

F-50-01
PROTOCOLO DE INVESTIGACIÓN



Universidad Autónoma de Querétaro

Facultad de Informática

C.U. a 014/08/2024



LICENCIATURA EN INGENIERÍA DE SOFTWARE

Me permito presentar ante el H. Consejo Académico de la Facultad de Informática la siguiente solicitud de titulación por tesis de acuerdo con la fracción VII, artículos: 45-54 y 55-58 del reglamento de titulación vigente.

Nombre del Alumno: Karel Alexis Vázquez Suárez

No. de Expediente: 297137

No. CVU: 2026735

Correo electrónico: karel.vs.alex@gmail.com

Teléfono: 4423701249

Línea de investigación sobre la cual se desarrollará el trabajo: Inteligencia artificial

Responsable de la Asesoría: Jorge Luis Pérez Ramos

Programas educativos que apoya: Sof 18

Línea de investigación del responsable: Análisis de rutas mínimas en espacios discretos

Firma de aceptación:

Responsable de la Co-Asesoría: Roberto Carlos Álvarez Martínez

Programas educativos que apoya: Sof 18

Línea de investigación del responsable: Sistemas complejos , redes y biología de sistemas.

Firma de aceptación:

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Número de veces que se ha sometido esta propuesta al H. Consejo Académico de la Facultad: Primera vez (X) 2da. () Más de 2 veces ()

Semestre o Cuatrimestre actual del alumno: 9

Fecha de Ingreso al programa: 25 de agosto del 2020

I. DATOS GENERALES

- Reconocimiento y clasificación de microbiomas con técnicas de fusión de redes de similitud
- Karel Alexis Vázquez Suárez
- Jorge Luis Pérez Ramos¹, Roberto Carlos Álvarez Martínez²
- (Centro de Investigación e Innovación en Ciencias de la Computación y Tecnología)¹ - FIF - UAQ, (Laboratorio de Biología Cuantitativa y Sistemas Complejos)² - FCN - UAQ
- Investigación aplicada
- Inteligencia artificial aplicada

II. ANTECEDENTES Y/O FUNDAMENTACIÓN TEÓRICA

Las redes neuronales artificiales (ANN, por sus siglas en inglés), están enfocadas en replicar el funcionamiento del cerebro humano por medio de conexiones de procesamiento de información, las cuales, se pueden asimilar a una neurona y principalmente actúan como nodos que permiten realizar las conexiones entre sí para mantener la comunicación entre los mismos y tener la capacidad de procesar la misma red en un sistema simple (Moreno, s/f). Ahora, se ha conseguido replicar el proceso de una neurona utilizando diversos métodos y algoritmos para poder modificar la información que recibimos, lo que nos permite tener una mejor percepción de cómo funcionan los datos y cómo es que estos pueden manejarse de una forma más conveniente. Es por ello, que este tipo de redes trabajan con

F-50-01
PROTOCOLO DE INVESTIGACIÓN

modelos de procesamiento sensorial, por lo que, con la información de esos primeros modelos se logró encontrar una aplicación utilizando algoritmos que puedan asemejarse a una neurona real (Krogh, 2008).

No solo las ANN nos han permitido tener una mayor comprensión sobre cómo funciona un modelo computacional, sino que también estas redes pueden tener un mejor procesamiento de unidades simples que se asemejen a neuronas, las cuales trabajan en paralelo en capas interconectadas (Cichy & Kaiser, 2019). Así mismo, las redes neuronales profundas (DNN, por sus siglas en inglés) nos ofrecen diversas herramientas para cada una de las diferentes áreas de estudio que existen, desde, procesamiento del lenguaje natural, visión por computadora, reconocimiento de voz, análisis y clasificación de imágenes, con el fin de ser más precisos en el procesamiento de datos y así obtener, una alta precisión a los interacciones buscadas. En donde, se busca tener un desempeño que en muchos de los casos está a la par de un humano (Mehrer et al., 2020; Montavon et al., 2018).

En otras palabras, las DNN actualmente nos ofrecen un sin fin de herramientas las cuales, podemos aplicar a nuestro proyecto, ya nos ofrece una alta fidelidad de los resultados; ahora bien, una DNN se refiere a su característica principal de utilizar múltiples capas, con este tipo de arquitectura, es posible abordar los problemas más complicados. Dado que, una DNN permite que se puedan solucionar problemas complejos distribuyendo la información de una única capa a múltiples capas de una nueva red, lo que nos permite profundizar en la conectividad de la red y así asimilar la información permitiendo el aprendizaje haciendo que este sea más eficiente (Kriegeskorte, 2015). Esta capacidad de aprendizaje profundo ha abierto un abanico de posibilidades, y una de las áreas que promete la integración de múltiples fuentes de datos (Angermueller et al., 2016).

F-50-01
PROTOCOLO DE INVESTIGACIÓN

En este contexto, la evolución tecnológica nos ha permitido recopilar múltiples conjuntos de datos con los que se ha abierto la posibilidad de utilizar diversos métodos de integración, que a su vez, deben superar diversas adversidades que aumentan la complejidad de dichos métodos. Es así que ha predominado la estrategia de realizar análisis independientes o aislados, que consiste en agrupar cada fuente de datos para centrarse en un solo grupo de patrones, que suele generar un sesgo en la información que procesamos. Siendo así que, la implementación de un modelo de fusión de redes de similitud (SNF, por sus siglas en inglés), se distingue de los análisis por grupo o clasificación integrando a las redes una única red de similitud haciendo que esta comparta la información y se complementen mutuamente, lo que hace más visible la identificación de patrones y agrupaciones (Mac Aogáin et al., 2021; Wang et al., 2014).

Por consiguiente, el uso de las redes neuronales no se ha enfocado en una única línea de investigación, sino que tienen un sin fin de aplicaciones como al área de la informática, ingeniería, economía, medicina, biología, entre muchas otras en las que se pueden usar para obtener los mejores resultados. Aunque, se busca trabajar principalmente en el área de la bioinformática, con el fin de desarrollar modelos de reconocimiento y clasificación a niveles de expresión genética utilizando datos de secuenciación masiva, con el fin de comprender las interacciones que influyen en un microbioma.

Una de las técnicas que nos permitirá recolectar estos datos es con la secuenciación de nueva generación, que es un grupo de tecnologías que fueron desarrolladas para secuenciar grandes segmentos de ADN de forma masiva (*Secuenciación de nueva generación (NGS) de ADN*, s/f). Como resultado de realizar la secuenciación del ADN con esta técnica, se permite obtener un gran conjunto de datos con los que se pueden clasificar en tablas de identificación OTU (Unidades Taxonómicas Operativas), donde, podemos trabajar para mejorar la comprensión del genoma de diferentes microorganismos existentes en un

F-50-01
PROTOCOLO DE INVESTIGACIÓN

microbioma (*Comparison of Next-Generation Sequencing Systems - Liu - 2012 - BioMed Research International - Wiley Online Library, s/f*).

III. JUSTIFICACIÓN

El propósito de este estudio va de la mano con la constante evolución tecnológica y lo rentable que se ha hecho recopilar los diversos datos del genoma (Wang et al., 2014). Es por ello que ha surgido la necesidad de utilizar diversos métodos y modelos computacionales que nos permitan combinar esta información para obtener una visión más amplia sobre un proceso biológico en específico y determinar cómo es que este nos puede afectar. Por esta razón, la fusión de redes de similitud (SNF) se presenta como una herramienta con la cual es más viable realizar una integración entre las bacterias, virus y hongos que puedan encontrar en la microbiota (Zhang et al., 2017).

Por consiguiente, la aplicación de SNF en la computación biológica es un modelo ideal para obtener resultados alternos a los obtenidos, esto debido que se hace un análisis integrativo de los datos, considerando que, estos no se aíslan cómo comúnmente se efectúa en diferentes investigaciones sobre la microbiota, aun así, tener una perspectiva diferente de cómo influyen los datos en conjunto lo que nos permite identificar alteraciones significativas que se produzcan al paso del tiempo, prediciendo así un comportamiento diferente al esperado (Jiang & Hu, 2014).

Teniendo en cuenta que la SNF es útil no únicamente en la computación biológica, sino que, esta tiene la flexibilidad de utilizarse en otras ramas, como lo son la investigación aplicada, ingeniería, biotecnología e incluso la medicina, con el objeto de ampliar la información que visualizamos con los datos recopilados.

IV. DESCRIPCIÓN DEL PROBLEMA

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Con la actual necesidad de innovar y comprender las funciones que puede desempeñar un microbioma, es de suma importancia identificar cómo impacta esencialmente en la salud humana (Ursell et al., 2012), es por ello que se ha encontrado una brecha de oportunidad utilizando las redes neuronales, permitiéndonos realizar análisis de datos de secuencias de nueva generación. Siendo así nos encontramos que con los datos de secuenciación de un gen son ideales para desarrollar una red neuronal que nos permita comprender y determinar cómo es que pueden interactuar las diferentes comunidades encontradas en este microbioma las cuales pueden estar integradas por bacterias, hongos y virus. Sin embargo, actualmente se pueden encontrar diversos estudios que realizan un análisis aislados de cada una de estas comunidades, siendo que este aspecto nos permite tener un conocimiento de cómo puede actuar en su propio ambiente.

,En consecuencia el uso de las redes neuronales tiene el propósito de procesar la información asemejando a una neurona, esto con el fin de identificar o reconocer patrones existentes en una serie de datos o imágenes, ya que, con la recolección de datos e información concisa podemos obtener modelos predictivos que sin necesidad de hacer suposiciones, estas nos den resultados consistentes, veraces y fiables. De modo que utilizando aplicaciones de biología computacional o bioinformática, se busca implementar el uso de las redes neuronales para poder procesar, identificar, clasificar y desarrollar diferentes modelos neurales; estos pueden ser para identificar patrones o realizar predicciones sobre cómo interactúan estos microorganismos (Angermueller et al., 2016). De tal modo que, utilizando el modelo de Fusión de Redes de Similitud (SNF), un modelo que nos permite “fusionar” dos o más redes de similitud lo que nos permite conocer e identificar diferentes perspectivas sobre los datos que se van a emplear. Por lo tanto, esta aplicación es ideal para poder descentralizar el análisis de la microbiota, que en la mayoría de aplicaciones busca hacer una investigación en

F-50-01
PROTOCOLO DE INVESTIGACIÓN

torno a aislar uno de los microorganismos presentes en la microbiota (Jarada et al., 2021).

Por lo tanto el objetivo principal es conocer cómo se integran las bacterias, virus y hongos en un microbioma con el fin de comprender cómo las interacciones influyen en la estabilidad, crecimiento y funcionamiento del microbioma.

V. COMITÉ DE BIOÉTICA

En el presente proyecto no se requiere un comité de ética, debido a que la información que se utilizó en el proyecto se encuentra en “MGnify”, la cual es una base de datos de libre acceso (Richardson et al., 2023). Sin embargo, es importante considerar que en la investigación se pueden relacionar ciertos aspectos a la ética del proyecto.

1. Exactitud de los datos: Para garantizar la precisión y confiabilidad de los datos utilizados, se verificó la procedencia y fuente de cada uno de ellos. De esta manera, se asegura que todos los datos sean precisos, fiables y verificables.
2. Responsabilidad: Es crucial realizar una investigación responsable, garantizando que cumpla con las normas éticas y legales para asegurar su calidad.
3. Privacidad y confidencialidad: Principalmente se debe de proteger la privacidad y confidencialidad de los datos, si y sólo si, los datos a utilizar contienen información confidencial de personas que puedan ser identificadas, de ser así se debe de realizar un consentimiento explícito de cada uno de los participantes.

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Al tener en cuenta estos factores y para garantizar la privacidad y confidencialidad de los datos, la investigación utilizará datos con identificadores para manejo de la información. Esto nos permite trabajar con los datos sin necesidad de integrar aspectos de privacidad y confidencialidad, lo que simplifica la manipulación y análisis de la información.

VI. PLANTEAMIENTO TEÓRICO

En el presente trabajo presentaremos puntualmente las preguntas de investigación que nos servirán como punto de partida en el desarrollo del proyecto. A continuación, planteamos las preguntas, hipótesis y los supuestos de investigación.

Preguntas de investigación

- ¿Es posible generar una clasificación de SNF en un microbioma?
- ¿Aplicar técnicas de SNF nos permitirá comprender más eficientemente la interacción de un microbioma?
- ¿Existe alguna diferencia entre utilizar técnicas de SNF y los métodos tradicionales de redes neuronales?

Hipótesis, Supuestos y/o proposiciones de investigación

Hi: Si se tiene una recolección de datos de secuencias de nueva generación de diferentes microbiomas, entonces se pueden aplicar redes de similitud de cada microbiota para así comparar, clasificar e identificar interacciones entre los diferentes microorganismos.

Ho: Si se tiene una recolección de datos de secuenciación de nueva generación de diferentes microbiomas, entonces no se pueden aplicar redes de similitud de cada microbiota para así comparar, clasificar e identificar interacciones entre los diferentes microorganismos.

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Supuestos:

- Suponemos que un análisis integral nos permitirá identificar cambios que no son visibles al realizar redes independientes para cada comunidad.
- Suponemos que puede existir interacciones que sean considerables al realizar las redes de similitud.
- Suponemos que en la forma en que se muestran los datos fusionados sufrirán un cambio en cómo se visualiza la información.

Proposiciones de investigación:

- Comparar los diferentes métodos y procesos en los que se analizan los genes de secuenciación para comprender e identificar si es que existe alguna diferencia notable.
- Investigar las diferentes técnicas y algoritmos que generan las secuencias de nueva generación utilizando diferentes áreas de muestra para comparar cómo influyen con respecto al área muestra.
- Visualizar y analizar los parámetros generados por las redes de similitud e inferir cómo es que las redes interactúan sobre los conjuntos de datos.
- Investigar diferentes algoritmos y técnicas que se utilizan para realizar redes de similitud y que datos de las secuencias son seleccionables para realizarlas.
- Investigar las diferentes clasificaciones que pueden tomar los datos para mejorar y optimizar la generación de modelos SNF.

VII. OBJETIVOS

Objetivo general:

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Generar un modelo de fusión de redes de similitud (SNF), utilizando datos de secuenciación masiva perteneciente a varios repositorios de libre acceso que incluyen al sitio EBI Metagenomics y diferentes artículos de investigación que comparten los datos utilizados. Para así, entender cómo interactúa un microbioma aplicando técnicas de SNF.

Objetivos específicos:

1. Seleccionar la información y filtrar los datos a utilizar, para que, posteriormente se pueda procesar y analizar.
2. Realizar un pre-procesamiento en los datos para generar las primeras redes de similitud, estas pueden ser de cualquier comunidad (bacterias, hongos y virus).
3. Desarrollar la fusión de las redes de similitud.
4. Analizar las relaciones generadas al fusionar las redes de similitud identificando si es que estas fluctúan con respecto a una tendencia o al tiempo.

VIII. METODOLOGÍA

Para explorar cada uno de los pasos a seguir, se propone un diagrama de flujo (Fig. 3) que abarca todos los puntos a desarrollar. Este diagrama permite visualizar cada una de las actividades esenciales para el progreso del proyecto. El proceso comienza con la recolección y recopilación de fuentes bibliográficas, seguida de la selección de datos a utilizar a lo largo del proyecto, por consiguiente, se realizará la descarga, procesado, y limpieza de los mismos datos, lo que nos permitirá identificar si cada una de los conjuntos de datos son ideales para la investigación o existe la necesidad de recolectar más conjuntos de datos.

Para continuar con el proceso técnicas de SNF, se realizará un análisis estadístico

F-50-01
PROTOCOLO DE INVESTIGACIÓN

con los datos obtenidos de las redes de similitud y se compararán con las redes fusionadas. De este modo, se busca generar redes de similitud, como se ejemplifica en el artículo de Wang et al. (2014). La Fig. 1, tomada de Wang et al. (2014), muestra la generación de las redes de similitud, que sirven como base para la creación de las SNF presentadas en la Fig. 2.

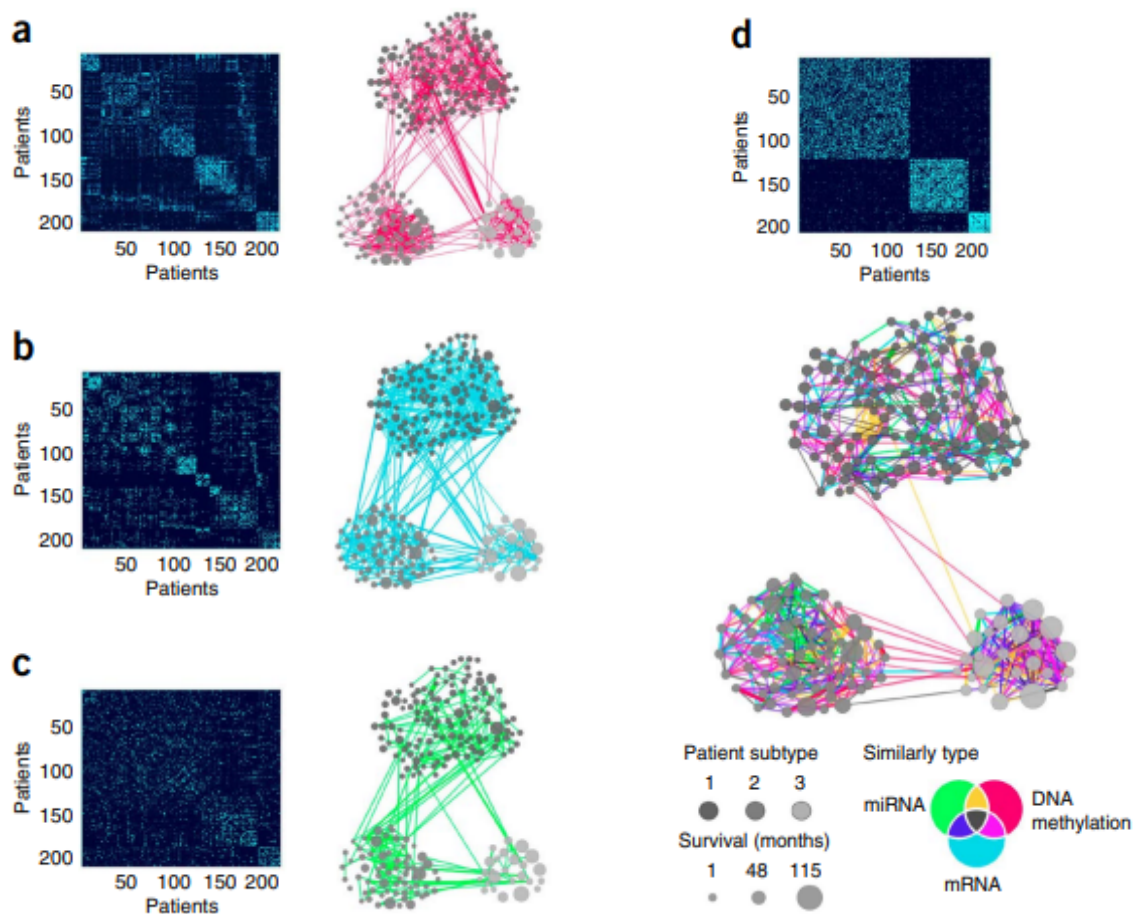


Fig 1. Generación de redes de similitud. Adaptado de Wang et al. (2014).

Fig 2. Redes de similitud fusionadas (SNF). Adaptado de Wang et al. (2014).

F-50-01
PROTOCOLO DE INVESTIGACIÓN

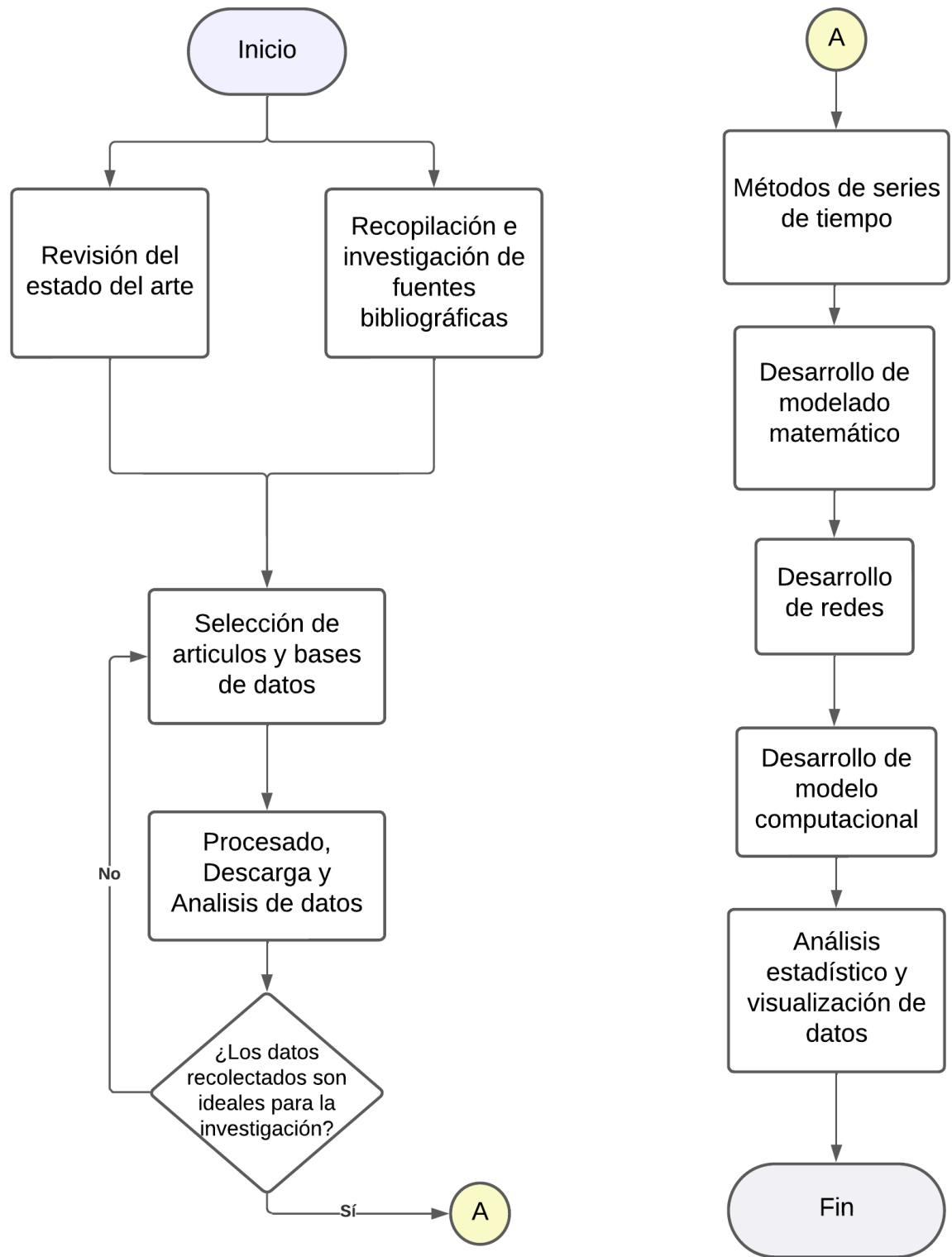


Fig 3. Diagrama de flujo de metodología.

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Preparación y Selección de Datos

Se realizará una búsqueda exhaustiva para seleccionar secuencias de artículos de investigación y bases de datos libres, recopilando artículos con datos e información necesarios. Se clasificarán estos artículos teniendo en cuenta diversos factores, como su procedencia, las comunidades encontradas en el bioma y su microbiota asociada. Esto permitirá una organización más coherente y precisa de los datos disponibles para su posterior análisis.

Procesado, Descarga y Análisis de Datos

La recopilación de datos para un análisis metagenómico de secuencias de nueva generación se llevará a cabo obteniendo datos de la microbiota intestinal humana a partir de bases de datos públicas, tales como el sitio EBI Metagenomics y el Student Microbiome Project. Posteriormente, se seleccionarán y filtrarán datos de secuencias de nueva generación de diversas muestras, utilizando un corte de disimilitud del 3% para la identificación de las tablas OTU. Este filtrado permitirá una selección más precisa y relevante de las secuencias a analizar.

Clasificación de Datos y Metaanálisis

Los datos obtenidos serán clasificados según sus metadatos, con el objetivo de identificar condiciones de salud alteradas. A través de un metaanálisis de las distribuciones de las abundancias relativas, se buscarán señales de criticalidad utilizando la transformada de Legendre. Esta metodología permitirá obtener una visión más profunda de cómo las distintas condiciones afectan la microbiota y cómo se manifiestan estas alteraciones en los datos recolectados.

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Métodos de Series de Tiempo

Para medir los cambios en la memoria a corto plazo y otras propiedades estadísticas como la varianza, la asimetría y la curtosis, se emplearán métodos métricos y basados en modelos. Estos métodos permitirán detectar de manera temprana los cambios repentinos, proporcionando información crucial sobre la dinámica de la microbiota a lo largo del tiempo.

Modelado Matemático

Se utilizarán herramientas computacionales avanzadas, como los algoritmos genéticos y la regresión simbólica con la herramienta Eureka, para abordar enfoques sin modelos predeterminados. Además, se aplicarán modelos clásicos como el modelo generalizado Lotka-Volterra y el modelo de consumo de recursos de McArthur, enriqueciendo el análisis con diversas perspectivas y técnicas matemáticas.

Redes Neuronales

Para inferir redes de interacciones sobre los conjuntos de datos, se emplearán paquetes R, como "limits" y "seqtime". Estas herramientas permitirán destacar una inferencia más detallada de las interacciones presentes en la microbiota, facilitando una comprensión más clara de las relaciones complejas entre los diferentes componentes del bioma utilizando redes neuronales.

Modelo Computacional

Se implementarán bibliotecas especializadas de Bioconductor en R y Python, junto con software libre, para evaluar si los datos satisfacen los supuestos de la teoría neutral. Esto permitirá hacer una validación rigurosa de los modelos computacionales utilizados y una mayor confianza en los resultados obtenidos.

F-50-01
PROTOCOLO DE INVESTIGACIÓN

Análisis Estadístico y Visualización de Datos

Finalmente, se realizarán diversos análisis estadísticos, tanto descriptivos como inferenciales, para comparar los diferentes modelos obtenidos. Estos análisis permitirán comprender cómo los datos se verán afectados por diferentes valores de muestra e identificar los factores que influyen en estos cambios. La visualización de los resultados facilitará una interpretación clara y accesible de los hallazgos, proporcionando una base sólida para futuras investigaciones y aplicaciones en el campo del análisis metagenómico de la microbiota.

La implementación de las SNF se centrará principalmente en el crecimiento, aprendizaje e interacción de las secuencias obtenidas de los diferentes microbiomas extraídos de las bases de datos utilizados, siendo que se analizará, comparará e identificarán las diferentes relaciones que se puedan generar en estas redes de similitud creadas.

Es por ello que con la evaluación de los datos resultantes durante el desarrollo de las SNF se recabarán los datos más influyentes de la generación de estos modelos y se interpretarán los datos obtenidos para verificar cuán contundentes son los cambios de un análisis aislado a uno integral.

F-50-01
 PROTOCOLO DE INVESTIGACIÓN

Cronograma

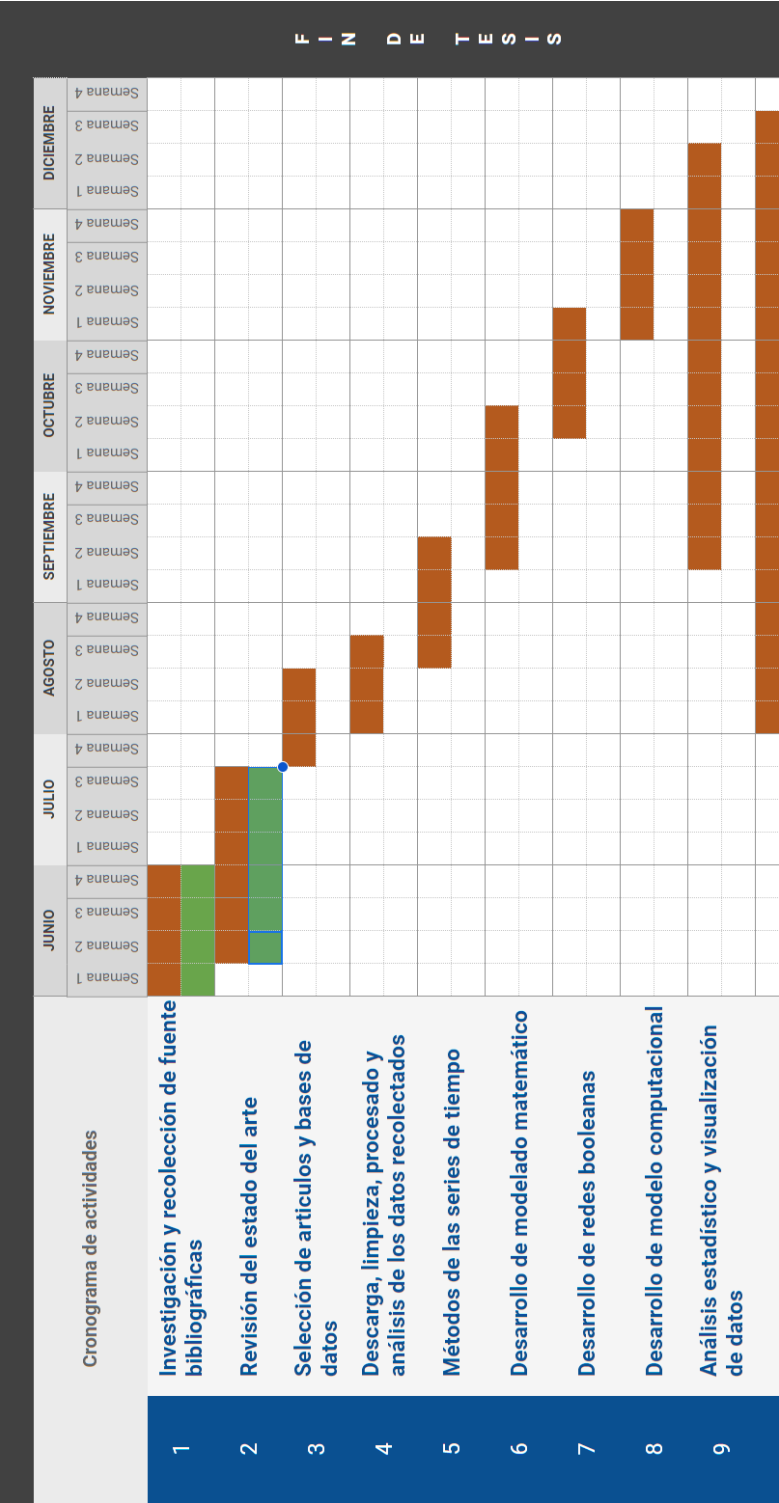


Fig 4. Cronograma de actividades

IX. RESULTADOS ESPERADOS, POSIBLES APLICACIONES Y USO DEL PROYECTO

- Desarrollar modelos matemáticos para comprender el comportamiento de la fusión de redes de similitud en la microbiota. Estos modelos permitirán realizar predicciones sobre cómo cambiará la microbiota en respuesta a diversos factores, como la dieta, los medicamentos o las enfermedades.
- Desarrollar redes de similitud con los datos de secuenciación masiva, para entender cómo interactúa un microbioma.
- Fusionar las redes de similitud generadas para realizar un análisis con respecto a las nuevas interacciones obtenidas
- Clasificar las redes de similitud para poder destacar y comprender cómo influyen las diversas colonias de un microbioma.
- La fusión de redes de similitud nos servirá para comprender cómo interactúan las diversas colonias en un microbioma.
- Utilizar en un ámbito biotecnológico los datos resultantes de las SNF, con el fin de generar patrones para desarrollar nuevas herramientas de diagnóstico.
- La información generada podría utilizarse para desarrollar nuevas estrategias para mantener o restaurar la salud de la microbiota.
- Evaluar y comparar si es que aplicar técnicas de SNF tiene mayor influencia en un análisis compuesto de dos o más microorganismos.
- El estudio propuesto proporcionará información de vanguardia sobre las aplicaciones de las estructuras supramoleculares (SNF). Con este conocimiento, podremos explorar nuevos usos para estas estructuras en diversos campos científicos y tecnológicos.

XI. REFERENCIAS BIBLIOGRÁFICAS

Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for

F-50-01
PROTOCOLO DE INVESTIGACIÓN

computational biology. *Molecular Systems Biology*, 12(7), 878.
<https://doi.org/10.15252/msb.20156651>

Cichy, R. M., & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in Cognitive Sciences*, 23(4), 305–317.
<https://doi.org/10.1016/j.tics.2019.01.009>

Comparison of Next-Generation Sequencing Systems—Liu—2012—BioMed Research International—Wiley Online Library. (s/f). Recuperado el 5 de agosto de 2024, de <https://onlinelibrary.wiley.com/doi/full/10.1155/2012/251364>

Jarada, T. N., Rokne, J. G., & Alhajj, R. (2021). SNF-NN: Computational method to predict drug-disease interactions using similarity network fusion and neural networks. *BMC Bioinformatics*, 22(1), 28.
<https://doi.org/10.1186/s12859-020-03950-3>

Jiang, X., & Hu, X. (2014). Microbiome data integration by robust similarity network fusion. *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 418–423. <https://doi.org/10.1109/BIBM.2014.6999194>

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science*, 1(Volume 1, 2015), 417–446.
<https://doi.org/10.1146/annurev-vision-082114-035447>

Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 26(2), 195–197. <https://doi.org/10.1038/nbt1386>

Mac Aogáin, M., Narayana, J. K., Tiew, P. Y., Ali, N. A. B. M., Yong, V. F. L., Jaggi,

F-50-01
PROTOCOLO DE INVESTIGACIÓN

- T. K., Lim, A. Y. H., Keir, H. R., Dicker, A. J., Thng, K. X., Tsang, A., Ivan, F. X., Poh, M. E., Oriano, M., Aliberti, S., Blasi, F., Low, T. B., Ong, T. H., Oliver, B., ... Chotirmall, S. H. (2021). Integrative microbiomics in bronchiectasis exacerbations. *Nature Medicine*, 27(4), 688–699. <https://doi.org/10.1038/s41591-021-01289-7>
- Mehrer, J., Spoerer, C. J., Kriegeskorte, N., & Kietzmann, T. C. (2020). Individual differences among deep neural network models. *Nature Communications*, 11(1), 5725. <https://doi.org/10.1038/s41467-020-19632-w>
- Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1–15. <https://doi.org/10.1016/j.dsp.2017.10.011>
- Moreno, J. J. M. (s/f). *Redes Neuronales Artificiales aplicadas al Análisis de Datos*.
- Richardson, L., Allen, B., Baldi, G., Beracochea, M., Bileschi, M. L., Burdett, T., Burgin, J., Caballero-Pérez, J., Cochrane, G., Colwell, L. J., Curtis, T., Escobar-Zepeda, A., Gurbich, T. A., Kale, V., Korobeynikov, A., Raj, S., Rogers, A. B., Sakharova, E., Sanchez, S., ... Finn, R. D. (2023). MGnify: The microbiome sequence data analysis resource in 2023. *Nucleic Acids Research*, 51(D1), D753–D759. <https://doi.org/10.1093/nar/gkac1080>
- Secuenciación de nueva generación (NGS) de ADN: Presente y futuro en la práctica clínica.* (s/f). Recuperado el 5 de agosto de 2024, de [https://revistas.javeriana.edu.co/files-articulos/UMED/61-2%20\(2020\)/231062391008/](https://revistas.javeriana.edu.co/files-articulos/UMED/61-2%20(2020)/231062391008/)
- Ursell, L. K., Metcalf, J. L., Parfrey, L. W., & Knight, R. (2012). Defining the human microbiome. *Nutrition Reviews*, 70(suppl_1), S38–S44.

F-50-01
PROTOCOLO DE INVESTIGACIÓN

<https://doi.org/10.1111/j.1753-4887.2012.00493.x>

Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., & Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods*, 11(3), 333–337.
<https://doi.org/10.1038/nmeth.2810>

Zhang, Y., Hu, X., & Jiang, X. (2017). Multi-View Clustering of Microbiome Samples by Robust Similarity Network Fusion and Spectral Clustering. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 14(2), 264–271. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.
<https://doi.org/10.1109/TCBB.2015.2474387>



Firma del alumno

Nivel de revisión: 02