

UACM

Universidad Autónoma
de la Ciudad de México

NADA HUMANO ME ES AJENO

COLEGIO DE CIENCIAS Y HUMANIDADES

MAESTRÍA EN CIENCIAS DE LA COMPLEJIDAD

**Mapeos acoplados espacio-temporales
en redes $C\alpha$ de aminoácidos en proteínas globulares.
Sincronización y acoplamientos**

TESIS QUE PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS DE LA COMPLEJIDAD

P R E S E N T A

HUGO SÁNCHEZ LOZADA

D I R E C T O R

DR. LUIS AGUSTÍN OLIVARES QUIROZ

Ciudad de México, diciembre de 2024.

SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN



UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO COORDINACIÓN ACADÉMICA

RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

DERECHOS RESERVADOS[©]

La presente obra y cada uno de sus elementos está protegido por la Ley Federal del Derecho de Autor; por la Ley de la Universidad Autónoma de la Ciudad de México, así como lo dispuesto por el Estatuto General Orgánico de la Universidad Autónoma de la Ciudad de México; del mismo modo por lo establecido en el Acuerdo por el cual se aprueba la Norma mediante la que se Modifican, Adicionan y Derogan Diversas Disposiciones del Estatuto Orgánico de la Universidad de la Ciudad de México, aprobado por el Consejo de Gobierno el 29 de enero de 2002, con el objeto de definir las atribuciones de las diferentes unidades que forman la estructura de la Universidad Autónoma de la Ciudad de México como organismo público autónomo y lo establecido en el Reglamento de Titulación de la Universidad Autónoma de la Ciudad de México.

Por lo que el uso de su contenido, así como cada una de las partes que lo integran y que están bajo la tutela de la Ley Federal de Derecho de Autor, obliga a quien haga uso de la presente obra a considerar que solo lo realizará si es para fines educativos, académicos, de investigación o informativos y se compromete a citar esta fuente, así como a su autor ó autores. Por lo tanto, queda prohibida su reproducción total o parcial y cualquier uso diferente a los ya mencionados, los cuales serán reclamados por el titular de los derechos y sancionados conforme a la legislación aplicable.

La materia no sólo interactúa, también se organiza. Conocemos básicamente todas las leyes de interacción de la materia, pero no sabemos casi nada sobre sus leyes de organización.

Albert L. Lehninger (1917-1986)

Agradecimientos

Quiero expresar mi más profundo agradecimiento a todas las personas que hicieron posible la realización de esta tesis.

A mi director de tesis, el Dr. Luis Olivares Quiroz, por su invaluable guía, paciencia y apoyo constante en cada etapa de este trabajo. Su conocimiento y orientación fueron fundamentales para alcanzar los objetivos planteados.

A los miembros de mi comité sinodal, el Dr Fernando Ramírez Alatraste, la Dra Rosa Margarita Alvarez González y el Mtro Gabriel Eloy Aguilar Pineda, por sus valiosas sugerencias y críticas constructivas, que enriquecieron significativamente este trabajo.

A mi amiga y colega Macchia Felix C, por su apoyo incondicional y aliento constante, quienes fueron fundamentales durante este proceso.

A todos, mi más sincero agradecimiento.

Índice general

0.1	Abstract	6
1	Fundamentos de Sistemas Complejos y Redes	7
1.1	Estudio de redes de aminoácidos en proteínas globulares	11
1.2	Metodología de análisis dinámico de la estructura de las PRN	12
2	Teoria de redes y modelos de mapeos acoplados	14
2.1	Estudio de las redes en el mundo real	14
2.2	Estructura y propiedades	16
2.3	Métricas de centralidad	17
2.4	Modelos de redes	26
2.5	Dinámica discreta y mapeos acoplados CML	30
3	Proteínas globulares y métodos computacionales basados en mapeos acoplados	32
3.1	Aminoácidos	33
3.2	Enlace peptídico	35
3.3	Estructuras	37
3.4	Proteínas globulares	38
3.5	Introducción a la Bioinformática y Métodos Computacionales	40
3.6	Construcción de redes y análisis de estructuras proteicas	42
4	Estructura y propiedades de redes de residuos de proteínas	45
4.1	Resultados de M^{pro} de SARS CoV-1	49
4.2	Resultados de M^{pro} de SARS CoV-2	52
4.3	Resultados promedio para M^{pro}	56
5	Propagación de Información en redes mediante Modelos de Mapeo Acoplado	63
5.1	Comportamiento espacio-temporal	66

5.2	Espectro de potencias	76
5.3	Interpretación de resultados y conclusiones	79

0.1. Abstract

In this study, a model of protein residue networks (PRNs) was developed using globular proteins, including the main protease of SARS CoV-2. The relevant topological properties of the PRNs were examined and compared to properties of other reference networks. Additionally, structurally important sites were identified using metrics of centrality and long-range communicability. The results confirmed differences between the PRNs of SARS CoV-1 and SARS CoV-2 proteases, indicating structural changes that enhance the efficiency of the SARS CoV-2 protease in long-range perturbation transmission. In the analysis, the coupled map lattice (DCML) mapping was used as a tool, and a correlation was found between high centrality sites and spatiotemporal synchronization. Furthermore, the impact of network topology differences on information flow in globular proteins was evaluated, using C_α carbon networks. In conclusion, this study provides a detailed understanding of how the structure of a protein influences its synchronization dynamics and long-range information flow in the network.

En este estudio, se desarrolló un modelo de redes de residuos de proteínas (PRNs) utilizando proteínas globulares, incluyendo la proteasa principal del SARS CoV-2. Se examinaron las propiedades topológicas relevantes de las PRNs y se compararon con las propiedades de otras redes de referencia. Además, se identificaron sitios estructuralmente importantes mediante métricas de centralidad y comunicabilidad de largo alcance. Los resultados corroboraron diferencias entre las PRNs de proteasas de SARS CoV-1 y SARS CoV-2, indicando cambios estructurales que mejoran la eficiencia de la proteasa de SARS CoV-2 en la transmisión de perturbaciones a larga distancia. En el análisis, se utilizó el mapeo acoplado dinámico (DCML) como herramienta y se encontró una correlación entre los sitios de alta centralidad de largo alcance y la sincronización espacio-temporal. Asimismo, se evaluó el impacto de las diferencias en la topología de las redes en el flujo de información en proteínas globulares, utilizando redes de carbonos C_α . En conclusión, este estudio proporciona una comprensión detallada de cómo la estructura de una proteína influye en su dinámica de sincronización y flujo de información a larga distancia en la red.

Capítulo 1

Fundamentos de Sistemas Complejos y Redes

El concepto de red se refiere a una estructura compuesta por elementos interconectados, de manera que se comunican o interactúan entre sí, matemáticamente cada elemento de la red es representado como un nodo que está vinculado a otros nodos a través de enlaces o conexiones que representan las relaciones o interacciones entre ellos. Las redes son parte de la amplia gama de sistemas que se componen de numerosos elementos dinámicos interconectados entre sí, los cuales suelen dar lugar a la emergencia de propiedades globales complejas. La conceptualización de sistemas complejos surge al considerar las propiedades emergentes y los comportamientos no triviales exhibidos globalmente por los elementos interconectados de un sistema. Estos sistemas compuestos y dinámicos se caracterizan por la interdependencia no lineal de sus partes, resiliencia, capacidad de evolucionar en respuesta a cambios internos y externos. Los fenómenos complejos aparecen y se estudian de manera no reduccionista e interdisciplinaria desde diversos campos como la física, la biología, la economía y las ciencias sociales. La dinámica de redes se estudia desde la perspectiva de los sistemas complejos debido a la capacidad, de este enfoque, para abordar la interconexión y la emergencia de propiedades globales de sistemas compuestos.

En el campo de los sistemas complejos, los nodos de una red representan unidades dinámicas y los enlaces simbolizan las interacciones entre ellas y se distingue por poseer propiedades estadísticas y topológicas no triviales que no se encuentran en las redes más simples. Las redes complejas representan sistemas que se encuentran en abundancia en la naturaleza y que se manifiestan en diversos niveles de escala de organización. Por ejemplo a nivel microscópico, podemos estudiar redes de regulación genética, redes neuronales que gobiernan la comunicación entre las células del cerebro, redes metabólicas que controlan

las reacciones químicas dentro de los organismos vivos y redes de proteínas que regulan las interacciones moleculares fundamentales en los sistemas biológicos.

A una escala macroscópica, encontramos una variedad de redes complejas, como las redes informáticas que conectan dispositivos y facilitan la transferencia de datos, las redes sociales que entrelazan a las personas en comunidades virtuales, las redes ecológicas que describen las interacciones entre especies en un ecosistema determinado, las redes de transporte que vinculan diferentes puntos geográficos para el movimiento eficiente de bienes y personas, y las redes de comercio que conectan empresas y facilitan el intercambio de productos y servicios.

Estos ejemplos ilustran la omnipresencia y la relevancia de las redes en múltiples dominios. El estudio de estas redes no solo nos permite comprender mejor la estructura y el funcionamiento de los sistemas complejos, sino que también nos proporciona herramientas para analizar su dinámica, predecir su comportamiento y diseñar estrategias óptimas para intervenir en ellos. En conclusión, las redes complejas son un campo fascinante y en constante evolución que nos brinda una perspectiva única para explorar y comprender la complejidad intrínseca de nuestro mundo.

Las redes brindan la oportunidad de investigar el comportamiento dinámico de sistemas extendidos cuyos elementos que interactúan a través de topologías complejas [1]. El análisis de las propiedades de las redes complejas tiene un impacto significativo en el estudio de fenómenos en equilibrio y no equilibrio, incluyendo el estudio de epidemias, un tema de suma relevancia en nuestro contexto actual [2]. Numerosas propiedades estadísticas y topológicas son comunes en la mayoría de los sistemas modelados mediante redes. Por ejemplo, se observa la presencia de distribuciones de grado que siguen leyes de potencia, donde algunos nodos tienen una conectividad excepcionalmente alta mientras que la mayoría de los nodos tienen una conectividad baja [3]. Además, se ha constatado la existencia de redes con características de “mundo pequeño” en los caminos que conectan los nodos, lo que significa que, en promedio, el número de pasos requeridos para viajar de un nodo a otro es sorprendentemente bajo en relación con el tamaño de la red [4]. También se ha descubierto que hay redes que exhiben una propiedad de autosimilaridad bajo transformaciones de escala, lo que implica que los patrones observados en una parte de la red se repiten en diferentes escalas [5]. Por último, se ha observado en algunas redes una tendencia de los nodos a conectarse con otros nodos que son similares o diferentes, lo que se conoce como patrones de mezcla [6].

Al investigar estas propiedades, los científicos pueden obtener una visión más profunda de cómo se propagan las epidemias, cómo se forman y extienden las redes sociales, cómo se distribuyen los recursos en las redes de transporte, entre otros fenómenos de interés. Además, el estudio de las redes complejas también contribuye al desarrollo de estrategias de intervención y control más efectivas en diversas áreas, como la salud pública, la ciberseguridad y la gestión de recursos. En resumen, las redes complejas representan una herramienta invaluable para analizar y comprender la complejidad presente en numerosos sistemas del mundo real.

Se ha aplicado la teoría de redes al estudio de la estructura de proteínas, obteniendo resultados prometedores [7]. En su artículo, Vendruscolo *et al.* [8] construyeron redes correspondientes a estructuras proteicas y demostraron por primera vez que estas redes exhiben características de “mundo pequeño”. En otro estudio, Atilgan *et al.* [9] representaron las proteínas como redes de residuos *Protein Residue Networks* (PRNs), es decir, las proteínas se representaron mediante secuencias de aminoácidos, y demostraron que los residuos en las proteínas plegadas también muestran una topología de “mundo pequeño”. En estas redes, los nodos representan átomos de $C\alpha$ de los aminoácidos, y las conexiones se establecen si dos nodos están separados por una distancia menor a cierto valor o radio de corte. Además, Atilgan *et al.* [9] encontraron que estos resultados son independientes de los diferentes valores de corte seleccionados dentro de cierto intervalo.

Bagler y Sinha [10], utilizando la misma representación propuesta por Atilgan *et al.* [9] (PRNs), descubrieron que las proteínas, independientemente de su clase de estructura de red, presentan propiedades de “mundo pequeño”. Por otro lado, Brinda y Vishvesh-wara [11] emplearon una medida parametrizada de la fuerza de interacción para determinar si dos nodos que representan residuos de proteínas están conectados. Observaron que estas redes exhiben una distribución de grado compleja, que combina características gaussianas, sigmoidales y un decaimiento de ley de potencia para diferentes valores de corte de interacción.

En general, las redes de residuos de proteínas muestran características topológicas universales, ya que pertenecen a la misma clase topológica de redes [12]. Estos estudios han permitido obtener una comprensión más profunda de la estructura y organización de las proteínas a nivel de red, lo cual es fundamental para entender su función y su relación con enfermedades y procesos biológicos.

Un enfoque utilizado para el estudio de la dinámica de redes complejas son los mapeos acoplados, también conocidos como *Coupled Map Lattices* (CML). Estos mapeos acoplados son sistemas dinámicos teóricos en los que existe alguna forma de interacción entre elementos que poseen un estado continuo, y evolucionan en un tiempo y espacio discretos [13]. Los CML se utilizan principalmente para investigar cualitativamente la dinámica caótica de sistemas espacialmente extendidos, lo que incluye la dinámica del caos espacio-temporal donde el número de grados de libertad efectivos diverge a medida que el sistema aumenta en tamaño [14].

Los mapeos acoplados han encontrado aplicaciones en diversas áreas de estudio. Por ejemplo, se han utilizado para investigar la formación de patrones [15], el crecimiento de cristales [16], la turbulencia en espiral en medios excitables, la ebullición [17], la convección térmica [18], la dinámica de las nubes [19], la formación de ondulaciones en la arena [20], entre otros ejemplos. Además de estas aplicaciones en dinámicas específicas, los mapeos acoplados también se utilizan en el procesamiento de información biológica. Por ejemplo, se han estudiado mediante CML la fibrilación en el ritmo cardíaco [21], la dinámica de la expresión génica [22], y se ha discutido su posible aplicación en la fluctuación del vacío en la física de altas energías [23].

En resumen los mapeos acoplados o CML representan un método utilizado para investigar la dinámica de redes complejas. Estos sistemas teóricos permiten explorar la dinámica caótica y la formación de patrones en sistemas espacialmente extendidos, y han encontrado aplicaciones en diversos campos, desde la física y la química hasta la biología y la informática.

Por todo lo anterior es claro que el análisis de las redes de proteínas puede ser abordado utilizando mapeos acoplados, ya que su dinámica puede ser representada en un espacio y tiempo discretos. Investigaciones previas han utilizado autómatas celulares, que son modelos dinámicos similares a los mapeos acoplados, para visualizar secuencias de proteínas, investigar sus procesos de evolución y predecir sus atributos [24]. Además, se ha logrado identificar y predecir la ubicación de sitios activos en proteínas globulares utilizando cadenas discretas de Markov [25]. En este enfoque, se representa una red tridimensional donde los nodos representan los átomos de C_α de la proteína, y las conexiones están determinadas por vértices que representan el grado de interacción entre los átomos de proteína.

El estudio de las proteínas es de gran importancia, ya que desempeñan un papel fundamental en la organización y regulación de sistemas biológicos en todas las escalas. Por lo tanto, este trabajo se enfoca en el estudio de las redes de residuos de proteína, con el objetivo de caracterizar sus propiedades estadísticas y topológicas, así como investigar su dinámica mediante el uso de mapeos acoplados. El resultado de este análisis proporcionará información valiosa para comprender mejor la estructura, función y comportamiento de las proteínas en los sistemas biológicos.

1.1. Estudio de redes de aminoácidos en proteínas globulares

Ernesto Estrada en su artículo [26] hace un profundo estudio topológico de redes de residuos de proteína (PRNs) de la proteasa principal del virus SARS CoV-2, la cual es una enzima esencial para la replicación del virus. La proteasa es responsable de cortar las proteínas polipéptidas virales en fragmentos más pequeños, lo que permite la producción de nuevas partículas virales. El análisis topológico presentado en el documento se basa en la representación de la estructura tridimensional de la proteína como una red de residuos, donde cada aminoácido está representado por un nodo y dos nodos están conectados si los correspondientes átomos de C_α están a no más de 8.0 Å.

El estudio compara las características topológicas de la proteasa del SARS CoV-2 con las del virus SARS CoV-1 y encuentra que hay una similitud significativa entre ambas proteasas. Además, el estudio utiliza el test U de Mann-Whitney para evaluar la significancia estadística de las diferencias entre las dos proteasas y encuentra que no hay diferencias significativas en la mayoría de las métricas topológicas analizadas. Sin embargo el estudio propone dos índices topológicos relevantes: “comunicabilidad” y “centralidad de subgrafos de largo alcance”, respectivamente.

La comunicabilidad es una métrica topológica que mide la capacidad de un nodo para comunicarse con otros nodos en una red. La centralidad de subgrafos de largo alcance (LR) mide la participación de un nodo en todos los subgrafos en la red, incluyendo subgrafos más grandes que los considerados en la comunicabilidad. Ambos índices son relevantes para evaluar la eficiencia topológica de una red en términos de su capacidad para transmitir información a través de sus nodos

El estudio de Estrada *et al.* encuentra diferencias significativas entre estos dos índices para las proteasas del SARS CoV-1 y SARS CoV-2, lo que sugiere grandes diferencias en la capacidad de transmisión de información entre las proteasas del SARS CoV-1 y SARS CoV-2, según los índices de comunicabilidad y centralidad de subgrafos de largo alcance (LR). En particular, muestra que la proteasa del SARS CoV-2 tiene 13 veces más capacidad para transmitir perturbaciones entre pares de nodos que la proteasa del SARS CoV-1, lo que indica una mayor eficiencia en la transmisión de información entre sus aminoácidos.

Los resultados sugieren que el análisis topológico puede ser útil para comprender mejor la estructura y función del virus, lo que podría contribuir al desarrollo de tratamientos y vacunas efectivos contra COVID-19.

1.2. Metodología de análisis dinámico de la estructura de las PRN

Las dinámicas espacio-temporales en redes de carbonos C_α de proteínas globulares están estrechamente relacionadas con su estructura topológica, y esta relación puede ser representada cualitativamente y visualizada mediante el uso de mapeos acoplados (CML).

Esperamos que el análisis mediante CML permita identificar patrones de dinámica en las redes de C_α , y que estos patrones se correspondan con los sitios activos en las proteínas globulares encontrados experimentalmente. Además, esperamos que las diferencias en la dinámica espacio-temporal entre las redes de carbonos C_α y el flujo de información en estas estructuras presenten una correlación con los puntos de mayor centralidad de la red.

La metodología utilizada en este estudio incluye el análisis de las propiedades topológicas de las redes de residuos de proteína (PRNs) de C_α de la proteína globular M^{pro} del virus SARS-CoV-2., en particular, se exploran las medidas de centralidad utilizando la biblioteca NetworkX en Python. Paralelamente se involucra el análisis de propiedades de sincronización espacio-temporal y el estado dinámico de los nodos de la red con el objetivo de identificar y predecir la ubicación de los sitios activos de la proteína. Finalmente se realiza un análisis comparativo entre las propiedades de esta red y las de otras redes de referencia utilizando mapeos acoplados.

Para llevar a cabo estos análisis, se utiliza el lenguaje de programación Python. Python ofrece una amplia gama de herramientas y bibliotecas especializadas en simulación y análisis de redes tales como NetworkX, lo que lo convierte en una elección adecuada para este tipo de estudios.

La proteína M^{pro} del virus SARS-CoV-2 se selecciona específicamente debido a su importancia como principal proteasa del coronavirus y su relevancia en la investigación sobre la inhibición viral mediante fármacos, especialmente en el contexto actual de la pandemia de COVID-19. Esta proteína desempeña un papel fundamental durante el inicio de la infección y la replicación viral, lo que la convierte en un objetivo clave para el desarrollo de posibles tratamientos [27].

En resumen la metodología propuesta se basa en el análisis de las propiedades topológicas, de sincronización y dinámica de las redes de residuos de proteína de M^{pro} utilizando técnicas computacionales y algoritmos específicos implementados en Python. La comparación con redes de referencia y el uso de mapeos acoplados permiten obtener información relevante sobre los sitios activos de la proteína y su relación con otras redes biológicas. Estos resultados contribuyen a la comprensión de la estructura y función de la proteína M^{pro} y su potencial como objetivo terapéutico en la lucha contra el virus SARS-CoV-2.

Los objetivos específicos de este estudio comprenden la identificación de propiedades topológicas en las redes de proteínas globulares, enfocándose especialmente en la búsqueda de puntos de centralidad. Además, se busca descubrir patrones de dinámica presentes en estas redes mediante la aplicación de mapeos acoplados. Se tiene como objetivo establecer relaciones entre los patrones dinámicos identificados, los mecanismos de transmisión de información y la estructura topológica de las proteínas. También se pretende evaluar cómo las diferencias en la topología de las redes afectan el flujo de información en proteínas globulares, realizando una comparación entre las propiedades y la dinámica de las redes de C_α . Finalmente, se busca proporcionar una representación gráfica de la dinámica de las redes mediante el uso de mapeos acoplados, con el objetivo de facilitar la comprensión de los mecanismos de transmisión de información en estas estructuras.

En conjunto, este estudio busca proporcionar una comprensión más profunda de cómo la dinámica espacio-temporal de las redes de proteínas globulares está relacionada con la topología en estas estructuras, y cómo esta relación puede ser representada y visualizada.

Capítulo 2

Teoría de redes y modelos de mapeos acoplados

2.1. Estudio de las redes en el mundo real

En la actualidad, la investigación sobre redes complejas ha experimentado una amplia expansión en diversos campos científicos [28]. Se ha reconocido que muchos sistemas pueden ser descritos y analizados utilizando modelos de redes. Por ejemplo, Internet se puede representar como una red de enrutadores o dominios interconectados, el cerebro humano puede ser estudiado como una red de neuronas interconectadas, las organizaciones se pueden analizar como redes de interacción entre personas, y la economía global puede ser entendida como una red compuesta por economías nacionales, las cuales a su vez son redes de mercados interconectados. Las redes también son una herramienta útil para representar relaciones en sistemas biológicos, como redes alimentarias y vías metabólicas. Incluso en ámbitos lingüísticos y de resolución de problemas matemáticos, las redes pueden utilizarse para representar las relaciones entre palabras en un idioma y las estrategias utilizadas para resolver problemas. Además, las redes desempeñan un papel crucial en el estudio de la propagación de enfermedades y virus informáticos. Las enfermedades se transmiten a través de redes de contacto social, mientras que los virus informáticos se propagan principalmente a través de Internet. Asimismo, las redes son fundamentales en la distribución de energía en organismos vivos y en infraestructuras humanas, como las redes eléctricas[29].

Existen diferentes enfoques y métodos para investigar las redes complejas. Uno de ellos es el análisis de las propiedades estructurales de las redes, como la distribución de grado, la densidad de enlaces, la centralidad y la modularidad. Otro enfoque es el modelado de

redes, donde se crean modelos matemáticos para reproducir las características de las redes reales. La simulación también se utiliza para estudiar las redes complejas, mediante el uso de algoritmos y herramientas computacionales. Se pueden analizar las dinámicas en las redes, como la propagación de enfermedades o la dinámica de sistemas complejos. Además, se pueden emplear técnicas de aprendizaje automático para analizar y extraer información de las redes. En general, el estudio de las redes complejas es un campo interdisciplinario que involucra varias disciplinas, como la física, las matemáticas, la informática, la biología y la sociología.

Matemáticamente, una red se define como un conjunto de elementos o nodos que se relacionan entre sí mediante enlaces o aristas. Estas relaciones pueden ser de diversa naturaleza, como las relaciones entre individuos en una red social, las relaciones entre sitios web en internet, o las relaciones entre proteínas en una red metabólica. El estudio de las redes complejas se basa en el análisis de estas relaciones y en la identificación de patrones y propiedades emergentes en el sistema global. A continuación se presenta una definición matemática formal de red o grafo:

Definición 1 Una red G consiste de un conjunto de nodos o vértices $V = \{v_1, v_2, \dots, v_N\}$, y un conjunto de parejas ordenadas o aristas $\epsilon = \{(v_i, v_j)\} \subset V \times V$. Cada pareja ordenada (v_i, v_j) se llama conexión dirigida del nodo v_i al nodo v_j . La red se llama no dirigida si para cada pareja $(v_i, v_j) \in \epsilon$ también existe la pareja $(v_j, v_i) \in \epsilon$. De lo contrario, la red se denomina dirigida. Llamaremos a todos los nodos que estén conectados directamente a un nodo v_i , los vecinos de v_i . Finalmente, el número k_i de vecinos del nodo v_i (es decir, el número de conexiones de v_i) se llama la conectividad de v_i , y el promedio de estas conectividades, $\langle k \rangle = N^{-1} \sum_{i=1}^N k_i$, es la conectividad de la red.

Toda red puede ser representado por una matriz (esto se ejemplifica en la figura 2.1), que llamamos matriz de adyacencia que se define de la siguiente manera:

Definición 2 Para un grafo con conjunto de vértices $V = \{v_1, v_2, \dots, v_N\}$, la matriz de adyacencia es una matriz A cuadrada $N \times N$ de tal manera que su elemento A_{ij} es uno cuando hay una arista (conexión) del vértice v_i al vértice v_j , y cero cuando no hay arista [30]. Los elementos diagonales de la matriz son todos cero, ya que las conexiones de un nodo consigo mismo (bucles) no están permitidos.

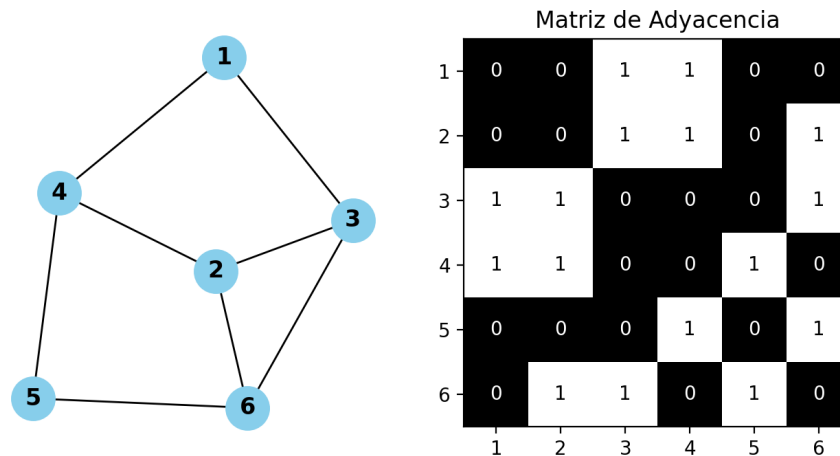


Figura 2.1: Representación de un grafo y su respectiva matriz de adyacencia.

2.2. Estructura y propiedades

Hasta el siglo pasado, el estudio de sistemas físicos y no físicos se basaba en la suposición implícita de que los patrones de interacción entre los componentes del sistema o proceso se pueden describir mediante una estructura regular y simple. [29]. Esto cambió a finales de 1950, cuando dos matemáticos, Erdős y Rényi (ER), describieron una red con una topología compleja mediante un grafo aleatorio [31]. Su trabajo estableció las bases de la teoría de redes aleatorias, a pesar de que en la vida real las redes no son completamente regulares ni completamente aleatorias, el modelo de red aleatorio de ER fue el único enfoque riguroso sobre redes complejas durante casi medio siglo, debido a la ausencia de poder computacional que permitiera extraer información topológica detallada sobre redes de gran escala del mundo real. En la era actual, el acceso a las tecnologías de la información ha abierto la posibilidad de adquirir información detallada sobre la estructura de diversas redes reales. Esto ha dado lugar al análisis de las características y propiedades de diferentes tipos de redes complejas.

A través del estudio de las redes reales, han surgido dos descubrimientos de gran relevancia: el fenómeno de “mundo pequeño” y las propiedades libres de escala. El fenómeno de mundo pequeño revela que, en muchas redes reales, la distancia entre dos nodos seleccionados al azar es sorprendentemente corta, lo que implica una alta eficiencia en la comunicación y la propagación de información. En 1998, para describir la transición de un grafo regular a un grafo aleatorio, Watts y Strogatz (WS) introdujeron el concepto de red de “mundo

pequeño”[4]. Una manifestación del efecto mundo pequeño es el llamado principio de los “seis grados de separación”, sugerido por un psicólogo social, Milgram, a fines de la década de 1960 [32]. Aunque este principio sigue siendo controvertido, se ha demostrado que el patrón de mundo pequeño está presente en muchas redes reales. Una característica común de las redes aleatorias de ER y del modelo de mundo pequeño de WS es que la distribución de conectividad de la red alcanza su punto máximo en un valor promedio y decae exponencialmente. Tales redes se denominan “redes exponenciales” o “redes homogéneas”, porque cada nodo tiene aproximadamente el mismo número de conexiones.

Por otra parte, se ha descubierto que muchas redes complejas son libres de escala, es decir, su conectividad sigue una distribución de ley de potencias que es independiente de la escala de la red [3, 33]. Las propiedades libres de escala señalan que en estas redes, algunos nodos tienen un grado de conectividad significativamente mayor que otros, lo que da lugar a la formación de “*hubs*” o centros con una influencia destacada. Es por esto que a diferencia de una red exponencial, una red sin escala no es homogénea por naturaleza: la mayoría de los nodos tienen muy pocas conexiones y sin embargo, algunos nodos tienen muchas conexiones. El descubrimiento de las redes libres de escala fue basado en la observación de que las distribuciones de grado de muchas redes reales que tienen una forma de ley de potencia, siendo que las distribuciones de ley de potencia se han investigado durante mucho tiempo en la física para muchos otros sistemas y procesos. Todos estos descubrimientos han revolucionado nuestro entendimiento de las redes y han tenido aplicaciones en diversos campos, como el estudio de epidemias, la difusión de información y el diseño de sistemas resilientes y eficientes.

2.3. Métricas de centralidad

Además de formular un modelo matemático adecuado para representar una red del mundo real, es fundamental comprender las propiedades y características estructurales de dicha red. Gracias a las tecnologías computacionales actuales es relativamente sencillo cuantificar y evaluar la importancia relativa de los nodos individuales dentro de la red. Para esto, se utilizan métricas de centralidad, que son medidas que permiten determinar qué nodos son más relevantes en función de diferentes criterios. La centralidad es un término utilizado para describir la importancia de los nodos en un grafo, y las métricas de centralidad proporcionan una forma de cuantificar esta importancia.

Las métricas de centralidad son herramientas clave en el análisis de redes complejas, ya que nos permiten comprender la estructura y las características de los nodos en términos de su influencia, posición y conexión con otros nodos. Algunas de las métricas de centralidad más importantes incluyen la centralidad de grado, que se basa en el número de conexiones directas de un nodo; la centralidad de intermediación, que mide la participación de un nodo en los caminos más cortos entre otros nodos; la centralidad de cercanía, que evalúa la distancia promedio de un nodo a todos los demás nodos en la red; y la centralidad de eigenvector, que cuantifica la importancia de un nodo en función de la importancia de los nodos con los que está conectado.

Al utilizar estas métricas, los investigadores pueden identificar los nodos más influyentes, los puntos de control clave y los posibles puntos críticos de falla en la red. Esto tiene implicaciones en una amplia gama de campos, como la identificación de líderes de opinión en redes sociales, el diseño de estrategias de marketing, la detección de anomalías en redes de infraestructura y mucho más.

Una primera categoría de medidas de centralidad corresponde a aquellas que se relacionan con la estructura más local alrededor de los nodos. Se puede comenzar con el grado k_i del nodo i que se define como el número total de sus conexiones. Siendo así, que cuanto mayor sea el grado, más “importante” es el nodo en la red. El promedio de k_i sobre todo i se denomina grado promedio de la red y se denota por $\langle k \rangle$. La distribución de los grados en una red se caracteriza por una función de distribución $P(k)$, que cuantifica la probabilidad de que un nodo seleccionado aleatoriamente tenga exactamente k conexiones. Una red regular tiene una secuencia de grado simple porque todos los nodos tienen el mismo número de aristas; por lo que su gráfica de distribución de grado contiene un solo pico agudo (distribución delta). En el caso límite de una red completamente aleatoria, la secuencia de grado obedece a la distribución de Poisson; y la forma de la distribución de Poisson cae exponencialmente lejos del valor pico $\langle k \rangle$. Debido a este decaimiento exponencial, la probabilidad de encontrar un nodo con k aristas se vuelve insignificanamente pequeño para $k \gg \langle k \rangle$ [34].

Muchos resultados empíricos muestran que para la mayoría de redes reales a gran escala la distribución de grados se desvía significativamente de la distribución de Poisson. En particular, para varias redes, la distribución de grados se puede describir mejor mediante una ley de potencias de la forma $P(k) \sim k^{-\gamma}$. Esta distribución de ley de potencias cae más gradualmente que una exponencial, permitiendo la existencia de unos pocos nodos

con grados muy grandes. Debido a que estas leyes de potencia son libres de cualquier escala característica, tales redes se denominan libres de escala o *scale-free*. En la figura 2.2 se ilustra una comparación entre la distribución de Poisson y la distribución de ley de potencia de grado k de una red de ER y de una red libre de escala.

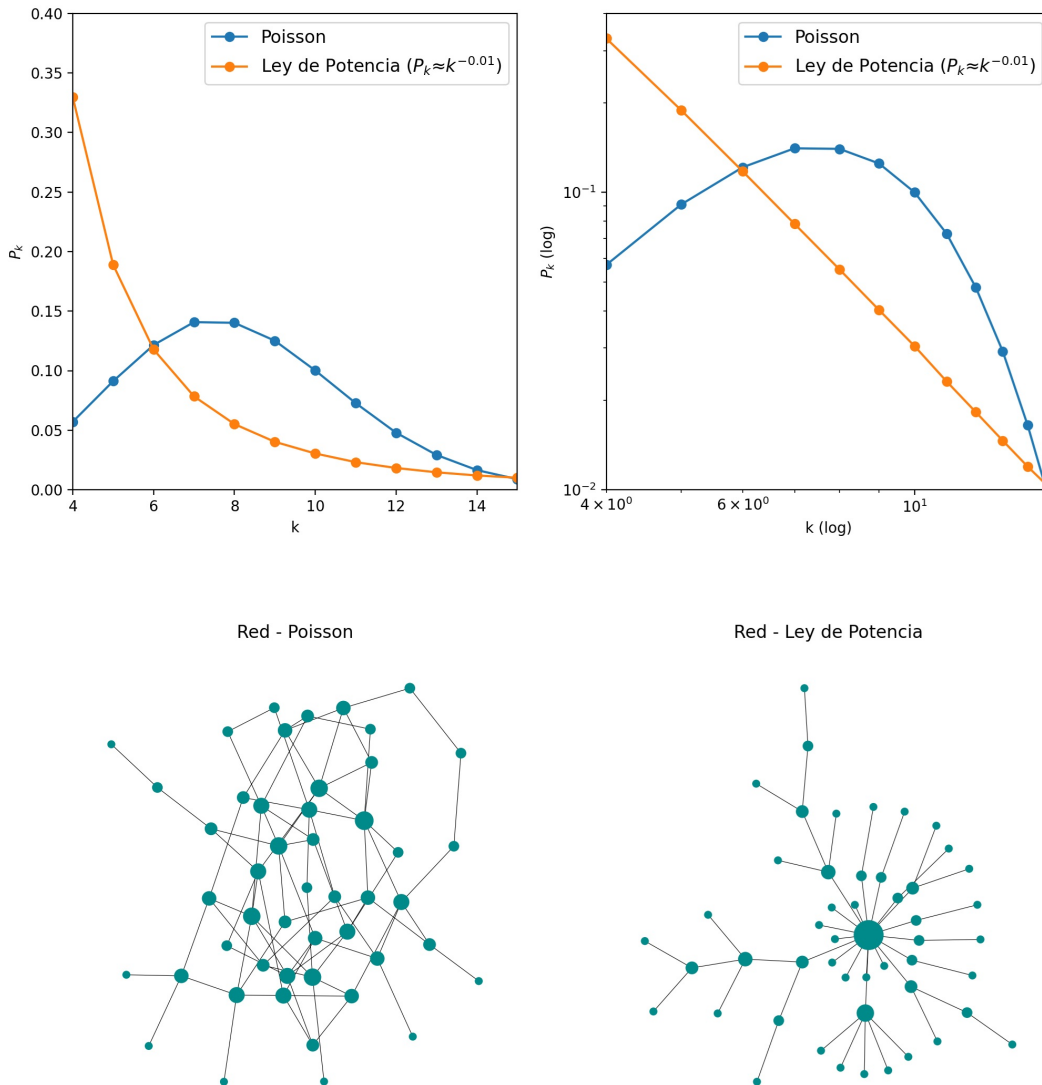


Figura 2.2: Comparación de una distribución de Poisson con una distribución de ley de potencia ($\gamma = -0.01$) en un gráfico lineal y en un gráfico log-log. Ambas distribuciones tienen $\langle k \rangle = 11$.

El grado representa el efecto inmediato de un nodo en su vecindario más cercano. Entre estas medidas, utilizamos aquí la densidad de aristas, que se define como:

$$\delta = \frac{2m}{n(n-1)} \quad (2.1)$$

donde m es el número de aristas y n es el número de nodos. Esta densidad está acotada por $0 \leq \delta \leq 1$, donde el límite inferior se alcanza para una red sin aristas y el límite superior se obtiene para la red completa. Y de manera similar el grado promedio también se puede definir como:

$$\langle k \rangle = \frac{2m}{n} \quad (2.2)$$

Se puede definir otra métrica relacionada con el grado llamada heterogeneidad de grado para la cual la suma se lleva a cabo sobre todos los pares de nodos conectados de la red. La heterogeneidad de grado es una medida de cuán heterogéneos son los grados de los nodos. Un grafo regular tendrá $\rho = 0$, de lo cual se siguen las redes con una distribución normal y luego conforme aumenta la heterogeneidad se puede llegar a redes en las que la distribución de grado sigue leyes de potencia [35].

$$\rho = \sum_{(i,j) \in E} (k_i^{-\frac{1}{2}} - k_j^{-\frac{1}{2}})^2 \quad (2.3)$$

El coeficiente de agrupamiento C se define como la fracción promedio de vecinos de un nodo que también son vecinos unos de otros [4], esto se esquematiza en la figura 2.3. Suponiendo que un nodo i en la red tiene k_i conexiones (nodos directamente conectados). Denotemos por E_i el número de enlaces existentes entre los vecinos de i . Claramente, como máximo solo pueden existir $k_i(k_i - 1)/2$ conexiones entre los vecinos de i , y esto ocurre cuando cada vecino del nodo i conecta a cada otro vecino del nodo i . De esta manera el coeficiente C_i de agrupamiento del nodo i es definido como la relación entre el número E_i de conexiones o aristas que en realidad existen entre estos k_i nodos y el número total posible $k_i(k_i - 1)/2$, de conexiones es decir:

$$C_i = 2E_i / (k_i(k_i - 1)) \quad (2.4)$$

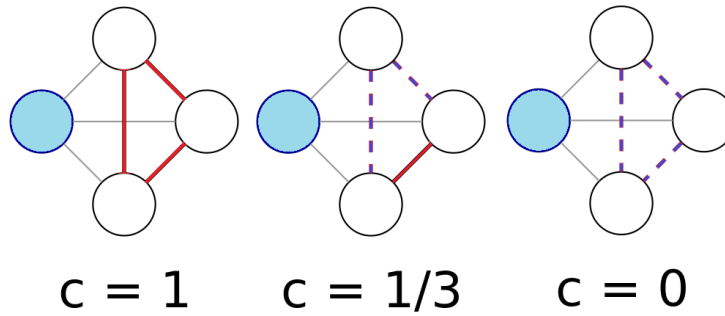


Figura 2.3: Coeficiente de agrupamiento para el nodo azul en un grafo no dirigido. Las líneas rojas son aristas y los segmentadas punteados son aristas inexistentes.

El coeficiente de agrupamiento $\langle C \rangle$ de toda la red es el promedio de C_i sobre todo nodo i es:

$$\langle C \rangle = \frac{1}{n} \sum_{i=1}^n C_i \quad (2.5)$$

Ciertamente, $C \leq 1$; y $C = 1$ si y solo si el grafo es completo, lo que significa que cada nodo en la red se conecta a todos los demás nodos. En una red aleatoria que consta de N nodos, $C \approx 1/N$, que es muy pequeño en comparación con la mayoría de las redes reales. La mayoría de las redes reales a gran escala tienen una tendencia hacia el agrupamiento, en el sentido de que sus coeficientes de agrupamiento son mucho mayores a $1/N$, aunque son todavía significativamente menores que uno (es decir, lejos de ser completas). Esto, a su vez, significa que la mayoría las redes reales no son completamente aleatorias.

Otra medida relacionada con el grado es el coeficiente de asortatividad del grado r [36], que es el coeficiente de correlación de Pearson de la correlación de grado a grado. Donde $r > 0$ (asortatividad de grado) indica una tendencia de que los nodos de alto grado se conecten con otros nodos de alto grado. $r < 0$ (disasortatividad de grado) indica la tendencia de que los nodos de alto grado se conecten con nodos de bajo grado.

En una red, la distancia $d_{i,j}$ entre dos nodos, con etiquetas i y j respectivamente, se define como el número de aristas en el camino más corto que los conecta. El diámetro de una red, por lo tanto, se define como la distancia máxima entre todas las distancias entre cualquier par de nodos en la red. La longitud de camino promedio D de la red, se define como la distancia media entre dos nodos, promediada sobre todos los pares de nodos. Aquí, D determina el “tamaño” efectivo de una red, es decir, la separación mas típica entre un par de nodos en la red. Esta medida asume que la ”información” se transmite en la red a través de las rutas topológicas más cortas. La longitud de camino promedio es:

$$\langle D \rangle = \frac{2}{n(n-1)} \sum_{i < j} d_{ij} \quad (2.6)$$

esta medida se utiliza típicamente como una medida de la “propiedad de pequeño mundo” (small-worldness) de la red [4].

También consideramos la centralidad de intermediación promedio BC (betweenness centrality) [37].

$$\langle BC \rangle = \frac{1}{n} \sum_{i \neq k \neq j} \frac{\rho_{ikj}}{\rho_{ij}} \quad (2.7)$$

donde ρ_{ikj} es el número de caminos más cortos entre los nodos i y j que pasan por el nodo k , y ρ_{ij} es el número total de caminos más cortos que van desde i hasta j . Esta medida representa la importancia de un nodo en el paso de información a través de él para conectar otros pares de nodos mediante el camino más corto.

La centralidad de la cercanía es una medida en la que la importancia de cada nodo está determinada por su cercanía a todos los demás nodos. Sea d_{ij} la longitud del camino más corto entre los nodos i y j , la distancia media l_i es: $\frac{1}{N} \sum_j d_{ij}$. Dado que se esta buscando el nodo más cercano, la centralidad de cercanía Cd_i es inversamente proporcional a la longitud promedio l_i , por lo que: $Cd_i = \frac{1}{l_i} = \frac{N}{\sum_j d_{ij}}$

$$\langle Cd \rangle = \frac{1}{n} \sum_{i=1}^n Cd_i \quad (2.8)$$

Existe otra categoría de medidas de centralidad que tiene en cuenta la transmisión de información no solo a través de las rutas más cortas, sino que utiliza cualquier ruta disponible que conecte un par de nodos. Estas medidas se basan en caminatas en lugar de caminos. Una caminata de longitud k en la red G es un conjunto de nodos $i_1, i_2, \dots, i_k, i_{k+1}$ tal que para todo $1 \leq l \leq k$, $(i_l, i_{l+1}) \in E$ enlaces. Entonces una caminata cerrada es aquella en la que el primer y último nodo son el mismo $i_1 = i_{k+1}$. Y el número de caminatas de longitud k entre los nodos i y j en una red se denota como $(A^k)_{i,j}$.

La centralidad de vector propio o de eigenvector (EC, por sus siglas en inglés), es una generalización de la centralidad de grado y cuantifica la importancia relativa de un nodo en función de su conexión con otros nodos en la red. Los nodos con una centralidad de vector propio alta son considerados más importantes en términos de su posición, contribución a la estructura y flujo de información en la red. La relación de este índice con las caminatas se expresa de la siguiente manera: Sea $N_k(i)$ el número de caminatas de longitud k que comienzan en el nodo i y terminan en otro lugar. Entonces, si la red no es bipartita [38]:

$$EC_i = \lim_{k \rightarrow \infty} \frac{N_k(i)}{\sum_{j=1}^n N_k(j)} \quad (2.9)$$

En otras palabras, la centralidad de vector propio de un nodo es la proporción entre el número de caminos de longitud infinita que comienzan en ese nodo y el número total de caminos de longitud infinita que comienzan en cualquier otro lugar. En consecuencia, la centralidad de vector propio promedio $\langle EC \rangle$ tiene en cuenta la propagación de la información desde los nodos más allá de los vecinos más cercanos y utilizando caminos de longitud infinita en el grafo.

Otro tipo de medidas de este tipo se basan en contar todos los caminos de cualquier longitud, pero dando más peso a los más cortos que a los más largos. Estas medidas se basan en la siguiente función matricial.

$$G = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \exp(A) \quad (2.10)$$

Donde $\exp(A)$ es la exponencial de la matriz. Luego, consideramos el promedio de las entradas diagonales de esta matriz, lo cual se conoce como la centralidad subgráfica promedio [39].

$$\langle SC \rangle = \frac{1}{n} \sum_{p=1}^n G_{pp} \quad (2.11)$$

Esta métrica tiene en cuenta la participación del nodo correspondiente en todos los subgrafos de la red, dándole más peso a los subgrafos más cortos que a los más largos. Estos subgrafos pueden incluir, aristas, triángulos, cuñas, cuadrados, etc.

Otra medida es el promedio de los elementos no diagonales de $\exp(A)$, que se conoce como la comunicabilidad promedio de la red [40].

$$\langle G_{pq} \rangle = \frac{2}{n(n-1)} \sum_{p < q} G_{pq} \quad (2.12)$$

Esta medida considera cuánto pueden comunicarse entre sí un par de nodos utilizando todas las rutas potenciales disponibles en la red, pero otorgando más importancia a las rutas más cortas que a las más largas.

Finalmente, en esta categoría, se incluye el ángulo promedio de comunicabilidad [41].

$$\langle \theta_{pq} \rangle = \frac{2}{n(n-1)} \sum_{p < q} \theta_{pq} \quad (2.13)$$

Donde el ángulo entre un par de nodos está definido como:

$$\theta_{pq} = \cos^{-1} \left(\frac{G_{pq}}{\sqrt{G_{pp}G_{qq}}} \right) \quad (2.14)$$

El ángulo promedio de comunicabilidad describe qué tan eficientemente una red transmite información entre sus pares de nodos utilizando todas las rutas disponibles.

Una ultima categoría de medidas de centralidad puede ser definida sobre los índices de "todos los caminos" de forma que penaliza menos a los caminos más largos que conectan pares de nodos en una red. Es decir, al igual que $G = \exp(A)$ tiene en cuenta todos los caminos que conectan cada par de nodos, pero penaliza menos a aquellos caminos de longitud relativamente grande.

$$Z = \sum_{k=0}^{\infty} \frac{A^k}{k!!} = \frac{1}{2} (\sqrt{2\pi} \operatorname{erf}(\frac{A}{\sqrt{2}}) - 2I) \exp(\frac{A^2}{2}) \quad (2.15)$$

En lugar de utilizar el factorial simple ($k!$) para penalizar los caminos de longitud k , se utiliza el doble factorial ($k!!$). El efecto principal de esta penalización con doble factorial es que los caminos de mayor longitud son menos penalizados en comparación con el uso del factorial simple. Esto se describe en la figura 2.4 que compra el crecimiento de la función doble factorial y factorial.

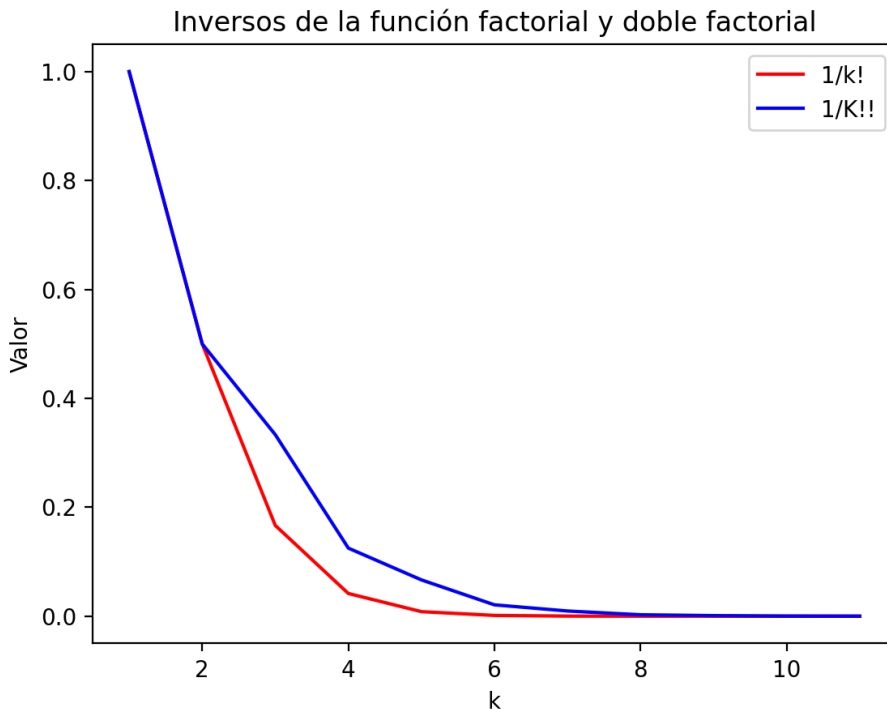


Figura 2.4: Crecimiento de la inversa de la función doble factorial y de la inversa de la función factorial.

Luego, podemos considerar el promedio de la diagonal principal de la matriz Z , de forma que se toman en cuenta los nodos i de todos los subgrafos de la red.

$$\langle Z_{ii} \rangle = \frac{1}{n} \sum_{i=1}^n Z_{ii} \quad (2.16)$$

De igual manera se puede considerar estos índices para tomar en cuenta la capacidad global de la red para transmitir información entre pares de nodos y permitir transmisiones a larga distancia, en mayor medida que la comunicabilidad.

$$\langle Z_{ij} \rangle = \frac{2}{n(n-1)} \sum_{i < j} Z_{ij} \quad (2.17)$$

Por estas razones, se propone llamar a los índices Z_{ii} como centralidad de subgrafos de largo alcance (LR) y a Z_{ij} como índices de comunicabilidad de largo alcance [26].

2.4. Modelos de redes

La ciencia de redes tiene como objetivo construir modelos que reproduzcan las propiedades de las redes reales. Si determinar algunas propiedades básicas de una red; como la longitud de camino promedio D , el coeficiente de agrupamiento C y la distribución de grados $P(k)$, es el primer paso hacia la comprensión de su estructura. El siguiente paso, es entonces, desarrollar un modelo matemático con una topología de propiedades estadísticas similares, obteniendo así una plataforma en la que es posible el análisis matemático.

Una red regular es un tipo de red determinista en la cual todos los nodos tienen el mismo número de conexiones o grado. En una red regular, cada nodo está conectado exactamente con el mismo número de vecinos, lo que resulta en una estructura altamente simétrica. Formalmente, una red regular se define de la siguiente manera: Sea una red con N nodos. Si cada nodo tiene el mismo número k de conexiones y cada par de nodos está conectado por un solo enlace, entonces la red se denomina k - regular.

Es posible definir una red completa como un caso especial de una red regular, donde todos los nodos están conectados directamente entre sí. Esta definición implica que cada par de nodos en la red completa está unido por una arista. Aunque la red completa comparte propiedades como el efecto de mundo pequeño y un alto coeficiente de agrupamiento con

muchas redes reales, es importante reconocer sus limitaciones. Por ejemplo, una red completa con N nodos tiene un número total de aristas igual a $N(N - 1)/2$. Sin embargo, la mayoría de las redes reales a gran escala son poco densas, lo que significa que no están completamente conectadas y el número de conexiones suele ser del orden de N , en lugar de N^2 .

Un modelo de red ampliamente estudiado que aborda esta limitación de densidad es la red de vecinos cercanos, también conocida como “lattice”. Esta red consiste en un grafo regular donde cada nodo está conectado solo a unos pocos de sus vecinos. Aunque el término lattice se asocia comúnmente con una cuadrícula cuadrada bidimensional, en realidad puede tener diversas geometrías. Un ejemplo común es una red en forma de anillo, donde cada nodo está conectado a sus dos vecinos más cercanos en una dimensión con una condición de frontera periódica consta de N nodos dispuestos en un anillo, donde cada nodo i está conectado a sus vecinos $i = 1, 2, \dots, k/2$, siendo k un número entero par.

En una red de vecinos cercanos con un valor alto de k , se observa una alta agrupación, lo que se refleja en un coeficiente de agrupamiento aproximado de $C = 3/4$. Esto significa que los nodos tienden a formar grupos o clústeres en la red de vecinos cercanos, lo cual contrasta con la conectividad completa de una red completa. No obstante, la red de vecinos cercanos se diferencia de una red de mundo pequeño, ya que su longitud de camino promedio es considerablemente mayor y tiende a infinito a medida que el tamaño de la red (N) se acerca al infinito. El hecho de que la longitud de camino promedio sea alta en una red de vecinos cercanos puede dificultar la realización de procesos dinámicos eficientes. Los caminos entre nodos distantes en la red son más largos y no existen atajos o conexiones directas que permitan acortar las distancias. Esto implica que la propagación de información o la coordinación global en la red se vuelven más desafiantes en comparación con las redes que exhiben propiedades de mundo pequeño. En consecuencia, la estructura de la red de vecinos cercanos puede dificultar el logro de una sincronización completa u otros procesos dinámicos que requieran una coordinación global rápida. La falta de atajos o rutas más cortas limita la capacidad de la red para alcanzar un consenso rápido o difundir información de manera eficiente. En resumen, a diferencia de las redes de mundo pequeño, la red de vecinos cercanos presenta una longitud de camino promedio considerablemente grande, lo que puede dificultar la realización de procesos dinámicos que necesiten una coordinación global, como la sincronización.

En contra partida de las redes regulares y deterministas se encuentran los grafos completamente aleatorios o estocásticos, que fueron originalmente propuestos y estudiados por

Erdős y Rényi (ER) hace unos 40 años [31]. Existen dos definiciones para una red aleatoria:

Definición 3 $G(N, p)$ es un grafo aleatorio con N nodos donde cada posible arista tiene probabilidad p de existir.

Este modelo fue introducido por Gilbert [42]. El número de grafos con exactamente N nodos y L aristas es igual al número de formas de elegir las aristas de los $\binom{N}{2} = N(N-1)/2$ pares de vértices distintos posibles. El número de aristas en un grafo $G(N, p)$ es una variable aleatoria con valor esperado $\langle L \rangle = \binom{N}{2} p$. El objetivo principal de la teoría de redes aleatorias consiste en determinar a qué probabilidad de conexión p es más probable que emerja una propiedad particular del grafo.

Definición 4 $G(N, L)$ es el modelo de todos los grafos posibles con N nodos y exactamente L aristas, seleccionadas al azar.

Erdős y Rényi utilizaron esta definición en su serie de artículos sobre redes aleatorias.

El modelo $G(N, p)$ fija la probabilidad p de que dos nodos estén conectados y el modelo $G(N, L)$ fija el número total de enlaces L . En el modelo $G(N, L)$ el grado promedio de un nodo es simplemente $\langle k \rangle = 2L/N$, y el grado promedio en $G(N, p)$ es $\langle k \rangle = 2\binom{N}{2} p/N = p(N-1) \sim pN$.

Sea D_{ER} la longitud de camino promedio de una red aleatoria. El grado promedio $\langle k \rangle^{D_{ER}}$ sobre los nodos de la red aleatoria están a una distancia D_{ER} o muy cerca de eso, es intuitivamente $N \sim \langle k \rangle^{D_{ER}}$, lo que significa que $D_{ER} \sim \ln N / \langle k \rangle$. Este aumento logarítmico en la longitud del camino promedio con el tamaño de la red es un efecto típico de mundo pequeño. Como $\ln N$ aumenta lentamente con N , permite que la longitud de camino promedio sea bastante pequeña incluso en una red bastante grande. Por otra parte, el coeficiente de agrupamiento del modelo ER es $C = p = \langle k \rangle / N \ll 1$. Esto significa que una red aleatoria a gran escala no muestra agrupamiento en general. De hecho, para un N grande, el algoritmo ER genera una red homogénea, donde la conectividad sigue aproximadamente una distribución de Poisson.

Como se mencionó previamente, las redes regulares tienden a exhibir agrupamiento pero generalmente no muestran el efecto de mundo pequeño. Por otro lado, las redes aleatorias muestran el efecto de mundo pequeño, pero carecen de agrupamiento en su estructura. Por lo tanto, no es sorprendente que ambos modelos no logren reproducir algunas características importantes presentes en muchas redes del mundo real. La mayoría de las redes reales

no se ajustan completamente a la regularidad o a la aleatoriedad.

Con el propósito de describir la transición de una red regular a una red aleatoria, Watts y Strogatz [4] introdujeron el modelo de mundo pequeño, conocido como modelo WS. Este modelo se construye inicialmente a partir de una red regular de vecinos cercanos que consta de N nodos. Cada nodo i está conectado a sus vecinos más cercanos, es decir, los nodos $i = 1, 2, \dots, K/2$, donde K es un número par. Posteriormente, se procede a "recablear" aleatoriamente los enlaces de un nodo en la red con una probabilidad p , variando p para generar una transición desde el orden ($p = 0$) hasta la aleatoriedad ($p = 1$). En el contexto de este modelo, el recableado implica cambiar uno de los extremos de una conexión por un nuevo nodo seleccionado aleatoriamente de toda la red, con la restricción de que dos nodos diferentes no pueden tener más de una conexión entre ellos y ningún nodo puede estar conectado consigo mismo. Este proceso introduce $pNK/2$ conexiones de largo alcance, que conectan nodos que, de otra manera, pertenecerían a vecindarios distintos.

Tanto el coeficiente de agrupamiento $C(p)$ como la longitud de camino promedio $D(p)$ en el modelo de mundo pequeño de WS pueden considerarse como funciones de la probabilidad de recableado p . Se ha observado que, para una probabilidad de recableado baja ($p \sim 0$), las propiedades locales de la red se mantienen prácticamente iguales a las de la red regular original, y el coeficiente de agrupamiento no difiere significativamente de su valor inicial ($C(p) \sim C(0)$). Sin embargo, la longitud de camino promedio disminuye rápidamente y se vuelve comparable a la de las redes aleatorias ($D(p) \ll D(0)$). En otras palabras, basta con realizar algunos recableados aleatorios para reducir de manera significativa la longitud de camino promedio, mientras que varios enlaces recableados no alteran de manera crucial la propiedad de agrupamiento local de la red.

En este sentido, el modelo de mundo pequeño también puede ser interpretado como una red homogénea, donde todos los nodos tienen aproximadamente el mismo número de conexiones, pero con una longitud de camino promedio considerablemente pequeña. Esto explica por qué los modelos de mundo pequeño son buenas representaciones de las redes sociales.

2.5. Dinámica discreta y mapeos acoplados CML

Dada la naturaleza discreta de las redes tiene sentido que su dinámica sea discreta por lo que para su estudio se pueden emplear las ecuaciones en diferencias o mapeos acoplados. Un mapeo acoplados o CML es un sistema dinámico con tiempo discreto (“mapa”), espacio discreto (“lattice” o retícula) y estado continuo. Por lo general, involucra elementos dinámicos dispuestos en una retícula que interactúan (acoplados) entre sí, conjuntos adecuadamente seleccionados pueden tratarse de una red.

Los mapeos acoplados se introdujeron originalmente para facilitar el estudio del caos espacio-temporal, es decir, la dinámica caótica en un sistema espacialmente extendido [43]. Los mapeos logísticos acoplados también pueden presentar transiciones de fase [44] constituyendo esta una propiedad de extrema relevancia de los sistemas dinámicos. El CML es comparable con otros tres modelos estándar para el estudio de sistemas dinámicos espacialmente extendidos, estos modelos son los siguientes: ecuaciones diferenciales ordinarias acopladas, en términos de discretización del espacio pero con un continuo de tiempo y estado; ecuaciones diferenciales parciales, en términos de su continuo de espacio, tiempo y estado; y autómatas celulares, en términos de su discretización de espacio, tiempo y estado.

La mayoría de fenómenos dinámicos que los CML modelan se describen mediante la combinación de algunas dinámicas locales elementales y acoplamientos espaciales[45]. Los rasgos característicos del CML son:

- Dinámica de tiempo discreta
- Naturaleza discreta de la red

Formalmente, los CML son sistemas dinámicos de tiempo discreto generado mediante mapeos que actúan sobre secuencias reales (vectoriales). El típico y más estudiado ejemplo es el modelo introducido por Kaneko en 1983 [43]:

$$u_s^{t+1} = (1 - \varepsilon)f(u_s^t) + \frac{\varepsilon}{2}(f(u_{s-1}^t) + f(u_{s+1}^t)) \quad t \in \mathbb{N}, \varepsilon \in [0, 1] \quad (2.18)$$

Donde $u_s^t \in \mathbb{R}$ y f es un mapeo real.

Dependiendo del contexto, las configuraciones u_s^t pueden representar el espacio fase de una concentración química, de una densidad de población, de un campo de velocidad, etc.

El modelo básico de la dinámica de un CML se rige por dos términos en competencia; una reacción individual (estado) no lineal representada por f y una interacción espacial (acoplamiento) con intensidad variable ε .

Su formulación simple hace del CML un paradigma de los sistemas dinámicos no lineales extendidos espacialmente. En particular, los CML están especialmente diseñados para facilitar las simulaciones por computadora en grandes dominios de espacio-tiempo. Las simulaciones exhiben un conjunto de comportamientos extraordinariamente grande ante cambios en el mapeo local y en la interacción (o simplemente en sus parámetros) [46].

El mapeo acoplado dinámico (DCML) [47] puede ser formulado como:

$$x_i(t+1) = (1 - \varepsilon_i(t))f(x_i(t)) + \frac{\varepsilon_i(t)}{2}(f(x_{i+1}(t)) + f(x_{i-1}(t))) \quad (2.19)$$

Donde el coeficiente de acoplamiento varía con el espacio-tiempo sobre los índices i y t . Mas específicamente el sistema DCML aprovecha un mapa logístico auxiliar de la forma:

$$\varepsilon_i(t+1) = f(\varepsilon_i(t)) = \mu^{aux} \varepsilon_i(t)(1 - \varepsilon_i(t)) \quad (2.20)$$

en el cual $t = 1, 2, \dots, T$ genera la secuencia de coeficientes de acoplamiento dinámico. El parámetro auxiliar μ se establece en 3,99 para lograr una dinámica caótica, y el valor inicial de ε_k se establece en e .

Capítulo 3

Proteínas globulares y métodos computacionales basados en mapeos acoplados

Las proteínas son las macromoléculas biológicas esenciales que desempeñan una amplia gama de funciones en los organismos. Estas moléculas son responsables de expresar la información genética en los seres vivos, componer las estructuras celulares y posibilitar las reacciones químicas del metabolismo celular. De hecho, en la mayoría de los organismos (excepto en las plantas, que contienen más celulosa), las proteínas representan más del 50 % de su peso en seco.

La diversidad de proteínas existente es asombrosa. Por ejemplo, una bacteria puede tener alrededor de 1000 proteínas diferentes, mientras que una célula humana puede albergar hasta 10,000 clases distintas de proteínas. Químicamente, las proteínas están constituidas por polímeros de aminoácidos, los cuales se unen mediante enlaces covalentes conocidos como enlaces peptídicos. Estos aminoácidos se disponen de manera lineal, formando una secuencia específica.

Es importante destacar que las células son capaces de producir proteínas con propiedades muy diversas a partir de tan solo 20 aminoácidos diferentes. La combinación de estos aminoácidos y su secuencia específica en la cadena polipeptídica permiten que cada proteína adquiera su estructura tridimensional única y sus características funcionales particulares.

3.1. Aminoácidos

Los componentes básicos de las proteínas se llaman aminoácidos. Los aminoácidos son moléculas orgánicas que poseen un grupo carboxilo ($COOH$) y un grupo amino (NH_2), cuya estructura es representada en la figura 3.1. El carbono alfa (C_α) es un átomo de carbono que se encuentra unido al grupo amino ($-NH_2$) y al grupo carboxilo ($-COOH$) del mismo aminoácido, formando así el esqueleto central de la proteína. Estas moléculas exhiben propiedades ácidas débiles en el caso del grupo carboxilo, y propiedades básicas débiles en el grupo amino.

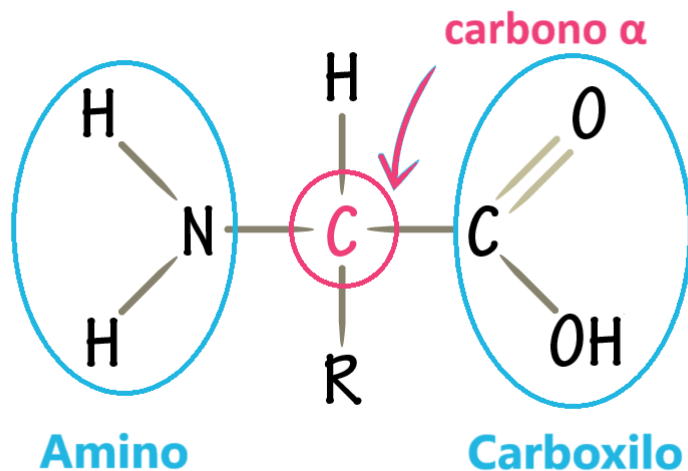


Figura 3.1: Estructura de un α – *aminocido*.

Las proteínas están compuestas por una combinación de 20 aminoácidos diferentes, aunque existen numerosos aminoácidos adicionales que no forman parte de las proteínas. Algunas proteínas también pueden contener moléculas distintas a estos 20 aminoácidos estándar, las cuales a menudo son aminoácidos modificados durante la formación de la proteína.

Los aminoácidos α son los únicos que forman proteínas en cualquier organismo. Sin embargo, algunos péptidos elaborados por microorganismos pueden contener una clase diferente de aminoácidos llamados aminoácidos D. Los organismos heterótrofos tienen la capacidad de sintetizar la mayoría de los aminoácidos por sí mismos. Aquellos aminoácidos que no pueden ser sintetizados se conocen como aminoácidos esenciales y deben ser incorporados a través de la dieta. En el caso de los seres humanos, existen 10 aminoácidos esenciales.

La estructura de un aminoácido α consiste en la unión del grupo amino y el grupo carboxilo al mismo átomo de carbono, conocido como carbono α . Este carbono α también se une a una cadena lateral, conocida como cadena R, y a un átomo de hidrógeno. La diversidad de los aminoácidos radica en sus cadenas laterales (cadenas R), las cuales determinan sus propiedades características, como la polaridad o el carácter ácido o básico.

Se han identificado cinco grupos principales que agrupan a los aminoácidos en base a propiedades estructurales y funcionales similares:

1. Aminoácidos no polares con cadenas R hidrofóbicas: Este grupo incluye aminoácidos con cadenas laterales no polares, lo que significa que no interactúan significativamente con el agua. Ejemplos de aminoácidos en este grupo son glicina, alanina, valina, leucina, isoleucina y metionina.

2. Aminoácidos polares con cadenas R no cargadas: Estos aminoácidos poseen cadenas laterales polares, lo que les permite interactuar con el agua, pero no tienen carga eléctrica. Algunos ejemplos de aminoácidos en esta categoría son serina, treonina, cisteína, tirosina, asparagina y glutamina.

3. Aminoácidos polares con cadenas R cargadas positivamente (básicos): Estos aminoácidos presentan cadenas laterales con carga positiva debido a la presencia de grupos amino adicionales. Incluyen a la lisina, arginina e histidina. Estos aminoácidos pueden formar enlaces iónicos con grupos cargados negativamente.

4. Aminoácidos polares con cadenas R cargadas negativamente (ácidos): Este grupo se compone de aminoácidos con cadenas laterales que tienen carga negativa debido a la presencia de grupos carboxilo adicionales. Los aminoácidos ácidos incluyen ácido aspártico (Asp) y ácido glutámico (Glu). Estos aminoácidos pueden formar enlaces iónicos con grupos cargados positivamente.

5. Aminoácidos especiales: Además de los grupos anteriores, existen algunos aminoácidos con características únicas. Por ejemplo, la prolina tiene una cadena lateral que se une al grupo amino, lo que crea un anillo y restringe la flexibilidad de la molécula. La glicina, por su parte, tiene una cadena lateral compuesta solo por un átomo de hidrógeno, lo que la convierte en el aminoácido más pequeño.

Esta clasificación basada en las cadenas laterales de los aminoácidos es útil para comprender sus propiedades y su papel en la estructura y función de las proteínas.

En el entorno celular, los aminoácidos tienden a encontrarse en forma ionizada, es decir, con carga eléctrica. Esto se debe a que tanto el grupo amino (NH_2) como el grupo carboxilo ($COOH$) pueden perder o aceptar protones (iones de hidrógeno) en función del pH del medio.

Cuando un aminoácido actúa como ácido, el grupo carboxilo dona un protón, lo que resulta en la formación de un ión negativo llamado anión carboxilato ($-COO^-$). Por otro lado, cuando un aminoácido actúa como base, el grupo amino acepta un protón, generando un ión positivo llamado catión amino ($-NH_3^+$).

La capacidad de un aminoácido para comportarse como ácido o base depende de su pKa, que es el valor de pH en el cual la forma ionizada y no ionizada del aminoácido están presentes en igual proporción. Cuando el pH del medio es inferior al pKa, el aminoácido predominará en su forma protonada (ácida). Por el contrario, cuando el pH del medio es superior al pKa, el aminoácido predominará en su forma desprotonada (básica).

Es importante destacar que la ionización de los aminoácidos en disolución acuosa es fundamental para su interacción con otras moléculas en el entorno celular, como en la formación de enlaces peptídicos durante la síntesis de proteínas o en la participación en reacciones químicas y enzimáticas. Además, la capacidad de los aminoácidos para actuar como ácidos o bases contribuye a su capacidad de regular el pH intracelular y mantener un equilibrio en el entorno celular.

3.2. Enlace peptídico

El enlace peptídico es la unión covalente que se forma entre dos aminoácidos mediante la eliminación de una molécula de agua. Este enlace se establece entre el grupo amino ($-NH_2$) de un aminoácido y el grupo carboxilo ($-COOH$) del otro aminoácido, dando como resultado un enlace covalente $CO - NH$. Es importante destacar que la formación del enlace peptídico implica la pérdida de una molécula de agua, lo que se conoce como reacción de condensación.

El enlace peptídico es responsable de la formación de estructuras lineales sin ramificaciones

conocidas como péptidos. Estas estructuras, que pueden variar en longitud, son altamente estables debido a la naturaleza covalente de los enlaces peptídicos. Todos los péptidos tienen un grupo amino en un extremo y un grupo carboxilo en el otro extremo, lo que les confiere polaridad y la capacidad de interactuar con otras moléculas.

Es importante mencionar que la reacción química en la que se forma un enlace peptídico se denomina condensación, ya que implica la unión de dos moléculas con la eliminación de una pequeña molécula de agua. Por otro lado, la descomposición del enlace peptídico, que implica la ruptura del enlace y la separación de los aminoácidos, se conoce como hidrólisis. Durante la hidrólisis, se añade una molécula de agua para romper el enlace peptídico y regenerar los aminoácidos individuales.

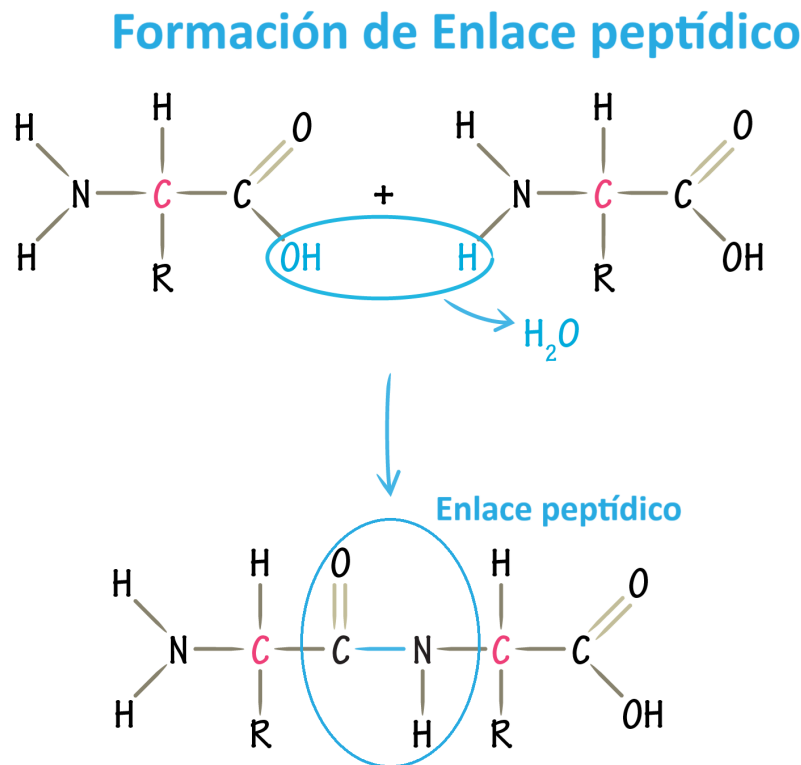


Figura 3.2: Enlace peptídico: condensación.

En función del número de aminoácidos que componen su cadena, los péptidos se pueden clasificar en tres categorías principales:

Péptidos pequeños u oligopéptidos: Estos péptidos están compuestos por un pequeño número de aminoácidos, generalmente entre 2 y 20. Ejemplos conocidos de oligopéptidos son

la oxitocina, que consta de 9 aminoácidos, y el tripéptido glutatión, formado por los aminoácidos glutamato, cisteína y glicina.

Péptidos medianos: Esta categoría abarca aquellos péptidos que contienen entre 20 y 50 aminoácidos. Algunos ejemplos de péptidos medianos son el péptido natriurético atrial (ANP), compuesto por 28 aminoácidos, y el péptido antimicrobiano dermcidina, que contiene 47 aminoácidos.

Péptidos grandes o polipéptido: Estos péptidos son más extensos y están formados por más de 50 aminoácidos. Un ejemplo de polipéptido es la insulina, que consta de 51 aminoácidos en los seres humanos y es conocida como una hormona de acuerdo a la función que tiene en el organismo.

Proteínas: Son cadenas extensas de aminoácidos que poseen una estructura tridimensional bien definida. Usualmente, se utiliza el término proteína para referirse a los polipéptidos cuya masa molecular supera los 10,000. Estas moléculas están compuestas generalmente por un rango de 100 a 300 aminoácidos, aunque algunas pueden llegar a tener más de mil aminoácidos.

Se llama residuo a cada uno de los aminoácidos que forman un péptido.

3.3. Estructuras

La estructura de una proteína se puede describir en cuatro niveles jerárquicos, conocidos como estructura primaria, estructura secundaria, estructura terciaria y estructura cuaternaria. Estos niveles representan diferentes niveles de organización y conformación de la proteína.

1. **Estructura primaria:** Es el nivel más básico y se refiere a la secuencia lineal de aminoácidos en la cadena polipeptídica de la proteína. La estructura primaria está determinada por la información genética codificada en el ADN y es fundamental para la función y características de la proteína.

2. **Estructura secundaria:** Se refiere a la conformación local adoptada por la cadena polipeptídica debido a interacciones entre los aminoácidos cercanos. Los principales tipos de estructura secundaria son las hélices alfa y las láminas beta plegadas. Estas estructuras se

mantienen por puentes de hidrógeno entre los grupos amida y carbonilo de los aminoácidos.

3. Estructura terciaria: Corresponde a la conformación tridimensional completa de una sola cadena polipeptídica. En este nivel, las interacciones entre los aminoácidos más distantes en la secuencia primaria dan lugar a la estructura tridimensional final de la proteína. Las interacciones pueden incluir puentes de hidrógeno, interacciones hidrofóbicas, fuerzas electrostáticas y enlaces disulfuro.

4. Estructura cuaternaria: Solo se aplica a las proteínas que están compuestas por múltiples subunidades polipeptídicas. La estructura cuaternaria se refiere a cómo se organizan y se unen estas subunidades para formar una proteína funcional. Las interacciones entre las subunidades pueden ser a través de enlaces no covalentes, como puentes de hidrógeno y fuerzas electrostáticas.

Estos cuatro niveles de estructura son esenciales para comprender la función y las propiedades de las proteínas, ya que determinan cómo se pliegan y cómo interactúan con otras moléculas en el entorno celular.

3.4. Proteínas globulares

Existen dos tipos principales de proteínas (ver figura 2.4) en función de su estructura terciaria:

1. Proteínas globulares: Estas proteínas tienen una estructura tridimensional compacta y generalmente adoptan una forma esférica o globular. Las proteínas globulares son solubles en agua y desempeñan una amplia gama de funciones en el organismo. Pueden actuar como enzimas, transportadores, receptores, anticuerpos, entre otros. Ejemplos conocidos de proteínas globulares incluyen la hemoglobina, la mioglobina y la insulina.

2. Proteínas fibrosas: A diferencia de las proteínas globulares, las proteínas fibrosas tienen una estructura alargada y fibrilar. Estas proteínas están compuestas por cadenas polipeptídicas que se organizan en una forma repetitiva y alineada, lo que les confiere una estructura fibrosa. Las proteínas fibrosas son menos solubles en agua y se encuentran principalmente en tejidos estructurales como el colágeno, la queratina y la elastina. Estas proteínas proporcionan soporte, resistencia y elasticidad a tejidos como la piel, el cabello y los tendones.

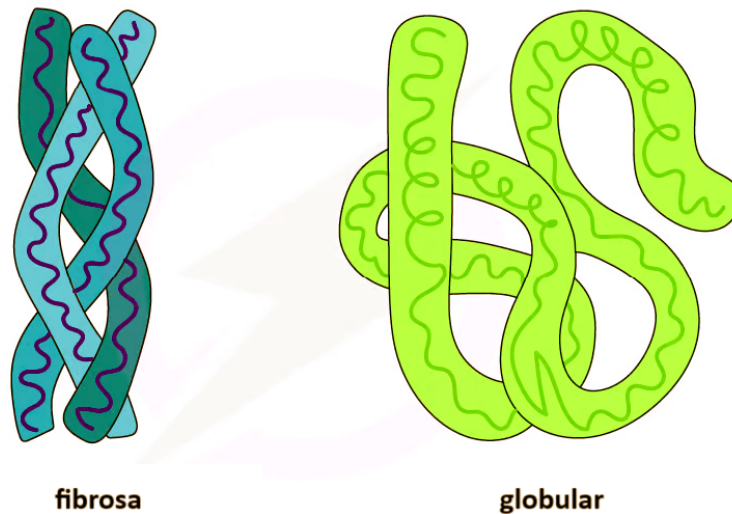


Figura 3.3: Esquema de proteínas fibrosas y globulares.

Ambos tipos de proteínas, globulares y fibrosas, cumplen funciones importantes en los organismos y su estructura terciaria determina sus propiedades físicas y químicas, así como sus interacciones con otras moléculas. Cabe destacar que esta clasificación es una simplificación, ya que algunas proteínas pueden tener características intermedias entre los dos tipos mencionados.

La estructura tridimensional de una proteína globular se compone de una o varias cadenas polipeptídicas, que están formadas por secuencias de aminoácidos. Estas cadenas se pliegan y se organizan en una estructura compacta debido a la interacción de diferentes fuerzas y enlaces químicos, como los puentes disulfuro, puentes de hidrógeno y fuerzas de Van der Waals.

Las proteínas globulares desempeñan diversos roles en los organismos, como enzimas, transportadores, receptores, anticuerpos, factores de coagulación, entre otros. Su estructura compacta les permite interactuar específicamente con otras moléculas, como sustratos, ligandos o proteínas complementarias, para llevar a cabo sus funciones biológicas. Además, la estructura globular proporciona estabilidad a la proteína y le permite realizar cambios conformacionales necesarios para su actividad.

Es importante destacar que la estructura tridimensional de una proteína globular es fundamental para su función biológica. Alteraciones en la estructura, ya sea por mutaciones

genéticas o por condiciones ambientales desfavorables, pueden afectar la capacidad de la proteína para llevar a cabo su función normal, lo que puede tener implicaciones en la salud y enfermedad.

En una proteína globular, los carbonos alfa (C_α) es el punto de referencia utilizado para definir la estructura tridimensional de una proteína, ya que se utiliza para determinar la ubicación espacial de los demás átomos en relación con él. Los carbonos alfa también son importantes en el estudio de la estructura y la dinámica de las proteínas, ya que su posición se utiliza para calcular distancias, ángulos y torsiones que ayudan a caracterizar y comprender la conformación de la proteína.

3.5. Introducción a la Bioinformática y Métodos Computacionales

En la actualidad la cantidad de información biológica está aumentando a un ritmo exponencial. Por ejemplo, hasta agosto de 2000, el repositorio GenBank contenía más de 8 millones de secuencias de ácidos nucleicos, mientras que la base de datos SWISS-PROT albergaba más de 88,000 secuencias de proteínas. Esta información se duplica aproximadamente cada 15 meses. Además, desde la divulgación del genoma de *Haemophilus influenzae*, se han hecho públicas secuencias completas de más de 40 organismos[48]. En conjunto con proyectos afines que investigan la expresión génica, las estructuras de proteínas y las interacciones entre los productos genéticos, se está generando una gran y variada cantidad de datos biológicos.

Debido al aumento considerable en la cantidad de bio-información, muchos desafíos en el ámbito de la biología se han transformado en desafíos relacionados con la informática. Gracias a la capacidad que tienen las computadoras para manejar grandes volúmenes de información y de modelar las dinámicas complejas que se observan en la naturaleza la aproximación informática de la biología resulta altamente conveniente. La bioinformática, que es el nombre de este enfoque se define generalmente como la aplicación de técnicas computacionales para comprender y organizar la información vinculada a las macromoléculas biológicas.

Los objetivos de la bioinformática son triples. En primer lugar, la bioinformática organiza

los datos de manera que permita a los investigadores acceder a la información existente y presentar nuevas entradas a medida que se producen, como por ejemplo el Banco de Datos de Proteínas para estructuras macromoleculares en 3D. Aunque la curación de datos es una tarea esencial, la información almacenada en estas bases de datos es prácticamente inútil hasta que se analice. Por lo tanto, el propósito de la bioinformática va mucho más allá del simple control de volumen. El segundo objetivo es desarrollar herramientas y recursos que ayuden en el análisis de datos. Por ejemplo, después de secuenciar una proteína en particular, es de interés compararla con secuencias previamente caracterizadas. Esto requiere más que una simple búsqueda en la base de datos. El tercer objetivo es utilizar estas herramientas para analizar los datos e interpretar los resultados de manera biológicamente significativa. Tradicionalmente, los estudios biológicos examinaban sistemas individuales en detalle y a menudo los comparaban con unos pocos relacionados. En bioinformática, también podemos realizar análisis globales de todos los datos disponibles con el objetivo de descubrir principios comunes que se apliquen en muchos sistemas y resaltar características únicas en algunos.

El Banco de Datos de Proteínas, PDB [48], ofrece un archivo primario de todas las estructuras 3D de macromoléculas como proteínas, ARN, ADN y varios complejos. La mayoría de las 214,121 estructuras (en el año de 2023) se resuelven mediante cristalografía de rayos X y RMN, pero también se incluyen algunos modelos teóricos. Los archivos de la Base de Datos de Proteínas (PDB) almacenan información sobre las estructuras tridimensionales de moléculas biológicas como proteínas y ácidos nucleicos. Contienen coordenadas atómicas, identificadores de aminoácidos/nucleótidos, detalles de estructuras secundarias, conformaciones, métodos de resolución, datos de autores y referencias, y anotaciones sobre interacciones y funciones biológicas. Los identificadores de aminoácidos son códigos de una letra que se utilizan para representar los diferentes tipos de aminoácidos en las secuencias de proteínas. Aquí hay algunos ejemplos de identificadores de aminoácidos comunes:

Es importante tener en cuenta que estos son solo algunos ejemplos y existen muchos otros aminoácidos y códigos de residuos utilizados en el archivo PDB, incluidos aminoácidos modificados y no estándar.

Dado que la información proporcionada en las entradas individuales del PDB puede ser difícil de organizar e interpretar, se erige una herramienta de invaluable utilidad: la simulación computacional. La simulación es una herramienta útil para analizar la estructura, propiedades y dinámica de una proteína, lo cual puede ser de gran relevancia en el estudio de su función y relaciones con otras moléculas en el contexto de la bioquímica y la biología

Aminoácido	Código de Residuo
Alanina (ALA)	A
Arginina (ARG)	R
Asparagina (ASN)	N
Aspartato (ASP)	D
Cisteína (CYS)	C
Glutamina (GLN)	Q
Glutamato (GLU)	E
Glicina (GLY)	G
Histidina (HIS)	H
Isoleucina (ILE)	I
Leucina (LEU)	L
Lisina (LYS)	K
Metionina (MET)	M
Fenilalanina (PHE)	F
Prolina (PRO)	P
Serina (SER)	S
Treonina (THR)	T
Triptófano (TRP)	W
Tirosina (TYR)	Y
Valina (VAL)	V

Tabla 3.1: Códigos de Residuo para Aminoácidos

estructural.

3.6. Construcción de redes y análisis de estructuras proteicas

El método computacional consiste en la organización y análisis de la información estructural de las proteínas con el propósito de interpretarlas como redes de residuos de proteínas (PRNs), utilizando un archivo en formato PDB. A continuación, se proporciona una breve descripción de la función de cada paso en el programa:

1. Lee el archivo PDB y lo convierte en un DataFrame: El programa utiliza la librería PandasPDB para leer el archivo PDB y convertir los datos en un DataFrame, una estructura de datos tabular que facilita la manipulación de la información.
2. Filtra los carbonos alfa de cada residuo: Se seleccionan únicamente los átomos de carbono alfa de la proteína y se toma el átomo de carbono alfa de cada residuo.

3. Filtra los residuos que se van a omitir: Se excluyen residuos específicos ubicados al inicio y al final de la proteína que no serán tenidos en cuenta en el análisis.
4. Obtiene el número de nodos: Se determina el número total de nodos (residuos) a representar en la red.
5. Calcula la matriz de distancia entre los carbonos alfa: Se calcula la matriz de distancia euclidiana entre los átomos de carbono alfa seleccionados.
6. Encuentra las parejas de carbonos alfa cercanos: Se identifican las parejas de átomos de carbono alfa cuya distancia es menor a 8 Ångstroms, lo cual indica una proximidad espacial.
7. Crea una matriz de adyacencia: Se crea una matriz de ceros del tamaño adecuado para representar la conectividad entre los nodos.
8. Llena la matriz con las conexiones: Se asignan valores de 1 en la matriz de adyacencia para indicar las conexiones entre los nodos correspondientes a las parejas de carbonos alfa encontradas.
9. Crea un objeto Graph y agrega nodos y aristas: Se utiliza la librería NetworkX para crear un objeto Graph y agregar los nodos y las aristas a partir de la matriz de adyacencia.
10. Imprime el resultado: Se muestra en pantalla el número de nodos de la red generada.
11. Retorna el objeto Graph: Se devuelve el objeto Graph que representa la red de interacciones entre los nodos.

Las siguientes partes del programa consisten en una función “centralidad”: Esta función recibe un objeto Graph (G) como parámetro. Realiza una serie de cálculos para obtener varias medidas de centralidad y propiedades estructurales de la red. Algunas de las medidas calculadas incluyen la densidad de la red, el grado de heterogeneidad, la centralidad de eigenvector, la longitud media del camino, el coeficiente de agrupamiento, el coeficiente de asortatividad, la centralidad de intermediación, la matriz de comunicabilidad y el ángulo de comunicabilidad.

Estas funciones son herramientas útiles para analizar las propiedades estructurales y topológicas de redes representadas como grafos, lo que puede brindar información importante sobre la organización y el funcionamiento de sistemas complejos en diversos campos como la biología, la sociología y la informática, entre otros.

Capítulo 4

Estructura y propiedades de redes de residuos de proteínas

El análisis de las redes de residuos de proteínas (PRNs) puede ayudar a comprender la estabilidad estructural de la proteína, las rutas de transferencia de información y la influencia de mutaciones o modificaciones en la función proteica. La representación en forma de red proporcionan una visión detallada de las interacciones internas dentro de una proteína y pueden revelar patrones de conectividad y comunicación entre los residuos. En general las PRNs se utilizan en estudios de bioquímica, bioinformática y diseño de proteínas, entre otros campos. Proporcionan una representación visual y cuantitativa de las interacciones entre los residuos de aminoácidos, lo que ayuda a desentrañar los mecanismos moleculares y las propiedades emergentes de las proteínas.

Estas redes se construyen utilizando las coordenadas cartesianas de la proteína reportada en el *Protein Data Bank* [49] o PDB. Los nodos de estas redes representan los aminoácidos de la proteína, centrados en sus átomos de C_α . Para conectar los nodos de la red se considera un radio de corte r_C , que representa un límite superior para la separación entre dos residuos en contacto. La distancia entre dos residuos se mide tomando la distancia entre los átomos de C_α de ambos residuos. Entonces, cuando la distancia es menor a r_C , se considera que ambos residuos interactúan. Bajo este criterios se seleccionan los nodos correspondientes en la red residual que estarán conectados.

Los elementos de la matriz de adyacencia de la red de residuos se obtienen de:

$$A_{ij} = \begin{cases} H(r_C - r_{ij}) & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases} \quad (4.1)$$

donde $H(x > 0) = 1$ y $H(x \leq 0) = 0$. Entonces, una proteína está representada por el grafo $G = (V, E)$, donde V representa el conjunto de aminoácidos y E representa el conjunto de interacciones entre ellos. Las redes de residuos aquí estudiadas se construyen utilizando un radio de corte de 8 Å. Atilgan et al.[9] han verificado en sus conclusiones generales sobre la topología de estas redes que no son afectadas cuando se usa un valor de corte de entre 7 y 8.5 Å.

La longitud de las proteasas de la proteína M^{pro} del SARS CoV-2 es de 306 aminoácidos. Sin embargo hay varias estructuras que solo están resueltas para los aminoácidos 3-300, por lo tanto para que se tenga una homogeneidad en el estudio se considera solamente desde el residuo 3 hasta el residuo 300. Esto no altera el análisis, ya que los dos extremos de la proteasa están desordenados y no participan en interacciones importantes [41].

La información estructural utilizada para construir la red de residuos, a partir de las coordenadas cartesianas de los componentes de una proteína, se obtiene del banco de datos PDB [50]: El banco de datos PDB es una base de datos de la estructura tridimensional de residuos de proteínas y ácidos nucleicos. Estos datos se obtienen mediante cristalografía de rayos X o resonancia magnética nuclear que son proporcionados por biólogos y bioquímicos de todo el mundo.

En general se utilizaron proteínas en su forma apo, que se refiere a la forma de una proteína que carece de su ligando o molécula cofactor. El término “apo” proviene del griego y significa “sin”. Cuando una proteína se encuentra en su forma apo, significa que no está unida a ningún ligando o cofactor específico.

La representación gráfica de una red obtenida de este proceso se muestra en la figura 4.1. El estudio de las propiedades estructurales de la red (primera categoría de medidas de centralidad descritas en la sección 2.2) es llevado a cabo mediante la paquetería de Python NetworkX, y los resultados son mostrados en la tabla 4.1.

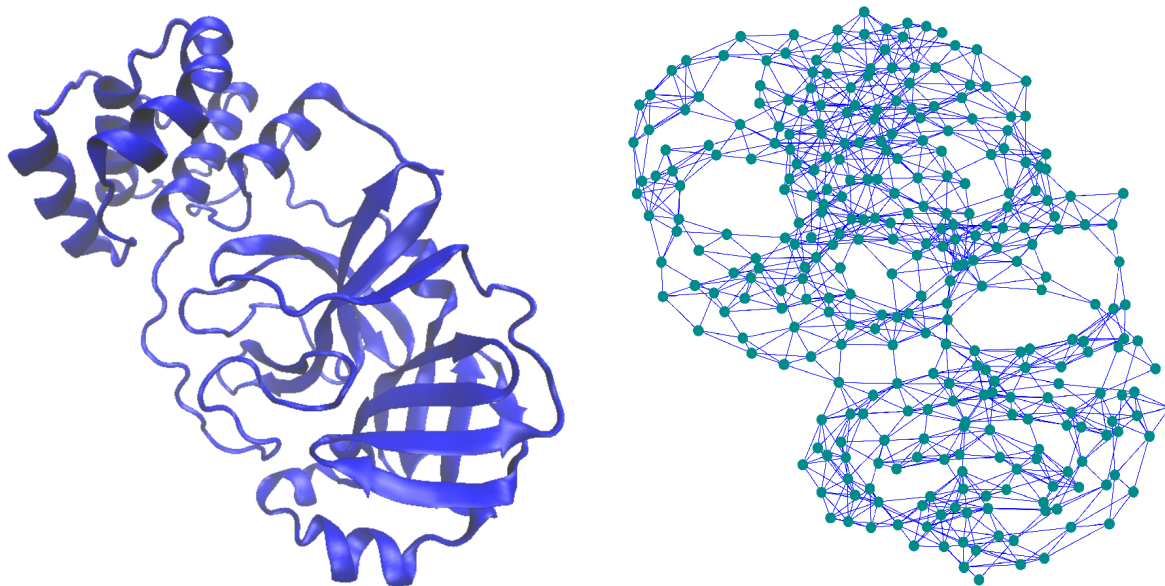


Figura 4.1: Representación (izquierda) de la proteína M^{pro} del SARS CoV-2 (PDB = 6Y2E) y la red correspondiente al residuo proteico (derecha), donde los nodos son aminoácidos conectados a través de vértices que representan la interacción entre átomos de C_α .

Con base al número de nodos y al grado promedio $\langle k \rangle$ (redondeado a 10 por cuestiones de limitaciones del método constructor de redes de Networkx) es posible crear una red regular con $\langle k \rangle = 10$, una red de Erdos-Renyi (ER) (modelo $G(N, L)$), y una red de mundo pequeño de Watts-Strogatz (WS) con $\langle k \rangle = 10$ y $p = 0.019$. Estas estructuras sirven como primer acercamiento a las redes de referencia para el estudio. En la tabla 4.1 se muestran las propiedades estructurales resultantes de cada red.

Propiedades de la red	M^{pro}	PV	WS	ER
Nodos N	298	298	298	298
Conexiones L	1492	1490	1490	1490
Agrupamiento promedio $\langle C \rangle$	0.530	0.667	0.631	0.031
Longitud de camino más corto promedio $\langle D \rangle$	5.244	15.353	5.232	2.710
Grado promedio $\langle k \rangle$	10.012	10	10	10

Tabla 4.1: Propiedades estructurales promedio de 75 redes de residuos de la proteína M^{pro} del SARS CoV-2 ($r_C = 8.16$ Å), la red de primeros vecinos (PV), la red de Watts-Strogatz (WS) y la red de Erdos-Renyi (ER). Para las redes de WS y ER se promedia sobre 1000 redes.

También fue posible obtener la gráfica de distribución de grado de las redes, presentada en la figura 4.2.

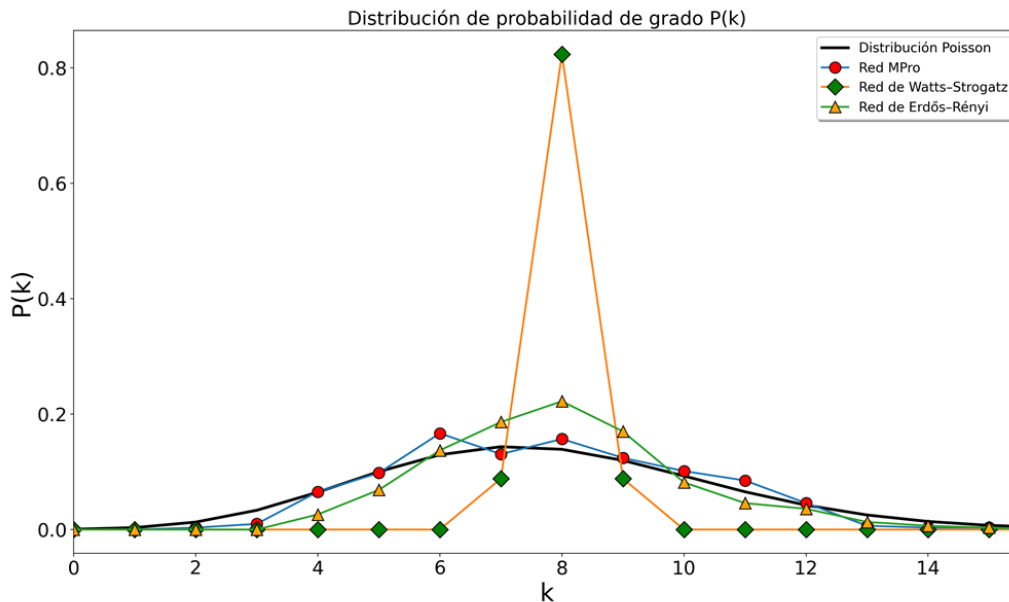


Figura 4.2: Distribución de probabilidad grado promedio de 75 redes de proteína M^{pro} del SARS CoV-2 (75 archivos PDB), 1000 redes de WS, 1000 redes de ER y distribución de Poisson.

Al comparar los datos de la tabla 4.1 en términos de su coeficiente de agrupamiento promedio $\langle C \rangle$ y longitud de camino promedio $\langle D \rangle$, se observa que la red que muestra mayor similitud a la red de residuos de proteína es la red de WS. Sin embargo, al examinar la distribución de probabilidad de grado de las redes de ER y WS, representadas en la figura 4.2, en la cual hay poca similitud con la distribución de probabilidad de grado de la red de residuos de proteína. Sorprendentemente, la distribución de probabilidad de grado que se asemeja más a la de la red de proteína es la de la red de ER, a pesar de las similitudes estructurales encontradas entre la red de proteína y la red de WS. Adicionalmente la distribución de Poisson se ajusta casi perfectamente a la distribución de probabilidad de grado de la red de residuos de proteína.

4.1. Resultados de M^{pro} de SARS CoV-1

Al igual que Estrada en su artículo (mencionar esto en algún capítulo anterior) [41] estudiaremos las cuatro estructuras de la proteasa principal del SARS CoV-1 presentadas en dicho artículo y calculamos las dos centralidades de subgrafo (de corto y largo alcance) definidas en el capítulo 2.2.

A continuación a modo de muestra una representación de la matriz de adyacencia de la red de residuos correspondiente a la proteína PDB = 2H2Z del SARS CoV-1.

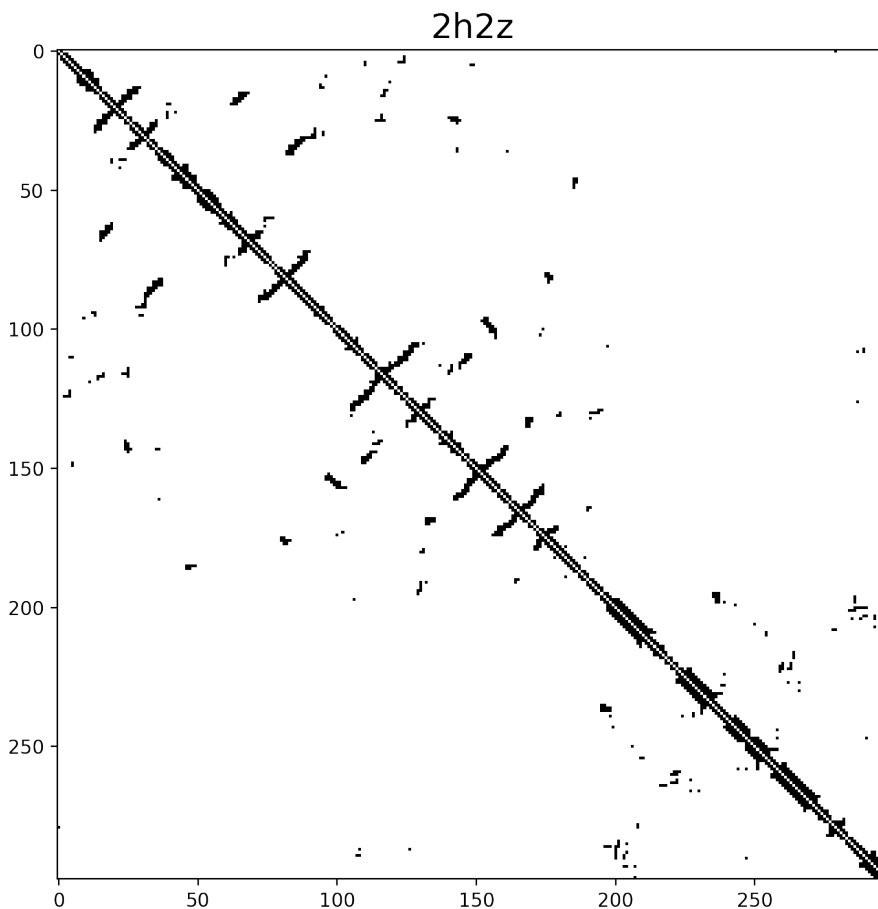


Figura 4.3: Representación de la matriz de adyacencia de la proteína M^{pro} del SARS CoV-1 (PDB = 2H2Z), donde los nodos son aminoácidos conectados a través de vértices que representan la interacción entre átomos de C_α .

Gracias a esta representación matricial de la red, es fácil encontrar la centralidad de subgrafo y de centralidad de subgrafo de LR (largo alcance) normalizadas de las cuatro proteínas M^{pro} del SARS CoV-1 estudiadas que se muestran a continuación.

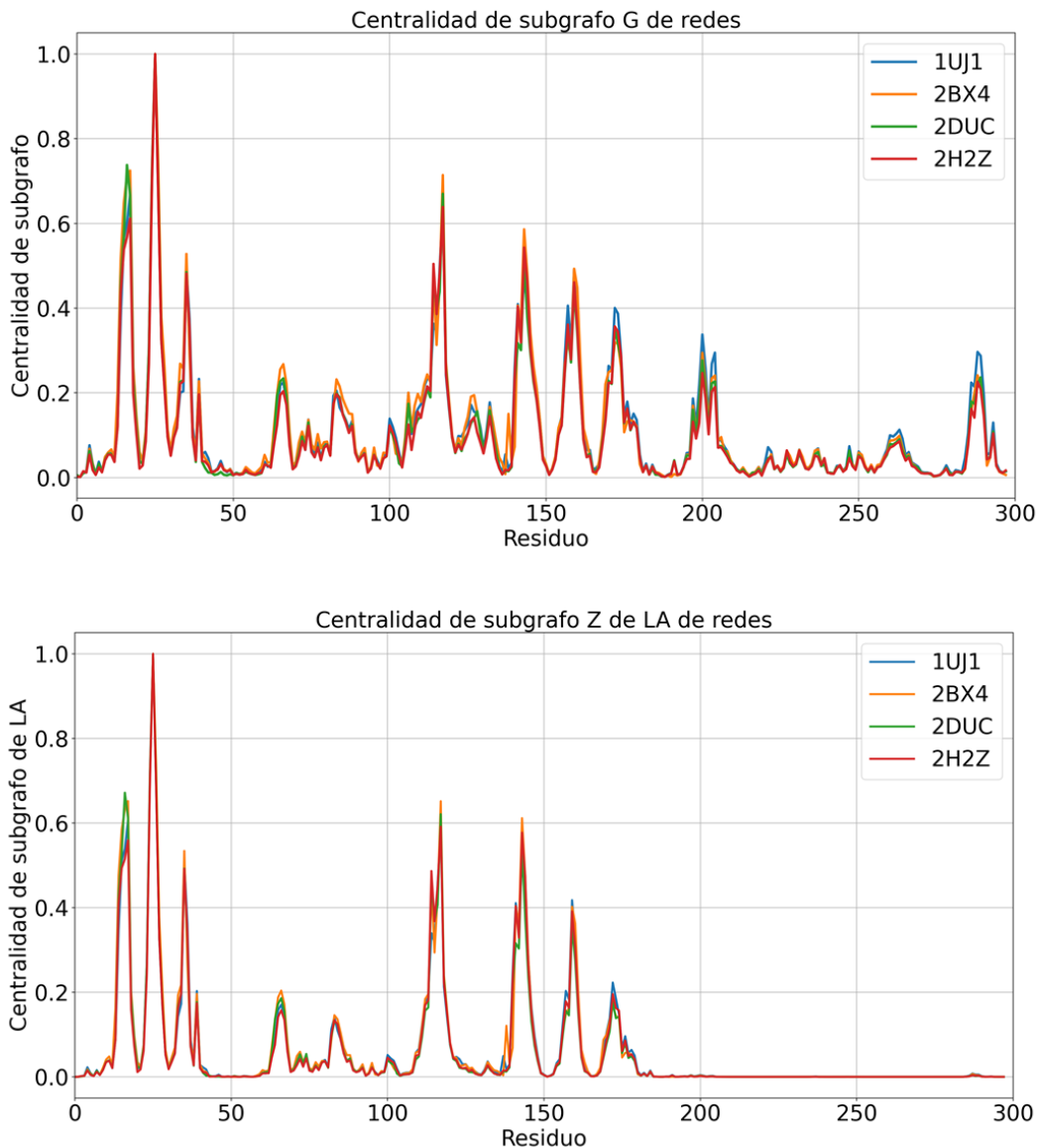


Figura 4.4: Gráfico de la centralidad de subgrafo (arriba) y de centralidad de subgrafo de LR (abajo) normalizadas de la proteína M^{pro} de SARS CoV-1 para cada archivo PDB (1UJ1, 2BX4, 2DUC y 2H2Z)

También se obtiene el promedio de ambas centralidades de las cuatro estructuras estudiadas. Donde se observan las veinte mayores centralidades de cada categoría y los residuos

pertenecientes al sitio catalítico de las proteínas estudiadas por Atilgan *et al.* [9].

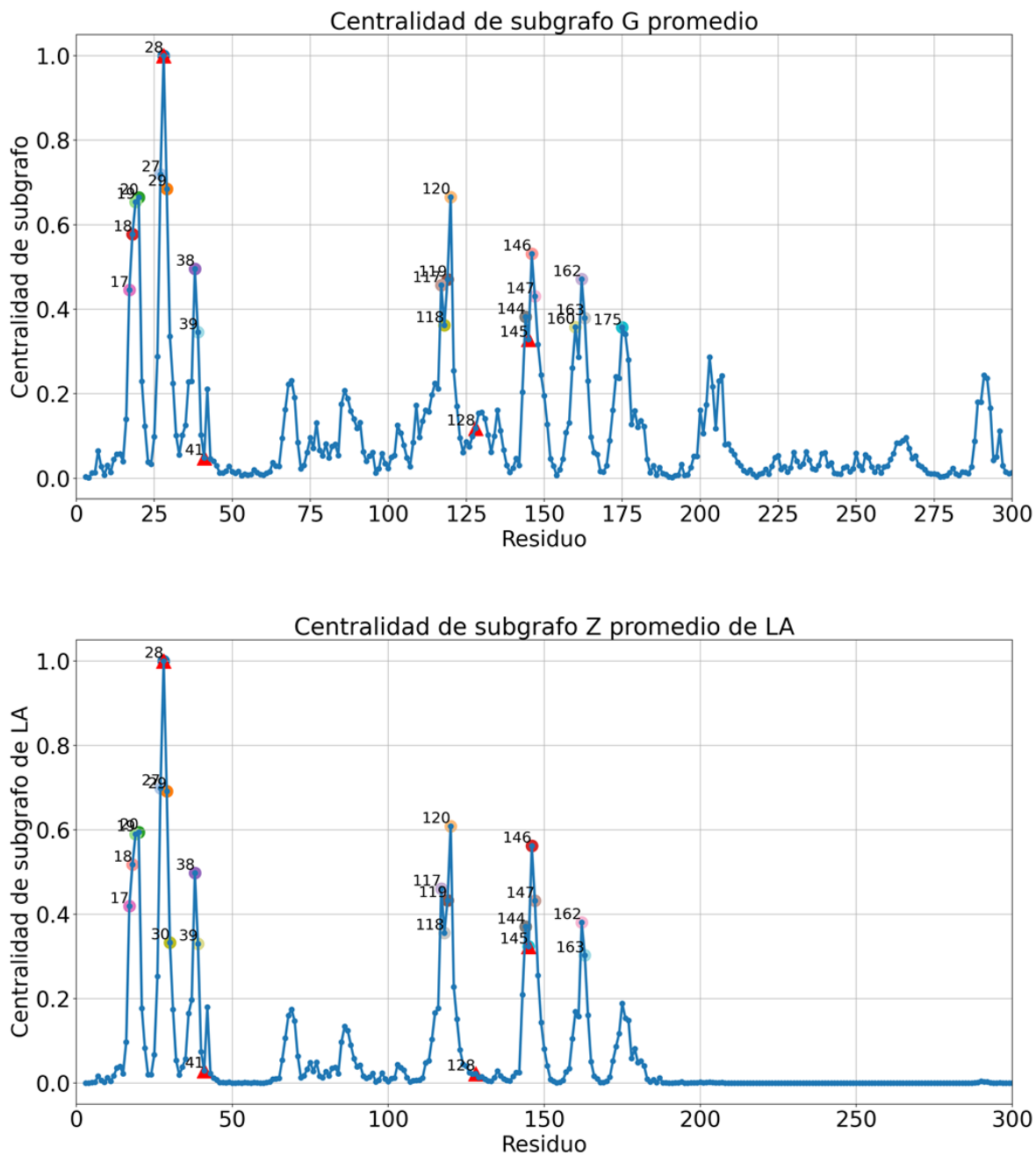


Figura 4.5: Gráfico de la centralidad de subgrafo (arriba) y de centralidad de subgrafo de LR (abajo) normalizadas promedio de la proteína M^{pro} del SARS CoV-1 de los PDB (1UJ1, 2BX4, 2DUC y 2H2Z). Y residuos pertenecientes al sitio catalítico de la proteína

De las anteriores gráficas se destaca que dos residuos del sitio activo de la proteína corres-

ponden con los nodos de mayor centralidad de subgrafo de la red.

4.2. Resultados de M^{pro} de SARS CoV-2

En correspondencia con el artículo de Estrada continuamos con el estudio de las tres proteasas de SARS CoV-2 reportadas. A continuación una representación de la matriz de adyacencia del grafo de la proteína PDB = 6Y2E.

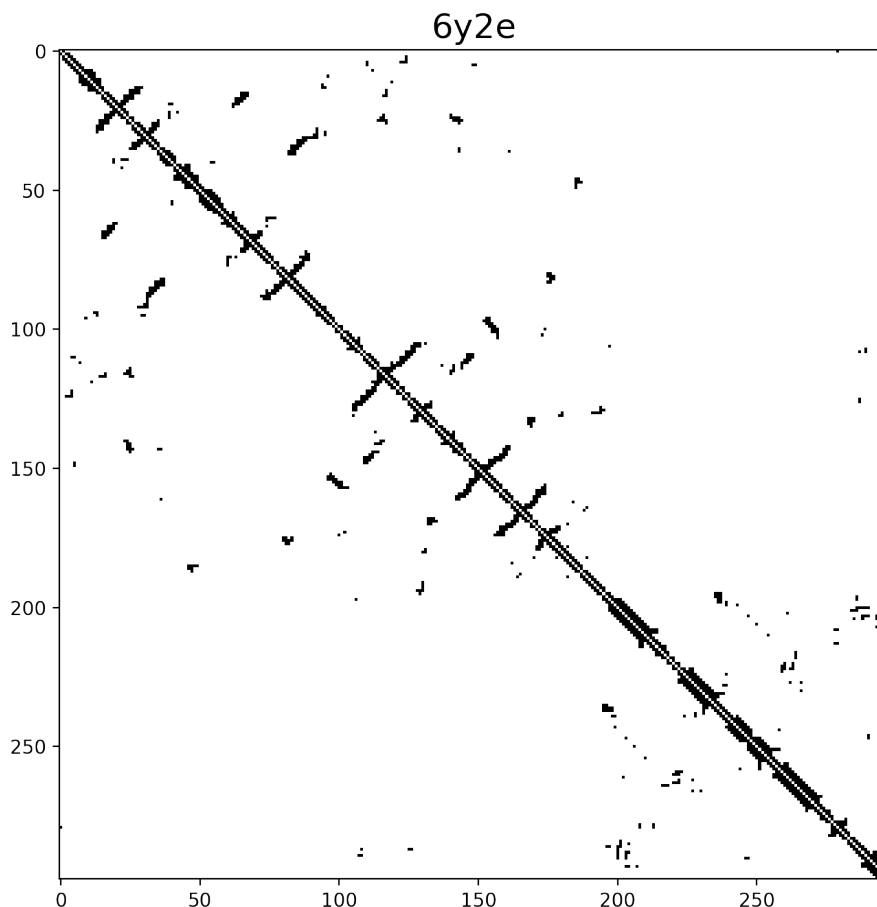


Figura 4.6: Representación de la matriz de adyacencia de la proteína M^{pro} del SARS CoV-2 (PDB = 6Y2E), donde los nodos son aminoácidos conectados a través de vértices que representan la interacción entre átomos de C_{α} .

La centralidad de subgrafo y de centralidad de subgrafo de LR (largo alcance) de las tres

proteínas M^{pro} del SARS CoV-2 estudiadas por Estrada.

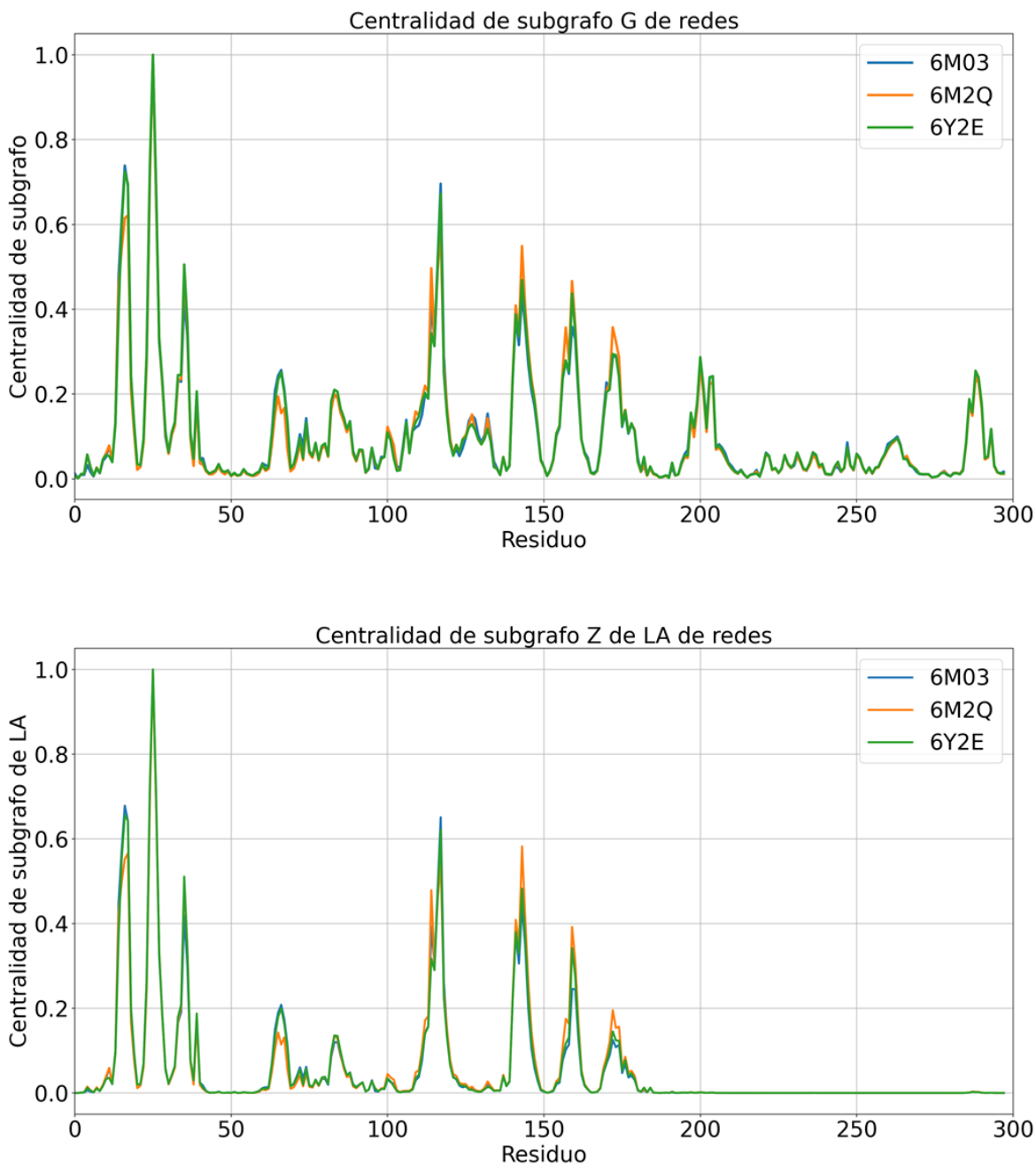


Figura 4.7: Gráfico de la centralidad de subgrafo (arriba) y de centralidad de subgrafo de LR normalizadas de la proteína M^{pro} del SARS CoV-2 para cada archivo PDB (6M03, 6M2Q y 2Y2E)

El promedio de las centralidades de las anteriores estructuras, indicando las veinte mayo-

res centralidades de cada categoría y los residuos pertenecientes al sitio catalítico de las proteínas estudiadas por Atilgan *et al.* [9].

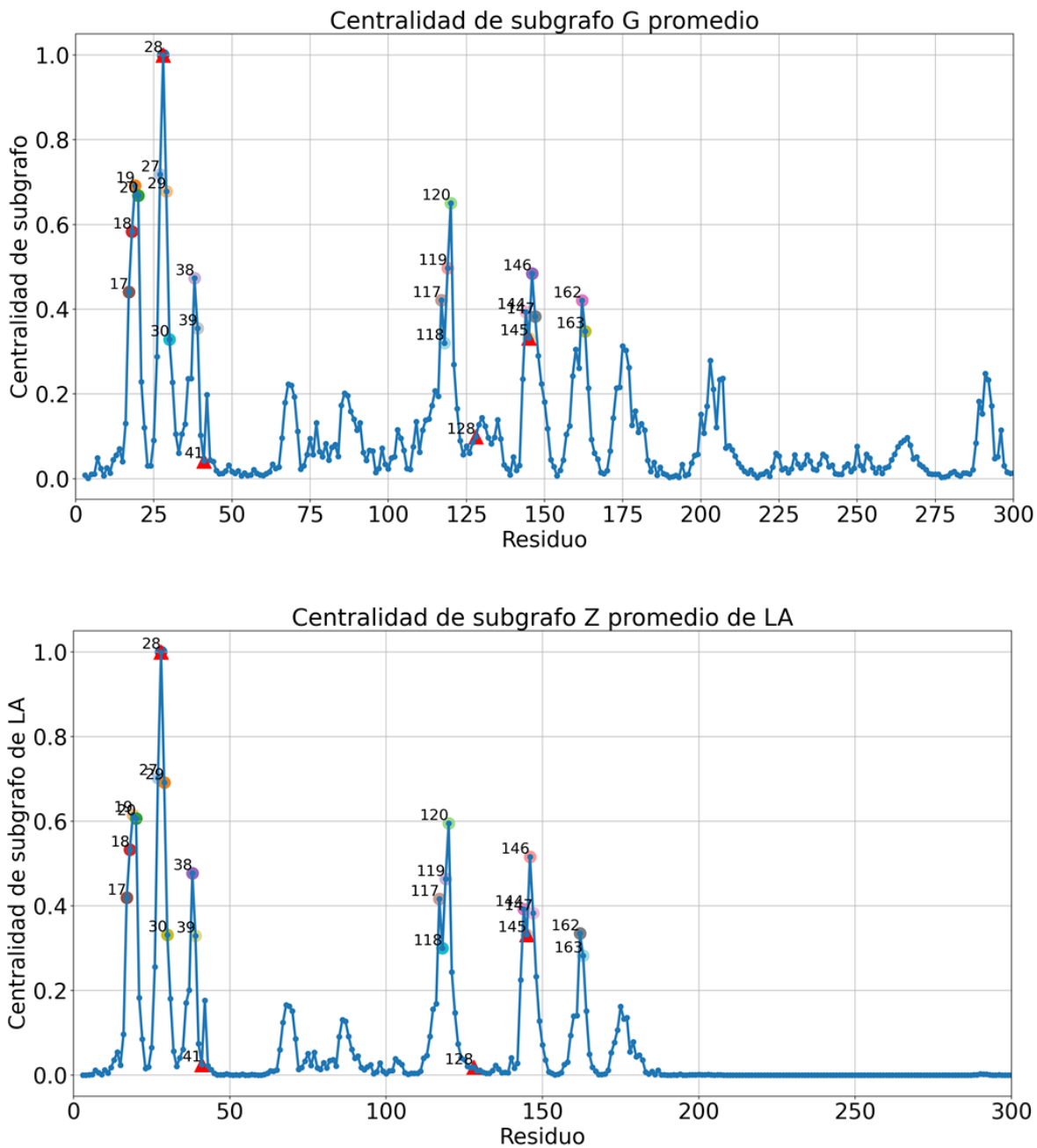


Figura 4.8: Gráfico de la centralidad de subgrafo (arriba) y de centralidad de subgrafo de LR normalizadas promedio de la proteína M^{pro} del SARS CoV-2 de los PDB (6M03, 6M2Q y 2Y2E). Y residuos pertenecientes al sitio catalítico de la proteína

Una característica notable de la Gráfica 4.8 es que el sitio catalítico del aminoácido Cys 145 de Mpro del SARS CoV-2 coincide con uno de los residuos con la mayor centralidad de subgrafo de LR de la proteasa de SARS CoV-2.

Haciendo una comparación de la centralidad de subgrafo y de la centralidad de subgrafo de LR promedio de la proteína M^{pro} del SARS CoV-2 (en la Figura 4.9), podemos distinguir que los aminoácidos del 0 al 180 coinciden respecto a sus máximos de centralidad, mientras que para la zona del aminoácido 180 a 300 hay una zona plana para la centralidad de subgrafo de LR.

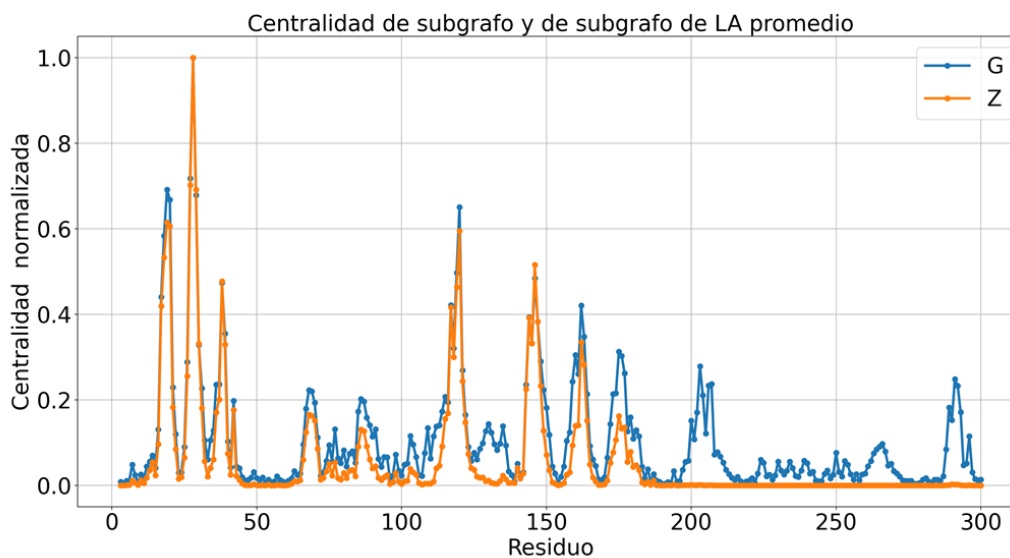


Figura 4.9: Gráfico de comparación de la centralidad de subgrafo y de centralidad de subgrafo de LR normalizadas promedio de la proteína M^{pro} del SARS CoV-2 de los PDB (1UJ1, 2BX4, 2DUC y 2H2Z)

En esta sección también replicamos los hallazgos reportados por Ernesto Estrada en su artículo [26]. Para ello, realizamos el cálculo de todas las medidas topológicas definidas en capítulo 2. Específicamente, consideramos las cuatro estructuras de la proteasa principal del SARS CoV-1 y las tres estructuras correspondientes al SARS CoV-2 vistas anteriormente. A continuación, presentamos los resultados promedio y la desviación estándar de estas medidas para ambos grupos de proteasas en la tabla 4.2

Métricas de la red	SARS CoV-1	SARS CoV-2	$\Delta(\%)$	U-stat
δ	0.0260	0.0262	0.906	0.00482
ρ	0.0162	0.0165	1.601	0.02609
$\langle L \rangle$	6.35	6.32	0.457	0.00867
$\langle C \rangle$	0.541	0.538	0.468	0.000592
r	0.392	0.392	0.174	0.00246
$\langle BC \rangle$	795.01	790.70	0.543	0.0103
$\langle EC \rangle$	0.0325	0.0318	2.324	0.0108
$\langle SC \rangle$	171.74	185.62	8.077	0.0077
$\langle G_{pq} \rangle$	22.35	24.93	11.560	0.0114
$\langle \theta \rangle$	82.31	82.03	0.340	0.00008
$\langle Z_{pp} \rangle$	1.044×10^{18}	4.697×10^{18}	349.904	0.231
$\langle Z_{pq} \rangle$	3.225×10^{17}	1.391×10^{18}	331.318	0.246

Tabla 4.2: Valores promedio de las propiedades topológicas globales de la M^{pro} del SARS CoV-1 (2H2Z, 1UJ1, 2DUC, 2BX4) y del SARS CoV-2 (6M03, 6Y2E, 6M2Q). También se proporciona la diferencia relativa entre ellos, expresada como porcentaje del cambio relativo al SARS CoV-1, y los valores de p de la prueba U de Mann-Whitney. La negrita denota los valores con la diferencia más significativa en relación con el SARS CoV-1.

Al analizar los datos de la tabla 4.2, se observa que la mayoría de las características topológicas de las redes de proteínas de ambos grupos de proteasas presentan similitudes significativas. Las diferencias relativas entre las propiedades analizadas son mínimas. Para evaluar la significancia de estas diferencias, se utiliza el test de Mann-Whitney U, una medida estadística propuesta para el análisis de medidas de redes, especialmente en redes de proteínas. Los valores obtenidos de este análisis indican que no existen diferencias significativas entre los dos grupos de proteasas en ninguna de las medidas evaluadas excepto para los valores de Z .

4.3. Resultados promedio para M^{pro}

Estos resultados respaldan y confirman los hallazgos previos de Ernesto Estrada, lo que indica que nuestra replicación es consistente con su trabajo. Esta conclusión también se confirma en la Figura 4.10, donde se presenta la lista de aminoácidos con los valores más altos de centralidad de subgrafo LR promedio de 75 proteasas M^{pro} del SARS-CoV-2. Los resultados obtenidos en esta tabla coinciden con los reportados por Estrada.

En las siguientes figuras, se presentan las centralidades de subgrafo y de subgrafo de LR para las redes de residuos de proteínas de C_α de las proteasas M^{pro} del SARS-CoV-2. Estas centralidades han sido promediadas a partir de 75 estructuras del banco de datos PDB. En las gráficas de centralidad de largo alcance de las redes de C_α , se puede observar que los residuos de mayor centralidad se encuentran en extremos opuestos de la red. Esta observación se muestra en detalle en la Figura 4.12.

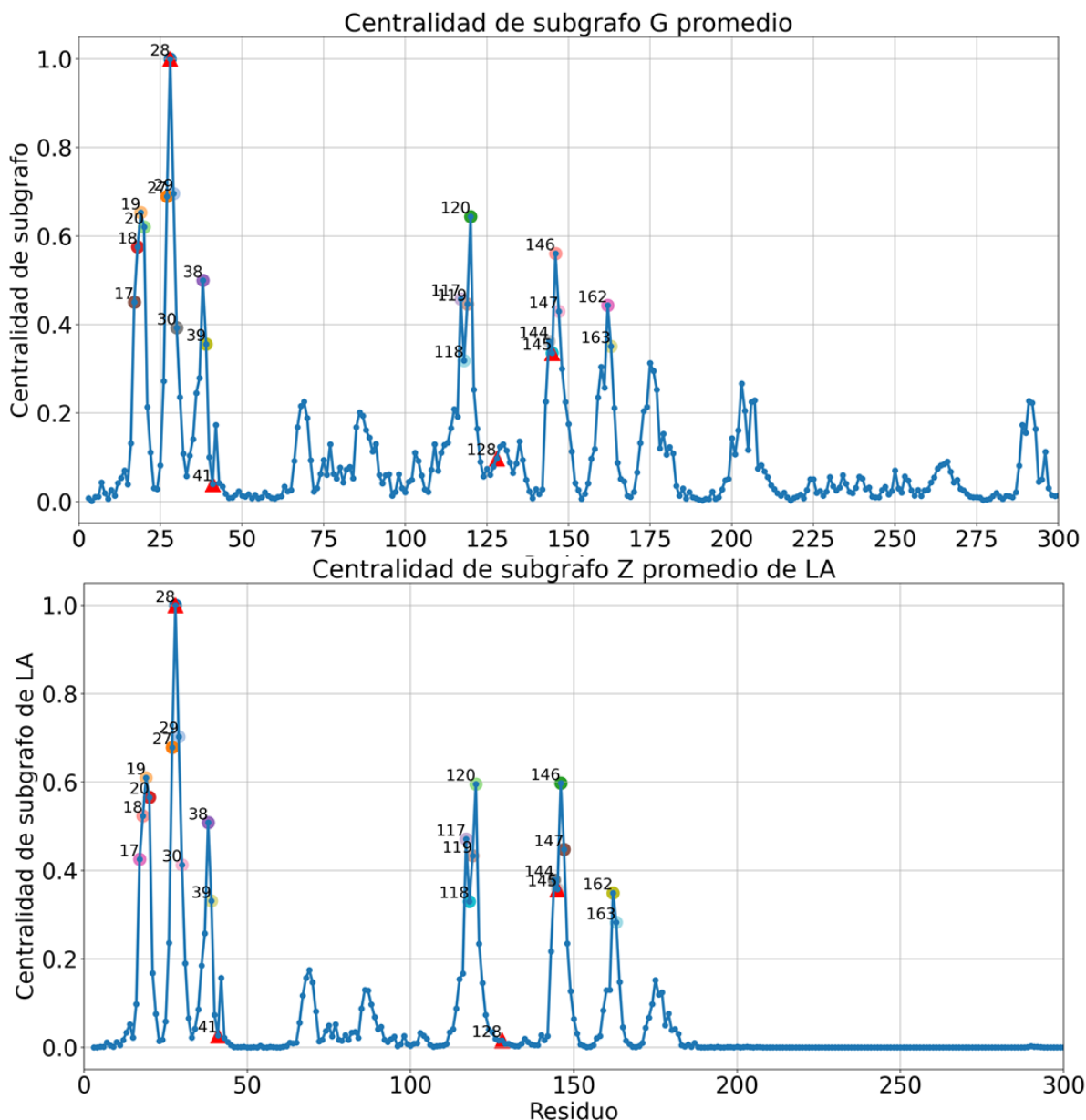


Figura 4.10: Gráfico de la centralidad de subgrafo promedio (arriba) y de centralidad de subgrafo de LR normalizadas promedio (abajo) de 75 redes de residuos de aminoácidos de C_α de la proteína M^{pro} del SARS CoV-2 extraídos del banco PDB.

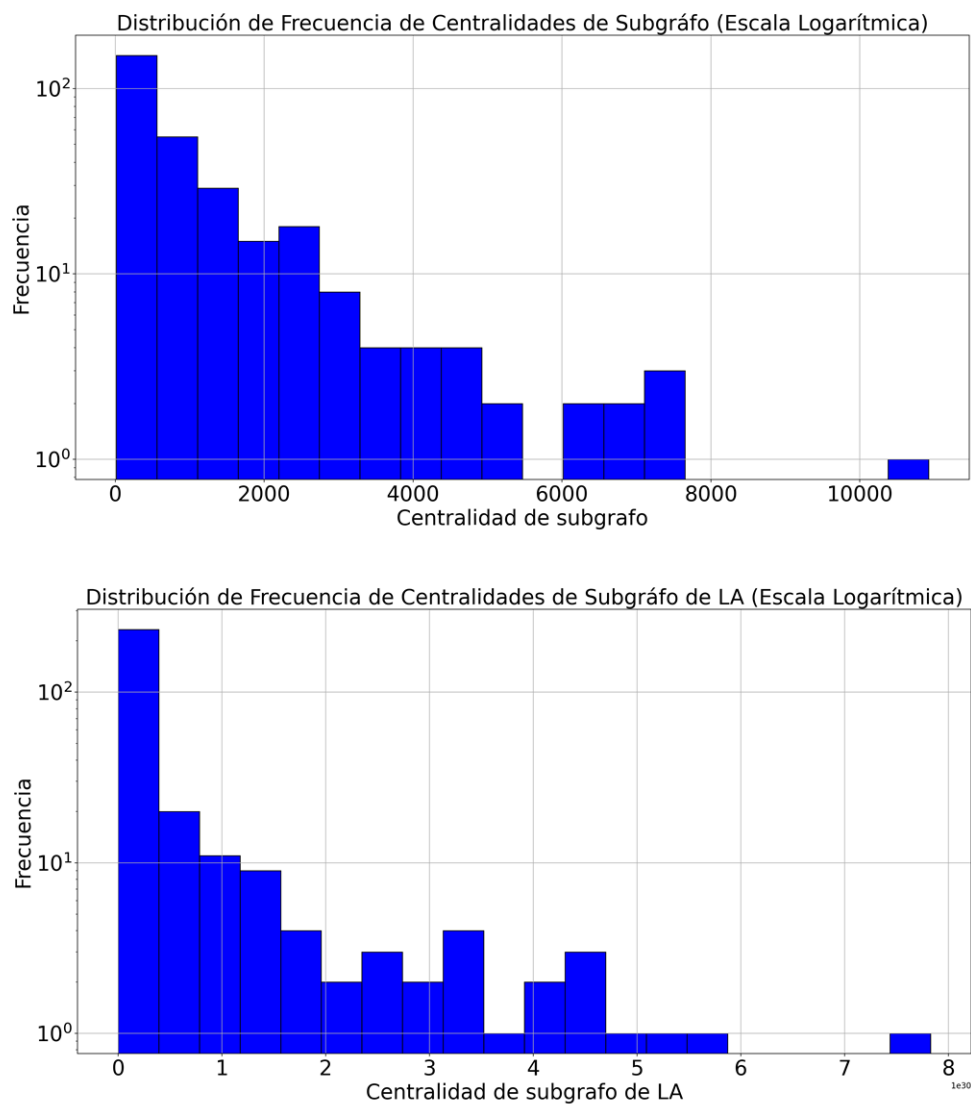


Figura 4.11: Gráficas de distribución de frecuencia de la centralidad de subgráfo promedio (arriba) y centralidad de subgráfo de largo alcance promedio (abajo) de 75 redes de residuos de aminoácidos de C_α de la proteína M^{pro} del SARS CoV-2 extraídos del banco PDB..

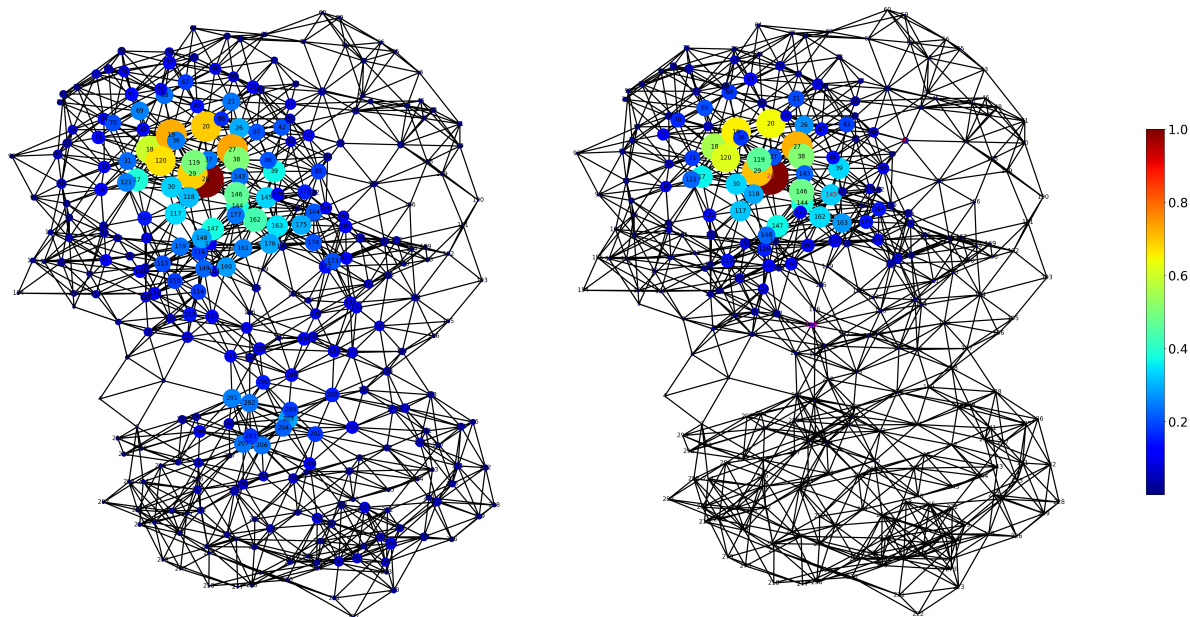


Figura 4.12: Ilustración de las centralidades de subgrafo (izquierda) y de subgrafo de LR (derecha) de la red de proteína de C_α de M^{pro} del SARS CoV-2 (PDB = 6Y2E). El tamaño de los nodos es proporcional a la centralidad correspondiente, normalizada al valor más grande en la proteasa analizada. Los colores también corresponden a los mismos valores en el código de colores "jet", con rojo para valores más altos y azul para valores más bajos.

Para ilustrar las distribuciones de los aminoácidos de mayor centralidad en las estructuras tridimensionales de las proteasas, se ha seleccionado la estructura 6Y2E del SARS-CoV-2. Es importante destacar que este proceso se ha llevado a cabo para las redes de residuos de C_α . Las estructuras se muestran en las Figuras 4.12, respectivamente. La primera observación notable sobre los aminoácidos con la mayor centralidad de subgrafo de LR es que todos están ubicados muy cerca entre sí en el espacio tridimensional en el sitio activo de la proteasa. Por otro lado, la centralidad del subgrafo se distribuyen de una manera más uniforme en toda la estructura.

También es posible realizar una gráfica de las centralidades descritas en la sección 2 (grado, cercanía, intermediación y eigenvector) de las redes de residuos de C_α en la Figura 4.13. Se observa una gran similitud entre las gráficas de centralidades de eigenvector y las gráficas de centralidad de subgrafo de LR anteriormente descritas. Sin embargo, en contraste, las gráficas de centralidad de grado, cercanía e intermediación no muestran una clara correspondencia entre sus máximos.

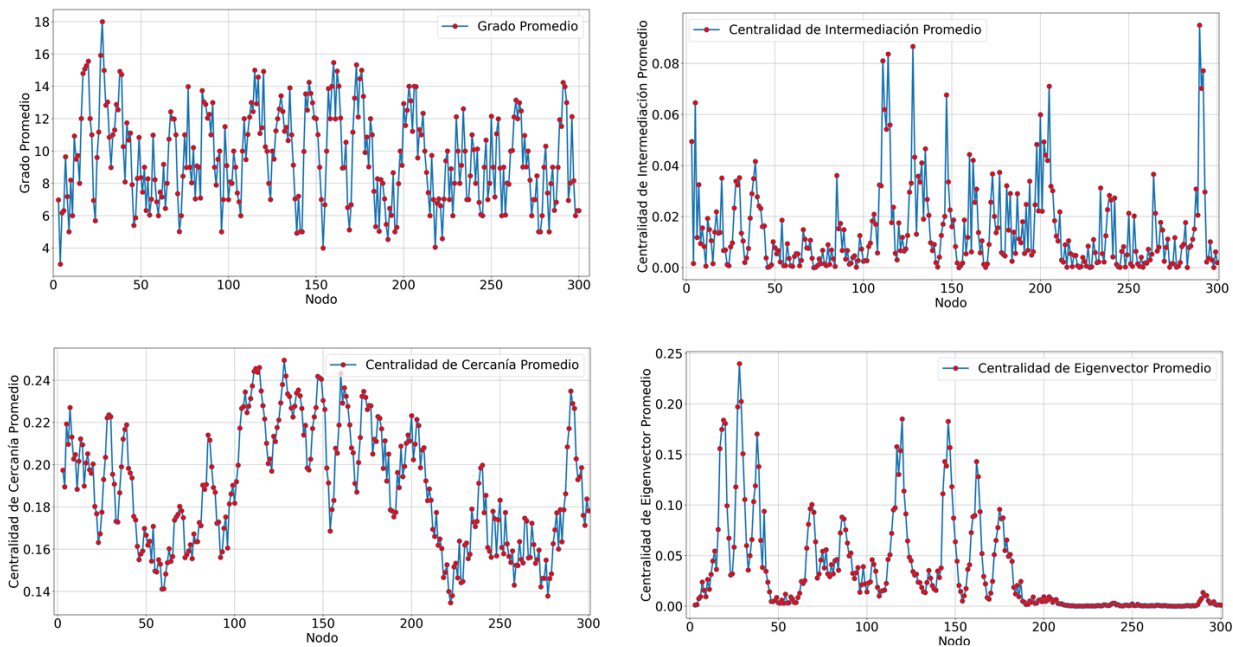


Figura 4.13: Gráficas de centralidad de grado, centralidad de cercanía, centralidad de intermediación y centralidad de eigen vector de los residuos de proteína de C_α en 75 redes de M^{pro} del SARS-CoV-2.

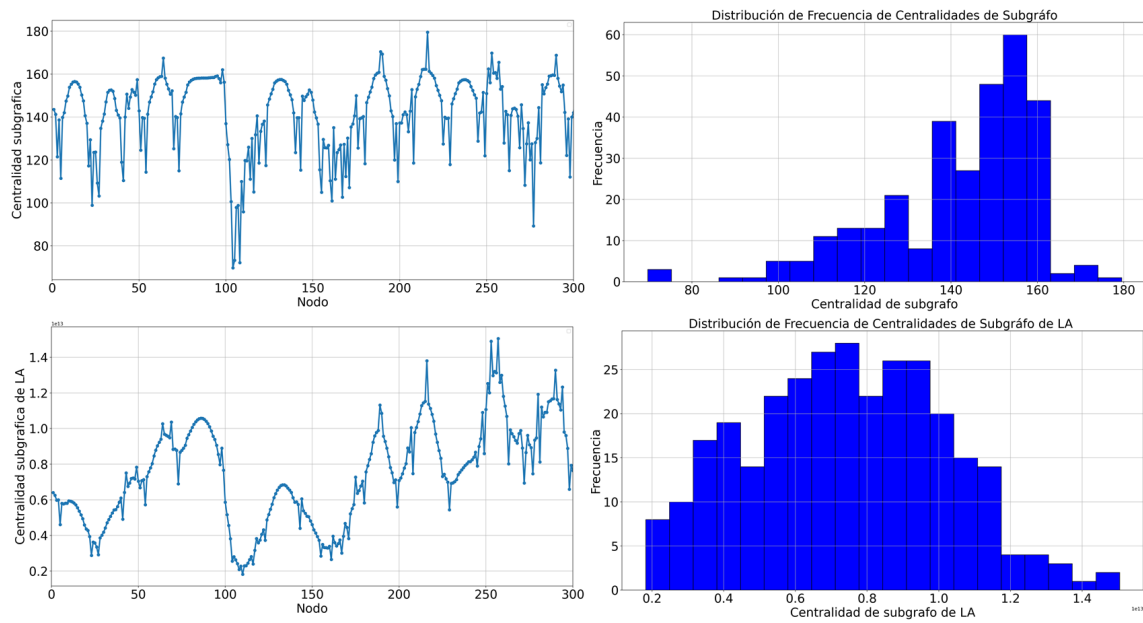


Figura 4.14: En orden descendente, gráficas de centralidad de subgráfo y centralidad de subgráfo de largo alcance de una red de WS (izquierda) y su distribución de frecuencia respectiva (derecha).

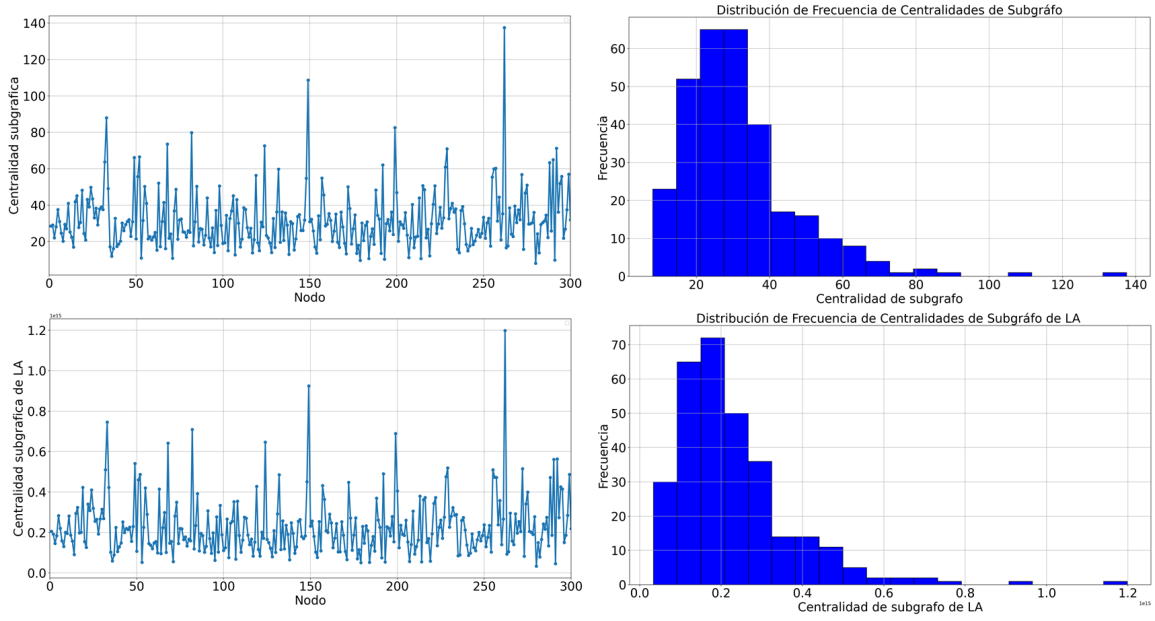


Figura 4.15: En orden descendente, gráficas de centralidad de subgrafo y centralidad de subgrafo de largo alcance de una red de ER (izquierda) y su distribución de frecuencia respectiva (derecha).

Capítulo 5

Propagación de Información en redes mediante Modelos de Mapeo Acoplado

El modelo de mapeos acoplados proporciona una descripción de la dinámica de la propagación y sincronización de información en la red de residuos de proteína. Para implementar este modelo de manera efectiva, utilizamos Python y la programación orientada a objetos (POO), que permite crear objetos con propiedades dinámicas y que pueden interactuar. Definimos un objeto llamado **nodo**, que representa un nodo en una red y tiene tres atributos:

- **índice:** Representa el número de residuo de la proteína que representa el nodo en la red.
- **conexiones:** Es una lista que almacena los índices de los nodos a los que cada nodo está conectado.
- **estado:** Es una lista que guarda los estados del nodo en diferentes pasos de tiempo. Estos estados representan una magnitud física o estado de información que evoluciona en el tiempo. Inicialmente, el estado es aleatorio y se actualiza con cada paso de tiempo del sistema.

Esta estrategia proporciona una buena forma de representación de una red, permitiendo agregar conexiones, asignar y obtener estados en diferentes pasos de tiempo para cada nodo de la red.

El estudio de la dinámica y la propagación de información en proteínas puede ser abordado por medio de los CML. La estrategia utilizada consiste en la simulación numérica computacional de modelos de CML sobre la red y su ejecución sucesiva en pasos de tiempo

discreto. De esta manera se puede obtener la evolución temporal de estados de cada sitio de la red y las trayectorias de sus estados en función de sus parámetros de acoplamiento. Para este estudio se emplean dos modelos de mapeo acoplado; el modelo básico y el modelo dinámico (DCML).

El modelo básico de CML de N nodos se puede describir de la siguiente manera:

$$x_i(t+1) = f_i(t) + \epsilon \left(\sum_{j=1, j \neq i}^N \frac{A_{ij} f_j(t)}{k(i)} - f_i(t) \right) \quad (5.1)$$

donde $x_i(t)$ es la variable de estado del i -ésimo nodo en el t -ésimo paso de tiempo, $i = 1, 2, \dots, N$, $k(i)$ es el grado del nodo i , y ϵ , donde $0 < \epsilon < 1$ es el parámetro de acoplamiento. En cuanto mayor es el valor de ϵ , más fuerte es la influencia de los vecinos directos en el valor de $x_i(t+1)$. Cuando $\epsilon = 1$ el valor de la variable de estado solo se ve afectada por los vecinos directos del nodo. La información de conexión de los nodos viene dada por una matriz de adyacencia $A = (A_{ij})_{NN}$ de la red. Aquí la función de mapeo $f(x)$ denota la dinámica local de cada sitio de la red a través del mapeo logístico $f(x) = Rx(1-x)$ donde $0 < R < 4$.

El mapeo acoplado dinámico (DCML) [47] puede ser formulado como:

$$x_i(t+1) = f_i(t) + \epsilon_i(t) \left(\sum_{j=1, j \neq i}^N \frac{A_{ij} f_j(t)}{k(i)} - f_i(t) \right) \quad (5.2)$$

donde el coeficiente de acoplamiento varía con el espacio-tiempo sobre los índices i y t . Más específicamente, el sistema DCML hace uso de un mapeo logístico auxiliar de la forma:

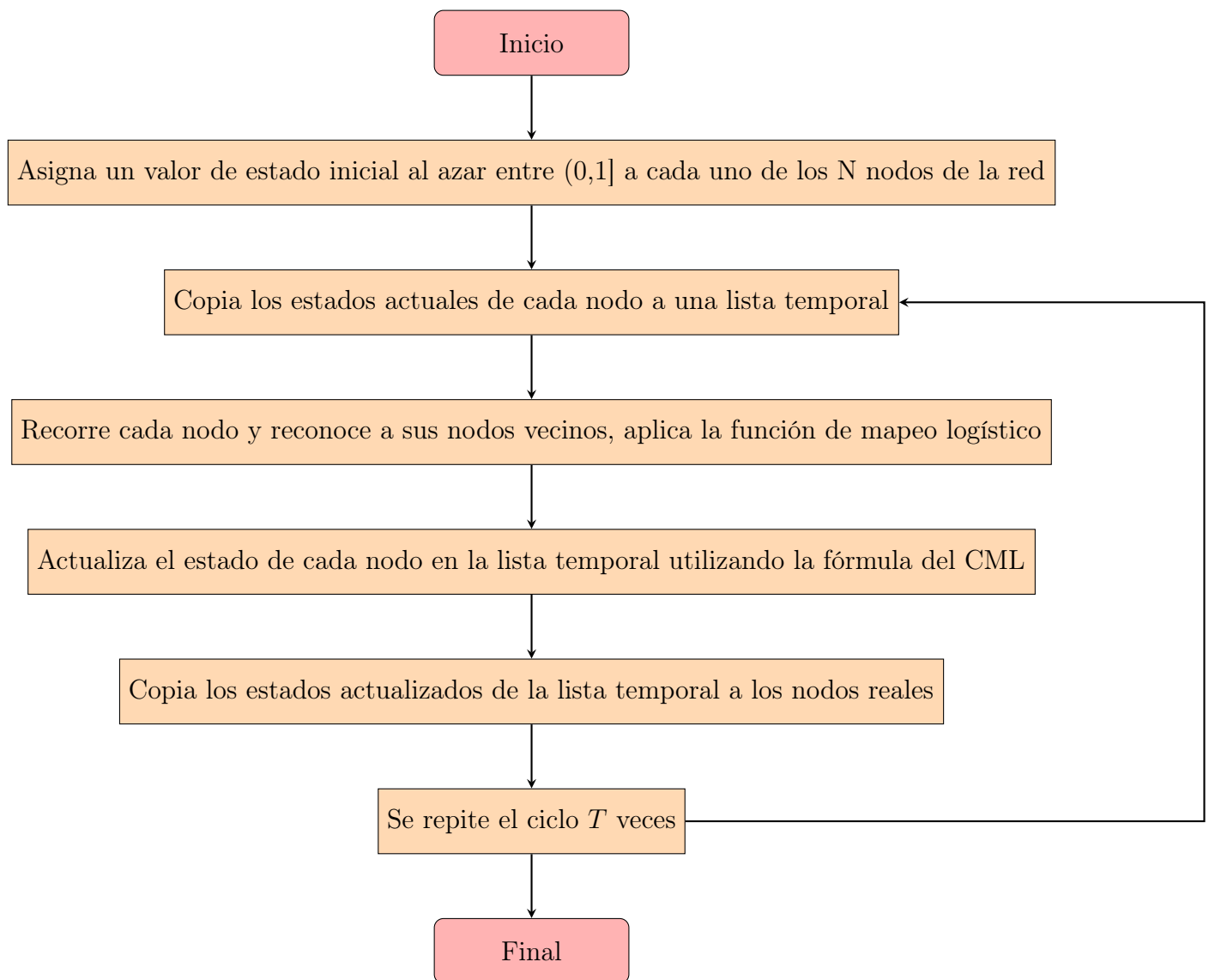
$$\epsilon_i(t+1) = f(\epsilon_i(t)) = \mu^{aux} \epsilon_i(t)(1 - \epsilon_i(t)) \quad (5.3)$$

en el cual $t = 1, 2, \dots, T$, genera la secuencia de coeficientes de acoplamiento dinámico. El parámetro auxiliar μ se establece en 3.99 para lograr una dinámica caótica, y el valor inicial de ϵ_i se establece en ϵ .

En la simulación computacional se definen varios parámetros:

- N : El número total de nodos en la red.
- ϵ : El parámetro de acoplamiento, que determina la influencia de los nodos vecinos en la actualización de cada nodo.
- R : El parámetro control del mapeo logístico.
- T : El número de pasos de tiempo.

El método general de la simulación en cada paso de tiempo, los siguientes pasos:



Para CML, el valor de ϵ se mantiene constante en cada iteración. En cambio, para DCML se utiliza la función de mapeo logístico auxiliar en función de ϵ para actualizar el valor de ϵ en cada paso de tiempo antes de calcular el estado del nodo.

5.1. Comportamiento espacio-temporal

Para cumplir el objetivo estudiar la propagación de perturbaciones en la red de residuos de proteína como parte general de su dinámica, se emplean métodos y técnicas exploratorias de la evolución de estado promedio del sistema creando M copias de este y mediante la comparación con otras redes de referencia.

Los mapeos, al ser aplicados a la PRN de C_α de M^{pro} del SARS CoV-2 (PDB = 6Y2E), proporcionan una representación numérica de su dinámica y permiten estudiar su evolución en el tiempo. Además, se compara la red de proteína con otras redes de referencia, como las redes regulares, las redes de Erdős-Rényi (ER) y las redes de Watts-Strogatz (WS), con el objetivo de identificar similitudes y diferencias cualitativa en sus comportamientos dinámicos.

A lo largo del análisis solo se consideran los aminoácidos comprendidos entre los residuos 3 y 300, lo que resulta en una longitud de 298, por lo que el número de nodos N se establece en 298; mientras que el número máximo de iteraciones T , es igual a 2000 (para descartar estados transitorios). Al estado inicial en $t = 0$ se tiene una distribución aleatoria de estados continuos entre 0 y 1 para cada nodo. Y el parámetro auxiliar μ^{aux} para el DCML es 3.99.

En este estudio, se emplean series de amplitud espacio-tiempo para analizar y visualizar las características dinámicas de los nodos en una red. Estas series de amplitud espacio-tiempo permiten explorar la presencia de dinámicas sincronizada o caóticas en los nodos y, al mismo tiempo, proporcionan una medida cualitativa de la turbulencia dinámica.

Al examinar las series de amplitud espacio-tiempo, se busca identificar la magnitud de los cambios de estado en cada nodo de la red. Cuanto mayor sea la amplitud de estos cambios, mayor será la turbulencia dinámica observada en el nodo en cuestión. Este enfoque nos brinda una valiosa herramienta para discernir si un nodo exhibe comportamientos caóticos o no.

En una primera simulación numérica realizada, se establece un valor fijo de ϵ igual a 0.5 con el propósito de simplificar el análisis. Por otro lado, el parámetro R se establece en tres valores distintos: 3.15, 3.57 (correspondiente al inicio de la zona caótica) y 3.99. Estos valores de R se seleccionan para explorar diferentes regímenes dinámicos en el sistema.

Esta metodología de simulación se aplica tanto a la red de residuos de proteína (PRNs) como a tres tipos de redes de referencia: una red regular con 8 vecinos más cercanos ($\langle k \rangle = 8$), una red de Erdős-Rényi (ER) y una red de Watts-Strogatz (WS). Al repetir la metodología para cada una de estas redes, se busca examinar y comparar la dinámica general de diferentes topologías de red bajo las mismas condiciones de simulación.

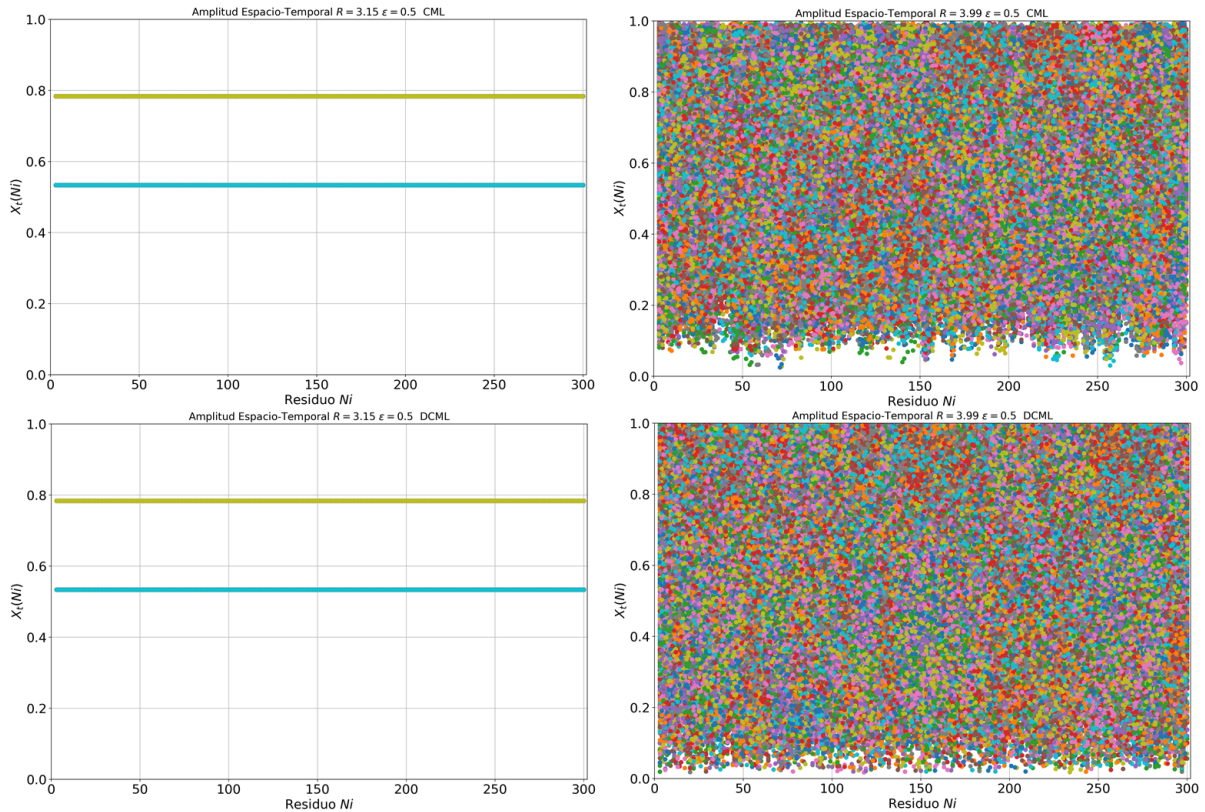


Figura 5.1: Gráficas de análisis del comportamiento espacio-temporal en la PRN usando CML (arriba) y DCML (abajo). El ajuste de R , de izquierda a derecha, es 3,15 y 3.99, respectivamente. Cada color indica el tiempo t en que se encontraba el sistema.

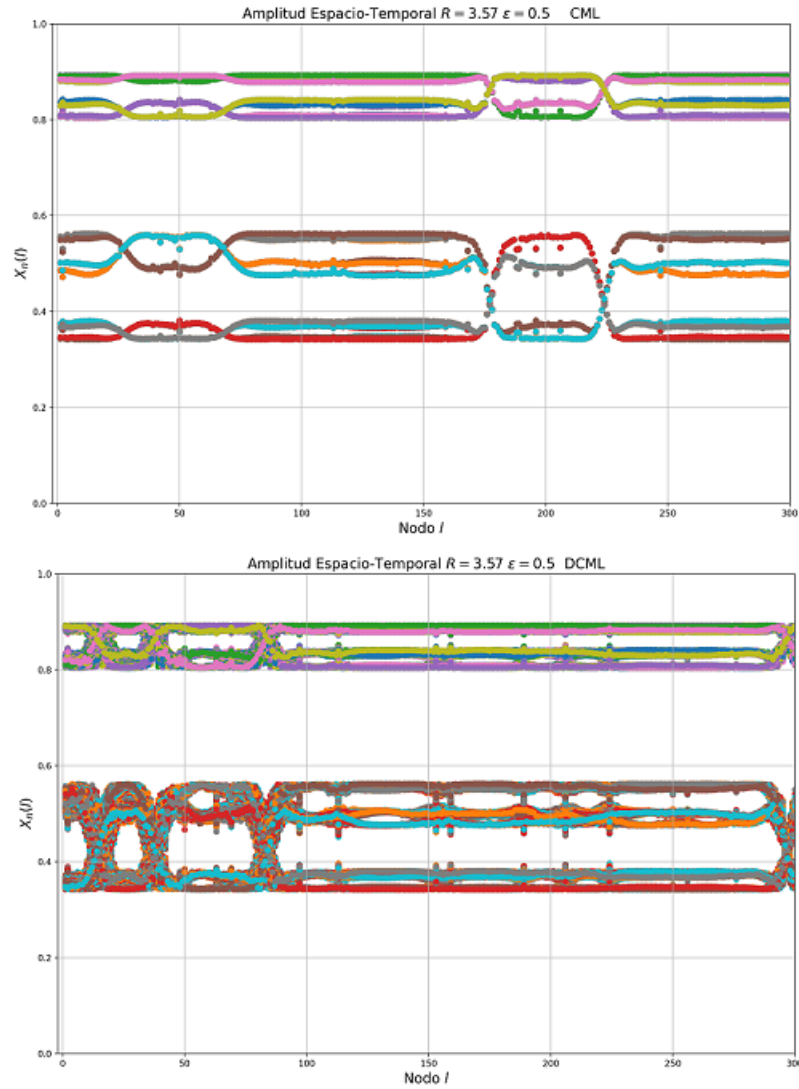


Figura 5.2: Gráficas de análisis del comportamiento espacio-temporal de la red de WS de $\langle k \rangle = 8$, $p = 0.025$ y $N = 306$. El ajuste de R es 3,57. Cada color indica el tiempo t en que se encontraba el sistema.

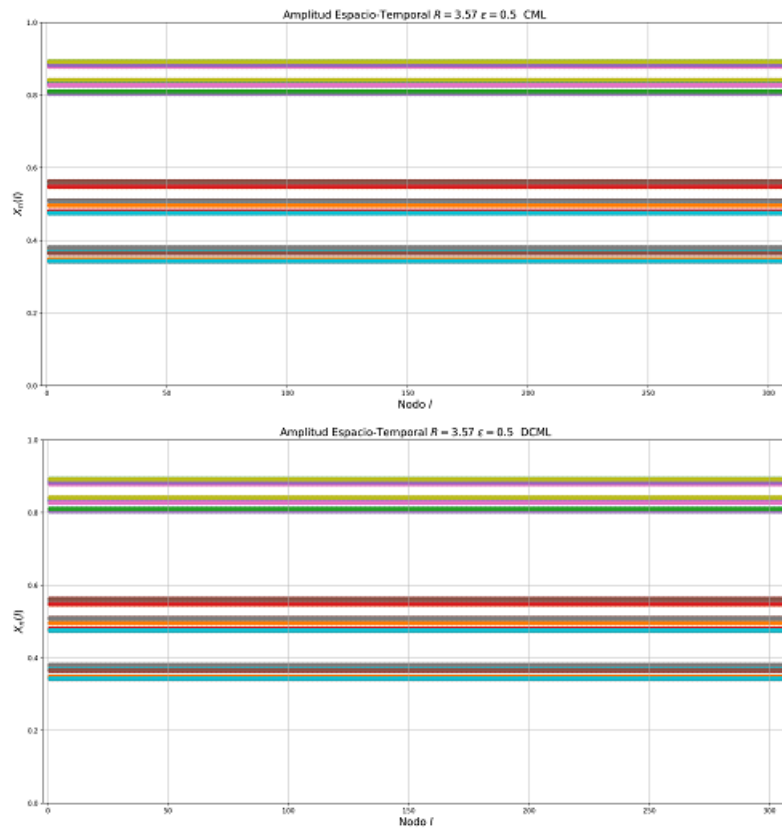


Figura 5.3: Gráficas de análisis del comportamiento espacio-temporal de la red de ER. El ajuste de R es 3,57. Cada color indica el tiempo t en que se encontraba el sistema.

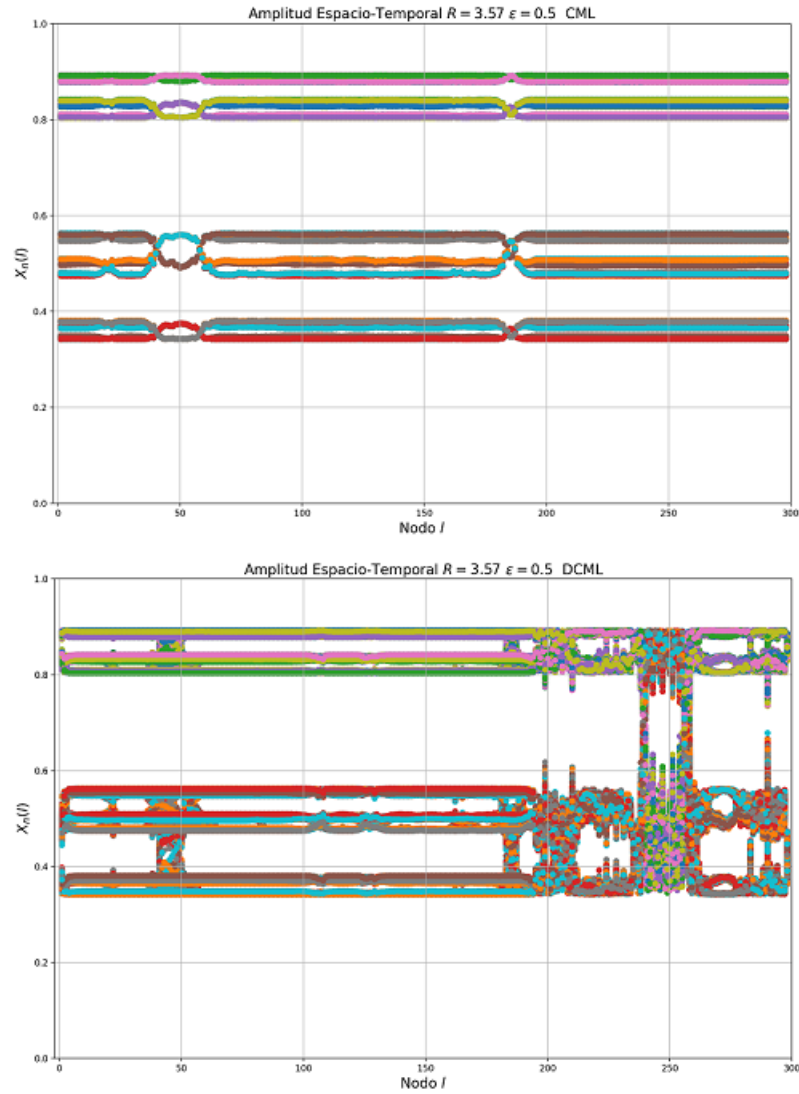


Figura 5.4: Gráficas de análisis del comportamiento espacio-temporal en la PRN. El ajuste de R es 3,57. Cada color indica el tiempo t en que se encontraba el sistema.

En términos generales, al considerar todas las redes con un valor de parámetro R igual a 3.15, tanto el modelo de mapeo acoplado (CML) como el modelo de mapeo acoplado dinámico (DCML) exhiben un comportamiento sincronizado en todo el sistema de periodo dos. Este comportamiento sincronizado queda representado visualmente por la presencia de dos líneas horizontales en los gráficos correspondientes, las cuales presentan los mismos valores de estado.

Si $R = 3.99$, los comportamientos se vuelven caóticos. El rango dinámico se extiende hacia los límites superior e inferior. Estos resultados revelan que todos los sistemas han entrado en el patrón completo de turbulencia.

Este tipo de comportamiento sincronizado en un período de dos unidades de tiempo es consistente con las características esperadas para el mapeo logístico en la región periódica, donde solo existen dos puntos fijos. En esta región, el sistema oscila alternativamente entre estos dos puntos fijos, generando un patrón repetitivo en la evolución temporal de los estados de los nodos.

Cuando el parámetro R toma el valor de 3.57, se pueden observar efectos de multiplicación de periodo en el sistema, lo que implica una bifurcación de estados que ocurre de manera regular. Sin embargo, es importante destacar que no todos los nodos se sincronizan en el mismo periodo, lo que da como resultado un patrón en forma de 'x' en la dinámica del sistema. Los puntos de torsión en este contexto representan las soluciones estables del sistema con el periodo más corto, cuyas posiciones están determinadas por los estados iniciales de los nodos.

Es interesante notar que según el estudio realizado por Kaneko (1989) [51], se han identificado respuestas aperiódicas en el sistema DCML cuando se aplica a una red regular. Esto implica que algunos nodos en la red muestran un comportamiento caótico en lugar de seguir un patrón periódico. Esta observación resalta la complejidad y diversidad de las dinámicas en las redes regulares.

En contraste, se ha encontrado que el sistema en la red de Erdős-Rényi (ER) no presenta patrones en forma de 'x', lo que indica que el sistema permanece en una fase periódica y se encuentra completamente sincronizado. Esto sugiere que la topología de la red ER favorece un comportamiento más ordenado y estable en comparación con las redes regulares. Esto podría deberse a la alta comunicabilidad de la red.

Si repetimos el análisis de la dinámica espacio temporal para el CML variando el acoplamiento ϵ de 0 a 0.5 para un parámetro de acoplamiento R correspondiente a estados periódicos (por ejemplo 3.45) se puede observar que cuando el mapeo esta desacoplado ($\epsilon = 0$) hay una sincronización de estados pero no de fase, si se incrementa ϵ hasta un valor de acoplamiento alto por ejemplo 0.5 la sincronización se vuelve completa. Esto se puede observar en figura 5.4.

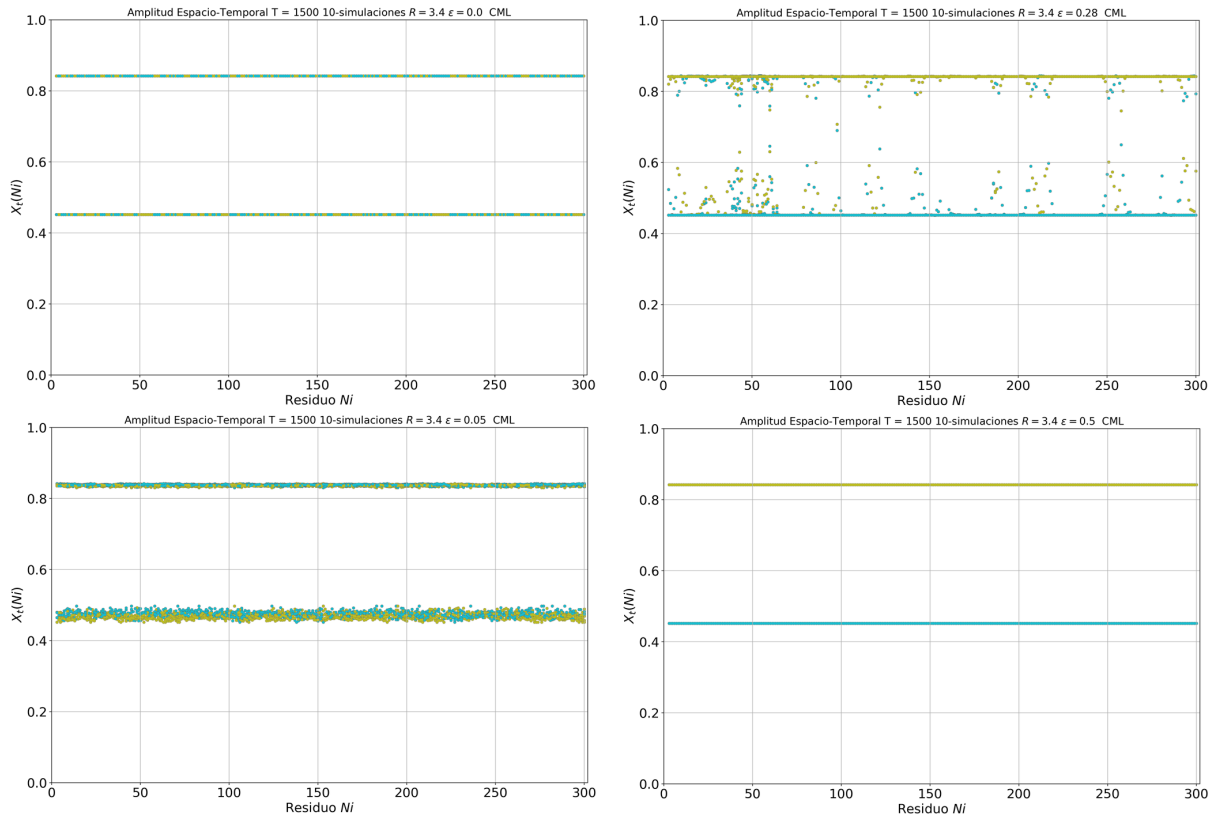


Figura 5.5: Gráficas de análisis del comportamiento espacio-temporal en la PRN. El ajuste de R es 3,4, y ϵ toma los valores 0, 0.05, 0.28 y 0.5 respectivamente

Por otro lado, el modelo DCML exhibe una combinación de dinámica periódica completamente sincronizada y dinámica turbulenta para valores de R correspondientes a la región periódica del atractor del mapeo logístico. Al variar R , se observan dos tipos de dinámicas, como se muestra en las figuras 5.6 y 5.7. En estas figuras, se realiza el conteo del número de órbitas distintas a lo largo de la simulación, excluyendo los primeros 950 pasos de tiempo transitorios. Este análisis permite identificar los nodos que mantienen un comportamiento turbulento de forma persistente.

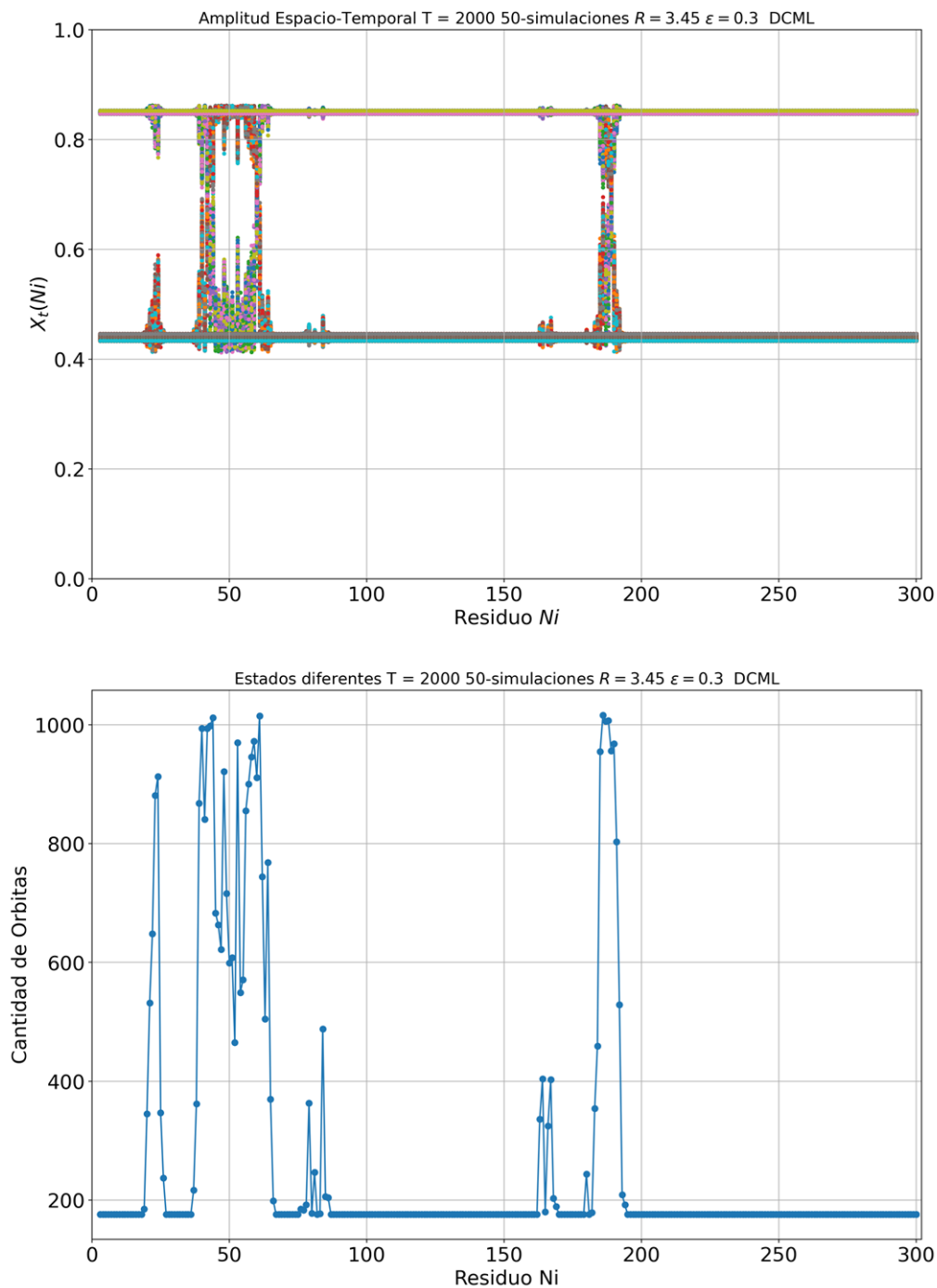


Figura 5.6: Gráficas de análisis del comportamiento espacio-temporal de la red de residuos de proteína de C_α para 50 simulaciones con diferentes estados iniciales aleatorios y cantidad de orbitas diferentes para cada nodo residuo representado como nodo en la red. El ajuste de R es 3,45.

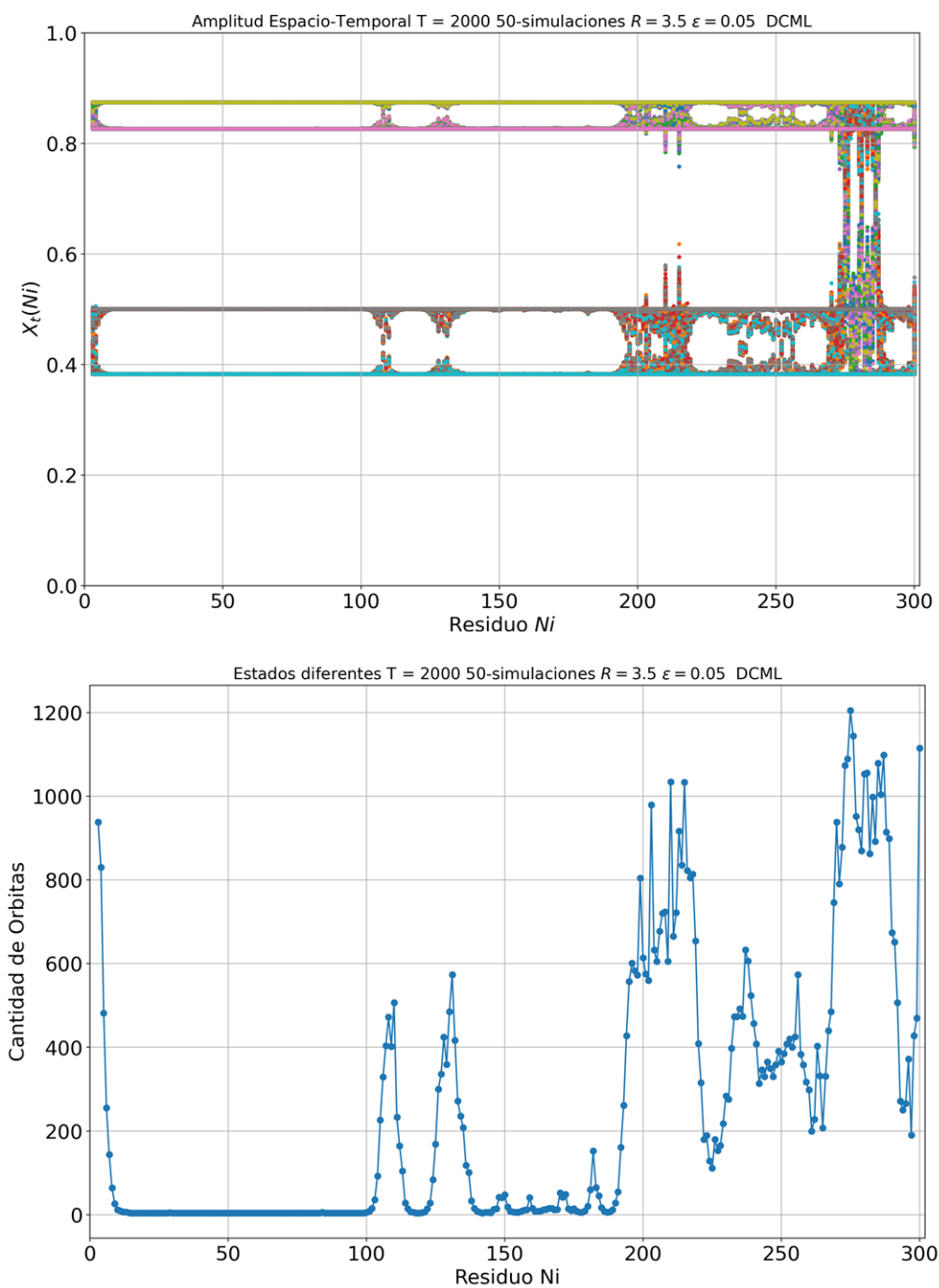


Figura 5.7: Gráficas de análisis del comportamiento espacio-temporal de la red de residuos de proteína de C_α para 50 simulaciones con diferentes estados iniciales aleatorios y cantidad de orbitas diferentes para cada nodo residuo representado como nodo en la red. El ajuste de R es 3,5.

La figura 5.8 muestra los nodos con dinámica no periódica, identificados a partir de 100 simulaciones consecutivas del modelo DCML, representadas en las figuras 5.6 y 5.7, junto con la gráfica de centralidad de subgrafo de largo alcance normalizada para la red de residuos de la proteína M^{pro} del SARS-CoV-2 (PDB = 6Y2E). Al analizar esta figura, se observa una coincidencia de 164 puntos que presentan una dinámica aperiódica, coincidiendo con los residuos de menor centralidad en la proteína.

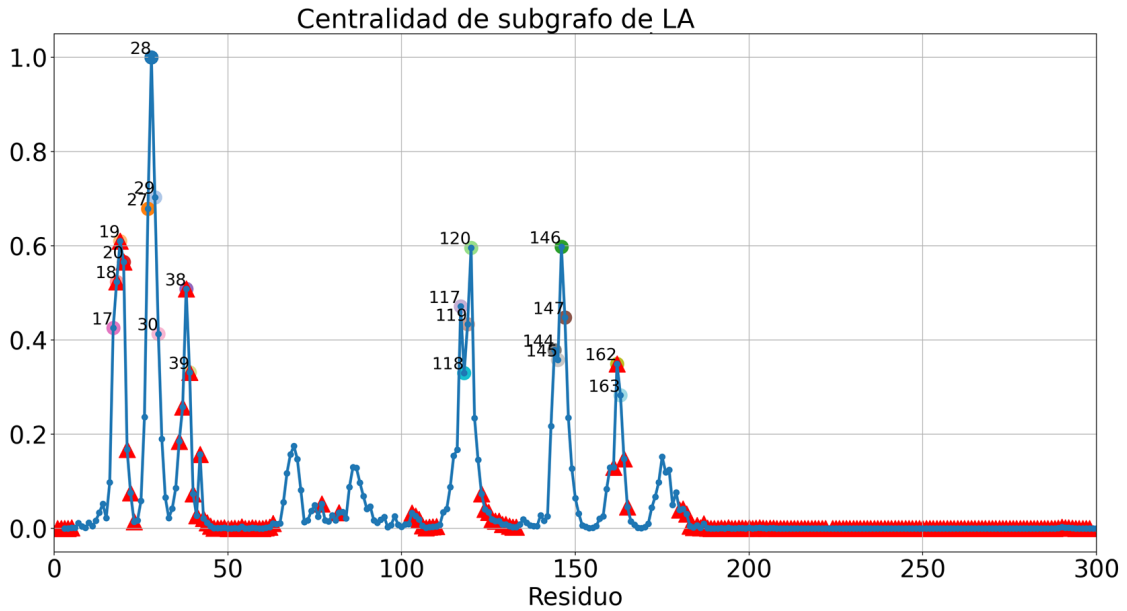


Figura 5.8: Correspondencia entre la centralidad de subgrafo de largo alcance normaliza y el comportamiento espacio-temporal no periódico de la PRN de la proteína M^{pro} del SARS CoV-2 (PDB = 6Y2E).

A lo largo de las simulaciones, algunos nodos de la red de proteínas muestran dinámicas aperiódicas de manera más marcada en comparación con otros nodos, lo que permite identificar dos zonas con patrones dinámicos distintos: una región de comportamiento aperiódico y otra de comportamiento periódico. El patrón aperiódico se asocia con los residuos de la PRN con menor comunicabilidad, como se observa en la figura 5.9 (triángulos rojos). Desde otra perspectiva, esto equivale a la aparición de una sincronización dinámica en la zona de la red formada por los puntos con mayor comunicabilidad a largo alcance, los cuales no aparecen marcados por triángulos rojos en la figura 5.9 y tienden hacia la sincronización dinámica según las simulaciones realizadas en DCML.

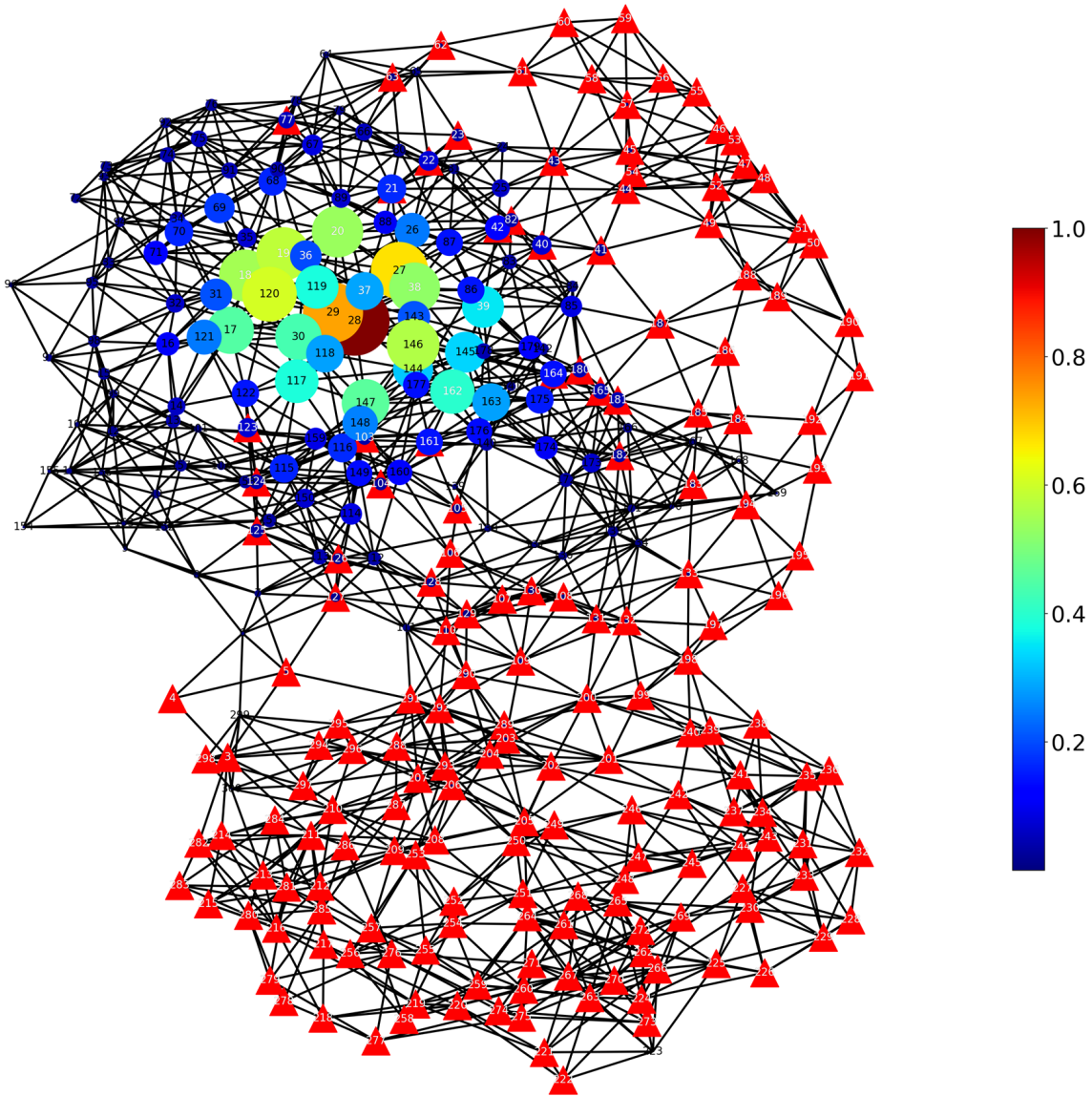


Figura 5.9: Correspondencia entre la centralidad de subgrafo de largo alcance y el comportamiento espacio-temporal en la red de residuos de C_α .

5.2. Espectro de potencias

El análisis del espectro de potencias es una herramienta crucial para determinar si una secuencia exhibe características caóticas. El teorema de Fourier establece que cualquier señal periódica puede descomponerse en una serie de Fourier, lo que da lugar a un espectro de potencias discreto. Por otro lado, una señal aperiódica se representa mediante una integral

de Fourier, lo que resulta en un espectro de potencias continuo. En este sentido, si el espectro muestra uno o más picos distintos, la señal se considera periódica o cuasi-periódica. En cambio, si el espectro de potencias presenta una distribución similar al ruido sin picos evidentes, la señal se considera caótica[52].

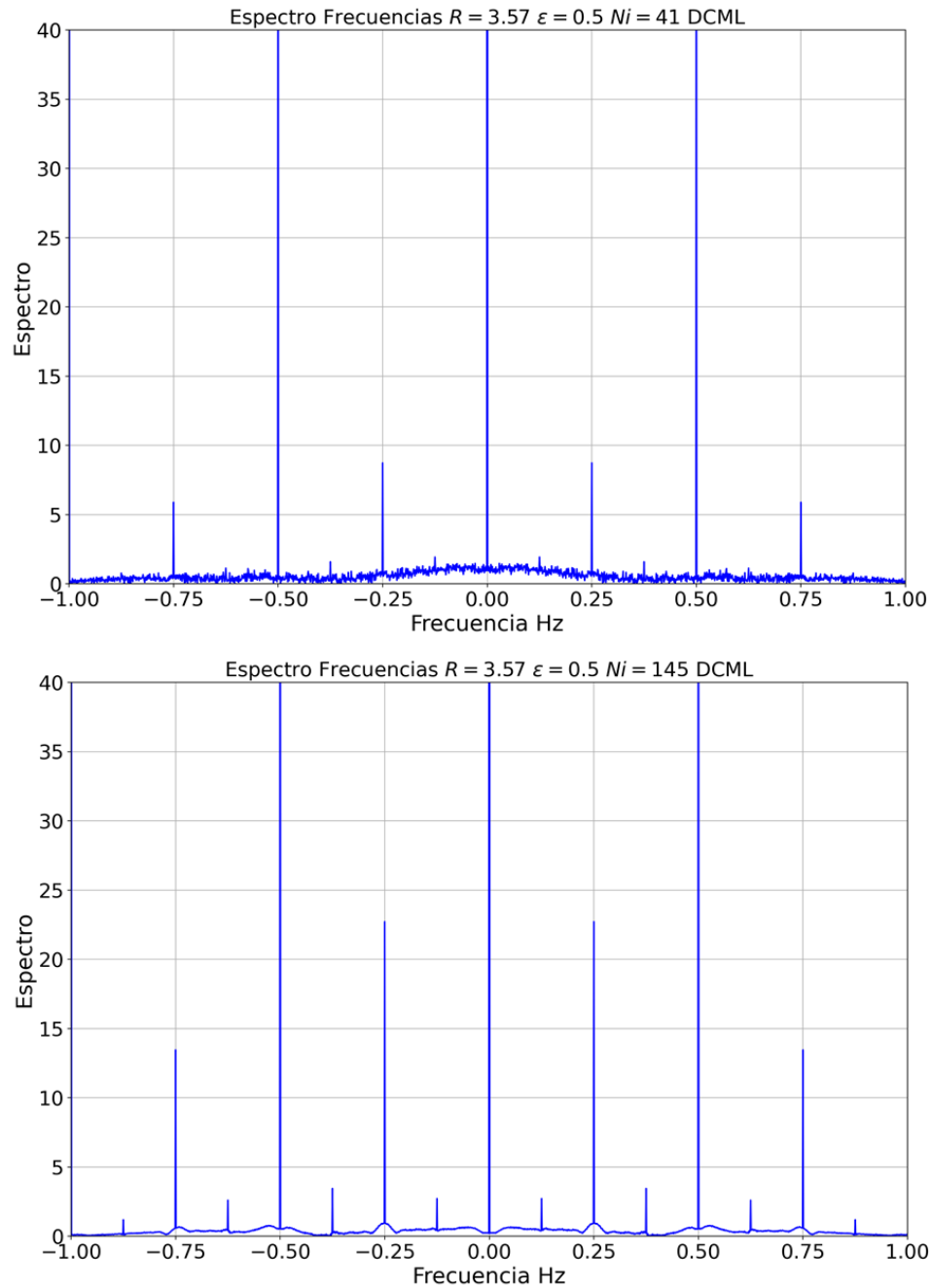


Figura 5.10: Espectro de potencias para los sitios catalíticos 41 y Cys-145 obtenidos de [49] de la red de residuos de proteína de C_α .

Por lo tanto, es posible determinar si un nodo de la red exhibe una dinámica caótica mediante el análisis del espectro de potencias. Esto se ejemplifica en la figura 5.9, donde se muestra el espectro de potencias correspondiente a los dos sitios activos de la red de proteínas extraídos de la base de datos PDB.

Los resultados obtenidos revelan diferencias significativas en la dinámica de los dos sitios catalíticos clave en la proteína. En primer lugar, se observó que el sitio catalítico Cys-145, ubicado entre los nodos de mayor centralidad en la subgráfica de largo alcance, exhibe una dinámica periódica o no caótica. Esto sugiere que este sitio desempeña un papel fundamental en la estabilidad y la función de la proteína, ya que su comportamiento está caracterizado por patrones predecibles y periódicos.

Por otro lado, el sitio catalítico 41, que se encuentra en una posición de menor centralidad en la red, muestra una dinámica aperiódica o caótica según el análisis del espectro de potencias. Esto indica que este sitio es más susceptible a fluctuaciones y cambios impredecibles en su comportamiento. La baja centralidad de este nodo sugiere que su contribución a la sincronización, estabilidad y la función de la proteína puede ser menos significativa en comparación con el otro sitio catalítico Cys-145.

En resumen, nuestros hallazgos destacan la importancia de la centralidad de los nodos en la determinación de la dinámica y la funcionalidad de los sitios catalíticos en la proteína. Mientras que el sitio catalítico Cys-145, con una alta centralidad, exhibe una dinámica periódica, el sitio catalítico 41, con una centralidad más baja, muestra una dinámica caótica. Estos resultados resaltan la relevancia de la estructura de la red y la posición relativa de los nodos en la comprensión de la función y la estabilidad de las proteínas a nivel molecular.

5.3. Interpretación de resultados y conclusiones

Para propósitos de esta investigación, desarrollamos exitosamente un modelo de redes de residuos de proteínas (PRNs) usando diferentes proteínas globulares de la base de datos PDB, incluyendo la proteasa principal del SARS CoV-2. El objetivo principal de esta construcción fue analizar y comparar la estructura tridimensional y la dinámica de estas proteínas. Cumpliendo con estos objetivos, pudimos encontrar propiedades topológicas clave, como la ubicación y distribución de los nodos de mayor centralidad de comunicabilidad en la proteína y sus propiedades de sincronización dinámica.

Al comparar con redes de referencia, tales como la red de primeros vecinos (PV), la red de Watts-Strogatz (WS) y la red de Erdős-Rényi (ER), descubrimos que la distribución de grados en la PRN se asemeja a la distribución de Poisson característica de la red ER (ver Figura ??). Además, el coeficiente de agrupamiento promedio y la longitud de camino promedio se acercan a los valores observados en la red WS (ver Tabla 4.1), lo cual indica que la proteína exhibe características de redes de mundo pequeño. No obstante, la distribución de los nodos de alta centralidad de subgrafo de largo alcance (LR) en la proteína difiere de las distribuciones de esta métrica en las redes ER y WS (ver Figuras ?? y ??). También identificamos que el comportamiento dinámico de las redes PV y WS es más similar al de la PRN (ver Figuras 5.2 y 5.4), mostrando una menor tendencia a la sincronización de estados en comparación con la red ER, lo cual se podría explicarse por la longitud de camino promedio más corta de las redes ER, que facilita la comunicación y sincronización en la red.

A partir de estos hallazgos, concluimos que la red de Watts-Strogatz, con parámetros específicos de promedio de grado $\langle k \rangle$ y probabilidad de recableado p , logra asemejarse notablemente a la estructura y dinámica de las PRNs, capturando algunas de sus características esenciales en términos de dinámica y estructura, aunque con las limitaciones mencionadas previamente.

El análisis de las propiedades topológicas de la proteasa principal del SARS CoV-2 demostró que la centralidad de subgrafo de largo alcance (LR) permite identificar puntos estructuralmente relevantes en las PRNs. Estas métricas de comunicabilidad brindan información sobre los sitios clave para la transmisión de información a larga distancia en la proteína, entre los cuales se destaca el sitio activo Cys-145, que se encuentra en la zona de mayor centralidad de subgrafo.

En línea con los hallazgos de Ernesto Estrada en su estudio [26], también encontramos que la comunicabilidad de largo alcance en la proteasa M^{pro} del SARS CoV-2 influye en la capacidad de transmitir perturbaciones a larga distancia entre aminoácidos. Para evaluar esta comunicabilidad en la dinámica espacio-temporal de la proteína, empleamos mapeos acoplados (CMLs) como herramienta de análisis. Los resultados indican que las zonas de alta comunicabilidad de la red tienden a una mayor sincronización, resaltando su papel en la dinámica global de la proteína.

Los CMLs fueron utilizados de manera efectiva como herramienta para modelar la dinámica de las redes y obtener información sobre el comportamiento espacio-temporal de diferentes puntos en las redes. En nuestro análisis, encontramos una posible correlación entre los sitios de alta centralidad de subgráfica de largo alcance en las PRNs estudiadas y su sincronización espacio-temporal. Esta correlación puede interpretarse cualitativamente como la existencia de órbitas periódicas en los sitios con alta centralidad de subgráfica de largo alcance en la red (ver Figura 5.7 y Figura 5.8). Además, es relevante destacar que el modelo de mapeo acoplado dinámico (DCML) exhibe una dinámica más diversa y compleja, sin importar la red que se esté estudiando, lo cual lo convierte en la elección más apropiada para esta parte de la investigación.

Los resultados de las simulaciones sugieren que la red de proteínas posee una organización estructural intrínseca en la que existen zonas de alta y baja comunicabilidad promueven distintos patrones dinámicos. La tendencia de los nodos con menor comunicabilidad a exhibir comportamientos aperiódicos, contrastan con la sincronización dinámica observada en los nodos de alta comunicabilidad, y sugiere una segmentación funcional en la red que corresponde al punto activo de la red.

Es importante destacar que también se puede explorar la variación en la estructura de la red mediante la modificación del valor del radio de corte r_C . En este estudio en particular, se estableció un valor de 8Å basado en la literatura de referencia[9]. Al ajustar este valor, se producen pequeños cambios en la estructura de la red, como la centralidad de los nodos y el número de enlaces.

El análisis topológico y dinámico de las PRNs nos permite concluir que la estructura de una proteína influye significativamente en su sincronización y en el flujo de información a través de la red. La capacidad de esta red para transmitir información a larga distancia destaca la presencia de nodos con alta comunicabilidad, los cuales juegan un rol esencial

en la dinámica global de la proteína, al corresponderse con su sitio activo. Es especialmente notable que en el punto activo de la proteína estudiada se encuentran nodos de alta centralidad, caracterizados por su estabilidad dinámica y comportamiento sincronizado, subrayando su relevancia funcional. Estos hallazgos son de gran importancia para el avance en biología molecular y podrían tener implicaciones significativas en el diseño de fármacos, al enfocarse en el sitio activo de la proteína. En conjunto, nuestros resultados subrayan la importancia de considerar la estructura de la red de residuos para obtener una visión más completa de las proteínas y su papel en los procesos biológicos.

Bibliografía

- [1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [2] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of modern physics*, vol. 87, no. 3, p. 925, 2015.
- [3] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [4] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [5] C. Song, S. Havlin, and H. A. Makse, “Self-similarity of complex networks,” *Nature*, vol. 433, no. 7024, pp. 392–395, 2005.
- [6] M. E. Newman, “Mixing patterns in networks,” *Physical review E*, vol. 67, no. 2, p. 026126, 2003.
- [7] C. Böde, I. A. Kovács, M. S. Szalay, R. Palotai, T. Korcsmáros, and P. Csermely, “Network analysis of protein dynamics,” *Febs Letters*, vol. 581, no. 15, pp. 2776–2782, 2007.
- [8] M. Vendruscolo, N. V. Dokholyan, E. Paci, and M. Karplus, “Small-world view of the amino acids that play a key role in protein folding,” *Physical Review E*, vol. 65, no. 6, p. 061910, 2002.
- [9] A. R. Atilgan, P. Akan, and C. Baysal, “Small-world communication of residues and significance for protein dynamics,” *Biophysical journal*, vol. 86, no. 1, pp. 85–91, 2004.
- [10] G. Bagler and S. Sinha, “Network properties of protein structures,” *Physica A: Statistical Mechanics and its Applications*, vol. 346, no. 1-2, pp. 27–33, 2005.

- [11] K. Brinda and S. Vishveshwara, “A network representation of protein structures: implications for protein stability,” *Biophysical journal*, vol. 89, no. 6, pp. 4159–4170, 2005.
- [12] E. Estrada, “Universality in protein residue networks,” *Biophysical journal*, vol. 98, no. 5, pp. 890–900, 2010.
- [13] K. Kaneko, “Theory and applications of coupled map lattices,” *Nonlinear science: theory and applications*, 1993.
- [14] K. Kaneko, “Overview of coupled map lattices,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 2, no. 3, pp. 279–282, 1992.
- [15] Y. Oono and S. Puri, “Computationally efficient modeling of ordering of quenched phases,” *Physical review letters*, vol. 58, no. 8, p. 836, 1987.
- [16] D. A. Kessler, H. Levine, and W. N. Reynolds, “Coupled-map lattice model for crystal growth,” *Physical Review A*, vol. 42, no. 10, p. 6125, 1990.
- [17] T. Yanagita, “Coupled map lattice model for boiling,” in *International Symposium on Imaging in Transport Processes.*, Begel House Inc., 1992.
- [18] T. Yanagita and K. Kaneko, “Coupled map lattice model for convection,” *Physics Letters A*, vol. 175, no. 6, pp. 415–420, 1993.
- [19] T. Yanagita and K. Kaneko, “Modeling and characterization of cloud dynamics,” *Physical Review Letters*, vol. 78, no. 22, p. 4297, 1997.
- [20] H. Nishimori and N. Ouchi, “Formation of ripple patterns and dunes by wind-blown sand,” *Physical Review Letters*, vol. 71, no. 1, p. 197, 1993.
- [21] H. Ito and L. Glass, “Spiral breakup in a new model of discrete excitable media,” *Physical review letters*, vol. 66, no. 5, p. 671, 1991.
- [22] F. Bignone, R. Livi, and M. Propato, “Long transients dynamics in biochemical networks,” *Il Nuovo Cimento D*, vol. 20, no. 1, pp. 91–102, 1998.
- [23] C. Beck, “Chaotic strings and standard model parameters,” *Physica D: Nonlinear Phenomena*, vol. 171, no. 1-2, pp. 72–106, 2002.
- [24] X. Xiao, P. Wang, and K.-C. Chou, “Cellular automata and its applications in protein bioinformatics,” *Current Protein and Peptide Science*, vol. 12, no. 6, pp. 508–519, 2011.

- [25] G. E. Aguilar-Pineda and L. Olivares-Quiroz, “Catalytic and binding sites prediction in globular proteins through discrete markov chains and network centrality measures,” *Physical Biology*, vol. 18, no. 6, p. 066002, 2021.
- [26] E. Estrada, “Topological analysis of sars cov-2 main protease,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 6, p. 061102, 2020.
- [27] A. Citarella, A. Scala, A. Piperno, and N. Micale, “Sars-cov-2 mpro: a potential target for peptidomimetics and small-molecule inhibitors,” *Biomolecules*, vol. 11, no. 4, p. 607, 2021.
- [28] S. H. Strogatz, “Exploring complex networks,” *nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [29] X. F. Wang and G. Chen, “Complex networks: small-world, scale-free and beyond,” *IEEE circuits and systems magazine*, vol. 3, no. 1, pp. 6–20, 2003.
- [30] N. Biggs, N. L. Biggs, and B. Norman, *Algebraic graph theory*. No. 67, Cambridge university press, 1993.
- [31] P. Erdős, A. Rényi, *et al.*, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, no. 1, pp. 17–60, 1960.
- [32] S. Milgram, “The small world problem,” *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.
- [33] A.-L. Barabási, R. Albert, and H. Jeong, “Mean-field theory for scale-free random networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1-2, pp. 173–187, 1999.
- [34] A.-L. Barabási, *Network Science*. Cambridge University Press, 2016.
- [35] E. Estrada, “Quantifying network heterogeneity,” *Physical Review E*, vol. 82, no. 6, p. 066102, 2010.
- [36] M. E. Newman, “Assortative mixing in networks,” *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.
- [37] L. C. Freeman *et al.*, “Centrality in social networks: Conceptual clarification,” *Social network: critical concepts in sociology*. Londres: Routledge, vol. 1, pp. 238–263, 2002.
- [38] E. Estrada, *The structure of complex networks: theory and applications*. Oxford University Press, 2012.

- [39] E. Estrada and J. A. Rodriguez-Velazquez, “Subgraph centrality in complex networks,” *Physical Review E*, vol. 71, no. 5, p. 056103, 2005.
- [40] E. Estrada and N. Hatano, “Communicability in complex networks,” *Physical Review E*, vol. 77, no. 3, p. 036111, 2008.
- [41] E. Estrada and N. Hatano, “Communicability angle and the spatial efficiency of networks,” *SIAM Review*, vol. 58, no. 4, pp. 692–715, 2016.
- [42] E. N. Gilbert, “Random graphs,” *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.
- [43] K. Kaneko, “Period-doubling of kink-antikink patterns, quasiperiodicity in antiferro-like structures and spatial intermittency in coupled logistic lattice: Towards a prelude of a “field theory of chaos”,” *Progress of Theoretical Physics*, vol. 72, no. 3, pp. 480–486, 1984.
- [44] G. Gielis and R. S. MacKay, “Coupled map lattices with phase transition,” *Nonlinearity*, vol. 13, no. 3, p. 867, 2000.
- [45] C. Storm and W. J. Freeman, “Complex systems: chaos and beyond. a constructive approach with applications in life sciences-kunihiko kaneko and ichiro tsuda; springer, berlin/heidelberg, 2001, 273pp., 7 chapters, isbn 3-540-67202-8,” *Engineering Applications of Artificial Intelligence*, vol. 1, no. 15, pp. 117–119, 2002.
- [46] J.-R. Chazottes and B. Fernandez, *Dynamics of coupled map lattices and of related spatially extended systems*, vol. 671. Springer Science & Business Media, 2005.
- [47] W. Xingyuan, F. Le, W. Shibing, C. Zhang, and Z. Yingqian, “Spatiotemporal chaos in coupled logistic map lattice with dynamic coupling coefficient and its application in image encryption,” *Ieee Access*, vol. 6, pp. 39705–39724, 2018.
- [48] A. D. Baxevanis, G. D. Bader, and D. S. Wishart, *Bioinformatics*. John Wiley & Sons, 2020.
- [49] “Protein data bank: the single global archive for 3d macromolecular structure data,” *Nucleic acids research*, vol. 47, no. D1, pp. D520–D528, 2019.
- [50] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, “The protein data bank,” *Nucleic acids research*, vol. 28, no. 1, pp. 235–242, 2000.

- [51] K. Kaneko, “Pattern dynamics in spatiotemporal chaos: Pattern selection, diffusion of defect and pattern competition intermittency,” *Physica D: Nonlinear Phenomena*, vol. 34, no. 1-2, pp. 1–41, 1989.
- [52] R. Huang, F. Han, X. Liao, Z. Wang, and A. Dong, “A novel intermittent jumping coupled map lattice based on multiple chaotic maps,” *Applied Sciences*, vol. 11, no. 9, p. 3797, 2021.