

¿Cómo obtener soluciones heurísticas buenas?
Caso de estudio en problemas de minimización de la
influencia en redes sociales.

**XV Congreso Nacional sobre Metaheurísticas,
Algoritmos Evolutivos y Bioinspirados (MAEB).**

Isaac Lozano-Osorio (isaac.lozano@urjc.es)

Jesús Sánchez-Oro (jesus.sanchezoro@urjc.es)

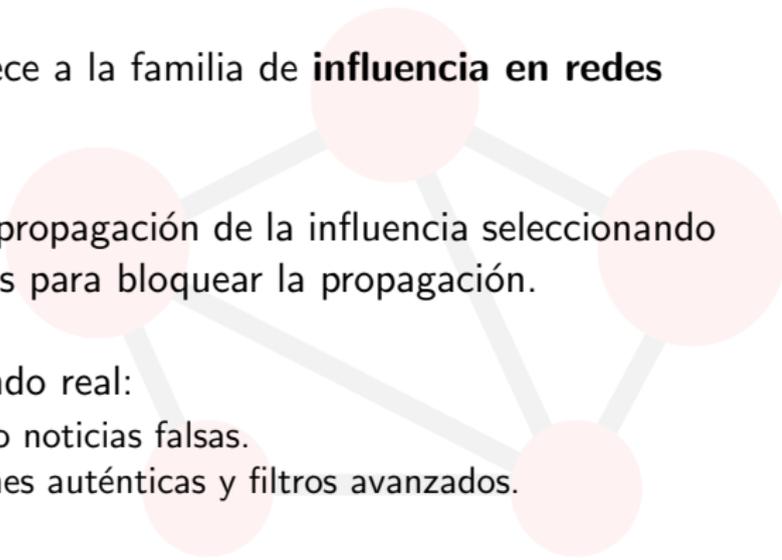
Abraham Duarte (abraham.duarte@urjc.es)

Kenneth Sörensen (kenneth.sorensen@uantwerpen.be)



Introducción

- El **problema** pertenece a la familia de **influencia en redes sociales**.
- Busca **minimizar** la propagación de la influencia seleccionando un conjunto de nodos para bloquear la propagación.
- Aplicaciones del mundo real:
 - Reducir rumores o noticias falsas.
 - Anuncios, opiniones auténticas y filtros avanzados.



Introducción

Representación de la solución

Dado:

- Una Red Social (RS) representada mediante $G = (V, A)$, dónde el conjunto de vértices V representa los usuarios, y un conjunto de aristas A indica la relación entre los usuarios.
- Un conjunto de Nodos Maliciosos (NM), con $|NM| \geq 1$.
- Un modelo de difusión de la influencia μ .

La solución consiste en seleccionar un conjunto de bloqueadores B que será responsable de reducir la propagación de información falsa ($B \subseteq V \setminus NM$, con $|B| = b$), dónde b tiene un tamaño fijo.

El objetivo de PMI es **minimizar** el total de nodos activados siguiendo un Modelo de Difusión de la Influencia (MDI).

Introducción

Modelos de Difusión de la Influencia

La evaluación de la influencia requiere la definición de un MDI. Este modelo se encarga de decidir qué nodos se ven afectados por la información recibida de sus nodos vecinos en la RS.

Los MDIs más extendidos son:

- Modelo de cascada independiente (ICM).
- **Modelo de cascada con pesos (WCM).**
- **Modelo de Tri-Valencia (TV).**

Todos ellos se basan en asignar una probabilidad de influencia a cada arista en la RS.

Introducción

Ejemplo: Problema de Minimización de la Influencia

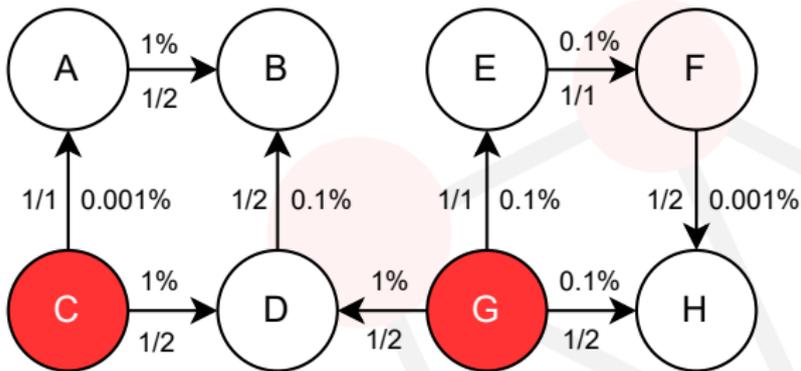


Figura 1: Red social inicial. Los NM están resaltados en rojo.

Para simplificar, se supone que se activarán todos los nodos con un valor de TV igual o superior al 0,1%.

Introducción

Ejemplo: Problema de Minimización de la Influencia

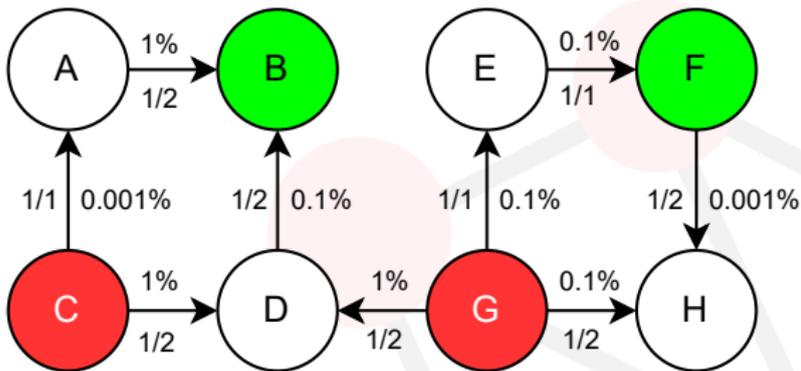


Figura 2: $S_1 = \{B, F\}$ y $PMI(S_1) = 3$.

Para simplificar, se supone que se activarán todos los nodos con un valor de TV igual o superior al 0,1%.

Introduction

Ejemplo: Problema de Minimización de la Influencia

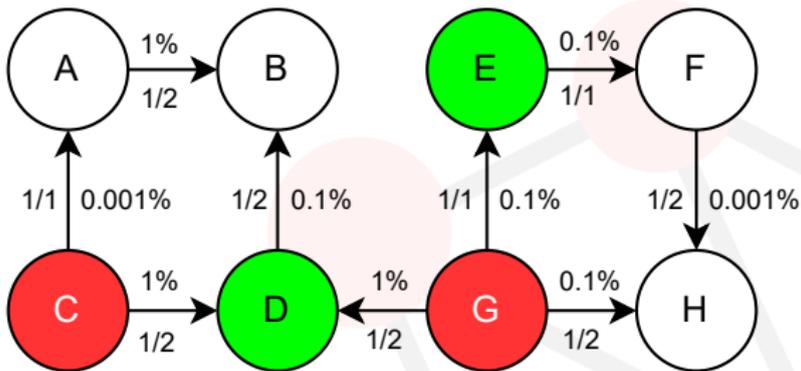


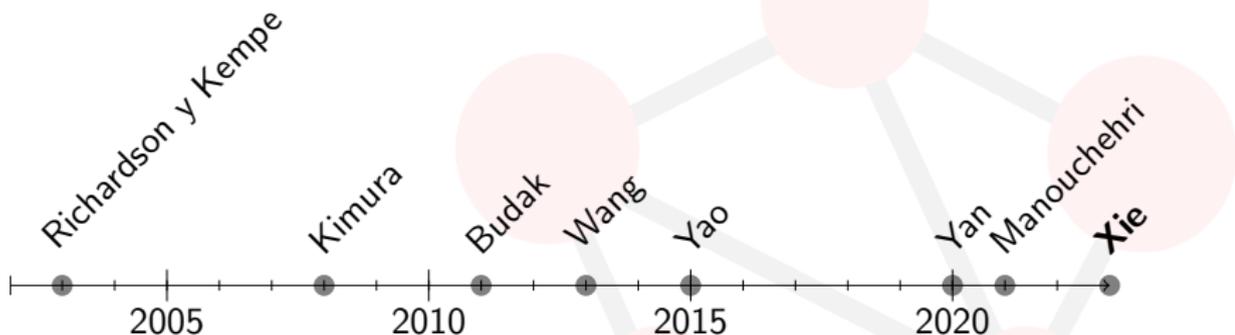
Figura 3: $S_2 = \{E, D\}$, $PMI(S_2) = 1$.

Para simplificar, se supone que se activarán todos los nodos con un valor de TV igual o superior al 0,1%.

Introducción

Revisión de la literatura

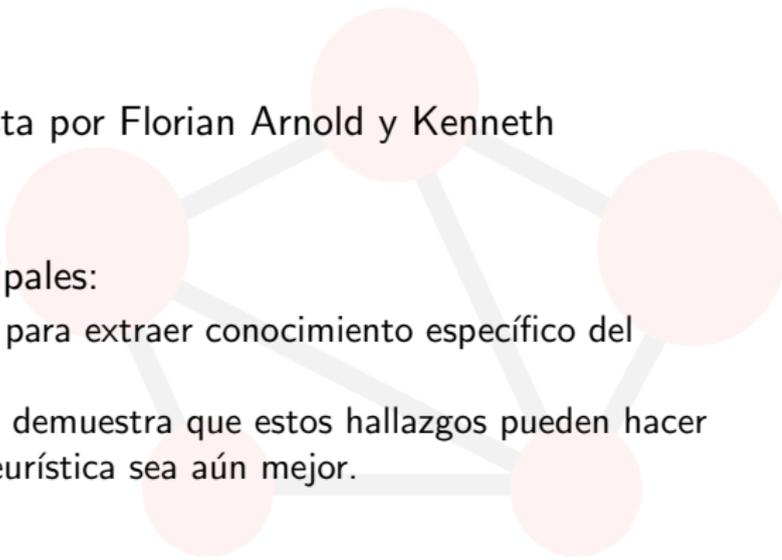
Se ha demostrado ser \mathcal{NP} -hard (Budak, 2011).



Los estudios de revisión del estado del arte destacan que los estudios sobre heurísticas y metaheurísticas son escasos en esta familia de problemas.

De los datos al conocimiento específico del problema

¿Cómo obtener soluciones heurísticas buenas?

- Metodología propuesta por Florian Arnold y Kenneth Sörensen en 2019.
 - Contribuciones principales:
 - 1 Una metodología para extraer conocimiento específico del problema.
 - 2 Un estudio dónde demuestra que estos hallazgos pueden hacer que una buena heurística sea aún mejor.
- 

De los datos al conocimiento específico del problema

Generación de características

Las características seleccionadas están basadas en métricas conocidas de problemas de análisis de redes sociales. Las características necesitan una normalización, para ello se definen los siguientes valores para normalizar:

- Total de nodos (I1).
- Total de aristas (I2).
- Total de componentes conexas (I3).
- Promedio de grado de entrada (I4).
- Promedio de grado de salida (I5).
- Total de comunidades (I6).

De los datos al conocimiento específico del problema

Generación de características

Las características consideradas en este trabajo son las siguientes (entre paréntesis el valor por el que se normalizará):

- S1: nodos influenciados (I1).
- S2: aristas que propagan la influencia (I2).
- S3: número de componentes conexas (I3).
- S4: promedio de grado de entrada (I4).
- S5: promedio de grado de salida (I5).
- S6: número de comunidades influenciadas (I6).
- S7: suma del ranking de B según el grado de salida (I1).
- S8: promedio de probabilidad de activación de los B.
- S9: mínima influencia desde un B a un NM.
- S10: mínima distancia de un B a un NM.

De los datos al conocimiento específico del problema

Conjunto de datos representativo

El conjunto de datos tiene 2 tipos de soluciones: buena calidad (1) y mala calidad (0).

- **Soluciones de mala calidad:** se seleccionan B de forma aleatoria.
- **Soluciones de buena calidad:** se escogen B usando el mejor algoritmo en la literatura con una pequeña aleatorización.

Obsérvese que si se conoce la función objetivo exacta, se pueden generar más clases, por ejemplo, según la desviación respecto a los valores exactos.

De los datos al conocimiento específico del problema

Clasificación

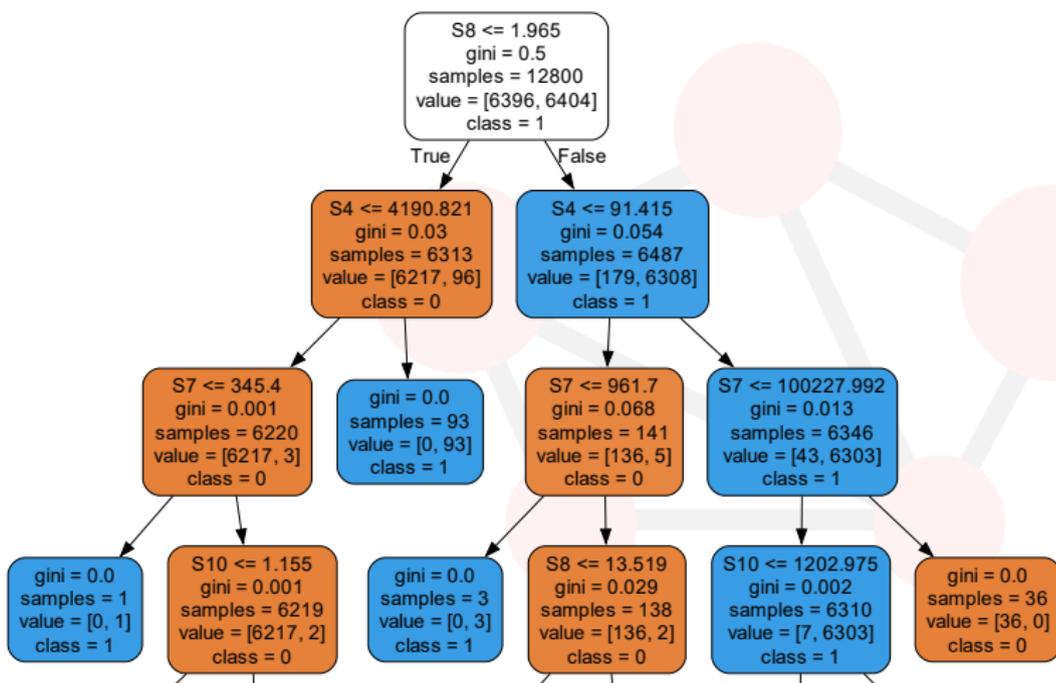


Figura 4: Árbol de decisión.

De los datos al conocimiento específico del problema

Propuesta heurística

La precisión de predicción obtenida con el clasificador de árbol de decisión es **99,87%**.

Promedio de influencia de los vecinos del nodo bloqueador (S8) y promedio de grado de entrada (S4) son las características claves para el PMI.

La heurística propuesta utiliza estas características como criterio voraz para generar soluciones de alta calidad.

B con una suma media de probabilidad de propagación de sus vecinos mayor que 1,965 ordenados de forma decreciente por grado de entrada.

Conclusiones

Entorno experimental

- Lenguaje de programación: **Java 17** y **Python 3.10**.
- *Metaheuristic Optimization framework* (MORK) 13.
- Características del servidor: AMD EPYC 7282 16 cores virtuales CPU con 32GB de RAM.
- Instancias: 8 por cada MDI (las mismas que el estado del arte cuyo rango está entre 4039 y 1134890 nodos).
- Métricas:
 - **Promedio**: valor de la función objetivo.
 - **Tiempo (s)**: tiempo de ejecución en segundos.
 - **Des. (%)**: desviación media de la mejor solución conocida.
 - **#B**: veces que el algoritmo es capaz de alcanzar la mejor solución en el experimento.

Resultados

Comparación entre el estado del arte y nuestro enfoque heurístico

		GR				g_{in}			
b		Promedio	Tiempo (s)	Des. (%)	#B	Promedio	Tiempo (s)	Des. (%)	#B
WCM (1)	20	10077,13	28,40	0,02	7	10077,02	28,34	0,00	8
	40	9750,55	55,49	0,01	7	9750,50	55,80	0,03	7
	60	9513,50	81,54	0,34	6	9512,33	81,99	0,05	7
	80	9301,72	106,63	0,44	6	9299,53	107,71	0,03	7
	100	9137,11	131,14	0,91	5	9132,98	132,87	0,00	8
		9556,00	80,64	0,34	31	9554,47	81,34	0,02	37
TV (2)	20	15716,42	211,56	0,00	8	15716,42	199,17	0,00	8
	40	14979,66	413,64	0,00	8	14979,66	387,90	0,00	8
	60	14465,62	604,81	0,00	8	14465,62	569,66	0,00	8
	80	13411,32	781,76	0,01	7	13411,30	739,35	0,00	8
	100	13004,12	952,28	0,00	8	13004,21	899,81	0,01	7
		14315,43	592,81	0,00	39	14315,44	559,18	0,00	39

Tabla 1: Comparación entre el estado del arte y nuestro enfoque heurístico basado en las mejores características identificadas por un algoritmo supervisado.

Conclusiones



El algoritmo propuesto obtiene resultados competitivos en comparación con el **estado del arte**.



Trabajo futuro: añadir algoritmos metaheurísticos y otros métodos de aprendizaje supervisado para profundizar en la comprensión de las soluciones.

¿Cómo obtener soluciones heurísticas buenas?
Caso de estudio en problemas de minimización de la
influencia en redes sociales.

**XV Congreso Nacional sobre Metaheurísticas,
Algoritmos Evolutivos y Bioinspirados (MAEB).**

Isaac Lozano-Osorio (isaac.lozano@urjc.es)

Jesús Sánchez-Oro (jesus.sanchezoro@urjc.es)

Abraham Duarte (abraham.duarte@urjc.es)

Kenneth Sörensen (kenneth.sorensen@uantwerpen.be)

