Optimization models for crime analytics applied to social networks

Richard Weber[a] Fredy Troncoso[b] Alex Barrales-Araneda[b]

[a]Department of Industrial Engineering, FCFM, University of Chile
[b]Department of Industrial Engineering, University of Bío-Bío

July 10, 2022. Santiago, Chile

IFORS2023
The 23rd Conference of the International Federation of Operational Research Societies
July 10-14 · SANTIAGO, CHILE

IFORS International Federation of Operational Research Societies

ICHIO INSTITUTO CHILENO DE INVESTIGACIÓN OPERATIVA

alio asociación latino iberoamericana de investigación operativa

UNIVERSIDAD DE CHILE

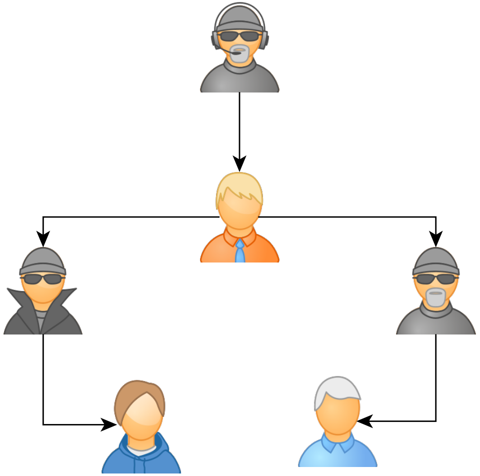
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

ISCI INSTITUTO SISTEMAS COMPLEJOS DE INGENIERÍA

Outline

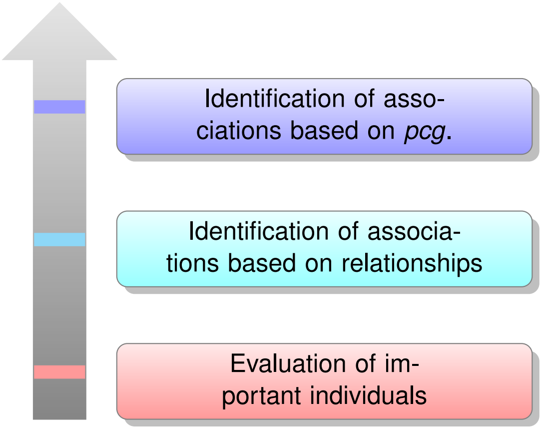
1. Criminal investigation usually involves the mobilization of large amounts of human and technical resources in search of persons responsible for a crime.
2. Knowledge about the structure and organization of criminal networks is important both for investigation and for the development of effective crime prevention strategies[4].

1. A criminal group can be understood as a social network in which the nodes represent the suspects and the arcs are the links between individuals.
2. These links act as channels for the transfer or flow of material and/or immaterial resources. [1]



Extraction of information from social networks

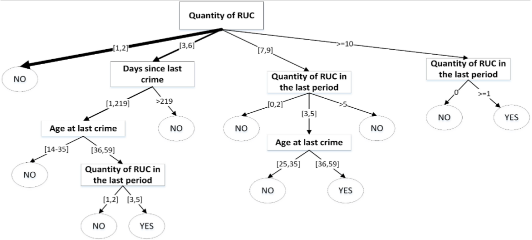
1. There are two approaches to extracting information from social networks: the evaluation of important individuals and the identification of associations.
2. The main evaluators of individuals are derived from Social Network Analysis (SNA). [1].
3. The techniques for identifying associations present in the literature are based on relationships between individuals (SPA).



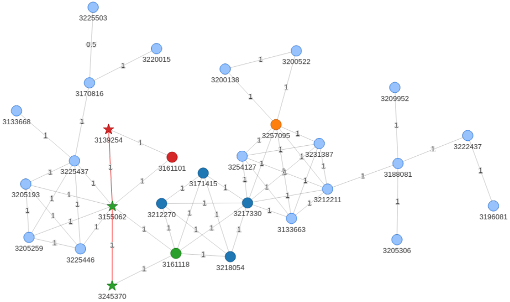
1. The indicator pcg_i is introduced as the propensity of each individual $i \in N$ to belong to a criminal group.
2. The general form of estimating pcg_i is given by:

$$pcg_i = f(s_i), \forall i \in N$$

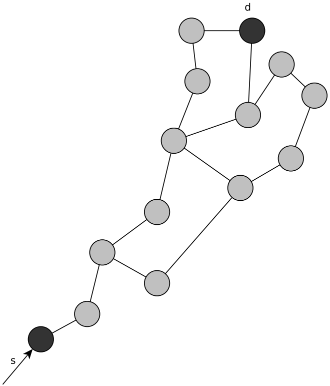
where s_i is the set of relevant attributes of individual i, and f is some function that transforms these attributes into a propensity score.



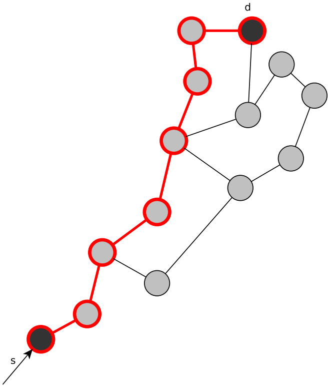
1. Criminal skills are represented by the criminal propensity *pcg* and trust is represented by the social distance between individuals d_{ij}.
2. The social distance between two individuals is represented by a value between 0 and 1, where 1 represents the maximum distance between them.
3. We define an arc when there is a common crime between a pair of individuals.



1. The search for the best association can be interpreted as the process of forming a criminal group in which a planner chooses individuals to participate based on their criminal capabilities and reliability, maximizing some utility function.



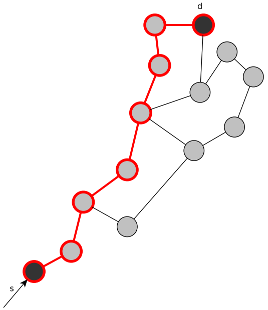
1. The search for the best association can be interpreted as the process of forming a criminal group in which a planner chooses individuals to participate based on their criminal capabilities and reliability, maximizing some utility function.
2. The LiRAM model determines the subset of individuals $i \in N$ and the set of arcs $(i, j) \in A$ that form the best association between individuals s (planner) and d (receiver).



Decision variables

$$y_i = \begin{cases} 1 & \text{if } i \in N \text{ is in the criminal group} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1 & \text{if } (i,j) \in A \text{ is in the solution} \\ 0 & \text{otherwise} \end{cases}$$



Objective function

- Utility function of a crime planner.

$$\max U = \frac{I}{pcg_{max}} \sum_{i \in N} pcg_i y_i - \frac{I\gamma}{d_{max}} \sum_{(i,j) \in A} d_{ij} x_{ij} - w \sum_{i \in N} pcg_i y_i$$

Constraints

- Predecessor constraint:

$$\sum_{i \in N} x_{ij} = y_j \qquad \forall j \in N \setminus \{s, d\} \quad (1)$$

- Flow conservation:

$$\sum_{i \in N} x_{ij} = \sum_{i \in N} x_{ji} \qquad \forall j \in N \setminus \{s, d\} \quad (2)$$

- Maximum criminal propensity:

$$\sum_{i \in N} pcg_i y_i \leq \varphi pcg_{max} \quad (3)$$

- Subtour elimination constraints:

$$\sum_{i,j \in L} x_{ij} = |L| - 1 \qquad \forall L \subseteq N \setminus \{s, d\} : |L| \geq 2 \quad (4)$$

- Initial and final flow:

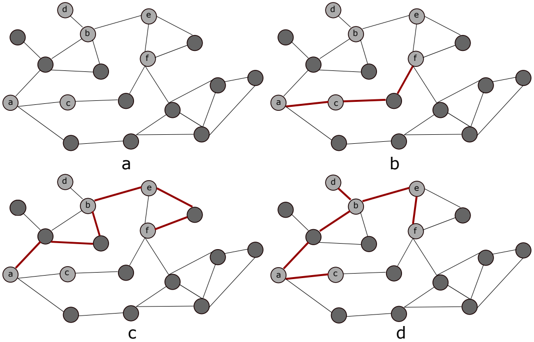
$$\sum_{j \in N} x_{sj} = \sum_{i \in N} x_{id} = 1 \quad (5)$$

- Variable's domain:

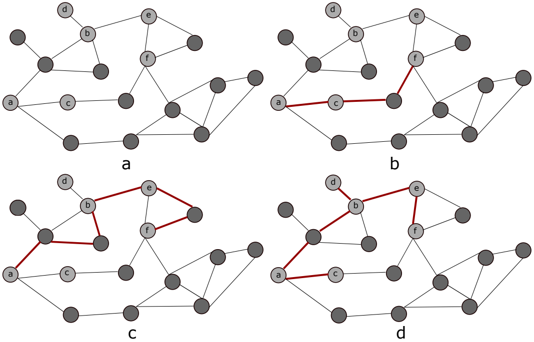
$$x_{ij} \in \{0, 1\} \qquad \forall (i, j) \in A \quad (6)$$

$$y_i \in \{0, 1\} \qquad \forall i \in N \quad (7)$$

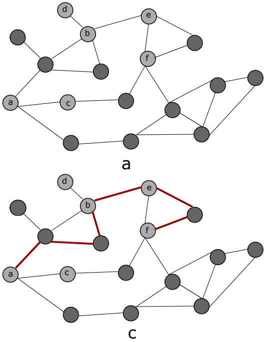
1. The planner is rational and chooses criminals with the skills that will ensure that the crime is carried out with the maximum utility.



1. The planner is rational and chooses criminals with the skills that will ensure that the crime is carried out with the maximum utility.
2. The StRAM model determines the Steiner tree, rooted in the crime planner $s \in N$.



1. The planner is rational and chooses criminals with the skills that will ensure that the crime is carried out with the maximum utility.
2. The StRAM model determines the Steiner tree, rooted in the crime planner $s \in N$.
3. Requires a single suspect to determine criminal groups



Decision variables

$$y_i = \begin{cases} 1 & \text{if } i \in N \text{ is in the criminal group} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1 & \text{if } (i,j) \in A \text{ is in the solution} \\ 0 & \text{otherwise} \end{cases}$$

$$f_{ij} = \text{flow through the arc } (i,j) \in A$$

Objective function

- Utility function of a crime planner

$$\max U = \frac{I}{pcg_{max}} \sum_{i \in N} pcg_i y_i - \frac{I\gamma}{d_{max}} \sum_{(i,j) \in A} d_{ij} x_{ij} - w \sum_{i \in N} pcg_i y_i$$

Constraints

- Predecessor constraint:

$$\sum_{i \in N} x_{ij} = y_j \qquad \forall j \in N \setminus \{s\} \qquad (8)$$

- Flow conservation:

$$\sum_{i \in N} f_{ij} - \sum_{i \in N} f_{ji} = y_j \qquad \forall j \in N \setminus \{s\} \qquad (9)$$

- Link of the variables:

$$f_{ij} \leq (|N| - 1)x_{ij} \qquad \forall (i,j) \in A \qquad (10)$$

- Maximum criminal propensity:

$$\sum_{i \in N} pcg_i y_i \leq \varphi pcg_{max} \qquad (11)$$

- Variable's domain:

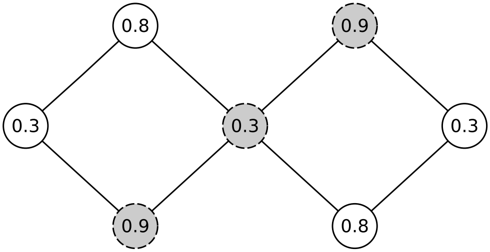
$$f_{ij} \geq 0 \qquad \forall (i,j) \in A \qquad (12)$$

$$x_{ij} \in \{0,1\} \qquad \forall (i,j) \in A \qquad (13)$$

$$y_i \in \{0,1\} \qquad \forall i \in N \qquad (14)$$

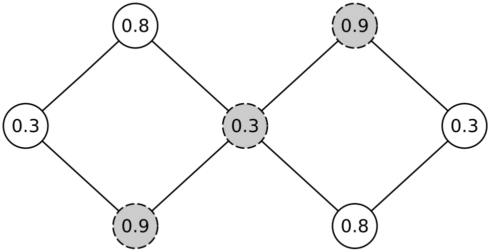
New criminal association model based on Steiner tree (NCAM)

1. The total social capital of the group is maximized, considering that interactions allow sharing part of the knowledge.



New criminal association model based on Steiner tree (NCAM)

1. The total social capital of the group is maximized, considering that interactions allow sharing part of the knowledge.
2. Allows to generate results in networks where there are no known suspects.



New criminal association model based on Steiner tree (NCAM)

Decision variables

$$y_i = \begin{cases} 1 & \text{if } i \in N \text{ is in the criminal group} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1 & \text{if } (i,j) \in A \text{ is in the solution} \\ 0 & \text{otherwise} \end{cases}$$

$f_{ij} =$ Flow through the arc $(i,j) \in A$

$w_{ij} =$ potential knowledge given by candidate $i \in N$ to $j \in N$
$z_i =$ knowledge provided by the candidate $i \in N$

Objective function

- Total band capital.

$$\max U = \sum_{i \in N} z_i$$

Constraints

- Predecessor constraint:

$$\sum_{i \in N \cup \{0\}} x_{ij} = y_j \qquad \forall j \in N \qquad (15)$$

- Flow conservation:

$$\sum_{i \in N \cup \{0\}} f_{ij} - \sum_{i \in N} f_{ji} = y_j \qquad \forall j \in N \qquad (16)$$

- Initial arc:

$$\sum_{j \in N} x_{0j} = 1 \qquad (17)$$

- Maximum size of the criminal group:

$$\sum_{i \in N} y_i = k_{max} \qquad (18)$$

- Initial flow:

$$\sum_{j \in N} f_{0j} = k_{max} \qquad (19)$$

- Link of the variables:

$$f_{ij} \leq (|N| - 1)x_{ij} \qquad \forall i \in N \cup \{0\}, j \in N \qquad (20)$$

Constraints

- Knowledge of the candidates:

$$z_i = pcg_i y_i + \sum_{(i,j)\in A} (e_{ij} w_{ij}) \qquad \forall i \in N \qquad (21)$$

- Linearization I:

$$w_{ij} \leq (x_{ij} + x_{ji})M \qquad \forall (i,j) \in A \qquad (22)$$

- Linearization II:

$$w_{ij} \geq z_j - (1 - x_{ij} - x_{ji})M \qquad \forall (i,j) \in A \qquad (23)$$

- Linearization III:

$$w_{ij} \leq z_j \qquad \forall (i,j) \in A \qquad (24)$$

- Variable's domain:

$$f_{ij} \geq 0 \qquad \forall i \in N \cup \{0\}, j \in N \qquad (25)$$

$$x_{ij} \in \{0,1\} \qquad \forall i \in N \cup \{0\}, j \in N \qquad (26)$$

$$y_i \in \{0,1\} \qquad \forall i \in N \qquad (27)$$

$$z_i \geq 0 \qquad \forall i \in N \qquad (28)$$

$$w_{ij} \geq 0 \qquad \forall (i,j) \in A \qquad (29)$$

- A network of 77 individuals provided by the Bío-Bío Regional Prosecutor's Office was used.
- The database contains 1666 crimes and a criminal group of 12 individuals.
- The models were implemented using the AMPL programming language, using the CPLEX solver.

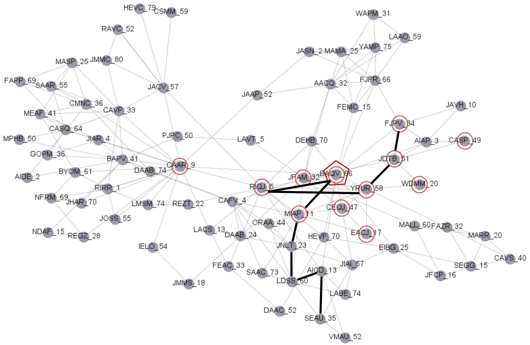


Figure: Criminal network of 77 suspects.

Results

1. We evaluated the performance of the *LiRAM*[2], *StRAM*[2] and *SPA*[3] models applied between pairs of suspect individuals.
2. According to the confidence intervals, there are no significant differences between the F-measure of *LiRAM* and *StRAM*.

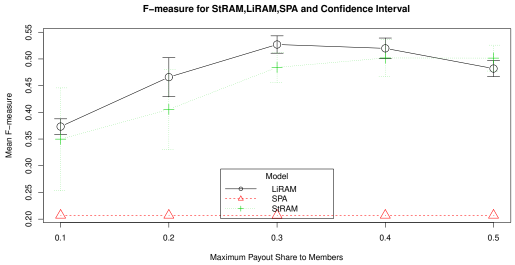


Figure: F-measure for *StRAM*, *LiRAM* and *SPA*

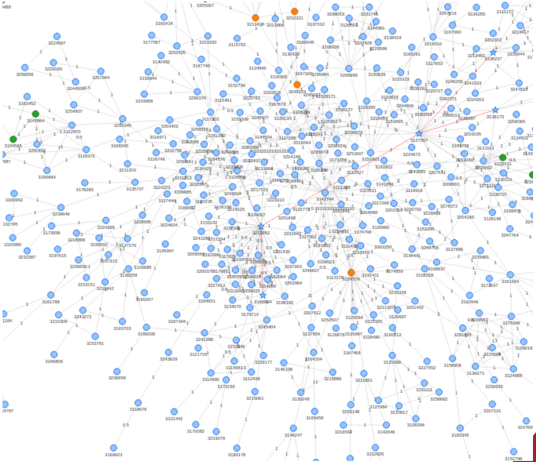
Results

1. At a significance level of 0.05, the Shapiro-Wilk test indicates that the F-Measure is not normally distributed.
2. Levene's test indicates that there is no homogeneity of variance ($\alpha = 0.05$).
3. The Kruskal-Wallis nonparametric test was applied. It is concluded that there are no significant differences between the results of the two models ($\alpha = 0.05$).

Table: Results to statistical tests for different values of φ.

Results to Statistical Tests					
	Maximum Payout Share to Members P-value				
Test	$\varphi = 0.1$	$\varphi = 0.2$	$\varphi = 0.3$	$\varphi = 0.4$	$\varphi = 0.5$
Shapiro-Wilk (LiRAM data)	0.0000000034111	0.0021649	0.0027877	0.0117392	0.2396275
Shapiro-Wilk (StRAM data)	0.02501584	0.00374289	0.00201161	0.04011305	0.00056471
Levene	0.00006102	0.1438	0.05163	0.03224	0.03611
Kruskal-Wallis	0.8504	0.4327	0.07577	0.6481	0.2407

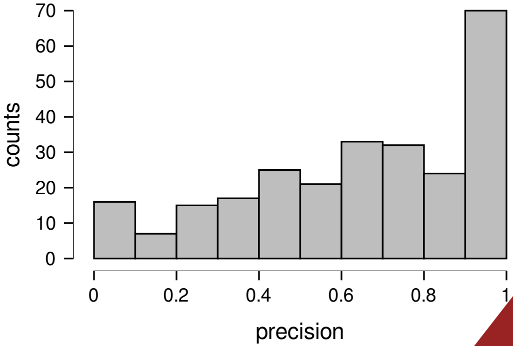
- The database of the Public Prosecutor's Office was used to validate and evaluate the models.
- Crimes committed between 2019 and 2021 were considered.
- The models were implemented using the Python programming language, using the Gurobipy library.



1. 260 instances generated from 13 networks were evaluated.
2. The number of individuals belonging to a *Phenomenon* in the solution was determined.
3. The *NCAM* model has an average accuracy of 0.657.

Table: Descriptive Statistics

	precision
Valid	260
Median	0.690
Mean	0.657
Std. Deviation	0.295
25th percentile	0.472
50th percentile	0.690
75th percentile	0.923



1. Currently, the models are implemented in the Public Prosecutor's Office.
2. The database of the Public Prosecutor's Office, with 20 years of crimes, is used.
3. Models are being used to derive possible criminal groupings.

Nacional

Viernes 12 mayo de 2023 | 11:33

"Operación influencer": Inteligencia Artificial permitió detener a banda que se dedicaba a encerronas

Por Florencia Ortiz

Relacionados

- "Operación Influencer": cae banda que realizaba encerronas para exhibir autos como trofeos en la web

Cedida a BBCL

8,424 visitas

Conclusions

1. Models provide excellent results with respect to existing approaches.
2. The proposed models open new approaches for applied research in criminal analysis.
3. The potential of the proposed models and their application to real cases is observed.

[1] Jeffrey Scott McIllwain. Organized crime: A social network approach. *Crime, Law and Social Change*, 32(4):301–323, 1999.

[2] Fredy Troncoso and Richard Weber. A novel approach to detect associations in criminal networks. *Decision Support Systems*, 128:113159, 2020.

[3] Jennifer J Xu and Hsinchun Chen. Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks. *Decision Support Systems*, 38(3):473–487, 2004.

[4] Jennifer J. Xu and Hsinchun Chen. Crimenet explorer: A framework for criminal network knowledge discovery. *ACM Trans. Inf. Syst.*, 23(2):201–226, apr 2005.

- FONDEF project ID20I10230ANID
- The Initiation Research Project 2060204IF/I
- Fondecyt Project 1181036
- Project ING 2030 I+D 20-34.
- Santiago based Complex Engineering Systems Institute (CONICYT PIA/BASALAFB180003).
- The Criminal Analysis Unit of the Public Prosecutor's Office

UNIVERSIDAD DEL BÍO-BÍO

UNIVERSIDAD DE CHILE

Universidad de los Andes

ADVANCED ANALYTICS FOR A BETTER WORLD

IFORS2023

The 23rd Conference of the International Federation of Operational Research Societies

July 10-14 · SANTIAGO, CHILE

ICHIO INSTITUTO CHILENO DE INVESTIGACIÓN OPERATIVA

alio asociación latino iberoamericana de investigación operativa

UNIVERSIDAD DE CHILE

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

ISCI INSTITUTO SISTEMAS COMPLEJOS DE INGENIERÍA