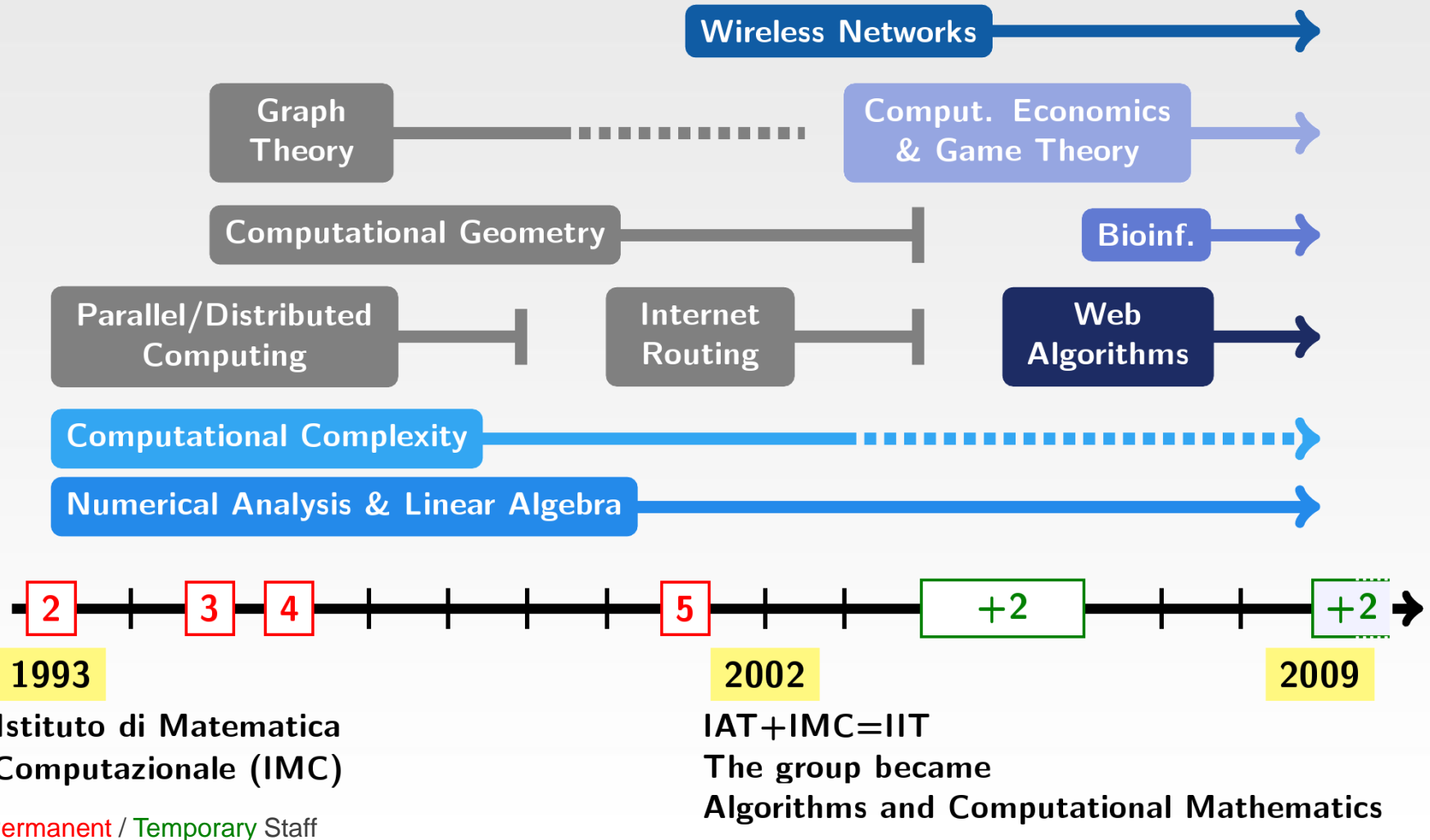


Pisa. 5/11/2009. Meeting of the ERCIM WG on Social Network Analysis

Algorithms and Computational Mathematics

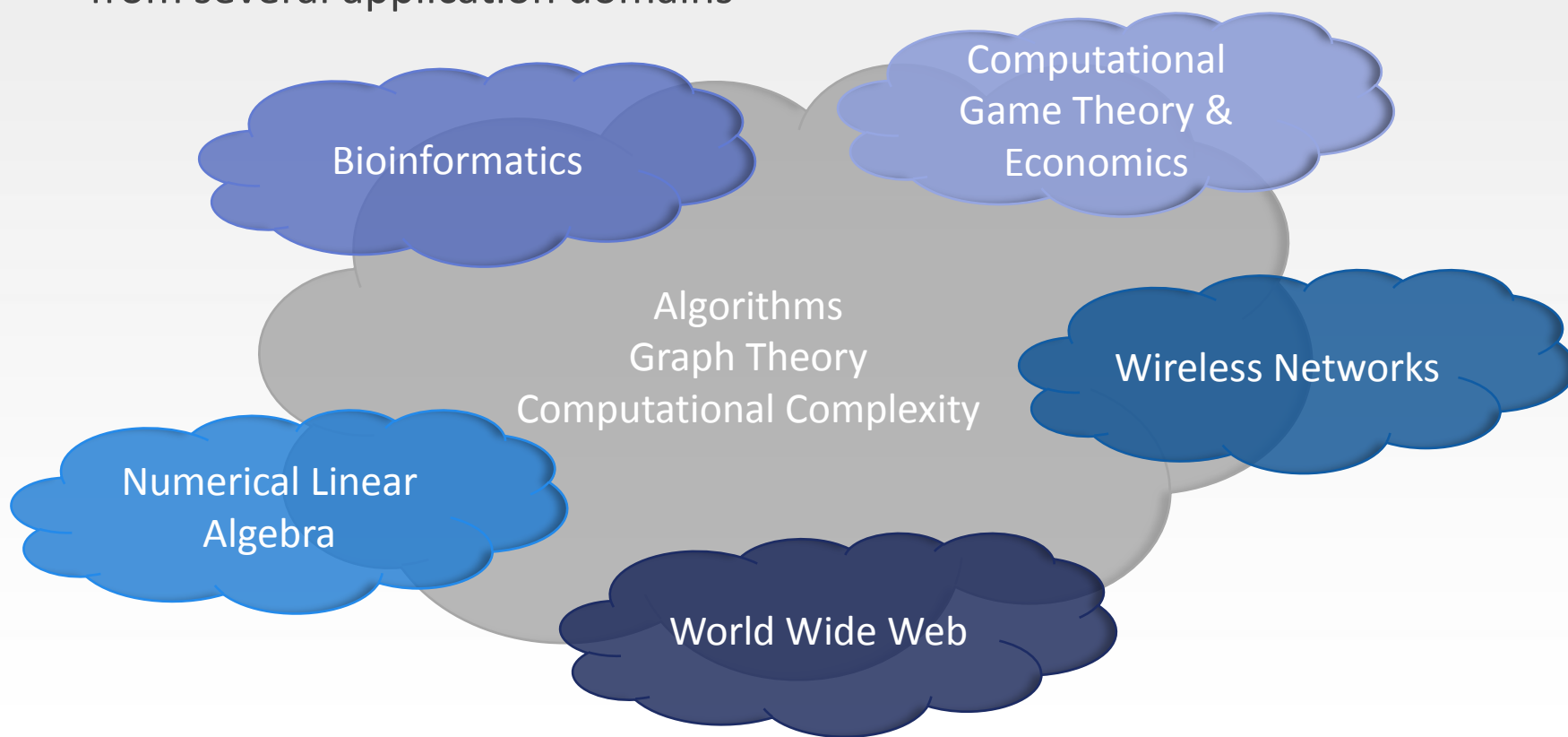
Marco Pellegrini
(Resp. Paola Favati)

Group's History



Research Areas

Established experience in Design and Analysis of Efficient Algorithms, Graph Theory, and Computational Complexity used to solve complex problems from several application domains



People

- Permanent Staff



Paola Favati
Research Director
Group's head



Bruno Codenotti
Research Director



Marco Pellegrini
Senior Researcher



Giovanni Resta
Senior Researcher



Paolo Santi
Senior Researcher

- Temporary Staff



Filippo Geraci
Researcher



M. Elena Renda
Researcher

- Term Contract Worker



Claudio Felicioli

- Associate Researchers



Mauro Leoncini
Univ. of Modena & Reggio Emilia, Italy



Giovanni Manzini
Univ. of Piemonte Orientale, Italy



Manuela Montangero
Univ. of Modena & Reggio Emilia, Italy

- Ph.D. Students

Michele Budinich

World Wide Web



- **Web Metasearch**
- **Personalized Web search**
- **Clustering and labeling of web snippets**
- **Automatic generation of digital video summaries**
- **Models for structural properties of the web graph**
- **Detection/Extraction of implicit cyber-communities in the web graph**
- **Tools development for the analysis and visualization of data retrieved from the WWW**
(Armil, ComWatch, PISA, Visto)



WebAlgo



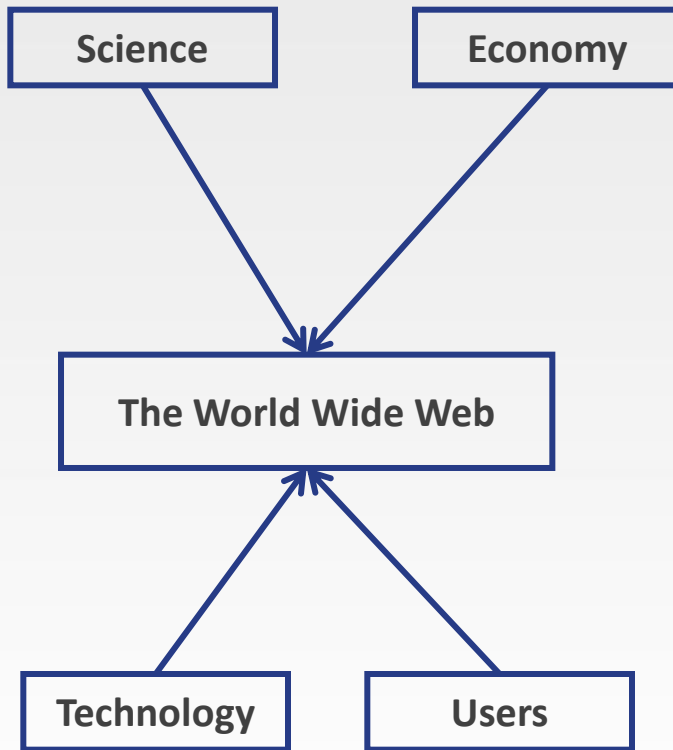
In collaboration with:

ISTI – CNR

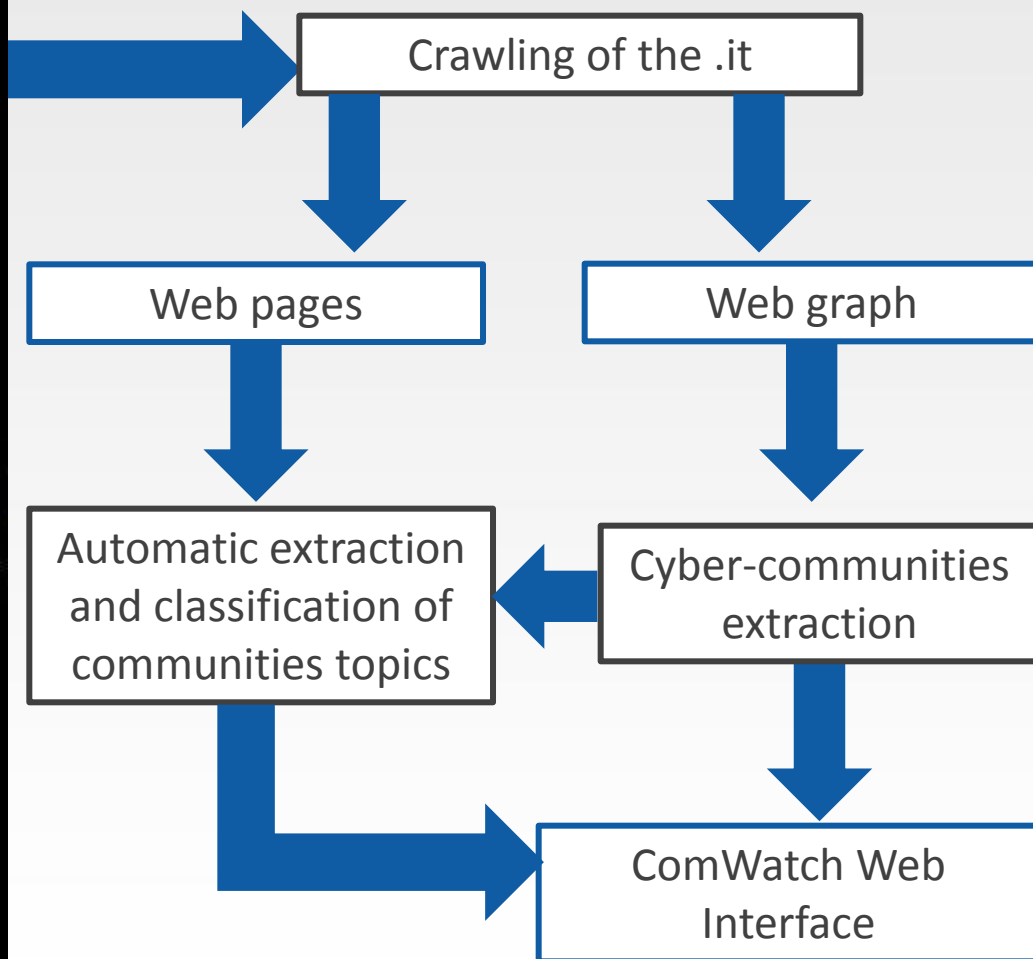
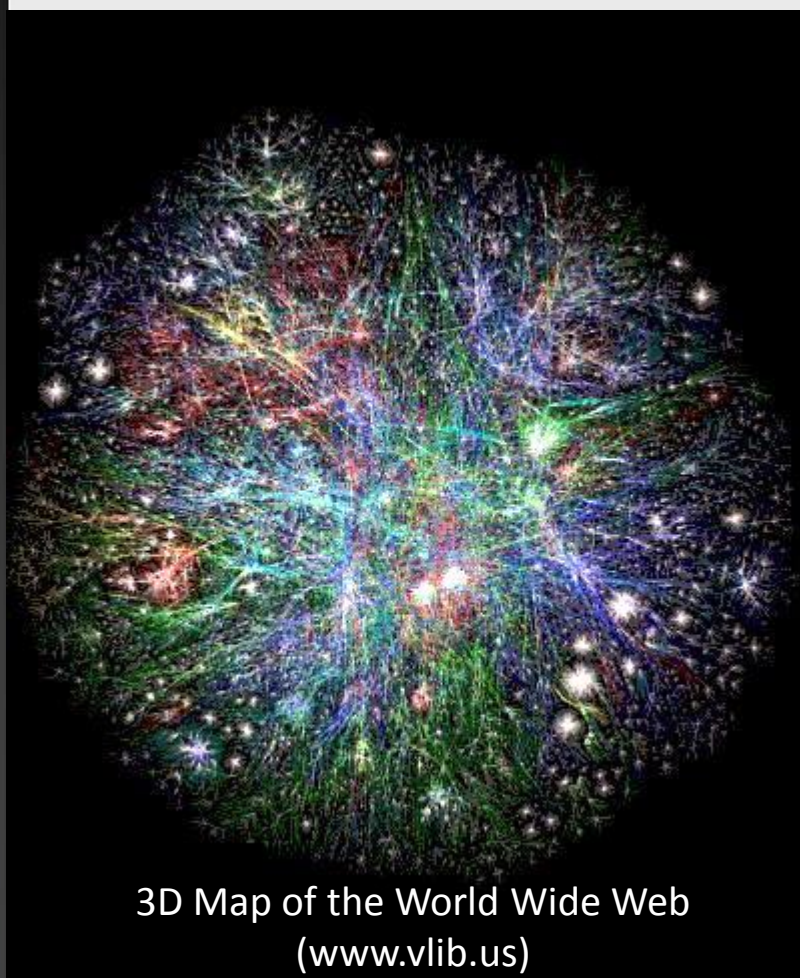
Dept. of Computer Engineering, Univ. of Siena

Web Business Intelligence

Mathematics is the Queen of the Sciences [Carl Friedrich Gauss]



ComWatch

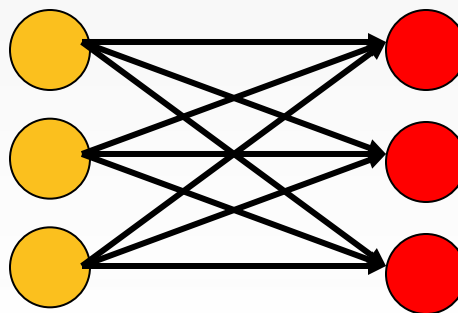


Dense subgraphs and communities

“Web communities are characterized by sufficiently dense directed bipartite sub-graphs”. Kumar-Raghavan-Rajagopalan-Tomkins (IJCTN 1999).

Fans

Centers



Experiments

- A) Experiments for Validation.
- B) Experiments on Real Data Sets.

.it 04	2004	41M pages	1.1G links
.uk 05	2005	39M pages	936M links
Web (Stanford)	2001	118M pages	1.0G links

- c) Data from crawls of national domains done by UbiCrawler (Vigna, Boldi, Santini, U. of Milano) and the Web-base (Stanford U.)

Effectiveness in Detection of Quasi-Cliques

We embed artificial dense subgraphs and find them blindly. Count the number of communities found (out of 10) per category (average of 70 experiments).

# Pages	40	9.6	9.8	9.7
	30	8.5	9.4	9.3
	20	3.6	7.6	8.3
	10	0	0.1	3.5
		Low	Med	High

Density

Detection of Quasi-bipartite graphs: /

- Number of communities found (out of 10) per category. Density in range [0.25..0.5] . Average of 53 experiments.

# Centers	80	0	5.2	9.6	10
	40	0	5.4	9.5	9.9
	20	0	2.7	5.4	6
	10	0	0	0	0
		10	20	40	80
		# of Fans			
		Low density			

Detection of Quasi-bipartite graphs://

- Number of communities found (out of 10) per category. Density in range [0.5..0.75] . Average of 53 experiments.

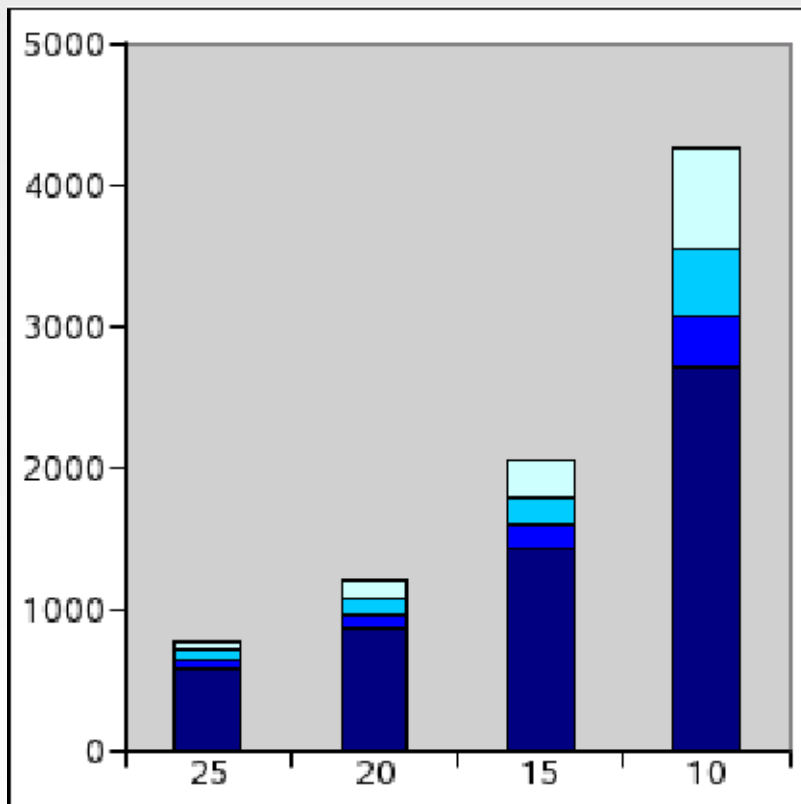
# Centers	80	1.2	8.4	9.7	10
	40	0.7	8	9.7	9.9
	20	0.9	7.9	9.6	9.9
	10	0.1	0.8	1.9	3.2
		10	20	40	80
		# of Fans			
		Med. density			

Detection of Quasi-bipartite graphs:III

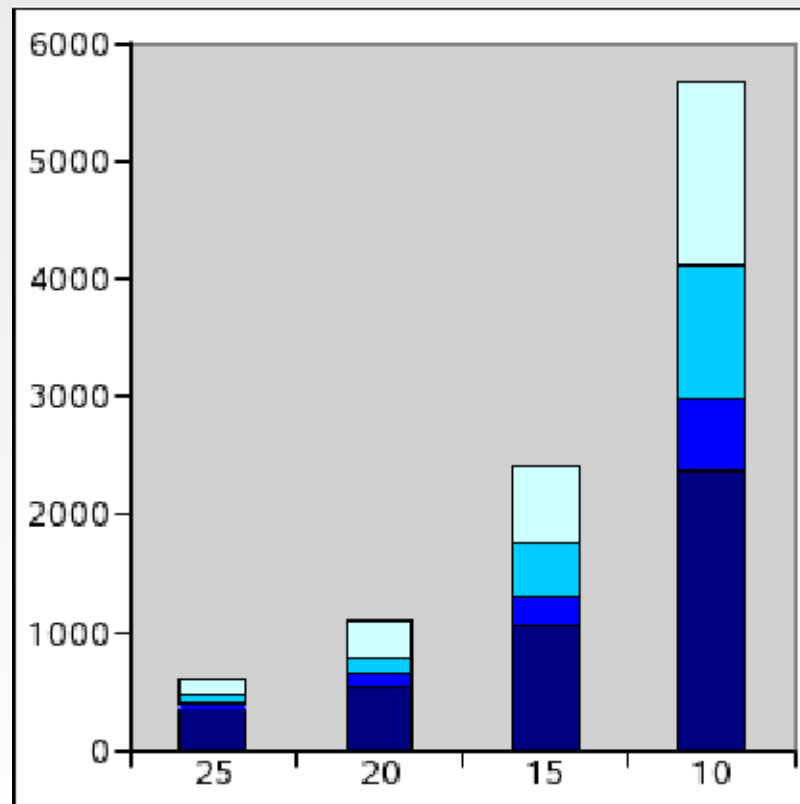
- Number of communities found (out of 10) per category. Density in range [0.75..1] . Average of 53 experiments.

# Centers	80	5.7	8.6	9.5	9.8
	40	5.4	8.6	9.7	9.8
	20	4.6	8.4	9.6	9.9
	10	3.3	6.5	9	9.7
		10	20	40	80
		# of Fans			
		High density			

Number and density of communities found.

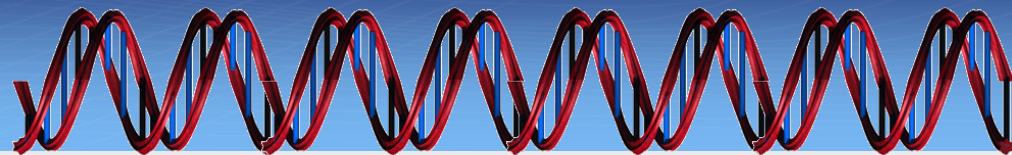


UK 2005



Web 2001

Bioinformatics



- **Biological sequences analysis/classification**
 - SNP haplotyping analysis
 - Microarray gene expression data analysis
 - Motif and Tandem Repeat identification/extraction
- **Diseases and gene expression profiling classification**
- **Web tools development for the analysis and visualization of biological sequences (AMIC@, ReHap, TReaDS)**



BioAlgo



In collaboration with:

IFC - CNR, Pisa

Dept. of Computer Engineering, Univ. of Pisa

Dept. of Computer Engineering , Univ. of Modena & Reggio Emilia

Microarray Lab, Dept. of Experimental Pathology, BMIE, Univ. of Pisa

Amic@: All Microarray Clusterings @ once

- **User friendly interface**
- **Benchmark for innovative clustering algorithms (K-Boost)**
- **A multi-method clustering engine for gene expression microarray data**
- **Suggestion of the number of clusters**
- **Automatic file format detection**
- **Measurements of cluster quality (homogeneity)**
- **Cluster visualization via heat-maps**
- **Off-line clustering wizard**

GUI

AMIC@

- SET DATA FILTERS
- START OVER
- WIZARD
- AMIC@ - How To

BIOALGO GROUP

- BIOALGO HOME
- BIOALGO@IIT.CNR.IT

RAW MATRIX

Zoom: 5

Show raw matrix

EXPRESSION LEVEL

563 3 d

COLOR LEGEND

- Under 102
- 102 --- 640.1
- 640.1 --- 1178.2
- 1178.2 --- 1716.3
- 1716.3 --- 2254.4
- 2254.4 --- 2792.5
- 2792.5 --- 3330.6
- 3330.6 --- 3868.7
- 3868.7 --- 4406.8
- 4406.8 --- 4944.9
- 4944.9 --- 5483
- Over 5483

K-Boost **FPF-SB** **K-means** **SOM** **HAC**

Metric

- Euclidean distance
- Pearson coefficient
- City block distance
- Cosine similarity
- Spearman rank correlation
- Kendall

Clustering 1

How Many clusters? 14 **Make clustering**

Suggested number of clusters 11 **Accept proposal** 2

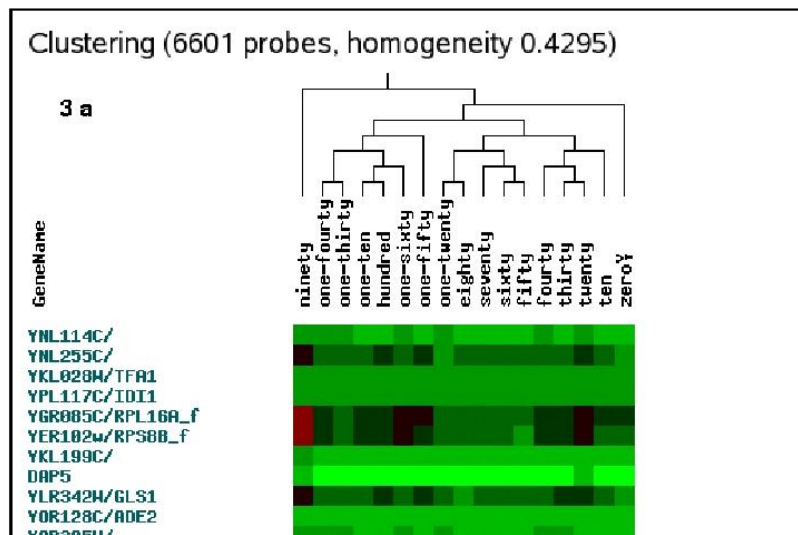
Show all clusters (6601 probes, homogeneity 0.4295) 3

Show Cluster Zoom: Sort columns 3

3 a 3 b 3 c

Homogeneity 0.4295 Clustering time 4.56

Download clustering 4



- Metric
- Number of clusters
- Cluster visualization
 - Show
 - Zoom
 - Sort
 - Expression Level
- Download
- Raw Matrix

Computational Game Theory & Economics

- Efficient algorithms for computing market equilibria with Leontief functions and extensions
- Convex programming techniques for computing equilibria in Markets
- Nash equilibria in two-persons games and equilibria in exchange markets (characterizations via computational equivalence)
- Applications to ad-words markets

In collaboration with:

Dept. of Computer Science, Univ. of Iowa
Dept. of Computer Science, Univ. of Chicago
TTI - Chicago

