

RESEARCH

Análisis de la Esclerosis Lateral Amiotrófica a través de la biología de sistemas

Santiago Juarroz Surballe^{*}, Gabriela Milenova Yordanova, Maga C. Chediack and Sebastián Rozenblum

^{*}Correspondence:
santiagojuarroz@uma.es

ETSI Informática, Universidad de
Málaga, Málaga, España

Full list of author information is
available at the end of the article

Abstract

Keywords: ELA; gen; interacción; ontología

1 Introducción

Se estima que más del 70% de las muertes en el mundo están relacionadas con enfermedades no transmisibles, es decir, patologías de larga duración y progresión lenta que no se transmiten entre personas y que incluyen afecciones como el cáncer, las enfermedades cardiovasculares o las neurodegenerativas [1]. Dentro de este último grupo, patologías como la esclerosis lateral amiotrófica (ELA) suponen un desafío particular, tanto por la complejidad de su diagnóstico como por la evolución impredecible de sus síntomas [2]. En este contexto, la detección temprana supone una ventaja y, al mismo tiempo, un desafío crucial para mitigar el impacto clínico en los pacientes y optimizar los sistemas de salud.

A día de hoy, los avances tecnológicos en el área de la biología de sistemas, abarcando campos más específicos como la genómica, la proteómica y la metabolómica, han demostrado producir resultados significativos en el diagnóstico y la detección de numerosas patologías [3]. Estos enfoques han sido aplicados con éxito en el análisis de enfermedades complejas como el cáncer [4], la diabetes tipo 2 [5] y las enfermedades neurodegenerativas [6], permitiendo identificar biomarcadores, redes moleculares disfuncionales y posibles dianas terapéuticas.

Para interpretar la gran cantidad de datos genómicos y vincularlos a manifestaciones clínicas, es fundamental utilizar vocabularios estandarizados. En este marco, el sistema Human Phenotype Ontology (HPO) [7] describe tanto signos como síntomas y sus características clínicas asociadas a las enfermedades humanas. La ELA, definida por un conjunto característico de síntomas, se representa en la HPO como un fenotipo específico bajo el identificador *HP:0007354*, el cual se asocia a un grupo determinado de genes [8].

La ELA se caracteriza por una degeneración de las neuronas motoras, tanto a nivel superior como a nivel inferior [9]. Esta patología tiene una tasa anual de diagnóstico a nivel mundial de entre 2 y 11 casos por cada 100.000 habitantes [10]. Específicamente, en Europa la incidencia anual es de 2.16 casos por cada 100.000 habitantes [10]. Esta afecta en mayor medida a los hombres (3 por cada 100.000 habitantes), por sobre las mujeres (2.4 por cada 100.00 habitantes) [10]. La enfermedad es mayoritariamente esporádica, si bien aproximadamente el 5-10% de los casos se presenta en formas familiares con herencia autosómica dominante [11, 12]. En las formas juveniles de ELA también se han descrito mutaciones con herencia autosómica recesiva, siendo estas menos frecuentes pero clínicamente relevantes [13].

Si bien todavía no se conocen completamente las causas de la ELA, se ha avanzado sustancialmente en comprender su patogénesis y base genética. Se han identificado mutaciones en múltiples genes —entre ellos *SOD1*, *C9orf72*, *TARDBP*, *FUS*, *NEK1*, *OPTN*, *TBK1*, entre otros— que pueden causar o aumentar el riesgo de ELA, tanto en formas familiares como, en algunos casos, esporádicas [14]. En particular, las mutaciones en *SOD1* fueron de las primeras identificadas y siguen siendo relevantes: estos cambios genéticos pueden alterar la superóxido dismutasa 1, una proteína que, al plegarse de forma incorrecta, genera toxicidad mediante mecanismos de disfunción mitocondrial [15]. Además, los fenotipos asociados a mutaciones en este gen son muy variables [16]. Se estima que mutaciones en cuatro genes —*SOD1*, *TARDBP*, *FUS* y *C9orf72*— explican más del 50% de los casos familiares de ELA [17]. No obstante, el número de genes implicados continúa aumentando conforme se producen avances en la investigación.

En la Table 1 se presenta un resumen de la clasificación funcional de los principales genes asociados a la ELA según el mecanismo biológico en el que participan:

Categoría funcional	Genes asociados	Función principal alterada
Procesamiento de ARN	<i>TARDBP</i> , <i>FUS</i> , <i>HNRNPA1</i> , <i>HNRNPA2B1</i> , <i>MATR3</i> , <i>SETX</i> , <i>TIA1</i> , <i>GLE1</i>	Alteraciones en transcripción, splicing y transporte de ARN.
Autofagia y degradación proteica	<i>SQSTM1</i> , <i>OPTN</i> , <i>TBK1</i> , <i>UBQLN2</i> , <i>VCP</i> , <i>CCNF</i> , <i>CHMP2B</i> , <i>CYLD</i> , <i>SPG11</i>	Acumulación de proteínas mal plegadas y fallo en los mecanismos de eliminación celular.
Estrés oxidativo y disfunción mitocondrial	<i>SOD1</i> , <i>CHCHD10</i> , <i>PPARGC1A</i> , <i>VAPB</i> , <i>SIGMAR1</i> , <i>DAO</i> , <i>GLT8D1</i> , <i>TRPM7</i>	Producción excesiva de especies reactivas de oxígeno y daño mitocondrial.
Metabolismo lipídico y oxidativo	<i>SPTLC1</i> , <i>PON1</i> , <i>PON2</i> , <i>PON3</i> , <i>ANG</i> , <i>ANXA11</i>	Disregulación del metabolismo lipídico y alteraciones en la homeostasis oxidativa.
Procesos neurodegenerativos	<i>MAPT</i> , <i>PSEN1</i> , <i>TREM2</i> , <i>ALS2</i>	Solapamiento con otras enfermedades neurodegenerativas (ej., Alzheimer o FTD).

Table 1 Principales genes asociados a la ELA, agrupados según su categoría funcional y la función alterada [2, 12, 18, 19, 20].

La enfermedad se diagnostica mediante criterios clínicos basados en la sintomatología, ya que no existen biomarcadores claros para la enfermedad [21]. Esta limitación causa un retraso en el diagnóstico y la intervención temprana. Los tratamientos actuales se basan en disminuir la progresión de la enfermedad y atenuar los síntomas lo máximo posible para mejorar la calidad de vida, ya que todavía no existe cura [22]. Desde el punto de vista farmacológico, Riluzole es uno de los medicamentos con mayor evidencia de prolongar la supervivencia, aunque su efecto es

más bien modesto [23, 24]. Además, existen otros fármacos autorizados en ciertos países, como Edaravone, que en poblaciones seleccionadas pueden ralentizar el deterioro funcional [25].

Por lo tanto, este trabajo tiene como objetivo analizar las relaciones entre los genes asociados a la ELA y sus fenotipos correspondientes para identificar patrones que puedan conducir a un análisis estadístico más robusto de la enfermedad. Mediante el uso de técnicas de biología de sistemas, incluyendo las herramientas de construcción y análisis de redes basadas en datos de interacción, se busca comprender mejor los mecanismos moleculares de la enfermedad.

2 Materiales y métodos

Para este trabajo se realizó un análisis exhaustivo de la red de interacción proteína-proteína (PPI, *protein-protein interaction*) del fenotipo de interés, en donde se utilizaron diferentes herramientas que posibilitaron el estudio de la red. Entre ellas destaca STRING (*Search Tool for the Retrieval of Interacting Genes/protein*), una base de datos biológica y un servidor web de acceso libre enfocado principalmente en las PPIs. Esta herramienta es útil para integrar las asociaciones funcionales entre las proteínas de un organismo.

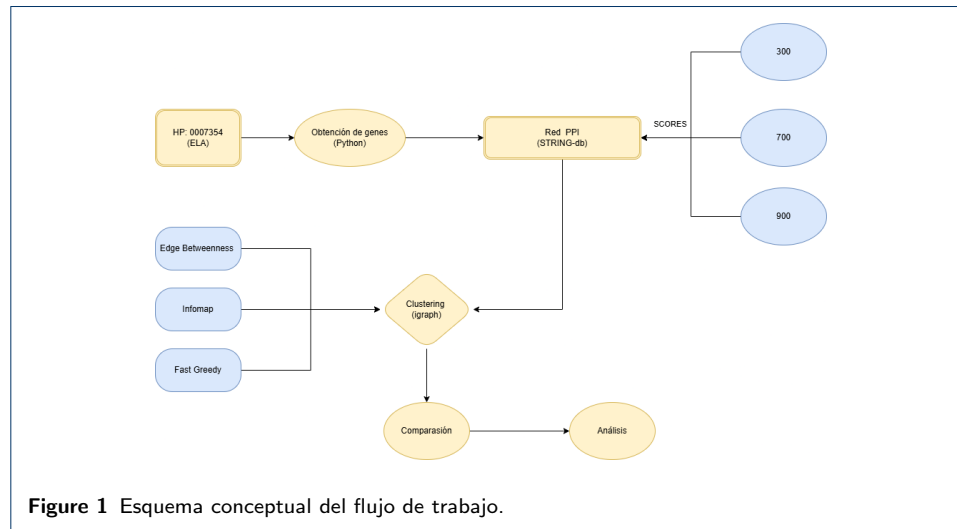
A su vez se utilizó HPO (Human Gene Ontology), un vocabulario estandarizado que utilizan médicos e investigadores para describir de manera uniforme y precisa anomalías fenotípicas observadas en enfermedades humanas.

La implementación se realizó en Python, debido a su versatilidad, simplicidad en la sintaxis y amplia disponibilidad de bibliotecas especializadas, que facilitan tanto la comunicación con las APIs de las bases de datos más utilizadas como la representación gráfica de redes complejas. Específicamente, se utilizaron las siguientes librerías:

- **Numpy** (v. 1.26.4): librería esencial para cálculos numéricos y operaciones matriciales.
- **Matplotlib** (v. 3.9.2): librería para visualizar datos e imágenes y generar gráficas.
- **Pandas** (v. 2.2.3): librería para manipular, limpiar y analizar datos en forma de tablas (*DataFrames*).
- **Requests** (v. 2.32.3): librería para realizar solicitudes HTTP y comunicarse con APIs o servicios web.
- **Json**: módulo para leer, escribir y procesar datos en formato JSON. Es parte de la biblioteca estándar de Python, por lo que no requiere instalación ni tiene versión propia.
- **NetworkX** (v 3.6.0): librería para crear, analizar y visualizar redes o grafos, con soporte para métricas y tareas adicionales.

El flujo de trabajo diseñado ([Figure 1](#)) consta de cinco etapas que abarcan desde la obtención de genes relacionados con el fenotipo de interés hasta la extracción

de conocimiento biológico de la red PPI. Para algunas de las etapas se realizaron múltiples ejecuciones, variando en cada una los parámetros principales. De esta forma, se obtuvieron múltiples combinaciones de resultados, con el objetivo de evaluar qué configuración presentaba los resultados más favorables para luego profundizar sobre ella.



En primer lugar, se recopiló la lista de genes asociados al fenotipo de interés. Para hacerlo, se adoptaron dos vías distintas: por un lado, se utilizaron los genes encontrados tras una revisión bibliográfica manual (Table 1, 35 genes); por el otro, se extrajeron programáticamente a partir de la API de la HPO (49 genes). Los 35 genes de la primera estrategia (la cual, a lo largo del trabajo, será referida como "manual") están contenidos dentro de los 49 genes de la segunda.

A continuación, mediante la API de STRING, se generó la red PPI para cada una de las dos listas de genes. En este punto se propusieron tres posibles umbrales de puntuación para dar lugar a la red: uno laxo, con un valor de 300; otro intermedio, de 700; y el más estricto, igual a 900. Esto permitió comparar cómo varía la estructura de la red en función del nivel de confianza asignado a las interacciones, evaluando si los patrones observados se mantienen consistentes al aumentar el umbral de evidencia requerido. Una vez obtenidas las seis redes, se representaron gráficamente mediante la librería *NetworkX*. Para cada una de ellas, se llevó a cabo un pequeño análisis exploratorio inicial, en el cual se midieron distintas propiedades básicas de topología como número de nodos, grado, densidad, centralidad, modularidad y dispersión.

Luego, resultó necesario establecer los posibles agrupamientos o *clusters* que tuvo cada red. Este proceso se llevó a cabo con el objetivo de descubrir grupos de genes altamente conectados entre sí, los cuales podrían corresponder a funciones biológicas relacionadas con la ELA. Nuevamente, se propusieron tres estrategias diferentes de agrupamiento:

- **Clustering basado en la centralidad de intermediación de enlace (Edge Betweenness Clustering)**: es un método de detección de comu-

nidades propuesto por Girvan y Newman en el año 2002 [26] Este algoritmo consiste en calcular la centralidad de intermediación para todas las aristas donde se eliminan aquellas con los valores más altos, dicho proceso se hace de manera iterativa con el objetivo de separar la red en subconjuntos mayormente conectados de forma interna.

Este proceso continúa hasta obtener la red con agrupamientos que presenten el valor más alto de la modularidad, la cual cuantifica la calidad de la conexión interna de los agrupamientos respecto al resto de la red.. Dicha métrica tiene un intervalo $[-1,1]$ donde el valor más cercano o igual a 1 determina que el agrupamiento está bien definido (Table 2).

Valor de Q	Interpretación
$Q \approx 1$	Agrupamiento fuerte y bien definido.
$0 < Q < 1$	Partición clara, aunque no completamente aislada, ya que existe cierta conectividad entre los agrupamientos.
$Q \approx 0$	Estructura compuesta principalmente por conexiones aleatorias.
$-1 < Q < 0$	La partición es peor que el azar, con más conexiones entre comunidades que dentro de ellas.
$Q \approx -1$	Agrupamiento totalmente incoherente.

Table 2 Significado de los valores de la modularidad (Q).

- **Clustering basado en el algoritmo InfoMap:** es un método de detección de comunidades propuesto por Martin Rosvall y Carl T. Bergstrom en el año 2008 [27], que no depende de una métrica estructural predeterminada y detecta comunidades de tamaño variable con submódulos jerárquicos. Este algoritmo detecta los agrupamientos analizando el flujo de información entre nodos usando caminatas aleatorias que simulan el movimiento dentro de una red y eligen el próximo nodo al azar. Las múltiples caminatas se utilizan para calcular la cantidad de bits que se necesitan para describir el recorrido: si el recorrido pasó una gran parte dentro de un mismo grupo de nodos se considera una comunidad densamente conectada internamente; en el caso contrario, se considera que hay una conexión más débil entre comunidades.
- **Clustering basado en el algoritmo Fast Greedy:** es un método jerárquico que optimiza directamente la modularidad de una red para detectar comunidades, introducido por Clauset, Newman y Moore (2004) [28]. Parte de la hipótesis de que una buena partición es aquella en la que los nodos se conectan con mayor frecuencia dentro de su propio grupo que con nodos externos. Comienza considerando cada nodo como una comunidad independiente y, de forma iterativa, fusiona los pares de comunidades cuya unión produce el mayor aumento en la modularidad global. Este criterio local permite construir una estructura jerárquica (dendrograma) que refleja niveles anidados de organización. Gracias a su implementación eficiente, es especialmente útil para redes grandes o densas, donde métodos más costosos como el Edge Betweenness se vuelven inviables.

A partir de estas ejecuciones, se generaron dieciocho resultados en total: tres conjuntos de *clusters* —uno por cada algoritmo— para cada una de las seis redes

analizadas. Luego, se llevó a cabo sobre cada uno de ellos un análisis de enriquecimiento funcional de sobrerrepresentación (ORA, *Over-Representation Analysis*) como aproximación inicial al posible significado biológico de los módulos individuales identificados en cada red. Para estos análisis de sobrerrepresentación, así como para todos los análisis posteriores, se consideraron únicamente aquellos *clusters* con al menos 3 genes, con el fin de evitar ruido derivado de agrupamientos demasiado pequeños y asegurar una mínima coherencia funcional. Además, se estableció como valor umbral de significancia estadística un *p*-valor ajustado de 0.05, y se recurrió a Gene Ontology (GO) Biological Process como base de datos de referencia.

Finalmente, se examinó el equilibrio entre el número de *clusters* y la media de funciones enriquecidas por *cluster*, buscando la configuración con el mejor compromiso entre estructura de la red y coherencia funcional.

Para llevar a cabo un estudio aún más exhaustivo de la red con mejor configuración según el equilibrio previamente explicado, se recurrió a dos enfoques opuestos pero complementarios. Por una parte, se utilizó una validación técnica, basada en principios de topología de redes (calculando las métricas ya mencionadas como densidad, centralidad y modularidad, entre otras), pero teniendo en cuenta el número de *clusters*. Por otro lado, para profundizar en el significado biológico, se ejecutaron enriquecimientos funcionales adicionales. A los resultados ya obtenidos con GO Biological Process, se añadieron dos bases de datos ampliamente utilizadas: Kyoto Encyclopedia of Genes and Genomes (KEGG) y Reactome. Estas bases de datos son complementarias entre sí, ya que GO se centra en procesos biológicos generales, KEGG incorpora rutas metabólicas y de señalización, y Reactome se especializa en eventos moleculares y vías celulares.

3 Resultados

A partir de la aplicación de la metodología, para el caso de la lista creada a partir del HPO obtuvimos los datos que se muestran en la tabla 3. Mientras que para la lista manual se obtuvieron los datos que se muestran en la tabla 4.

ID	Edge Betweenness	Infomap	Fast Greedy
300	12	2	4
700	24	2	4
900	8	6	6

Table 3 Resultados de agrupamiento por umbral y algoritmo HPO.

ID	Edge Betweenness	Infomap	Fast Greedy
300	2	2	4
700	15	2	5
900	7	4	4

Table 4 Resultados de agrupamiento por umbral y obtenidos manualmente.

Como se muestra en las tablas, el análisis de las redes mostró un balance óptimo al utilizar el score de **700**. En contraste el umbral de 300 generó una red con una estructura laxa, mientras que el score 900 resultó estricto, limitando las interacciones.



figures/COMPARATIVA DE ALGORITMOS DE AGRUPAMIENTO.png

Figure 2 Esquema conceptual del flujo de trabajo.

En la figura se ilustra el rendimiento de los métodos de agrupamiento utilizados. En ella se revela que el algoritmo de Edge Betweenness ofreció los resultados más favorables, tanto para la red generada manualmente como para la basada en HPO.

4 Discusión

5 Conclusiones

Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

References

1. World Health Organization: Noncommunicable diseases. <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases> (2023)
2. Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E.M., Logroscino, G., Robberecht, W., Shaw, P.J., Simmons, Z., van den Berg, L.H.: Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers* **3**(1), 17071 (2017). doi:10.1038/nrdp.2017.71
3. Kitano, H.: Systems biology: a brief overview. *Science* **295**(5560), 1662–1664 (2002). doi:10.1126/science.1069492
4. Barabási, A.-L., Gulbahce, N., Loscalzo, J.: Network medicine: a network-based approach to human disease. *Nature Reviews Genetics* **12**(1), 56–68 (2011). doi:10.1038/nrg2918
5. Nielsen, J.: Systems biology of metabolism: a driver for developing personalized and precision medicine. *Cell Metabolism* **25**(3), 572–579 (2017). doi:10.1016/j.cmet.2017.02.002
6. Tian, M., Chen, J., Yang, Z., Peng, X.: Systems biology approaches to understanding neurodegenerative diseases. *Cellular and Molecular Life Sciences* **79**(4), 1–22 (2022). doi:10.1007/s00018-022-04261-9
7. The Human Phenotype Ontology Consortium: Human Phenotype Ontology. <https://hpo.jax.org/>. Accessed: 2024-10-16 (2025)

8. Human Phenotype Ontology Consortium: HP:0007354: Amyotrophic lateral sclerosis. <https://hpo.jax.org/browse/term/HP:0007354>. Accessed: October 16, 2024
9. Xie, M., Pallegar, P.N., Parusel, S., Nguyen, A.T., Wu, L.-J.: Regulation of cortical hyperexcitability in amyotrophic lateral sclerosis: focusing on glial mechanisms. *Molecular Neurodegeneration* **18**, 75 (2023). doi:[10.1186/s13024-023-00665-w](https://doi.org/10.1186/s13024-023-00665-w)
10. Beghi, E., Logroscino, G.: Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. *The Lancet* **376**(9743), 513–523 (2010). doi:[10.1016/S0140-6736\(10\)61156-7](https://doi.org/10.1016/S0140-6736(10)61156-7)
11. Barberio, J., Lally, C., Kupelian, V., Hardiman, O., Flanders, W.D.: Estimated familial amyotrophic lateral sclerosis proportion: A literature review and meta-analysis. *Neurology: Genetics* **9**(6), 200109 (2023). doi:[10.1212/NXG.0000000000200109](https://doi.org/10.1212/NXG.0000000000200109)
12. Kiernan, M.C., Vucic, S., Talbot, K., McDermott, C.J., Hardiman, O., Shefner, J.M., Al-Chalabi, A., Shaw, P.J.: Amyotrophic lateral sclerosis. *The Lancet* **397**(10280), 394–406 (2021). doi:[10.1016/S0140-6736\(20\)30959-9](https://doi.org/10.1016/S0140-6736(20)30959-9)
13. Wu, C., Fan, D.: A novel missense mutation of the *ddhd1* gene associated with juvenile amyotrophic lateral sclerosis. *Frontiers in Aging Neuroscience* **8**, 291 (2016). doi:[10.3389/fnagi.2016.00291](https://doi.org/10.3389/fnagi.2016.00291)
14. Van Daele, S.H., Moisse, M., van Vugt, J.J.F.A., Zwamborn, R.A.J., van Rheeën, W., Van Eijk, K., Kenna, K., Corcia, P., Couratier, P., Hardiman, O., *et al.*: Genetic variability in sporadic amyotrophic lateral sclerosis. *Brain* **146**(9), 3760–3769 (2023). doi:[10.1093/brain/awad120](https://doi.org/10.1093/brain/awad120)
15. Gagliardi, D., Ripellino, P., Meneri, M., *et al.*: Clinical and molecular features of patients with amyotrophic lateral sclerosis and *sod1* mutations: a monocentric study. *Frontiers in Neurology* **14**, 1169689 (2023). doi:[10.3389/fneur.2023.1169689](https://doi.org/10.3389/fneur.2023.1169689)
16. Berdyński, M., Safranow, K., Andersen, P.M., Żekanowski, C.: Phenotypic characterization of *als*-causing *sod1* mutations affecting polypeptide length. *Human Mutation* **2025**, 9792233 (2025). doi:[10.1155/humu/9792233](https://doi.org/10.1155/humu/9792233)
17. Rummens, J., Da Cruz, S.: Rna-binding proteins in *als* and *ftd*: from pathogenic mechanisms to therapeutic insights. *Molecular Neurodegeneration* **20**, 64 (2025). doi:[10.1186/s13024-025-00851-y](https://doi.org/10.1186/s13024-025-00851-y)
18. Chia, R., Chiò, A., Traynor, B.J.: Genetics and pathogenesis of amyotrophic lateral sclerosis. *Nature Reviews Neurology* **14**(11), 651–664 (2018). doi:[10.1038/s41582-018-0047-0](https://doi.org/10.1038/s41582-018-0047-0)
19. Mejzini, B., Flynn, L.L., Pitout, I.L., Fletcher, S., Wilton, S.D., Akkari, P.A.: *Als* genetics, mechanisms, and therapeutics: where are we now? *Frontiers in Neuroscience* **13**, 1310 (2019). doi:[10.3389/fnins.2019.01310](https://doi.org/10.3389/fnins.2019.01310)
20. Renton, A.E., Chio, A., Traynor, B.J.: State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience* **17**(1), 17–23 (2014). doi:[10.1038/nn.3584](https://doi.org/10.1038/nn.3584)
21. Bjornevik, K., Zhang, Z., Berry, J.D., Clish, C.B., Matthews, P., Schwarzschild, M.A., Cudkowicz, M.E., Weisskopf, M.G., Ascherio, A.: Prediagnostic neurofilament light chain levels in amyotrophic lateral sclerosis. *Neurology* **101**(1), 13–23 (2023). doi:[10.1212/WNL.0000000000207296](https://doi.org/10.1212/WNL.0000000000207296)
22. Lynch, K.: Optimizing pharmacologic treatment for *als* to improve outcomes and quality of life. *The American Journal of Managed Care* **29**(7 Suppl), 112–119 (2023). doi:[10.37765/ajmc.2023.89389](https://doi.org/10.37765/ajmc.2023.89389)
23. Miller, R.G., Mitchell, J.D., Moore, D.H., Swash, M.: Riluzole for amyotrophic lateral sclerosis (*als*)/motor neuron disease (*mnd*). *Cochrane Database of Systematic Reviews* (3) (2012). doi:[10.1002/14651858.CD001447.pub3](https://doi.org/10.1002/14651858.CD001447.pub3)
24. Hinchcliffe, M., Smith, A., *et al.*: Riluzole: real-world evidence supports significant extension of median survival times in patients with amyotrophic lateral sclerosis. *Degenerative Neurological and Neuromuscular Disease* **10**, 111–118 (2020). doi:[10.2147/DNND.S268523](https://doi.org/10.2147/DNND.S268523)
25. Brooks, B.R., Wynn, D., Mitumoto, H., *et al.*: Efficacy of intravenous edaravone in slowing disability progression in patients with amyotrophic lateral sclerosis: an individual patient data meta-analysis. *Therapeutic Advances in Neurological Disorders* **17**, 17562864231220996 (2024). doi:[10.1177/17562864231220996](https://doi.org/10.1177/17562864231220996)
26. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Physical Review E* **69**(2), 026113 (2004). doi:[10.1103/PhysRevE.69.026113](https://doi.org/10.1103/PhysRevE.69.026113)
27. Rosvall, M., Bergstrom, C.T.: Maps of information flow reveal community structure in complex networks. *Proceedings of the National Academy of Sciences* **105**(4), 1118–1123 (2008). doi:[10.1073/pnas.0706851105](https://doi.org/10.1073/pnas.0706851105)
28. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. *Physical Review E* **70**(6), 066111 (2004). doi:[10.1103/PhysRevE.70.066111](https://doi.org/10.1103/PhysRevE.70.066111)