - \frac{k_1 \cdot 2}{14} \right) \cdot \delta(v_1, v_2) \right. \\
 &\quad + \left(a_{21} - \frac{k_2 \cdot k_1}{14} \right) \cdot \delta(v_2, v_1) + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
 &\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
 &\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
 &\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
 &\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \tag{E.1} \\
 &= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \right. \\
 &\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
 &= -0,092
 \end{aligned}$$

Cambiando la partición de forma que v_1 y v_3 forman una única comunidad, la modularidad Q de G es $-0,092$, y su cálculo se observa en la Ecuación E.2.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{13} - \frac{k_1 \cdot k_3}{14} \right) \cdot \delta(v_1, v_3) \right. \\
&\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
&\quad + \left(a_{31} - \frac{k_3 \cdot k_1}{14} \right) \cdot \delta(v_3, v_1) + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
&\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
&\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \tag{E.2} \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \right. \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= -0,092
\end{aligned}$$

Cambiando la partición de forma que v_1 y v_4 forman una única comunidad, la modularidad Q de G es $-0,122$, y su cálculo se observa en la Ecuación E.3.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{14} - \frac{k_1 \cdot k_4}{14} \right) \cdot \delta(v_1, v_4) \right. \\
&\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
&\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{41} - \frac{k_4 \cdot k_1}{14} \right) \cdot \delta(v_4, v_1) + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
&\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
&\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \tag{E.3} \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= -0,122
\end{aligned}$$

Cambiando la partición de forma que v_4 y v_5 forman una única comunidad, la modularidad Q de G es $-0,092$, y su cálculo se observa en la Ecuación E.4.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) \right. \\
&\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
&\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) + \left(a_{45} - \frac{k_4 \cdot k_5}{14} \right) \cdot \delta(v_4, v_5) \\
&\quad + \left(a_{54} - \frac{k_5 \cdot k_4}{14} \right) \cdot \delta(v_5, v_4) + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
&\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \tag{E.4} \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= -0,092
\end{aligned}$$

Cambiando la partición de forma que v_4 y v_6 forman una única comunidad, la modularidad Q de G es $-0,092$, y su cálculo se observa en la Ecuación E.5.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) \right. \\
&\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
&\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) + \left(a_{46} - \frac{k_4 \cdot k_6}{14} \right) \cdot \delta(v_4, v_6) \\
&\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
&\quad \left. + \left(a_{64} - \frac{k_6 \cdot k_4}{14} \right) \cdot \delta(v_6, v_4) + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \tag{E.5} \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= -0,092
\end{aligned}$$

Cambiando la partición de forma que v_5 y v_6 forman una única comunidad, la modularidad Q de G es $-0,071$, y su cálculo se observa en la Ecuación E.6.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) \right. \\
&\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
&\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
&\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) + \left(a_{56} - \frac{k_5 \cdot k_6}{14} \right) \cdot \delta(v_5, v_6) \\
&\quad \left. + \left(a_{65} - \frac{k_6 \cdot k_5}{14} \right) \cdot \delta(v_6, v_5) + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \tag{E.6} \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= -0,071
\end{aligned}$$

Cambiando la partición de forma que v_1, v_2 , y v_4 forman una única comunidad, la modularidad Q de G es $-0,087$, y su cálculo se observa en la Ecuación E.7.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{12} - \frac{k_1 \cdot 2}{14} \right) \cdot \delta(v_1, v_2) + \left(a_{14} - \frac{k_1 \cdot k_4}{14} \right) \cdot \delta(v_1, v_4) \right. \\
&\quad + \left(a_{21} - \frac{k_2 \cdot k_1}{14} \right) \cdot \delta(v_2, v_1) + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) + \left(a_{24} - \frac{k_2 \cdot k_4}{14} \right) \cdot \delta(v_2, v_4) \\
&\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{41} - \frac{k_4 \cdot k_1}{14} \right) \cdot \delta(v_4, v_1) + \left(a_{42} - \frac{k_4 \cdot k_2}{14} \right) \cdot \delta(v_4, v_2) + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
&\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
&\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= -0,087
\end{aligned} \tag{E.7}$$

Cambiando la partición de forma que $v_4, v_5,$ y v_6 forman una única comunidad, la modularidad Q de G es 0,092, y su cálculo se observa en la Ecuación E.8.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) \right. \\
&\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
&\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) + \left(a_{45} - \frac{k_4 \cdot k_5}{14} \right) \cdot \delta(v_4, v_5) + \left(a_{46} - \frac{k_4 \cdot k_6}{14} \right) \cdot \delta(v_4, v_6) \\
&\quad + \left(a_{54} - \frac{k_5 \cdot k_4}{14} \right) \cdot \delta(v_5, v_4) + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) + \left(a_{56} - \frac{k_5 \cdot k_6}{14} \right) \cdot \delta(v_5, v_6) \\
&\quad \left. + \left(a_{64} - \frac{k_6 \cdot k_4}{14} \right) \cdot \delta(v_6, v_4) + \left(a_{65} - \frac{k_6 \cdot k_5}{14} \right) \cdot \delta(v_6, v_5) + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= 0,092
\end{aligned} \tag{E.8}$$

Cambiando la partici3n de forma que v_1, v_2 y v_3 forman una 3nica comunidad, la modularidad Q de G es 0,02, y su c3lculo se observa en la Ecuaci3n E.9.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{12} - \frac{k_1 \cdot 2}{14} \right) \cdot \delta(v_1, v_2) + \left(a_{13} - \frac{k_1 \cdot k_3}{14} \right) \cdot \delta(v_1, v_3) \right. \\
&\quad + \left(a_{21} - \frac{k_2 \cdot k_1}{14} \right) \cdot \delta(v_2, v_1) + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) + \left(a_{23} - \frac{k_2 \cdot k_3}{14} \right) \cdot \delta(v_2, v_3) \\
&\quad + \left(a_{31} - \frac{k_3 \cdot k_1}{14} \right) \cdot \delta(v_3, v_1) + \left(a_{32} - \frac{k_3 \cdot k_2}{14} \right) \cdot \delta(v_3, v_2) + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
&\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
&\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \right. \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= 0,02
\end{aligned} \tag{E.9}$$

Cambiando la partición de forma que v_1, v_2, v_3 y v_4 forman una única comunidad, la modularidad Q de G es 0,01, y su cálculo se observa en la Ecuación E.10.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{12} - \frac{k_1 \cdot 2}{14} \right) \cdot \delta(v_1, v_2) \right. \\
&\quad + \left(a_{13} - \frac{k_1 \cdot k_3}{14} \right) \cdot \delta(v_1, v_3) + \left(a_{14} - \frac{k_1 \cdot k_4}{14} \right) \cdot \delta(v_1, v_4) \\
&\quad + \left(a_{21} - \frac{k_2 \cdot k_1}{14} \right) \cdot \delta(v_2, v_1) + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
&\quad + \left(a_{23} - \frac{k_2 \cdot k_3}{14} \right) \cdot \delta(v_2, v_3) + \left(a_{24} - \frac{k_2 \cdot k_4}{14} \right) \cdot \delta(v_2, v_4) \\
&\quad + \left(a_{31} - \frac{k_3 \cdot k_1}{14} \right) \cdot \delta(v_3, v_1) + \left(a_{32} - \frac{k_3 \cdot k_2}{14} \right) \cdot \delta(v_3, v_2) \\
&\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) + \left(a_{34} - \frac{k_3 \cdot k_4}{14} \right) \cdot \delta(v_3, v_4) \\
&\quad + \left(a_{41} - \frac{k_4 \cdot k_1}{14} \right) \cdot \delta(v_4, v_1) + \left(a_{42} - \frac{k_4 \cdot k_2}{14} \right) \cdot \delta(v_4, v_2) \\
&\quad + \left(a_{43} - \frac{k_4 \cdot k_3}{14} \right) \cdot \delta(v_4, v_3) + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
&\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
&\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 \right. \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 2}{14} \right) \cdot 0 + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= 0,01
\end{aligned} \tag{E.10}$$

Cambiando la partici3n de forma que v_1, v_4, v_5 y v_6 forman una 3nica comunidad, la modularidad Q de G es 0,02, y su c3lculo se observa en la Ecuaci3n E.11.

$$\begin{aligned}
 Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
 &= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{14} - \frac{k_1 \cdot k_4}{14} \right) \cdot \delta(v_1, v_4) \right. \\
 &\quad + \left(a_{15} - \frac{k_1 \cdot k_5}{14} \right) \cdot \delta(v_1, v_5) + \left(a_{16} - \frac{k_1 \cdot k_6}{14} \right) \cdot \delta(v_1, v_6) \\
 &\quad + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) \\
 &\quad + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
 &\quad + \left(a_{41} - \frac{k_4 \cdot k_1}{14} \right) \cdot \delta(v_4, v_1) + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) \\
 &\quad + \left(a_{45} - \frac{k_4 \cdot k_5}{14} \right) \cdot \delta(v_4, v_5) + \left(a_{46} - \frac{k_4 \cdot k_6}{14} \right) \cdot \delta(v_4, v_6) \\
 &\quad + \left(a_{51} - \frac{k_5 \cdot k_1}{14} \right) \cdot \delta(v_5, v_1) + \left(a_{54} - \frac{k_5 \cdot k_4}{14} \right) \cdot \delta(v_5, v_4) \\
 &\quad + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) + \left(a_{56} - \frac{k_5 \cdot k_6}{14} \right) \cdot \delta(v_5, v_6) \\
 &\quad + \left(a_{61} - \frac{k_6 \cdot k_1}{14} \right) \cdot \delta(v_6, v_1) + \left(a_{64} - \frac{k_6 \cdot k_4}{14} \right) \cdot \delta(v_6, v_4) \\
 &\quad \left. + \left(a_{65} - \frac{k_6 \cdot k_5}{14} \right) \cdot \delta(v_6, v_5) + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
 &= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 2}{14} \right) \cdot 1 \right. \\
 &\quad + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad \left. + \left(0 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
 &= 0,02
 \end{aligned} \tag{E.11}$$

Cambiando la partición de forma que v_1, v_2, v_3, v_4, v_5 y v_6 forman una única comunidad, la modularidad Q de G es 0., y su cálculo se observa en la Ecuación E.12.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \dots + \left(a_{16} - \frac{k_1 \cdot k_6}{14} \right) \cdot \delta(v_1, v_6) \right. \\
&\quad + \left(a_{21} - \frac{k_2 \cdot k_1}{14} \right) \cdot \delta(v_2, v_1) + \dots + \left(a_{26} - \frac{k_2 \cdot k_6}{14} \right) \cdot \delta(v_2, v_6) \\
&\quad + \left(a_{31} - \frac{k_3 \cdot k_1}{14} \right) \cdot \delta(v_3, v_1) + \dots + \left(a_{36} - \frac{k_3 \cdot k_6}{14} \right) \cdot \delta(v_3, v_6) \\
&\quad + \left(a_{41} - \frac{k_4 \cdot k_1}{14} \right) \cdot \delta(v_4, v_1) + \dots + \left(a_{46} - \frac{k_4 \cdot k_6}{14} \right) \cdot \delta(v_4, v_6) \\
&\quad + \left(a_{51} - \frac{k_5 \cdot k_1}{14} \right) \cdot \delta(v_5, v_1) + \dots + \left(a_{56} - \frac{k_5 \cdot k_6}{14} \right) \cdot \delta(v_5, v_6) \\
&\quad \left. + \left(a_{61} - \frac{k_6 \cdot k_1}{14} \right) \cdot \delta(v_6, v_1) + \dots + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \tag{E.12} \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \dots + \left(0 - \frac{3 \cdot 2}{14} \right) \cdot 1 \right. \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \dots + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \dots + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \dots + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \dots + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(0 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \dots + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= 0
\end{aligned}$$

Cambiando la partici3n de forma que v_1 , v_2 y v_3 forman una comunidad, v_4 y v_5 otra y v_6 otra distinta, la modularidad Q de G es 0,087, y su c3lculo se observa en la Ecuaci3n E.13.

$$\begin{aligned}
 Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
 &= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{12} - \frac{k_1 \cdot 2}{14} \right) \cdot \delta(v_1, v_2) + \left(a_{13} - \frac{k_1 \cdot k_3}{14} \right) \cdot \delta(v_1, v_3) \right. \\
 &\quad + \left(a_{21} - \frac{k_2 \cdot k_1}{14} \right) \cdot \delta(v_2, v_1) + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) + \left(a_{23} - \frac{k_2 \cdot k_3}{14} \right) \cdot \delta(v_2, v_3) \\
 &\quad + \left(a_{31} - \frac{k_3 \cdot k_1}{14} \right) \cdot \delta(v_3, v_1) + \left(a_{32} - \frac{k_3 \cdot k_2}{14} \right) \cdot \delta(v_3, v_2) + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
 &\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) + \left(a_{45} - \frac{k_4 \cdot k_5}{14} \right) \cdot \delta(v_4, v_5) \\
 &\quad + \left(a_{54} - \frac{k_5 \cdot k_4}{14} \right) \cdot \delta(v_5, v_4) + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) \\
 &\quad \left. + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
 &= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \right. \\
 &\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \\
 &\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
 &\quad \left. + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
 &= 0,087
 \end{aligned} \tag{E.13}$$

Cambiando la partición de forma que v_1, v_2 y v_3 forman una comunidad y v_4, v_5 y v_6 forman otra distinta, la modularidad Q de G es 0,357, y su cálculo se observa en la Ecuación E.14.

$$\begin{aligned}
Q &= \frac{1}{2 \cdot 7} \cdot \sum_{i=1}^6 \sum_{j=1}^6 \left(a_{ij} - \frac{k_i \cdot k_j}{2 \cdot 7} \right) \cdot \delta(v_i, v_j) \\
&= \frac{1}{14} \cdot \left[\left(a_{11} - \frac{k_1 \cdot k_1}{14} \right) \cdot \delta(v_1, v_1) + \left(a_{12} - \frac{k_1 \cdot 2}{14} \right) \cdot \delta(v_1, v_2) + \left(a_{13} - \frac{k_1 \cdot k_3}{14} \right) \cdot \delta(v_1, v_3) \right. \\
&\quad + \left(a_{21} - \frac{k_2 \cdot k_1}{14} \right) \cdot \delta(v_2, v_1) + \left(a_{22} - \frac{k_2 \cdot k_2}{14} \right) \cdot \delta(v_2, v_2) + \left(a_{23} - \frac{k_2 \cdot k_3}{14} \right) \cdot \delta(v_2, v_3) \\
&\quad + \left(a_{31} - \frac{k_3 \cdot k_1}{14} \right) \cdot \delta(v_3, v_1) + \left(a_{32} - \frac{k_3 \cdot k_2}{14} \right) \cdot \delta(v_3, v_2) + \left(a_{33} - \frac{k_3 \cdot k_3}{14} \right) \cdot \delta(v_3, v_3) \\
&\quad + \left(a_{44} - \frac{k_4 \cdot k_4}{14} \right) \cdot \delta(v_4, v_4) + \left(a_{45} - \frac{k_4 \cdot k_5}{14} \right) \cdot \delta(v_4, v_5) + \left(a_{46} - \frac{k_4 \cdot k_6}{14} \right) \cdot \delta(v_4, v_6) \\
&\quad + \left(a_{54} - \frac{k_5 \cdot k_4}{14} \right) \cdot \delta(v_5, v_4) + \left(a_{55} - \frac{k_5 \cdot k_5}{14} \right) \cdot \delta(v_5, v_5) + \left(a_{56} - \frac{k_5 \cdot k_6}{14} \right) \cdot \delta(v_5, v_6) \\
&\quad \left. + \left(a_{64} - \frac{k_6 \cdot k_4}{14} \right) \cdot \delta(v_6, v_4) + \left(a_{65} - \frac{k_6 \cdot k_5}{14} \right) \cdot \delta(v_6, v_5) + \left(a_{66} - \frac{k_6 \cdot k_6}{14} \right) \cdot \delta(v_6, v_6) \right] \\
&= \frac{1}{14} \cdot \left[\left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \right. \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(0 - \frac{3 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{3 \cdot 2}{14} \right) \cdot 1 \\
&\quad + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 \\
&\quad \left. + \left(1 - \frac{2 \cdot 3}{14} \right) \cdot 1 + \left(1 - \frac{2 \cdot 2}{14} \right) \cdot 1 + \left(0 - \frac{2 \cdot 2}{14} \right) \cdot 1 \right] \\
&= 0,357
\end{aligned}
\tag{E.14}$$

Bibliografía

- [1] Lada A. Adamic y Bernardo A. Huberman. «Power-Law Distribution of the World Wide Web». En: *Science* 287.5461 (2000), págs. 2115-2115. DOI: 10.1126/science.287.5461.2115a. eprint: <https://www.science.org/doi/pdf/10.1126/science.287.5461.2115a>. URL: <https://www.science.org/doi/abs/10.1126/science.287.5461.2115a>.
- [2] Réka Albert y Albert-László Barabási. «Statistical mechanics of complex networks». En: *Rev. Mod. Phys.* 74.1 (2002), págs. 47-97. DOI: 10.1103/RevModPhys.74.47. URL: <https://link.aps.org/doi/10.1103/RevModPhys.74.47>.
- [3] Maximino Aldana. «Redes complejas». En: *Recuperado a partir de http://www.fis.unam.mx/~max/English/notasredes.pdf* (2006). URL: <https://shre.ink/gVOS>.
- [4] Ginestra Bianconi y Albert-László Barabási. «Bose-Einstein Condensation in Complex Networks». En: *Phys. Rev. Lett.* 86.24 (2001), págs. 5632-5635. DOI: 10.1103/PhysRevLett.86.5632. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.86.5632>.
- [5] Ismael Crehuet Lucas. «El problema del viajante con grafos». Trabajo Fin de Grado. Grado en Matemáticas. Trabajo Fin de Grado. Universidad de Valladolid, 2022. URL: <https://uvadoc.uva.es/bitstream/handle/10324/57985/TFG-G5977.pdf?sequence=1&isAllowed=y>.
- [6] Autor Desconocido. *Estudio sobre la red del universo de Marvel*. <http://dmi.uib.es/~joe/marvel.html>. En la red del universo de Marvel, dos superhéroes están conectados si han aparecido por lo menos una vez en el mismo cómic. De este estudio resulta que es el hombre araña el superhéroe con el máximo número de conexiones.
- [7] S. N. Dorogovtsev y J. F. F. Mendes. «Evolution of networks». En: *Advances in Physics* 51.4 (2002), págs. 1079-1187. DOI: 10.1080/00018730110112519. eprint: <https://doi.org/10.1080/00018730110112519>. URL: <https://doi.org/10.1080/00018730110112519>.
- [8] Joaquín Muñoz García. *Tema 7: Modelos de variables aleatorias discretas y absolutamente continuas*. Cálculo de Probabilidades y Estadística Matemática. Universidad de Sevilla.
- [9] Andrea Garritano. *Twitch Social Networks Dataset*. 2024. URL: <https://www.kaggle.com/datasets/andreagarritano/twitch-social-networks>.
- [10] Gephi. *Gephi - Users*. 2024. URL: <https://gephi.org/users/>.
- [11] P. L. Krapivsky, S. Redner y F. Leyvraz. «Connectivity of Growing Random Networks». En: *Phys. Rev. Lett.* 85 (2000), págs. 4629-4632. DOI: 10.1103/PhysRevLett.85.4629. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.85.4629>.
- [12] José Ignacio Farrán Martín. *Grafos y Redes Sociales*. Matemática Discreta. Universidad de Valladolid.

- [13] José Ignacio Farrán Martín. *MATEMÁTICA DISCRETA Grafos y Árboles*. Matemática Discreta. Universidad de Valladolid.
- [14] J. McAuley y J. Leskovec. «Learning to Discover Social Circles in Ego Networks». En: *Advances in Neural Information Processing Systems (NIPS)*. NeurIPS, 2012.
- [15] Neo4j. *HITS Algorithm*. 2024. URL: <https://neo4j.com/docs/graph-data-science/current/algorithms/hits/>.
- [16] Mark. E. J. Newman. «Spread of Epidemic Disease on Networks». En: *Phys. Rev. E* 66.1 (2002), pág. 16128. DOI: 10.1103/PhysRevE.66.016128. URL: <https://link.aps.org/doi/10.1103/PhysRevE.66.016128>.
- [17] Mark. E. J. Newman. «The Structure and Function of Complex Networks». En: *SIAM Review* 45.2 (2003), págs. 167-256. DOI: 10.1137/S003614450342480. eprint: <https://doi.org/10.1137/S003614450342480>. URL: <https://doi.org/10.1137/S003614450342480>.
- [18] Romualdo Pastor-Satorras y Alessandro Vespignani. «Epidemic Spreading in Scale-Free Networks». En: *Phys. Rev. Lett.* 86.14 (2001), págs. 3200-3203. DOI: 10.1103/PhysRevLett.86.3200. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.86.3200>.
- [19] Rodrigo Aldecoa Rodríguez. «Detección de comunidades en redes complejas». Trabajo Fin de Máster. Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital. Trabajo Fin de Máster. Universitat Politècnica de València, 2012. URL: https://riunet.upv.es/bitstream/handle/10251/15337/TFM_RodrigoAldecoa.pdf.
- [20] Statista. *Ranking mundial de redes sociales por número de usuarios*. 2023. URL: <https://es.statista.com/estadisticas/600712/ranking-mundial-de-redes-sociales-por-numero-de-usuarios/>.
- [21] VadeNumeros. *Combinatoria*. 2024. URL: <https://www.vadenumeros.es/sociales/combinatoria.htm>.