IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

Modeling crime events by d-separation method

To cite this article: R Aarthee and D Ezhilmaran 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **263** 042145

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042145 doi:10.1088/1757-899X/263/4/042145

Modelling crime events by d-separation method

R Aarthee and D Ezhilmaran

Department of Mathematics, School of Advanced Sciences, VIT University, Vellore.

aarsi229@gmail.com.

Abstract. Problematic legal cases have recently called for a scientifically founded method of dealing with the qualitative and quantitative roles of evidence in a case [1].To deal with quantitative, we proposed a d-separation method for modeling the crime events. A d-separation is a graphical criterion for identifying independence in a directed acyclic graph. By developing a d-separation method, we aim to lay the foundations for the development of a software support tool that can deal with the evidential reasoning in legal cases. Such a tool is meant to be used by a judge or juror, in alliance with various experts who can provide information about the details. This will hopefully improve the communication between judges or jurors and experts. The proposed method used to uncover more valid independencies than any other graphical criterion.

1. Introduction

A d-separation, a graphical criterion for identifying independencies in a DAG, is shown to uncover more valid independencies than any other criterion [8]. A d-separation remains for directional (or) dependence - separation and brings out a sense of the causal directionality fundamental the forecast of statistical independence or separation following the introduction of particular controls. According to geneticist Wright, Directed Acyclic Graphs using networks have a long and rich custom. He built up a technique called path analysis. After some time casual models play a major role in financial aspects, human science, and psychology. Howard et.al stated that by using DAG we can produce decision analysis and contain both chance nodes and decision nodes. Some analysts used d-separation for modelling of offender behaviour for criminal profiling [2].Because of significant and powerful deteriorations, some researchers had given the name as Recursive models to such networks. For evidential thinking Pearl named the system as Bayesian Belief Networks.

In 1980s PC analysts in the laboratory wearing down the problem of securing and getting ready questionable data adequately in erroneously clever masters handled this numerical issue. Pearl and his accomplices comprehended that unverifiable information could be secured fundamentally more adequately by exploiting conditional independence and to encode probabilities and the conditional independence relations between probabilities and conditional independence they used directed acyclic graph. For detailing conditional independence relations involved by the graph is studied by the method called d-separation [9]. In a special class of discrete causal models d-separation accurately encodes the independencies involved by directed graphs with or without cycles which are demonstrated by Pearl in 1996.Additionally, some researchers demonstrated for linear models with correlated errors also d-separation method works out very well. Fenton et al used for decision making [3].So it ought to be evident that d-separation is a focal thought in the hypothesis of graphical causal models. A d-separation connects to control variables, partial correlations, casual structuring, and even to a potential mistake in regression [4].

2. Preliminaries

2.1. Definition [Pearl 1988]
If J, K, and L are three disjoint subsets of nodes in a DAG D, then L is said to d -separate J from K, denoted I (J, L, K) $_D$, if and only if there is no trail t between a node in J and a node in K along which every head-to-head node (with respect to r) either is or has a descendant in L and every node that delivers an arrow along t is outside L. A trail satisfying the two conditions above is said to be active, otherwise it is said to be blocked (by L).

2.2. Example
In figure 1, for instance, 1 = {A} and 2 = {B} are d-separated by 3 = {C}; the path B←A→C is blocked by A∈ 3 while the path B →D←C is blocked because D and all its descendants are outside 3. Thus I(B, A, C) is graphically-verified by F. Be that as it may, 1 and 2 are not d- separated by R'={ A, E} because the path B →D ←C is rendered active. Consequently, I(B,{A,E},C) is graphically-unverified by F; by virtue of E, a descendant of D, being in 3. Learning the value of the consequence E, renders its causes B and C dependent, like opening a pathway along the converging arrows at D.

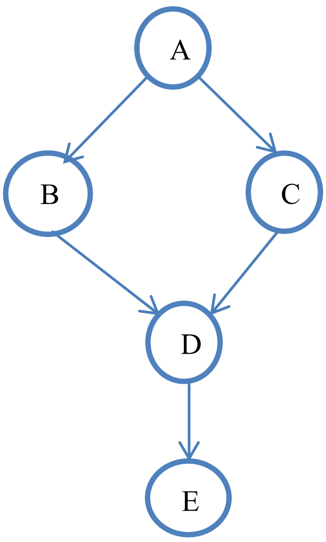
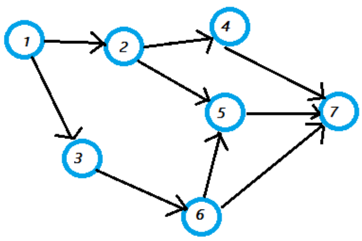


Figure 1. Example graph for D-separation

2.3. Directed acyclic graph
Directed acyclic graphs (DAGs) are used to model probabilities, connectivity, and causality. A "graph" in this sense means a structure made of nodes and edges.



IOP Conf. Series: Materials Science and Engineering **263** (2017) 042145 doi:10.1088/1757-899X/263/4/042145

Figure 2. Directed acyclic graph

2.4. The inverted-V graph

Let x and y is conditionally independent given z:

$$\mathrm{P(x, y/ z)} = P\mathrm{(x/ y, z)\,P(y/ z)} = P\mathrm{(x/ z)\,P(y/ z)} \tag{1}$$

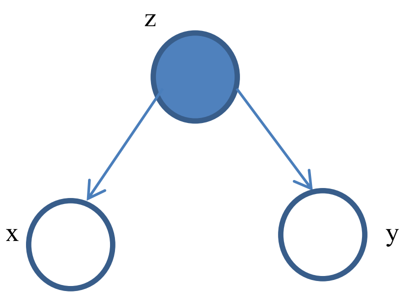


Figure 3.x and y are conditionally independent given z

The above-mentioned graph satisfies the condition z is the parent of x and y. By using the Bayes chain rule we have,

$$\mathrm{P(x, y/ z)} = P\mathrm{(x/ y, z)\,P(y/ z)} \tag{2}$$

1. z is the parent of x.
2. y is non-descendant of x. So we can say that:

$$\Rightarrow \quad \mathrm{P(x/ y, z)} = P\mathrm{(x/ z)} \tag{3}$$

$$\Rightarrow \quad \mathrm{P(x, y/ z)} = P\mathrm{(x/ z)\,P(y/ z)} \tag{4}$$

Hence x and y are conditionally independent given z.

Now consider z is not an observed value.

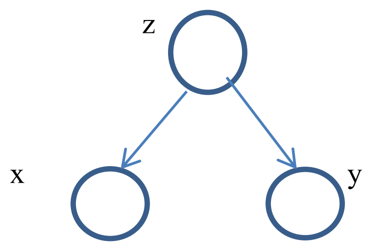


Figure 4. z is not an observed value

$$\mathrm{P(x, y, z)} = P\mathrm{(x/ z)\,P(y/ z)\,P(z)} \tag{5}$$

$P(x, y, z) = P(x/z)\,P(y/z)\,P(z)$ .In general, this does not factorize into the product of $P(x) \times P(y)$, and so x and y are not conditionally independent$|\,z$.

2.5. *The V graph problem*

2.5.1. *Case 1: z is not an observed variable*

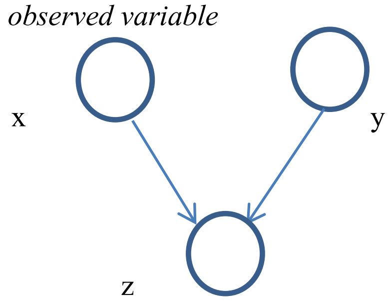


Figure 5. All value unobserved

$$P(x, y, z) = P(x)\,P(y/x)\,P(z/x, y) \tag{6}$$

But since x is not a parent or descendant of y, we have,

$$P(x, y, z) = P(x)\,P(y)\,P(z/x, y) \tag{7}$$

Now we can find $P(x, y)$ $P(x, y)$

$$P(x, y) = \sum_z P(x, y, z) \tag{8}$$

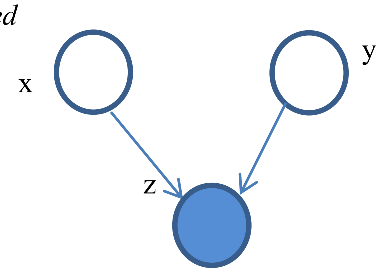
$$= \sum_z P(x)\,P(y)\,P(z/x, y)$$

$$= \sum_z P(x)\,P(y)\,P(z/x, y)$$

$$= P(x)\,P(y) \sum_z P(z/x, y)$$

$$P(x, y) = P(x)\,P(y) \tag{9}$$

2.5.2. *Case 2: z is observed*



IOP Conf. Series: Materials Science and Engineering **263** (2017) 042145 doi:10.1088/1757-899X/263/4/042145

Figure 6. All value unobserved

$$P(x, y/z) = \frac{P(x, y, z)}{P(z)} \tag{10}$$

$$P(x, y/z) = \frac{P(x)\,P(y)\,P(z/x, y)}{P(z)} \tag{11}$$

Which is general does not factorize into the product of, $P(x) \times P(y)$ and so x and y are not conditionally independent$|\,z$.

If we have to solve all the marginalizing summations to reach the final result means, the role of graphs is lacking. For understanding without any calculations, we should understand the concept for d-separation. A d-separation is nothing but graph terminology to comment on independence [2].

3. D-separation

The "d" in d-separation and d-association remain for directional and dependence. Accordingly, if two factors are d-separated with respect to an arrangement of factors S in a directed graph, at that point they are conditional independent on S most likely disseminations such a graph can represent [6]. Approximately, two factors P and Q are conditional independent on S if learning of P provides you no additional details of Q in one case you know about S.

3.1. Conditioning set is empty

Now let us concentrate on what makes a path active or inactive. If each vertex on the path is active is said to be path is active. With respect to the arrangement of different vertices Z, paths and vertices on these paths are active or inactive. To start with, let look at when things are dynamic or inert in respect to an unfilled Z. If we take a bigger part of the possible undirected paths between a pair of variables X and Y that experience a third variable Z [5].

$$A.\,X \rightarrow Z \rightarrow Y$$
$$B.\,X \leftarrow Z \leftarrow Y$$
$$C.\,X \leftarrow Z \rightarrow Y$$
$$D.\,X \rightarrow Z \leftarrow Y$$

Let us consider X, Y and Z are three paths. In the above mentioned first case, X and Y are directly passing through Z so it is said to be directed path. Similarly, in the second case Y to X are passing through Z that is also called directed path, and in the third case a pair of directed paths from Z to X and from Z to Y. In the event that we decipher these ways causally, in the main case X is an indirect cause for Y, in the next case Y is an indirect cause for X, and in the final case, Z is a typical cause for X and Y. Every one of the three of these causal circumstances offers ascent to affiliation, or reliance, amongst X and Y, and each of the three of these undirected paths is dynamic in the hypothesis of d-separation. On the off chance that we translate the fourth case casually, at that point X and Y have a typical impact in Z, however no causal association between them. In the hypothesis of d-separation, the fourth way is latent. Hence, when the conditioning set is empty, just ways that relate to the causal association are active. In the initial three, Z is a non-collider on the path, and in the fourth Z is a collider. At the point when the conditioning set is vacant, non-colliders are active. Naturally, non-colliders transmit data (reliance). At the point when the conditioning set is unfilled, colliders are inactive. Naturally, colliders don't transmit information.

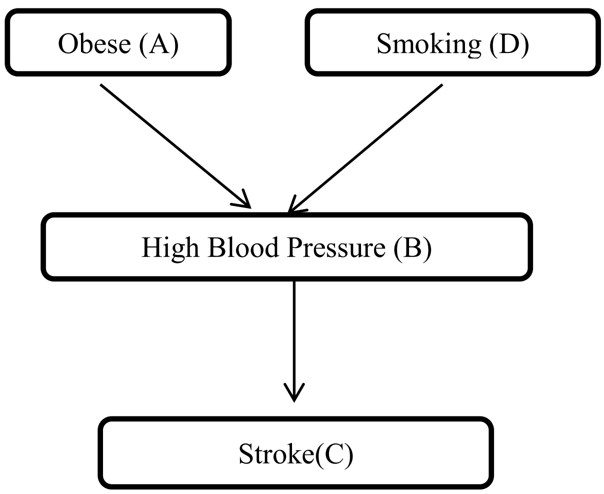
3.2. Conditioning set is not empty
Now consider, conditioning set is not empty. At the point when a vertex is in the conditioning set its status as for being dynamic or inert flip-flops. Consider the four paths above once more, however now gives us a chance to see whether the variables X and Y are d-separated by Z

$$\text{A.}X \rightarrow Z \rightarrow Y$$
$$\text{B.}X \leftarrow Z \leftarrow Y$$
$$\text{C.}X \leftarrow Z \rightarrow Y$$
$$\text{D.}X \rightarrow Z \leftarrow Y$$

In the initial three ways, Z was active when the conditioning set was empty, so now Z is latent on these paths. To settle instincts, again decipher the paths causally. In the main case the path from X to Y is obstructed by conditioning on the intermediary Z, comparable, on the off chance that 2, and on the off chance that 3, you are conditioning on a typical reason, which makes the impacts free. In the fourth case, Z is a collider and hence inactive when the conditioning set is void, so is currently active. This can likewise be made instinctive by considering what happens when you take a gander at the connection between two free causes after you condition on a typical impact. Presently by utilizing the d-separation idea, we will attempt to make a model for crime events.

4. Modelling of d-separation

The meaning of d-separation is best propelled by seeing DAGs as a portrayal of causal connections. Assigning a node for each factor and appointing a connection between each reason to each of its immediate outcomes characterizes a graphical portrayal of a causal progressive system [3]. For instance, "Obese" (A), "High Blood Pressure" (B) and "Stroke" (C) c are all around spoke to by a three hub chain, from A through B to C; it shows that either obese or high blood pressure could cause stroke, yet high blood pressure is assigned as the immediate reason. Obese is one of the reasons for stroke, but high blood pressure is the main reason. In addition, knowing the state of the high blood pressure restore "obese" and "stroke" independent, and this is spoken to graphically by a d – separation condition, $I(A, C, B)_D$, demonstrating node A and B isolated from each other by node C. Consider, "Smoking" (D) is considered another direct cause for high blood pressure, as in figure 1. An induced dependency exists between the two events that may cause the High blood pressure: "obese" and "Smoking". In spite of the fact that they seem associated in Figure 1, these announcements are possibly free and wind up noticeably subordinate once we discover that because of high blood pressure or because of heat waves. An expansion in our confidence in either cause would diminish our faith in alternate as it would "clarify away" the perception.



IOP Conf. Series: Materials Science and Engineering **263** (2017) 042145 doi:10.1088/1757-899X/263/4/042145

Figure 7. Graphical Representation of d-separation

5. Case Study

The complainant was accepting debilitating and vulgar messages from obscure individuals. The messages contained the complainant's revolting transformed photos. The denounced debilitated to post these on obscene Web sites.

From the above definition, we deal the above case as follows. Let us consider the girl who has a complaint about the accused is financially strong (α).The accused has a love affair with the girl (β). The blamed debilitated to post these on explicit sites (γ).These are illustrated by a three node chain from α through β to γ; it indicates either lavishness or attractiveness could cause threatening. Yet love affair is destined as the direct cause; lavishness could cause someone to make threatening for making money. Moreover knowing the background of the girl, renders lavishness and threatening are free and this is spoken to graphically by a d-separation condition $I(\alpha,\gamma,\beta)_D$ showing node α and β isolated from each other by node γ. Assume that attractiveness(δ) is considered another direct cause for a love affair. An initiated reliance exists between the two occasions that may cause for the love affair: “Lavishness” and “Attractiveness”. Although they appear connected, the above case study is marginally independent and become dependent. When we discover that the accused was truly loved the girl or to revenge the girl for rejection. When we analyze the above case study we conclude that first we increase our belief by assuming Lavishness is the main reason for a love affair and slowly we decrease our belief and assuming Attractiveness is also another reason.

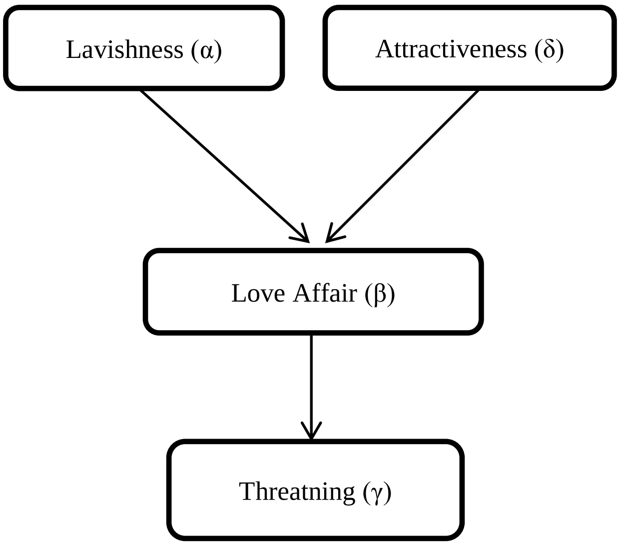


Figure 8.Graphical representation for the case study

6. Conclusions

Legitimate cases incorporate discouraging evidence and with the progression of a software support instrument at the top of the priority list, a formal establishment for evidential thinking is required. In this work, a model called d-separation is introduced for modeling the crime events. The subsequent model is intended to help a judge or jury, keeping up a decent review of the connections between important factors for a situation. In future, we can extend our work by using d-separation algorithms and we can able to model in a very precise way.

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042145 doi:10.1088/1757-899X/263/4/042145

References

[1] Vlek C, Prakken H, Renooij S and Verheij B 2013 Modeling Crime Scenarios in a Bayesian Network *In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law* pp 150-159
[2] Baumgartner K C, Ferrari S and Salfati C G 2005 Bayesian Network Modeling of Offender Behavior for Criminal Profiling *European Control Conference CDC-ECC'05 44th IEEE Conference on IEEE* pp 2702-2709
[3] Fenton N and Neil M 2011 The Use of Bayes and Causal Modelling in Decision Making, Uncertainty and Risk CEPIS Upgrade **12(5)** pp 10-21
[4] Hayduk L *et al* 2003 *Pearl's D-separation One more step into causal thinking. Structural Equation Modeling* **10(2)** pp 289-311
[5] Geiger D and Pearl J 2013 On the Logic of Causal Models *arXiv preprint* pp 1304-2355
[6] Shachter R D 1998 *In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* pp 480-487
[7] Geiger D, Verma S T S and Pearl J 2013 *arXiv.org* pp 1304-1505
[8] https://www.andrew.cmu.edu/user/scheines/tutor/d-sep.html