

Un enfoque para la conformación de comisiones de individuos independientes utilizando redes sociales y algoritmos genéticos

Eduardo Zamudio, Luis S. Berdun, y Analía A. Amandi

ISISSTAN (CONICET - UNCPBA)

Campus Universitario, Paraje Arroyo Seco, Tandil, Argentina

eduardo.zamudio@isistan.unicen.edu.ar, {lberdun, amandi}@exa.unicen.edu.ar

Resumen La selección de un conjunto de individuos a partir de una población para la conformación de comisiones es una actividad propia de varios ámbitos organizativos. Una de las características deseables de estas comisiones es que sus integrantes presenten la mayor independencia entre sí. Este problema, que en poblaciones pequeñas de candidatos podría considerarse trivial, resulta poco manejable cuando esta población supera las decenas de integrantes. En este trabajo se propone un enfoque de selección de candidatos a conformar comisiones tomando en cuenta criterios de independencia basado en las relaciones entre estos individuos, y posteriormente la aplicación de un algoritmo genético para la búsqueda de una conformación que resulte óptima en relación a la mayor independencia entre ellos. El enfoque propuesto es evaluado mediante un caso de estudio en el que se propone la construcción de una red de investigadores y posteriormente la definición de un algoritmo genético para la búsqueda de una conformación óptima de una comisión. La evaluación experimental da cuenta de la posible aproximación a un óptimo global en las configuraciones propuestas.

Palabras clave: algoritmo genético, red social, selección de miembros, independencia de miembros, grupo.

Abstract The selection of a set of individuals from a population to the conformation of commissions is an activity of various organizational environments. One of the desirable characteristics of these committees is that their members submit the largest independence. This problem, which in small populations of applicants could be considered trivial, is troublesome when this population exceeds tens of members. This paper presents an approach for selecting candidates to form committees considering independence criteria based on the relationships between these individuals and then applying a genetic algorithm to search for optimal conformation that results in relation to the proposed higher independence between them. The proposed approach is evaluated through a case study involving the construction of a network of researchers and then the definition of a genetic algorithm for finding the optimal conformation of a commission is proposed. The experimental evaluation reports on the possible approach to a global optimum in the proposed configurations.

Keywords: genetic algorithm, social network, member selection, member independence, group.

1. Introducción

Al momento de conformar una comisión, una característica deseable es la independencia de los miembros candidatos. Cómo seleccionar estos candidatos en base a criterios objetivos puede resultar un trabajo complejo, ya sea en la definición misma de estos criterios, como en el análisis de la población. Este problema requiere la selección de un conjunto de candidatos de una población en la que no interesa el orden de

éstos, aunque es requisito que dichos individuos no se repitan en una misma comisión, y a la vez presenten mayor independencia entre sí.

Son varias las disciplinas que requieren una alternativa para la generación automática o semiautomática de un conjunto de candidatos para la conformación de comisiones. Algunos trabajos han dedicado su esfuerzo en el análisis de las comisiones conformadas en diversos ámbitos, como el análisis de comités de auditoría [2], de directorios de organizaciones [22, 27], o de organismos gubernamentales [18]. Otros se han enfocado en la selección de algoritmos de clasificación [5, 3, 16, 29], otros basados en agrupamiento [12, 24], y otros han propuesto alternativas mediante enfoques evolutivos [25, 10]. Algunos presentan un enfoque basado en la diversidad de sus miembros [30, 15, 12, 21, 14, 4], mientras que otros trabajos tienen en cuenta la elección de la comisión mediante votación de un grupo de electores [6, 7, 11].

Inicialmente, el problema de la conformación de comisiones podría representarse como una combinación, lo que tendría como consecuencia un espacio de búsqueda que fácilmente podría resultar en dimensiones no adecuadas para su tratamiento computacional (mucho menos manual). Esto último debido a que para una población n , la cantidad de grupos posibles está dada por $2^n - 1$ con una complejidad de orden exponencial $O(2^n)$. Adicionalmente, si requerimos que las comisiones se conformen en r grupos, la cantidad de soluciones posibles estará dada por la aplicación del coeficiente binomial ${}_nC_r$ con complejidad de orden factorial $O(n!)$.

Ante la complejidad del problema y la consecuente inviabilidad de diseñar una solución polinómica, es posible adoptar un enfoque no determinístico que permita aproximar a una mejor solución. Cabe mencionar que una vez conformadas las comisiones, sería necesario determinar las mejores soluciones, quizás mediante la implementación de un ranking evaluado a partir de la aplicación de una función que incluya los criterios de independencia entre los miembros candidatos.

Como alternativa a la revisión exhaustiva de todas las soluciones posibles, podría adoptarse un enfoque estocástico mediante la generación aleatoria de n soluciones a las que se le aplicara la función de independencia, y posteriormente se generaría un ranking de los resultados. Este enfoque no sería adecuado debido a que presentaría la subjetividad de la elección aleatoria de los miembros de la comisión, quedando ésta determinada por la probabilidad de cada candidato de ser elegido para la conformación de la comisión, y la probabilidad conjunta de sus miembros. Esta probabilidad solo expresa la posibilidad de que un candidato sea elegido en la comisión. Si consideramos cuál es la probabilidad de conformar la mejor comisión, dicho valor es inversamente proporcional al número de alternativas.

Sin embargo, los problemas de estas características pueden ser abordados alternativamente mediante la implementación de un algoritmo genético (AG), cuyo objetivo consiste en aproximar soluciones a una solución óptima. Este enfoque resulta prometedor para el contexto de selección de individuos para la conformación de comisiones, ya que lo que se busca es maximizar la independencia entre sus miembros, pero no necesariamente conformar la mejor comisión. Es decir, es suficiente con aproximar a una buena solución.

Habiendo determinado el enfoque para el proceso de búsqueda de una solución óptima, queda por definir qué se pretende optimizar, que en este caso se corresponde con la maximización de la independencia entre los miembros de la comisión.

Si la población de candidatos es considerada como un conjunto de individuos que se encuentran relacionados entre sí, es posible determinar cuáles de estas relaciones pueden resultar relevantes para analizar la independencia entre ellos. Teniendo en cuenta estos elementos (individuos, y relaciones entre individuos) podemos representar la población como una red social, y en consecuencia aplicar técnicas de análisis sobre la misma que permitan seleccionar un conjunto de candidatos en base a criterios establecidos.

En este sentido, la construcción de una red social requiere un conjunto de datos mediante los cuales se puedan representar los actores, las relaciones entre los actores, el tipo de red, y el objeto de análisis.

En [26] se presenta una clasificación de las notaciones utilizadas para representar redes sociales. Estas representaciones se clasifican en: grafos, sociomatrices (matrices de adyacencia), o representaciones algebraicas. La elección de la notación se encuentra relacionada con el tipo y las características de la red que se pretende representar, así como también por los medios para analizarla.

La notación usualmente preferida para la representación y análisis por computadora de redes sociales son las sociomatrices, debido a que éstas utilizan matrices para representar las relaciones entre cada par de nodos.

Las técnicas actuales de análisis de redes sociales concentran su esfuerzo en la identificación de la

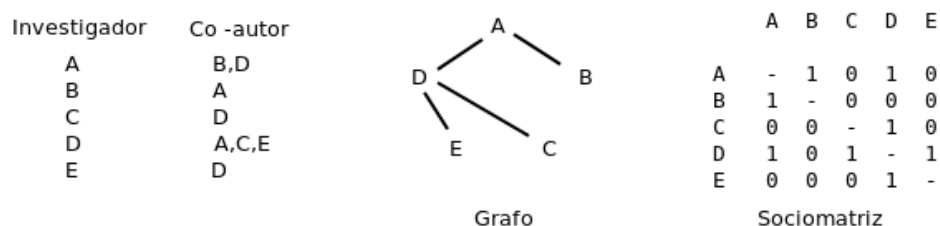


Figura 1: Entradas de la relación *publicó con* representada mediante un grafo y una sociomatriz.

cantidad o valor de sus relaciones, el rol de un nodo en la red, la importancia de dicho nodo en la estructura de la red, el descubrimiento de grupos ocultos, o la identificación de grupos cohesivos.

El objetivo de este trabajo consiste en seleccionar un conjunto de individuos representativos de una población con menor relación entre sí para conformar una comisión. Para ello, se propone primero, la representación de esta población como una red social para determinar las relaciones entre estos individuos, y luego la aplicación de un AG con el objetivo de determinar el grupo de n candidatos más distantes para la conformación de una comisión. Este trabajo corresponde a una versión extendida de una presentación realizada en el XIV Argentine Symposium of Artificial Intelligence (ASAI) en el marco de las 42th Jornadas Argentinas de Informática (JAIIO), respecto del cual se ha ampliado el análisis de la problemática abordada en relación a su complejidad, permitiendo a su vez una revisión del caso de estudio y la oportunidad de extender la descripción del enfoque propuesto.

El documento se encuentra organizado de la siguiente manera. La sección 2 describe la construcción de la red social y sus características. La sección 3 describe la implementación del AG. La sección 4 presenta las condiciones de ejecución del AG y sus resultados. La sección 5 presenta los trabajos relacionados. Finalmente, la sección 6 presenta las conclusiones y trabajos futuros.

2. Construcción de la red social

En este trabajo se estableció el objetivo de seleccionar un conjunto de candidatos de una población que presenten menor relación entre sí, para lo cual se optó por la construcción de una red social que permita analizar las relaciones entre los individuos de la población.

Una red social se compone de individuos (representados como actores) y relaciones (representadas como enlaces entre los actores), donde generalmente interesa analizar las relaciones entre individuos y los grupos que conforman [26].

En particular, nos interesa la construcción de la red social por la capacidad de representar criterios de análisis a partir de sus relaciones. Para clarificar este concepto se propone a modo de ejemplo, la construcción de una red de investigadores relacionados mediante la coautoría de publicaciones, y su pertenencia al mismo lugar de trabajo. En esta red, los investigadores representan los candidatos para la conformación de comisiones, y las relaciones representan los criterios de análisis para determinar la distancia entre cada par de investigadores en la red.

Como se mencionó anteriormente, las relaciones entre los actores determinan lo que puede analizarse de la red. En este caso, el objetivo del análisis es determinar la distancia entre un grupo de actores. Para ello, se propone representar las relaciones como sociomatrices, las cuales establecen en forma binaria las relaciones entre actores (por ejemplo, si dos investigadores han publicado juntos, o bien, si trabajan en el mismo lugar). La figura 1 muestra la relación *publicó con* de un conjunto de 5 investigadores (A,B,C,D,E), mediante un grafo y una sociomatriz, en la que se representan relaciones de tipo binaria (está presente o no), no dirigidas (no interesa el sentido de la relación), y no reflexivas (un investigador no publica consigo mismo). Estas mismas características están presentes en la relación que representa si dos investigadores trabajan en un mismo lugar.

Una vez definidas las sociomatrices que corresponden a las relaciones entre los actores, se propone la construcción de una sociomatriz que incorpore los criterios definidos, es decir, todos los tipos de relaciones definidas. Este procedimiento consiste en realizar un solapamiento mediante la aplicación del operador lógico *OR* para las celdas de posiciones iguales de las matrices de criterios en una matriz de criterios

	A	B	C	D	E		A	B	C	D	E		A	B	C	D	E		
A	-	1	0	1	0		A	-	1	0	0	0		A	-	1	0	1	0
B	1	-	0	0	0	OR	B	1	-	1	0	0	=	B	1	-	1	0	0
C	0	0	-	1	0		C	0	1	-	0	0		C	0	1	-	1	0
D	1	0	1	-	1		D	0	0	0	-	1		D	1	0	1	-	1
E	0	0	0	1	-		E	0	0	0	1	-		E	0	0	0	1	-
	publicó con						trabaja en mismo lugar						Criterios acumulados						

Figura 2: Generación de la sociomatrix de criterios acumulados a partir de sociomatrices de relaciones entre investigadores mediante la aplicación del operador lógico OR.

acumulados. Suponiendo que se tienen dos matrices de criterios $C1$ y $C2$, las celdas de estas matrices están representadas por su posición en la fila i y la columna j , y la matriz de criterios acumulados (CA) está dada por $CA = C1 OR C2$, donde cada celda $ca_{i,j}$ perteneciente a CA está dada por $ca_{i,j} = c1_{i,j} OR c2_{i,j}$ para $\forall i, j \in S$ donde $i \neq j$, y S representa el conjunto de investigadores. La figura 2 presenta la aplicación del operador lógico sobre las matrices de criterios del ejemplo y en consecuencia la generación de la matriz de criterios acumulados.

Este trabajo propone que la menor relación entre cada par de individuos puede obtenerse alternativamente en base a la distancia entre éstos. En consecuencia es necesario determinar la distancia entre cada par de nodos en la red, para lo cual se utilizaron las métricas de camino más corto [8], y distancia geodésica (longitud del camino más corto) sobre la matriz de criterios acumulados.

Un requisito previo a la determinación de las distancias es que debe garantizarse que la red esté conectada, es decir, que cada nodo de la red sea alcanzable desde cualquier otro nodo de la red. Esto puede determinarse mediante una matriz de alcanzabilidad, la cual se obtiene a partir del producto de matrices [26].

La distancia entre cada par de nodos se representa mediante una matriz de distancia, generada a partir de la potenciación [26] de la matriz de criterios acumulados. Esta matriz de distancia contiene los datos de entrada de un algoritmo cuyo objetivo es conformar una comisión en la cual los integrantes se encuentren mínimamente relacionados. Particularmente se trabajó con un AG, en el cual la entrada es una función que pretende ser optimizada para determinar el grupo de individuos que presenten menor relación entre sí o (su equivalente para este trabajo) mayor distancia entre sus miembros.

La figura 3 presenta un diagrama de flujo para el enfoque propuesto. En este diagrama se resume el proceso de ingreso de los individuos que conforman la población junto con los criterios de independencia. Con estos datos se genera la matriz de criterios acumulados con objeto de determinar las relaciones y conformar la red social. Una vez obtenida la red es posible determinar la matriz de distancias utilizando la distancia geodésica. Luego, tanto la matriz de distancias como la información de los nodos de la red son ingresados en el algoritmo genético para la generación de las posibles soluciones y su evolución en busca de una solución óptima.

3. Algoritmo genético

Un AG es un tipo de algoritmo evolutivo que puede ser considerado como un método de optimización de funciones [23]. Si bien no existe un AG definitivo, es posible adaptar uno a partir de un conjunto de representaciones y operadores que pueden ser adecuados a las necesidades particulares de una aplicación. El elemento con el que trabaja el AG es el cromosoma, el cual contiene la información genética representada por la disposición y valor de sus genes.

Con objeto de seleccionar un conjunto de individuos de la red social, se ha definido una función de relación ad-hoc, la cual establece la suma acumulada de las distancias entre todos los candidatos de la comisión conformada. Consecuentemente se ha diseñado un AG, con el fin de obtener soluciones aproximadas a una mejor solución mediante la maximización de dicha función aplicada a un conjunto de n candidatos.

Asimismo, el desarrollo del AG requiere definir su representación, la función de fitness, los mecanismos de selección de padres y de selección de supervivientes a utilizar, así como los operadores de cruce y de

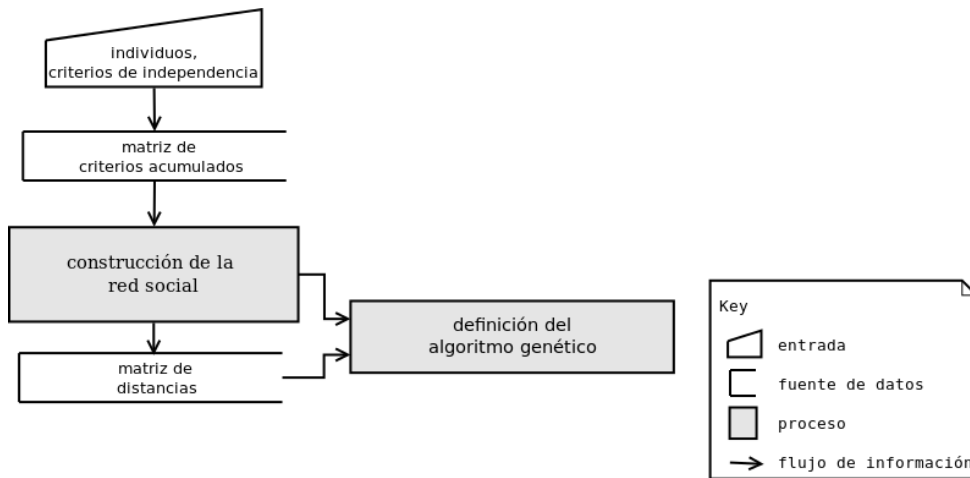


Figura 3: Diagrama de flujo de la primera parte del enfoque propuesto en el que se presentan las entradas, la generación de las fuentes de datos, y los procesos vinculados a la construcción de la red social y la definición del algoritmo genético.

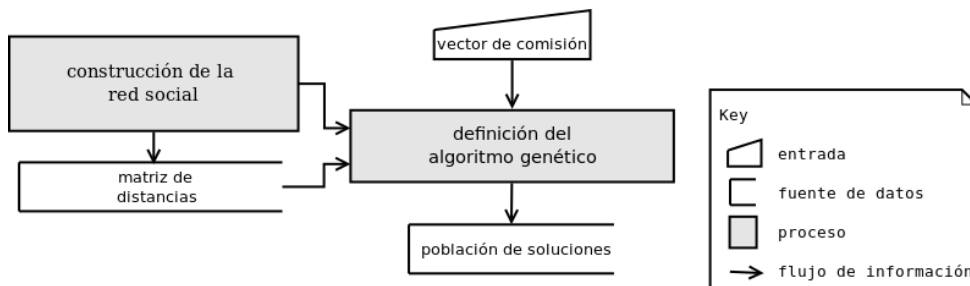


Figura 4: Diagrama de flujo de la segunda parte del enfoque propuesto en el que se presentan las entradas, la generación de las fuentes de datos, y los procesos vinculados a la definición del algoritmo genético y la generación de la población de soluciones.

mutación. A continuación se presenta cada uno de estos componentes.

Representación

La representación del problema requiere definir cómo se compone un cromosoma. Este trabajo se basó en permutaciones de un vector de enteros (cromosoma) donde cada elemento del mismo representa una referencia a un investigador (gen), y en cuyo vector están incluidos todos los investigadores de la población. Es decir, un cromosoma tiene tantos genes como investigadores contenga la población. La inclusión o exclusión de los candidatos en la comisión está dada por un vector de igual tamaño al vector de investigadores, haciendo corresponder las posiciones de ambos vectores para indicar la conformación de comisiones. De esta manera, el conjunto de candidatos propuestos para la conformación de una comisión presenta ocurrencias únicas de los candidatos, es decir, un candidato no puede repetirse en una misma comisión. El orden en que se presentan dichos candidatos dentro de la comisión no tiene relevancia para el problema. La figura 5 presenta en forma gráfica dicha representación en vectores de longitud n .

El problema requiere determinar el tamaño de las comisiones, de modo que el tamaño de la comisión debe ser establecido de antemano, para lo cual es definido el *vector de comisión*. Este vector es único para cada ejecución y es utilizado por el algoritmo genético para indicar las posiciones de los vectores solución que conformarán la comisión como lo presenta la figura 4.

La codificación de la permutación queda representada mediante el elemento i , el cual denota el evento que sucede en ese lugar en la secuencia. Por ejemplo, para cuatro investigadores $[I1, I2, I3, I4]$, la

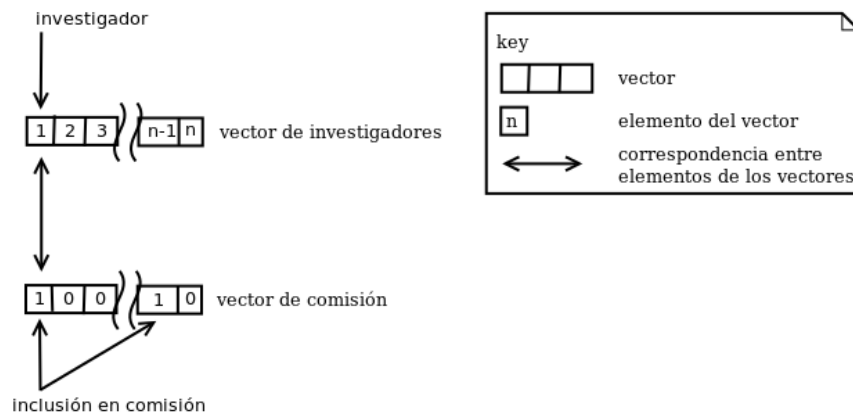


Figura 5: Representación del algoritmo genético mediante un vector de investigadores, que contiene todos los investigadores de la población, y un vector de comisión, que indica los elementos del vector de investigadores incluidos en la comisión.

permutación $[2, 1, 4, 3]$ queda representada por $[I2, I1, I4, I3]$.

Fitness

La función de fitness tiene por objetivo determinar el valor de una solución, el cual se busca optimizar de acuerdo al problema planteado. Para este trabajo, donde se requiere maximizar las distancias entre los investigadores del grupo candidato, se desarrolló una función de modo ad-hoc representando la suma acumulada de las distancias entre cada par de candidatos dentro de una comisión (solución individual), quedando expresada dicha función de la siguiente manera:

$$f = \sum d(i, j)$$

$$\forall i, j / i \neq j \text{ y } i, j \in S$$

Donde d es la función de distancia entre dos miembros, i y j son miembros de la comisión, y S representa el conjunto de investigadores.

Como se estableció previamente, es condición de la construcción de la red que ésta debe estar conectada, lo que implica que todos los miembros de cualquier comisión presentan una distancia entre sí distinta de ∞ .

Selección de padres

La información genética de las nuevas generaciones se obtiene a partir de padres, los cuales son cromosomas (soluciones) de la generación inmediatamente anterior. Con este fin, es necesario definir una estrategia de selección de padres, en la que se puede utilizar alguno de los mecanismos adecuados al tipo de problema que se pretende resolver.

Los mecanismos de selección utilizados en este trabajo incluyen los algoritmos: *Stochastic Universal Sampling (SUS)* debido a que deben seleccionarse varios padres de una población; y *Tournament*, ya que en ambos casos se pretende obtener resultados sin conocer el fitness global.

Cruce

La información genética de la nueva generación es determinada por la información genética de los padres de los cuales desciende. Este proceso de recombinación genética se realiza mediante mecanismos de cruce. Por ejemplo, teniendo dos cromosomas indicando soluciones distintas, el cruce implica que la descendencia obtendrá información genética de ambas soluciones.

Con el objetivo de mantener la permutación válida se utilizaron los operadores de recombinación para permutaciones *Partially Mapped Crossover (PMX)* y *Order Crossover (OX)*. Ya que *PMX*, al ser un algoritmo diseñado para problemas de adyacencia, resulta aplicable al problema presentado. Por su parte, *OX* está diseñado para problemas de orden, sin embargo, la información del orden del segundo padre puede resultar beneficiosa en la generación del nuevo cromosoma.

Mutación

Otro mecanismo de recombinación genética utilizado en este trabajo es la mutación, el cual implica alterar los genes de un cromosoma. En permutaciones, como es el caso del ejemplo de la representación de los investigadores, la mutación implica alterar la disposición de los valores presentes en el vector solución de la nueva generación.

Los operadores de mutación seleccionados para ser aplicados en la descendencia fueron *Swap Mutation* e *Insert Mutation*, ya que ambos son operadores aceptados para mantener la permutación válida.

Selección de supervivientes

Para la selección de supervivientes se utilizó el mecanismo *Steady-state* con objeto de no perder los individuos con mejor fitness en las sucesivas generaciones.

4. Caso de estudio

El enfoque propuesto ofrece una alternativa ante la inviabilidad de un análisis determinístico del problema. Esta propuesta está diseñada para ser compatible con redes completas, que adicionalmente son consideradas pequeñas. Si bien esta característica del tamaño no presenta una definición formal, se tomó como límite la generación de una red de no más de 1000 nodos.

Con esto en cuenta, se ha diseñado un caso de estudio en el cual se presentan investigadores candidatos a conformar una comisión, con objeto de demostrar que dichos candidatos presentarán menor relación entre sí, lo que puede asumirse como un criterio de independencia en la actuación de estos investigadores dentro de una comisión.

Los datos utilizados para la experimentación fueron obtenidos de la base de datos publicada por The Auton Lab [1], parte del School of Computer Science de Carnegie Mellon University, la cual contiene información de coautoría de trabajos presentados en el Neural Information Processing conference (NIPS). Este conjunto de datos fue reducido a los primeros 1001 registros con objeto de facilitar su procesamiento, y cuya selección incluyó 720 investigadores, 443 publicaciones, la cual se enriqueció con 11 lugares de trabajo que fueron asignados en forma aleatoria a los investigadores.

Como se mencionó previamente, el enfoque de aplicación del AG para la determinación de los miembros de la comisión se presenta como una alternativa a otros métodos, como podría ser una combinatoria de 720 individuos tomados en comisiones de 5 integrantes, lo que resultaría en ${}_{720}C_5 = 1,590145128 \times 10^{12}$ soluciones posibles; o como la selección aleatoria de un conjunto de soluciones posibles donde la probabilidad de elección de cada candidato estaría dada por $P\left(\frac{1}{720}\right) = 0,1388889 \times 10^{-2}$, y la probabilidad conjunta de la comisión de 5 candidatos estaría dada por $\sum_{n=0}^4 P\left(\frac{1}{720-n}\right) = 0,2779712 \times 10^{-2}$. La probabilidad de encontrar un óptimo asumiendo la existencia de m posibles conformaciones de comisiones óptimas, es igual a $m/{}_{720}C_5$, lo que equivale a un valor próximo a 0.

Con objeto de evaluar la solución propuesta mediante el enfoque del AG, se planteó un caso experimental en el que se tuvieron en cuenta los siguientes aspectos:

- Parámetros de ejecución:
 - Tamaño de población: El número de soluciones candidatas en cualquier punto de tiempo se calculó mediante P/n , donde P representa el conjunto de todos los investigadores, y n el tamaño de las comisiones. En el caso de estudio se utilizó $P = 720$ y $n = 5$.
 - Probabilidad de cruce: Se seleccionó el valor 0,7, tomado del rango $[0,6; 0,9]$.

Configuración	Selección		Cruce		Mutación	
	<i>Tournament</i>	<i>SUS</i>	<i>PMX</i>	<i>OX</i>	<i>Swap</i>	<i>Insert</i>
1	X		X		X	
2	X		X			X
3	X			X	X	
4	X			X		X
5		X	X		X	
6		X	X			X
7		X		X	X	
8		X		X		X

Cuadro 1: Configuraciones del algoritmo genético en la que se indica para cada configuración el método de selección, el operador de cruce, y el operador de mutación.

Configuración	Distancia de acuerdo al N ^o de ejecución					Distancia promedio ¹	σ
	1	2	3	4	5		
2	17	12	18	12	20	15,8	3,6332
5	12	11	19	12	19	14,6	4,0373
3	12	18	12	12	12	13,2	2,6833
7	12	12	18	12	12	13,2	2,6833
4	12	12	11	12	18	13,0	2,8284
6	12	20	10	12	11	13,0	4,0000
1	12	17	12	12	11	12,8	2,3875
8	12	12	12	11	12	11,8	0,4472

Cuadro 2: Ranking de ejecuciones para cada configuración en la que se presentan 5 ejecuciones, la distancia promedio, y la desviación estándar (se resaltan los máximos obtenidos en las ejecuciones).

- Probabilidad de mutación: Se seleccionó el valor 0,15, tomado del rango [0,01; 0,15].
- Condición de corte: Se establecieron 25 generaciones como límite de corte.
- Configuraciones: Se tomaron 8 configuraciones distintas, como resultado de las posibles combinaciones de los mecanismos adoptados en este trabajo (selección (2), operadores de cruce (2), y operadores de mutación (2)). Adicionalmente, se optó por el mecanismo de selección *Steady-state*. La tabla 1 presenta estas configuraciones.
- Ejecuciones: Se realizaron 40 ejecuciones, correspondientes a 5 ejecuciones por cada configuración propuesta, luego se promediaron los resultados de las ejecuciones, y posteriormente se calculó la desviación estándar (σ) de las mismas.

Resultados

La tabla 2 presenta el ranking de las ejecuciones donde se indica para cada configuración: el fitness de las ejecuciones, el fitness promedio, y la desviación estándar. El fitness logrado en las ejecuciones se encuentra en el intervalo [11;20], a partir de lo cual se plantea la hipótesis de que el fitness global se aproxima a 20 para este caso de estudio. Con objeto de corroborar dicha hipótesis, se amplió el número de generaciones a 50, 100, 250, 500, y 1000. La figura 6 presenta el fitness promedio de 5 ejecuciones para las 8 configuraciones de acuerdo a la cantidad de generaciones indicada, donde se evidencia que con el incremento de generaciones en la ejecución del algoritmo, el fitness promedio se aproxima a 20.

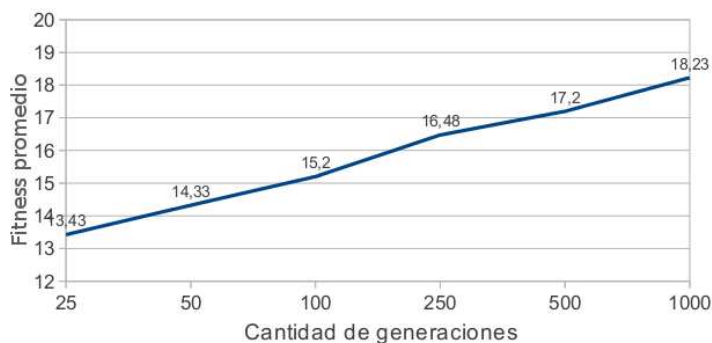


Figura 6: Fitness promedio obtenido a partir de las configuraciones de acuerdo a la cantidad de generaciones.

5. Trabajos relacionados

Algunos trabajos presentes en la literatura se enfocan en la determinación de la importancia de un nodo en la red, como PageRank [20], AuthorRank [17] (en un ámbito de coautoría de trabajos académicos). Otros trabajos identifican métricas de centralidad de los nodos de una red como *degree*, *closeness*, y *betweenness* [9]. En [19] se propone un modelo que generaliza las métricas de centralidad en redes donde se incorporan el número de relaciones y el peso de éstas en las métricas propuestas originalmente en [9].

Otros trabajos se enfocan en la identificación de vecinos más cercanos en la red. En [13] se define una red de coautoría de trabajos académicos para medir la cercanía de los autores y de este modo recomendar, en base a la proximidad, trabajos de otros autores. Por otra parte, en [28] se definen dos medidas de fortaleza del camino, aplicable en grafos valuados.

En [10] se presenta una experiencia de aplicación de un algoritmo genético en el análisis de la tendencia a la conformación de grupos relativamente pequeños de individuos que presentan relaciones recurrentes, particularmente en el descubrimiento de grupos encubiertos mediante el análisis de estas relaciones.

6. Conclusiones

Este trabajo presenta dos aportes principales al problema tratado. El primero de ellos es la utilización de redes sociales para la representación de relaciones entre individuos que permita determinar los más distantes entre sí. Esta propuesta resulta innovadora debido a que, como se menciona anteriormente, la mayoría de los estudios se enfocan en la conformación de las comisiones sin tener en cuenta criterios de independencia. El segundo aporte consiste en el método de identificación de la distancia entre los individuos candidatos a conformar la comisión mediante la implementación de un algoritmo genético. Este trabajo demuestra que la utilización de un algoritmo genético, junto con la definición de una red social, resulta una alternativa posible y adecuada frente a otros enfoques (como la selección manual, o la selección aleatoria de miembros) cuando se tiene por objetivo la conformación de grupos de candidatos que presenten menor relación entre sí. Esto se da gracias a que la definición de la red social permite la generación de matrices de distancia, las cuales son utilizadas por el algoritmo genético en la búsqueda de la maximización de una función que determine las distancias entre los miembros de una comisión candidata.

Este enfoque se encuentra inicialmente limitado en la definición de la red social, ya que exige que la misma se corresponda con un grafo conectado, es decir, que exista al menos un camino entre cada par de nodos de la red, lo que es una característica propia del dominio presentado en el caso de estudio. Este es un aspecto que debe tenerse en cuenta si se pretende analizar el enfoque en otros ámbitos de aplicación. Consecuentemente, en contextos donde la red presente características como relaciones valuadas o dirigidas, o redes que presenten grupos desconectados, será necesario evaluar alternativas a la función de relación propuesta.

El mismo enfoque presentado en este trabajo podría mejorar su desempeño si se desarrollara una

función de fitness que no se encuentre tan afectada por valores extremos, por ejemplo utilizando las variables estadísticas de media o mediana en lugar de una sumatoria de las distancias, aunque debe tenerse en cuenta que la modificación de la función puede afectar su simplicidad y en consecuencia su rendimiento.

Trabajos futuros tienen por objetivo ampliar las dimensiones de la red en términos de cantidad de nodos, y tanto cantidad como tipo de relaciones entre los nodos, con objeto de analizar la escalabilidad del enfoque propuesto en este trabajo, como así también su aplicabilidad en otros dominios. Adicionalmente, se pretende analizar la conformación de grupos de candidatos mediante otras técnicas de inteligencia artificial.

Referencias

- [1] SBNS datasets. 4
- [2] Lawrence J. Abbott and Susan Parker. Auditor selection and audit committee characteristics. *AUDITING: A Journal of Practice & Theory*, 19(2):47–66, September 2000. 00253. 1
- [3] Matti Aksela. Comparison of classifier selection methods for improving committee performance. In *Multiple Classifier Systems*, page 84–93. Springer, 2003. 00043. 1
- [4] Matti Aksela and Jorma Laaksonen. Using diversity of errors for selecting members of a committee classifier. *Pattern Recognition*, 39(4):608–623, April 2006. 00074. 1
- [5] S. Argamon-Engelson and I. Dagan. Committee-based sample selection for probabilistic classifiers. arXiv e-print 1106.0220, June 2011. 00139. 1
- [6] Hans-Hermann Bock, William H.E. Day, and F.R. McMorris. Consensus rules for committee elections. *Mathematical Social Sciences*, 35(3):219–232, May 1998. 1
- [7] Peter C Fishburn. Majority committees. *Journal of Economic Theory*, 25(2):255–268, October 1981. 1
- [8] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, March 1977. ArticleType: research-article / Full publication date: Mar., 1977 / Copyright © 1977 American Sociological Association. 2
- [9] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978. Cited by 0000. 5
- [10] Linton C. Freeman. Finding groups with a simple genetic algorithm*. *Journal of Mathematical Sociology*, 17(4):227–241, 1993. 1, 5
- [11] William V. Gehrlein. The condorcet criterion and committee selection. *Mathematical Social Sciences*, 10(3):199–209, December 1985. 1
- [12] Stefan T. Hadjitodorov, Ludmila I. Kuncheva, and Ludmila P. Todorova. Moderate diversity for better cluster ensembles. *Information Fusion*, 7(3):264–275, 2006. 00110. 1
- [13] San-Yih Hwang, Chih-Ping Wei, and Yi-Fan Liao. Coauthorship networks and academic literature recommendation. *Electronic Commerce Research and Applications*, 9(4):323–334, July 2010. 5
- [14] Ludmila I. Kuncheva. Using diversity measures for generating error-correcting output codes in classifier ensembles. *Pattern Recognition Letters*, 26(1):83–90, January 2005. 00038. 1
- [15] Ludmila I. Kuncheva and Christopher J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning*, 51(2):181–207, 2003. 01013. 1
- [16] Leijun Li, Bo Zou, Qinghua Hu, Xiangqian Wu, and Daren Yu. Dynamic classifier ensemble using classification confidence. *Neurocomputing*, 99:581–591, January 2013. 1

-
- [17] Xiaoming Liu, Johan Bollen, Michael L. Nelson, and Herbert Van de Sompel. Co-authorship networks in the digital library research community. *Information Processing & Management*, 41(6):1462–1480, December 2005. 5
- [18] Gerhard Loewenberg. *Handbook of Legislative Research*. Harvard University Press, 1985. 1
- [19] Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, July 2010. 5
- [20] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web., November 1999. 5
- [21] H.W. Shin and S.Y. Sohn. Selected tree classifier combination based on both accuracy and error diversity. *Pattern Recognition*, 38(2):191–197, February 2005. 00024. 1
- [22] Anil Shivdasani and David Yermack. CEO involvement in the selection of new board members: An empirical analysis. *The Journal of Finance*, 54(5):1829–1853, 1999. 00734. 1
- [23] J. E. Smith and Agoston E. Eiben. *Introduction to Evolutionary Computing*. Springer, October 2008. 3
- [24] Hui Tao, Xiao-ping Ma, and Mei-ying Qiao. Subspace selective ensemble algorithm based on feature clustering. *Journal of Computers*, 8(2), February 2013. 00000. 1
- [25] Xiao Wang and Han Wang. Classification by evolutionary ensembles. *Pattern Recognition*, 39(4):595–607, April 2006. 00028. 1
- [26] Stanley Wasserman and Katherine Faust. *Social network analysis: methods and applications*. Cambridge University Press, Cambridge; New York, 1994. 1, 2, 2
- [27] James D. Westphal and Edward J. Zajac. Who shall govern? CEO/Board power, demographic similarity, and new director selection. *Administrative Science Quarterly*, 40(1):60, March 1995. 00560. 1
- [28] Song Yang and David Knoke. Optimal connections: strength and distance in valued graphs. *Social Networks*, 23(4):285–295, October 2001. 5
- [29] Zijian Zheng. Naive bayesian classifier committees. Springer, 1998. 00002. 1
- [30] H ela Zouari, Laurent Heutte, and Yves Lecourtier. Controlling the diversity in classifier ensembles through a measure of agreement. *Pattern Recognition*, 38(11):2195–2199, November 2005. 00027. 1