



重庆邮电大学

# 三支决策与聚类

于洪

计算机科学与技术学院  
重庆邮电大学

<http://cs.cqupt.edu.cn/yuhong/>

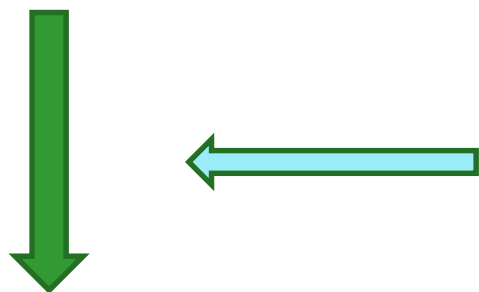
Email: [yuhong@cqupt.edu.cn](mailto:yuhong@cqupt.edu.cn)

2016.03

# 决策过程

问题求解

决策过程

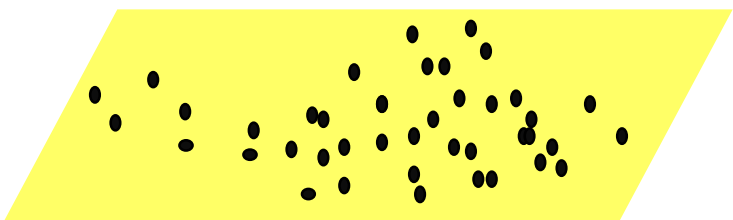
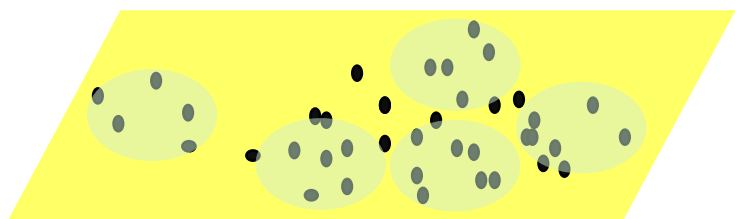
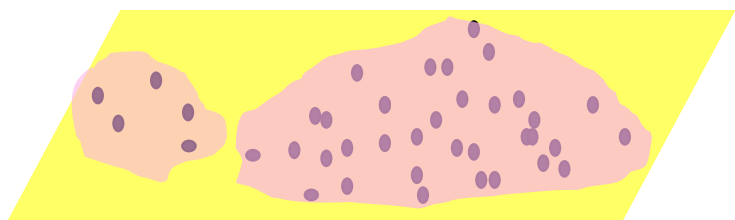
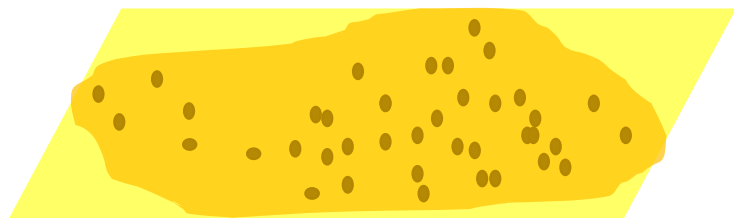


在认知和处理现实世界的问题时，常常采用从不同层次观察问题的策略，往往从极不相同的粒度上观察和分析同一问题

适应不同层次问题求解的决策过程



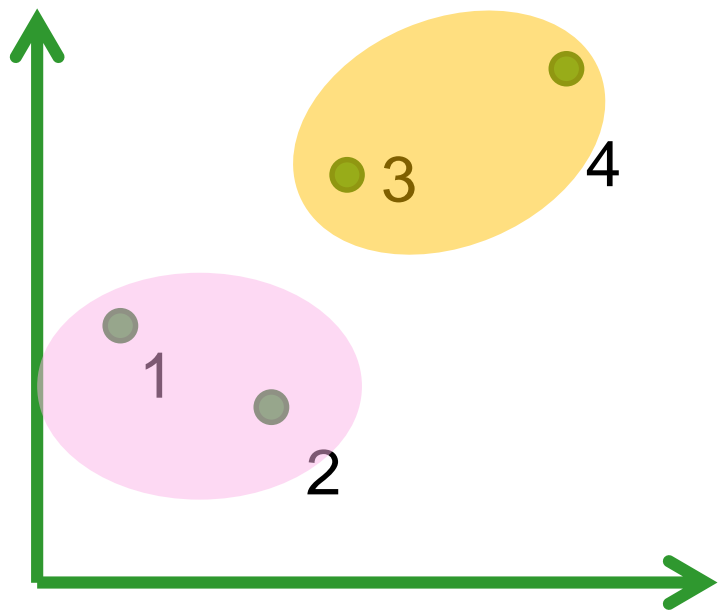
# 聚类过程



聚类过程反映的就是适应不同层次问题求解的决策过程，即聚类过程就是在某个粒度上决策对象元素是否属于某个类簇的过程。



# 二支决策聚类与三支决策聚类



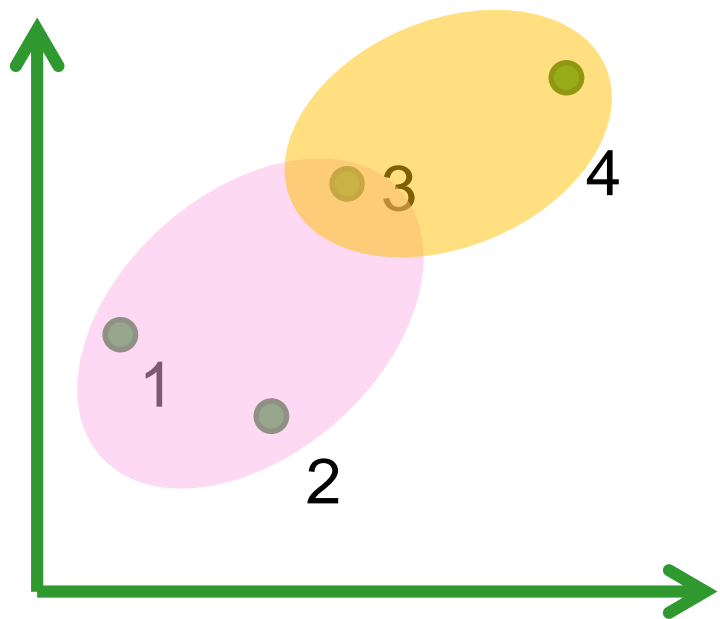
$$C_1 = \{1, 2\}$$

$$C_2 = \{3, 4\}$$

Hard Clustering



# 二支决策聚类与三支决策聚类



$$C_1 = \{1, 2\}$$

$$C_2 = \{3, 4\}$$

Hard Clustering

$$C_1 = \{1, 2, \mathbf{3}\}$$

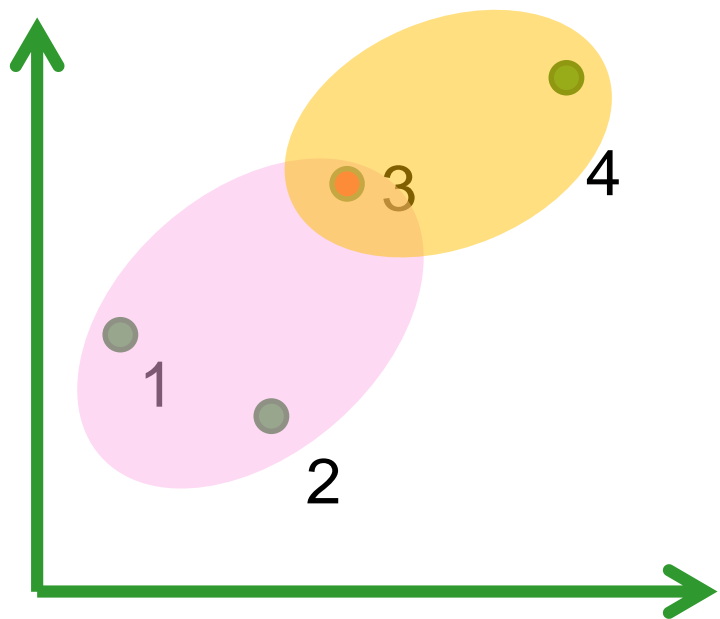
$$C_2 = \{\mathbf{3}, 4\}$$

Soft Clustering

## 二支决策的思想



# 二支决策聚类与三支决策聚类



针对那些目前知识体系下还难以决策的对象，

- 可以在已有知识体系下给出博弈后的一个**决策结果**
- 也可以等待新信息以帮助**进一步**决策

## 三支决策的思想

采用这种动态决策的思路可望为大数据处理带来新的理论模型和计算方法。



# 二支决策聚类与三支决策聚类

## 三支决策聚类的含义

决策结果是三支的：

类簇的三支表示

决策过程是三支的：

动态渐进的三支决策聚类

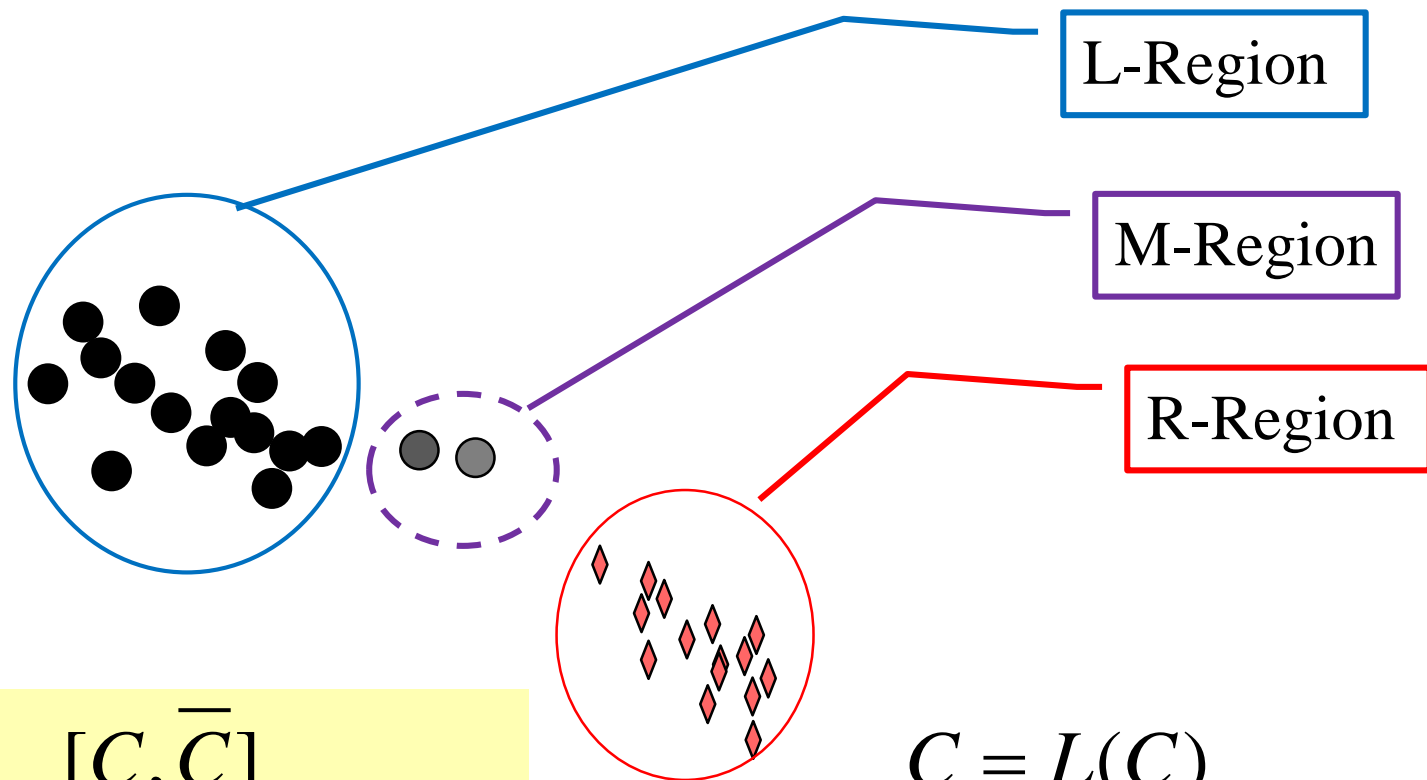
## 三支决策的思想

采用这种动态决策的思路可望为大数据处理带来新的理论模型和计算方法。



# 类簇的表示

## Three-way decision with two sets



$$\begin{array}{l} C \quad [\underline{C}, \bar{C}] \\ C \quad (L(C), M(C)) \end{array}$$

$$\begin{array}{l} \underline{C} = L(C) \\ \bar{C} = L(C) \cup M(C) \end{array}$$



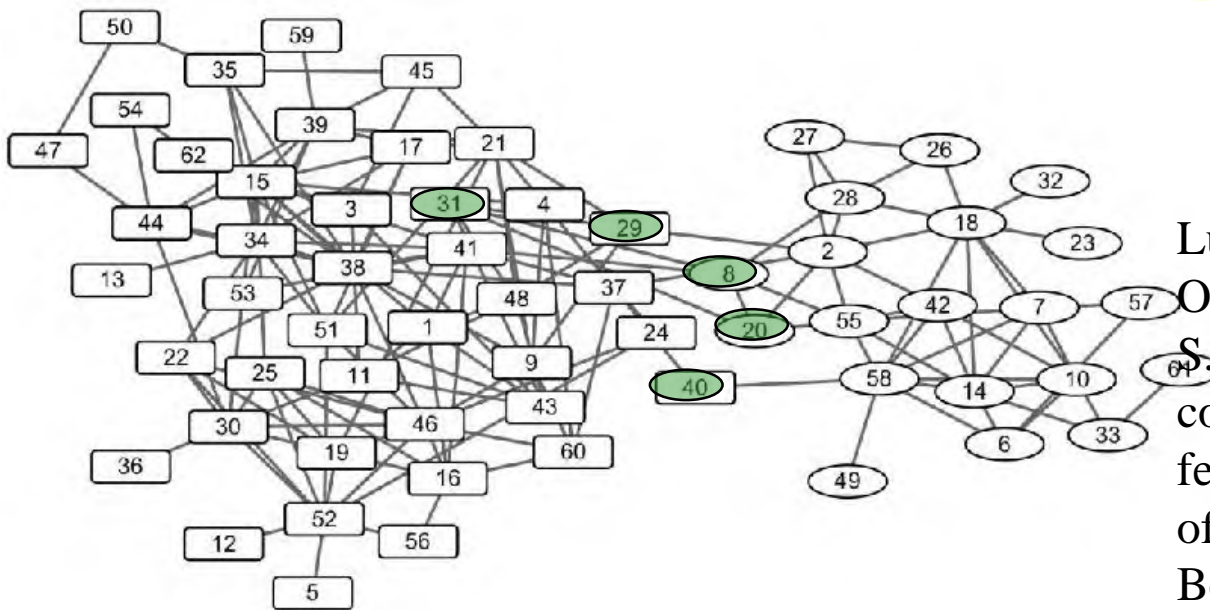


# A single set

$C_1 = \{1, 3, 4, 5, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 29, 30, 31, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 50, 51, 52, 53, 54, 56, 59, 60, 62\}$

$C_2 = \{2, 6, 7, 8, 10, 14, 18, 20, 23, 26, 27, 28, 29, 31, 32, 33, 40, 42, 49, 55, 57, 58, 61\}$

**{8, 20, 29, 31, 40}**

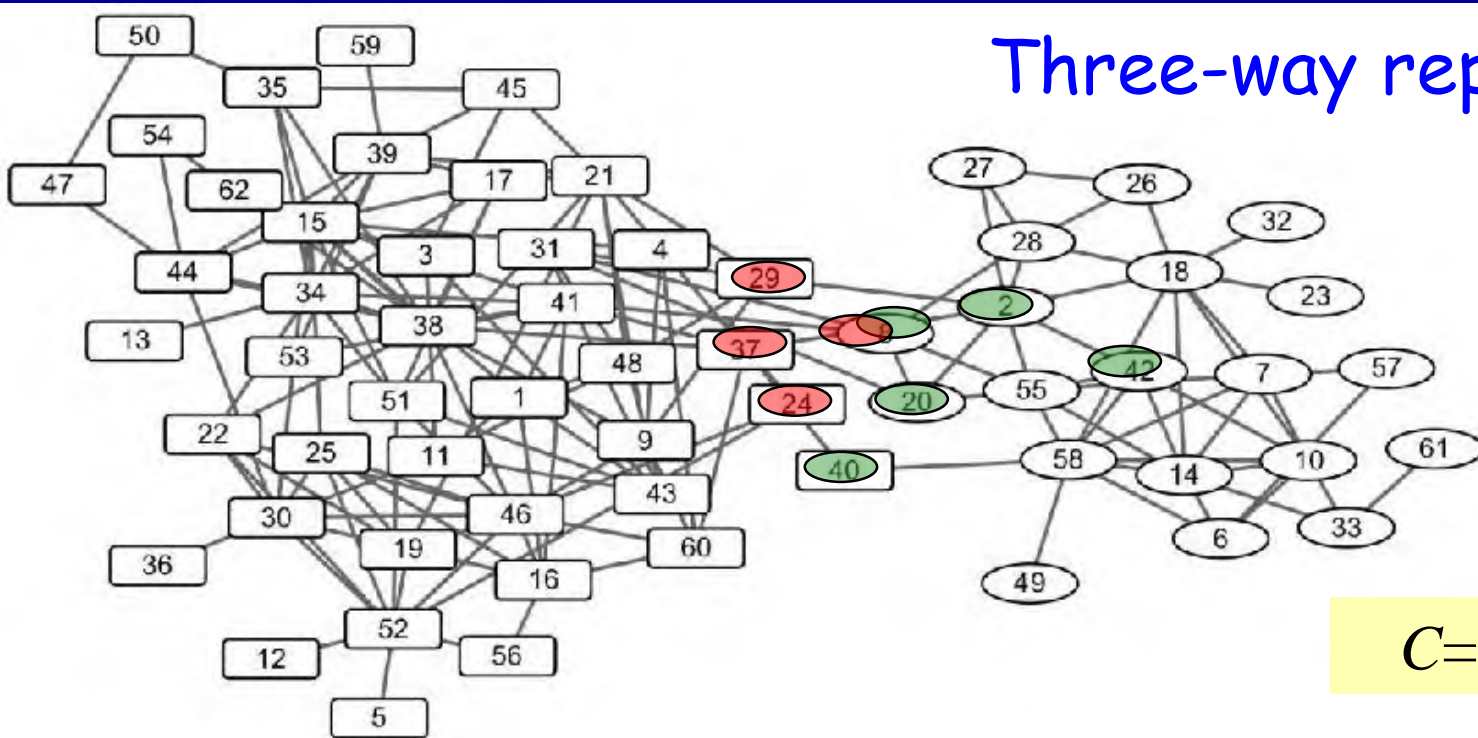


Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Sloaten, E., Dawson, S.M.: The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Behavioral Ecology and Sociobiology, 54(4): 396-405 (2003)



The dolphins social networkst

# Three-way representation



$$C = (L, M)$$

$L(C_1) = \{1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 24, 25, 29, 30, 31, 34, 35, 36, 37, 38, 39, 41, 43, 44, 45, 46, 47, 48, 50, 51, 52, 53, 54, 55, 56, 59, 60, 62\}$ ,

$M(C_1) = \{2, 8, 40, 42\}$ ,

$L(C_2) = \{2, 6, 10, 14, 18, 20, 23, 26, 27, 28, 32, 33, 40, 42, 49, 55, 57, 58, 61\}$ ,

$M(C_2) = \{8, 24, 29, 37\}$ .



1

**An automatic method to determine the number of clusters using decision-theoretic rough set**

2

**A tree-based incremental overlapping clustering method using the three-way decision theory**

3

**Detecting and refining overlapping regions in complex networks by three-way decisions**



# Our work

1

An automatic method to determine the number of clusters using decision-theoretic rough set

2

A tree-based incremental overlapping clustering method using the three-way decision theory

3

**Detecting and refining overlapping regions in complex networks by three-way decisions**



# In working: Detecting and refining overlapping regions in complex networks by three-way decisions



the relationships/roles among these members (objects) in overlapping regions is different

refining overlapping regions

detecting different types of overlapping regions

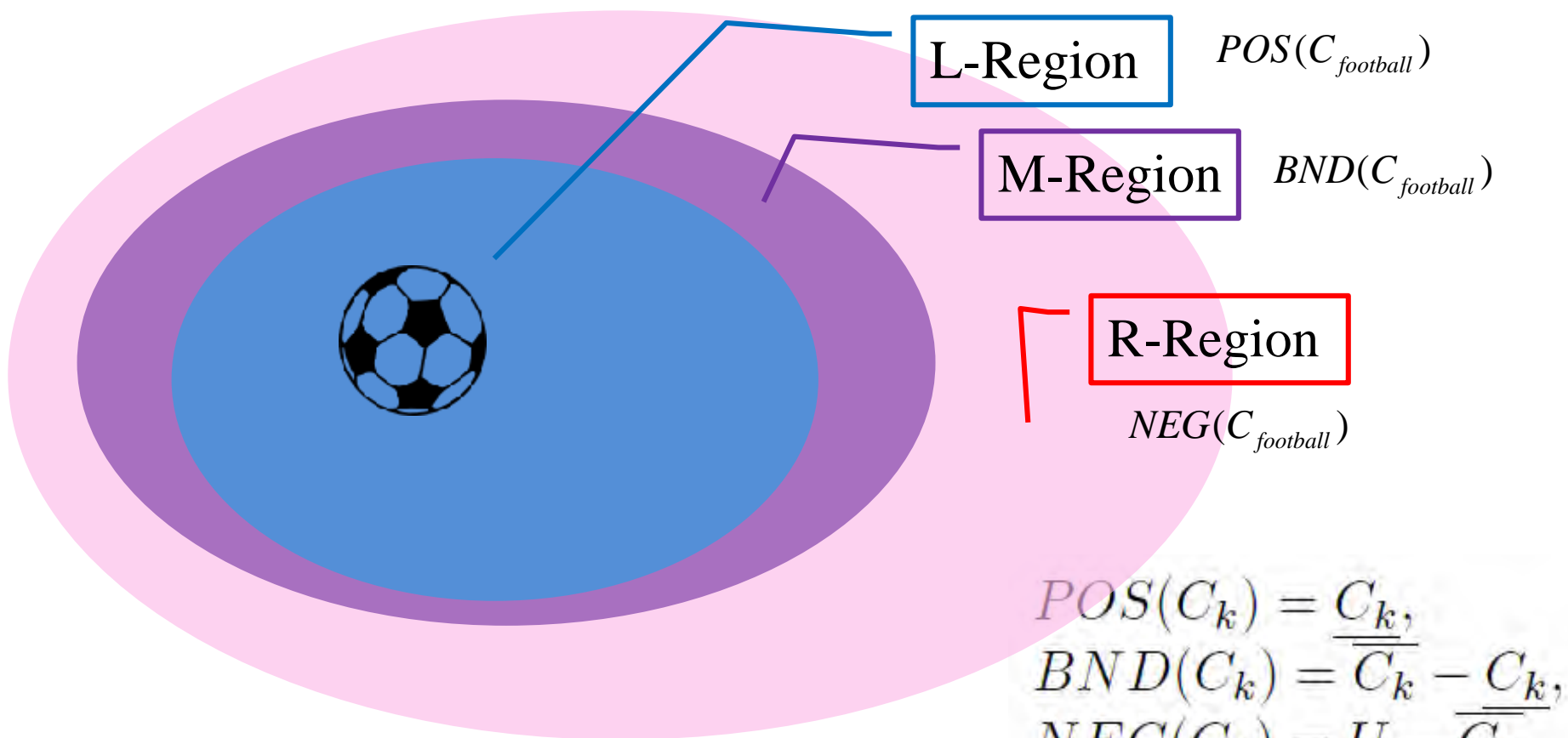
+

Represent a community with an interval set

Propose a multi-stage algorithm based on three-way decisions



# Represent a community



$$\begin{aligned}
 POS(C_k) &= \underline{C}_k, \\
 BND(C_k) &= \overline{C}_k - \underline{C}_k, \\
 NEG(C_k) &= U - \overline{C}_k.
 \end{aligned}$$

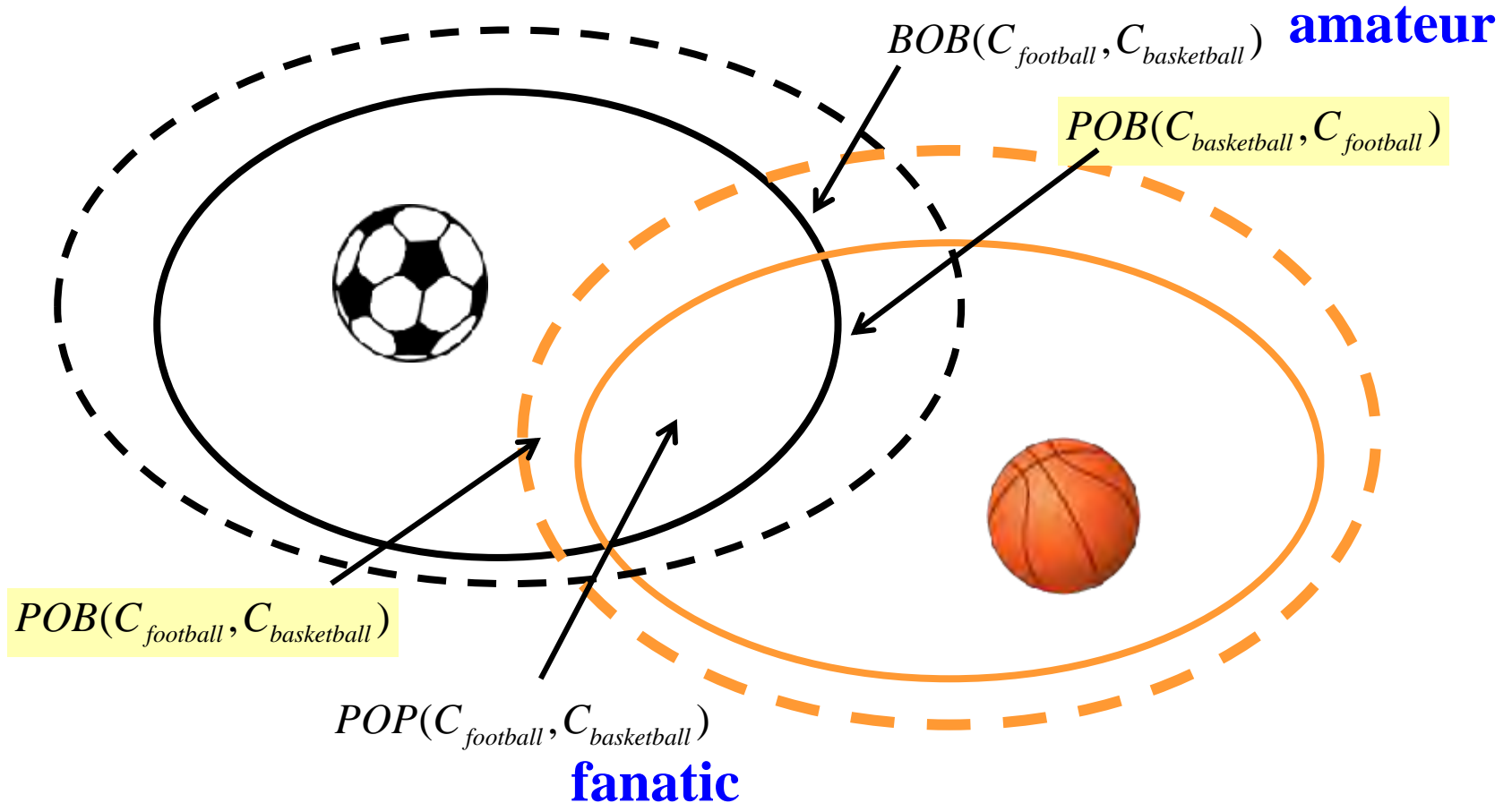
$C_{football}$

$[\underline{C}, \overline{C}]$

$[POS(C_{football}), POS(C_{football}) \cup BND(C_{football})]$



# Refining Overlapping Regions



- the example about football fans and basketball fans described in previous can be explained in this model



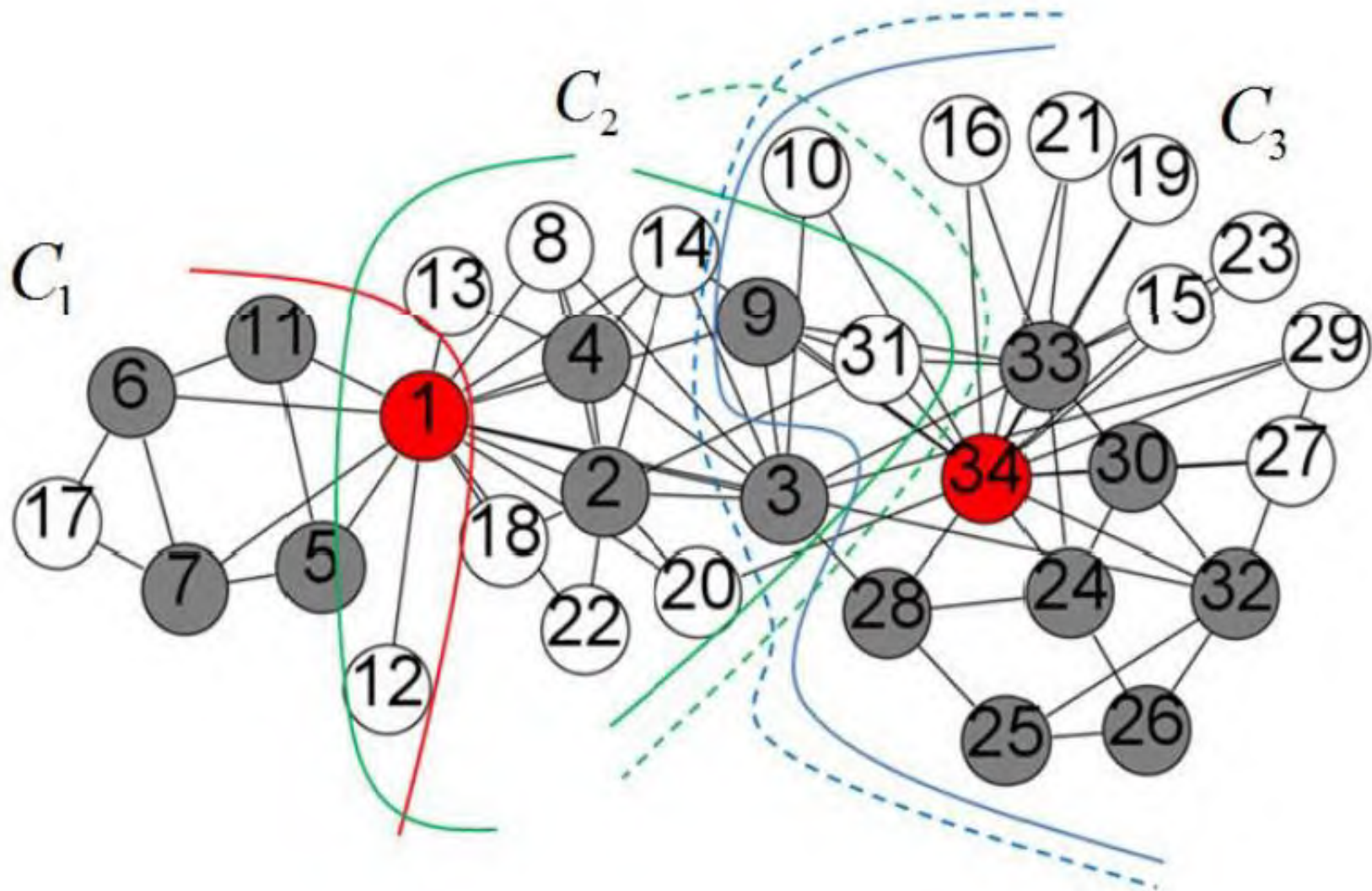
# Categorization of overlapping regions

Traditional	Categorization for overlapping regions				
Overlapping	Macro Type	Micro Type	POP	POB	BOB
	TYPE.1	A	○	○	○
		B	○	○	×
		C	○	×	○
		D	○	×	×
	TYPE.2	E	×	○	○
		F	×	○	×
TYPE.3	G	×	×	○	
Non Overlapping	TYPE.4	H	×	×	×

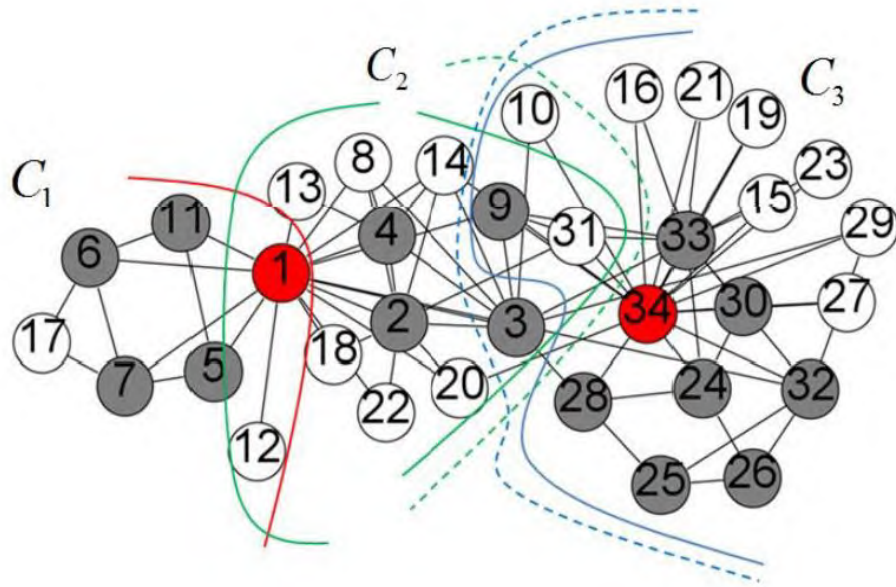




# Zachary's karate club network



# Zachary's karate club network



◆ Result of DOC-TWD in Zachary's karate club

◆ Overlapping vertices between  $C_2$  and  $C_3$  in Zachary's karate club

Algorithm	Overlapping Vertices	
DOC-TWD	$POP(C_2, C_3)$	9,31
	$POB(C_2, C_3)$	3
	$POB(C_3, C_2)$	10
LFM[1]	3,9,10,14,31	
DenShrink[2]	10,20	
EM-BOAD[3]	3	
Sun[4]	9,10,14,20	

[1]Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks[J]. **New Journal of Physics**, 2009, 11(3): 033015.

[2]Huang J, Sun H, Han J, et al. Density-based shrinkage for revealing hierarchical and overlapping community structure in networks[J]. **Physica A: Statistical Mechanics and its Applications**, 2011, 390(11): 2160-2171.

[3]Li J, Wang X, Eustace J. Detecting overlapping communities by seed community in weighted complex networks[J]. **Physica A: Statistical Mechanics and its Applications**, 2013, 392(23): 6125-6134.

[4]Sun P G, Gao L, Shan Han S. Identification of overlapping and non-overlapping community structure by fuzzy clustering in complex networks[J]. **Information Sciences**, 2011, 181(6): 1060-1071.



1

An automatic method to determine **the number of clusters** using decision-theoretic rough set

2

A tree-based incremental overlapping clustering method using the three-way decision theory

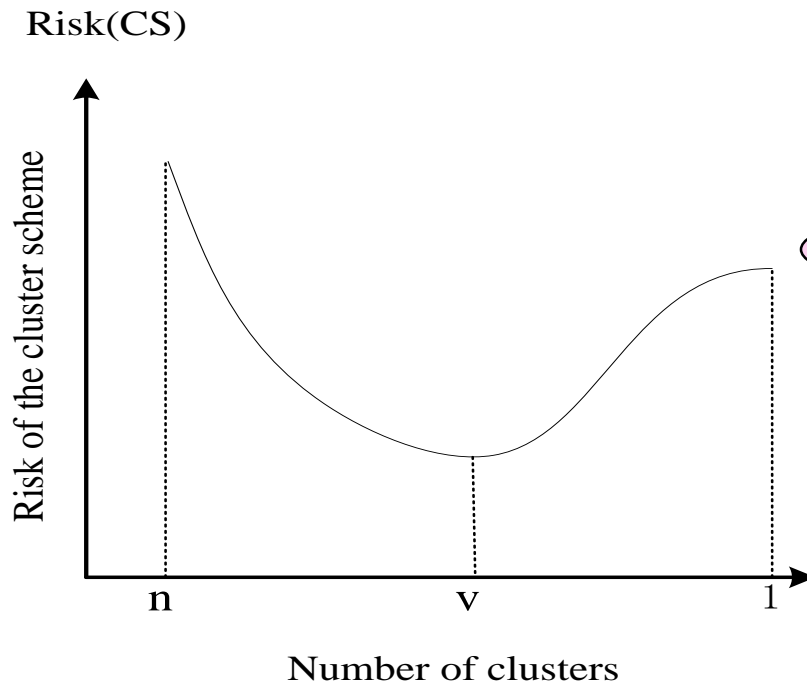
3

Detecting and refining overlapping regions in complex networks by three-way decisions



# Determining the Number of Clusters Using DTRS

a new clustering validity evaluation function based on the **extended DTRS**



get the curve of the clustering quality

determine the number of clusters corresponding to the extremum of the curve

**Hong Yu**, Zhanguo Liu, Guoyin Wang. An automatic method to determine the number of clusters using decision-theoretic rough set.

*International Journal of Approximate Reasoning*, 2014, 55(1): 101-115.



# Our work

1

An automatic method to determine the number of clusters using decision-theoretic rough set

2

A tree-based **incremental overlapping clustering method** using the **three-way decision theory**

3

Detecting and refining overlapping regions in complex networks by three-way decisions



# Main objective of the study

the original data set



the incremental data



To propose a new soft clustering algorithm using **three-way decisions**



To propose strategies based on **three-way decisions**

Representative points and construct the tree of representative points

Searching and updating the tree

Clustering the incremental data using **three-way decisions**

**To propose a soft incremental clustering approach**

Hong Yu, Cong Zhang, Guoyin Wang. A tree-based incremental overlapping clustering method using the three-way decision theory, *Knowledge-Based Systems*, Vol.91, Jan. 2016, Pages 189–203

<http://cs.cqupt.edu.cn/yuhong/>



# 大数据算法

在给定的**资源约束**下，以大数据为输入，在**给定时间约束内**可以生成**满足给定约束结果**的算法。

不是

仅仅在电子计算机上运行的算法

精确算法

内存算法

串行算法

不仅仅是

云计算

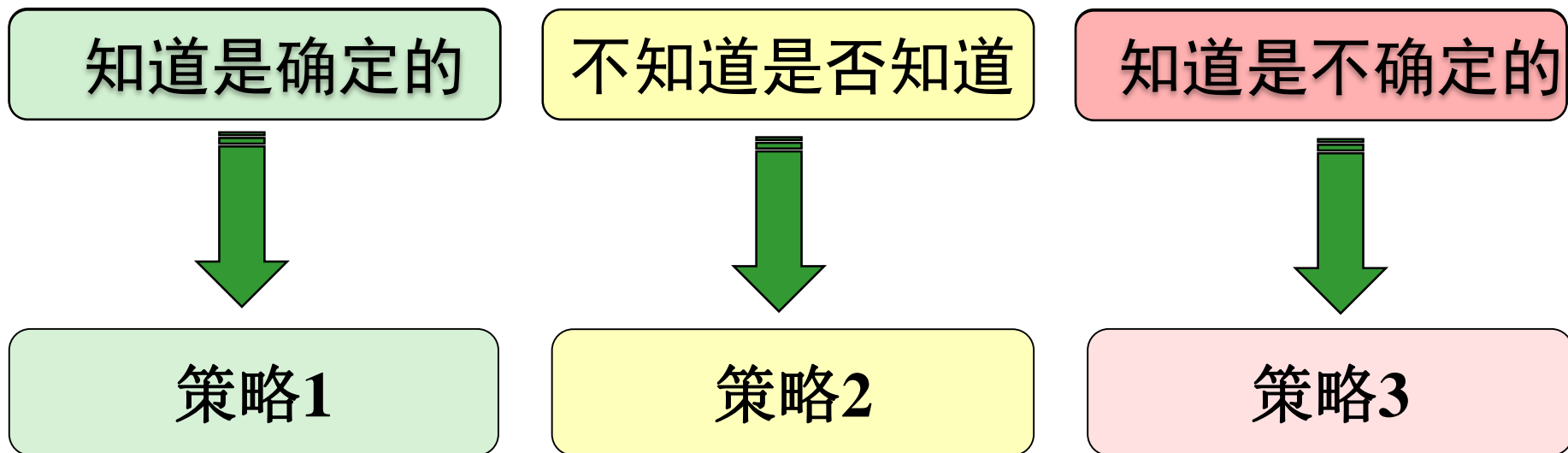
MapReduce

大数据分析和

挖掘的算法



## BIG DATA – BIG PROBLEM



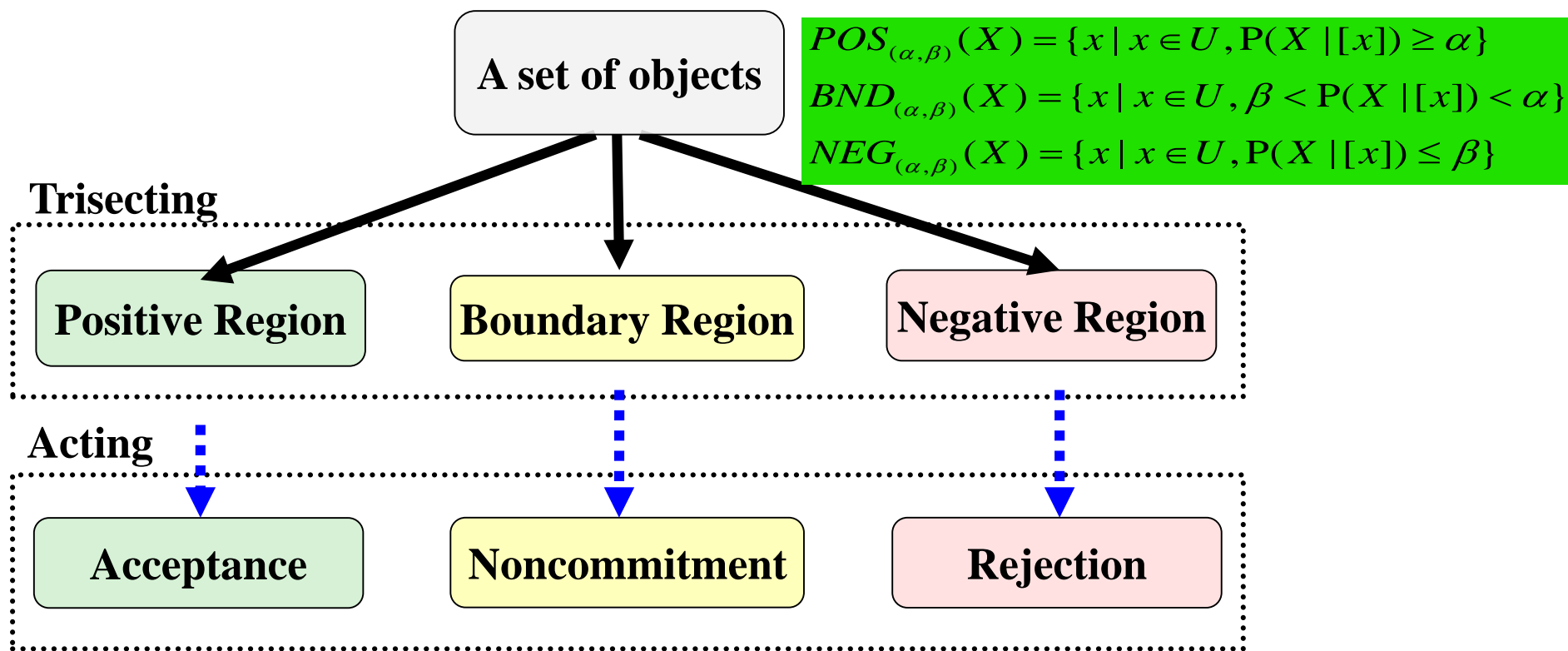
一分为三（知道是确定的、知道是不确定的、不知道是否知道）：**给出不同的解决方案**





# 大数据 — 三支决策 三支分类模型-不确定性

**概率三支分类模型**：从决策粗糙集模型出发，构造三个决策域，从而用于分类。  
A three-way classification model



一分为三：给出不同的解决方案（并行化）

解决过程可以是三支的：构建渐进式的三支决策模型



在给定的资源约束下，以大数据为输入，在给定时间约束内可以生成**满足给定约束结果**的算法。

可以从“**评价指标/标准**”是三支的来考虑：

完全/高标准  
满足约束

某种程度  
满足约束

不满足约束



WELCOME

THANK YOU

**关于三支决策的座谈会：**

**地点：信科大厦19楼会议室**

**时间：14:30-16:00**



**三支决策微信群：**

**三支决策主页：**

**<http://www2.cs.uregina.ca/~twd/>**

WELCOME

THANK YOU

**关于三支决策的座谈会：**

**地点：信科大厦19楼会议室**

**时间：14:30-16:00**



**三支决策主页：**

<http://www2.cs.uregina.ca/~twd/>

WELCOME

THANK YOU

关于三支决策的座谈会：

地点：信科大厦19楼会议室

时间：14:30-16:00



三支决策微信群：



三支决策主页  
网页版