
Archetypes of Wildfire Arsonists: An Approach by Using Bayesian Networks

Rosario Delgado, José-Luis González,
Andrés Sotoca and Xavier-Andoni Tibau

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.72615>

Abstract

Wildfires are a phenomenon of great importance because of their environmental and economic consequences, as well as the human losses they cause. The rate of resolution of arson-caused wildfires is extremely low when compared to other criminal activities. This fact highlights the importance of developing methodologies to assist investigators in the criminal profiling. For that we propose the use of Bayesian networks (BNs), which are a methodology belonging to the field of machine learning. BNs are probabilistic models that have only recently been applied to criminal profiling. We learn a BN model from real data of solved arson-caused wildfires in Spain, and after validation we use it to construct archetypes of the forest fires/arsonists with the aim of better understanding of this phenomenon and help in the task of identification of the culprits. We characterize five different archetypes around author motivation from a quantitative and objective point of view, which are in correspondence with the modes of operation in criminal activities of Shye.

Keywords: provoked wildfire, arsonist, archetype, profiling, Bayesian networks

1. Introduction

According to the Food and Agriculture Organization of the United Nations (FAO) survey [1], “[...] every year, wildfires destroy millions of hectares of forests, woodlands and other vegetation, causing the loss of many human and animal lives and immense economic damage, both in terms of resources destroyed and the costs of suppression. There are also impacts on society and the environment [...]”. Mediterranean countries are especially sensitive to this phenomenon due to the characteristics of their vegetation, land use, and climate. On the average, 50,000 fires burn 400,000 hectares every year in these regions (San-Miguel-Ayanz, Moreno and Camia [2]), and the situation is worsening due to the effect of climate change (Turco, Llasat, von Hardenberg and

Provenzale [3]). According to the Ministry of Agriculture and Fishery, Food and Environment of Spain [4], in period 2006–2015, a yearly average of 13, 126 forest fires burned 133, 060 hectares. As a consequence, this phenomenon is one of the major environmental problems in Spain.

In this work, we are interested in the arson-caused wildfire, understood as “the uncontrolled fire on forest land caused by humans that spreads quickly out control over woodland or brush, affecting vegetation that was not destined to burn” (this definition does not include the burning of stubble, grass, or scrub for the removal of forest residues, unless they are carried out where they are prohibited).

From a quantitative point of view, wildfires have been studied mainly from the point of view of risk assessment. Just to mention some studies, Thompson, Scott Helmbrecht and Calvin [5] present an integrated and systematic risk assessment framework to better manage wildfires and to mitigate losses to highly valued resources and assets, with application to an area in Montana, United States, while Penman, Bradstock and Price [6] study the patterns of wildfires in south-eastern Australia in relation to risk of ignition, and Adab, Kanniah and Solaimani [7] consider different fire risk indices in northeastern Iran. In the criminological context, Cozens and Christensen [8] analyze how environmental criminology can help to prevent arson-caused wildfires in Australia, where this phenomenon also represents a serious problem.

Although arson is one potential cause of many fires, yet the rate of clarification of arson-caused wildfires is extremely low when compared to other criminal activities. According to the interim report of the Ministry of Agriculture and Fishery, Food and Environment of Spain [9], 11, 928 wildfires were committed in 2015 in Spain, of which 429 offenders have been identified, representing a resolution rate of 6–6.5% since the estimated percentage of wildfires in Spain that were deemed arson in 2015 ranges from 55 to 60%. This fact highlights the difficulty in identifying the authors of provoked forest fires. Therefore, any help in developing methodologies that can aid investigators to better understand motivation of arsonists in order to solve and, if possible, to prevent these crimes, is welcome. In this sense, our main aim is to find predictive relationships between different typologies of forest fire and the characteristics of the perpetrators, by constructing archetypes taking into account both author features (behavioral, criminological, socio-demographic, and of personality) and evidences obtained from the fire, in order to assist people with responsibilities in the judicial investigation, increasing the rate of clarification of crimes and misdemeanors. Our work is framed into a project led by the Prosecution Office of Environment and Urbanism of Spain, which is carried out by a team in which members of the Crime Behavior Analysis Section of the Technical Unit of the Judicial Police of the Civil Guard participate.

Apart from some few descriptive studies as Soeiro and Guerra [10], to our knowledge the only quantitative approaches to this question stem from the works of Sotoca, González, Fernández, Kessel, Montesinos and Ruz [11] and Delgado, González, Sotoca and Tibau [12]. More specifically, the approach followed in Sotoca et al. [11] consists in the application of different techniques of statistical multivariate analysis (mainly, cluster analysis) to criminal profiling, based on the premise that the crime scene contains clues that if properly collected and interpreted, could say something about the person who set the fire. Otherwise, in Delgado et al. [12], the methodology of Bayesian networks (from now on, BNs) was applied for the first

time to profiling of wildfire arsonists. BNs had only recently been applied to criminal profiling (see, for instance, Baumgartner, Ferrari and Palermo [13] and Baumgartner, Ferrari and Salfati [14]) and as far as we know, never before for profiling of any kind of arsonist.

The unpredictability of human behavior adds a component of randomness to all our activities, the criminal among them. BNs are an increasingly popular methodology in the field of machine learning for modeling uncertain in complex domains, and in the opinion of many Artificial Intelligence researchers, the most significant contribution in this area in the last years (Korb and Nicholson [15]). Indeed, BNs are of the most effective machine learning techniques and fall in the field of supervised learning, along with other techniques such as support vector machines, kernels, or neural networks.

BNs were introduced in the 1920s as a probabilistic tool to model the relationships among different variables. Usefulness of this methodology has been shown in many decision-making procedures and in different areas. In particular, it has been used with a great success in risk analysis in ecology (Ticehurst, Newham, Rissik, Letcher and Jakeman [16]), economy (Adusei-Poku [17]), emerging diseases (Walshe and Burgman [18]), environmental sciences (Borsuk, Stow and Reckhow [19] and Pollino, Woodberry, Nicholson, Korb and Hart [20]), medicine (Spiegelhalter [21], and Cruz-Ramrez, Acosta-Mesa, Carrillo-Calvet, Alonso Nava-Fernández and Barrientos-Martnez [22]), or nuclear waste accidents (Lee and Lee [23]). And with respect to criminology, for example, BNs have been introduced as a novel methodology for assessing the risk of recidivism of sex offenders in Delgado and Tibau [24].

Regarding wildfires, Papakosta and Straub [25] study a wildfire building damage consequences assessment system constructed from a BN, and applies it to spatial datasets from the Mediterranean island of Cyprus. Dlamini develops a BN model in [26] from satellite and geographic information systems (GIS), with variables of biotic, abiotic, and human kind, in order to determine factors that influence wildfire activity in Swaziland (see also Dlamini [27]). As mentioned above, Delgado et al. [12] is the only previous study on the use of BN for profiling of the author of a forest fire. The authors also implement this methodology for criminal profiling in an Internet computer application to be used by the Prosecution Office of Environment and Urbanism.¹

In this chapter, we set two objectives: in the first place, we intend to introduce BN and explain their application to the study of profiles of forest arsonists. Secondly, we go beyond Delgado et al. [12] into the use of this methodology for a better understanding of wildfire arsonists motivation, constructing archetypes which will help to identify the culprits. For that, we learn a BN model from the updated available data provided by the Spanish government, and use it to study motivation and for the construction of archetypes from the characteristics of an arson-caused wildfire and offender features. Roughly speaking, we construct the most probable BN given the observed cases (learning procedure), and this model provides information on the relationships between the considered variables, which are both fire features and author

¹Delgado R, Tibau XA. "PerfilNet.Pyros: Expert System based on Bayesian networks for the prediction of criminal profiles in forest fires". Register on June 10, 2016 of authorship at the "Benelux Office for Intellectual Property" (BOIP), with reference number i-depot number: 088029.

characteristics, allowing us to carry out predictions about some of them (query variables) from other (evidences).

The organization of the chapter is as follows. In Section 2, we introduce the research methods we use, starting with an introduction to the theoretical framework that supports profiling and archetypes, a description of the dataset on which we rely to construct our BN model, and a description of the model itself. Complementary and more technical information of the latter topic can be found in Appendix A. In Section 3, we apply the previously constructed BN to develop archetypes for forest fires/arsonists based on motivation. The chapter finishes with a conclusion section.

2. Research methods

2.1. Theoretical framework

As the comprehensive literature review, Dowden, Bennell and Bloomfield [28] showed that most criminal profiling publications do not provide any clear theoretical framework on the rationale of the profiling process, and only a few articles reported the use of statistical techniques (most of them multivariate). For this reason, some authors criticize the use of profiling and call it “pseudoscientific practice”, as Snook, Cullen, Bennell, Taylor and Gendreau [29], while police officers see it with some skepticism (Snook, Haines, Taylor and Bennell [30]) and the mental health professionals of the forensic environment also show their doubts about it (Torres, Boccaccini and Miller [31]).

In the United Kingdom, however, scientific literature that overcome previous criticisms has been available for more than 20 years, and has led to a new methodological approach to profiling known as “Behavioral Investigative Advice”. This approach takes into account evidence-based knowledge to aid decision-making by the police investigator, and includes many other tasks such as crime scene assessment, case-link analysis, suspect prioritization matrices, counseling in the police interview, etc. (see Alison and Rainbow [32]). The origin of this new perspective began with the studies of Canter, in which multidimensional scaling was applied to datasets of solved crimes in order to obtain clusters or profiles, in the first place of the crimes themselves, and later of the authors, to finally calculate the statistical correlation with each other. In this way, depending on how the crime was committed, it could be assigned to a profile, which would automatically report the characteristics of the author who most often commits this type of crime. In addition, Canter offered a theoretical model that helped interpret the results: *Shye’s model of action system* (Shye [33]). This methodology was applied to the elaboration of profiles of arsonists (Canter and Fritzon [34]; Fritzon, Canter and Wilton [35]) and was continued in other works, such as Fritzon [36], in which it was applied to study the relationship between the distance traveled by the arsonists and their motivation; Kocsis and Cooksey [37], which is focused on serial arsonists; and Wachi, Watanabe, Yokota, Suzuki, Hoshino, Sato and Fujita [38], in which the incendiary women in Japan are studied.

However, in spite of so many antecedents, all these authors address the incendiary phenomenon in general, not the forest fire in particular. The only work specifically forestry previous to

the studies carried out in Spain is the aforementioned Viegas and Soeiro [39] where, taking into account the model of action system and using multiple correspondence analysis, four profiles of forest arsonist in Portugal were proposed, denominated: “expressive with clinical history”, “expressive with attraction by the fire”, “vengeful instrumental”, and “instrumental to obtain profit”. Each of these profiles involves a series of identifying characteristics of its authors and a distinctive way of committing forest fires, depending on whether the main motivation was revenge, psychiatric problems, pathological attraction for fire, or obtaining an economic profit.

The work carried out in Spain Sotoca et al. [11] is inspired by the aforementioned Portuguese study and explores other data analysis methodologies, specifically techniques of multivariate statistical analysis, to establish an a priori classification of forest fires according to their cause or motivation, resulting in the following basic archetypes: “negligence”, which opposes “intentional”, being mutually exclusive. Intentional fires were grouped into four subtypes, also mutually exclusive: “profit”, “revenge”, “impulsive”, and “inadequate traditional practice”. This classification is consistent with the four modes of operation of the theoretical framework of criminal activities of Shye, and the correspondence among them is shown in **Table 1**.

As in Delgado et al. [12], in this chapter, we consider a slight modification of the archetypes constructed in Sotoca et al. [11]: we stack “negligence” and “inadequate traditional practice” into “negligence”, since in both cases the fire occurs as a consequence of a recklessness, but distinguishing between “slight negligence” and “gross negligence”, depending on whether the perpetrator remains on site and helps extinguishing services, in the first case, or not. The rest of archetypes have not been modified. Then, the list of updated archetypes and their correspondence with modes of operation is given in **Table 2**. This is in line with the proposal of the five

Former forest fire classification	Mode of operation
Negligence	Adaptive
Inadequate traditional practice	Adaptive
Impulsive	Integrative
Profit	Expressive
Revenge	Conservative

Table 1. Equivalence between former classification given in Sotoca et al. [11] and mode of operation in Shye [33].

Present forest fire classification	%	Mode of operation
Slight negligence	47.64	Adaptive
Gross negligence	31.30	Adaptive
Impulsive	10.05	Integrative
Profit	7.59	Expressive
Revenge	3.42	Conservative
	100.00	

Table 2. Equivalence between present classification and mode of operation in Shye [33].

main profiles of forest fire from an “operational” character, each one with its own author profile, found in previous years and confirmed by the most recent statistical analysis carried out by the team working in this project. It is important to note that “impulsive”, “profit”, and mainly “revenge” are uncommon compared to the rest. Motivation has been recorded in 1,463 of the 1,597 solved cases in our database, and in **Table 2** we show the percentages of each motivation type.

In Delgado et al. [12] and in this chapter, the use of BN is proposed as an alternative to the analysis used in Sotoca et al. [11], since BNs allow to know not only if the way of committing a forest fire is associated with some characteristic of the author, but to quantify this association, which gives the fire investigator far more accurate information. BNs are a machine learning methodology of self-learning from the data that can be used with success in the social sciences, where efforts to find scientific laws on human behavior often fail to establish a conceptual framework to guide empirical observation and the method of analysis corresponding to that framework.

As mentioned in Section 1, our aim is to present BN as a methodology to improve understanding of the different types of motivations from a quantitative and objective point of view, helping in the construction of archetypes.

2.2. The dataset

Statistical information on the phenomenon of forest fires has been collected in Spain since 1968, generating one of the most complete databases in Europe and been pioneer worldwide. This information is currently managed by the General Directorate of Natural Environment and Forestry Policy of the Ministry of Agriculture and Fishery, Food and Environment of Spain. However, our database consists of policing clarified arson-caused wildfires (for which the alleged offenders have been identified), has been feeding since 2008 by the Secretary of State for Security throughout the entire Spanish territory, under the leadership of the Prosecution Office of Environment and Urbanism of the Spanish state, and contains information obtained from a specific questionnaire concerning authors that have been arrested or imputed.

As mentioned above, adding certain and supposed causes it seems that the percentage of wildfires in Spain that were intentional ranges from 55 to 60% (close to other countries like Australia, Cozens and Christensen [8]), while it was only possible to identify 6–6.5% of the arsonists. Given these numbers, it could be said that the intentional forest fire is a criminal activity with very low rate of clarification, which explains the interest of the involved authorities and the society in general, in increasing the rate of clarification.

This subset conforms our dataset, which contains 1597 solved cases. According to the expert’s knowledge, $n = 25$ variables have been chosen of the total set of 32 initial variables, because of their usefulness and predictive relevance. The choice is the result of a balance between the benefits of having a high number of variables (more realistic model with higher accuracy) and the drawbacks arising from the corresponding increasing complexity (implying the need for more data to learn the model properly). The chosen variables refer to crime (C_1, \dots, C_{10}) and to the arsonist (A_1, \dots, A_{15}), and are described in **Table 3**, where their possible outcomes are also

Variables	Outcomes
C_1 = season	Spring/winter/summer/autumn
C_2 = risk level	High/medium/low
C_3 = start time	Morning/afternoon/evening
C_4 = starting point	Pathway/road/houses/crops/interior/forest track/others
C_5 = main use of burned surface	Agricultural/forestry/livestock/interface/recreational
C_6 = number of seats	One/more
C_7 = related offense	Yes/no
C_8 = pattern	Yes/no
C_9 = traces	Yes/no
C_{10} = who denounces	Guard/particular/vigilance
A_1 = age	$\leq 34/35 - 45/46 - 60/> 60$
A_2 = way of living	Parents/in couple/single/others
A_3 = kind of job	Handwork/qualified
A_4 = employment status	Employee/unemployed/sporadic/retired
A_5 = educational level	Illiterate/elementary/middle/upper
A_6 = income level	High/medium/low/without incomes
A_7 = sociability	Yes/no
A_8 = prior criminal record	Yes/no
A_9 = history of substance abuse	Yes/no
A_{10} = history of psychological problems	Yes/no
A_{11} = stays in the scene	No/remains there/remains and gives aid
A_{12} = distance home-scene	Short/medium/long/very long
A_{13} = displacement means	On foot/ by car/all terrain/others
A_{14} = residence type	Village/house/city/town
A_{15} = motivation	Slight negligence/gross negligence/impulsive/profit/revenge

Table 3. Outcomes of the variables in the dataset.

shown. The incendiary variables A_1, \dots, A_{15} correspond to aspects that are easily observable and have some police relevance, which is very convenient since they are intended to guide the police activity to clarify the crime. We use exclusively categorical variables, by discretizing the (few) continuous variables in the original database. Approximately 78% of cases have missing values in at least one of the variables, mostly variable authors, which are the ones that have the most missing cases. Because it is a very high percentage, instead of omitting cases containing at least one missing value, which is a standard practice, we replace missing values by a new value different from the rest of the outcomes (a “blank”, in our case), treating missing values as a unique value and not mapping them into any other. In this way we do not lose information. Once obtained the predictions for each query variable, the “blank” value is eliminated from prediction and its probability is proportionally divided among the rest of its outcomes.

2.3. Constructing the BN

BNs are graphical structures for representing the probabilistic relationships among the variables describing a random phenomenon, such as in our setting provoked forest fires, and for performing probabilistic inference with them. Given a set of random variables $V = \{X_1, \dots, X_n\}$, a BN is a model that represents the joint probability distribution P over those variables. In our case, $V = \{C_1, \dots, C_{10}, A_1, \dots, A_{15}\}$ and $n = 25$. The graphical representation of the BN consists of a *directed acyclic graph* (DAG), whose n nodes represent the random variables (from now on, we identify a node with the variable that represents). The directed arcs among the nodes represent conditional dependencies between variables. **Figure 1** shows the DAG corresponding to the BN that has been constructed (learned from data).

We can use the BN to help in characterizing a provoked wildfire in terms of the relationships between different variables. These relationships are expressed in a very simple way in the BN, through the absence/presence of directed arcs in its DAG, taking into account the *Markov condition*, which stays the following: “knowing the values that its **parents** take, which are the nodes sending a directed arc to it in the DAG, any variable is independent of any other which is not a parent nor a **descendant** of it (a “descendant” of a node is any other node to which is possible to arrive from it

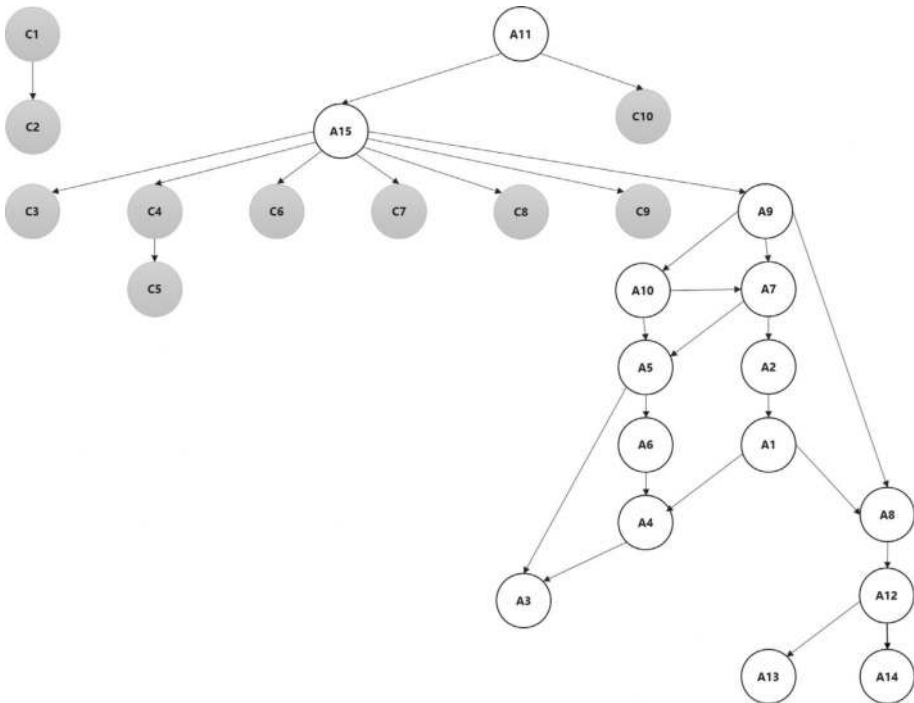


Figure 1. Learned structure (DAG) of the BN from the dataset.

by following a path linking directed arcs)". For example, observing **Figure 1** we can see that known the value of variable A_{15} , C_4 is independent of any other variable except C_5 , since C_5 is its unique descendant. Just to mention another example, if we know the outcome of variable A_8 , then A_{12} is independent of the rest of variables except A_{13} and A_{14} .

Once learned the BN model from the dataset, both the structure (DAG) and the *parameters* (the probability distribution of each variable conditioned to its parents), we can use it to compute any *a posteriori* probability we are interested in: we can consider an evidence concerning some variables of the model and use the BN to update the (*a priori*) probability distribution of any of the rest of variables, knowing the evidence. More specifically, from an evidence of the form $E = \{X_{i_1} = x_{i_1}, \dots, X_{i_t} = x_{i_t}\}$, where $\{X_{i_1}, \dots, X_{i_t}\} \subset V$ are the *evidence* variables, we could be interested in computing the *a posteriori* (conditioned) probability $P(X_{j_1} = x_{j_1}, \dots, X_{j_s} = x_{j_s} / E)$ with $\{X_{j_1}, \dots, X_{j_s}\} \subset V \setminus \{X_{i_1}, \dots, X_{i_t}\}$ the set of *query* variables. This probability is the update when we noticed an additional piece of knowledge, of the corresponding *a priori* probability, which would be the same but without conditioning with respect evidence E . Given an evidence E , the prediction of the query variable X is chosen to be the instantiation of X that maximizes the *a posteriori* probability. In a more formal way, if x_1, \dots, x_r are the possible instantiations of X , then $x^* = \arg \max_{k=1, \dots, r} P(X = x_k / E)$ is the prediction for X knowing evidence E , and $P(X = x^* / E)$ is said to be the *confidence level (CL)* of the prediction. We will apply this procedure to our setting in the following way: given an evidence in terms of the crime (evidence) variables for a given provoked forest fire, we will predict the value of the query arsonist features (query variables), which form the predicted profile of the arsonist. Interested readers can find technical details about the construction and validation of the BN in Appendix A.

All calculations, as well as the process of model construction, validation, and inference, have been carried out with **R**, which is "GNU S", a freely available language and environment for statistical computing and graphics, which provides a wide variety of statistical and graphical techniques. It can be obtained from the CRAN site <https://cran.r-project.org/>. Different packages of R has been adopted:

- **bnlearn**: Bayesian Network Structure Learning, Parameter Learning and Inference, by Marco Scutari and Robert Ness, <http://www.bnlearn.com/>

We use this package for Bayesian network structure learning and parameter learning, using the score-based Hill-Climbing structure learning algorithm and maximum likelihood parameter estimation, respectively.

- **gRain**: Graphical Independence Networks, by Søren Højsgaard, <http://people.math.aau.dk/sorenh/software/>

We use this package for making inference by probability propagation with the BN learned by using the bnlearn package.

- **sna**: Tools for Social Network Analysis, by Carter T. Butts, <http://www.statnet.org>.

From this package, we use some social network analysis measures in Section 3.1.

3. Archetypes

In this section we use the BN model learned from the dataset and described in Section 2, to construct forest fire archetypes related to arsonist motivation.

First of all, note that author variables A_8 = “prior criminal record”, A_9 = “history of substance abuse”, and A_{10} = “history of psychological problems” are operative variables of practical use so that the investigators can identify the author of a provoked fire. Fortunately, these variables have a good accuracy in prediction with the BN model, higher than 80%. See **Table 4**, where accuracies, both individual for the prediction of each author variable (IPA), as well as overall (OPA), are consigned.

3.1. Why motivation?

We use motivation (A_{15}) as a cornerstone from which to construct the archetypes by two reasons: (1) from a viewpoint of the theoretical framework, motivation plays a key role in criminological investigations (see Collin [40]), and as explained in Section 2.1, in order to meet Shye’s classification for criminal activities, motivation should be taken as classification criterion, and (2) A_{15} is the author variable with the most central role and explanatory capacity of fire characteristics in the model.

Indeed, 8 of the 10 nodes representing crime variables are directly related to it. More concretely, 7 of them are descendants (from C_3 to C_9 , being all of them “sons” of A_{15} , except C_5 ,

Author variable	IPA (%)
A_1 = age	33.13
A_2 = way of living	60.40
A_3 = kind of job	72.28
A_4 = employment status	44.62
A_5 = educational level	46.24
A_6 = income level	46.96
A_7 = sociability	97.02
A_8 = prior criminal record	80.19
A_9 = history of substance abuse	90.58
A_{10} = history of psychological problems	89.56
A_{11} = stays in the scene	60.49
A_{12} = distance home-scene	45.13
A_{13} = displacement means	34.38
A_{14} = residence type	47.12
A_{15} = motivation	56.36
Total OPA (%)	58.12

Table 4. Individual predictive accuracy (IPA) and overall predictive accuracy (OPA).

which is a “grandson”), and one, C_{10} , is a “brother”, that is, it is a son of the father of A_{15} , which is A_{11} (see **Figure 1**). The main role of A_{15} in the model can be quantified by using centrality and/or betweenness measures borrowed from the Network Analysis area. In Graph Theory and Network Analysis, indicators of centrality identify the most important nodes within a graph. Here, “importance” is conceived as involvement in the cohesiveness of the network. Applications of centrality include identifying the most influential person(s) in a social network, key infrastructure nodes in the Internet or urban networks, and super-spreaders of a disease. Concretely, for each author variable we computed two measures, which are shown in **Table 5**, both normalized in order to sum up 100:

- a. *Freeman’s degree of centrality* (Freeman [41]), which counts paths which pass through each node, that is, directed arcs which arrive at or depart from it. **Table 5** points out A_{15} as the author variable with the most central role, doubling the value of the following in the ranking.
- b. *Borgatti and Everett’s betweenness measure* (Borgatti and Everett [42]). Betweenness quantifies the number of times a node acts as a “bridge” along the shortest path between two other nodes (which we will call “geodesic” from now on). Nodes that have a high probability to occur on a randomly chosen geodesic between two randomly chosen nodes, have a high betweenness. Borgatti and Everett’s betweenness is a modification of a basic

Author Variables	Freeman’s Centrality (%)	Borgatti and Everett’s Betweenness (%)
A_1 = age	6.67	9.36
A_2 = way of living	4.44	8.17
A_3 = kind of job	4.44	0.00
A_4 = employment status	6.67	2.90
A_5 = educational level	8.88	9.74
A_6 = income level	4.44	3.07
A_7 = sociability	8.88	11.36
A_8 = prior criminal record	6.67	12.68
A_9 = history of substance abuse	8.88	14.94
A_{10} = history of psychological problems	6.67	3.33
A_{11} = stays in the scene	4.44	0.00
A_{12} = distance home-scene	6.67	9.72
A_{13} = displacement means	2.23	0.00
A_{14} = residence type	2.23	0.00
A_{15} = motive	17.77	14.68
	100.00	100.00

Table 5. (Normalized) Freeman’s degree of centrality and Borgatti and Everett’s betweenness measure of the author variables.

standard betweenness measure, which was defined for a node v as $\sum_{i,j \text{ nodes}} \frac{g_{vi}}{g_{ij}}$ (with the convention $0/0 = 0$), where g_{ij} is the number of geodesics from i to j in the graph, and g_{ivj} is the number of geodesics in the subset of those that pass through v . The modification proposed by Borgatti and Everett is as follows:

$$\sum_{i,j \text{ nodes}} \frac{1}{d_{ij}} \frac{g_{ivj}}{g_{ij}} \quad (1)$$

where d_{ij} is the geodesic distance from i to j (that is, the number of directed arcs that compose any geodesic from i to j). Conceptually, using the basic standard betweenness measure, high-betweenness nodes lie on a large number of non-redundant geodesics between other nodes; they can thus be thought of as “bridges”. Borgatti and Everett’s betweenness adjusts the basic standard by down-weighting long geodesics, and attending to it we see in **Table 5** that A_{15} is the second most important after, but very close, to A_9 .

3.2. Constructing archetypes

The explained above justifies the decision to base our archetypes of provoked forest fires on A_{15} . Therefore, we construct some archetypes around motivation, and comparing them with that in Sotoca et al. [11], we see they are consistent. To carry this out, we predict query variables $C_1, \dots, C_{10}, A_1, \dots, A_{14}$ by introducing as evidence the different possible outcomes of variable A_{15} . Some of the crime variables, and most of the author variables are insensitive, that is, they coincide for the consigned five possible criminal motivations, and for any of them always have the same predicted values, which are collected in **Table 6**.

In case of C_1 and C_2 , it is not surprising since, as can be seen in **Figure 1**, they are not related neither with A_{15} nor with any other variable in our model. Coinciding with common sense, for each of these two variables the most probable value is chosen, independently of the evidence variable A_{15} .

Explanation for each of the variables appearing in **Table 6** that are sons of A_{15} , which are C_6, C_7 , and C_9 , is straightforward: we just have to have a look at its conditional probability table (CPT so on), whose values are parameters of the BN model that have been learned from data when constructing it, and observe that conditioned to the different outcomes of A_{15} , the most probable value of any of them does not vary. Simply to illustrate, **Table 7** is the conditional probability table of variable C_6 conditioned to A_{15} . The maximum probability corresponds to the same row when we vary from one column to another, that is, conditioned to any of the possible outcome of A_{15} the prediction of our model for C_6 is always “one”.

For the rest of variables in **Table 6**, intuition is no longer reliable since their relation with A_{15} is modeled through a chain of oriented arcs (a path). We can say that, in general, the longer the path linking them, the lesser the mutual influence is between two nodes, which would explain the presence of the author variables in **Table 6**.

Variables not appearing in **Table 6** take different values according to motivation, as **Table 8** shows, and they are those from which we will describe our archetypes. Of the crime variables,

Variable	Predicted value
C_1 = season	Spring
C_2 = risk level	High
C_6 = number of seats	One
C_7 = related offense	No
C_9 = traces	No
C_{10} = who denounces	Particular
A_1 = age	46 – 60
A_2 = way of living	In couple
A_3 = kind of job	Handwork
A_4 = employment status	Employee
A_5 = educational level	Elementary
A_6 = income level	Medium
A_7 = sociability	Yes
A_8 = prior criminal record	No
A_{10} = history of psychological problems	No
A_{12} = distance home-scene	Medium
A_{14} = residence type	Town

Table 6. Common predicted values for all the five archetypes.

$C_6 \downarrow A_{15} \rightarrow$	Slight negligence	Gross negligence	Impulsive	Profit	Revenge
More	0.06	0.10	0.33	0.41	0.37
One	0.94	0.90	0.67	0.59	0.63
	1.00	1.00	1.00	1.00	1.00

Table 7. Conditional probability table (CPT) of C_6 to A_{15} . For example, if A_{15} were “slight negligence”, the estimated probability for C_6 = “one” is 0.94, that is, $P(C_6 = \text{“one”} / A_{15} = \text{“slight negligence”}) = 0.94$, which is the maximum value of its column, being then “one” the prediction for C_6 conditioned to $A_{15} = \text{“slight negligence”}$.

C_3 , C_4 , and C_8 are sons of A_{15} , while C_5 is a grandson. CPTs of C_3 , C_4 , and C_8 conditioned to A_{15} , whose values are parameters of the model which are learned from data, give a straightforward prediction for each of these variables, which is the most likely predicted value conditioned to each motivation type. With C_5 and A_{13} we have to be more cautious. It is recommended to the interested readers to delve into this aspect, to consult Appendix B.

3.3. Checking, improving, and reducing archetypes

It seems convenient to check the constructed archetypes given in **Table 8**, and we will carry it out as follows. We could ask if using as evidence the values of the variables in **Table 8** for each of the archetypes, and as query variable motivation, the model will predict the concordant archetype. If so, the archetype would be strengthened and would, in a certain sense, be

	Negligence		Intentional		
Variables↓ Archetypes →	(1) Slight negli.	(2) Gross negli.	(3) Impulsive	(4) Profit	(5) Revenge
C ₃ = start time	Afternoon	Afternoon	Afternoon	Afternoon	Evening
C ₄ = starting point	Crops	Crops	Pathway	Pathway	Pathway
C ₅ = main use surface	Agricultural	Agricultural	Forestry	Forestry*	Forestry
C ₈ = pattern	No	No	Yes	Yes	No
A ₉ = history subst. Abuse	No	No	No	No	Yes
A ₁₁ = stays in the scene	Gives aid	No	No	No	No
A ₁₃ = displacement means	By car	By car	On foot	By car	On foot

*the second most likely outcome, "agricultural", has a very close probability to that of "forestry", as can be seen in **Table 15**, Appendix B.

Table 8. Specific predicted values for each of the five archetypes (extended version).

validated. But it may not happen, because we do not obtain the same probabilities conditioning C₄ by A₁₅ for example, that vice versa. Indeed, to exemplify this fact, we set specific values for these variables, say "pathway" and "impulsive", respectively, and we will see that

$$P(C_4 = \text{"pathway"} / A_{15} = \text{"impulsive"}) \neq P(A_{15} = \text{"impulsive"} / C_4 = \text{"pathway"}). \quad (2)$$

The reason appears clearly when using Bayes' Theorem we relate these two probabilities:

$$P(A_{15} = \text{"impulsive"} / C_4 = \text{"pathway"}) = \frac{P(C_4 = \text{"pathway"} / A_{15} = \text{"impulsive"}) P(A_{15} = \text{"impulsive"})}{P(C_4 = \text{"pathway"})}, \quad (3)$$

That is, these probabilities are related by means of the multiplicative factor

$$\frac{P(A_{15} = \text{"impulsive"})}{P(C_4 = \text{"pathway"})} \cong \frac{0.1005}{0.1490} \cong 0.6745 \quad (4)$$

in this way:

$$P(A_{15} = \text{"impulsive"} / C_4 = \text{"pathway"}) = 0.6745 \times P(C_4 = \text{"pathway"} / A_{15} = \text{"impulsive"}). \quad (5)$$

Table 9 shows the CPT of A₁₅ to the evidences given by the values of variables in **Table 8**. The predicted (most likely) value for A₁₅ appears in boldface

Looking at **Table 8**, we note that the only difference between the archetypes **impulsive** and **profit** is given by A₁₃. Will this difference propagate to A₁₅? **Table 9** tells no, since the conditional probability tables of A₁₅ for the corresponding evidences match, and we see that **impulsive** is the only archetype given by **Table 8** that has not been confirmed by **Table 9**. Could we modify this archetype in some sense to better adapt to data and result in an improved version? Actually yes.

Archetype	Evidence variables	Value	Conditioned distrib. of A_{15} to evidence
(1)	C_3 = wildfire start time	Afternoon	Profit (0.25%)
	C_4 = starting point	Crops	Gross negligence (0.00%)
	C_5 = main use of surface	Agricultural	Slight negligence (99.72%)✓
	C_8 = pattern	No	Impulsive (0.03%)
	A_9 = traces	No	Revenge (0.00%)
	A_{11} = stay in the scene	Gives aid	
	A_{13} = displacement means	By car	
(2)	C_3 = wildfire start time	Afternoon	Profit (2.79%)
	C_4 = starting point	Crops	Gross negligence (96.75%)✓
	C_5 = main use of surface	Agricultural	Slight negligence (0.00%)
	C_8 = pattern	No	Impulsive (0.31%)
	A_9 = traces	No	Revenge (0.15%)
	A_{11} = stay in the scene	No	
	A_{13} = displacement means	By car	
(3)	C_3 = wildfire start time	Afternoon	Profit (39.39%)χ
	C_4 = starting point	Pathway	Gross negligence (22.02%)
	C_5 = main use of surface	Forestry	Slight negligence (0.00%)
	C_8 = pattern	Yes	Impulsive (32.85%)
	A_9 = traces	No	Revenge (5.74%)
	A_{11} = stay in the scene	No	
	A_{13} = displacement means	On foot	(It should be Impulsive)
(4)	C_3 = wildfire start time	Afternoon	Profit (39.39%)✓
	C_4 = starting point	Pathway	Gross negligence (22.02, %)
	C_5 = main use of surface	Forestry/agricultural	Slight negligence (0.00%)
	C_8 = pattern	Yes	Impulsive (32.85%)
	A_9 = traces	No	Revenge (5.74%)
	A_{11} = stay in the scene	No	
	A_{13} = displacement means	By car	
(5)	C_3 = wildfire start time	Evening	Profit (4.56%)
	C_4 = starting point	Pathway	Gross negligence (5.15, %)
	C_5 = main use of surface	Forestry	Slight negligence (0.00%)
	C_8 = pattern	no	Impulsive (24.52%)
	A_9 = traces	Yes	Revenge (65.77%)✓
	A_{11} = stay in the scene	No	
	A_{13} = displacement means	On foot	

Table 9. Checking archetypes given in Table 8.

Let us go back for a moment to **Table 8**. Given an evidence as, for example, A_{15} = "profit", we predict query variables appearing in the table (and the rest as well) *as if they were independents*. This assumption make the calculations for predictions feasible, since if this assumption were not made, calculations would be so large that they would easily overflow the calculating capacity of a personal computer. But is it realistic? By the Markov condition, given A_{15} known, the independency among variables appearing in **Table 8** can be assumed (approximately in case of A_9 and A_{13} , because although A_{13} is a descendant of A_9 , the length of the geodesic that connects them weakens dependency) except in one case: C_4 and C_5 . Fortunately, it is feasible to carry on the calculations to obtain the joint probability distribution of C_4 and C_5 conditioned to A_{15} , and making the joint prediction of both (that is, taking the values that maximize this joint distribution), this prediction improves that made separately assuming an independence that is far from certain. For example, conditioned to A_{15} = "impulsive", the combination of values of C_4 and C_5 that maximizes the joint probability distribution is: C_4 = "road" and C_5 = "forestry". By replacing C_4 = "pathway" by C_4 = "road" in archetype (3) of **Table 9**, we obtain the conditioned distribution of A_{15} to the evidence given by the evidence variables in **Table 10**.

For the rest of archetypes, the joint predictions of C_4 and C_5 are exactly the same as the separated ones assuming independency, except for **revenge**. In this case, the joint prediction is C_4 = "forest track" and C_5 = "forestry". If substitute C_4 = "pathway" by C_4 = "forest track" while maintaining C_5 = "forestry" in **Table 9**, archetype (5), the probability of predict **revenge** increases from 65.77 to 76.45%.

Finally, for each archetype we can eliminate some of the variables without a great loss, those that are superfluous in the sense that if we do not include them as part of the evidence, the conditioned probability of A_{15} does not change excessively, maintaining the same prediction (value that maximizes probability). The improved and reduced version of the archetypes are given in **Table 11**. Naturally, the archetypes with the highest confidence level (CL) are those that correspond to both types of negligence, which are the most frequently consigned motivations in the dataset. We summarize the main distinctive features of each archetype:

- **Negligence** is characterized because the starting point of the fire is **crops**, and the main use of the burned surface is **agricultural**. The only difference between **slight** and **gross** negligence is that in the first case arsonist stays at the scene and **gives aid** while in the

Modified archetype	Evidence variables	Value	Conditioned distrib. of A_{15} to evidences
(3)	C_3 = wildfire start time	Afternoon	Profit (25.62%)
	C_4 = starting point	Road	Gross negligence (13.22%)
	C_5 = main use of surface	Forestry	Slight negligence (0.00%)
	C_8 = pattern	Yes	Impulsive (54.32%)✓
	A_9 = traces	No	Revenge (6.84%)
	A_{11} = stay in the scene	No	
	A_{13} = displacement means	On foot	

Table 10. Checking modified archetype **impulsive**.

second he does **not**. This is consistent with intuition, given that these type of fires are mainly accidentally caused by farmers.

- **Impulsive** is characterized by the starting point of the fire, which is a **road**, and the main use of the burned surface, which is **forestry**. As for **profit**, there is a **pattern** of action of the incendiary in the criminal activity. In this case, the arson has no specific objective beyond the arsonist momentum, so the forest is usually burned but not other types of surfaces. A road as starting point of the fire is characteristic in this archetype because it is a fast escape route after causing the fire.
- **Profit** is mainly characterized because the starting point of the fire is a **pathway**, and there is **no history of substance abuse** by the arsonist, which is logical from the point of view that, contrary to the previous archetypes, this type of wildfires are premeditated. The existence of a **pattern** of action is shared with **impulsive**.
- **Revenge** is the only archetype in which wildfire start time matters, and it occurs in the **evening**. Moreover, it is just the opposite as **profit** in the sense that for this archetype, there is **no pattern** of action but the author does have a **history of substance abuse**. This would tell us that usually this type of provoked forest fire is not the consequence of deliberate action, rather, it is carried out by a person under the effects of drugs and who could be swayed by an impulsive feeling of rage.

4. Conclusion

By using an ad hoc BN model learned from a dataset, we construct five archetypes for provoked forest fires. These archetypes are structured from arsonist motivation, which is the most central author variable in the model and plays an important role in psychological criminology, in accordance with the modes of operation in criminal activities of *Shye's model of action system* [33]. We see that the constructed model from the dataset of solved provoked Spanish forest fires conforms to this theoretical model. Two archetypes correspond to the mode of operation *adaptive*: **slight negligence** and **gross negligence**, which are distinguished in that while for the first the author stays at the crime scene and helps firefighting equipment, for the second he does not. The rest of archetypes are **impulsive**, **profit** and **revenge**, and correspond respectively to the modes of operation *integrative*, *expressive* and *conservative*.

In addition, we obtain a ratification of the five archetypes introduced in Sotoca et al. [11] in general terms, but with some specificities obtained thanks to the great potentiality of the used methodology. Indeed, the constructed BN models the relationships of dependency between the different variables (features of the wildfire and characteristics of the arsonist, including motivation), and it is precisely the understanding of these dependencies that allows to obtain predictions about some variables (queries) from others (evidences), without having to give up to take into account the complex relations that exist among them. As a matter of fact, the BN model captures these complexity and use it in an efficient way.

The specificities of each archetype are given by the values of a reduced set of variables that characterize each one, as stated in **Table 11**, where the confidence level or each archetype,

which is the probability of the prediction given the corresponding set of evidences, is also consigned. As expected, the best results in terms of the predictive capacity of the model correspond to both types of negligence, which are the most common consigned motivations in the dataset, far ahead of the other three archetypes, much less frequent.

With this work we hope to highlight the usefulness of BN as an objective and quantitative methodology to obtain valuable information from the dataset, and its applicability in the study of criminal motivation and behavior in general and, in particular, of forest arsonists, helping to identify the authors and to study this phenomenon, so complex and with such serious consequences for the environment.

Archetype	Evidence variables	Value	Conditioned distr. of A_{15}	CL
(1) Slight negl.			Profit (0.87%)	98.92%
	C_4 = starting point	Crops	Gross negligence (0.00%)	
	C_5 = main use of surface	Agricultural	Slight negligence (98.92%)✓	
	A_{11} = stay in the scene	Gives aid	Impulsive (0.19%) Revenge (0.02%)	
(2) Gross negl.			Profit (6.78%)	90.98%
	C_4 = starting point	Crops	Gross negligence (90.98%)✓	
	C_5 = main use of surface	Agricultural	Slight negligence (0.00%)	
	A_{11} = stay in the scene	No	Impulsive (1.69%) Revenge (0.56%)	
(3) Impulsive			Profit (16.39%)	59.32%
	C_4 = starting point	Road	Gross negligence (5.60%)	
	C_5 = main use of surface	Forestry	Slight negligence (9.23%)	
	C_8 = pattern	Yes	Impulsive (59.32%)✓ Revenge (9.46%)	
(4) Profit			Profit (32.63%)✓	32.63%
	C_4 = starting point	Pathway	Gross negligence (10.75, %)	
	C_8 = pattern	Yes	Slight negligence (19.49%)	
	A_9 = history subst. abuse	No	Impulsive (31.79%) Revenge (5.34%)	
(5) Revenge			Profit (5.22%)	51.81%
	C_3 = wildfire start time	Evening	Gross negligence (11.03, %)	
	C_8 = pattern	No	Slight negligence (4.10%)	
	A_9 = history subst. abuse	Yes	Impulsive (27.84%) Revenge (51.81%)✓	

Table 11. Final archetypes: an improved and reduced version. The confidence level (CL) for each archetype is the probability of the outcome predicted for A_{15} .

Acknowledgements

The authors wish to express their acknowledgment to the Prosecution Office of Environment and Urbanism and to the Secretary of State for Security of the Spanish state for providing data and promote research.

R. Delgado and X.A. Tibau are supported by Ministerio de Economía y Competitividad, Gobierno de España, project ref. MTM2015 67802-P (MINECO/FEDER, UE).

A. Appendix A

A.1. Learning the BN

For the learning process of the BN we adopt the *score-based* structure learning method (“*Greedy search-and-score*”), which is an algorithm that attempts to find the structure that maximizes the score function. We choose, as usual, the **Bayesian Information Criterion** (BIC) as score function, since it is intuitively appealing because contains a term that shows how well the model predicts the observed data when the parameter set is equal to its MLE estimation, which is the log-likelihood function, and a term that punishes for model complexity. This algorithm searches through the space of possible structures of the network; in each step, it considers the addition, elimination, or the reverse of an arc, given the structure of the previous step (with the constraint that the resultant graph be acyclic), and “greedily” choose the option that maximizes the score function, stopping when no increase is possible. In order to compute the score of the model in each step, this algorithm only needs to recompute few scores from the previous step (*local scoring updating*), which represents a huge calculation advantage. The problem with this algorithm is that we could obtain a solution that is a local (but not global) maximum of the score function. For that, we use the “iterated hill-climbing” algorithm, which carries out a local search until a local maximum is obtained, randomly perturbing it for then repeat the process. Finally, the maximum over local maxima is used as a better approximation of the global maximum.

A.2. Validation

We perform a cross-validation procedure, which is a technique for assessing how the BN model performs in the sense of correctly predicting a query variable (author variable) from an evidence given in terms of the variables of an independent (future) wildfire. That is, we want to estimate the accuracy in prediction in practice using our model. Concretely, we use *leave-one-out cross-validation*. Each round of the cross-validation procedure involves choosing a case (one different every time) and learn the corresponding BN model from the training set which is the complementary of the choosing case in the dataset, which is then used to validate the BN model. Indeed, for that case, we use as evidence the values of the crime variables C_1, \dots, C_{10} in order to predict each of the query variables A_1, \dots, A_{15} , and take note of the matches between predictions and real values of these variables in the case. We perform, then, $N = 1597$ rounds of the cross-validation, one for each of the cases in the dataset. We take into account the matches over the N rounds in combination in order to estimate predictive accuracy for each

of the author variables individually (“IPA” *Individual Predictive Accuracy* values), as well as globally (“OPA” *Overall Predictive Accuracy* value).

For each query variable, the IPA value is obtained by dividing the number of correct predictions by the total number of predictions (excluding blanks). The OPA value is obtained by dividing the total number of matches (10,543) by the total number of predictions (excluding blanks), which is 18,141. The result shows an OPA of 58.12%, that is, the 58.10% of times we predict correctly an offender characteristic. Note that in total $n \times N = 15 \times 1,597 = 23,955$ is the number of predictions (number of variables that are predicted multiplied by the number of cases in the dataset), but only 18,141 of them are recorded, which are those in which the corresponding author variable outcome was **not** a missing value. Of these, 10,543 match and the rest do not. Both the IPA and OPA values are recorded in **Table 4**.

From this table we can see which are the wildfire arsonist characteristics that are typically correctly predicted ($IPA \geq 70\%$): $A_3, A_7, A_8, A_9,$ and A_{10} . Note that all the author variables are predicted correctly more often than simply by chance, taking into account the number of levels of each one. Then, they can be used to narrow the list of suspects in an unsolved wildfire. It should also be borne in mind that, as predictions are made with our model, we choose as prediction for a variable the outcome that maximizes the probability, causing failures in prediction when the second most likely outcome has a probability close to the first one, what is really happening with some of the variables, making the accuracy not as high as would be desirable.

Finally, we also compute the “DIPA” (*Disincorporate Individual Predictive Accuracy*), which is the percentage of correct predictions, for each author variable, according to the prediction that we made for it from the evidence given by the crime variables. For example, for A_{15} , the IPA (accuracy rate) is 56.36%. If the prediction for A_{15} were “slight negligence”, what happens 60.38% of the times, then accuracy rate would be 61.29%, as consigned in **Table 12**, while if the prediction for A_{15} were “revenge”, what instead happens only 0.75% of the times, this rate plummets to 20.00%. We note that the most popular prediction for A_{15} is “slight negligence”, which is the type of motivation with which prediction is most accurate. At the opposite end, the less popular prediction is “revenge”, which is the type of motivation with the less accurate prediction.

If prediction for A_{15} were...		% Accuracy in predicting A_{15} (DIPA)
Slight negligence	(60.38%)	61.29
Gross negligence	(22.51%)	47.67
Impulsive	(9.90%)	53.33
Profit	(6.46%)	41.05
Revenge	(0.75%)	20.00

Table 12. Disincorporate Individual Predictive Accuracy (DIPA) for A_{15} . For each outcome of A_{15} , the percentage of times that the prediction for A_{15} is that value is consigned in parentheses.

A.3. The final model

The final BN model is that obtained learning from the whole dataset with $N = 1597$ cases, after validation process. The corresponding structure is that given by the DAG in **Figure 1**.

It is known that the performance of the algorithms used for learning BN is unsatisfactory if the database set does not have a sufficiently high number of cases. When can we say that the number of cases is big enough? It depends on the number of nodes and on the size of their domain, which is the set of different possible instantiations of the set formed by all the nodes. Both, number of nodes and size of their domain, are known in practice. But the sufficiency of the number of cases also depends on the underlying probability distribution, which a priori used to be unknown.

Are our $N = 1597$ cases sufficient to learn the BN model? In order to study this issue, we generate subset samples of size ranging from $m = 25$ to $m = N$ in increments of 5, at random, and from each one we learn the model and compute the BIC score function. Then, we plot the BIC score as a function of the size of the subset sample (see **Figure 2**). In this case, before attaining N a saturation point is reached (approximately at 1250), from which the BIC score does not improve significantly by increasing the size of the subset sample. As a consequence, we can say that it does seem the number of cases of the database set is big enough to learn the BN.

B. Appendix B

In Section 3.2, we have discussed the main idea in constructing archetypes by illustrating it with a simple example. There we mentioned that it was very important to be cautious applying intuition since otherwise, we could naively make the following erroneous reasoning: since the prediction for C_4 is “crops” if A_{15} is any type of negligence, and “pathway” for the rest of

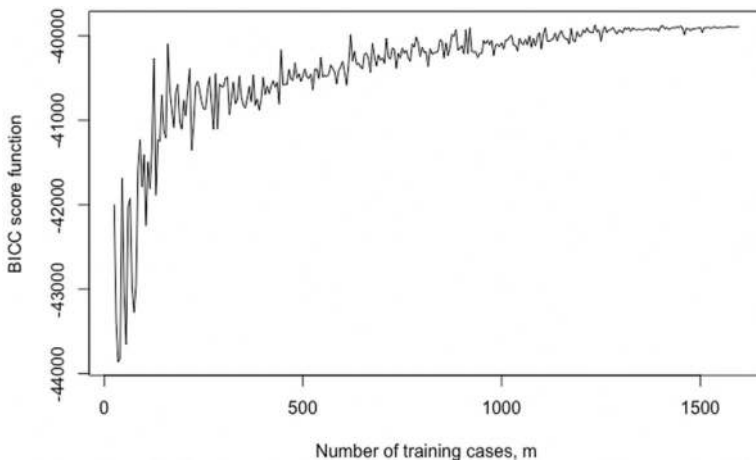


Figure 2. Evolution of the BIC score function as the number of training cases, m , increase to N .

values of A_{15} , as can be seen in **Table 13**, and since prediction for C_5 in both cases is “agricultural” (**Table 14**), then the prediction for C_5 would be the same, “agricultural”, independently of the motivation. Actually this is not so. Indeed, since the geodesic joining A_{15} and C_5 has distance 2, passing through the only one intermediate node C_4 , we can easily compute the probability of each value of C_5 conditioned to A_{15} from the CPT of C_5 conditioned to C_4 (**Table 14**), and that of C_4 conditioned to A_{15} (**Table 13**).

In this simple case it is possible to show the calculations, and we do it “by hand” to exemplify the procedure. For example, we can compute $P(C_5 = \text{“agricultural”}/A_{15} = \text{“slightnegligence”})$ by using the Conditioned Law of Total Probability in the following way, by conditioning to all the possible outcomes of C_4 :

$$\begin{aligned}
 &P(C_5 = \text{“agricultural”}/A_{15} = \text{“slightnegligence”}) = \\
 &P(C_5 = \text{“agricultural”}/C_4 = \text{“pathway”})P(C_4 = \text{“pathway”}/A_{15} = \text{“slightnegligence”}) + \\
 &P(C_5 = \text{“agricultural”}/C_4 = \text{“road”})P(C_4 = \text{“road”}/A_{15} = \text{“slightnegligence”}) + \\
 &P(C_5 = \text{“agricultural”}/C_4 = \text{“houses”})P(C_4 = \text{“houses”}/A_{15} = \text{“slightnegligence”}) + \\
 &P(C_5 = \text{“agricultural”}/C_4 = \text{“crops”})P(C_4 = \text{“crops”}/A_{15} = \text{“slightnegligence”}) + \\
 &P(C_5 = \text{“agricultural”}/C_4 = \text{“interior”})P(C_4 = \text{“interior”}/A_{15} = \text{“slightnegligence”}) + \\
 &P(C_5 = \text{“agricultural”}/C_4 = \text{“foresttrack”})P(C_4 = \text{“foresttrack”}/A_{15} = \text{“slightnegligence”}) + \\
 &P(C_5 = \text{“agricultural”}/C_4 = \text{“others”})P(C_4 = \text{“others”}/A_{15} = \text{“slightnegligence”}) = \\
 &0.38 \times 0.10 + 0.18 \times 0.04 + 0.30 \times 0.07 + 0.75 \times 0.38 \\
 &+ 0.20 \times 0.15 + 0.11 \times 0.05 + 0.22 \times 0.21 \cong 0.43
 \end{aligned} \tag{6}$$

Similarly, we can find the rest of conditioned probabilities and write the CPT of C_5 conditioned to A_{15} (**Table 15**), which coincides with the product of matrices given by **Tables 14** and **13**, in

$C_4 \downarrow A_{15} \rightarrow$	Slight negligence	Gross negligence	Impulsive	Profit	Revenge
Pathway	0.10	0.11	0.31	0.29	0.33
Road	0.04	0.04	0.29	0.11	0.22
Houses	0.07	0.05	0.02	0.01	0.04
Crops	0.38	0.35	0.03	0.14	0.02
Interior	0.15	0.16	0.09	0.16	0.04
Forest track	0.05	0.07	0.14	0.16	0.27
Others	0.21	0.22	0.12	0.13	0.08
	1.00	1.00	1.00	1.00	1.00

Table 13. Conditional probability table (CPT) of C_4 to A_{15} .

$C_5 \downarrow C_4 \rightarrow$	Pathway	Road	Houses	Crops	Interior	Forest track	Others
Agricultural	0.38	0.18	0.30	0.75	0.20	0.11	0.22
Forestry	0.21	0.45	0.17	0.12	0.48	0.51	0.30
Livestock	0.29	0.18	0.09	0.11	0.26	0.28	0.27
Interface	0.07	0.18	0.34	0.01	0.03	0.01	0.12
Recreational	0.05	0.01	0.10	0.01	0.03	0.09	0.09
	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 14. Conditional probability table (CPT) of C_5 to C_4 .

$C_5 \downarrow A_{15} \rightarrow$	Slight negligence	Gross negligence	Impulsive	Profit	Revenge
Agricultural	0.43	0.42	0.26	0.32	0.25
Forestry	0.26	0.27	0.35	0.33	0.36
Livestock	0.19	0.20	0.24	0.24	0.25
Interface	0.07	0.07	0.10	0.06	0.09
Recreational	0.05	0.04	0.05	0.05	0.05
	1.00	1.00	1.00	1.00	1.00

Table 15. Conditional probability table (CPT) of C_5 to A_{15} computed by using the conditional law of total probability.

this order. The highest probability in each column is in boldface and corresponds to the prediction for C_5 given each evidence in terms of A_{15} , as stated in **Table 8**. We can see that the prediction for C_5 is “agricultural” only if motivation is “negligence” (either slight or gross), being “forestry” otherwise.

On the other hand, for A_{13} the dependency chaining is more subtle and much more harder to follow *by hand*, so we give up on it and only carry out predictions by using the BN model with **R**.

Author details

Rosario Delgado^{1*}, José-Luis González², Andrés Sotoca³ and Xavier-Andoni Tibau⁴

*Address all correspondence to: delgado@mat.uab.cat

1 Departament de Matemàtiques, Universitat Autònoma de Barcelona, Cerdanyola del Vallés, Spain

2 Gabinete de Coordinación y Estudios, Secretaría de Estado de Seguridad, Madrid, Spain

3 Sección de Análisis del Comportamiento Delictivo, Unidad Técnica de Policía Judicial, Madrid, Spain

4 Research group “Quantitative Methods in Criminology” of the Universitat Autònoma de Barcelona, Cerdanyola del Vallés, Spain

References

- [1] FAO: Fire Management—Global Assessment 2006. A thematic study prepared in the framework of the global forest resources assessment 2005. Rome: FAO Forestry Paper 151; 2007
- [2] San-Miguel-Ayanz J, Moreno JM. Camia a (2013) analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management*. 2013;**294**:11-22. DOI: 10.1016/j.foreco.2012.10.050
- [3] Turco M, Llasat M, von Hardenberg J, Provenzale A. Climate change impacts on wildfires in a mediterranean environment. *Climatic Change* 2014;**125**(3–4):369-380. DOI: <http://dx.doi.org/10.1007/s10584-014-1183-3>
- [4] Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente, Los Incendios Forestales en España: Avance informativo. 1 de enero al 31 de diciembre de 2016; 2017. (In Spanish.) http://www.mapama.gob.es/es/desarrollo-rural/estadisticas/iiff_2016_def_tcm7-454599.pdf
- [5] Thompson MP, Scott Helmbrecht JD, Calvin DE. Integrated wildfire risk assessment: Framework development and application on the Lewis and Clark National Forest in Montana, USA. *Integrated Environmental Assessment and Management*. 2012;**9**(2):329-342
- [6] Penman TD, Bradstock RA, Price O. Modelling the determinants of ignition in the Sydney Basin, Australia: Implication for future management. *International Journal of Wildland Fire*. 2013;**22**:469-478
- [7] Adab H, Kanniah KD, Solaimani K. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Natural Hazards*. 2013;**65**:1723-1743
- [8] Cozens P, Christensen W. Environmental criminology and the potential for reducing opportunities for bushfire arson. *Crime Prevention and Community Safety*. 2011;**13**(2):119-133. DOI: 10.1057/cpcs.2010.24
- [9] Ministerio de Agricultura, Alimentación y Medio Ambiente, Los Incendios Forestales en España: Avance informativo. 1 de enero al 31 de diciembre de 2015; 2016. (In Spanish.) http://www.mapama.gob.es/es/desarrollo-rural/estadisticas/iiff_2015_def_tcm7-416547.pdf
- [10] Soeiro C, Guerra R. Forest arsonists: Criminal profiling and its implications for intervention and prevention. *European Police Science and Research Bulletin*. Winter 2014/15; Issue 11:34-40
- [11] Sotoca A, González JL, Fernández S, Kessel D, Montesinos O, Ruz MA. Perfil del incendiario forestal español: aplicación del perfilamiento criminal inductivo. *Anuario de Psicología Jurídica*. 2013;**23**:31-38. (In Spanish.)
- [12] Delgado R, González JL, Sotoca A, Tibau XA. A Bayesian network profiler for wildfire arsonists. In: Pardalos P., Conca P., Giuffrida G., Nicosia G. (editors.) *Machine Learning, Optimization and Big Data*. MOD 2016. Lecture Notes in Computer Science, Cham: Springer; 2016;**10122**:379-390. DOI: 10.1007/978-3-319-51469-7_31

- [13] Baumgartner KC, Ferrari S, Palermo G. Constructing Bayesian networks for criminal profiling from limited data. *Knowledge-Based Systems*. 2008;**21**:563-572
- [14] Baumgartner KC, Ferrari S, Salfati CG. Bayesian network modeling of offender behavior for criminal profiling. In: *Proceedings of the 44th IEEE Conference on Decision and Control 2005 and 2005 European Control Conference*. 2005. pp. 2702-2709. DOI: 10.1109/CDC.2005.1582571
- [15] Korb KB, Nicholson AE. *Bayesian Artificial Intelligence*. 2nd ed. Taylor & Francis Group: CRC Press; 2011
- [16] Ticehurst JL, Newham LTH, Rissik D, Letcher RA, Jakeman AJ. A BN approach for assessing the sustainability of coastal lakes in New South Wales, Australia. *Environmental Modelling and Software*. 2007;**22**(8):1129-1139
- [17] Adusei-Poku K. *Operational Risk management – Implementing a BN for Foreign Exchange and Money Market Settlement [PhD thesis]*. Göttinger University; 2005. www.statistics.uni-goettingen.de/fileadmin/cfs/Dokumente/Dissertations/diss_adusei-poku.pdf
- [18] Walshe T, Burgman MA. Framework for assessing and managing risks posed by emerging diseases. *Risk Analysis*. 2010;**30**(2):236-249
- [19] Borsuk ME, Stow CA, Reckhow KH. A BN of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modeling*. 2004;**173**:219-239
- [20] Pollino CA, Woodberry O, Nicholson A, Korb K, Hart BT. Parameterization and evaluation of a BN for use in an ecological risk assessment. *Environmental Modelling and Software*. 2007;**22**:1140-1152
- [21] Spiegelhalter DJ. Incorporating Bayesian ideas into healthcare evaluation. *Statistical Science*. 2004;**19**:156-174
- [22] Cruz-Ramrez N, Acosta-Mesa HG, Carrillo-Calvet H, Alonso Nava-Fernández L, Barrientos-Martinez RE. Diagnosis of breast cancer using BN: A case study. *Computers in Biology and Medicine*. 2007;**37**:1553-1564
- [23] Lee C, Lee KJ. Application of BN to the probabilistic risk assessment of nuclear waste disposal. *Reliability Engineering and System Safety*. 2006;**91**(5):515-532
- [24] Delgado R, Tibau XA. Las Redes Bayesianas como herramienta para la evaluación del riesgo de reincidencia: Un estudio sobre agresores sexuales. *Revista Española de Investigación Criminológica*. 2015;**13**, paper 1. (In Spanish.)
- [25] Papakosta P, Straub D. A Bayesian network approach to assessing wildfire consequences. In: *Proceedings ICOSSAR 2013*. New York; 2013. www.era.bgu.tum.de/fileadmin/w00bkd/www/Papers/2013_ICOSSAR_PapakostaStraub.pdf
- [26] Dlamini WM. A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling & Software*. 2010;**25**:199-208
- [27] Dlamini WM. Application of Bayesian networks for fire risk mapping using GIS and remote sensing data. *GeoJournal*. 2011;**76**:283-296

- [28] Dowden C, Bennell C, Bloomfield S. Advances in offender profiling: A systematic review of the profiling literature published over the past three decades. *Journal of Police and Criminal Psychology*. 2007;**22**:44-56
- [29] Snook B, Cullen RM, Bennell C, Taylor PJ, Gendreau P. The criminal profiling illusion: What's behind the smoke and mirrors? *Criminal Justice and Behavior*. 2008;**35**:1257-1276
- [30] Snook B, Haines A, Taylor P, Bennell C. Criminal profiling belief and use: A study of Canadian police officer opinion. *Canadian Journal of Police and Security Services*. 2007; **5**(3/4):1-11
- [31] Torres A, Boccaccini M, Miller H. Perceptions of the validity and utility of criminal profiling among forensic psychologists and psychiatrists. *Professional Psychology: Research and Practice*. 2006;**37**(1):51-58
- [32] Alison L, Rainbow L. *Professionalizing Offender Profiling* London. UK: Routledge; 2011
- [33] Shye S. Nonmetric multivariate models for behavioural actions systems. In Canter DV, editor. *Facet Theory: Approaches to Social Research*. New York: Springer Verlag; 1985. p. 97-148
- [34] Canter D, Fritzon K. Differentiating arsonists: A model of firesetting actions and characteristics. *Legal and Criminological Psychology*. 1998;**3**:73-96
- [35] Fritzon K, Canter D, Wilton Z. The application of an action system model to destructive behaviour: The examples of arson and terrorism. *Behavioural Sciences & the Law*. 2001; **19**:657-690
- [36] Fritzon K. An examination of the relationship between distance travelled and motivational aspects of firesetting behaviour. *Journal of Environmental Psychology*. 2001;**21**:45-60
- [37] Kocsis RN, Cooksey RW. Criminal psychological profiling of serial arson crimes. *International Journal of Offender Therapy and Comparative Criminology*. 2002;**46**(6):631-656
- [38] Wachi T, Watanabe K, Yokota K, Suzuki M, Hoshino A, Sato A, Fujita G. Offender and crime characteristics of female serial arsonists in Japan. *Journal of Investigative Psychology and Offender Profiling*. 2007;**4**:29-52
- [39] Viegas E, Soeiro C. Perfis psicossociais dos incendiários portugueses. Propostas para prevenção. *Jornadas sobre Investigación Criminal de Incendios Forestales*, marzo 2007, Santiago de Compostela. 2007. (In Portuguese)
- [40] Collin C. Criminological psychology. In: Maguire M, Morgan R, Reiner R (editors). *The Oxford Handbook of Criminology*, 5th edition. USA: Oxford University Press; 2017. pp. 81-112
- [41] Freeman LC. A set of measures of centrality based upon betweenness. *Sociometry*. 1977; **40**:35-41
- [42] Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. *Social Networks*. 2006;**28**:466-484