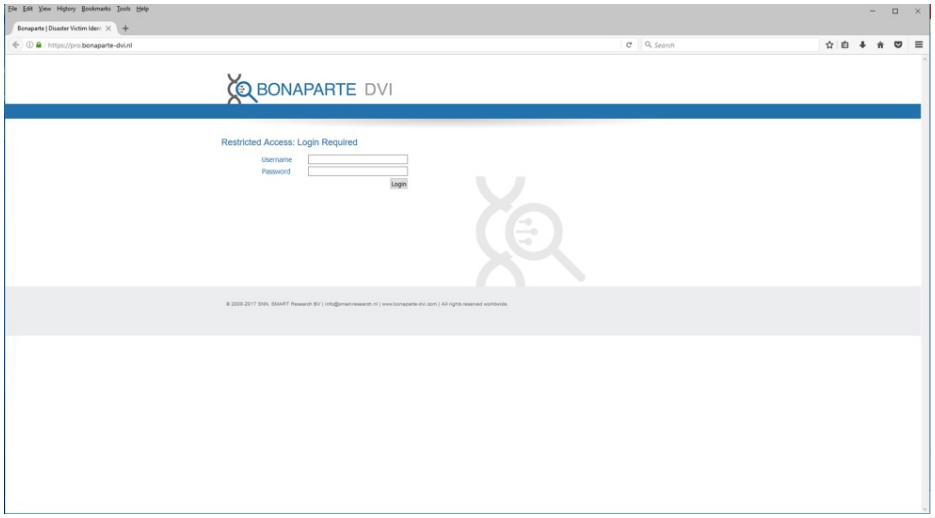


BONAPARTE

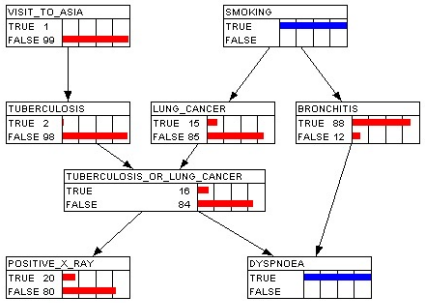
Forensic software for large scale DNA matching

Wim Wiegerinck
SMART Research BV, Nijmegen, The Netherlands

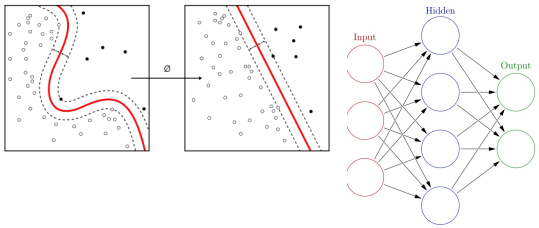
ABOUT US

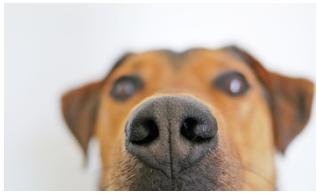
Statistical Models and Algorithms

- **innovating applications**



$$f_X(x|Y=y) = \frac{f_Y(y|X=x)\, f_X(x)}{\int_{-\infty}^{\infty} f_Y(y|X=\xi)\, f_X(\xi)\, d\xi}$$



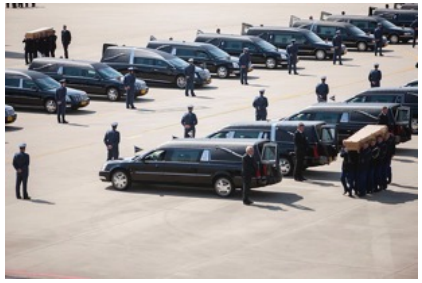


ABOUT BONAPARTE

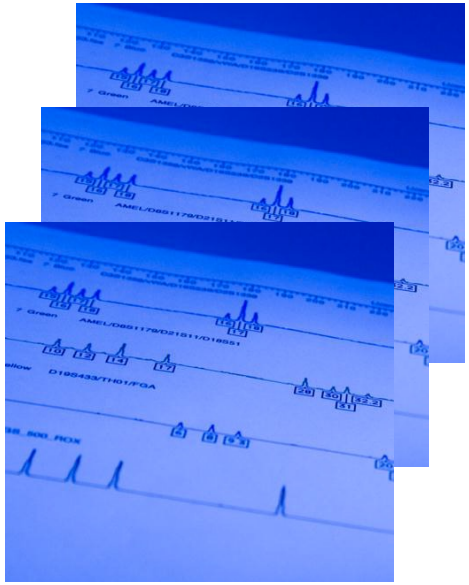
Netherlands Forensics Institute NFI

Identification of unidentified human remains

DVI: Disaster Victim Identification

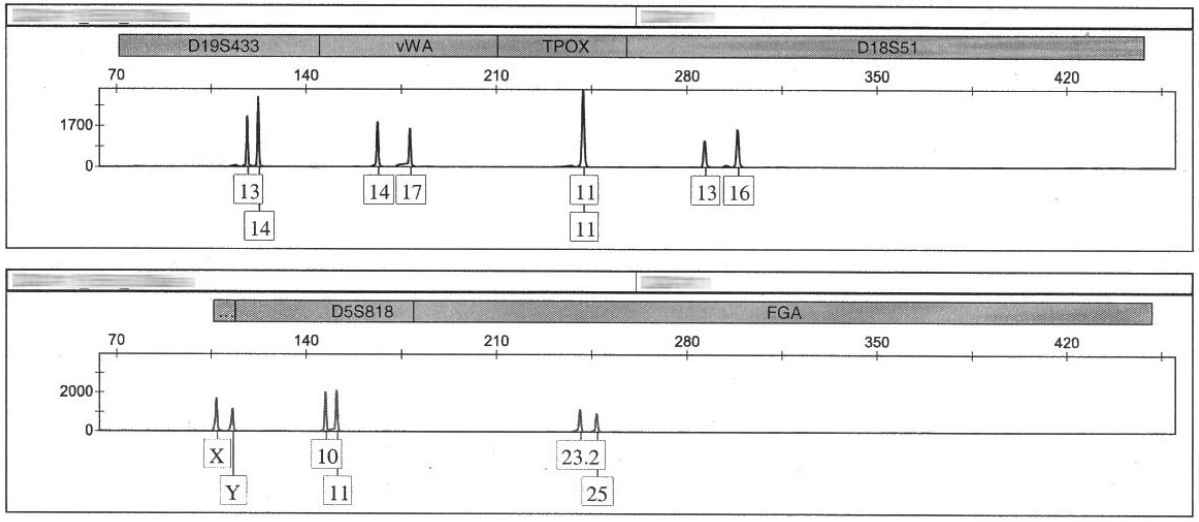


DNA victim identification workflow



victims + relatives → DNA → LAB → Processing → DNA profiles → **Analysis**

DNA Profile

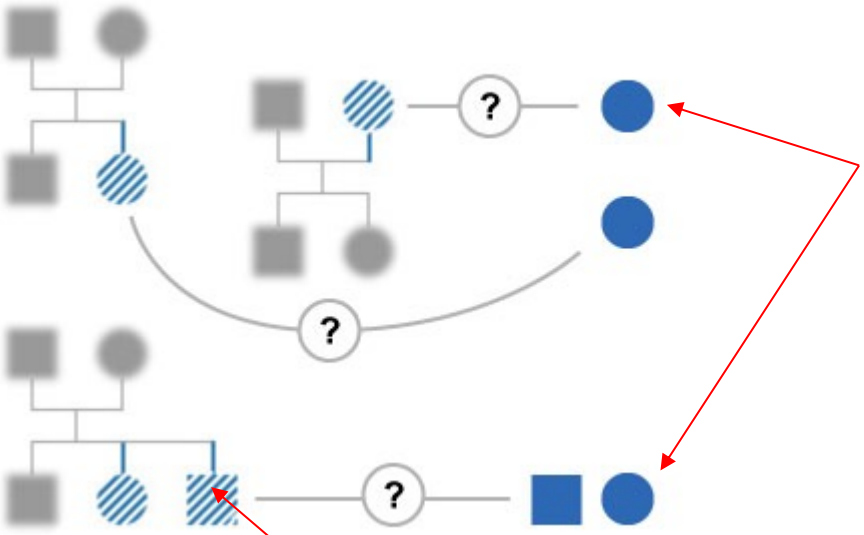


D19s433	13,14
VWA	14,17
TPOX	11,11
D18S51	13,16
D5S818	10,11
FGA	23.2,25
AMEL	X,Y

- D19S433 = {13,14} , …
- Chromosomes come in pairs: one from father, one from mother

The DVI matching problem:

which victim matches with which family?



Unknown VICTIMS
with DNA profiles

Matching
VICTIMS with MISSINGS

requires:

Statistical analysis
of DNA profiles
in family trees

Family trees of
relatives *with*
reference DNA profiles

MISSING persons
without DNA profile

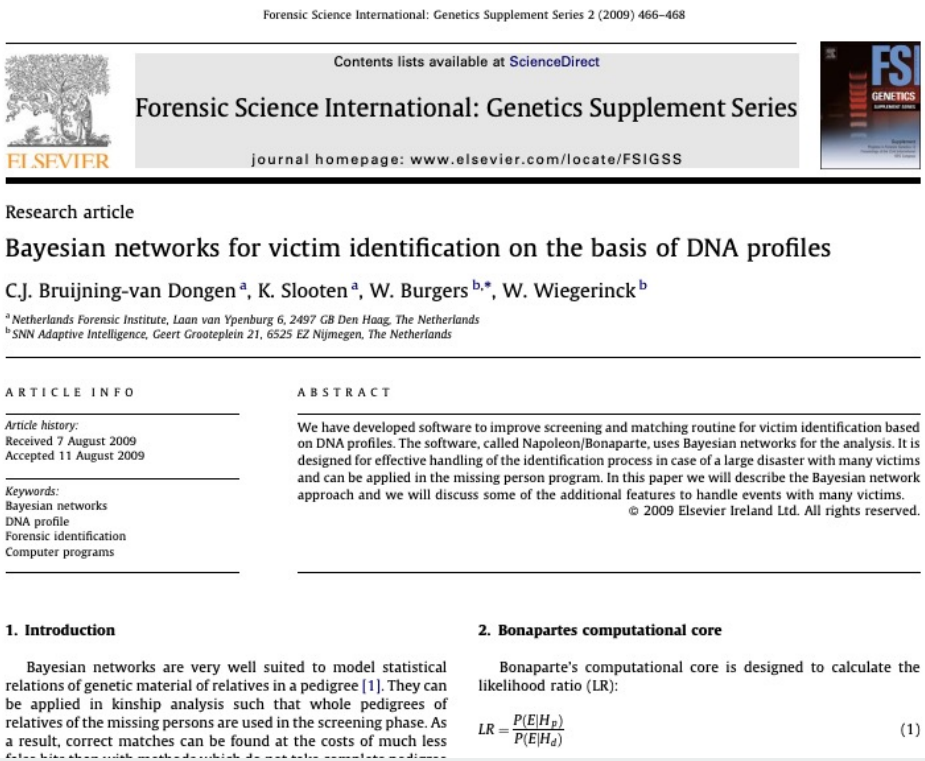
- National CBRN/e project
- NFI needed to be prepared for disaster > 500 victims

  - NFI required statistical software for DNA analysis

BONAPARTE

developed by SMART Research

In **close collaboration** with
NFI domain experts
and **end-users**



Forensic Science International: Genetics Supplement Series 2 (2009) 466–468

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

Research article

Bayesian networks for victim identification on the basis of DNA profiles

C.J. Bruijning-van Dongen [a], K. Slooten [a], W. Burgers [b,*], W. Wiegerinck [b]

[a] Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB Den Haag, The Netherlands
[b] SNN Adaptive Intelligence, Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 7 August 2009
Accepted 11 August 2009

Keywords:
Bayesian networks
DNA profile
Forensic identification
Computer programs

ABSTRACT

We have developed software to improve screening and matching routine for victim identification based on DNA profiles. The software, called Napoleon/Bonaparte, uses Bayesian networks for the 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. In this paper we will describe the Bayesian network approach and we will discuss some of the additional features to handle events with many victims.

1. Introduction

Bayesian networks are very well suited to model statistical relations of genetic material of relatives in a pedigree [1]. They can be applied in kinship analysis such that whole pedigrees of relatives of the missing persons are used in the screening phase. As a result, correct matches can be found at the costs of much less

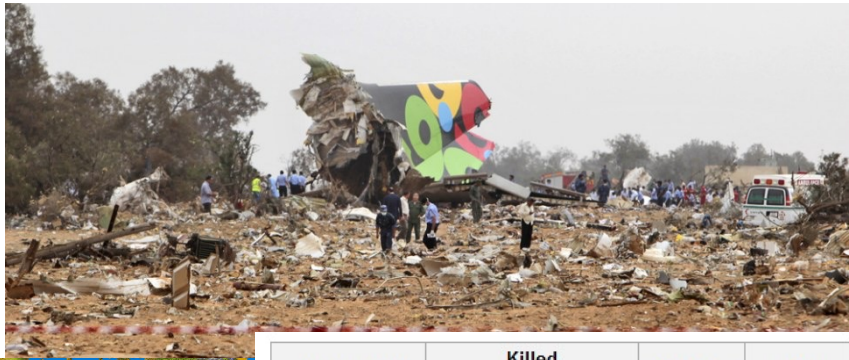
2. Bonapartes computational core

Bonaparte's computational core is designed to calculate the likelihood ratio (LR):

$$LR = \frac{P(E|H_p)}{P(E|H_d)} \quad (1)$$

March 2010: Bonaparte operational

May 12, 2010: Tripoli, Libya, air disaster
103 victims, 1 survivor



Nationality	Killed		Survivors	Total
	Passengers	Crew		
Dutch	67	–	1	68
Libyan	2	11	–	13
South African	13*	–	–	13*
Belgian	4	–	–	4
Austrian	2	–	–	2
British	1	–	–	1
French	1	–	–	1
German	1	–	–	1
Zimbabwean	1	–	–	1
Total	92*	11	1	104*

Analysis using Bonaparte
- **84 identifications**

The whole project took NFI **26 days**
Match computation takes few minutes

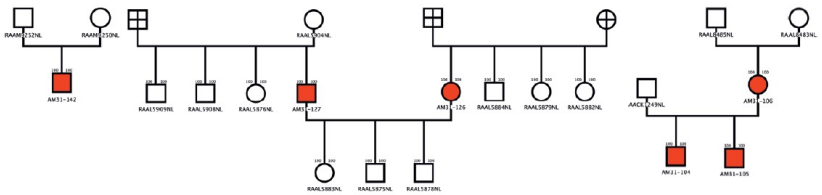
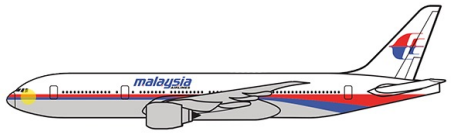


Figure 1. Examples of pedigrees with one, two and three MPs.

Malaysia Airlines Flight 17
MH17, July 17 2014



- **298** victims

- **6,000+** body parts
- Degraded DNA
- **500+** reference profiles
- 300+ family trees

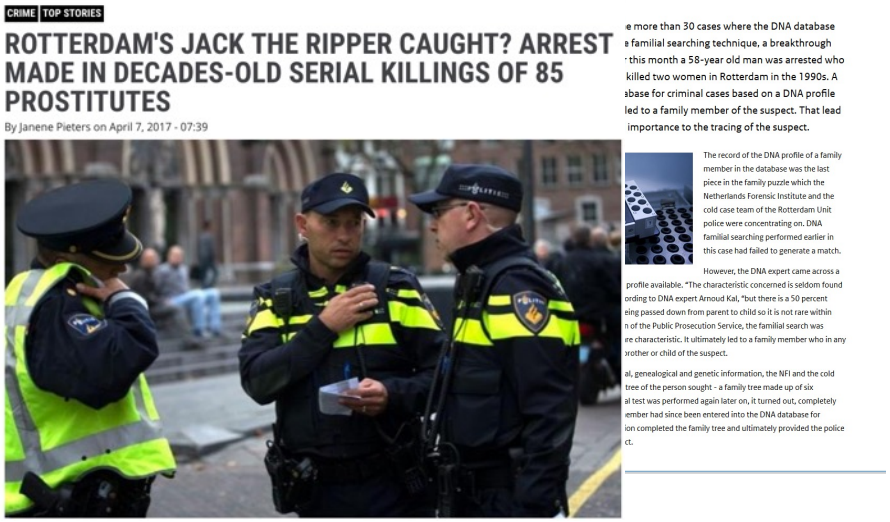
- **296 victims identified**
- **Analysis with Bonaparte**

Nation	Number
Netherlands[f]	193
Malaysia[e]	43
Australia	27
Indonesia	12
United Kingdom[g]	10
Belgium	4
Germany[d]	4
Philippines	3
Canada[c][34]	1
New Zealand	1
Total	298

Flight deck window cutout. Paul Sonne/The Wall Street Journal

Familial search

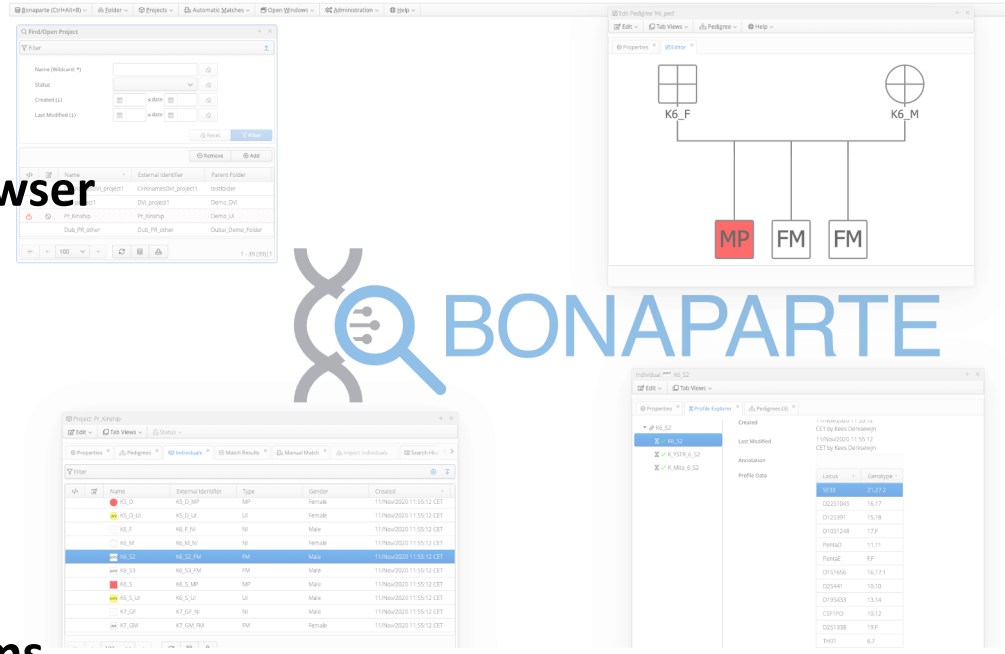
high profile criminal (cold) cases



BONAPARTE

is software for **large scale** DNA matching

- **Client-server** structure
  - Platform independent
- User friendly interface via **web browser**
  - (No need to install client software)
  - Concurrent users
  - Window like structure
- Integrated **data base**
  - Versioning (audit: who what when)
- Programmable **business rules**
  - Levels of security and user rights
- Can be **integrated** into **other systems**

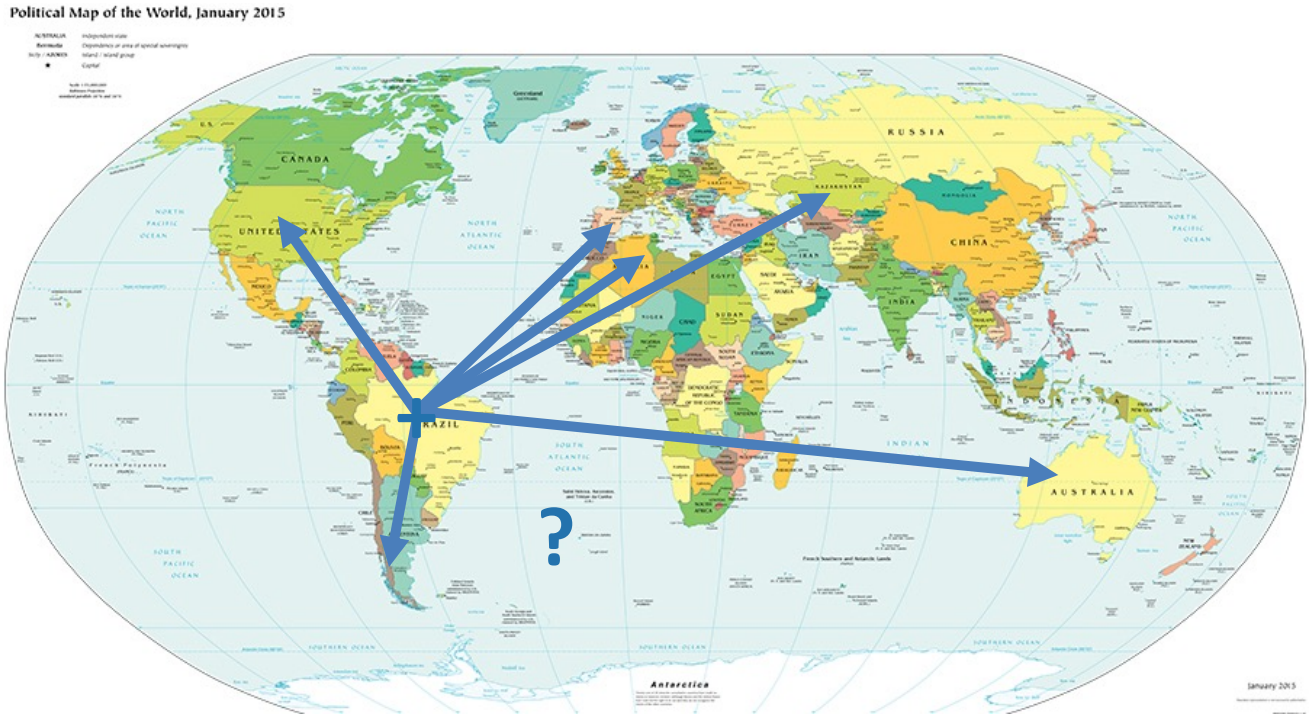


Australian government

- **ACIC (Australian Criminal Intelligence Commission)**
  - National information-sharing
  - National Criminal Investigation DNA Database
- DNA investigations across state/territory borders
  - 9 police agencies + other law enforcement
  - About 100 concurrent users
  - Working at different locations
  - business rules, security
- Criminal DNA data base
- Missing persons and unidentified bodies

INTERPOL I-Familia

- I-Familia: global database for identifying missing persons based on international DNA kinship matching
- Victim and families in different member countries
- Runs with Bonaparte: matching to families possible



Vietnamese government

- Identification of war victims "project 150"
- Largest victim identification project ever
- Degraded DNA
- Remote family members

DE VOLKSKRANT 25

Wetenschap

Reportage Dna-identificatie

Software helpt Vietnam bij megapuzzel

Een enorm project gaat ruim een half miljoen doden uit de Vietnamoorlog aan een naam helpen – en familieleden aan gemoedsrust. De software ervoor is ontwikkeld in Nijmegen.

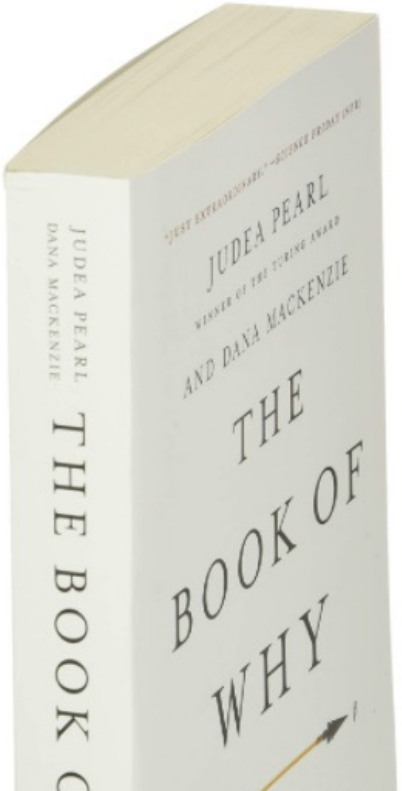
Mickey Steijaert

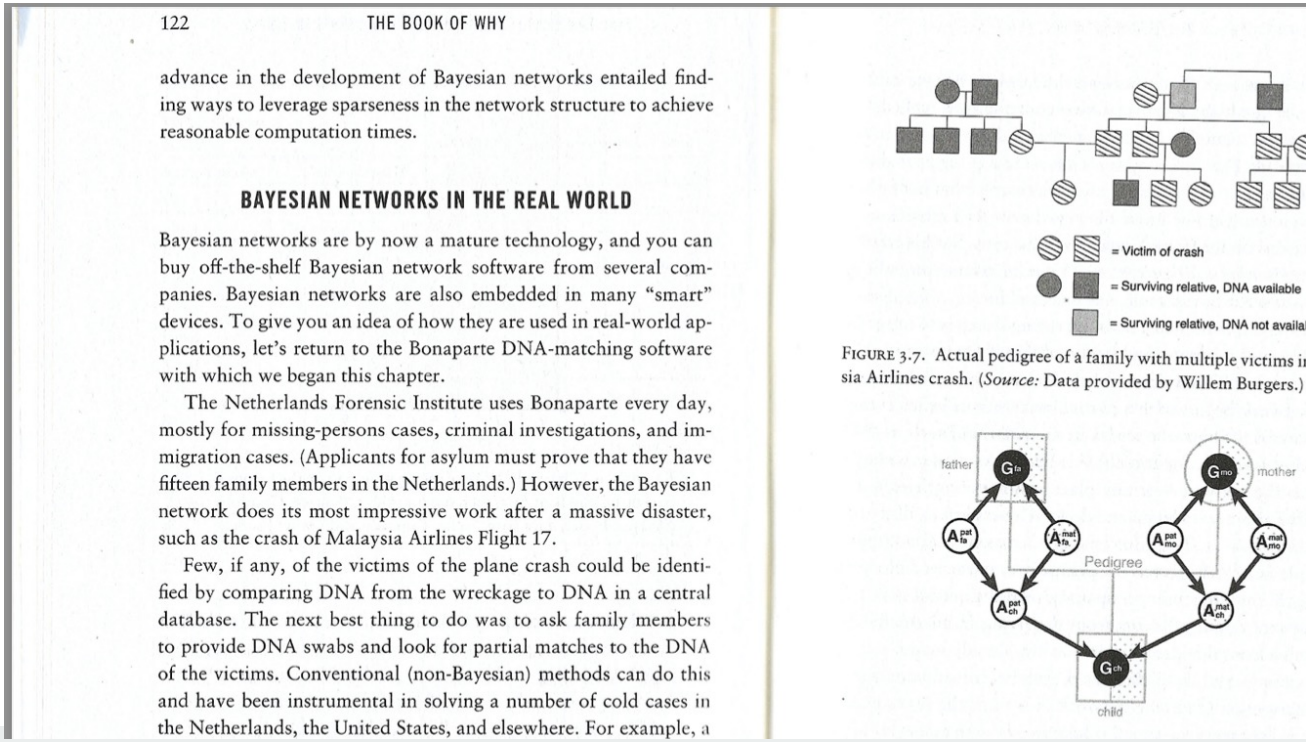
Dood familielid ten grave dragen is in Vietnam zeer belangrijk

Huyen Tran

Conclusions

- Bonaparte: application of AI/Bayesian networks in the real world





122 THE BOOK OF WHY

advance in the development of Bayesian networks entailed finding ways to leverage sparseness in the network structure to achieve reasonable computation times.

BAYESIAN NETWORKS IN THE REAL WORLD

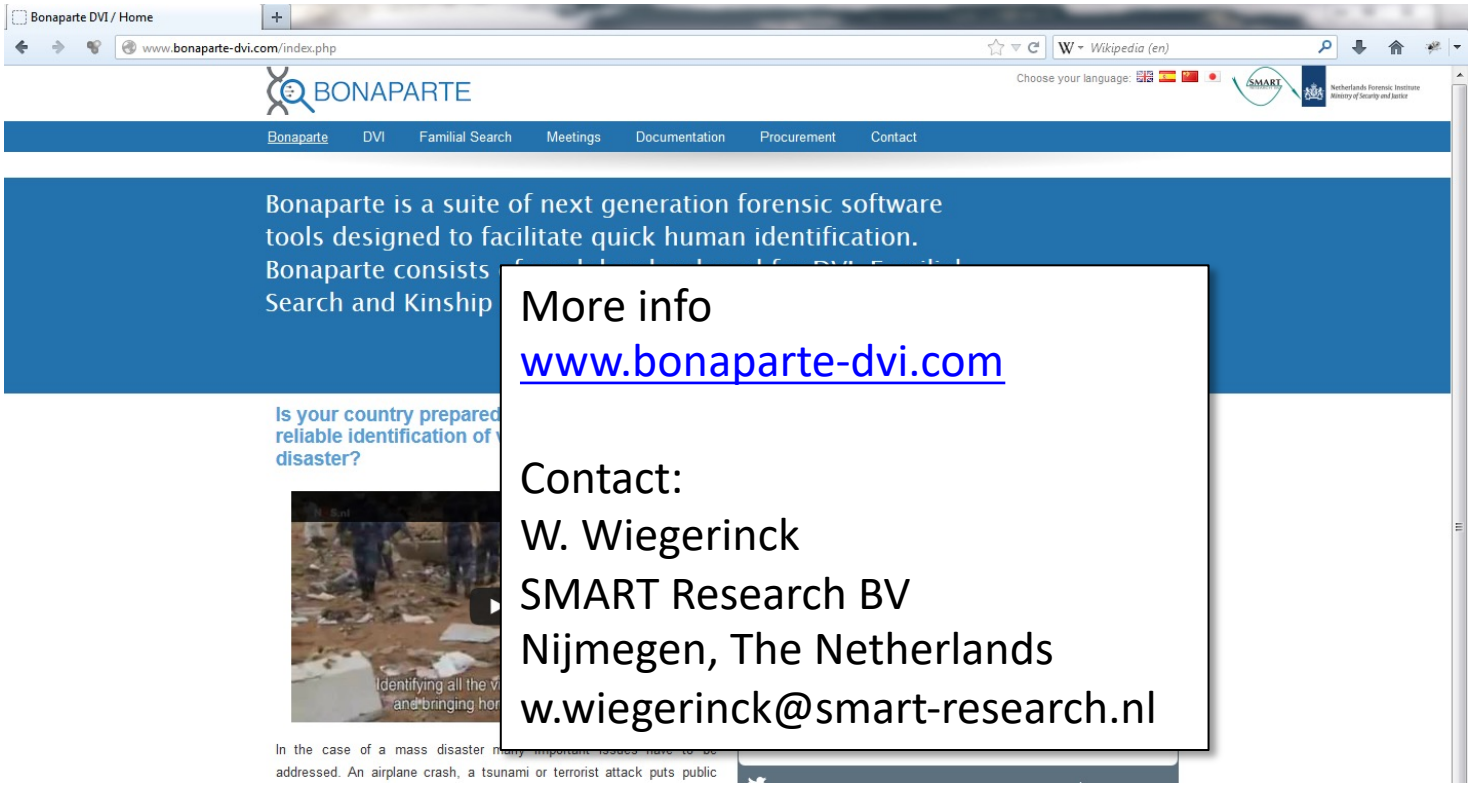
Bayesian networks are by now a mature technology, and you can buy off-the-shelf Bayesian network software from several companies. Bayesian networks are also embedded in many "smart" devices. To give you an idea of how they are used in real-world applications, let's return to the Bonaparte DNA-matching software with which we began this chapter.

The Netherlands Forensic Institute uses Bonaparte every day, mostly for missing-persons cases, criminal investigations, and immigration cases. (Applicants for asylum must prove that they have fifteen family members in the Netherlands.) However, the Bayesian network does its most impressive work after a massive disaster, such as the crash of Malaysia Airlines Flight 17.

Few, if any, of the victims of the plane crash could be identified by comparing DNA from the wreckage to DNA in a central database. The next best thing to do was to ask family members to provide DNA swabs and look for partial matches to the DNA of the victims. Conventional (non-Bayesian) methods can do this and have been instrumental in solving a number of cold cases in the Netherlands, the United States, and elsewhere. For example, a

FIGURE 3.7. Actual pedigree of a family with multiple victims i
sia Airlines crash. (*Source:* Data provided by Willem Burgers.)

More information?



Bonaparte is proprietary software of SNN, Nijmegen, The Netherlands
SMART Research BV a subsidiary of SNN
Bonaparte is developed by SMART Research BV and SNN