WELCOME

to Pula

To the 3rd School / Workshop

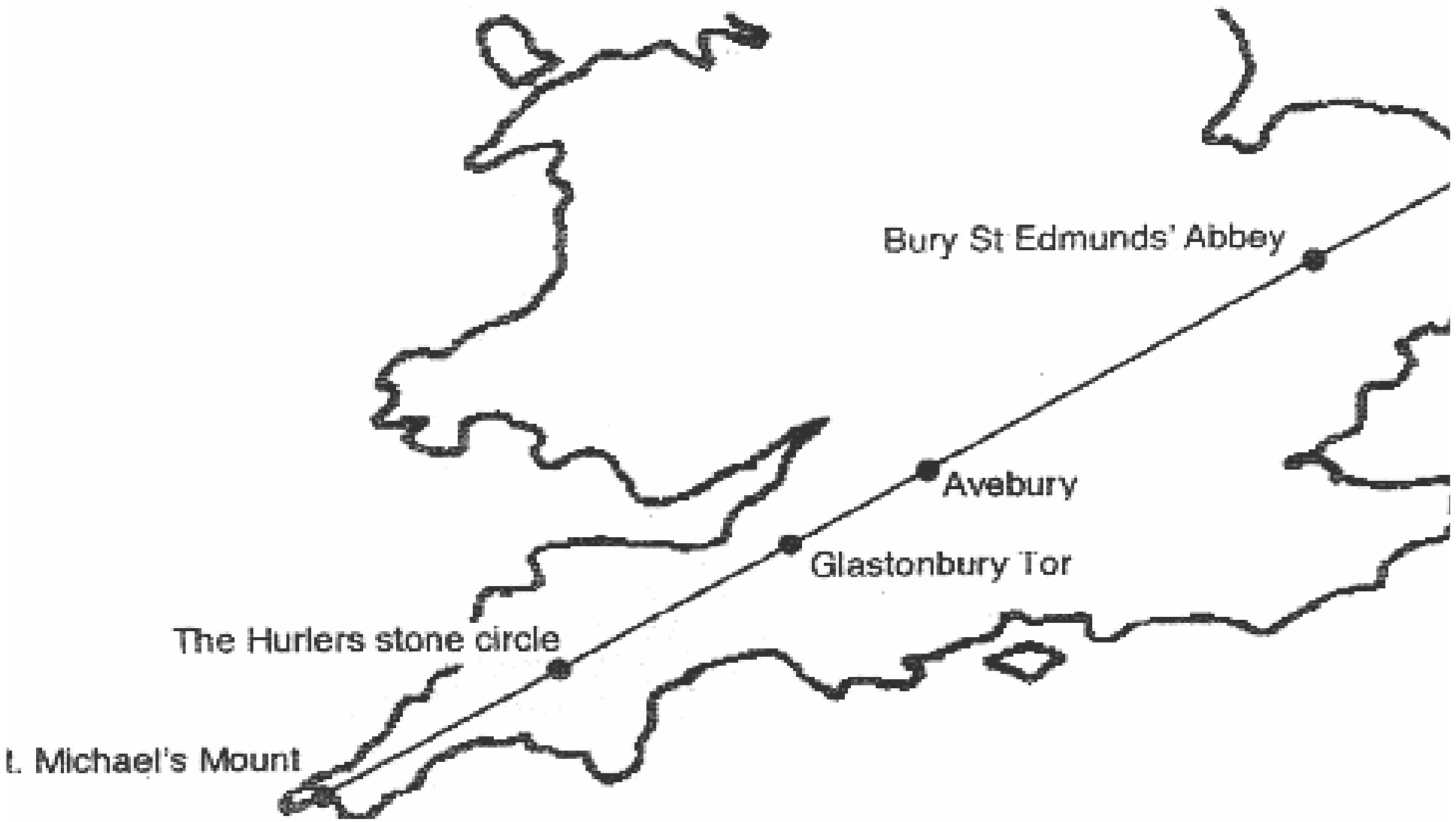
“The Analysis of Patterns”

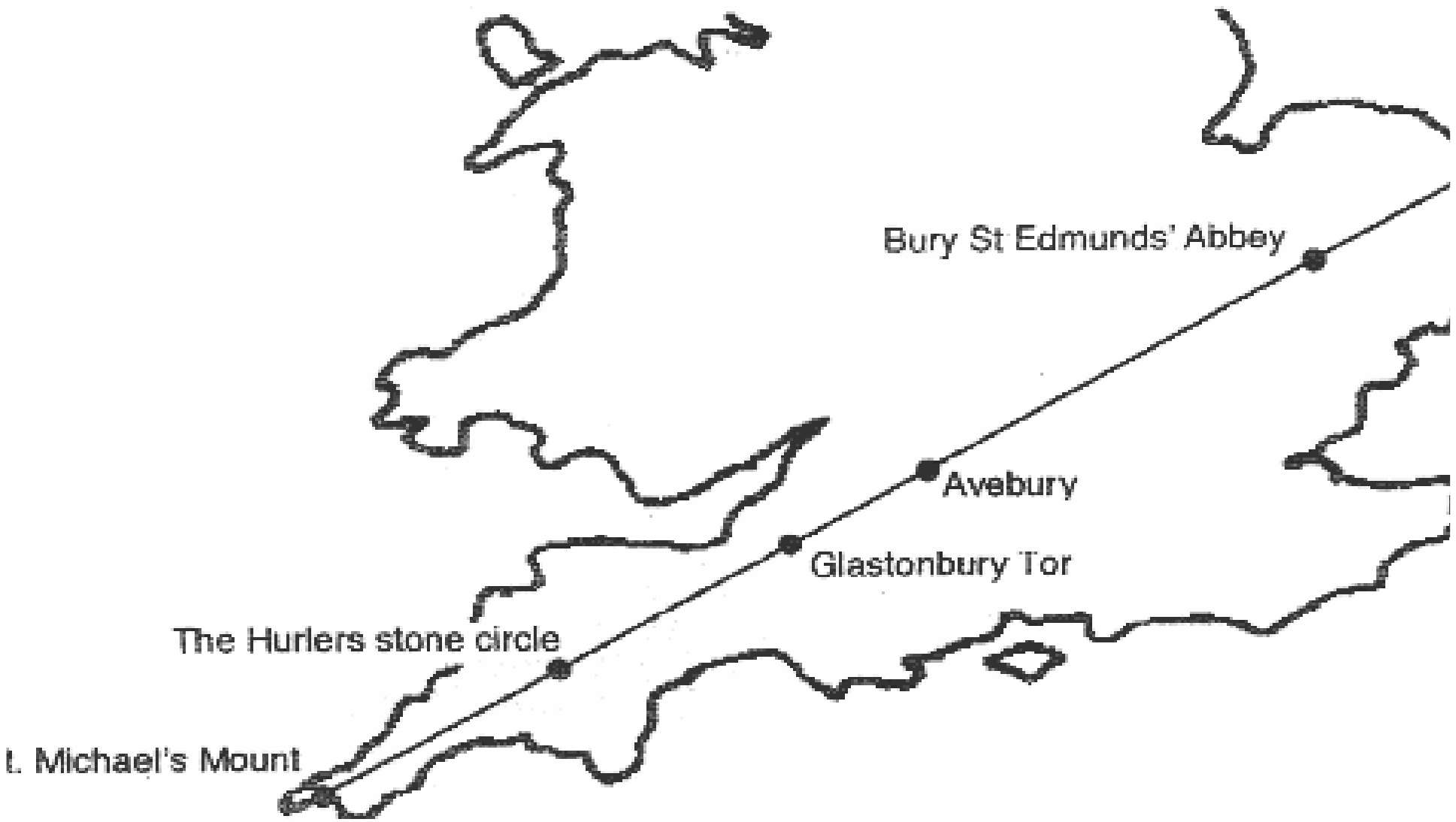
Who Needs Patterns?

Nello Cristianini
University of Bristol


©2007


Wales
Mersea
Canvey Island
(Londra)
Isles of Sh
Stonehenge, Amesbury, Wiltshire, Regno Unito
Hayling Island
Brownsea Island
Isle of Wight
© 2007 Europa Technologies
Image NASA
© 2007 Tele Atlas
©2007


Bury St Edmunds' Abbey
Avebury
Glastonbury Tor
The Hurlers stone circle
t. Michael's Mount



These points are aligned – they form a pattern on the map.

Pattern: Alignments of Points

- Pattern discovered in 1921 by the amateur archaeologist Alfred Watkins, whose book *The Old Straight Track* brought the alignments to the attention of the wider public.

Leylines

- **Leylines:** alignments of a number of places of geographical interest, such as ancient monuments and megaliths.

- The existence of alignments between sites is easily demonstrated. However, the causes of these alignments are disputed.

Leylines

- Are these alignments “significant”?
- And if they are, how can we “explain” them?

Leylines

- Are these alignments significant?
- And if they are, how can we explain them?

- ***Archaeological****:* the product of ancient surveying, property markings, or commonly travelled pathways. Numerous societies, ancient and modern, employ straight lines between points of use;

- ***Cultural:*** many cultures use straight lines across the landscape. In South America, such lines often are directed towards mountain peaks; the Nazca lines are a famous example of lengthy lines made by ancient cultures. Straight lines connect ancient pyramids in Mexico…

- ***New Age:*** the ley lines and their intersection points resonate a special psychic or magical energy, often including elements such as geomancy, dowsing or UFOs, stating that, for instance, UFO's travel along ley lines (in the way that one might observe that cars use roads and highways). These points on lines have electrical or magnetic forces associated with them.

- ***Skeptical****:* Skeptics of the actuality of ley lines often classify them as pseudoscience. Such skeptics tend to doubt that ley lines were planned or made by ancient cultures, and argue that apparent ley lines **can be readily explained without resorting to extraordinary** or pseudoscientific ideas.

Leylines

- How to choose the locations
  to include in a leyline?

- Choosing locations
  AFTER the leyline has been hypothesized
  can lead to problems…

Knockraheen

Stone Circle

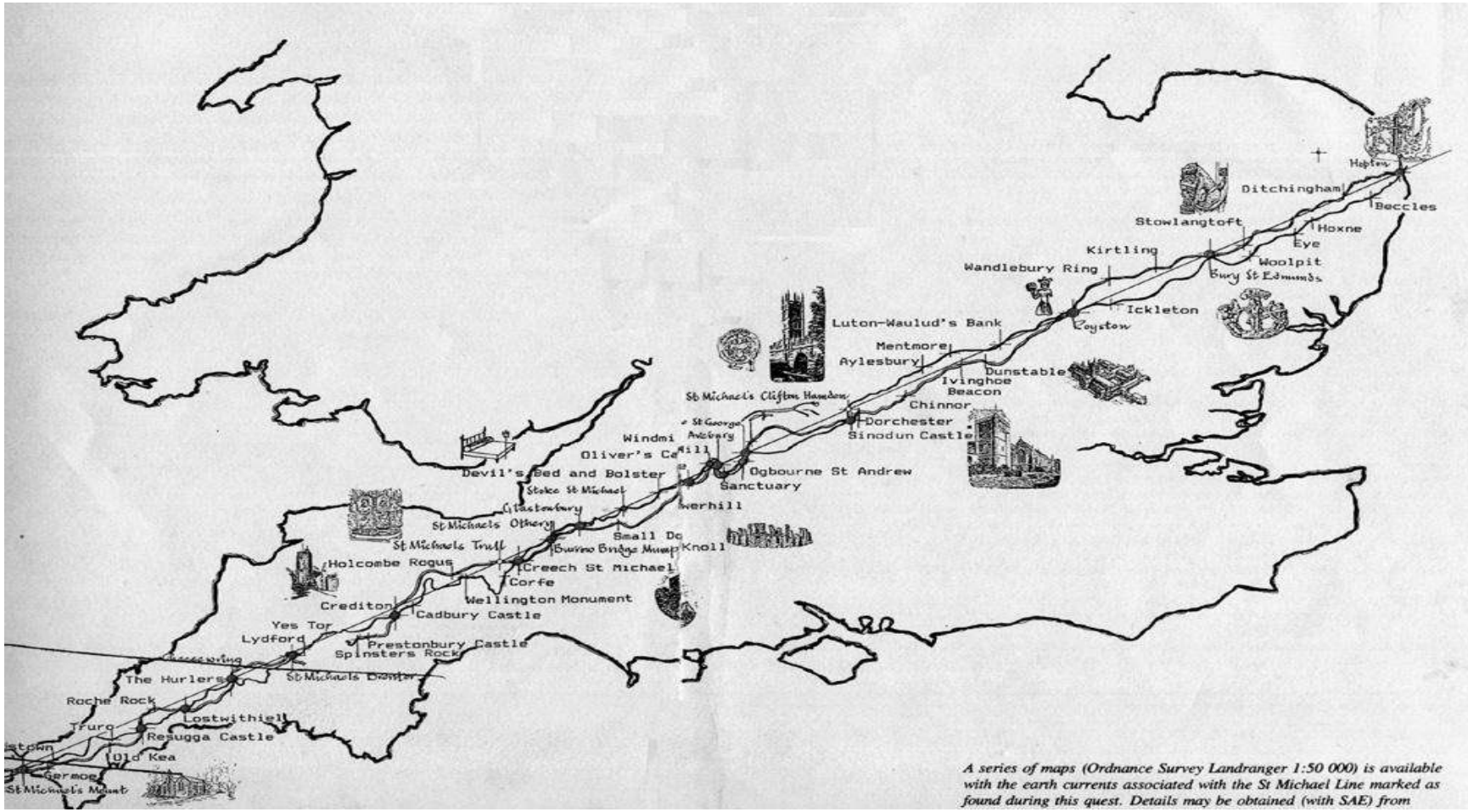
Eire

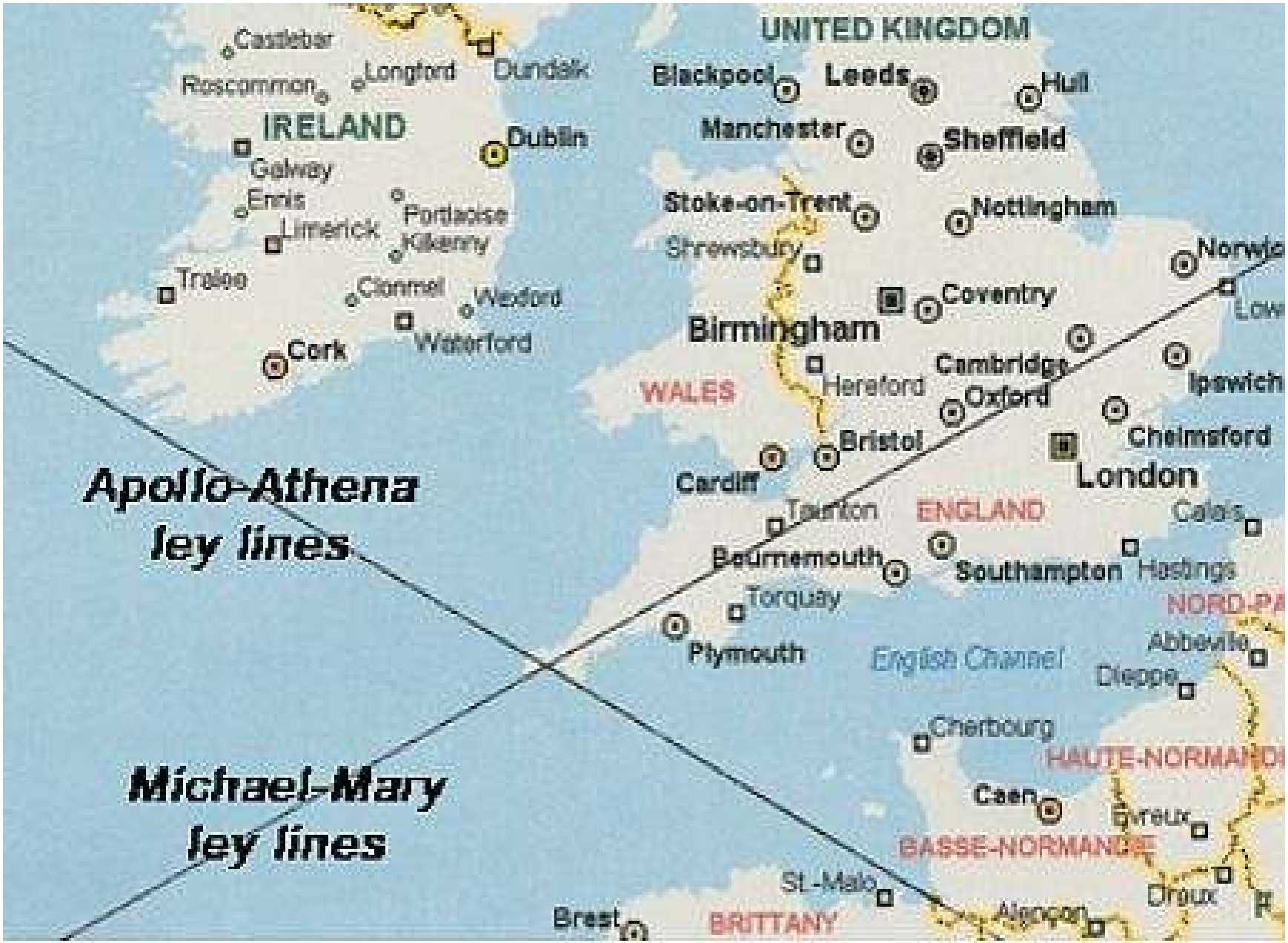
region: Cork
Grid Ref: W303802, Latitude: 51.969687, Longitude: -9.014272
This site is on the map **OSI Discovery sheet 79, Cork/Kerry** which you can buy from Amazon.co.uk

Stone circle, with quartz alignment in background

Attractive and complete Five-stone circle, with an associated alignment of strikingly white quartz. All the stones are small, less than 1m high.

A Rich Leyline

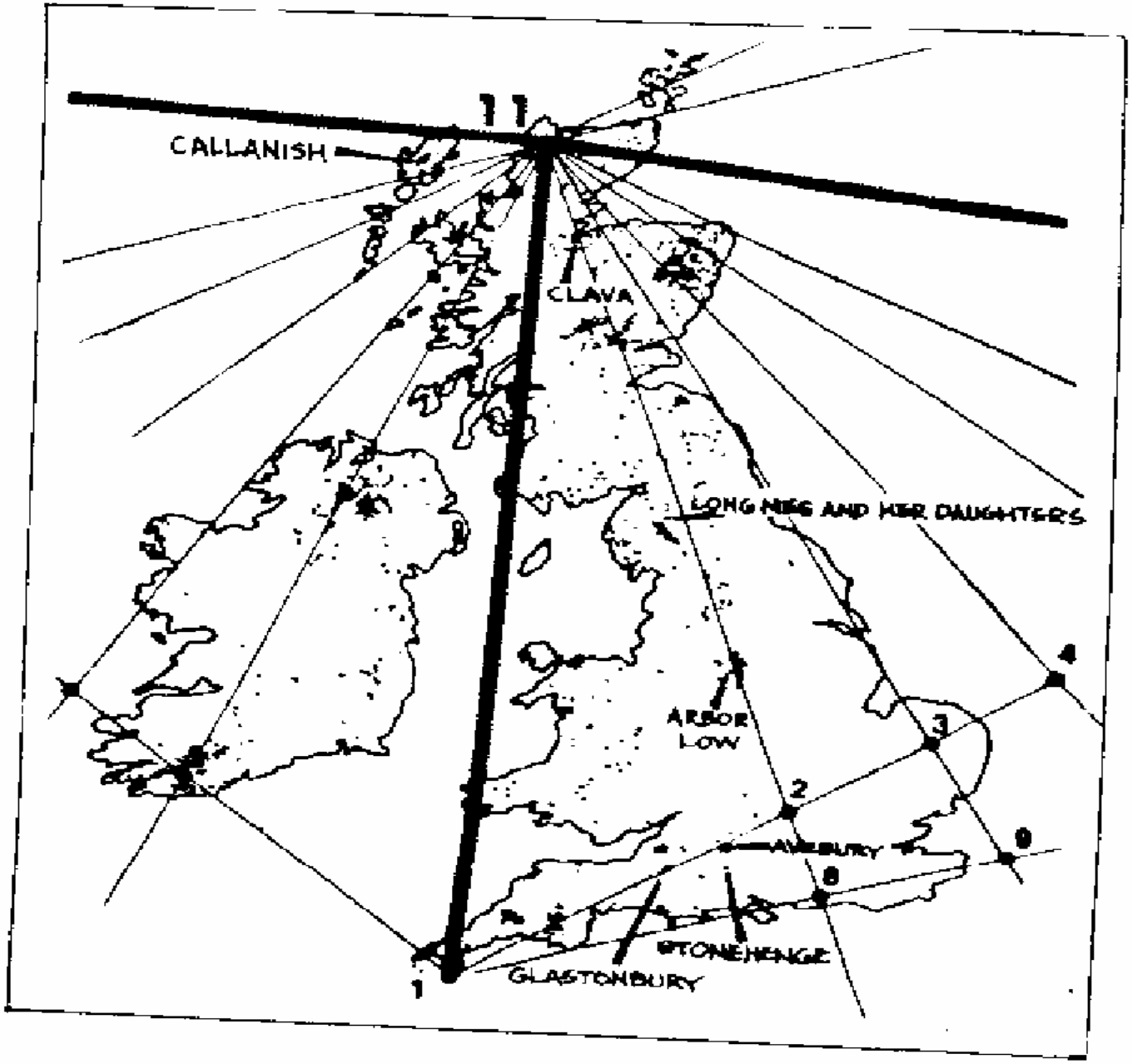


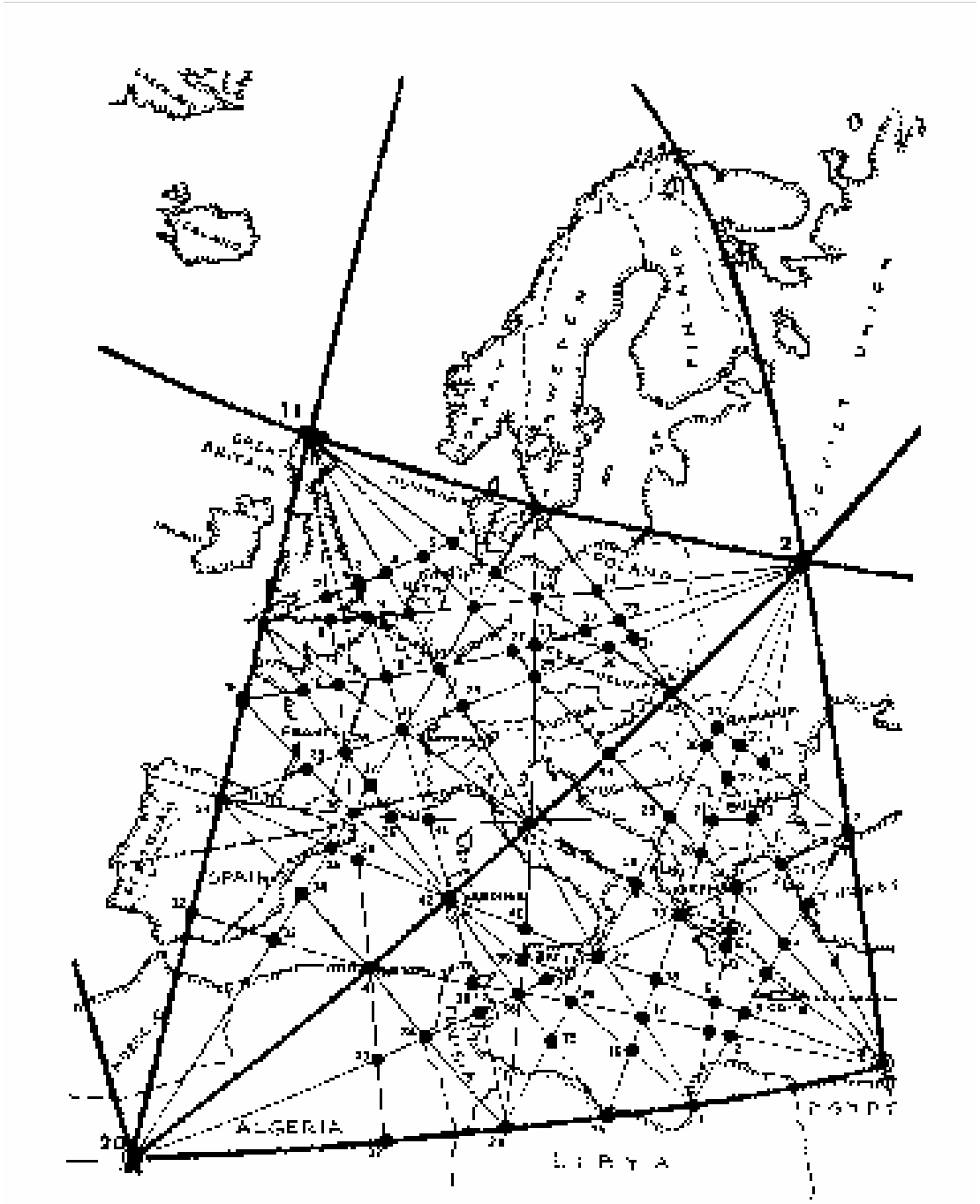

UNITED KINGDOM
Castlebar
Roscommon
Longford
Dundalk
IRELAND
Dublin
Galway
Ennis
Limerick
Portlaoise
Kilkenny
Tralee
Clonmel
Wexford
Waterford
Cork
Blackpool
Leeds
Hull
Manchester
Sheffield
Stoke-on-Trent
Nottingham
Shrewsbury
Coventry
Birmingham
Cambridge
Norwich
Ipswich
WALES
Hereford
Oxford
Bristol
Chelmsford
Cardiff
London
Taunton
ENGLAND
Calais
Bournemouth
Southampton
Hastings
Torquay
Plymouth
English Channel
Abbeville
Dieppe
Cherbourg
Caen
Evreux
St-Malo
Brest
BRITTANY
Granville
Apollo-Athena
ley lines
Michael-Mary
ley lines

Leylines

- Choosing the data after the line has been postulated can lead to serious statistical issues (when proving its significance)

- But also choosing a line connecting some locations out of a large set, can lead to serious statistical problems


11
CALLANISH
CLAVA
LONG MEG AND HER DAUGHTERS
ARBOR LOW
AVEBURY
STONEHENGE
GLASTONBURY
1
2
3
4
8
9

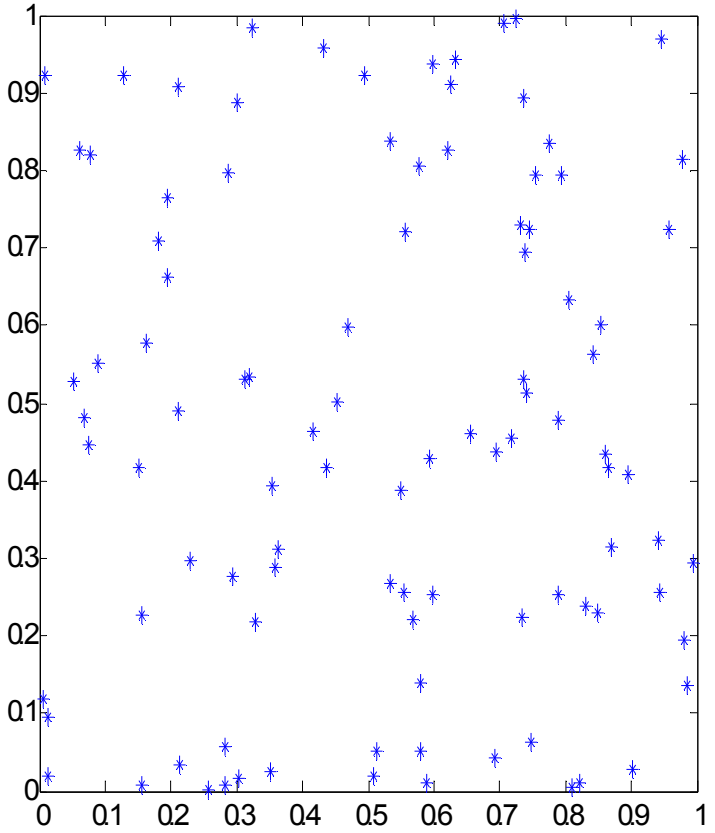

SPAIN
ALGERIA
LIBYA
EGYPT


http://www.easyjet.com/EN/routemap/
Go
JS-weekdays
Latest Headlines
Inverness
Glasgow
Edinburgh
Belfast
Newcastle
IRELAND
UK
Gdansk
Amsterdam
Warsaw
Bristol
Berlin
MADEIRA ISLANDS
Madeira (Funchal)
CANARY ISLANDS
Krakow
Prague
CZECH REPUBLIC
SLOVAKIA
Paris CDG
FRANCE
Innsbruck
AUSTRIA
HUNGARY
ROMANIA
La Rochelle
Geneva
Milan Malpensa
Venice
CROATIA
Rijeka
Bordeaux
Grenoble
Toulouse
Pisa
Marseille
Nice
BULGARIA
Barcelona
Rome
Madrid
Lisbon
Mahon
Valencia
Palma
Ibiza
Alicante
Murcia
Faro
Malaga
GREECE
TURKEY
MOROCCO
easyJet.com
the web's favourite airline
Visit the West of England

Alignments of Random Points

- **Skeptics believe that their null hypothesis of ley-line-like alignments as due to random chance is consistent with the evidence.** They believe that this consistency removes the need to explain the alignments in any other way.

- Given the high density of historic and prehistoric sites in Britain and other parts of Europe, finding straight lines that "connect" sites is trivial, and may be easily ascribed to coincidence.



What is a Pattern?

- Any **relation** in the **data** that is of interest to us.
- Typically we are interested in relations that are **not accidental** (significant).

- We often look for patterns because we expect them *to have **predictive power***.

Pattern Discovery

- Given some DATA, we can look for patterns belonging to a pre-specified class of candidates (pattern class).

- This can be seen as a search problem, often addressed with optimization techniques, or by resorting to clever data structures that help the search.

- Example: all substrings in a text, etc. etc.

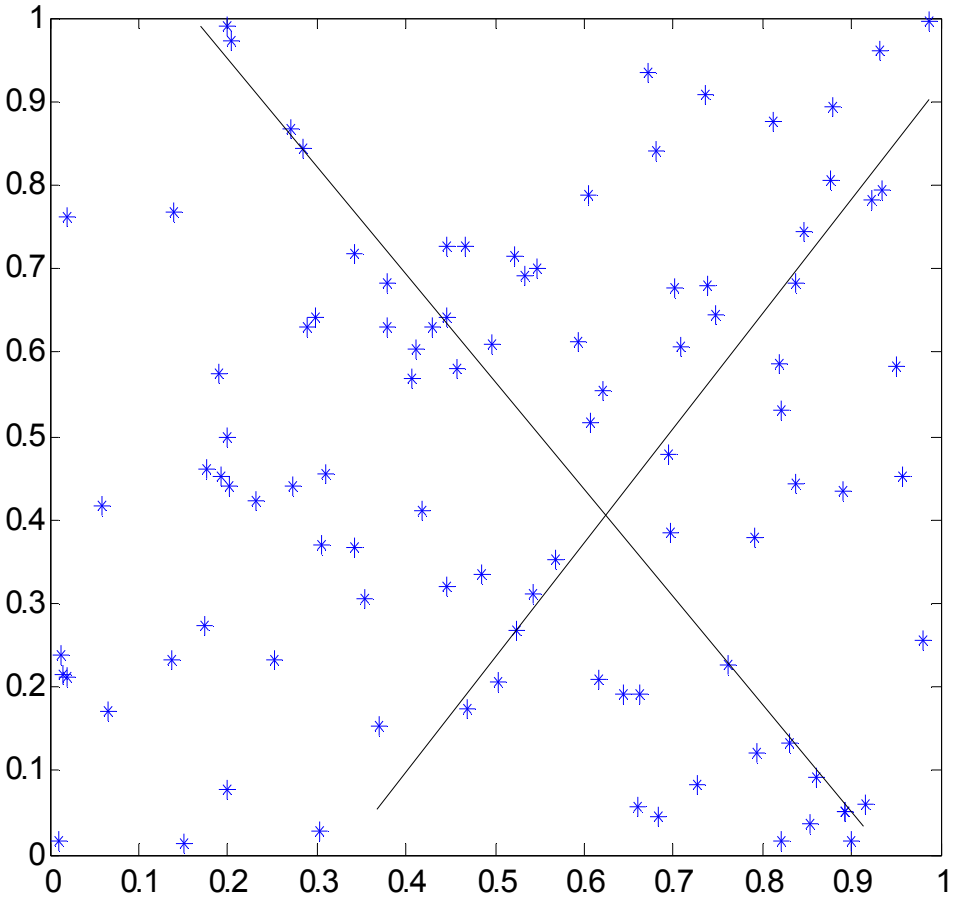
Pattern Discovery

- This part of the problem, does not distinguish between significant and non-significant relations
- It uses algorithms to discover relations.

Finding Patterns

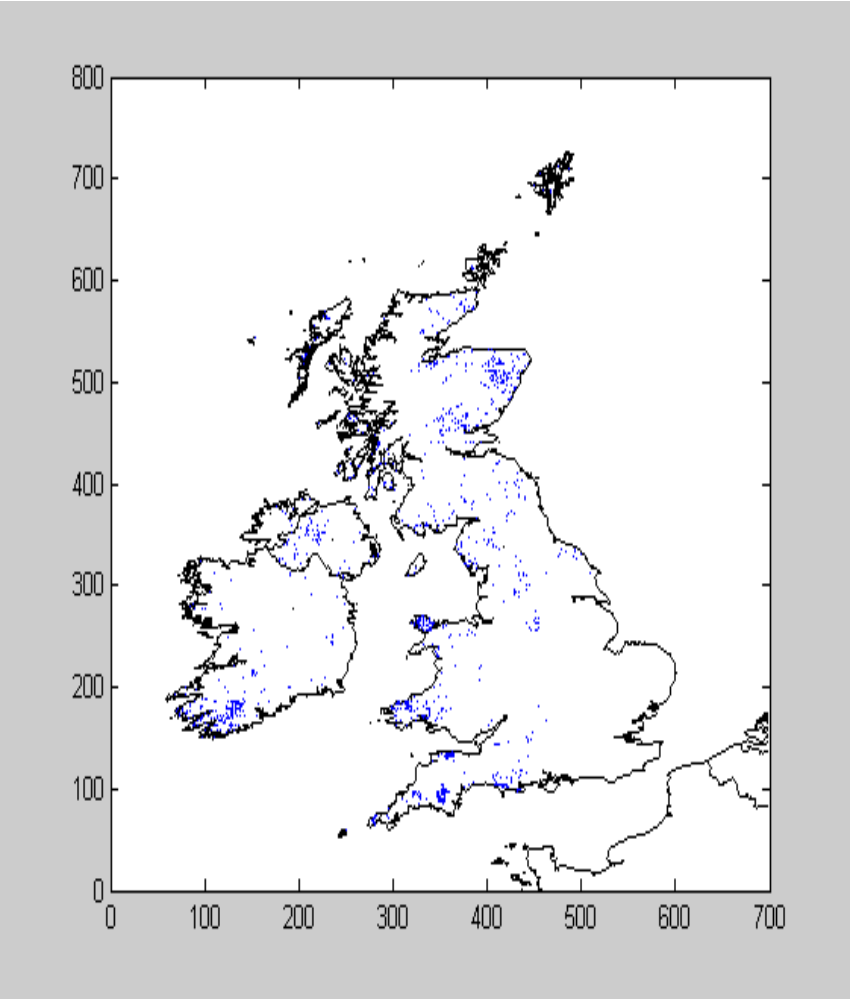
- Hough transform is an algorithm to detect linear configurations of points

- We can count how many k-tuples of points are (approximately) aligned



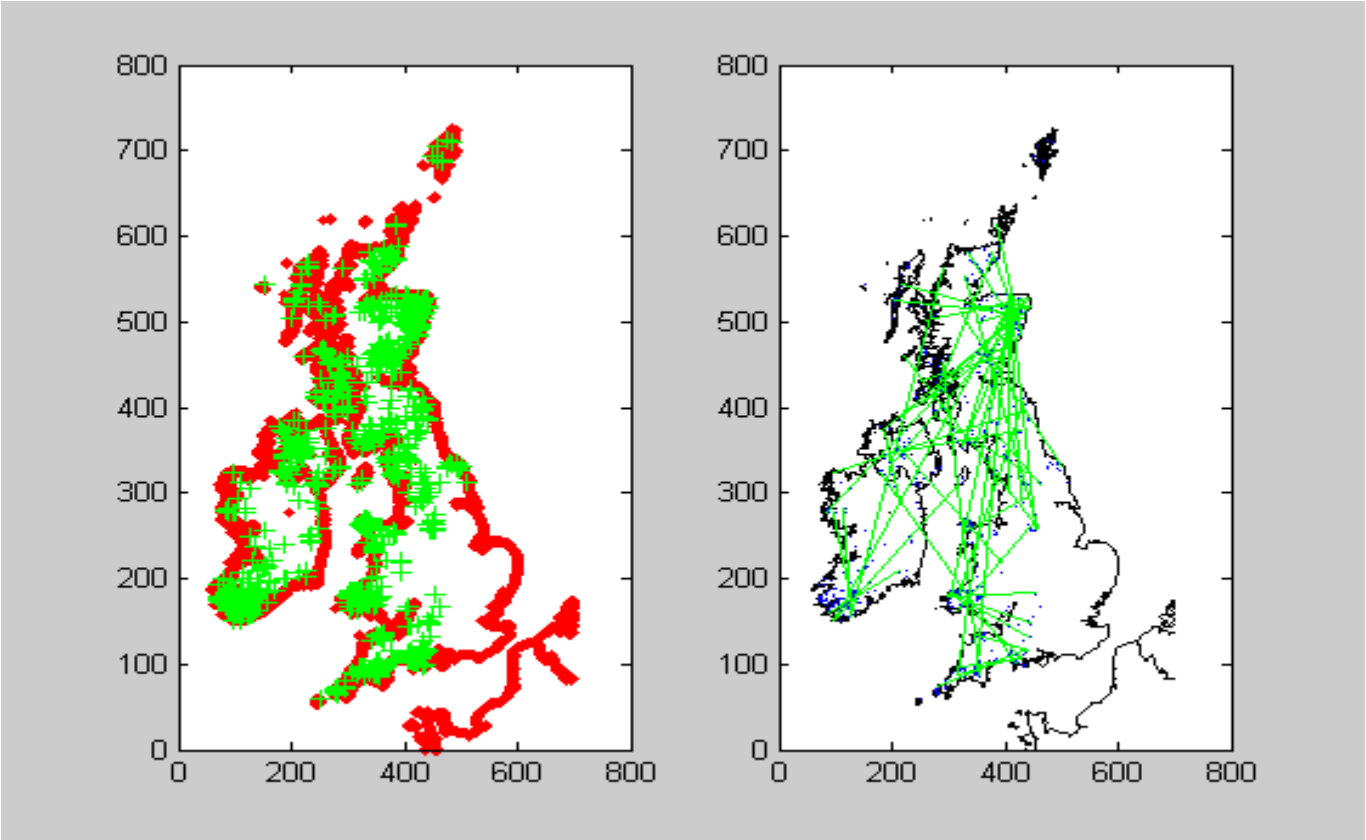
Data

- Megalithic sites in UK



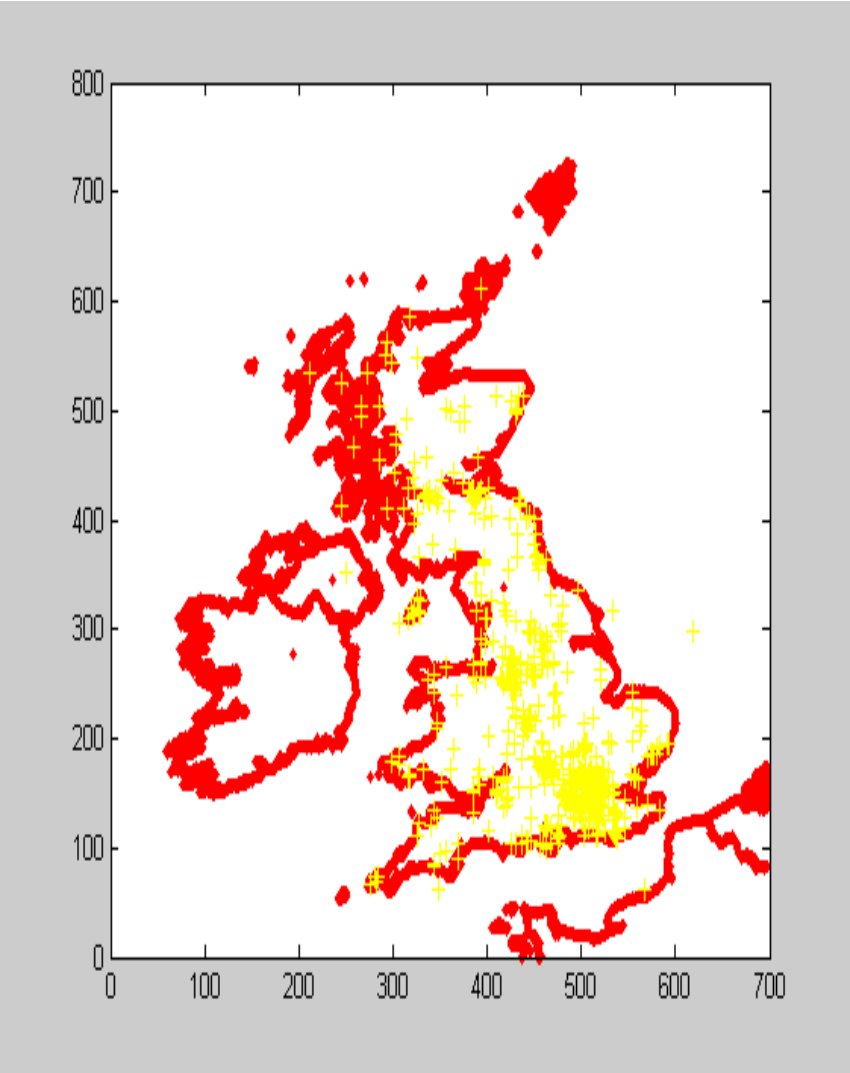
Patterns

Caveat:
- our dataset did not have locations in London
-We did not constrain Lines to Be within coastlines
- we need an objective way to validate
…



Null Model

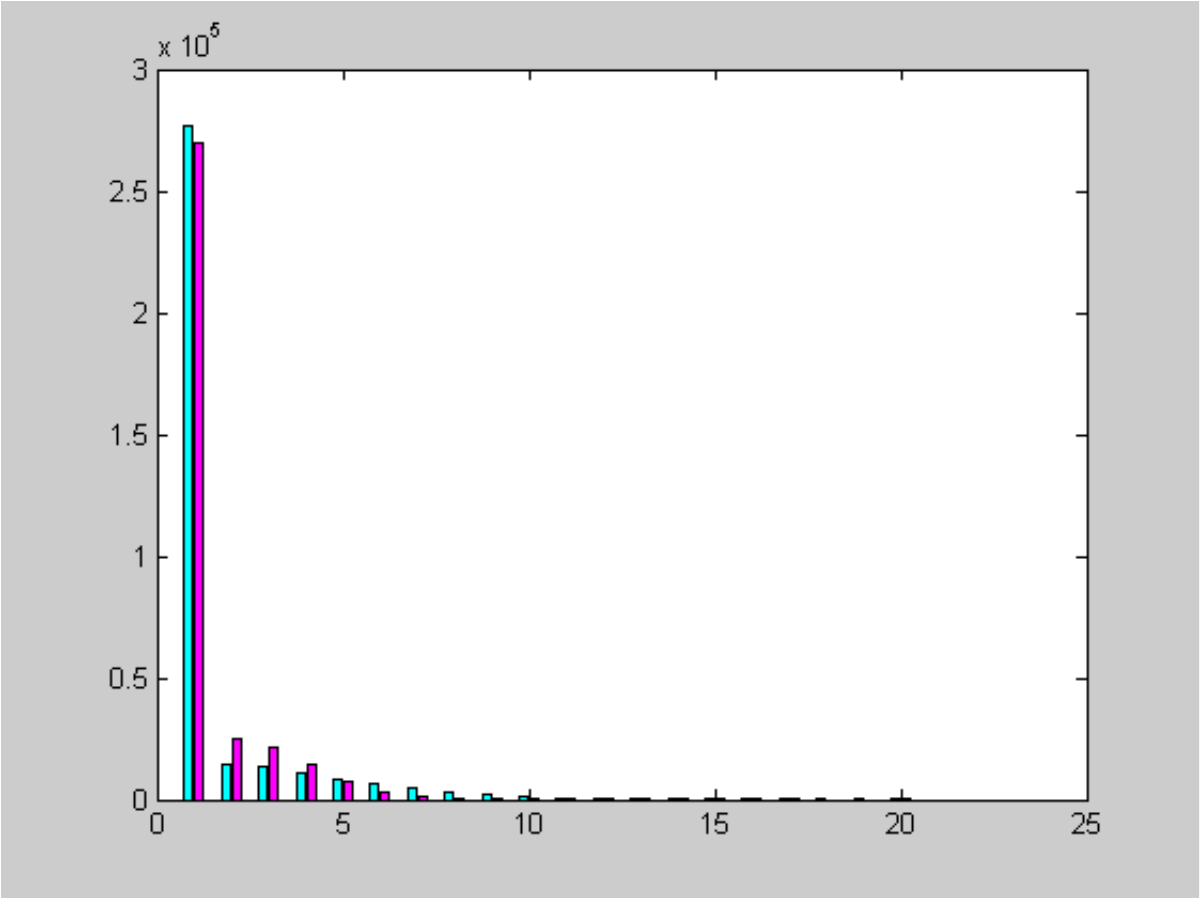
- Postcode locations in UK?
- (not a very good one)



Strength of a Pattern

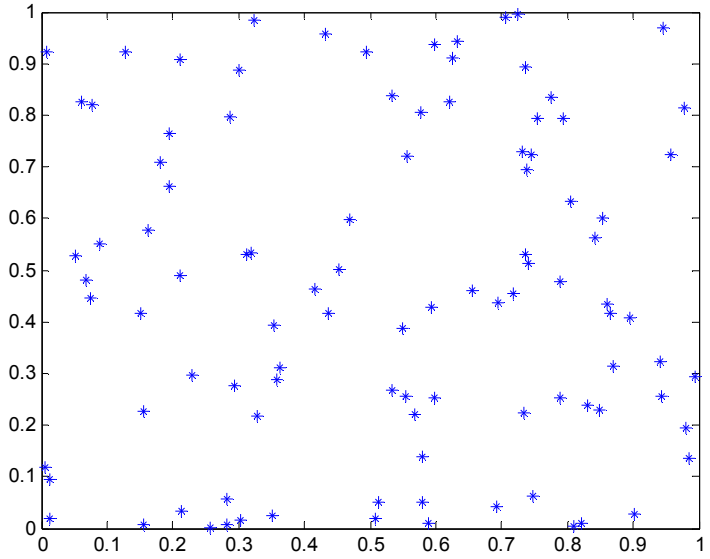
- Quantify:
  how many points are aligned, and at what level of precision?

- Histogram of: number of k-points alignments in data…



Predictive vs. Significant

- Can we *predict* the presence of a point from the rest of the data? (redundancy…)
- Can we *explain* the relation as the effect of chance? (significance)



Null Hypothesis: Random Set

Two ways to ***validate*** discovered patterns

General Pattern Finding

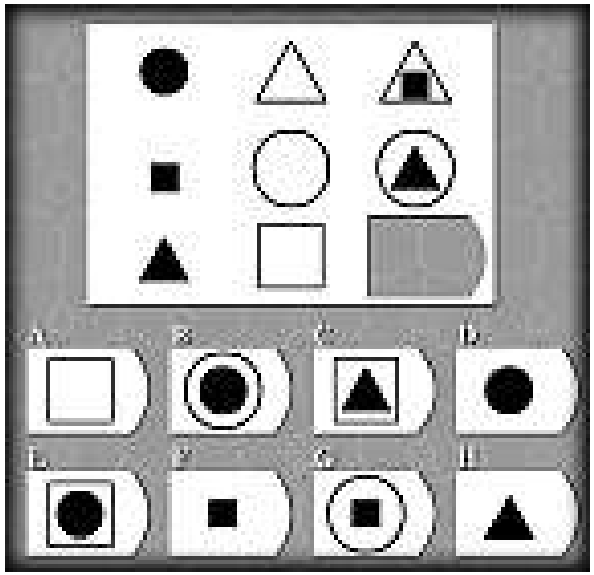
- The same problem
  arises in all areas of pattern analysis
  - Network Patterns
  - Biosequences
  - Images
  - Numeric data analysis
    (eg time series)
  - Transactions in DataBases
  - …

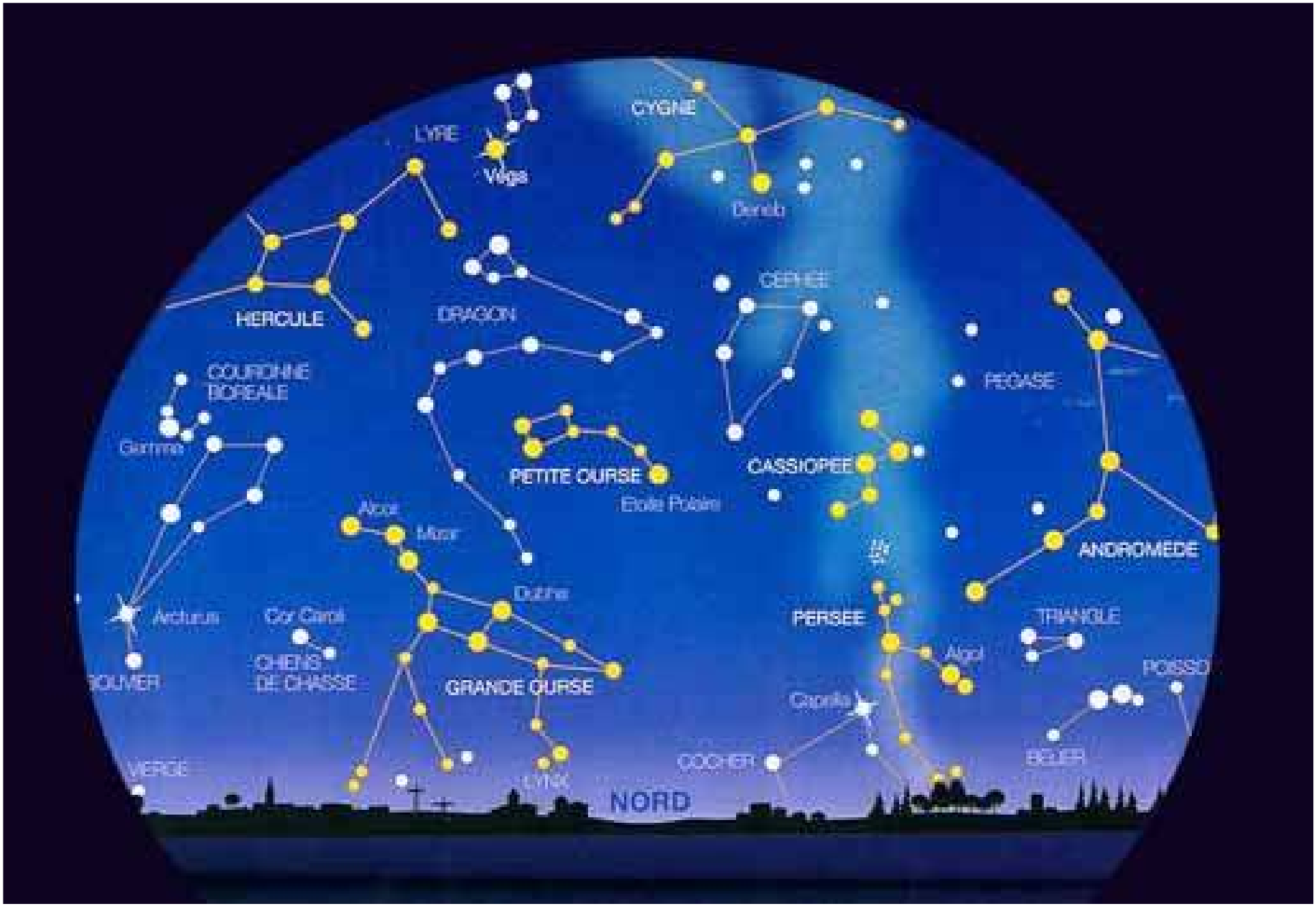
Why Patterns?

- Patterns in IQ testing
- Patterns in Science
- Patterns in ART ?
- Patterns in the sky?

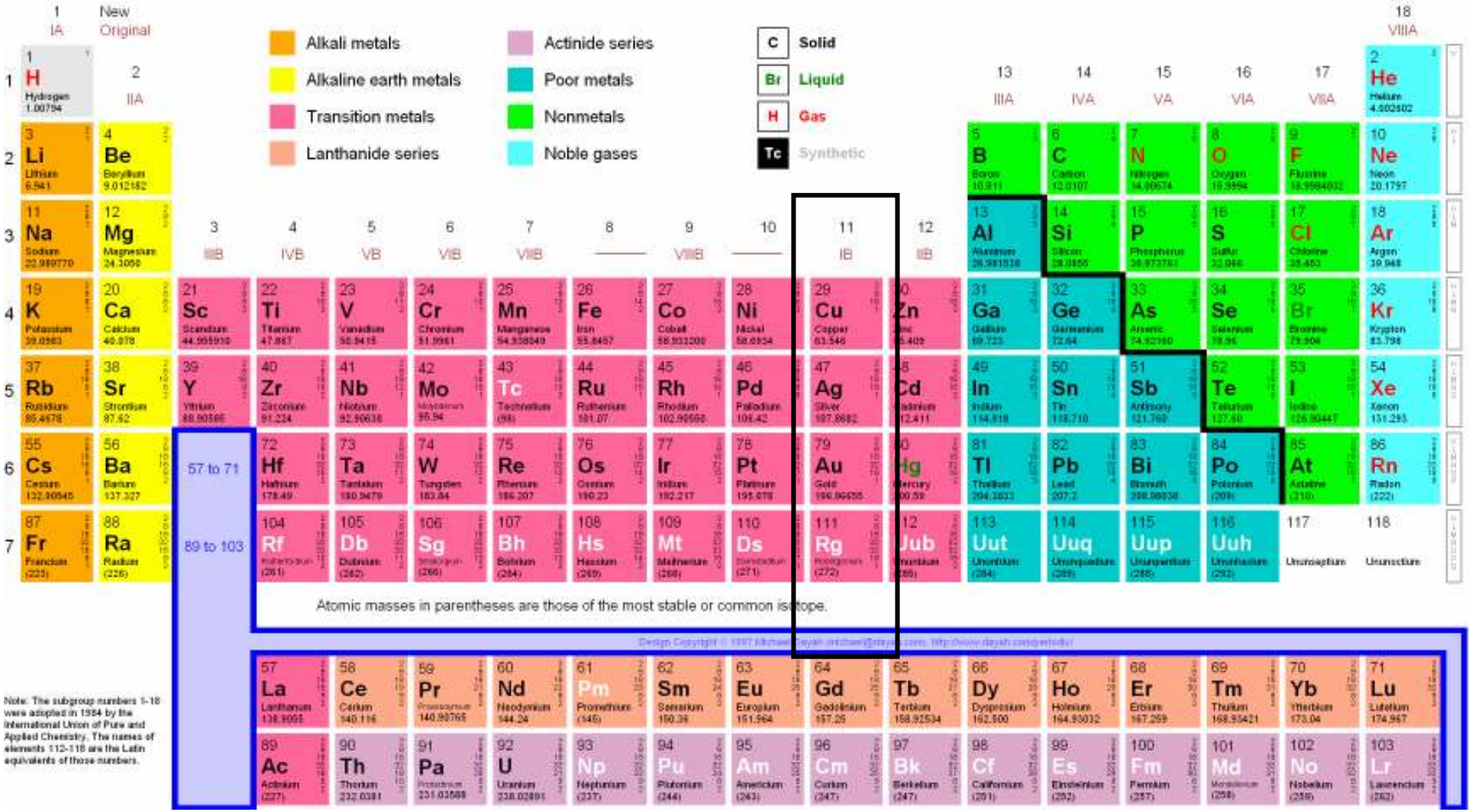
Patterns and Intelligence

- We care so much about pattern finding skills, that we even use them (partly) to quantify intelligence…




LYRE
Vega
CYGNE
Deneb
HERCULE
DRAGON
CEPHEE
COURONNE BOREALE
Gemma
PETITE OURSE
Etoile Polaire
PEGASE
CASSIOPEE
Alcor
Mizar
ANDROMEDE
Dubhe
Arcturus
Cor Caroli
PERSEE
TRIANGLE
CHIENS DE CHASSE
GRANDE OURSE
Algol
Capella
POISSON
BELIER
COCHER
VIERGE
LYNX
NORD

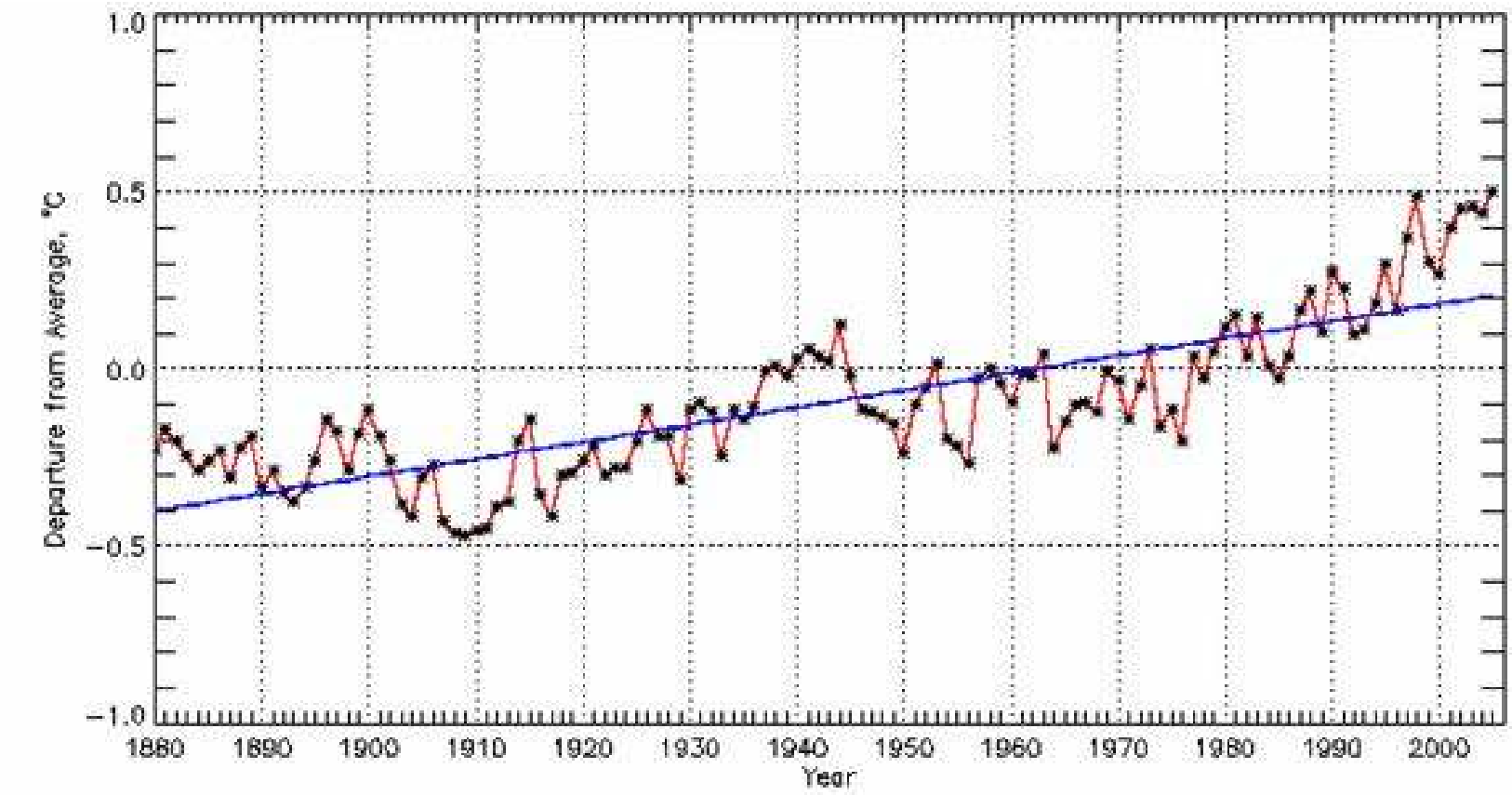
Periodic Table of the Elements

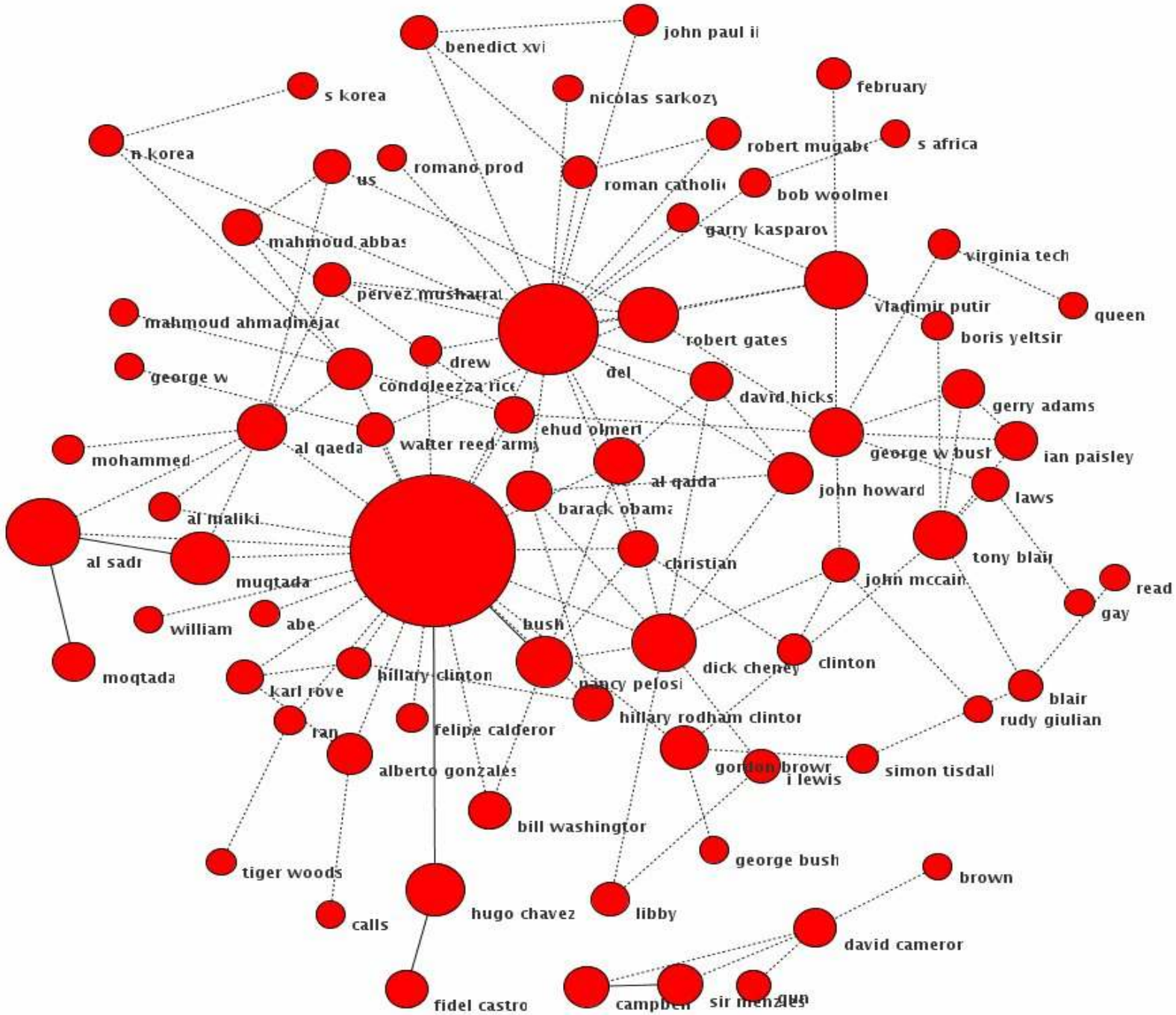


Note: The subgroup numbers 1-18 were adopted in 1984 by the International Union of Pure and Applied Chemistry. The names of elements 112-118 are the Latin equivalents of those numbers.

Time Series Analysis

- Are we warming up?




john paul ii
benedict xvi
february
s korea
nicolas sarkozy
s africa
robert mugabe
n korea
romano prod
us
roman catholi
bob woolmer
garry kasparov
mahmoud abbas
virginia tech
pervez musharraf
vladimir putin
queen
mahmoud ahmadinejac
robert gates
boris yeltsin
george w
drew
condoleezza rice
del
david hicks
gerry adams
ehud olmert
george w bush
ian paisley
mohammed
al qaeda
walter reed army
al qaida
john howard
laws
al maliki
barack obama
tony blair
al sadr
christian
john mccain
read
mugtada
gay
william
abe
bush
dick cheney
clinton
moqtada
karl rove
hillary clinton
nancy pelosi
blair
rudy giulian
hillary rodham clinton
iran
felipe calderon
gordon brown
simon tisdall
alberto gonzales
i lewis
bill washington
george bush
tiger woods
brown
hugo chavez
libby
calls
david cameron
fidel castro
campbell
sir menzies
duo

DNA Analysis

TGACTGTAATGTGCTATGTACGGTAAA
TGGCTTTATGTGCTATGTACTGTTAAG
GGTGGGTAGGTTTGTTGGTATCCTAGT
GGGTGAGGGGTGGCTTTGGAGTTGCA
GTTGATGTGTGATAGTTGGGGGTTGAT
TGCTGTACTTGCTTGTAAGCATGGGGA
GGGGGTTTTGATGTGGATTGGGTA

Why Patterns?

- Patterns can be anywhere
- Finding them is an algorithmic and statistical task
- Data can include sequences, numbers, graphs, maps, images, text, etc. etc. …

- *Methods for search and statistical analysis will be discussed in the next talks*

The Big Picture

- Next talk: model the process of finding patterns in data

- **Pattern Discovery**: given a hypothesis space Π, and some data X, find …
- **Pattern Matching**: given a pattern Π and some data X, detect occurrences of the pattern in the data …

- The expression *Pattern Recognition* has been used in both senses, and in many others, and will not be used here for clarity…

Task for you: write down DATA and Pattern Space behind every method explained in this school

The Search Problem

- Optimization / Operations Research / Theoretical Computer Science provide an excellent language and framework to study searches in pattern spaces

Task for you: identify search strategy behind every method explained in this school?

The Statistical Problem

- Statistics (both classical and modern learning-theory) provides an excellent framework to analyze significance and predictive power of relations found in data

Task for you: identify statistical validation strategies behind
every method explained in this school.
(Significance? Predictive power? Other?)

Pattern Analysis

- PA: all issues concerned with efficient discovery, matching, validation of significant patterns in any kind of data

- Hence this meeting and other meetings of the same type:
  to bring together various strands of this discipline

Back to Leylines: pattern testing…

- Are leylines in need of an explanation?
- Is the number of aligned locations comparable with what is found in random data?
- Have key locations been found as the result of leyline analysis?

- Difficult to answer, unless we have an objective way to select candidate locations

Who Needs Patterns?

- Modern science is BASED on our capability to automatically detect patterns in data
  - Genomics
  - Astronomy
  - ?

Pattern Analysis

- Can we understand in a unified way the task of finding relations in data?

- There are general principles and methods that do not distinguish between…
  - types of data,
  - types of applications,
  - types of relations,
  - algorithmic vs statistical problems

Some Modern Challenges

- Structured input / output
- Statistical inference for complex data
- Predictive power for non-statistical data
- Integration / fusion of heterogeneous types of data
- Understanding search / inference problems in a unified way

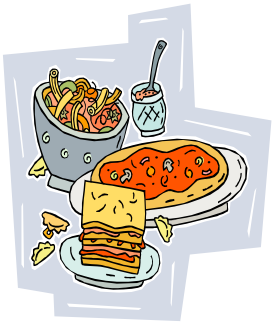
- Applications are setting the agenda for theoretical research
(bioinformatics, web, robotics, translation, etc)

This School

- Third edition
  after Erice 2005 and Bertinoro 2007

- General framework for Pattern Analysis and Impact of Pattern Analysis on science / society / technology
  - ASSOCIATION PATTERNS
  - COMBINATORIAL STRING PATTERNS
  - NETWORK PATTERNS AND BIOINFORMATICS
  - PATTERNS IN VECTORS (SETS OF POINTS) (Includes KERNEL METHODS)
  - STATISTICAL LEARNING THEORY
  - METHODS
  - APPLICATIONS

The Programme



09.00 **10.30**
10.30 11.00
11.00 **12.30**
12.30 14.30
14.30 **16.00**
16.00 16.30
16.30 **18.00**

Logistics

MAPS

- A map here

WELCOME TO
PULA (CAGLIARI)
&
HAVE FUN