Mathematical Analysis of Complex Networks and Databases

Ph. Blanchard
Mathematical Physics & BiBoS
Bielefeld University
Oldenburg, April 27, 2015

Axiom 0

Real World $\Longrightarrow$ Models of parts of the world

Human thinking

The human mind thinks about relations between objects, thinks, concepts, agents, . . .

$$\text{Binary relations} \begin{cases} i \to j & \text{oriented} \\ i \sim j & \text{not oriented} \end{cases}$$

Relations + Objects $\to$ Network, graph

– Emergence creates new structures and functions by cooperation

– Nonlocal phenomena arise out of local interactions

Bridges between Probability Theory and Graph Theory

– **Bridge 1:** The Probabilistic Method

Ramsey 1930 Erdös-Renyi 1959

"Complete disorder is impossible"

$(\Omega, \mathcal{A}, \mathbf{P}) \sim$ probability space of random graphs

$P \sim$ graph property

$\mathbf{P}(P) > 0 \Rightarrow$ Platonic existence proof of graphs satisfying property P

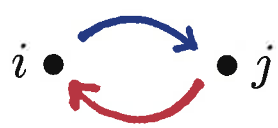
– **Bridge 2:** Graphs (and non negative symmetric real matrices) as probabilistic manifolds

$G = (V, E)$ $(M_{T,G}, g)$

$M_{T,G} \approx \mathbf{R}^{N-1}$

$|V| = N$ $T \sim$ random walk

$g_{ij} \sim$ commute time



Göteborg 27. IV. 2015

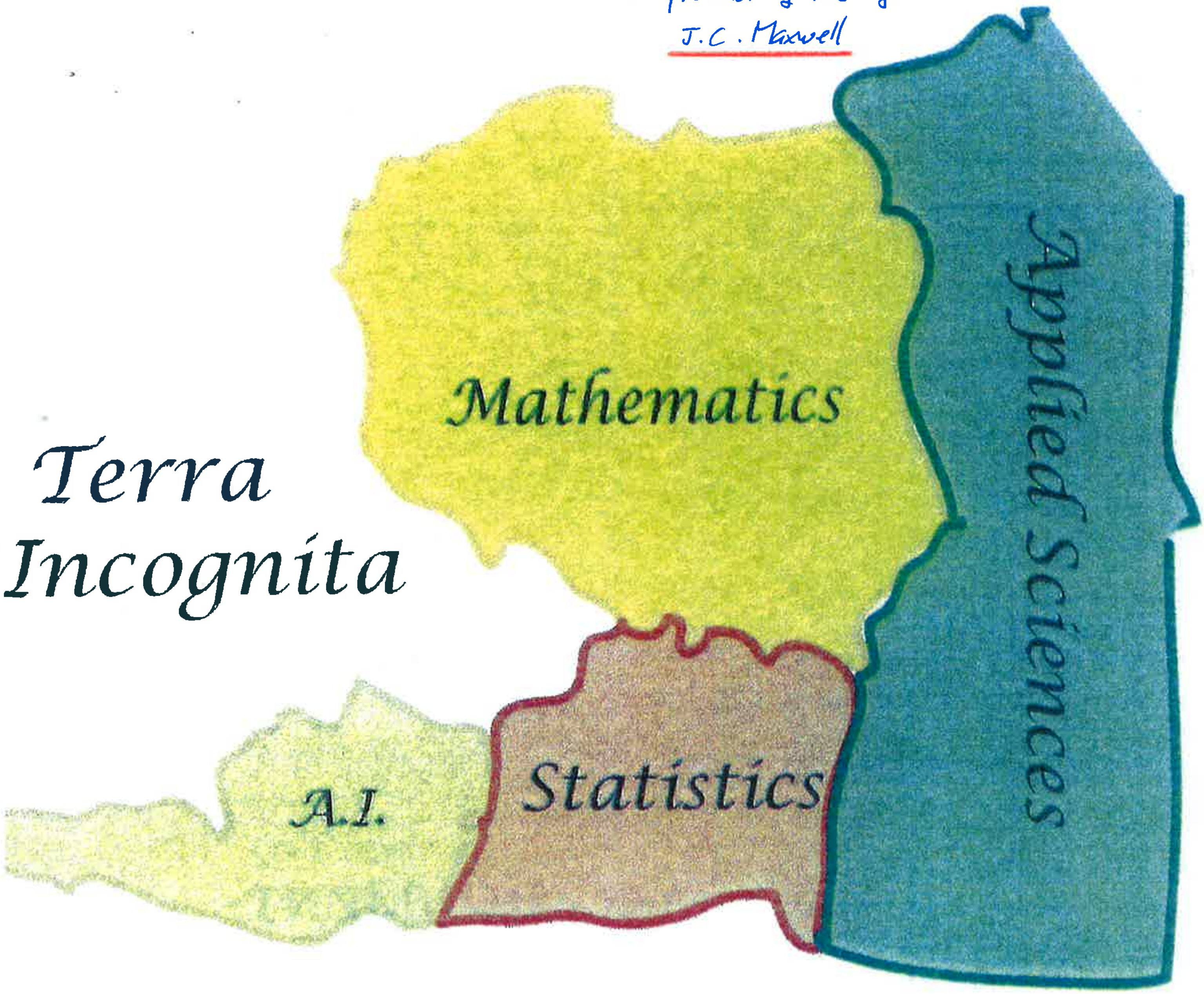
Mathematical Analysis of complex networks and data base

Ph. B, D. Volchenkov, T. Krüger, ...

Preaxiom The true logic of this world is probability theory

J.C. Maxwell

Terra Incognita



The greater world of mathematics and science

THEORY

- Probabilistic method (Erdős 1947, ...)

simple reasoning $\Rightarrow$ highly non trivial ex results

Ex 1

$$A \subset \mathbb{N} = \{1, 2, 3, \dots\}$$

$$|A| = n$$

A sum-free if

$a + b = c$ $a, b, c \in A$ has no solution

$x \in (0,1)$ random

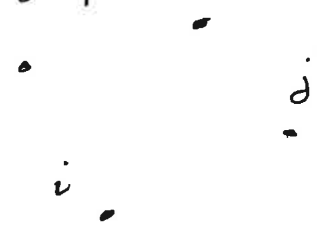
$$A_0 = \{a \in A \mid ax \bmod 1 \in \{\tfrac{1}{3}, \tfrac{2}{3}\}$$

A_0 is sum-free and $\mathbb{E}(A_0) = \frac{n}{3}$

- Random graphs theory

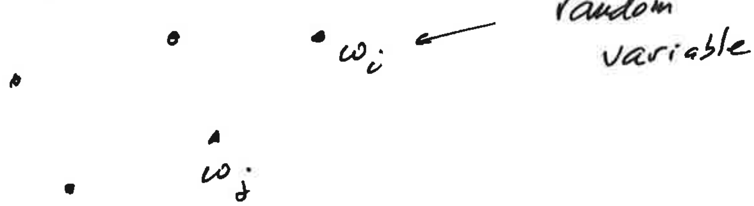
1) Classical Random Graphs

$\mathcal{G}(N, p)$



$$\mathrm{Prob}(i \sim j) = \frac{c}{N}$$

2) Inhomogeneous Random graphs



$$\mathrm{Prob}(i \sim j) = \frac{\omega_i \omega_j}{N}$$

Cameo graphs
Ph. B. Tyll Krüger

- Structure and functions

- Dynamics of graphs   How they growth

- Dynamics on graphs   How information propagates

PRAXIS

Generalized "communication" processes

- Urban spatial networks
- Linguistic, speech, writen language
- Music Pitches
- Architecture of human movements
- Epidemic spreading

a) Standard epidemic process :
classical infection spread : HIV, . . .

b) "Generalized" contagion process :
Knowledge, innovation, spread of t. rumors,
Corruption, terrorism,

Graph

$$G = (V, E)$$

$$A_{ij} = \begin{cases} 1 & i \sim j \\ 0 & \text{otherwise} \end{cases}$$

$$d(i) = \sum_j a_{ij}$$

Databases

$$W_{ij} \geq 0$$

affinity, correlation, relation

Strategy

- USE RANDOMNESS AS A TOOL TO REVEAL SYSTEM PROPERTY

  Randomness $\Leftrightarrow$ Exploration of G by random walks

- Why RW?

  RANDOM WALKS PROBE GLOBAL PROPERIES BUT USE ONLY LOCAL INFORMATION

- 1) Emergence creates new structures and functions by cooperation

  2) Nonlocal phenomena arise out of local interactions

α) EXPLOITING SYMMETRIES

From graph automorphisme Π

$$[\Pi, A] = 0$$

to RW

$$T = D^{-1} A \qquad D = \mathrm{diag}(d(1), d(2), \dots d(N))$$

j

i

$$\mathrm{Prob}(i \to j) = \frac{A_{ij}}{d(i)} = \begin{cases} \frac{1}{d(i)} & \text{or} \\ 0 \end{cases}$$

β) EXPLOITING SPECTRAL PROPERTIES OF THE LAPLACIAN ON G

$$L \equiv \mathbb{1} - T$$

$$Lu(i) = \frac{1}{d(i)} \sum_{j \sim i} u_j - u_i$$

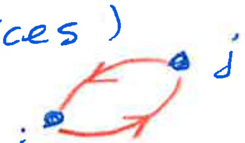
$$Lu(i) = 0 \iff u(i) = \frac{1}{d(i)} \sum_{j \sim i}$$

u is harmonic

Theorem $G = (V, E)$ can be embedded in a Riemannian manifold (M_g, g)

$$M_G \simeq \mathbb{R}^{N-1} \qquad N = \#(\text{vertices})$$

Locally $\qquad g_{ij} =$ commute time



γ) REDUCTION OF COMPLEXITY

From $N-1$ to 3 possible if we have "mass gaps" in $\sigma(L)$

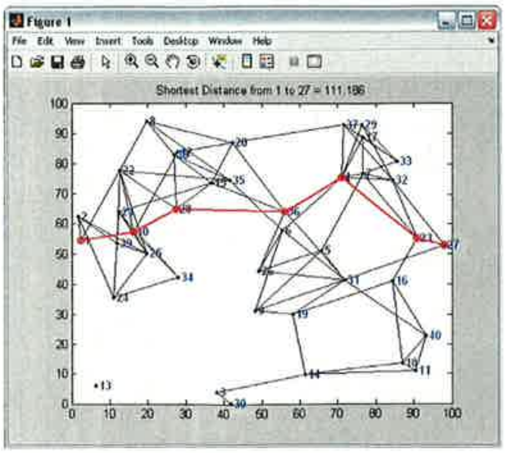
$$\sigma(L) = \{0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N \leq 2\}$$

connected

bipartite

$\Rightarrow$ Patter recognition, visualization

In classical graph theory:

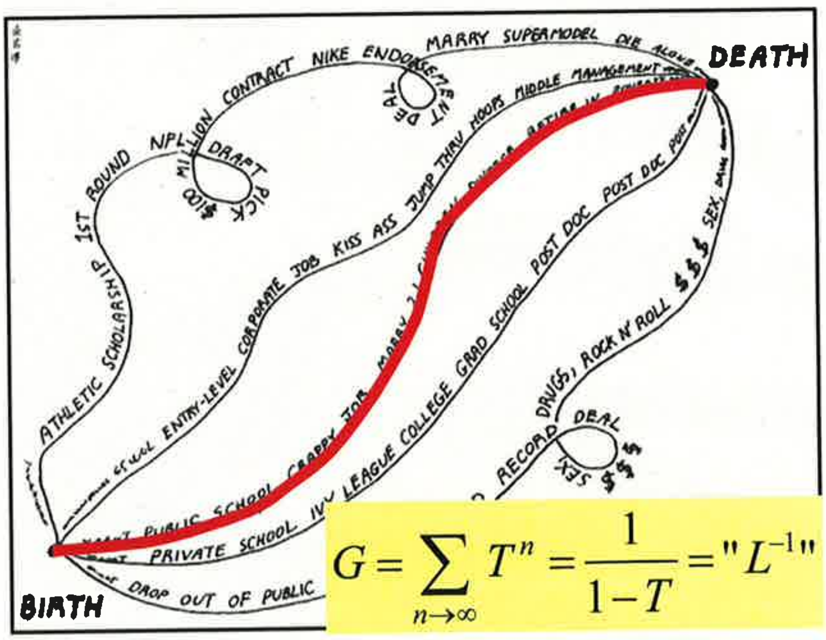


The shortest-path distance, insensitive to the structure of the graph:

$$d(i,j) = \min_{\hat{W}(\Gamma)} l\left(\hat{W}\,(i \to j)\right).$$

The distance = "a Feynman path integral"

sensitive to the global structure of the graph.



$$G = \sum_{n \to \infty} T^n = \frac{1}{1-T} = "L^{-1}"$$

Fractional automorphisms and structure learning

$$A \longrightarrow T_1 = D_1^{-1} A \longrightarrow$$ Centrality measure

$$D = \mathrm{diag}\,(d(1) \dots d(N))$$

$$\pi_i = \#\,\text{paths}\,(\dots \to i \to) = \frac{d(i)}{2E}$$

$$A^2 \longrightarrow T_2 = D_2^{-1} A^2 \longrightarrow \pi_i^{(2)} = \frac{\#\,\text{paths}_2\,(\to i \to)}{\#\,\text{paths}_2\,(G)}$$

$$A^n$$

All paths γ_{ij} of length n are equiprobable

$$T_{n\,ij} = \frac{(A^n)_{ij}\ \varphi_j}{\lambda_{max}\ \varphi_i}$$

$$A\varphi = \lambda_{max}\,\varphi$$

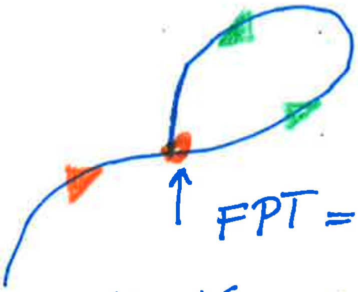
$n \to \infty$ All paths are equiprobable

$$T_{\infty\,ij} = \sum_{n=0}^{\infty} \frac{\varphi_j}{\varphi_i} \left(\frac{A}{\lambda_{max}}\right)^n$$

$$= \left(\frac{\lambda_{max}}{\lambda_{max}\mathbb{1} - A}\right)_{ij} \frac{\varphi_j}{\varphi_i}$$

Nearest neighbor RW flow $\neq$ Maximal entropy RW flow

Intelligibility / Predictability



FRT = first return time

FPT = first passage time

M. Kac : Distributions of FRT (recurrence time) are quite regular $FRT(i) = \frac{2|E|}{d_i}$

Distributions of FPT(i) are very irregular

An ergodic process returns infinitely many times with probability 1. The hitting event occurs only once

Ollivier's Ricci curvature

On a metric space $\forall x, y$ $x \neq y$ the Ricci curvature is defined along (x, y) by

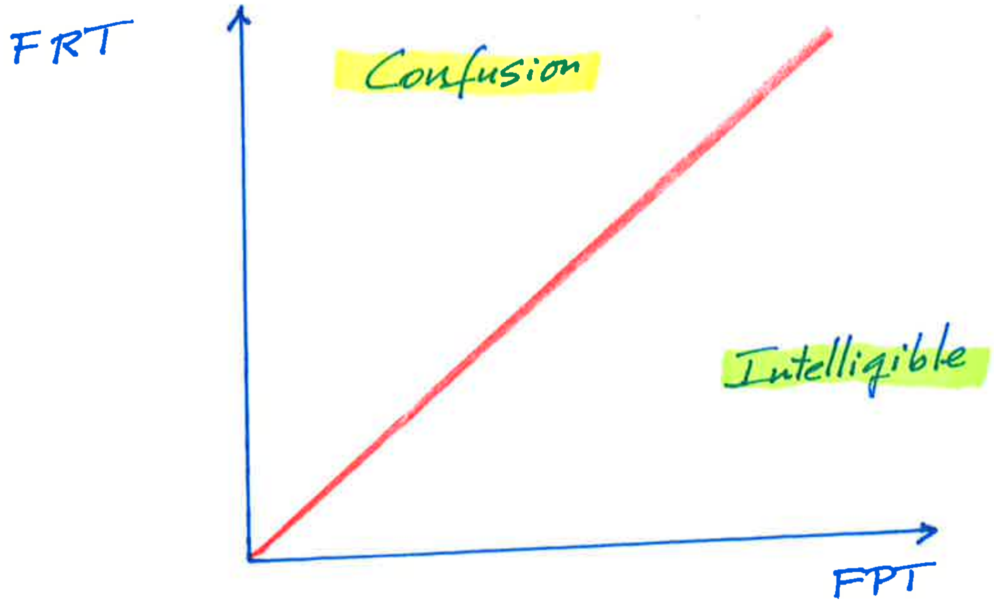
$$\mathcal{K}(x, y) = 1 - \frac{W(\mu_x, \mu_y)}{d(x, y)}$$

W Wasserstein distance between probability measures focused at x and y

$\sim$ problem of transport with minimal cost
Monge - Kantorovich problem

C. Villani Topics in optimal transportation
AMS (2003)

$$\mathcal{K}(x, y) = 1 - \frac{\|e_x - e_y\|_T^2}{d(x, y)}$$



$$\hat{\mathcal{X}}(x) = 1 - \frac{FPT(x)}{FRT(x)}$$

Confusion $\quad FRT > FPT \Rightarrow \hat{\mathcal{X}}(x) < 0$

Intelligibility $\quad FPT > FRT \Rightarrow \hat{\mathcal{X}}(x) > 0$

Evolution of networks

Weight $\quad W_{ij}(t) \Rightarrow g_{ij}(t)$

$$\dot{g}_{ij} = -R_{ij} - \frac{R}{N-1} g_{ij}$$

Ricci flow tends

to expand negatively curvated region

to contract positively curved region

$R > 0 \Rightarrow$ densification of the network of positive curvature

$\sim$ preferential attachment

$R < 0 \Rightarrow$ collapse and decomposition of the network of negative curvature

Examples

$L \psi_i = \lambda_i \psi_i \qquad L \psi_1 = 0$

$\psi_1 \Leftrightarrow \pi$ invariant measure of the RW on G

$\pi_i = \psi_1^2(i) > 0 \quad \forall i$ Perron – Frobenius $\psi_1(i) > 0 \quad \forall i \in V$

FRT(i) first return time

∼ regular distributions in general ergodicity

FPT(i) first passage time ∼ very irregular distributions

$$FRT(i) = \frac{1}{\pi(i)} = \frac{d(i)}{2|E|}$$

$$FPT(i) = \frac{1}{\pi(i)} \sum_{s=2}^{N} \frac{\psi_s^2(i)}{\lambda_s}$$

$$g(i,j) = H(i,j) + H(j,i)$$

- ACCESSABILITY $\qquad$ INTELLIGIBILITY

$FRT(i) \approx FPT(i)$

Ollivier's Rici curvature

$$R(i) = 1 - \frac{FPT(i)}{FRT(i)}$$

Intelligibility $\qquad R(i) \approx 0$

quasi flat

LEONARD FREED/MAGNUM

Bridge to nowhere

The future poverty hiding in cities

MARCUS CHOWN

IF YOU worry that your neighbourhood is going downhill, there could be a way to spot the signs before it happens. You might unwittingly be living in an area designed to foster crime, deprivation and ghettoisation, according to two mathematicians who have developed a method to spot hidden areas of geographical isolation in the urban landscape.

Many neighbourhoods are cut off from other parts of the city by poor transport links and haphazard urban planning, which can often lead to social ills. "Geographical isolation is a prime cause of social deprivation, economic inactivity and crime," says Dimitry Volchenkov at the University of Bielefeld in Germany.

Sociologists think that isolation worsens an area's economic prospects by reducing opportunities for commerce, and engenders a sense of isolation in inhabitants, both of which can fuel poverty and crime. For example, Laura Vaughan at University College London analysed street-by-street poverty in London over the past century and showed that inaccessible areas attract poorer inhabitants (*World Architecture*, vol 185, p 88).

Unfortunately, urban planners and governments have often failed to take such isolation into account when shaping the city landscape, not least because isolation can sometimes be difficult to quantify in the complex fabric of a major city.

Now Volchenkov and colleague Philippe Blanchard have created an algorithm that aims to capture a neighbourhood's inaccessibility, which they claim could expose hidden islands of future deprivation in cities (http://arxiv.org/abs/0710.3021).

To test their equations, Volchenkov and Blanchard analysed how easy it is to get to various places on the labyrinthine network of canals in Venice, Italy. They chose the city because its 96 canals, which snake around 122 small islands, provided a simple model to test the method. For their calculations, they imagined gondoliers dispersing randomly along the canals, as if they were drunk. This allowed them to work out the average number of random turns at junctions it would take to reach any particular place in Venice from various starting points.

Not surprisingly, the Grand Canal, the giant Giudecca Canal and the Venetian Lagoon were the most connected, says Volchenkov (see Map). In contrast, the researchers found that one district – the Venetian Ghetto – jumped out as by far the most isolated, despite being apparently well connected to the rest of the city. "On average, it took 300 random steps to reach, far more than the average of 100 steps for other places in Venice," says Volchenkov.

The Ghetto was created in March 1516 to separate Jews from the Christian majority of Venice. It persisted until 1797, when Napoleon conquered the city and demolished the Ghetto's gates. "You would never guess that the Ghetto was the most isolated district from the geography of Venice," says Volchenkov.

Although Venice was a simple place to analyse, Volchenkov says that their method could easily be used to identify isolated neighbourhoods in big cities with a complex web of roads, walkways and public transport systems.

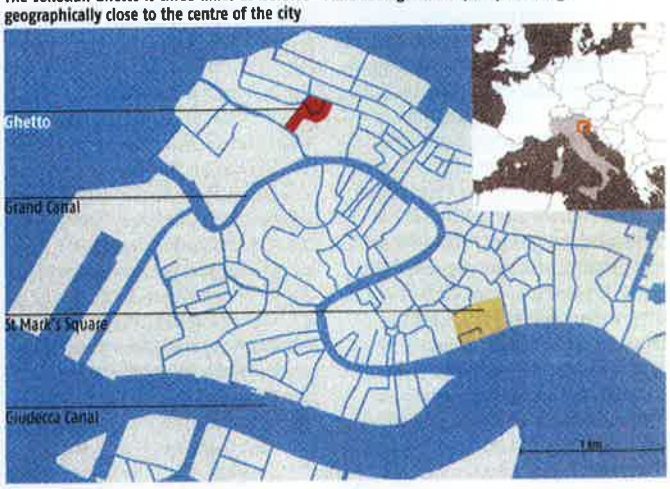
For example, he believes that the Bowery, which was a deprived district of New York for most of the 20th century, might have been isolated from nearby areas at the time. "In existing cities, efforts should be made to reconnect isolated districts, perhaps by building tunnels and bridges," Volchenkov says.

"Geographical isolation is a prime cause of social deprivation, economic inactivity and crime, but can be hard to quantify"

Geoffrey Ingarfield, a housing expert at Middlesex University in London, points to an example in the UK where planners did the opposite in the 1970s, upgrading a major road and splitting a neighbourhood in two. "The widening of the A13 in Newham, London, completely isolated the area to the immediate south, causing shops and the last secondary school to close," he says. "Unable to cross the road, the people to the south became more isolated and socially deprived." ●

GETTING LOST IN VENICE

The Venetian Ghetto is three times as isolated as the average district, despite being geographically close to the centre of the city



Geometric Representations of Language Taxonomies

w_1, w_2 2 words

$$D(w_1, w_2) = \frac{\| w_1, w_2 \|}{\max(|w_1|, |w_2|)}$$

$\| w_1, w_2 \|$ edit distance
$\equiv$
minimal number of insertions, deletions or substitutions needed to transform $w_1 \to w_2$

$|w| = \#(\text{characters in } w)$

$$D(\underline{mil}k, \underline{Mil}ch) = \frac{2}{5}$$

Swadesh's list = 200 meanings essentially resistant to change

l_1, l_2 2 languages

$$d(l_1, l_2) = \sum_{\alpha=1}^{200} D(w_\alpha^{(l_1)}, w_\alpha^{(l_2)})$$

α "meaning" $\approx$ chat

$N = \#(\text{languages})$

$$\Delta = \operatorname{diag}(\delta_{l_1}, \dots \delta_{l_N})$$

$$\delta_{l_i} = \sum_j \delta(l_i, l_j)$$

$$P(l_i, l_2) = \lim_{n \to \infty} \sum_{k=0}^{n} T^k(l_i, l_2)$$

$$= L^{-1}$$

$$T(l_i, l_j) = \Delta^{-1} d(l_i, l_j)$$

L^{-1} generalized inverse of the Laplacian L

$(L^{-1})_{ii}$ ~ first passage time

$(L^{-1})_{ij}$ ~ quantify the interference of 2 random walks ending at l_i and l_j respectively

$\{\psi_k\}_{1 \leq k \leq N}$ eigenvectors of L^{-1}

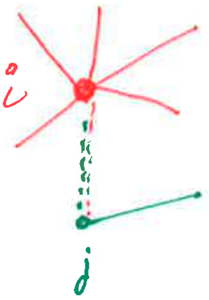
ONB

$l_i \longrightarrow (\psi_1(i), \psi_2(i), \dots \psi_N(i))$

- **Communication index**

Knowledge, innovation transfer ....

Time is a ressource !



$$C_{ij} = \min\left(\frac{1}{d(i)}, \frac{1}{d(j)} \right) = \frac{1}{\max(d(i), d(j))}$$

- **Inhomogeneity of the population**

$i \to$ Random variable of distribution $\rho(i)$

$$\mathrm{Prob}(i \sim j) = \frac{\chi(\omega_i, \omega_j)}{N}$$

$$E(\#\text{edges}) \simeq \frac{N}{2} \int \chi(\omega_i, \omega_j)\, \rho(\omega_i)\, \rho(\omega_j)$$

Sparse graphs

2 different social structures

Multiplicative $\quad \chi^*(\omega_i, \omega_j) = \psi(\omega_i)\, \psi(\omega_j)$

Additive $\quad \chi^*(\omega_i, \omega_j) = \psi(\omega_i) + \psi(\omega_j)$

- $\| T_{\chi^*} \| \geq \| T_{\chi^+} \|$

- local clustering $\begin{cases} \text{slows down SEP} \\ \text{accelerates GEP} \end{cases}$

Terrorism and Phase Transition Nature 462 911 - 914 (2009)

0 ~ neutral about terrorism 1 ~ passive supporter

2 ~ active terrorist ..

Phase transition $0 \to 1$ induced by collateral damages

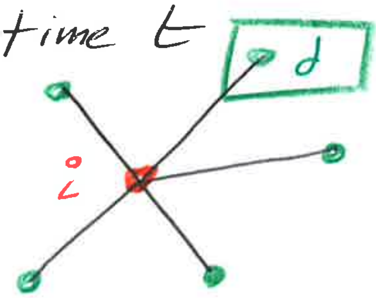
$x \sim \#(\text{casualties})$ $\quad p(x \geq) \sim x^{-5/2}$ Afganistan

No military solution ....

EPIDEMIC PROCESSES : SEP & GEP

26. April 2015

SEP $x_t(i)$ infection states of i at time t



$x_t(i) \in \{0, 1, 2, \dots, p\}$

0 non infected, 1 infected, ...

$B_1(i) = \{ j \mid j \sim i \}$

$$\text{Prob}\{x_t(i) = k\} = f_\alpha\left(x_t(B_1(i))\right) + g_\alpha\left(\sum w_i x_t(i)\right)$$

local dynamics

global dynamics

j infected

Prob(j infects i) $= \lambda$

$$f_\alpha(i) = 1 - (1-\lambda)^{\#(\text{infected in } B_1(i))}$$

NEMO FP6 2006-2009

$$\simeq \lambda \, \#(\text{infected in } B_1(i))$$

GEP social infection

- Threshold effet

$$\#(\text{infected neighbors}) < \Delta$$

$$\lambda = \varepsilon \quad \text{very small}$$

$$\#(\text{infected neighbors}) \geq \Delta$$

$$\lambda = 1 - \varepsilon' \qquad \varepsilon' \text{ small}$$

- Mean field dependence

$$g_\alpha = g_\alpha\left(b_t = \frac{\#(\text{infected at time } t)}{N}\right)$$

$b_t \equiv$ densition of infected at time t

$=$ prevalence

NEMO

NETWORK MODELS, GOVERNANCE AND R&D COLLABORATION NETWORKS

Newsletter #5 October 2008

R&D NETWORKS ACROSS EUROPEAN REGIONS

What determines the spatial configuration of R&D collaboration networks in Europe at a regional level? A deeper understanding of this issue is urgently needed for future governance of science, technology and innovation policies in the European Union. Within the NEMO project, this issue is addressed by means of exploratory spatial data analysis as well as spatial interaction modelling.

The spatial region-by-region R&D network of the Fifth EU Framework Programme (1998-2002) is visualised in Figure 1. The nodes represent regions (NUTS2), and their size is related to the number of links connected to a region. The central hub in this spatial network is Île-de-France. A high density can also be observed for south-eastern regions of the UK, northern Italian regions, southern and western regions in Germany, the Netherlands and Switzerland as well as for the capital regions in Greece and Spain. The number of links to Eastern European regions is quite low. The maximum number of collaborations is 6,152, referring to intraregional collaborations within the region of Île-de-France, while the maximum interregional collaboration activity (1,609 collaborations)

(continued on page 2)

Figure 1: Cross-region R&D collaborations in Europe as captured by research projects funded by EU FP5. [Scherngell T., Barber M. (2008), Spatial Interaction Modelling of Cross-Region R&D Collaborations. Empirical Evidence from the EU Framework Programmes. Papers in Regional Science, forthcoming]

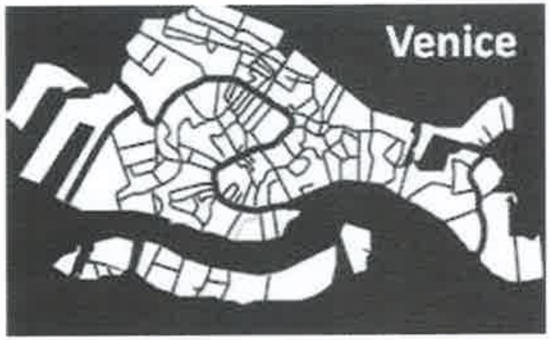
THE NEMO NEWSLETTER

The objective of the periodic NEMO Newsletter is to provide a platform of interdisciplinary information exchange and discourse for all sciences concerned with complex interorganisational R&D collaboration networks, and to promote the NEMO project worldwide. The Newsletter will offer regular insights into the NEMO project and document its results including previews of NEMO publications.

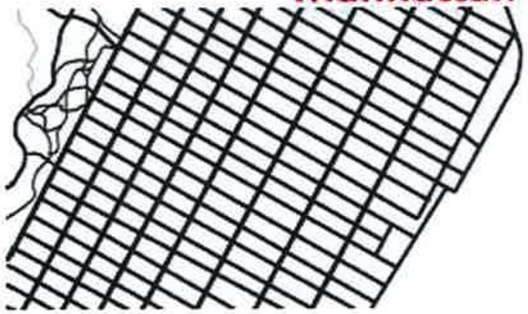
Other continuous features include the publication of short articles, comments on interesting links, and information about events or publications which are located in the area of our research. External contributions are welcome and should be addressed to the editor. The Newsletter is published quarterly on the NEMO website http://www.nemo-net.eu

CONTENTS OF THIS ISSUE

Examples: cities & languages

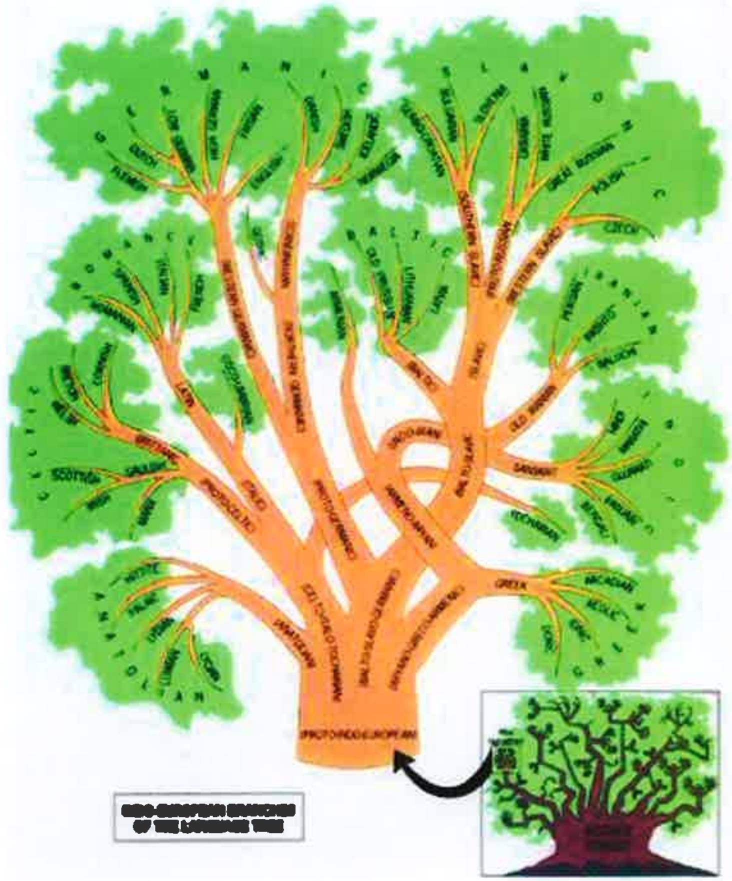


Manhattan



Neubeckum

Language Taxonomy

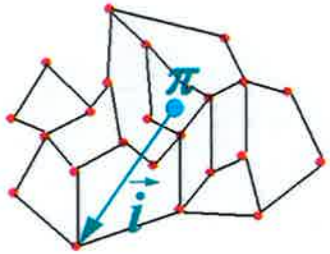


FIRST-PASSAGE AND COMMUTE TIMES

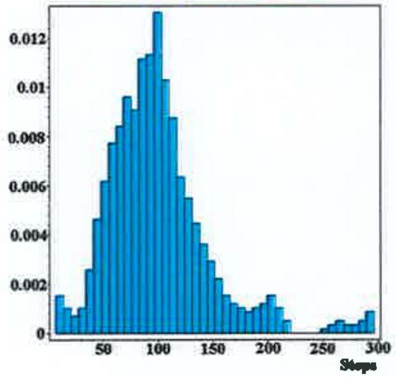
First-Passage time

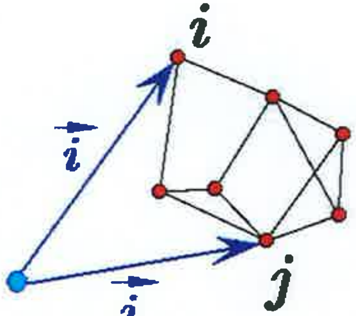
$$\mathbf{e}_i = \langle 0,0,\ldots 1_i,\ldots 0\rangle, \quad i \in V \qquad \varphi T = \lambda \varphi,$$

$$\|\mathbf{e}_i\|_T^2 = \frac{1}{\pi_i}\sum_{s=2}^{N}\frac{\varphi_{s,i}^2}{1-\lambda_s}$$



FPT distribution in Venice





Commute time

$$\|\mathbf{e}_i - \mathbf{e}_j\|_T^2 = \sum_{s=2}^{N}\frac{1}{1-\lambda_s}\left(\frac{\varphi_{s,i}}{\sqrt{\pi_i}} - \frac{\varphi_{s,j}}{\sqrt{\pi_j}}\right)^2$$

$$= \|\mathbf{e}_i\|_T^2 + \|\mathbf{e}_j\|_T^2 - 2(\mathbf{e}_i,\mathbf{e}_j)_T,$$

FPT akin to the famous path integrals of R. Feynman accounting for all paths possible in the graph.

Springer Series in Synergetics

Philippe Blanchard · Dimitri Volchenkov

Random Walks and Diffusions on Graphs and Databases

An Introduction

Most networks and databases that humans have to deal with contain large, albeit finite number of units. Their structure, for maintaining functional consistency of the components, is essentially not random and calls for a precise quantitative description of relations between nodes (or data units) and all network components.

This book is an introduction, for both graduate students and newcomers to the field, to the theory of graphs and random walks on such graphs. The methods based on random walks and diffusions for exploring the structure of finite connected graphs and databases are reviewed (Markov chain analysis). This provides the necessary basis for consistently discussing a number of applications such diverse as electric resistance networks, estimation of land prices, urban planning, linguistic databases, music, and gene expression regulatory networks.

Physics
ISSN 0172-7389
ISBN 978-3-642-19591-4
9 783642 195914
▶ springer.com

Blanchard · Volchenkov

Random Walks and Diffusions on Graphs and Databases

Springer Series in Synergetics

Springer Complexity

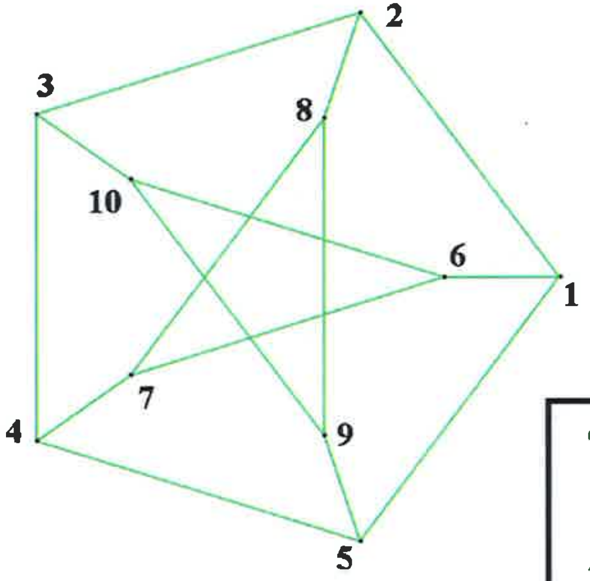
Philippe Blanchard
Dimitri Volchenkov

Random Walks and Diffusions on Graphs and Databases

An Introduction

Springer

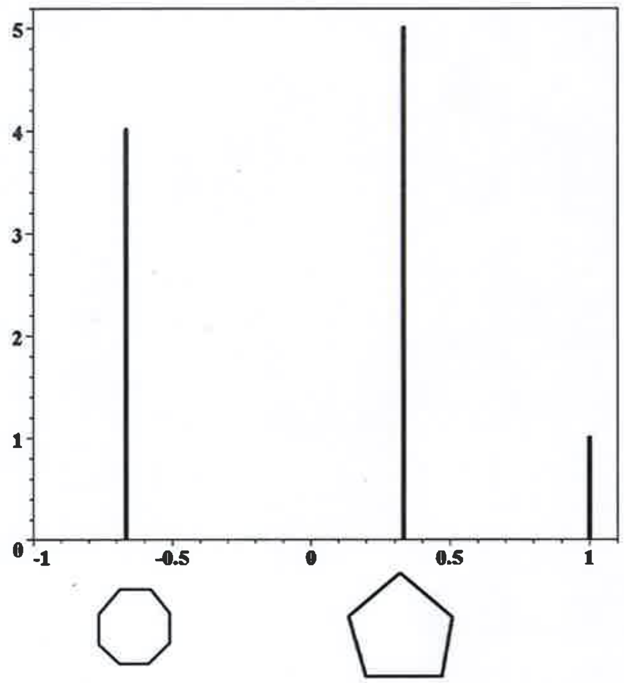
Petersen graph

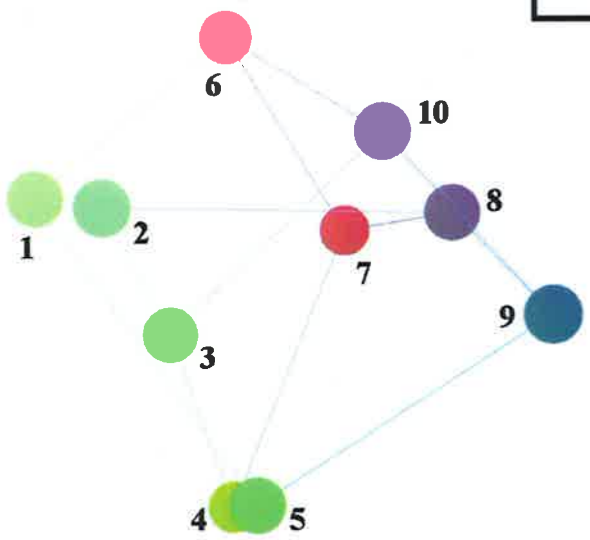


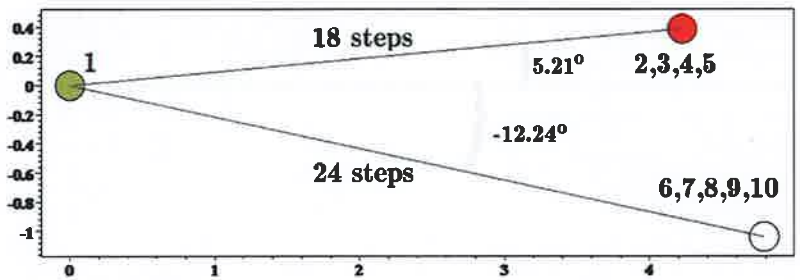
$$T_{ij} = \frac{A_{ij}}{3}, \quad \varphi T = \lambda \varphi,$$

$$\mathrm{T}_i \equiv \|i\|_T^2 = 9.9 \text{ steps}$$

$$\tau_i = \frac{1}{\pi} = 10 \text{ steps}$$

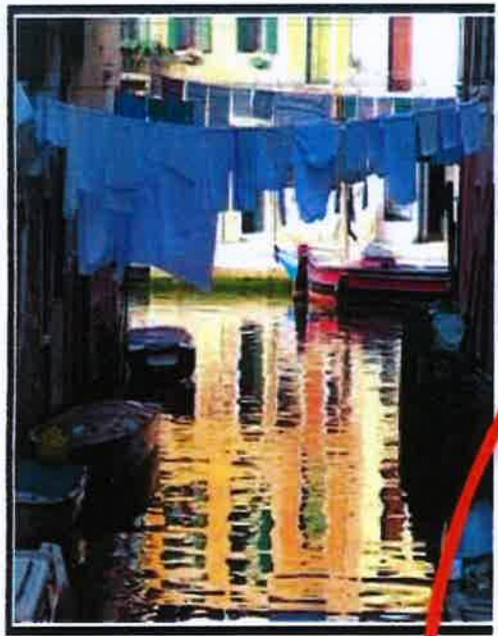






Social isolation vs. structural isolation

Ghetto di Venezia



This week

The future poverty hiding in cities

fuel poverty and crime. For example, Laura Vaughan at University College London analysed street-by-street poverty in London over the past century and showed that inaccessible areas attract poorer inhabitants (*World Architecture*, vol 185, p 88).

Unfortunately, urban planners and governments have often failed to take such isolation into account when shaping the city landscape, not least because isolation can sometimes be difficult to quantify in the complex fabric of a major city.

Now Volchenkov and colleague Philippe Blanchard have created an algorithm that aims to capture a neighbourhood's inaccessibility, which they claim could expose hidden islands of future deprivation in cities (http://arxiv.org/abs/0710.3021).

To test their equations, Volchenkov and Blanchard analysed how easy it is to get to various places on the labyrinthine network of canals in Venice, Italy. They chose the city because its

300 random steps to reach, far more than the average of 100 steps for other places in Venice," says Volchenkov.

The Ghetto was created in March 1516 to separate Jews from the Christian majority of Venice. It persisted until 1797, when Napoleon conquered the city and demolished the Ghetto's gates

says that their method could easily be used to identify isolated neighbourhoods in big cities with a complex web of roads, walkways and public transport systems.

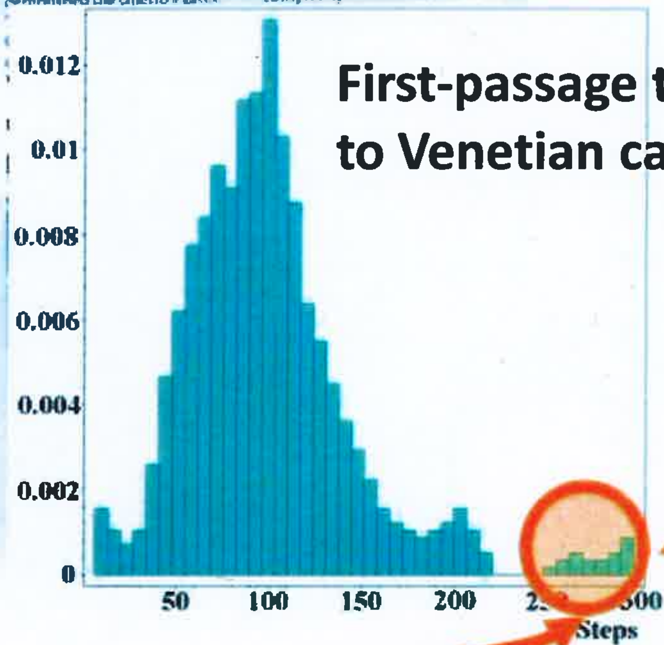
For example, he believes that the Bowery, which was a deprived district of New York for most of the 20th century, might have been isolated from nearby areas at the time. "In existing cases, efforts should be made to reconnect isolated districts, perhaps by building tunnels and bridges," Volchenkov says.

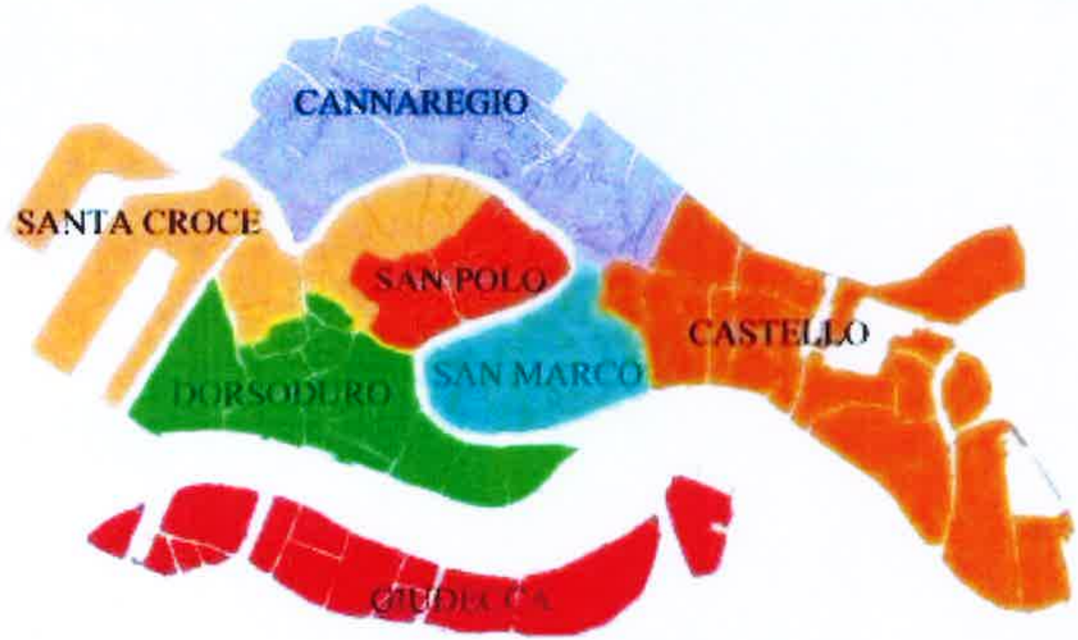
"Geographical isolation is a prime cause of social deprivation, economic inactivity and crime, but can be hard to quantify"

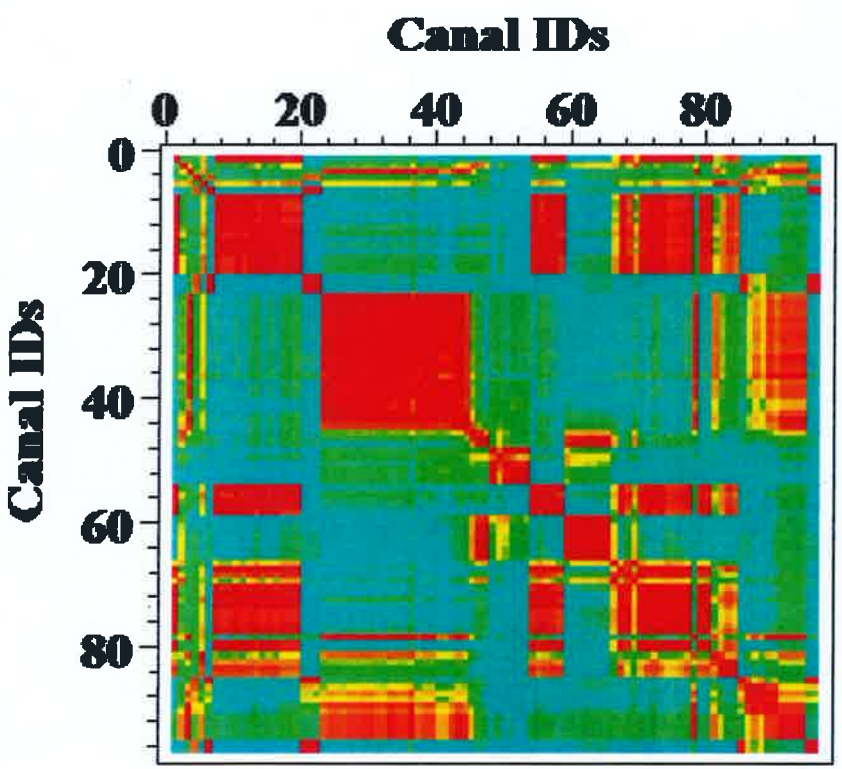
Geoffrey Ingarfield, a housing expert at Middlesex University in London, points to an example in the UK where planners did the opposite in the 1970s, upgrading a major road and splitting a neighbourhood in two. "The widening of the A13 in Newham, London, completely isolated the area to

Bridge to nowhere

First-passage times to Venetian canals




CANNAREGIO
SANTA CROCE
SAN POLO
CASTELLO
SAN MARCO
DORSODURO


Canal IDs
0
20
40
60
80
Canal IDs
0
20
40
60
80

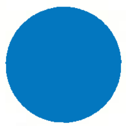

GIUDECCA
CASTELLO
DORSODURO
SAN POLO
SANTA CROSE
SAN MARCO
CANNAREGIO

A variety of random walks at different scales

An example of equivalence relation:

walks of the given length n starting at the same node are equivalent.

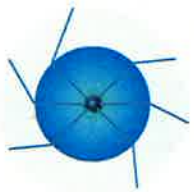
Equiprobable walks: | *Stochastic matrices:* | Left eigenvectors (μ=1) *Centrality measures*:



$$\mathbf{A} \rightarrow \mathbf{T}_1 = \mathbf{D}_1^{-1}\mathbf{A} \rightarrow \pi_i^{(1)} = \frac{\#\,\mathrm{Paths}_1(\rightarrow i \rightarrow)}{\#\,\mathrm{Paths}_1(G)} = \frac{\deg(i)}{2E}$$

The "stationary distribution" of the nearest neighbor RW

$$\downarrow \qquad \downarrow$$



$$\mathbf{A}^2 \rightarrow \mathbf{T}_2 = \mathbf{D}_2^{-1}\mathbf{A}^2 \rightarrow$$

$$\downarrow \qquad \downarrow$$

...

$$\cdots \qquad \cdots$$

$$\mathbf{A}^n \rightarrow \mathbf{T}_n = \mathbf{D}_n^{-1}\mathbf{A}^n \rightarrow \pi_i^{(n)} = \frac{\#\,\mathrm{Paths}_n(\rightarrow i \rightarrow)}{\#\,\mathrm{Paths}_n(G)}$$

$$\downarrow \qquad \downarrow$$

...

$$\cdots \qquad \mathbf{T}_\infty = \left[\varphi_x \Big/ \sum_{y\in V}\varphi_y\right]_{x=1,..N} \rightarrow \pi_x^{(\infty)} = \frac{\varphi_x}{\sum_{y\in V}\varphi_y}$$

Distances between words & languages

200 words

I
you (singular)
he
we
you (plural)
they
this
that
here
there
who
what
where
when
how
not
all
many ....

- The ***Swadesh list*** contains terms which are common to all cultures and which concern the basic activities of humans

MIL~~CH~~ K = MILK   Levenshtein (Milch, Milk) =2

Levenshtein distance (edit distance) is a measure of the similarity between two strings, the number of deletions, insertions, or substitutions required to transform one into another.

MIL~~CH~~ K = MILK   Normalized Levenshtein (Milch, Milk) =2/5

the edit distance divided by the number of characters of the longer of the two

$$\mathrm{Dist}(L_1, L_2) = \frac{1}{200}\sum_{n=1}^{200} \mathrm{Norm.Levenshtein}(u_n, v_n) \in [0,1],$$

u and v have the same meaning.

Distance between languages by lexicostatistics

- <u>**Levenshtein distance (edit distance)**</u>

  is a measure of the similarity between two strings - the number of deletions, insertions, or substitutions required to transform one into another.

MIL~~CH~~ K ***= MILK*** Normalized Levenshtein distance (Milch, Milk) by the maximal length of the two **= 2/5**

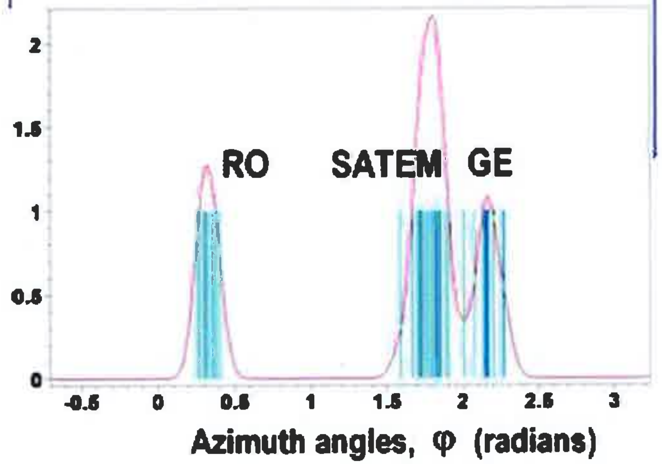
- The ***Swadesh list*** *contains terms which are common to all cultures and which concern the basic activities of humans.*

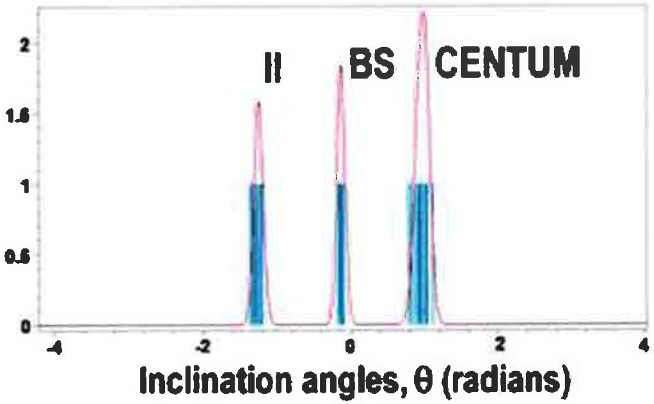
200 words ***I, you (singular),he, we, you (plural), they, this, etc…***

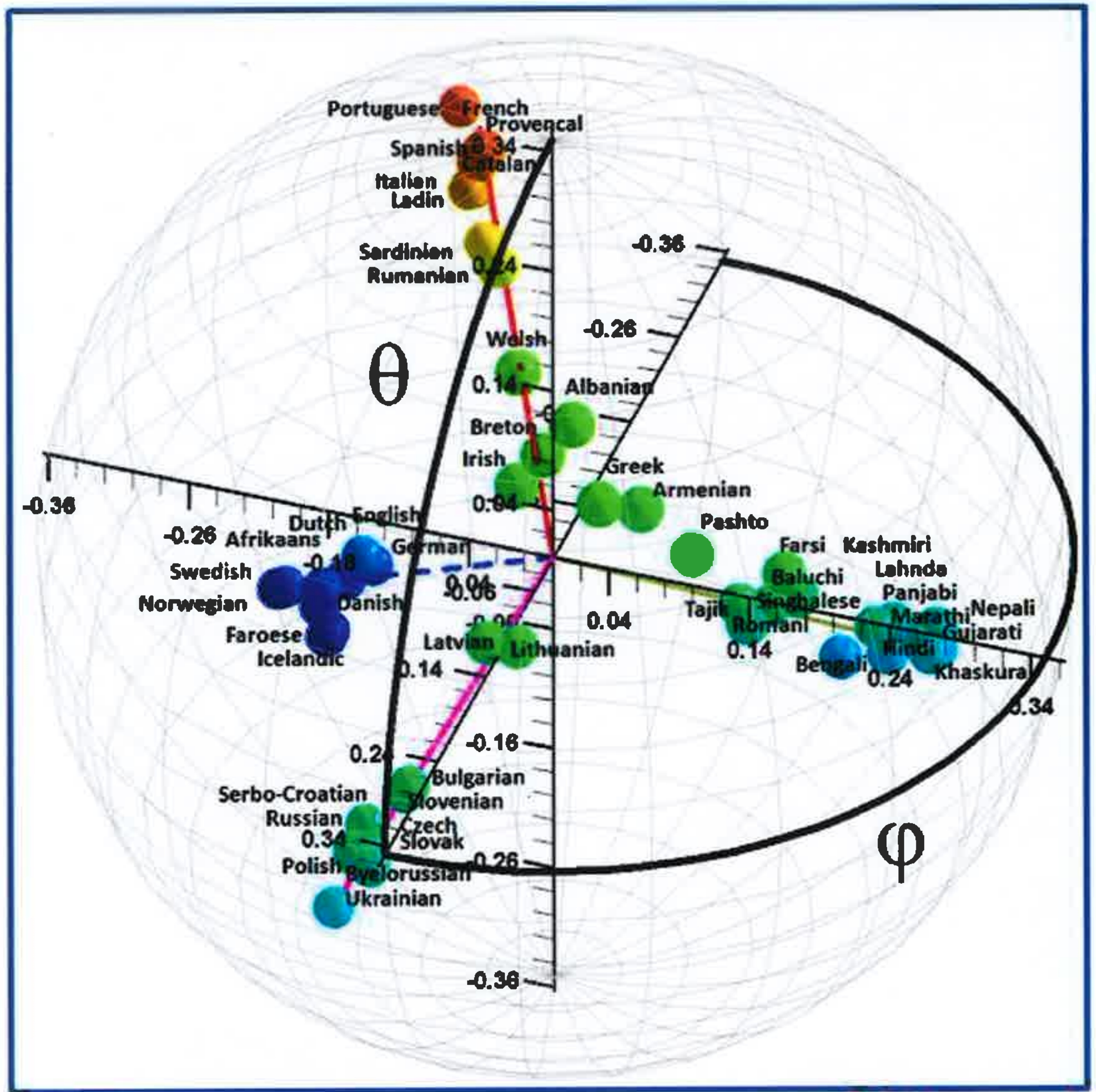
$$\mathrm{Dist}(\mathrm{L}_1,\mathrm{L}_2)=\frac{1}{200}\sum_{n=1}^{200}\mathrm{Norm.Levenshtein}(u_n,v_n)\in[0,1],$$

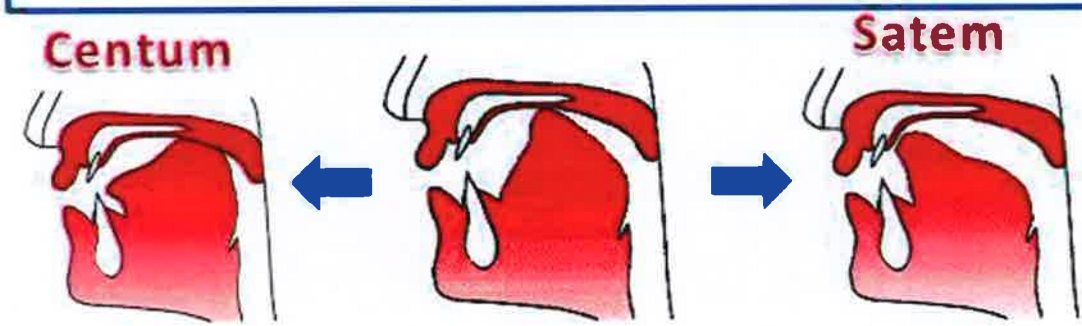
u and v have the same meaning.

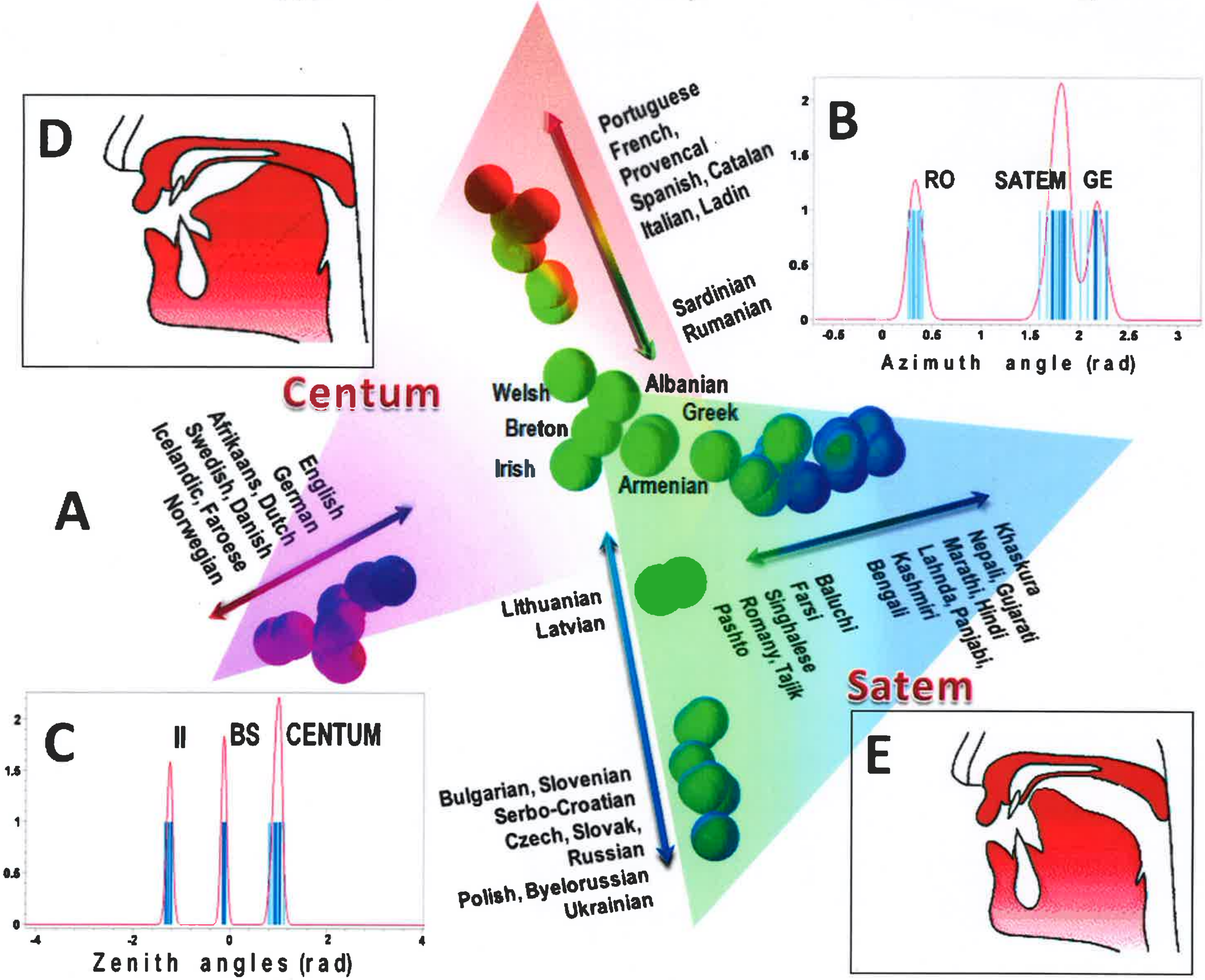
	L_1	L_2	...	L_N
L_1	0.87693	0.86967	0.85137	0.86254
L_2	0.8425	0.62306	0.86193	0.86479
	0.8573	0.8699	0.86146	0.89153
...	0.88004	0.877 17	0.8731	0.89119
	0.87077	0.89438	0.85755	0.89021
...	0.83333	0.89629	0.85294	0.85179



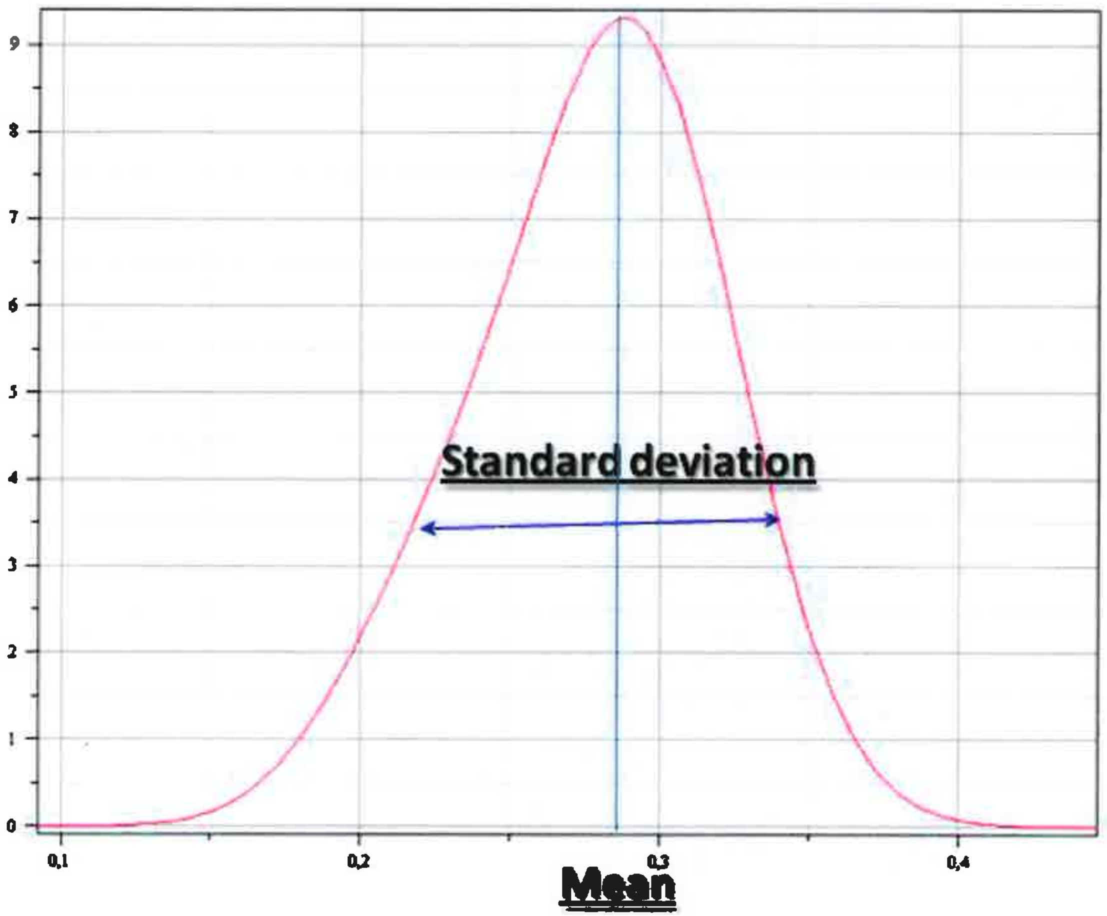






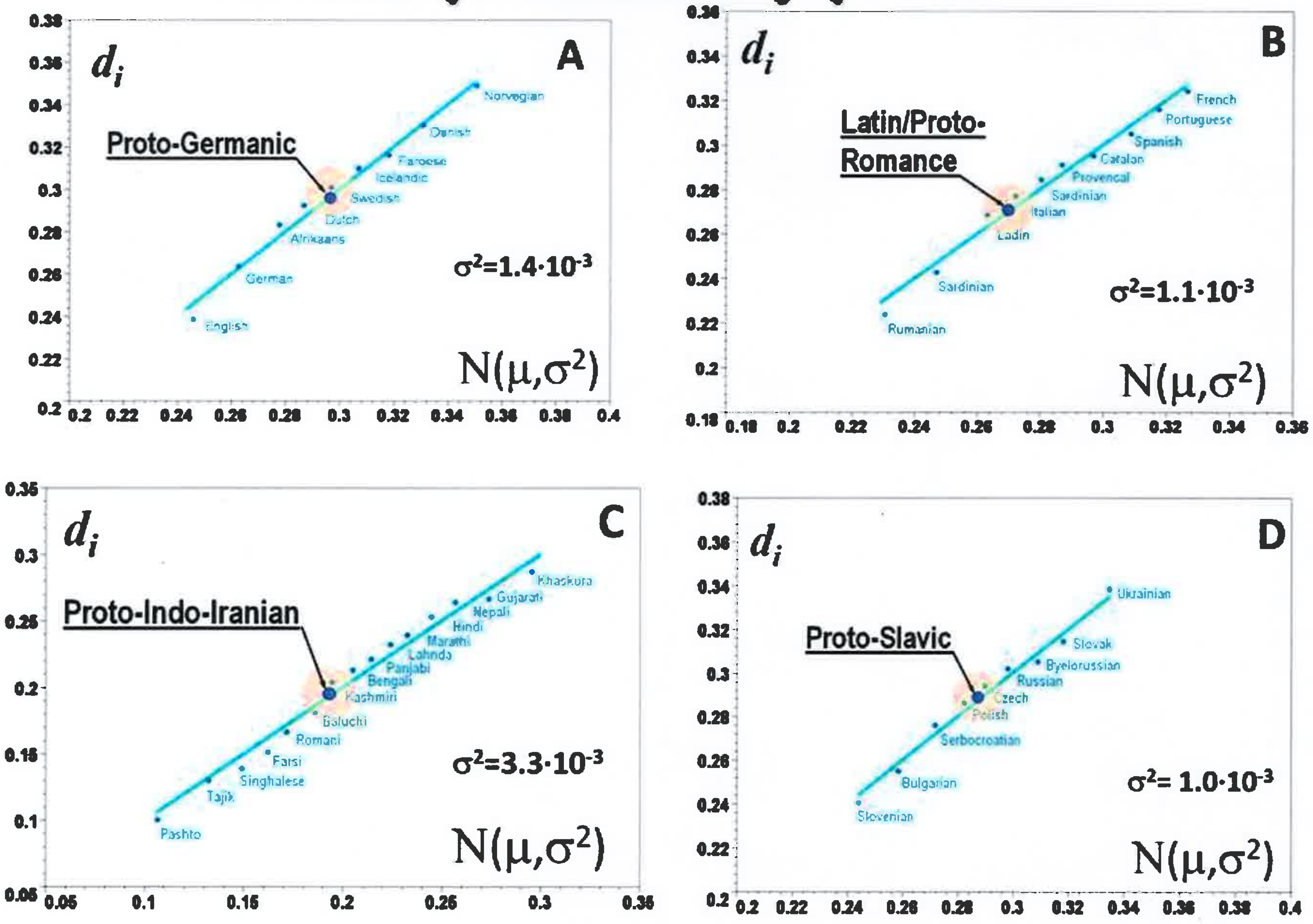


D
Centum
English
German
Afrikaans, Dutch
Swedish, Danish
Icelandic, Faroese
Norwegian
A
Portuguese
French,
Provencal
Spanish, Catalan
Italian, Ladin
Sardinian
Rumanian
Welsh
Breton
Irish
Albanian
Greek
Armenian
B
RO
SATEM
GE
Azimuth angle (rad)
Lithuanian
Latvian
Khaskura
Nepali, Gujarati
Marathi, Hindi
Lahnda, Panjabi,
Kashmiri
Bengali
Baluchi
Farsi
Singhalese
Romany, Tajik
Pashto
Satem
C
II
BS
CENTUM
Zenith angles (rad)
Bulgarian, Slovenian
Serbo-Croatian
Czech, Slovak,
Russian
Polish, Byelorussian
Ukrainian
E



$$\frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \rightarrow \frac{\exp\left(-\frac{(x-\mu)^2}{2t}\right)}{\sqrt{2\pi t}}$$

Normal probability plots



Attesting historical events

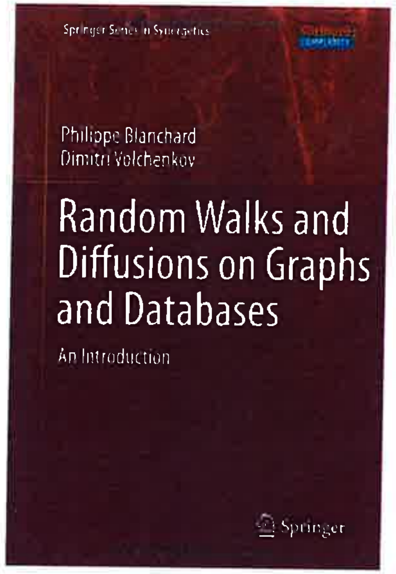
$$\frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \rightarrow \frac{\exp\left(-\frac{(x-\mu)^2}{2t}\right)}{\sqrt{2\pi t}}$$

Anchor events:

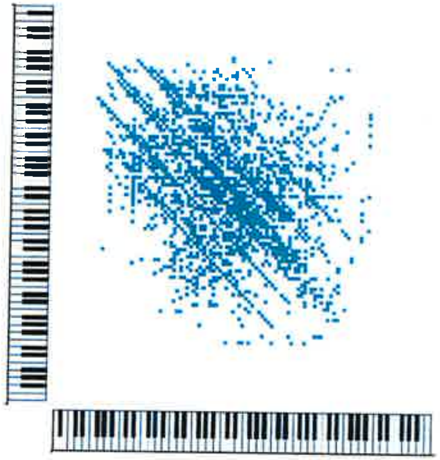
1. the last Celtic migration (to the Balkans and Asia Minor) (300 BC),
2. the division of the Roman Empire (500 AD),
3. the migration of German tribes to the Danube River (100 AD),
4. the establishment of the Avars Khaganate (590 AD) overspreading Slavic people who did the bulk of the fighting across Europe.

$$t/\sigma^2 = (1.367 \pm 0.002) \cdot 10^6$$

Can we hear first-passage times?

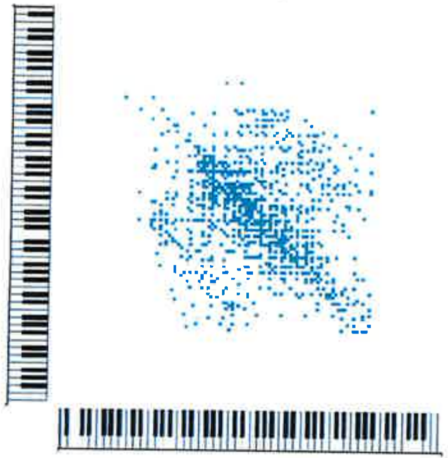


F. Liszt Consolation-No1

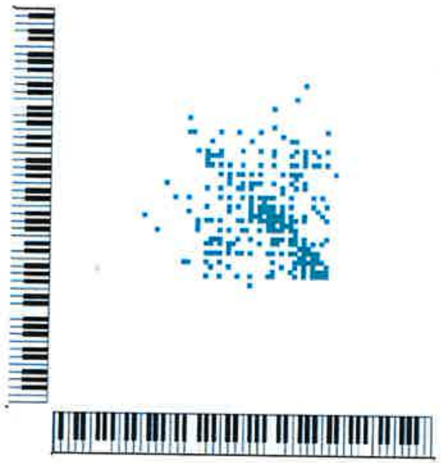


Bach_Prelude_BWV999

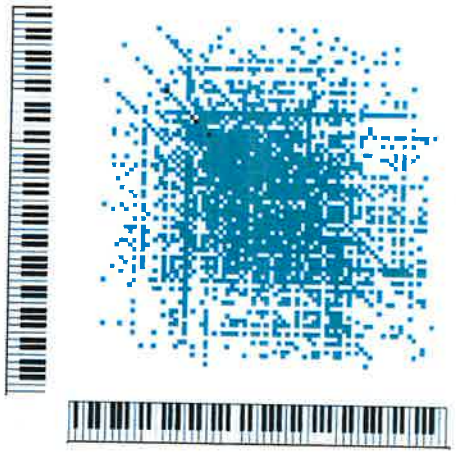
P. Tchaikovsky, Danse Napolitaine

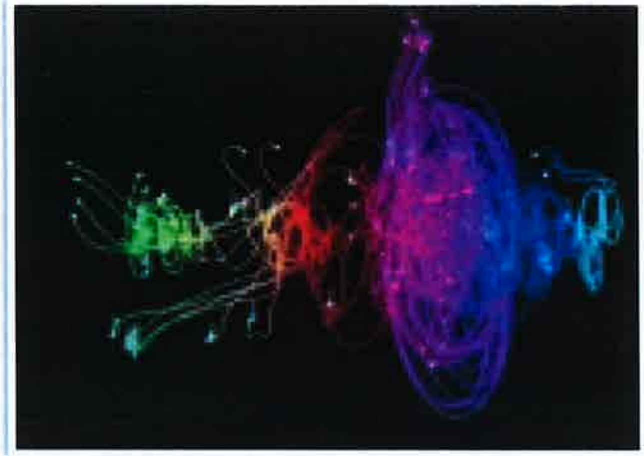


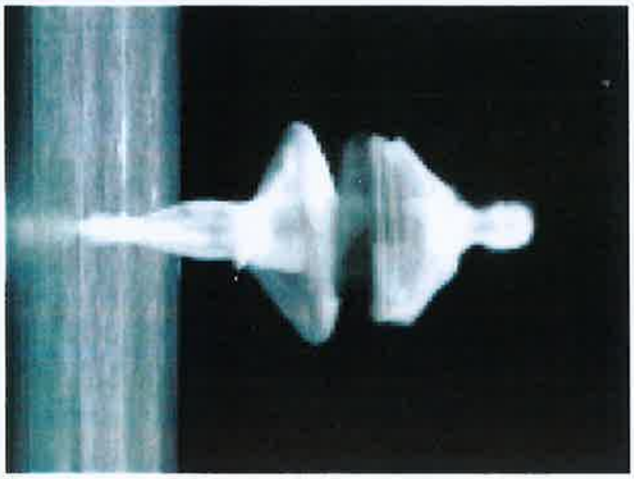
V.A. Mozart, Eine Kleine Nachtmusik



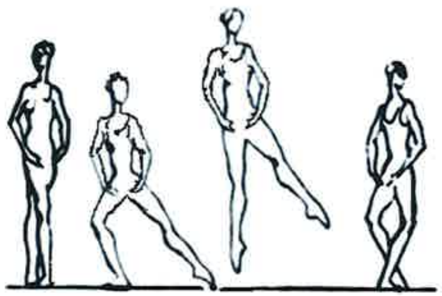
R. Wagner, Das Rheingold
(Entrance of the Gods)

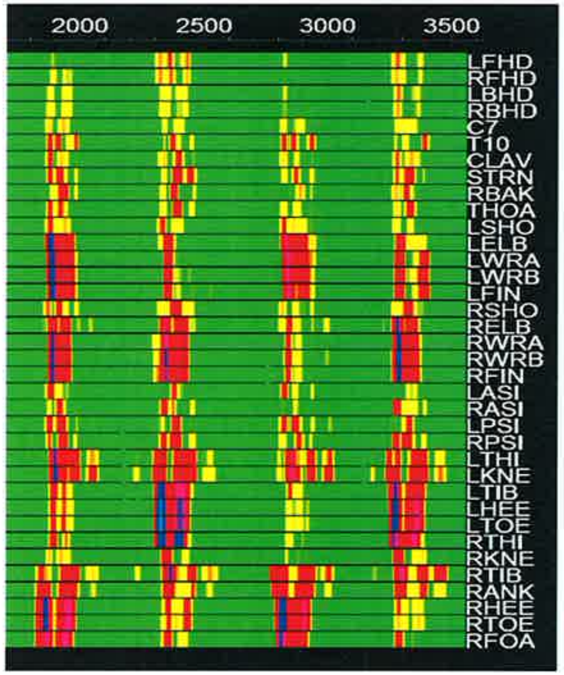






Assemblé





Sauté