

# Un enfoque inteligente para la selección de grupos de expertos mediante redes sociales

Dr. Eduardo Zamudio (*autor*)

Prof. Dra. Analía A. Amandi (*director*)

Prof. Dr. Luis S. Berdún (*co-director*)

*Tesis doctoral defendida el 10 de marzo de 2017*

*Doctorado en Ciencias de la Computación, FCE, UNICEN*

## Resumen

La selección de grupos de expertos generalmente considera la evaluación de los criterios de selección que deben cumplir los candidatos y los grupos en sí mismos. Los criterios de selección representan un aspecto clave de los procesos de selección de expertos, ya que son estos criterios los que establecen qué candidato es un experto, o cómo se debe conformar un grupo de expertos.

En esta tesis se propone un enfoque inteligente e integral para el problema de la selección de grupos de expertos compuesto por dos partes. Primero, se propone un método para la identificación y evaluación de criterios de selección de candidatos elegibles a ocupar posiciones de expertos, utilizando descripciones de los candidatos como fuente de información. Segundo, se propone un método para la selección óptima de conformaciones de grupos de expertos, utilizando información relacional de los candidatos.

Esta tesis introduce importantes contribuciones en el área de selección de expertos, incluyendo la aplicación de Aprendizaje Automático (ML) en la identificación de evidencia de experiencia; la aplicación de técnicas de Procesamiento de Lenguaje Natural (NLP) para determinar la similitud de criterios de selección; una nueva métrica de Análisis de Redes Sociales (SNA) para determinar la independencia de grupos; la aplicación de una estrategia multicriterio para la evaluación de grupos en redes sociales; y la implementación de un algoritmo evolutivo para la selección óptima de grupos de expertos.

Las evaluaciones experimentales indican que la elegibilidad de un conjunto de candidatos puede

ser determinada a partir del nivel de correspondencia semántica entre las evaluaciones de los candidatos y los criterios de selección de referencia. Asimismo, los resultados indican que es posible recomendar grupos de expertos con mejor desempeño al compararlos con los comités actuales, a partir del uso de información relacional.

## 1. Introducción

Un grupo de expertos es un conjunto de personas con reconocido dominio de uno o varios temas. Los grupos de expertos se constituyen por varios motivos, como el tratamiento de asuntos específicos, definición de políticas, asignación de recursos, y asesoramiento, entre otros.

Por lo general, los grupos de expertos son seleccionados mediante algún proceso de selección de grupos, en el cual se evalúan tanto los requisitos que deben cumplir los candidatos, como así también los requisitos de los grupos en sí mismos. Estos requisitos se definen a partir de los criterios de selección que, cuando están disponibles, se pueden encontrar definidos en los lineamientos generales de los mismos procesos de selección.

La falta de acceso a los criterios de selección, o la subjetividad con la que éstos son definidos, limitan la posibilidad de realizar análisis objetivos de dichos procesos. Esto implica un desafío, principalmente en relación con las fuentes de información disponibles para realizar cualquier análisis sobre procesos de selección.

Por una parte, en varios procesos de selección de expertos, la única fuente de información disponible se constituye a partir de las evaluaciones de los candidatos. Estas evaluaciones son documentos de texto en los que los evaluadores argumentan la adecuación del candidato a los criterios de selección.

Por otra parte, la selección de grupos de expertos difícilmente se acompañe de un documento que describa la evaluación de los criterios de selección del grupo. En estos casos, la primera fuente de información disponible se constituye a partir del listado de miembros del grupo seleccionado.

En esta tesis se propone un enfoque integral para el problema de la selección de grupos de expertos, el cual consiste de dos aspectos. Por una parte, se propone determinar la elegibilidad de candidatos a ocupar posiciones de experto, mediante un método para la identificación y evaluación de los

criterios de selección utilizados. Por otra parte, se propone un método para la selección óptima de conformaciones de grupos de expertos, basada en las relaciones sociales de candidatos elegibles.

La hipótesis general del trabajo sostiene que es posible desarrollar un método objetivo de selección de grupos de expertos utilizando técnicas del área de la inteligencia artificial aplicada sobre un conjunto de candidatos para la conformación de dichos grupos.

Esta tesis introduce importantes contribuciones en el área de selección de expertos, incluyendo:

- Un enfoque de aprendizaje supervisado para la extracción de descripciones de candidatos, como una nueva fuente de evidencia de experiencia para procesos de selección de expertos.
- Una nueva aplicación de técnicas de Procesamiento de Lenguaje Natural (NLP) para determinar la similitud del contenido semántico de criterios de selección de expertos.
- Una nueva métrica de Análisis de Redes Sociales (SNA) para determinar la independencia de miembros de grupos [17, 18, 19].
- Un método de selección óptima de grupos a partir de la integración de métricas de selección de grupos basadas en redes sociales y la implementación de un algoritmo genético.

Este documento se organiza de la siguiente manera. En la sección 2 se presenta una visión general del método propuesto para la selección inteligente de grupos de expertos, separados en las etapas de (1) identificación de candidatos elegibles para la conformación de grupos de expertos, y (2) la representación del problema de selección de grupos mediante redes sociales, junto con el método propuesto para la optimización de conformaciones posibles de grupos de expertos. En la sección 3 se presentan las evaluaciones experimentales. En la sección 4 se presentan los antecedentes del trabajo. Finalmente, en la sección 5 se presentan las conclusiones.

## 2. Visión general de la propuesta

El enfoque inteligente e integral propuesto para el problema de la selección de expertos consiste de dos partes. Por una parte, con objeto de determinar la elegibilidad de candidatos a ocupar po-

siciones de experto, se propone un método para la identificación y evaluación de criterios de selección aplicados en la evaluación de dichos candidatos. Por otra parte, con objeto de seleccionar grupos de expertos, se propone un método para la selección óptima de conformaciones, basada en las relaciones sociales de sus miembros.

La figura 1 presenta el enfoque general de la propuesta con sus procesos principales. El proceso de elegibilidad de candidatos toma información de los candidatos a conformar los grupos de expertos, junto con información de los criterios de selección. La salida de este proceso es un subconjunto de candidatos elegibles, los cuales cumplen con los criterios del proceso de selección de expertos. Luego, el proceso de selección de grupos de expertos toma la información del subconjunto de candidatos elegibles, y de las relaciones sociales de éstos, con objeto de determinar las mejores conformaciones posibles.

### 2.1. Elegibilidad de candidatos mediante criterios de selección de expertos

Las evaluaciones de los candidatos, cuando son expresadas en lenguaje natural, suelen contener *descripciones*, las cuales se definen en este trabajo como las características destacadas de los candidatos desde el punto de vista de la persona responsable de una evaluación. Una descripción en un contexto de selección de expertos podría ser: “The candidate have demonstrated skills for conflict resolution”.

Considerando que los candidatos de un proceso de selección de expertos deben ser evaluados en base a un conjunto de criterios de selección, es posible asumir que los criterios de selección (explícitos o no) se encuentran implícitos en las descripciones de los candidatos contenidas en sus evaluaciones. En forma complementaria, puede ocurrir que los criterios de selección se encuentren explícitamente, por ejemplo, en guías, estatutos, lineamientos, etc. Se establece entonces una distinción entre los criterios de selección *explícitos* e *implícitos*.

Por lo tanto, resulta necesario un mecanismo que permita determinar en qué medida se asemejan dos conjuntos de criterios de selección, uno correspondiente a criterios de referencia (como los explícitos) y otro correspondiente a los criterios de correspondencia (como los implícitos).

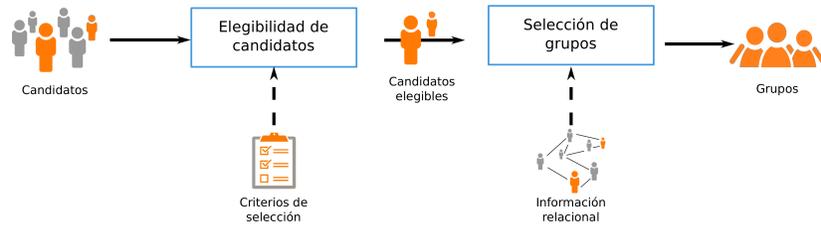


Figura 1: Enfoque general propuesto de selección de grupos de expertos.

Esta situación conduce a la siguiente hipótesis:

Los criterios de selección explícitos e implícitos (de un mismo proceso de selección) presentan un contenido semántico similar entre sí.

De esta manera, sería posible analizar la similitud semántica entre criterios (explícitos e implícitos), y así determinar en qué medida se aplican estos criterios en la evaluación de candidatos. Así, la *elegibilidad* de un candidato estaría dada cuando los criterios implícitos en sus evaluaciones alcanzan un nivel de similitud con el conjunto de criterios de selección de referencia.

La figura 2 presenta un esquema general del enfoque propuesto, cuyo proceso principal es el análisis de similitud de criterios de selección. Este análisis de similitud tiene como datos de entrada al conjunto de criterios explícitos, extraídos en forma manual a partir de documentos, tales como guías, estatutos, lineamientos, y otros documentos que definan los requerimientos de las posiciones de expertos. El análisis de similitud también toma como datos de entrada al conjunto de contenidos con criterios implícitos (descripciones), extraídos en forma automática a partir de las evaluaciones de los candidatos. El análisis de similitud extrae un conjunto de conceptos semánticos de cada conjunto de criterios (explícitos e implícitos), y calcula la frecuencia de los mismos. Finalmente, se determina el nivel de correlación de frecuencias de los conceptos semánticos correspondientes a cada grupo de criterios. El resultado final es un puntaje de similitud, asociado al nivel de correlación entre criterios explícitos e implícitos, o entre criterios implícitos.

### 2.1.1. Similitud semántica entre criterios de selección

El enfoque de elegibilidad de candidatos implica determinar la similitud entre conjuntos de criterios

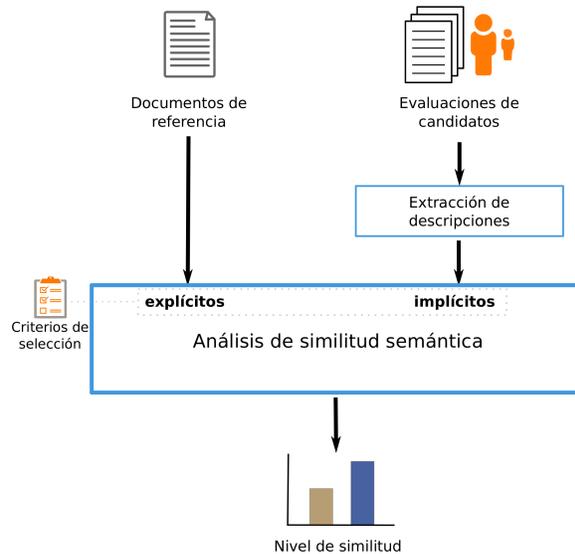


Figura 2: Enfoque general propuesto para el análisis de criterios de selección.

de selección mediante un análisis de las propiedades semánticas de éstos. El proceso propuesto incluye las siguientes etapas:

1. **Extraer criterios de selección contenidos en descripciones** de candidatos  
Mediante el entrenamiento de un modelo de aprendizaje supervisado de clasificación binaria, cuyo objetivo es etiquetar sentencias en dos clases distintas: “descripciones” y “no descripciones”.
2. **Identificar las propiedades semánticas** de los criterios de selección  
Mediante la aplicación de Etiquetado de Roles Semánticos (SRL)[4] a un conjunto de descripciones de candidatos, considerados contenedores de criterios de selección.
3. **Evaluar la relevancia de las propiedades**

**semánticas** de los criterios de selección

Mediante la determinación de relevancia de cada una de propiedades semánticas expresadas como la frecuencia de ocurrencia de los Roles Semánticos etiquetados en las descripciones.

#### 4. **Evaluar la similitud entre criterios de selección**

Mediante el uso de un coeficiente de correlación de rankings sobre las frecuencias de las propiedades semánticas de los criterios de selección.

## 2.2. Selección de grupos de expertos mediante redes sociales

A diferencia de la selección de candidatos individuales, la selección de grupos no suele estar acompañada por evaluaciones de los grupos, que permitan analizar la aplicación de los criterios de selección. Sin embargo, una característica importante de la selección de grupos, es que los criterios de selección suelen considerar las relaciones entre los miembros del grupo. Por ejemplo, un grupo interdisciplinario con miembros “representativos” relacionaría los miembros con sus disciplinas, áreas geográficas, etc. La versatilidad de la información relacional representa una alternativa importante ante la falta de fuentes de información para la evaluación de criterios de selección de grupos de expertos.

Adicionalmente al problema de representación, la selección de grupos de expertos implica un elevado costo de procesamiento por la naturaleza combinatoria del problema, determinado por el coeficiente binomial  ${}_nC_r$ , con complejidad  $O(n!)$ .

En esta segunda parte del enfoque propuesto, se presenta un método para la selección de grupos de expertos, a partir del análisis de las redes sociales de un conjunto de candidatos elegibles. El método propuesto, utiliza un conjunto de métricas de SNA que permiten evaluar la conformidad de grupos a algunos criterios de selección de grupos. Asimismo, el método propone una estrategia de optimización para la búsqueda de conformaciones óptimas, mediante la implementación de un algoritmo genético.

### 2.2.1. Criterios de selección de grupos

El enfoque de selección de grupos de expertos presenta una estrategia para evaluar criterios de

selección de grupos en base a las relaciones sociales de sus miembros. Esta estrategia implica los siguientes aspectos principales:

#### 1. **Representar información relacional** de los candidatos a conformar grupos de expertos

Mediante una estrategia de unificación de tipos de relaciones, para permitir la aplicación de varias de las métricas actuales de SNA.

#### 2. **Determinar la independencia** entre los miembros de un grupo

A partir de una métrica de distancia social de dentro del grupo, como uno de los criterios más importantes para la selección de grupos de expertos.

La función para la métrica de grupos con miembros independientes queda definida como:

$$f = \frac{\left[ \left( \sum_{i,j=0}^k d_{ij} \right) / k \right] + m}{2 * D} \quad (1)$$

donde  $d$  es la distancia geodésica entre dos miembros  $i$  y  $j$  del grupo, para  $\forall i, j / i \neq j$  y  $i, j \in N$ , dado un conjunto de nodos  $N$ , donde  $k$  es el número de distancias entre miembros del grupo,  $D$  es el diámetro de la red, y  $m$  es la mínima distancia geodésica entre cada par de miembros del grupo.

#### 3. **Integrar criterios** de selección de grupos basados en redes sociales

Mediante un enfoque de Toma de Decisiones con Múltiples Atributos (MADM).

El enfoque presenta la aplicación de una suma ponderada, junto con la identificación de pesos asignados a cada criterio mediante el método de Entropía [16].

La figura 3 presenta un esquema general de la propuesta de integración de criterios de selección de grupos.

#### 4. **Optimizar el proceso de búsqueda**

Mediante la implementación de un Algoritmo Genético (GA).

El GA propuesto tiene por objetivo la búsqueda de conformaciones óptimas de grupos, mediante la evaluación de una función basada en la integración de métricas de SNA.

La representación de las soluciones utilizada para procesar el algoritmo genético es un vector de enteros que contiene el índice de nodos

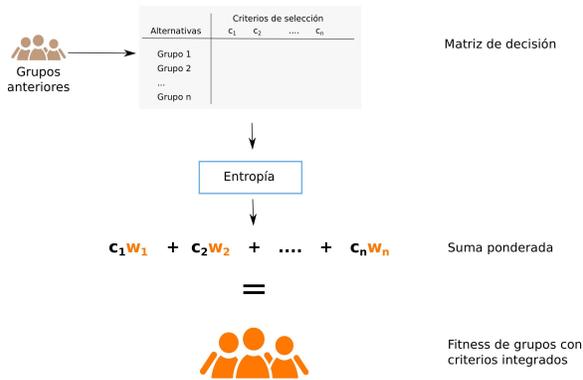


Figura 3: Esquema general de integración de criterios de selección con suma ponderada.

de la red social conformada por los candidatos, y un vector binario que contiene la información de inclusión en el grupo por parte de cada candidato.

### 3. Experimentos

Esta sección comienza con una descripción de los datos utilizados para la evaluación experimental del enfoque de elegibilidad de candidatos. Luego se presentan las actividades de clasificación de descripciones, y la conformidad de las evaluaciones respecto de los criterios de selección. Posteriormente, se presentan los datos utilizados para la evaluación experimental del enfoque de selección de grupos de expertos. Finalmente, se presentan los resultados de selección de grupos de expertos a partir de la métrica de independencia propuesta, y de la integración de criterios de selección basados en métricas de Análisis de Redes Sociales.

#### 3.1. Datos para la evaluación de elegibilidad

Se desarrolló un nuevo conjunto de datos a partir de 65 documentos de texto correspondientes a audiencias de nominación del Committee of Commerce, Science, and Transportation del Congreso de los Estados Unidos. Estas nominaciones tuvieron lugar desde marzo de 2000 hasta julio de 2013<sup>1</sup>. El conjunto de datos se constituyó de 7738 sentencias extraídas de las declaraciones de senadores, luego etiquetadas manualmente en clases

<sup>1</sup><http://www.gpo.gov/fdsysinfo/aboutfdsys.htm> (en línea en octubre de 2016)

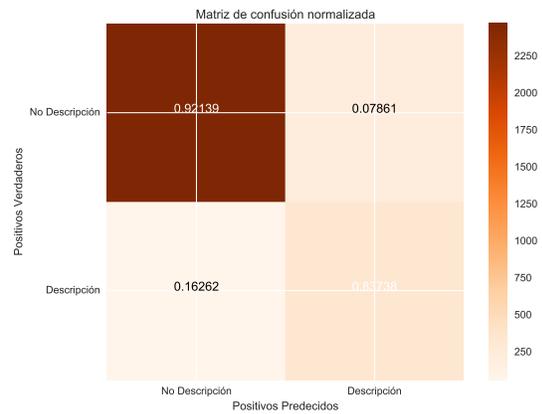


Figura 4: Matriz de confusión normalizada.

positivas (descripciones) y negativas (no descripciones).

Finalmente, el conjunto de datos presentó una distribución de clases asimétrica ( $\approx 13\%$  positivas,  $\approx 87\%$  negativas).

#### 3.2. Clasificación de descripciones

El proceso de clasificación implicó la extracción de features (n-grams,  $n=\{2,3,4\}$ ) dando 415176 n-gramas. y la selección de features mediante Chi cuadrado  $\chi^2$  (threshold=1000 features).

Se entrenó un modelo de aprendizaje supervisado mediante Support Vector Machines (SVM-RBF), y se optimizaron sus parámetros mediante Grid Search Cross Validation (GS-CV).

El tiempo total requerido para el entrenamiento del modelo demandó 270 segundos (4,5 minutos) aproximadamente en una computadora de escritorio estandar, compuesta por un procesador Intel I3 530 de 2.93GHz.

La figura 4 presenta la matriz de confusión normalizada.

El entrenamiento del modelo de clasificación propuesto alcanzó un desempeño de  $f1\ score = 0,92$  y  $ROC\ AUC = 0,88$  con los parámetros  $C = 2$  y  $gamma = 2e^3$ .

La figura 5 presenta las curvas de aprendizaje (learning curves), las cuales presentan una reducción en el error al incrementar el tamaño del conjunto de datos de entrenamiento.

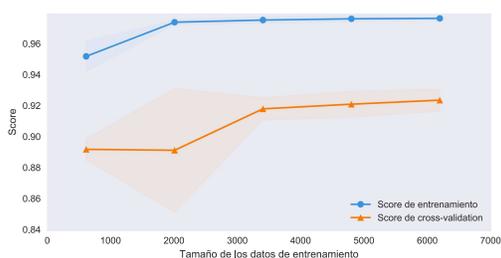


Figura 5: Curvas de aprendizaje del modelo entrenado para la clasificación de descripciones.

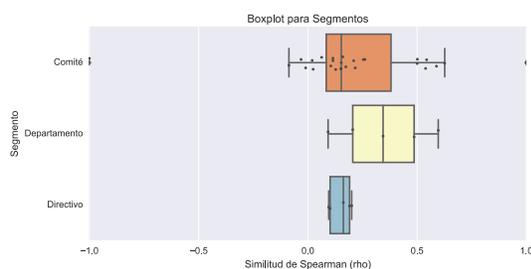


Figura 6: Diagramas de caja para los segmentos Comité, Departamento, y Directivo.

### 3.3. Conformidad de evaluaciones a los criterios de selección

Para determinar la conformidad de los criterios de selección se utilizaron dos conjuntos de datos, el de referencia asociado a los criterios explícitos, y el de correspondencia asociado a los criterios implícitos.

El conjunto de datos de criterios explícitos, se generó a partir de la información relacionada con las posiciones a cubrir por los candidatos. Mientras que el conjunto de datos de criterios implícitos, se generó a partir de las descripciones extraídas de las audiencias de nominación (ver cuadro 1).

Se evaluó la similitud entre los criterios implícitos en referencia a los criterios explícitos de tres segmentos: Comité, Departamento, y Directivo.

La figura 6 presenta los diagramas de caja de los coeficientes similitud para los segmentos Comité, Departamento, y Directivo. En la figura se observa un mejor desempeño en la similitud de los criterios de selección del segmento Departamento. La desviación estándar de los segmentos Comité ( $\approx 0,35$ ) y Directivo ( $\approx 0,05$ ) presentan una amplia diferencia. Sin embargo, debe considerarse la diferencia en la cantidad de nominaciones correspondientes a cada segmento (27 en Comité y 5 en Directivo).

### 3.4. Datos para la selección de grupos de expertos

Se implementó una red social de investigadores a partir de información pública extraída del sitio web de CONICET. Se aplicaron estrategias de unificación y desambiguación, ya que la mayor parte de la información disponible en el sitio web es ingresada por los mismos investigadores.

La elegibilidad de los candidatos, para la evaluación del enfoque propuesto de selección de grupos, se determinó a partir un listado de candidatos elegibles que cumplían con los requisitos para conformar las comisiones. Sin embargo, se mantuvo la información de las relaciones con otros investigadores no elegibles, ya que todos ellos conforman la estructura de la red social.

Finalmente, la red social para el caso de estudio quedó conformada por un conjunto de 1293 nodos (investigadores) y 4322 enlaces (publicaciones y lugares de trabajo en común), conformando 74 componentes. Para la evaluación, se tomó el componente de mayor tamaño incluyó 1058 ( $\approx 82\%$ ) de los investigadores (75 de ellos calificados para conformar comisiones), y 3878 ( $\approx 90\%$ ) de relaciones.

### 3.5. Selección de grupos con miembros independientes

La primera evaluación experimental se realizó con la métrica de independencia de grupos como función de fitness del GA. Con objeto de comparar el fitness de las comisiones actuales con el fitness de las comisiones generadas por el algoritmo genético, se generaron comisiones de 3, 4, y 5 miembros.

La figura 7 presenta el fitness de las comisiones generadas y las actuales de acuerdo a las 16 configuraciones. En todas las ejecuciones se obtuvieron comisiones generadas con fitness superior al de las comisiones actuales.

La figura 8 muestra la red social utilizada en la experimentación, en la cual se identifica que los miembros de una comisión generada se encuentran más alejados entre sí que los miembros de la comisión actual. Esta representación muestra una mejora en el balance de las distancias entre los

Segmento	Fuentes para <b>criterios explícitos</b>	Fuentes para <b>criterios implícitos</b>
<b>Comité</b>	definiciones de criterios de selección del Federal Advisory Committee Act (FACA) <sup>2</sup>	nominaciones a posición en mesas directivas o grupos consultivos
<b>Departamento</b>	definiciones de visión y misión de las dependencias correspondientes	nominaciones a dirección de sectores dentro de la misma dependencia
<b>Directivo</b>	definiciones de responsabilidades de cargos	nominaciones a posiciones dentro de una junta directiva

Cuadro 1: Orígenes de datos para la identificación de los criterios de selección explícitos (referencia) e implícitos (correspondencia) utilizados en la evaluación de similitud de criterios de selección aplicados.

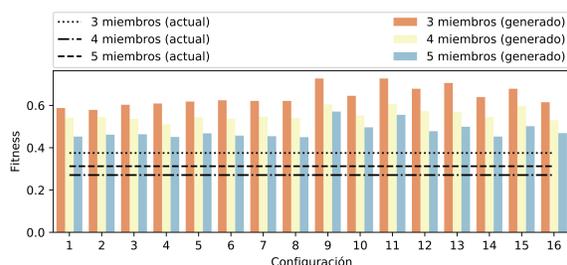


Figura 7: Fitness promedio para comisiones generadas de 3, 4, y 5 miembros.

miembros de las comisiones generadas al comparlas con las comisiones actuales.

### 3.6. Selección de grupos con múltiples criterios

La segunda evaluación experimental se realizó mediante una suma ponderada de la métrica propuesta de Independencia de grupos y KPP-Pos[3]. Para ello, se determinaron los pesos correspondientes a cada métrica, mediante la aplicación del método de Entropía a los fitness de las comisiones actuales. Los pesos obtenidos para las métricas evaluadas indican una mayor relevancia dada a la selección de grupos con miembros independientes, dada por  $w_{Independencia} = 0,567966$ , que la relevancia dada a la selección de grupos que maximicen la difusión en la red, dada por  $w_{KPP-Pos} = 0,432034$ .

Al igual que en la primera evaluación, se comparó el fitness de las comisiones actuales con el fitness de las comisiones generadas por el algoritmo genético, por lo que se generaron comisiones de 3, 4, y 5 miembros.

Para la configuración del GA, se utilizó la configuración que arrojó mejores resultados en la primera evaluación. Esta configuración utilizó los me-



Figure 8: Comisión de 5 miembros. Actual (círculos grandes) y generada (estrellas).

canismos PMX, SWAP, SUS, y Generational.

Posteriormente se comparó el fitness de las comisiones generadas con 3, 4, y 5 miembros respecto del fitness de las comisiones actuales de Ingresos, Informes, y Becas del área de Informática y Comunicaciones, respectivamente.

La figura 9 presenta los fitness promedio, basado en la integración de las métricas KPP-pos e Independencia, de las comisiones generadas con 3, 4, y 5 miembros, junto con el fitness evaluado en las comisiones actuales. En la figura se observa que, al igual que en el caso de la evaluación de la métrica de Independencia en forma individual, al evaluar el fitness con las métricas integradas el método propuesto generó comisiones con mejores fitness que las comisiones actuales.

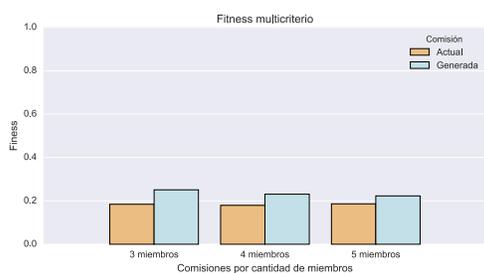


Figura 9: Fitness multicriterio de comisiones generadas y comisiones actuales.

## 4. Trabajos relacionados

El área de Recuperación de Expertos (RE), una subárea de Recuperación de Información (IR), define dos objetivos relacionados con el descubrimiento de asociaciones entre expertos y áreas de conocimiento [2]: el Perfilado de Expertos (Expert Profiling) y Hallazgo de Expertos (Expert Finding). Sin embargo, el objetivo de los sistemas de Hallazgo de Expertos es asociar expertos en un tema dado, mientras que el objetivo del Perfilado de Expertos es descubrir temas asociados a un experto.

Independientemente de su objetivo, los sistemas de RE necesitan encontrar temas o áreas de conocimiento mediante el análisis de evidencia que demuestre la experiencia del experto. En RE, esta evidencia de experiencia es generalmente representada por distintas fuentes de información como curriculum vitae, publicaciones científicas, y listas de áreas de conocimiento, entre otras.

La importancia de la evidencia de experiencia, así como el contexto, son requeridos para asistir a los usuarios de los sistemas de RE. En [1] se introduce la tarea de determinar los perfiles de expertos y proponen dos modelos para su resolución. Ambos modelos se basan en la construcción de perfiles basados en evidencia de experiencia tales como documentos y áreas de conocimiento.

La evidencia de experiencia es usualmente almacenada en documentos de texto en forma estructurada o semi estructurada. Por lo tanto, la extracción de evidencia de experiencia es uno de los principales desafíos de la RE. En [9] se reconoce la necesidad de transformar descripciones textuales de proyectos en un conjunto de habilidades que los candidatos expertos deben cumplir para ser seleccionados. Los autores proponen el concepto de *noción implícita de experiencia* en descripciones

de proyectos.

La mayoría de los enfoques para la extracción de evidencia de experiencia están basados en propiedades sintácticas de los documentos. En [10] se usan los curriculum vitae de un grupo de investigadores, como evidencia de experiencia en un modelo de recuperación de expertos basado en etiquetas. Los autores comparan el algoritmo utilizado con otros algoritmos para la recomendación de etiquetas que utilizan títulos, resúmenes, y palabras claves de contribuciones científicas. Otros trabajos se enfocan en la extracción de temas o tópicos, como representativos de áreas de conocimiento. En este sentido, en [14] se propone la extracción de temas de experiencia usando una adaptación de otros enfoques basados en extracción de términos y en extracción de frases clave. En [8] se argumenta que los enfoques actuales de RE, basados en la extracción de temas a partir de documentos, no consideran las relaciones ocultas entre los candidatos y las consultas. Con este fin, los autores proponen un enfoque para el Hallazgo de Expertos basado en Latent Dirichlet Allocation (LDA).

Otros trabajos en la extracción de experiencia tienen en cuenta otras propiedades de los documentos, además de las sintácticas o las semánticas. En [13] se aplica ML a la tarea de Perfilado de Expertos mediante el etiquetado automático de personas dentro de un contexto empresarial. Los autores proponen un listado de etiquetas con un orden de relevancia o ranking. Este enfoque utiliza un vocabulario de etiquetas creado por un conjunto de usuarios para este problema en particular.

Además de la evidencia de experiencia, otro aspecto importante sobre la creación de perfiles de expertos radica en la organización de los datos de dichos perfiles, así como de su inherente cambio. En [12] se introduce la tarea de identificar y caracterizar los cambios en la experiencia de los individuos en el tiempo. Los autores proponen el concepto de *perfil de experiencia jerárquica* para organizar la taxonomía de la experiencia, representada mediante un árbol ponderado. En esta línea, en [15] se considera la estructura jerárquica de las áreas de conocimiento relacionadas a un experto mediante un modelo de clasificación.

Algunos trabajos explotan particularidades del dominio de aplicación para la RE. En este sentido, en [7] se utilizan aspectos específicos del dominio de los sitios de pregunta-respuesta. Entre estos aspectos se incluyen la relevancia del tema, reputación del usuario, y la autoridad de una categoría.

En [6, 5] se modelan múltiples aspectos de la experiencia de personas en el marco del problema de Asignación de Revisores en Comités (CRA). Los autores proponen la extracción de temas a partir de documentos asociados a los candidatos expertos, como el caso de las contribuciones científicas.

En [11] se identifican un conjunto de criterios de selección, tanto incluyentes como excluyentes, de expertos para la conformación de paneles de asesores. El trabajo propone principios como la transparencia, disponibilidad, y público acceso de los criterios de elegibilidad y selección. Una diferencia fundamental con el trabajo propuesto es que la identificación de criterios se realiza sobre los lineamientos de las organizaciones relevadas, sin considerar la opinión de los responsables del proceso de selección. Por otra parte, los autores reconocen la importancia de la parcialidad (bias), y menciona algunas alternativas para su tratamiento, entre las que se encuentra el análisis de las declaraciones públicas de los candidatos, aunque no de las declaraciones de terceros.

## 5. Conclusiones

En este trabajo se propone un abordaje integral al problema de la selección de grupos de expertos basado en Aprendizaje Automático (ML) Procesamiento de Lenguaje Natural (NLP) y Análisis de Redes Sociales (SNA).

Con este trabajo, se pretende disponer de una estrategia para resolver la costosa tarea de analizar procesos de selección de expertos, y a su vez, sugerir conformaciones posibles de grupos de expertos en forma repetible y con pretensiones de objetividad. De esta manera, el enfoque propuesto es un aporte para dar claridad a los procesos de selección de grupos de expertos.

Los resultados obtenidos en la evaluación experimental indican que la correspondencia entre evaluaciones de los candidatos y los criterios de selección puede ser evaluada a partir de herramientas automatizadas de extracción y análisis de contenido semántico. Asimismo, la evaluación del método de selección de grupos de expertos basado en análisis de redes sociales, indican que es posible generar grupos con un mejor propiedades estructurales al compararlos con los comités actuales.

Independientemente de las contribuciones del trabajo, resulta pertinente indicar algunas consideraciones, como el costo del enfoque de aprendi-

zaje supervisado, el costo y estado actual de desarrollo de las alternativas de análisis semántico automático de textos, la representatividad del modelo de red social utilizado.

Actualmente, esta línea de investigación se continúa en el marco de una investigación posdoctoral, con especial interés en la construcción de perfiles de expertos en grandes conjuntos de datos, particularmente en el ámbito científico.

## References

- [1] Balog, K., Azzopardi, L., & de Rijke, M. (2009). A language modeling framework for expert finding. *Information Processing & Management*, 45(1), 1–19.
- [2] Balog, K. & De Rijke, M. (2007). Determining Expert Profiles (With an Application to Expert Finding). In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, volume 7 of *IJCAI '07* (pp. 2657–2662).
- [3] Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, 12(1), 21–34.
- [4] Gildea, D. & Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational linguistics*, 28(3), 245–288.
- [5] Karimzadehgan, M. & Zhai, C. (2012). Integer Linear Programming for Constrained Multi-Aspect Committee Review Assignment. *Information processing & management*, 48(4), 725–740.
- [6] Karimzadehgan, M., Zhai, C., & Belford, G. (2008). Multi-aspect expertise matching for review assignment. In *Proceedings of the 17th ACM conference on Information and knowledge management*, CIKM '08 (pp. 1113–1122).
- [7] Liu, D.-R., Chen, Y.-H., Kao, W.-C., & Wang, H.-W. (2013). Integrating expert profile, reputation and link analysis for expert finding in

- question-answering websites. *Information Processing & Management*, 49(1), 312–329.
- [8] Momtazi, S. & Naumann, F. (2013). Topic modeling for expert finding using latent Dirichlet allocation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(5), 346–353.
- [9] Neshati, M., Beigy, H., & Hiemstra, D. (2014). Expert group formation using facility location analysis. *Information Processing & Management*, 50(2), 361–383.
- [10] Ribeiro, I. S., Santos, R. L., Gonçalves, M. A., & Laender, A. H. (2015). On tag recommendation for expertise profiling: A case study in the scientific domain. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15 (pp. 189–198).
- [11] Rowe, S., Alexander, N., Weaver, C. M., Dwyer, J. T., Drew, C., Applebaum, R. S., Atkinson, S., Clydesdale, F. M., Hentges, E., Higley, N. A., & Westring, M. E. (2013). How experts are chosen to inform public policy: Can the process be improved? *Health Policy*, 112(3), 172–178.
- [12] Rybak, J., Balog, K., & Nørnvåg, K. (2014). Temporal expertise profiling. In *European Conference on Information Retrieval*, ECIR '14 (pp. 540–546).
- [13] Serdyukov, P., Taylor, M., Vinay, V., Richardson, M., & White, R. W. (2011). Automatic people tagging for expertise profiling in the enterprise. In *European Conference on Information Retrieval*, ECIR '11 (pp. 399–410).
- [14] Silvello, G., Bordea, G., Ferro, N., Buitelaar, P., & Bogers, T. (2016). Semantic representation and enrichment of information retrieval experimental data. *International Journal on Digital Libraries*, (pp. 1–28).
- [15] Wu, T., Wang, Q., Zhang, Z., & Si, L. (2015). Determining expert research areas with multi-instance learning of hierarchical multi-label classification model. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI '07 (pp. 2305–2511).
- [16] Xu, X. (2004). A note on the subjective and objective integrated approach to determine attribute weights. *European Journal of Operational Research*, 156(2), 530–532.
- [17] Zamudio, E., Berdun, L. S., & Amandi, A. (2013). Un algoritmo genetico para la conformacion de grupos de individuos distantes en redes sociales. In *Argentine Symposium on Artificial Intelligence*, 42 JAIIO Cordoba, Argentina.
- [18] Zamudio, E., Berdún, L. S., & Amandi, A. (2014). An approach to the creation of commissions of independent individuals using social networks and genetic algorithms. *Revista Iberoamericana de Inteligencia Artificial*, 17(53), 24–34.
- [19] Zamudio, E., Berdún, L. S., & Amandi, A. A. (2016). Social Networks and Genetic Algorithms to Choose Committees with Independent Members. *Expert Syst. Appl.*, 43(C), 261–270.