Lecture 3

Applications in Search and Discovery

and the final part,

New Perspectives and New Approaches

Fionn Murtagh

In summary

We use random projections of our data – of our cloud of points. When in high dimensions, such projections can even approximate a projection into a lower dimension orthonormal space. (This is what we want in PCA or CA, etc.)

But we have pursued a new objective: to use random projection in order to approximate our data cloud such that it is rescaled well. What we mean by that: that its clustering properties are well respected – the interrelationships among points in our cloud of points.

In summary

Having rescaled our data, based on random projections, we next show how we can simplify that mapping of our data cloud.

Then we want to read off the clusters. We show that the Baire metric, that is also an ultrametric, is an excellent framework for this. (The Baire metric, as will be shown, is the "longest common prefix metric".)

Applications in Search and Discovery

- First, agglomerative hierarchical clustering; or: "hierarchical encoding" of data.
- Ultrametric topology, Baire distance.
- Clustering of large data sets.
- Hierarchical clustering via Baire distance using SDSS (Sloan Digital Sky Survey) spectroscopic data.
- Hierarchical clustering via Baire distance using chemical compounds.

Next: the Baire (ultra)metric

Baire, or longest common prefix distance – and also an ultrametric

An example of Baire distance for two numbers (x and y) using a precision of 3:

$x = 0.425$

$y = 0.427$

Baire distance between x and y:

$d_{\mathcal{B}}(x, y) = 10^{-2}$

Base ($\mathcal{B}$) here is 10 (suitable for real values)

Precision here = |K| = 3

That is:

k=1 -> $x_k = y_k$ -> 4

k=2 -> $x_k = y_k$ -> 2

k=3 -> $x_k \neq y_k$ -> 5≠7

On the Baire (ultra)metric

– Baire space consists of countable infinite sequences with a metric defined in terms of the longest common prefix *[A. Levy. Basic Set Theory, Dover, 1979 (reprinted 2002)]*

– The longer the common prefix, the closer a pair of sequences.

– The Baire distance is an ultrametric distance. It follows that a hierarchy can be used to represent the relationships associated with it. Furthermore the hierarchy can be directly read from a linear scan of the data. (Hence: hierarchical hashing scheme.)

– We applied the Baire distance to: chemical compounds, spectrometric and photometric redshifts from the Sloan Digital Sky Survey (SDSS), and various other datasets.

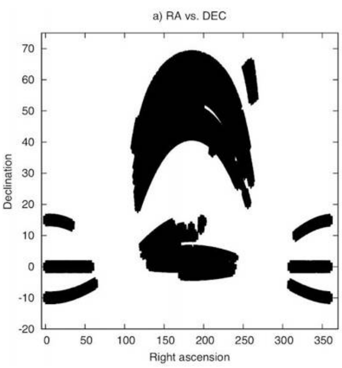
Data – sample

RA	DEC	spec. redshift	phot. redshift
145.4339	0.56416792	0.14611299	0.15175095
145.42139	0.53370196	0.145909	0.17476539
145.6607	0.63385916	0.46691701	0.41157582
145.64568	0.50961215	0.15610801	0.18679948
145.73267	0.53404553	0.16425499	0.19580211
145.72943	0.12690687	0.03660919	0.06343859
145.74324	0.46347806	0.120695	0.13045037

- A subset was taken of approximately 0.5 million data points from the SDSS release 5.
- These were objects with RA and Dec (Right Ascension and Declination, and spectrometric redshift, and photometric redshift). Problem addressed: regress one redshift (spectro.) on the other (photo.).
- Baire approach used, and compared with k-means.
- 1.2 million chemical compounds, each characterized by 1052 boolean presence/absence values.
- Random projections used on normalized compound/attribute values.
- Baire approach used; also another approach based on restricting the precision of the normalized compound/attribute values.

- Motivation - regress z_spect on z_phot
- Furthermore: determine good quality mappings of z_spect onto z_phot, and less qood quality mappings
- I.e., cluster-wise nearest neighbour regression
- Note: cluster-wise not spatially (RA, Dec) but rather within the data itself

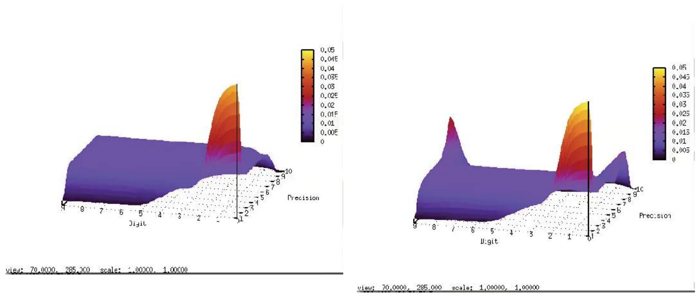
SDSS (Sloan Digital Sky Survey) Data



- We took a subset of approximately 0.5 million data points from the SDSS release 5 *[reference: D'Abrusco et al]*
- declination (Dec)
- right ascension (RA)
- spectrometric redshift
- photometric redshift.

- Dec vs RA are shown

Perspective Plots of Digit Distributions



- On the left we have z_spec where three data peaks can be observed. On the right we have z_phot where only one data peak can be seen.

Framework for Fast Clusterwise Regression

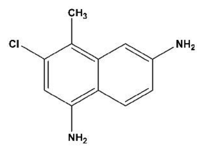
- 82.8% of z_spec and z_phot have at least 2 common prefix digits.
- I.e. numbers of observations sharing 6, 5, 4, 3, 2 decimal digits.
- We can find very efficiently where these 82.8% of the astronomical objects are.
- 21.7% of z_spec and z_phot have at least 3 common prefix digits.
- I.e. numbers of observations sharing 6, 5, 4, 3 decimal digits.

Matching of Chemical Structures

- - Clustering of compounds based on chemical descriptors or chemical representations, in the pharmaceutical industry.
- - Used for screening large corporate databases.
- - Chemical warehouses are expanding due to mergers, acquisitions, and the synthetic explosion brought about by combinatorial chemistry.

- Next - another case study, using chemoinformatics - which is high dimensional.
- Since we are using digits of precision in our data (re)coding, how do we handle high dimensions?

Binary Fingerprints

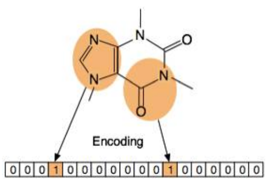


Encode

1 0 0 0 1 0 0 0 1 ...1

Fixed length bit strings with encoding schemes
Daylight, MDL
BCI *(We will be using this)*

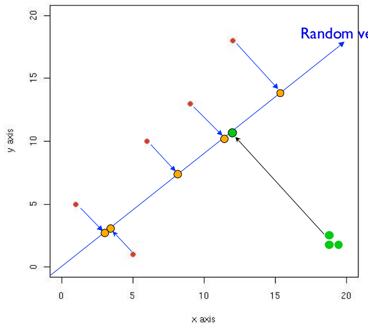
Baire Distance Applied to Chemical Compounds



Chemoinformatics clustering

- 1.2 million chemical compounds, each characterized by 1052 boolean presence/absence values.
- Firstly we note that precision of measurement leads to greater ultrametricity (i.e. the data are more hierarchical).
- From this we develop an algorithm for finding equivalence classes of specified precision chemicals. We call this: data “condensation”.
- Secondly, we use random projections of the 1052-dimensional space in order to find the Baire hierarchy. We find that clusters derived from this hierarchy are quite similar to k-means clustering outcomes.

Random projection and hashing



In fact random projection here works as a class of hashing function.

Hashing is much faster than alternative methods because it avoids the pairwise comparisons required for partitioning and classification.

If two points (p , q) are close, they will have a very small |p-q| (Euclidean metric) value; and they will hash to the same value with high probability; if they are distant, they should collide with small probability.

- Quite a different starting point:
- Using Apache Lucene and Solr for indexing, storage, and query support
- The following slide is showing where we used 152,998 cooking recipes, with 101,060 unique words in them.

- Normalize chemical compounds by dividing each row by row sum (hence “profile” in Correspondence Analysis terms).
- Two clustering approaches studied:
- Limit precision of compound / attribute values. This has the effect of more compound values becoming the same for a given attribute. Through a heuristic (e.g. interval of row sum values), read off equivalence classes of 0-distance compounds, with restricted precision. Follow up if required with further analysis of these crude clusters. We call this “data condensation”. For 20000 compounds, 1052 attributes, a few minutes needed in R.
- Second approach: use random projections of the high dimensional data, and then use the Baire distance.

Summary Remarks on Search and Discovery

- We have a new way of inducing a hierarchy on data
- First viewpoint: encode the data hierarchically and essentially read off the clusters
- Alternative viewpoint: we can cluster information based on the longest common prefix
- We obtain a hierarchy that can be visualized as a tree
- We are hashing, in a hierarchical or multiscale way, our data
- We are targeting clustering in massive data sets
- The Baire method - we find - offers a fast alternative to k-means and a fortiori to traditional agglomerative hierarchical clustering
- At issue throughout this work: embedding of our data in an ultrametric topology

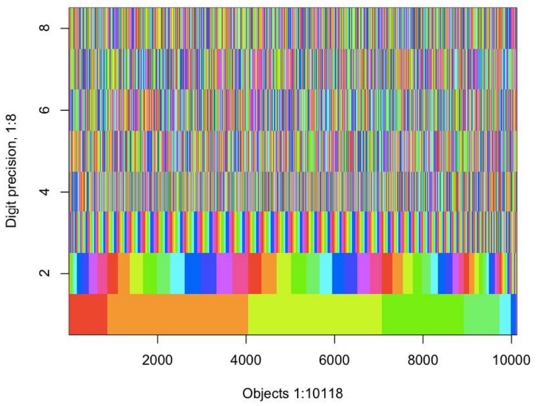
From random projection to the Baire hierarchical clustering

Selection of 10,317 funding proposals, out of set of 34,352, were indexed in Apache Solr. Their similarities were determined, using Solr’s MLT (“more like this”) score. (This uses weights for fields in the proposal documents, and is analogous to a chi squared, or tf-idf-based similarity.) *(tf-idf: term frequency – inverse document frequency)*

We used a very sparse similarity matrix of dimensions 10317 x 34252. Through random projection, we obtained a unidimensional scaling of the 10317 proposals.

In the following the mean of 99 random projections was used.
The projection values were rescaled to the interval 0,1.
Layer 1 clusters: the same first digit.
Layer 2 clusters: given the same first digit, having the same second digit.
Layer 3 clusters: given the same first two digits, having the same third digit.
And so on.
This is a regular 10-way tree.

Abscissa: 10118 documents sorted by random projection value. Ordinate: 8 digits comprising random projection value.
Layer 1: 8 clusters are very evident. Layer 2: there are 87 clusters (maximum is 100). Layer 3: here 671 clusters (maximum is 1000).



Reduced dimensionality: k << d
Below: Johnson-Lindenstrauss Lemma
Distance changes by a fraction 1+\epsilon

$$F(x) : \mathbb{R}^d \rightarrow \mathbb{R}^k$$

Lemma 1. *For any $0 < \varepsilon < 1$ and any integer n, let k be a positive integer such that*

$$k \geq 4(\varepsilon^2/2 - \varepsilon^3/3)^{-1} \ln n. \tag{2}$$

Then for any set V of any points in $\mathbb{R}^d$, there is a map $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that for all $u, v \in V$,

$$(1-\varepsilon) \parallel u - v \parallel^2 \;\leq\; \parallel f(u) - f(v) \parallel^2 \;\leq\; (1+\varepsilon) \parallel u - v \parallel^2.$$

Furthermore, this map can be found in randomized polynomial time.

Low dimensional goodness of fit to our data, versus linear rescaling

Conventional use of random projections:

Project data into lower dimension subspace, of dimension > 1.

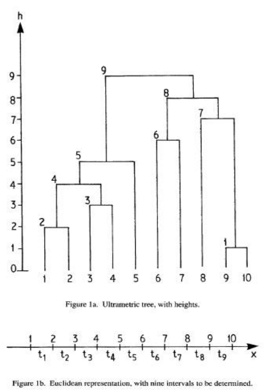
Aim is to have proximity relations respected in the low dimensional fit to the high dimensional cloud of points.

In the work presented here, we seek a consensus one-dimensional mapping of the data, that represents relative proximity.

Two following slides: Our aim is relative clustering properties. Cf. the now conventional use of the Johnson-Lindenstrauss lemma.

S. Kaski, "Dimensionality reduction by random mapping: fast similarity computation for clustering", Proceedings of The 1998 IEEE International Joint Conference on Neural Networks, pp. 413-418, 1998.

In random projection matrix, each column is of unit norm. Values are 0-mean Gaussian. So – random Gaussian vectors.
Reduced dimensionality space is not guaranteed to be in an orthonormal coordinate system.
Distortion of the variances/covariances relative to orthogonality of the random projections has approximate variance 2/m where m is low dimensionality.
For sufficient m, orthonormal system is mapped into a near-orthonormal system.
Kaski cites Hecht-Nielsen: the number of almost orthogonal directions in a coordinate system, that is determined at random in a high dimensional space, is very much greater than the number of orthogonal directions.



F. Critchley and W. Heiser, "Hierarchical trees can be perfectly scaled in one dimension", Journal of Classification, 5, 5-20, 1988.

Conventional random projections: random vectors that are iid 0-mean Gaussian. This is only necessary condition for preserving pairwise distances (Li, Hastie, Church, Proc. 12th ACM SIGKDD, 2006).
Other work has used 0 mean, 1 variance, 4th moment =3.
Also elements of random projection matrix from {-1,0,1} with different (symmetric in sign) probabilities.
It is acknowledged that: "a uniform distribution is easier to generate than normals, but the analysis is more difficult".

The non-conventional approach to random projections that is at issue in the case studies described here

Uniform [0,1) valued vectors in the random projection matrix.
Projections are rescaled to be in [0,1), i.e. closed/open interval.
Take mean (over random projections) of projected values.
It is known from the central limit theorem, and the concentration, or data piling, effect of high dimensional data clouds, that: pairwise distances become equidistant, and orientation tends to be uniformly distributed.
We find also: norms of the target space axes are Gaussian. (That is, before taking the mean of the projections.) As typifies sparsified data, the norms of the points themselves are negative exponential, or power law, distributed.

Visualization of Baire hierarchy.

Means of 99 random projections.

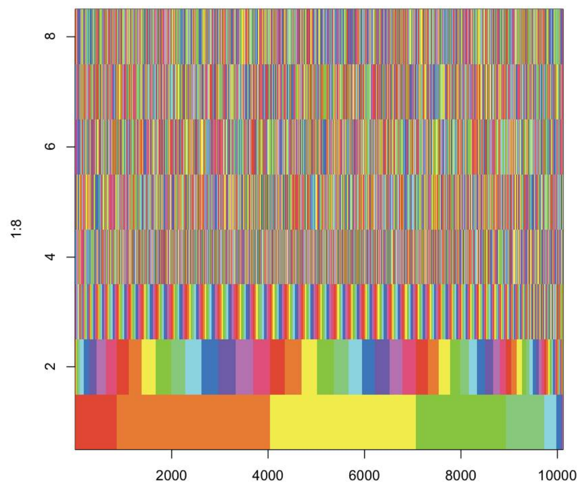
Abscissa: the 10118 (non-empty) documents are sorted (by random projection value).

Ordinate: each of 8 digits comprising random projection values.

Scaling followed by clustering

Correlation between most projection vectors > 0.99. We also found very high correlation between first principal component loadings and the mean random projection (> 0.999999).

Our objective is less to determine or model cluster properties as they are in very high dimensions, than it is to extract useful analytics by "re-representing" the data. That is to say, we are having our data coded (or encoded) in a different way.



Summary Remarks on Reading Baire Distance Properties from the (Mean) Random Projected Values

- We have a new way of inducing a hierarchy on data
- First viewpoint: encode the data hierarchically and essentially read off the clusters
- Alternative viewpoint: we can cluster information based on the longest common prefix
- We obtain a hierarchy that can be visualized as a tree
- We are hashing, in a hierarchical or multiscale way, our data
- We are targeting clustering in massive data sets
- The Baire method - we find - offers a fast alternative to k-means and a fortiori to traditional agglomerative hierarchical clustering
- At issue throughout this work: embedding of our data in an ultrametric topology

Traditional clustering: use pairwise distances, determine clustering structure (hierarchy or optimization of criterion). Often: then a partition is determined.

Here we build a series of partitions. Then the hierarchy is determined from them.

To be followed by – a look at accompanying web site content

- Data sets used in this presentation.
- Software used – mainly in R.
- The processing carried out.
- Samples of the output produced.

Tracking and Displaying Narratives of Behaviour, Actions, Activities, Living

- Unsupervised.
- Excellent for case studies:
- Film scripts, novels, social media, focused dialogue or monologue, e.g. court cases
- Generally, too, new applications for IoT, forensics.

Final Section of Lecture 3
New Perspectives and New Applications

- Towards narratives of IoT (Internet of Things), Big Data, and related applications.
- Ultrametricity – its role in tracking and displaying the unconscious.

General and Broad Perspectives

- Integrate mathematical underpinnings, cross-disciplinarity, company (commercialization) plans.
- Consider the overall computational science and engineering that support cloud computing and PaaS (Platform as a Service), analytics of massive data and big data, complex systems, IoT (Internet of Things) and cyber systems.

Historical perspective on challenges and potential.

Historical Periods Focused On:

Compute: Better computer infrastructure, including processor power and memory, up to the early 1990s.
Network: especially from the release of the Mosaic web browser in early 1993. Followed later by search engines.
Data: from the late 2000s.

Addressing the New Data Challenges:

The new Data Science: the visualization and verbalization of data. Extending data mining, knowledge discovery.

Achieving Excellence, Impact and Implementation (H2020 terminology) in:

Narratives of computation, and computational narratives in IoT (Internet of Things) and Cloud contexts.

Narratives of interaction, behaviour, experience. (Between lives of narratives and narratives of lives.)

Opportunity Abounds…
1) Connected health
2) Environment and Energy
3) Smart cities
4) Agriculture and environment

"Data is really the next form of energy … I view data as just a more processed form of energy."
(Christian Belady, General Manager, Data Center Services, Microsoft)

Ultrametric embedding: setting the scene.

(Then to follow: perspectives arising out of Matte Blanco's work, and their application.)

- Measuring metric content of data.
- Enhancing, inducing metric geometry in data.
- Measuring the ultrametric content of data.
- Enhancing, inducing an ultrametric topology.

Fingerprinting Using Ultrametricity

1) Wide range of time series signals

2) Wide range of texts

Quantifying and enhancing metric or ultrametric properties

- Classical multidimensional – embed in a metric space. Non-negative eigenvalues indicate how metric the data is.
- Can add constant to dissimilarities to enhance metric property.
- Ultrametricity: extent of respect of strong triangular, or ultrametric, inequality.

Assessing the ultrametricity of time series - I

- Fingerprint the time series signals based on their ultrametricity.
- Approach used: Take "sliding window" of fixed length. Used "window" sizes m = 5, 10, 15, … , 105, 110. Look at distance between each pair of values in the window. Encode as high/low distance. Test ultrametricity of all these indicators of local variability, and accumulate ultrametricity index over all such "windows".
- In "window" code each value as 1 if there is no/small change; and 2 if there is large change (up or down). Small/large defined relative to threshold $\max_{jj'} d_{jj'}^2/2$, $j,j' \in$ "window". Recoded values are metric.

Pervasive Ultrametricity

- As dimensionality increases, so does ultrametricity.
- In very high dimensional spaces, the ultrametricity approaches being 100%.
- Relative density is important: high dimensional and spatially sparse mean the same in this context.

Ultrametricity of time series - II

- So in a local region (window) we map pairwise dissimilarities onto relative (i.e. local) "change = 2" versus "no change = 1" distance.
- This is our "change/no change metric.

- Used signals: FTSE, USD/EUR, sunspot, stock, futures, eyegaze, Mississippi, www traffic, EEG/sleep/normal, EEG/petit mal epilepsy, EEG/irreg. epilepsy, quadratic chaotic map, uniform.

- Signals can be clearly distinguished. Extremes are: EEG and uniform.

Assessing the ultrametricity of text

Semantic networks defined from texts, through shared words.

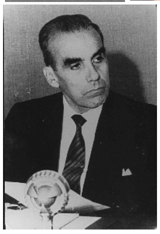
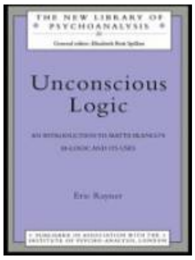
- Used as texts: 209 tales of Brothers Grimm; 266 Jane Austen chapters (full/partial) from 3 novels from 1811, 1813, 1817; 50 air accident reports; 384 dream reports. In all: nearly 1000 texts, over 1 million words.
- Use words as found (no stemming). Define χ^2 distance between profiles of frequency of occurrence table.
- We "euclideanized" by mapping into correspondence analysis factor space. E.g. for dream reports, 384 texts crossed by 11,441 words.
- Then we determined ultrametricity of text collections in factor space.
- We found dream reports to be highest in ultrametricity (albeit with fairly small coefficient of ultrametricity); and air accident reports similar to Grimm texts.

Other assessments were carried out on Aristotle's Categories; and James Joyce's Ulysses (304,414 words).

Ignacio Matte Blanco (1908-1995)

Born Santiago, Chile. Trained in London, Maudsley Hospital in psychiatry, London Institute in psychoanalysis. Worked in US, Chile, Italy. Died in Rome.



Ultrametricity (i.e. hierarchical substructure) for various text collections

- 209 Grimm Brothers tales, 209 x 7443, ultrametricity coefficient 0.1147
- 266 Jane Austen chapters or partial chapters, 266 x 9723, ultrametricity coefficient 0.1404
- 50 aviation accident reports, 50 x 4261, ultrametricity coefficient 0.1154
- 385 dream reports, 385 x 11441, ultrametricity coefficient 0.1933
- 171 Barbara Sanders dream reports, 171 x 7044, ultrametricity coefficient 0.2603 [We use this data later.]

Matte Blanco: There are "two fundamental types of being which exist within the unity of every man: that of the 'structural' id (or unrepressed unconscious or system unconscious or symmetrical being) which becomes understandable with the help of the principle of symmetry; and that visible in conscious thinking, which can roughly be comprehended in Aristotelian logic.'"

- Within a class, when symmetry logic applies: no contradiction, absence of negation, displacement, space and time vanish, no relations of contiguity, no order. "the unconscious does not know individuals but only classes or propositional functions which define the class".
- "Consciousness ... when confronted by a whole class can only consider it in two ways: either it focuses on the limits (or definition) of the class, that is, on those precise features which characterize it and distinguish it from all other classes, or it concentrates on the individuals which form the class.
- These are the clopen properties of a set or class or ball in an ultrametric space. Or a cluster.
- A class comes about through condensation.
- The principle of generalization relates different classes.
- "Symmetrical being alone is not observable in man." Even delineating it is "already an asymmetrical ... activity". So the symmetrical (and unconscious) is measurable but only in the context of the asymmetrical (and physical or empirical world).
- "We must ... keep in mind the possibility that if things are viewed in terms of multidimensional space, symmetrical being can actually unfold into an infinite number of asymmetrical relations."

"Approximation has two subtly different aspects: metric and topology. Metric tells how close our ideal point is to a specific wrong one. Topology tells how close it is to the combination of all unacceptable (non-neighboring) points".

2004 Kolmogorov Medal

N. Gower, L.Levin, A. Gammerman

- Leonid A. Levin, The tale of one-way functions, Problems of Information Transmission (Problemy Peredachi Informatsii), 39 (1), 92-103, 2003.
- (Also arXiv.org/abs/cs.CR/0012023)

Expressing Matte Blanco's symmetric mental processes

- Cluster members, as members of the cluster are conflated, they are identical.
- Every member of a cluster can be considered its centre. (And the radius of a cluster/ball is equal to its diameter.)
- Each cluster/ball is topologically open and closed. This is referred to as sets being clopen.

Words, and language, are tracers for what lies behind

- "Consciousness cannot exist without asymmetrical relations, because the essence of consciousness is to distinguish and to differentiate and that cannot be done with symmetrical relations alone."
- "Symmetrical being is translated into asymmetrical terms by means of words. Words (i.e. their meanings) are the asymmetrical tools of the translating-unfolding function." (Italics in the original.)
- We have that "words, abstract things, fulfill the function of differentiating between concepts and also between other things. They are bound to be, therefore, highly asymmetrical in their structure."

Metric Properties Are Closely Associated with Conscious Mental Processes i.e. our 3D ambient space, and the time dimension

- Semantics of interrelationships are expressible through metric mapping.
- (Whereas Matte Blanco's asymmetric thought processes requires classes and their relationships. Semantics are of relevance there too, e.g. anomaly and exception, as will be exemplified.)

How Matte Blanco's approach can be related to empirical data: through adoption of appropriate representations

- Matte Blanco s asymmetric logic (Aristotelian logic), asymmetric being. Conscious thought processes.
- ... Metric representation.
- Matte Blanco's symmetric being. Unconscious reasoning processes. Identity within a class. Only classes (or propositional functions) are at issue. A class comes about through (Freudian) condensation. Principle of generalization relates classes.
- ... Partially ordered sets (posets) of classes, implying: Ultrametric representation.
- Note however the earlier Matte Blanco indication that the symmetrical (and unconscious) is measurable but only in the context of the asymmetrical (and physical or empirical world).

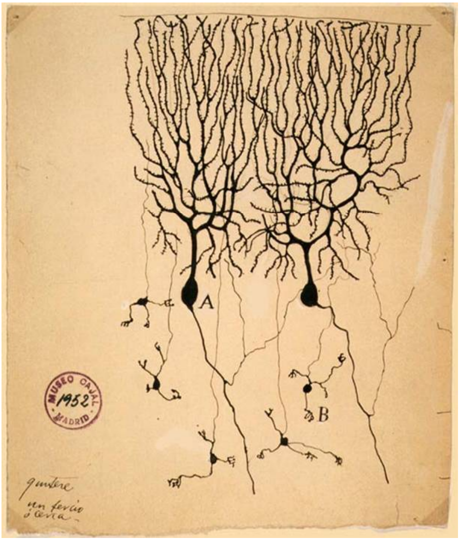
Ultrametric Properties Are Closely Associated with generative processes, and anomaly and exception.

(Potentially of relevance for Matte Blanco's symmetric logic, reflective of unconscious thought processes.)

- An ultrametric topology - a "tree topology".
- As a visual, or representational, model, it does rather well - very well, in fact - in expressing and encapsulating the context of Matte Blanco's unconscious reasoning.
- Also it - the hierarchy, the ultrametric topology that underpins the hierarchy - can be induced from the data.

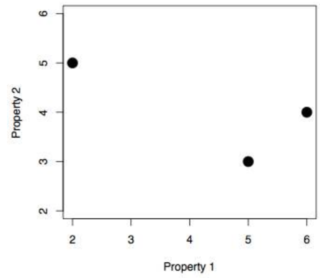
- Then the question becomes: what data?
- In line with Bourdieu's work, for example, we need to find the data that supports the investigation that we want to pursue.
- Induction (and transduction, and maybe even some deduction) is key.

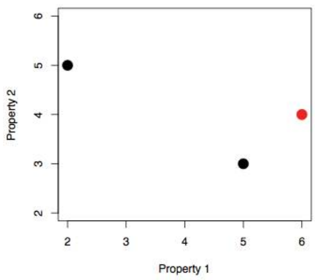
- Data? - written, or verbal, or other activity expressed by the subject, encompassing conscious reasoning, and unconscious reasoning processes.

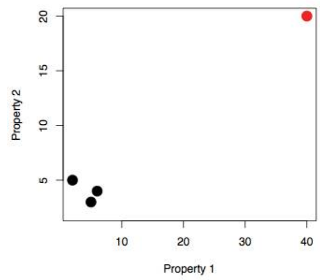


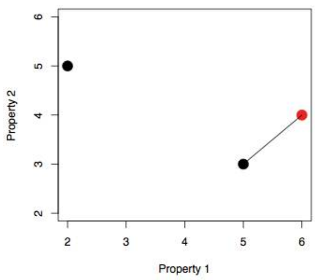
Cell structures from pigeon cerebellum, by Santiago Ramón y Cajal, 1899, Madrid

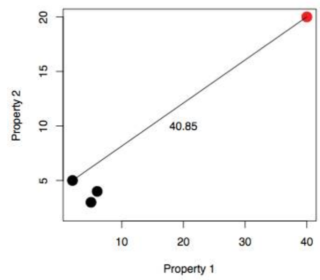
How a hierarchy expresses anomaly or change.

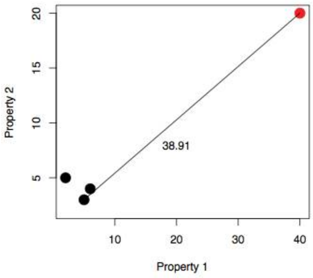






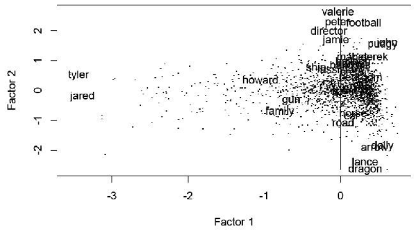


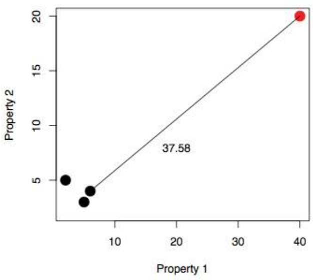




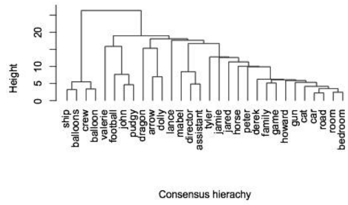
Determining vestiges of the subconscious through ultrametric component analysis

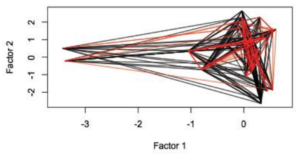
- DreamBank repository, 139 Barbara Sanders dream reports (1980 to 1997).
- 2000 terms, 30 of them indicated below - see how Tyler and Jared; Lance and dragon; Valerie, football, Peter, director, Jamie; etc. come out close. Note this is a planar approximation – 2.2% and 1.37% of the inertia explained by these factors, respectively.





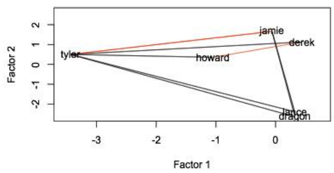
Determine ultrametric-respecting components in a consensus hierarchy





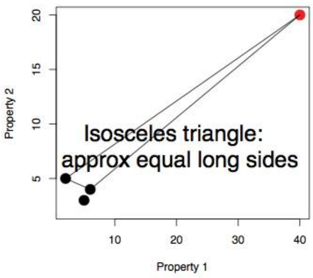
Upper left: consensus ultrametric (of two different agglomerative criteria, Ward's, single link).

Upper right: in principal factor plane, 163 ultrametric triplets determined in the (full dimensional, Euclidean) factor space.



Left: three of these triplets. Small base of isosceles (in full dimensional space) in red.

Howard: ex-husband, divorced, died.
Derek: had an affair with.
Jamie: friend, homosexual.
Tyler: co-worker.
Lance: married, former city manager assistant to disability rights group.
dragon: (no information about this).



Consensus hierarchy in an ultrametric topology

- Consider two hierarchies constructed on the same object set. Consider all triplets of objects, i,j,k.
- A morphologically consistent isosceles-with-small-base triplet means that the apex vertex, i, has the same label in both hierarchies, and the base pair, j and k, have the same labels.
- Count these matches between the hierarchies. The count is that of isosceles-with-small-base that are consistent for the two hierarchies. The total number of triangles considered, for n observations (hence n terminal nodes in the hierarchical tree), is n(n-1)(n-2)/(3.2.1). This furnishes a coefficient of ultrametric consensus between the two hierarchies, or ultrametric embedding of the same set of observations.
- Consensus hierarchy: For all isosceles-with-small-base triples that are morphologically consistent, use the minimum, between the two hierarchies considered, pairwise distances.
- For morphologically inconsistent triplets, we put all in the triplet considered to be equal to the minimum of all triplet pair distances, i.e. the single minimum value.
- We use the minimum in view of the maximal inferior ultrametric properties that ensue, i.e. the optimal fit from below. The consensus hierarchy is commutative over the pair of hierarchies, and unique for a given pair of hierarchies.

Conclusion

- We can pick out patterns that respect an ultrametric topology. Then we need to analyze them.
- Cf. earlier work where novels and other texts were quantified in terms of inherent ultrametric content. Following that, the task becomes (i) determining the ultrametric patterns, and (ii) interpreting such patterns.

- Next slide - some computational perspectives on this.

Far Greater Computational Power of Unconscious Mental Processes

- Conscious thought can process between 10 and 60 bits per second. E.g. reading, about 45 bits per second.
- But human visual system alone: about 10 million bits per second.
- Conscious thinking is both limited and limiting.
- "conscious thought is guided by expectancies and schemas" (Dijksterhuis and Nordgren, 2006). Limited capacity therefore goes hand in hand with use of stereotypes or schemas.
- Stereotypes may be "activated automatically (i.e., unconsciously)", but "they are applied while we consciously think about a person or group". Conscious thought is therefore more likely to (unknowingly) attempt "to confirm an expectancy already made".
- Unconscious thought is less biased in this way, and more slowly integrates information. "Unconscious thought leads to a better organization in memory", arrived at through "incubation" of ideas and concepts. "The unconscious works ... aschematically, whereas consciousness works ... schematically". "... conscious thought is more like an architect, whereas unconscious thought behaves more like an archaeologist".
- Through Matte Blanco' symmetry logic, and ultrametric topology, we know just how unconscious thought processes can be so superior from the computational aspect - they are hierarchical and generative.

(See my paper, "On ultrametric algorithmic information", Computer Journal, 2010.)