Network-based Distribution Model for the Reviewer Assignment Problem

Carlos Casanova[1] and Lucas Forni[1]

Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Grupo de Investigación sobre Inteligencia Computacional y Optimización de Sistemas
{casanovac, fornil}@frcu.utn.edu.ar

Abstract The assignment of reviewers to manuscripts is crucial in the peer-review process, which serves as the primary mechanism for ensuring quality within academic fields. Making this assignment is no easy task, as reviewers typically have specialized backgrounds and limited time, which they must divide among many other responsibilities. In this regard, achieving a balanced assignment that meets specific constraints is a challenging task, one that has been addressed on numerous occasions through a computational approach. This paper proposes a comprehensive solution that covers the three phases of the assignment process: candidate search, calculation of the degree of match, and optimal assignment. An algorithm is presented that is capable of obtaining profiles from Google Scholar and retrieving the reviewers' publications, which are processed using an embedding model to calculate their semantic similarity, thereby assessing a reviewer's suitability for evaluating a manuscript. This suitability measure is used to construct a distribution network-based model that performs optimal assignment, taking into account an appropriate balance in reviewers' workloads and supporting reassignment. Finally, a reviewer evaluation framework is presented for its use in future instances. The results obtained demonstrate the effectiveness and scalability of the approach.

Keywords: Reviewer Assignment Problem, Distribution Networks, Semantic Similarity

Modelo basado en Redes de Distribución para el Problema de Asignación de Evaluadores

Resumen La asignación de revisores a manuscritos es crucial en el proceso de revisión por pares, principal mecanismo para asegurar la calidad dentro de ámbitos académicos. Realizar esta asignación no es una tarea sencilla, ya que los revisores suelen tener perfiles especializados y disponen de una cantidad escasa de tiempo, compartida entre otras muchas

ocupaciones. En este sentido, lograr una asignación equilibrada y que cumpla con restricciones específicas es una tarea desafiante, que ha sido abordada en numerosas ocasiones mediante un enfoque computacional. En este trabajo se propone una solución integral que abarca las tres fases del proceso de asignación: búsqueda de candidatos, cálculo del grado de coincidencia y asignación óptima. Se presenta un algoritmo capaz de obtener perfiles de Google Scholar y recuperar publicaciones de los evaluadores, que son procesadas mediante un modelo de embeddings para calcular su similitud semántica que pondera la adecuación de un revisor para evaluar un manuscrito. Esta medida de adecuación se utiliza en la construcción de un modelo basado en redes de distribución que realiza la asignación óptima, considerando un balance adecuado de la carga de trabajo de los evaluadores, y que contempla la reasignación. Finalmente, se presenta un esquema de evaluación de los revisores para su uso en ediciones futuras. Los resultados obtenidos demuestran la efectividad y escalabilidad del enfoque.

Keywords: Problema de Asignación de Evaluadores, Redes de Distribución , Similitud Semántica

1. Introducción

El proceso de revisión por pares es el principal mecanismo para garantizar la calidad de las publicaciones científicas. Durante este proceso, se envían manuscritos que deben ser evaluados por un panel de expertos —conocidos como evaluadores o revisores— quienes determinan la competencia, importancia y originalidad de la investigación, y recomiendan si el manuscrito debe ser aceptado o rechazado. Identificar y asignar revisores adecuados, cuyo perfil de investigación se alinee con la temática del manuscrito, es una de las tareas más desafiantes en las actividades científicas, abarcando contextos tan diversos como revistas académicas, congresos, workshops, proyectos de investigación, entre otros. La asignación adecuada y eficiente de revisores impacta directamente en la reputación de los investigadores y en la credibilidad de los eventos y publicaciones científicas. Este problema de recomendación y asignación de revisores a manuscritos es conocido en la literatura como *Reviewer Assignment Problem* (RAP), y puede concebirse como una variante del clásico *Generalized Assignment Problem* (GAP)(Cattrysse & van Wassenhove, 1992).

En el proceso típico de revisión por pares, los manuscritos son enviados y categorizados según su temática. Un *responsable* —ya sea un editor, coordinador de área o chair— selecciona evaluadores, usualmente en función de su experiencia en el área correspondiente, basándose en criterios como publicaciones, participación en proyectos de investigación, trayectoria académica, desempeño histórico y la opinión de otros expertos. Dependiendo del contexto, este proceso puede incorporar mecanismos de anonimato entre autores y revisores para garantizar

mayor objetividad, como el doble ciego (*double-blind*), lo que significa que tanto los evaluadores desconocen quiénes son los autores, como los autores ignoran quiénes evalúan su manuscrito. Según las evaluaciones recibidas, el *responsable* decide aceptar, rechazar o solicitar más revisiones para el manuscrito.

Una asignación deficiente de evaluadores tiene consecuencias concretas en múltiples dimensiones del proceso. En primer lugar, compromete la garantía de calidad: solo un evaluador con experiencia directa en el área del manuscrito puede identificar con precisión sus fortalezas, debilidades, originalidad y relevancia. Además, una distribución inequitativa de la carga de trabajo genera sobrecarga en ciertos evaluadores, lo que reduce la profundidad y calidad de sus revisiones al disponer de menos tiempo para cada manuscrito. Asimismo, los autores pierden la oportunidad de recibir retroalimentación constructiva y precisa que les permita mejorar sus investigaciones. Del mismo modo, y más allá de lo operativo, la percepción externa sobre el rigor del proceso de evaluación influye directamente en la credibilidad y el prestigio del evento, revista o institución financiadora, afectando la calidad de los manuscritos que se presentan en ediciones futuras. A esto se suma que la correcta asignación debe contemplar y prevenir posibles conflictos de interés, ya que mecanismos como el doble ciego solo funcionan efectivamente cuando los evaluadores no tienen vinculación previa con los autores o sus instituciones. Finalmente, en particular en el caso de eventos, las demoras en el proceso de asignación tienen un efecto cascada que compromete toda la organización posterior.

2. Definición del problema

Según un análisis del área (Wang et al., 2010), el proceso se divide formalmente en tres fases: búsqueda de candidatos evaluadores, cálculo del grado de coincidencia, y optimización de la asignación. En la práctica estas fases muchas veces presentan límites difusos, por ello algunos trabajos posteriores (Ribeiro et al., 2023), consolidan las dos primeras fases en una sola. A continuación se describen estas fases, siguiendo principalmente la estructura original de Wang.

El objetivo de la fase de búsqueda de candidatos evaluadores es construir una lista de perfiles de revisores cualificados. Información de los posibles revisores (nombre, publicaciones, citas, filiaciones e intereses de investigación) por lo general es accesible libremente por distintas fuentes. Entre las más frecuentemente utilizadas en la literatura se encuentran bases de datos bibliográficas como DBLP, ACM Digital Library e IEEE Xplore, así como motores de búsqueda académica como AMiner ("AMiner: Academic Search and Mining System", s.f.) y Google Scholar. Algunas instituciones utilizan una base de datos propia, donde los voluntarios se insertan manualmente, y otras, por ejemplo la National Science Foundation (NSF) de Estados Unidos, optan por restringir los candidatos a quienes hayan sometido propuestas previamente.

Por lo general, a los evaluadores que entran en la selección y están interesados, se les pide proporcionar o seleccionar palabras claves, las cuales sirven para determinar de forma rudimentaria su *expertise*. Cuando las palabras claves

no son suficientes porque se requiere mayor precisión a la hora de caracterizar al evaluador, se recurre a técnicas de minería de datos y recuperación de información para extraerlas automáticamente (Biswas & Hasan, 2007; Larsen & Chinatsu, 1999; Willett, 1998).

En una situación ideal para asegurar la cualificación de un revisor, Wang et al. (2010) identifican cinco criterios clave, desglosados en atributos específicos que permiten una valoración integral del experto. Estos cinco criterios son: palabras clave, publicaciones, proyectos, desempeño histórico en selección de proyectos como evaluador y la opinión de otros expertos

Por último, un aspecto relevante que por lo general se realiza en esta fase es la detección de conflictos de interés entre revisores y autores. Para ello se han explorado enfoques basados en redes semánticas y grafos de coautoría, que permiten identificar vínculos entre investigadores con mayor precisión que los métodos simplificados basados únicamente en filiación institucional (Wang et al., 2010). Esta detección es crítica para garantizar la integridad del proceso, en particular en esquemas de revisión doble ciego (Aleman-Meza et al., 2006; Kruk & Decker, 2005; Li et al., 2006; Papagelis et al., 2005).

2.1. Cálculo del grado de coincidencia

Dado un conjunto $M = \{1, \ldots, |M|\}$ de manuscritos y un conjunto $R = \{1, \ldots, |R|\}$ de revisores, la adecuación o correspondencia de un revisor para evaluar un manuscrito es modelada mediante una función de similitud $s : R \times M \to [0, 1]$, donde $s(r, m) = 1$ significa que el revisor r es completamente adecuado para evaluar el manuscrito m, mientras que $s(r, m) = 0$, que no es para nada adecuado. Este escalar mide el grado de similitud entre el contenido del manuscrito y la experiencia del revisor.

Ya que la selección de candidatos debe considerar su nivel profesional, el cual puede evaluarse a través de criterios como sus publicaciones, proyectos de investigación, desempeño histórico y la opinión de otros expertos, estos aspectos deberían ser tenidos en cuenta en la función de similitud. Una cuidadosa definición de esta función es clave, ya que en el RAP, una asignación óptima construida sobre un cálculo de coincidencia deficiente difícilmente producirá resultados satisfactorios.

Wang et al. (2010) identificaron dos grandes enfoques para esta fase: la calificación discreta y la recuperación de información. Ribeiro et al. (2023), en su revisión sistemática más reciente, amplían este panorama incorporando enfoques basados en aprendizaje automático y modelos de lenguaje que emergieron con posterioridad.

Dentro de estos enfoques, el más simple consiste en una correspondencia binaria: si el manuscrito i y el revisor j comparten al menos una palabra clave y no existe conflicto de interés, entonces $s_{ij} = 1$, de lo contrario $s_{ij} = 0$. Una alternativa más expresiva es la calificación discreta en una escala ordinal (típicamente de 0 a 5 o de 0 a 10). donde valores más altos indican mayor correspondencia.

También pueden encontrarse modelos de *"bidding"*, donde los revisores califican su preferencia sobre cada manuscrito a partir del resumen.

Se puede definir la recuperación de información como encontrar datos no estructurados dentro de una gran colección para satisfacer una necesidad de información particular (IBM, 2026). Wang et al. (Wang et al., 2010) clasifican estos métodos en tres grupos: basados en contenido, filtrado colaborativo e híbridos. Uno de los trabajos pioneros en la asignación automática de revisores es el de Dumais y Nielsen (Dumais & Nielsen, 1992) quienes aplicaron *Latent Semantic Indexing*. Este enfoque considera únicamente las características propias de un documento o elemento, por ejemplo, las palabras que contiene un texto. En cambio, el filtrado colaborativo, que tiene en cuenta las opiniones, elecciones o preferencias de otras personas con intereses similares respecto de esos elementos, es muy utilizado en sistemas de recomendación. En los últimos años, arquitecturas como las Redes Neuronales Convolucionales (CNN), las Redes Recurrentes (LSTM) y los modelos basados en Transformers han sido aplicadas con resultados superadores respecto de los métodos tradicionales.

2.2. Modelo de Asignación de Revisores

El problema de asignación de revisores consiste en asignar un conjunto finito de revisores a un conjunto finito de manuscritos enviados a un evento o publicación científica, de modo que cada manuscrito reciba una cantidad determinada a_i de evaluaciones y cada revisor no supere su capacidad máxima de revisión b_j, de modo de maximizar una medida global de calidad del proceso. Sean los conjuntos M de manuscritos y R de revisores como se definió en secciones anteriores. Sea además la variable binaria x_{ij} que toma el valor 1 si el manuscrito i es evaluado por el revisor j, y 0 en caso contrario. El problema de asignación de revisores se formula como el siguiente problema de programación entera (Wang et al., 2010):

$$\text{máx } RAP = \sum_{i \in M} \sum_{j \in R} s_{ij}\, x_{ij} \tag{1}$$

sujeto a

$$\sum_{j \in R} x_{ij} = a_i \qquad \forall i \in M \tag{2}$$

$$\sum_{i \in M} x_{ij} \leq b_j \qquad \forall j \in R \tag{3}$$

$$x_{ij} = 0 \qquad \forall (i,j) \in M \times R, s_{ij} < T \tag{4}$$

$$x_{ij} \in \{0,1\} \qquad \forall i \in M, \forall j \in R \tag{5}$$

La función objetivo (1) maximiza el grado de correspondencia total de la asignación. La restricción (2) garantiza que cada manuscrito posea exactamente a_i revisiones. La restricción (3) establece que cada revisor evalúe como máximo

b_j manuscritos. La restricción (4) impide que un revisor sea asignado a un manuscrito cuando su grado de correspondencia es inferior al umbral T, asegurando un nivel mínimo de competencia en todas las asignaciones.

Una revisión sobre distintos enfoques para el abordaje de este problema puede, a su vez, ser consultada en (Ribeiro et al., 2023; Wang et al., 2010), entre los cuales se encuentran variantes utilizando otras formulaciones de programación entera, redes de flujo y cobertura de conjuntos.

Por lo tanto, en el diseño de la asistencia para las decisiones del proceso de revisión por pares, deben resolverse estos desafíos:

- Cómo medir cuantitativamente el grado de potencial adecuación de un revisor en la evaluación de un manuscrito.
- Cómo asignar equilibradamente los evaluadores a los manuscritos, de modo de maximizar la calidad de las revisiones, no sobrecargar los evaluadores y lograr un mínimo de evaluaciones por manuscrito.
- Cómo medir cuantitativamente criterios de desempeño de los revisores en su rol, e incorporarlos en los mecanismos anteriores.

3. Enfoque Propuesto

La implementación práctica de un sistema de asignación de revisores debe adaptarse a las restricciones y particularidades del proceso de revisión concreto. En este sentido, en esta propuesta se incluye una primera fase de búsqueda de candidatos evaluadores:

1. **Perfilado de expertos:** se emplean técnicas de minería de datos y recuperación de información en Google Scholar para extraer y procesar los artículos más representativos de cada evaluador.
2. **Detección de Conflictos de Interés:** la detección de conflictos se centra en la filiación institucional. Si bien la literatura sugiere que métodos más complejos podrían ser empleados, la validación por pertenencia a la misma entidad constituye un primer filtro robusto.
3. **Criterios de Cualificación:** la cualificación de los revisores se fundamenta en el criterio experto del comité organizador. Se asume la idoneidad de una eventual base de datos histórica, centrando el esfuerzo tecnológico en la precisión de la asignación más que en la re-evaluación de la trayectoria del investigador.

3.1. Cálculo del grado de coincidencia

Se propone un enfoque multidimensional donde el grado de coincidencia es el resultado de integrar el *expertise* del revisor sobre el tema con su desempeño histórico. Esto permite que el modelo de optimización priorice no solo a los revisores más idóneos en términos de conocimiento, sino también a aquellos que tienden a no comprometer la integridad y agilidad del proceso. A continuación se detallan los cálculos de ambos componentes.

Cálculo de similitud basada en *embeddings* multilingües Una característica usual en procesos de revisión por pares es la presencia de manuscritos redactados en distintos idiomas, lo cual debe ser tenido en cuenta en el cálculo del grado de coincidencia. Luego de evaluar las alternativas descritas en los trabajos relacionados, se optó por un enfoque de *machine learning*, basado en un modelo de *embedding* , dado que los modelos de este tipo han demostrado resultados superiores en tareas de recuperación de información y matching semántico de documentos (Ribeiro et al., 2023).

Para calcular la similitud entre un manuscrito y un evaluador, se generan *embeddings*, es decir, representaciones vectoriales de palabras o (como en nuestro caso) textos que capturan relaciones semánticas, particularmente del título, resumen y palabras clave tanto de los manuscritos como de los artículos de cada evaluador. Formalmente, si $E(m)$ es el *embedding* del manuscrito m y $\{E(a_1), \ldots, E(a_n)\}$ es el conjunto de vectores que representan los artículos previos del revisor r, la similitud $s(r, m)$ se calcula como el valor máximo de similitud coseno entre el manuscrito y cada uno de los artículos del revisor:

$$s(r, m) = \text{máx} \left\{ \frac{E(m) \cdot E(a_i)}{\|E(m)\| \|E(a_i)\|} \right\}_{i \in \{1,\ldots,n\}} \tag{6}$$

La elección del máximo responde a que un evaluador resulta adecuado para revisar un manuscrito cuando al menos uno de sus trabajos previos aborda una temática estrechamente relacionada, aun cuando el resto de su producción corresponda a otras líneas de investigación. Alternativas como el promedio resultan menos apropiadas en este contexto, ya que penaliza a los revisores con perfiles amplios o multidisciplinarios, ya que diluye la cercanía de su trabajo más pertinente entre el resto de sus publicaciones.

El modelo seleccionado fue **BAAI/bge-m3** (Chen et al., 2024), un modelo de embedding multilingüe con soporte nativo para más de 100 idiomas, incluyendo español, inglés y portugués. Su elección se fundamentó en múltiples criterios técnicos y contextuales:

- **Capacidad multilingüe:** El soporte nativo para los tres idiomas del congreso garantiza una representación semántica uniforme independientemente del idioma del manuscrito (*Cross-Lingual Retrieval*).
- **Arquitectura:** BGE-M3 integra tres modalidades de recuperación. Si bien en la implementación actual se utiliza únicamente la modalidad densa, las modalidades restantes constituyen una vía de refinamiento disponible para trabajos futuros.
- **Capacidad para documentos largos:** El modelo admite entradas de hasta 8192 tokens, lo que permite procesar abstracts extendidos y secciones completas sin truncamiento.
- **Dimensionalidad del espacio vectorial:** La dimensión de 1024 proporciona un equilibrio entre expresividad semántica y eficiencia computacional.
- **Disponibilidad:** Al tratarse de un modelo de código abierto, se garantiza su integración directa con el *pipeline* de procesamiento desarrollado, evitando dependencias externas o APIs comerciales que podrían comprometer

la sostenibilidad del sistema, y garantizando así la posibilidad de realizar pruebas.

Puntaje Se propone un sistema de puntuación para recompensar e incentivar a aquellos evaluadores que dan devoluciones de calidad y desalentar comportamientos que comprometen el proceso de evaluación. Para determinar la calidad de una devolución, el sistema utiliza las puntuaciones otorgadas por dos fuentes complementarias: los Autores del artículo y los Responsables (chair, editores) ($P_{a,u}$ y $P_{c,u}$). Los autores evalúan la utilidad práctica de las correcciones para mejorar su trabajo, mientras que los responsables en la organización evalúan la profundidad técnica y pertinencia académica de las revisiones. Una 'buena devolución' será aquella valorada positivamente por ambas partes, caracterizada por retroalimentación constructiva, detallada y justificada. Esto se incorpora mediante la asignación de un puntaje de penalización para dos escenarios, uno para aquellos evaluadores que no aceptaron o declinaron la solicitud de revisión dentro del plazo establecido, y otro para aquellos que aceptaron la asignación, pero no dieron una devolución a tiempo. Esta decisión se fundamentó en la necesidad de mantener la integridad temporal del proceso de revisión. Este puntaje castigo reemplaza la fórmula $P_{i,t}$ detallada abajo, cuando esto sucede.

Además, se incorpora un factor de decaimiento temporal (θ) considerando que el desempeño reciente de un evaluador es más predictivo de su comportamiento futuro que su historial lejano. Un puntaje inicial (γ) se calibra para equilibrar la participación de evaluadores novatos versus experimentados. Valores cercanos a 1 favorecen la incorporación de nuevos evaluadores, promoviendo la renovación del cuerpo evaluador, mientras que valores menores priorizaron la experiencia de participaciones previas, esto ayuda a mitigar el problema de arranque en frío (Cold-Start). En futuras oportunidades se podría integrar la cualificación definida por Wang y otros que se comentó en la Sección 2.

Para el resto de los casos el cálculo del puntaje del evaluador se calcula de la siguiente manera:

El puntaje por período del evaluador en el periodo t, $P_{i,t}$ se calcula como un promedio de todas sus evaluaciones en ese periodo, dividido la cantidad de evaluaciones:

$$P_{i,t} = \begin{cases} 0, & \text{si } C_{i,t} = 0 \\ \dfrac{\displaystyle\sum_{u=1}^{C_{i,t}} \frac{P_{cu} + P_{au}}{2}}{C_{i,t}}, & \text{si } C_{i,t} > 0 \end{cases}$$

donde:
$C_{i,t}$: cantidad de evaluaciones del evaluador i en el período t.
$P_{a,u} \in [0, 1]$: puntuación dada por el autor sobre la devolución u al evaluador.
$P_{c,u} \in [0, 1]$: puntuación dada por el encargado sobre la devolución u.

El puntaje final ponderado $P_{f,i}$ del evaluador i se determina mediante:

$$P_{fi} = \begin{cases} \gamma, & \text{si } C_{f,i} = 0 \\ \displaystyle\sum_{t=1}^{5} \frac{\theta^{t-1}\delta_{i,a-t}P_{i,a-t}}{\displaystyle\sum_{t'=1}^{5}\theta^{t'-1}\delta_{i,a-t'}}, & \text{si } C_{f,i} > 0 \end{cases}$$

donde:
C_{fi}: cantidad de participación del evaluador i en los últimos 5 años.
$\delta_{i,t}$: 1 si el evaluador i participó en la edición t, 0 si no.
a: período correspondiente a la edición actual.

3.2. Modelo de asignación basado en redes de distribución

Con el fin de asegurar la escalabilidad del enfoque se construyó un modelo basado en una red de distribución, utilizando como base el problema del flujo capacitado de costo mínimo (FCCM). El "bien" que se distribuye por la red son las revisiones. Este problema posee un algoritmo de tiempo polinomial especializado para su resolución (símplex de redes) (Orlin, 1997), y ofrece la garantía de soluciones enteras cuando sus parámetros son enteros, de modo que es un candidato ideal para ser utilizado en este contexto. A continuación se describe la red que se deriva del problema RAP.

Sean los conjuntos M y R como se definieron en secciones anteriores. Adicionalmente sea el conjunto T, que clasifica los manuscritos en temáticas o tracks. Los revisores pueden no evaluar trabajos de todos los tracks, sino solo de un subconjunto de ellos. Llamemos RT al conjunto formado por todos los pares (r_i, t_k) tales que el evaluador r_i revisa trabajos en el track t_k.

Los nodos de la red resultan, en principio, de la unión de estos 3 conjuntos de nodos: $R \cup RT \cup M$. De estos conjuntos, los nodos $r_i \in R$ representan a los evaluadores y son nodos oferta, con flujo neto positivo igual a la cantidad máxima de evaluaciones $CMax_i$ que el revisor está dispuesto a realizar. Los nodos $m_j \in M$ representan los manuscritos y son nodos demanda, con flujo neto negativo igual a la cantidad de evaluaciones requerida por el proceso de revisión por pares. El resto de nodos son nodos trasbordo con flujo neto nulo.

Adicionalmente, para lograr un mejor balance de la carga de trabajo, se agregan dos conjuntos de nodos: Rm y Rr. Ambos conjuntos tienen la misma cardinalidad que R, y se utilizan para dividir el flujo saliente de cada r_i en dos: el arco (r_i, rm_i) tiene una capacidad de $Cmin$ revisiones y un costo de 0. La capacidad restante $CMax_i - Cmin$ viaja por el arco (r_i, rr_i), y tiene un costo unitario de 1 (más adelante se explica la elección de estos costos). El parámetro $Cmin$ puede configurarse de manera artesanal, aunque en algunos eventos es usual requerir un número mínimo de evaluaciones para que pueda ser expedido un certificado de revisor.

Luego, los nodos en $Rm \cup Rr$ son adyacentes hacia los nodos en $(r_i, t_k) \in RT$ que corresponden al mismo evaluador r_i. Estos arcos no poseen costo ni restricción de capacidad.

Finalmente, existen arcos desde los nodos $(r_i, t_k) \in RT$ hacia los manuscritos m_j que pertenecen a la temática t_k. En este punto pueden incorporarse restricciones de conflicto de interés (por ejemplo, que autores y evaluadores pertenezcan a la misma institución), eliminando los arcos cuya asignación no sea factible. Además, es aquí que se integra la similitud y el sistema de puntuación que se describió en la sección anterior. La capacidad de cada uno de estos arcos es de 1 (ya que el evaluador no puede evaluar múltiples veces el mismo manuscrito) y el costo se calcula mediante el término $\frac{1-s_{r_i m_j}}{P_{f,r_i}}$. $1 - s_{r_i m_j}$ convierte la similitud en una medida de distancia entre r_i y m_j, y P_{f,r_i} ayuda a disminuir este "costo" cuanto mayor es la puntuación del evaluador. Aquí cobra mayor sentido la elección de los costos anteriores, ya que puede verse que el peor costo que puede surgir de este término es 1, de modo que el algoritmo intentará priorizar la asignación de las primeras $Cmin$ revisiones de cada evaluador antes de pasar a las restantes. Esta característica constituye un aporte y un elemento diferenciador respecto de enfoques anteriores basados en redes de flujo (Wang et al., 2010).

Naturalmente, si la oferta y la demanda totales no están equilibradas se agrega un revisor ficticio o un artículo ficticio para balancear el problema, con flujo neto, costos y capacidades apropiados.

En la Figura 1 se muestra una red de ejemplo con 3 evaluadores y 4 manuscritos utilizando la convención de (Hillier & Lieberman, 2010). Puede apreciarse que los evaluadores tienen una capacidad de 4, 3 y 3 respectivamente, y los 4 manuscritos necesitan 2 revisiones. Dado que la oferta total (10) supera a la demanda total (8), el flujo neto no es nulo, por lo que se introduce un manuscrito ficticio, representado por el nodo m_f, que absorbe las revisiones sobrantes y balancea la red. A diferencia de los manuscritos reales, m_f recibe arcos directamente desde los nodos de los evaluadores, sin pasar por los nodos de capacidad ni de track, sin costo asociado. Su demanda es igual a la diferencia entre la oferta y la demanda totales.

4. Experimentación

Se realizaron tres experimentos como validación del enfoque propuesto. El primero se relaciona con la función de similitud. Para evaluar su rendimiento se han realizado pruebas con datos extraídos del **arXiv Dataset** el cual provee acceso abierto a artículos académicos de muchos ejes temáticos. El dataset contiene una entrada por cada artículo. Se tomaron los 5000 primeros artículos y se seleccionaron 50 artículos de este conjunto.

Se parafrasearon los resúmenes (abstracts) de estos 50 artículos, tanto en español como inglés (relevante para nuestro contexto dado que el dataset se encuentra en idioma inglés) y se creó una base de datos.

Luego, se generaron los embeddings de los resúmenes originales de 5000 artículos, así como de los 50 resúmenes parafraseados.

Posteriormente, se midió la efectividad del modelo utilizando la métrica propuesta, determinando en cuántos casos el artículo original correspondiente a un

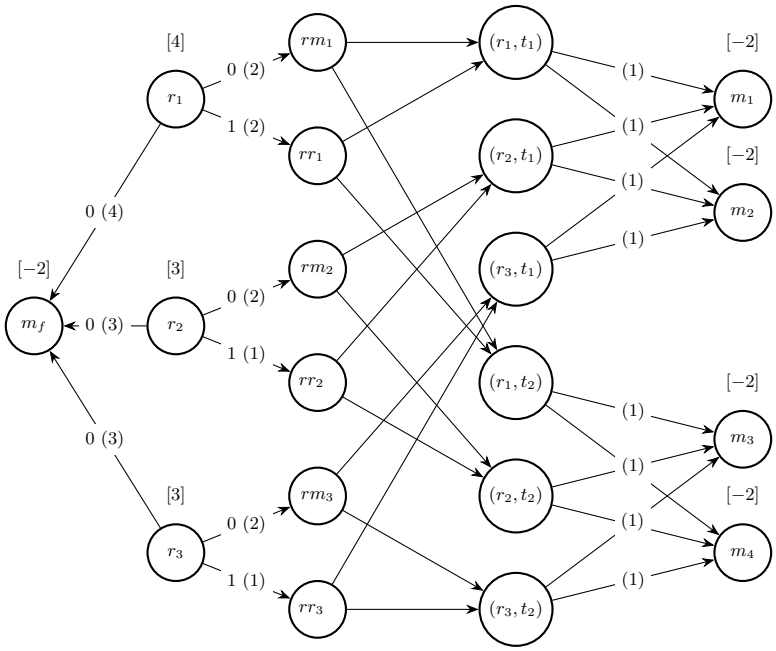

Figura 1. Red de ejemplo

resumen parafraseado se identificó como el más similar (primer puesto) y en cuántos se encontraba entre los cinco más similares (top 5). Los experimentos mostraron que en todos los casos se devuelve el *abstract* original como el más similar al parafraseado.

El alto puntaje de las similitudes, que alcanzan hasta 0.9, valida la efectividad del modelo BAAI/bge-m3 para capturar relaciones semánticas relevantes.

Respecto del modelo basado en redes de distribución, se realizaron dos experimentos más. Para el primero, se utilizó el mismo conjunto de datos que para el experimento anterior. Se dividieron 5000 embeddings a la mitad, formando dos grupos iguales, uno representó el conjunto de artículos presentados al congreso y el otro el conjunto de evaluadores
Cada "evaluador" tiene una oferta de 3 y cada "artículo" una demanda de 3.

A partir de esta configuración, los resultados obtenidos fueron los siguientes:

- Se asignaron todos los evaluadores a los artículos con similitudes mínimas de 0.40 . Las mejores similitudes se aproximan al 0.81.
- Matriz de similitud obtenida con forma: (2500, 2500). Tiempo: 76.0368 seg.
- Tiempo de ejecución del solver: 8.9045 segundos.
- Tiempo de ejecución del modelo: 50.8218 segundos.

- Tiempo total de la asignación 126.8586 segundos.

El siguiente histograma representa la distribución de los valores de similitud que se obtuvieron al aplicar el modelo a las 15000 asignaciones realizadas.

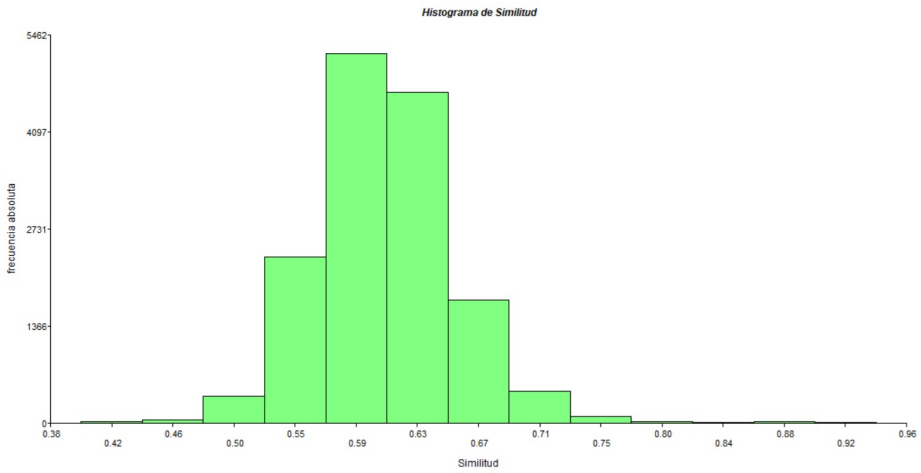

Figura 2. Histograma de distribución de valores de similitud.

El eje x representa los valores de la similitud que van desde 0.40 hasta 0.96. El eje y representa la frecuencia absoluta de las asignaciones en cada intervalo de similitud, con valores desde 0 hasta 5462.
La distribución tiene un deseable sesgo a la derecha (positiva), con un pico claro cerca de 0.65, lo que indica que el modelo tiende a asignar valores de similitud moderadamente altos en la mayoría de los casos, superando el baseline de una asignación aleatoria teórica con media en 0.5.

Respecto del tiempo utilizado (126.86 segundos), es muy bajo considerando la gran cantidad de evaluaciones, demostrando que el sistema es escalable.

Finalmente, el último experimento se realizó sobre la edición del Congreso Nacional de Ingeniería Informática y Sistemas de Información CONAIISI 2025 ("RIISIC Red de Ingeniería en Informática e Ingeniería en Sistemas de Información del CONFEDI", s.f.). Esta edición recibió 300 artículos que fueron evaluados por 409 evaluadores distribuidos en 8 ejes temáticos. Utilizando los datos de esta instancia se compararon los tiempos insumidos por el modelo propuesto y el modelo (1)-(5). Se realizaron 10 ejecuciones utilizando la biblioteca de Google OR Tools de Python para el primero, y COIN-OR CBC para el segundo. En promedio el modelo de red resultó 13 veces más rápido (2.7 contra 36.5 segundos). Las asignaciones son prácticamente idénticas en ambos modelos, con diferencias menores debido a errores de precisión de punto flotante. Según la opinión de los chairs que usaron el modelo, las asignaciones fueron muy adecuadas según el ex-

pertise de los evaluadores. En el modelo propuesto, el promedio de asignaciones fue de 2.2 artículos por evaluador, con un máximo de 5 revisiones asignadas a un mismo revisor.

5. Conclusiones

Se ha presentado el problema de asignación de revisores como el principal mecanismo para asegurar la calidad dentro de ámbitos académicos. Se ha propuesto una solución integral que abarca las tres fases del proceso de asignación: búsqueda de candidatos, cálculo del grado de coincidencia y asignación óptima. Se ha presentado un algoritmo capaz de obtener perfiles de Google Scholar y recuperar publicaciones de los evaluadores, procesándolas mediante un modelo de *embeddings* que sirve para calcular su similitud semántica y ponderar la adecuación de un revisor para evaluar un manuscrito. Esta medida de adecuación, junto a un esquema de puntuación de evaluadores se utiliza en la construcción de un modelo basado en redes de distribución que realiza la asignación óptima, considerando un balance equitativo en la carga de trabajo de los evaluadores, y que puede usarse también para la reasignación o segunda vuelta. Los resultados obtenidos han demostrado la efectividad tanto de la función de similitud propuesta como del modelo de optimización, así como la eficiencia y escalabilidad del enfoque.

Dentro de las líneas de trabajos futuros, se espera poder probar el esquema de puntuación de evaluadores en sucesivas ediciones. Además, la incorporación en una herramienta utilizable que se vincule con sistemas de gestión de conferencias sería de mucha utilidad para la realización de más validaciones empíricas. Finalmente, la exploración de otros mecanismos de detección de conflictos de interés, así como los aspectos éticos en el uso de Inteligencia Artificial Generativa son aspectos que deben seguir siendo analizados e incorporados al enfoque propuesto.

Referencias

Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A. P., Arpinar, I. B., Joshi, A., & Finin, T. (2006). Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. *Proceedings of the 15th International World Wide Web Conference*, 407-416.

AMiner: Academic Search and Mining System [Accessed: 2026-04-11]. (s.f.).

Biswas, H. K., & Hasan, M. M. (2007). Using Publications and Domain Knowledge to Build Research Profiles: An Application in Automatic Reviewer Assignment. *Proceedings of the 2007 International Conference on Information and Communication Technology*.

Cattrysse, D. G., & van Wassenhove, L. N. (1992). A Survey of Algorithms for the Generalized Assignment Problem. *European Journal of Operational Research*, *60*(3), 260-272. https://doi.org/10.1016/0377-2217(92)90077-M

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024, agosto). M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation. En L.-W. Ku, A. Martins & V. Srikumar (Eds.), *Findings of the Association for Computational Linguistics: ACL 2024* (pp. 2318-2335). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.137

Dumais, S. T., & Nielsen, J. (1992). Automating the assignment of submitted manuscripts to reviewers. *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, 233-244.

Hillier, F., & Lieberman, G. (2010). *Introduccion a la investigacion de operaciones*. McGraw-Hill Interamericana, S.A.

IBM. (2026). What is Information Retrieval? [Accedido: 03-04-2026]. https://www.ibm.com/mx-es/think/topics/information-retrieval

Kruk, S. R., & Decker, S. (2005). Semantic Social Collaborative Filtering with FOAFRealm. *First Semantic Desktop Workshop*.

Larsen, B., & Chinatsu, A. (1999). Fast and Effective Text Mining Using Linear-Time Document Clustering. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 16-22.

Li, J., Boley, H., Bhavsar, V. C., & Mei, J. (2006). Expert Finding for eCollaboration Using FOAF with RuleML Rules. *Conference on eTechnologies*.

Orlin, J. B. (1997). A polynomial time primal network simplex algorithm for minimum cost flows. *Mathematical Programming*, *78*(2), 109-129.

Papagelis, M., Plexousakis, D., & Nikolaou, P. N. (2005). CONFIOUS: Managing the Electronic Submission and Reviewing Process of Scientific Conferences. *Proceedings of the 6th International Conference on Web Information Systems Engineering*.

Ribeiro, A. C., Sizo, A., & Reis, L. P. (2023). Investigating the reviewer assignment problem: A systematic literature review. *Journal of Information Science*, *52*(1), 39-59. https://doi.org/10.1177/01655515231176668

RIISIC Red de Ingeniería en Informática e Ingeniería en Sistemas de Información del CONFEDI [Accessed: 2024-03-13]. (s.f.).

Wang, F., Shi, N., & Chen, B. (2010). A Comprehensive Survey of the Reviewer Assignment Problem. *International Journal of Information Technology & Decision Making*, *9*(4), 645-668. https://doi.org/10.1142/S0219622010003993

Willett, P. (1998). Recent Trends in Hierarchic Document Clustering: A Critical Review. *Information Processing and Management*, *24*, 577-597.