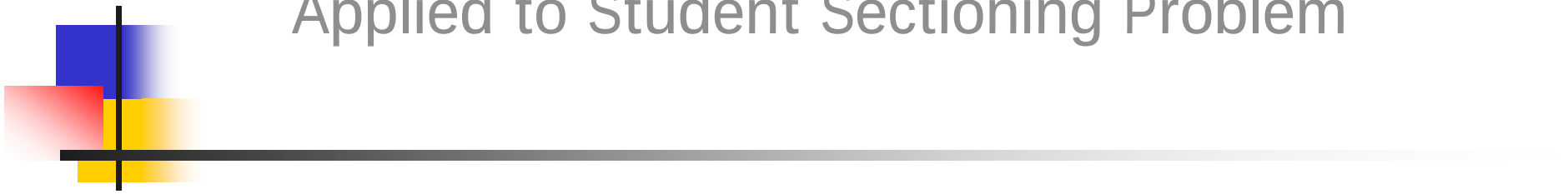
In the name of God

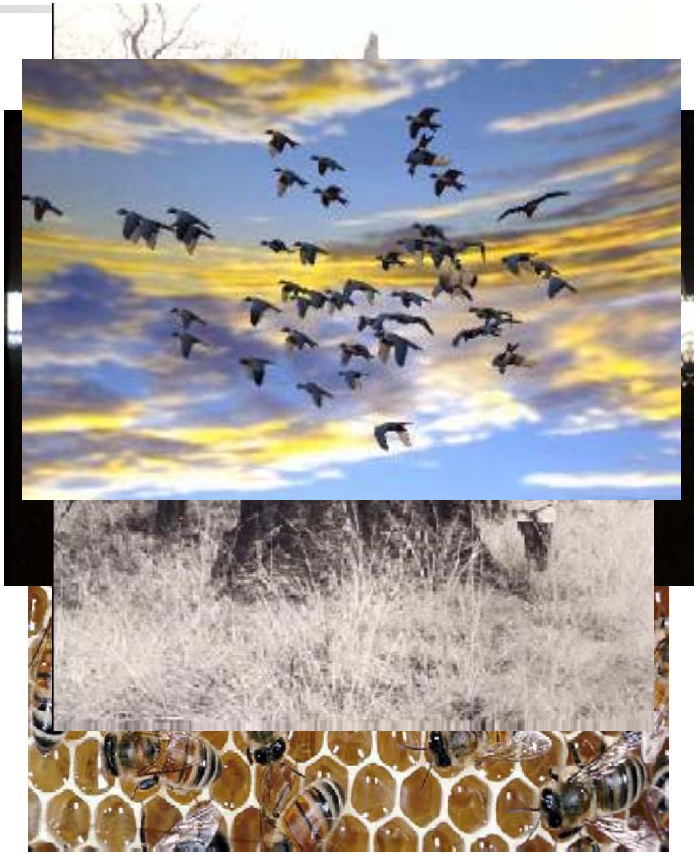
A Fish School Clustering Algorithm:
Applied to Student Sectioning Problem

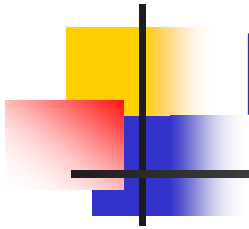


By:
Mahmood Amintoosi,
Mahmoud Fathy, Naser Mozayani, Adel T. Rahmani,
Iran University of Science and Technology

Introduction

- Inspiration from biology
- Swarm Intelligence
  - Ants' colonies, flocks of birds, termites, swarms of bees
- One application of “Swarm Intelligence” is “*Clustering Problem*”.
- Clustering: Process of grouping samples so that the samples are similar within each group.
- The groups are called *clusters*.





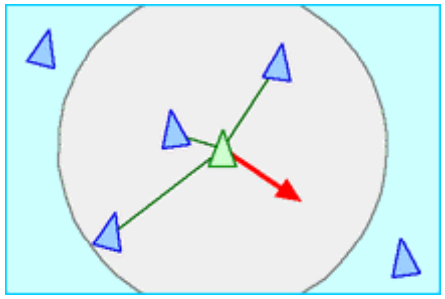
Fish School Behavior

- Agent, fish, flock-mate, individual have equal meaning in literatures.
- Flock meaning: herd, a group of birds, fish school

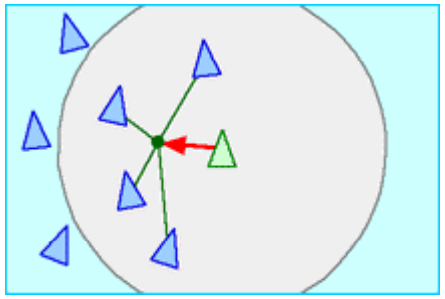
Flocking algorithms

- An example of emergent collective behavior.
  - No global control
  - Every agent has a limited visibility
  - The collective behavior emerges only by local interaction, following these three simple rules:

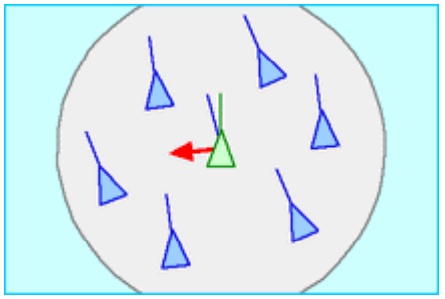
Separation



Cohesion



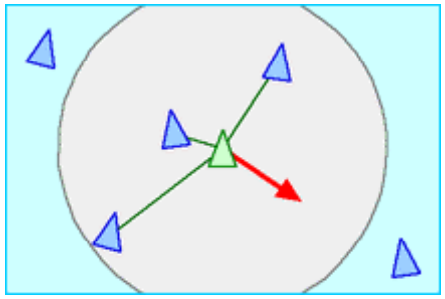
Alignment



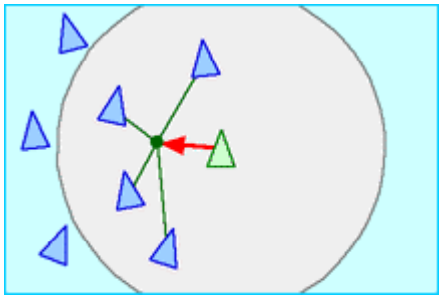
Flocking algorithms

- **Separation** gives an agent the ability to maintain a certain distance from others nearby.
- **Cohesion** gives an agent the ability to cohere (approach and form a group) with other nearby agents.
- **Alignment** gives an agent the ability to align with other nearby characters.

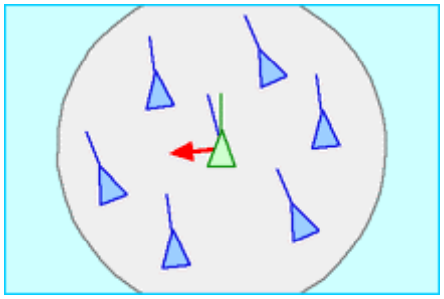
Separation



Cohesion



Alignment



Fish School Behavior

The proposed method is based on three observations in fish school and other flocks:

- Existence of leaders in flocks.
- Joining groups of individuals.
- Group shattering.

Existence of leaders in flocks

- Some researchers assumed a leader's existence and presumed it directed the movement of the whole flock, and some others don’t agree with this idea.
- The proposed method is based on this assumption that a group of animates have a leader.

Existence of leaders in flocks,
Definitions

In the proposed algorithm we considered two types of fishes:

- Leader fish
- Follower fish

Existence of leaders in flocks,
Definitions

In the proposed algorithm we considered two types of fishes:

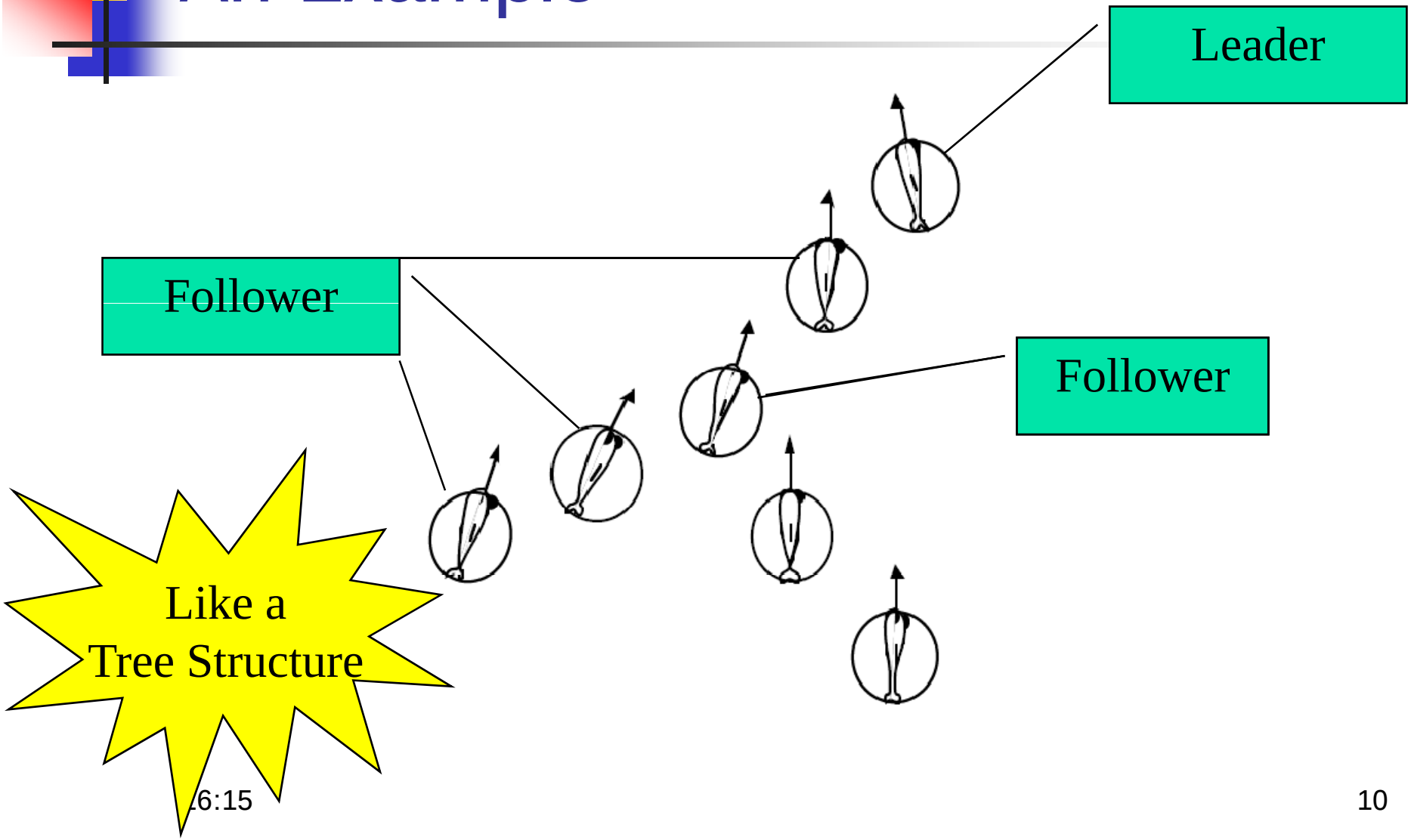
- Leader fish

  a fish that is not influenced by any of its flockmates

- Follower fish

  A follower fish is a fish, which is influenced by at least one of its flockmates.

Leader Fish & Follower Fish, An Example



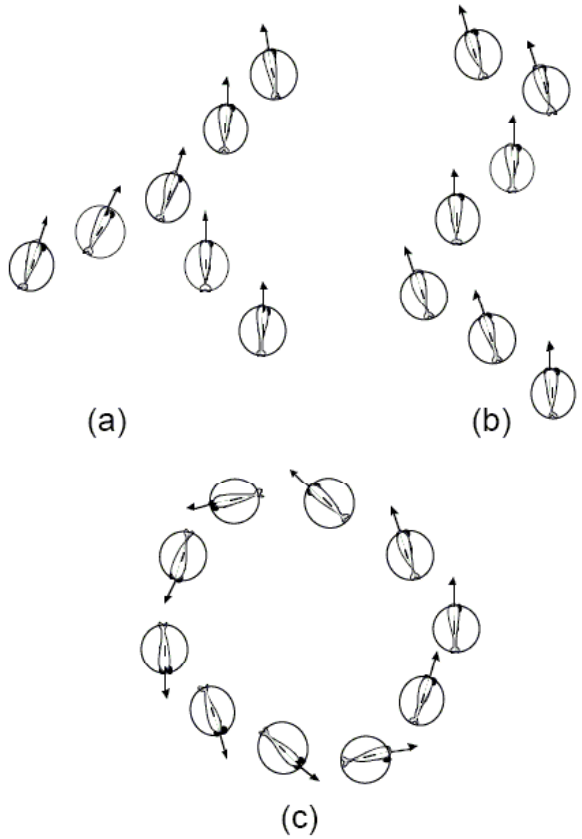
Some Form of Groups

a) An example of a 'Y' formation.

b) An example of a 'Z' formation.

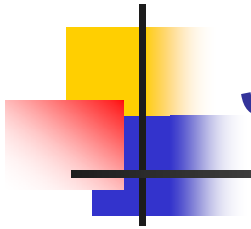
c) An example of a carousel structure

we avoid using carousel groups



Fish School Behavior

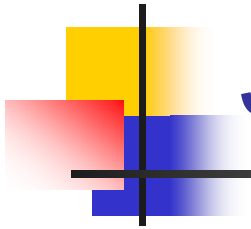
- Existence of leaders in flocks.
- **Joining groups of individuals.**
- Group shattering.



Joining groups of flockmates

According to Peterson's article when a number of individuals are initially placed at random in the environment:

- They quickly aggregate into several groups, generally consisting of four to five individuals.
- Over time these groups will themselves aggregate, if they are confined within a bounded area.
- The other individuals in the aggregation follow each other and the leader.



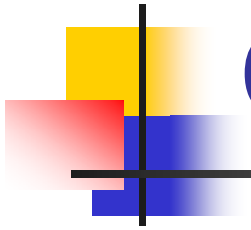
Joining groups of flockmates

leadership changing

- Over time leadership of the group will switch from one individual to another. A common leadership change occurs when a group leader sees another group, and begins to follow one of that group's members, bringing the two groups together.

Fish School Behavior

- Existence of leaders in flocks.
- Joining groups of individuals.
- **Group shattering.**



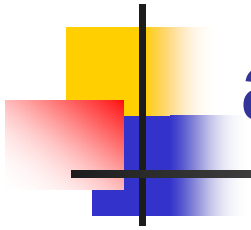
Group shattering

- Reynolds in his pioneer paper mentioned that real flocks sometimes split apart to go around an obstacle. Sparrows might flock around a group of obstacles that is in fact a herd of elephants.
- This behavior can be seen in Fish Schools.

The proposed fish school clustering algorithm (Assumptions)

- In the proposed method we assume that the whole flock contains several groups.
- Each group has a leader and other members of the group follow him.
- Every fish is corresponds to one data point.
- Fishes may form groups according to their similarity.
- Joining and shattering of the groups is the core of our algorithm.

The proposed fish school clustering algorithm (Assumptions), continue...



- In the proposed method, each cluster considered as a cave with a predefined capacity,
- Thus a sample's class label is corresponds to the fish's cave number.
- Each group can goes into a cave and stay there at time t.

The proposed fish school clustering algorithm (at a glance)

Initializing:

- Label each fish's class randomly.
- Mark each fish as a leader fish.

Repeat

1-Each leader fish A looks at all fishes he can see and select the nearest fish (B) as his leader.

2-If a leader found, A follows the leader by setting its class label and his followers' class labels to this new leader's class label. Mark A as a follower (Group Joining).

3-If the new class (cave) has not enough capacity, the group will be split. (Group Shattering).

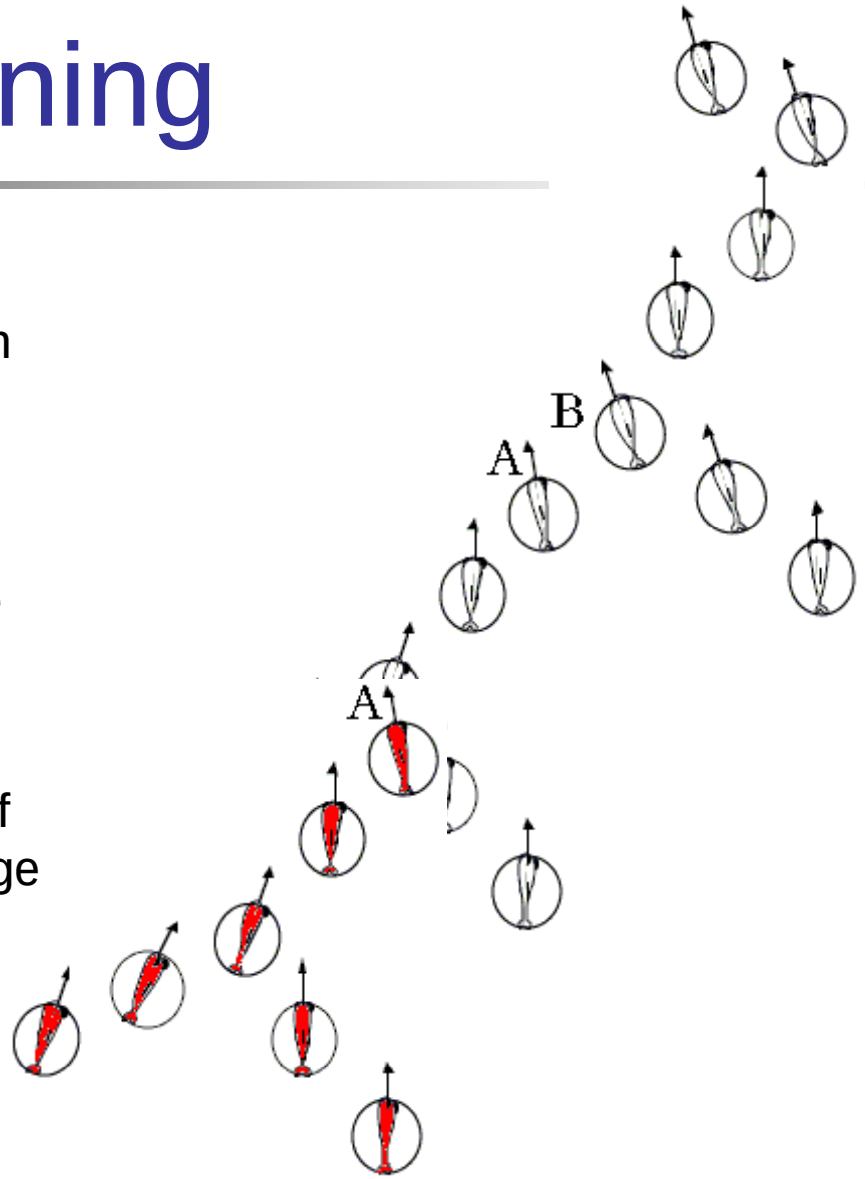
Until the labels remain unchanged in two consecutive iterations or reaching to maximum Iteration.

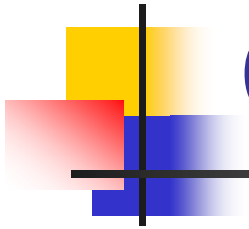
Group Joining

Suppose that we have two fish group. One group of ‘white’ class and another group of ‘red’ class.

Leader fish A of red group see and selects the nearest fish B as his leader.

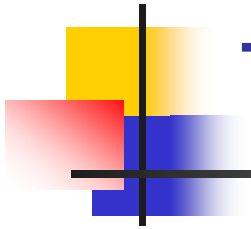
After Joining two groups, all of the fishes in ‘red’ group, change their classes and become members of ‘white’ class.





Checking cave capacity

- When a leader fish leaves its current cave and enters to a new cave, all of his followers follow him and go into the same cave , unless the cave reaches to its maximum capacity.



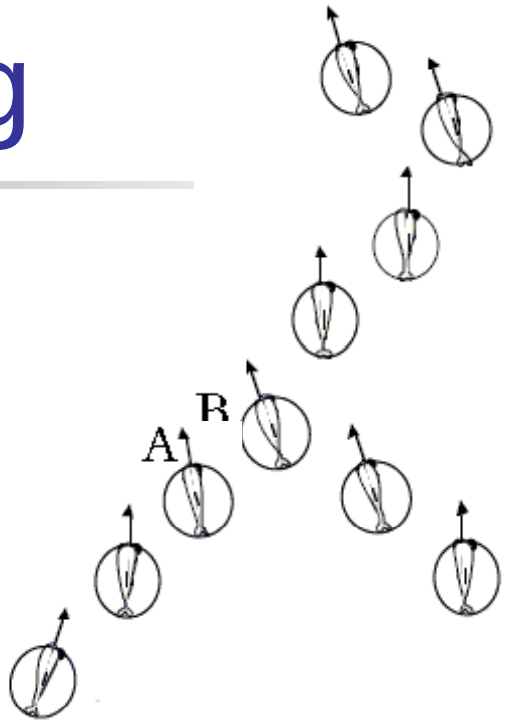
The Proposed Method, continue...

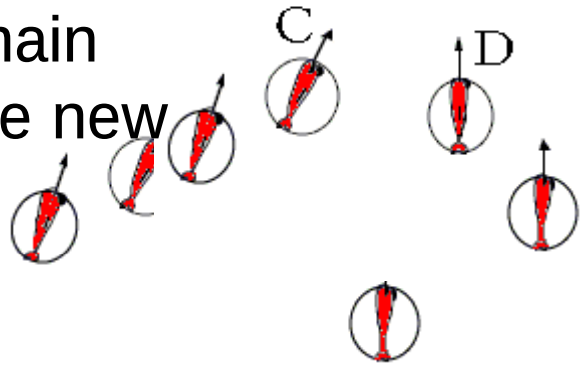
- What happens if the cave has not enough capacity, when a leader tries to enter it?
- The leader and a few of its followers – as much as possible – enter to the cave and the remaining fishes, stay on the previous cave.

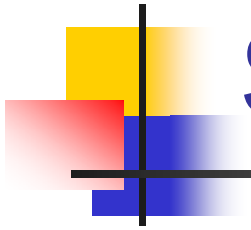
Group Shattering

- Now, suppose that the ‘white’ class only have 3 extra capacity.
- ‘A’ and two of his followers join to ‘white’ class.
- Group shattering occur.

Other 4 red fishes, remain ‘red’. ‘C’ and ‘D’ become new leaders.







Sight

- We used the term “he can see” in the algorithm. From this phrase we intend a criteria, which we restrict our self with a smaller domain in searching the neighborhoods of a fish.
- In our implementation we considered it as the half of the mean value of the similarities between each pair of fishes.

//Fish School Clustering algorithm

Initializing:

- maxClustSize: maximum allowed cluster size
- Label each fish's class randomly.
- Mark each fish as a leader fish.

Repeat

1-Each leader fish A looks at all fishes he can see and select the nearest fish (B) as his leader. He can only make another fish his leader, if B does not already followed A, directly or indirectly (for preventing carousel structure).

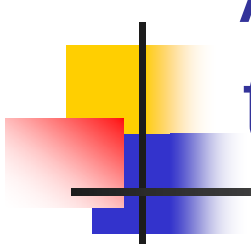
2-If a leader found, A follows the leader by setting its class label and his followers' class labels to this new leader's class label, in a breadth first search manner. Mark A as a follower.

3-Before changing a follower's class label, it is checked whether the size of the new class exceeds maxClustSize or not. If it exceeds, the labels of the remaining followers of A, will remain unchanged, and the headings fishes mark as leaders (shattering).

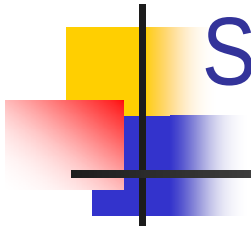
4-The fish remembers who his leader is.

Until the labels remain unchanged in two consecutive iterations or reaching to maximum Iteration.

Application of the proposed method to Student Sectioning Problem



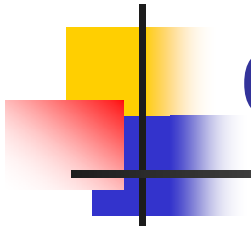
- Student Sectioning Problem (SSP) is a particular problem related to timetabling problems.
- The course timetabling problem involves the assignment of weekly lectures to time periods and lecture room in such a way that a set of constraints satisfy.
- SSP is due to courses, which involve a large number of students. For some reasons splitting these students to a few smaller sections is desirable



Student Sectioning Problem definition

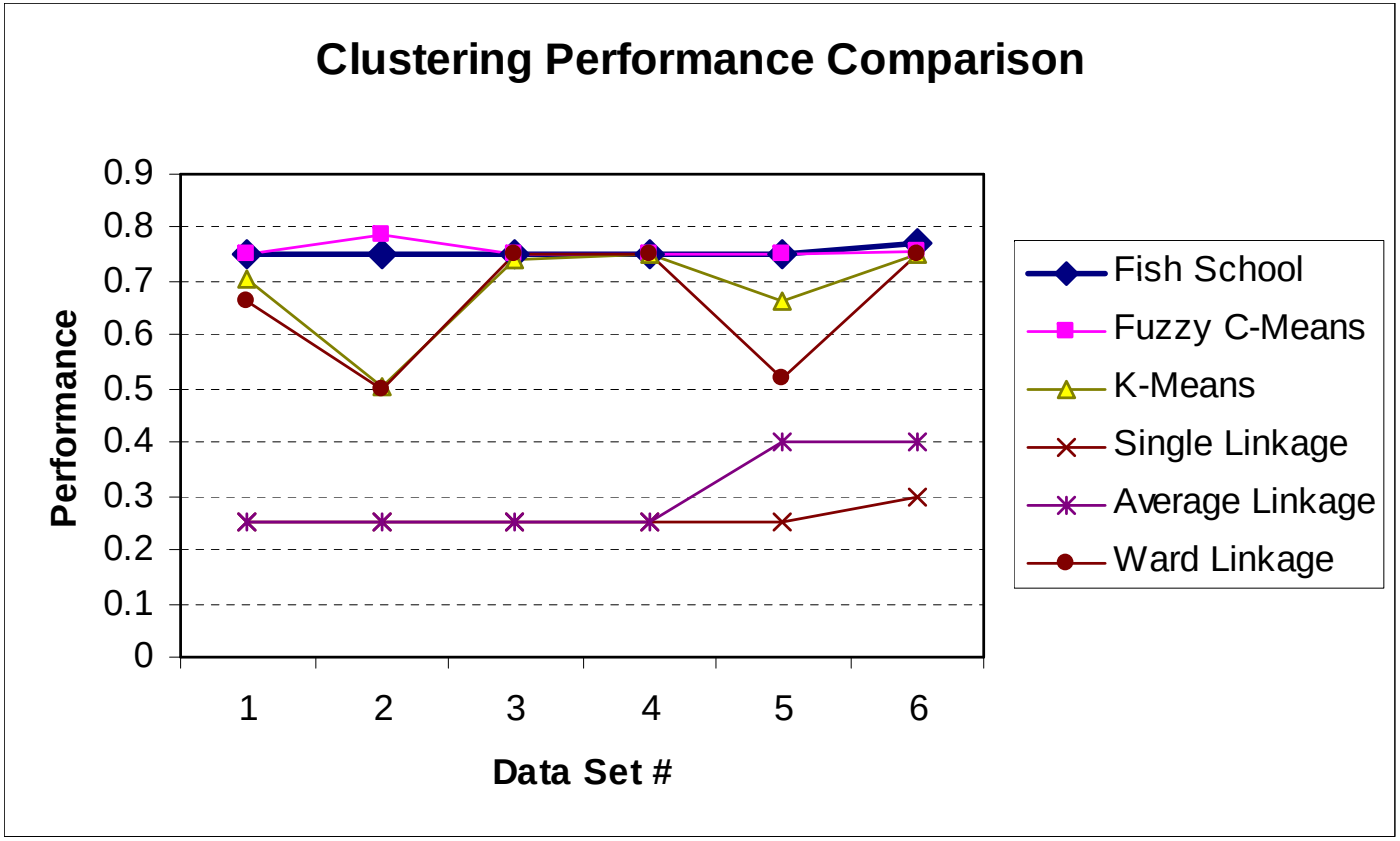
- The aim of the Student Sectioning Problem is to allocate students of a course to smaller sections for satisfying the following criteria:
  - Student course selections must be respected.
  - Section enrollments should be balanced, i.e. all sections of the same course should have roughly the same number of students;
  - Section capacities and policies of institute should not be exceeded.
  - Student schedules in each section would be the same as each other (as much as possible).

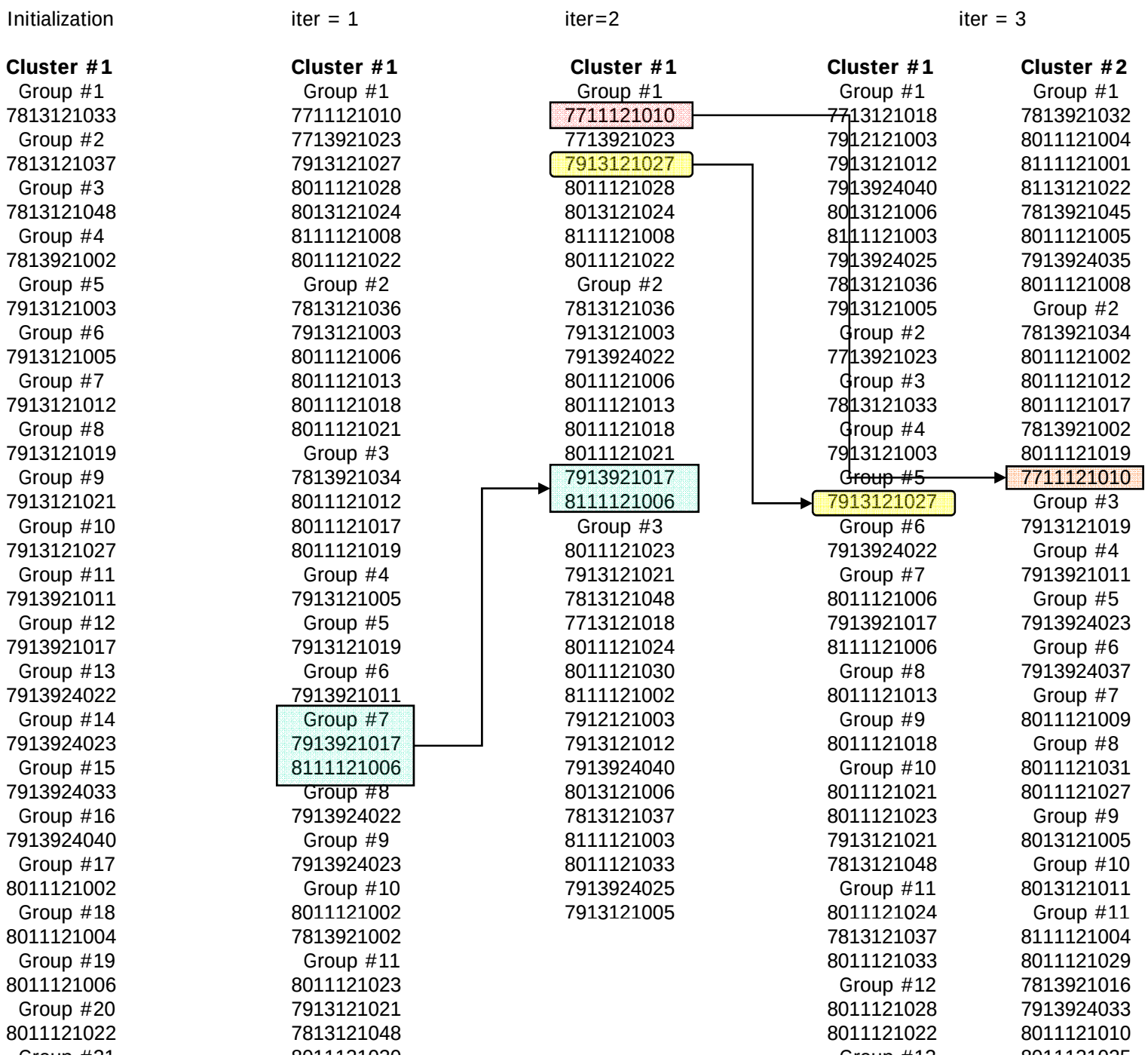
Applying the proposed clustering algorithm to SSP



- In fact students in our universities have behaviors like our artificial fishes.
- It is usual in our universities that a group of students (perhaps friends) have had same course schedules; and often the members of each group are influenced by a specified person.
- We suppose that each student is a fish in our algorithm.
- We need a metric for measuring similarity of two students. The total number of common courses of two students is defined as their similarity measure.

Comparison of the proposed Fish School clustering with some other clustering methods.




Initialization
iter = 1
iter=2
iter = 3
Cluster #1
Group #1
7813121033
Group #2
7813121037
Group #3
7813121048
Group #4
7813921002
Group #5
7913121003
Group #6
7913121005
Group #7
7913121012
Group #8
7913121019
Group #9
7913121021
Group #10
7913121027
Group #11
7913921011
Group #12
7913921017
Group #13
7913924022
Group #14
7913924023
Group #15
7913924033
Group #16
7913924040
Group #17
8011121002
Group #18
8011121004
Group #19
8011121006
Group #20
8011121022
Cluster #1
Group #1
7711121010
7713921023
7913121027
8011121028
8013121024
8111121008
8011121022
Group #2
7813121036
7913121003
8011121006
8011121013
8011121018
8011121021
Group #3
7813921034
8011121012
8011121017
8011121019
Group #4
7913121005
Group #5
7913121019
Group #6
7913921011
Group #7
7913921017
8111121006
Group #8
7913924022
Group #9
7913924023
Group #10
8011121002
7813921002
Group #11
8011121023
7913121021
7813121048
Cluster #1
Group #1
7711121010
7713921023
7913121027
8011121028
8013121024
8111121008
8011121022
Group #2
7813121036
7913121003
7913924022
8011121006
8011121013
8011121018
8011121021
7913921017
8111121006
Group #3
8011121023
7913121021
7813121048
7713121018
8011121024
8011121030
8111121002
7912121003
7913121012
7913924040
8013121006
7813121037
8111121003
8011121033
7913924025
7913121005
Cluster #1
Group #1
7713121018
7912121003
7913121012
7913924040
8013121006
8111121003
7913924025
7813121036
7913121005
Group #2
7713921023
Group #3
7813121033
Group #4
7913121003
Group #5
7913121027
Group #6
7913924022
Group #7
8011121006
7913921017
8111121006
Group #8
8011121013
Group #9
8011121018
Group #10
8011121021
8011121023
7913121021
7813121048
Group #11
8011121024
7813121037
8011121033
Group #12
8011121028
8011121022
Cluster #2
Group #1
7813921032
8011121004
8111121001
8113121022
7813921045
8011121005
7913924035
8011121008
Group #2
7813921034
8011121002
8011121012
8011121017
7813921002
8011121019
7711121010
Group #3
7913121019
Group #4
7913921011
Group #5
7913924023
Group #6
7913924037
Group #7
8011121009
Group #8
8011121031
8011121027
Group #9
8013121005
Group #10
8013121011
Group #11
8111121004
8011121029
7813921016
7913924033
8011121010