Statistical Mechanics of Networks

TROISIEME CYCLE DE LA SUISSE ROMANDE

Guido Caldarelli

Together with

C.Caretta, M. Catanzaro, F. Colaiori, D. Garlaschelli, L. Pietronero, V. Servedio

INFM and Dipartimento di Fisica, Università "La Sapienza" Roma, Italy

L. Laura, S. Leonardi, S. Millozzi, A. Marchetti-Spaccamela

Dipartimento di Sistemistica e Elettronica, Università "La Sapienza" Roma Italy

P. De Los Rios[a], G. Bianconi[b], A. Capocci[b]

[a]Université de Lausanne, [b]Université de Fribourg, Switzerland

S. Battiston, A. Vespignani

Ecole Normale Supériéure and Université de Paris Sud, Paris France

•Contents

Part 1 *20-11-2003*

BASICS

A. Networks as complex structures
B. Fractals, Self-similarity
C. Self-organization
D. Evidence of scale-free networks
E. Basic of Graphs

Part 2 *27-11-2003*

REAL GRAPHS

A. Technological data: Internet, WWW
B. Social data: Finance and Board of Directors
C. Biological data: Proteins

Part 3 *4-12-2003*

REAL TREES

A. Food Webs
B. Geophysical data: the River Networks
C. Biological data: Taxonomy and Community Structures

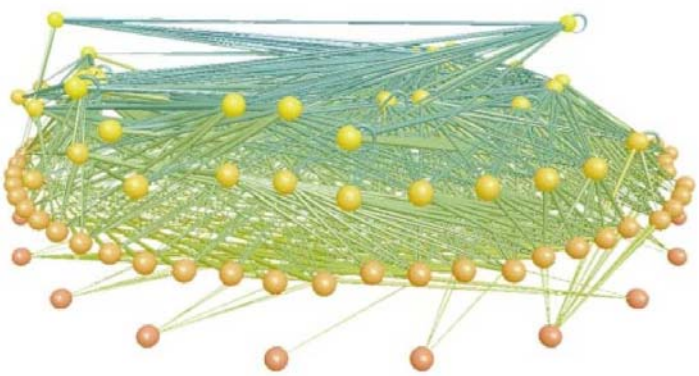
Part 4 *11-12-2003*

MODELS

A. Random Graphs (Erdös-Renyi)
B. Small world
C. Preferential attachment
D. Fitness models

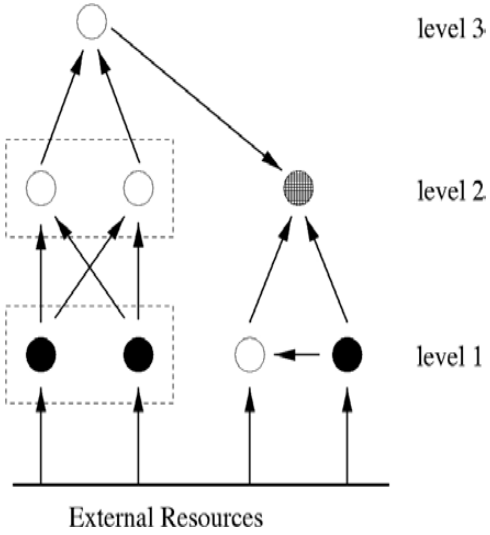
•3A Food Webs (1)

Species and their predation relationships form a very peculiar hierarchical graph.



Little Rock Lake (Wisconsin)

188 species



Top Species:

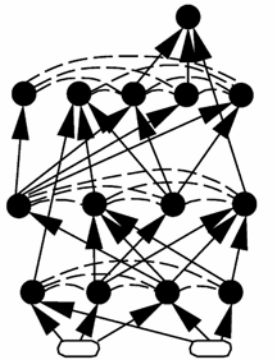
i.e. no predators

Intermediate Species:

i.e. both predators and prey

Basal Species:

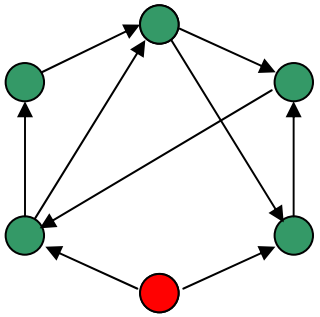
No prey



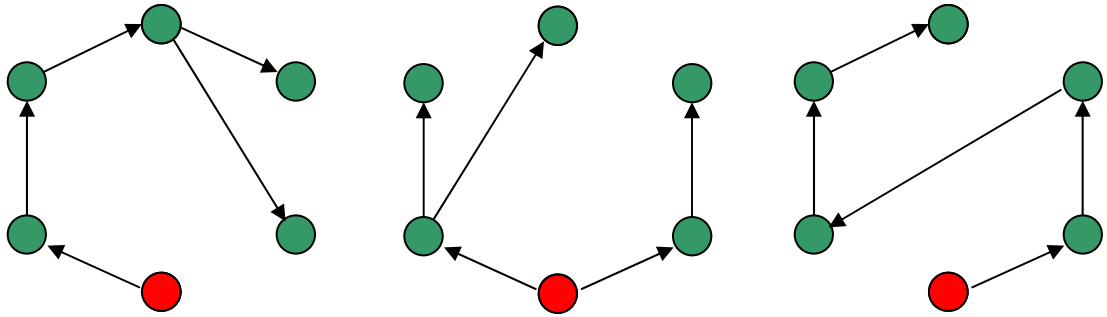
Pamlico Esthuary (North Carolina)

14 species

•3A Food Webs (2)

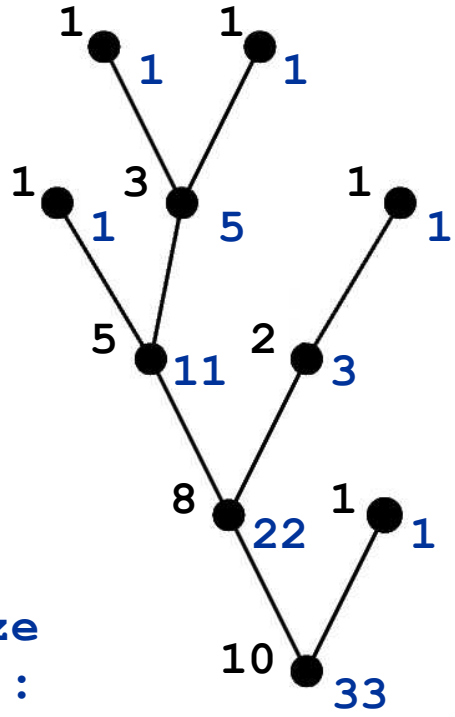


A *spanning tree* of a connected directed graph is any of its connected directed subtrees with the same number of vertices.



In general, the same graph can have more spanning trees with different topologies.

•3A Tree Topology (3)



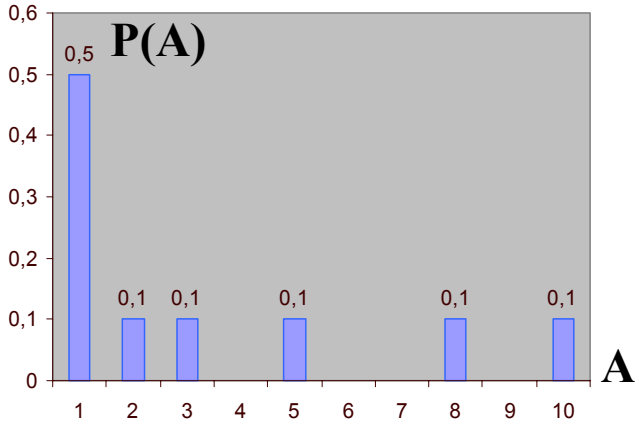
Out-component size:

$$A_X = \sum_{Y \in nn(X)} w_{XY} A_Y + 1$$

Sum of the sizes:

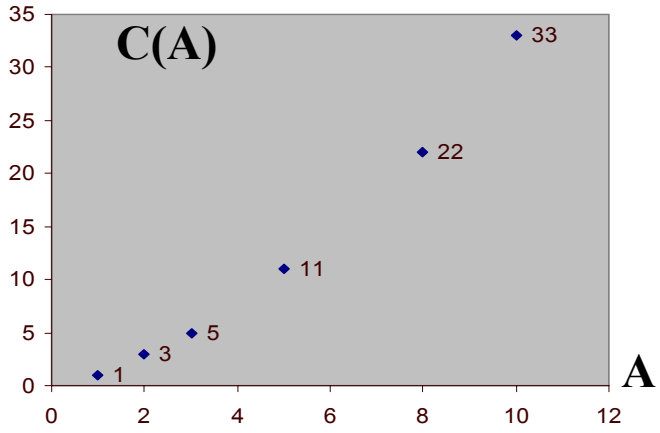
$$C_X = \sum_{Y \in \gamma(X)} A_Y$$

Out-component size distribution P(A):



Allometric relations:

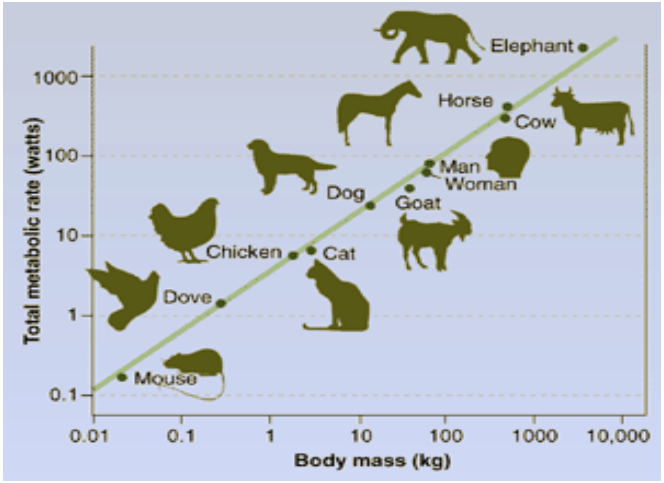
$$C_X = C_X(A_X) \Rightarrow C = C(A)$$



• 3A Allometric Relations (4)

A_0: metabolic rate B

C_0: blood volume $\sim M$



Kleiber's Law:

$$B(M) \propto M^{3/4}$$

$$C(A) \propto A^{\eta} \qquad \eta = \frac{4}{3}$$

General Case (tree-like transportation system embedded in a D-dimensional metric space):

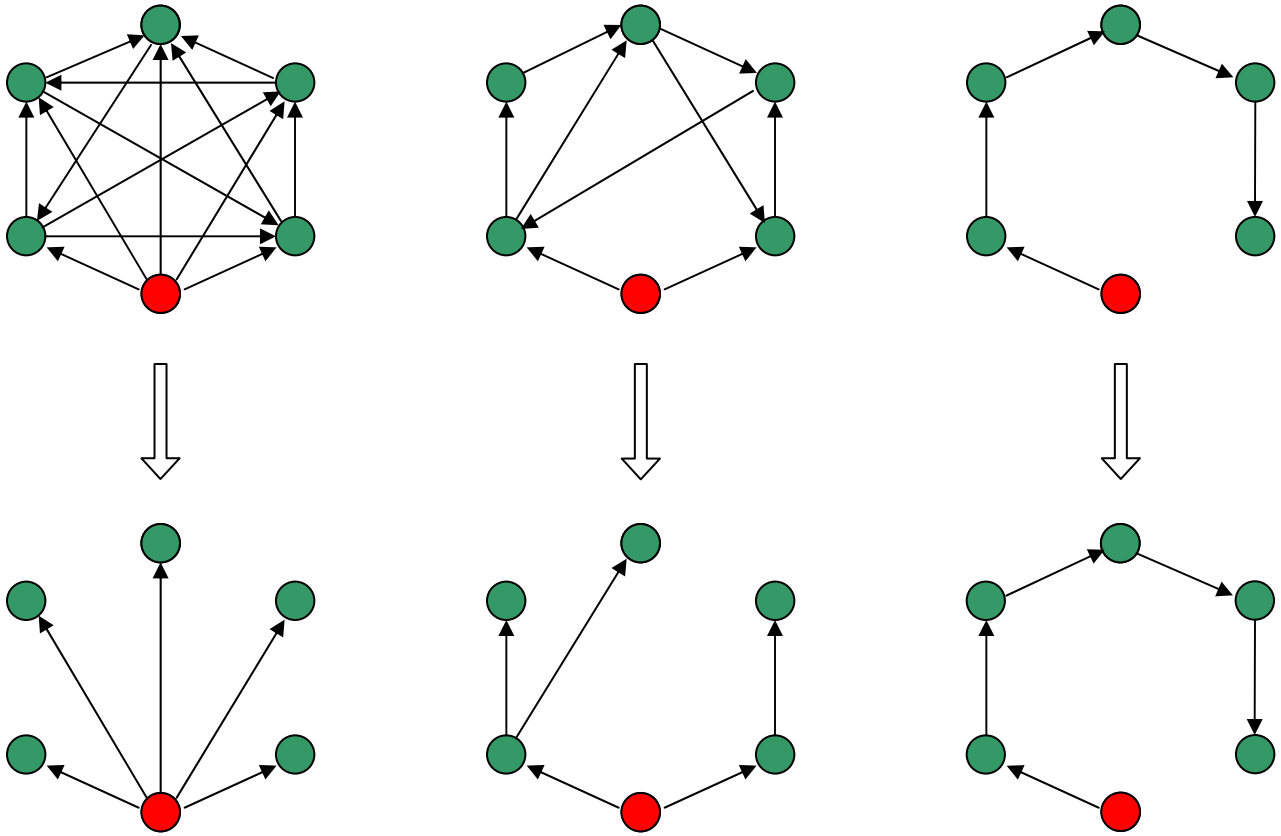
the most efficient scaling is $C(A) \propto A^{\eta} \qquad \eta = \dfrac{D+1}{D}$

West, G. B., Brown, J. H. & Enquist, B. J. *Science* **284**, 1677-1679 (1999)

Banavar, J. R., Maritan, A. & Rinaldo, A. *Nature* **399**, 130-132 (1999). |

• 3A Food Webs (5)

$\delta >> 1$ | $0 < \delta < 1$ | $\delta \to 0$



$C(A) \propto A$

efficient

$P(A) = \delta_{A1}$

stable

$C(A) \propto A^{\eta} \quad 1 < \eta < 2$

$P(A) \propto A^{-\tau} \quad 0 < \tau < \infty$

$C(A) \propto A^{2}$

inefficient

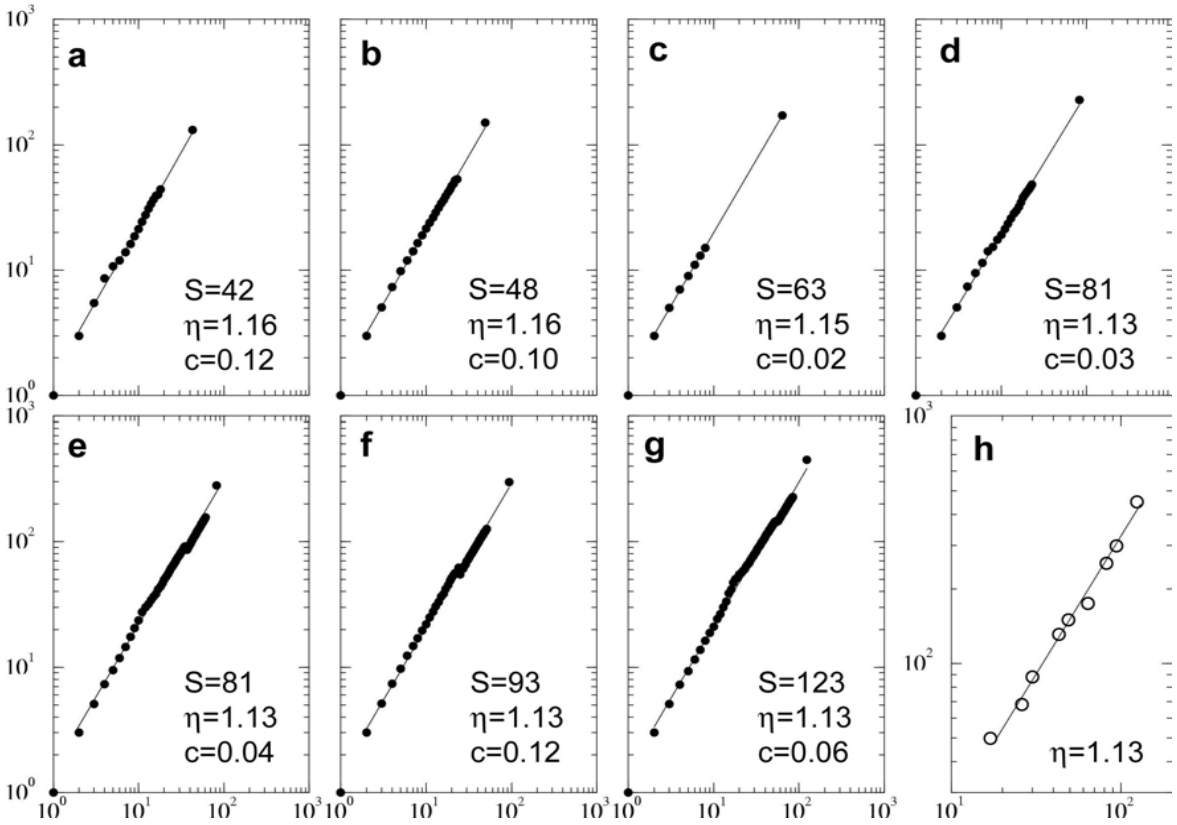
$P(A) = cost$

unstable

• 3A Food Webs (6)

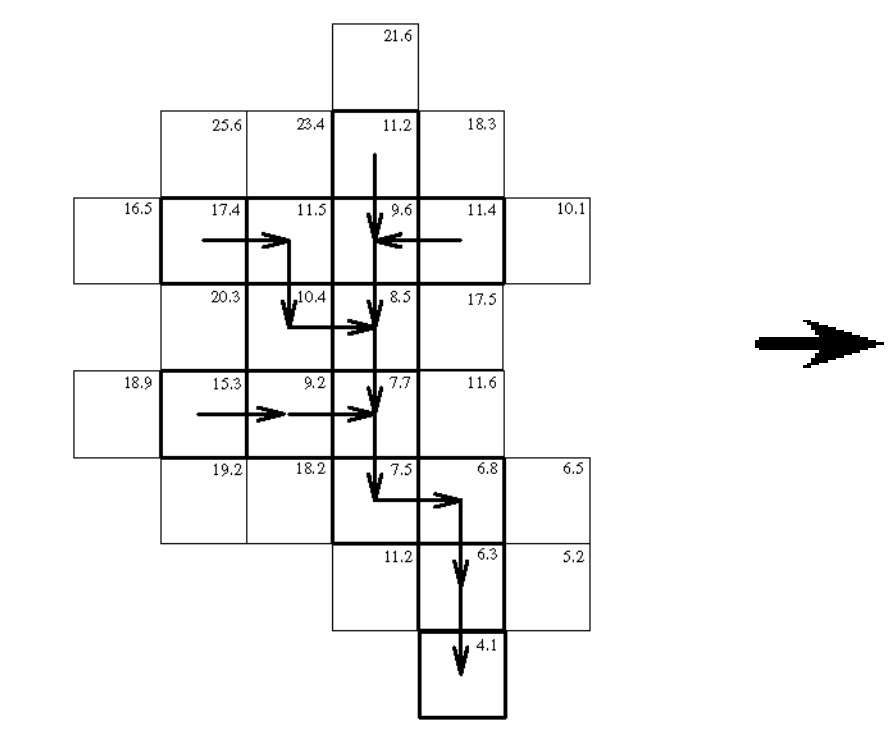
Results in this field are that the spanning tree of the Food web show Some sort of optimization in resources transfer.

For this case the most efficient tree should behave like $C(A) \propto A$



(D.Garlaschelli, G. Caldarelli, L. Pietronero *Nature* **423** *165* (2003))

•3B River Network Theory (1)

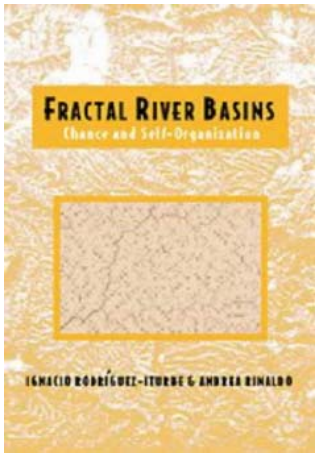


Law of steepest descent produces the complex Network structure of rivers drainage basins.

•3B River Network Theory (2)

River Networks study the interplay between Fluvial erosion and Landscape Evolution

This interplay produces "**UNIVERSAL**" non trivial properties.



The first quantitative measure of that universality was introduced by Hack who studied the shape of the basins.
In particular the Drainage Basins **for any River on Earth** display similar Fractal Properties

We investigate the statistical properties of such basins and map these Properties with the above interplay.
A possible application presented is trying to infer the presence of erosion From drainage basins of other planet (Mars).

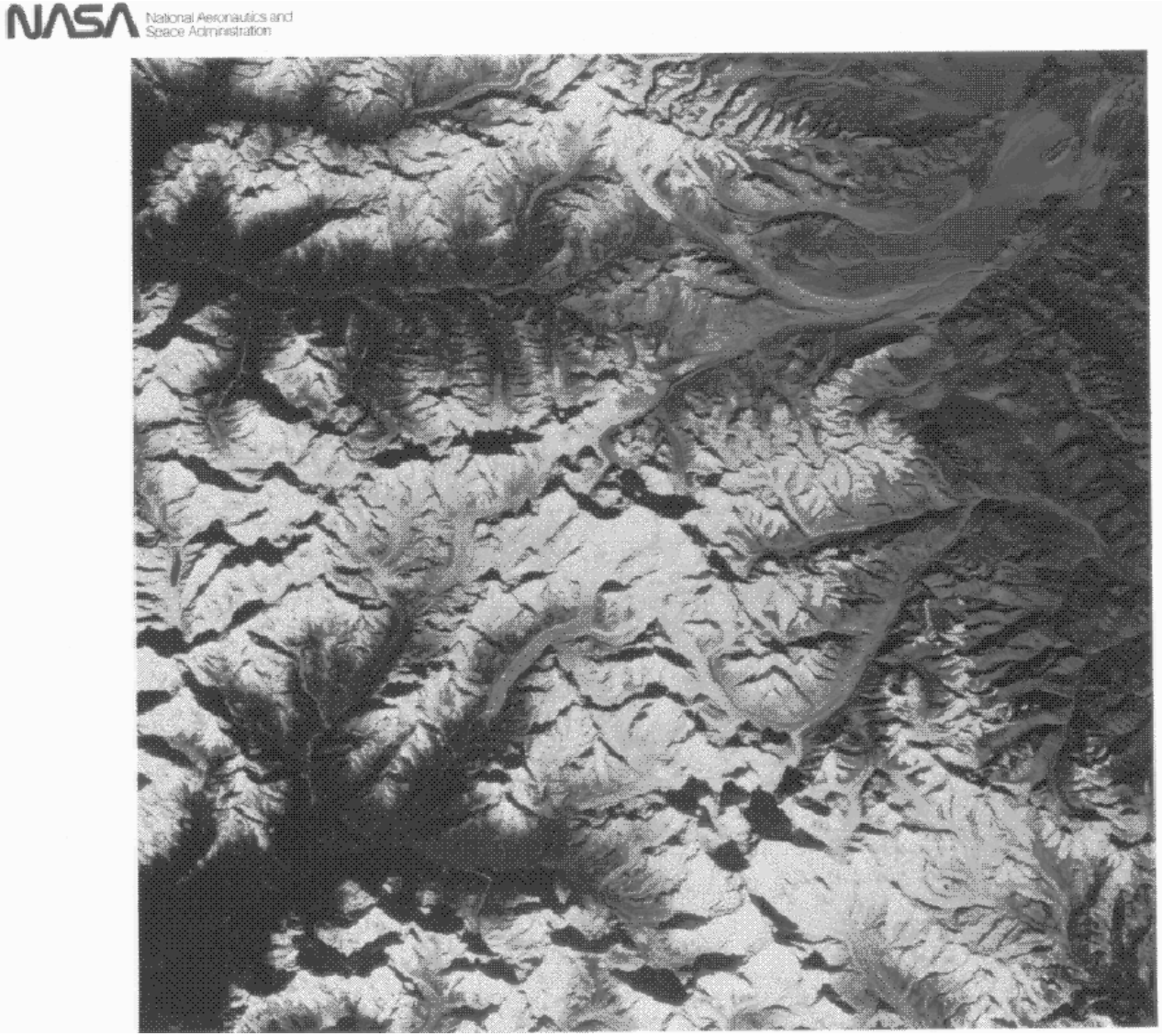
•3B River Network Theory (3)

Some phenomenological facts are at the basis of the River Network Theory

1. In river basins one can consider precipitation of water as nearly constant. Therefore the mass of water collected in the outlet is proportional to the **area of the basin A_i.**
2. Water follows the **steepest descent path (Δh_i)**
3. The erosion of water modifies the landscape and can cause terrain instability. Empirically, the stable landscape are those for which

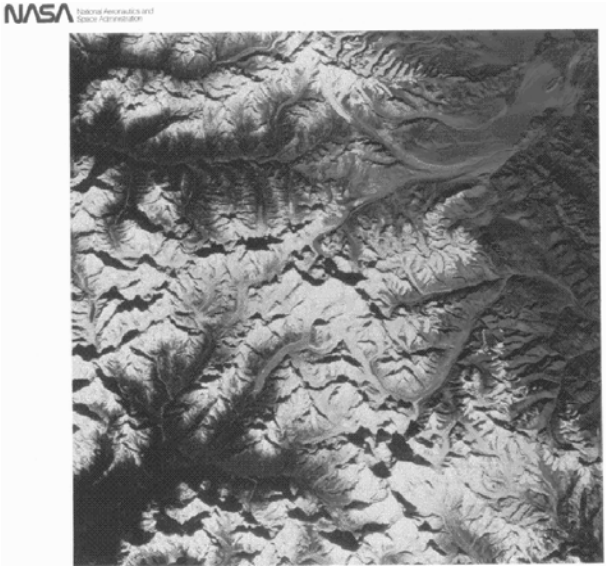
$$\Delta h_i \, A_i^{0.5} \leq \text{constant}$$

•3B Shape of Landscapes (4)



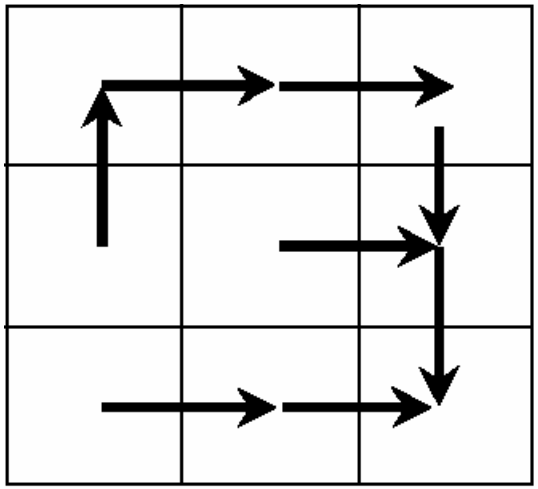
Satellite image of
Himalaya
Courtesy of NASA

•3B Digital Elevation Model (5)



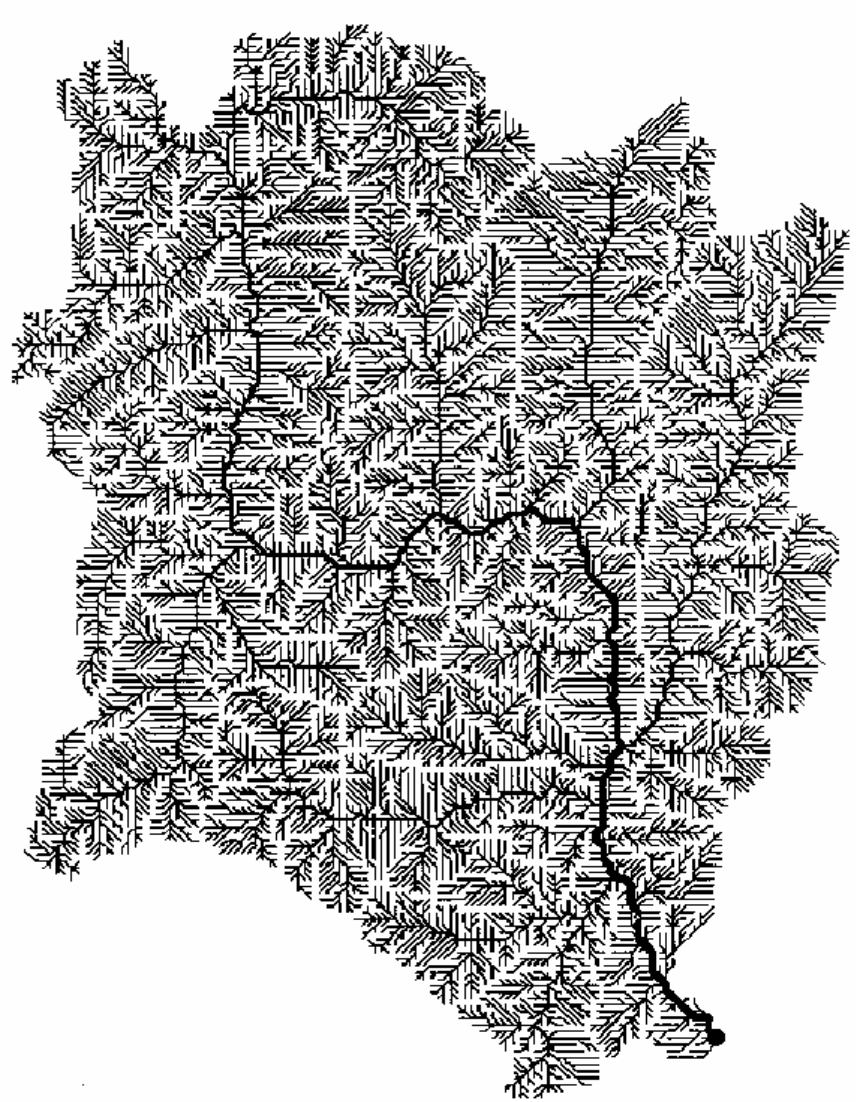
From satellite images one gets Digital Elevation Models (DEM)
From DEM a spanning tree is computed (via steepest descent)
From spanning tree, the number of points uphill is computed

156.4	132.4	111.4
170.8	161.3	108.2
182.4	154.5	106.0



2	3	4
1	1	6
1	2	9

•3B Real Drainage Basin (6)

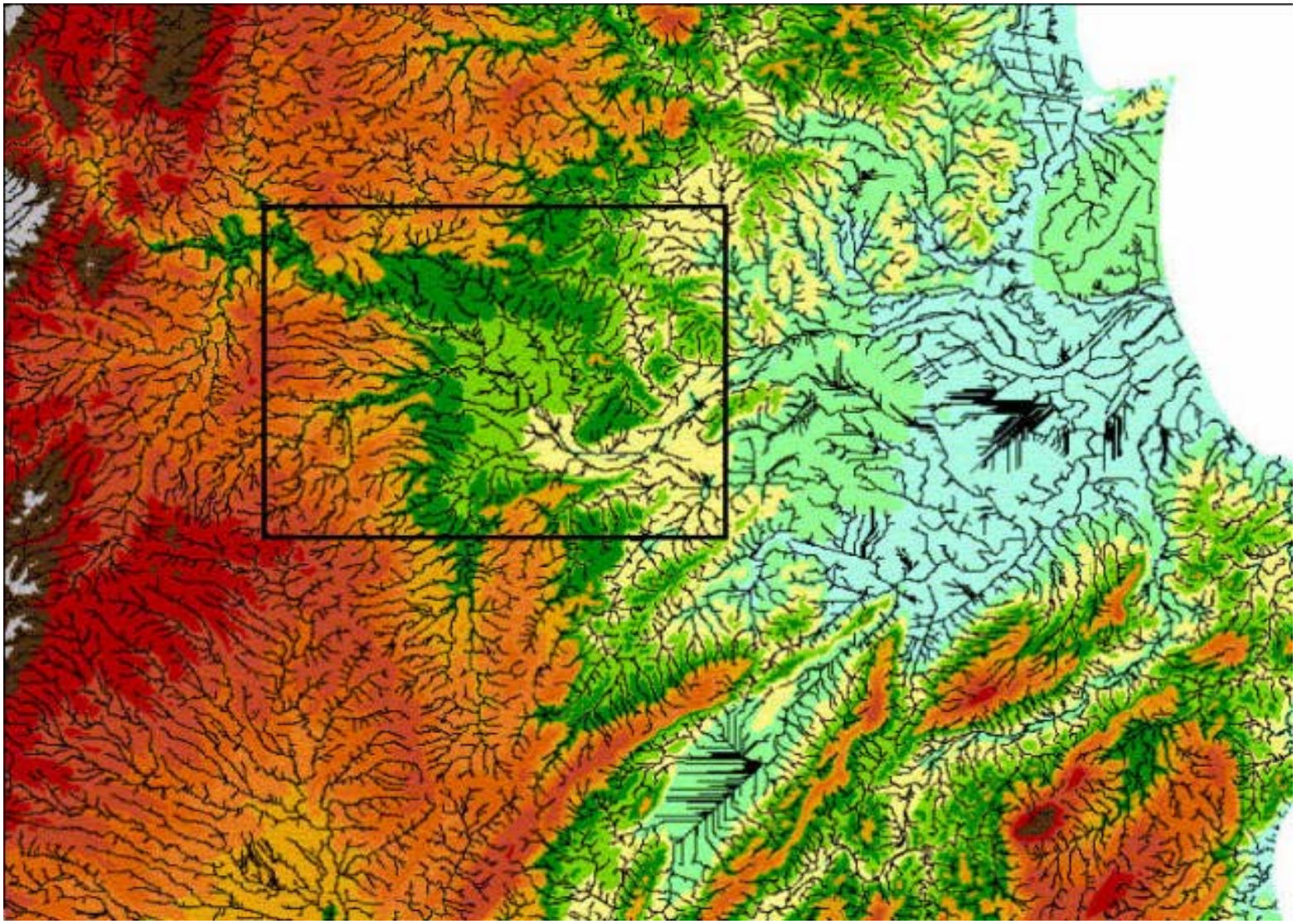


Drainage Basin of Fella river tributary of Tagliamento,

Northern Italy

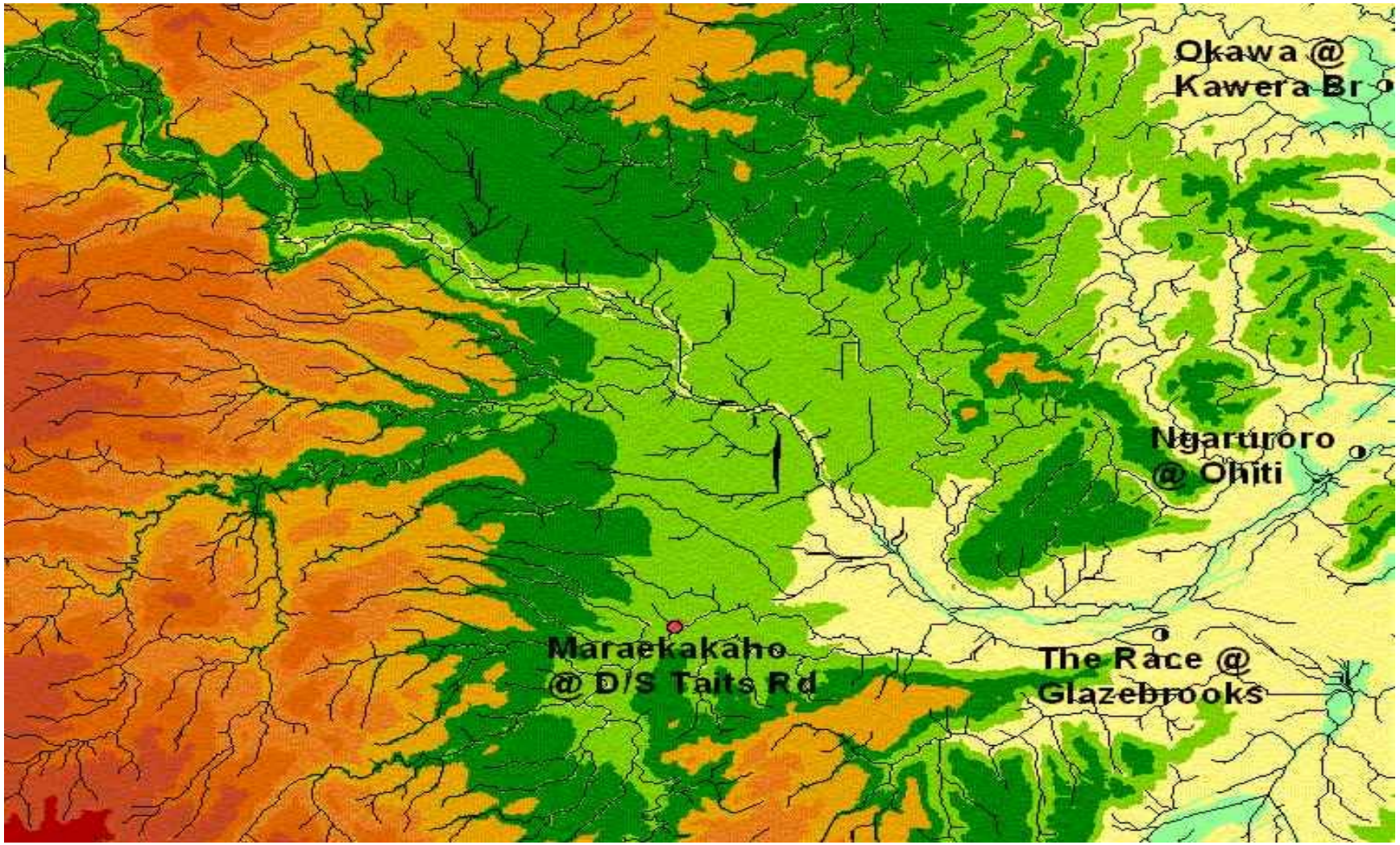
•3B Real Drainage Basin (7)

Basin of Ngaruroro River, New Zealand



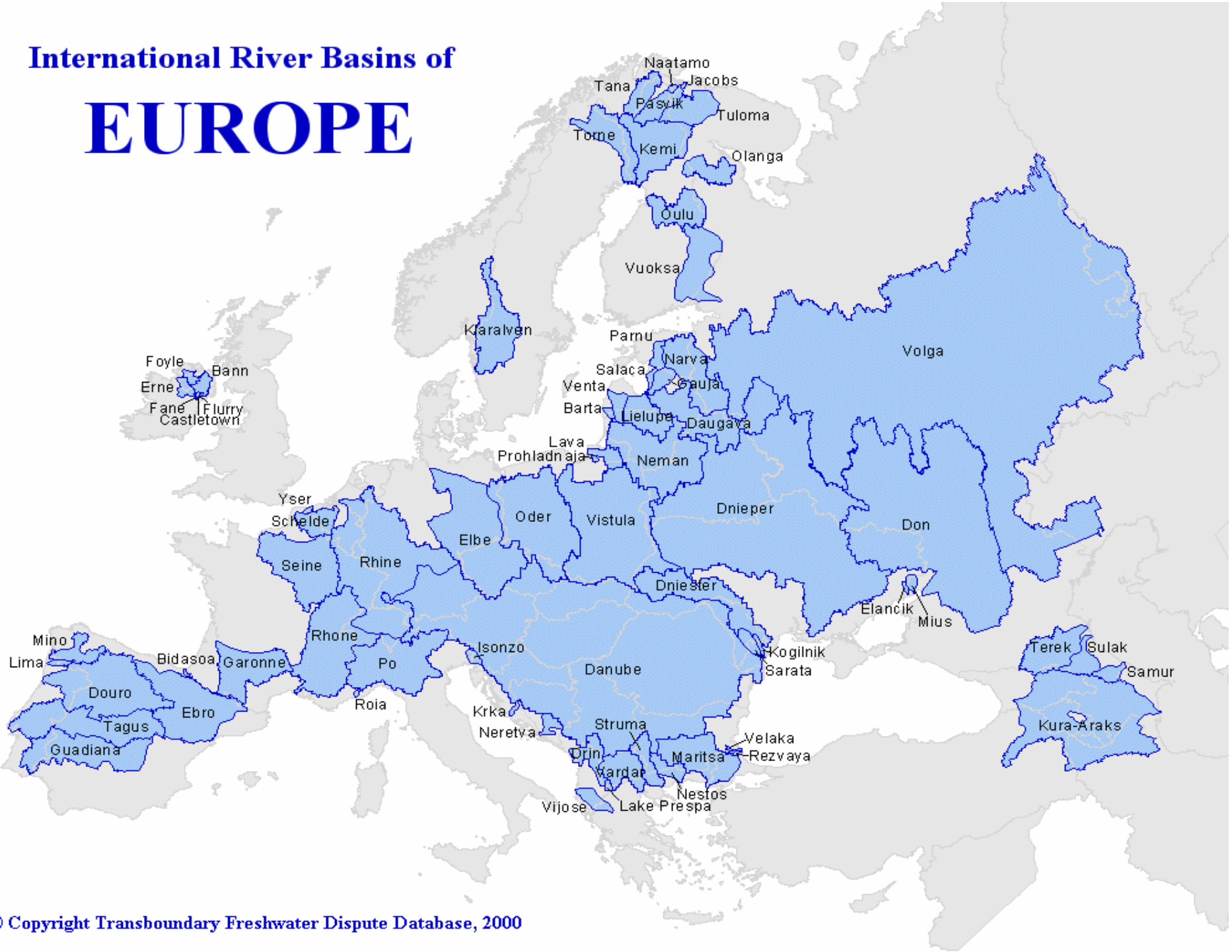
•3B Real Drainage Basin (8)

Larger view of the previous subbasin



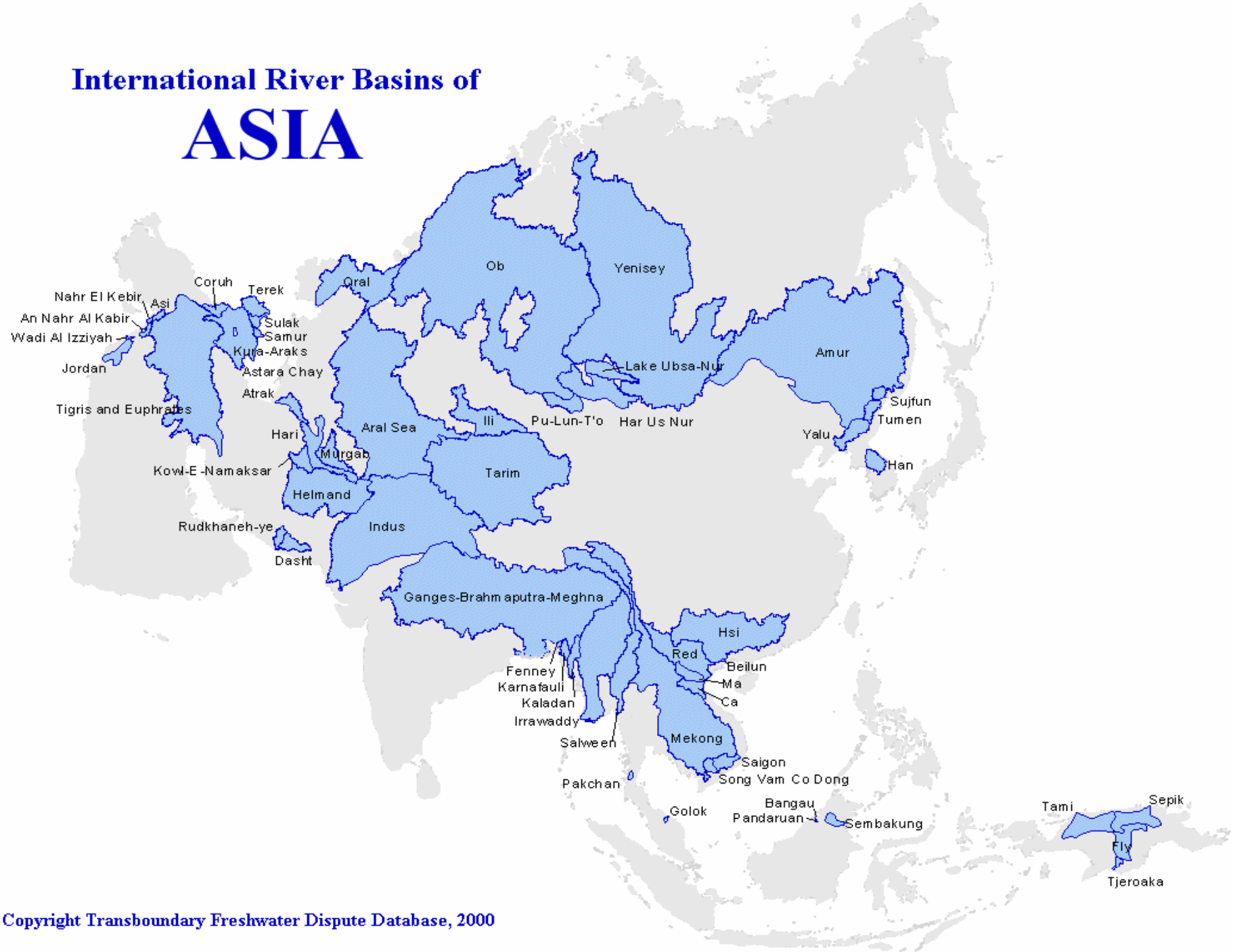
International River Basins of

EUROPE




Medjerda
Oued Bon Naima
Tafna
Daoura
Guir
Dra
Atui
Senegal
Gambia
Geba
Corubal
Great Scarcies
Little Scarcies
Moa
Mana-Morro
Loffa
St. Paul
St. John
Cestos
Cavally
Komoe
Sassandra
Bia
Tano
Mono
Volta
Oueme
Niger
Lake Chad
Nile
Baraka
Gash
Awash
Tagipi Swamp
Juba-Shibeli
Lake Turkana
Lake Natron
Umba
Cross
Akpa Yaf
Ntem
Benito
Utamboni
Mbe
Ogooue
Congo
Nyanga
Chiloango
Kunene
Zambezi
Ruvuma
Etosha
Okavango
Sabi
Buzi
Limpopo
Incomati
Umbeluzi
Maputo
Orange
International River Basins of
AFRICA
© Copyright Transboundary Freshwater Dispute Database, 2000

International River Basins of
ASIA



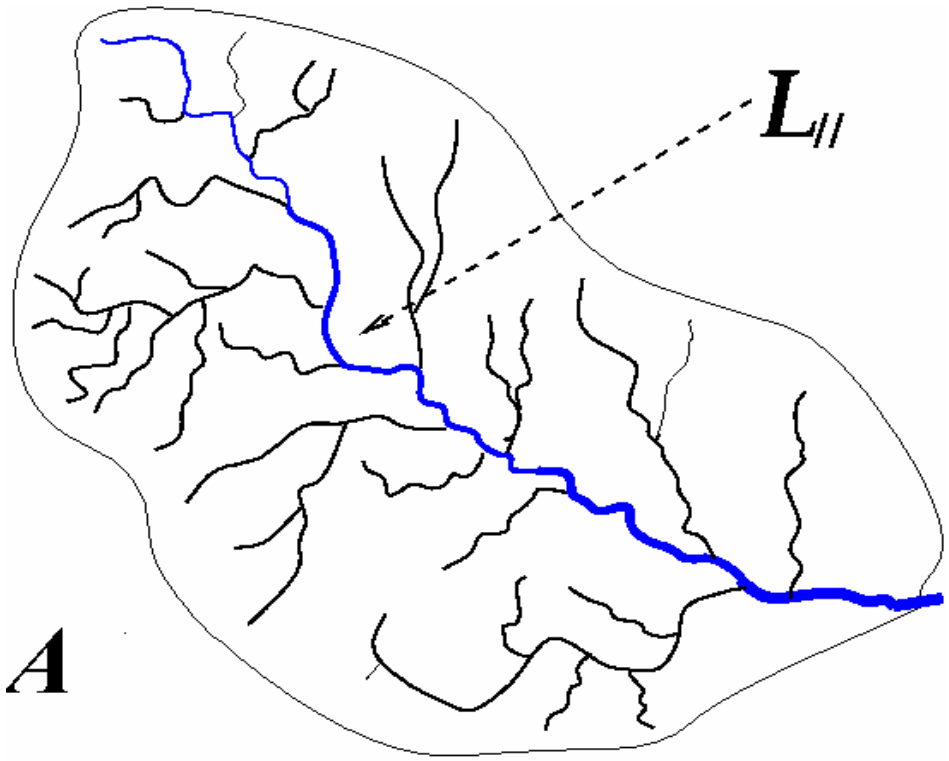
International River Basins of

SOUTH AMERICA

International River Basins of

NORTH AMERICA

•3B Statistical Measures: Hack's law (9)



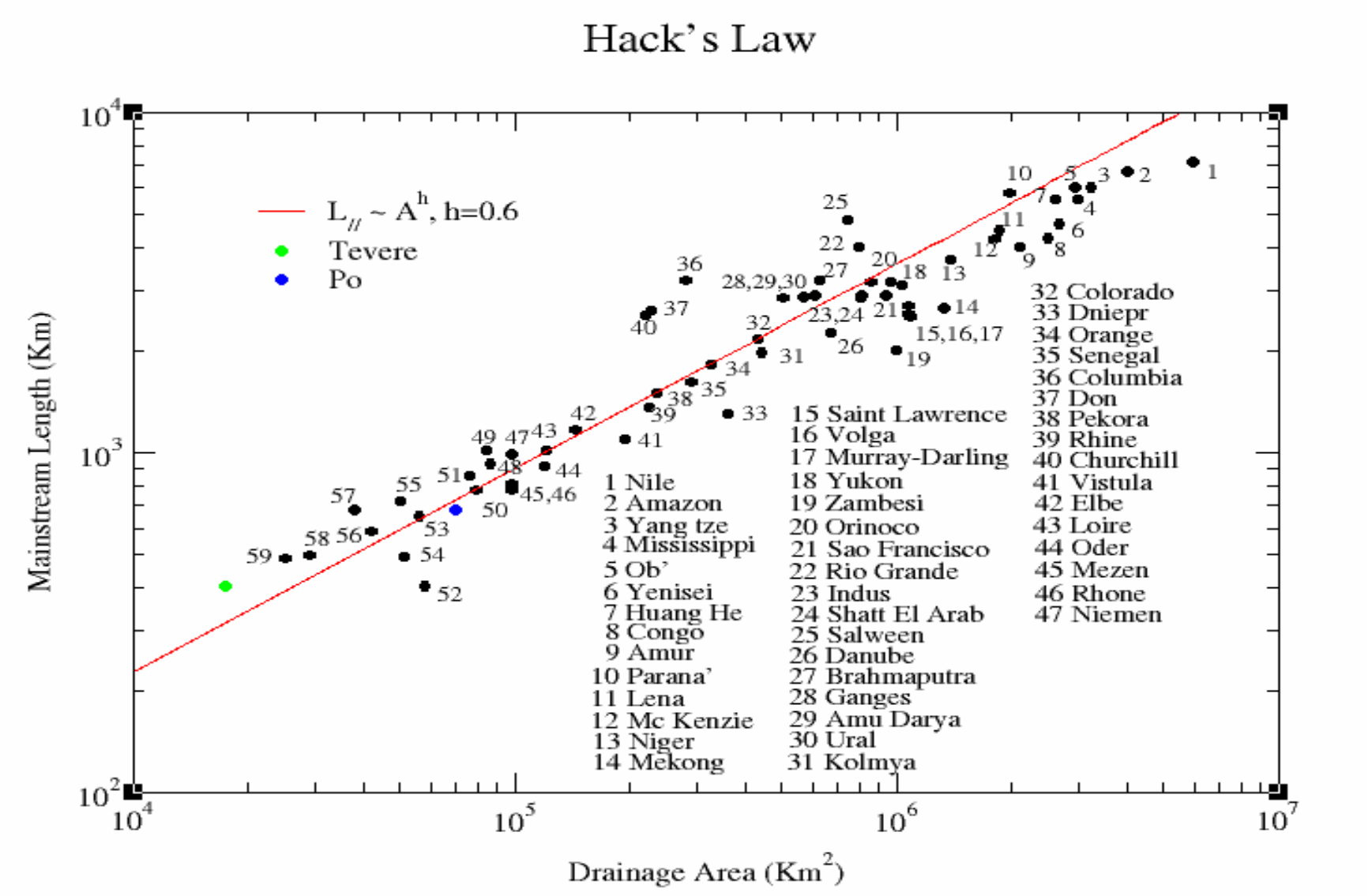
$$L_{//} \sim A^h, \; h \sim 0.6$$

This self-affinity results in power-law distributions of

-Number of points uphill

-Stream lengths

•3B Statistical Measures: Hack's law (10)



•3B Statistical Measures (11)

Different frequency distributions can be computed

•Frequency distribution **P(n)** to have **n** points uphill

For real rivers $\mathbf{P(n) \sim n^{-\tau} f(n/L^{\phi})}$ $\tau=1.43(1)$

•Frequency distribution **Π(l)** to have upstreams of length **l**

For real rivers $\mathbf{\Pi(l) \sim l^{-\chi} g(l/L)}$ $\chi=1.7(1)$

ALL THESE QUANTITIES ARE RELATED TO h!

One can describe networks by considering the P(n) only

•3B Scaling relations (12)

For self-affine river networks

$$\left.\begin{array}{l} A^h \sim L_{//} \quad (h>0.5) \\ L_o \sim L^H_{//} \quad (H<1) \end{array}\right\} \quad h = \frac{1}{1+H}$$

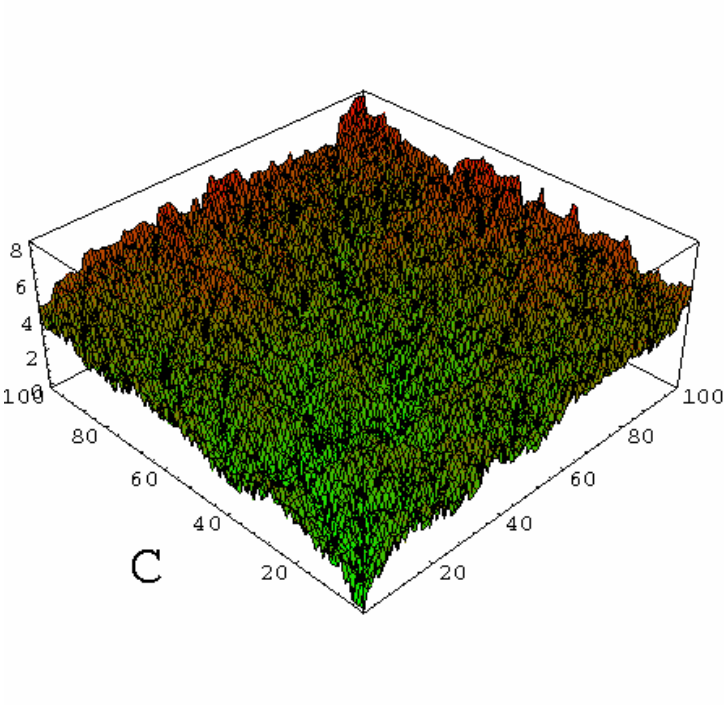
$$\tau = \frac{1+2H}{1+H} \qquad \gamma = \frac{1}{1+H} \qquad \phi = \frac{1}{1+H}$$

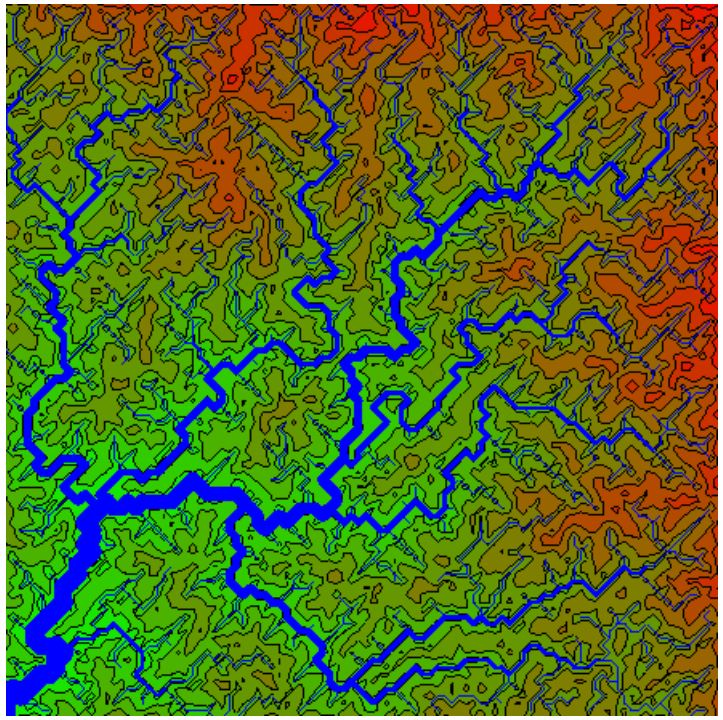
In particular, H ≅ 0.72, h ≅ 0.58, τ ≅ 1.42, γ ≅ 0.58, ϕ ≅ 0.58

A.Maritan et al., *Physical Review E* **53**, 1510 (1996)

•3B Computer Simulations (13)

Computer simulations of a Self-Organised Critical model reproduce the data

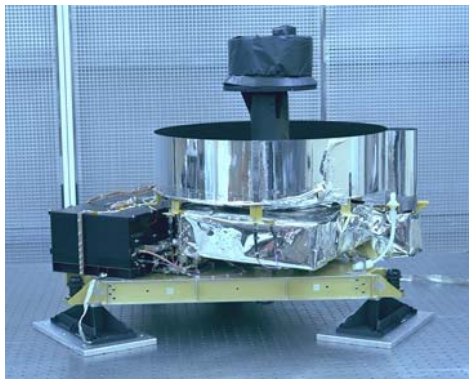


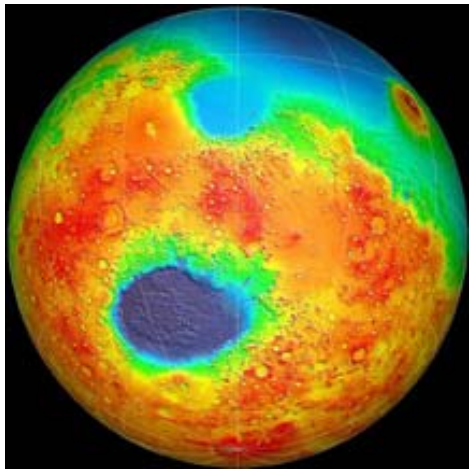


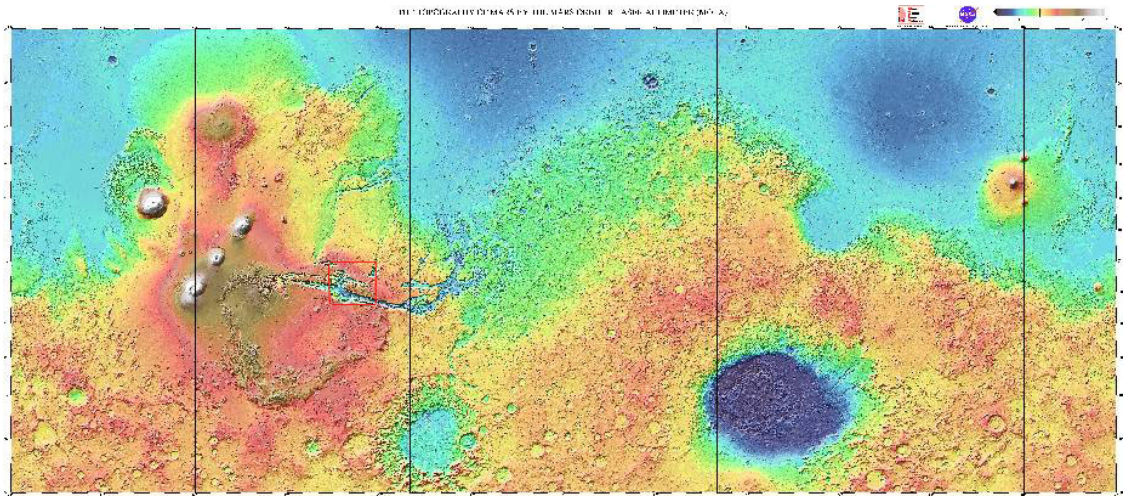
Can one applies these results to other cases?

•3B Martian Landscape (14)

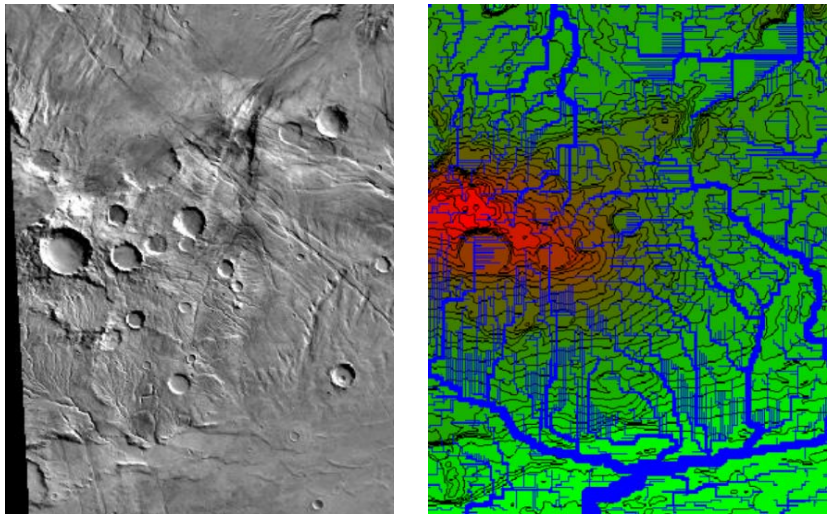
Data on Mars topography were collected through the Mars Orbiter Laser Altimeter (MOLA)





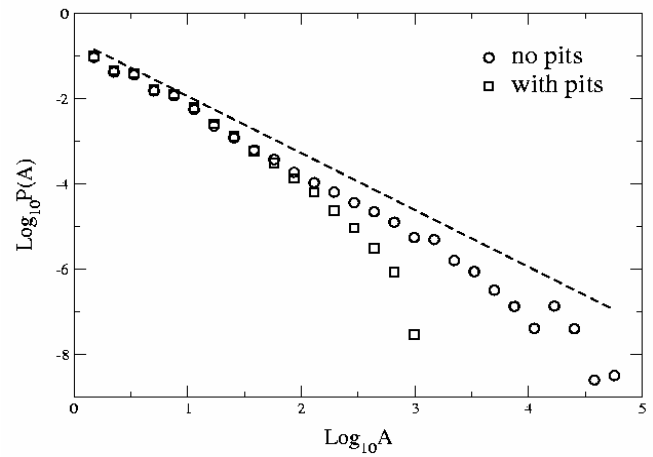


•3B Martian Landscape (15)



Computer Analysis of Warego Valley DEM from MOLA

Results are that we can distinguish regions whose DEM networks have properties similar to River Networks on Earth.



Dotted line correspond to

$P(A) \propto A^{-1.42}$

•3C Taxonomical Trees (1)

Ecosystem = **Set of all living organisms and environmental properties of a restricted geographic area**

→ **we focus our attention on plants**

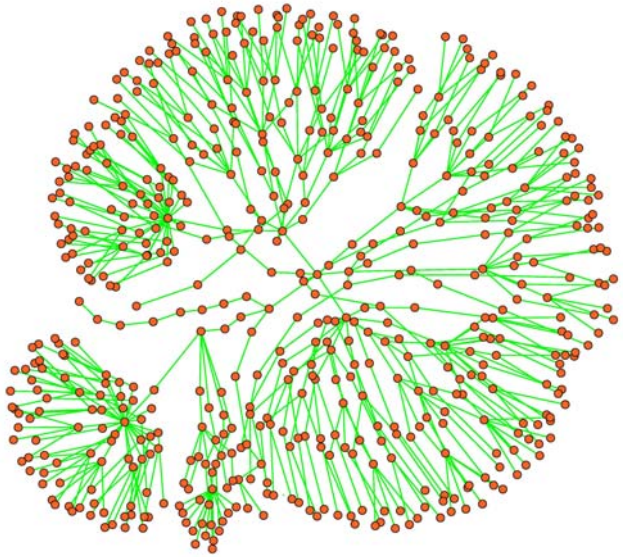
→ **in order to obtain a good universality of the results we have chosen a great variety of climatic environments**

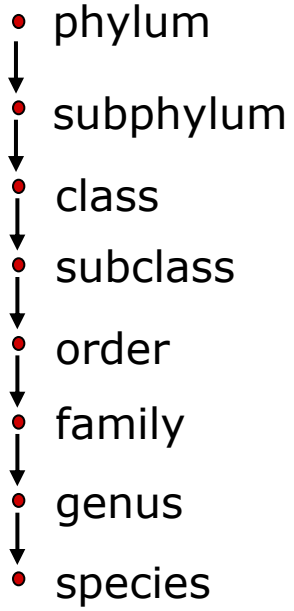
•3C Taxonomical Trees (2)

Phylogenetic Tree = **hierarchical structure organized on different levels, called *taxonomic levels*, representing:**

- **classification and identification of different plants**
- **history of the evolution of different species**

A phylogenetic tree already has the topological structure of a tree graph





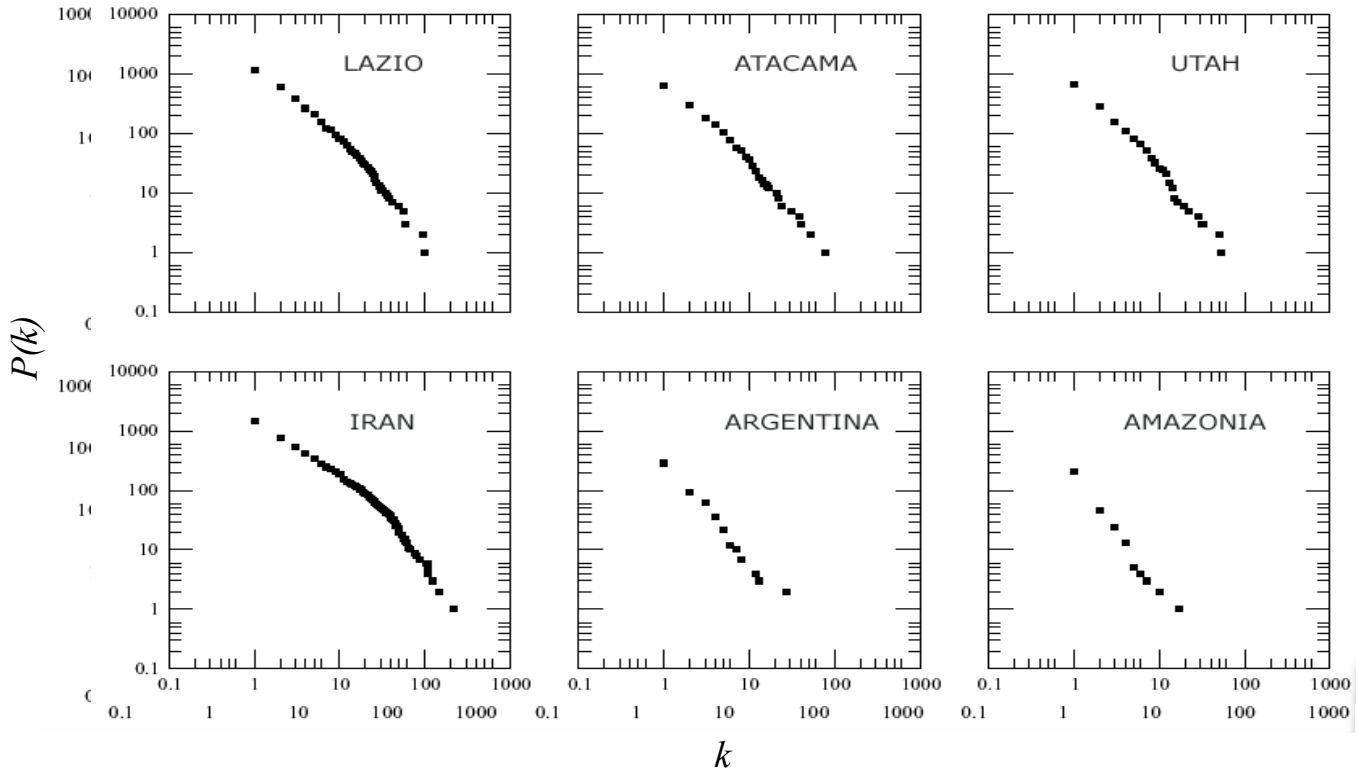
- **each node in the graph represents a different taxa (specie, genus, family, and so on). All nodes are organized on levels representing the taxonomic one**

- **all link are up-down directed and each one represents the belonging of a taxon to the relative upper level taxon**

→ ***Connected graph without loops or double-linked nodes***

•3C Scale-free trees (3)

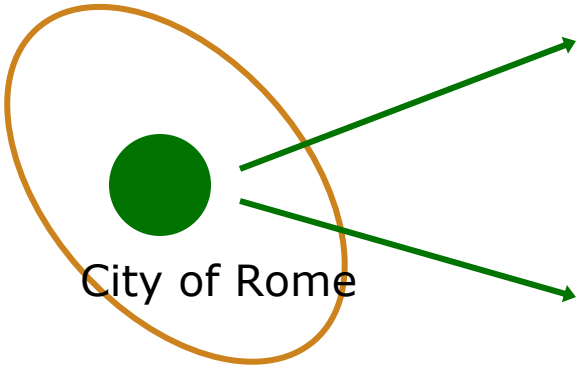
Degree distribution:

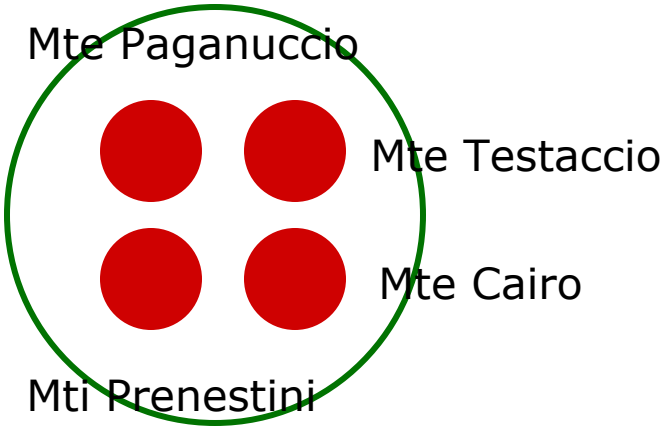


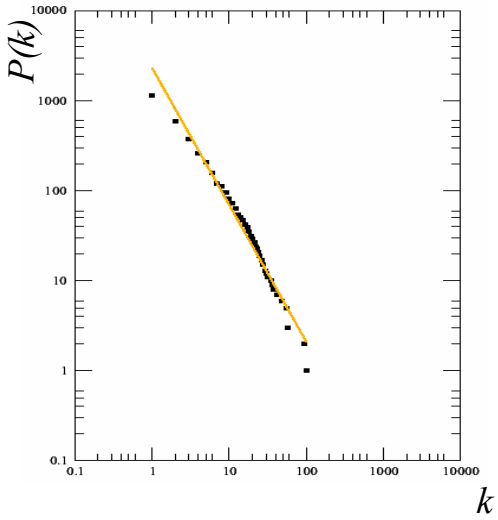
$$P(k) \propto k^{-\gamma} \qquad \gamma \sim 2.5 \pm 0.2$$

The best results for the exponent value are given by ecosystems with greater number of species. For smaller networks its value can increase reaching $\gamma = 2.8$ - 2.9**.**

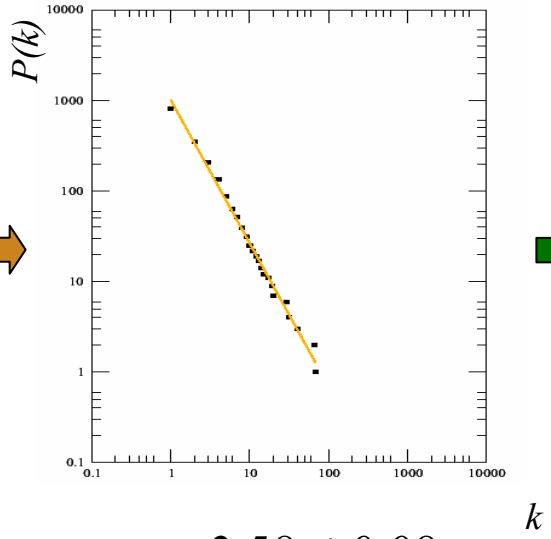
•3C Geographical Species subset (4)



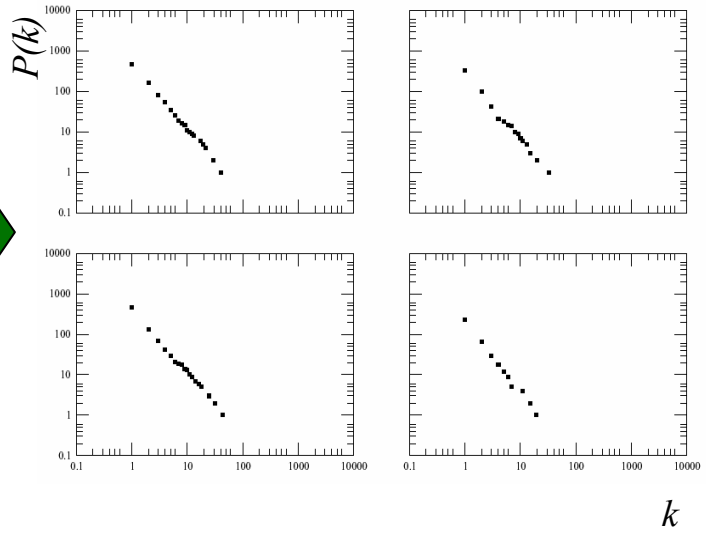




$\gamma = 2.52 \pm 0.08$



$\gamma = 2.58 \pm 0.08$

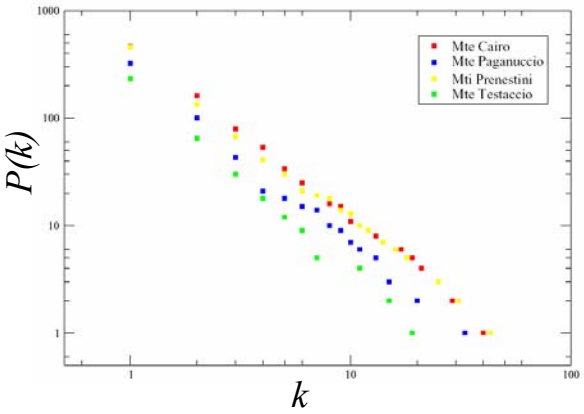


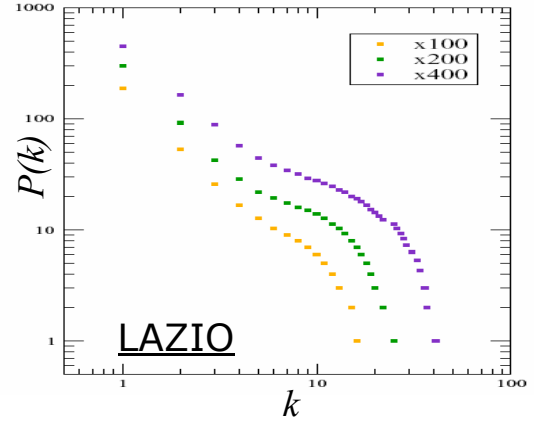
$2.6 \leq \gamma \leq 2.8$

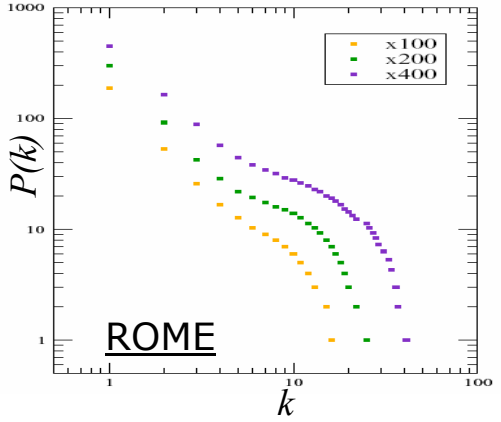
•3C Random subsets (5)

In spite of some slight difference in the exponent value, a subset which represents on its own a geographical unit of living organisms still show a power-law in the connectivity distribution.

random extraction of 100, 200 and 400 species between those belonging to the big ecosystems and reconstruction of the phylogenetic tree

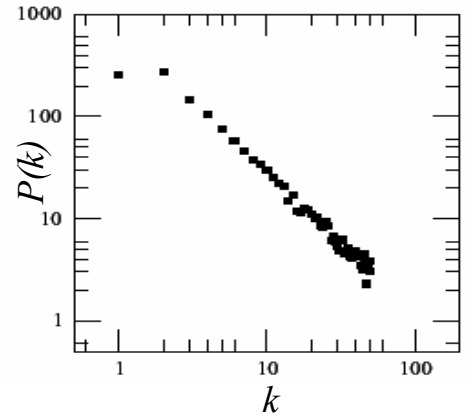


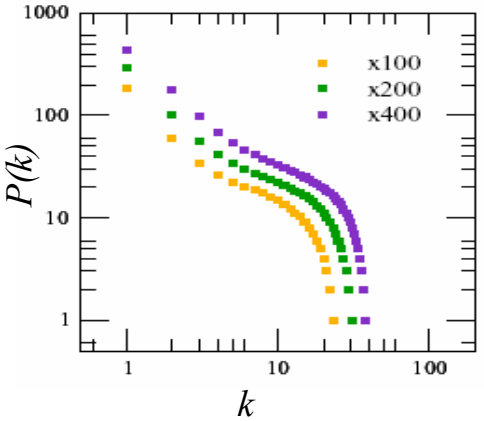




- **Simulation:**

$$P(k) = k^{-2.6}$$

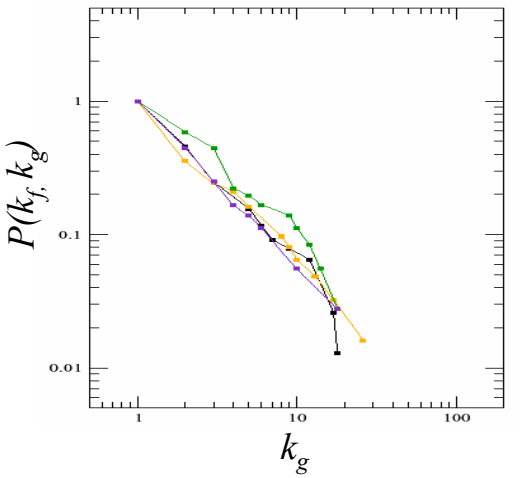




•3C Memory (6)

Particular rule to put a species in a genus, a genus in a family....? ***NO!***

$P(k_f, k_g)$ that a **genus** with degree k_g belongs to a **family** with degree k_f



- $k_f=1$
- $k_f=2$
- $k_f=3$
- $k_f=4$

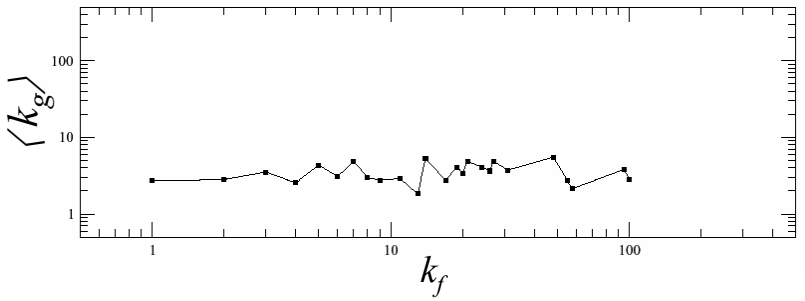
$$P(k_f,k_g) \propto k_g^{-\gamma}$$

fixed

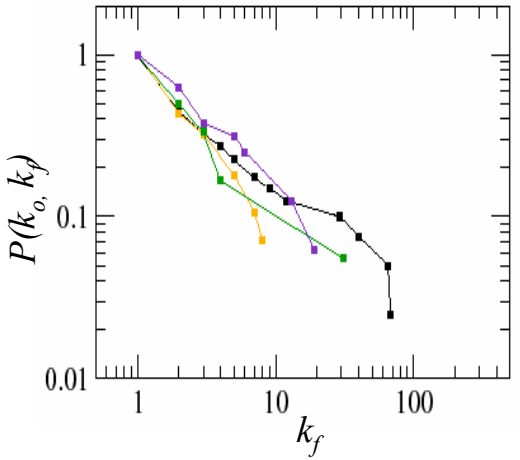
$$\gamma \sim 2.2 \pm 0.2$$

$$\langle k_g \rangle = \sum_g k_g P(k_f,k_g)$$

fixed



$P(k_o,k_f)$ that a **family** with degree k_f belongs to an **order** with degree k_o



- $k_o=1$
- $k_o=2$
- $k_o=3$
- $k_o=4$

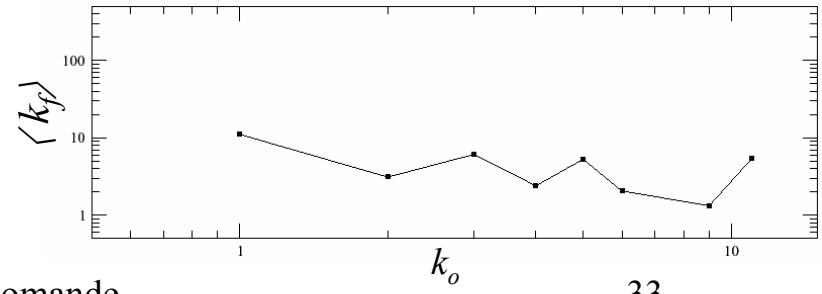
$$P(k_o,k_f) \propto k_f^{-\gamma}$$

fixed

$$\gamma \sim 1.8 \pm 0.2$$

$$\langle k_f \rangle = \sum_f k_f P(k_o,k_f)$$

fixed



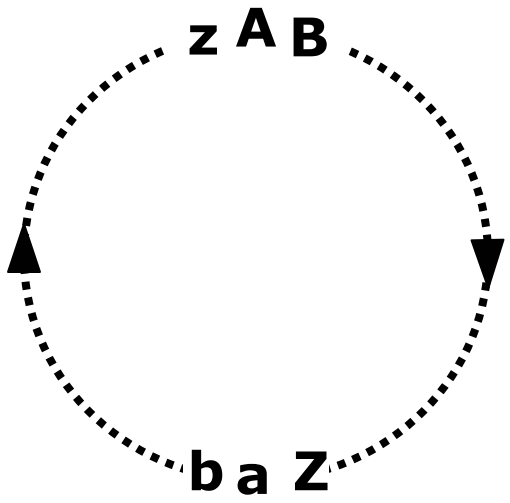
•3C A simple model (7)

1) create N species to build up an ecosystem

2) Group the different species in genus, the genus in families, then families in orders and so on realizing a **phylogenetic tree**

- Each species is represented by a string with **40** characters representing **40** properties which identify the single species (genes);
- Each character is chosen between **94** possibilities: all the characters and symbols that in the ASCII code are associated to numbers from **33** to **126**:

P	g	H	C	)	%	o	r	?	L	8	e	s	/	C	c	W	&	I	y	4	!	t	G	j	4	2	£	)	k	,	!	d	q	2	=	m	:	f	V



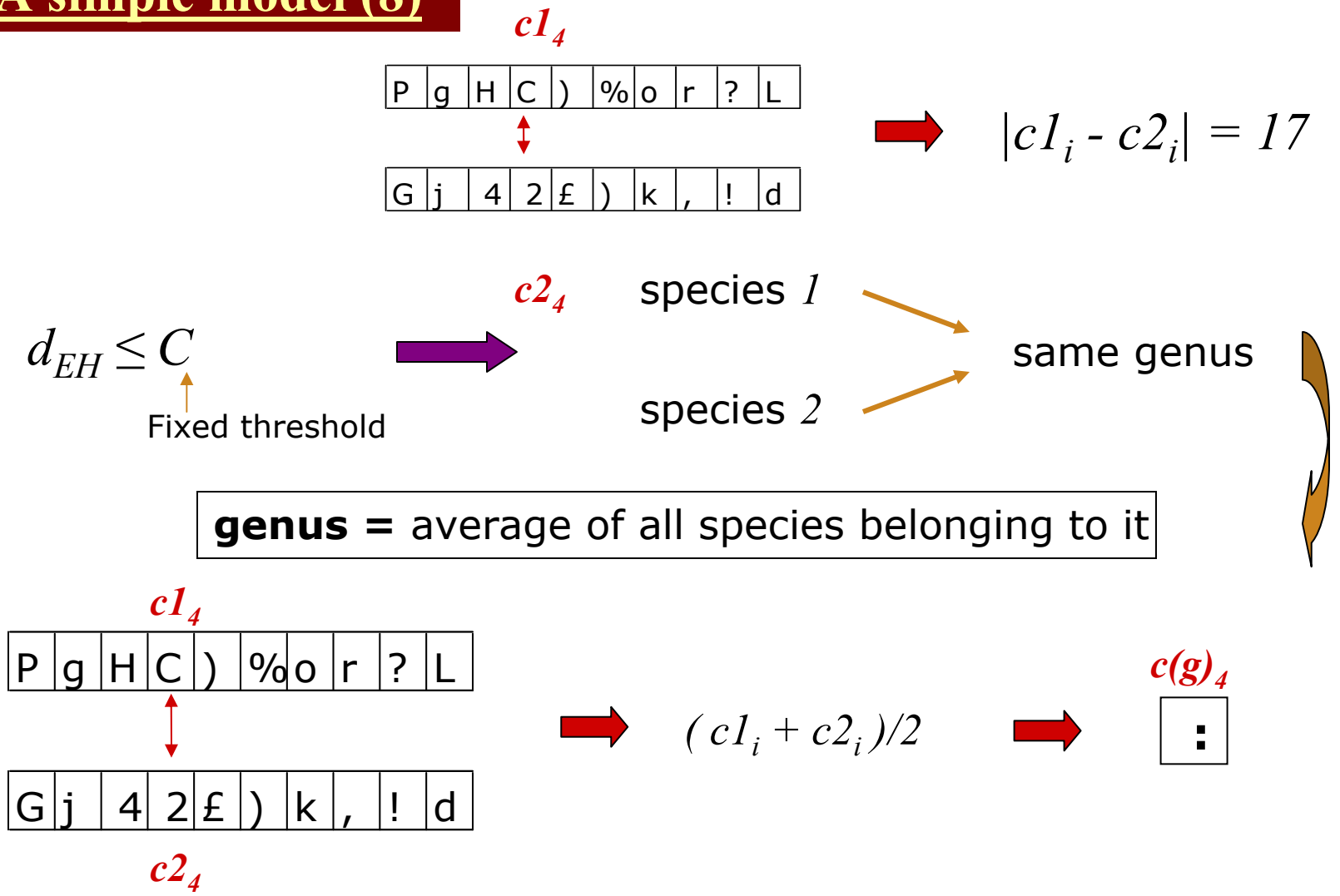
Two species are grouped in the same genus according to the ***extended Hamming distance*** d_{WH}:

$c1_i$ = character of species *1* with $i=1,\ldots\ldots\ldots,40$

$c2_i$ = character of species *2* with $i=1,\ldots\ldots\ldots,40$

$$\rightarrow \quad d_{EH} = \left(\sum_{i=1,40} |c1_i - c2_i| \right)/40$$

•3C A simple model (8)



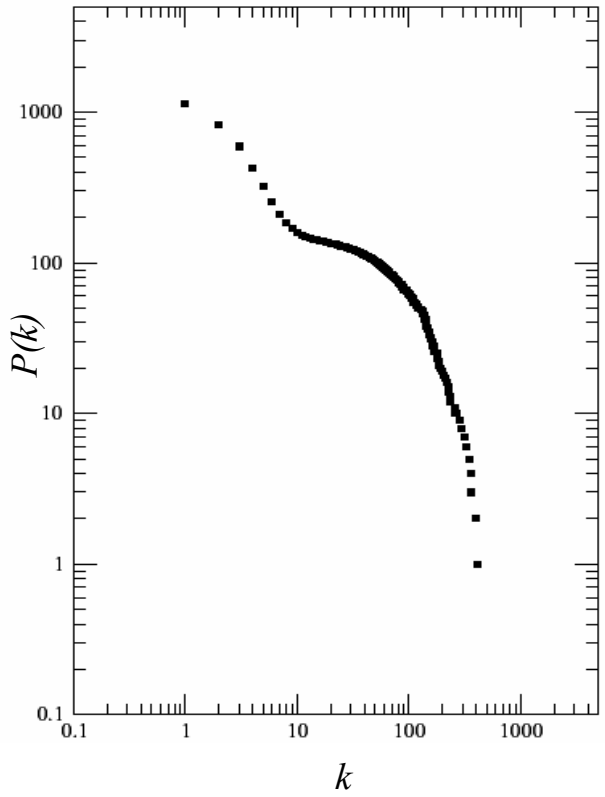
- ***Same proceedings at all levels with a fixed threshold for each one***
- ***At the last level (8) same phylum for all species (source node)***

•3C New species (9)

No correlation: species randomly created with no relationship between them

Genetic correlation: species are no more independent but descend from the same ancestor

- ***No correlation:***



- **ecosystems of 3000 species**
- **each character of each string is chosen at random**
- **quite big distance between two different species:**

$$\langle d_{EH} \rangle \sim 20$$

(S. `Ud~j <@a~Nf KMgXw´*:*4"j °zG9/Fy2J´R_x5

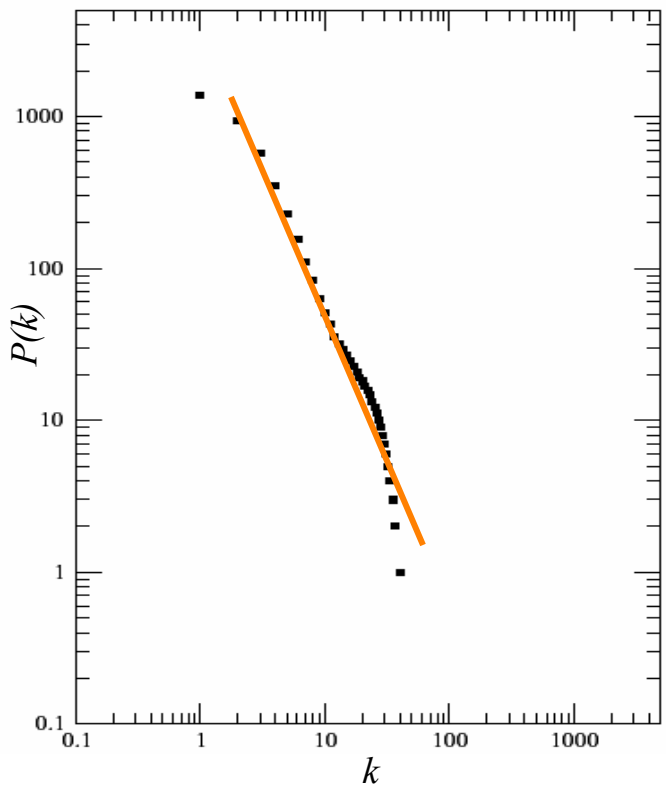
KL`<G´DQbnVUW;dLUxogZk*8yuNvDKZ+{Cx6I6dz

$$(\gamma_{top} \sim 1.7 \pm 0.2 \qquad \gamma_{bottom} \sim 3.0 \pm 0.2\ )$$

•3C New species (9)

- **single species ancestor of all species in the ecosystem**
- **at each *time step t* a new species appear:**
  - **chose (randomly) one of the species already present in the ecosystem**
  - **change one of its character**
- **3000 *time steps***

natural selection → *Environment* = **average of all species present in the the ecosystem at each time step *t*.**



- **At each time step *t* we calculate the distance between the environment and each species:**

$d_{EH} < C_{sel}$ → ***survival***

$d_{EH} > C_{sel}$ → ***extinction***

- **small distance between different species:**

$$\langle d_{EH} \rangle \sim 0.5$$

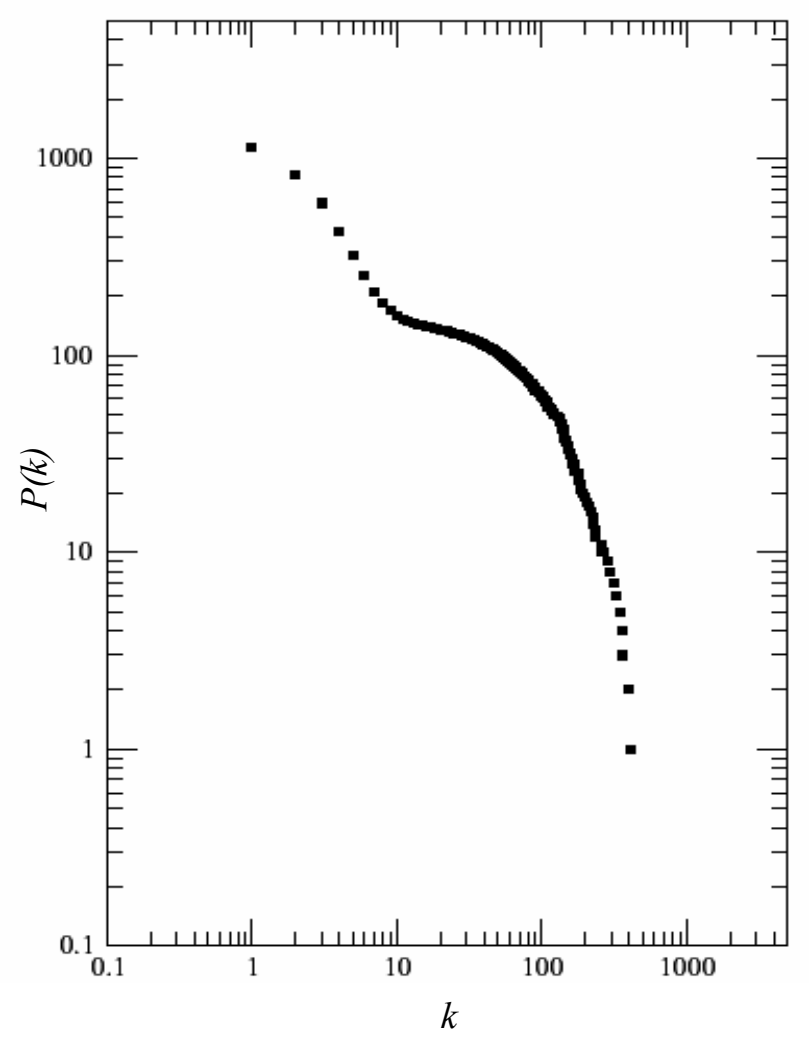
g50_"&y=Eo[lRC(xzG?g=X%W@@/Xr]TKg?6YG^Qz

g50_"&y=Eo[:RC(xzG?0=/%W´S/Xr]TKg?6K^^Qz

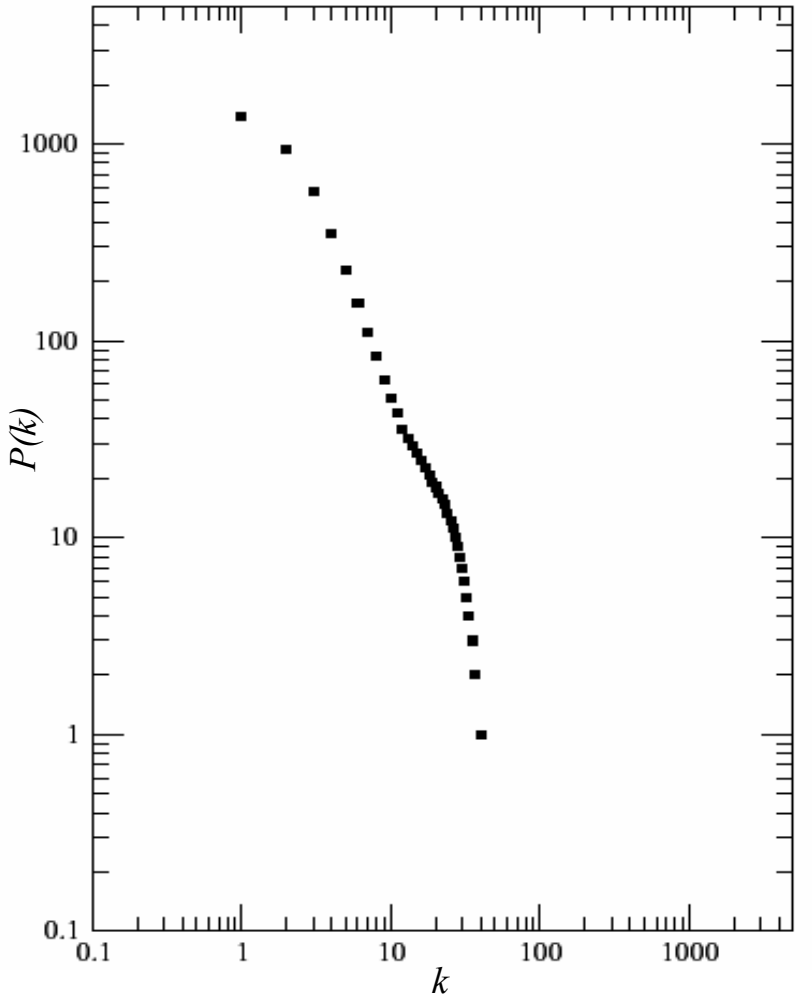
$$P(k) \sim k^{-\gamma} \qquad \gamma \sim 2.8 \pm 0.2$$

•3C A comparison (10)

Not Correlated:



Correlated:



•3C Conclusions (11)

Results:

- ***universality*** **(same statistical properties for ecosystems with different number of species and climatic environment) and *scale-free properties***
- **comparison between *geographical* and *random subsets*: evidence of the existence of a *correlation* between species in a same ecosystem due to some *self-organizing properties***
- ***simple model*** **presented shows the importance of a *genetic correlation***

Future:

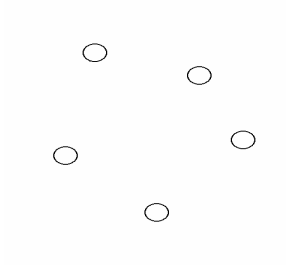
- **model presented is only a beginning: *improvement* with particular attention to *environment* and *natural selection***
- ***new data*** **and *chinese box ecosystem***

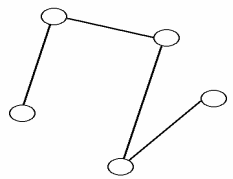
Applications:

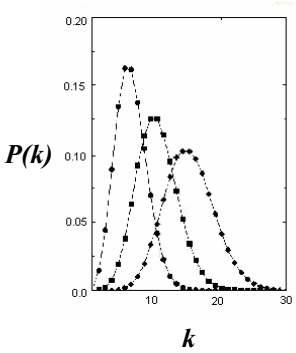
- **prevent *correlated plants extinction* due to human influence**
- **plant ecosystems *structure* and *reafforestation***

•3D Models (1)

Standard Theory of Random Graph
(Erdös and Rényi 1960)







Random Graphs are composed by starting with N vertices.
With probability p two vertices are connected by an edge

Degrees are Poisson distributed

$$P(k) = e^{-pN}\frac{(pN)^k}{k!}$$

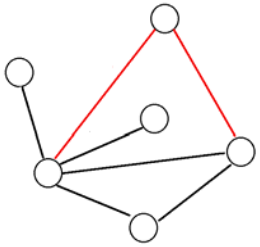
Model of Growing Networks
(Barabási 1999)

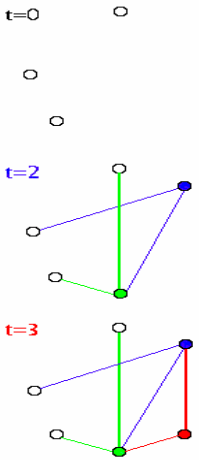
1) Growth

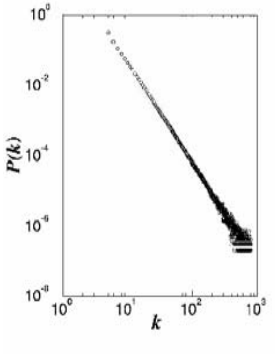
Every time step new nodes enter the system

2) Preferential Attachment

The probability to be connected depends on the degree P(k) ∝ k







Degrees are Power law distributed

$$P(k) \propto k^{-\gamma}$$

Exploration algorithm (P. De Los Rios, T. Petermann)

Starting from a highly connected node, we move along each of its links with probability p: if a link is lost, it is lost forever.

We iterate the procedure from every reached node, and so on until there are no further links to explore.

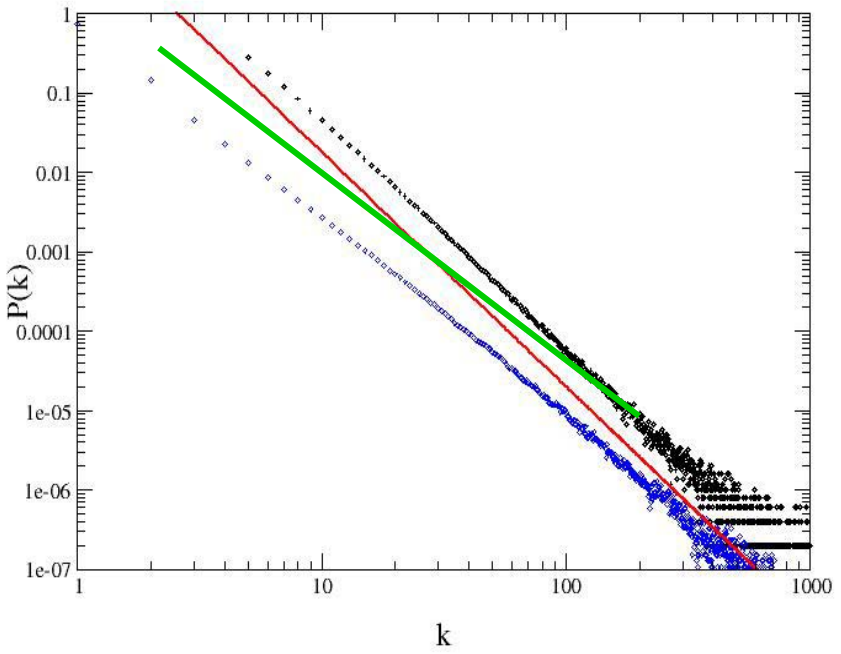
We then consider the network of the explored links and nodes as the "measured" network.

Exploration of a BA model

Every link is explored with probability p=0.5

The true exponent is 3…

the "measured" exponent is 2.5



An analytical derivation I

We use the lack of correlations in BA networks:
a node of degree k can be connected with nodes of any other degree (i.e. of any possible age).

So, we can reformulate the exploration problem as a process evolving in parallel to the growth of the network.

1. Start from m_0 nodes, all of them "detected"
2. Add a new node, and connect it to one of the existing nodes with preferential attachment
3. This new node is "detected" if it connects to a "detected" node AND if the connection is explored (with probability p)
4. Repeat from 2. till completion

An analytical derivation II

The total number *N(t)* of "detected" nodes at time *t* evolves according to the rate eq.

$$\frac{dN(t)}{dt} = p\int_0^t \frac{dN(t')}{dt'}\frac{k(t',t)}{2t}dt'$$

Density of detected nodes at time t

Probability to connect to a node born at time t'

An analytical derivation III

We assume that $N(t)$ grows algebraically
$\Rightarrow dN(t)/dt \sim t^{\alpha}$

We remember that $k(t',t) = (t/t')^{1/2}$

Plugging everything in the rate equation we obtain

$$\alpha = (p-1)/2$$

An analytical derivation IV – and last

We use the usual relation P(k)dk = dt, but now we modify it to $P(k)dk = dN(t)$, to obtain

$$P(k) \sim k^{-(2+p)}$$

$p=0.5$ implies $P(k) \sim k^{-2.5}$

Message: imperfect exploration changes the exponent!