# Multivariate analysis and complex networks

## 5: Complex networks for data-driven medicine

Antonio Scala

CNR-ISC, the Institute for Complex Systems

INFN School of Statistics
Ischia, May 2017

# Introduction

Motivation: Mining knowledge from Medical Records

# Introduction

Motivation: Mining knowledge from Medical Records

Methods: Network Analysis for Case-Features dataset

# Introduction

| | |
|---|---|
| Motivation: | Mining knowledge from Medical Records |
| Methods: | Network Analysis for Case-Features dataset |
| Case-study: | Childhood orthodontics |

# Data-driven science

- new knowledge will simply emerge as plausible patterns from data-mining

# Data-driven science

- new knowledge will simply emerge as plausible patterns from data-mining
- extend the reach of computers from analysis to hypothesis

# Data-driven science

- new knowledge will simply emerge as plausible patterns from data-mining
- extend the reach of computers from analysis to hypothesis
- new questions that must be answered (why *that* pattern?)

# Data-driven science

- new knowledge will simply emerge as plausible patterns from data-mining
- extend the reach of computers from analysis to hypothesis
- new questions that must be answered (why *that* pattern?)

The aim is to generate knowledge

“”Machine Science” James Evans and Andrey Rzhetsky

Science **329**. no. 5990, pp. 399 (2010)

# Complex Networks

- *systems* cannot be understood in terms of simple atomic components

# Complex Networks

- *systems* cannot be understood in terms of simple atomic components
- *data mining* can find simple relations and reduce the dimensionality of a problem

# Complex Networks

- *systems* cannot be understood in terms of simple atomic components
- *data mining* can find simple relations and reduce the dimensionality of a problem
- *data-mining* enriches data with *meta-data* (classification)

# Complex Networks

- *systems* cannot be understood in terms of simple atomic components
- *data mining* can find simple relations and reduce the dimensionality of a problem
- *data-mining* enriches data with *meta-data* (classification)

complex systems "resist" data-mining as they could not be easily broken in pieces

# Complex Networks

- *systems* cannot be understood in terms of simple atomic components
- *data mining* can find simple relations and reduce the dimensionality of a problem
- *data-mining* enriches data with *meta-data* (classification)

complex systems “resist“ data-mining as they could not be easily broken in pieces

- *network science* looks globally at the relations among the components of a system

# Complex Networks

- *systems* cannot be understood in terms of simple atomic components
- *data mining* can find simple relations and reduce the dimensionality of a problem
- *data-mining* enriches data with *meta-data* (classification)

complex systems “resist“ data-mining as they could not be easily broken in pieces

- *network science* looks globally at the relations among the components of a system
- complex network analysis reveals new conceptual classes *emerging due to the the interaction among the data*

# Data

medicine: clinical cases - clinical features

2d cromatography: experiment - substance is (not) detected

test/marketing pools: people - answer to question

human geography people - locations

coauthoring: author - publication

textual co-occurrence: word - query/sentence/paragraph/book

community matrices: taxa - sample

......

# Data

medicine: clinical cases - clinical features

2d cromatography: experiment - substance is (not) detected

test/marketing pools: people - answer to question

human geography people - locations

coauthoring: author - publication

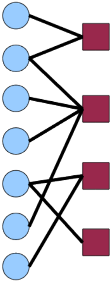
textual co-occurrence: word - query/sentence/paragraph/book

community matrices: taxa - sample

......

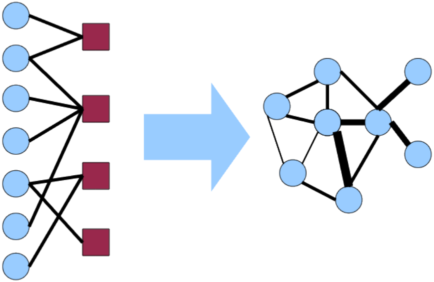
BIPARTITE GRAPH / BINARY MATRICES

# Methodology



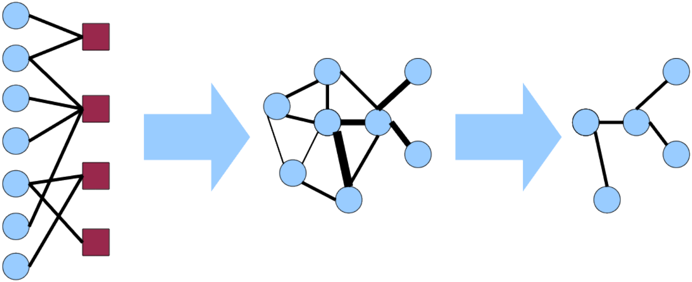
**DATA**

# Methodology



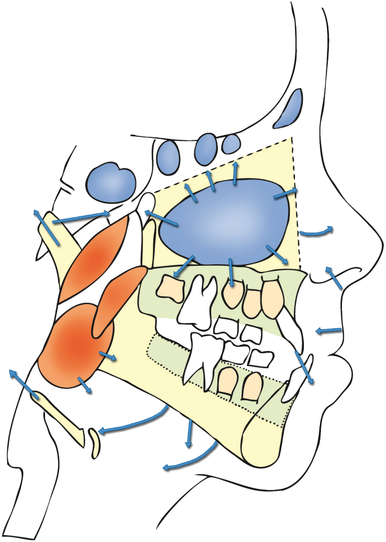
PROJECT

# Methodology



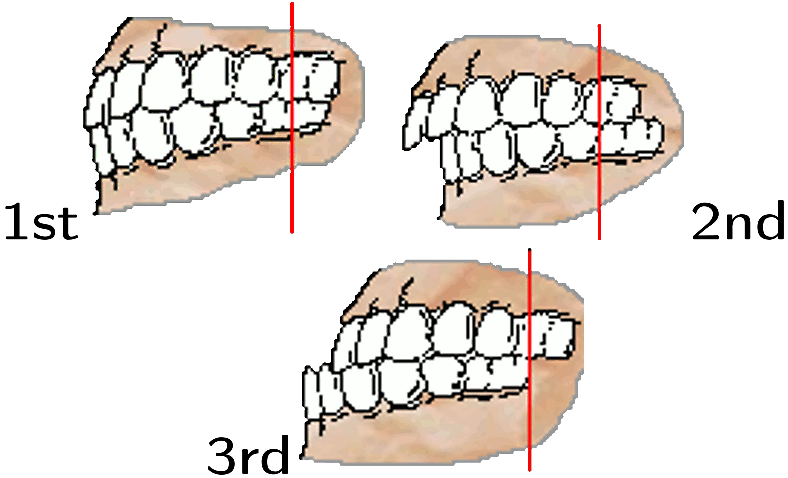
SELECT

# The Complex Oro-Facial System



- components
- relations
- interactions
- dynamics

# Dental Classes



## Cases

- 97 subjects (49 males, 48 females)
- age 8-13 (mean 10.2)
- 28 1st class, 44 2nd class, 25 3rd class patients
- Uniform age and gender

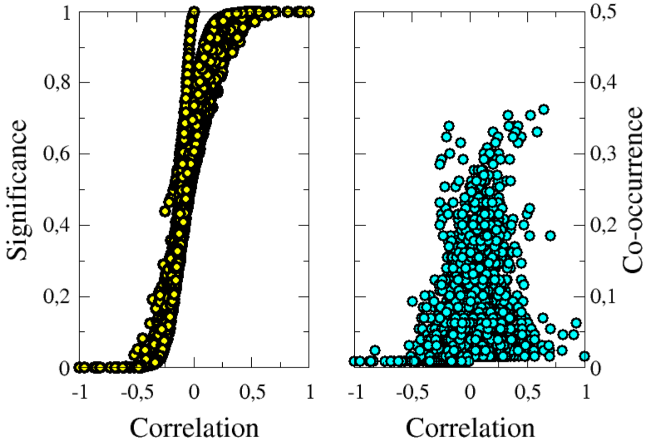
## Cases

- 97 subjects (49 males, 48 females)
- age 8-13 (mean 10.2)
- 28 1st class, 44 2nd class, 25 3rd class patients
- Uniform age and gender

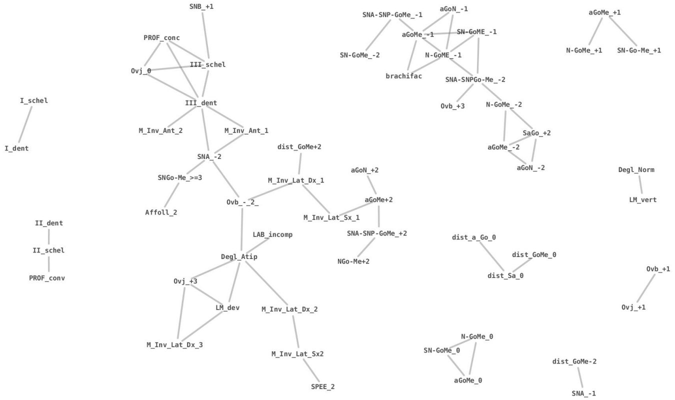
## Features

- 33 clinical, anatomic, functional and radiographic features
- 16 landmarks on the cephalograms
- 17 functional and clinical signs or oral habits
- score from -3 to + 3 = standard deviations from the mean

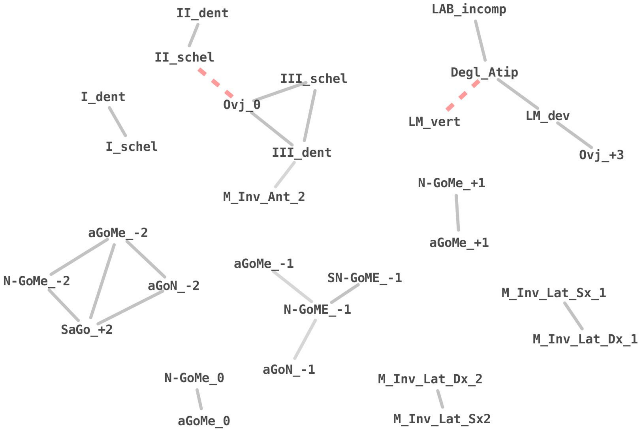
# Correlation vs Co-occurrence



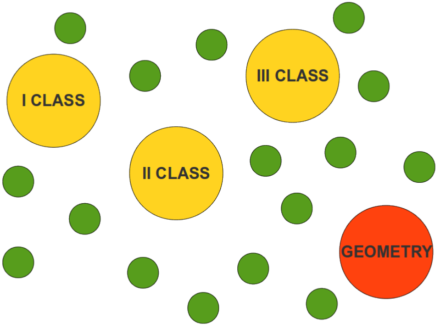
# Positive Correlations >40%



# All Together >50%

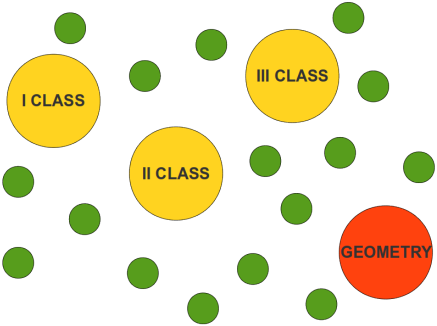


# Emergence of dental classes




- Emergence of dental classes

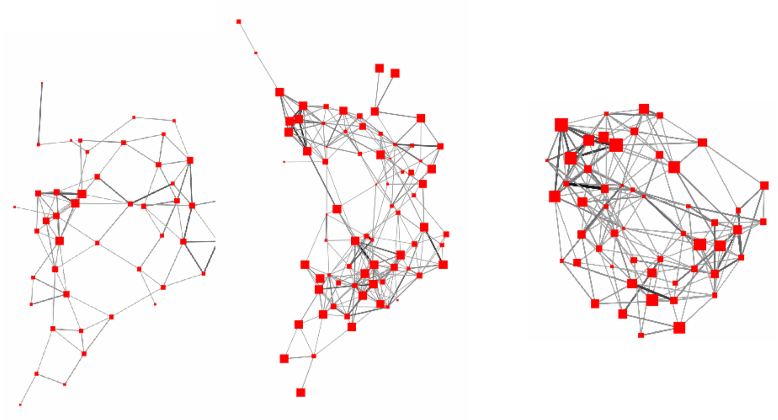
# Emergence of dental classes




- Emergence of dental classes
- Emergence of geometrical constraints

# Networks for single Classes

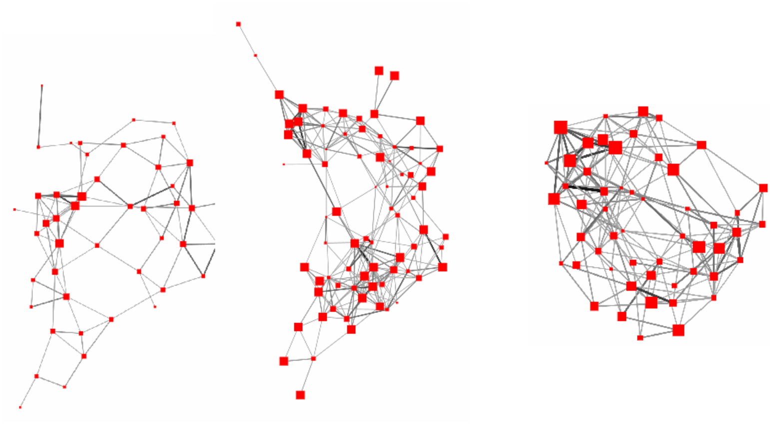
$\varphi \geqq 30\%$ (treshold for medium & high correlations)



- 3rd class malocclusion strongly connected but devoid of distinctive peculiar hubs

# Networks for single Classes

$\varphi \geqq 30\%$ (treshold for medium & high correlations)



- 3rd class malocclusion strongly connected but devoid of distinctive peculiar hubs
- multiple therapeutic attack on the many highly connected nodes

# Networks for single classes

| | Average Degree | Clustering coefficient | Mean Shortest |
|---|---|---|---|
| 1st (normal) | 4.04 | 0.28 | 3.43 |
| 2nd | 6.45 | 0.36 | 3.13 |
| 3rd | 7.09 | 0.31 | 2.39 |

- 2nd and 3rd class features are more connected than those of the control patients.

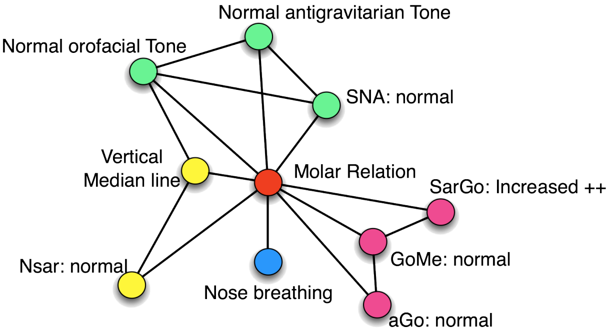
# Networks for single classes

| | Average Degree | Clustering coefficient | Mean Shortest |
|---|---|---|---|
| 1st (normal) | 4.04 | 0.28 | 3.43 |
| 2nd | 6.45 | 0.36 | 3.13 |
| 3rd | 7.09 | 0.31 | 2.39 |

- 2nd and 3rd class features are more connected than those of the control patients.
- 3rd class patients shows a much higher connection and closeness: this topology allows a high transmission of the bite forces and neuromuscular inputs

# Hubs in 2nd Class

Network local picture (simplified)




- peculiar hubs as starting point for an orthodontic selective treatment
- hubs do not necessarily correspond to the most evident clinical signs

# Conclusions I

- Dental classes emerge as clusters of interrelated feature

# Conclusions I

- Dental classes emerge as clusters of interrelated feature
- Features can be considered in the light of the appropriate network specific for a particular malocclusion

# Conclusions I

- Dental classes emerge as clusters of interrelated feature
- Features can be considered in the light of the appropriate network specific for a particular malocclusion
- Networks represent the system in a visually intuitive way, focus on most important features

# Conclusions I

- Dental classes emerge as clusters of interrelated feature
- Features can be considered in the light of the appropriate network specific for a particular malocclusion
- Networks represent the system in a visually intuitive way, focus on most important features
- Valuable tool for evidence-based diagnosis in primary orthodontic care

# Conclusions I

- Dental classes emerge as clusters of interrelated feature
- Features can be considered in the light of the appropriate network specific for a particular malocclusion
- Networks represent the system in a visually intuitive way, focus on most important features
- Valuable tool for evidence-based diagnosis in primary orthodontic care
- *Could also be applied to other clinical problems*

# Conclusions II

- we can extend the reach of computers from analysis to assist hypothesis

# Conclusions II

- we can extend the reach of computers from analysis to assist hypothesis
- new knowledge simply emerges as plausible patterns from network-based data-mining

# Conclusions II

- we can extend the reach of computers from analysis to assist hypothesis
- new knowledge simply emerges as plausible patterns from network-based data-mining

Complex Networks can contribute to mine new knowledge

## next step: Medical data for virtual experiments

- ICT for data consolidation, interfaces & algorithms

# next step: Medical data for virtual experiments

- ICT for data consolidation, interfaces & algorithms
- STATISTICS for the design of statistical test

# next step: Medical data for virtual experiments

- ICT for data consolidation, interfaces & algorithms
- STATISTICS for the design of statistical test
- COMPLEX NETWORKS for emerging knowledge representation, data mining

## next step: Medical data for virtual experiments

- ICT for data consolidation, interfaces & algorithms
- STATISTICS for the design of statistical test
- COMPLEX NETWORKS for emerging knowledge representation, data mining
- MEDICINE for
  - clinical records
  - designing experiments

# Application to medical diagnostic

The classification of human diseases builds on observed correlations between pathological analysis and clinical syndromes (observational skills to define the syndromic phenotype)

# Application to medical diagnostic

The classification of human diseases builds on observed correlations between pathological analysis and clinical syndromes (observational skills to define the syndromic phenotype)

**Problem** Classic diagnostic strategy is naturally limited by the lack of sensitivity in identifying pre-clinical disease and by the lack of specificity in defining disease unequivocally

# Application to medical diagnostic

The classification of human diseases builds on observed correlations between pathological analysis and clinical syndromes (observational skills to define the syndromic phenotype)

**Problem** Classic diagnostic strategy is naturally limited by the lack of sensitivity in identifying pre-clinical disease and by the lack of specificity in defining disease unequivocally

**GOAL** infer syndromic phenotypes from clinical data via complex networks methods