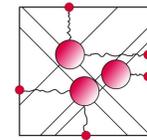


# Bayesian networks for victim identification on the basis of DNA profiles



C.J. Bruijning-van Dongen<sup>1</sup>, K. Slooten<sup>1</sup>, W. Burgers<sup>2</sup>, W. Wiegeler<sup>2</sup>

<sup>1</sup>Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB Den Haag, The Netherlands

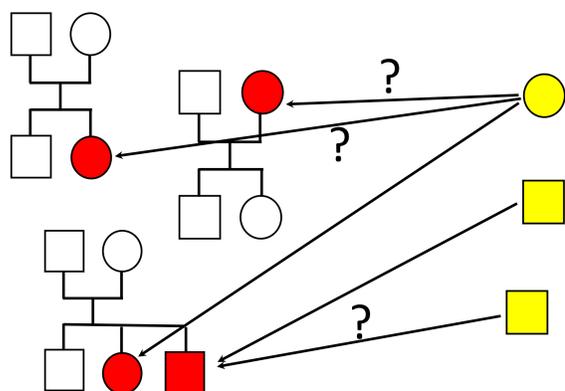
<sup>2</sup>SNN Adaptive Intelligence, Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands

## ABSTRACT

We have developed software to improve screening and matching routine for victim identification based on DNA profiles. The software (“Bonaparte”) uses Bayesian networks for analysis. It is designed for effective handling of the identification process in case of a large disaster with many victims and can be applied in the missing person program.

## INTRODUCTION

In DVI, the problem is finding matches between lists of pedigrees (which contain profiles from relatives) and lists of unidentified individuals as depicted by the figure below—where yellow indicates an Unidentified Individual (UI) and red a Missing Person (MP)



A measure for how well the UI fits in a MP slot (the match) is given by the likelihood ratio (LR)

$$LR = \frac{P(E|H_p)}{P(E|H_d)}$$

where each  $P(E|H_x)$  denotes a conditional probability;  $H_p$  (cq.  $H_d$ ) is the hypothesis that missing person MP is (cq. is not related to) UI;  $E$  contains a DNA-profile of UI, a pedigree of which MP is a member and which contains at least one typed relative. Bonaparte uses so-called Bayesian networks to compute these quantities.

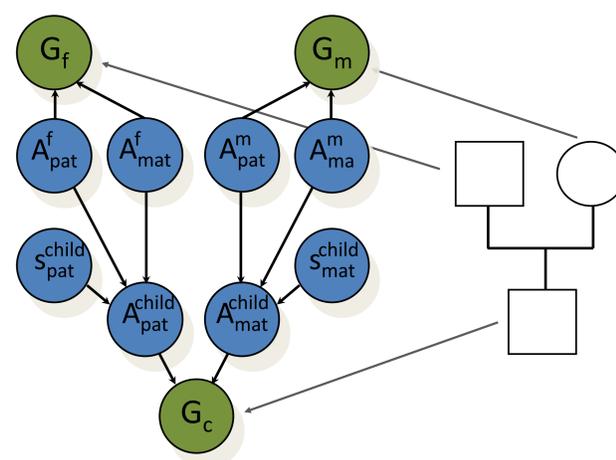
## BAYESIAN NETWORKS

- Very well suited to model statistical relations of genetic material in pedigrees
- Whole pedigrees of relatives of the missing persons are used in the screening phase (fewer false hits)
- Analysis tool is transparent and flexible; easily incorporate other factors such as new models for mutation, size bias corrections, measurement error probabilities, missing data, etc.

The joint distribution of a Bayesian network is the product of the conditional probability distributions

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | x_{pa(i)}) \quad (1)$$

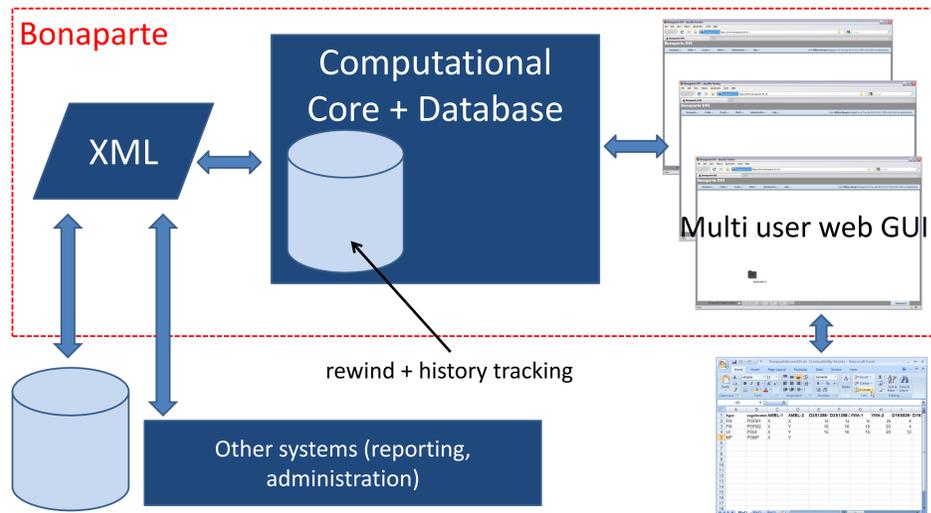
This joint distribution is used for calculating the LR. As an example of a Bayesian network consider a simple father-mother-child pedigree:



The network on the left hand side is a graphical representation of equation (1) and is derived from the pedigree on the right. It visualizes the dependency of the genotype of the child ( $G_c$ ) on the alleles of the parents. The model variables here are the alleles of the individuals ( $A$ 's) and segregation indicators ( $s$ ) that pick the paternal or maternal allele from the respective parent. The nodes in the network are probability tables; they specify a probability that the node is in a certain state given the state of its parents. Networks like this can be analyzed using a junction tree algorithm. Bonaparte generates these networks automatically.

## BONAPARTE

The clients interact with Bonaparte via web browsers. Bonaparte has its own internal database that can be rewind to any point back in time and it also keeps track of object modification (who, when and what). The system can import data from any data source via it's XML import protocol, or by using Excel files via the web GUI. Pedigrees can be created using a drag and drop pedigree editor (or imported as XML).



More information about Bonaparte and a demonstration version can be found at <http://www.bonaparte-dvi.com>