

## Detección de comunidades en redes sociales por medio de un algoritmo de agrupamiento dinámico en alta definición

Christian Iván Ledesma Bermúdez<sup>1</sup>, Abel García Najera<sup>2</sup>

<sup>1</sup> Universidad Autónoma Metropolitana Unidad Cuajimalpa,  
Posgrado en Ciencias Naturales e Ingeniería,  
México

<sup>2</sup> Universidad Autónoma Metropolitana Unidad Cuajimalpa,  
Departamento de Matemáticas Aplicadas y Sistemas,  
México

2163807450@alumnos.cua.uam.mx, agarcian@correo.cua.uam.mx

**Resumen.** Las redes sociales han sido foco de un creciente interés por el amplio espectro de aplicaciones que han demostrado tener, desde marketing hasta política. Por este motivo todos los enfoques de estudio relacionados a éstas tienen una gran importancia. Existen trabajos relacionados con la detección de comunidades que estudian el problema desde varios enfoques: algunos lo tratan como un problema estático, mientras que los más recientes lo tratan como un problema dinámico. Ambos enfoques presentan ventajas y desventajas que hacen que la selección del enfoque se decida en parte por la naturaleza del problema. En este trabajo se sigue el enfoque de AI-NSGA-II y, para mejorar la forma en que se detectan las comunidades, se considera una función objetivo adicional para lograr definir mejor el agrupamiento. Posteriormente, el desempeño de la propuesta HD-AI-NSGA-II se compara con AI-NSGA-II usando conjuntos de datos artificiales y un conjunto de datos real bien conocido que contiene ID de videos de YouTube. Finalmente se demuestra que nuestro enfoque HD-AI-NSGA-II destaca al lograr una mejor definición de los grupos y, como resultado directo, una detección sobresaliente de comunidades gracias a la forma en que guía la búsqueda del frente de Pareto.

**Palabras clave:** Algoritmos evolutivos multiobjetivo, agrupamiento multiobjetivo, detección de comunidades, redes sociales.

### Community Detection in Social Networks by a High-Definition Dynamic Clustering Algorithm

**Abstract.** Social networks have been focus of growing interest for the wide spectrum of applications they have, from marketing to politics. For this reason all approaches related to them have a great importance. There are several studies

related to community detection that tackle the problema from several perspectives: some of them study it as a static problem, while the more recent ones consider it as a dynamic problem. In this work the AI-NSGA-II approach is followed and, in order to improve the way in which communities are detected, an additional objective function is considered, which manages to better define the clustering. Subsequently, the performance of our proposed HD-AI-NSGA-II is compared against AI-NSGA-II using artificial data sets and a well-known real instance containing YouTube videos ID's. Finally, it is demonstrated that the proposed HD-AI-NSGA-II improves the other approach by achieving a better definition of the clusters and, as a direct result, a better community detection thanks to the way in which it guides the search of the Pareto front.

**Keywords:** Multi-objective evolutionary algorithms, multi-objective clustering, community detection, social networks.

## 1. Introducción

En la actualidad, la detección de comunidades es uno de los tópicos con mayor importancia por las aplicaciones que puede tener en redes sociales, economía, marketing, política, búsqueda de información, minería de datos, adquisición de conocimiento, análisis de datos, informática, física, sociología, etc. [1-8]. Algunas de las aplicaciones se han combinado para desarrollar un papel principal en los últimos años, como el caso de las redes sociales y la política. Por estos motivos, la detección de comunidades es un problema ampliamente investigado, abierto y de relevancia [1, 9-12]. Este problema puede ser modelado como uno de optimización multiobjetivo, ya que comúnmente se plantean varios objetivos que están en conflicto para lograr un buen conjunto de soluciones, en particular, se modela como un problema de partición de un grafo. En este sentido, los algoritmos genéticos son una de las herramientas más importantes para la optimización de problemas multiobjetivo NP-difíciles [2, 9]. Tomando en cuenta la estructura temporal del grafo, hay dos enfoques que abordan el tema, el primero considera el problema como un caso estático, en donde el estado del grafo no cambia con el tiempo [7, 11, 13-17], mientras que el segundo considera al problema como dinámico e implementa algún mecanismo que considera los cambios del grafo a través del tiempo [2, 4, 18-23].

El problema de agrupamiento trata de encontrar la mejor partición  $C$  de un grafo  $G(V, A)$  con nodos  $V$  y aristas  $A$ . Algunos algoritmos hacen suposiciones *a priori* de la estructura del grafo y esto puede afectar el rendimiento y la calidad de las soluciones. Por otro lado, agrega un mejor rendimiento [10]. En este trabajo se adopta el enfoque evolutivo de AI-NSGA-II [24] para formar las soluciones. HD-AI-NSGA-II, que es el algoritmo propuesto, puede trabajar tanto de manera estática como temporal, ya que no está enfocado a la calidad del agrupamiento conforme el paso del tiempo. En su lugar, crea soluciones de alta definición para cada estado en el tiempo. HD-AI-NSGA-II considera un objetivo adicional para aumentar la calidad y rendimiento a comparación del enfoque adoptado. Este objetivo trata de maximizar el número de grupos y así poder diferenciar dos soluciones, que en teoría podrían ser medidas como iguales por los otros dos objetivos, logrando una mayor diversificación del frente de Pareto.

El resto del trabajo está estructurado de la siguiente forma. En la Sección 2 se hace una comparativa con algunos trabajos relacionados. En la Sección 3 se plantea el problema y la nuestra propuesta para mejorar el agrupamiento, posteriormente, en la Sección 4 se presentan los resultados de comparar el enfoque adoptado contra nuestra propuesta. Finalmente, en la Sección 5 se presentan algunas conclusiones y se describe el trabajo futuro.

## **2. Trabajo relacionado**

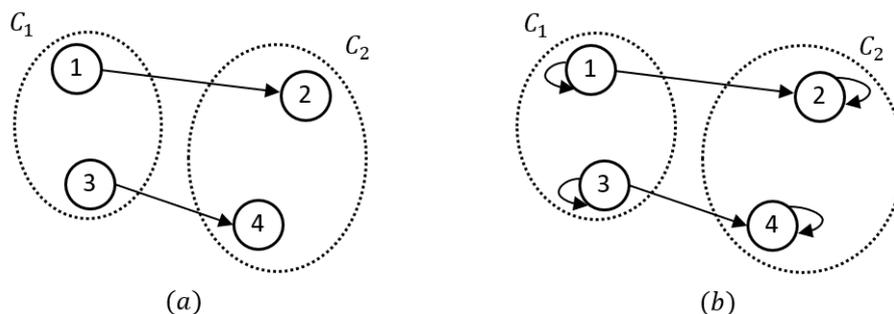
Uno de los trabajos relacionados al enfoque adoptado es el de GrasSC [25], el cual hace agrupación de usuarios para realizar recomendación de contenido. Para lograrlo utiliza el historial de páginas visitadas por cada usuario como nodo y la similitud de un historial a otro como arista. En este sentido, generan un grafo no dirigido y ponderado, dando como resultado un problema de partición de grafo, el cual es NP-difícil. Utilizan las mismas funciones objetivo que AI-NSGA-II, las cuales son  $1/(1 + M_{cut})$  y *Gobal Silhouette*, pero a diferencia AI-NSGA-II, GrasSC utiliza una codificación basada en etiquetas, además de estar basado en SPEA2 [26] y opera sobre tipos de datos categóricos y matrices de distancia.

Otro trabajo que está muy relacionado con AI-NSGA-II es DYN-MOGA [27], ya que ambos utilizan representación basada en grafo de adyacencia, adoptan el enfoque de NSGA-II [28] y operan sobre tipos de datos que representan grafos. Sin embargo, tiene una diferencia significativa y es que una de sus funciones objetivo prioriza suavizar la diferencia de los agrupamientos de dos momentos consecutivos en lugar de maximizar la precisión del agrupamiento. DYN-MOGA utiliza *Community Score* y *Normalized Mutual Information*, como funciones objetivo. El primer objetivo de DYN-MOGA trata de maximizar la calidad del agrupamiento en el estado actual de tiempo, mientras que el segundo objetivo considera el cambio que sufre una solución que pasa del tiempo  $t$  al tiempo  $t + 1$ . Estos dos objetivos entran en conflicto cuando el estado del grafo es dinámico y no se mantiene idéntico para cada estado de tiempo.

## **3. Detección de comunidades de alta definición**

El problema de detección de comunidades en redes sociales lo abordamos desde una estrategia multiobjetivo por la naturaleza de las redes sociales, ya que es difícil encontrar una red social representada por grafos disjuntos, lo común es tener una red social representada por un grafo conexo que contenga la mayoría de las relaciones entre todos los nodos. En el caso de que pudiéramos garantizar que el grafo es disjunto, podría aplicarse un método de agrupamiento de un solo objetivo. Pero en el caso de un grafo conexo, es deseable tener más de una métrica para medir la calidad del agrupamiento. Y como se demuestra en esta sección, a veces es necesario agregar nuevos objetivos para elevar la calidad de los resultados.

Usualmente, el problema de detección de comunidades trata de encontrar, en un grafo  $G(A, V)$ , una partición  $C$  compuesta por  $k$  conjuntos disjuntos de nodos  $v$  tal que cada  $v \in V$ , que minimice el número de aristas  $a$  entre grupos y maximice el número



**Fig. 1.** En (a) se muestra el caso especial de un agrupamiento donde  $W(C_1) = 0$  y  $W(C_2) = 0$ , ya que para  $C_1$  y  $C_2$  no hay relaciones internas. Para evitar la división entre cero en los términos de  $Mcut_k$  se considera implícitamente el bucle  $W_{uu} = 1$ , para todos los nodos de  $G$  como se muestra en (b), de esta forma se tiene  $W(C_1) = 2$  y  $W(C_2) = 2$ .

de aristas  $a$  dentro de cada grupo para cada  $a \in A$ . La ecuación (1) muestra el primer objetivo a maximizar, que es una modificación a la ecuación (2)  $Mcut$  [29], ya que  $Mcut$  por sí sola debe ser minimizada:

$$f_1 = \frac{1}{1+Mcut}, \tag{1}$$

en donde:

$$Mcut_k = \frac{cut(C_1, \bar{C}_1)}{W(C_1)} + \frac{cut(C_2, \bar{C}_2)}{W(C_2)} + \dots + \frac{cut(C_k, \bar{C}_k)}{W(C_k)}, \tag{2}$$

$$W(C_i) = cut(C_i, C_i), \tag{3}$$

$$cut(C_i, \bar{C}_i) = \sum_{u \in C_i, v \in \bar{C}_i} W_{uv}, \tag{4}$$

para toda  $i \in \{1, \dots, k\}$  y  $k > 1$ .  $f_1$  toma valores en el intervalo  $(0,1]$ , ya que  $Mcut_k$  toma valores en el intervalo  $[0, \infty)$ . El mejor caso de  $Mcut_k$  ocurre cuando la ecuación (4) de cada termino es igual a cero, es decir,  $cut(C_i, \bar{C}_i) = 0$ , para toda  $i \in \{1, \dots, k\}$ . Esto quiere decir que no existe ninguna arista  $a$  que vaya desde algún nodo a otro nodo de otro grupo, lo cual puede verse como que ninguna arista es cortada por la agrupación. El caso neutral está dado cuando el número de aristas entre grupos es igual al número de aristas internas de cada grupo, con lo cual cada término  $\frac{cut(C_i, \bar{C}_i)}{W(C_i)} = 1$ , teniendo como resultado  $Mcut_k = m$ , en donde  $m$  es el número de nodos de  $G$ . Finalmente, se tendría que  $f_1 = \frac{1}{1+m}$ . Esta observación es importante ya que  $f_1$  toma valores en el intervalo  $(0,1]$  y sería fácil pensar que el caso neutral estaría dado por  $f_1 = 0.5$ .

La última consideración con respecto a esta función se muestra en la Fig. 1 y hace evidente un caso especial en la ecuación (3) y por el cual se considera  $W_{uu} = 1$ , para toda  $u \in \{1, \dots, m\}$ .

El segundo objetivo por maximizar es *Global Silhouette* [30], está definido en la ecuación (5):

$$f_2 = Global\ Silhouette = \frac{\sum_{i \in V} S(i)}{|V|}, \tag{5}$$

en donde:

$$S(i) = \begin{cases} \frac{a(i)-b(i)}{\max\{a(i),b(i)\}}, & \text{para } |C_l| > 1 \\ 0, & \text{para } |C_l| = 1 \end{cases}, \forall i \in \{1, \dots, |V|\}, \quad (6)$$

$$a(i) = \frac{\sum_{j \in C_l} W_{ij}}{|C_l|}, i \in C_l, \quad (7)$$

$$b(i) = \max\{d(i, C_m)\}, m \in \{1, \dots, |C|\}, m \neq l, \quad (8)$$

$$d(i, C_m) = \frac{\sum_{j \in C_m} W_{ij}}{|C_m|}, \quad (9)$$

$f_2$  toma valores en el intervalo  $(-1,1]$ , ya que cada término de la ecuación (5) también toma valores en el intervalo  $(-1,1]$  y  $f_2$  es el promedio los mismos. En la ecuación (7) se calcula el promedio de aristas dentro del grupo al que pertenece el nodo  $i$ . La ecuación (8) selecciona el promedio máximo de aristas que existen entre el nodo  $i$  y el resto de los grupos, cada promedio es calculado con la ecuación (9). El mejor caso de *Global Silhouette* ocurre cuando la ecuación (6) toma el valor de uno y la ecuación (8) toma el valor de cero, es decir,  $S(i) = 1$ , dado  $b(i) = 0$ , para toda  $i \in \{1, \dots, |V|\}$ . Esto quiere decir que no existe ninguna arista  $a$  que vaya desde algún nodo a otro nodo de otro grupo. El caso neutral está dado cuando el número de aristas entre grupos es igual al número de aristas internas de cada grupo, con lo cual para cada término  $S(i)$  de la ecuación (5),  $a(i) = b(i)$ , teniendo como resultado *Global Silhouette* = 0. Para el caso que el grupo  $C_l$  solo contará con un nodo se le asigna a  $S(i)$  el valor cero.

La última consideración con respecto a esta función se muestra en la Fig. 2 y hace evidente un caso especial por el cual, al igual que el primer objetivo, se considera  $W_{uu} = 1$ , para toda  $u \in \{1, \dots, m\}$ .

### 3.1 HD-AI-NSGA-II

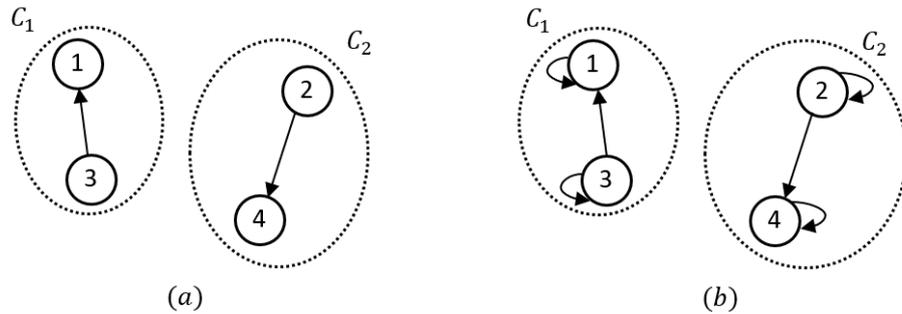
Los dos primeros objetivos deben maximizarse para encontrar el conjunto de soluciones que formen el frente de Pareto, pero por sí solos tienen una limitante ya que existen soluciones que representan diferentes agrupamientos con un número diferente de grupos, pero con mismos valores de aptitud, como se muestra en la Fig. 3.

Como se puede apreciar en la Fig. 3, la solución (b) es mejor que la (a) al estar mejor definida, por lo tanto, proponemos como tercer objetivo a maximizar la ecuación (10), que es el número de grupos  $k$ :

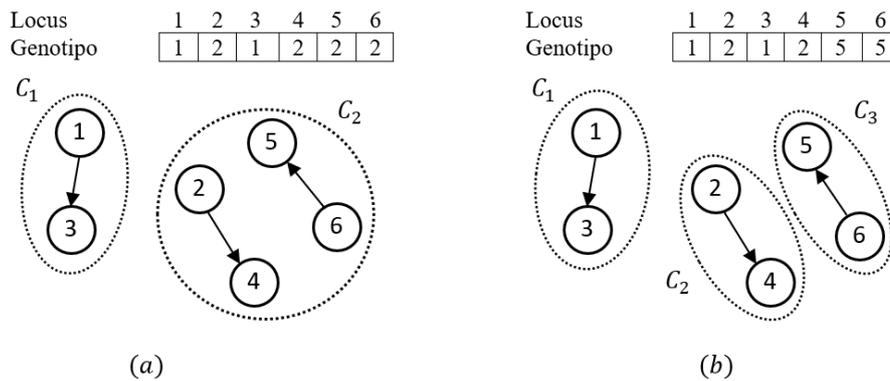
$$f_3 = k. \quad (10)$$

### 3.2 Mecanismos genéticos

Para poder hacer una comparación estricta entre AI-NSGA-II y HD-AI-NSGA-II, se utilizan los mismos mecanismos genéticos propuestos por el enfoque seguido. Para la codificación utilizamos *locus-based adjacency* [31] como forma de representar cada solución. En esta representación el genotipo es del tamaño del número de nodos que tiene el grafo y cada locus del genotipo indica una arista del nodo representado por el



**Fig. 2.** En (a) se muestra el caso especial de un agrupamiento donde tenemos  $a(1) = 0$  y  $b(1) = 0$ , es decir, que el nodo 1 no tiene relación con ningún otro, ya sea de su mismo grupo o de otro grupo. Para evitar la división entre cero en  $S(i)$ , se considera implícitamente el bucle  $W_{uu} = 1$ , para todos los nodos de  $G$  como se muestra en (b), en donde tenemos  $a(1) = 0.5$  y  $b(1) = 0$ .

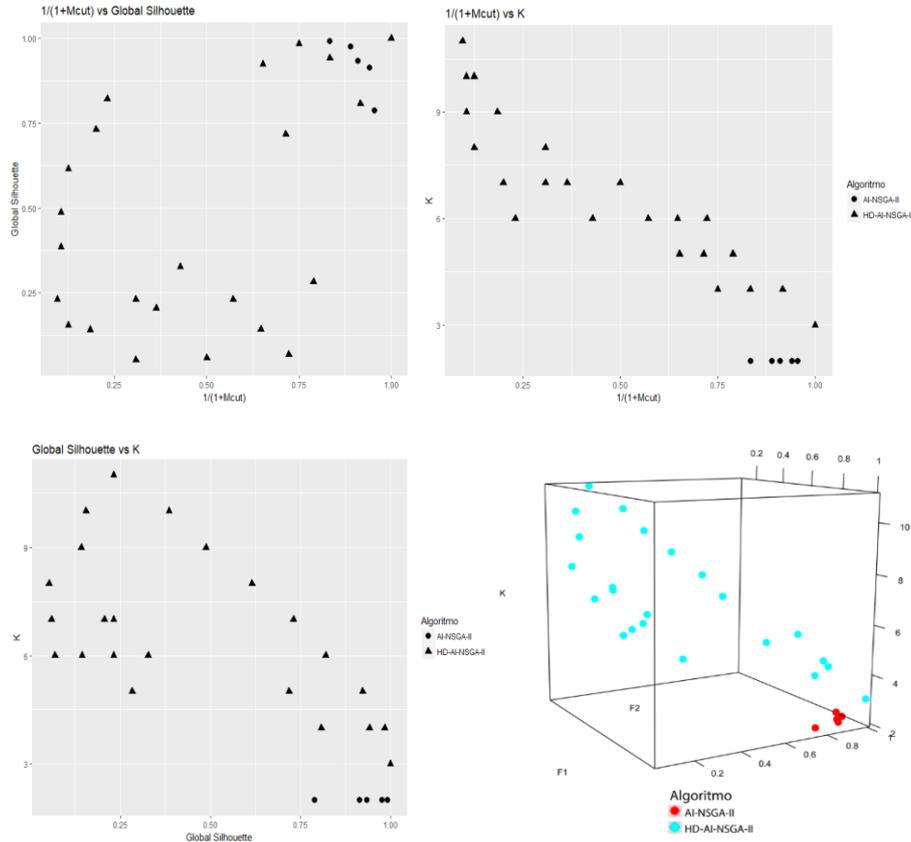


**Fig. 3.** En (a) se muestra la solución de un agrupamiento con valores de aptitud  $f_1 = 1, f_2 = 1$ , la cual forma dos grupos. En (b) se muestra la solución de un agrupamiento que también tiene los valores de aptitud  $f_1 = 1, f_2 = 1$ , pero formada por tres grupos.

locus al nodo representado por el valor del locus. La selección se hace por torneo binario [32], en donde se escogen al azar dos individuos de la población, su nivel en el frente de Pareto y la probabilidad para seleccionar uno de los dos. Se utiliza cruce uniforme por no ser sesgado [24] y poder generar cualquier solución posible.

#### 4. Resultados

Para medir el desempeño de nuestra propuesta, usamos dos conjuntos de datos artificiales y el conjunto bien conocido de datos reales de Cheng et al. [33] obtenidos de YouTube. Los parámetros de ejecución (ver Tabla 1) son los mismos reportados por el enfoque seguido [24].

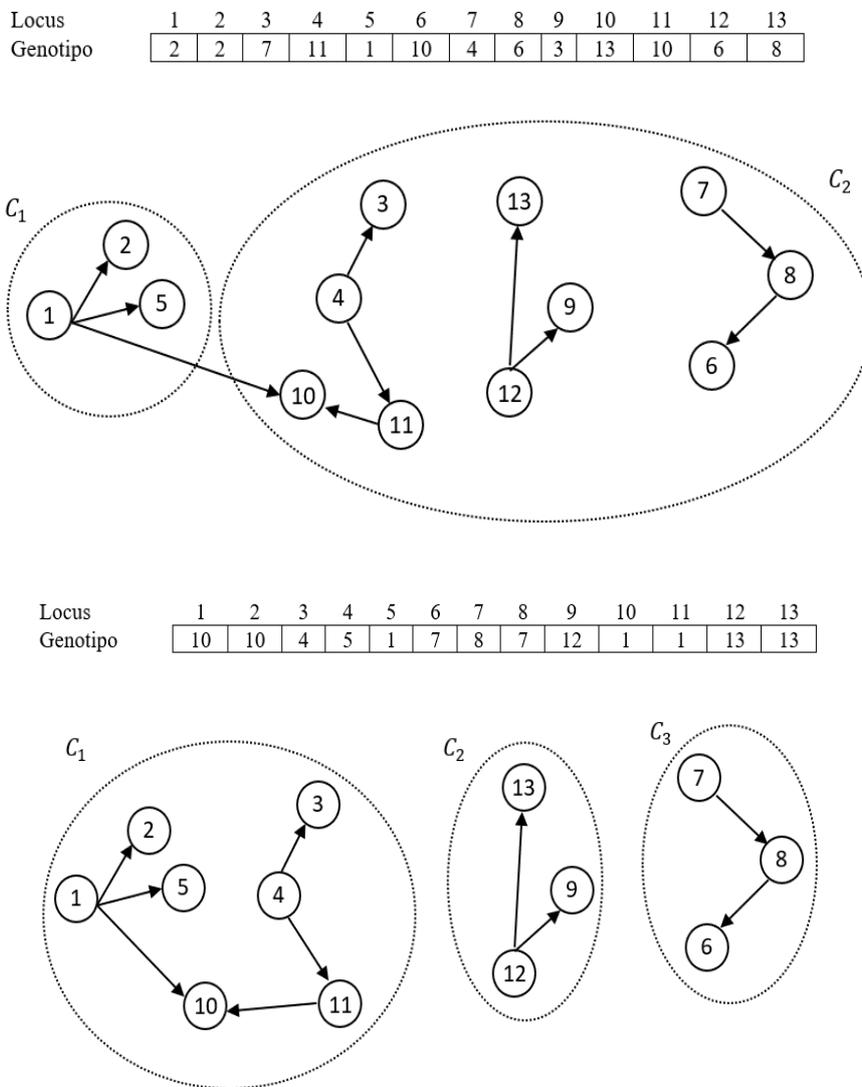


**Fig. 4.** Comparación, por pares de funciones objetivo y de los tres objetivos, de las soluciones de AI-NSGA-II y HD-AI-NSGA-II, para el primer conjunto de datos.

**Tabla 1.** Parámetros de ejecución.

	Número máximo de generaciones	Población P	Población Q	Selección Ps	Cruza Pc	Mutación Pm
AI-NSGA-II	200	30	60	0.6	1.0	0.01
HD-AI-NSGA-II	200	30	60	0.6	1.0	0.01

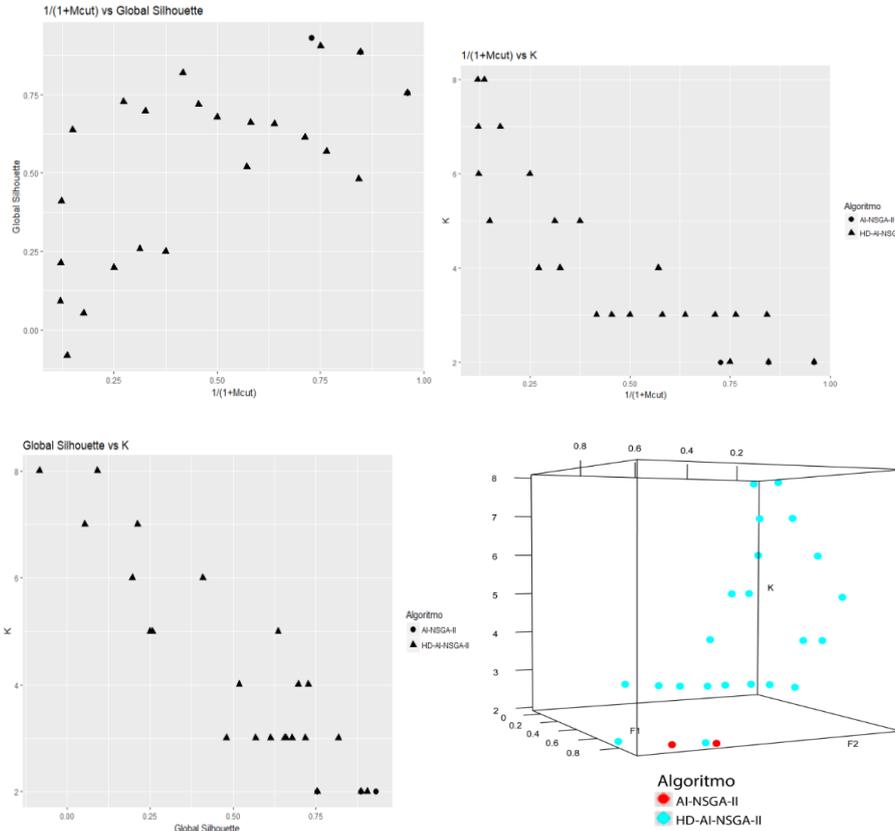
El primer conjunto de datos artificial consta de tres grafos disjuntos, por lo tanto, las soluciones óptimas son las que tengan  $\frac{1}{1+Mcut} = 1$ ,  $GS = 1$  y  $k = 3$ , es decir, las soluciones óptimas deben representar tres grupos sin aristas entre grupos. En la Fig. 4 se puede ver cómo únicamente HD-AI-NSGA-II encuentra las soluciones óptimas para



**Fig. 5.** Superior: la mejor solución de AI-NSGA-II. Inferior: la mejor solución de HD-AI-NSGA-II.

este conjunto de datos, las cuales son precisamente las soluciones  $(\frac{1}{1+Mcut}, GS, k) = (1,1,3)$ .

En la Fig. 4 se puede observar cómo las soluciones de AI-NSGA-II se acercan a valores de  $\frac{1}{1+Mcut} = 1$ ,  $GS = 1$ , las cuales indudablemente alcanzaría aumentando el número de iteraciones, pero se estanca en valores de  $k = 2$ , motivo por el cual no deja que este enfoque obtenga agrupaciones óptimas de  $k = 3$ .

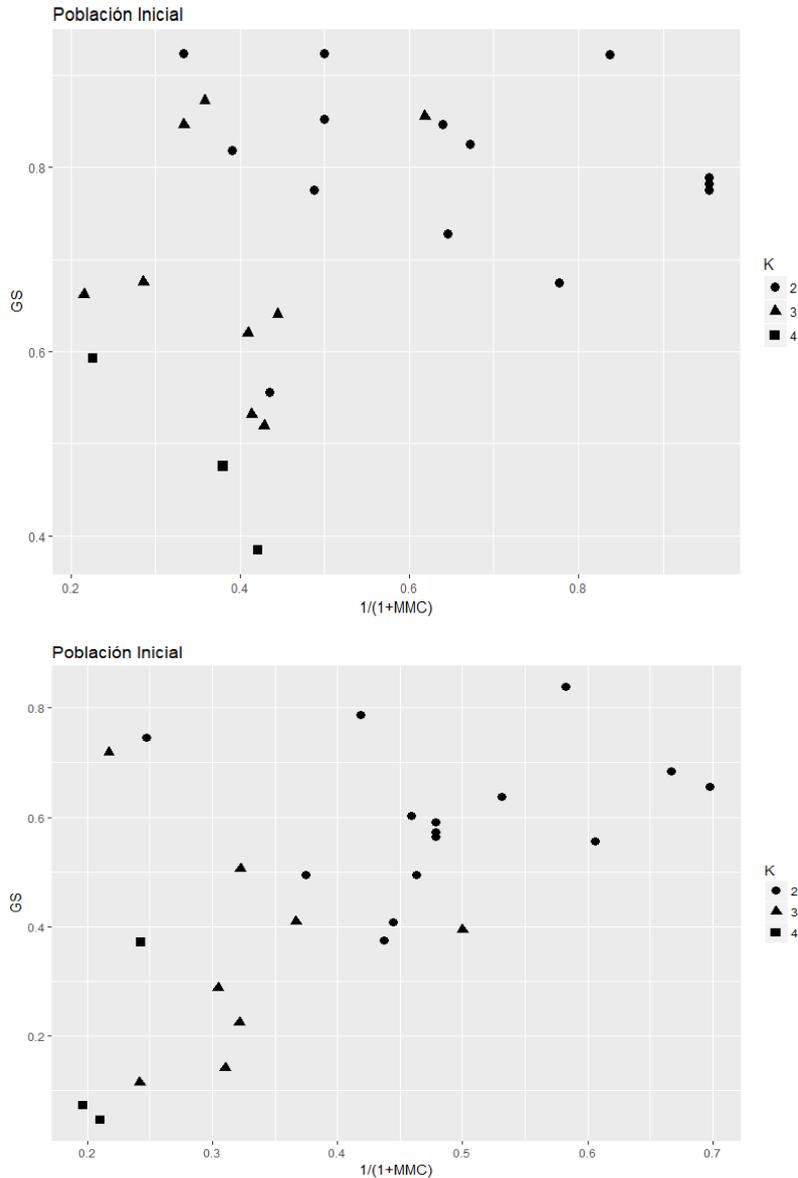


**Fig. 6.** Comparación, por pares de las funciones objetivo y de los tres objetivos, de las soluciones de AI-NSGA-II y HD-AI-NSGA-II, para el segundo conjunto de datos.

En la Fig. 5 se muestra la representación y comparación de la mejor solución del enfoque adoptado AI-NSGA-II y la mejor solución de nuestra propuesta HD-AI-NSGA-II, respectivamente. La mejor solución de AI-NSGA-II únicamente corta una arista, pero al no ser guiada la búsqueda en el sentido del número de grupos  $k$  le es difícil definir los tres grupos que existen. En contraste, nuestra propuesta sí logra identificar perfectamente los tres grupos.

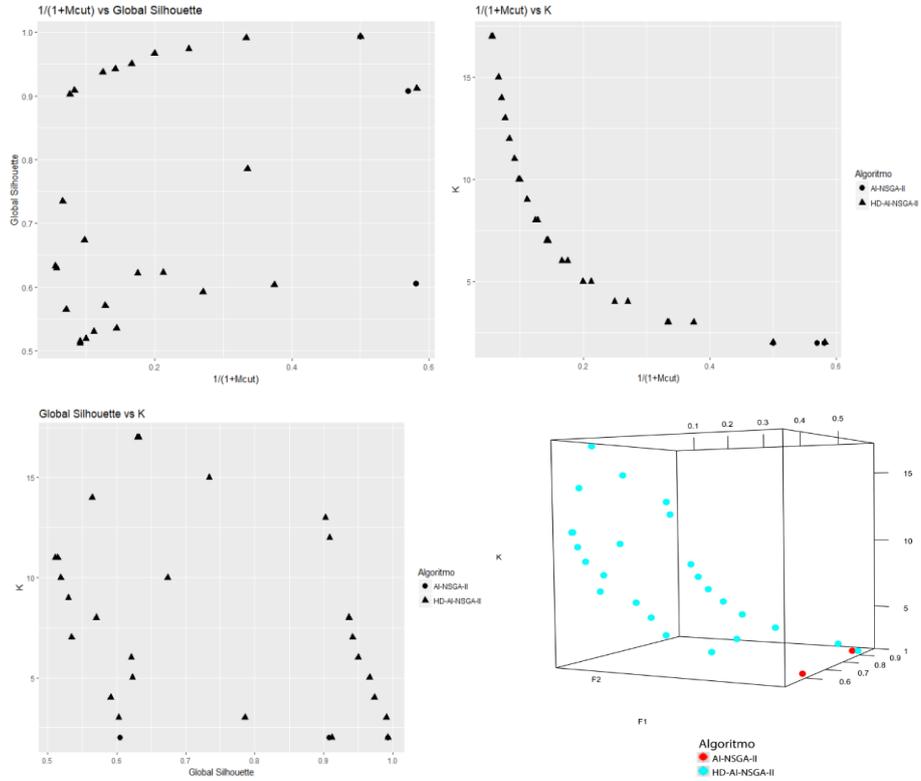
El segundo conjunto de datos artificial consta de un grafo conexo que contiene tres comunidades, por lo tanto, las mejores soluciones son las que tengan  $\frac{1}{1+Mcut} \approx 1$ ,  $GS \approx 1$  y  $k = 3$ . En la Fig. 6 se puede ver que también únicamente HD-AI-NSGA-II encuentra las soluciones óptimas para este tipo de grafo.

En la Fig. 6 se puede observar cómo las soluciones de AI-NSGA-II se acercan a valores de  $\frac{1}{1+Mcut} = 1$ ,  $GS = 1$ , pero se estanca en valores de  $k = 2$  motivo por el cual no deja que este enfoque obtenga agrupaciones óptimas con  $k = 3$ .

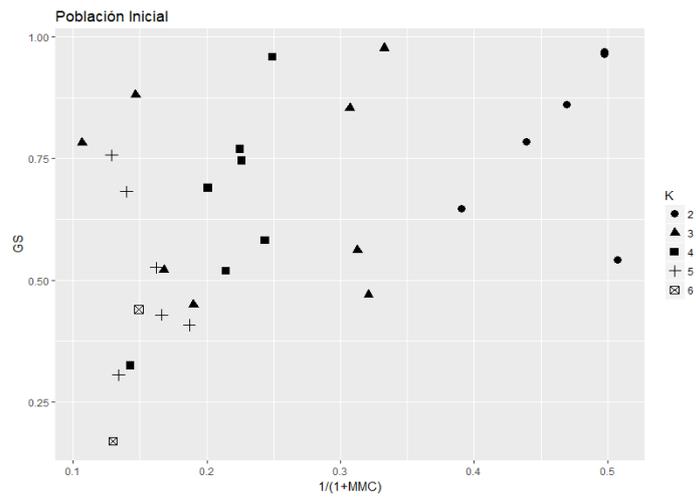


**Fig. 7.** Número de conjuntos en las poblaciones iniciales para el primer (superior) y segundo (inferior) conjunto artificial.

El motivo por el que sólo HD-AI-NSGA-II encuentre soluciones compuestas de tres grupos, se debe a que las soluciones iniciales (ver Fig. 7) con mejores valores de aptitud  $f_1$  y  $f_2$  generalmente son las de menor valor  $f_3$  y, por lo tanto, AI-NSGA-II va reemplazando las soluciones con altos valores de  $f_3$  por soluciones con valores menores para dicha función.



**Fig. 8.** Comparación, por pares de las funciones objetivo y de los tres objetivos, de las soluciones de AI-NSGA-II y HD-AI-NSGA-II para el conjunto de datos de YouTube que contiene tres grafos disjuntos.



**Fig. 9.** Número de conjuntos en las poblaciones iniciales para el conjunto de datos de YouTube.

Por último, en la Fig. 8 se muestran las soluciones encontradas para el conjunto de datos que contienen los ID de videos de YouTube, en el cual se puede apreciar que HD-AI-NSGA-II es el único en encontrar soluciones con un mayor número de grupos.

El motivo por el que solo HD-AI-NSGA-II encuentre soluciones compuestas de tres grupos, se debe a que las soluciones iniciales (ver Fig. 9) con mejores valores de aptitud  $f_1$  y  $f_2$  generalmente son las de menor valor  $f_3$  y, por lo tanto, AI-NSGA-II va reemplazando las soluciones con altos valores de  $f_3$  por soluciones con valores menores para dicha función.

## 5. Conclusiones

En el presente trabajo presentamos una alternativa al enfoque propuesto por AI-NSGA-II para la detección de comunidades. Nuestra propuesta agrega una función objetivo que permite guiar a soluciones que definen mejor los grupos hacia el frente de Pareto. Este objetivo maximiza el número de grupos formados por cada solución, dando como resultado una mayor exploración y evitar estancarse en óptimos locales que contengan un bajo número de grupos. Ambos enfoques fueron probados y comparados con dos conjuntos de datos artificiales y uno bien conocido, llegando a la conclusión de que, tanto para los grafos disjuntos como para los conexos que contienen comunidades internamente, trabaja mejor nuestra propuesta logrando definir mejor los grupos y haciendo una mejor detección de comunidades.

Actualmente se está trabajando con un conjunto de datos compuesto por ID de perfiles de Facebook creado por la empresa tecnológica Cad&Lan, los cuales forman un grafo conexo al estar construido de forma recursiva. Se pretende demostrar de igual manera la efectividad del enfoque propuesto sobre otro conjunto de datos reales. Posteriormente, se pretende adaptar los demás objetivos de optimización e implementar métricas multiobjetivo para conocer el desempeño.

**Agradecimientos.** El presente trabajo está patrocinado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) a través la beca otorgada en el marco del Programa Nacional de Posgrado de Calidad. También agradecemos el apoyo prestado por la empresa tecnológica Cad&Lan que a través de su departamento de investigación y desarrollo nos facilitó su base de datos de Facebook IDs.

## Referencias

1. Menendez, D.H., Llorente, J.L.: The Combination of Graph Theory and Unsupervised Learning Applied to Social Data Mining. In: Cavalcante, A.(eds.) Graph Theory: New Research, Nova Science Publishers Inc (2013)
2. Gong, M., Cai, Q., Ma, L., Wang, S., Lei, Y.: Introduction. In: Computational Intelligence for Network Structure Analytics. Springer (2017)
3. Leu, G., Abbass, H.: A multi-disciplinary review of knowledge acquisition methods: From human to autonomous eliciting agents. Knowledge-Based Systems, 105, pp: 1–22 (2016)

4. Rana, C., Jain, S.K.: An extended evolutionary clustering algorithm for an adaptive recommender system. *Social Network Analysis and Mining*, 4(1) pp. 1–13 (2014)
5. Mukhopadhyay, A., Maulik, U.: Survey of Multi-Objective Evolutionary Algorithms for Data Mining: Part-II. *IEEE Transactions on Evolutionary Computation*, 18(1), pp. 20–35 (2014)
6. Mukhopadhyay, A., Maulik, U.: A Survey of Multi-Objective Evolutionary Algorithms for Data Mining: Part-I. *IEEE Transactions on Evolutionary Computation*, 18(1), pp. 4–19 (2014)
7. Menendez, H.D., Barrero, D.F., Camacho, D.: A genetic graph-based approach for partitional clustering. *International journal of neural systems*, 24(03) (2014)
8. Fan, L., Wu, W., Lu, Z., Xu, W., Du, D.Z.: Influence diffusion, community detection, and link prediction in social network analysis. *Springer Proceedings in Mathematics and Statistics*, 51, pp. 305–325 (2013)
9. Cai, Q., Ma, L., Gong, M., Tian, D.: A survey on network community detection based on evolutionary computation. *International Journal of Bio-Inspired Computation* 8(2), pp. 84–98 (2016)
10. Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S.: A Survey of Multiobjective Evolutionary Clustering. *ACM Computing Surveys* 47(4), pp. 1–46 (2015)
11. Delgado C.C.: Utilización de técnicas de clustering para mejorar la detección de metatopics en conjuntos de datos extraídos de Twitter. MS thesis (2015)
12. Gong, M, et al.: Big Network Analytics Based on Nonconvex Optimization. In: Emrouznejad A. (ed.), *Big Data Optimization: Recent Developments and Challenges. Studies in Big Data*, 18, pp. 345–373 (2016)
13. Bucur, D., Iacca, G., Marcelli, A., Squillero, G., Tonda, A.: Multi-objective Evolutionary Algorithms for Influence Maximization in Social Networks. In: Squillero G., Sim K. (eds.), *Applications of Evolutionary Computation. Evo Applications, Lecture Notes in Computer Science*, 10199, Springer (2017)
14. Kirkland, O.: Multi-objective evolutionary algorithms for data clustering. Diss., University of East Anglia (2014)
15. Delgado C.C.: Modelo de identificación de Meta-Topics a través de análisis semántico de conjuntos de datos extraídos de twitter. BS thesis (2014)
16. Bello-Orgaz, G., Menéndez, H.D., Camacho D.: Adaptive K-Means Algorithm for Overlapped Graph Clustering. *International Journal of Neural Systems*, 22(5), 1250018 (2012)
17. Menéndez, H.D.: A genetic approach to the graph and spectral clustering problem. MS thesis (2012)
18. Li, Q., Zou, J., Yang, S. et al.: A predictive strategy based on special points for evolutionary dynamic multi-objective optimization. *Soft Computing* (2018)
19. Zou, J., Li, Q., Yang, S., Bai, H., Zheng, J.: A prediction strategy based on center points and knee points for evolutionary dynamic multi-objective optimization. *Applied Soft Computing Journal*, 61, pp. 806–818 (2017)
20. Rong, M., Gong, D., Zhang Y.: A Multi-direction Prediction Approach for Dynamic Multi-objective Optimization. In: Huang DS., Han K., Hussain A. (eds) *Intelligent Computing Methodologies, (ICIC) Lecture Notes in Computer Science*, 9773, Springer, Cham (2016)
21. Zadeh, P. M., Kobti, Z.: A multi-population cultural algorithm for community detection in social networks. *Procedia Computer Science* 52(1), pp. 342–349 (2015)
22. Hartmann T., Kappes A., Wagner, D.: Clustering Evolving Networks. In: Kliemann L., Sanders P. (eds), *Algorithm Engineering, Lecture Notes in Computer Science*, 9220 (LNCS), pp. 280–329 (2016)

23. Held, P., Dannies, K.: Clustering on dynamic social network data. *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*. 190, pp. 563–571 (2013)
24. Kim, K., McKay, R.I., Moon, B.R.: Multiobjective evolutionary algorithms for dynamic social network clustering. In: *Proceedings of the 12th Annual Genetic and Evolutionary Computation Conference, (GECCO)*, pp. 1179–1186 (2010)
25. Demir, G., Uyar, A., Oguducu, S.: Graph-based sequence clustering through multiobjective evolutionary algorithms for web recommender systems. In: *Proceedings of Genetic and Evolutionary Computation Conference (GECCO)*, pp.1943–1950 (2007)
26. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: Improving the Strength Pareto Evolutionary Algorithm. *Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems*, pp. 95–100 (2001)
27. Folino, F., Pizzuti, C.: A Multiobjective and Evolutionary Clustering Method for Dynamic Networks. In: *Advances in Social Networks Analysis and Mining (ASONAM), International Conference*, pp. 256–263 (2010)
28. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), pp. 182–197 (2002)
29. Ding, C.H., He, X., Zha, H., Gu, M., Simon, H.D.: A min-max cut algorithm for graph partitioning and data clustering. In: *Data Mining (ICDM), Proceedings IEEE International Conference*, pp. 107–114 (2001)
30. Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), pp. 53–65 (1987)
31. Park, Y., Song, M.: A genetic algorithm for clustering problems. In: *Proceedings of the third annual conference on genetic programming*, pp. 568–575 (1998)
32. Holland, J.H., Goldberg, D.: *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesley (1989)
33. Cheng, X., Dale, C., Liu, J.: Statistics and social network of youtube videos. In: *Quality of Service, IWQoS 16th International Workshop*, pp. 229–238 (2008)