

Quarterly Newsletter

CHAIR
iDANAE

INTELLIGENCE · DATA · ANALYSIS · STRATEGY

2Q24

Graph theory and neural networks
in the pharmaceutical industry



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID

MS *Management
Solutions*
Making things happen

Introduction

In a world where technology and science are advancing vertiginously, the graph theory is emerging as a powerful and versatile tool for its application in the pharmaceutical and healthcare industries. From its origin in the 18th century with the work of Euler to its modern application in bioinformatics, graphs have demonstrated their ability to solve complex problems and optimise processes in various areas. Graph-based artificial intelligence techniques, such as complex networks or graph neural networks, have also been developed.

Nowadays, these techniques can have a relevant application in the prediction of medical interactions, the analysis of genetic mutations and drug repositioning, critical areas for the development of more effective and safer treatments.

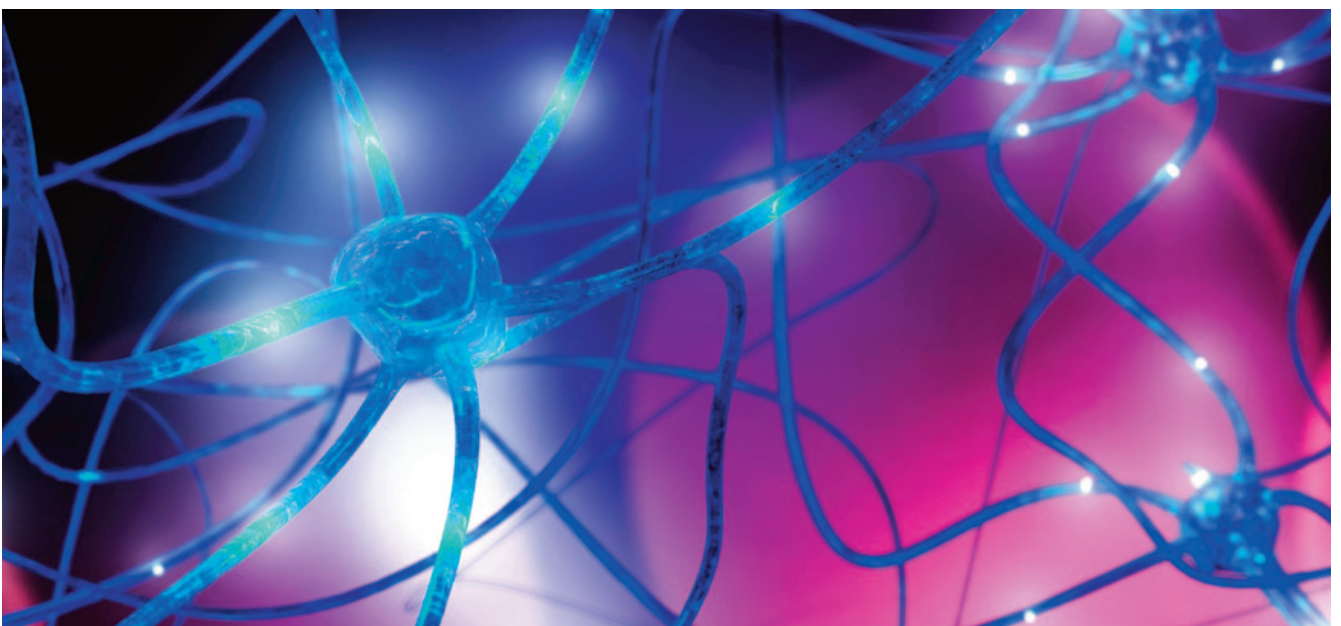
The combination of advances in genomic sequencing, artificial intelligence and big data has enabled further development in clinical practice, for example, by optimising drug therapies or improving the safety of treatments. In recent years, several studies have been published in this regard.

For example, mCSM [1] emerges as a tool that uses graph-based signatures to predict the effects of protein mutations. In this paper, mCSM is demonstrated to be effective in predicting changes in the stability of mutations in the p53 protein, outperforming other methods and showing its usefulness in complex disease contexts such as cancer.

Convolutional neural networks have also been used to extract local features and attention mechanisms to obtain important information in the field of drug-protein interactions (DPIs) [2].

In tests with datasets such as C.elegans, humanos, and BindingDB [3], the proposed approach was shown to improve the effectiveness in predicting DPIs compared to conventional machine learning methods. It is worth noting that such studies need to be fed by good databases, an area that has also been recently developed with datasets such as DDInter [4]. This is a curated database of drug-drug interactions (DDI) designed to help clinicians identify dangerous drug combinations and improve healthcare systems. In addition to basic queries, it incorporates a prescription verification function to help physicians decide whether drug combinations can be used safely.

This publication explores how graph theory and neural networks (both complex and graph neural networks) are transforming pharmaceutical research. It addresses current challenges in drug development, such as drug-protein interaction and drug-drug interaction prediction, and how methods based on these artificial intelligence techniques are providing new solutions. In addition, a concrete use case of drug repositioning is presented, highlighting the potential of these technologies to accelerate innovation and improve health outcomes.



Connection between networks and the pharmaceutical industry

Current overview of the pharmaceutical industry

In recent decades, significant advances in human genome sequencing and bioinformatics have catalysed an unprecedented transformation in the field of medicine. One of the main beneficiaries of this phenomenon is the pharmaceutical industry, which has emerged as a key player in Spain's growth and development. The growth of Spain's Gross Domestic Product (GDP) to 3% by 2023 [5], according to data from the National Statistics Institute (INE), is a clear indicator of the positive impact this industry has had on the country's economy. This economic growth is largely attributed to the pharmaceutical industry's ability to capitalise on scientific and technological advances, as well as to meet the health needs of the population through the production and marketing of innovative medicines.

In the field of clinical research into innovative medicines, national pharmaceutical laboratories play a leading role, with an investment of more than 750 million euros. Spain has a remarkably diverse pharmaceutical industry, with more than 100 manufacturers of basic pharmaceutical products and more than 200 companies specialising in the preparation of highly complex medicines. In addition, the country is home to major global pharmaceutical companies that have established subsidiaries in Spain [6].

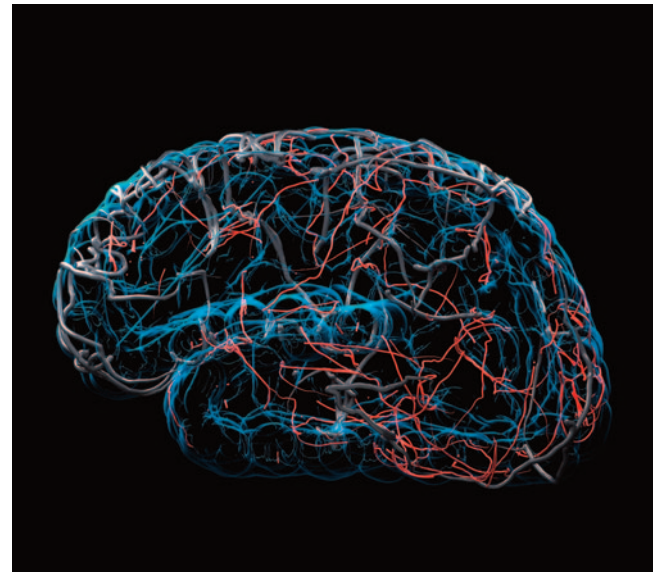
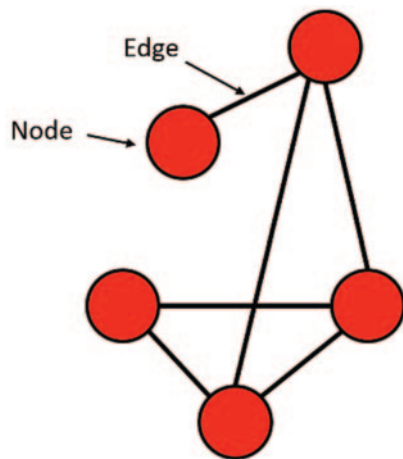
The development of new drugs in the pharmaceutical industry, a process that can take 10-15 years and require investments of up to \$800 million, faces considerable challenges in terms of time and resources [7]. To speed up this process, reduce costs and improve the effectiveness of existing drugs, numerous computational approaches based on artificial intelligence and massive use of data are being integrated into clinical practice. These approaches allow optimisation of drug therapy, improving the safety and efficacy of treatment for the benefit of patients. Within the pharmaceutical industry, numerous lines of research are being explored to achieve the objectives, with three main approaches: drug-protein interaction prediction, amino acid mutation prediction and drug-drug interaction prediction. This convergence between technological innovation and biomedical research promises to further revolutionise the field of medicine, driving significant advances in treating diseases and improving people's quality of life.

Graph theory

Graph theory [8] is one of the fundamental pillars of discrete mathematics, whose origin is attributed to the famous Swiss mathematician, Leonhard Euler. It was he who introduced the concept of graphs in his work *Solutio problematis ad geometriam situs pertinentis* (1736) [9], to solve the problem of the seven bridges of Königsberg. Thus began a whole new branch of mathematics, which continued to develop during the 19th and early 20th century, thanks to the contributions of mathematicians such as Arthur Cayley and Gustav Kirchhoff in the study of connected trees and graphs [10] [11]. In the second half of the 20th century, graph theory began to be applied in computer science, especially in network algorithms, optimisation and theory of computation. This was consolidated with works such as Dijkstra's [12] [13] [14] and his algorithm for finding the shortest path in a graph. Starting in the 1990s, and including the study of real interaction networks, the science of complex networks [15] [16] emerged, extending graph theory by incorporating the dynamics and evolution of these graphs over time. A complex network refers to a network, modelled as a graph, that possesses certain non-trivial statistical and topological properties. Complex networks today are used in the study of critical phenomena in statistical physics, in bio-inspired problem solving or in the social sciences [17], exploring how systems behave under different conditions. Thus, graph theory and complex networks coexist in such a way that the former serves as a mathematical foundation and the latter extends its scope to address dynamic and applied problems in the real world.

A graph [15] is a mathematical object consisting of a set of nodes (or vertices) and a set of edges (or links) connecting pairs of nodes. They are represented by diagrams like the one in Fig. 1, where nodes are associated with points and edges with lines connecting connected nodes. They allow very complex connectivity structures to be represented in a simple way. This is why graphs are essential in the study of complex networks, as they provide the mathematical basis necessary to model, understand and analyse the phenomenon under study.

Fig. 1 Diagram of the structure of a graph. The circles (red) are the nodes and the edges (black) are the links between nodes.



It is important to understand the concepts of nodes and edges. **Nodes** are the basic elements of a network and represent the individual entities within the system. Their definition depends largely on the network under study and its context. It is not the same to study the brain considering each neuron as a node, or defining each node as a brain region. **Edges**, on the other hand, represent the interactions or links given between nodes. These relationships can be physical, as in transport networks, or abstract, as in friendship networks.

Graph theory and complex networks have become tools with great potential for understanding and analysing systems in a wide variety of fields.

In **social science**, social network analysis allows to understand how information is disseminated, how communities are structured and the influence of certain individuals within a social network. This can be applied in politics, to understand the dynamics of social movements and the propagation of ideas, or in the field of digital social networks, to understand how connections between users are formed and evolve, allowing for improved algorithms for content recommendation and personalisation.

Complex networks have become essential in various research works in **biology and medicine**. They can be used to study interactions between proteins, neural networks or the propagation of diseases. The latter are crucial for modelling the spread of diseases, allowing to anticipate and design more effective intervention strategies, such as vaccination campaigns or quarantine measures.

Technological networks are another example. They include internet and transport networks. These networks help to optimise the efficiency and robustness of these systems to failures. The structure and operation of the global Internet, for example, is based on the principles of graphs and complex networks.

In **economics and finance**, network theory provides tools for modelling the complex interactions in financial markets and economic transactions. This allows economists and financial analysts to identify potential points of systemic risk and develop strategies to mitigate the impact of financial crises.

Finally, in **ecology**, they are used to model interactions within ecosystems, such as food webs. This helps to understand the stability of ecosystems and how different species may depend on each other, allowing the detection of keystone species for the preservation of the natural environment.

A common denominator in all these fields is the idea that complex networks allow to understand the underlying interactions and dynamics of the case study. While in other modelling domains it is necessary to resort to feature engineering techniques to understand the relationship between variables, graph networks inherently carry that relationship because of the way they are constructed. Initially, each node houses the information and characteristics that define it exclusively, but the very structure of the network predetermines what and how the different interactions with its environment will be, so that, ultimately, the feature vectors of each node will incorporate the information relating to its environment and how it interacts with its neighbours. In summary, graph theory and complex networks not only offer a way to represent and analyse complex structures and systems, but also provide crucial insights for predicting behaviour and designing effective interventions in a variety of fields.

Complex networks and graph neural networks

The evolution of graph theory has led to the development of complex networks and graph neural networks, tools for modelling and analysing dynamic systems and in various fields, including pharmaceutical research. These networks provide a robust framework for understanding complex interactions between different elements of a system, such as molecules, proteins and drugs, through the use of advanced artificial intelligence methods.

Complex networks

Complex networks are graphs characterised by non-trivial connection patterns that reflect the properties and relationships inherent in real systems. Unlike simple graphs, complex networks include features such as modularity, robustness and adaptability over time. These properties are essential for modelling biological and chemical systems where interactions are neither static nor homogeneous:

- 1. Structure and Dynamics.** Complex networks integrate both the topology (static structure) and dynamics (behaviour over time) of systems. In the pharmaceutical context, this allows modelling how drugs interact with multiple proteins and how these interactions may change due to mutations or the presence of other molecules.
- 2. Applications in Biomedicine.** Applications of complex networks in biomedicine include identifying new gene-disease relationships, modelling the spread of infectious diseases and optimising combination therapies. These models help predict side effects and personalise medical treatments.

Graph Neural Networks

Graph neural networks (GNNs) represent an evolution of machine learning methods that can operate directly on graph structures. GNNs can learn representations of nodes and edges that reflect the characteristics and relationships of these elements in the graph.

- 1. GNN capabilities.** Graph neural networks are particularly effective for tasks where structural relationships between data are crucial. For example, in predicting drug-protein interactions, GNNs can integrate molecular structure and bioactivity data to make accurate predictions about how a drug will interact with a target protein.
- 2. Integration with Biomedical Data.** Combining GNNs with biomedical databases allows the extraction of complex features and the identification of patterns not evident at first sight. This is useful for predicting the effects of mutations, where GNNs can model how an alteration in amino acid sequence affects the structure and function of a protein.



Applications in pharmaceutical research

In recent decades, complex networks have emerged as a powerful modelling tool. Some applications stand out, such as percolation, synchronization, epidemiological processes, and phase transitions, among others. The development that complex networks have undergone in these fields has also allowed them to expand their scope of application, being also used in the modelling and analysis of interactions between molecules, proteins and diseases, which offers new perspectives and innovative solutions for drug discovery and drug design. Some of the most promising proposals in this field are outlined below.

Drug-protein prediction

When a drug (a chemical molecule) binds to a biological target (such as proteins), it modulates its behaviour with the aim of returning it to its normal state. Predicting this drug-target interaction (DPI) is a very important step in the process of discovering new drugs and understanding their side effects [2] [19]. A clear example of the importance of investigating drug-protein interactions is the emergence of the COVID-19 pandemic. In the absence of specific drugs to combat this disease, detailed investigation of the interactions between SARS-CoV-2 proteins and various existing drugs led to the identification of compounds with therapeutic potential. Detailed study of the SARS-CoV-2 spike protein, as well as other key viral proteins such as the 3CLpro protease, enabled the development of antiviral therapies and guided the design of new targeted inhibitors [20].

Within this field of study, drug-protein interaction networks have emerged, networks aimed at representing the interactions between drugs and target proteins in the human body. The nodes correspond to the drugs and proteins, and the edges represent the different interactions between them. The development of these networks can help predict the efficacy and side effects of drugs, as well as identify possible drug combinations for new therapies.

Most approaches rely on drug and protein information represented by feature vectors. For example, one of the common approaches often formulated is to obtain DTI predictions as a binary classification task. These approaches make use of databases such as BindingDB that contain information on binding interactions between proteins and ligands, which are small molecules such as drugs. Their main goal is to provide detailed experimental data on binding affinity, which is a measure of how strongly a ligand binds to a protein [3]. KIBA (Kinase Inhibitor BioActivity) is a collection of data designed to facilitate research in the field of drug discovery, particularly in relation to kinase inhibitors. KIBA integrates information from several bioactivity databases and standardises interaction data between kinases and their inhibitors [21].

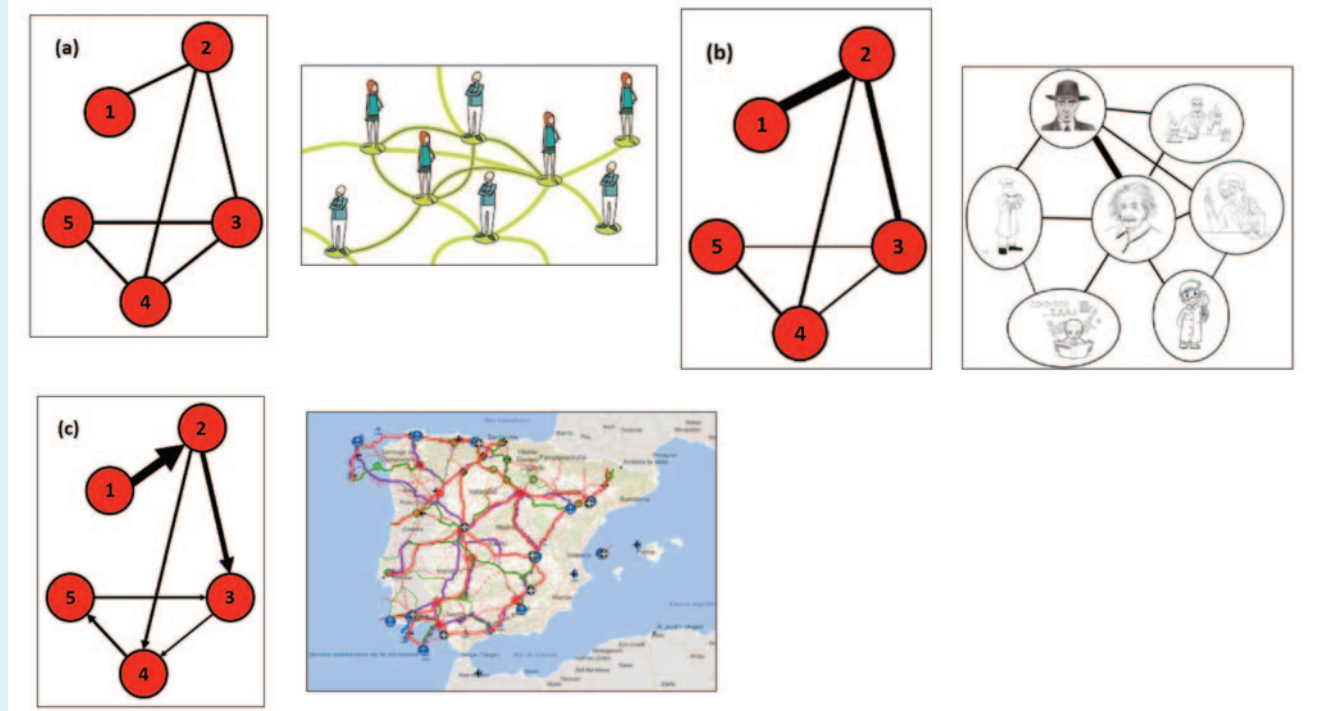


Properties of graphs

Graphs exhibit several properties that allow the characterisation and study of network behaviour in a variety of practical and theoretical applications. The main properties of graphs include:

- Degree of a node: indicates how many edges impinge on it.
 - Connectivity: measures the robustness of the graph to node or edge deletion.
 - Clustering coefficient: evaluates the tendency of nodes to form clusters.
 - Centrality: measures the importance of a node in the network.
 - Communities: are densely connected subsets of nodes.
 - Modularity: quantifies the structure of the network in communities.
 - Paths: represent sequences of edges between nodes. There are also those known as cycles or closed paths.
 - Average path length: is the average of the distances between all pairs of nodes.
 - Diameter: the longest distance between two nodes.
 - Pearson correlation: a metric by which the relationship between the properties of connected nodes is assessed.
- Directed or non-directed: the latter are bidirectional, while in the former it is of relevance which is the starting and end node in a link, as in a social network where one user can follow another without reciprocity.
 - Weighted or non-weighted: depending on whether all links are treated equally or whether some are given more importance than others.
 - Connected or unconnected: if at least one pair of nodes is not connected by a path, the network is unconnected.
 - Homogeneous or heterogeneous: depending on whether different types of elements are present, such as a biological network that includes different types of molecules and interactions.
 - Static or dynamic: if the network structure changes over time.
 - Scale-free: these are those that follow a power law; some nodes have many connections, while most have few.

Fig. 2. Examples of different types of complex networks.



(a) Network of friends at school. It is an undirected network - being friends is a two-way relationship - and not weighted - all friendships are of equal value. It can be non-connected if there are people in class who do not have any friends. It is homogeneous and dynamic - interpersonal relationships vary over time. It does not normally follow a power law, so it is not scale-free.

(b) Network of scientific collaborations. It is an undirected network - the fact of collaborating in a publication implies both - and weighted - those collaborations that are more cited are considered more relevant. Moreover, it is a connected network, homogeneous -one type of nodes or scientists and one type of links or collaborations-, dynamic and scale-free -scientists usually have few citations, although some are highly cited-.

(c) Spanish transport network [18]. In this case, we are dealing with a directed network - journeys have an origin and destination - and a weighted network - there are routes that are more relevant than others. It is also a connected network, heterogeneous -different types of passengers and transport methods-, dynamic -it develops over the years- and scale-free -there are places with more connections than others.



Prediction of amino acid mutations

Sequencing of the human genome has revealed great genetic diversity in human populations, including mutations that alter the sequence of amino acids in proteins. These mutations, called nonsense variants, can be pathogenic, affecting protein function and patient health, or benign, with minimal effects [22]. More than 15 million nucleotide variations have been catalogued in the human population, yet many variants are still unknown, and many others still have an unknown impact, representing a challenge in human genetics [23]. The human immunodeficiency virus (HIV), responsible for 630,000 deaths in 2022 [24], has a high rate of mutations in its proteins. Antiretroviral drugs approved to combat HIV are designed to inhibit two specific proteins. However, mutations in the amino acids of these proteins can alter their structure, causing the drugs to lose their effectiveness and stop working against the virus [25].

Computational methods are being used to study the molecular mechanism of mutation-induced drug resistance and to develop predictive tools to detect mutations that affect amino acid sequence and thus protein structure and function [26]. One approach that has already achieved good results is the mCSM method [1] [27] (mutation cutting scanning matrix), which uses graph-based signatures to represent the wild-type structural environment and machine learning to predict the effect of mutations on protein stability. In this context, proteins are represented as networks where the nodes are atoms and the edges are the interactions between these atoms. By analysing how mutations affect the local and global properties of these networks (e.g. changes in connectivity or in the distribution of interactions), mCSM can predict the impact of mutations on protein stability and function. Extensions of the mCSM method have been developed in Cambridge to predict the impact of mutation on protein-ligand (mCSM-lig) [28] and protein-protein interactions (mCSM-PPI2) [29], and more recently in Fiocruz, Brazil and Melbourne, Australia, by Douglas Pires and David Ascher for protein-nucleic acid (mCSM-NA) [30], antigen-antibody (mCSM-AB) [31] interactions and conformations and dynamics of proteins in combination based on normal mode dynamics (DynaMut) [32].

To develop this line of research, there are several databases, among the most prominent of which are UniProtKB, which houses an extensive collection of detailed and curated information on proteins. It provides data on sequences, functions, three-dimensional structures, cellular localization, interactions with other molecules and biological annotations

[33]. DbSNP focuses on genetic variations, including mutations and single nucleotide polymorphisms (SNPs) [34]. The Human Gene Mutation Database contains detailed information on human genetic mutations, including data on the exact location of the mutation in the genome, the nature of the mutation (e.g. whether it is a single nucleotide substitution, insertion or deletion), and any known effects of the mutation on gene function or human health [35].

Drug-drug interaction prediction

The combination of two or more drugs is known as combination therapy and is a common strategy to improve therapeutic efficacy and reduce side effects. However, inappropriate drug selection may result in adverse reactions. Therefore, knowledge of drug-drug interactions (DDIs) is of particular interest [36]. DDIEffects are an important risk factor for hospitalization, especially among elderly outpatients. In fact, it is estimated that DDIs contribute to 5-14% of adverse reactions in hospitalized patients [37]. An example of this dangerous IDD could be the combination therapy of warfarin (anticoagulant that prevents thrombus formation) with non-steroidal anti-inflammatory drugs (reduce inflammation, pain and fever) such as ibuprofen, which can cause bleeding by inhibiting the metabolism of the anticoagulant [38].

Drug-drug interaction networks can be used in this line of research. The underlying idea is similar to the previously exposed, with the particularity that the nodes represent drugs and diseases, and the edges indicate the association between them. In this case, the intention is to look for possible side effects and, once these side effects are located, and we understand why they arise, to find out how to prevent them or find a possible solution. Predictive algorithms take advantage of pharmacokinetic and pharmacodynamic data to predict how drugs will interact in the human body [39], so they can identify patterns and correlations that may not be apparent to traditional methods.

The data used in studies such as the one mentioned above are extracted from large interaction databases known as DrugBank and DDInter. DrugBank has a total of 570,091 pharmaceuticals including approved small molecule drugs and experimental drugs, among others [40]. It provides more than 200 data fields for each drug, with half of the information devoted to chemical, pharmacological, pharmaceutical and other aspects of the drug and the other half devoted to documenting the sequence, structure and route of the target drug [41]. DDInter contains more than 236,834 IDD involving 1833 drugs and documents detailed information on each DDI such as mechanisms, risk levels, recommendations for drug matching, etc. [4] [42].

Use case: Drug repositioning

Drug repositioning consists of identifying new therapeutic uses for drugs that have previously been approved for different medical purposes [43]. Until now, these studies relied on manual screening and statistical techniques, methods that required considerable time and effort. However, new computational methods make it possible to analyse large volumes of genetic and clinical data, discovering new drug targets and predicting the efficacy and toxicity of compounds with high accuracy. This accelerates the drug development process.

Deep learning algorithms such as those based on graph neural networks are used to study drug repositioning. These tools have been used to identify Janus Kinase 2 (JAK2) inhibitor drugs [44]. Impaired JAK function is associated with various inflammatory disorders, such as rheumatoid arthritis and psoriasis, among others [45]. Inhibition of this enzyme reduces the negative consequences of these diseases.

In this particular use case, data from the DUD-E database [46], which contains physicochemical properties of different drugs already approved by the FDA (Food and Drug Administration), have been used for training a graph convolutional network (GCN) provided by DeepChem [47], an open-source Python library with the GraphConvMol model. These data have been used to train a graph convolutional network (GCN) provided by DeepChem [47], an open-source Python library with the GraphConvMol model. This model is intended to determine the inhibitory capacity of drugs based on their molecular structure.

This graph model converts each atom of the drug into a node and each covalent bond into an edge, such that each atom communicates its unique characteristics to adjacent atoms [48]. The JAK2 dataset was split into training, validation and test sets in a ratio of 8:1:1 and then subjected to the GraphConvMol model using cross-validation with 5 iterations.

The trained model, using DeepChem's GraphConvMol algorithm, processed drugs approved by the US FDA to assess their potential JAK2 inhibitory activity. 20 active compounds were identified as having JAK2 inhibitory activity, including some already known. The JAK2 inhibitory activity of the 20 detected drugs was evaluated experimentally and all of them showed inhibition of JAK2 enzyme activity.

By integrating NGS with molecular docking techniques and applying it to a database of active compounds, a more comprehensive screening of potential JAK2 inhibitors has been achieved. This approach has enabled the efficient identification of new drug candidates that had not previously been considered. These compounds include ribociclib, amodiaquine, topiroxostat and gefitinib, all of which have shown promising JAK2 inhibitory potential. Furthermore, experimental validation has corroborated the results obtained by deep learning and molecular docking. Therefore, this procedure is proposed for drug repositioning across a broad spectrum of therapeutic targets.



Conclusions

vGraph theory and its evolution through artificial intelligence techniques have proven to be a tool with great potential in the pharmaceutical industry, successfully tackling complex challenges in drug-drug interaction prediction, genetic mutation analysis and drug repositioning. Its integration with advanced technologies and the analysis of large volumes of data has significantly optimised drug research and development.

Recent studies have validated the effectiveness of graph-based methods for predicting the effects of protein mutations and improving the stability of complex mutations. In addition, convolutional neural networks and other machine learning approaches have improved the prediction of drug-protein and drug-drug interactions, which are crucial for the development of safe and effective therapies.

In addition, the creation of databases has facilitated the identification of dangerous drug combinations and the optimisation of drug therapies. These advances underline the importance of continued collaboration between bioinformatics, graph theory and the pharmaceutical industry to improve health outcomes and accelerate innovation.

In short, graph theory has not only revolutionised the field of pharmaceutical research but has also opened new opportunities for the development of more effective and safer treatments. With their ability to model and analyse complex systems, graphs will continue to play a crucial role in the evolution of medicine and public health.



References

- [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, May 14). The challenges facing the Spanish pharmaceutical industry - Salud35. *Salud 35*. https://www.consalud.es/salud35/economia/industria-farmaceutica-espanola-crecimiento-economico_143765_102.html
- [6] Impact and relevance of the pharmaceutical industry in Spain. *Laboratorios Rubió*. <https://www.laboratoriosrubio.com/industria-farmaceutica-en-espana/>
- [7] The pharmaceutical industry in Spain - Statistical data (n/d). *Statista*. Retrieved June 11, 2024, from <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 (n.d.). <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] Global HIV data - EresVIHda (2023, 26 October). 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). Prediction of mutations in HIV reverse transcriptase protein sequences using new methods for Structural Learning Bayesian Networks. *Avances en Sistemas e Informática Journal*, 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, Aadhavya 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 (n.d.). <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. (n. d.). <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

Authors

Ernestina Menasalvas (UPM)
 Manuel Ángel Guzmán (Management Solutions)
 Sergio Ruiz (Management Solutions)
 Guillermo Fragio (Management Solutions)



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID

MSO *Management
Solutions*
Making things happen

The Universidad Politécnica de Madrid is a multi-sector and multi-disciplinary Public Law Entity, which carries out teaching, research and scientific and technological development activities.

www.upm.es

Management Solutions is an international consulting firm, focused on business, finance, risk, organization, technology and process consulting, operating in more than 50 countries and with a team of close to 4,000 professionals working for more than 2,000 clients worldwide.

www.managementsolutions.com

For more information visit

blogs.upm.es/catedra-idanae/