



POLITÉCNICA



UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA

AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

GRADO EN BIOTECNOLOGÍA

**Modelización de datos de geolocalización de especies
vegetales mediante redes complejas**

TRABAJO DE FIN DE GRADO

Autor: Leonardo Claudín González

Tutores: José Ángel Capitán Gómez, Juan Carlos Sanz Nuño



POLITÉCNICA



UNIVERSIDAD POLITÉCNICA DE MADRID

**Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de
Biosistemas**

GRADO EN BIOTECNOLOGÍA

**MODELIZACIÓN DE DATOS DE GEOLOCALIZACIÓN DE ESPECIES VEGETALES
MEDIANTE REDES COMPLEJAS**

TRABAJO FIN DE GRADO

Leonardo Claudín González

MADRID, 2021

Tutores: José Ángel Capitán Gómez, Juan Carlos Sanz Nuño

Departamento de Matemática Aplicada, Universidad Politécnica de Madrid



POLITÉCNICA



TÍTULO DEL TFG - Modelización de datos de
geolocalización de especies vegetales mediante redes complejas

Memoria presentada por **Leonardo Claudín González**
para la obtención del título de Graduado en
Biotecnología por la Universidad Politécnica de
Madrid

Fdo: Leonardo Claudín González

VºBº Cotutor y Codirector del TFG

José Ángel Capitán Gómez

Profesor Contratado Doctor

Dpto. de Matemática Aplicada

ETSIAAB - Universidad Politécnica de Madrid

VºBº Cotutor y Codirector del TFG

Juan Carlos Sanz Nuño

Profesor Titular de Universidad

Dpto. de Matemática Aplicada

ETSIAAB - Universidad Politécnica de Madrid

Madrid, 8, Julio, 2021

Índice

Índice de figuras	vi
Índice de cuadros	vii
Abreviaturas	vii
CAPÍTULO 1. Introducción y objetivos	1
1.1 Abstract	1
1.2 Introducción	1
1.3 Objetivos	2
CAPÍTULO 2. Materiales y métodos	4
2.1 Principales propiedades topológicas de redes complejas.	5
CAPÍTULO 3. Resultados	11
3.1 Redes sintéticas. El tamaño de la muestra puede importar	11
3.1.1 Redes aleatorias	11
3.1.2 Efecto del muestreo en redes geométricas aleatorias	15
3.2 Redes reales	19
CAPÍTULO 4. Discusión	25
CAPÍTULO 5. Conclusiones	29
CAPÍTULO 6. Bibliografía	30
Anexo I. Repositorio GitHub	32
Anexo II. Tablas de resultados de redes completas.	34

Índice de figuras

1	Ocurrencias de alcornoques	5
2	Ocurrencias de chumberas	5
3	Ocurrencias de olivos	5
4	Ocurrencias de viñedos	5
5	Ejemplos de redes aleatorias, geométricas aleatorias y reales.	7
6	Tamaño de la componente conexa mayor (LCC) del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n). Las líneas verticales corresponden con los puntos críticos —ecuación (3)— de cada red aleatoria; Cuando el tamaño del subgrafo cumple que $np = 1$.	11
7	Modularidad del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n)	12
8	Escalado de la Modularidad del muestreo de redes Erdős-Renyi para valores de la modularidad que han superado el punto crítico.	13
9	Coefficiente de clustering del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n).	13
10	Diámetro del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n).	14
11	Escalado de la figura 10. La línea vertical representa el punto crítico definido en la ecuación (3).	14
12	Densidad de conexiones del muestreo de redes Erdős-Renyi en función del tamaño de muestreo y la probabilidad de conexión (p)	15
13	Tamaño de la componente conexa mayor del muestreo de una red geométrica aleatoria (RGG) en función del radio de vecindad (r) y del tamaño de muestreo (n).	16
14	Modularidad del muestreo de una red geométrica aleatoria (RGG) en función del radio de vecindad (r) y del tamaño del muestreo (n).	16
15	Coefficiente de clustering del muestreo de una red geométrica aleatoria (RGG) en función del tamaño del muestreo (n) y del radio de vecindad (r)	17

16	Diámetro del muestreo de una red geométrica aleatoria (RGG) en función del tamaño muestreado (n) y el radio de vecindad (r)	18
17	Densidad de conexiones de una red geométrica aleatoria (RGG) en función del tamaño muestreado (n) y el radio de vecindad (r)	18
18	En vertical, el tamaño $C_{G(n^*,\theta)}$ de LCC. En horizontal, la distancia umbral (θ). Los valores se agrupan por tamaño muestral (n^*) y se clasifican por especie. . .	19
19	En vertical, el tamaño $C_{G(n^*,\theta)}$ de LCC. En horizontal, el tamaño muestral (n^*). Los valores se agrupan por distancias umbrales (θ) y se clasifican por especie. .	20
20	En vertical, la modularidad $Q(G(n^*,\theta))$. En horizontal, la distancia umbral (θ). Los valores se agrupan por tamaño muestral (n^*) y se clasifican por especie. . .	20
21	En vertical, la modularidad $Q(G(n^*,\theta))$. En horizontal, el tamaño muestral (n^*). Los valores se agrupan por distancias umbrales (θ) y se clasifican por especie.	21
22	En vertical, el coeficiente de clustering $CC(G(n^*,\theta))$. En horizontal, la distancia umbral (θ). Agrupados por tamaño muestral (n^*) y se clasifican por especie. . .	21
23	En vertical, el coeficiente de clustering $CC(G(n^*,\theta))$. En horizontal, el tamaño muestral (n^*). Agrupados por distancias umbrales (θ) y se clasifican por especie. .	22
24	En vertical, el diámetro $Diam(G(n^*,\theta))$. En horizontal, la distancia umbral (θ). Los valores se agrupan por tamaño muestral (n^*) y se clasifican por especie. . .	22
25	En vertical, el diámetro $Diam(G(n^*,\theta))$. En horizontal, el tamaño muestral (n^*). Los valores se agrupan por distancias umbrales (θ) y se clasifican por especie.	23
26	En vertical, la densidad de conexiones $EdgeD(G(n^*,\theta))$. En horizontal, la distancia umbral (θ). Agrupados por tamaño muestral (n^*) y se clasifican por especie.	23
27	En vertical, la densidad de conexiones $EdgeD(G(n^*,\theta))$. En horizontal, el tamaño muestral (n^*). Agrupados por distancias umbrales (θ) y se clasifican por especie.	24

Índice de cuadros

1	¿Afecta el muestreo de nodos a las propiedades estructurales de las redes reales estudiadas?	28
2	Determinación de cada propiedad en las redes reales	28
3	Propiedades de los Alcornosques	34
4	Propiedades de las Chumberas	34
5	Propiedades de los Olivos	35
6	Propiedades de los Viñedos	35

Abreviaturas

GBIF	Global Biodiversity Information Facility
MMU	Minimum mapping unit
QGIS	Quantum Geographic Information System
GRASS	Geographic Resources Analysis Support System
WGS84	Sistema Geodésico Global de 1984
LCC	Largest Connected Component
Q	Modularidad
CC	Clustering Coefficient
Diam	Diámetro
EdgeD	Densidad de Conexiones
RGG	Random Geometric Graph

CAPÍTULO 1. Introducción y objetivos

1.1. Abstract

We analyse the properties of real plant-based networks constructed from geospatial databases, in particular vineyards, olive trees, cork oak and prickly pear in the Iberian Peninsula. Due to the large scale of such networks, we study the sampling dependence of the networks measures: giant component, clustering coefficient, modularity, diameter and degree distribution. To compare with, we review the sampling properties of classical synthetic networks, namely random Erdős-Renyi networks and its spatial variant, the random geometric networks. Concerning sampling, the main results obtained were: 1) Clustering coefficient and Degree Distribution are independent of the sampling size, both on synthetic and real networks. 2) While modularity depends on the sampling size in synthetic networks, it behaves similarly for every size of the plant-based complex networks. Observed topological properties of real plant-based networks are framed and contextualized according to our results (Table 2).

1.2. Introducción

En la actualidad tenemos a nuestra disposición bases de datos suficientemente completas para obtener una aproximación de la estructura de las redes de plantas del mundo. En particular, en repositorios oficiales de instituciones europeas se puede encontrar información de geolocalización de muchas especies de interés agrícola o forestal. Estas bases de datos empiezan a tener unos tamaños que las hacen ciertamente inaccesibles para su análisis en ordenadores personales. Ante esta dificultad computacional, se plantean dos opciones inmediatas: (i) solicitar el acceso a un supercomputador, como el que dispone la UPM o (ii) confiar en el muestreo de parte de los datos para inferir información de la red completa. Esta última opción se ha adoptado en este TFG. Desde el punto de vista académico, probablemente sea el enfoque más adecuado, ya que en el mundo actual se manejan grandes volúmenes de datos, con lo cual un tratamiento estadístico de los mismos se hace indispensable cuando es imposible acceder a ellos o almacenarlos en su totalidad.

Como explican *Dan Frank et al.*, nos enfrentamos a uno de los grandes problemas en comunidades ecológicas: la definición de escalas que preserven las propiedades de la red original en

la mayor medida posible [1]. Pero, ¿qué ocurre si las propiedades observadas en el muestreo son significativamente diferentes de las que pudieran presentar las redes originales? En estos casos, el objetivo no es definir un tamaño de muestreo a partir del cual las propiedades se asemejan a las de la red real. Más bien se trata de definir una tendencia de las propiedades para poder anticipar sus valores en el sistema [2].

La solución más frecuente a las que se ha recurrido consiste en realizar estudios del efecto que tiene el muestreo [1], que es por lo que se ha optado en este trabajo fin de grado.

En este trabajo se estudia la estructura del sistema espacial de varias poblaciones de plantas que crecen en la península ibérica. El problema, como anticipamos, es el tamaño poblacional de las especies consideradas en este trabajo: olivos, vides, alcornoques y chumberas. Los datos se han extraído de la base de datos GBIF (<http://www.gbif.org/>). El elevado número de individuos (169.906 para olivos, 21.178 para vides, 433.540 para alcornoques, y 5.611 para chumberas), supone un factor limitante para su estudio.

Para sortear estas limitaciones, primero se analizó el efecto del muestreo en redes sintéticas de referencia (redes Erdős-Renyi y su variante espacial, redes geométricas aleatorias) y, posteriormente, se repitió el mismo análisis para nuestras redes reales de ocurrencias de plantas.

1.3. Objetivos

El objetivo principal de este trabajo fin de grado es la determinación de la red espacial que forman algunas comunidades de plantas a partir de los datos de geolocalización disponibles en organismos oficiales. Como hito intermedio para conseguir este objetivo, el trabajo aborda el problema del muestreo en redes complejas. Más allá del ámbito agroforestal, la aplicación de las técnicas de muestreo a redes complejas es de gran importancia, sobre todo debido al gran tamaño de, por ejemplo, las redes sociales [3]. Para lograr una primera valoración del efecto del muestreo en redes, este trabajo analiza redes sintéticas simples, en concreto, las redes aleatorias de Erdős-Renyi, y su variante espacial, las denominadas redes geométricas. Sobre estas redes se analizan algunos de sus principales índices: modularidad, coeficiente de clustering, tamaño de la mayor componente conexa, grado de conexión y diámetro.

Una vez entendido el alcance del muestreo sobre redes sintéticas completas, se aplican estas

técnicas sobre la red real y se obtienen sus propiedades derivadas de las leyes de muestreo determinadas. Por último, estos resultados se discuten en el contexto de la epidemiología de plantas, objetivo importante que ha quedado fuera de este TFG, pero que se planea abordar en el futuro.

CAPÍTULO 2. Materiales y métodos

Bases de Datos de geolocalización de especies vegetales. La información de los sitios de cultivo de vid se ha obtenido del repositorio de 2018 (CORINE land-cover) de Land Copernicus (<https://land.copernicus.eu>), con una resolución de 25Ha en la escala de MMU. Este formato almacena los sitios de cultivo como polígonos en WGS84. Para obtener las coordenadas geográficas de cada ocurrencia, el archivo fue procesado con el programa QGIS 3.10.14 with GRASS 7.8.5 (<https://www.qgis.org/es/site/forusers/download.html>). Se extrajeron los centroides de los polígonos que representan los cultivos de vid con una herramienta de geometría vectorial. En el caso de chumberas (*Opuntia ficus-indica.*), olivos (*Olea europea*) y alcornoques (*Quercus suber*), el censo extraído de GBIF incluye coordenadas geográficas (<https://www.gbif.org/occurrence>). La resolución media en la mayoría de las ocurrencias es de 10 metros, trabajando con una precisión de 6 decimales.

Elaboración de redes. A partir de los datos de geolocalización obtenidos de las bases de datos descritas anteriormente, que se representan en las figuras [1](#), [2](#), [3](#) y [4](#), se construyen las redes o grafos de interacción para cada especie. Se considera que dos individuos están conectados entre sí, de forma bidireccional, si están relacionados respecto alguna propiedad. Los individuos en la red son representados por nodos o vértices y la relación entre ellos, cuando existe, mediante arcos o enlaces. Por ejemplo, dos individuos que se encuentran a menos de una distancia de interacción notan la presencia uno del otro, bien porque compiten por recursos o bien porque quedan dentro de la distancia típica de dispersión de un patógeno. En este caso, los correspondientes nodos estarían conectados por un arco bidireccional. La sección [2.1](#) desarrolla con más detalle la descripción de este tipo de redes.

De acuerdo con esta definición, la red correspondiente a una especie depende del radio de interacción. Para valores del radios pequeños, los individuos (nodos) quedarán muy desconectados, mientras que para radios grandes, casi todos los individuos (nodos) estarán conectados. Ésta transición entre una situación y otra depende, como veremos más adelante, de la especie y de la propiedad estructural que se analice de las redes.

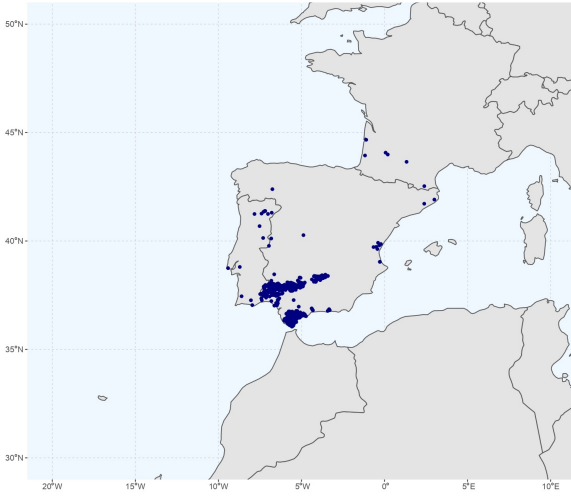


Figura 1: Ocurrencias de alcornoques

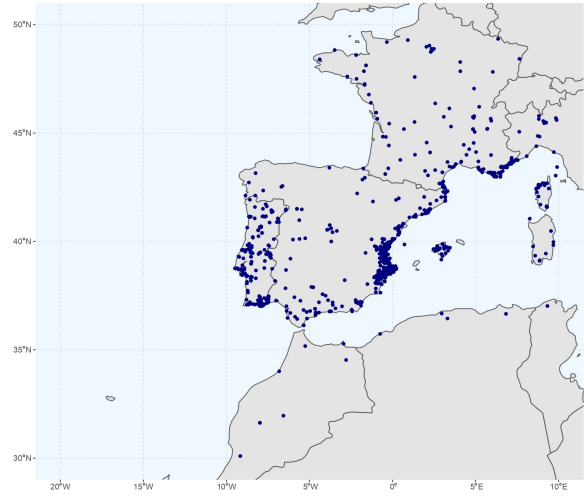


Figura 2: Ocurrencias de chumberas

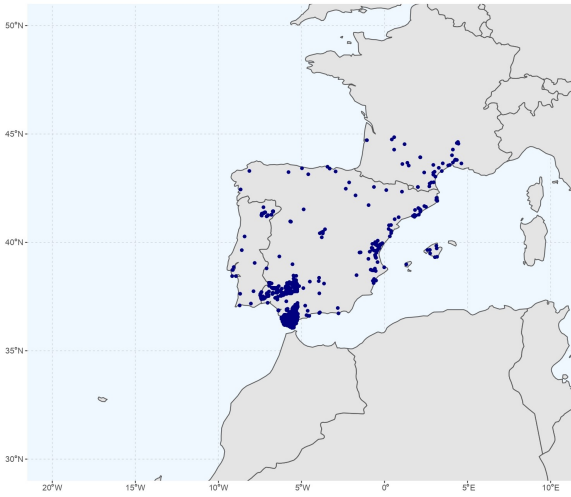


Figura 3: Ocurrencias de olivos

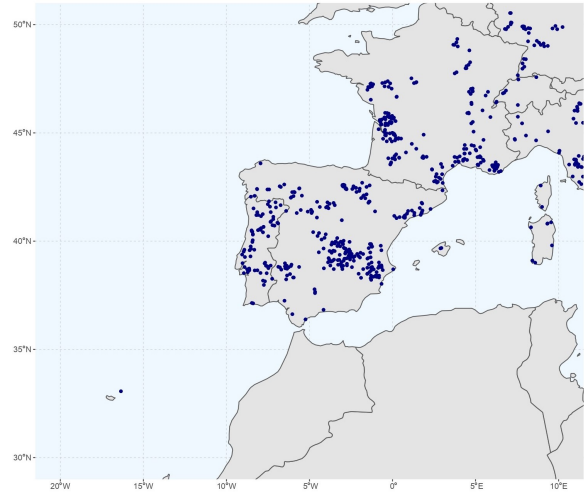


Figura 4: Ocurrencias de viñedos

2.1. Principales propiedades topológicas de redes complejas.

Definición de nodo y conectividad de la red. En este estudio, cada individuo de una misma especie se trata como un nodo de la red. Al tener geolocalizados todos los individuos, podemos asociar a cada nodo un vector \mathbf{r}_i de dos componentes,

$$\mathbf{r}_i = (X_i, Y_i), \quad (1)$$

siendo X_i e Y_i la latitud y longitud correspondientes al individuo i . Utilizando coordenadas esféricas (WGS84), calculamos la distancia $d(\mathbf{r}_i, \mathbf{r}_j)$ sobre la superficie esférica entre cada par

de individuos. Si esta distancia, medida en kilómetros, es menor que la distancia umbral θ , definiremos una conexión entre ambos nodos. Esta conexión tiene el mismo peso en ambos sentidos, es decir, no existe direccionalidad en las conexiones (redes no dirigidas).

La matriz de adyacencia $A = (A_{ij})$ recoge las conexiones entre todos los pares de nodos de la red. Se define, por tanto, como

$$A_{ij} = \begin{cases} 1 & \text{si } d(\mathbf{r}_i, \mathbf{r}_j) \leq \theta, \\ 0 & \text{si } d(\mathbf{r}_i, \mathbf{r}_j) > \theta, \end{cases} \quad (2)$$

de forma que los nodos i y j están conectados si $A_{ij} = 1$ y no lo están si $A_{ij} = 0$.

Redes sintéticas de referencia. Se han empleado dos modelos de generación de redes aleatorias (Figura 5).

- El modelo de *Erdős-Renyi*, en el que se define un número de nodos n y una probabilidad p de que un par de nodos esté conectado [4]. Se genera un grafo aleatorio que denotaremos como $G(n, p)$.
- Las *redes geométricas aleatorias*, que son una variante espacial de las redes Erdős-Renyi. Los nodos se localizan de forma aleatoria en un cuadrado de lado unidad, y se establecen arcos si la distancia euclídea entre nodos es inferior a una distancia de vecinaje, r [5]. El grafo generado se denotará por $G(n, r)$.

Los comandos empleados son `erdos.renyi.game()` y `sample_grg()`, respectivamente. Pertenecen al paquete `igraph` de **R** [6].

Tamaño de la componente conexa mayor. Es el número de nodos de la red que están conectados en una componente gigante (LCC). Quedan conectados entre sí a través de caminos de tamaño finito sobre los enlaces de la red.

En redes aleatorias $G(n, p)$, una vez superado el punto crítico de conectividad ($p > p_c \approx 1/n$), aparece una componente gigante que tiende a englobar todos los nodos de la red [7]. El tamaño C de dicha componente verifica que, si

$$np > 1, \quad (3)$$

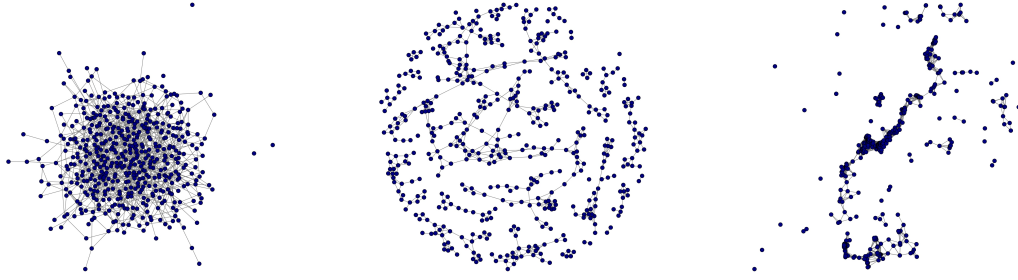


Figura 5: Ejemplos de redes aleatorias, geométricas aleatorias y reales.

La figura (a) es una red Erdős-Renyi $G(500, p = 0.008)$. La figura (b) es una red geométrica aleatoria $G(500, r = 0.05)$. La figura (c) es un muestreo de una red real de chumberas $G(500, \theta = 100km)$.

entonces $C_{G(n,p)}$ tiende a n .

En redes geométricas aleatorias $G(n, r)$, el tamaño de la componente gigante se acerca al tamaño total cuando el radio de vecindad (r) supera al radio crítico de conectividad [8]:

$$r_c \approx \sqrt{\frac{\log(n)}{n\pi}}. \quad (4)$$

Modularidad de la red. La modularidad es un valor escalar comprendido entre 0 y 1 que mide la densidad de enlaces dentro de comunidades (o módulos) en comparación con la densidad de enlaces entre comunidades distintas [9]. Es una cantidad que se calcula a partir de la descomposición de la red en módulos, que se espera que estén densamente conectados. La modularidad del grafo G asociada a una partición P de nodos en comunidades $P = \{C^{(i)}\}$ se define como

$$Q(G, P) = \frac{1}{2m} \sum_{i,j=1}^n \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C^{(i)}, C^{(j)}), \quad (5)$$

donde $2m$ es el número total de enlaces de la red, A_{ij} es la matriz de adyacencia de la red, k_i y k_j son los grados de conexiones de los nodos i y j , respectivamente. $C^{(i)}$ es la comunidad a la que pertenece el nodo i y la función delta se define como $\delta(C^{(i)}, C^{(j)}) = 1$ si $C^{(i)} = C^{(j)}$, y $\delta(C^{(i)}, C^{(j)}) = 0$ si $C^{(i)} \neq C^{(j)}$. De esta forma, si los nodos i y j no pertenecen a la misma comunidad, el elemento del sumatorio es nulo [9].

Hay algoritmos heurísticos que optimizan el valor de la modularidad encontrando una partición

G que divide los nodos en diferentes comunidades. Se ha utilizado el algoritmo de Louvain [10], implementado a través de la herramienta `clusterlouvain()` de `igraph`.

En redes aleatorias $G(n, p)$, de acuerdo con *Mediarmaid et al.*, la modularidad alcanza un máximo en $np = 1$. Superado el punto crítico p_c de conectividad, la modularidad decae según la ley potencial

$$Q \approx (np)^{-\lambda} \quad (6)$$

donde λ es un parámetro positivo [11].

En redes geométricas aleatorias $G(n, r)$, la modularidad alcanza su máximo en el radio crítico de conectividad dado por la ecuación (4) [12]. Para radios de vecindad por encima del radio crítico, la modularidad desciende según una ley potencial similar al caso de redes aleatorias.

Coefficiente de clustering. Es una medida de cómo de representativos son los triángulos en la red. En un grafo completo (todos los nodos están conectados con todos), aparecen todos los triángulos posibles entre ternas, de forma que el “clustering” es máximo. En redes reales, se espera que se mantenga constante o decaiga logarítmicamente con el número medio de enlaces [13]. El coeficiente de clustering local para el nodo i se calcula como

$$c_i = \frac{t_i}{\binom{k_i}{2}}, \quad (7)$$

donde t_i es el número de enlaces que existen entre los vecinos del nodo i y k_i es el grado (número de vecinos) de i , de forma que el número de enlaces que podrían existir entre los vecinos de i es $\binom{k_i}{2}$. El coeficiente local de clustering se puede interpretar como el número de tripletes en los que participa el nodo i dividido por el número de tripletes que podría formar. El coeficiente de clustering global se define como el promedio de los coeficientes locales,

$$CC(G) = \frac{1}{n} \sum_{i=1}^n c_i. \quad (8)$$

La mayoría de las redes reales tiene un mayor coeficiente de clustering que las redes aleatorias (ER o geométricas) debido, en parte, a la estructura modular que suelen mostrar las redes

reales [14].

En redes aleatorias $G(n, p)$, el coeficiente de clustering de cada nodo c_i coincide con la probabilidad p de enlaces (que, a su vez, es igual a el grado medio esperado para cada nodo). Al ser ésta la misma para todos, se cumple que el promedio es igual a p ,

$$CC(G) = c_i = p. \quad (9)$$

En una red geométrica aleatoria $G(n, r)$ (ver el Teorema 3 de [15]), para radios por debajo del radio crítico $r < r_c$, el coeficiente de clustering vale $C \approx 0.5865$. Para una red geométrica aleatoria de 5000 nodos, $r_c = 0.015$. Para radios r tales que $r < r_c$, el coeficiente de clustering valdrá 0.5865.

Diametro de la red. Es la distancia geodésica máxima sobre la red. La distancia geodésica entre dos nodos es el camino que involucra conexiones con menor distancia. El diámetro de la red corresponde, por tanto, con el camino entre dos nodos i y j de distancias mínimas que involucra el mayor número de nodos posible (*longest shortest path*, [16]):

$$\text{Diam}(G) = \max_{i,j} d(\mathbf{r}_i, \mathbf{r}_j), \quad (10)$$

donde el máximo recorre todos los pares de nodos posibles. En una red aleatoria $G(n, p)$, el comportamiento del diámetro viene determinado, al igual que la modularidad y la componente gigante, por el punto crítico $np = 1$. Por debajo del punto crítico, la red está desconectada y el diámetro de la red equivale al diámetro del árbol mayor [17]. Superado el punto crítico, se forma la componente gigante. Su diámetro determina el de la red, y decae a medida que aumenta la densidad de enlaces de la componente gigante [18], de la forma

$$\text{Diam}(G) = k \frac{\log(n)}{\log(np)}, \quad (11)$$

donde k es una constante positiva. El diámetro da una noción de la conectividad de la red. Cuanta mayor conectividad, menor es el diámetro. Las redes con una probabilidad de conexión muy alta tienen un diámetro muy bajo para cualquier tamaño del subgrafo. Sin embargo,

las redes desconectadas (por debajo del punto crítico $np = 1$) tienen un diámetro muy grande.

En una red geométrica aleatoria $G(n, r)$, el comportamiento del diámetro es similar al del modelo de Erdős-Renyi, con dos rangos o intervalos. De acuerdo con [19], por debajo del punto crítico (4), la red está desconectada. En esta situación, el diámetro de la red será el diámetro del árbol mayor (al igual que en redes aleatorias). Superado el punto crítico (4), se forma la componente gigante y el diámetro decae de manera similar al caso de redes aleatorias.

Densidad de enlaces. Es el cociente entre el número total de enlaces entre pares de nodos frente al número total de conexiones posibles entre nodos de una red. Es decir,

$$\text{EdgeD}(G) = \frac{\sum_{i,j=1}^n e_{ij}}{\binom{N}{2}}. \quad (12)$$

En una red aleatoria $G(n, p)$, la densidad de enlaces en un subgrafo $G^*(n, p)$ coincide con la probabilidad de conexión p [20].

En una red geométrica aleatoria $G(n, r)$, la densidad de conexión se mantiene constante para cualquier subgrafo $G^*(n, r)$ de la red. La dependencia de esta densidad no es inmediata, como en el caso de la red aleatoria.

Todas estas propiedades se han calculado usando el paquete *igraph* de **R** [21]. Los programas implementados para obtener los resultados se encuentran en el siguiente repositorio de GitHub:

<https://github.com/LeonardoClaudin/Trabajo-Fin-de-Grado-Biotecnologia-ETSIAB>.

CAPÍTULO 3. Resultados

3.1. Redes sintéticas. El tamaño de la muestra puede importar

3.1.1. Redes aleatorias

Se han generado redes Erdős-Renyi de 5000 nodos con probabilidades de conexión (p):

(0.1, 0.01, 0.005, 0.0033, 0.002, 0.001, 0.0007, 0.0006, 0.0005, 0.0004, 0.0003, 0.0002, 0.0001)

A partir de estas redes, se han obtenido subgrafos de tamaño creciente desde 100 a 5000 nodos. Todas las funciones utilizadas para el estudio de las probabilidades en redes aleatorias se encuentran en el archivo *funciones_redes_aleatorias.R*. Se han realizado promedios de 10 muestreos aleatorios para cada tamaño muestral.

Tamaño de la componente conexas mayor (LCC). El comportamiento de la componente gigante (sección 2.1) está representado en la figura 6.

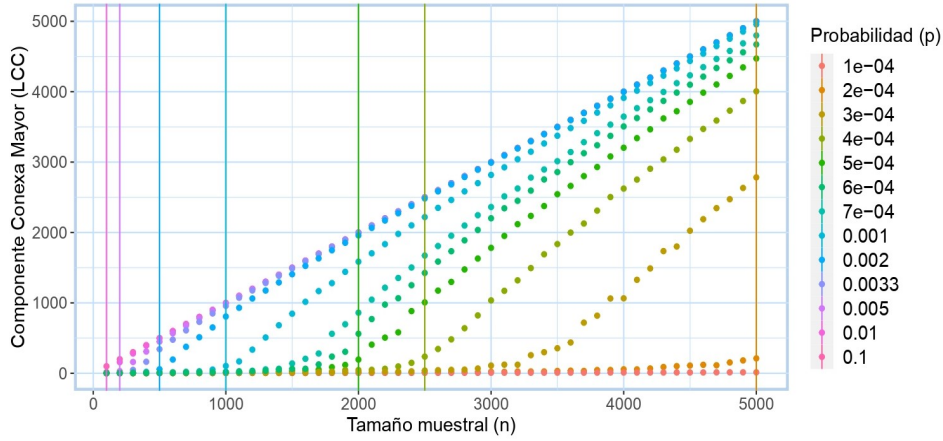


Figura 6: Tamaño de la componente conexas mayor (LCC) del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n). Las líneas verticales corresponden con los puntos críticos —ecuación (3)— de cada red aleatoria; Cuando el tamaño del subgrafo cumple que $n p = 1$.

De las redes $G(5000, p)$ generadas, aquellas cuyo muestreo verifique la ecuación (3) estarán completamente conectadas. Es decir, redes con $p > p_c = 0.0002$ están completamente conectadas. Se han estudiado varios muestreos de estas redes que no reflejan la conectividad real. En una red completamente conectada, el tamaño $C_{G(n,p)}$ de LCC tiende a n . Tomando como

ejemplo $G(5000, 0.001)$, una red conectada que, por el contrario, presenta muestreos que no reflejan esta conectividad para tamaños inferiores a 3000 nodos (para este tamaño, LCC se acerca a la bisectriz). Esta red presenta una diferencia de tamaño de, aproximadamente, 2000 nodos entre el punto crítico y el muestreo válido.

Modularidad. El comportamiento de la modularidad [2.1](#) queda representado en las curvas de la figura [7](#).

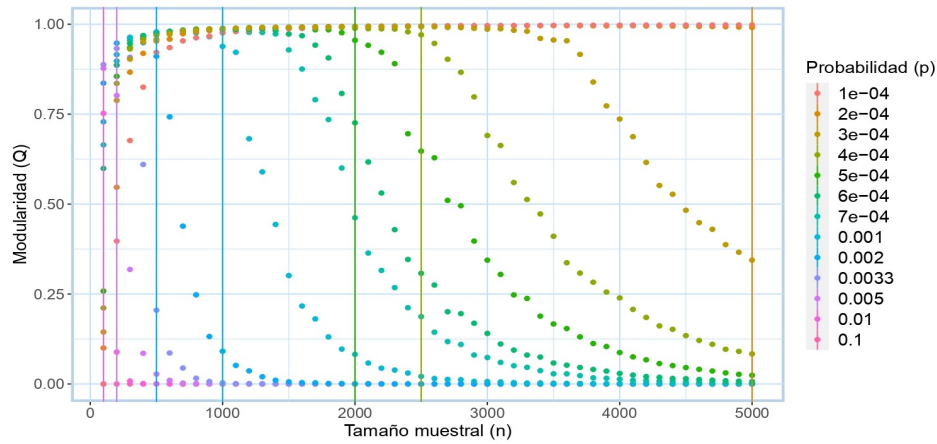


Figura 7: Modularidad del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n)

Las redes con $p > 0.0002$ presentan modularidad nula o descendiente según la ecuación [\(6\)](#). De la misma forma que en la LCC, existen muestreos de estas redes que no reflejan la modularidad real. En el caso de la red $G(n^*, 0.0004)$ con modularidad nula; los subgrafos de tamaño $n^* \in [0, 4500]$ presentan una modularidad mayor de la real. No es hasta que el muestreo tiene un tamaño $n^* = 2500$ - verifica la ecuación [\(3\)](#) - que la modularidad decrece. El muestreo aleatorio tiene un gran efecto sobre la modularidad, siendo válidas solo aquellas muestras con un tamaño cercano al de la red completa.

El descenso de la modularidad sigue la ecuación [\(6\)](#). La figura [8](#) representa este descenso exponencial de la modularidad como función del producto del tamaño del subgrafo por la probabilidad de conexión. Tras realizar un ajuste no lineal de mínimos cuadrados [$nls()$ en **R**], la modularidad se comporta como

$$Q(n, p) = (np)^{-2.83}. \quad (13)$$

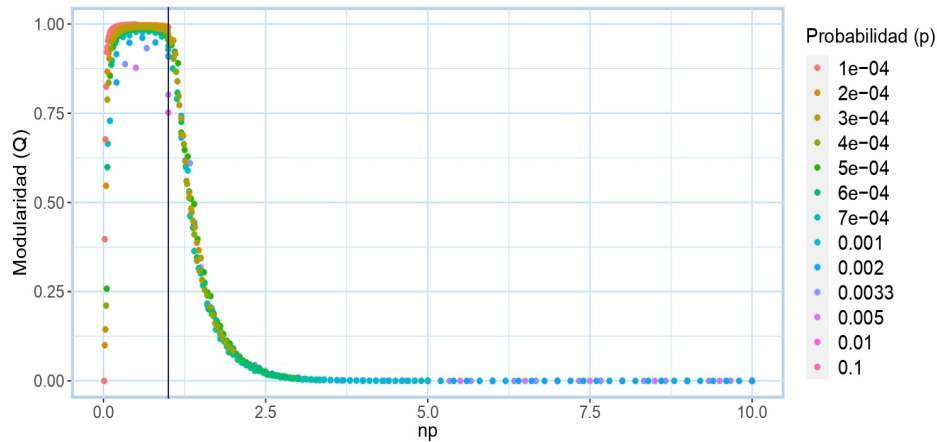


Figura 8: Escalado de la Modularidad del muestreo de redes Erdős-Renyi para valores de la modularidad que han superado el punto crítico.

La línea vertical, al igual que en la figura 6, corresponde con el punto crítico $np = 1$ de la ecuación 3. A la izquierda de dicha línea, los muestreos no han superado el punto crítico y la modularidad es no nula.

Coefficiente de clustering. Los resultados representados en la figura 9 corroboran la ecuación 9 que define el comportamiento del coeficiente de clustering en redes aleatorias. Su valor es el de la probabilidad de conexión.

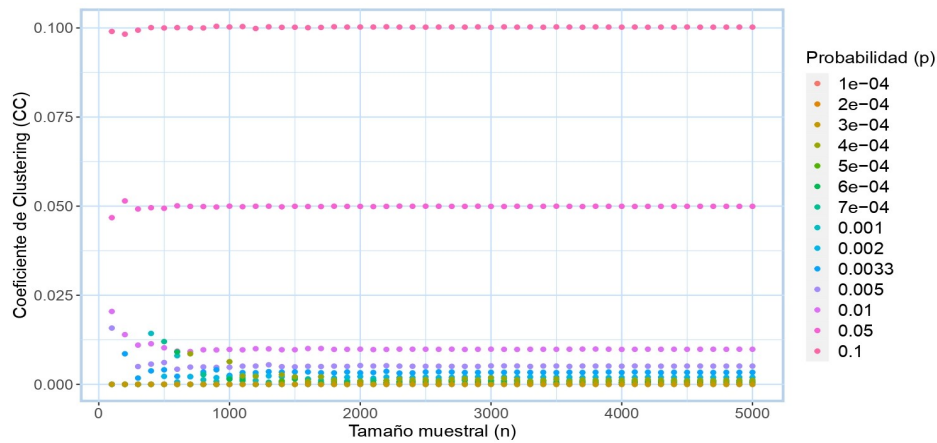


Figura 9: Coeficiente de clustering del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n).

Diámetro. Los resultados representados en la figura 10 confirman el comportamiento del diámetro en redes aleatorias definido en la ecuación 11).

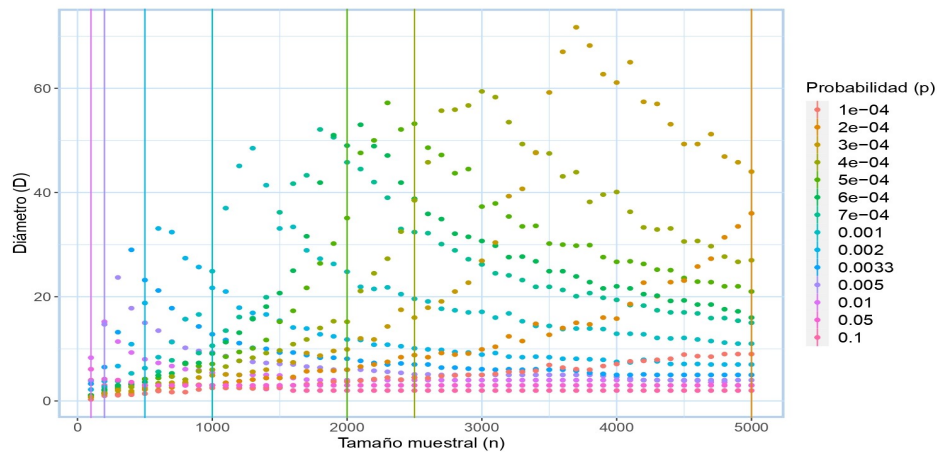


Figura 10: Diámetro del muestreo de redes Erdős-Renyi en función de la probabilidad de conexión (p) y del tamaño de muestreo (n).

Una manera de captar el comportamiento dado por la ecuación (11) consiste en representar el diámetro en un sistema de ejes escalados, tal y como se ve en la figura 11. Un ajuste de las curvas da lugar a una constante de proporcionalidad $k = 1$, esto es

$$\text{Diam}(n, p) = \frac{\log(n)}{\log(np)}. \quad (14)$$

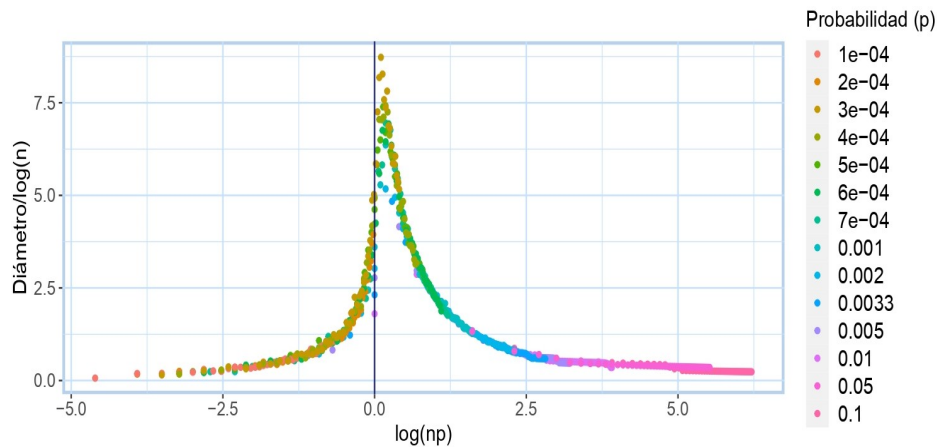


Figura 11: Escalado de la figura 10. La línea vertical representa el punto crítico definido en la ecuación (3).

Densidad de conexiones. Se cumple lo postulado en la sección [2.1](#) acerca de la densidad de conexión para redes aleatorias en la figura [12](#). Es decir,

$$\text{EdgeD}(n, p) = p. \quad (15)$$

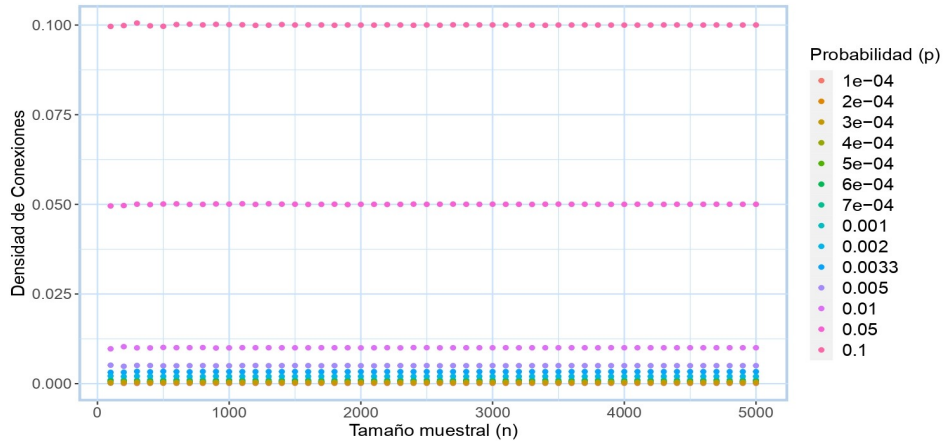


Figura 12: Densidad de conexiones del muestreo de redes Erdős-Renyi en función del tamaño de muestreo y la probabilidad de conexión (p)

Al igual que el coeficiente de clustering para redes aleatorias Erdős-Renyi, el muestreo no tiene efecto alguno sobre la densidad de conexiones de la red.

3.1.2. Efecto del muestreo en redes geométricas aleatorias

Se han generado redes *Random Geometric* de 5000 nodos y distancias de vecindad (r):

$$(1, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.01, 0.001)$$

A partir de estas redes, se han obtenido subgrafos de tamaño creciente desde 100 a 5000 nodos. Todas las funciones utilizadas para el estudio de las probabilidades en redes aleatorias se encuentran en el archivo *funciones_redes_geometricas.R*.

Se han realizado promedios de 10 muestreos aleatorios para cada tamaño muestral.

Tamaño de la componente conexa mayor (LCC). Según la sección [2.1](#) para redes geométricas aleatorias, el radio crítico de la conectividad es $r_c = \sqrt{\frac{\log(5000)}{5000\pi}} = 0.015$.

Disponemos de dos rangos de redes $G(n, r)$:

- $r > r_c$: la red está completamente conectada y el tamaño de LCC $C_{G(n^*,p)}$ tiende a n^*
- $r < r_c$: la red está desconectada y el tamaño de LCC es $C_{G(n^*,p)} \ll n^*$.

La representación en la figura [13](#) confirma estos dos rangos de redes.

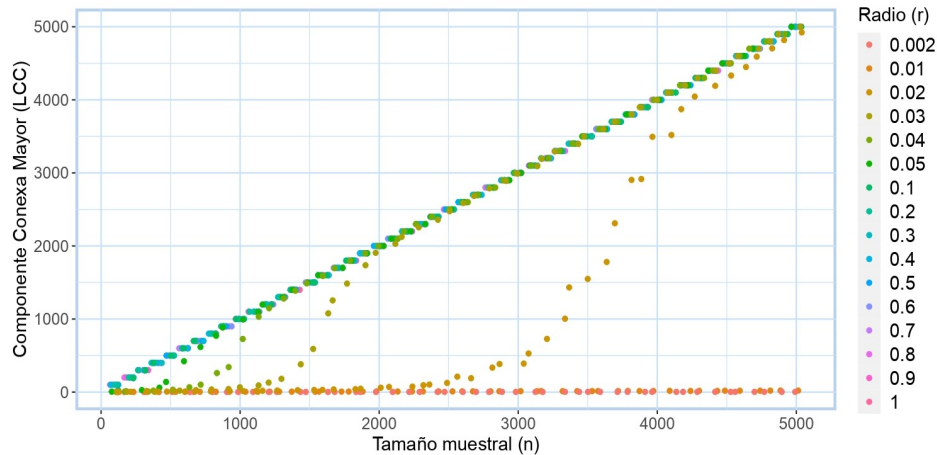


Figura 13: Tamaño de la componente conexa mayor del muestreo de una red geométrica aleatoria (RGG) en función del radio de vecindad (r) y del tamaño de muestreo (n).

Paralelamente a redes aleatorias, si un subgrafo está completamente conectado, también lo estará la red completa de la que se muestreó. Sin embargo, para redes completas y completamente conectadas, existen muestreos que no reflejan esta conectividad.

Modularidad. La modularidad también depende del radio crítico definido en la ecuación [\(4\)](#), como vemos en la figura [14](#).

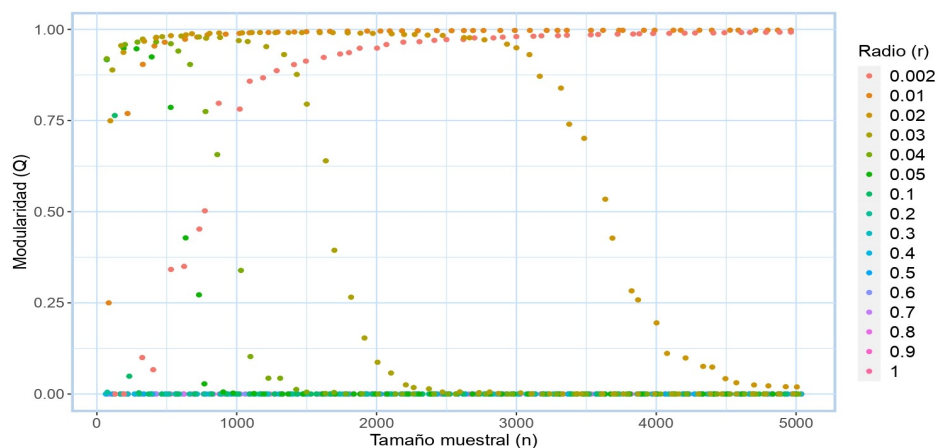


Figura 14: Modularidad del muestreo de una red geométrica aleatoria (RGG) en función del radio de vecindad (r) y del tamaño del muestreo (n).

El muestreo tiene un efecto similar al caso de redes aleatorias. Las redes con radio $r > r_c$ tienen modularidad nula. Por el contrario, las redes $r < r_c$ tienen modularidad no nula (máxima). Del primer caso de redes existen muestreos con una modularidad significativamente distinta. En el caso de $G(n^*, 0.02)$, una red con modularidad nula, para muestreos con tamaño $n^* \in [0, 4500]$ la modularidad reflejada no es nula, a diferencia de la real. Es decir, el efecto del muestreo es tal que, para estimar la modularidad real de la red, se deben tomar muestreos con tamaños cercanos al máximo.

Coefficiente de clustering. La representación de la figura 15 corrobora el comportamiento definido para el CC en la sección 2.1. El coeficiente es independiente del muestreo, y, además, en redes con $r < r_c = 0.015$, CC vale 0.5868.

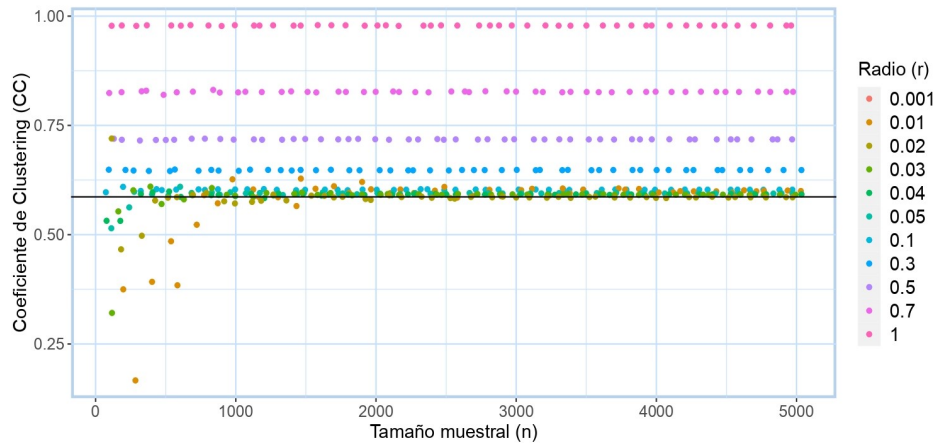


Figura 15: Coeficiente de clustering del muestreo de una red geométrica aleatoria (RGG) en función del tamaño del muestreo (n) y del radio de vecindad (r)

Diámetro. Su comportamiento, determinado por la ecuación (11), se verifica en la representación de la figura 16. El efecto que el muestreo tiene sobre esta propiedad es análogo al caso de redes aleatorias. Aquellas redes que han superado la conectividad crítica ($r > 0.015$), y están completamente conectadas tienen un diámetro mínimo (igual a 2). Sin embargo, el diámetro de las redes desconectadas depende del muestreo. Especialmente si existe un tamaño de muestreo n^* de estas redes, tal que $0 < n^* < 5000$, que verifica la ecuación (4).

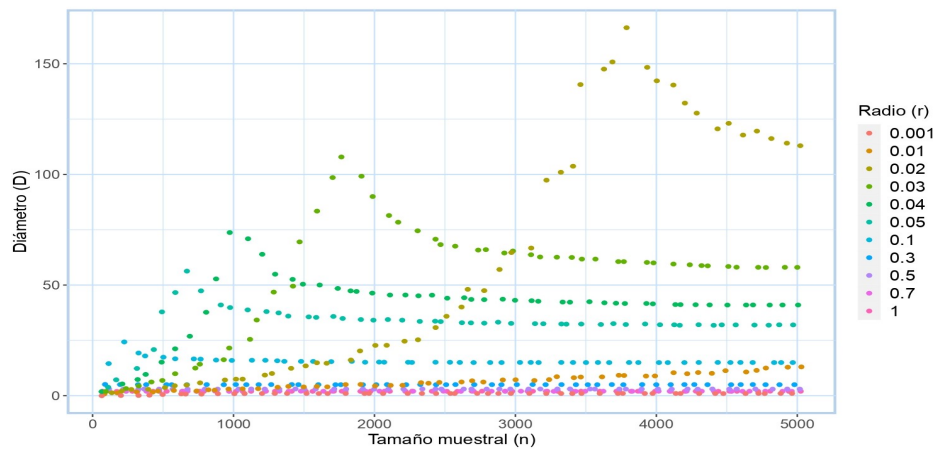


Figura 16: Diámetro del muestreo de una red geométrica aleatoria (RGG) en función del tamaño muestreado (n) y el radio de vecindad (r)

Densidad de conexiones. La densidad de conexiones se mantiene constante para todos los subgrafos de la red geométrica $G(n, r)$, como vemos en la representación de la figura. [17](#). Esta figura ratifica el comportamiento del diámetro en redes geométricas aleatorias definido en la sección [2.1](#).

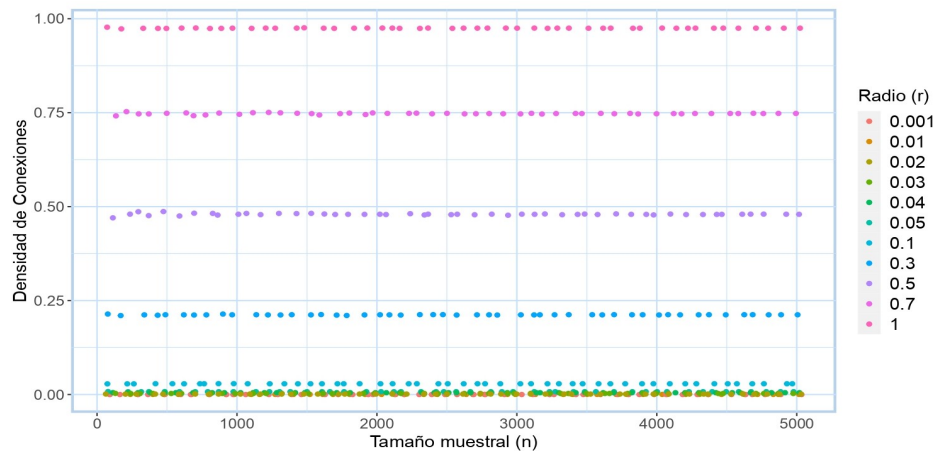


Figura 17: Densidad de conexiones de una red geométrica aleatoria (RGG) en función del tamaño muestreado (n) y el radio de vecindad (r)

3.2. Redes reales

Una vez comprobada la dependencia del tamaño de las principales propiedades topológicas en redes aleatorias Erdős-Renyi (sección [3.1.1](#)) y redes espaciales geométricas (sección [3.1.2](#)), en esta sección se repite el mismo análisis enfocado en las redes reales elaboradas a partir de las bases de datos (sección [1.3](#)).

De cada red, se han tomado muestras aleatorias de 100, 500, 1000, 1500 y 2000 nodos y se han fijado distancias umbrales (θ) con valores crecientes desde 5km hasta 300km. Los resultados son el promedio de 10 muestreos aleatorios. Las funciones empleadas para el cálculo están contenidas en *funciones_redes_reales.R*.

Tamaño de la componente conexa mayor (LCC). En la representación de la figura [18](#) distinguimos los siguientes comportamientos de la componente gigante:

- alcornoques, olivos y chumberas: todas ellas, dejando de lado la dispersión, tienen regiones con elevada densidad de nodos, como vemos en las figuras [1](#), [2](#) y [3](#). En estas regiones es donde se forma la componente gigante. El crecimiento de esta componente no es lineal, sino por adhesión de clusters. Primero en $\theta \approx [50, 80]$ y después en $\theta \approx 240km$, manteniéndose constante entre saltos.
- viñedos: la distribución es muy regular, por eso que el tamaño de la componente gigante crece casi linealmente con el radio.

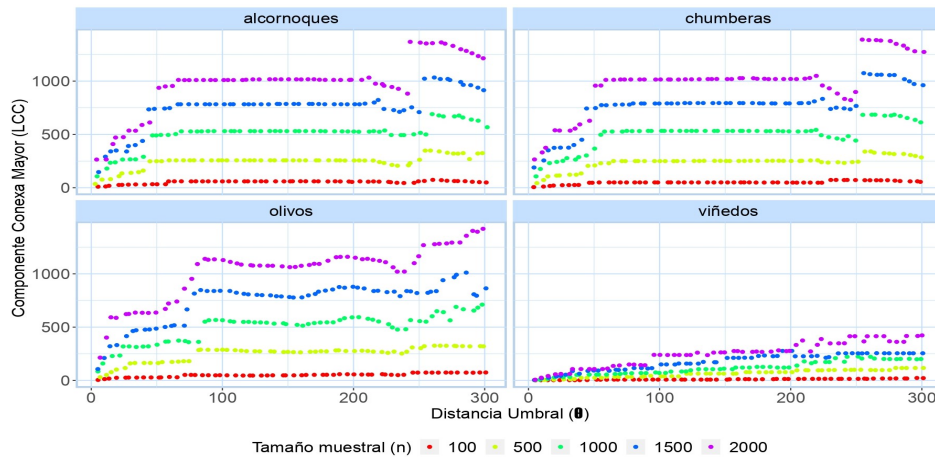


Figura 18: En vertical, el tamaño $C_{G(n^*, \theta)}$ de LCC. En horizontal, la distancia umbral (θ). Los valores se agrupan por tamaño muestral (n^*) y se clasifican por especie.

En la representación de la figura [19](#) se observa la dependencia lineal de la componente gigante

con el muestreo cuando se deja como parámetro la distancia umbral. Esto permitiría extrapolar los valores de redes completas.

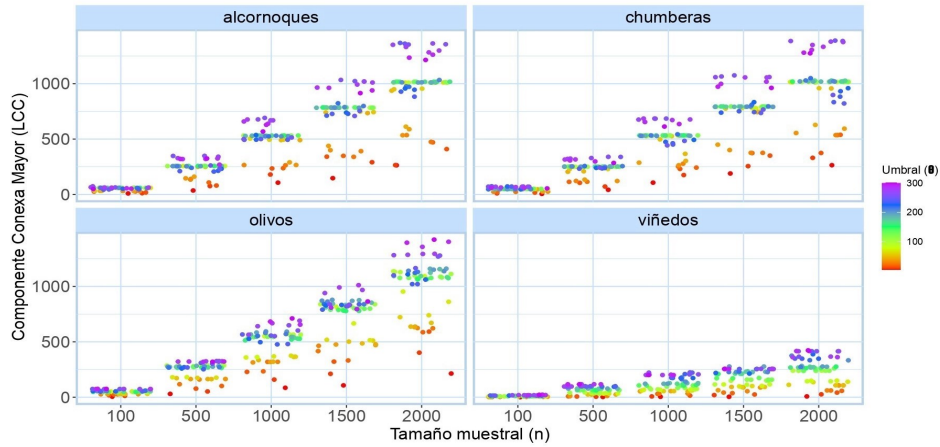


Figura 19: En vertical, el tamaño $C_{G(n^*, \theta)}$ de LCC. En horizontal, el tamaño muestral (n^*). Los valores se agrupan por distancias umbrales (θ) y se clasifican por especie.

Modularidad. En la figura 20 se distinguen dos comportamientos:

- alcornoques y olivos (menos extendidos): Q decae con θ . Al aumentar la distancia umbral, crece la conectividad entre módulos.
- chumberas y viñedos (más extendidos): Q no decae tan rápido con el aumento de θ . Una menor proximidad media favorece las conexiones intramodulares frente a las intermodulares, manteniéndose Q por encima de 0.5 en ambos casos.

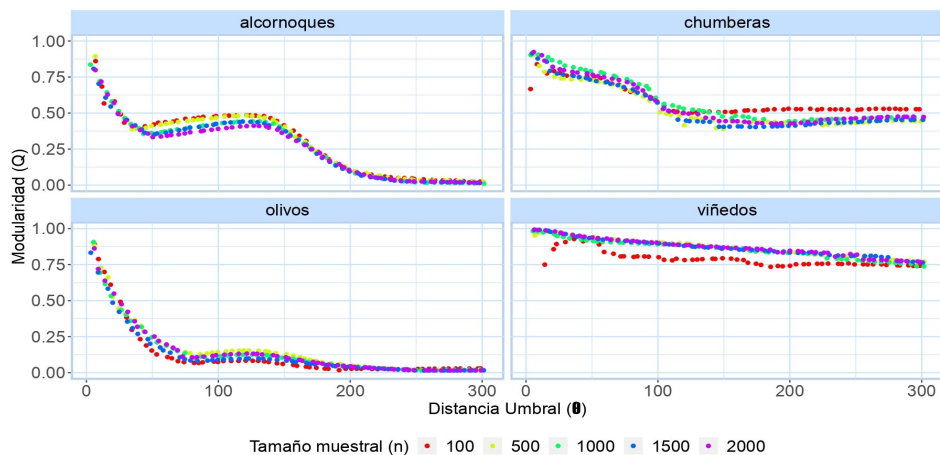


Figura 20: En vertical, la modularidad $Q(G(n^*, \theta))$. En horizontal, la distancia umbral (θ). Los valores se agrupan por tamaño muestral (n^*) y se clasifican por especie.

También podemos observar en la figura 20 que el tamaño muestral no tiene, en principio,

efecto sobre la modularidad. En la figura [21](#) se constata esta condición.

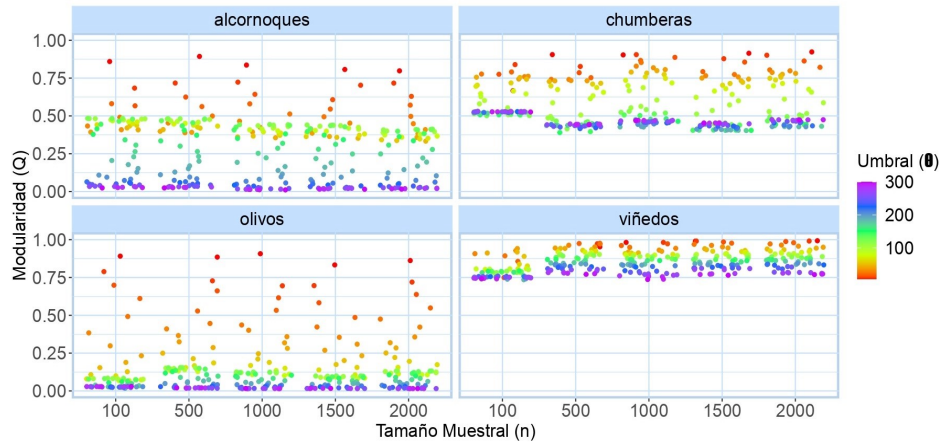


Figura 21: En vertical, la modularidad $Q(G(n^*, \theta))$. En horizontal, el tamaño muestral (n^*). Los valores se agrupan por distancias umbrales (θ) y se clasifican por especie.

Coefficiente de Clustering. Su comportamiento, representado en la figura [22](#), es:

- chumberas y viñedos: estas redes mantienen una conectividad constante para cualquier distancia (θ). Por ello que el coeficiente se mantiene prácticamente constante. Además, como se aprecia en la figura [23](#) vemos que no hay efecto del muestreo sobre esta medida.
- alcornoques y olivos: crece con el umbral de distancias (θ) porque aparecen nuevas conexiones, y se mantiene constante con el tamaño muestral (ver figura [23](#)).

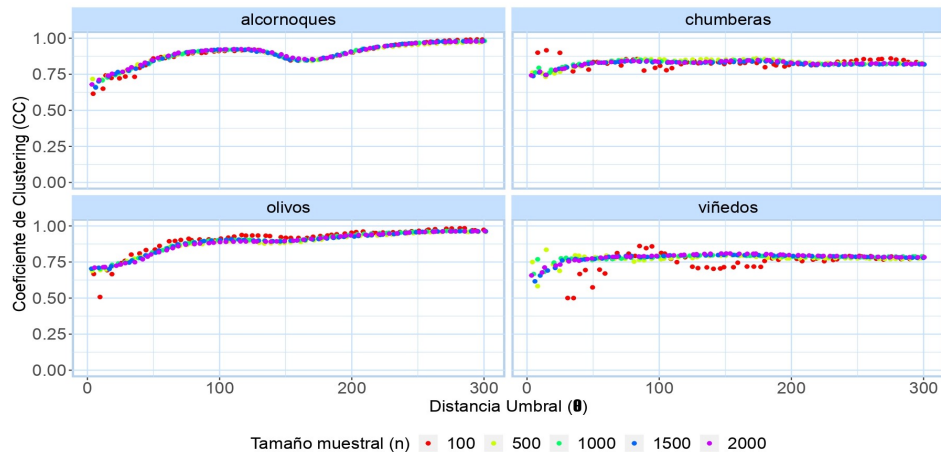


Figura 22: En vertical, el coeficiente de clustering $CC(G(n^*, \theta))$. En horizontal, la distancia umbral (θ). Agrupados por tamaño muestral (n^*) y se clasifican por especie.

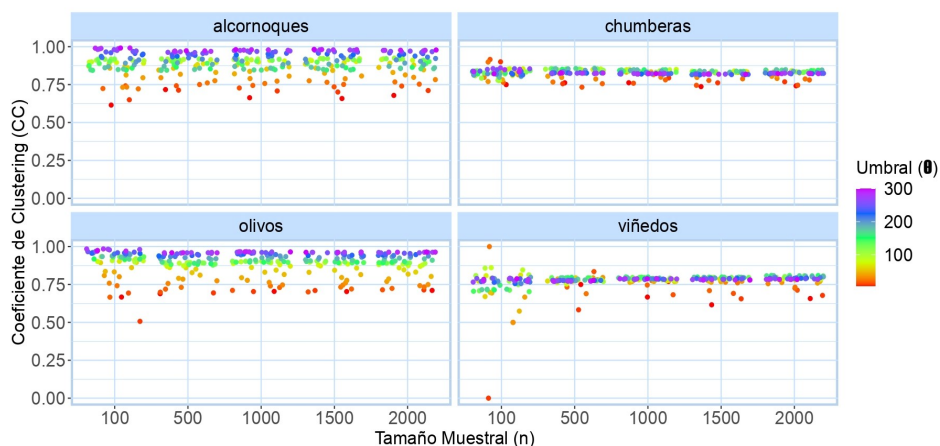


Figura 23: En vertical, el coeficiente de clustering $CC(G(n^*, \theta))$. En horizontal, el tamaño muestral (n^*). Agrupados por distancias umbrales (θ) y se clasifican por especie.

Diámetro de la red. En la figura 24 se representa el diámetro de los subgrafos, de la que se han extraído las siguientes conclusiones:

- chumberas y viñedos: el diámetro aumenta hasta un umbral crítico a partir del cual decae. Este punto crítico, que podría corresponder con una conectividad crítica, parece depender del tamaño muestral.
- alcornoques y olivos: para redes completamente conectadas, el diámetro es mínimo (vale 4 en alcornoques). En olivos, cuanto mayor sea el umbral (θ), menor es el diámetro.

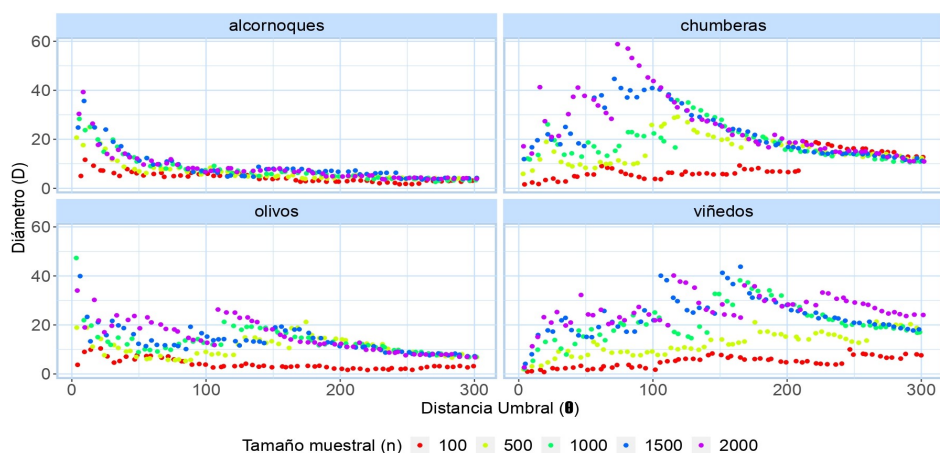


Figura 24: En vertical, el diámetro $Diam(G(n^*, \theta))$. En horizontal, la distancia umbral (θ). Los valores se agrupan por tamaño muestral (n^*) y se clasifican por especie.

En redes completamente conectadas, el diámetro no se ve afectado por el muestreo (ver alcornoques en la figura 25). Sin embargo, en redes con menor conectividad, un mayor tamaño muestreado tiene, en consecuencia, un mayor diámetro de red. En un principio, se diría que no es posible la determinación del diámetro de la red completa a partir del muestreo.

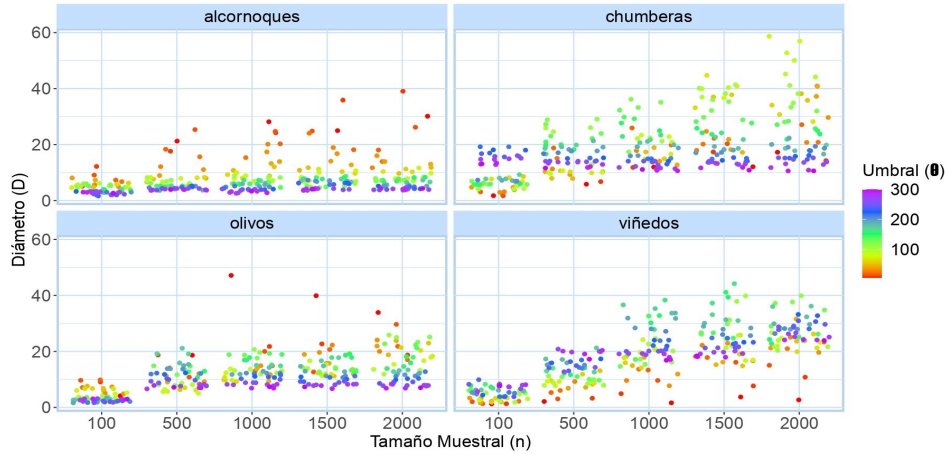


Figura 25: En vertical, el diámetro $\text{Diam}(G(n^*, \theta))$. En horizontal, el tamaño muestral (n^*). Los valores se agrupan por distancias umbrales (θ) y se clasifican por especie.

Densidad de conexiones. Los resultados obtenidos en la figura 26 concuerdan con lo resultados anteriores. La conectividad en alcornoques y olivos aumenta con el threshold (θ), mientras que en chumberas y olivos se mantiene constante.

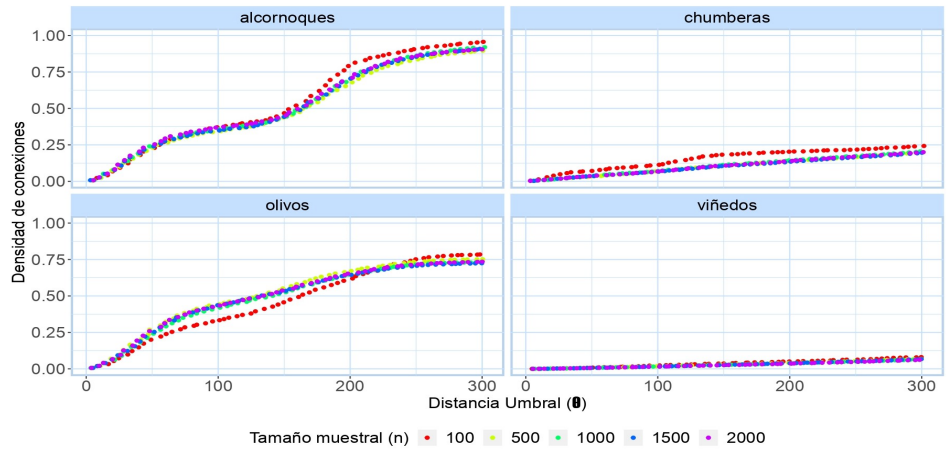


Figura 26: En vertical, la densidad de conexiones $\text{EdgeD}(G(n^*, \theta))$. En horizontal, la distancia umbral (θ). Agrupados por tamaño muestral (n^*) y se clasifican por especie.

La figura 27 muestra la independencia del diámetro y el muestreo. La densidad de conexiones toma valores similares en en cada muestreo y para todas las redes complejas.

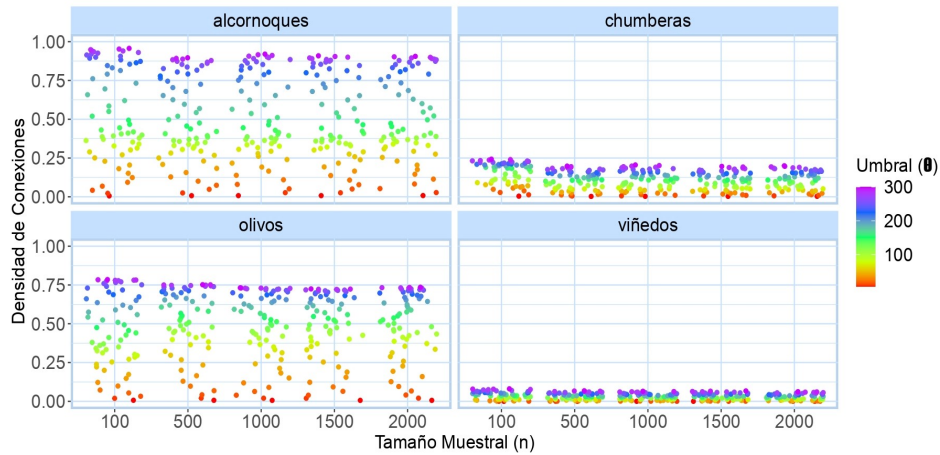


Figura 27: En vertical, la densidad de conexiones $\text{EdgeD}(G(n^*, \theta))$. En horizontal, el tamaño muestral (n^*). Agrupados por distancias umbrales (θ) y se clasifican por especie.

El comportamiento de estas propiedades para los muestreos realizados nos permiten concretar los siguientes resultados para el tamaño completo de estas redes reales de plantas.

Asumiendo que los individuos de una red se agrupan por regiones y que, para tamaños mayores de muestreo, no aparecen nuevas localizaciones sino más individuos en una región ya poblada de la red, estas medidas se comportarían de la siguiente forma para redes completas:

1. La componente gigante de alcornoques, olivos y chumberas (redes con alguna aglomeración) no incluye a todos los nodos de la red. En viñedos, es similar pero de menor tamaño, al tener menor aglomeración.
2. En la modularidad no se aprecia ningún efecto del muestreo. Esperamos unos valores similares de modularidad para cada red completa.
3. El coeficiente de clustering, al igual que las redes sintéticas, permanece invariante con respecto al tamaño muestral. Los valores de las redes completas serán los mismos.
4. El diámetro de la red en redes completamente conectadas no crece con el tamaño del muestreo (alcornoques). Sin embargo, en redes menos conectadas, crece linealmente con el tamaño muestral. Además, para radios pequeños, este crecimiento es aún mayor (ver figura [25](#) para chumberas). El diámetro esperado en redes completas será mayor.
5. La conectividad de la red es independiente del muestreo. Esperamos los mismos porcentajes en redes completas.

Para mayor detalle, los valores concretos de cada una de estas propiedades para cada red vegetal están en los cuadros [3](#), [4](#), [5](#) y [6](#).

CAPÍTULO 4. Discusión

En este trabajo se plantea el estudio de redes de especies vegetales generadas a partir de las bases de datos de geolocalización disponibles en los organismos oficiales europeos. El primer objetivo consistió en elaborar, a partir de esta información para las viñas, alcornoques, olivos y chumberas, los correspondientes grafos de interacción en términos de la escala espacial. Como el tamaño de las redes impide un análisis computacional completo, para determinar las principales propiedades estructurales de la topología de estas redes se realizaron muestreos de distintos tamaños que permitiesen establecer leyes de escala.

La limitación computacional a nuestro alcance ha requerido la realización de muestreos y, en consecuencia, el estudio de cómo éstos pueden servir para inferir propiedades estructurales de las redes reales.

En este sentido, con la intención de conocer el efecto del muestreo, se han analizado redes sintéticas de las que se conoce bien su comportamiento con el tamaño. Estas redes muestran un efecto del muestreo y un comportamiento similar en todas las propiedades. Hay que resaltar que algunas de las redes reales muestran indicios de un comportamiento análogo al de las redes sintéticas.

La conectividad de la red es una medida de gran importancia para inferir la topología de las redes complejas [22]. En las figuras 12 y 17 vemos que esta propiedad es invariante con respecto al tamaño del subgrafo, y está más conectada cuanto más alta sea la probabilidad de conexión en aleatorias o la distancia de vecindad en geométricas.

En redes reales, la densidad de conexiones es independiente del tamaño muestral al igual que en las redes de referencia, como vemos en la figura 27. Se distinguen dos grupos: primero, alcornoques y olivos con una densidad de conexiones entre 75 – 100 % para cualquier umbral; por otro, chumberas y los viñedos con apenas el 25 % de la conectividad máxima en todos los umbrales. Es decir, la red de estos dos sustratos estaría desconectada para la mayoría de umbrales estudiados.

La figuras 6 y 13 de redes aleatorias y geométricas aleatorias, respectivamente, muestran su comportamiento con el tamaño muestral. En subgrafos con tamaños por encima del punto

de conectividad crítico - ver ecuaciones (3) y (4) - la componente gigante incluye a todos los nodos. No así para tamaños inferiores.

En el caso de las redes reales, el comportamiento no se rige por la existencia de un punto crítico. En alcornoques, olivos y chumberas la LCC no está incluye a todos los nodos (las redes nunca quedan completamente conectadas). Aun así, el crecimiento es lineal con el tamaño del muestreo. A partir de la figura 19 se puede deducir que:

En alcornoques, para una muestra de 2000 nodos (0.46 % del total), quedan conectados en una componente gigante 1000 nodos cuando $\theta = 200km$. En la red completa esperaríamos valores similares, es decir, que la mitad de los nodos queden completamente conectados para ese umbral. Lo mismo ocurre para el resto de umbrales y en el caso de chumberas y olivos (2000 nodos representan el 35.64 % y 1.17 %, respectivamente). En estos casos, podríamos esperar que al menos el 50 % de los nodos formen la componente gigante de las redes completas para cualquier umbral (θ). En viñedos, una muestra de 2000 nodos (9.44 % del total) para $\theta = 200km$ muestra una componente gigante de menos de 500 nodos. Esta relación es lineal con el tamaño muestral y para los diferentes umbrales estudiados. En una red completa podríamos esperar una componente gigante con el 25 % de los nodos totales. Para cada valor de distancia umbral ver los cuadros 3, 4, 5 y 6.

La modularidad de redes sintéticas, según las figuras 7 y 14, se acerca al máximo antes de superar el punto crítico y, posteriormente, desciende con el tamaño hasta valores casi nulos. En redes reales y, en contraste con las sintéticas, la modularidad de la red no depende del muestreo, como se observa en la figura 21. En una red completa, para las distancias umbrales estudiadas, esperamos una modularidad nula en alcornoques y olivos cuando $\theta = 3000km$. Por otro lado, en chumberas y viñedos, la modularidad es elevada para cualquier umbral en las redes completas. Es decir, la escala de valores de modularidad en función de la distancia se conserva con el tamaño.

El coeficiente de clustering es una invariante en ambas redes sintéticas. Como hemos podido comprobar en la literatura 23, 15, este coeficiente es independiente del tamaño muestral en redes sintéticas. Esto se observa figuras 9 y 15

En el caso de redes reales, el comportamiento es similar. Las redes con mayor conectividad

tienen un coeficiente de clustering mayor. Lo interesante de esta propiedad, sin embargo, es que es independiente del tamaño muestral. Esto se cumplía en redes sintéticas pero se verifica también en redes reales, como queda representado en la figura [23](#). Existe una propiedad estructural invariante al tamaño de la estructura. Las redes completas de las diferentes redes vegetales presentarán los mismos valores de coeficiente de clustering.

Sabemos, gracias a [17](#), [18](#) que, mientras la red no esté completamente conectada, el diámetro crece. Una vez queda completamente conectada, el diámetro disminuye a medida que aparecen conexiones entre nodos de la componente gigante.

En redes reales el comportamiento con respecto al muestreo difiere en cada red. En alcornoques, la red completamente conectada, el diámetro no muestra una dependencia con el tamaño. Los valores son similares para los diferentes muestreos, especialmente en umbrales (θ) altos. En olivos, una red menos conectada, el diámetro crece ligeramente con el tamaño muestral. En chumberas, esta dependencia lineal se hace más evidente, especialmente para umbrales (θ) menores. Viñedos, la red más desconectada, muestra una dependencia lineal clara. Da la sensación de que, cuanto menos conectada esté una red, mayor es la dependencia del diámetro con el muestreo.

En definitiva, esperaríamos que la red completa de alcornoques no presente unos valores de diámetro para cada umbral (θ) muy alejados de los que se han obtenido. Sin embargo, los valores serán gradualmente más grandes para el resto de redes, exhibiendo la red de viñedos y de chumberas diámetros mucho mayores que los obtenidos en el muestreo. En el caso de olivos, el efecto del muestreo no será tan acentuado, lo cual no quiere decir que no sea mucho mayor. Hay que tener en cuenta que estos muestreos suponen el 0.46 % de la red completa en alcornoques, 1.17 % en olivos, 35.64 % en chumberas y 9.44 % en viñedos.

El efecto del muestreo sobre cada propiedad estructural se podría resumir en el cuadro [1](#):

Cuadro 1: ¿Afecta el muestreo de nodos a las propiedades estructurales de las redes reales estudiadas?

		Propiedades					
		LCC	Q	CC	Diam	EdgeD	
Redes	aleatorias	Sí	Sí	No	Sí	No	
	geométricas	Sí	Sí	No	Sí	No	
	Reales	alcornoques	Sí	No	No	No*	No
		olivos	Sí	No	No	Sí	No
		chumberas	Sí	No	No	Sí	No
viñedos		Sí	No	No	Sí	No	

*El diámetro en alcornoques no muestra dependencia con el muestreo para umbrales (θ) grandes.

Un ejemplo de propiedades estructurales para $\theta = 200km$ en el cuadro [2](#):

Cuadro 2: Determinación de cada propiedad en las redes reales

		Propiedades				
		LCC	Q	CC	Diam	EdgeD
Redes Reales	alcornoques	52.34%	0.103	0.903	5	0.719
	olivos	57.51%	0.041	0.936	-	0.647
	chumberas	50.967%	0.45	0.827	-	0.152
	viñedos	14.823%	0.818	0.788	-	0.041

Los valores que presentarán las redes completas para cada medida. En el caso del diámetro, debido al efecto del muestreo, no es posible determinar su valor.

En los cuadros [3](#), [4](#), [5](#) y [6](#) aparecen los resultados de estas propiedades para distancias umbrales comprendidas entre $5km$ y $300km$.

Para finalizar esta sección comentar que, una vez analizada la estructura de la red y superadas las limitaciones técnicas, el siguiente paso sería utilizar la red como espacio topológico sobre el que se define una dinámica, por ejemplo, una epidemia. Lo primero sería encontrar la escala espacial que percibe un patógeno de esta red. Esta escala dependerá de sus capacidades infecciosas, entre ellas, la distancia de salto. Tendremos, por tanto, que la distancia umbral (θ) determina esta escala. Las dinámicas estudiadas en la bibliografía [\[24\]](#), [\[25\]](#) muestran una capacidad dispersiva muy limitada. Para estos rangos de valores θ , en nuestras redes de estudio, el patógeno percibiría una red modular, con un diámetro elevado, y prácticamente desconectada en todos los sustratos.

CAPÍTULO 5. Conclusiones

1. Se han elaborado redes de plantas a partir de coordenadas de geolocalización de los individuos de una misma especie.
2. Se han encontrado limitaciones en la obtención de muestras representativas de redes vegetales con poblaciones que, por su tamaño, son inabarcables con los ordenadores disponibles.
3. Como referencia para el estudio de estas redes reales, se ha analizado el efecto del muestreo en redes sintéticas.
 - Se ha demostrado la independencia del coeficiente de clustering y la densidad de conexiones con el tamaño del muestreo.
 - En redes aleatorias se ha definido una ley de escala para la modularidad y el diámetro con el tamaño del muestreo. Para la componente gigante, se ha definido una tendencia con respecto al tamaño.
 - Para definir leyes de escala en redes geométricas aleatorias se requiere trabajar con mayores tamaños de red. No hemos podido definir leyes de escala para la modularidad y el diámetro en redes geométricas aleatorias. Para la componente gigante, se ha encontrado una relación al igual que en aleatorias.
4. Se han estudiado medidas estructurales de muestreos de las redes vegetales, probando cómo de representativo es el muestreo en estas redes reales.
 - En redes reales, al igual que en sintéticas, existen propiedades invariantes con respecto al muestreo: coeficiente de clustering y densidad de conexiones.
 - En estas redes se han encontrado, además, propiedades invariantes que, en redes sintéticas, sí dependen del muestreo: modularidad.
 - Se han observado propiedades que dependen del muestreo y de las que se podría definir una ley de escala para extrapolar su comportamiento para redes completas: diámetro de la red y tamaño de la componente gigante.
 - El comportamiento de las propiedades de las redes reales, con la excepción del diámetro, no depende de una conectividad crítica, a diferencia de las redes sintéticas, que sí dependen de los valores críticos que determinan la transición de percolación.

CAPÍTULO 6. Bibliografía

- [1] Dan Frank, Zhiheng Huang y Alvin Chyan. “Sampling A Large Network: How Small Can My Sample Be?” En: (dic. de 2012). <https://doi.org/10.1088/0305-4470/38/1/007>.
- [2] T. Petermann y P. De Los Rios. “Exploration of scale-free networks”. En: *The European Physical Journal B - Condensed Matter and Complex Systems* 38.2 (2004), págs. 201-204. DOI: [10.1140/epjb/e2004-00021-5](https://doi.org/10.1140/epjb/e2004-00021-5).
- [3] Tianyi Wang y col. “Understanding Graph Sampling Algorithms for Social Network Analysis”. En: *2011 31st International Conference on Distributed Computing Systems Workshops*. 2011, págs. 123-128. DOI: [10.1109/ICDCSW.2011.34](https://doi.org/10.1109/ICDCSW.2011.34).
- [4] Paul Erdős y Alfréd Rényi. “On random graphs I”. En: *Publ. Math. Debrecen* 6 (dic. de 1959), págs. 290-297.
- [5] Mathew Penrose. *Random Geometric Graphs*. 2003. DOI: [10.1093/acprof:oso/9780198506263.001.0001](https://doi.org/10.1093/acprof:oso/9780198506263.001.0001).
- [6] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA, 2020. URL: <http://www.rstudio.com/>.
- [7] V Sood, S Redner y D ben-Avraham. “First-passage properties of the Erdős-Rényi random graph”. En: *Journal of Physics A: Mathematical and General* 38.1 (dic. de 2004), págs. 109-123. DOI: [10.1088/0305-4470/38/1/007](https://doi.org/10.1088/0305-4470/38/1/007). URL: <https://doi.org/10.1088/0305-4470/38/1/007>.
- [8] Ghurumuruhan Ganesan. “Size of the giant component in a random geometric graph”. En: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 49.4 (2013). DOI: [10.1214/12-aihp498](https://doi.org/10.1214/12-aihp498).
- [9] Aaron Clauset, M. E. J. Newman y Christopher Moore. “Finding community structure in very large networks”. En: *Physical Review E* 70.6 (2004). DOI: [10.1103/physreve.70.066111](https://doi.org/10.1103/physreve.70.066111).
- [10] *A generalized louvain method for community detection implemented in MATLAB*. GitHub. URL: <https://github.com/GenLouvain/GenLouvain>.
- [11] Colin Mcdiarmid y Fiona Skerman. “Modularity of Erdős-Rényi random graphs”. En: *Random Structures Algorithms* 57.1 (2020), págs. 211-243. DOI: [10.1002/rsa.20910](https://doi.org/10.1002/rsa.20910).
- [12] Ernesto Estrada. *Random models of networks*. 2011, págs. 232-256. DOI: [10.1093/acprof:oso/9780199591756.003.0012](https://doi.org/10.1093/acprof:oso/9780199591756.003.0012).
- [13] Sara Nadiv Soffer y Alexei Vázquez. “Network clustering coefficient without degree-correlation biases”. En: *Physical Review E* 71.5 (2005). DOI: [10.1103/physreve.71.057101](https://doi.org/10.1103/physreve.71.057101).
- [14] Sang Hoon Lee, Pan-Jun Kim y Hawoong Jeong. “Statistical properties of sampled networks”. En: *Physical Review E* 73.1

- (2006). DOI: [10 . 1103 / physreve . 73 . 016102](https://doi.org/10.1103/physreve.73.016102).
- [15] Chen Avin. “Distance graphs: from random geometric graphs to Bernoulli graphs and between”. En: *Proceedings of the fifth international workshop on Foundations of mobile computing - DIAL M-POMC 08* (2008). DOI: [10.1145/1400863.1400878](https://doi.org/10.1145/1400863.1400878).
- [16] Eric Weisstein. “Making MathWorld”. En: *The Mathematica Journal* 10.3 (2007). DOI: [10.3888/tmj.10.3-3](https://doi.org/10.3888/tmj.10.3-3).
- [17] Fan Chung y Linyuan Lu. “The Diameter of Sparse Random Graphs”. En: *Advances in Applied Mathematics* 26.4 (2001), págs. 257-279. DOI: [10.1006/aama.2001.0720](https://doi.org/10.1006/aama.2001.0720).
- [18] Rick Durrett. “Random Graph Dynamics”. En: (2006), págs. 47-50. DOI: [10 . 1017 / cbo9780511546594](https://doi.org/10.1017/cbo9780511546594).
- [19] Robert B. Ellis, Jeremy L. Martin y Catherine Yan. “Random Geometric Graph Diameter in the Unit Ball”. En: *Algorithmica* 47.4 (2007), págs. 421-438. DOI: [10 . 1007 / s00453-006-0172-y](https://doi.org/10.1007/s00453-006-0172-y).
- [20] Adam Sealfon y Jonathan Ullman. “Efficiently Estimating Erdos-Renyi Graphs with Node Differential Privacy”. En: *Journal of Privacy and Confidentiality* 11.1 (2021). DOI: [10.29012/jpc.745](https://doi.org/10.29012/jpc.745).
- [21] Gabor Csardi y Tamas Nepusz. *The Igraph Software Package for Complex Network Research*. Nov. de 2005, pág. 1695.
- [22] Benjamin Shargel y col. “Optimization of Robustness and Connectivity in Complex Networks”. En: *Physical Review Letters* 90.6 (2003). DOI: [10.1103/physrevlett.90.068701](https://doi.org/10.1103/physrevlett.90.068701).
- [23] Tore Opsahl. “Triadic closure in two-mode networks: Redefining the global and local clustering coefficients”. En: *Social Networks* 35.2 (2013), págs. 159-167. DOI: [10.1016/j.socnet.2011.07.001](https://doi.org/10.1016/j.socnet.2011.07.001).
- [24] I. L. Boyd y col. “The Consequence of Tree Pests and Diseases for Ecosystem Services”. En: *Science* 342.6160 (2013), págs. 1235773-1235773. DOI: [10 . 1126 / science.1235773](https://doi.org/10.1126/science.1235773).
- [25] Kevin Schneider y col. “Impact of *Xylolla fastidiosa* subspecies pauca in European olives”. En: *Proceedings of the National Academy of Sciences* 117.17 (2020), págs. 9250-9259. DOI: [10 . 1073 / pnas . 1912206117](https://doi.org/10.1073/pnas.1912206117).

Anexo I. Repositorio GitHub

En la dirección:

<https://github.com/LeonardoClaudin/Trabajo-Fin-de-Grado-Biotecnologia-ETSIAB>

se encuentran todos los programas empleados en el estudio. Consta de 3 carpetas:

Complex Networks. Aquí se encuentran las funciones y herramientas de graficación que se han desarrollado para estimar las propiedades del muestreo de redes complejas basadas en ocurrencias de especies vegetales. Los archivos que contiene son:

1. **Funciones_redes_reales.R:**

- a) Componente gigante $LCC_real_network(doc.csv, nrep, sample_size, thres_vector)$
- b) Modularidad $Mod_real_network(doc.csv, nrep, sample_size, thres_vector)$
- c) Coeficiente Clustering $CCoef_real_network(doc.csv, nrep, sample_size, thres_vector)$
- d) Diámetro $Diameter_real_network(doc.csv, nrep, sample_size, thres_vector)$
- e) Densidad de Conexión $EdgeD_real_network(doc.csv, nrep, sample_size, thres_vector)$

2. **Resultados_redes_reales.rmd:** graficación de los resultados de redes complejas.

Synthetic Networks. A su vez dividio en dos carpetas, que corresponden con los dos tipos de redes modelo con las que se ha realizado la comparación de los resultados.

1. **Erdos-Renyi:** funciones y graficación desarrolladas para obtener los resultados de redes aleatorias $G(n, p)$:

a) **Funciones_redes_aleatorias.R:**

- 1) Componente gigante $LCC_erdos_renyi(size, prob, prom)$
- 2) Modularidad $Mod_erdos_renyi(size, prob, prom)$
- 3) Coeficiente Clustering $CCoef_erdos_renyi(size, prob, prom)$
- 4) Diámetro $Diameter_erdos_renyi(size, prob, prom)$
- 5) Densidad de Conexión $EdgeD_erdos_renyi(size, prob, prom)$

b) **Resultados_redes_aleatorias.rmd:** graficación de los resultados de redes Erdős-Renyi

2. **Random Geometric Graph:** funciones y graficación desarrolladas para obtener los resultados de redes geométricas aleatorias $G(n, r)$:

a) **Funciones_redes_geométricas.R:**

- 1) Componente gigante $LCC_grg(size, rad, prom)$
 - 2) Modularidad $Mod_grg(size, rad, prom)$
 - 3) Coeficiente Clustering $CCoef_grg(size, rad, prom)$
 - 4) Diámetro $Diameter_grg(size, rad, prom)$
 - 5) Densidad de Conexión $EdgeD_grg(size, rad, prom)$
- b) **Resultados_redes_geométricas.rmd**: graficación de los resultados de redes aleatorias de distancia.

Other Programmes. Están incluidos los programas para generar las redes y la disposición espacial de los individuos de las especies vegetales representados en el apartado **1.3** de Materiales y Métodos.

Además, incluye un programa interactivo realizado con Shiny Web App <https://shiny.rstudio.com/> para estudiar la disposición espacial de las redes vegetales y sus propiedades a medida que se modifican el tamaño de muestreo y la distancia umbral (θ).

1. **Ejemplo_redes_geométricas_aleatorias_reales.rmd**
2. **Mapa_sustratos.rmd**
3. **App_sustratos.R**

Anexo II. Cuadros de resultados de redes completas.

Cuadro 3: Propiedades de los Alcornoques

Umbral (θ)	Alcornoques				
	LCC (%)	Q	CC	EdgeD	Diameter
5	10.18	0.839	0.666	0.007	21.8
25	21.963	0.499	0.754	0.105	15.2
50	41.017	0.375	0.847	0.239	9.4
75	43.38	0.41	0.898	0.316	9.2
100	43.44	0.437	0.917	0.358	6.6
125	43.5	0.452	0.918	0.388	6.2
150	53.55	0.387	0.865	0.449	6.2
175	53.497	0.225	0.853	0.574	5.4
200	52.34	0.103	0.903	0.719	5
225	50.443	0.05	0.944	0.81	4.6
250	55.843	0.03	0.965	0.864	3.6
275	65.597	0.024	0.976	0.896	3.6
300	58.7	0.018	0.981	0.917	3.6

Cuadro 4: Propiedades de las Chumberas

Umbral (θ)	Chumberas			
	LCC (%)	Q	CC	EdgeD
5	10.243	0.862	0.751	0.003
25	24.813	0.79	0.808	0.024
50	44.63	0.754	0.83	0.043
75	51.21	0.694	0.847	0.061
100	51.343	0.57	0.834	0.077
125	51.457	0.486	0.837	0.102
150	51.477	0.444	0.838	0.123
175	51.36	0.449	0.845	0.137
200	50.967	0.45	0.827	0.152
225	49.557	0.456	0.82	0.166
250	52.227	0.463	0.827	0.178
275	68.22	0.474	0.83	0.193
300	60.227	0.473	0.823	0.208

Cuadro 5: Propiedades de los Olivos

Umbral (θ)	Olivos			
	LCC (%)	Q	CC	EdgeD
5	7.467	0.876	0.695	0.006
25	28.027	0.454	0.736	0.09
50	31.47	0.22	0.816	0.244
75	43.23	0.112	0.882	0.352
100	55.123	0.109	0.899	0.413
125	52.537	0.119	0.904	0.46
150	51.937	0.097	0.897	0.521
175	54.457	0.062	0.915	0.59
200	57.51	0.041	0.936	0.647
225	54.23	0.028	0.945	0.69
250	60.623	0.02	0.958	0.725
275	64.877	0.019	0.967	0.74
300	67.76	0.02	0.965	0.746

Cuadro 6: Propiedades de los Viñedos

Umbral (θ)	Viñedos			
	LCC (%)	Q	CC	EdgeD
5	0.523	0.98	0.673	0
25	2.81	0.951	0.697	0.003
50	5.81	0.929	0.733	0.007
75	7.12	0.892	0.788	0.012
100	8.98	0.88	0.79	0.018
125	10.537	0.863	0.778	0.023
150	11.823	0.852	0.783	0.029
175	13.197	0.831	0.78	0.035
200	14.823	0.818	0.788	0.041
225	16.98	0.811	0.78	0.047
250	19.017	0.794	0.785	0.053
275	18.57	0.776	0.782	0.061
300	20.73	0.757	0.783	0.068