

RESEARCH

Explorando la Demencia Frontotemporal mediante Biología de Sistemas: Un Enfoque Integrado

Mario Pascual González*, Ainhoa Nerea Santana Bastante, Carmen Rodríguez González, Gonzalo Mesas Aranda and Ainhoa Pérez González

*Correspondence:
pascualgonzalez.mario@uma.es
ETSI Informática, Universidad de
Málaga, Málaga, España
Full list of author information is
available at the end of the article

Abstract

Keywords: HPO; Demencia Frontotemporal; Biología de Sistemas

1 Introducción

La demencia se define como un síndrome caracterizado por un deterioro cognitivo que produce alteraciones en la memoria, el pensamiento y el comportamiento de una persona. Esto dificulta la capacidad del paciente para realizar sus actividades sociales o laborales. [1] Se estima que hay alrededor de 44 millones de personas con demencia, se prevé que esta cifra será más del triple para 2050. [2]. La *Enfermedad de Alzheimer* (EA), es la enfermedad más común donde se presenta este síndrome (45-55%), seguida de la demencia vascular y la demencia por cuerpos de Lewy. La demencia frontotemporal no supera el 5% en las frecuencias relativas. [3, 4]. En este proyecto se estudiará un fenotipo concreto presente en varios casos de demencia, la denominada *Demencia Frontotemporal* (FTD).

La FTD es un fenotipo clínico caracterizado por la degeneración progresiva de funciones cognitivas relacionadas con el comportamiento, la personalidad, y el lenguaje, resultado de procesos neurodegenerativos que afectan principalmente a los lóbulos frontal y temporal del cerebro [5, 6, 7]. Es una de las principales causas de demencia en personas menores de 65 años, presentándose con mayor frecuencia entre los 45 y 65 años [5, 6]. Las variantes clínicas más comunes de la FTD son la variante conductual (bvFTD), experimentada por el 70% de los pacientes y que afecta la regulación del comportamiento y las emociones [5, 7], y la afasia progresiva primaria (PPA), que afecta las habilidades del lenguaje y se subdivide en variantes no fluente/agramática (nfvPPA), semántica (svPPA) y logopénica (lvPPA) [8].

La FTD, al tratarse de un conjunto heterogéneo de fenotipos, muestra conexiones significativas con otras enfermedades neurodegenerativas. En estudios de coocurrencia de términos, *Esclerosis Lateral Amiotrófica* (ELA) y EA son conceptos frecuentemente asociados con FTD, lo que sugiere una relación estrecha [9]. Se puede deducir entonces otras enfermedades neurodegenerativas presentan el fenotipo clínico de la FTD, indicando que hay una neuropatología subyacente que las relaciona, manifestada bajo este conjunto de fenotipos. La degeneración lobar frontotemporal (FTLD) es el término general que agrupa a los fenotipos patológicos que dan

lugar al fenotipo clínico de la FTD [10]. La FTLD se clasifica según la acumulación de proteínas anormales en las neuronas. Los subtipos principales incluyen: FTLD-tau, caracterizado por la acumulación de tau hiperfosforilada y asociado fuertemente con la bvFTD [11]; FTLD-TDP, que afecta a más del 50% de los pacientes tau-negativos, estrechamente relacionada con la bvFTD y la svPPA [12]; y FTLD-FUS [13]. El ejemplo más sonado en la literatura se relaciona con la ELA, una forma común de enfermedad de la motoneurona (MND) [14], la cual es una enfermedad neurodegenerativa que comparte causas genéticas y neuropatologías con la FTD [15]. Mutaciones en genes como *C9orf72* [16], *TARDBP* [17] y *OPTN* [18] se han identificado en pacientes que presentan el fenotipo FTD, padecen ELA, o con manifestaciones de ambas. Las expansiones en *C9orf72* una causa frecuente en ambos casos [19, 16]. Patológicamente, se han observado disfunciones en el sistema autofagia-lisosoma [20] e inclusiones citoplásmicas neuronales tau-negativas pero ubiquitina-positivas [17] en ambas enfermedades. Pacientes con MND pueden desarrollar afectación cognitiva y evolucionar a FTD [21], e incluso mostrar síntomas típicos de bvFTD [22]. Asimismo, el parkinsonismo, especialmente el síndrome corticobasal, presenta solapamientos considerables con la FTD, incluyendo trastornos motores y cognitivos [23]

En cuanto al diagnóstico, la neuroimagen permite distinguir de manera confiable los subtipos de FTLD de otras demencias, ayudando a correlacionar los síntomas neuropsiquiátricos con los patrones de atrofia cerebral. Técnicas como la Image por Resonancia Magnética (MRI) permiten detectar la atrofia focal en los lóbulos frontal y temporal, típicamente observada en pacientes con bvFTD [8]. Las técnicas de neuroimagen funcional, como la tomografía por emisión de positrones (FDG-PET) y la tomografía por emisión de fotón único (SPECT), se utilizan para identificar áreas de hipometabolismo cerebral, dado que bvFTD muestra hipometabolismo en las regiones frontales [24, 25] (Figura 1). Estos métodos también son efectivos para caracterizar las variantes de la PPA. La nvPPA muestra atrofia en la región fronto-insular, mientras que la svPPA afecta a los lóbulos temporales anteriores [8] (Figura 1). Técnicas avanzadas como la PET con amiloide- β , han mostrado ser prometedoras para discriminar entre casos atípicos de AD y FTLD [26]. Además, se ha avanzado en la comprensión de las bases genéticas de la FTD, con mutaciones en genes como MAPT, GRN y C9orf72, que están implicadas en aproximadamente el 30-50% de los casos familiares de FTD [27].

Siguiendo en esta línea, en torno al 40% de los casos en los que se expresa la FTD son familiares, es decir tienen un patrón hereditario [28, 27]. De los genes que más frecuentemente se ven implicados, uno es el MAPT, que está implicado en la producción de la proteína tau, un importante mediador en la polimerización y estabilización de los microtúbulos cerebrales [29]. Algunas mutaciones en el gen GRN provocan una producción reducida de progranulina, que causa neurodegeneración y se asocia típicamente con bvFTD, aunque también se han reportado casos de PPA [30]. La causa genética más común de la FTD, viene dada por el gen C9ORF72, el cual sufre normalmente una expansión del hexanucleótido GGGGCC en una región no codificante del cromosoma 9 [31].

Otros genes implicados son el gen TARDBP, del cual se han encontrado mutaciones tanto en pacientes con FTD esporádico como familiar [32]; el gen VCP, que

interviene en diversos procesos celulares, del cual se han encontrado mutaciones en pacientes con FTD negativos en MAPT, GRN y C9ORF72 [33]; o el gen CHMP2B, que sufre mutaciones de truncamiento o sin sentido, relacionado con la demencia frontotemporal del cromosoma 3 (FTD-3) [34].

En relación con los factores ambientales que afectan al fenotipo, un estudio sugiere una relación entre el trauma craneoencefálico y la FTD [35, 36]. Además, no se encontró relación con factores como el tabaco, el alcohol, exposición a químicos, pesticidas o insecticidas [36].

Cabe destacar que la esperanza de vida promedio de un paciente que presenta los fenotipos clínicos de la FTD es de 6 a 8 años, variando entre 3 y 20 años según la gravedad y las mutaciones genéticas presentes [37]. Aunque no existen tratamientos aprobados por la FDA, las estrategias terapéuticas actuales se adaptan al fenotipo predominante en cada paciente, combinando intervenciones farmacológicas y no farmacológicas para mejorar la calidad de vida y controlar los síntomas [38].

Entre los tratamientos farmacológicos más prometedores destacan las terapias génicas, como PBFT02, que corrige mutaciones en el GRN [39], y otros medicamentos experimentales dirigidos a pacientes con mutaciones en granulina, como el FRM-0334 [40]. También se emplean estimulantes del sistema nervioso central y antipsicóticos atípicos para manejar los síntomas conductuales [41]. En cuanto a las terapias no farmacológicas, se incluyen intervenciones en el estilo de vida y terapias ocupacionales, del habla y físicas para mejorar la funcionalidad y la comunicación de los pacientes.

En este estudio se busca desentrañar las bases moleculares que subyacen a este fenotipo utilizando enfoques propios de la biología de sistemas, como el análisis de redes biológicas. Nuestro objetivo es identificar interacciones clave y procesos biológicos implicados, con el fin de proponer posibles nuevas estrategias terapéuticas.

2 Materiales y métodos

2.1 Datos

Datos del fenotipo

Para este estudio, se utilizó el fenotipo Demencia frontotemporal, identificado con el término HP:0002145 en Human Phenotype Ontology (HPO). A partir de este término HPO, se han extraído 52 genes asociados al fenotipo. Estos genes se obtuvieron mediante la API Ontology Annotation Network [43] de HPO, que permite acceder programáticamente a las anotaciones entre términos fenotípicos y genes. A través de esta API, se descargaron los datos en formato JSON, que luego fueron procesados para extraer los nombres de los genes y guardarlos en un archivo TSV.

Estos genes, relacionados con el desarrollo de la demencia frontotemporal y otras patologías neurodegenerativas, representan el conjunto inicial de genes sobre el que se construirá la red de interacciones para el análisis posterior. Cada uno de ellos se identifica mediante su ID único en la base de datos de NCBI Gene, lo cual facilita el acceso y la referencia a los datos genéticos específicos.

Para asegurar la reproducibilidad, se utilizó la versión 2.0.4 de HPO [44], para obtener el término fenotípico y descargar los genes relacionados con el fenotipo de estudio. En la siguiente sección se proporciona una descripción detallada de HPO.

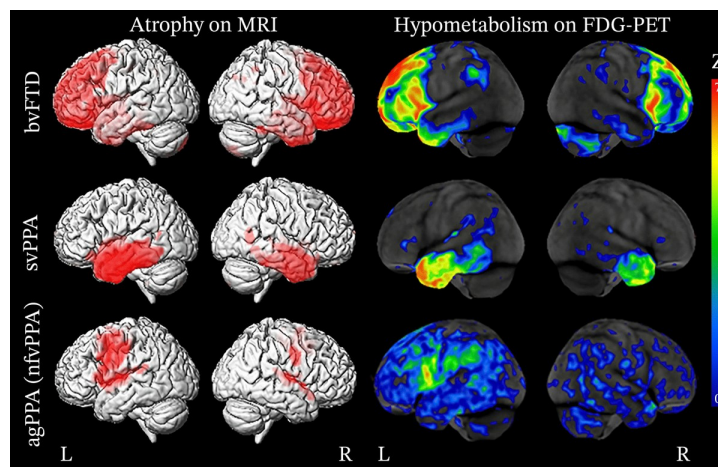


Figure 1: Comparación entre los patrones de atrofia cerebral y el hipometabolismo en diferentes variantes de FTD. En la columna izquierda, reconstrucciones 3D de MRI, que muestran las áreas de atrofia cerebral (en rojo) en tres subtipos de FTD: bvFTD, svPPA, y nfvPPA. En la columna derecha, reconstrucciones 3D de FDG-PET, las cuales indican el hipometabolismo cerebral (en colores) en las mismas regiones afectadas, reflejando la disminución del metabolismo en las áreas afectadas. Los colores cálidos (rojo/amarillo) indican mayor hipometabolismo, mientras que los colores fríos (verde/azul) indican menor hipometabolismo. Imagen tomada de [42].

Human Phenotype Ontology (HPO)

HPO proporciona una ontología estandarizada que describe anomalías fenotípicas observadas en enfermedades humanas, facilitando la identificación y análisis de genes asociados a diversas características clínicas. Cada término en HPO representa una anomalía específica, como la demencia frontotemporal, y está diseñado para facilitar la caracterización precisa de los fenotipos en el contexto de enfermedades hereditarias. La ontología se desarrolla y actualiza de forma continua utilizando fuentes como la literatura médica, así como bases de datos como Orphanet, DECIPHER y OMIM. Actualmente, HPO contiene más de 18,000 términos y ofrece más de 156,000 anotaciones asociadas a enfermedades hereditarias [44].

Datos de interacción

Los datos de interacción representan conexiones funcionales y físicas entre proteínas, y constituyen la base para construir redes de interacción en el análisis de procesos biológicos. En este estudio, los datos de interacción proteína-proteína (PPI) fueron extraídos de la base de datos STRING mediante su API REST [45], que permite recuperar programáticamente redes de interacción específicas basadas en listas de genes o proteínas de interés.

A través de esta API, se obtuvieron las interacciones entre los genes asociados al fenotipo FTD (HP:0002145) en formato TSV. En este archivo, cada fila representa una interacción entre dos proteínas y contiene las siguientes columnas.

- **protein1**: ID de la primera proteína en la interacción, precedido por el código taxonómico del organismo (por ejemplo, "9606" para proteínas humanas).
- **protein2**: ID de la segunda proteína en la interacción, también con el prefijo de organismo.
- **combined_score**: Puntuación de confianza combinada para cada interacción proteína-proteína, con valores que oscilan entre 0 y 1000. Esta puntuación refleja la probabilidad de que una interacción sea real, basada en una integración de diversas fuentes de evidencia, como co-ocurrencia filogenética, co-expresión, minería de texto y datos experimentales. Cada tipo de evidencia se evalúa y puntúa individualmente, y luego se combina en el "combined_score", proporcionando así un indicador global de confiabilidad para cada interacción funcional o física [46].

Estos datos obtenidos se utilizarán para construir una red de interacciones entre los genes asociados al fenotipo FTD, permitiendo analizar las relaciones funcionales entre proteínas en este contexto. Esta red servirá como base para el análisis de clustering, facilitando la identificación de módulos de genes potencialmente implicados en funciones biológicas específicas.

Para asegurar la reproducibilidad del análisis, se utilizó la versión 12.0 de STRING, junto con su API REST de STRING para la extracción de interacciones.

STRING

La base de datos STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) es un recurso bioinformático diseñado para recopilar, organizar y analizar redes de interacciones proteína-proteína y asociaciones funcionales en cualquier genoma secuenciado. STRING integra información de diversas fuentes, como minería de texto científico, predicciones computacionales basadas en coexpresión y contexto genómico, y datos experimentales obtenidos de estudios de interacciones proteicas. Además, los usuarios pueden acceder a la base de datos para explorar redes de interacción, realizar análisis de enriquecimiento funcional y generar redes personalizadas para genomas específicos, facilitando así la investigación en biología celular y molecular. Actualmente, STRING cubre 59.309.604 proteínas provenientes de 12.535 organismos. [46].

2.2 Software

Para el análisis funcional y la construcción de redes genéticas en este estudio, se seleccionaron herramientas especializadas que permiten tanto la exploración bioinformática como la visualización de datos complejos. Dado que el objetivo principal es investigar la interacción entre genes y módulos específicos asociados a la demencia frontotemporal, se ha optado por una combinación de paquetes en Python y R que ofrecen un balance entre precisión analítica y capacidades visuales avanzadas.

Paquetes de Python para el análisis funcional y otras funciones

- **Pandas (versión 2.2.3)**: Este paquete proporciona estructuras de datos eficientes y flexibles, como DataFrames, que facilitan el procesamiento y manipulación de datos complejos. En el contexto de este estudio, Pandas permite organizar, filtrar y procesar resultados de enriquecimiento funcional, simplificando el manejo de grandes volúmenes de datos bioinformáticos [47].

- **Matplotlib (versión 3.8.1)**: Matplotlib es una librería de visualización muy versátil que soporta múltiples tipos de gráficos en 2D, lo que resulta útil para representar tendencias y relaciones entre genes en gráficos de líneas, barras, dispersión, y más. Este paquete se utilizará para visualizar los resultados de enriquecimiento y las interacciones génicas [48].
- **Scienceplots (versión 2.1.1)**: Este paquete extiende Matplotlib proporcionando estilos de gráficos estéticamente optimizados para publicaciones científicas. Con Scienceplots, se puede lograr una presentación visual de alta calidad, ideal para gráficos que requieren una apariencia profesional [49].
- **Requests (versión 2.31.0)**: Requests es una biblioteca para realizar solicitudes HTTP de manera simple y efectiva. En este estudio, se utiliza para conectar con APIs externas como la de STRINGdb, permitiendo la descarga automatizada de datos de redes y enriquecimiento [50].
- **GOATOOLS (versión 1.2.3)**: Este paquete permite realizar análisis de enriquecimiento funcional en términos de Gene Ontology (GO), lo que facilita la identificación de términos GO sobre-representados en conjuntos de genes. Es una herramienta clave para comprender funciones biológicas asociadas a los genes estudiados [51].
- **G:Profiler (versión 1.4.0)**: G:Profiler realiza análisis de enriquecimiento funcional abarcando varias bases de datos (GO, KEGG, Reactome, entre otras). La versatilidad de este paquete lo convierte en una excelente opción para explorar las funciones biológicas de conjuntos de genes con una amplia cobertura de recursos [52].
- **Statsmodels (versión 0.14.0)**: Statsmodels es una biblioteca de estadística que permite aplicar ajustes de p-valor, como el método de Benjamini-Hochberg, para reducir el impacto de falsos positivos en los resultados de enriquecimiento. Esto asegura una interpretación estadísticamente robusta de los resultados [53].

Paquetes de R para clustering y visualización de redes

- **Cluster (versión 2.1.5)**: Este paquete ofrece una variedad de algoritmos clásicos de clustering (como k-means y clustering jerárquico), facilitando la agrupación de genes según patrones de expresión similares, lo cual es fundamental para identificar posibles módulos relacionados con la demencia frontotemporal [54].
- **Factoextra (versión 1.0.8)**: Factoextra permite una visualización clara y accesible de resultados de clustering, proporcionando gráficos intuitivos para explorar la estructura de los datos y resaltar posibles módulos o patrones de coexpresión [55].
- **WGCNA (versión 1.71)**: Este paquete es ideal para detectar módulos de genes coexpresados mediante la construcción de redes de coexpresión genética ponderada, lo cual es útil en estudios de expresión génica complejos como el de la demencia frontotemporal [56].
- **Igraph (versión 1.5.0)**: Igraph facilita la construcción, manipulación y visualización de redes genéticas y módulos de coexpresión. Este paquete es útil para analizar la conectividad y relación entre genes en un contexto de red [57].

- **STRINGdb (versión 2.10.0)**: STRINGdb es un paquete que conecta con la base de datos STRING, permitiendo realizar enriquecimiento funcional en redes y visualizar interacciones proteicas y genéticas. Es útil para identificar las relaciones funcionales entre genes [58].
- **Pathview (versión 1.36.0)**: Este paquete permite integrar datos de expresión génica en rutas de KEGG, ofreciendo representaciones gráficas que muestran la implicación de los genes en rutas metabólicas y de señalización, lo cual es relevante para entender su rol en la enfermedad [59].
- **ClusterProfiler (versión 4.15.0)**: ClusterProfiler permite realizar análisis de enriquecimiento funcional de alta precisión en términos de GO, KEGG y Reactome. Ofrece opciones avanzadas de visualización y ajuste de p-valores, facilitando la interpretación de los resultados [60].

2.3 Análisis de enriquecimiento de vías biológicas

El análisis de enriquecimiento permite identificar vías biológicas o pathways que están significativamente representados en una lista de genes de interés, por medio de pruebas estadísticas. Un pathway es un conjunto de genes que trabajan en conjunto para llevar a cabo un proceso biológico específico.[61]

Se realizó un análisis de enriquecimiento mediante la API de STRINGdb. Mediante el mapeo de las anotaciones funcionales de los genes en ontologías como Gene Ontology (GO), UniProt, Reactome y KEGG y comparando la distribución de estos términos en el conjunto de genes de interés con su distribución en un conjunto de referencia, se identificaron estadísticamente los términos sobrerrepresentados en la lista de genes asociados a la enfermedad. [62]

STRINGdb, por tanto, permite analizar diversas categorías de anotaciones funcionales, incluyendo los términos de Gene Ontology (procesos biológicos, funciones moleculares y componentes celulares), rutas de KEGG y Reactome [46]. Se realizó un filtrado según el tipo de término para obtener los diferentes enriquecimientos funcionales de interés.

Para calcular el enriquecimiento, STRINGdb utiliza un test hipergeométrico que evalúa si la representación de cada término funcional en el conjunto de genes de interés es mayor que la esperada por azar. Para controlar el error debido a comparaciones múltiples, se aplica la corrección de Benjamini-Hochberg para ajustar la tasa de descubrimiento falso (FDR) [46].

El análisis de enriquecimiento se realizó tanto en el conjunto completo de genes asociados a la enfermedad, extraído de términos de HPO, como en los genes agrupados en clústeres.

2.4 Clustering

Al aplicar algoritmos de clustering, nuestro objetivo es descubrir comunidades funcionales dentro de la red de genes asociada a la demencia frontotemporal. Estas comunidades, módulos, o *clusters* funcionales pueden representar procesos biológicos específicos, vías celulares, o mecanismos asociados al fenotipo FTD. Encontrar estos clusters podría revelar posibles dianas terapéuticas o grupos de biomarcadores dentro del conjunto de genes, lo cual podría abrir las puertas a nuevos tratamientos para los pacientes de FTD.

2.4.1 Red

En la red de interacción proteína-proteína (PPI) obtenida mediante la API de STRINGdb, cada nodo representa un gen y cada arista representa una interacción entre genes, ponderada por un puntaje de confianza otorgado por STRINGdb [46]. Para enfocarnos en interacciones más sólidas y confiables, se aplicó un umbral mínimo de confianza en las aristas, lo que permitió refinar la red y mejorar la relevancia de los clústeres detectados

2.4.2 Algoritmos

A continuación, se detallan los tres algoritmos de clustering, proporcionados por la librería *iGraph*, elegidos para este estudio, los cuales pretenden cubrir diferentes enfoques teóricos en la detección de comunidades funcionales [57].

Walktrap: Algoritmo de clustering jerárquico aglomerativo que detecta comunidades basándose en la probabilidad de que recorridos aleatorios dentro de la red permanezcan dentro de clústeres cohesivos [63]. El algoritmo aplica recorridos aleatorios para capturar la proximidad de los nodos, asumiendo que los nodos que pertenecen a la misma comunidad tienen mayores probabilidades de estar conectados mediante recorridos cortos. Comienza con cada nodo como su propio clúster, fusionando sucesivamente los clústeres cercanos hasta alcanzar una partición óptima. Se ajustó la longitud de los pasos (*steps*).

Algoritmo de Leiden: Se basa en optimizar el Constant Potts Model (CPM) de la red [64, 65]. Esta función, mostrada en la Ecuación 1, evalúa la calidad de una partición en comunidades considerando un balance entre la densidad interna de cada clúster y el parámetro de resolución γ , que controla el tamaño preferido de las comunidades finales. El término n_c se refiere al número de nodos, mientras que e_c es el número de aristas internas, ambos referidos a la comunidad c .

$$\mathcal{H} = \sum_c \left[e_c - \gamma \left(\frac{n_c(n_c - 1)}{2} \right) \right] \quad (1)$$

Se ajustó el parámetro γ , así como el número de iteraciones del algoritmo, permitiendo que el Leiden refinara iterativamente la partición. El resto de parámetros se fijaron a su valor por defecto.

Algoritmo de Louvain: Este algoritmo optimiza la modularidad de la red, una métrica que mide la calidad de una partición en comunidades al comparar la densidad de conexiones internas con la densidad esperada si las conexiones fueran aleatorias [66]. La modularidad Q , definida en la Ecuación 2, es un valor escalar entre -1 y 1 que representa la diferencia entre la densidad de aristas dentro de las comunidades y la densidad de aristas esperada. El algoritmo Louvain maximiza esta modularidad en dos fases, cuyo resultado es la formación una nueva red sobre la cual se repite el proceso hasta que la modularidad no mejore más. En redes ponderadas, los autores del algoritmo expresaron la modularidad a optimizar como:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \quad (2)$$

donde A_{ij} es el peso de la arista entre los nodos i y j , k_i y k_j son las sumas de los pesos de las aristas conectadas a los nodos i y j , c_i y c_j representan las comunidades a las que pertenecen los nodos i y j , y $\delta(c_i, c_j)$ es una función delta de Kronecker que es 1 si $c_i = c_j$ y 0 en caso contrario. El término m es la suma total de los pesos de las aristas en la red.

Se ajustó el parámetro de resolución, que controla el tamaño final de las comunidades. El resto de parámetros se dejaron con sus valores por defecto.

2.4.3 Medidas de Rendimiento

2.4.4 Optimización

Como se explicó en la Sección 2.4.2, Louvain y Leiden son algoritmos de clustering con múltiples parámetros configurables. Ajustar adecuadamente estos parámetros es crucial para mejorar tanto la interpretación biológica de la red como la detección de comunidades. Este proceso, conocido como ajuste de hiperparámetros, busca optimizar el rendimiento del algoritmo según diversas métricas. En este apartado, se detalla el procedimiento empleado para ajustar el parámetro de resolución (γ) en Leiden y Louvain, con el objetivo de maximizar las métricas descritas en la Sección 2.4.3.

La estadística bayesiana es un enfoque probabilístico que utiliza el teorema de Bayes para actualizar las creencias sobre un modelo a medida que se incorporan nuevos datos. El ajuste bayesiano de hiperparámetros (BHO) aplica este enfoque para optimizar los parámetros de un modelo, construyendo y actualizando iterativamente un modelo probabilístico de la función objetivo en función de los hiperparámetros. Este modelo probabilístico sugiere los hiperparámetros a probar en cada iteración, permitiendo enfocar eficientemente la búsqueda en las regiones más prometedoras del espacio de hiperparámetros.

Existen diversos algoritmos basados en BHO, cada uno caracterizado por el modelo probabilístico que construye. En este proyecto, hemos optado por el Tree-structured Parzen Estimator (TPE). En el TPE, consideramos que y es el valor obtenido al evaluar la función objetivo utilizando un conjunto de hiperparámetros θ . Definimos un umbral y^* que nos permite dividir nuestros datos en dos conjuntos: los hiperparámetros que resultan en un rendimiento mejor que y^* y los que resultan en un rendimiento peor. Es decir:

$$\begin{aligned} \mathcal{C}_1 &= \left\{ \theta^{(i)} \mid y^{(i)} \leq y^* \right\} \\ \mathcal{C}_2 &= \left\{ \theta^{(i)} \mid y^{(i)} > y^* \right\} \end{aligned} \tag{3}$$

Utilizando estimación de densidad por kernel (KDE), en este caso, el método de ventana deslizante de Parzen-Rosenblatt -ver detalles en Sección 6- estimamos las distribuciones de probabilidad para estos dos conjuntos:

$$\begin{aligned} l(\theta) &= p(\theta \mid \theta \in \mathcal{C}_1) \\ g(\theta) &= p(\theta \mid \theta \in \mathcal{C}_2) \end{aligned} \tag{4}$$

El objetivo es seleccionar nuevos hiperparámetros θ que maximicen el Expected Improvement (EI), que es proporcional al cociente de estas dos densidades:

$$\text{EI}(\theta) \propto \frac{l(\theta)}{g(\theta)} \quad (5)$$

Al maximizar $\text{EI}(\theta)$, favorecemos los hiperparámetros que son más probables en el conjunto de buen rendimiento \mathcal{C}_1 y menos probables en el conjunto de peor rendimiento \mathcal{C}_2 . Este proceso se repite iterativamente; en cada iteración, los nuevos datos de rendimiento actualizan $l(\theta)$ y $g(\theta)$, permitiendo explorar eficientemente el espacio de hiperparámetros.

Para Louvain y Leiden, se han ejecutado, respectivamente, 150 iteraciones, explorando resoluciones entre 0.1 y 2.0 que maximicen tanto el coeficiente Q como el puntaje FES, descritos en la Sección 2.4.3. Los datos se han almacenado en una base de datos SQLite, lo que permite extraer informes tabulados y reanudar el ajuste desde el punto de guardado si se desea, facilitando la gestión y continuidad del proceso.

Como se comentó anteriormente, el principal objetivo del clustering en nuestra red PPI es determinar las comunidades funcionales subyacentes con la finalidad de identificar rutas metabólicas compatibles con la elaboración de dianas o biomarcadores para el tratamiento de la FTD. Por esto, no se debe perder la interpretabilidad biológica de la red, a la vez que se intenta maximizar la calidad del clustering aplicado. Dadas estas condiciones, se propone una optimización multiobjetivo, repartiendo las tareas de significancia biológica y calidad del clustering entre las siguientes métricas:

- *Modularidad (Q)*. Destinada a la evaluación de la calidad del clustering aplicado. Valores altos indican una conectividad interna densa dentro de los clústeres y conexiones dispersas entre ellos. La alta modularidad sugiere comunidades bien definidas.
- *Puntaje de Enriquecimiento Funcional*: Para cada clúster, se realizó un análisis de enriquecimiento funcional mediante STRINGdb. Este puntaje refleja la significancia estadística de pathways, términos de la Ontología de Genes u otras anotaciones funcionales enriquecidas dentro de los clústeres. Un puntaje alto indica mayor relevancia biológica, ya que los clústeres enriquecidos en términos específicos probablemente representan procesos biológicos significativos.

Optimización

Se generaron varias configuraciones de modelos al ajustar sistemáticamente los parámetros de los algoritmos y evaluamos cada configuración en base a modularidad y puntaje de enriquecimiento funcional. Al trazar estos puntajes en un frente de Pareto, identificamos las configuraciones que representan el mejor equilibrio entre coherencia estructural (modularidad) y relevancia biológica (enriquecimiento funcional) [67, 68, 69]. Las configuraciones a lo largo de esta frontera representan soluciones Pareto-eficientes de clustering, donde mejorar una métrica no compromete significativamente la otra, permitiendo explorar soluciones con un balance diferente de las métricas de rendimiento elegidas.

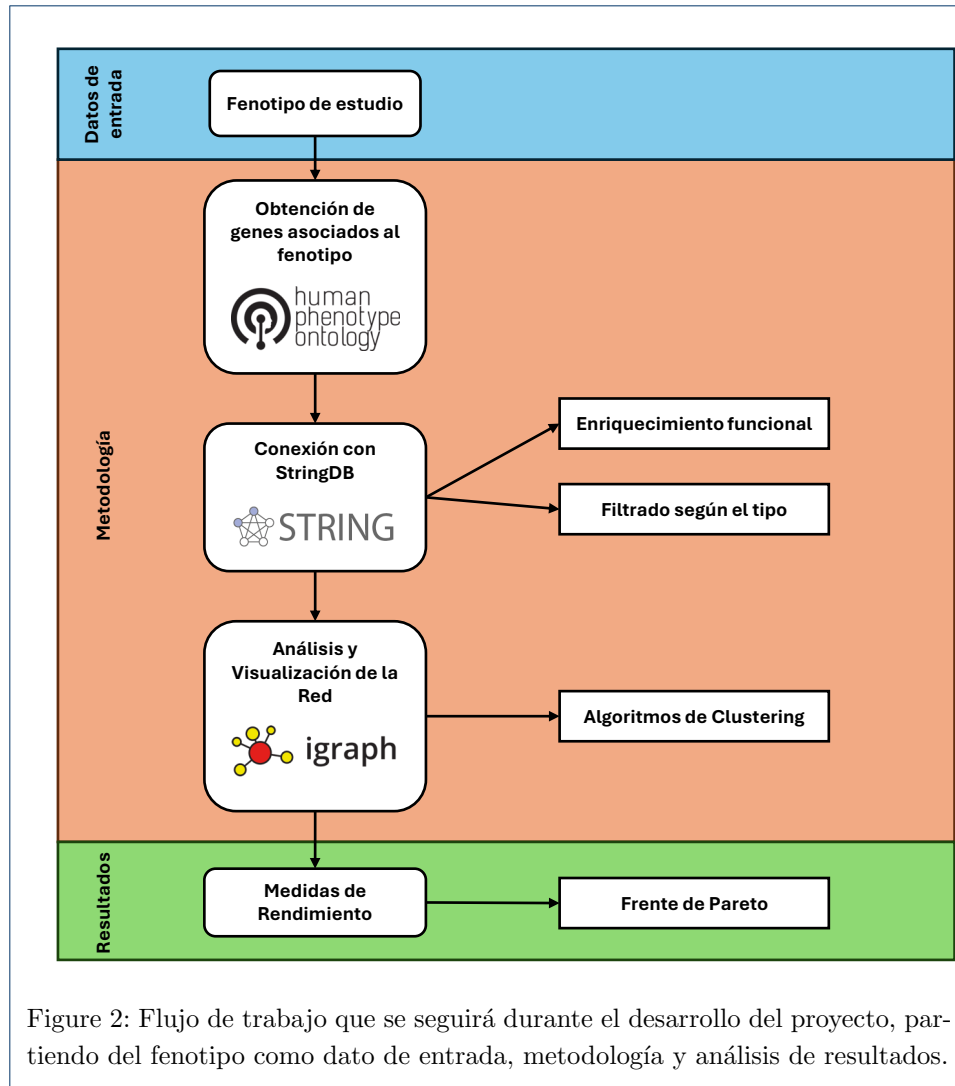


Figure 2: Flujo de trabajo que se seguirá durante el desarrollo del proyecto, partiendo del fenotipo como dato de entrada, metodología y análisis de resultados.

3 Resultados

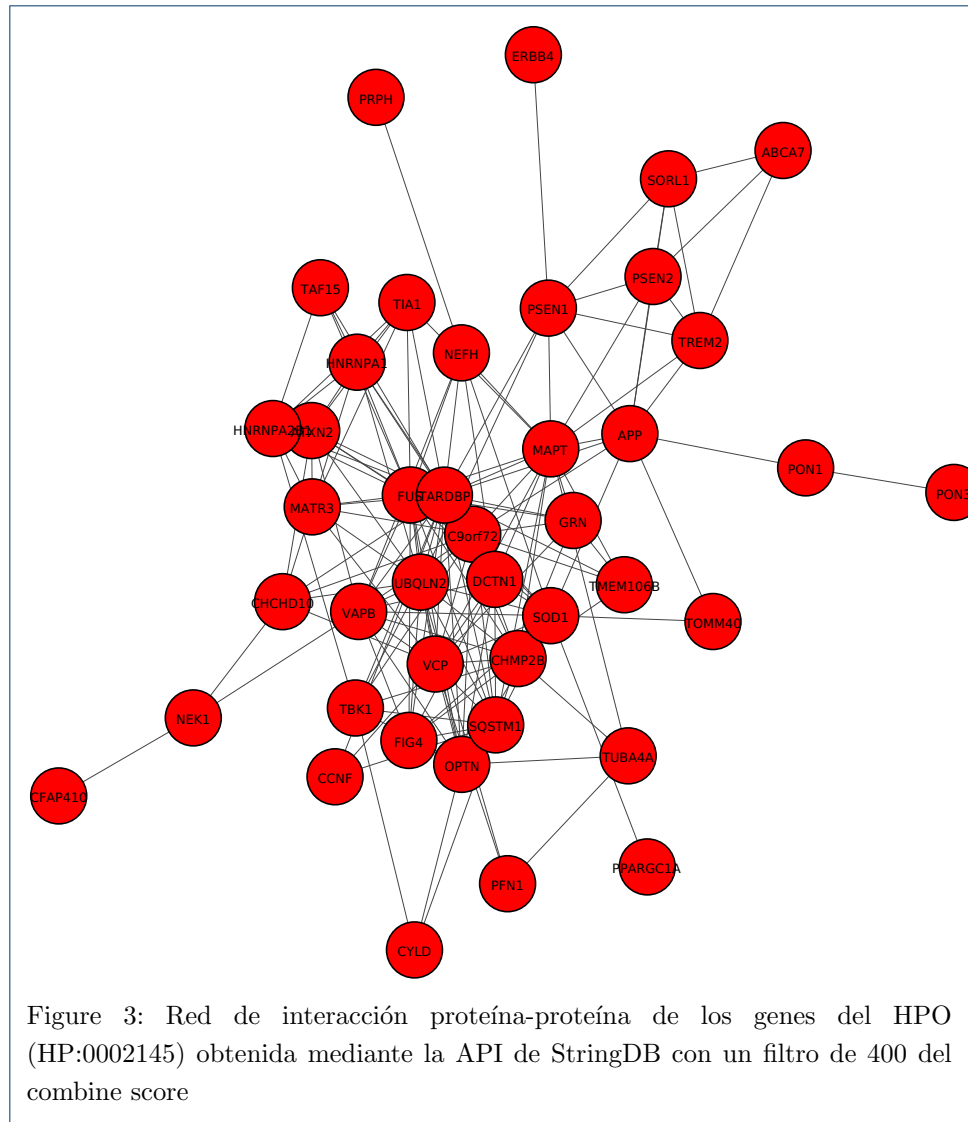
Se ha explorado distintas técnicas de biología de sistemas con el objetivo de entender el fenotipo *Frontotemporal Dementia* y sus relaciones con genes y las funciones de estas. Mediante el análisis de redes de interacción proteína-proteína, de los genes relacionados con el fenotipo, obtenidos mediante StringDB.

3.1 Red PPI y sus propiedades

HPO permite extraer los genes anotados a un fenotipo, por lo que estos han sido descargados y utilizando la API de StringDB se han obtenido la red de interacción (Figura 3). Se ha establecido el umbral de *combine_score* a 700.

En la Figura 3 observamos que proteínas con un número de conexiones respecto a otras. Un análisis en mayor profundidad de la red se presenta en la Tabla 1.

En HPO se tenían 52 genes relacionados con nuestro fenotipo, pero en la red filtrado tan solo hay 42 proteínas. Por tanto hay diez genes, que no interaccionan con las otras proteínas de la red con un alto umbral de confianza. Con un umbral



inferior al establecido, estos genes estarían en la red, pero la significancia biológica de las interacciones no sería fiable.

El grado medio de los nodos, correspondiente al número de conexiones de un nodo, es 8.14, un valor considerablemente alto si tenemos en cuenta el número de nodos presentes en la red. En total, la red tiene 171 conexiones entre proteínas. Hay un valor significativo de desviación típica del grado de los nodos. Un mejor análisis de la distribución de los grados se presenta en la *Figura 4*. En esta figura, observamos que la gran mayoría de las proteínas, tienen grados relativamente bajos, inferiores a la media. Encontramos un par de proteínas, con grados superiores a 20, lo que indica que están conectados a más de la mitad de las proteínas de la red. Estos nodos altamente conectados podrían ser hubs (FUS, TARDBP). Si nos fijamos en la densidad del grafo, vemos que tenemos una red con baja densidad, hay pocas conexiones con respecto a las posibles.

La red obtenida es un grafo conexo (ver *Figura 3*), el análisis de la conectividad llevado a cabo muestra que existe una conectividad tanto de vértice (nodo) como de

Table 1: Resumen de Métricas de la Red PPI

Categoría	Métrica y Valor
Tamaño de la Red	Número de nodos: 42 Número de aristas: 171
Grado	Grado promedio: 8.14 Desviación estándar del grado: 5.93
Conectividad	Grafo conectado: Sí Conectividad de nodos: 1 Conectividad de aristas: 1
Densidad y Sparsity	Densidad: 0.199 Esparsidad: 0.801
Cercanía (Closeness)	Cercanía promedio: 0.456 Desviación estándar de cercanía: 0.093
Centralidad (Betweenness)	Betweenness promedio: 26.50 Desviación estándar de betweenness: 39.56
Transitividad	Transitividad local promedio: 0.550 Desviación estándar de transitividad local: 0.322 Transitividad global: 0.520
Otras Propiedades	Asortatividad: -0.018 Diámetro: 6 Longitud de camino promedio: 2.293

arista de 1. Esto indica que existen uno o varios nodos/aristas que si son eliminados desconectan el grafo. Se ha determinado que estos nodos son las proteínas NEK1, SOD1, APP, PON1, NEFH Y PSEN1.

Las siguientes métrica analizadas son la centralidad y cercanía, en la *Tabla 1* se muestra la media y desviación, ya que se tratan de métricas que se miden por cada nodo. Para ambas métricas destaca una de las dos proteínas antes mencionadas con el grado, TARDBP.

Finalmente podemos mencionar la transitividad, probabilidad de que los vecinos de un nodo estén también conectados entre sí. Tanto la transitividad local como la global tienen valores parecidos, entorno a 0.5. Puntualizar que para el cálculo de la transitividad local, se han obviado aquellos nodos que dan valores de NaN debido a que solo presentan un vecino.

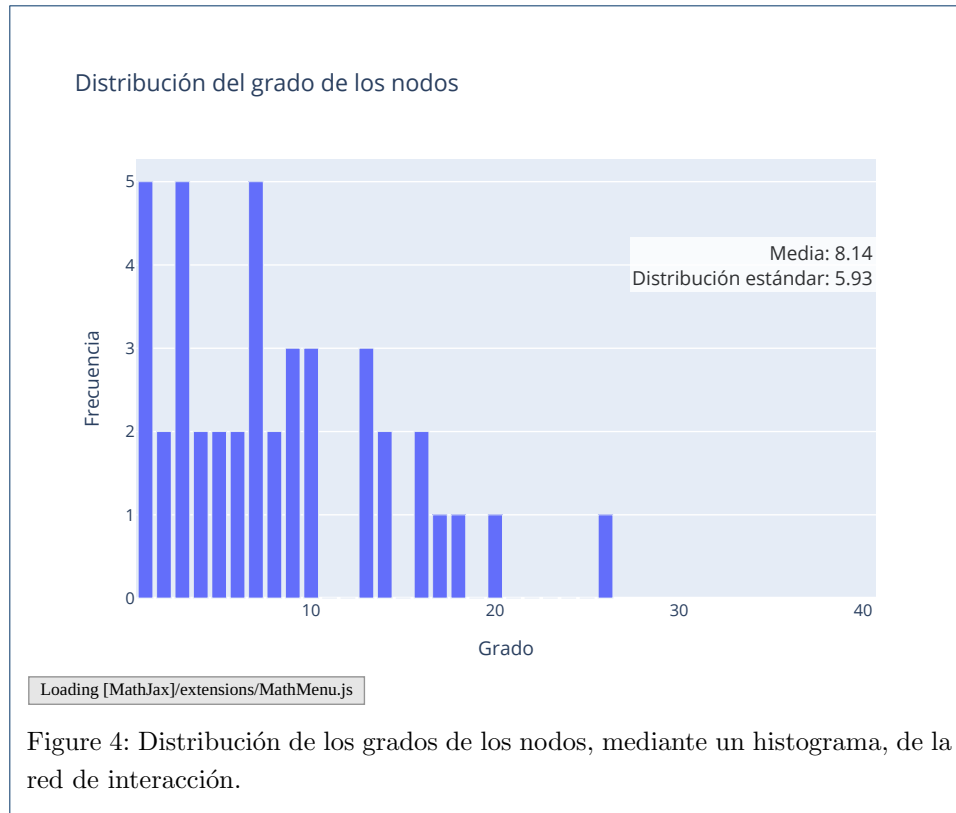
A lo largo de esta sección se presentan los resultados de los procedimientos descritos en la metodología.

3.2 Análisis de la Red

3.3 Optimización de Hiperparámetros

El proceso de optimización de hiperparámetros mediante el BHO se analizará mostrando el frente de Pareto de ambos algoritmos y una visualización del rendimiento marginal de ambas métricas en base al valor de resolución (γ) evaluado.

3.4 Clustering



3.5 Análisis Funcional

4 Discusión

5 Conclusiones

6 Anexo A: KDE mediante ventana deslizante de Parzen-Rosenblatt en el TPE

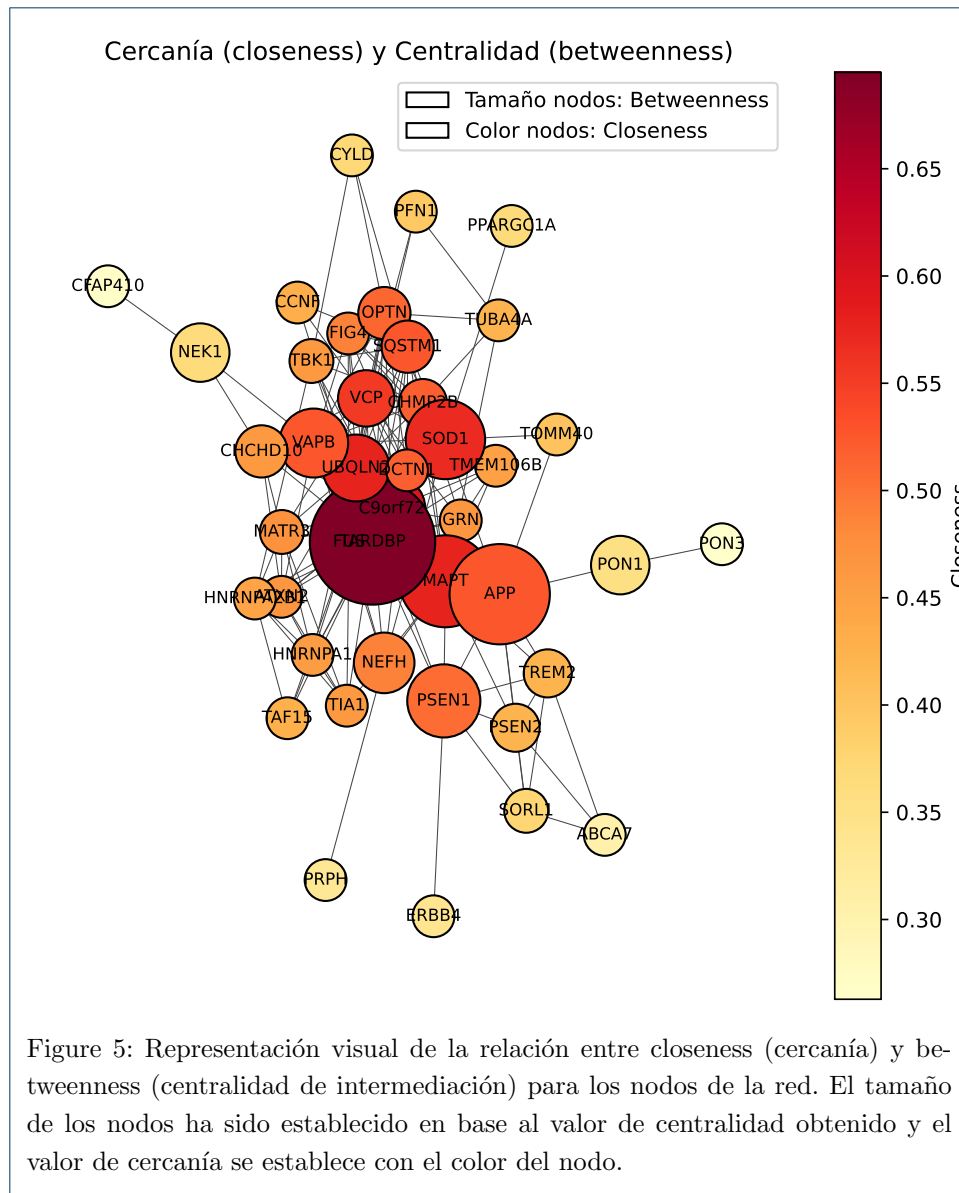
La Estimación de Densidad por Kernel (KDE) mediante el método de ventana deslizante de Parzen-Rosenblatt es una técnica no paramétrica para estimar la función de densidad de probabilidad (PDF) de una variable aleatoria a partir de una muestra de datos. La estimación en un punto x se realiza sumando las contribuciones de cada dato x_i mediante una función kernel K centrada en x_i :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (6)$$

Donde n es el número de muestras, h es el ancho de banda (parámetro de suavizado) y $K(u)$ es la función kernel, comúnmente el kernel gaussiano:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (7)$$

En el contexto del TPE, este método se utiliza para estimar las distribuciones de probabilidad de los hiperparámetros en los conjuntos de buen rendimiento \mathcal{C}_1 y



peor rendimiento \mathcal{C}_2 . Los datos de entrada son los hiperparámetros observados θ_i en cada conjunto, y la salida es una estimación suave de las densidades $l(\theta)$ y $g(\theta)$:

$$\begin{aligned}
 l(\theta) &= \frac{1}{|\mathcal{C}_1|h} \sum_{\theta_i \in \mathcal{C}_1} K\left(\frac{\theta - \theta_i}{h}\right) \\
 g(\theta) &= \frac{1}{|\mathcal{C}_2|h} \sum_{\theta_i \in \mathcal{C}_2} K\left(\frac{\theta - \theta_i}{h}\right)
 \end{aligned} \tag{8}$$

Al utilizar el método de ventana deslizante de Parzen-Rosenblatt con kernel gaussiano, el TPE obtiene una estimación flexible y no paramétrica de las distribuciones de los hiperparámetros. Esto facilita la exploración eficiente del espacio de búsqueda

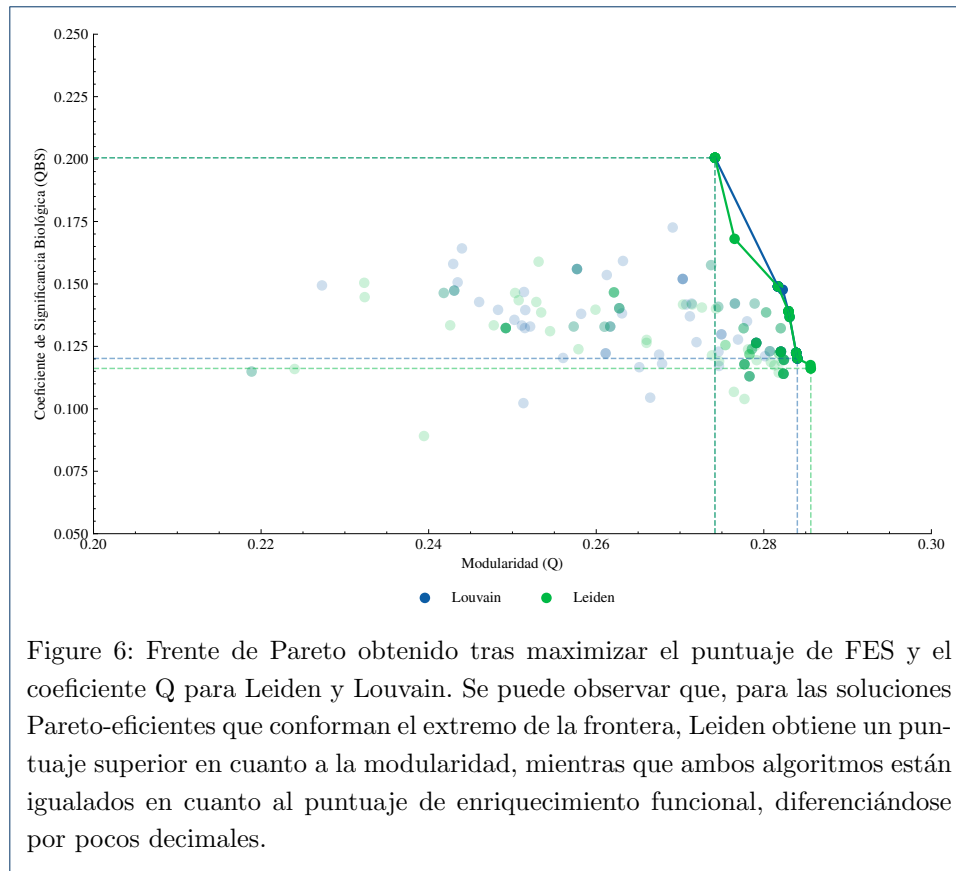


Figure 6: Frente de Pareto obtenido tras maximizar el puntaje de FES y el coeficiente Q para Leiden y Louvain. Se puede observar que, para las soluciones Pareto-eficientes que conforman el extremo de la frontera, Leiden obtiene un puntaje superior en cuanto a la modularidad, mientras que ambos algoritmos están igualados en cuanto al puntaje de enriquecimiento funcional, diferenciándose por pocos decimales.

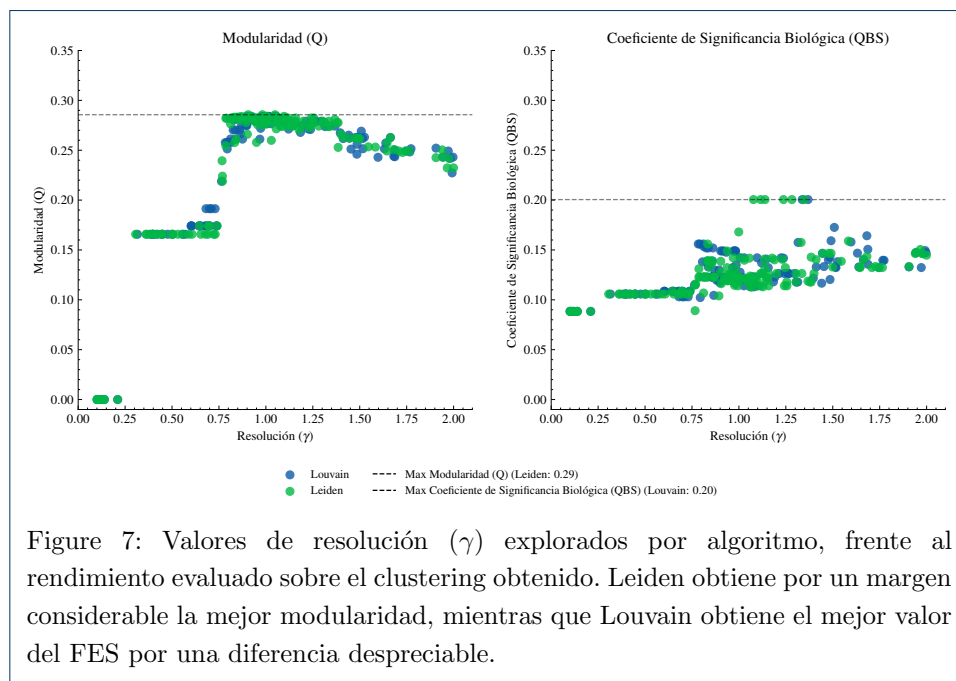
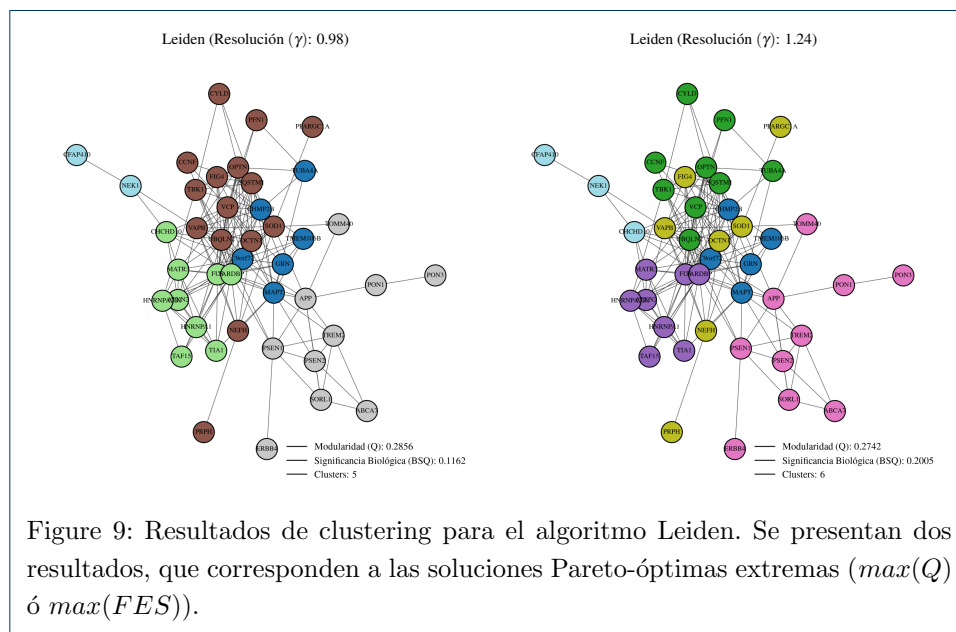
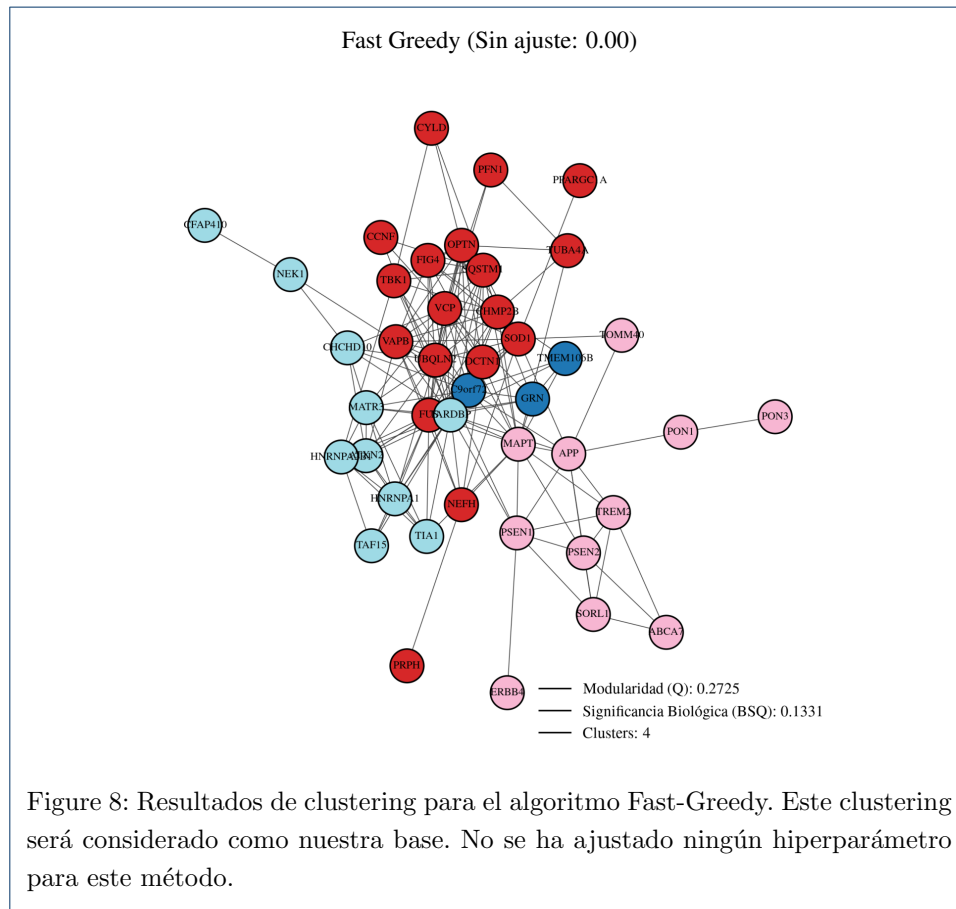
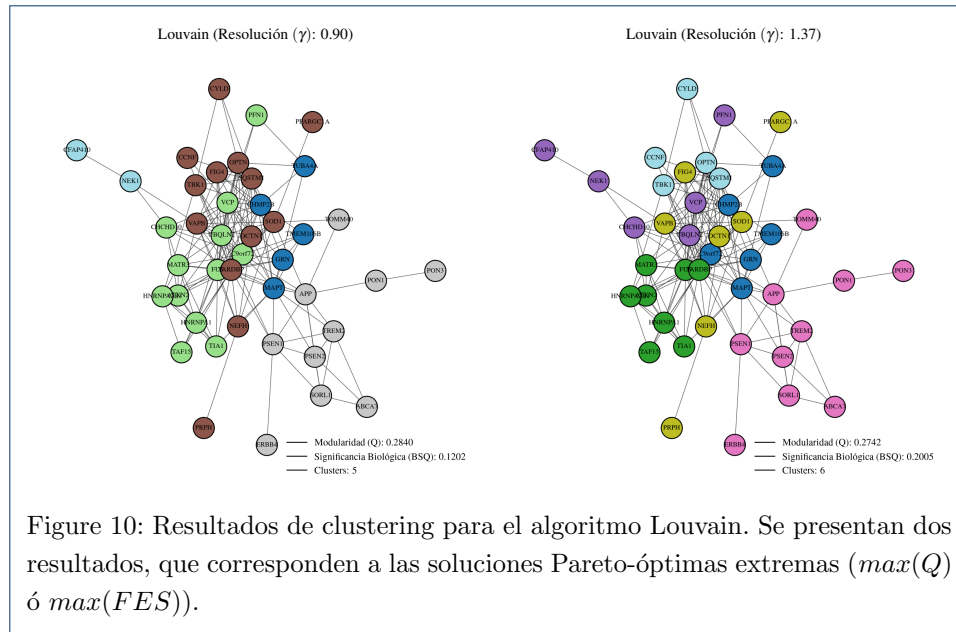


Figure 7: Valores de resolución (γ) explorados por algoritmo, frente al rendimiento evaluado sobre el clustering obtenido. Leiden obtiene por un margen considerable la mejor modularidad, mientras que Louvain obtiene el mejor valor del FES por una diferencia despreciable.





y la identificación de regiones donde es más probable encontrar hiperparámetros que mejoren el rendimiento del modelo.

Abreviaciones

EA → Enfermedad de Alzheimer
 FTD → Demencia Frontotemporal
 bvFTD → Variante del Comportamiento de la Demencia Frontotemporal
 PPA → Afasia Progresiva Primaria
 nvPPA → Variante no fluente/agramática de la PPA
 svPPA → Variante semántica de la PPA
 lvPPA → Variante logopéica de la PPA
 MND → Enfermedad de la Motoneurona
 ELA → Esclerosis Lateral Amiotrófica
 FTLD → Degeneración Lobar Frontotemporal
 MRI → Imagen por Resonancia Magnética
 PET → Tomografía por emisión de positrones
 SPECT → Tomografía por Emisión de Fotón Único
 FDA → Administración de Alimentos y Medicamentos
 GRN → Gen precursor de granulina
 GO → Gene Ontology
 FDR → Funtional False Discovery (tasa de descubrimiento falso)
 BHO → Bayesian Hyperparameter Optimization (ajuste bayesiano de hiperparámetros)
 PDF → Probability Density Function (Función de Densidad de Probabilidad)
 KDE → Kernel Density Estimation (Estimación de la PDF mediante Kernel)

Disponibilidad de datos y materiales

<https://github.com/MarioPasc/project.template.git>

Contribución de los autores

Usando las iniciales que habéis definido al comienzo del documento, debeis indicar la contribución al proyecto en el estilo: J.E : Encargado del análisis de coexpresión con R, escritura de resultados; J.R.S : modelado de red con python y automatizado del código, escritura de métodos; ... OJO: que sea realista con los registros que hay en vuestros repositorios de github.

Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

References

1. Formiga, F., Robles, M.J., Fort, I.: Demencia, una enfermedad evolutiva: demencia severa. identificación de demencia terminal. *Revista Española de Geriatria y Gerontología* **44**, 2–8 (2009). doi:10.1016/j.regg.2009.05.009

2. Long, S., Benoist, C., Weidner, W.: World Alzheimer Report 2023: Reducing Dementia Risk: Never Too Early, Never Too Late. Published by Alzheimer's Disease International (ADI), London.
3. Goodman, R.A., Lochner, K.A., Thambisetty, M., Wingo, T.S., Posner, S.F., Ling, S.M.: Prevalence of dementia subtypes in united states medicare fee-for-service beneficiaries, 2011–2013. *Alzheimer's & Dementia* **13**(1), 28–37 (2017). doi:10.1016/j.jalz.2016.04.002
4. Garre-Olmo, J., et al.: Memòria D'activitat del Registre de Demències de Girona 2016. Institut d'Assistència Sanitària, Girona. <https://www.ias.cat/adjunts/files/MemoriaReDeGi2016.pdf>
5. Snowden, J.S., Neary, D., Mann, D.M.: Frontotemporal dementia. *The British journal of psychiatry* **180**(2), 140–143 (2002)
6. Ratnavalli, E., Brayne, C., Dawson, K., Hodges, J.: The prevalence of frontotemporal dementia. *Neurology* **58**(11), 1615–1621 (2002)
7. Piguet, O., Hornberger, M., Mioshi, E., Hodges, J.R.: Behavioural-variant frontotemporal dementia: diagnosis, clinical staging, and management. *The Lancet Neurology* **10**(2), 162–172 (2011)
8. Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., Ogar, J.M., Rohrer, J.D., Black, S., Boeve, B.F., et al.: Classification of primary progressive aphasia and its variants. *Neurology* **76**(11), 1006–1014 (2011)
9. Chen, X., Chen, Y., Ni, B., Huang, C.: Research trends and hotspots for frontotemporal dementia from 2000 to 2022: a bibliometric analysis. *Frontiers in Neurology* **15** (2024). doi:10.3389/fneur.2024.1399600
10. Freedman, M., Robert, P., et al.: Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* **51**(6), 1546–1554 (1998)
11. Mackenzie, I.R., Rademakers, R., Neumann, M.: Tdp-43 and fus in amyotrophic lateral sclerosis and frontotemporal dementia. *The Lancet Neurology* **9**(10), 995–1007 (2010)
12. Mackenzie, I.R., Foti, D., Woulfe, J., Hurwitz, T.A.: Atypical frontotemporal lobar degeneration with ubiquitin-positive, tdp-43-negative neuronal inclusions. *Brain* **131**(5), 1282–1293 (2008)
13. Mackenzie, I.R., Rademakers, R., Neumann, M.: Tdp-43 and fus in amyotrophic lateral sclerosis and frontotemporal dementia. *The Lancet Neurology* **9**(10), 995–1007 (2010)
14. NHS Inform: Motor Neurone Disease (MND). Accessed: 2024-10-04 (n.d.). <https://www.nhsinform.scot/illnesses-and-conditions/brain-nerves-and-spinal-cord/motor-neurone-disease-mnd/>
15. Burrell, J.R., Kiernan, M.C., Vucic, S., Hodges, J.R.: Motor Neuron dysfunction in frontotemporal dementia. *Brain* **134**(9), 2582–2594 (2011). doi:10.1093/brain/awr195. <https://academic.oup.com/brain/article-pdf/134/9/2582/17869835/awr195.pdf>
16. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.C.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.Y.R., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R., Rademakers, R.: Expanded ggggcc hexanucleotide repeat in noncoding region of c9orf72 causes chromosome 9p-linked ftd and als. *Neuron* **72**, 245–256 (2011). doi:10.1016/J.NEURON.2011.09.011/ATTACHMENT/095EB372-EA80-49C0-886D-91026D6299DE/MMC1.PDF
17. Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y., Hashizume, Y., Oda, T.: Tdp-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochemical and Biophysical Research Communications* **351**, 602–611 (2006). doi:10.1016/j.bbrc.2006.10.093
18. Bussi, C., Ramos, J.M.P., Arroyo, D.S., Gallea, J.I., Ronchi, P., Kolovou, A., Wang, J.M., Florey, O., Celej, M.S., Schwab, Y., Ktistakis, N.T., Iribarren, P.: Alpha-synuclein fibrils recruit tbk1 and optn to lysosomal damage sites and induce autophagy in microglial cells. *Journal of Cell Science* **131** (2018). doi:10.1242/JCS.226241/VIDEO-13
19. Balendra, R., Isaacs, A.M.: C9orf72-mediated als and ftd: multiple pathways to disease. *Nature Reviews Neurology* **14**:9 **14**, 544–558 (2018). doi:10.1038/s41582-018-0047-2
20. Casterton, R.L., Hunt, R.J., Fanto, M.: Pathomechanism heterogeneity in the amyotrophic lateral sclerosis and frontotemporal dementia disease spectrum: Providing focus through the lens of autophagy. *Journal of Molecular Biology* **432**, 2692–2713 (2020). doi:10.1016/j.jmb.2020.02.018
21. Devenney, E., Vucic, S., Hodges, J.R., Kiernan, M.C.: Motor neuron disease-frontotemporal dementia: a clinical continuum. *Expert Review of Neurotherapeutics* **15**, 509–522 (2015). doi:10.1586/14737175.2015.1034108
22. Devenney, E.M., Ahmed, R.M., Hodges, J.R.: Frontotemporal dementia. *Handbook of Clinical Neurology* **167**, 279–299 (2019). doi:10.1016/B978-0-12-804766-8.00015-7
23. Orphanet: Enfermedad neurológica poco frecuente. Accessed: 2024-10-04 (n.d.)
24. Varma, A., Adams, W., Lloyd, J., Carson, K., Snowden, J., Testa, H., Jackson, A., Neary, D.: Diagnostic patterns of regional atrophy on mri and regional cerebral blood flow change on spect in young onset patients with alzheimer's disease, frontotemporal dementia and vascular dementia. *Acta neurologica scandinavica* **105**(4), 261–269 (2002)
25. Kanda, T., Ishii, K., Uemura, T., Miyamoto, N., Yoshikawa, T., Kono, A.K., Mori, E.: Comparison of grey matter and metabolic reductions in frontotemporal dementia using fdg-pet and voxel-based morphometric mr studies. *European journal of nuclear medicine and molecular imaging* **35**, 2227–2234 (2008)
26. Rowe, C.C., Ng, S., Ackermann, U., Gong, S.J., Pike, K., Savage, G., Cowie, T.F., Dickinson, K.L., Maruff, P., Darby, D., et al.: Imaging β -amyloid burden in aging and dementia. *Neurology* **68**(20), 1718–1725 (2007)
27. Sirkis, D.W., Geier, E.G., Bonham, L.W., Karch, C.M., Yokoyama, J.S.: Recent advances in the genetics of frontotemporal dementia. *Current genetic medicine reports* **7**, 41–52 (2019)
28. Pan, X.-d., Chen, X.-c.: Clinic, neuropathology and molecular genetics of frontotemporal dementia: a mini-review. *Translational neurodegeneration* **2**, 1–9 (2013)
29. Caillet-Boudin, M.-L., Buée, L., Sergeant, N., Lefebvre, B.: Regulation of human mapt gene expression. *Molecular neurodegeneration* **10**, 1–14 (2015)

30. Bateman, A., Bennett, H.P.: The granulin gene family: from cancer to dementia. *Bioessays* **31**(11), 1245–1254 (2009)
31. Mackenzie, I.R., Frick, P., Neumann, M.: The neuropathology associated with repeat expansions in the c9orf72 gene. *Acta neuropathologica* **127**, 347–357 (2014)
32. Borroni, B., Archetti, S., Del Bo, R., Papetti, A., Buratti, E., Bonvicini, C., Agosti, C., Cosseddu, M., Turla, M., Di Lorenzo, D., *et al.*: Tardbp mutations in frontotemporal lobar degeneration: frequency, clinical features, and disease course. *Rejuvenation research* **13**(5), 509–517 (2010)
33. Wong, T.H., Pottier, C., Hondius, D.C., Meeter, L.H., Van Rooij, J.G., Melhem, S., bank, N.B., van Minkelen, R., van Duijn, C.M., Rozemuller, A.J., *et al.*: Three vcp mutations in patients with frontotemporal dementia. *Journal of Alzheimer's Disease* **65**(4), 1139–1146 (2018)
34. M Isaacs, A., Johannsen, P., Holm, I., E Nielsen, J., *et al.*: Frontotemporal dementia caused by chmp2b mutations. *Current Alzheimer Research* **8**(3), 246–251 (2011)
35. Granadillo de Luque, J.L., Zarante, I.: Genética de la demencia frontotemporal. *Revista colombiana de psiquiatría* **37**(1), 94–112 (2008)
36. Rosso, S., Landweer, E., Houterman, M., Kaat, L.D., Van Duijn, C., Van Swieten, J.: Medical and environmental risk factors for sporadic frontotemporal dementia: a retrospective case–control study. *Journal of Neurology, Neurosurgery & Psychiatry* **74**(11), 1574–1576 (2003)
37. Hodges, J.R., Davies, R., Xuereb, J., Casey, B., Broe, M., Bak, T., Kril, J., Halliday, G.: Survival in frontotemporal dementia. *Neurology* **61**(3), 349–354 (2003)
38. Neurotherapeutics: New approaches to the treatment of frontotemporal dementia. SpringerLink (2023)
39. Bio, P.: upliFT-D: A Clinical Trial for Frontotemporal Dementia with Progranulin Gene Mutations (2023). <https://passagebioclinicaltrials.com/Frontotemporal-Dementia/en/about-this-trial>
40. Clinic, M.: Frontotemporal Dementia Clinical Trials - Mayo Clinic Research (2023). <https://www.mayo.edu/research/clinical-trials/diseases-conditions/frontotemporal-dementia>
41. of Neurological Disorders, N.I., (NINDS), S.: Treatment Study of Frontotemporal Dementia (2023). <https://www.ninds.nih.gov/health-information/clinical-trials/treatment-study-frontotemporal-dementia>
42. Peet, B.T., Spina, S., Mundada, N., La Joie, R.: Neuroimaging in frontotemporal dementia: heterogeneity and relationships with underlying neuropathology. *Neurotherapeutics* **18**(2), 728–752 (2021)
43. The Human Phenotype Ontology Consortium: HPO Ontology Annotation Network API. Accessed: 2024-11-05 (2024). <https://ontology.jax.org/api/network/docs>
44. The Human Phenotype Ontology Consortium: Human Phenotype Ontology. Accessed: 2024-11-05 (2024). <https://hpo.jax.org/>
45. Szklarczyk, Damian, Gable, Annika L., Lyon, David, Junge, Alexander, Wyder, Stefan, Huerta-Cepas, Jaime, Simonovic, Milan, Doncheva, Nadezhda T., Morris, John H., Bork, Peer, *et al.*: STRING API. Accessed: 2024-11-06 (2024). <https://string-db.org/help/api/>
46. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryar, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J., Mering, C.V.: The string database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research* **51**, 638–646 (2023). doi:10.1093/NAR/GKAC1000
47. pandas development team, T.: pandas: Powerful data structures for data analysis in Python. Versión 2.2.3 (2024). <https://pandas.pydata.org>
48. Hunter, J.D.: Matplotlib: A 2D Graphics Environment. Versión 3.8.1 (2007). <https://matplotlib.org>
49. Nelson, G.L.: SciencePlots: Matplotlib styles for scientific figures. Versión 2.1.1 (2021). <https://github.com/garrettj403/SciencePlots>
50. Reitz, K., colaboradores de Requests: Requests: HTTP for Humans. Versión 2.31.0 (2024). <https://docs.python-requests.org>
51. Klopfenstein, D.V., Dangl, L., Wilfert, A., Lombardot, E., Mariadassou, H.S., Goetz, W.: GOATOOLS: A Python library for Gene Ontology analyses. Versión 1.2.3 (2018). <https://github.com/tanghaibao/goatools>
52. Rätty, R., y otros, T.L.: gProfiler: A web server for functional enrichment analysis and conversions of gene lists. Versión 1.4.0 (2019). <https://github.com/biit/gprofiler>
53. y Josef Perktold, S.S.: Statsmodels: Econometric and statistical modeling with Python. Versión 0.15.0 (2010). <https://www.statsmodels.org>
54. Martin Maechler, A.S.M.H.y.K.H. Peter Rousseeuw: Cluster: Cluster Analysis Basics and Extensions. (2022). Versión 2.1.5. <https://cran.r-project.org/package=cluster>
55. y Fabian Mundt, A.K.: Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. (2017). Versión 1.0.8. <https://cran.r-project.org/package=factoextra>
56. y Steve Horvath, P.L.: WGCNA: Weighted Correlation Network Analysis. (2008). Versión 1.71. <https://cran.r-project.org/package=WGCNA>
57. y Tamás Nepusz, G.C.: Igraph: Network Analysis and Visualization. (2006). Versión 1.5.0. <https://igraph.org/r/>
58. Damian Szklarczyk, D.L.A.J.S.W.J.H.-C.M.S.N.T.D.J.H.M.P.B.y.o. Annika L. Gable: STRINGdb: Functional Protein Association Networks. (2019). Versión 2.10.0. <https://cran.r-project.org/package=STRINGdb>
59. y Cory Brouwer, W.L.: Pathview: An R/Bioconductor Package for Pathway-based Data Integration and Visualization. (2013). Versión 1.36.0. <https://bioconductor.org/packages/pathview/>
60. Guangchuang Yu, Y.H.y.Q.-Y.H. Li-Gen Wang: clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. (2012). Versión 4.15.0. <https://bioconductor.org/packages/clusterProfiler/>
61. Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D., Bader, G.D.: Pathway enrichment analysis and visualization of omics data using g:profiler, gsea, cytoscape and enrichmentmap. *Nature Protocols* **14**(2), 482–517 (2019). doi:10.1038/s41596-018-0103-9
62. Tipney, H., Hunter, L.: An introduction to effective use of enrichment analysis software. *Human Genomics* **4**(3), 202 (2010). doi:10.1186/1479-7364-4-3-202

63. Pons, P., Latapy, M.: Computing communities in large networks using random walks. In: Computer and Information Sciences-ISCIS 2005: 20th International Symposium, Istanbul, Turkey, October 26-28, 2005. Proceedings 20, pp. 284–293 (2005). Springer
64. Traag, V.A., Waltman, L., Van Eck, N.J.: From louvain to leiden: guaranteeing well-connected communities. *Scientific reports* **9**(1), 1–12 (2019)
65. Traag, V.A., Van Dooren, P., Nesterov, Y.: Narrow scope for resolution-limit-free community detection. *Phys. Rev. E* **84**, 016114 (2011). doi:10.1103/PhysRevE.84.016114
66. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008** (2008). doi:10.1088/1742-5468/2008/10/p10008. Algoritmo de Louvain en clustering
67. Goodarzi, E., Ziaei, M., Hosseinipour, E.Z.: *Introduction to Optimization Analysis in Hydrosystem Engineering*, pp. 111–148. Springer, Berlin/Heidelberg (2014)
68. Jahan, A., Edwards, K.L., Bahraminasab, M.: *Multi-criteria Decision Analysis*, 2nd edn., pp. 63–65. Elsevier, Amsterdam (2013)
69. Costa, N.R., Lourenço, J.A.: Exploring pareto frontiers in the response surface methodology. In: Yang, G.-C., Ao, S.-I., Gelman, L. (eds.) *Transactions on Engineering Technologies: World Congress on Engineering 2014*, pp. 399–412. Springer, Berlin/Heidelberg (2015)