

Un enfoque de Inteligencia Artificial para la predicción de fármacos antivirales para el COVID-19

Rafael Alí Adorno Alvarez¹, Dr. Diego H. Stalder², Dr. Diego Galeano²

¹Alumno, Facultad de Ingeniería. Universidad Nacional de Asunción. San Lorenzo, Paraguay - rafaliaa8@gmail.com

²Orientador, Facultad de Ingeniería. Universidad Nacional de Asunción. San Lorenzo, Paraguay

Palabras clave:

- Sistemas de Recomendación
- Reposicionamiento de Fármacos
- Descomposición Matricial
- GcNMF

Resumen

Enfermedades infecciosas emergentes como el COVID-19, causado por el virus SARS-CoV-2, requieren de estrategias sistemáticas que ayuden al descubrimiento de curas y tratamientos. El reposicionamiento de fármacos, el proceso de búsqueda de nuevas indicaciones terapéuticas para los fármacos comercializados, es una alternativa prometedora al desarrollo de nuevos fármacos, con menores costos y tiempos de desarrollo más cortos. En este estudio propongo un sistema de recomendación llamado *geometric confidence non-negative matrix factorization* (GcNMF) para asistir en el reposicionamiento de fármacos antivirales de amplio espectro. GcNMF modela la estructura no-euclidiana de los fármacos mediante el uso de grafos y produce una lista priorizada de fármacos para virus específicos. Mis experimentos revelan que GcNMF posee un rendimiento de predicción superior a otros métodos de descomposición matricial. Esto sugiere que GcNMF podría asistir a expertos farmacólogos en la búsqueda de fármacos eficaces contra el COVID-19 específicamente y más en general contra enfermedades virales.

Keywords

- Recommender Systems
- Drug Repositioning
- Matrix Decomposition
- GcNMF

Abstract

Emerging infectious diseases such as COVID-19, caused by the SARS-CoV-2 virus, require systematic strategies to assist in the discovery of effective treatments. Drug repositioning, the process of finding new therapeutic indications for commercialized drugs, is a promising alternative to the development of new drugs, with lower costs and shorter development times. In this paper, we propose a recommendation system called *geometric confidence non-negative matrix factorization* (GcNMF) to assist in the repositioning of 126 broad spectrum antiviral drugs for 80 viruses, including SARS-CoV-2. GcNMF models the non-Euclidean structure of the space using graphs, and produces a ranked list of drugs for each virus. My experiments reveal that GcNMF significantly outperforms other matrix decomposition methods at predicting missing drug-virus associations. My analysis suggests that GcNMF could assist pharmacological experts in the search for effective drugs against viral diseases.

1. Introducción y Motivación

En los inicios del año 2020 la humanidad ha tenido que enfrentar una pandemia, producto de la aparición de un nuevo coronavirus. Una afección a la salud que ha ocasionado millones de muertes debido a que al iniciar la pandemia no existían fármacos eficaces

para tratarla [1]. La Organización Mundial de la Salud (OMS) ha declarado al COVID-19, la enfermedad causada por el virus del SARS-CoV-2, como pandemia, el 11 de Marzo del 2020. Luego de casi 4 años de su aparición, COVID-19 se ha expandido a más de 230 países alrededor del mundo y ha causado más de 6,9 millones de muertes a nivel mundial, con mas de

769 millones de infecciones [2].

Al considerar además las consecuencias a la economía que ocasiona una pandemia tan extendida en el tiempo, se observan cifras realmente abrumadoras. Según datos de Banco Mundial [3] la pobreza mundial se incrementó por primera vez en el curso de una generación, el desempleo temporal se elevó en el 70 % en algunos países, la pérdida de ingresos fue mayor entre los jóvenes, entre otros efectos colaterales. Por lo que la naturaleza atípica de la pandemia, su expansión acelerada y permanencia de esta enfermedad en el tiempo, ha revelado la necesidad de un rápido desarrollo, prueba y despliegue de nuevos medicamentos y curas [4]. Sin embargo el desarrollo de nuevos fármacos presenta varios desafíos, que incluyen costos elevados —en el orden de los billones de dólares— y tiempos largos para los ensayos requeridos a fines de certificar su uso seguro en seres humanos, proceso que normalmente puede prolongarse de 10 a 15 años [5–7]. Y si bien se han desarrollado vacunas y se han habilitado procesos de certificación de fármacos para uso de emergencia en plazos significativamente más cortos, aún la eficacia de estos a largo plazo está demostrada que disminuyen [8], por lo que el desarrollo de fármacos nuevos no puede ser la única solución factible a corto plazo. Una estrategia eficiente en el tiempo debe basarse en el reposicionamiento de fármacos [1, 4].

El reposicionamiento de fármacos ha surgido como una alternativa prometedora al desarrollo de fármacos nuevos [9, 10], y consiste en el proceso de búsqueda de nuevas indicaciones terapéuticas para fármacos ya comercializados [11, 12]. El reposicionamiento de fármacos implicaría menores costos y tiempos de desarrollo mucho mas reducidos [10].

El reposicionamiento de fármacos ha sido uno de los focos de la atención de la comunidad de Inteligencia Artificial (IA) desde el inicio de la pandemia. Varios enfoques basados en redes neuronales y medicina de redes han sido propuestos para el reposicionamiento de fármacos [14]. Por ejemplo, Ke et al. (2020) [15] propusieron un enfoque de deep learning para predecir compuestos con actividades potenciales que inhiban el SARS-CoV-2. Beck et al. (2020) [16] utilizaron un modelo de predicción de interacción fármaco-objetivo¹ para identificar medicamentos antivirales comerciales que podrían alterar los componentes virales del SARS-CoV-2. La medicina en redes por su parte involucra la construcción de grafos de conocimiento médico, que contienen relaciones entre diferentes tipos de entidades médicas (por ejemplo, enfermedades, medicamentos y proteínas) y predice nuevos vínculos entre medicamentos y enfermedades existentes [14]. Gysi et al. (2020) [17] desarrollaron un método que se basa en una red neuronal de grafos

(del inglés *Graph Neural Networks* o GNNs) y presentaron un estudio de caso sobre el SARS-CoV-2. Be-nevolentAI [14] construyó un grafo de conocimiento integral de COVID-19 (denominado CoV-KGE) que incluía 15 millones de aristas en 39 tipos de relaciones que conectan medicamentos, enfermedades, proteínas, genes, vías y expresiones de genes y proteínas.

En este documento presento un sistema de recomendación para asistir en el reposicionamiento de antivirales de amplio espectro contra virus específicos. Los sistemas de recomendación han sido históricamente desarrollados para recomendar películas a usuarios en plataformas digitales como Netflix [18–21]. Estos funcionan analizando las similitudes en preferencias de los usuarios hacia ciertas películas para sugerirles nuevos contenidos. Mi sistema recomienda fármacos a los virus, basándose en las asociaciones terapéuticas entre ellos, modelando de esa manera las preferencias de los virus hacia los fármacos.

Recientemente Santos et al. (2021) han desarrollado un sistema de recomendación basado en descomposición matricial para predecir antivirales efectivos contra virus específicos. Este modelo [22], denominado cNMF (del inglés *confidence Non Negative Matrix Factorization*) se basa en la idea de modelar las probabilidades de éxito de una asociación fármaco-virus dependiendo de su etapa de desarrollo. La limitación de aquel modelo, que se basa en el modelo de descomposición matricial no-negativa de Lee & Seung [23], es que no modela potenciales relaciones no-euclidianas entre los puntos en el espacio, es decir, fármacos. Aquí propongo *Geometric confidence NMF* (GcNMF), un modelo de descomposición matricial basado en (Santos et al., 2021) que modela la estructura no-euclidiana del espacio. GcNMF se inspira en trabajos previos que utilizaron grafos en un marco de regularización para incorporar información sobre la estructura geométrica intrínseca de los datos en modelos de aprendizaje de máquina [24–27]. Mediante una serie de simulaciones computacionales, demuestro la superioridad de GcNMF sobre otros métodos de descomposición matricial, incluyendo el método cNMF [22] desarrollado recientemente (Santos et al., 2021), para predecir antivirales efectivos contra virus específicos, en particular contra el Sars-CoV-2.

El resumen técnico se estructura de la siguiente forma. En la siguiente sección se presentan el conjunto de datos que relacionan antivirales de amplio espectro con diversos virus, el modelo de descomposición matricial cNMF (Santos et al., 2021) y mi método: *Geometric confidence NMF*. A continuación se presenta la metodología adoptada para la evaluación y selección de los conjuntos de entrenamiento y prueba, los principales resultados obtenidos y en particular el análisis de las predicciones para COVID-19. Finalmente, en la última sección se discuten algunas con-

¹Las proteínas objetivo son biomoléculas funcionales que son tratadas y controladas por compuestos biológicamente activos

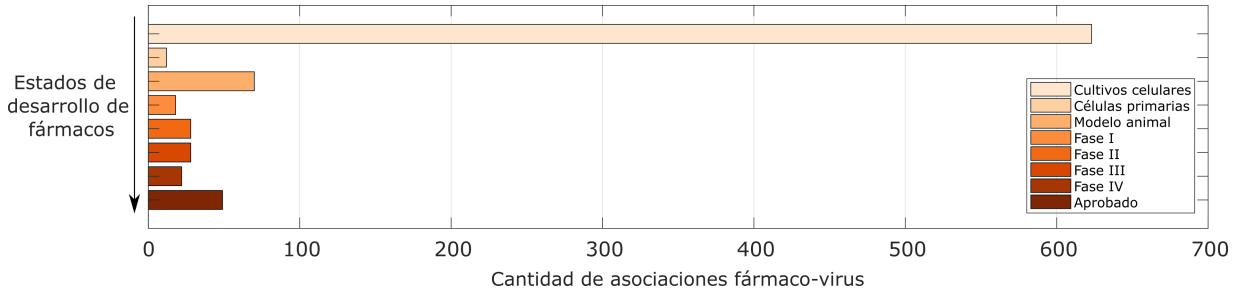


Figura 1: Estadística de los datos: Para este trabajo utilizamos el conjunto de datos manualmente curado por Andersen et al. (2020) [13]. Este gráfico muestra los números de asociaciones fármaco-virus para cada estado de desarrollo. El término “asociaciones” se refieren a la existencia de evidencia de actividad antiviral de un fármaco sobre un virus. **Fuente:** Elaboración propia.

sideraciones finales.

2. El Modelo Geometric confidence NMF

2.1. Datos

El conjunto de datos utilizado es el manualmente curado por Andersen et al. (2020) [13]. El mismo consiste en fármacos antivirales de amplio espectro (BSAA, del inglés *Broad Spectrum Antiviral Agents*) que aún están en desarrollo o han sido aprobados para distintos virus. Al codificar asociaciones conocidas de fármacos y virus con el número entero 1, ensamblamos una matriz Y de $n \times m$ que contiene 850 asociaciones (ver Figura 1) entre $n = 126$ fármacos y $m = 80$ virus. Las asociaciones restantes en Y se completaron con ceros.

2.2. El modelo de descomposición matricial

Los sistemas de recomendación de películas basados en descomposición matricial se basan en la premisa fundamental de que cada usuario y cada película pueden representarse como vectores de características latentes en un espacio de baja dimensión, tal que el valor de preferencia dado por un usuario a una película se obtiene mediante el producto escalar de dichos vectores. Intuitivamente, los vectores de características latentes capturan las preferencias de los usuarios hacia las películas, ej. ciencia ficción, comedia, acción. El reciente método (Santos et al., 2021) adopta un modelo similar para la matriz Y donde, cada fármaco i es representado con un vector w_i de dimensión q , y cada virus j es representado con un vector h_j de la misma dimensión, tal que el producto escalar entre ambos modela la eficacia del fármaco i contra el virus j (ver Figura 2).

De una forma mas concisa, dada una matriz de datos $Y = [y_1, \dots, y_m]^T \in \mathbb{R}^{m \times n}$, donde cada fila de Y es un vector de muestra. Se tiene como objetivo encontrar dos matrices $W = [w_{iq}] \in \mathbb{R}^{m \times q}$ y $H = [h_{qj}] \in \mathbb{R}^{q \times n}$ cuyo producto puede aproximarse a la matriz original Y , como sigue:

$$Y \approx \hat{Y} = WH, \quad (1)$$

donde q representa la cantidad de características latentes en el espacio generado al factorizar la matriz Y , y \hat{Y} representa la matriz de predicciones.

La tarea de la factorización matricial consiste por lo tanto en completar una matriz incompleta Y . Las predicciones del modelo constituyen los valores reales en \hat{Y} correspondientes a los elementos iguales a cero en Y .

Este modelo de descomposición matricial aprende W y H de la ecuación (1) mediante la minimización de la siguiente función de costo:

$$\min \mathcal{L} = \sum_s \frac{\alpha_s}{2} \|M_s \circ (Y - WH)\|_F^2, \quad (2)$$

tal que $W, H \geq 0$,

donde $\|\cdot\|_F$ es la norma Frobenius de una matriz o también denominada norma Euclidiana de una matriz, y \circ es el producto de Hadamard entre matrices. M_s es una matriz binaria utilizada para seleccionar subconjuntos de elementos en Y de acuerdo a su estado de I+D en el proceso de desarrollo de un fármaco. Los subconjuntos son $s \in \{A, B, C, D, E, F\}$, y representan los fármacos en estado: aprobados o fase IV de ensayos clínicos (subconjunto A), fase III (subconjunto B), fase II (subconjunto C), fase I (subconjunto D), cultivos celulares/células primarias/modelo animal (subconjunto E). Finalmente, los ceros en Y corresponden al subconjunto F . Los parámetros α_s codifican probabilidades de éxito, es decir, probabilidades de que dichos subconjuntos resulten en fármacos eficaces para los virus. La contribución principal de (Santos et al., 2021) consiste en modelar las probabilidades de éxito de diferentes conjuntos de elementos en Y de acuerdo a sus etapas de I+D. Es bien sabido que las probabilidades de éxito de un fármaco candidato depende de su nivel de evidencia experimental. Por ejemplo, aquellos fármacos con evidencia experimental de eficacia en cultivos celulares son menos probables de ser aprobados por la FDA (del inglés *Food and Drug Administration*) que otros que ya

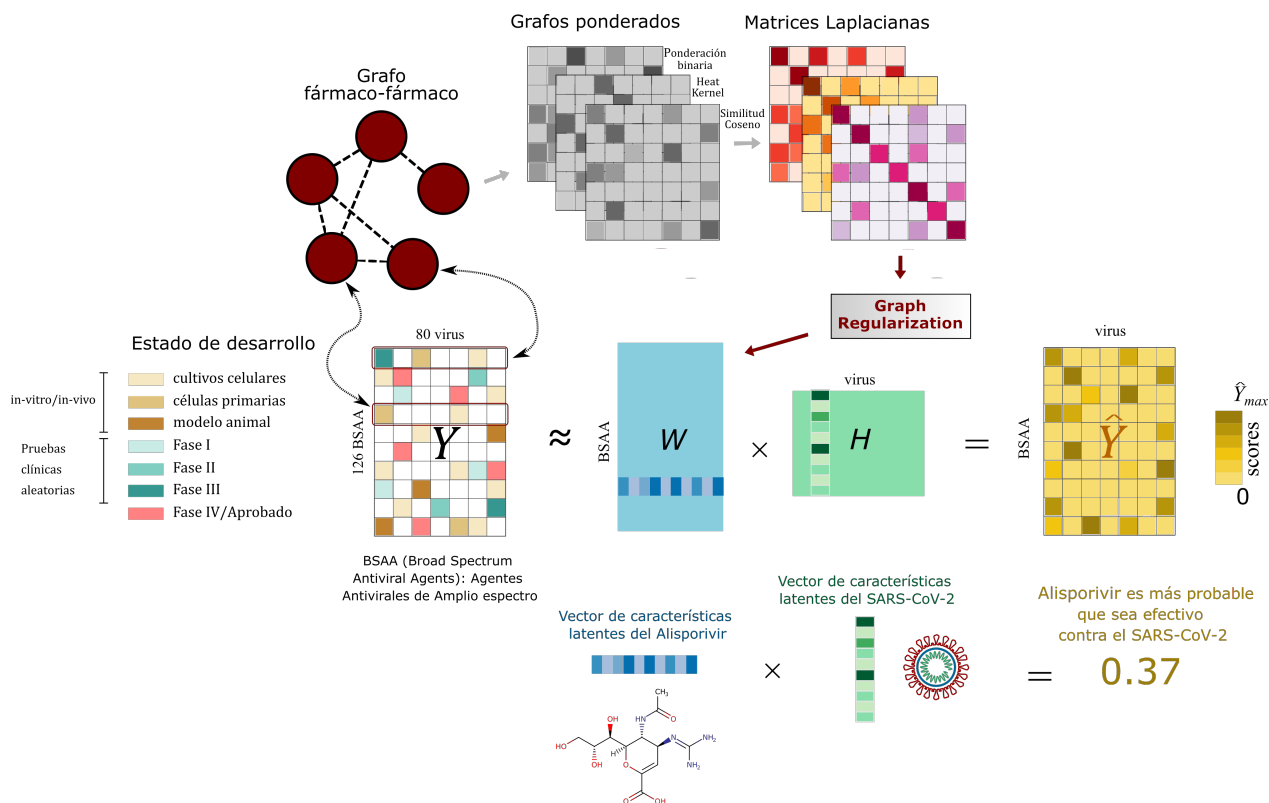


Figura 2: Mi modelo de descomposición matricial: Del conjunto de datos de Andersen et al. (2020) [13] utilizado para nuestro enfoque, se recogieron 850 asociaciones para $m = 126$ diferentes agentes antivirales de amplio espectro (BSAA) y $n = 80$ virus distintos. Las asociaciones observadas se organizaron en una matriz Y de $m \times n$ estableciendo $y_{ij} = 1$. Las asociaciones no observadas se codificaron con ceros. Mi algoritmo descompone la matriz Y en el producto de dos matrices, W (de tamaño $m \times q$) y H (de tamaño $q \times n$). Al multiplicar las matrices W y H , obtenemos \hat{Y} —que modela la matriz Y —, donde todos los elementos cero de Y fueron reemplazados por números reales, estos corresponden a nuestras puntuaciones predichas. La ilustración superior muestra el marco de regularización implementado. Construí un grafo para capturar las relaciones no-euclidianas entre los puntos del espacio, es decir, entre los fármacos. Este grafo contiene m vértices, donde cada vértice corresponde a un fármaco que lo conectamos a otros k vértices (o vecinos) más cercanos, es decir, a otros fármacos. Obtenemos así un grafo fármaco-fármaco cuyas conexiones entre sí ponderamos de distintas maneras, y que nos ayuda a incorporar información no lineal a nuestro modelo de descomposición matricial y mejorar su rendimiento. Las filas de W son los vectores de características latentes del fármaco; las columnas de H son los vectores de características latentes del virus. La ilustración inferior muestra cómo nuestro modelo descubre un vector latente de baja dimensión para el fármaco antiviral Alisporivir, y un vector de latente de baja dimensión para el SARS-CoV-2, de modo que el producto punto de estas dos firmas modela la eficacia predicha de fármaco contra el SARS-CoV-2. **Fuente:** Elaboración propia.

se encuentran en alguna fase de ensayo clínico. La última restricción que (Santos et al., 2021) impone en las matrices W y H es la no-negatividad. Esta restricción tiene como objetivo favorecer la interpretabilidad biológica de los resultados [28].

2.3. Mi método: Geometric cNMF (GcNMF)

Una de las limitaciones del modelo de descomposición matricial de (Santos et al., 2021) es que no modela la geometría intrínseca de los datos, asumiendo que los datos se encuentran en una superficie no-euclidiana de baja dimensión. Muchas veces, los conjuntos de datos poseen una estructura subyacente de menor dimensión en un espacio euclidiano de alta dimensión [29]. El estudio de las técnicas para recuperar una estructura dimensional inferior se denomina *manifold learning* o reducción de dimensionalidad no lineal [29]. Los algoritmos que utilizan esta técnica se basan en la denominada idea de invariancia local [30], es decir, la suposición es que si dos puntos y_i y y_l son cercanos en la geometría intrínseca de la distribución de datos, también los mismos puntos w_i y w_l , representados en sus nuevas bases, serán cercanos [27]. Se ha demostrado que el rendimiento de modelos de descomposición matricial puede mejorarse significativamente si se explota la estructura geométrica intrínseca del conjunto de datos y se considera la invariancia local [27].

Cai et al. (2010) [27] demostró que la estructura geométrica local se puede modelar eficazmente a través de un grafo construido con técnicas denominadas de vecinos cercanos (kNN, del inglés *k-nearest neighbors*). Considere un grafo con m vértices, donde cada vértice corresponde a un fármaco (definido como un vector fila en la matriz Y) –ver Figura 2. La idea de kNN consiste en conectar vértices en el grafo utilizando los k -vecinos cercanos de cada vértice. La distancia utilizada para definir los vecinos cercanos es la distancia euclidiana entre las filas en Y . Hasta aquí, obtenemos un grafo fármaco-fármaco binario basado en kNN.

Aquellos vértices conectados en el grafo pueden ser ponderados de diversas formas. Para esto, definimos una matriz de ponderación P de $m \times m$, y siguiendo a (Cai et al., 2010) [27], utilizamos tres formas diferentes de definir P :

1. **Ponderación Binaria 0-1:** $P_{il} = 1$ si existe una conexión entre los nodos i y l . En caso contrario $P_{il} = 0$. Es el caso de ponderación más simple.
2. **Heat Kernel:** si existe una conexión entre los nodos i y l entonces ponderamos

$$P_{il} = e^{-\frac{\|y_i - y_l\|^2}{\sigma}} \quad , \quad \sigma = 2 \frac{\sum_{i,l} E_{il}}{m^2} \quad ,$$

donde E es una matriz cuadrada de $m \times m$ que agrupa las distancias euclidianas de todos los fármacos entre sí. Es decir, cada elemento de la matriz E es la distancia entre el fármaco y_i con otro y_l .

3. **Similitud Coseno:** si existe una conexión entre los nodos i y l entonces ponderamos

$$P_{il} = y_i^T y_l \quad ,$$

donde cada vector y está normalizado a 1. Esto es la similitud coseno.

Para el término de regularización de nuestra función de costo utilizamos la propuesta de (Cai et al., 2010) [27]. Siendo $w_i = [w_{i1}, \dots, w_{ik}]$ la representación de baja dimensión de y_i , en el espacio de características latentes generadas por sus nuevas bases, se puede utilizar la distancia euclidiana y constituir una función para medir cómo estas nuevas bases captan la información geométrica de nuestro conjunto de datos. Aquí \mathcal{R} representa la función regulador de grafo que será adherida a nuestra nueva función de costo.:

$$\mathcal{R}(W) = \frac{1}{2} \sum_{i,l=1}^m \|w_i - w_l\|^2 P_{il},$$

$$\mathcal{R}(W) = \text{Tr}(W^T L W),$$

donde $L = D - P$ es denominado grafo laplaciano, D es una matriz diagonal cuyos valores son las sumas de las columnas de la matriz P . Note que el término \mathcal{R} ayuda a que dos fármacos i y l conectados en el grafo (es decir $P_{il} > 0$) tengan características latentes similares, ya que se penaliza la diferencia entre w_i y w_l . Intuitivamente, el aprendizaje de las características latentes de los fármacos en W es condicionado a la estructura geométrica del grafo de fármacos, de allí el término *geometric* en GcNMF.

La función de costo de GcNMF es entonces:

$$\min \mathcal{L} = \sum_s \frac{\alpha_s}{2} \|M_s \circ (Y - WH)\|_F^2 + \frac{\lambda}{2} \mathcal{R}(W), \quad (3)$$

tal que $W, H \geq 0$,

cuyos términos tienen los mismos significados que los de la ecuación (2), exceptuando el parámetro $\lambda > 0$ que representa el control de la influencia del regularizador de grafo.

2.4. Algoritmo multiplicativo

Para minimizar la Ecuación (3), seguimos el procedimiento en [23, 28] utilizando las condiciones de Karush-Khun-Tucker (KKT). Las siguientes reglas multiplicativas minimizan la función de costo en (3):

$$W := W \circ \frac{\{\sum_s \alpha_s [M_s \circ Y]\} H^T + \lambda P W}{\{\sum_s \alpha_s [M_s \circ WH]\} H^T + \lambda D W}, \quad (4)$$

$$H := H \circ \frac{W^T \{\sum_s \alpha_s [M_s \circ Y]\}}{W^T \{\sum_s \alpha_s [M_s \circ WH]\}}, \quad (5)$$

partiendo de condiciones iniciales no negativas para W y H (distribución uniforme $[0, 0.1]$), las iteraciones de estas reglas de actualización encuentran una W y H tal que sean una factorización aproximada para Y al converger a un mínimo local de la función objetivo de la ecuación (3).

3. Entrenamiento del modelo y Análisis de los Resultados

La metodología adoptada para la evaluación y selección de los conjuntos de entrenamiento y prueba, además de los resultados principales, se describen a continuación.

3.1. Evaluación y comparación

Para medir la eficiencia predictiva de Geometric cNMF (GcNMF) he implementado el método de validación cruzada de LOOCV (del inglés *Leave One Out Cross Validation*). En este procedimiento, una sola asociación fármaco-virus conocida en Y es puesta en el conjunto de prueba a la vez. Note que en la matriz de entrenamiento dicha asociación será puesta en cero. Luego, para el virus en el conjunto de prueba, se obtiene una lista ordenada de fármacos basados en las predicciones en \hat{Y} . El rendimiento del modelo se mide en su capacidad de recuperar el fármaco asociado al virus que fue oculto en el conjunto de prueba entre los primeros N fármacos predichos (o top- N).

Siguiendo el procedimiento el modelo de descomposición matricial cNMF [22], el conjunto de prueba consiste en aquellas asociaciones de Y que corresponden a las fases IV de ensayos clínicos y aprobados (subconjunto A). Estos fueron seleccionados debido a que su eficacia antiviral ha sido comprobada. 71 asociaciones cumplen este criterio, donde fármacos han sido ya aprobados o se encuentran en la última fase de desarrollo para 26 virus. Siguiendo LOOCV, hemos removido cada una de estas asociaciones, una a la vez, entrenando el modelo con las asociaciones restantes, y generando las predicciones correspondientes.

Para el entrenamiento del modelo, se procedió a una búsqueda en grillas para los distintos parámetros, asignándose los siguientes conjuntos, $q = \{1, 5, 10, 15, 20, 25, 30\}$, $\alpha_E = \{0.01, 0.05, 0.1, 0.5\}$, $\alpha_F = \{0.5, 1, 2, 5, 10\}$ y $\lambda = \{0.1, 0.5, 1, 5, 10, 30, 50, 100, 150\}$. Estableciéndose así en el entrenamiento los siguientes parámetros para GcNMF: $q = 15$, $\alpha_E = 0.05$, $\alpha_F = 5$, y para los fármacos aprobados se asigna $\alpha_A = 1$. Los demás parámetros de mi modelo se establecieron en función de las probabilidades de éxito para los fármacos antiinfecciosos en distintas fases de los ensayos clínicos, datos proveídos por Dowden y

Munro [32]. Es decir, $\alpha_B = 0.73$ (fase III), $\alpha_C = 0.27$ (fase II), $\alpha_D = 0.15$ (fase I). Además, uso el esquema de ponderación de la similitud coseno una vez construido el grafo a través kNN. El número de vecinos más cercanos k se estableció en 13 y el parámetro de regularización λ en 0.1. La selección del esquema de ponderación se discutirá en una sección posterior.

Para medir el rendimiento de GcNMF, es decir, si recuperó el fármaco ya conocido que trata al virus, en la lista de clasificación de N fármacos recomendados, uso la función de recuperación definida como:

$$rec(N) = \frac{\# \text{ predicciones correctas en la lista de } N \text{ fármacos}}{T},$$

donde $T = 71$ es el número total de casos que se evaluaron y N es el parámetro que indica el número de fármacos mejor clasificados resultantes de nuestras predicciones.

Antes de hablar de cómo demuestro que mi modelo GcNMF puede mejorar el rendimiento predictivo de modelos basados en el método de descomposición matricial deberíamos prestar atención al reciente artículo de Sosnina et al. (2020) [33]. En el mismo se ha demostrado que la factorización matricial no negativa [23] (NMF, por sus siglas en inglés) y la descomposición de valores singulares truncados [31] (tSVD, por sus siglas en inglés) proporcionan un mejor rendimiento en la predicción de interacciones faltantes, en lugar de otros algoritmos utilizados para predecir datos faltantes en sistemas de recomendación. Cabe destacar que aunque (Sosnina et al., 2020) utilizaron un conjunto de datos diferente, que consiste básicamente en interacciones de actividad antiviral de molécula pequeña derivadas de ensayos celulares, pensamos que sería instructivo comparar el rendimiento de nuestro método contra tSVD, NMF y (Santos et al., 2021).

Finalmente se demuestra la mejoría al compararlo con los siguientes 4 algoritmos, también basados en el método de descomposición matricial: i) tSVD [31], ii) NMF [23], i) GNMF [27] y iv) (Santos et al., 2021) (cNMF, el modelo desarrollado recientemente –ver ecuación (2)).

Debido a las condiciones iniciales aleatorias en el problema de factorización de matrices, podríamos llegar a una solución diferente por cada vez que el algoritmo es ejecutado. Para lidiar con esta cuestión, decidimos ejecutar, tanto el algoritmo (Santos et al., 2021) como el GcNMF, 10 veces. Y calcular con ellos el performance promedio y su desviación estándar, para visualizar qué tan estables son los algoritmos. Este experimento podemos observarlo en la Figura 3, donde se obtuvo que el performance para (Santos et al., 2021) es: $20.5 \pm 2.6\%$ para el top-1, $45.7 \pm 3\%$ para el top-5, $58 \pm 2.8\%$ para el top-10, $68 \pm 4.2\%$ para el top-20, $77 \pm 3\%$ para el top-30. Y el performance para el GcNMF: $52.6 \pm 1.5\%$ para el top-1, $60 \pm 1.4\%$ para el top-5, $70 \pm 2.8\%$ para el top-10, $74 \pm 2\%$ para el top-20, $79 \pm 2\%$ para el top-30.

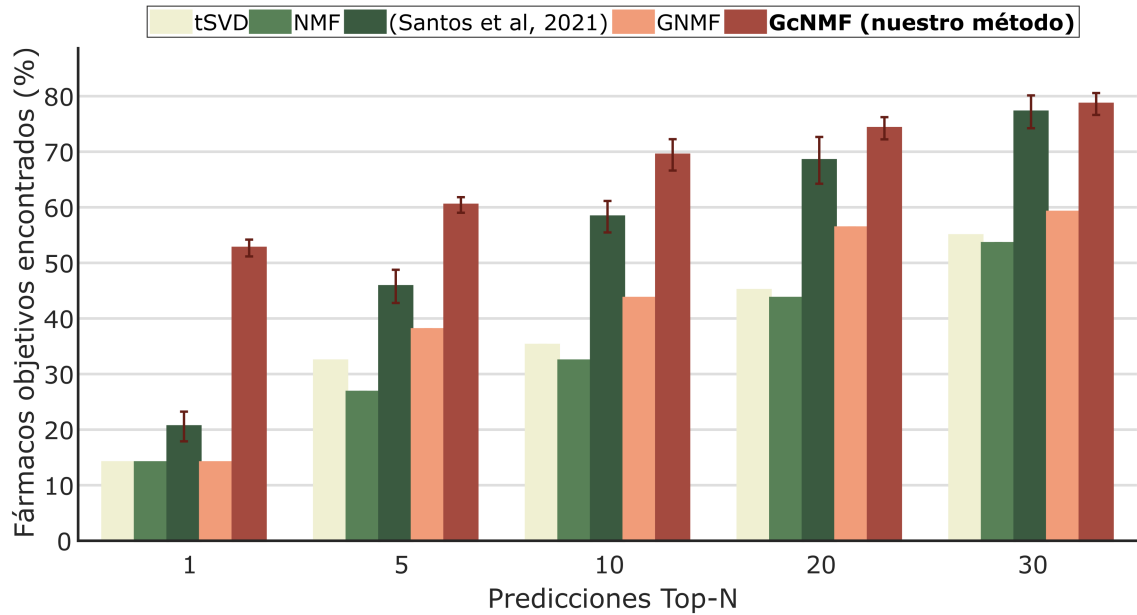


Figura 3: Evaluación del rendimiento de la predicción de antivirales: el gráfico nos muestra la recuperación de fármacos conocidos que tratan a los virus, de la lista de recomendación top- N de fármacos predichos, para cada uno de los métodos y principales top- N . Los métodos basados en descomposición matricial comparados con el propuesto en este artículo son: *truncated singular value decomposition* (tSVD) [31], *non-negative matrix factorization* (NMF) [23], *confidence non-negative matrix factorization* (cNMF) (Santos et al., 2021) (ver ecuación (2)) y *geometric non-negative matrix factorization* (GNMF) [27]). **Fuente:** Elaboración propia.

Los resultados visualizados en la Figura 3 muestran la recuperación de fármacos por virus en diferentes predicciones top- N . Podemos ver que mi modelo propuesto supera a todos los competidores en cada una de las predicciones: en un 32 – 38 % en el top-1, en un 14 – 33 % en el top-5, en un 11 – 37 % en el top-10, en un 5 – 30 % en el top-20 y en un 1,5 – 25 % en el top-30. La mejora más notoria de mi modelo puede verse en los primeros top- N . Por ejemplo, observamos que podría recuperar hasta el 52 % de los fármacos de fase IV o aprobados, para 26 virus distintos, en la predicción principal. Es decir, en el top-1. Y de manera general, en el top-10 podría recuperar hasta el 70 % de los fármacos antivirales eficaces.

La superioridad de mi método por encima de los demás, especialmente de su más inmediato perseguidor, el modelo descomposición matricial recientemente desarrollado (Santos et al., 2021), se debe principalmente a la discriminación de la estructura geométrica de la nube de puntos de los datos mediante el modelado que realiza a través del grafo. Esto le posibilita a mi método una comprensión mayor de la estructura de los datos.

3.2. Influencia del parámetro k en el rendimiento de GcNMF

El parámetro k hace referencia a la cantidad de vecinos cercanos considerados para construir el gra-

fo. En mis experimentos pude notar que dicho parámetro ejerce una influencia en el rendimiento de nuestro modelo al variarlo de manera creciente.

En la Figura 4 se puede observar cómo aumenta el rendimiento del modelo con todos los métodos de ponderación de los grafos en la medida que k aumenta, para los principales top- N . Si bien los resultados pueden indicar que el aumento no sucede de manera estable o monótonica, se puede ver claramente como la tendencia es de aumento. Por ejemplo, en la Figura 4.a, que hace referencia a las puntuaciones obtenidas para el top-1, vemos que con todas las ponderaciones con $k = 5$ el modelo posee un rendimiento por debajo del 50 %, pero con $k = 15$ se eleva por encima, incluso llegando el método de Similitud Coseno hasta el 55 %. Es esa la misma tendencia observada para los demás tops.

Esto nos indica que valorar la capacidad que tiene el parámetro k de alterar el rendimiento de GcNMF, y ajustarlo de manera óptima, es vital para lograr un rendimiento maximizado.

3.3. Elección del esquema de ponderación para los grafos

De los resultados observados en la Figura 4 y Figura 5 no se constatan diferencias amplias entre los rendimientos que brindan los distintos esquemas de ponderación de los grafos.

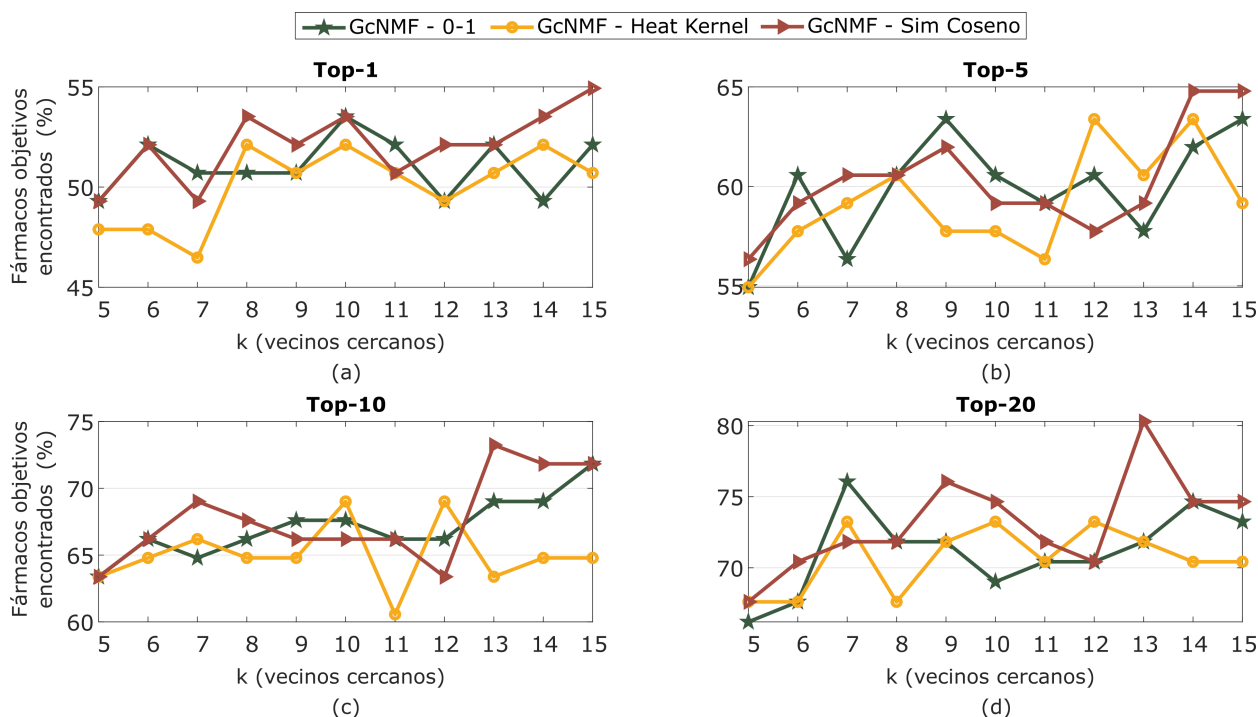


Figura 4: Variación del rendimiento de mi modelo de acuerdo a la variación del parámetro k del método kNN: El parámetro k hace referencia a la cantidad de vecinos cercanos considerados para construir el grafo. Estos experimentos muestran la influencia ejercida sobre nuestro modelo en cada esquema de ponderación al aumentar k , aquí los gráficos nos muestran que el rendimiento tiende a crecer. (a) Puntuaciones obtenidas para el top-1, con $k = 5$ el modelo con sus esquemas de ponderación está por debajo del 50 %, pero con $k = 15$ se eleva por encima, incluso llegando hasta el 55 %; (b) puntuaciones obtenidas para el top-5, vemos del mismo modo una tendencia a aumentar el rendimiento; (c) puntuaciones obtenidas para el top-10; (d) puntuaciones obtenidas para el top-20. **Fuente:** Elaboración propia.

Para escoger el esquema de ponderación con el rendimiento mayor comparo primero dentro del mismo esquema el rendimiento promedio obtenido con cada parámetro k . De los experimentos realizados obtuve que para la ponderación binaria $k = 15$ es el parámetro óptimo que maximiza su rendimiento promedio, $k = 10$ para el Heat kernel, y $k = 13$ para el de Similitud Coseno.

En la Figura 5 comparo los rendimientos de los tres esquemas para los top- N principales, de acuerdo a sus parámetros óptimos k seleccionados. Además, como referencia coloqué líneas de trazo horizontal que nos muestran el rendimiento promedio de cada uno, esto nos ayuda a notar que el esquema de Similitud Coseno es superior a los otros dos. Sin embargo, esta superioridad no es muy amplia, es decir, de solo 2 – 2,5 % mayor.

Los resultados visualizados en la Figura 5 nos muestran que independiente del método de ponderación utilizado en nuestro modelo GcNMF, los rendimientos promedios de los tres son superiores al rendimiento promedio de (Santos et al., 2021). Lo cual demuestra la superioridad de mi método.

3.4. Análisis de la robustez del mejor modelo

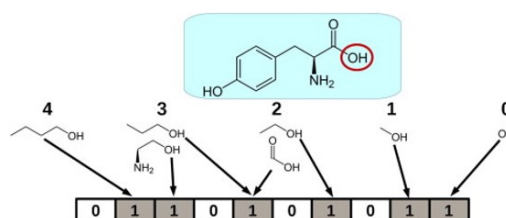


Figura 6: MACCS FP (del inglés *Molecular Access System Fingerprints*) a partir de las notaciones SMILES (del inglés *Simplified Molecular Input Line Entry System*): notación química que permite representar una estructura química tal que pueda ser utilizada por la computadora. Está formada por 167 elementos en un vector, los mismos codifican la estructura molecular de los fármacos. **Fuente:** (Cereto-Massagué et al. 2015) [34]

Luego de obtener los mejores resultados, realice un experimento que denominamos “similitudes químicas”, con el modelo ganador. Como bien sabemos por (Galeano, 2020) [28], las similitudes químicas entre fármacos podría afectar el rendimiento del modelo. Y queremos saber en qué medida GcNMF está sien-

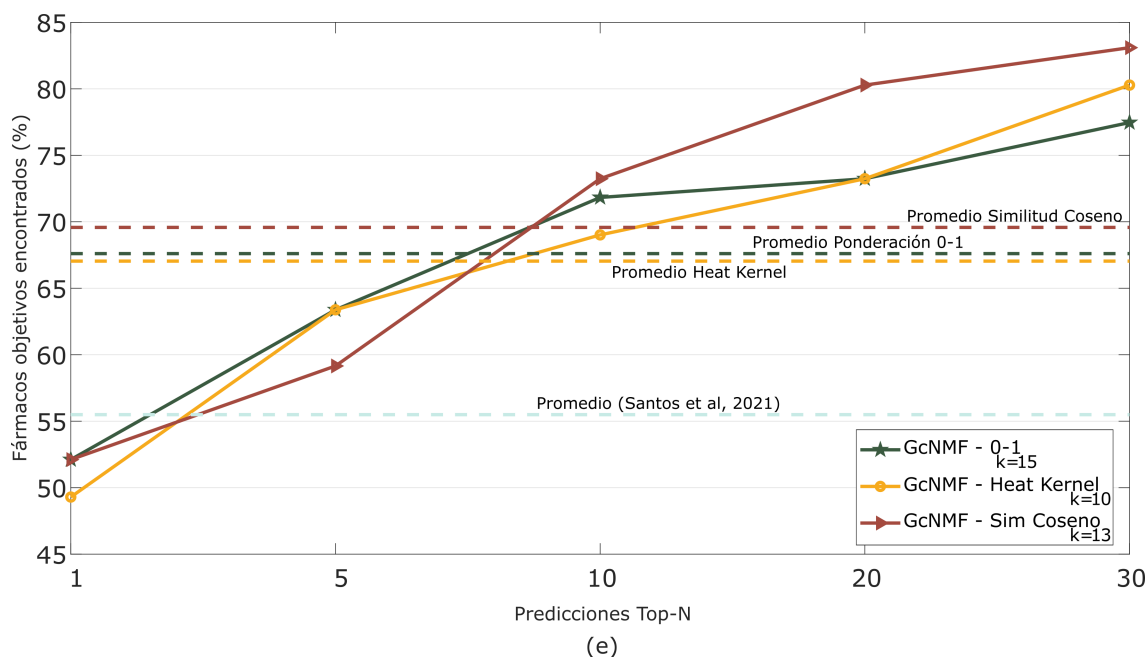


Figura 5: Mejor rendimiento de mi modelo para cada esquema de ponderación del grafo: los rendimientos de mi modelo para cada esquema de ponderación son muy parecidos, ganando en promedio el de Similitud Coseno por tan solo 2 – 2,5 %. Destaca el hecho de que en promedio todos superan en por lo menos 11,5 % al rendimiento promedio de (Santos et al., 2021), lo que demuestra la superioridad de GcNMF. **Fuente:** Elaboración propia.

do influenciado por esta información secundaria.

Para este experimento primero obtenemos las informaciones químicas de los fármacos, denominadas “huellas dactilares químicas” (o Chemical FP, del inglés *Chemical Fingerprints*, estas se ensamblan a partir de las notaciones SMILES (ver Figura 6) de los fármacos que nos interesan. Estos datos los obtuvimos consultando las bases de datos de DrugBank [35] y PubChem [36]. Las Chemical FP son vectores ensamblados a partir de la estructura molecular de los fármacos, se codifican con 167 elementos –ver Figura 6 [34].

De los 126 fármacos del conjunto de datos de (Andersen et al., 2020) [13], se pueden acceder a la información molecular química de 110 de ellos. Antes de realizar el experimento, para que pueda tenerse en cuenta las similitudes químicas, eliminamos de la matriz Y (ver Sección 2.1) los fármacos que no tienen información secundaria, es decir las que no tienen información química molecular disponible, quedando así una matriz Y de $m = 110 \times n = 80$.

Finalmente realizamos nuevamente el experimento de LOOCV, validando las asociaciones con el subconjunto de pruebas A (ver Sección 3.1). En cada iteración que dejamos una asociación fármaco-virus afuera (ver procedimiento LOOCV en Sección 3.1) revisamos todos los fármacos químicamente similares al que contiene la asociación, por encima de un umbral definido, y los eliminaremos de la matriz principal Y y solo luego hallamos las predicciones \hat{Y} . De ese

Tabla 1: Rendimiento de GcNMF al variar las similitudes químicas. **Fuente:** Elaboración propia

Umbral	Top 1 (%)	Top 5 (%)	Top 10 (%)	Top 20 (%)	Top 30 (%)
$\leq 0,2$	$-20,2 \pm 1,2$	$-35,6 \pm 2,1$	$-29,1 \pm 2,6$	$-29,5 \pm 2,4$	$-21,2 \pm 4,7$
$\leq 0,3$	$-9,8 \pm 3,2$	$-23,7 \pm 2,8$	$-21,2 \pm 2,3$	$-16,4 \pm 3,3$	$-11,2 \pm 2,3$
$\leq 0,4$	$-9,6 \pm 3,8$	$-18,5 \pm 2,1$	$-5,1 \pm 3,1$	$-14,2 \pm 2,5$	$-9,8 \pm 2,4$
$\leq 0,5$	$-8,6 \pm 3,8$	$-9,8 \pm 2,1$	$-4,4 \pm 1,9$	$-3,4 \pm 1,5$	$-1,7 \pm 1,7$
$\leq 0,6$	$-10 \pm 3,0$	$-7,1 \pm 2,8$	$-4,6 \pm 1,8$	$-0,8 \pm 2,1$	$-1,0 \pm 1,6$
$\leq 0,7$	$-1,9 \pm 5,0$	$0 \pm 2,9$	$+0,8 \pm 2,3$	$+1,2 \pm 1,3$	$+1,2 \pm 2,1$
$\leq 0,8$	$-3,5 \pm 3,3$	$+0,7 \pm 2,4$	$+1,2 \pm 1,7$	$+0,8 \pm 1,2$	$+0,7 \pm 1,6$
$\leq 0,9$	$-0,7 \pm 4,3$	$-0,8 \pm 2,3$	$-0,2 \pm 2,3$	$+1 \pm 0,8$	$-0,5 \pm 1,2$
≤ 1	$52,2 \pm 3$	$74,9 \pm 3,5$	$78,6 \pm 2,3$	$83,2 \pm 1,7$	$84,9 \pm 2,4$

modo exploraremos la influencia de esta información en el rendimiento del modelo.

Al observar la Tabla 1 podemos ver resultados bastante reveladores de este experimento. Observamos que si tenemos en cuenta todos los fármacos con similitudes hasta 0,6 (teniendo en cuenta que la literatura nos indica este valor como referencia de análisis [28]), que pertenecen al 95 % de los datos (percentil 95), el modelo GcNMF sigue siendo bastante robusto, perdiendo solo el 10 % de su rendimiento para Top-1 y tan solo el 1 % para el Top-30. Ganando aún en todos los Top-N a cNMF (ver Sección 3.1), con un 6,39 % de fármacos removidos. Lo que nos indica que nuestro conjunto es químicamente disimilar y es probable que está aprendiendo informaciones no relacionadas necesariamente a las similitudes entre fármacos.

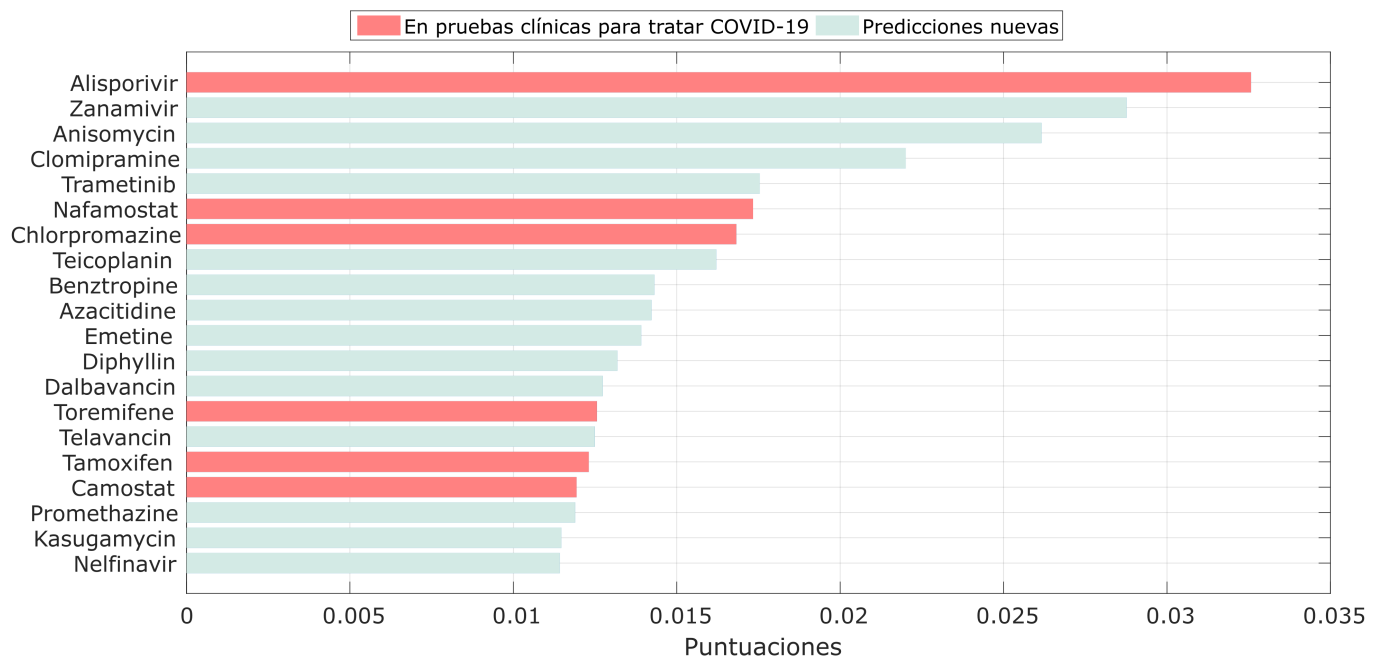


Figura 7: Lista de los 20 primeros fármacos recomendados por GcNMF potencialmente efectivos para COVID-19. El eje- x muestra las puntuaciones predichas obtenidas de la matriz de predicciones \hat{Y} para SARS-CoV-2. El eje- y muestra los fármacos (nombre genérico) ordenado por puntuación. En rojo se marcan aquellos fármacos predichos que ya se encuentran en ensayos clínicos para COVID-19. **Fuente:** Elaboración propia.

3.5. Análisis de predicciones para COVID-19

La eficiencia de GcNMF para predecir antivirales efectivos contra diversos virus nos impulsa a preguntar cuáles son las recomendaciones que realiza para SARS-CoV-2.

Para obtener las predicciones para SARS-CoV-2, entrenamos el modelo utilizando todos los datos disponibles en Y y luego extraje las puntuaciones correspondientes a SARS-CoV-2 de la matriz de predicciones \hat{Y} . Cabe señalar que solo analizo aquellas predicciones que no se encontraban en la matriz de entrenamiento para SARS-CoV-2. Finalmente, ordeno los fármacos basados en las puntuaciones y me enfoco en los top-20 fármacos predichos, que representan predicciones novedosas.

La Figura 7 nos muestra la lista de los 20 fármacos con mayor puntuación para SARS-CoV-2. Mediante una búsqueda manual en clinicaltrials.gov, he constatado que seis de las predicciones realizadas por GcNMF en el top-20 ya se encuentran actualmente en ensayos clínicos para COVID-19. Estos son: Alisporivir [37], Nafamostat [38], Chlorpromazine [39], Toremifene [40], Tamoxifen [41, 42] y Camostat [43–50]. La validación de las predicciones nuevas requiere análisis farmacológico de expertos, y podría ayudar a la elaboración de hipótesis científicas que permitan encontrar tratamientos efectivos contra COVID-19.

4. Conclusiones

El descubrimiento sistemático de antivirales contra virus específicos es crítico para el desarrollo de nuevos tratamientos para enfermedades infecciosas emergentes, como COVID-19. En este trabajo he propuesto el nuevo modelo de aprendizaje máquina GcNMF para predecir antivirales efectivos contra virus específicos. GcNMF extiende el modelo reciente de (Santos et al, 2021) [22] para modelar estructuras no-euclidianas en los datos.

Ordenando las conclusiones obtenidas:

- Se seleccionó una base de datos de entrenamiento creada por especialistas del área clínica.
- Se comparó detalladamente los desempeños y rendimientos de los algoritmos de descomposición más competitivos según el estado de arte. GcNMF puede ser un método eficiente basado en la factorización matricial capaz de recomendar nuevas relaciones fármacos-virus de los cuales no se tiene evidencia de eficacia y así apoyar al descubrimiento sistemático de nuevos fármacos antivirales de amplio espectro para virus específicos. Mi método logra obtener un rendimiento significativamente mejor que otros competidores, especialmente en los primeros top- N , recuperando hasta el 52 % de los fármacos ya conocidos que tratan a los virus en el top-1 y de manera

general hasta el 70 % en el top-10. GcNMF podría ser útil en el descubrimiento de antivirales de amplio espectro potencialmente efectivos contra enfermedades virales.

- Debido a que (Santos et al, 2021) es incapaz de modelar potenciales relaciones no-euclidianas entre los puntos en el espacio, es decir, fármacos, y que mis resultados muestran una mejora a través de GcNMF, se sigue que probablemente los datos –las interacciones fármaco-virus– se encuentran en una superficie no-euclidiana en un espacio multidimensional.
- Los experimentos de “similaridades químicas” nos muestran que nuestro modelo es bastante robusto frente a remociones de fármacos químicamente similares. Se sigue que es probable que GcNMF no se basa solo en similaridades químicas para recomendar fármacos. Lo que quiere decir que el conjunto de datos es disimilar. Lo que interesa a los científicos farmacológicos ya que el estado del arte ya contempla larga y tradicionalmente modelos que se enfocan en las similitudes.

Referencias

- [1] M. B. Serafin, A. Bottega, V. S. Foletto, T. F. da Rosa, A. Hörner, and R. Hörner, “Drug repositioning is an alternative for the treatment of coronavirus covid-19,” *International Journal of Antimicrobial Agents*, vol. 55, no. 6, p. 105969, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924857920301266>
- [2] W. H. Organization. Coronavirus disease (covid-2019) situation reports. [Accessed 01-11-2023]. [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
- [3] W. Bank, *Emerging risks to the recovery*, pp. 49–73. [Online]. Available: https://elibrary.worldbank.org/doi/abs/10.1596/978-1-4648-1730-4_ch1
- [4] V. Benavides-Cordoba, “Drug Repositioning for COVID-19,” *Colombia Médica*, vol. 51, 06 2020. [Online]. Available: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-95342020000200002&nrm=iso
- [5] T. T. Ashburn and K. B. Thor, “Drug repositioning: identifying and developing new uses for existing drugs,” *Nature Reviews Drug Discovery*, vol. 3, no. 8, pp. 673–683, Aug 2004. [Online]. Available: <https://doi.org/10.1038/nrd1468>
- [6] R. A. Hodos, B. A. Kidd, K. Shameer, B. P. Readhead, and J. T. Dudley, “In silico methods for drug repurposing and pharmacology,” *Wiley interdisciplinary reviews. Systems biology and medicine*, vol. 8, no. 3, p. 186–210, May 2016. [Online]. Available: <https://europepmc.org/articles/PMC4845762>
- [7] N. A. M. Tamimi and P. Ellis, “Drug development: From concept to marketing!” *Nephron Clinical Practice*, vol. 113, no. 3, pp. c125–c131, 2009. [Online]. Available: <https://www.karger.com/DOI/10.1159/000232592>
- [8] C. for Disease Control and P. (CDC). Covid-19 vaccine effectiveness update. [Accessed 05-01-2024]. [Online]. Available: <https://covid.cdc.gov/covid-data-tracker/#vaccine-effectiveness>
- [9] M. Rudrapal, S. J. Khairnar, and A. G. Jadhav, “Drug repurposing (dr): An emerging approach in drug discovery,” in *Drug Repurposing-Hypothesis, Molecular Aspects and Therapeutic Applications*. IntechOpen, 2020.
- [10] S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Williams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, and M. Pirmohamed, “Drug repurposing: progress, challenges and recommendations,” *Nature Reviews Drug Discovery*, vol. 18, no. 1, pp. 41–58, Jan 2019. [Online]. Available: <https://doi.org/10.1038/nrd.2018.168>
- [11] Y. li and S. Jones, “Drug repositioning for personalized medicine,” *Genome medicine*, vol. 4, p. 27, 03 2012.
- [12] S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Williams, J. Latimer, C. McNamee et al., “Drug repurposing: progress, challenges and recommendations,” *Nature reviews Drug discovery*, vol. 18, no. 1, pp. 41–58, 2019.
- [13] P. I. Andersen, A. Ianevski, H. Lysvand, A. Vitkauskienė, V. Oksenysh, M. Bjørås, K. Telling, I. Lutsar, U. Dumpis, Y. Irie, T. Tenson, A. Kantele, and D. E. Kainov, “Discovery and development of safe-in-man broad-spectrum antiviral agents,” *International Journal of Infectious Diseases*, vol. 93, pp. 268 – 276, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S120197122030076X>
- [14] Y. Zhou, F. Wang, J. Tang, R. Nussinov, and F. Cheng, “Artificial intelligence in covid-19 drug repurposing,” *The Lancet Digital Health*, 2020.
- [15] Y.-Y. Ke, T.-T. Peng, T.-K. Yeh, W.-Z. Huang, S.-E. Chang, S.-H. Wu, H.-C. Hung, T.-A. Hsu, S.-J. Lee, J.-S. Song et al., “Artificial intelligence approach fighting covid-19 with repurposing drugs,” *Biomedical journal*, vol. 43, no. 4, pp. 355–362, 2020.

- [16] B. R. Beck, B. Shin, Y. Choi, S. Park, and K. Kang, "Predicting commercially available antiviral drugs that may act on the novel coronavirus (sars-cov-2) through a drug-target interaction deep learning model," *Computational and structural biotechnology journal*, vol. 18, pp. 784–790, 2020.
- [17] D. M. Gysi, Í. Do Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, H. Sanchez, R. M. Baron, D. Ghiassian, J. Loscalzo *et al.*, "Network medicine framework for identifying drug repurposing opportunities for covid-19," *ArXiv*, 2020.
- [18] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [19] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The adaptive web*. Springer, 2007, pp. 291–324.
- [20] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 287–296.
- [21] X. Amatriain and J. Basilico, "Recommender systems in industry: A netflix case study," in *Recommender systems handbook*. Springer, 2015, pp. 385–419.
- [22] S. Santos, M. Torres, D. Galeano, M. Sánchez, L. Cernuzzi, and A. Paccanaro, "Computational approaches to assist drug repositioning for covid-19," *SSRN Electronic Journal*, 01 2021.
- [23] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [24] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.
- [25] F. Monti, M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 3697–3707.
- [26] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *arXiv:1709.05584*, 2017.
- [27] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1548–1560, 2010.
- [28] D. Galeano, S. Li, M. Gerstein, and A. Paccanaro, "Predicting the frequencies of drug side effects," *Nature Communications*, vol. 11, no. 1, p. 4575, Sep 2020. [Online]. Available: <https://doi.org/10.1038/s41467-020-18305-y>
- [29] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, p. 18–42, Jul 2017. [Online]. Available: <http://dx.doi.org/10.1109/MSP.2017.2693418>
- [30] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, June 2006, pp. 1735–1742.
- [31] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of recommender algorithms on top-n recommendation tasks," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 39–46.
- [32] H. Dowden and J. Munro, "Trends in clinical success rates and therapeutic focus," *Nature Reviews Drug Discovery*, vol. 18, pp. 495–496, 2019.
- [33] E. A. Sosnina, S. Sosnin, A. A. Nikitina, I. Nazarov, D. I. Osolodkin, and M. V. Fedorov, "Recommender systems in antiviral drug discovery," *ACS Omega*, vol. 5, no. 25, pp. 15039–15051, Jun 2020. [Online]. Available: <https://doi.org/10.1021/acsomega.0c00857>
- [34] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé, and G. Pujadas, "Molecular fingerprint similarity search in virtual screening," *Methods*, vol. 71, pp. 58–63, 2015, virtual Screening. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1046202314002631>
- [35] Drugbanks. [Accessed 01-11-2023]. [Online]. Available: <https://www.drugbank.com/>
- [36] Pubchem. [Accessed 01-11-2023]. [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/>
- [37] U. N. L. of Medicine. Evaluation of alisporivir for the treatment of hospitalised patients with infections due to sars-cov-2 (covid-19) (cyclovid). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04608214>
- [38] —. A study evaluating the efficacy and safety of ckd-314 in hospitalized adult patients diagnosed with covid-19 pneumonia. [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04608214>

//clinicaltrials.gov/ct2/show/NCT04628143?
cond=COVID-19&draw=3&rank=12

- [39] ——. Repurposing of chlorpromazine in covid-19 treatment (recovery). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04366739>
- [40] ——. Selective estrogen modulation and melatonin in early covid-19 (sentinel). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04531748>
- [41] ——. Combination of chemopreventive agents (all- trans retinoic acid and tamoxifen) as potential treatment for the lung complication of covid-19. [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04568096>
- [42] ——. Combination therapy with isotretinoin and tamoxifen expected to provide complete protection against severe acute respiratory syndrome coronavirus (combination). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04389580>
- [43] ——. The impact of camostat mesilate on covid-19 infection (camoco-19). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04321096>
- [44] ——. The potential of oral camostat in early covid-19 disease in an ambulatory setting to reduce viral load and disease burden. [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04625114>
- [45] ——. Camovid : Evaluation of efficacy and safety of camostat mesylate for the treatment of sars-cov-2 infection - covid-19 in ambulatory adult patients (camovid). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04608266>
- [46] ——. The dawn camostat trial for ambulatory covid-19 patients (dawn). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04730206>
- [47] ——. Camostat with bicalutamide for covid-19 (combo). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04652765>
- [48] ——. Reconvalescent plasma/camostat mesylate early in sars-cov-2 q-pcr (covid-19) positive high-risk individuals (res-q-hr). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04681430>
- [49] ——. A trial looking at the use of camostat to reduce progression of symptoms of coronavirus (covid-19) in people who have tested positive but are able to stay at home. [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04455815>
- [50] ——. Camostat mesilate treating patients with hospitalized patients with covid-19 (recover). [Accessed 01-01-2024]. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT04470544>