What characteristics define a Good Solution in Social Influence Minimization Problems?

15th Metaheuristics International Conference

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)

ANT/OR
University of Antwerp
Operations Research Group

MIC 2024

grafo
RESEARCH

MINISTERIO
DE CIENCIA
E INNOVACIÓN
AGENCIA
ESTATAL DE
INVESTIGACIÓN

Table of contents

Introduction

- The **problem** belongs to the family of **Social Network Influence**.
- **Minimize** the influence propagation by selecting a set of nodes that block propagation.
- Real-world applications:
  - Reduce rumors or fake news.
  - Advertising, authentic reviews, and advanced filters.

Introduction

Solution Representation

Given a:

- Social Network (SN), $G = (V, E)$, where the set of vertices V represents the users, and the set of edges E indicates the relations among users in the SN.
- A set of Malicious Nodes ($\texttt{MN}$), with $|\texttt{MN}| \geq 1$.
- A certain diffusion model μ.

The solution consists of selecting a set of blockers $\texttt{B}$ will be responsible for reducing the propagation of misinformation ($\texttt{B} \subseteq \texttt{V} \setminus \texttt{MN}$, with $|\texttt{B}| = \texttt{b}$), where $\texttt{b}$ is a fixed constraint
The aim of IMP is to **minimize** the number of active nodes following a specific Influence Diffusion Model (IDM).

Introduction

Influence Diffusion Model

Evaluation of the influence requires the definition of an Influence Diffusion Model (IDM). This model is responsible for deciding which nodes are affected by the information received from their neighboring nodes in the SN.

The most extended IDMs are:

- Independent Cascade Model (ICM)
- **Weighted Cascade Model (WCM)**
- **Tri-Valency Model (TV)**

All of them are based on assigning an influence probability to each relational link in the SN.

Introduction

Influence Diffusion Model

- ICM, which is one of the most widely used IDMs, considers that the probability of influence is the same for each link and is usually a small probability, being 1% a widely accepted value.
- WCM considers that the probability that a user v will be influenced by an user u is proportional to the in-degree of user v, i.e., the number of users that can eventually influence the user v. Therefore, the probability of influencing the user v is defined as $1/d_{in}(v)$, where $d_{in}(v)$ is the in-degree of user v.
- TV randomly selects the edge probabilities from the set $(1\%, 0.1\%, 0.001\%)$.

Introduction
Objective Function

Given a solution S, the objective function of IMP is evaluated as follows:

$$IMP(S) \leftarrow \underset{\mathtt{B} \subseteq \mathtt{V} \setminus \mathtt{MN}}{\arg\min}\ \varphi_\mu(G, \{\mathtt{MN}, \mathtt{B}\}) \quad (1)$$

where $\varphi_\mu(G, \{\mathtt{MN}, \mathtt{B}\})$ represents the spread ability of $\mathtt{MN}$ when both sets $\mathtt{MN}$ and $\mathtt{B}$ spread two opposite messages.

Introduction

Example: Influence Minimization Problem

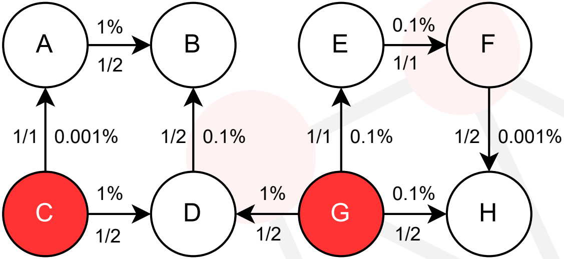


Figure 1: Initial Social Network. MN are highlighted in red.

For the sake of simplicity, it is assumed that all nodes with 0.1% or more TV value will be activated.

Introduction

Example: Influence Minimization Problem

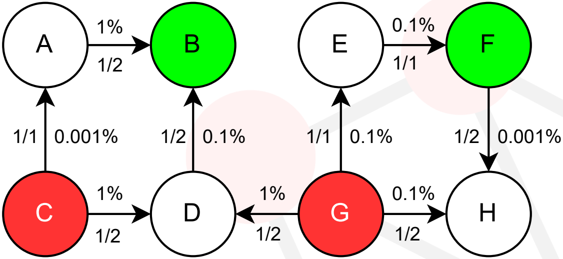


Figure 2: $S_1 = \{\texttt{B},\texttt{F}\}$ and $IMP(S_1) = 3$.

For the sake of simplicity, it is assumed that all nodes with 0.1% or more TV value will be activated.

Introduction

Example: Influence Minimization Problem

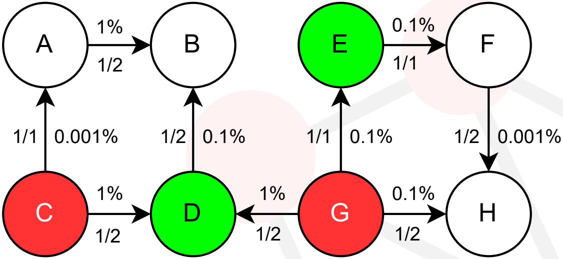


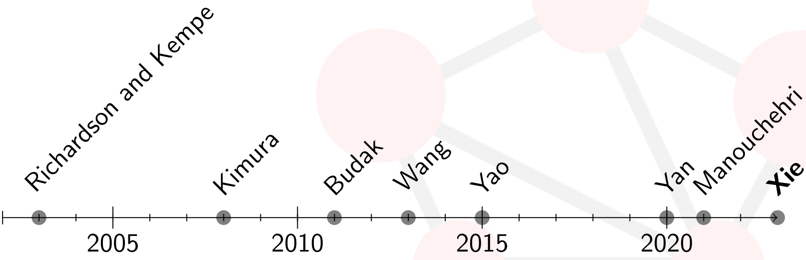
Figure 3: $S_2 = \{\mathtt{E},\mathtt{D}\}$, $IMP(S_2) = 1$.

For the sake of simplicity, it is assumed that all nodes with 0.1% or more TV value will be activated.

Introduction

Literature Review

It has been shown to be $\mathcal{NP}$-hard (Budak, 2011).



Surveys highlight that studies on heuristics and metaheuristics are scarce in this family of problem.

From data to problem-specific knowledge

What distinguishes a good from a bad solution?

- Methodology proposed by Florian Arnold and Kenneth Sörensen in 2019.
- Main contributions:
  1. A framework to derive problem-specific knowledge.
  2. A case study demonstrates that these findings can make a good heuristic even better.

From data to problem-specific knowledge

Features Generation

The features selected based on well-known metrics in Social Network Analysis. Features are dependent on the instance normalization is needed. The following metrics have been used for normalization:

- Number of nodes (I1)
- Number of edges (I2)
- Number of total connected components (I3)
- Average in-degree (I4)
- Average out-degree (I5)
- Number of communities (I6)

From data to problem-specific knowledge

Features Generation

The features considered to characterize a solution in this work are the following (normalization metric between parentheses):

- S1: nodes that can be influenced (I1)
- S2: edges that can propagate influence (I2)
- S3: total connected components influenced (I3)
- S4: average sum of blockers in-degree (I4)
- S5: average sum blockers out-degree (I5)
- S6: total number of communities influenced (I6)
- S7: sum of blockers ranking in the instance according to its out-degree (I1)
- S8: average sum of blockers probability to neighbors
- S9: average activation probability to `MN` from blockers
- S10: sum of distance to `MN` from blockers

From data to problem-specific knowledge

Representative Dataset Generation

The dataset contains two classes of solutions: good (1) and bad quality solutions (0).

- **Bad quality solutions**: are created by a random blockers selection.
- **Good-quality solutions**: using the best algorithm in the literature that includes some randomization.

Notice that if the exact objective function is known, more classes can be generated, for instance, according to deviation to the exact values.

From data to problem-specific knowledge

Dataset Classification Results

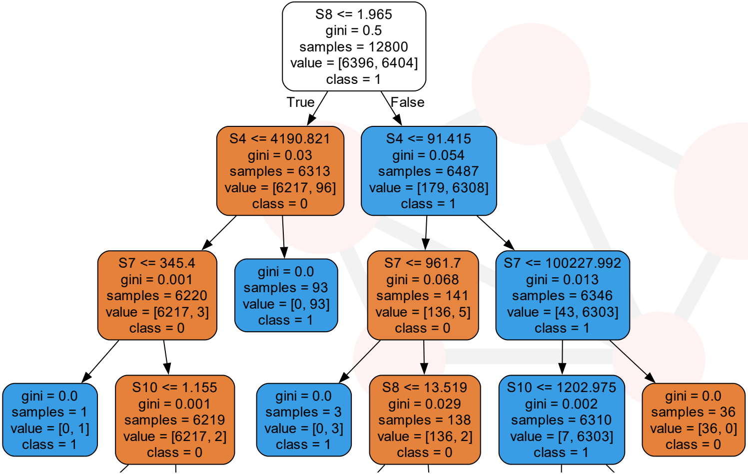


Figure 4: Decision Tree.

From data to problem-specific knowledge

Heuristic Proposal

The prediction accuracy obtained with decision tree classifier is **99.87%**.

Average sum of the propagation blockers neighbors (S8) and the average sum blockers in-degree (S4) are the key features for IMP.

The proposed heuristic uses this features as the greedy criterion to generate high-quality solutions.

Blockers with more than 1.965 average sum of the propagation blockers neighbors in decreasing order by in-degree.

Results

Experimental environment

- Programming language: **Java 17** and **Python 3.10**.
- *Metaheuristic Optimization framewoRK* (MORK) 13.
- Features of the experimental machine: AMD EPYC 7282 16 virtual cores CPU with 32GB of RAM.
- Instances: 8 per each IDM (state-of-the-art instances range between 4039 and 1134890 nodes).
- Performance metrics:
  - **Avg.**: objective function value.
  - **Time (s)**: run time measured in seconds.
  - **Dev.**: average deviation from the best known solution.
  - **#B**: times that the algorithm is able to reach the best solution in the experiment.

Results

Competitive test between state-of-the-art and our heuristic approach

		GR				g_{in}			
	b	Avg.	Time (s)	Dev.	#B	Avg.	Time (s)	Dev.	#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**

Table 1: A comparison between state-of-the-art method and a heuristic approach based on the best features identified by a supervised algorithm.

Conclusions

- The proposed algorithm obtains competitive results compared to the **state of the art**.
- **Future work**: further exploration of metaheuristic algorithms and additional supervised learning algorithms to deepen the understanding of the solutions.

What characteristics define a Good Solution in Social Influence Minimization Problems?

15th Metaheuristics International Conference

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)