

Path Relinking para el problema Target Set Selection

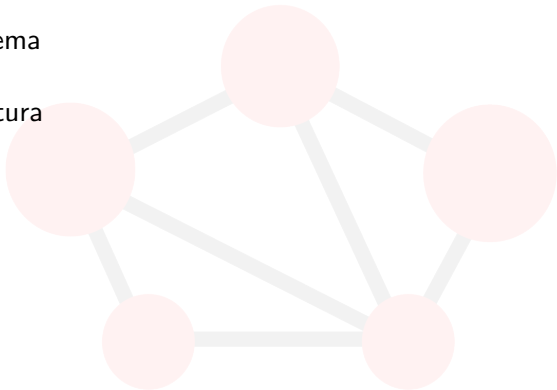
XL Congreso Nacional de Estadística e Investigación Operativa

Isaac Lozano-Osorio (isaac.lozano@urjc.es)
Andrea Oliva-García (andrea.oliva@urjc.es)
Jesús Sánchez-Oro (jesus.sanchezoro@urjc.es)



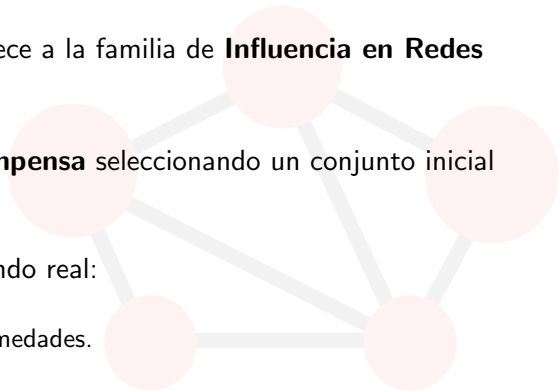
Índice

- 1 Introducción**
 - Definición del problema
 - Target Set Selection
 - Revisión de la literatura
- 2 Path Relinking**
 - Constructivo
 - Búsqueda Local
- 3 Resultados**
- 4 Conclusiones**



Introducción

- El **problema** pertenece a la familia de **Influencia en Redes Sociales**.
- **Maximizar** la **recompensa** seleccionando un conjunto inicial de nodos.
- Aplicaciones del mundo real:
 - Marketing Viral.
 - Análisis de enfermedades.



Introducción

Representación de la solución

Definimos x_v^t como una variable binaria que toma el valor 1 si el nodo v se activa en el paso t y 0 en caso contrario.

$$x_v^{t-1} \leq x_v^t, \quad \forall v \in V, 1 \leq t \leq T$$

T indica el número máximo de pasos en el proceso de propagación de la influencia, es decir, el número de iteraciones en las que se activan nuevos nodos.

La solución consiste en seleccionar conjuntos de semillas donde la suma de los esfuerzos de los nodos activados inicialmente es menor o igual que K , es decir:

$$\sum_{v \in V} \alpha(v) \cdot x_v^0 \leq K$$

Introducción

Función Objetivo

Un conjunto de usuarios activados S^t influye en un usuario no activado u si y solo si la suma de la influencia potencial de todos los usuarios en S^t es mayor o igual a 1:

$$\sum_{v \in S^t} \psi(v, u) \geq 1$$

El objetivo del TSS es **maximizar** las recompensas que devuelve cada uno de los nodos activos en la última iteración del proceso de propagación de influencias.

$$TSS(S) = \sum_{v \in V} \beta(v) \cdot x_v^T$$

Introducción

Función Objetivo

El TSS busca entonces una solución S^* con el máximo valor de la función objetivo.

$$S^* \leftarrow \arg \max_{S \in \mathbb{S}} TSS(S)$$

donde \mathbb{S} representa el conjunto de todas las soluciones factibles, es decir, todas las posibles combinaciones de nodos cuya suma de esfuerzo es menor o igual a K .

Introducción

Ejemplo del Target Set Selection

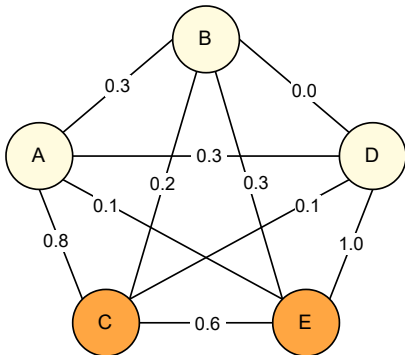


Figura 1: Paso I: solución original con $S = C, E$.

Introducción

Ejemplo del Target Set Selection

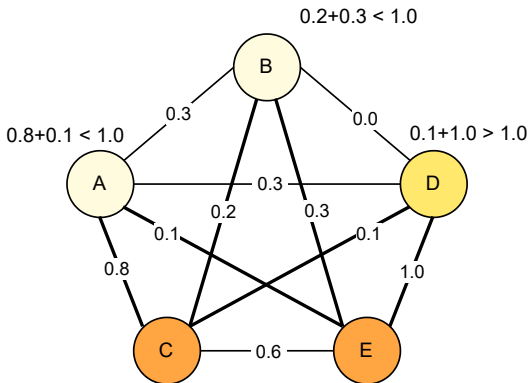


Figura 2: Paso II: solución original con $S = C, E, D$.

Introducción

Ejemplo del Target Set Selection

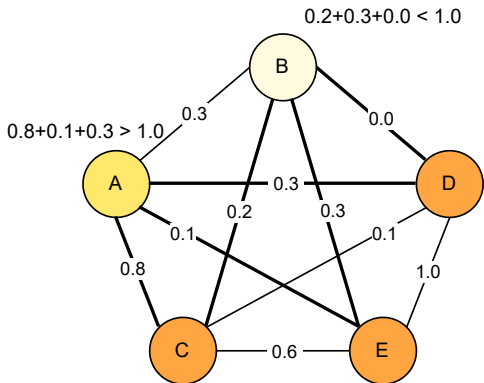


Figura 3: Paso III: solución original con $S = C, E, D, A$.

Introducción

Ejemplo del Target Set Selection

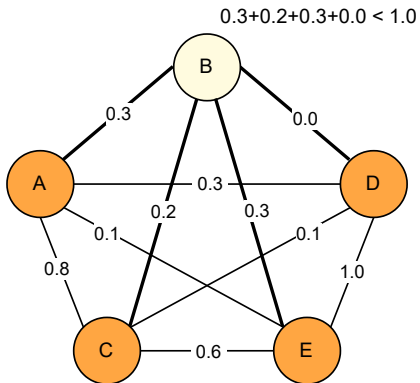


Figura 4: Paso IV: solución original con $S = C, E, D, A$.

Introducción

Revisión de la Literatura

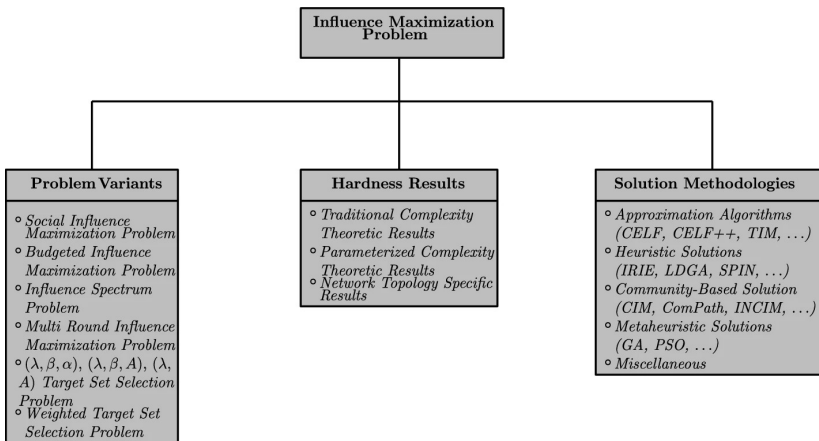
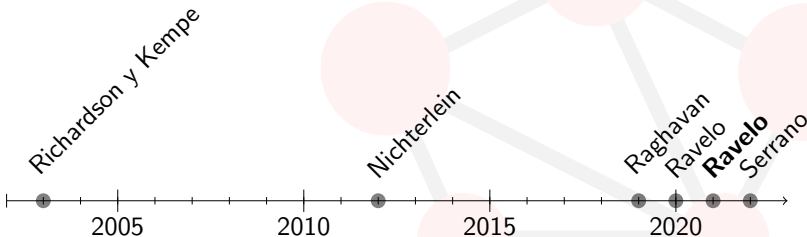


Figura 5: Un estudio sobre la maximización de la influencia en una red social, Banerjee et al (2020).

Introducción

Revisión de la literatura

Es una variante del problema clásico Target Set Selection (TSS).
Se ha demostrado ser \mathcal{NP} -hard (Ravelo, 2020).



Diferentes estudios resaltan escasos estudios sobre heurísticas y metaheurísticas en este tipo de problemas.

Propuesta

Path Relinking (PR)

- Propuesto por Fred Glover y Pierre Hansen en 1997.
- Contribuciones principales:
 - 1 Algoritmo de búsqueda que combina soluciones de calidad similar para generar nuevas soluciones potencialmente mejores.
 - 2 Es capaz de combinar diferentes heurísticas o métodos de búsqueda en una sola estructura, lo que le brinda flexibilidad y adaptabilidad para abordar una amplia gama de problemas de optimización.
 - 3 Utiliza información de soluciones previas para guiar la búsqueda, lo que permite ir hacia regiones prometedoras del espacio de búsqueda.

Propuesta

Static Path Relinking (SPR)

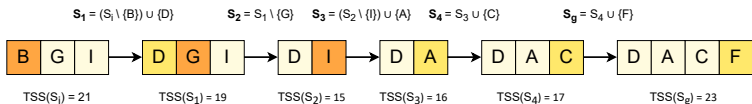


Figura 6: Ejemplo entre caminos de soluciones $S_i = \{B, G, I\}$ y $S_g = \{D, A, C, F\}$.

Propuesta

Constructivo

GRASP + PR (Laguna y Marti) basado en el **grado** de los nodos donde resaltan las optimizaciones realizadas.

- Factorización de la función objetivo donde se reduce de $O(N^2)$ a $O(N)$.
- Evitar soluciones ya exploradas.
- Limitar los nodos en la búsqueda local, de tal forma que, si se ordenan y no cumplen el presupuesto, se ignoran.

Propuesta

Búsqueda Local - Movimiento

Definimos el operador de movimiento propuesto en dos pasos: eliminar y añadir.

El operador de movimiento elimina un nodo de la solución y, a continuación, añade nodos de forma iterativa hasta alcanzar el presupuesto máximo:

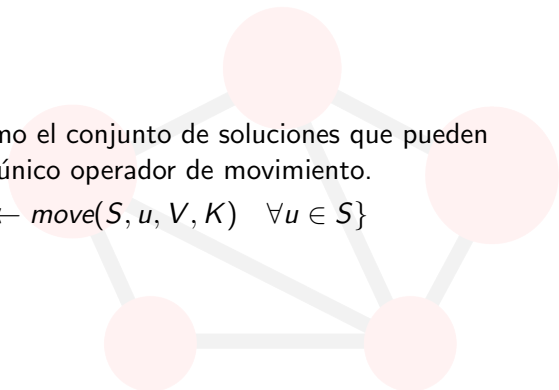
$$\text{move}(S, u, V, K) = (S \setminus \{u\}) \cup \left\{ v \in V \setminus (S \cup \{u\}) : \sum_{s \in S \cup \{u\}} \alpha(s) + \alpha(v) \leq K \right\} \quad (1)$$

Propuesta

Búsqueda Local - Definición de la vecindad

La vecindad se define como el conjunto de soluciones que pueden alcanzarse realizando un único operador de movimiento.

$$N(S) \leftarrow \{S' \leftarrow \text{move}(S, u, V, K) \quad \forall u \in S\}$$



Resultados

- Lenguaje de programación: **Java 17**.
- *Metaheuristic Optimization framewoRK* (MORK) 13.
- Características de la máquina experimental: AMD EPYC 7282 16 cores virtuales CPU con 32GB de RAM.
- Instancias: 90 (ampliadas, las instancias preliminares tenían *nodos* ≤ 53 , ahora la más grande supera los 10.000 nodos).
- Métricas de rendimiento:
 - **Avg.**: el valor de la función objetivo.
 - **Gap.**: medida para evaluar la calidad de una solución en comparación con el valor óptimo conocido.
 - **Time (s)**: tiempo de ejecución medido en segundos.
 - **#Exacto**: el número de veces que el algoritmo es capaz de alcanzar el valor exacto en el experimento.

Resultados

Instancias del previo que se conoce el exacto

Algoritmo	Avg.	Gap. (%)	Time (s)	#Exacto
Gurobi	45.38	0.00	117.14	82
SPR	44.54	1.82	0.01	79

Tabla 1: Comparación de SPR y el solver de Gurobi cuando se considera el conjunto de datos original en el que Gurobi es capaz de alcanzar el valor óptimo.

Resultados

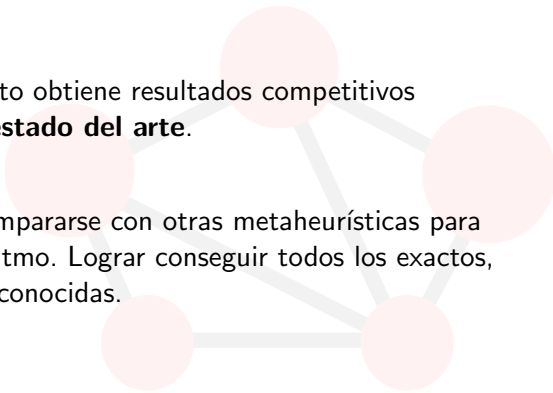
Nuevas instancias

Instancia	Gurobi				SPR			
	Avg.	Dev. (%)	Time (s)	#Mejor	Avg.	Dev. (%)	Time (s)	#Mejor
PRISON	270	0.00	0.16	1	270	0.00	0.07	1
EMAIL-EU-CORE	4757	0.00	262.59	1	4757	0.00	84.48	1
EGO-FACEBOOK	19462	0.00	769.70	1	19462	0.00	279.47	1
CA-GRQC	-	-	-	-	23630	0.00	442.07	1
TWITCH_EN	-	-	-	-	24853	0.00	680.77	1
LASTFM_ASIA	-	-	-	-	24556	0.00	760.97	1
CA-HEPTH	-	-	-	-	44909	0.00	2140.92	1
BLOG_CATALOG3	-	-	-	-	46595	0.00	1336.91	1
Resumen	-	-	-	3	23629.00	0.00	715.71	8

Tabla 2: Comparación de Gurobi y SPR sobre el conjunto de instancias más grande.

El símbolo - denota que la instancia no pudo ser cargada en memoria ni obtener ningún valor por Gurobi.

Conclusiones

- 
- ✓ El algoritmo propuesto obtiene resultados competitivos comparando con el **estado del arte**.
 - 🕒 **Trabajo Futuro:** compararse con otras metaheurísticas para validar nuestro algoritmo. Lograr conseguir todos los exactos, frente las instancias conocidas.

Path Relinking para el problema Target Set Selection

XL Congreso Nacional de Estadística e Investigación Operativa

Isaac Lozano-Osorio (isaac.lozano@urjc.es)
Andrea Oliva-García (andrea.oliva@urjc.es)
Jesús Sánchez-Oro (jesus.sanchezoro@urjc.es)



Path Relinking para el problema Target Set Selection

XL Congreso Nacional de Estadística e Investigación Operativa

Isaac Lozano-Osorio (isaac.lozano@urjc.es)
Andrea Oliva-García (andrea.oliva@urjc.es)
Jesús Sánchez-Oro (jesus.sanchezoro@urjc.es)



Path Relinking para el problema Target Set Selection

XL Congreso Nacional de Estadística e Investigación Operativa

Isaac Lozano-Osorio (isaac.lozano@urjc.es)
Andrea Oliva-García (andrea.oliva@urjc.es)
Jesús Sánchez-Oro (jesus.sanchezoro@urjc.es)



Dynamic Path Relinking

Combina creando dinámicamente nuevas soluciones y caminos entre ellas.

Instancia	DPR				SPR			
	Avg.	Dev. (%)	Time (s)	#Mejor	Avg.	Dev. (%)	Time (s)	#Mejor
PRISON	270	0.00	0.16	1	270	0.00	0.07	1
EMAIL-EU-CORE	4757	0.00	262.59	1	4757	0.00	84.48	1
EGO-FACEBOOK	19462	0.00	769.70	1	19462	0.00	279.47	1
CA-GRQC	23684	0.00	2166.29	1	23630	0.23	442.07	0
TWITCH_EN	25116	0.00	2230.12	1	24853	1.06	680.77	0
LASTFM_ASIA	24780	0.00	2409.55	1	24556	1.80	760.97	0
CA-HEPTH	44972	0.00	7743.06	1	44909	0.14	2140.92	0
BLOG_CATALOG3	46692	0.00	8705.44	1	46595	0.29	1336.91	0
Resumen	23716.63	0.00	3035.87	8	23629.00	0.44	715.71	3

Tabla 3: Comparación de DPR y SPR sobre el conjunto de instancias más grande.