Combinatorial Optimization Approaches for Clustering and Biclustering

Paola Festa

Dipartimento di Matematica e Applicazioni “R. Caccioppoli”

Università degli Studi di Napoli FEDERICO II

http://www.dma.unina.it/∼festa/

E-mail: paola.festa@unina.it

Outline

- Introduction to Data Clustering:
  - definitions and notation;
  - problem formulation;
  - state-of-the-art methods;
  - our recent proposal: a hybrid GRASP with Path Relinking;
  - analysis of a case study for Biological Data on 5 datasets.

Outline

- Introduction to Data Clustering:
    - definitions and notation;
    - problem formulation;
    - state-of-the-art methods;
    - our recent proposal: a hybrid GRASP with Path Relinking;
    - analysis of a case study for Biological Data on 5 datasets.

- Introduction to Data BiClustering:
    - definitions and notation;
    - problem formulation;
    - state-of-the-art methods;
    - our recent proposal: a GRASP-like algorithm;
    - analysis of a case study for Biological Data on 2 datasets.

Outline

- ❍ Introduction to Data Clustering:
  - ☞ definitions and notation;
  - ☞ problem formulation;
  - ☞ state-of-the-art methods;
  - ☞ our recent proposal: a hybrid GRASP with Path Relinking;
  - ☞ analysis of a case study for Biological Data on 5 datasets.
- ❍ Introduction to Data BiClustering:
  - ☞ definitions and notation;
  - ☞ problem formulation;
  - ☞ state-of-the-art methods;
  - ☞ our recent proposal: a GRASP-like algorithm;
  - ☞ analysis of a case study for Biological Data on 2 datasets.
- ❍ Conclusions and Future directions.

Data Clustering

Collaborations

Material on Data clustering presented in this seminar is based on joint work with:

✔ Mauricio G.C. Resende

AT&T Labs Research, Florham Park, NJ, USA

✔ Ricardo M.A. Silva

Universidade Federal de Lavras, Lavras, MG, Brazil

✔ Rafael M.D. Frinhani and Geraldo R. Mateus

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

Outline

Data Clustering
Collaborations
Description
Applications
Problem Formulation, ①
Problem Formulation, ②
Problem Formulation, ③
Problem Formulation, ④
Problem Formulation, ⑤
Problem Formulation, ⑥
Problem Formulation, ⑦
Problem Formulation, ⑧
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

Description

Task: to group data (viewed as a set of *objects*) s.t.

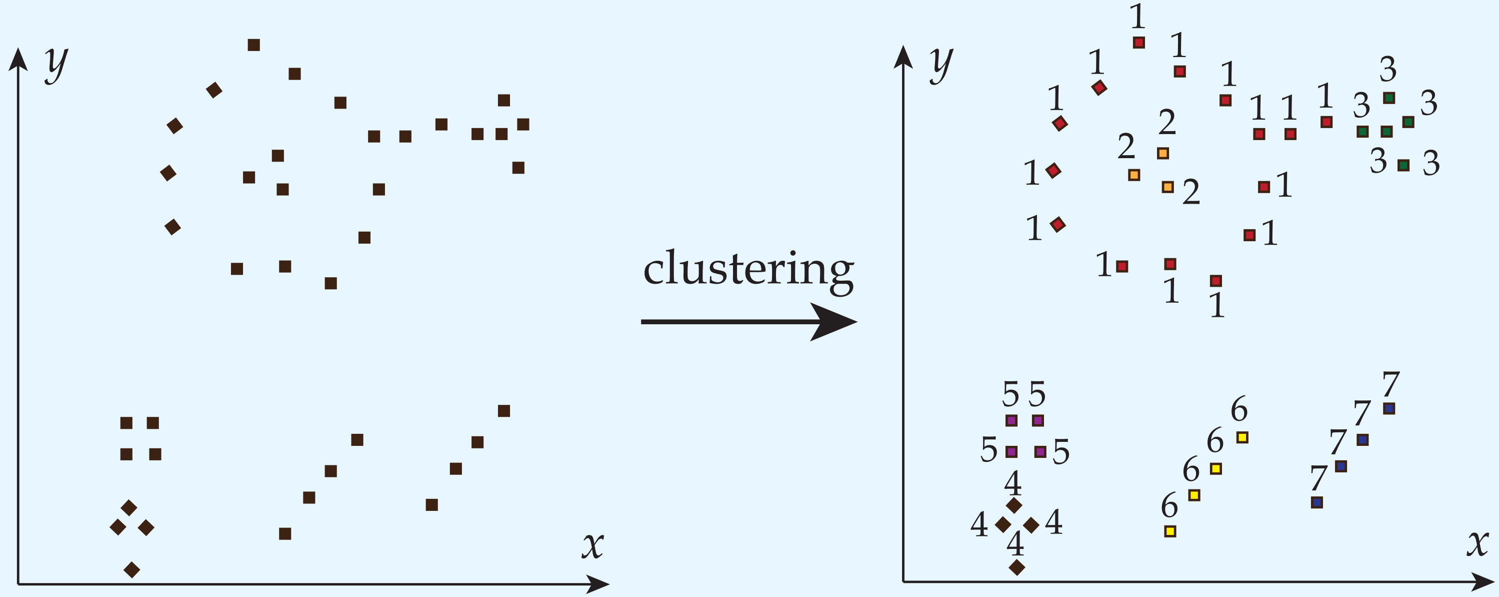
- ✔ the most similar objects belong to the same group or *cluster*, and
- ✔ the dissimilar objects are assigned to different clusters.

Description

Task: to group data (viewed as a set of *objects*) s.t.

- ✔ the most similar objects belong to the same group or *cluster*, and
- ✔ the dissimilar objects are assigned to different clusters.

Example for a 2-dimensional data set (“easy” for humans):

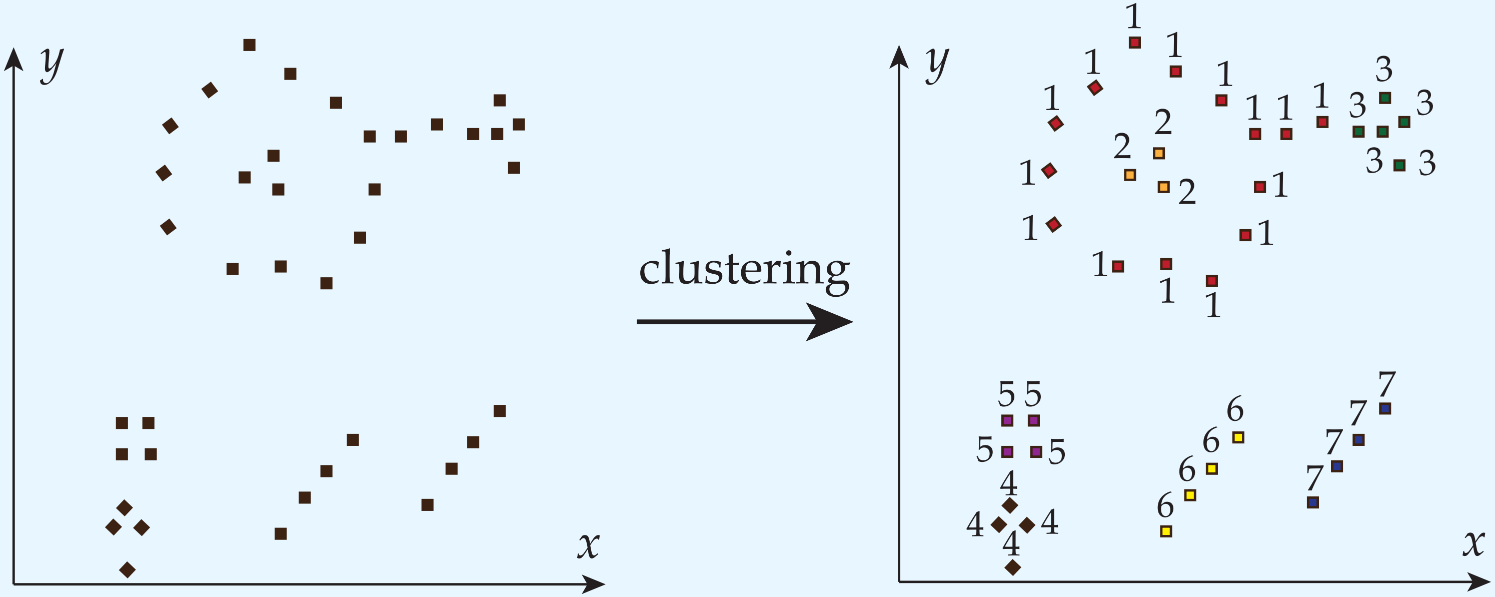


Description

Task: to group data (viewed as a set of *objects*) s.t.

- ✔ the most similar objects belong to the same group or *cluster*, and
- ✔ the dissimilar objects are assigned to different clusters.

Example for a 2-dimensional data set ("easy" for humans):



Bad new: most real–world problems involve clustering in higher dimensions!

Outline

Data Clustering
Collaborations
Description
Applications
Problem Formulation, ①
Problem Formulation, ②
Problem Formulation, ③
Problem Formulation, ④
Problem Formulation, ⑤
Problem Formulation, ⑥
Problem Formulation, ⑦
Problem Formulation, ⑧
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

Applications

Applications include:

- ➪ natural language processing [Ushioda et al, 1996];
- ➪ galaxy formation [Wu et al, 1993];
- ➪ image segmentation [White et al, 1991];
- ➪ biological data.
  [Jain et al, 1999 – Jiang et al, 2004 – Nascimento et al, 2010].

Problem Formulation, ①

We are given

☞ a set of N objects $\mathcal{O} = \{o_1, \ldots, o_N\}$;

☞ a set of M of pre-assigned clusters $\mathcal{S} = \{S_1, \ldots, S_M\}$;

☞ a function $d : \mathcal{O} \times \mathcal{O} \mapsto \mathbb{R}$ that assigns to each $o_i, o_j \in \mathcal{O}$ a "distance" or "similarity" $d_{ij} \in \mathbb{R}$

(usually, $d_{ij} \geq 0$, $d_{ii} = 0$, $d_{ij} = d_{ji}$, for $i, j = 1, \ldots, N$);

Problem Formulation, ①

We are given

- ☞ a set of N objects $\mathcal{O} = \{o_1, \ldots, o_N\}$;
- ☞ a set of M of pre-assigned clusters $\mathcal{S} = \{S_1, \ldots, S_M\}$;
- ☞ a function $d : \mathcal{O} \times \mathcal{O} \mapsto \mathbb{R}$ that assigns to each $o_i, o_j \in \mathcal{O}$ a "distance" or "similarity" $d_{ij} \in \mathbb{R}$

  (usually, $d_{ij} \geq 0$, $d_{ii} = 0$, $d_{ij} = d_{ji}$, for $i, j = 1, \ldots, N$);

By introducing

- ☞ a set of $N \times M$ decision variables $x_{ik} \in \{0, 1\}$ s.t.

$$x_{ik} = \begin{cases} 1, & \text{if } o_i \in \mathcal{O} \text{ is in cluster } S_k; \\ 0, & \text{otherwise.} \end{cases}$$

Problem Formulation, ②

Data clustering can be formulated as a **non-linear 0-1 problem**: [**Nascimento et al's (2010)**]

$$\text{(DC)} \quad \min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \sum_{k=1}^{M} x_{ik} \cdot x_{jk}$$

s.t.

$$(1) \quad \sum_{k=1}^{M} x_{ik} = 1, \qquad i = 1, \ldots, N$$

$$(2) \quad \sum_{i=1}^{N} x_{ik} \geq 1, \qquad k = 1, \ldots, M$$

$$(3) \quad x_{ik} \in \{0, 1\}, \qquad i = 1, \ldots, N, \ k = 1, \ldots, M.$$

(DC) is a non-linear 0-1 problem:

$$\min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \sum_{k=1}^{M} x_{ik} \cdot x_{jk} \Longrightarrow \boxed{\text{Minimize the distance between objects in the same cluster}}$$

Problem Formulation, ③

Data clustering can be formulated as a **non-linear 0-1 problem**: [**Nascimento et al's (2010)**]

$$(DC) \quad \min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \sum_{k=1}^{M} x_{ik} \cdot x_{jk}$$

s.t.

$$(1) \quad \sum_{k=1}^{M} x_{ik} = 1, \qquad i = 1, \ldots, N$$

$$(2) \quad \sum_{i=1}^{N} x_{ik} \geq 1, \qquad k = 1, \ldots, M$$

$$(3) \quad x_{ik} \in \{0, 1\}, \qquad i = 1, \ldots, N, \; k = 1, \ldots, M.$$

$$\sum_{k=1}^{M} x_{ik} = 1, i = 1, \ldots, N \Longrightarrow$$

They assure that each o_i belongs to only one cluster

Problem Formulation, ④

Data clustering can be formulated as a **non-linear 0-1 problem**: [**Nascimento et al's (2010)**]

$$\text{(DC)} \quad \min \quad \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \sum_{k=1}^{M} x_{ik} \cdot x_{jk}$$

s.t.

$$(1) \quad \sum_{k=1}^{M} x_{ik} = 1, \qquad i = 1, \ldots, N$$

$$(2) \quad \sum_{i=1}^{N} x_{ik} \geq 1, \qquad k = 1, \ldots, M$$

$$(3) \quad x_{ik} \in \{0, 1\}, \qquad i = 1, \ldots, N, \; k = 1, \ldots, M.$$

$$\sum_{i=1}^{N} x_{ik} \geq 1, k = 1, \ldots, M \Longrightarrow$$

They guarantee that each cluster S_k contains at least one object

Problem Formulation, ⑤

Remedy to the non-linear o.f. = linearization [Nascimento et al, 2010]:

$\forall\, i, j = 1, \ldots, N, \quad y_{ij} = 1 \quad \Leftrightarrow \quad o_i, o_j \in \mathcal{O}$ are in the same cluster.

$$
\begin{array}{lll}
\text{(LDC)} & \min & \displaystyle\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \cdot y_{ij} \\
& \text{s.t.} & \\
& (1) \quad \displaystyle\sum_{k=1}^{M} x_{ik} = 1, & i = 1, \ldots, N \\
& (2) \quad \displaystyle\sum_{i=1}^{N} x_{ik} \geq 1, & k = 1, \ldots, M \\
& (3) \quad x_{ik} \in \{0, 1\}, & i = 1, \ldots, N, \ k = 1, \ldots, M \\
& (4) \quad y_{ij} \geq x_{ik} + x_{jk} - 1, & i = 1, \ldots, N, \ j = i+1, \ldots, N, \ k = 1, \ldots, M \\
& (5) \quad y_{ij} \geq 0, & i = 1, \ldots, N, \ j = i+1, \ldots, N.
\end{array}
$$

Outline

Data Clustering
Collaborations
Description
Applications
Problem Formulation, ①
Problem Formulation, ②
Problem Formulation, ③
Problem Formulation, ④
Problem Formulation, ⑤
Problem Formulation, ⑥
Problem Formulation, ⑦
Problem Formulation, ⑧
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

Problem Formulation, ⑥

Linearization [Nascimento et al, 2010]:

$$\text{(LDC)} \quad \min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \cdot y_{ij}$$

s.t.

$$
\begin{array}{lll}
(1) & \displaystyle\sum_{k=1}^{M} x_{ik} = 1, & i = 1, \ldots, N \\
(2) & \displaystyle\sum_{i=1}^{N} x_{ik} \geq 1, & k = 1, \ldots, M \\
(3) & x_{ik} \in \{0, 1\}, & i = 1, \ldots, N, \; k = 1, \ldots, M \\
(4) & y_{ij} \geq x_{ik} + x_{jk} - 1, & i = 1, \ldots, N, \; j = i+1, \ldots, N, \; k = 1, \ldots, M \\
(5) & y_{ij} \geq 0, & i = 1, \ldots, N, \; j = i+1, \ldots, N.
\end{array}
$$

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \cdot y_{ij} \Longrightarrow$$

Minimize the distance between objects in the same cluster

Outline

Data Clustering
Collaborations
Description
Applications
Problem Formulation, ①
Problem Formulation, ②
Problem Formulation, ③
Problem Formulation, ④
Problem Formulation, ⑤
Problem Formulation, ⑥
Problem Formulation, ⑦
Problem Formulation, ⑧
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

Problem Formulation, ⑦

Linearization [Nascimento et al, 2010]:

$$\text{(LDC)} \quad \min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \cdot y_{ij}$$

s.t.

$$
\begin{array}{lll}
(1) & \displaystyle\sum_{k=1}^{M} x_{ik} = 1, & i = 1, \ldots, N \\
(2) & \displaystyle\sum_{i=1}^{N} x_{ik} \geq 1, & k = 1, \ldots, M \\
(3) & x_{ik} \in \{0, 1\}, & i = 1, \ldots, N, \; k = 1, \ldots, M \\
(4) & y_{ij} \geq x_{ik} + x_{jk} - 1, & i = 1, \ldots, N, \; j = i+1, \ldots, N, \; k = 1, \ldots, M \\
(5) & y_{ij} \geq 0, & i = 1, \ldots, N, \; j = i+1, \ldots, N.
\end{array}
$$

(4) + (5) $\Longrightarrow$ **They guarantee that $y_{ij} = 1$ if $x_{ik} = x_{jk} = 1$, i.e. $o_i, o_j \in \mathcal{O}$ are in the same cluster**

Problem Formulation, ⑧

Linearization [Nascimento et al, 2010]:

$$\text{(LDC)} \quad \min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \cdot y_{ij}$$

s.t.

$$\begin{array}{lll}
(1) & \displaystyle\sum_{k=1}^{M} x_{ik} = 1, & i = 1, \dots, N \\
(2) & \displaystyle\sum_{i=1}^{N} x_{ik} \geq 1, & k = 1, \dots, M \\
(3) & x_{ik} \in \{0, 1\}, & i = 1, \dots, N, \; k = 1, \dots, M \\
(4) & y_{ij} \geq x_{ik} + x_{jk} - 1, & i = 1, \dots, N, \; j = i+1, \dots, N, \; k = 1, \dots, M \\
(5) & y_{ij} \geq 0, & i = 1, \dots, N, \; j = i+1, \dots, N.
\end{array}$$

Note:
(LDC) has $\frac{N^2}{2}$ more variables and $\frac{N \cdot (N-1) \cdot (M+1)}{2}$ more constraints than (DC) but it is "easier".

Outline
Data Clustering
Collaborations
Description
Applications
Problem Formulation, ①
Problem Formulation, ②
Problem Formulation, ③
Problem Formulation, ④
Problem Formulation, ⑤
Problem Formulation, ⑥
Problem Formulation, ⑦
Problem Formulation, ⑧
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
Experimental results and Biological Significance
To conclude...

Graph representation

Datasets can be represented via a weighted undirected graph.

Given:

- ☞ the set of objects $\mathcal{O} = \{o_1, \ldots, o_N\}$;
- ☞ the function $d : \mathcal{O} \times \mathcal{O} \mapsto \mathbb{R}$ that assigns to each $i, j \in \mathcal{O}$ a "distance" or "similarity" $d_{ij} \in \mathbb{R}$ (usually, $d_{ij} \geq 0$, $d_{ii} = 0$, $d_{ij} = d_{ji}$, for $i, j = 1, \ldots, N$),

the following weighted undirected graph $G = (V, E, w)$ can be defined:

Graph representation

Datasets can be represented via a weighted undirected graph.

Given:

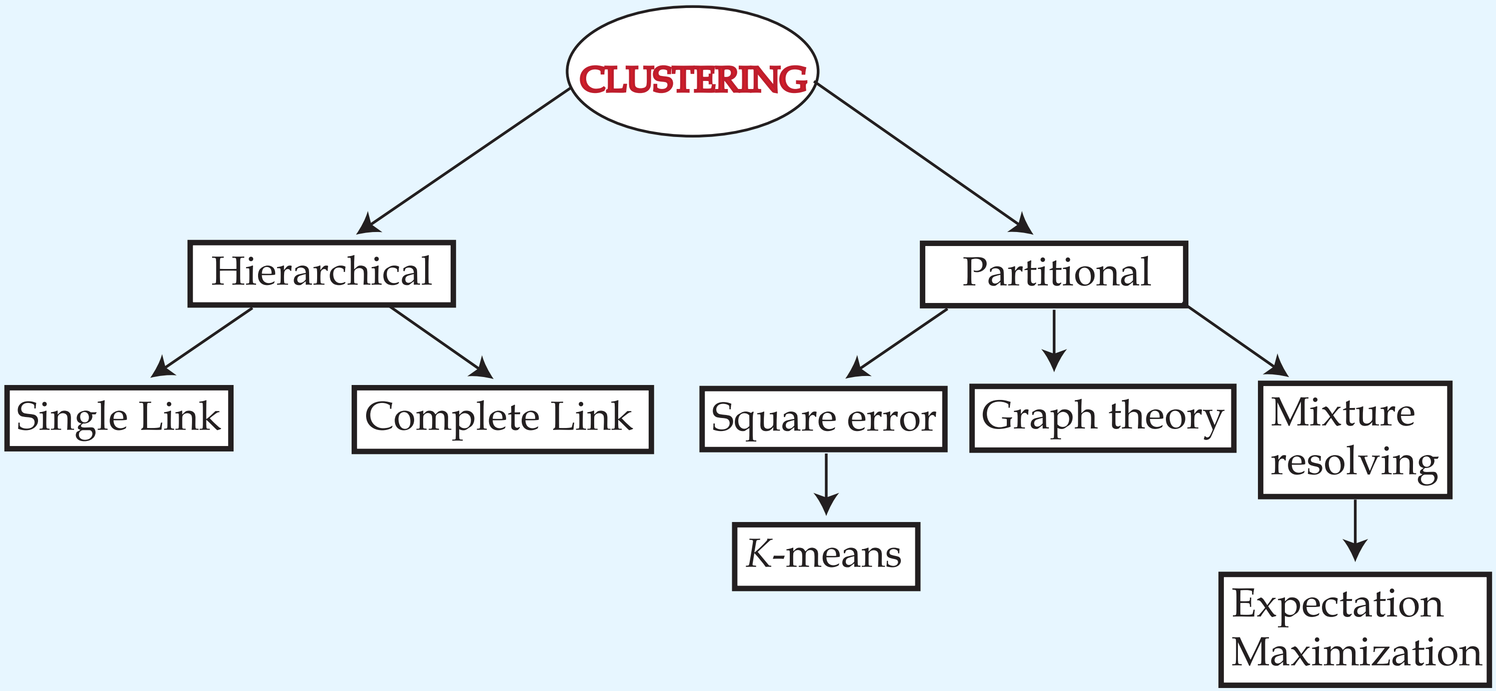
- ☞ the set of objects $\mathcal{O} = \{o_1, \dots, o_N\}$;
- ☞ the function $d : \mathcal{O} \times \mathcal{O} \mapsto \mathbb{R}$ that assigns to each $i, j \in \mathcal{O}$ a "distance" or "similarity" $d_{ij} \in \mathbb{R}$
  (usually, $d_{ij} \geq 0$, $d_{ii} = 0$, $d_{ij} = d_{ji}$, for $i, j = 1, \dots, N$),

the following weighted undirected graph $G = (V, E, w)$ can be defined:

- ❒ $V = \mathcal{O}$;
- ❒ Edges in E indicate the relationship between objects;
- ❒ $w_{ij} = d_{ij}, \forall\, i, j \in V$ (i.e., $o_i, o_j \in \mathcal{O}$).

State-of-the-art, ①

A **taxonomy of clustering approaches**:



CLUSTERING

Hierarchical

Partitional

Single Link

Complete Link

Square error

Graph theory

Mixture resolving

K-means

Expectation Maximization

Outline

Data Clustering

Collaborations

Description

Applications

Problem Formulation, ①

Problem Formulation, ②

Problem Formulation, ③

Problem Formulation, ④

Problem Formulation, ⑤

Problem Formulation, ⑥

Problem Formulation, ⑦

Problem Formulation, ⑧

Graph representation

State-of-the-art, ①

State-of-the-art, ②

State-of-the-art, ③

State-of-the-art, ④

State-of-the-art, ⑤

State-of-the-art, ⑥

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

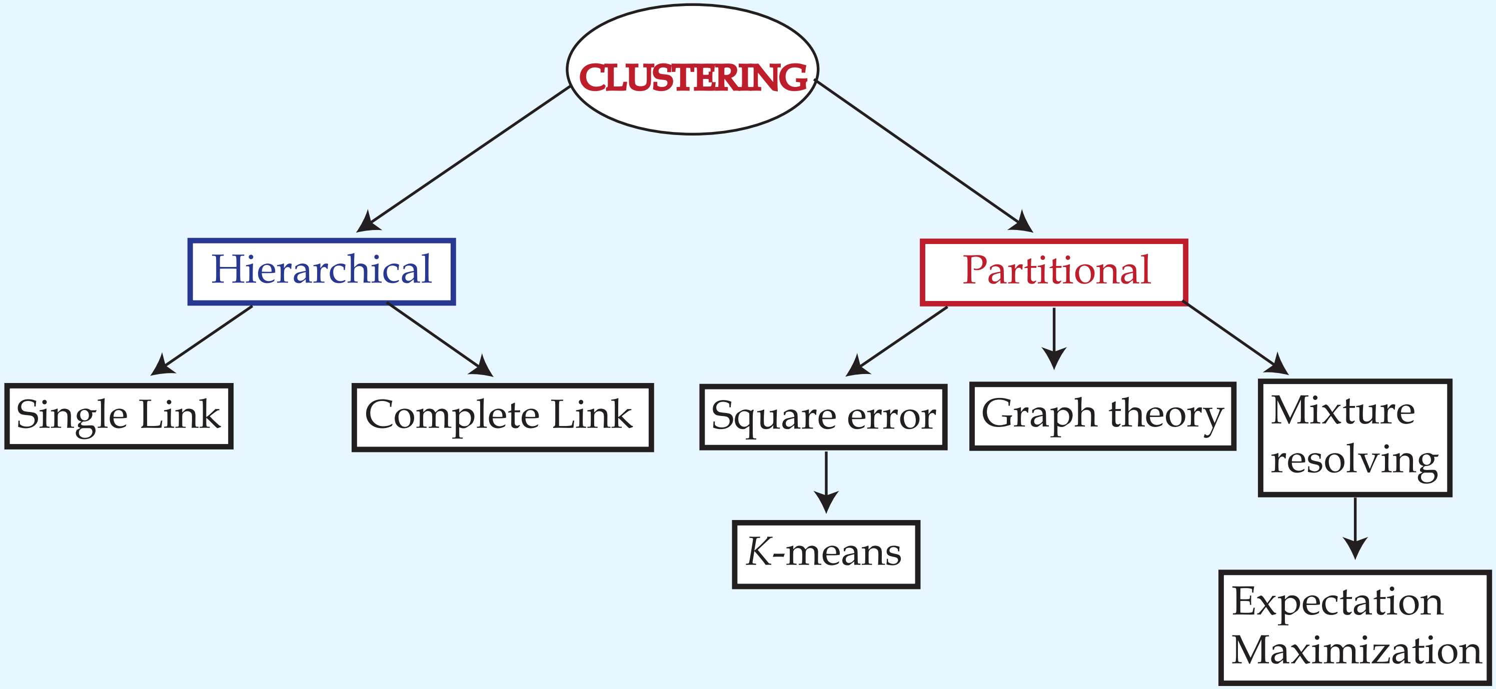
A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

State-of-the-art, ②

A **taxonomy of clustering approaches**:

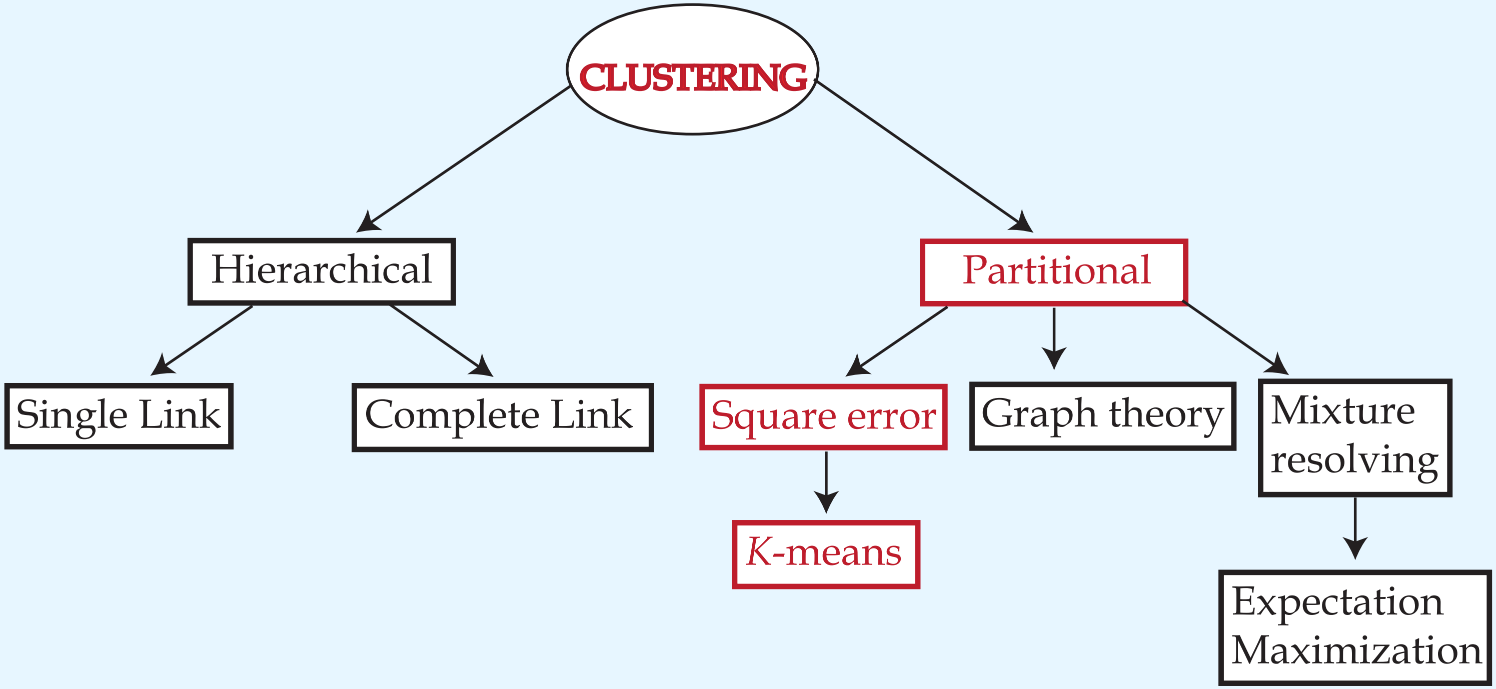


Hierarchical versus Partitioning Algorithms:

- ✔ Hierarchical methods produce a nested series of partitions;
- ✔ Partitional methods produce only one.

State-of-the-art, ③

A **taxonomy of clustering approaches**:

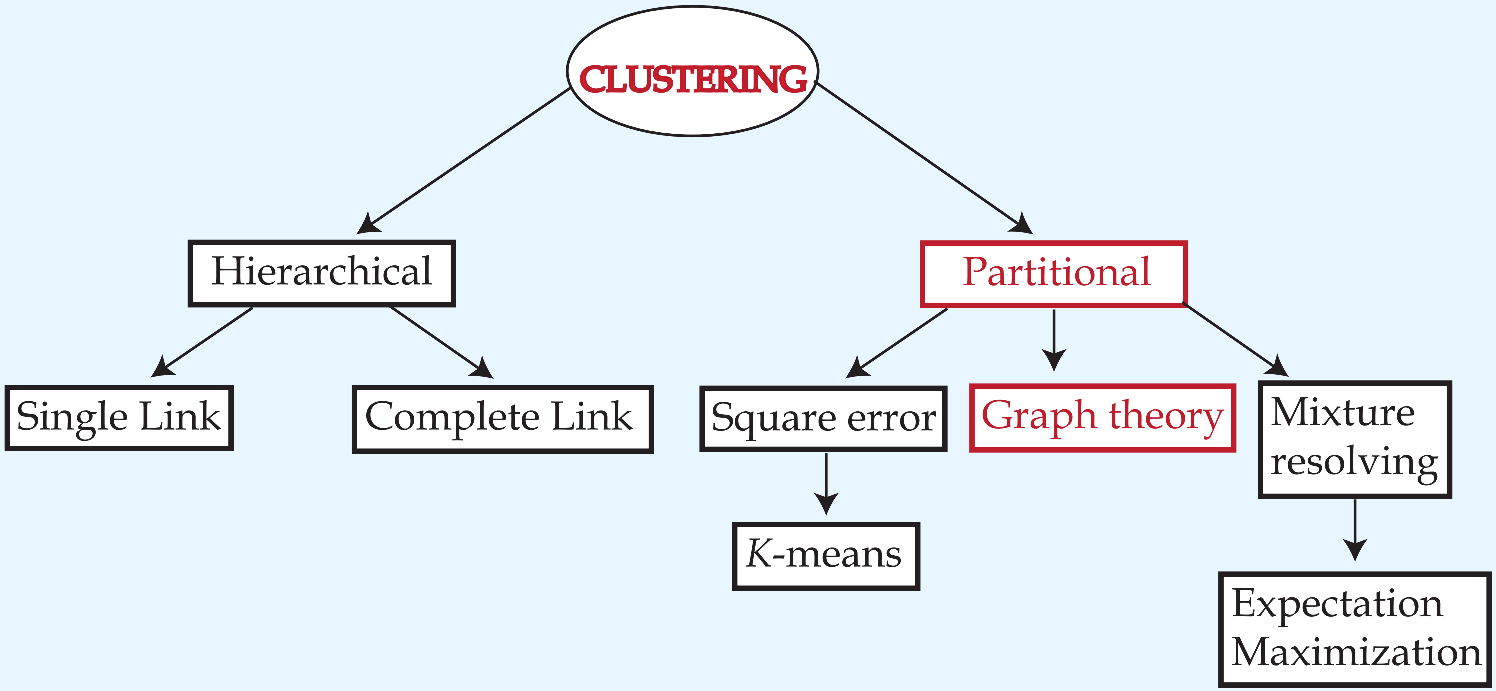


Partitional Algorithms:

- ✔ *K*-means: it starts with a random initial partition and keeps reassigning objects to "close"clusters until a convergence criterion is met.

State-of-the-art, ④

A **taxonomy of clustering approaches**:

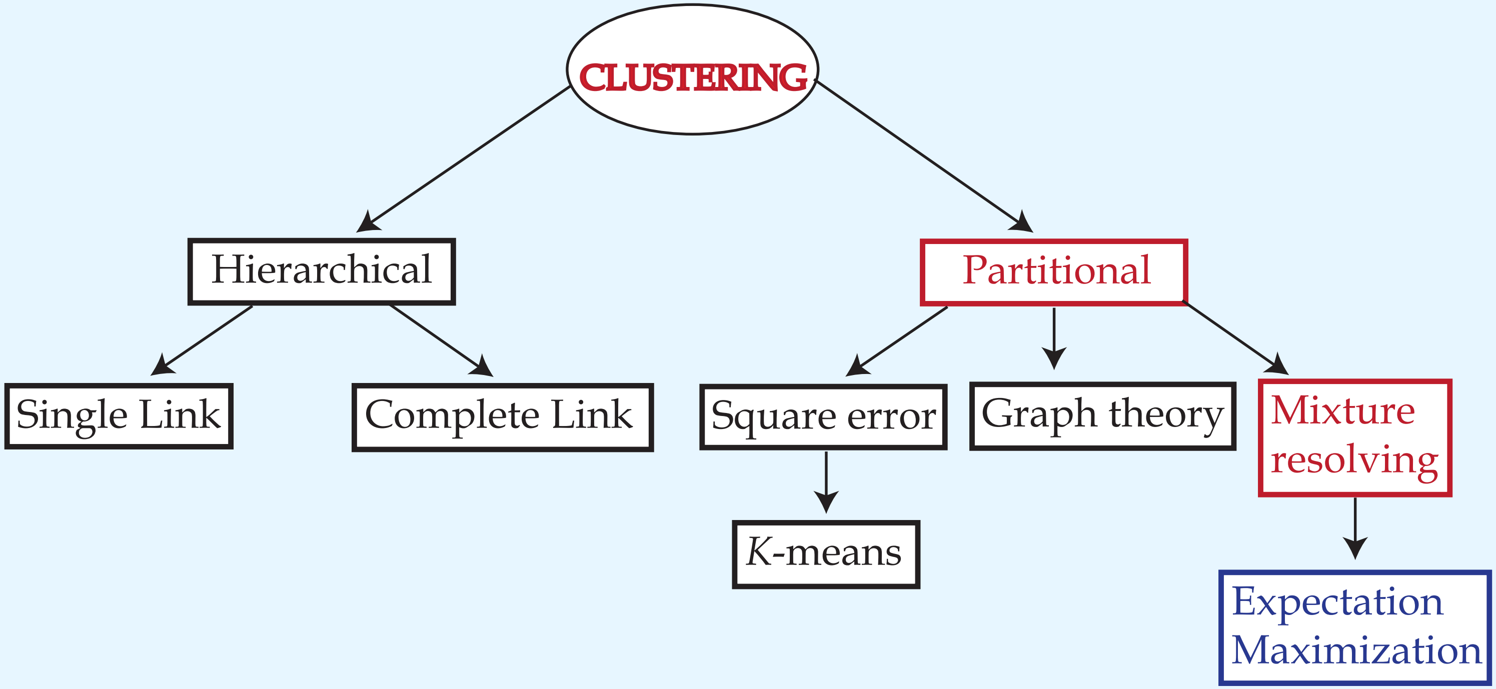


Graph-Theoretic Algorithms:

✔ They are *divisive* algorithms is based on construction of a MST of the data and then the deletion of the MST edges with the largest lengths to generate clusters.

State-of-the-art, ⑤

A **taxonomy of clustering approaches**:

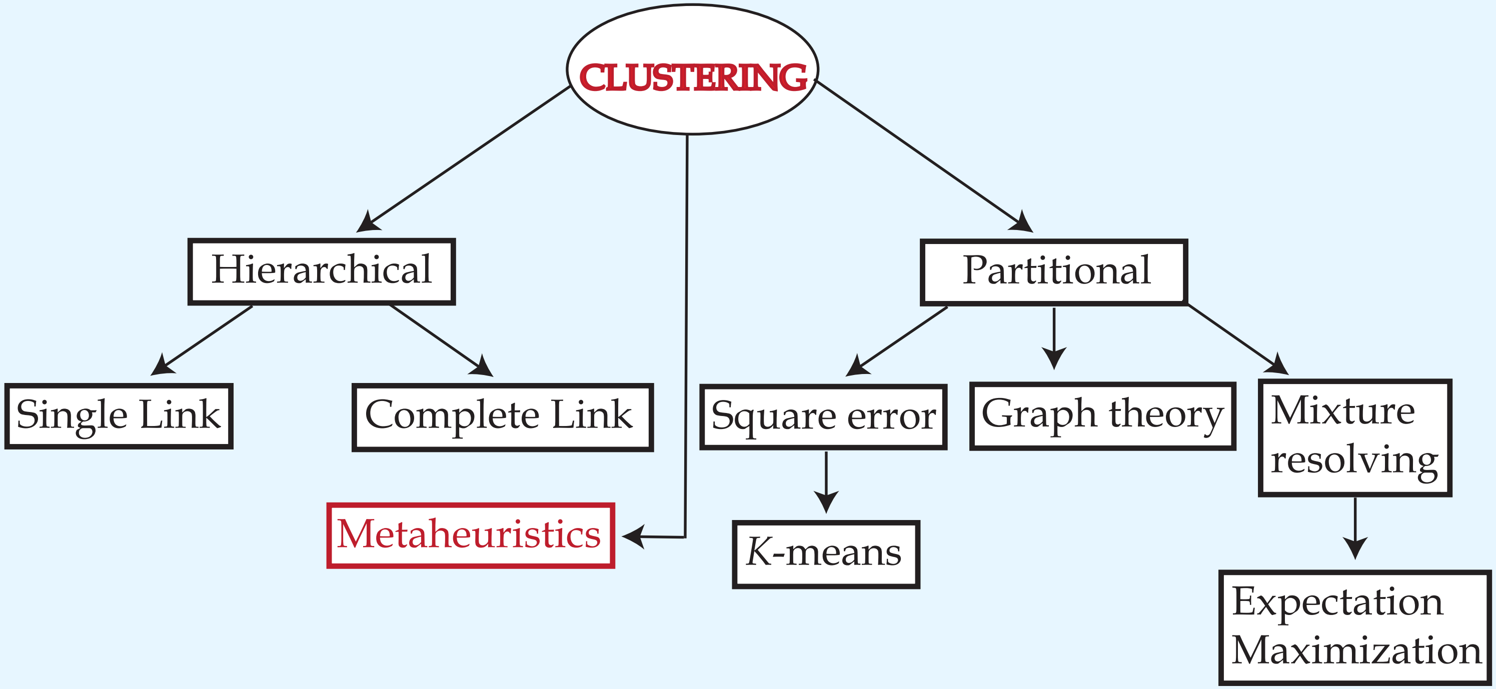


Mixture-Resolving Algorithms:

- ✔ The underlying assumption is that the objects are drawn from one of several distributions (usually, Gaussian), and the goal is to identify the parameters of each (e.g., a maximum likelihood estimate).

State-of-the-art, ⑥

A **taxonomy of clustering approaches**:



Metaheuristic approaches, including

- tabù search [Sultan, 1995];
- evolutionary algorithms [Bandyopadhyay et al, 2002; Ma et al, 2006];
- GRASP [Nascimento et al, 2010].

GRASP + Path Relinking for Data Clustering

Our proposal: GRASP + PR

Our proposal for Data Clustering: **GRASP + Path Relinking**.

Our proposal: GRASP + PR

Our proposal for Data Clustering: **GRASP + Path Relinking**.

- As Nascimento et al (2010) and graph theoretic algorithms, we have represented datasets as a weighted undirected graph $G = (V, E, w)$.
- We have been inspired by Nascimento et al.'s GRASP adopting the max number of its without improvement as stopping criterion.
- At each GRASP iteration, we apply path relinking as intensification procedure.

GRASP

GRASP (Greedy Randomized Adaptive Search Procedure) is a multi-start metaheuristic, where each iteration consists of two phases.

```
algorithm GRASP($f(\cdot)$,$g(\cdot)$,$\mathcal{N}$,Seed)
    $x_{best}$:=$\emptyset$;   $f(x_{best})$:=$+\infty$;
    while (stopping criterion not satisfied) do
        $x$:=ConstructGreedyRandomizedSolution(Seed, $g(\cdot)$);
        if ($x$ not feasible) then
            $x$:=repair($x$);
        endif
        $x$:=LocalSearch($x$, $f(\cdot)$, $\mathcal{N}$);
        if ($f(x) < f(x_{best})$) then
            $x_{best}$:=$x$;
        endif
    endwhile;
    return($x_{best}$);
end GRASP
```

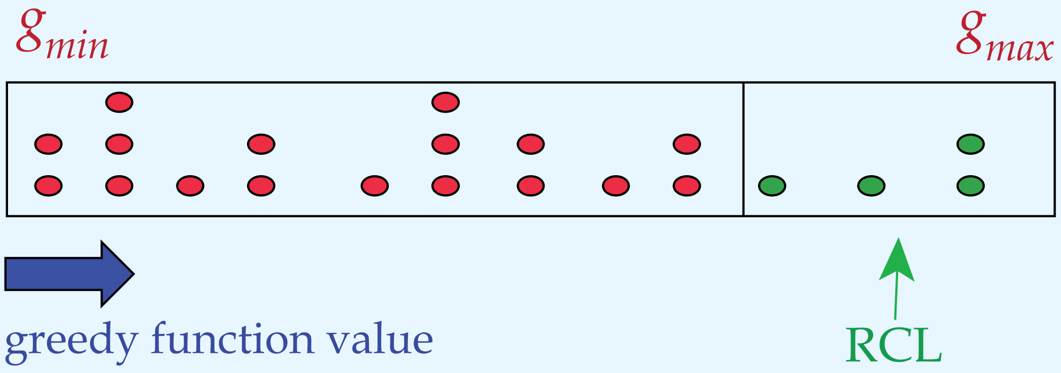
GRASP Construction, ①

In a typical iteration let S be a partial solution.

Let g_{min} and g_{max} be the smallest and the largest greedy values among the $|L|$ candidates, respectively, i.e.

$$g_{min} = \min_{e \in L} g(e), \qquad g_{max} = \max_{e \in L} g(e).$$

A restricted candidate list RCL is made up of all elements $e \in L$ with the best greedy values $g(e)$.



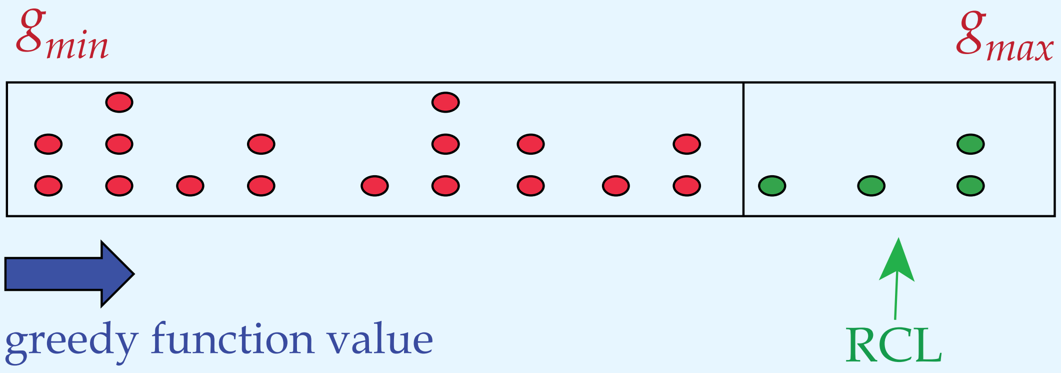
GRASP Construction, ①

In a typical iteration let $\mathcal{S}$ be a partial solution.

Let g_{min} and g_{max} be the smallest and the largest greedy values among the $|L|$ candidates, respectively, i.e.

$$g_{min} = \min_{e \in L} g(e), \qquad g_{max} = \max_{e \in L} g(e).$$

A restricted candidate list RCL is made up of all elements $e \in L$ with the best greedy values $g(e)$.



g_{min}

g_{max}

greedy function value

RCL

Random component: $e := \texttt{select}(\text{RCL}); \mathcal{S} := \mathcal{S} \cup \{e\};$

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Our proposal: GRASP + PR
GRASP
GRASP Construction, ①
GRASP Construction, ②
GRASP Construction, ③
GRASP Construction, ④
Local Search
GRASP Local search
Path relinking, ①
Path relinking, ②
Path relinking, ③
Path relinking, ③
Path relinking, ③
Path relinking, ③
Our proposal: GRASP+PR
GRASP+PR variants
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
Experimental results and Biological Significance
To conclude...

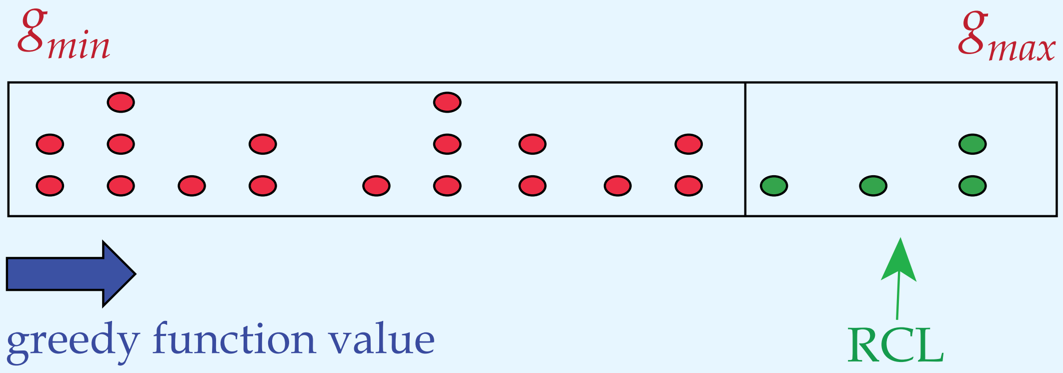
GRASP Construction, ①

In a typical iteration let $\mathcal{S}$ be a partial solution.

Let g_{min} and g_{max} be the smallest and the largest greedy values among the $|L|$ candidates, respectively, i.e.

$$g_{min} = \min_{e \in L} g(e), \qquad g_{max} = \max_{e \in L} g(e).$$

A restricted candidate list RCL is made up of all elements $e \in L$ with the best greedy values $g(e)$.



g_{min}

g_{max}

greedy function value

RCL

Random component: $e := \texttt{select}(\text{RCL}); \mathcal{S} := \mathcal{S} \cup \{e\};$

Adaptive component: greedy function values depend on the partial solution constructed so far.

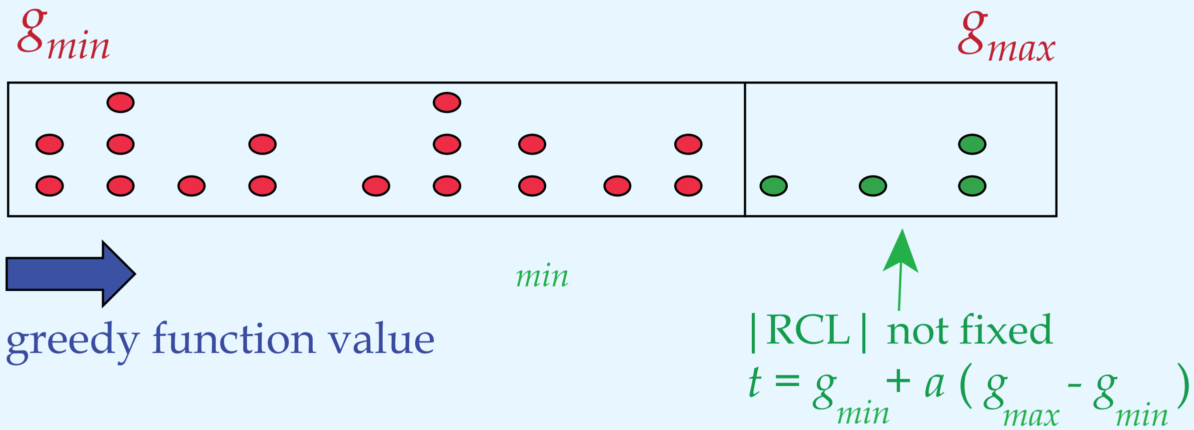
Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Our proposal: GRASP + PR
GRASP
GRASP Construction, ①
GRASP Construction, ②
GRASP Construction, ③
GRASP Construction, ④
Local Search
GRASP Local search
Path relinking, ①
Path relinking, ②
Path relinking, ③
Path relinking, ③
Path relinking, ③
Path relinking, ③
Our proposal: GRASP+PR
GRASP+PR variants
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
Experimental results and Biological Significance
To conclude...

GRASP Construction, ②

To build the RCL we have adopted a *value-based* (VB) mechanism:

RCL is associated with a parameter $a \in [0,1]$ and a threshold value $t = g_{min} + a \cdot (g_{max} - g_{min})$:

$$\text{RCL} = \{e \in L : \; g(e) \geq t\}.$$



g_{min}

g_{max}

min

greedy function value

|RCL| not fixed

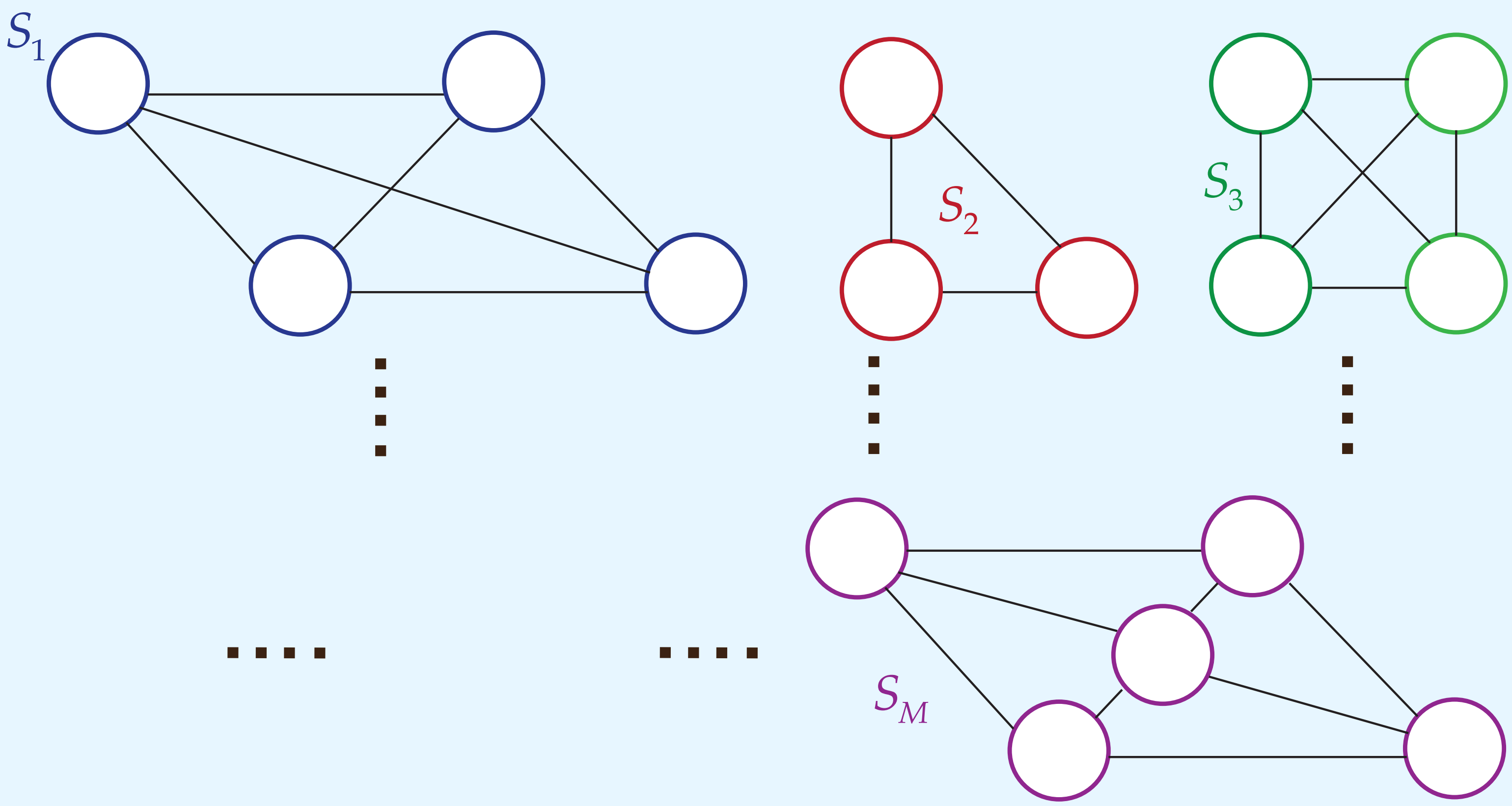
$t = g_{min} + a\,(g_{max} - g_{min})$

GRASP Construction, ③

```
procedure build-grasp-sol($N$,$M$,$\mathcal{O}$)
  $V := \mathcal{O}$;   $E := \{(i,j) \mid i,j \in V,\ i < j\}$;
  $L$ :=sort($E$);        /* w.r.t. distances/weights (non decreasing) */
    for $k = 1$ to $M - 1$ do        /* a set of $M$ clusters */
      $g_{min} := \operatorname{argmin}_{(i,j)\in L} d_{ij}$;   $g_{max} := \arg\max_{(i,j)\in L} d_{ij}$;
      $a$ :=select([0, 1]);   $t := g_{max} + a \cdot (g_{min} - g_{max})$;
      RCL:= $\{(i,j) \in L \mid d_{ij} \geq t\}$;   $(i,j)$ :=select(RCL);
      $S_i := S_j := \emptyset$;
      for each $v \in V$ s.t. $(v,i)$, $(v,j) \in E$ do
        if ($d_{vi} < d_{vj}$) then $S_i := S_i \cup \{v\}$;
        else $S_j := S_j \cup \{v\}$;
      endfor
      for each $u_i \in S_i$ and $u_j \in S_j$ do
        $E := E \setminus \{(u_i, u_j)\}$;   $L := L \setminus \{(u_i, u_j)\}$;
      endfor
    endfor
  return ($V$, $E$);
end build-grasp-sol
```

GRASP Construction, ④

Output of `build-grasp-sol`: a **set of M clusters**.



Local Search

To define **local search**, one needs to specify a local neighborhood structure $N(\mathcal{S})$ of a solution $\mathcal{S}$:

$$N(\mathcal{S}) = \{\overline{\mathcal{S}} \mid \overline{\mathcal{S}} \text{ is an elementary modification of } \mathcal{S}\}.$$

Local Search

To define **local search**, one needs to specify a local neighborhood structure $N(\mathcal{S})$ of a solution $\mathcal{S}$:

$$N(\mathcal{S}) = \{\overline{\mathcal{S}} \mid \overline{\mathcal{S}} \text{ is an elementary modification of } \mathcal{S}\}.$$

A generic local search algorithm

① takes as input a solution $\mathcal{S}$ that is considered as *current solution* $\overline{\mathcal{S}}$;

Local Search

To define **local search**, one needs to specify a local neighborhood structure $N(\mathcal{S})$ of a solution $\mathcal{S}$:

$$N(\mathcal{S}) = \{\overline{\mathcal{S}} \mid \overline{\mathcal{S}} \text{ is an elementary modification of } \mathcal{S}\}.$$

A generic local search algorithm

① takes as input a solution $\mathcal{S}$ that is considered as *current solution* $\overline{\mathcal{S}}$;

② iteratively, explores $\mathcal{N}(\overline{\mathcal{S}})$:

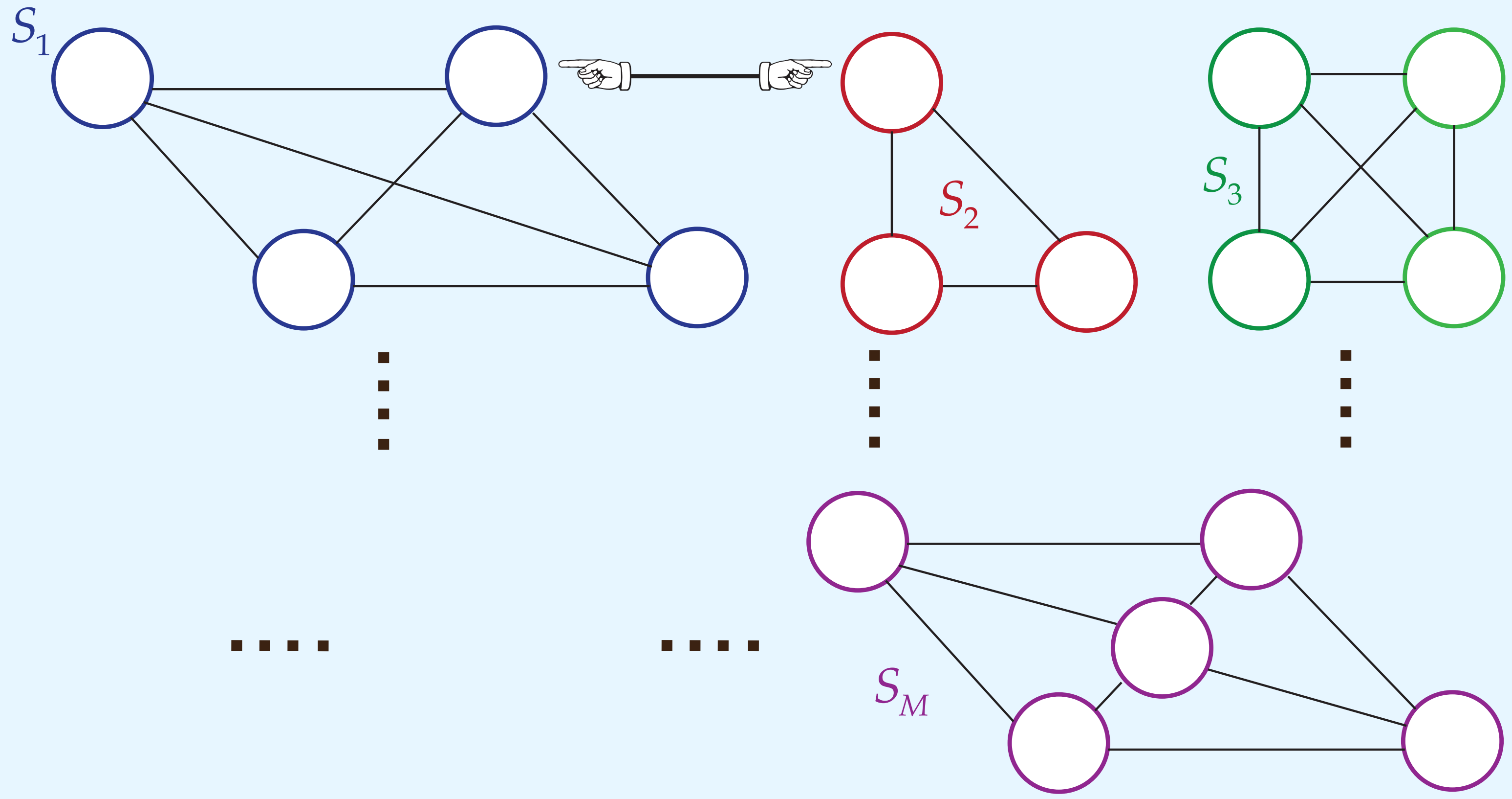
- ✍ if there exists $\hat{\mathcal{S}} \in \mathcal{N}(\overline{\mathcal{S}})$ better than $\overline{\mathcal{S}}$, then $\overline{\mathcal{S}} := \hat{\mathcal{S}}$ and the procedure continues exploring $\mathcal{N}(\overline{\mathcal{S}})$;
- ✍ otherwise, it outputs a *locally optimal solution* $\overline{\mathcal{S}}$.

Computational complexity of each iteration: $O(|\mathcal{N}(\overline{\mathcal{S}})|)$.

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Our proposal: GRASP + PR
GRASP
GRASP Construction, ①
GRASP Construction, ②
GRASP Construction, ③
GRASP Construction, ④
Local Search
GRASP Local search
Path relinking, ①
Path relinking, ②
Path relinking, ③
Path relinking, ③
Path relinking, ③
Path relinking, ③
Our proposal: GRASP+PR
GRASP+PR variants
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
Experimental results and Biological Significance
To conclude...

GRASP Local search

Modification of $\mathcal{S}$ consists of transferring an object from a cluster to another one in order to improve the solution:

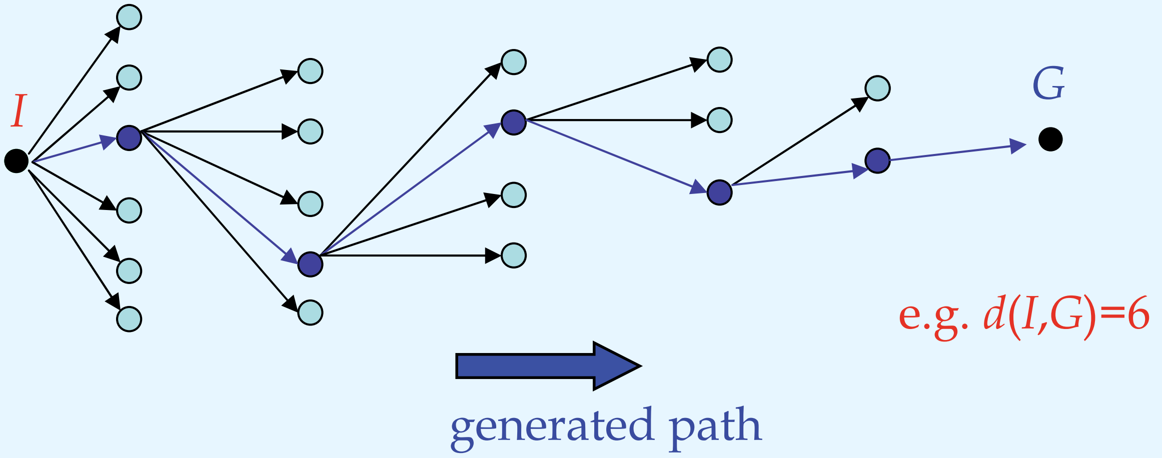


Path relinking, ①

It consists in exploring trajectories that connect high quality solutions (members of a "small" population P, called Elite Set).

Path is generated by selecting modifications (moves) that introduce attributes of the guiding solution G in the initial solution I.

At each step, all moves ($d(I, G)$) that incorporate attributes of the guiding solution are analyzed and best move is taken.



Path relinking, ②

It consists in exploring trajectories that connect high quality solutions (members of a "small" population P, called Elite Set).

Path is generated by selecting modifications (moves) that introduce attributes of the guiding solution G in the initial solution I.

At each step, all moves ($d(I, G)$) that incorporate attributes of the guiding solution are analyzed and best move is taken.

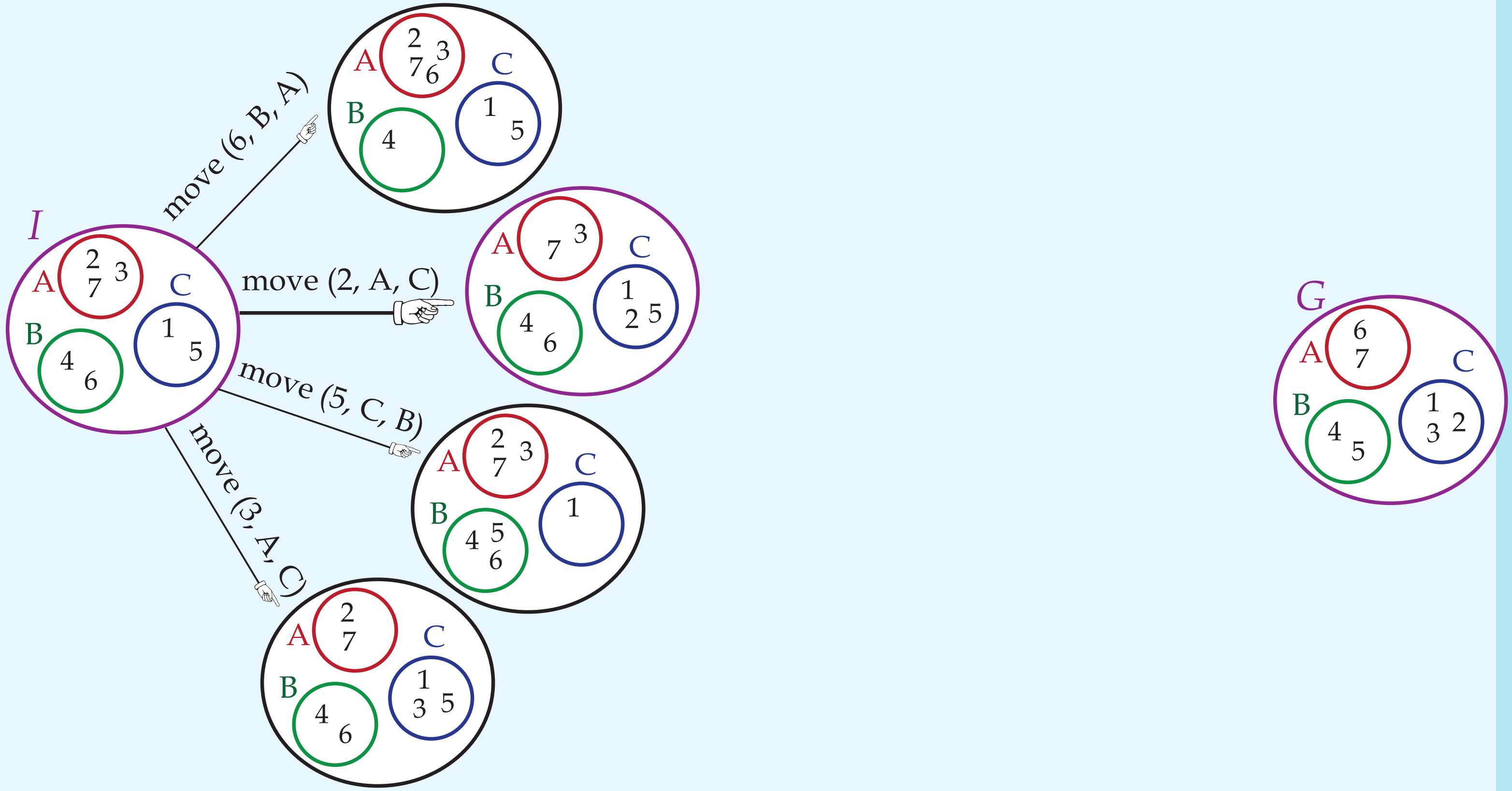
Theorem.
For any instance $\mathcal{I}$ of (DC) and for any pair of solutions I and G for $\mathcal{I}$ such that $d(I, G) = k$ there exists at least one path

$$\mathcal{P}_{I,G} = \{I = w^0, w^1, \ldots, w^k = G\}$$

connecting I to G in the solution space.

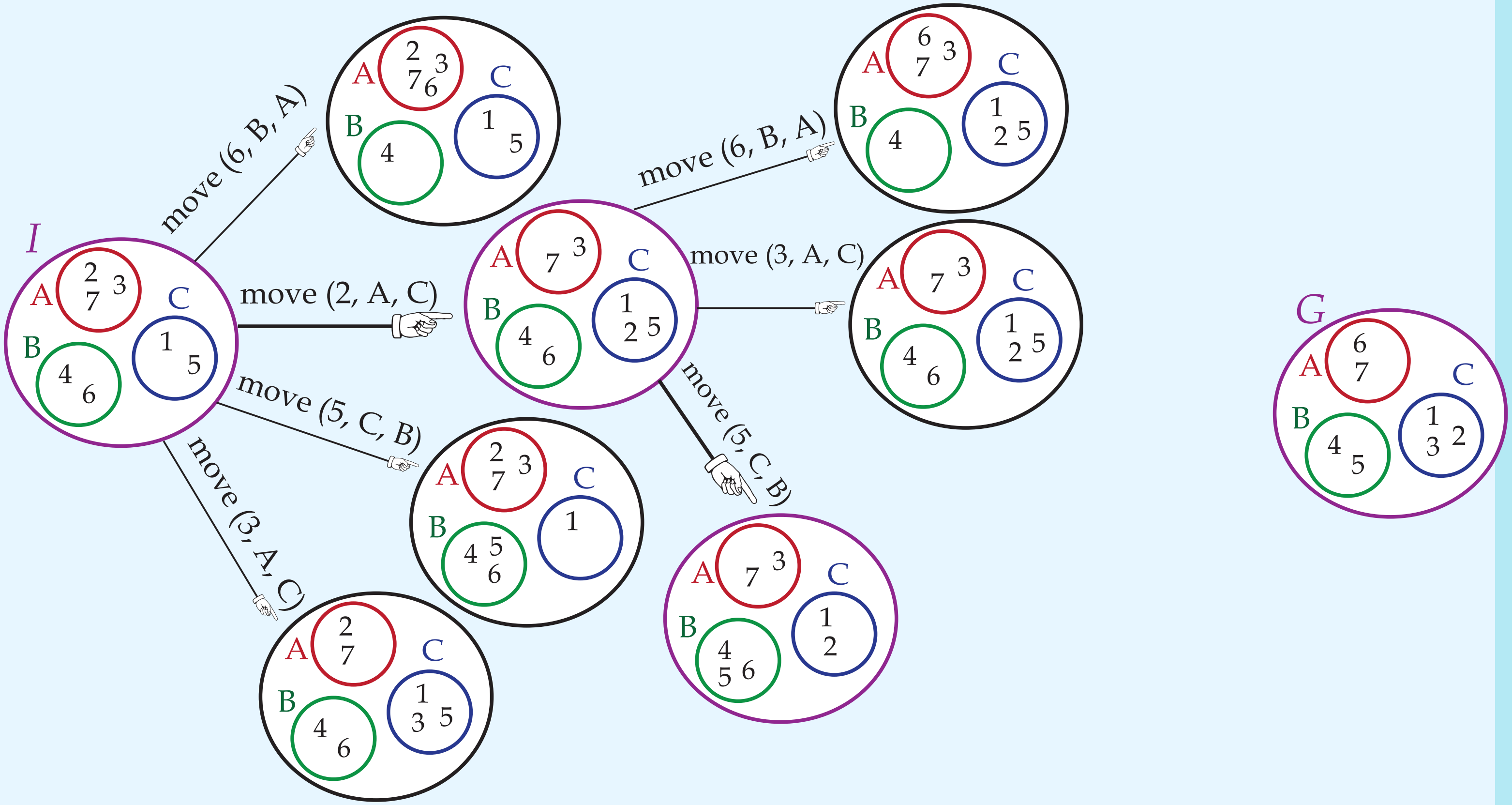
Path relinking, ③

Path relinking for Data Clustering:



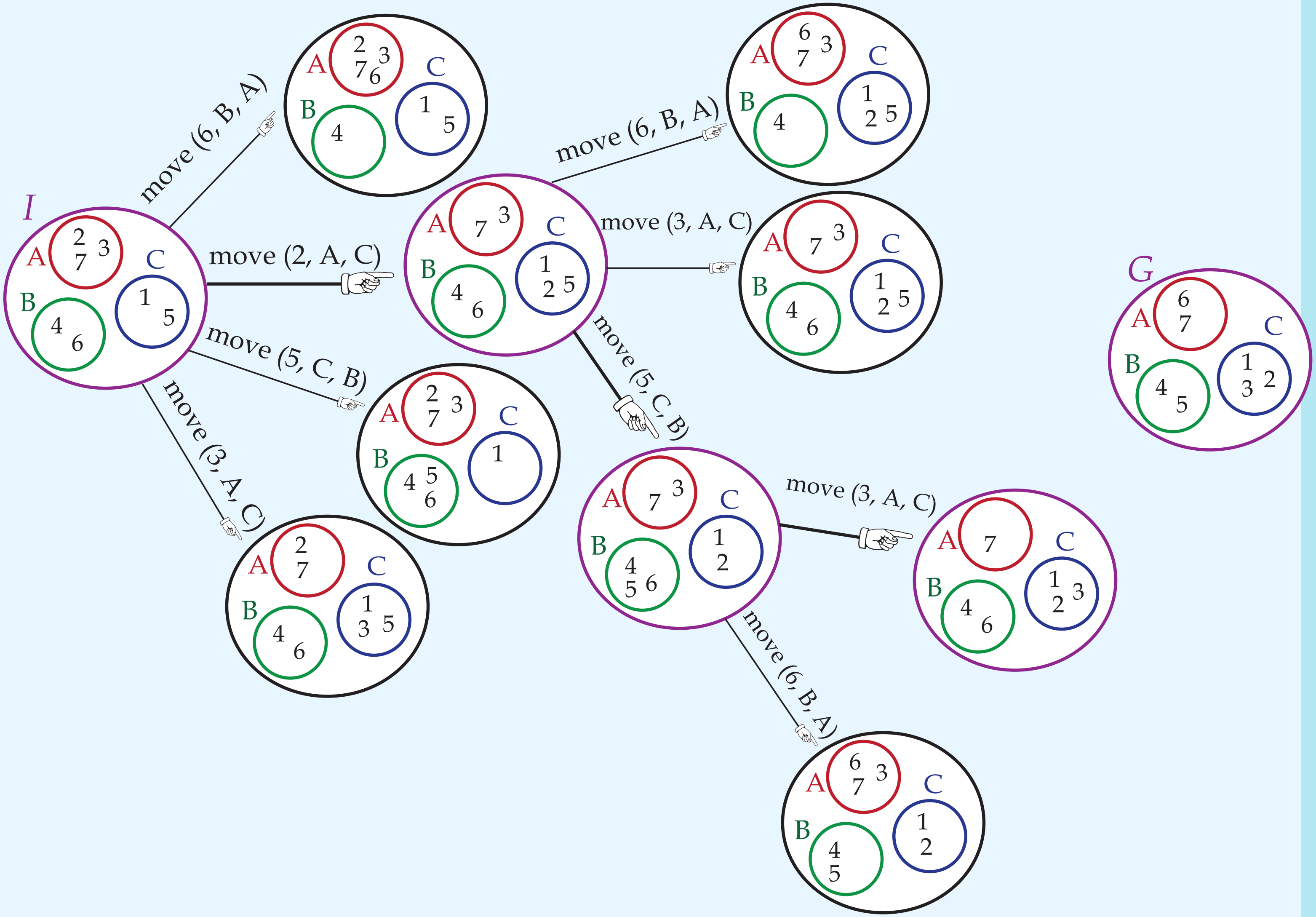
Path relinking, ③

Path relinking for (DC):



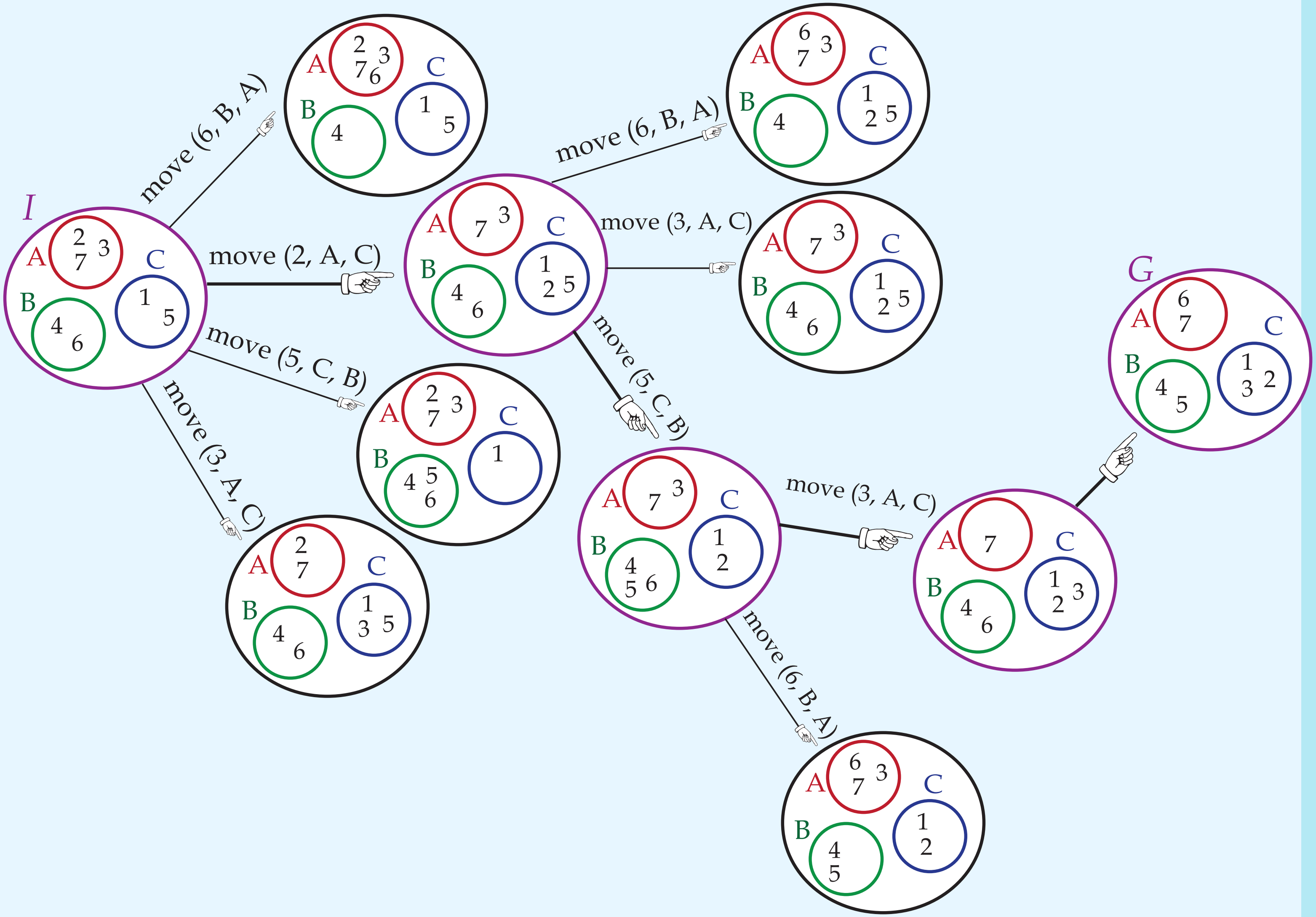
Path relinking, ③

Path relinking for (DC):



Path relinking, ③

Path relinking for (DC):



Our proposal: GRASP+PR

At each GRASP iteration, we apply path relinking as intensification.

```
algorithm GRASP+PR($f(\cdot)$,$g(\cdot)$,$\mathcal{N}$,Seed)
  $P := \emptyset$;
  while (stopping criterion not satisfied) do
    $\mathcal{S}$:=ConstructGreedyRandomizedSolution(Seed, $g(\cdot)$);
    $\mathcal{S}$:=LocalSearch($\mathcal{S}$, $f(\cdot)$, $\mathcal{N}$);
    if ($P$ not full) then $P := P \cup \{\mathcal{S}\}$;
     else
       $\hat{\mathcal{S}}$ :=select($P$);   $\hat{\mathcal{S}}$ :=path-relinking($\mathcal{S}$,$\hat{\mathcal{S}}$);
       update($P$,$\hat{\mathcal{S}}$);
    endif
  endwhile;
  $\mathcal{S}_{best}$ :=select-best($P$);
  return($\mathcal{S}_{best}$);
end GRASP+PR
```

Our proposal: GRASP+PR

At each GRASP iteration, we apply path relinking as intensification.

```
algorithm GRASP+PR($f(\cdot)$,$g(\cdot)$,$\mathcal{N}$,Seed)
  $P := \emptyset$;
  while (stopping criterion not satisfied) do
    $\mathcal{S}$:=ConstructGreedyRandomizedSolution(Seed, $g(\cdot)$);
    $\mathcal{S}$:=LocalSearch($\mathcal{S}$, $f(\cdot)$, $\mathcal{N}$);
    if ($P$ not full) then $P := P \cup \{\mathcal{S}\}$;
     else
       $\hat{\mathcal{S}}$ :=select($P$);   $\hat{\mathcal{S}}$ :=path-relinking($\mathcal{S}$,$\hat{\mathcal{S}}$);
       update($P$,$\hat{\mathcal{S}}$);
    endif
  endwhile;
  $\mathcal{S}_{best}$ :=select-best($P$);
  return($\mathcal{S}_{best}$);
end GRASP+PR
```

update(P,$\hat{\mathcal{S}}$):

$P := P \cup \{\hat{\mathcal{S}}\}$, if $\hat{\mathcal{S}}$ better than the worst elite solution and sufficiently different from all elite solutions.

GRASP+PR variants

Several **different GRASP+PR variants** have been designed:

☞ a forward path relinking:

$$\texttt{worst}(\mathcal{S},\hat{\mathcal{S}}) \quad \overset{\texttt{path-relinking}}{\Longrightarrow} \quad \texttt{best}(\mathcal{S},\hat{\mathcal{S}})$$

☞ a backward path relinking:

$$\texttt{worst}(\mathcal{S},\hat{\mathcal{S}}) \quad \overset{\texttt{path-relinking}}{\Longleftarrow} \quad \texttt{best}(\mathcal{S},\hat{\mathcal{S}})$$

☞ a mixed relinking:

$$\texttt{worst}(\mathcal{S},\hat{\mathcal{S}}) \quad \overset{\texttt{path-relinking}}{\Longrightarrow} \quad \overline{\mathcal{S}} \overset{\texttt{path-relinking}}{\Longleftarrow} \quad \texttt{best}(\mathcal{S},\hat{\mathcal{S}})$$

☞ a randomized relinking: instead of selecting the best yet unselected move, randomly selects one from among a candidate list with the most promising moves in the path being investigated.

Experimental results on Biological Data

Tested algorithms

- 3 known clustering algorithms:
  - `K-means`: deterministic, minimizes the dissimilarities between an object and the centroid of its cluster;
  - `K-medians`: deterministic, minimizes the dissimilarities between an object and the medoid of its cluster;
  - `PAM`: deterministic, 2 stages: ① BUILD: defines a set of initial *medoids*; ② SWAP: tunes the medoids by swapping objects between the clusters;
- `GRASP-L`: Nascimento et al, 2010;
- `GRASP`: our implementation of `GRASP-L`;
- GRASP+PR variants:
  - `GRASP-PRf`: GRASP + PR forward;
  - `GRASP-PRb`: GRASP + PR backward;
  - `GRASP-PRm`: GRASP + PR mixed;
  - `GRASP-PRrnd`: GRASP + PR greedy randomized.

Distance (dissimilarity) metrics, ①

For all algorithms, we used the **same distance (dissimilarity) metrics** between 2 objects using their attribute values:

- Euclidean: $d_{ij} = \sqrt{\sum_{k=1}^{L}(a_{ik} - a_{jk})^2}$;
- City-block or Manhattan (city road grid): $d_{ij} = \sum_{k=1}^{L}|a_{ik} - a_{jk}|$;

Distance (dissimilarity) metrics, ①

For all algorithms, we used the **same distance (dissimilarity) metrics** between 2 objects using their attribute values:

- ❍ Euclidean: $d_{ij} = \sqrt{\sum_{k=1}^{L} (a_{ik} - a_{jk})^2}$;

- ❍ City-block or Manhattan (city road grid): $d_{ij} = \sum_{k=1}^{L} |a_{ik} - a_{jk}|$;

- ❍ Cosine or uncentered correlation: $D_{ij} \in [-1, 1]$

$$d_{ij} = 1 - |D_{ij}|, \quad D_{ij} = \frac{\sum_{k=1}^{L} a_{ik} \cdot a_{jk}}{\sum_{k=1}^{L} a_{ik}^2 \sum_{k=1}^{L} a_{jk}^2};$$

Note:

- ❒ $D_{ij} = 1 \Longrightarrow$ angle 0^{o};
- ❒ $D_{ij} = -1 \Longrightarrow$ angle 90^{o}.

Outline

Data Clustering

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Tested algorithms

Distance (dissimilarity) metrics, ①

Distance (dissimilarity) metrics, ②

Test environment

Datasets, ①

Datasets, ②

Experimental Design, ①

Experimental Design, ②

Numerical results, ①

Numerical results, ②

Numerical results, ③

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

Distance (dissimilarity) metrics, ②

For all algorithms, we used the **same distance (dissimilarity) metrics** between 2 objects using their attribute values:

- Euclidean: $d_{ij} = \sqrt{\sum_{k=1}^{L} (a_{ik} - a_{jk})^2};$
- City-block or Manhattan (city road grid): $d_{ij} = \sum_{k=1}^{L} |a_{ik} - a_{jk}|;$
- Cosine or uncentered correlation: $D_{ij} \in [-1, 1];$
- Pearson's correlation: $d_{ij} = 1 - |r_{ij}|;\ r_{ij} \in [-1, 1]$

$$r_{ij} = \frac{L \cdot \sum_{k=1}^{L} a_{ik} \cdot a_{jk} - \sum_{k=1}^{L} a_{ik} \cdot a_{jk}}{\sqrt{L \cdot \sum_{k=1}^{L} a_{ik}^2 - (\sum_{k=1}^{L} a_{jk})^2}\sqrt{L \cdot \sum_{k=1}^{L} a_{jk}^2 - (\sum_{k=1}^{L} a_{jk})^2}}.$$

Note:

- ❒ $r_{ij} = 1 \Longrightarrow$ perfect association;
- ❒ $r_{ij} = -1 \Longrightarrow$ perfect negative linear relationship.

Outline

Data Clustering

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Tested algorithms

Distance (dissimilarity) metrics, ①

Distance (dissimilarity) metrics, ②

Test environment

Datasets, ①

Datasets, ②

Experimental Design, ①

Experimental Design, ②

Numerical results, ①

Numerical results, ②

Numerical results, ③

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

Test environment

- Dell computer with Core 2 Duo 2.1 GHz T8100 Intel processor and 3 Gb of memory;
- Windows XP Professional version 5.1 2002 SP3 x86;
- Java language, Javac compiler ver.1.6.0.20;
- Random-number generator: Mersenne Twister algorithm (Matsumoto and Nishimura, 1998) from the COLT2 library.

Datasets, ①

Datasets:

① fold protein classification: Protein [Ding et al, 2001];

② prediction of protein localization sites: Yeast [Nakai et al, 1991];

③ 7 cancer diagnosis data sets:

- ✓ Breast [Bennett et al, 1992];
- ✓ Novartis [Su et al, 2002];
- ✓ BreastA [Veer et al, 2002];
- ✓ BreastB [West et al, 2001];
- ✓ DLBCLA [Monti et al, 2005];
- ✓ DLBCLB [Rosenwald et al, 2002];
- ✓ MultiA [Su et al, 2002];

④ a benchmark dataset: Iris [Fisher et al, 1936].

Datasets, ②

Characteristics of datasets used in the experiments.

Data Set	N	# Structures (M)	# Attributes
Protein	698	2 (4,27)	125
Yeast	1484	1 (10)	8
Breast	699	2 (2,8)	9
Novartis	103	1 (4)	1000
BreastA	98	1 (3)	1213
BreastB	49	2 (2,4)	1213
DLBCLA	141	1 (3)	661
DLBCLB	180	1 (3)	661
MultiA	103	1 (4)	5565
Iris	140	1 (3)	4

Outline

Data Clustering

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Tested algorithms

Distance (dissimilarity) metrics, ①

Distance (dissimilarity) metrics, ②

Test environment

Datasets, ①

Datasets, ②

Experimental Design, ①

Experimental Design, ②

Numerical results, ①

Numerical results, ②

Numerical results, ③

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

Experimental Design, ①

Tuning phase – values of the parameters for GRASP+PR heuristics used for each dataset:

Pool size (PS), elements in pool before start PR (EPBS), symmetrical difference (SD), and Iterations without Improvement (IWI).

	Iris	Novartis	BrstA	BrstB1	BrstB2	DLBCLA
PS	3	5	4	3	3	5
EPBS	1	3	1	1	1	2
SD	4	70	4	30	30	100
IWI	15	15	15	15	15	15

	DLBCLB	MultA	Brst1	Brst2	Prt1	Prt2	Yeast
PS	5	5	3	6	5	5	7
EPBS	2	2	1	3	2	3	3
SD	100	70	4	550	450	450	1200
IWI	15	15	15	15	15	15	5

Experimental Design, ②

Measure to evaluate the results:

- ✔ CRand – Corrected (adjusted) Rand index [Hubert and Arabie, 1985].

To compare 2 partitions P and Q on the same set X, compute

$$\mathrm{CRand}(P, Q) = \frac{r - \mathrm{Exp}(r)}{\mathrm{Max}(r) - \mathrm{Exp}(r)},$$

where

- ✗ r is the number of common joined pairs in P and Q;
- ✗ $\mathrm{Exp}(r)$ is the expected value of r;
- ✗ $\mathrm{Max}(r)$ is the maximum value of r.

Numerical results, ①

Euclidean distance

Out of 10 datasets

- ✓ GRASP-PRrnd found best results for 9;
- ✓ GRASP-PRb found best results for 8;
- ✓ GRASP-PRm found best results for 8;
- ✓ GRASP-PRf found best results for 6;
- ✓ GRASP found best results for 6;
- ✓ GRASP-L for 2;
- ✓ K-medians found the best solution for 2;
- ✓ K-means found the best solution for only 1.

Numerical results, ②

City-block or Manhattan distance

Out of 10 datasets

- ✓ `GRASP-PRrnd` found best results for 8;
- ✓ `GRASP-PRb` found best results for 8;
- ✓ `GRASP-PRm` found best results for 8;
- ✓ `GRASP-PRf` found best results for 7;
- ✓ `GRASP` found best results for 6;
- ✓ `GRASP-L` for 2;
- ✓ `K-medians` found the best solution for 2;
- ✓ `K-means` found the best solution for only 1.

Numerical results, ③

Cosine distance

Out of 10 datasets

- ✓ GRASP-PRrnd found best results for 6;
- ✓ GRASP-PRb found best results for 6;
- ✓ GRASP-PRf found best results for 6;
- ✓ GRASP-PRm found best results for 5;
- ✓ GRASP found best results for 4;
- ✓ GRASP-L for 2;
- ✓ K-medians found the best solution for 4;
- ✓ K-means found the best solution for only 1.

Data BiClustering

Collaborations

Material on Data Biclustering presented in this seminar is based on joint work with:

✔ Angelo Facchiano

Institute of Food Science – CNR, Italy

✔ Francesco Musacchia

Dept. of Mathematics and Applications “R. Caccioppoli”
University of Napoli FEDERICO II

✔ Anna Marabotti and Luciano Milanesi

Institute of Biomedical Technologies – CNR, Italy

Description and Applications

Input: the input data comes from two domain sets and some relation over the Cartesian product of these two sets is given.

Task: to partition each of the sets s.t.

- ✔ the subsets from one domain exhibit similar behavior across the subsets of the other domain, or, in other words,
- ✔ simultaneously, data clustering and feature selection.

Description and Applications

Input: the input data comes from two domain sets and some relation over the Cartesian product of these two sets is given.

Task: to partition each of the sets s.t.

- ✔ the subsets from one domain exhibit similar behavior across the subsets of the other domain, or, in other words,
- ✔ simultaneously, data clustering and feature selection.

As Clustering, **applications** include

- ➪ galaxy formation;
- ➪ image segmentation;
- ➪ ...;
- ➪ biological data.

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
Collaborations
Description and Applications
Problem Formulation, ①
Problem Formulation, ②
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥
State-of-the-art, ⑦
A new GRASP-like algorithm for Data Biclustering
Experimental results and Biological Significance
To conclude...

Problem Formulation, ①

We are given a gene expression matrix $\mathcal{A} \in \mathbb{R}^{n \times m}$

$$\mathcal{A} = \left[\begin{array}{c|ccccc} & \text{Condition } 1 & \cdots & \text{Condition } j & \cdots & \text{Condition } m \\ \hline \text{Gene } 1 & a_{11} & \cdots & a_{1j} & \cdots & a_{1m} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \text{Gene } i & a_{i1} & \cdots & a_{ij} & \cdots & a_{im} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \text{Gene } n & a_{n1} & \cdots & a_{nj} & \cdots & a_{nm} \end{array}\right],$$

where a_{ij} represents the expression level of gene i under condition j.

Outline

Data Clustering

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

Collaborations

Description and Applications

Problem Formulation, ①

Problem Formulation, ②

Graph representation

State-of-the-art, ①

State-of-the-art, ②

State-of-the-art, ③

State-of-the-art, ④

State-of-the-art, ⑤

State-of-the-art, ⑥

State-of-the-art, ⑦

A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

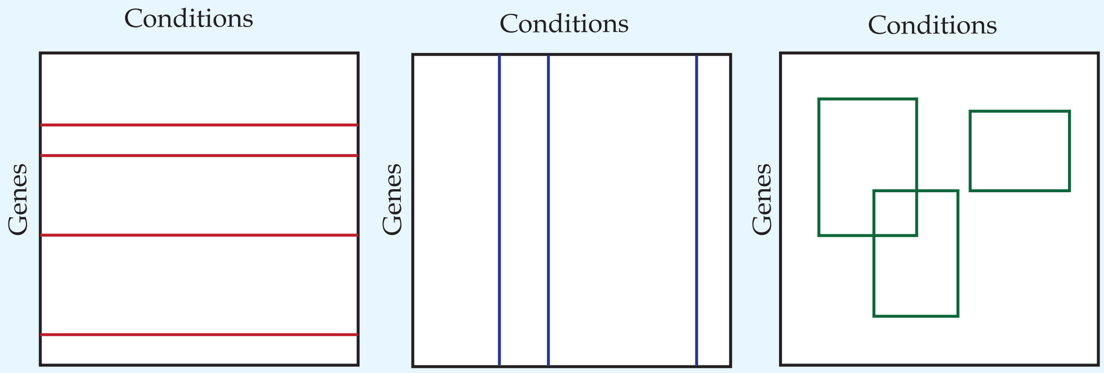
Problem Formulation, ①

We are given a gene expression matrix $\mathcal{A} \in \mathbb{R}^{n \times m}$

$$\mathcal{A} = \left[\begin{array}{c|ccccc} & \text{Condition } 1 & \cdots & \text{Condition } j & \cdots & \text{Condition } m \\ \hline \text{Gene } 1 & a_{11} & \cdots & a_{1j} & \cdots & a_{1m} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \text{Gene } i & a_{i1} & \cdots & a_{ij} & \cdots & a_{im} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \text{Gene } n & a_{n1} & \cdots & a_{nj} & \cdots & a_{nm} \end{array}\right],$$

where a_{ij} represents the expression level of gene i under condition j.

Goal of biclustering:
to identify subgroups of genes and subgroups of conditions, by performing simultaneous clustering of both n **rows** and m **columns**.



Gene clusters

Condition clusters

Biclusters

Problem Formulation, ②

We considered the **general case of a data matrix** $\mathcal{A} = (X, Y)$, where

☞ $X = \{x_1, \ldots, x_n\}$ is the set of rows;

☞ $Y = \{y_1, \ldots, y_m\}$ is the set of columns, and

☞ the element a_{ij}, $i \in X$, $j \in Y$, corresponds to a value representing the relation between row i and column j.

Problem Formulation, ②

We considered the **general case of a data matrix** $\mathcal{A} = (X, Y)$, where

☞ $X = \{x_1, \ldots, x_n\}$ is the set of rows;

☞ $Y = \{y_1, \ldots, y_m\}$ is the set of columns, and

☞ the element a_{ij}, $i \in X$, $j \in Y$, corresponds to a value representing the relation between row i and column j.

Definitions:

➜ a *cluster of rows* $\mathcal{A}_{IY}$ is a $k \times m$ submatrix of $\mathcal{A}$, where $I = \{x_{i_1}, \ldots, x_{i_k}\} \subseteq X$, i.e. it is a subset of $k \leq n$ rows defined over the set of all columns Y;

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
Collaborations
Description and Applications
Problem Formulation, ①
Problem Formulation, ②
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥
State-of-the-art, ⑦
A new GRASP-like algorithm for Data Biclustering
Experimental results and Biological Significance
To conclude...

Problem Formulation, ②

We considered the **general case of a data matrix $\mathcal{A} = (X, Y)$**, where

- ☞ $X = \{x_1, \ldots, x_n\}$ is the set of rows;
- ☞ $Y = \{y_1, \ldots, y_m\}$ is the set of columns, and
- ☞ the element a_{ij}, $i \in X$, $j \in Y$, corresponds to a value representing the relation between row i and column j.

Definitions:

- ➜ a *cluster of rows* $\mathcal{A}_{IY}$ is a $k \times m$ submatrix of $\mathcal{A}$, where $I = \{x_{i_1}, \ldots, x_{i_k}\} \subseteq X$, i.e. it is a subset of $k \leq n$ rows defined over the set of all columns Y;
- ➜ a *cluster of columns* $\mathcal{A}_{XJ}$ is a $n \times s$ submatrix of $\mathcal{A}$, where $J = \{y_{j_1}, \ldots, y_{j_s}\} \subseteq Y$, i.e. it is a subset of $s \leq m$ columns defined over the set of all rows X;

Problem Formulation, ②

We considered the **general case of a data matrix** $\mathcal{A} = (X, Y)$, where

☞ $X = \{x_1, \ldots, x_n\}$ is the set of rows;

☞ $Y = \{y_1, \ldots, y_m\}$ is the set of columns, and

☞ the element a_{ij}, $i \in X$, $j \in Y$, corresponds to a value representing the relation between row i and column j.

Definitions:

➜ a *cluster of rows* $\mathcal{A}_{IY}$ is a $k \times m$ submatrix of $\mathcal{A}$, where $I = \{x_{i_1}, \ldots, x_{i_k}\} \subseteq X$, i.e. it is a subset of $k \leq n$ rows defined over the set of all columns Y;

➜ a *cluster of columns* $\mathcal{A}_{XJ}$ is a $n \times s$ submatrix of $\mathcal{A}$, where $J = \{y_{j_1}, \ldots, y_{j_s}\} \subseteq Y$, i.e. it is a subset of $s \leq m$ columns defined over the set of all rows X;

➜ a *bicluster* $\mathcal{B} = \mathcal{A}_{IJ}$ is a $k \times s$ submatrix of $\mathcal{A}$, where $I = \{x_{i_1}, \ldots, x_{i_k}\} \subseteq X$ and $J = \{y_{j_1}, \ldots, y_{j_s}\} \subseteq Y$, i.e. it is a subset of $k \leq n$ rows defined over a subset of $s \leq m$ columns or, equivalently, **a subset of $s \leq m$ columns defined over a subset of $k \leq n$ rows**.

Graph representation

Data matrices can be naturally represented via a complete weighted bipartite graph $G = (V, E, w)$:

- ☞ $V = X \cup Y$ (clearly, $X \cap Y = \emptyset$);
- ☞ $E = \{[x_i, y_j] \mid x_i \in X,\ y_j \in Y\}$;
- ☞ $w:\ E \mapsto \mathbb{R}$ s.t. $\forall\, [x_i, y_j] \in E$, $w_{ij} = a_{ij} \in \mathbb{R}$.

Graph representation

Data matrices can be naturally represented via a complete weighted bipartite graph $G = (V, E, w)$:

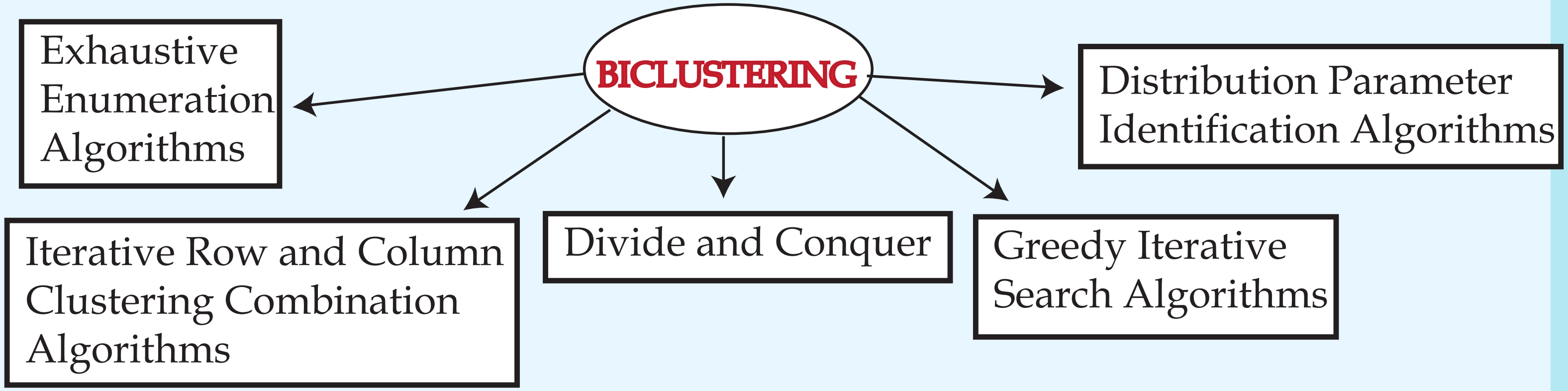
- ☞ $V = X \cup Y$ (clearly, $X \cap Y = \emptyset$);
- ☞ $E = \{[x_i, y_j] \mid x_i \in X,\ y_j \in Y\}$;
- ☞ $w : E \mapsto \mathbb{R}$ s.t. $\forall\ [x_i, y_j] \in E$, $w_{ij} = a_{ij} \in \mathbb{R}$.

Bad new: even in its **simplest form where** $\mathcal{A} \in \{0,1\}^{n \times m}$, the problem of finding a maximum size bicluster in a data matrix $\mathcal{A}$ is **NP**-complete.

In fact, it reduces to finding the maximum edge biclique in the corresponding bipartite graph G.
[Peeters, 2003]

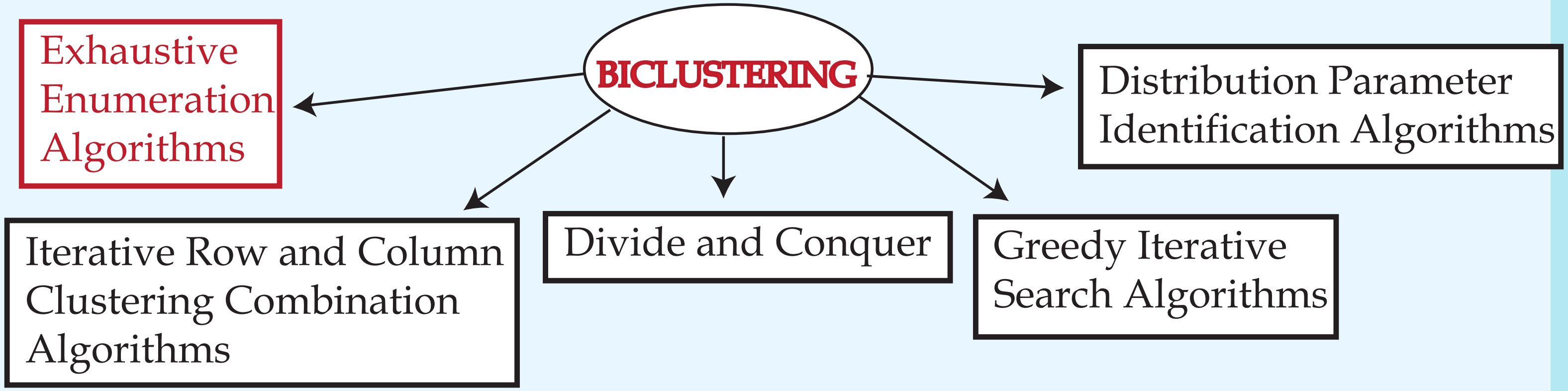
State-of-the-art, ①

A **taxonomy of biclustering approaches**:


BICLUSTERING
Exhaustive Enumeration Algorithms
Distribution Parameter Identification Algorithms
Iterative Row and Column Clustering Combination Algorithms
Divide and Conquer
Greedy Iterative Search Algorithms

State-of-the-art, ②

A **taxonomy of biclustering approaches**:

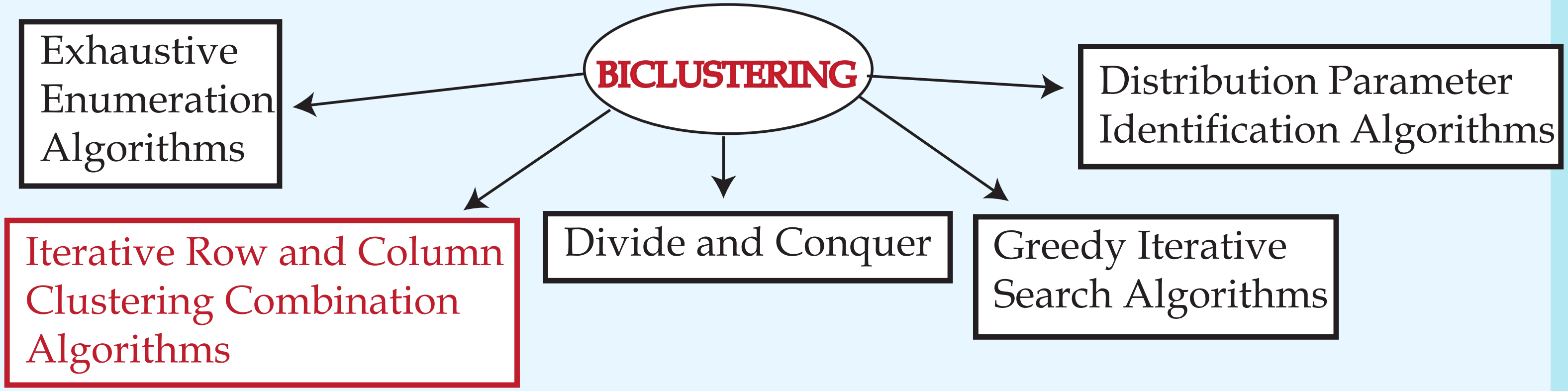

BICLUSTERING
Exhaustive Enumeration Algorithms
Distribution Parameter Identification Algorithms
Iterative Row and Column Clustering Combination Algorithms
Divide and Conquer
Greedy Iterative Search Algorithms

Exhaustive enumeration algorithms:

- ✔ they exhaustively search in the input matrix the best biclusters with very high computational running times.
  `[Tanay, Sharan, and Shamir, 2002]`
  `[Wang, Wang, Yang, and Yu, 2002]`

State-of-the-art, ③

A **taxonomy of biclustering approaches**:


BICLUSTERING
Exhaustive Enumeration Algorithms
Distribution Parameter Identification Algorithms
Iterative Row and Column Clustering Combination Algorithms
Divide and Conquer
Greedy Iterative Search Algorithms

Iterative row and column clustering combination algorithms:

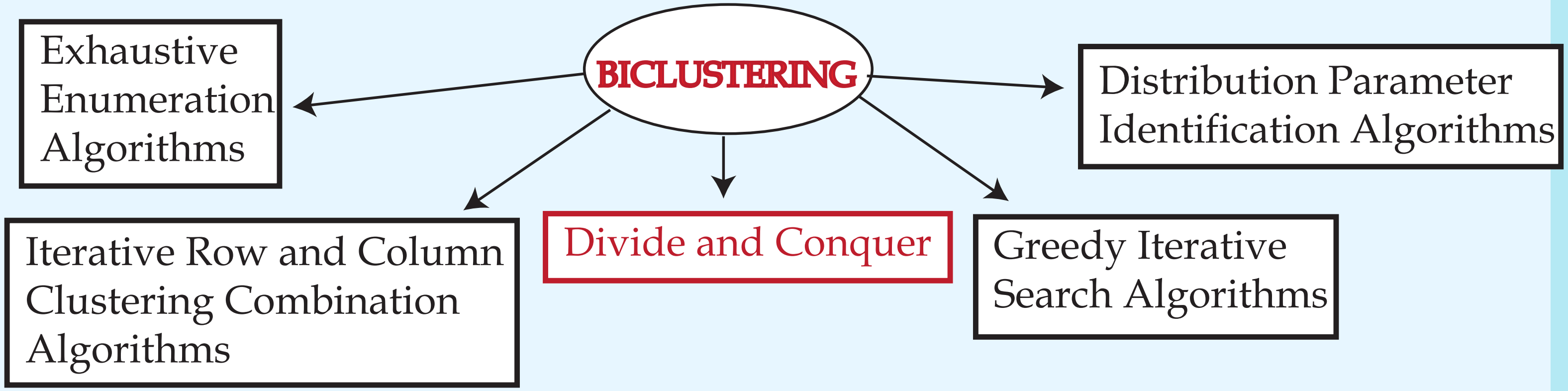
✔ they first apply separately clustering algorithms to the rows and columns of the data matrix and then combine the results using some sort of iterative procedure.

[Getz, Levine, and Domany, 2000]

[Tang, Zhang, Zhang, and Ramanathan, 2001]

State-of-the-art, ④

A **taxonomy of biclustering approaches**:


BICLUSTERING
Exhaustive Enumeration Algorithms
Distribution Parameter Identification Algorithms
Iterative Row and Column Clustering Combination Algorithms
Divide and Conquer
Greedy Iterative Search Algorithms

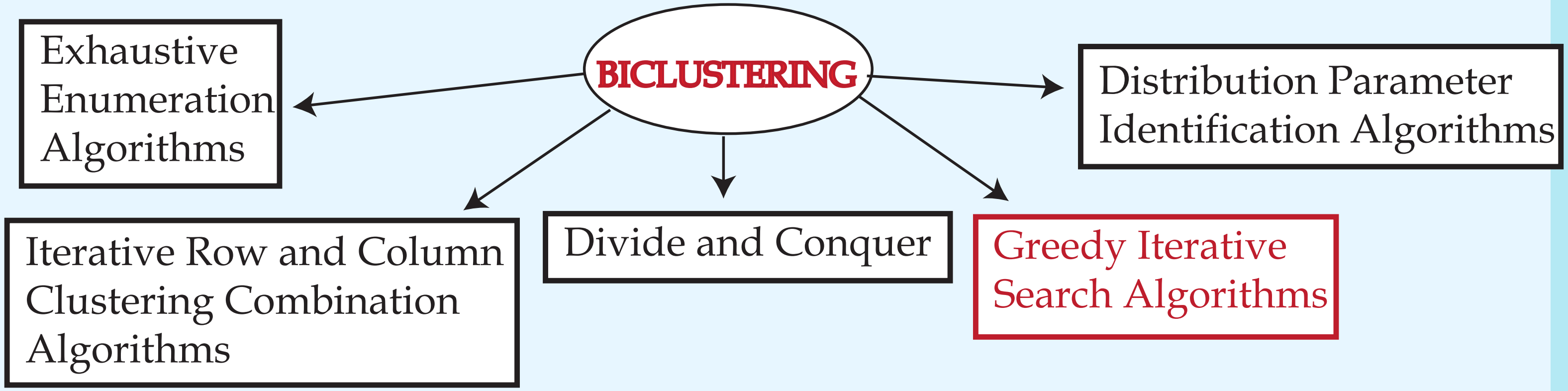
Divide and conquer algorithms:

- ✔ they divide the problem in subproblems and are potentially very fast but usually split good biclusters before they can be identified.

  [Duffy and Quiroz, 1991]

State-of-the-art, ⑤

A **taxonomy of biclustering approaches**:



BICLUSTERING

Exhaustive Enumeration Algorithms

Distribution Parameter Identification Algorithms

Iterative Row and Column Clustering Combination Algorithms

Divide and Conquer

Greedy Iterative Search Algorithms

Greedy iterative search algorithms:

- ✔ based on the steepest descent idea, they create biclusters by adding and/or removing rows and columns optimizing a local gain criterion.
  `[Yang, Wang, Wang, and Yu, 2002, 2003]`
  `[Cho, Dhillon, Guan, and Sra, 2004]`

Outline

Data Clustering

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering
Collaborations
Description and Applications
Problem Formulation, ①
Problem Formulation, ②
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥
State-of-the-art, ⑦

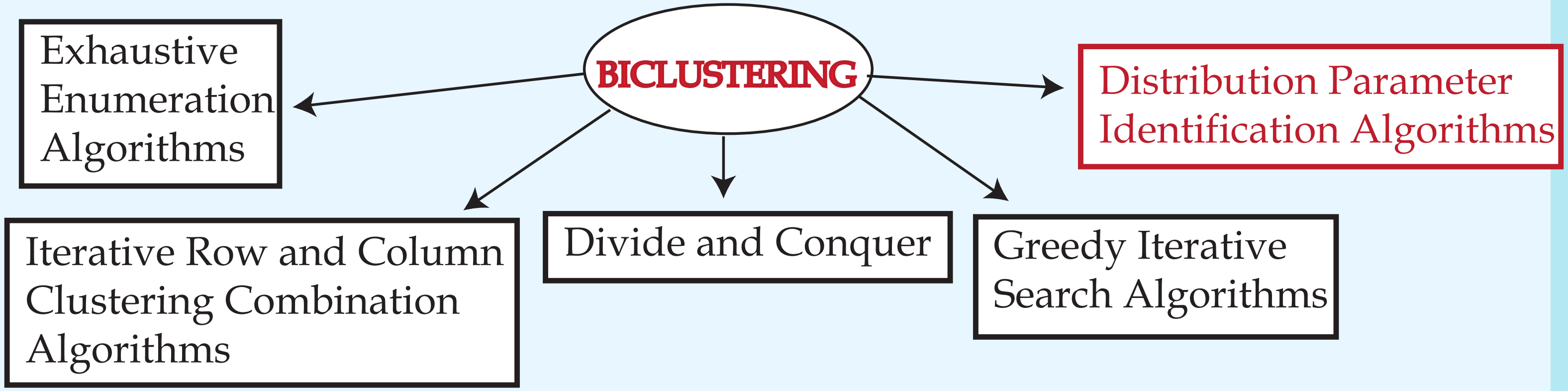
A new GRASP-like algorithm for Data Biclustering

Experimental results and Biological Significance

To conclude...

State-of-the-art, ⑥

A **taxonomy of biclustering approaches**:

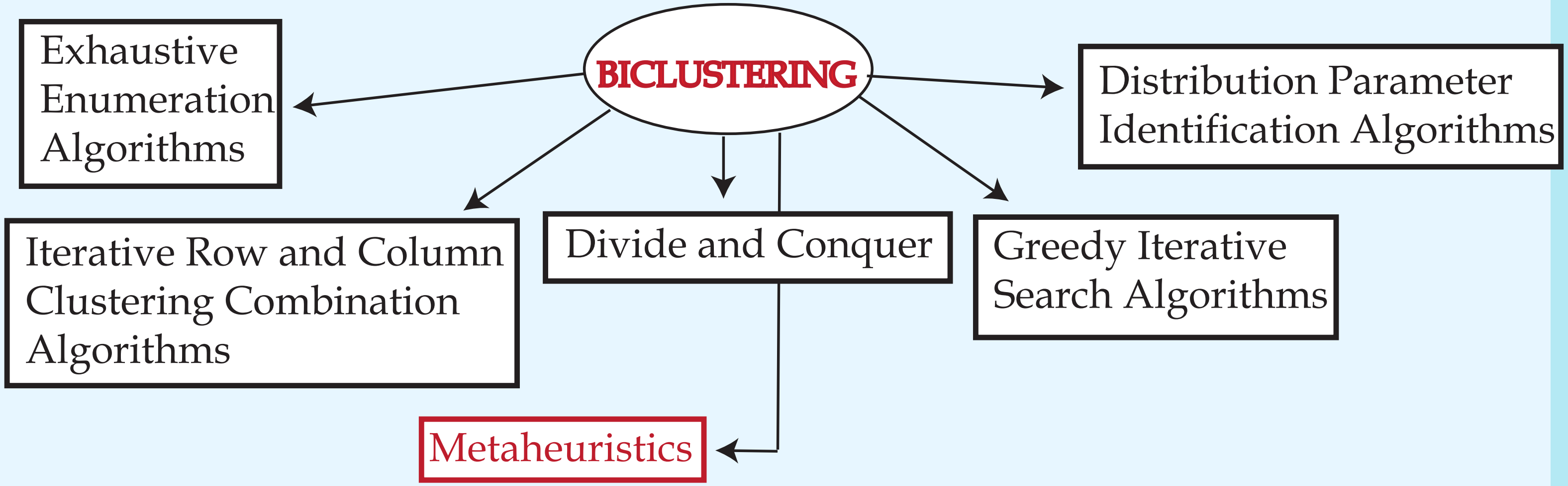

BICLUSTERING
Exhaustive Enumeration Algorithms
Distribution Parameter Identification Algorithms
Iterative Row and Column Clustering Combination Algorithms
Divide and Conquer
Greedy Iterative Search Algorithms

Distribution parameter identification algorithms:

- ✔ they try to identify the distribution parameters used to generate the data. `[Klugar, Basri, Chang, and Gerstein, 2003]` `[Sheng, Moreau, and De Moor, 2003]`

State-of-the-art, ⑦

A **taxonomy of biclustering approaches**:


BICLUSTERING
Exhaustive Enumeration Algorithms
Distribution Parameter Identification Algorithms
Iterative Row and Column Clustering Combination Algorithms
Divide and Conquer
Greedy Iterative Search Algorithms
Metaheuristics

Metaheuristic approaches:

- ✔ a Simulated Annealing [Bryan, Cunningham, and Bolshakova, 2006];
- ✔ a Genetic Algorithm [Mitra and Banka, 2006];
- ✔ a Reactive GRASP [Dharan and Nair, 2009].

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
Collaborations
Description and Applications
Problem Formulation, ①
Problem Formulation, ②
Graph representation
State-of-the-art, ①
State-of-the-art, ②
State-of-the-art, ③
State-of-the-art, ④
State-of-the-art, ⑤
State-of-the-art, ⑥
State-of-the-art, ⑦
A new GRASP-like algorithm for Data Biclustering
Experimental results and Biological Significance
To conclude...

A new GRASP-like algorithm for Data Biclustering

A solution and objective function

Given a gene expression matrix $\mathcal{A} \in \mathbb{R}^{n \times m}$ s.t. a_{ij} represents the expression level of gene i under condition j,

a **solution** is **a set of biclusters**

$$\{\mathcal{B}_1 = (I_1, J_1), \ldots, \mathcal{B}_k = (I_k, J_k)\}$$

s.t. **each bicluster** $\mathcal{B}_q$, $q = 1, \ldots, k$, satisfies some **specific characteristics of "homogeneity"**.

A solution and objective function

Given a gene expression matrix $\mathcal{A} \in \mathbb{R}^{n \times m}$ s.t. a_{ij} represents the expression level of gene i under condition j,

a **solution** is **a set of biclusters**

$$\{\mathcal{B}_1 = (I_1, J_1), \ldots, \mathcal{B}_k = (I_k, J_k)\}$$

s.t. **each bicluster** $\mathcal{B}_q$, $q = 1, \ldots, k$, satisfies some **specific characteristics of "homogeneity"**.

In our approach, we wanted

- to analyze directly the numeric values in the data matrix $\mathcal{A}$ and
- try to find subsets of rows and subsets of columns with similar behaviors;

A solution and objective function

Given a gene expression matrix $\mathcal{A} \in \mathbb{R}^{n \times m}$ s.t. a_{ij} represents the expression level of gene i under condition j,

a **solution** is **a set of biclusters**

$$\{\mathcal{B}_1 = (I_1, J_1), \ldots, \mathcal{B}_k = (I_k, J_k)\}$$

s.t. **each bicluster** $\mathcal{B}_q$, $q = 1, \ldots, k$, satisfies some **specific characteristics of “homogeneity”**.

In our approach, we wanted

- to analyze directly the numeric values in the data matrix $\mathcal{A}$ and
- try to find subsets of rows and subsets of columns with similar behaviors;
- according to `[Cheng and Church, 2000]`, we have used as a measure of the coherence of the rows and columns in the bicluster the so called *mean squared residue score* to be minimized.

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
A solution and objective function
Mean squared residue score
Our proposal, ①
Our proposal, ②
Our proposal, ③
Our proposal, ④
Small example, ①
Small example, ②
Small example, ③
Experimental results and Biological Significance
To conclude...

Mean squared residue score

Given a data matrix $\mathcal{A} = (X, Y)$, where

$a_{ij}, (i \in X, j \in Y)$, represents the relation between row i and column j,

given a bicluster $\mathcal{B} = (I, J)$, $I \subseteq X$, $J \subseteq Y$, and given

➜ the **mean of the ith row in $\mathcal{B}$**: $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij}$;

Mean squared residue score

Given a data matrix $\mathcal{A} = (X, Y)$, where

$a_{ij}, (i \in X, j \in Y)$, represents the relation between row i and column j,

given a bicluster $\mathcal{B} = (I, J)$, $I \subseteq X$, $J \subseteq Y$, and given

- ➜ the **mean of the ith row in $\mathcal{B}$**: $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij}$;
- ➜ the **mean of the jth column in $\mathcal{B}$**: $a_{Ij} = \frac{1}{|I|} \sum_{i \in I} a_{ij}$;

Mean squared residue score

Given a data matrix $\mathcal{A} = (X, Y)$, where

$a_{ij}, (i \in X, j \in Y)$, represents the relation between row i and column j,

given a bicluster $\mathcal{B} = (I, J)$, $I \subseteq X$, $J \subseteq Y$, and given

- ➜ the **mean of the i^{th} row in $\mathcal{B}$**: $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij};$
- ➜ the **mean of the j^{th} column in $\mathcal{B}$**: $a_{Ij} = \frac{1}{|I|} \sum_{i \in I} a_{ij};$
- ➜ the **mean of all the elements in $\mathcal{B}$**:

$$a_{IJ} = \frac{1}{|I| \cdot |J|} \sum_{i \in I,\, j \in J} a_{ij}; \quad a_{IJ} = \frac{1}{|I|} \sum_{i \in I} a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{Ij};$$

Mean squared residue score

Given a data matrix $\mathcal{A} = (X, Y)$, where

$a_{ij}, (i \in X, j \in Y)$, represents the relation between row i and column j,

given a bicluster $\mathcal{B} = (I, J)$, $I \subseteq X$, $J \subseteq Y$, and given

- ➜ the **mean of the ith row in $\mathcal{B}$**: $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij};$
- ➜ the **mean of the jth column in $\mathcal{B}$**: $a_{Ij} = \frac{1}{|I|} \sum_{i \in I} a_{ij};$
- ➜ the **mean of all the elements in $\mathcal{B}$**:

$$a_{IJ} = \frac{1}{|I| \cdot |J|} \sum_{i \in I,\, j \in J} a_{ij}; \quad a_{IJ} = \frac{1}{|I|} \sum_{i \in I} a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{Ij};$$

- ➜ the ***residue* of element a_{ij}**, i.e. the **difference between the actual value of a_{ij} and its expected value predicted from the corresponding row mean, column mean, and bicluster mean**:
  $r(a_{ij}) = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}; \quad a_{ij} = r(a_{ij}) + a_{iJ} + a_{Ij} - a_{IJ};$

Outline

Data Clustering

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

A new GRASP-like algorithm for Data Biclustering
A solution and objective function
Mean squared residue score
Our proposal, ①
Our proposal, ②
Our proposal, ③
Our proposal, ④
Small example, ①
Small example, ②
Small example, ③

Experimental results and Biological Significance

To conclude...

Mean squared residue score

Given a data matrix $\mathcal{A} = (X, Y)$, where

$a_{ij}, (i \in X, j \in Y)$, represents the relation between row i and column j,

given a bicluster $\mathcal{B} = (I, J)$, $I \subseteq X$, $J \subseteq Y$, and given

- ➜ the **mean of the i^{th} row in $\mathcal{B}$**: $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij}$;
- ➜ the **mean of the j^{th} column in $\mathcal{B}$**: $a_{Ij} = \frac{1}{|I|} \sum_{i \in I} a_{ij}$;
- ➜ the **mean of all the elements in $\mathcal{B}$**:

$$a_{IJ} = \frac{1}{|I| \cdot |J|} \sum_{i \in I,\, j \in J} a_{ij}; \quad a_{IJ} = \frac{1}{|I|} \sum_{i \in I} a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{Ij};$$

- ➜ the ***residue* of element a_{ij}**, i.e. the **difference between the actual value of a_{ij} and its expected value predicted from the corresponding row mean, column mean, and bicluster mean**:
  $r(a_{ij}) = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}; \quad a_{ij} = r(a_{ij}) + a_{iJ} + a_{Ij} - a_{IJ};$

the ***mean squared residue*** $H(\mathcal{B})$ is the **sum of the squared residues**:

$$H(\mathcal{B}) = \frac{1}{|I| \cdot |J|} \sum_{i \in I,\, j \in J} r(a_{ij})^2. \qquad \text{[To be minimized.]}$$

Our proposal, ①

A **new Reactive GRASP-like algorithm** with **a learning mechanism**: at each it., the RCL parameter $\alpha \in \Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_\ell\}$.

Our proposal, ①

A **new Reactive GRASP-like algorithm** with **a learning mechanism**: at each it., the RCL parameter $\alpha \in \Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_\ell\}$.

```
algorithm GRASP-like-bicluster($\mathcal{A}$,MaxNoImpr,MaxDist,$\delta$)
    $\Delta := \{\alpha_1, \ldots, \alpha_\ell\}$;        /* $\alpha_i \in [0,1]$, $i = 1, \ldots, \ell$ */
    for $i = 1$ to $\ell$ do
        $p_{\alpha_i} := \frac{1}{\ell}$;
    endfor
    $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ :=filtered-Kmeans($\mathcal{A}$);  /* $H(\mathcal{B}_q) \leq \delta$, $q = 1, \ldots, k$ */
    for $q = 1$ to $k$ do
        $\hat{\mathcal{B}}_q$ :=grasp($\mathcal{B}_q$,$\Delta$,$\mathcal{A}$,MaxNoImpr,MaxDist);
    endfor
    return ($\hat{\mathcal{B}} = \{\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_k\}$);
end
```

At the first GRASP it.: $p_{\alpha_i} = \frac{1}{\ell}$, $i = 1, \ldots, \ell$.

Our proposal, ①

A **new Reactive GRASP-like algorithm** with **a learning mechanism**: at each it., the RCL parameter $\alpha \in \Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_\ell\}$.

```
algorithm GRASP-like-bicluster($\mathcal{A}$,MaxNoImpr,MaxDist,$\delta$)
    $\Delta := \{\alpha_1, \ldots, \alpha_\ell\}$;        /* $\alpha_i \in [0,1]$, $i = 1, \ldots, \ell$ */
    for $i = 1$ to $\ell$ do
        $p_{\alpha_i} := \frac{1}{\ell}$;
    endfor
    $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ :=filtered-Kmeans($\mathcal{A}$);  /* $H(\mathcal{B}_q) \leq \delta$, $q = 1, \ldots, k$ */
    for $q = 1$ to $k$ do
        $\hat{\mathcal{B}}_q$ :=grasp($\mathcal{B}_q$,$\Delta$,$\mathcal{A}$,MaxNoImpr,MaxDist);
    endfor
    return ($\hat{\mathcal{B}} = \{\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_k\}$);
end
```

At the first GRASP it.: $p_{\alpha_i} = \frac{1}{\ell}$, $i = 1, \ldots, \ell$.

At any subsequent it., let $\hat{z}$ be the incumbent o.f. value and let A_i be the average o.f. value of all solutions found using $\alpha = \alpha_i$, $i = 1, \ldots, \ell$, then

$$p_i = \frac{q_i}{\sum_{j=1}^{\ell} q_j}, \quad q_i = \hat{z}/A_i, \; i = 1, \ldots, \ell.$$

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
A solution and objective function
Mean squared residue score
Our proposal, ①
Our proposal, ②
Our proposal, ③
Our proposal, ④
Small example, ①
Small example, ②
Small example, ③
Experimental results and Biological Significance
To conclude...

Our proposal, ②

A **new Reactive GRASP-like algorithm** with **a learning mechanism**: at each it., the RCL parameter $\alpha \in \Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_\ell\}$.

```
algorithm GRASP-like-bicluster($\mathcal{A}$,MaxNoImpr,MaxDist,$\delta$)
    $\Delta := \{\alpha_1, \ldots, \alpha_\ell\}$;        /* $\alpha_i \in [0,1], i = 1, \ldots, \ell$ */
    for $i = 1$ to $\ell$ do
        $p_{\alpha_i} := \frac{1}{\ell}$;
    endfor
    $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ :=filtered-Kmeans($\mathcal{A}$);  /* $H(\mathcal{B}_q) \leq \delta, q = 1, \ldots, k$ */
    for $q = 1$ to $k$ do
        $\hat{\mathcal{B}}_q$ :=grasp($\mathcal{B}_q$,$\Delta$,$\mathcal{A}$,MaxNoImpr,MaxDist);
    endfor
    return ($\hat{\mathcal{B}} = \{\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_k\}$);
end
```

It starts from a partial solution made of a set $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ of k biclusters found by applying a k-means procedure and retaining only biclusters with small mean squared residue (δ is a given input parameter).

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
A solution and objective function
Mean squared residue score
Our proposal, ①
Our proposal, ②
Our proposal, ③
Our proposal, ④
Small example, ①
Small example, ②
Small example, ③
Experimental results and Biological Significance
To conclude...

Our proposal, ③

A **new Reactive GRASP-like algorithm** with **a learning mechanism**: at each it., the RCL parameter $\alpha \in \Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_\ell\}$.

```
algorithm GRASP-like-bicluster($\mathcal{A}$,MaxNoImpr,MaxDist,$\delta$)
    $\Delta := \{\alpha_1, \ldots, \alpha_\ell\}$;        /* $\alpha_i \in [0,1]$, $i = 1, \ldots, \ell$ */
    for $i = 1$ to $\ell$ do
        $p_{\alpha_i} := \frac{1}{\ell}$;
    endfor
    $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ :=filtered-Kmeans($\mathcal{A}$);  /* $H(\mathcal{B}_q) \leq \delta$, $q = 1, \ldots, k$ */
    for $q = 1$ to $k$ do
        $\hat{\mathcal{B}}_q$ :=grasp($\mathcal{B}_q$,$\Delta$,$\mathcal{A}$,MaxNoImpr,MaxDist);
    endfor
    return ($\hat{\mathcal{B}} = \{\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_k\}$);
end
```

It proceeds in the attempt of finding a larger and/or better solution iteratively replacing a bicluster in the current solution by a larger and/or better bicluster.

Our proposal, ③

A **new Reactive GRASP-like algorithm** with **a learning mechanism**: at each it., the RCL parameter $\alpha \in \Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_\ell\}$.

```
algorithm GRASP-like-bicluster($\mathcal{A}$,MaxNoImpr,MaxDist,$\delta$)
    $\Delta := \{\alpha_1, \ldots, \alpha_\ell\}$;        /* $\alpha_i \in [0,1], i = 1, \ldots, \ell$ */
    for $i = 1$ to $\ell$ do
        $p_{\alpha_i} := \frac{1}{\ell}$;
    endfor
    $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ :=filtered-Kmeans($\mathcal{A}$);  /* $H(\mathcal{B}_q) \leq \delta, q = 1, \ldots, k$ */
    for $q = 1$ to $k$ do
        $\hat{\mathcal{B}}_q$ :=grasp($\mathcal{B}_q$,$\Delta$,$\mathcal{A}$,MaxNoImpr,MaxDist);
    endfor
    return ($\hat{\mathcal{B}} = \{\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_k\}$);
end
```

It proceeds in the attempt of finding a larger and/or better solution iteratively replacing a bicluster in the current solution by a larger and/or better bicluster.

As soon as `MaxNoImpr` its are performed without improving the current better solution, this solution is returned.

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
A solution and objective function
Mean squared residue score
Our proposal, ①
Our proposal, ②
Our proposal, ③
Our proposal, ④
Small example, ①
Small example, ②
Small example, ③
Experimental results and Biological Significance
To conclude...

Our proposal, ④

Given a bicluster $\bar{\mathcal{B}}_q = (\bar{I}_q, \bar{J}_q)$, `grasp` iteratively

☞ replaces it by a larger and/or better bicluster in its neighborhood

$$\mathcal{N}(\bar{\mathcal{B}}_q) = \left\{ \hat{\mathcal{B}}_q \quad \middle| \quad \begin{array}{l} \hat{\mathcal{B}}_q \text{ has one more element and/or} \\ \text{one less element (row or column)} \end{array} \right\};$$

Our proposal, ④

Given a bicluster $\bar{\mathcal{B}}_q = (\bar{I}_q, \bar{J}_q)$, `grasp` iteratively

- ☞ replaces it by a larger and/or better bicluster in its neighborhood

$$\mathcal{N}(\bar{\mathcal{B}}_q) = \left\{ \hat{\mathcal{B}}_q \quad | \quad \begin{array}{l} \hat{\mathcal{B}}_q \text{ has one more element and/or} \\ \text{one less element (row or column)} \end{array} \right\};$$

- ☞ the element to be removed and/or added is chosen on the basis either of the diversity or of the improvement in terms of mean squared residue and a RCL mechanism;

Our proposal, ④

Given a bicluster $\bar{\mathcal{B}}_q = (\bar{I}_q, \bar{J}_q)$, `grasp` iteratively

- ☞ replaces it by a larger and/or better bicluster in its neighborhood

$$\mathcal{N}(\bar{\mathcal{B}}_q) = \left\{ \begin{array}{ccl} \hat{\mathcal{B}}_q & | & \hat{\mathcal{B}}_q \text{ has one more element and/or} \\ & & \text{one less element (row or column)} \end{array} \right\};$$

- ☞ the element to be removed and/or added is chosen on the basis either of the diversity or of the improvement in terms of mean squared residue and a RCL mechanism;

- ☞ if a better mean squared residue neighbor bicluster is found, then the selection probabilities of the α's in Δ are accordingly reevaluated.

Small example, ①

Suppose a matrix $\mathcal{A}$ of 10 genes (rows) and 5 conditions (columns) is given:

$$\mathcal{A} = \left[\begin{array}{c|ccc} & \text{Condition 1} & \cdots & \text{Condition 5} \\ \hline \text{Gene 1} & a_{11} & \cdots & a_{15} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \text{Gene 10} & a_{10\,1} & \cdots & a_{10\,5} \end{array}\right],$$

Outline
Data Clustering
GRASP + Path Relinking for Data Clustering
Experimental results on Biological Data
Data BiClustering
A new GRASP-like algorithm for Data Biclustering
A solution and objective function
Mean squared residue score
Our proposal, ①
Our proposal, ②
Our proposal, ③
Our proposal, ④
Small example, ①
Small example, ②
Small example, ③
Experimental results and Biological Significance
To conclude...

Small example, ①

Suppose a matrix $\mathcal{A}$ of 10 genes (rows) and 5 conditions (columns) is given:

$$\mathcal{A} = \left[\begin{array}{c|ccc} & \text{Condition 1} & \cdots & \text{Condition 5} \\ \hline \text{Gene 1} & a_{11} & \cdots & a_{15} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \text{Gene 10} & a_{10\,1} & \cdots & a_{10\,5} \end{array}\right],$$

Fixed as input

- the number of sets of genes = 3, and
- the number of sets of conditions = 2,

`k-means` outputs the required sets and biclusters seeds are created:

$$\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_6\}.$$

Note: $6 = 3 \times 2$ combinations to match each set of genes with each set of conditions.

Outline

Data Clustering

GRASP + Path Relinking for Data Clustering

Experimental results on Biological Data

Data BiClustering

A new GRASP-like algorithm for Data Biclustering

A solution and objective function

Mean squared residue score

Our proposal, ①

Our proposal, ②

Our proposal, ③

Our proposal, ④

Small example, ①

Small example, ②

Small example, ③

Experimental results and Biological Significance

To conclude...

Small example, ①

Suppose a matrix $\mathcal{A}$ of 10 genes (rows) and 5 conditions (columns) is given:

$$\mathcal{A} = \left[\begin{array}{c|ccc} & \text{Condition 1} & \cdots & \text{Condition 5} \\ \hline \text{Gene 1} & a_{11} & \cdots & a_{15} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \text{Gene 10} & a_{10\,1} & \cdots & a_{10\,5} \end{array}\right],$$

Fixed as input

- ✍ the number of sets of genes = 3, and
- ✍ the number of sets of conditions = 2,

`k-means` outputs the required sets and biclusters seeds are created:

$$\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_6\}.$$

Note: $6 = 3 \times 2$ combinations to match each set of genes with each set of conditions.
Among the 6 combinations, only those whose mean squared residue is less than or equal to a given threshold δ are saved.
Suppose that

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}.$$

Small example, ②

Suppose that

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}.$$

$\mathcal{B}$ is given as input to an iterative refinement procedure that tries to add and/or remove items, considering first the columns and then the rows.

Small example, ②

Suppose that

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}.$$

$\mathcal{B}$ is given as input to an iterative refinement procedure that tries to add and/or remove items, considering first the columns and then the rows.

Suppose that $\mathcal{B}_1 = (I_1, J_1)$, with $|I_1| = 6$ and $J_1 = \{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_5\}$.

Small example, ②

Suppose that

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}.$$

$\mathcal{B}$ is given as input to an iterative refinement procedure that tries to add and/or remove items, considering first the columns and then the rows.

Suppose that $\mathcal{B}_1 = (I_1, J_1)$, with $|I_1| = 6$ and $J_1 = \{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_5\}$.

Suppose that

- ✍ RCL=$\{\mathcal{A}_2, \mathcal{A}_4\}$ (hScore);
- ✍ $\mathcal{A}_4$:=`select`(RCL);
- ✍ $J_1 := J_1 \cup \mathcal{A}_4$.

Therefore,

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\},$$

$\mathcal{B}_1 = (I_1, J_1)$, $|I_1| = 6$ and $J_1 = \{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}$.

Small example, ③

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\},\ \mathcal{B}_1 = (I_1, J_1),\ |I_1| = 6,\ J_1 = \{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}.$$

The local search tries to improve $\mathcal{B}_1$, by performing the following 3 steps, until a certain number of its without improvement are performed.

Small example, ③

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\},\ \mathcal{B}_1 = (I_1, J_1),\ |I_1| = 6,\ J_1 = \{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}.$$

The local search tries to improve $\mathcal{B}_1$, by performing the following 3 steps, until a certain number of its without improvement are performed.

① Randomly select a column not included: in our example, $\mathcal{A}_2$.
If the distance of $\mathcal{A}_2$ from the column previously extracted from RCL ($\mathcal{A}_4$) is at most a threshold given in input (`MaxDist`), $\mathcal{A}_2$ is added to J_1.
Let us suppose this is the case: $J_1 = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}$.

② From J_1 the column that makes worst the hScore is then eliminated.
Suppose that this column is $\mathcal{A}_3 \Longrightarrow J_1 = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_4, \mathcal{A}_5\}$.

③ A further column is selected at random from J_1.
It will be removed only if an improvement in terms of hScore is obtained.
Supposing that this happens for $\mathcal{A}_5 \Longrightarrow J_1 = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_4\}$.

Small example, ③

$$\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\},\ \mathcal{B}_1 = (I_1, J_1),\ |I_1| = 6,\ J_1 = \{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}.$$

The local search tries to improve $\mathcal{B}_1$, by performing the following 3 steps, until a certain number of its without improvement are performed.

① Randomly select a column not included: in our example, $\mathcal{A}_2$.
If the distance of $\mathcal{A}_2$ from the column previously extracted from RCL ($\mathcal{A}_4$) is at most a threshold given in input (`MaxDist`), $\mathcal{A}_2$ is added to J_1.
Let us suppose this is the case: $J_1 = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}$.

② From J_1 the column that makes worst the hScore is then eliminated.
Suppose that this column is $\mathcal{A}_3 \Longrightarrow J_1 = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_4, \mathcal{A}_5\}$.

③ A further column is selected at random from J_1.
It will be removed only if an improvement in terms of hScore is obtained.
Supposing that this happens for $\mathcal{A}_5 \Longrightarrow J_1 = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_4\}$.

These steps are applied on each selected bicluster $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$.

Experimental results and Biological Significance

Test environment

- MacBookPro 2GHz Intel Core Duo running MAC OSX 10.6;
- C language, compiled with the Apple Xcode 3.1;
- Stopping criterion: a maximum number of iterations without improvement of the incumbent solution.

Datasets

Datasets:

① Yeast (Saccharomyces cerevisiae) cell cycle expression [S. Tavazoie et al, 1999]:

it includes 2884 genes and 17 conditions, with **the expression level reported as an integer value in the range 0 to 600**.

Missing values are represented by -1.

② Lymphoma/Leukemia Molecular Profiling Project [A.A. Alizadeh et al, 2000]:

it includes 4026 genes and 96 conditions, with **the expression level reported as an integer value in the range -300 to 300**.

Statistics, ①

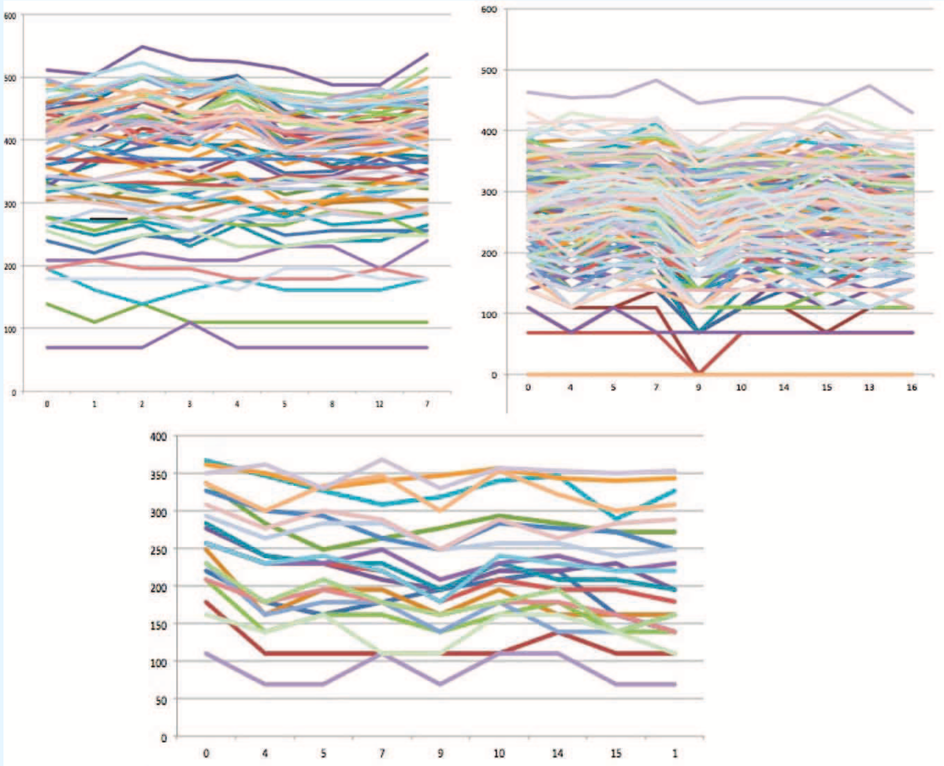
Results for a set of **33 biclusters generated for Yeast Dataset** and **11 biclusters generated for Lymphoma Dataset**:

Statistics (10 trials)	Yeast	Lymphoma
mean number of genes	97,33	59,63
mean number of conditions	10,52	8,18
mean volume	1000,06	478,93
mean H value	195,73	0,03
mean running time (in secs)	4044,43	5012,03
mean H_r value	1821,76	0,56

Our proposal is outperforming a simple random approach, since H_r is in both cases about one order of magnitude larger than the H.

Statistics, ②

Bicluster plots on Yeast:
gene behaviour on the rows; conditions on the columns.



Genes in sample biclusters present a similar behavior under a set of conditions ⟹ Our method is able to identify coherent biclusters from gene expression data.

Same on the Lymphoma Dataset.

To conclude...

Conclusions and Future Directions

- ✔ We have designed several GRASP+PR algorithms for Data Clustering:
  - ✧ `GRASP-PRf`: GRASP + PR forward;
  - ✧ `GRASP-PRb`: GRASP + PR backward;
  - ✧ `GRASP-PRm`: GRASP + PR mixed;
  - ✧ `GRASP-PRrnd`: GRASP + PR greedy randomized

  and tested on 5 datasets.
- ✔ We have designed a Reactive GRASP-like algorithm for Data BiClustering tested on 2 datasets.
- ✔ For all datasets, the proposed algorithms outperformed the state-of-the-art approaches and were able to identify coherent clusters/biclusters.

Conclusions and Future Directions

As future work, we intend

- ✔ to perform further validation with other datasets from literature;
- ✔ to further investigate the robustness and efficiency of our proposals by performing the so called TTT-plots;
- ✔ to include the automatic parameter tuning procedure for GRASP+PR heuristics based on a biased random-key genetic algorithm [Festa, Gonçalves, Resende, and Silva, 2010].

Conclusions and Future Directions

As future work, we intend

- ✔ to perform further validation with other datasets from literature;
- ✔ to further investigate the robustness and efficiency of our proposals by performing the so called TTT-plots;
- ✔ to include the automatic parameter tuning procedure for GRASP+PR heuristics based on a biased random-key genetic algorithm [Festa, Gonçalves, Resende, and Silva, 2010].

THANK YOU!