

Newsletter Trimestral

CÁTEDRA
iDANAE

INTELIGENCIA · DATOS · ANÁLISIS · ESTRATEGIA

2T24

Teoría de grafos y redes neuronales
en la industria farmacéutica



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID

MS  **Management
Solutions**
Making things happen

Introducción

En un mundo donde la tecnología y la ciencia avanzan de forma vertiginosa, la teoría de grafos se erige como una herramienta poderosa y versátil para su aplicación en el campo de la industria farmacéutica y la salud. Desde su origen en el siglo XVIII con los trabajos de Euler hasta su aplicación moderna en la bioinformática, los grafos han demostrado su capacidad para resolver problemas complejos y optimizar procesos en diversas áreas. Asimismo, se han desarrollado técnicas de inteligencia artificial basadas en grafos, como las redes complejas o las redes neuronales de grafos.

Hoy en día, estas técnicas pueden tener una aplicación relevante en la predicción de interacciones medicamentosas, el análisis de mutaciones genéticas y el reposicionamiento de fármacos, áreas críticas para el desarrollo de tratamientos más efectivos y seguros.

La combinación de avances en secuenciación genómica, inteligencia artificial y grandes volúmenes de datos ha permitido un mayor desarrollo en la práctica clínica; por ejemplo, optimizando terapias farmacológicas o mejorando la seguridad de los tratamientos. En los últimos años se han publicado diversos estudios en este sentido.

Por ejemplo, mCSM [1] surge como una herramienta que utiliza firmas basadas en grafos para predecir los efectos de las mutaciones en proteínas. Aquí se demuestra que mCSM es eficaz para predecir cambios en la estabilidad de mutaciones en la proteína p53, superando a otros métodos y mostrando su utilidad en contextos de enfermedades complejas como el cáncer.

También se han utilizado redes neuronales convolucionales para extraer características locales y mecanismos de atención para obtener información importante en el campo de interacciones fármaco-proteína (DPIs) [2].

En pruebas con datasets como C.elegans, humanos, y BindingDB [3], se demostró que el enfoque propuesto mejoraba la efectividad en la predicción de DPIs en comparación con métodos de aprendizaje automático convencionales. Cabe destacar que este tipo de estudios necesita alimentarse de buenas bases de datos, ámbito que también se ha desarrollado recientemente con datasets como DDInter [4]. Esta es una base de datos curada de interacciones entre fármacos (DDI) diseñada para ayudar a los médicos a identificar combinaciones de medicamentos peligrosas y mejorar los sistemas de salud. Además de las consultas básicas, incorpora una función de verificación de prescripción para ayudar a los médicos a decidir si las combinaciones de fármacos pueden utilizarse de forma segura.

En esta publicación se explora cómo la teoría de grafos y las redes neuronales (complejas y de grafos) están transformando la investigación farmacéutica. Se abordan los desafíos actuales en el desarrollo de medicamentos, como la predicción de la interacción medicamento-proteína y la interacción medicamento-medicamento, y cómo los métodos basados en estas técnicas de inteligencia artificial están proporcionando nuevas soluciones. Además, se presenta un caso de uso concreto de reposicionamiento de fármacos, destacando el potencial de estas tecnologías para acelerar la innovación y mejorar los resultados en salud.



Conexión entre los grafos y la industria farmacéutica

Panorama actual de la industria farmacéutica

En las últimas décadas, los avances significativos en la secuenciación del genoma humano y la bioinformática han catalizado una transformación sin precedentes en el ámbito de la medicina. Uno de los principales beneficiarios de este fenómeno es la industria farmacéutica, la cual ha emergido como un actor fundamental en el crecimiento y desarrollo de España. El crecimiento del Producto Interior Bruto (PIB) español hasta alcanzar un 3% en 2023 [5], según datos del Instituto Nacional de Estadística (INE), es un claro indicador del impacto positivo que esta industria ha tenido en la economía del país. Este crecimiento económico se atribuye en gran medida a la capacidad de la industria farmacéutica para capitalizar los avances científicos y tecnológicos, así como para satisfacer las necesidades de salud de la población mediante la producción y comercialización de medicamentos innovadores.

En el ámbito de la investigación clínica de medicamentos innovadores, los laboratorios farmacéuticos nacionales desempeñan un papel destacado, con una inversión que supera los 750 millones de euros. España cuenta con una industria farmacéutica notablemente diversa, con más de 100 fabricantes de productos farmacéuticos básicos y más de 200 empresas especializadas en la preparación de medicamentos de alta complejidad. Además, importantes empresas farmacéuticas globales han establecido filiales en el territorio español [6].

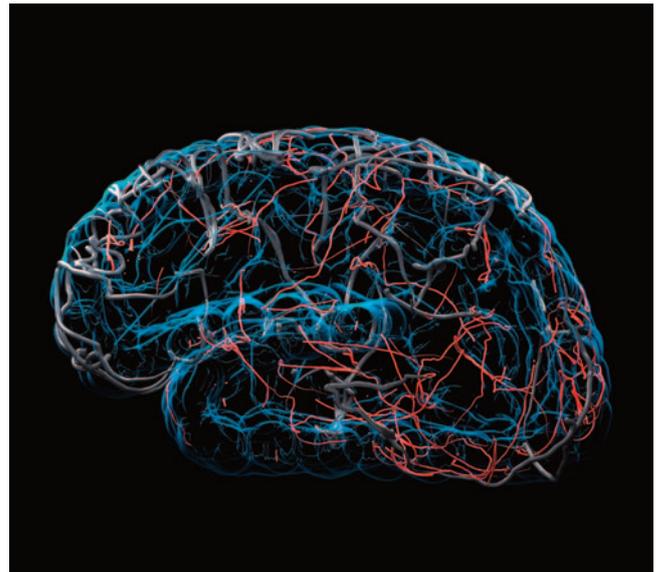
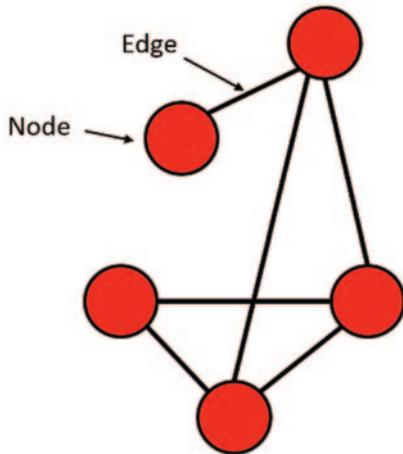
El desarrollo de nuevos medicamentos en la industria farmacéutica, un proceso que puede tardar entre 10 y 15 años y requerir inversiones de hasta 800 millones de dólares, enfrenta desafíos considerables en términos de tiempo y recursos [7]. Para acelerar este proceso, reducir costes y mejorar la efectividad de los fármacos existentes, se están integrando en la práctica clínica numerosos enfoques computacionales basados en inteligencia artificial y el uso masivo de datos. Estos enfoques permiten optimizar la terapia farmacológica, mejorando la seguridad y eficacia del tratamiento en beneficio de los pacientes. Dentro de la industria farmacéutica, se exploran numerosas líneas de investigación destinadas a lograr los objetivos mencionados, destacando tres enfoques principales: la predicción de la interacción medicamento-proteína, la predicción de mutaciones en aminoácidos y la predicción de interacción medicamento-medicamento. Esta convergencia entre la innovación tecnológica y la investigación biomédica promete revolucionar aún más el campo de la medicina, impulsando avances significativos para el tratamiento de enfermedades y mejorando la calidad de vida de las personas.

La teoría de grafos

La teoría de grafos [8] es uno de los pilares fundamentales de las matemáticas discretas, cuyo origen se atribuye al famoso matemático suizo, Leonhard Euler. Él fue quien introdujo el concepto de grafos en su trabajo *Solutio problematis ad geometriam situs pertinentis* (1736) [9], para resolver el problema de los siete puentes de Königsberg. Así se inició toda una nueva rama de las matemáticas, que siguió desarrollándose durante el siglo XIX y principios del siglo XX, gracias a las contribuciones de matemáticos como Arthur Cayley y Gustav Kirchhoff en el estudio de árboles y grafos conectados [10] [11]. En la segunda mitad del siglo XX, la teoría de grafos comenzó a aplicarse en informática, especialmente en algoritmos de redes, optimización y teoría de la computación. Esto se consolidó con trabajos como los de Dijkstra [12] [13] [14] y su algoritmo para encontrar el camino más corto de un grafo. A partir de la década de 1990, y con la inclusión del estudio de redes de interacción reales, surgió la ciencia de redes complejas [15] [16], que extiende la teoría de grafos al incorporar la dinámica y evolución de estos grafos en el tiempo. Una red compleja se refiere a una red, modelada como grafo, que posee ciertas propiedades estadísticas y topológicas no triviales. Las redes complejas de hoy en día se emplean en el estudio de fenómenos críticos de la física estadística, en la resolución de problemas bioinspirados o en las ciencias sociales [17], explorando cómo se comportan los sistemas bajo diferentes condiciones. Así, la teoría de grafos y las redes complejas coexisten de forma que la primera sirve como fundamento matemático y la segunda amplía su alcance para abordar problemas dinámicos y aplicados en el mundo real.

Un grafo [15] es un objeto matemático que consiste en un conjunto de nodos (o vértices) y un conjunto de aristas (o enlaces) que conectan pares de nodos. Se representan mediante diagramas como el de la Fig. 1, de manera que los nodos se asocian con puntos y las aristas con líneas que unen nodos conectados entre sí. Permiten representar de manera simple estructuras de conectividad muy complejas. Es por esto que los grafos son esenciales en el estudio de redes complejas, ya que proporcionan la base matemática necesaria para modelar, entender y analizar el fenómeno de estudio.

Fig. 1 Diagrama de estructura de un grafo. Los círculos (rojo) son los nodos y las aristas (negro) son los enlaces existentes entre nodos.



Es importante entender los conceptos de **nodos** y aristas. Los **primeros** son los elementos básicos de una red y representan las entidades individuales dentro del sistema. Su definición depende en gran medida de la red de estudio y su contexto. No es lo mismo estudiar el cerebro si se considera cada neurona como un nodo, o si se define cada nodo como una región cerebral. Las **aristas**, por otro lado, representan las interacciones o enlaces dados entre los nodos. Estas relaciones pueden ser físicas, como en redes de transporte, o abstractas, como en redes de amistad.

La teoría de grafos y las redes complejas se han convertido en herramientas con un gran potencial para entender y analizar sistemas en una amplia variedad de campos.

En **ciencias sociales**, el análisis de redes sociales permite comprender cómo se difunde la información, cómo se estructuran las comunidades y cuál es la influencia de ciertos individuos dentro de una red social. Esto se puede aplicar en política, para entender la dinámica de movimientos sociales y la propagación de ideas, o en el ámbito de las redes sociales digitales, para entender cómo se forman y evolucionan las conexiones entre usuarios, permitiendo mejorar los algoritmos de recomendación y personalización de contenido. Las redes complejas se han convertido en esenciales en diversos estudios de **biología y medicina**. Con ellas se pueden estudiar interacciones entre proteínas, redes neuronales o propagación de enfermedades. Estas últimas son cruciales para modelar la propagación de enfermedades, lo que permite anticiparse y diseñar estrategias de intervención más efectivas, como las campañas de vacunación o las medidas de cuarentena. Las **redes tecnológicas** son otro ejemplo. Incluyen internet y redes de transporte. Estas redes ayudan a optimizar la eficiencia y la robustez de estos sistemas frente a fallos. La estructura y funcionamiento de la red global de Internet, por ejemplo, se basa en principios de grafos y redes complejas.

En el **ámbito económico y financiero**, la teoría de redes proporciona herramientas para modelar las complejas interacciones en los mercados financieros y las transacciones económicas. Esto permite a los economistas y analistas financieros identificar posibles puntos de riesgo sistémico y desarrollar estrategias para mitigar el impacto de crisis financieras.

Por último, **en ecología**, se utilizan para modelar las interacciones dentro de ecosistemas, como las redes alimentarias. Esto ayuda a entender la estabilidad de los ecosistemas y el cómo diferentes especies pueden depender unas de otras, lo que permite detectar especies clave para la preservación del medio natural.

Un denominador común en todos estos campos es la idea de que las redes complejas permiten entender las interacciones y la dinámica subyacentes del caso de estudio. Mientras que en otros ámbitos de modelización es necesario recurrir a técnicas de *feature engineering* para comprender la relación entre variables, las redes de grafos llevan inherente esa relación por cómo está concebida su construcción misma. En un principio, cada nodo alberga la información y características que lo definen exclusivamente a él, pero la propia estructura de la red ya predetermina cuáles y cómo van a ser las diferentes interacciones con su entorno, de modo que, en última instancia, los vectores de características de cada nodo pasarán a incorporar la información relativa a su entorno y a cómo interactúa con sus vecinos. En resumen, la teoría de grafos y las redes complejas no solo ofrecen una forma de representar y analizar estructuras y sistemas complejos, sino que también proporcionan conocimientos cruciales para predecir comportamientos y diseñar intervenciones efectivas en una variedad de campos.

Redes complejas y redes neuronales de grafos

La evolución de la teoría de grafos ha dado lugar al desarrollo de redes complejas y redes neuronales de grafos, herramientas que permiten modelar y analizar sistemas dinámicos y en diversos campos, incluyendo la investigación farmacéutica. Estas redes proporcionan un marco robusto para entender las interacciones complejas entre distintos elementos de un sistema, como moléculas, proteínas y medicamentos, a través del uso de métodos avanzados de inteligencia artificial.

Redes complejas

Las redes complejas son grafos caracterizados por patrones de conexión no triviales que reflejan las propiedades y relaciones inherentes a los sistemas reales. A diferencia de los grafos simples, las redes complejas incluyen características como la modularidad, la robustez y la capacidad de adaptación a través del tiempo. Estas propiedades son esenciales para modelar sistemas biológicos y químicos donde las interacciones no son estáticas ni homogéneas:

1. Estructura y dinámica. Las redes complejas integran tanto la topología (estructura estática) como la dinámica (comportamiento en el tiempo) de los sistemas. En el contexto farmacéutico, esto permite modelar cómo los medicamentos interactúan con múltiples proteínas y cómo estas interacciones pueden cambiar debido a mutaciones o la presencia de otras moléculas.

2. Aplicaciones en biomedicina. Las aplicaciones de redes complejas en la biomedicina incluyen la identificación de nuevas relaciones entre genes y enfermedades, la modelización de la propagación de enfermedades infecciosas y la optimización de terapias combinadas. Estos modelos ayudan a predecir efectos secundarios y a personalizar tratamientos médicos.

Redes neuronales de grafos

Las redes neuronales de grafos (GNN, por sus siglas en inglés) representan una evolución de los métodos de aprendizaje automático que pueden operar directamente sobre estructuras de grafos. Las GNN son capaces de aprender representaciones de los nodos y las aristas que reflejan las características y las relaciones de estos elementos en el grafo.

- 1. Capacidades de las GNN.** Las redes neuronales de grafos son particularmente efectivas para tareas donde las relaciones estructurales entre datos son cruciales. Por ejemplo, en la predicción de interacciones medicamento-proteína, las GNN pueden integrar datos de estructura molecular y bioactividad para hacer predicciones precisas sobre cómo un fármaco interactuará con una proteína objetivo.
- 2. Integración con datos biomédicos.** La combinación de GNN con bases de datos biomédicas permite la extracción de características complejas y la identificación de patrones no evidentes a simple vista. Esto es útil para la predicción de efectos de mutaciones, donde las GNN pueden modelar cómo una alteración en la secuencia de aminoácidos afecta la estructura y función de una proteína.



Aplicaciones en la investigación farmacéutica

En las últimas décadas, las redes complejas han emergido como una potente herramienta de modelización. Destaca su aplicación en procesos de percolación, sincronización, epidemiológicos o transiciones de fase, entre otros. El desarrollo que las redes complejas han experimentado en estos campos ha permitido, también, que se expanda su ámbito de aplicación, llegando a utilizarse en la modelización y análisis de interacciones entre moléculas, proteínas y enfermedades, lo cual ofrece nuevas perspectivas y soluciones innovadoras para el descubrimiento de fármacos y el diseño de medicamentos. A continuación, se exponen algunas de las propuestas más prometedoras en este campo.

Predicción medicamento-proteína

Cuando un fármaco (una molécula química) se une a un objetivo biológico (como las proteínas), modula su comportamiento con el objetivo de devolverlo a su estado normal. Predecir esta interacción entre fármacos y objetivos (Drug-protein interaction, DPI) es una etapa muy importante en el proceso de descubrimiento de nuevos medicamentos y la comprensión de sus efectos secundarios [2] [19]. Un ejemplo claro de la importancia de investigar las interacciones fármaco-proteína es el surgimiento de la pandemia de COVID-19. Ante la ausencia de medicamentos específicos para combatir esta enfermedad, la investigación detallada de las interacciones entre las proteínas del virus SARS-CoV-2 y diversos fármacos existentes permitió identificar compuestos con potencial

terapéutico. El estudio detallado de la proteína espiga del SARS-CoV-2, así como de otras proteínas virales clave como la proteasa 3CLpro, permitió desarrollar terapias antivirales y guiar el diseño de nuevos inhibidores específicos [20].

Dentro de este campo de estudio, han surgido las redes de interacción medicamento-proteína, redes orientadas a representar las interacciones entre fármacos y proteínas objetivo en el cuerpo humano. Los nodos se corresponden con los fármacos y proteínas, y las aristas representan las diferentes interacciones entre ellos. El desarrollo de estas redes puede ayudar a predecir la eficacia y los efectos secundarios de los fármacos, así como a identificar posibles combinaciones de fármacos para nuevas terapias.

La mayoría de los enfoques se basan en la información sobre fármacos y proteínas representada por vectores de características. Por ejemplo, uno de los enfoques comunes que suele formularse es el de obtener predicciones de DTI como una tarea de clasificación binaria. Estos enfoques hacen uso de bases de datos como BindingDB, que contiene información sobre las interacciones de unión entre proteínas y ligandos, que son moléculas pequeñas como los fármacos. Su objetivo principal es proporcionar datos experimentales detallados sobre la afinidad de unión, que es una medida de cuán fuertemente un ligando se une a una proteína [3]. KIBA (Kinase Inhibitor BioActivity) es una colección de datos diseñada para facilitar la investigación en el ámbito del descubrimiento de fármacos, particularmente en relación con los inhibidores de quinasas. KIBA integra información de varias bases de datos de bioactividad y estandariza los datos de interacción entre quinasas y sus inhibidores [21].



Propiedades de los grafos

Los grafos presentan una serie de propiedades que permiten caracterizar y estudiar el comportamiento de redes en diversas aplicaciones prácticas y teóricas. Las principales propiedades de los grafos incluyen:

- Grado de un nodo: indica cuántas aristas inciden en él.
- Conectividad: mide la robustez del grafo frente a la eliminación de nodos o aristas.
- Coeficiente de clustering: evalúa la tendencia de los nodos a formar clústeres.
- Centralidad: mide la importancia de un nodo en la red.
- Comunidades: son subconjuntos densamente conectados de nodos.
- Modularidad: cuantifica la estructura de la red en comunidades.
- Caminos: representan secuencias de aristas entre nodos. También existen los conocidos como ciclos o caminos cerrados.
- Longitud de camino promedio: es la media de las distancias entre todos los pares de nodos.
- Diámetro: es la distancia más larga entre dos nodos.
- Correlación de Pearson: métrica mediante la cual se evalúa la relación entre las propiedades de los nodos conectados.

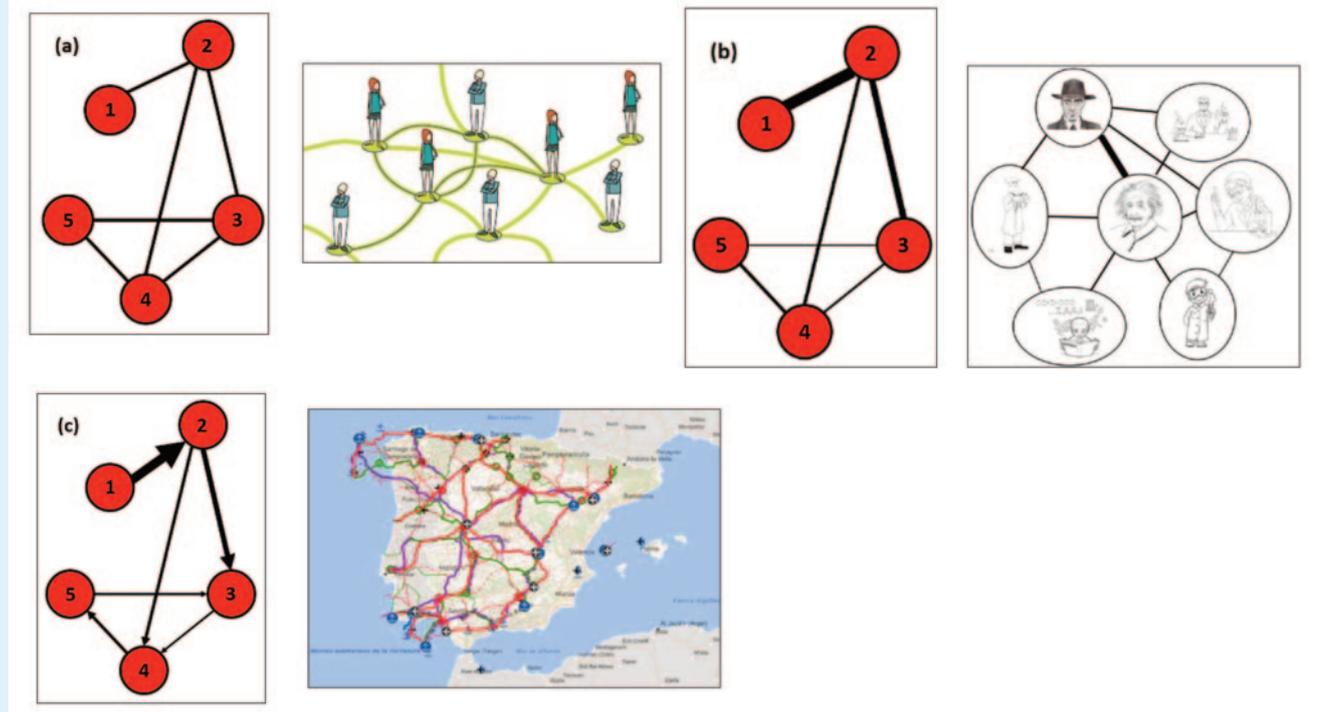
En función de las propiedades y características de la red en estudio, y de cómo quedan definidos los nodos y aristas de la misma, se pueden distinguir diferentes tipos de redes, a saber:

- Dirigidas o no dirigidas: las segundas son bidireccionales, mientras que en las primeras cobra especial relevancia cuál es el nodo de partida y cuál el de llegada en un enlace, como en una red social donde un usuario puede seguir a otro sin reciprocidad.
- Pesadas o no pesadas: según se traten todos los enlaces por igual o se otorgue más importancia a unos que a otros.
- Conexas o no conexas: si al menos un par de nodos no están conectados por un camino, la red es conexa.
- Homogéneas o heterogéneas: según aparezcan elementos de distintos tipos, como una red biológica que incluye diferentes tipos de moléculas e interacciones.

Estáticas o dinámicas: si la estructura de la red varía con el paso del tiempo.

- Libres de escala: son aquellas que siguen una ley de potencias; algunos nodos tienen muchas conexiones, mientras que la mayoría tienen pocas.

Fig. 2. Ejemplos de diferentes tipos de redes complejas.



(a) Red de amigos en la escuela. Es una red no dirigida -ser amigos es una relación bidireccional- y no pesada -todas las relaciones de amistad tienen el mismo valor-. Puede ser no conexa si hay gente en clase que no tiene ningún amigo. Es homogénea y dinámica -las relaciones interpersonales varían con el tiempo-. Normalmente no sigue una ley de potencias, luego no es libre de escala.

(b) Red de colaboraciones científicas. Es una red no dirigida -el hecho de colaborar en una publicación implica a ambos- y pesada -aquellas colaboraciones que son más citadas se consideran más relevantes-. Además, es una red conexa, homogénea -un tipo de nodos o científicos y un tipo de enlaces o colaboraciones-, dinámica y libre de escala -habitualmente los científicos tienen pocas citas, aunque hay algunos muy citados-.

(c) Red de transportes española [18]. En este caso, estamos ante una red dirigida -los viajes tienen origen y destino- y pesada -hay rutas más relevantes que otras-. También es una red conexa, heterogénea -diferentes tipos de viajeros y de métodos de transporte-, dinámica -se desarrolla con el paso de los años- y libre de escala -hay lugares con más conexiones que otros-.

Predicción mutaciones en aminoácidos

La secuenciación del genoma humano ha revelado una gran diversidad genética en las poblaciones humanas, incluyendo mutaciones que alteran la secuencia de aminoácidos en las proteínas. Estas mutaciones concretas, llamadas variantes sin sentido, pueden ser patógenas, afectando la función de las proteínas y la salud de los pacientes, o benignas, con efectos mínimos [22]. Se ha conseguido catalogar más de 15 millones de variaciones de nucleótidos en la población humana, sin embargo, muchas variantes aún se desconocen y muchas otras aún tienen un impacto desconocido, representando un desafío en la genética humana [23]. El virus de la inmunodeficiencia humana (VIH), responsable de 630.000 muertes en 2022 [24], presenta una alta tasa de mutaciones en sus proteínas. Los fármacos antirretrovirales aprobados para combatir el VIH están diseñados para inhibir dos proteínas específicas. Sin embargo, las mutaciones en los aminoácidos de estas proteínas pueden alterar su estructura, haciendo que los fármacos pierdan su efectividad y dejen de actuar contra el virus [25].

Los métodos computacionales se están usando para estudiar el mecanismo molecular de la resistencia a los medicamentos inducida por mutaciones y para desarrollar herramientas predictivas para detectar mutaciones que afectan a la secuencia de aminoácidos y, por tanto, a la estructura y función de la proteína [26]. Un enfoque que ya ha obtenido buenos resultados es el del método mCSM [1] [27] (matriz de escaneo de corte de mutaciones), que utiliza firmas basadas en grafos para representar el entorno estructural de tipo salvaje y el aprendizaje automático para predecir el efecto de las mutaciones en la estabilidad de las proteínas. En este contexto, las proteínas se representan como redes donde los nodos son átomos y las aristas son las interacciones entre estos átomos. Al analizar cómo las mutaciones afectan las propiedades locales y globales de estas redes (por ejemplo, cambios en la conectividad o en la distribución de las interacciones), mCSM puede predecir el impacto de las mutaciones en la estabilidad y función de las proteínas. En Cambridge se han desarrollado extensiones del método mCSM para predecir el impacto de la mutación en proteína-ligando (mCSM-lig) [28] e interacciones proteína-proteína (mCSM-PPI2) [29], y más recientemente en Fiocruz, Brasil y en Melbourne, Australia, por Douglas Pires y David Ascher para proteína-ácido nucleico (mCSM-NA) [30], antígeno anticuerpo (mCSM-AB) [31] interacciones y conformaciones y dinámicas de proteínas en combinación basadas en la dinámica de modo normal (DynaMut) [32].

Para poder desarrollar esta línea de investigación existen diversas bases de datos; entre las más destacadas se encuentran: UniProtKB alberga una extensa colección de información detallada y curada sobre proteínas, y proporciona datos sobre secuencias, funciones, estructuras tridimensionales, localización celular, interacciones con otras moléculas y anotaciones biológicas [33]; DbSNP se centra en las variaciones genéticas, incluyendo mutaciones y polimorfismos de nucleótido único (SNPs) [34]; Human Gene Mutation Database

contiene información detallada sobre mutaciones genéticas humanas, e incluye datos sobre la ubicación exacta de la mutación en el genoma, la naturaleza de la mutación (por ejemplo, si es una sustitución de un solo nucleótido, una inserción o una delección), y cualquier efecto conocido de la mutación en la función génica o en la salud humana [35].

Predicción interacción medicamento-medamento

La combinación de dos o más fármacos se conoce como terapia combinada, y es una estrategia común para mejorar la eficacia terapéutica y reducir los efectos secundarios. Sin embargo, una selección inadecuada de medicamentos puede resultar en reacciones adversas. Por ello, es de especial interés conocer las interacciones entre medicamentos (IDD) [36]. Los efectos que producen las IDD son un factor de riesgo importante para la hospitalización, especialmente entre pacientes ambulatorios de edad avanzada. De hecho, se estima que las IDD contribuyen del 5% al 14% de las reacciones adversas en pacientes hospitalizados [37]. Un ejemplo de esta IDD peligrosa podría ser la terapia combinada de Warfarina (anticoagulante que previene formación de trombos) con antiinflamatorios no esteroideos (reducen inflamación, dolor y fiebre) como el ibuprofeno, que pueden causar hemorragia al inhibir el metabolismo del anticoagulante [38].

En esta línea de investigación se pueden emplear redes de interacción medicamento-medamento. La idea subyacente es similar a las anteriores, con la particularidad de que los nodos representan fármacos y enfermedades, y las aristas indican la asociación entre ellos. En este caso, la intención es la de buscar posibles efectos secundarios y, toda vez localizados esos efectos secundarios, y entendido por qué surgen, conocer cómo prevenirlos o encontrar una posible solución. Los algoritmos predictivos aprovechan datos farmacocinéticos y farmacodinámicos para predecir cómo los medicamentos interactuarán en el cuerpo humano [39], así pueden identificar patrones y correlaciones que quizás no sean evidentes para los métodos tradicionales.

Los datos utilizados en estudios como el mencionado anteriormente se extraen de grandes bases de datos de interacciones conocidas como DrugBank y DDInter. DrugBank cuenta con un total de 570.091 productos farmacéuticos donde se incluyen medicamentos de molécula pequeña aprobados y medicamentos experimentales, entre otros [40]. Proporciona más de 200 campos de datos para cada medicamento, con la mitad de la información dedicada a aspectos químicos, farmacológicos, farmacéuticos y otros aspectos del medicamento, y la otra mitad dedicada a documentar la secuencia, estructura y ruta del medicamento objetivo [41]. DDInter contiene más de 236.834 IDD que involucran 1.833 medicamentos y documenta la información detallada sobre cada IDD, como mecanismos, niveles de riesgo, recomendaciones para el ajuste de medicamentos, etc. [4] [42].

Caso de uso: reposicionamiento de fármacos

El reposicionamiento de medicamentos consiste en identificar nuevos usos terapéuticos para medicamentos que previamente han sido aprobados para diferentes propósitos médicos [43]. Hasta ahora, estos estudios dependían de exámenes manuales y técnicas estadísticas, métodos que requerían considerable tiempo y esfuerzo. Sin embargo, los nuevos métodos computacionales permiten analizar grandes volúmenes de datos genéticos y clínicos, descubriendo nuevos objetivos farmacológicos y prediciendo la eficacia y toxicidad de los compuestos con alta precisión; acelerando así el proceso de desarrollo de fármacos.

Algoritmos de Deep Learning, como los basados en redes neuronales de grafos, se utilizan para el estudio de reposicionamiento de fármacos. Estas herramientas han sido utilizadas para identificar fármacos inhibidores de Janus Kinase 2 (JAK2) [44]. La alteración de la función JAK se asocia con diversos trastornos inflamatorios, como la artritis reumatoide y la psoriasis, entre otras [45]. La inhibición de esta enzima reduce las consecuencias negativas de estas enfermedades.

En este caso de uso concreto se han utilizado los datos procedentes de la base de datos DUD-E [46], que contiene propiedades fisicoquímicas de distintos medicamentos ya aprobados por la FDA. Estos datos han sido utilizados para el entrenamiento de una red convolucional gráfica (GCN) proporcionada por DeepChem [47], una biblioteca Python de código abierto que cuenta con el modelo GraphConvMol. Con este modelo se pretende determinar, con base en la estructura molecular, la capacidad inhibitoria de los medicamentos.

Este modelo de grafos convierte cada átomo del medicamento en un nodo y cada enlace covalente en una arista, de tal manera que cada átomo comunica sus características únicas a los

átomos adyacentes [48]. El conjunto de datos JAK2 se dividió en conjuntos de entrenamiento, validación y prueba en una proporción de 8:1:1 y luego se sometió al modelo GraphConvMol usando validación cruzada con 5 iteraciones.

El modelo entrenado, utilizando el algoritmo GraphConvMol de DeepChem, procesó medicamentos aprobados por la Administración de Alimentos y Medicamentos del Gobierno de los Estados Unidos (FDA) para evaluar su potencial de actividad inhibitoria de JAK2. Se han identificado 20 compuestos activos que presentan actividad inhibitoria de JAK2, entre los cuales se encuentran algunos ya conocidos. La actividad inhibitoria de JAK2 de los 20 fármacos detectados se evaluó experimentalmente y todos ellos mostraron una inhibición de la actividad enzimática JAK2.

Mediante la integración de GCN con técnicas de acoplamiento molecular y su aplicación a una base de datos de compuestos activos, se ha logrado una detección más exhaustiva de posibles inhibidores de JAK2. Este enfoque ha permitido identificar de manera eficiente nuevos candidatos a fármacos que no habían sido considerados anteriormente. Entre estos compuestos, se encuentran ribociclib, amodiaquina, topiroxostat y gefitinib, los cuales han demostrado un potencial inhibidor prometedor de JAK2. Además, la validación experimental ha corroborado los resultados obtenidos mediante aprendizaje profundo y acoplamiento molecular. Por consiguiente, se propone este procedimiento para el reposicionamiento de fármacos en un amplio espectro de objetivos terapéuticos.



Conclusiones

La teoría de grafos y su evolución a través de técnicas de inteligencia artificial han demostrado ser una herramienta con un gran potencial en la industria farmacéutica, abordando con éxito desafíos complejos en la predicción de interacciones medicamentosas, análisis de mutaciones genéticas y reposicionamiento de fármacos. Su integración con tecnologías avanzadas y el análisis de grandes volúmenes de datos ha optimizado significativamente la investigación y el desarrollo de medicamentos.

Los estudios recientes han validado la eficacia de métodos basados en grafos para predecir los efectos de mutaciones en proteínas y mejorar la estabilidad de mutaciones complejas. Asimismo, las redes neuronales convolucionales y otros enfoques de aprendizaje automático han mejorado la predicción de interacciones fármaco-proteína y fármaco-fármaco, cruciales para el desarrollo de terapias seguras y efectivas.

Además, la creación de bases de datos ha facilitado la identificación de combinaciones de medicamentos peligrosas y la optimización de las terapias farmacológicas. Estos avances subrayan la importancia de una colaboración continua entre la bioinformática, la teoría de grafos y la industria farmacéutica para mejorar los resultados en salud y acelerar la innovación.

En resumen, la teoría de grafos no solo ha revolucionado el campo de la investigación farmacéutica, sino que también ha abierto nuevas oportunidades para el desarrollo de tratamientos más efectivos y seguros. Con su capacidad para modelar y analizar sistemas complejos, los grafos seguirán desempeñando un papel crucial en la evolución de la medicina y la salud pública.



Bibliografía

- [1] Pires, D. E., Ascher, D. B., & Blundell, T. L. (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, 30(3), 335-342.
- [2] Zhang Y, Hu Y, Li H, Liu X. Drug-protein interaction prediction via variational autoencoders and attention mechanisms. *Front Genet*. 2022 Oct 14;13:1032779. doi: 10.3389/fgene.2022.1032779. PMID: 36313473; PMCID: PMC9614151.
- [3] Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., & Chong, J. (2016). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic acids research*, 44(D1), D1045-D1053.
- [4] Xiong, G., Yang, Z., Yi, J., Wang, N., Wang, L., Zhu, H., ... & Cao, D. (2022). DDInter: an online drug-drug interaction database towards improving clinical decision-making and patient safety. *Nucleic acids research*, 50(D1), D1200-D1207.
- [5] Azpiroz, A. (2024, mayo 14). Los desafíos de la industria farmacéutica española - Salud35. *Salud 35*. https://www.consalud.es/salud35/economia/industria-farmaceutica-espanola-crecimiento-economico_143765_102.html
- [6] Impacto y relevancia de la industria farmacéutica en España. Laboratorios Rubió. <https://www.laboratoriosrubio.com/industria-farmaceutica-en-espana/>
- [7] La industria farmacéutica en España - Datos estadísticos. (s/f). Statista. Recuperado el 11 de junio de 2024, de <https://es.statista.com/temas/5603/la-industria-farmaceutica-en-espana/>
- [8] Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2), 167-256. <http://www.jstor.org/stable/25054401>
- [9] Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 128-140.
- [10] Cayley, A. (1857). XXVIII. On the theory of the analytical forms called trees. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(85), 172-176.
- [11] Kirchhoff, G. (1850). LXIV. On a deduction of Ohm's laws, in connexion with the theory of electro-statics. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 37(252), 463-468.
- [12] Richards, H. (2017). Edsger Wybe Dijkstra. AM Turing Award. Association for Computing Machinery.
- [13] Frana, P. L., & Misa, T. J. (2010). An interview with edsger w. dijkstra. *Communications of the ACM*, 53(8), 41-47.
- [14] Dijkstra, E. W. (2022). A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: His Life, Work, and Legacy* (pp. 287-290).
- [15] Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.
- [16] Dorogovtsev, S. N., & Mendes, J. F. (2002). Evolution of networks. *Advances in physics*, 51(4), 1079-1187.
- [17] Perc, M., Gómez-Gardenes, J., Szolnoki, A., Floría, L. M., & Moreno, Y. (2013). Evolutionary dynamics of group interactions on structured populations: a review. *Journal of the royal society interface*, 10(80), 20120997.
- [18] <https://mapas.fomento.gob.es/VisorTENT/>
- [19] Ye, Q., Hsieh, CY., Yang, Z. et al. A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. *Nat Commun* 12, 6775 (2021). <https://doi.org/10.1038/s41467-021-27137-3>
- [20] Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science*. 2020 Apr 24;368(6489):409-412. doi: 10.1126/science.abb3405. Epub 2020 Mar 20. PMID: 32198291; PMCID: PMC7164518.
- [21] KIBA Dataset. (s. f.). <https://paperswithcode.com/dataset/kiba>
- [22] Reimand J, Wagih O, Bader GD. Evolutionary constraint and disease associations of post-translational modification sites in human genomes. *PLoS Genet*. 2015 Jan 22;11(1):e1004919. doi: 10.1371/journal.pgen.1004919. PMID: 25611800; PMCID: PMC4303425.

- [23] 1000 Genomes Project Consortium. (2010). A map of human genome variation from population scale sequencing. *Nature*, 467(7319), 1061.
- [24] Datos del VIH en el mundo - EresVIHda. (2023, 26 octubre). EresVIHda. <https://www.eresvihda.es/que-es-el-vih/vih-una-historia-de-40-anos/datos-en-el-mundo/>
- [25] del Carmen Chávez, M., Casas, G., Moreira, J., Silveira, P., Moya, I., Bello, R., & Grau, R. (2008). Predicción de mutaciones en secuencias de la proteína transcriptasa inversa del VIH usando nuevos métodos para Aprendizaje Estructural de Redes Bayesianas. *Revista Avances en Sistemas e Informática*, 5(2), 77-85.
- [26] Cao ZW, Han LY, Zheng CJ, Ji ZL, Chen X, Lin HH, Chen YZ. Computer prediction of drug resistance mutations in proteins. *Drug Discov Today*. 2005 Apr 1;10(7):521-9. doi: 10.1016/S1359-6446(05)03377-5. PMID: 15809198.
- [27] Pandurangan, A. P., & Blundell, T. L. (2020). Prediction of impacts of mutations on protein structure and interactions: SDM, a statistical approach, and mCSM, using machine learning. *Protein Science*, 29(1), 247-257.
- [28] Pires DE, Blundell TL, Ascher DB. mCSM-lig: Quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci Rep*. 2016;6:29575
- [29] Rodrigues CHM, Myung Y, Pires DEV, Ascher DB. mCSM-PPI2: Predicting the effects of mutations on protein-protein interactions. *Nucleic Acids Res*. 2019;47:W338–W344
- [30] Pires DEV, Ascher DB. mCSM-NA: Predicting the effects of mutations on protein-nucleic acids interactions. *Nucleic Acids Res*. 2017;45:W241–W246
- [31] Pires DE, Ascher DB. mCSM-AB: A web server for predicting antibody-antigen affinity changes upon mutation with graph-based signatures. *Nucleic Acids Res*. 2016;44:W469–W473.
- [32] Rodrigues CH, Pires DE, Ascher DB. DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res*. 2018;46:W350–W355.
- [33] Coudert E, Gehant S, de Castro E, Pozzato M, Baratin D, Neto T, Sigrist C J A, Redaschi N, Bridge A, UniProt Consortium. Annotation of biologically relevant ligands in UniProtKB using ChEBI
Bioinformatics 39:btac793(2023)
- [34] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001 Jan 1;29(1):308-11. doi: 10.1093/nar/29.1.308. PMID: 11125122; PMCID: PMC29783.
- [35] Stenson et al (2003), The Human Gene Mutation Database (HGMD®): 2003 Update. *Hum Mutat* (2003) 21:577-581. Articles describing the HGMD facility have also appeared in *Trends Genet* (1997) 13:121-122, *Nucleic Acids Res* (1998) 26: 285-287 and *Hum Mutat* (2000) 15: 45-51
- [36] Zhao Y, Yin J, Zhang L, Zhang Y, Chen X. Drug-drug interaction prediction: databases, web servers and computational models. *Brief Bioinform*. 2023 Nov 22;25(1):bbad445. doi: 10.1093/bib/bbad445. PMID: 38113076; PMCID: PMC10782925.
- [37] Villa Zapata L, Subbian V, Boyce RD, Hansten PD, Horn JR, Gephart SM, Romero A, Malone DC. Overriding Drug-Drug Interaction Alerts in Clinical Decision Support Systems: A Scoping Review. *Stud Health Technol Inform*. 2022 Jun 6;290:380-384. doi: 10.3233/SHTI220101. PMID: 3673040.5[38] Ray WA, Chung CP, Murray KT, Smalley WE, Daugherty JR, Dupont WD, Stein CM. Association of Proton Pump Inhibitors With Reduced Risk of Warfarin-Related Serious Upper Gastrointestinal Bleeding. *Gastroenterology*. 2016 Dec;151(6):1105-1112.e10. doi: 10.1053/j.gastro.2016.08.054. Epub 2016 Sep 14. PMID: 27639805; PMCID: PMC5124401.5[39] Hauben M. Artificial Intelligence and Data Mining for the Pharmacovigilance of Drug-Drug Interactions. *Clin Ther*. 2023 Feb;45(2):117-133. doi: 10.1016/j.clinthera.2023.01.002. Epub 2023 Jan 31. PMID: 36732152.5[40] Knox C, Wilson M, Klinger CM, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res*. 2024 Jan 5;52(D1):D1265-D1275. doi: 10.1093/nar/gkad976.5[41] Craig Knox, Mike Wilson, Christen M Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun (Lucy) Chin, Seth A Strawbridge, Marysol Garcia-Patino, Ray Kruger, Aadhavaya Sivakumaran, Selena Sanford, Rahil Doshi, Nitya Khetarpal, Omolola Fatokun, Daphnee Doucet, Ashley Zubkowski, Dorsa Yahya Rayat, Hayley Jackson, Karxena Harford, Afia Anjum, Mahi Zakir, Fei Wang, Siyang Tian, Brian Lee, Jaanus Liigand, Harrison Peters, Ruo Qi (Rachel) Wang, Tue Nguyen, Denise So, Matthew Sharp, Rodolfo da Silva, Cyrella Gabriel, Joshua Scantlebury, Marissa Jasinski, David Ackerman, Timothy Jewison, Tanvir Sajed, Vasuk Gautam, David S Wishart, DrugBank 6.0: the DrugBank Knowledgebase for 2024, *Nucleic Acids Research*, Volume 52, Issue D1, 5 January 2024, Pages D1265–D1275, <https://doi.org/10.1093/nar/gkad976>

- [42] DDInter. (s. f.). <http://ddinter.scbdd.com/>
- [43] Jourdan JP, Bureau R, Rochais C, Dallemagne P. Drug repositioning: a brief overview. *J Pharm Pharmacol.* 2020 Sep;72(9):1145-1151. doi: 10.1111/jphp.13273. Epub 2020 Apr 17. PMID: 32301512; PMCID: PMC7262062.
- [44] Yasir M, Park J, Han ET, Park WS, Han JH, Chun W. Drug Repositioning via Graph Neural Networks: Identifying Novel JAK2 Inhibitors from FDA-Approved Drugs through Molecular Docking and Biological Validation. *Molecules.* 2024 Mar 19;29(6):1363. doi: 10.3390/molecules29061363. PMID: 38542998; PMCID: PMC10974395.
- [45] Spiewak, T. A., & Patel, A. (2022). User's guide to JAK inhibitors in inflammatory bowel disease. *Current Research in Pharmacology and Drug Discovery*, 3, 100096.
- [46] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK J. *Med. Chem.*, 2012, Jul 5. doi 10.1021/jm300687e
- [47] DeepChem. (s. f.). <https://deepchem.io/models/>
- [48] Duvenaud, David K., et al. "Convolutional networks on graphs for learning molecular fingerprints." *Advances in neural information processing systems*. 2015

Autores

Ernestina Menasalvas (UPM)

Manuel Ángel Guzmán (Management Solutions)

Sergio Ruiz (Management Solutions)

Guillermo Fraggio (Management Solutions)



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID

MSO *Management
Solutions*
Making things happen

La Universidad Politécnica de Madrid es una Entidad de Derecho Público de carácter multisectorial y pluridisciplinar, que desarrolla actividades de docencia, investigación y desarrollo científico y tecnológico.

www.upm.es

Management Solutions es una firma internacional de consultoría, centrada en el asesoramiento de negocio, finanzas, riesgos, organización, tecnología y procesos, que opera en más de 50 países y con un equipo de cerca de 4.000 profesionales que trabajan para más de 2.000 clientes en el mundo.

www.managementsolutions.com

Para más información visita

blogs.upm.es/catedra-idanae/