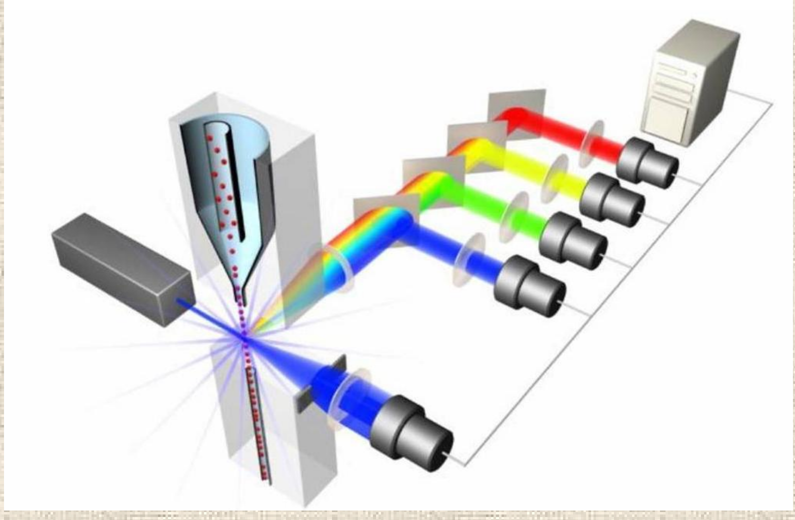
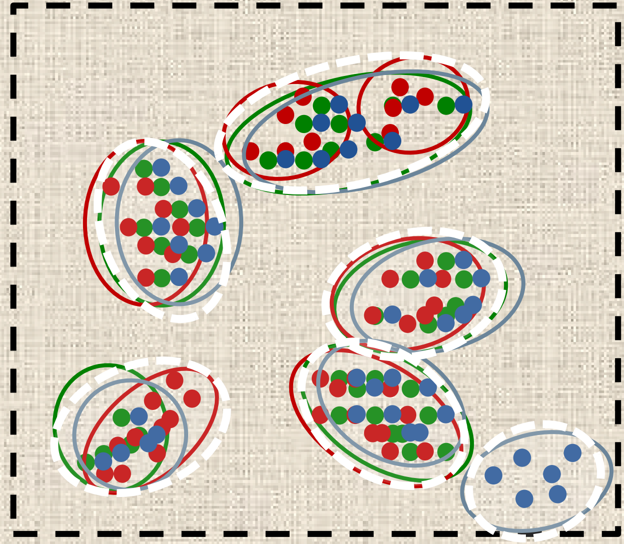
IDENTIFYING CROSS CLASS PHOSPHORYLATION DIFFERENCES IN COMPARATIVE FLOW CYTOMETRY





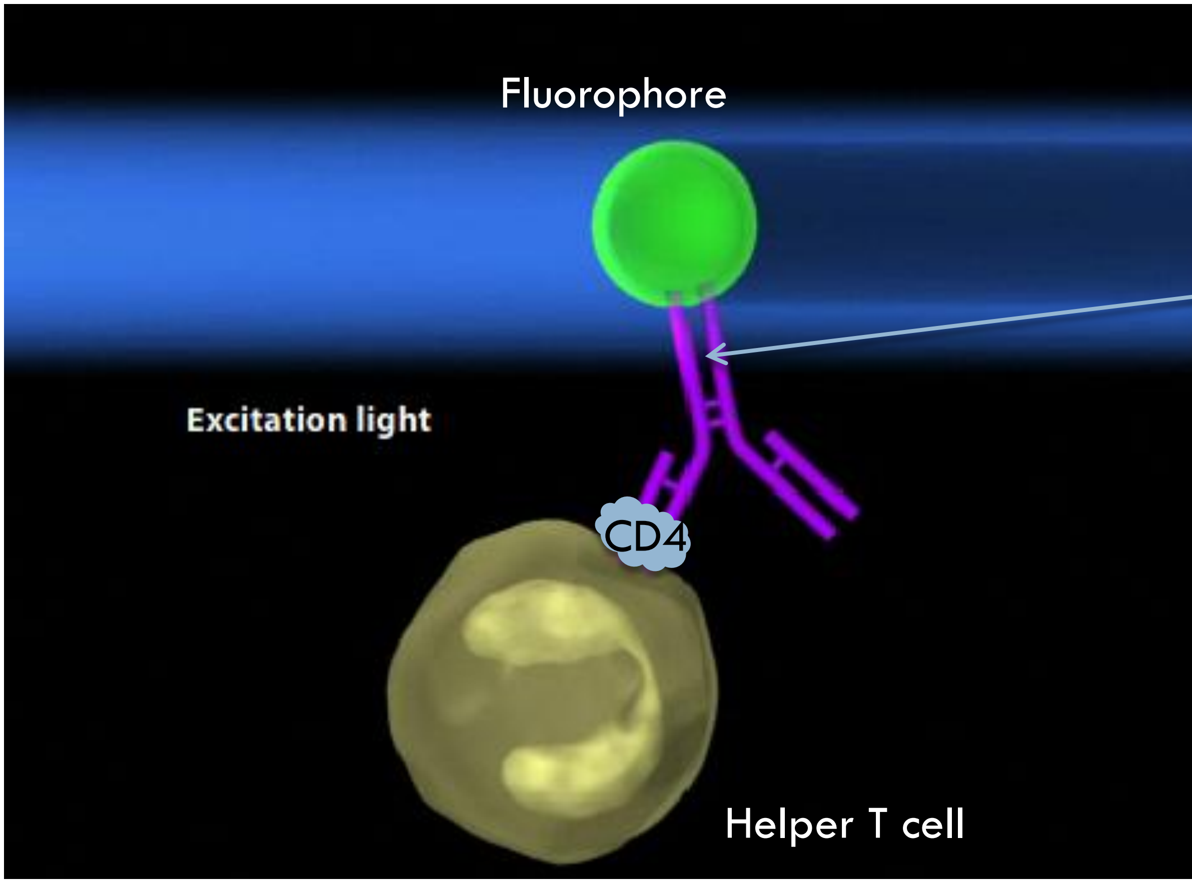
05/03/2011

Ariful Azad, Alex Pothen, Department of Computer Science
Purdue University, Indiana

Outline

- Flow cytometry data collection overview
- Problem description
- Methods
- Results

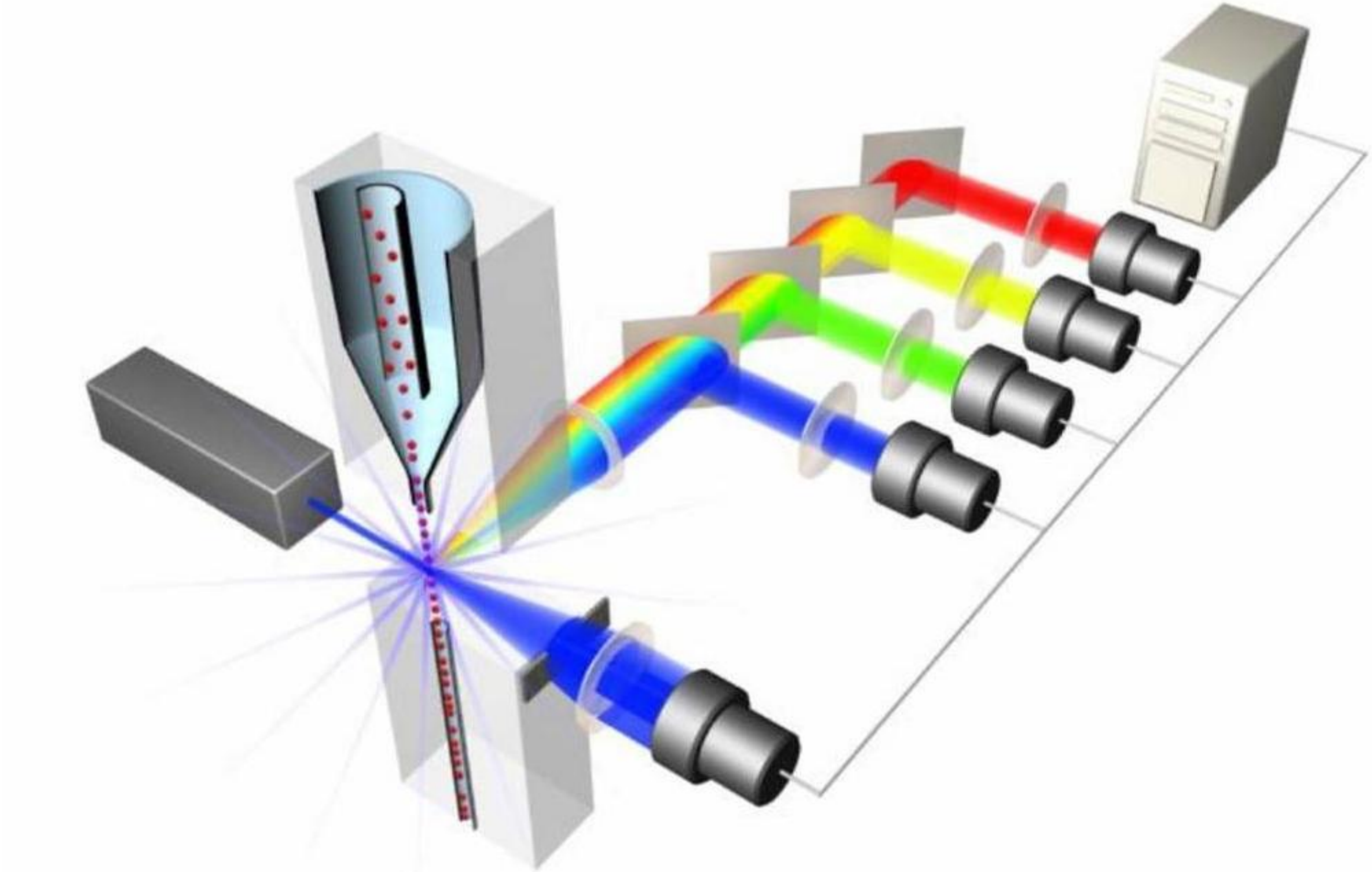
Flow cytometer data collection



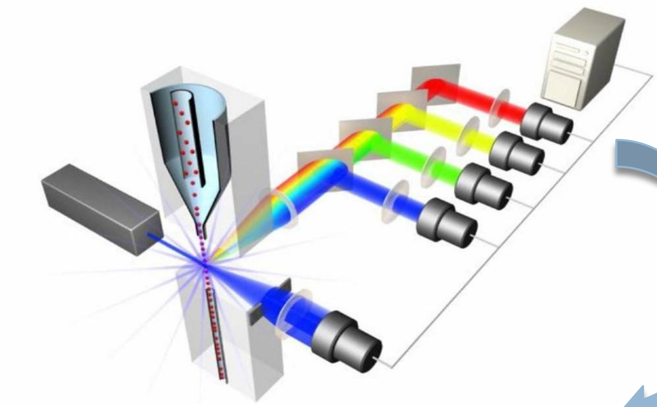
Antibody that is able to bind to CD4

http://probes.invitrogen.com/resources/education/tutorials/4Intro_Flow/player.html (slide 15,16,17)

Flow cytometer data collection



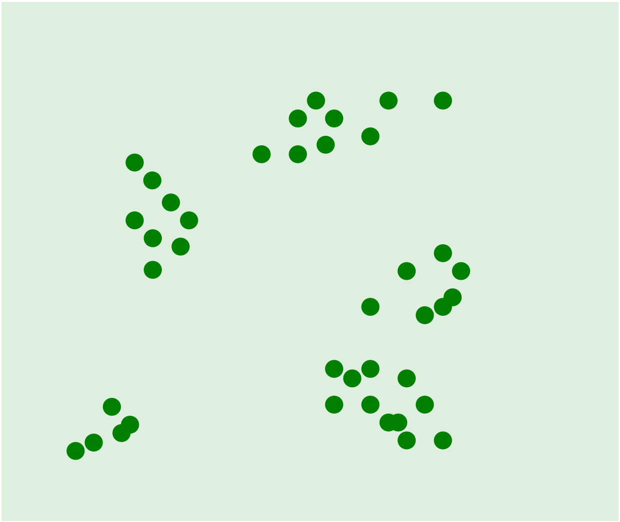
Flow cytometer data collection



	CD4	CD8
Cell-1	3	1
Cell-2	12	1
Cell-3	3.3	4.4
Cell-4	3.1	3.3
…..	…	…

Matrix format

Can be treated as an image!



Scatter plot

Assessing T cell phosphorylation effect

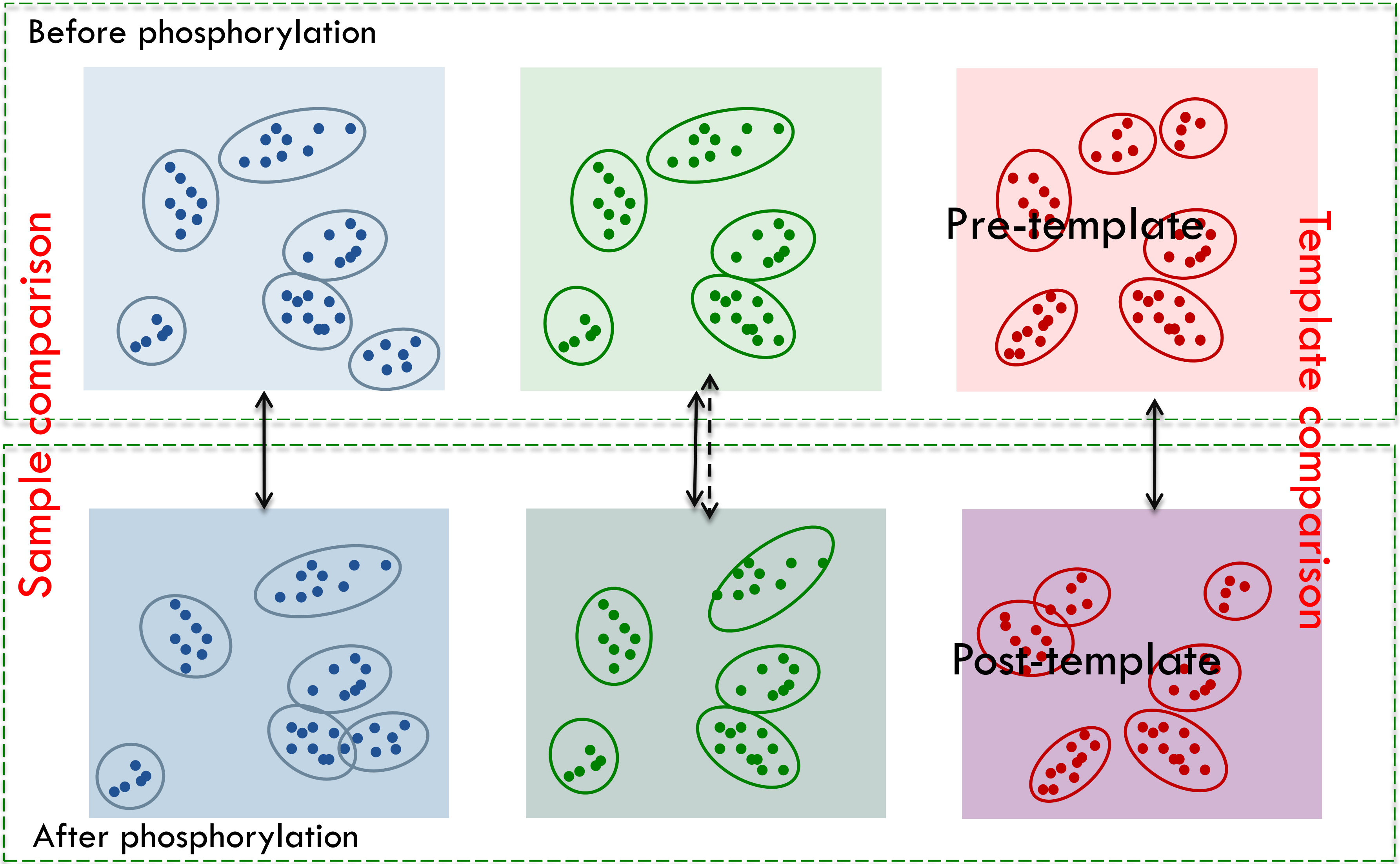
- Adding a Phosphate group (PO_4) into a protein.
- May activate or deactivate the protein .
- May impact on the level of expression of other protein .
- It is known that in T cell the expression level of ZAP70 and SLP76 increases in a sample after phosphorylation event (Maier et. al 2007, Pyne et. al 2009)

Objectives

- Input data
  - 30 blood samples stained with 4 fluorophores before and after phosphorylation .
  - Total 60 samples (each with 4 dimension)

- Compare phosphorylation pattern in general before phosphorylation and after phosphorylation (not particular pair of samples)
  - Identification of a template of all populations before phosphorylation and after phosphorylation.
  - Compare the two templates to find similarity or dissimilarity.

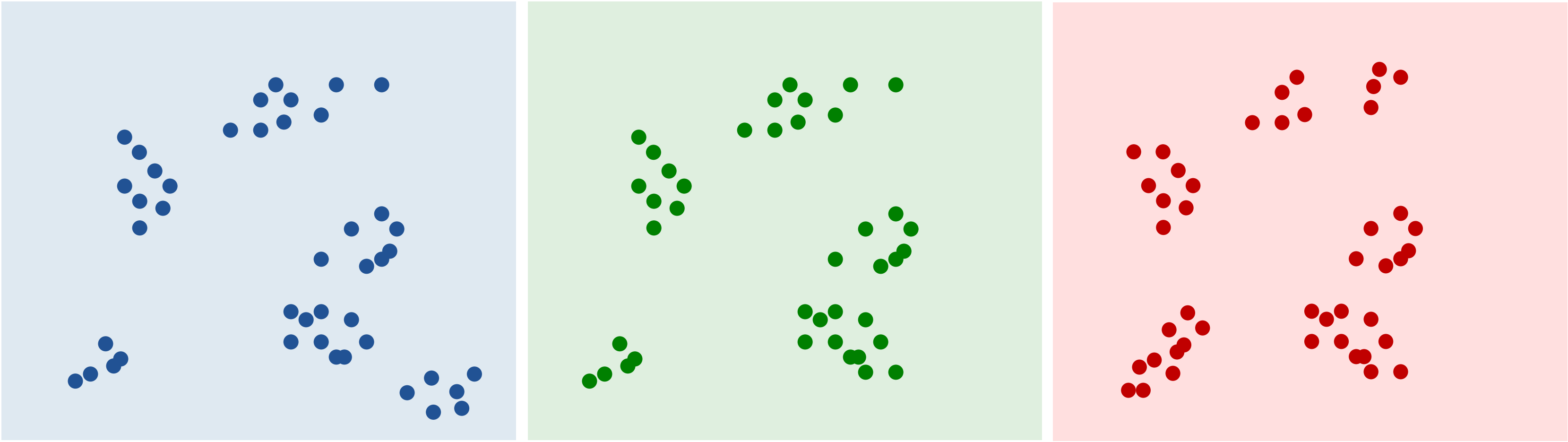
Effect of phosphorylation



Method

Formation of Class Templates

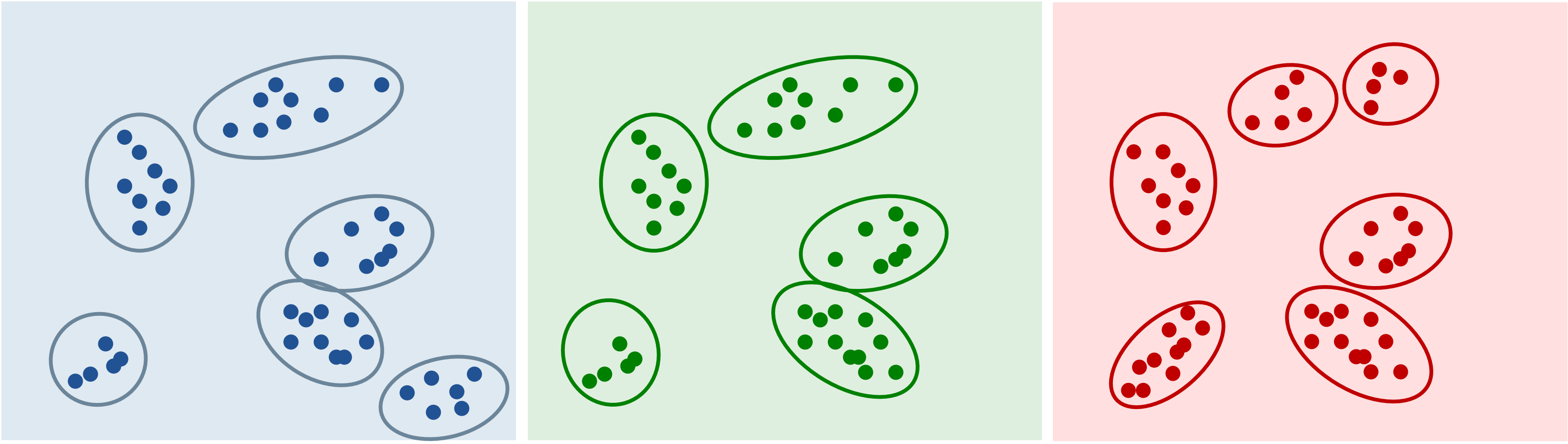
Formation of class template



Sample -1 Sample -2 Sample -3

Visual representation of samples (2D)

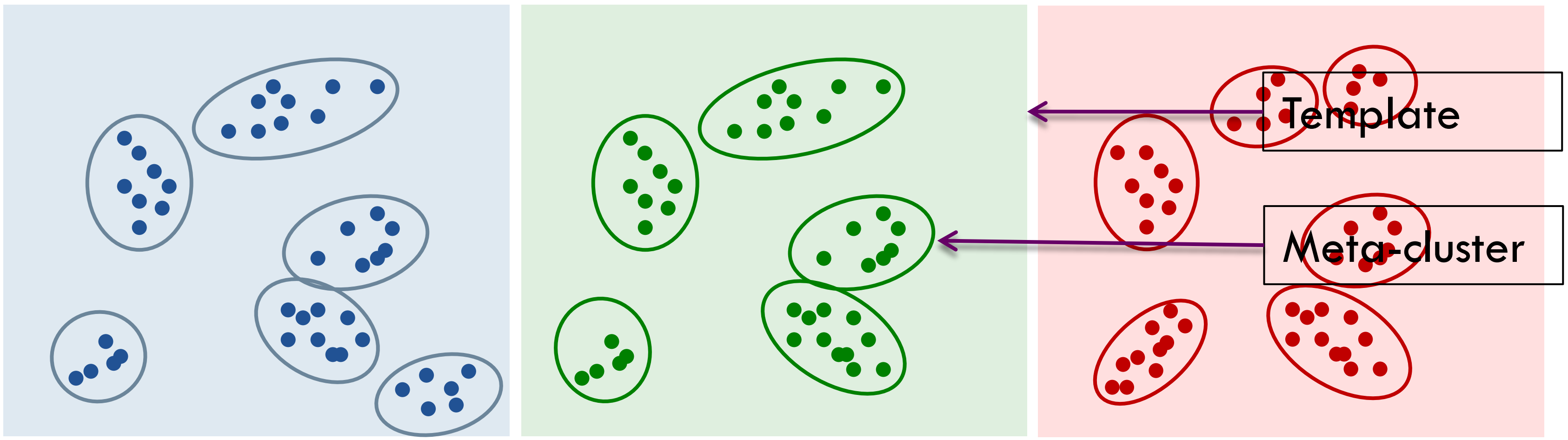
Formation of class template - clustering



Sample -1 Sample -2 Sample -3

Input: N samples. Each sample already clustered.
Sample i has n_i clusters $\{c^i_1, c^i_2, \ldots, c^i_{n\_i}\}$

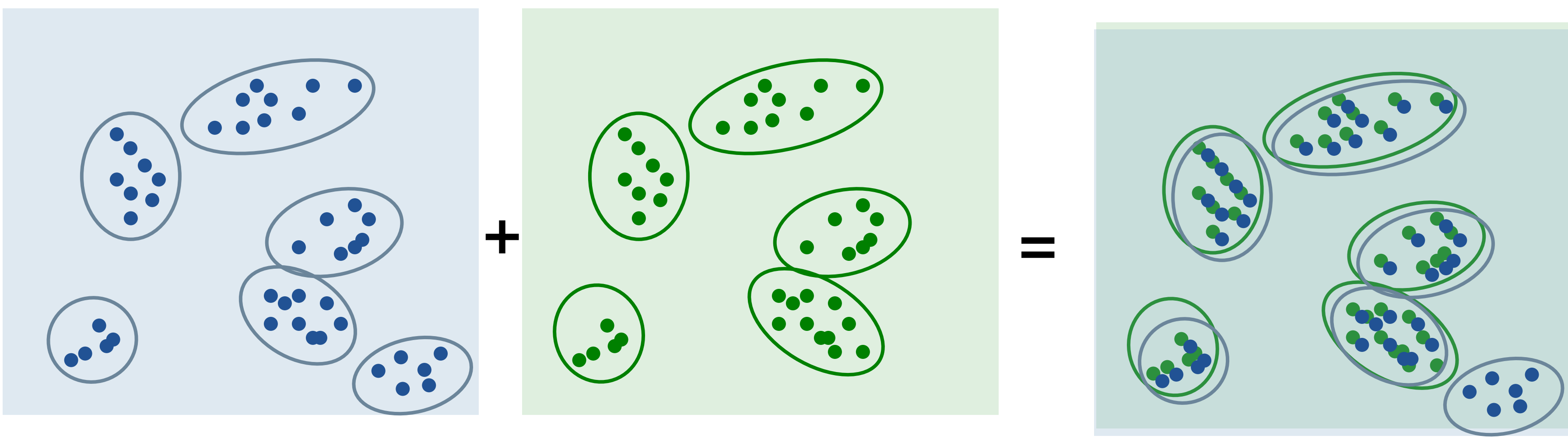
Meta-cluster : cluster of clusters



Output: From the N samples construct a template with k meta-clusters $\{m_1, m_2, \ldots m_k\}$ where every cluster from each samples are assigned to a unique meta-cluster.

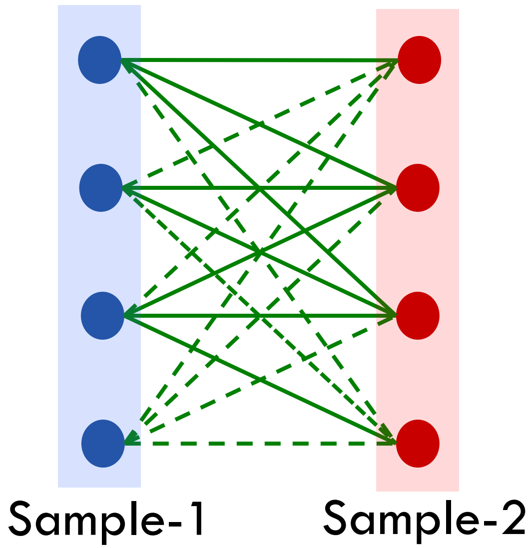
Building block

- Assume we have only two samples.
- We want to merge the two samples to form a template.



Template from two samples:
Generalized Edge Cover in a Bipartite Graph

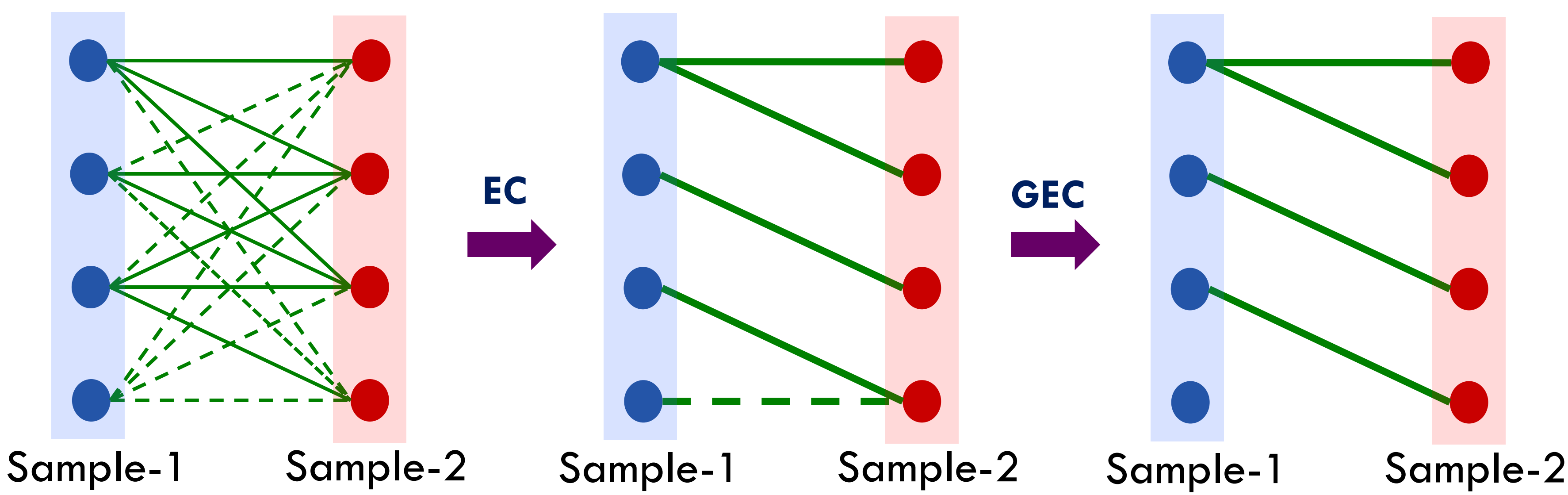
- Construct a complete bipartite graph with clusters with clusters from Sample-1 on one side and from Sample-2 on the other side.
- Edge weights are KL-divergence between corresponding clusters.
- Edges with large weights are shown in dashed line.



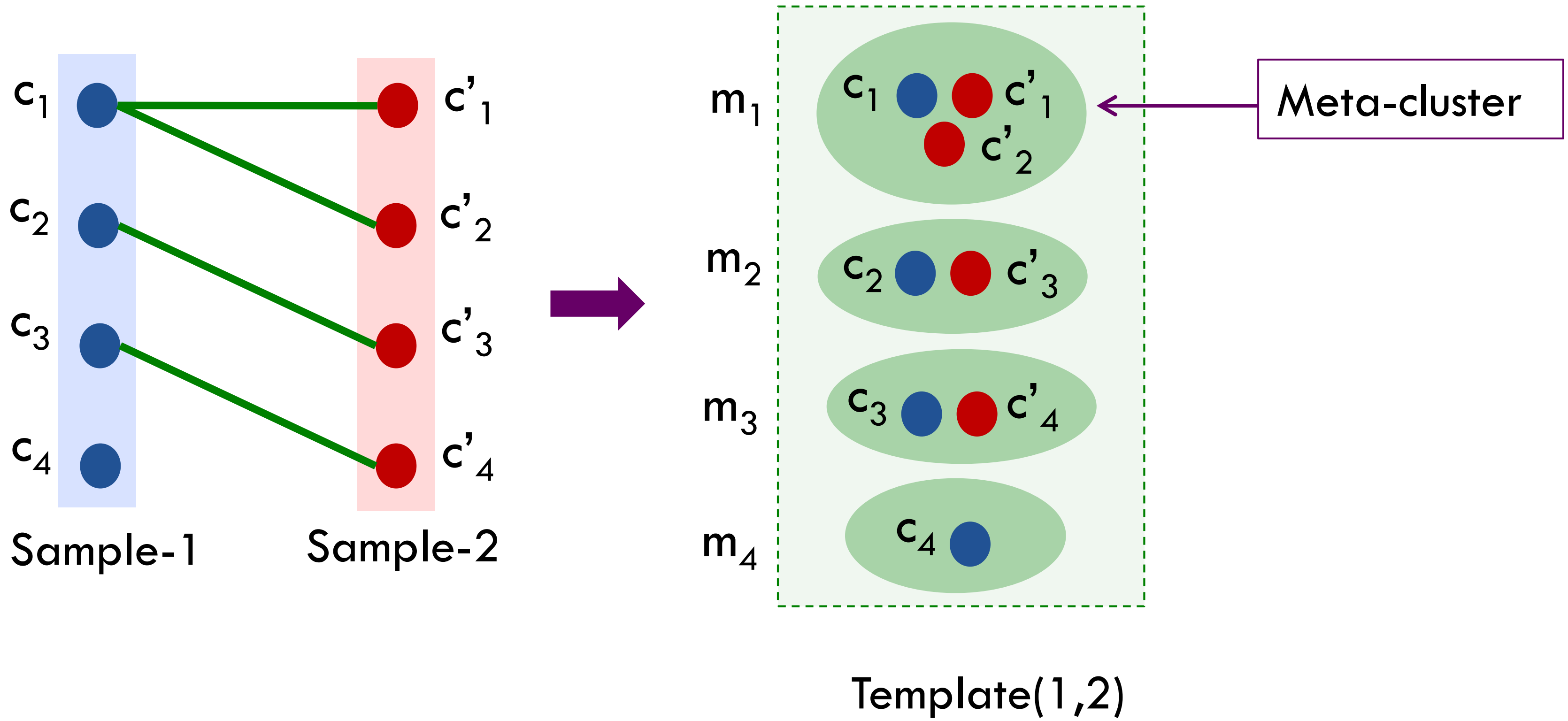
Template from two samples:

Generalized Edge Cover in a Bipartite Graph

- A generalized edge cover (GEC) is an edge cover which allows few uncovered vertices at the cost of a penalty (λ).
- Objective function: $\min\left(\sum_{(v_i,v_j)\in EC} c_{ij} + \lambda * |V_{uc}| \right)$

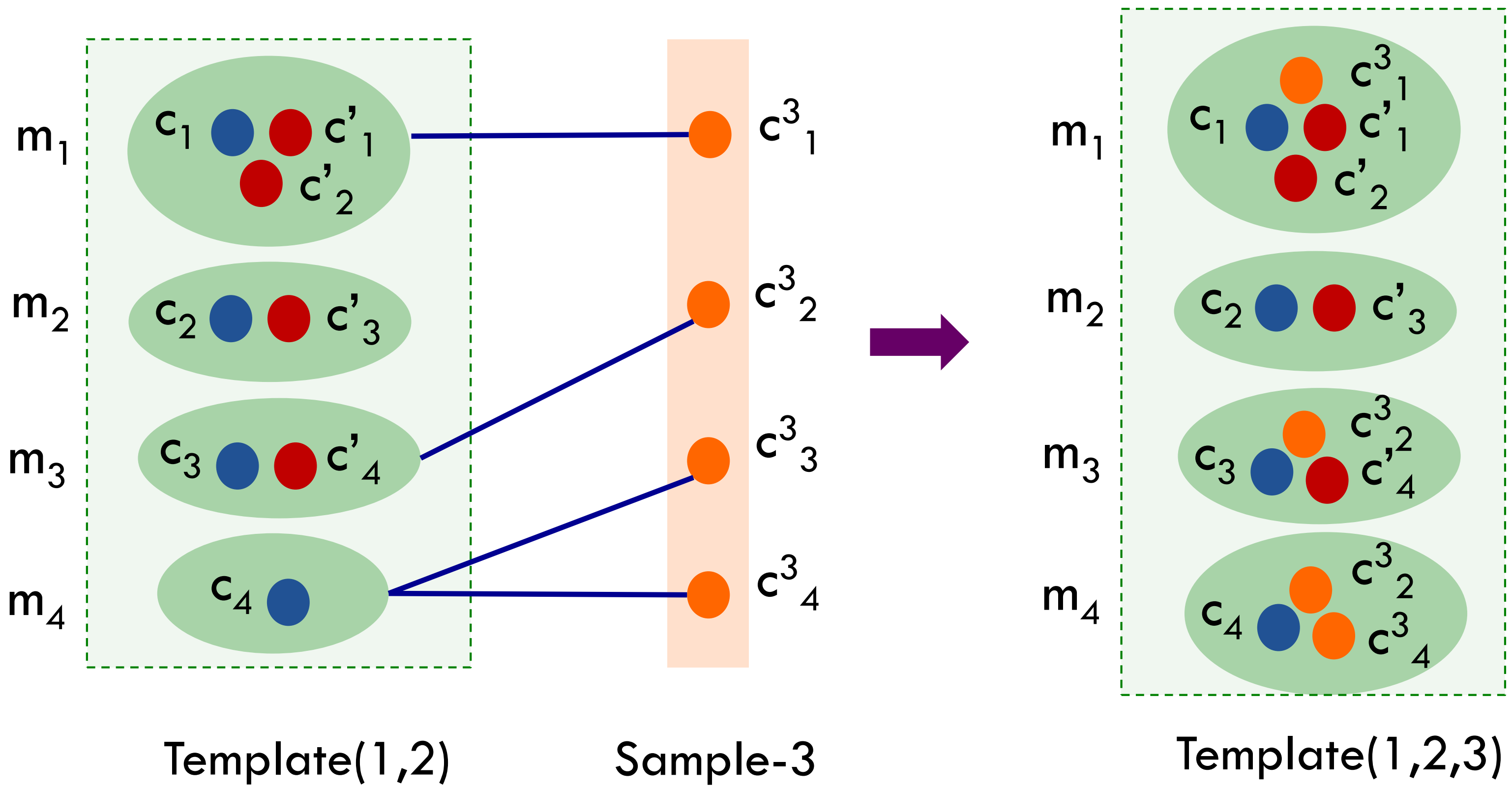


Template from two samples:
Generalized Edge Cover in a Bipartite Graph

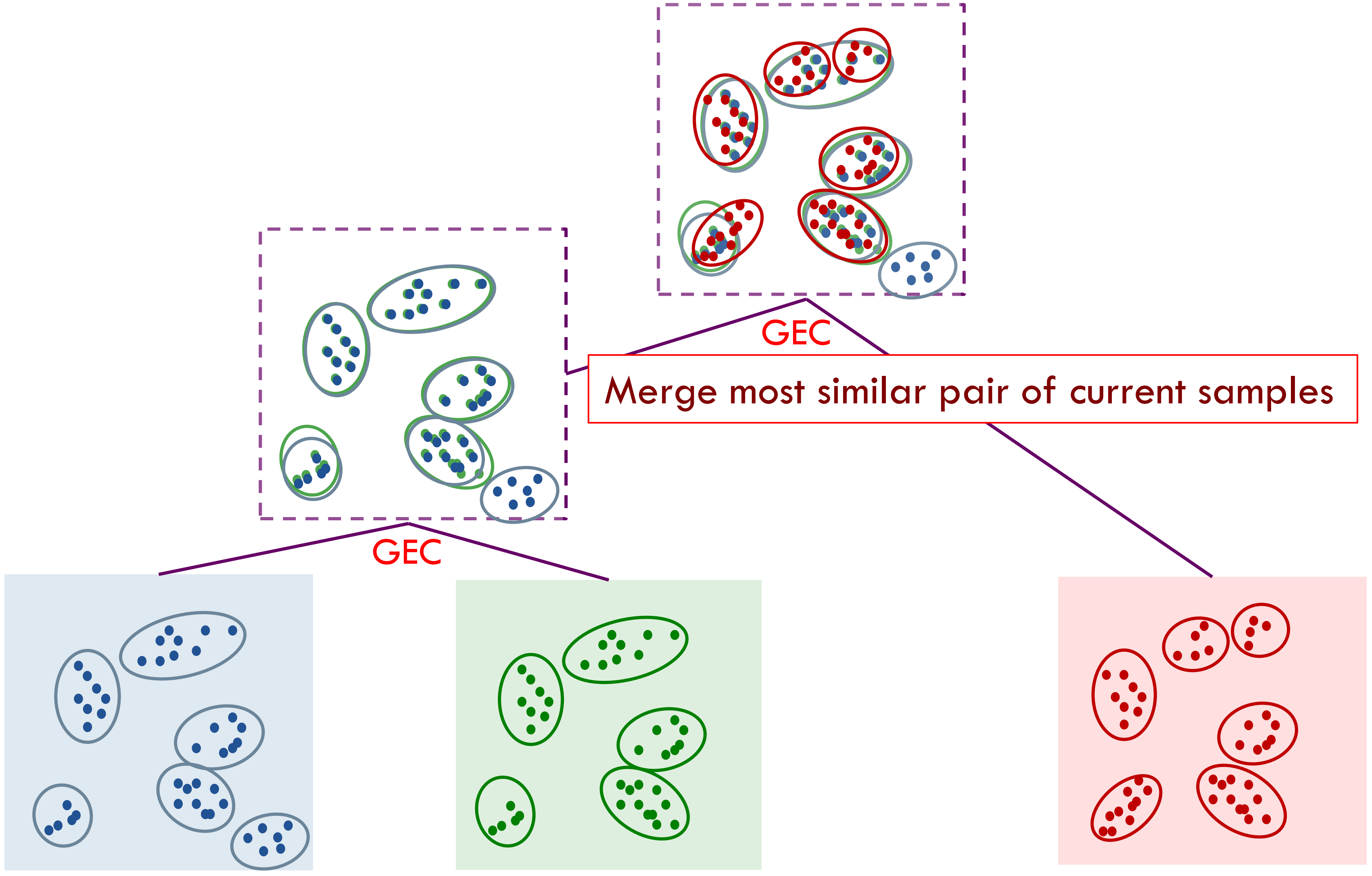


Cascaded Merging of templates

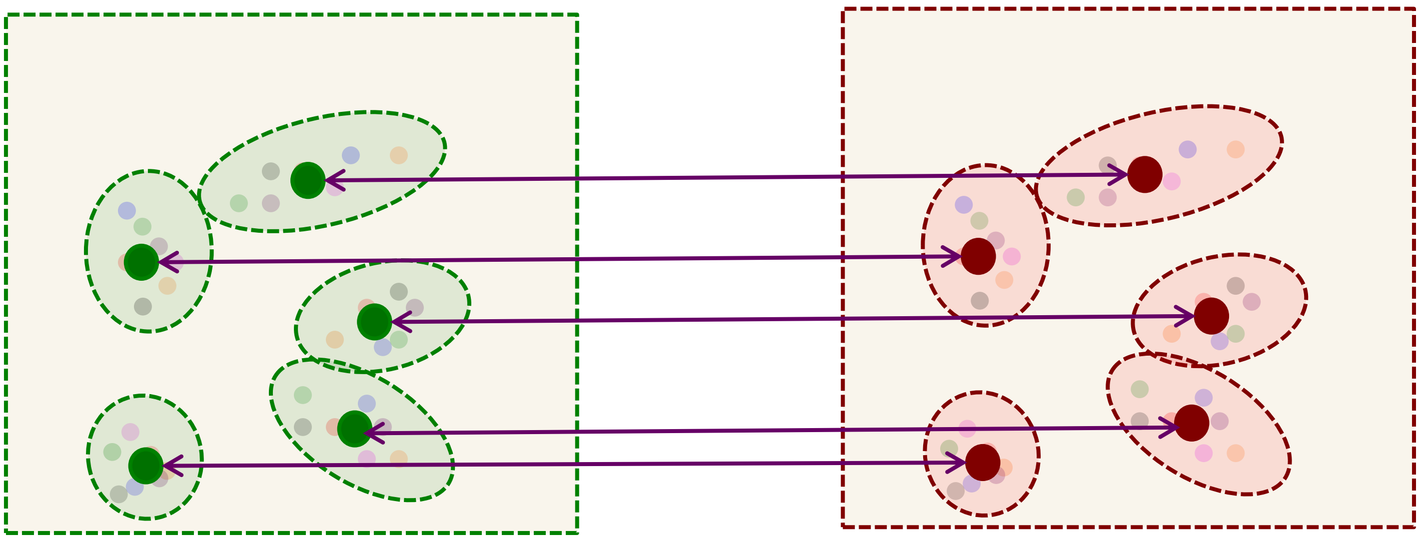
- Template from three samples can done by merging Template(1,2) and Sample-3.



Hierarchical merging to form template



Cross class matching



Template-1
Before phosphorylation

Template-2
After phosphorylation

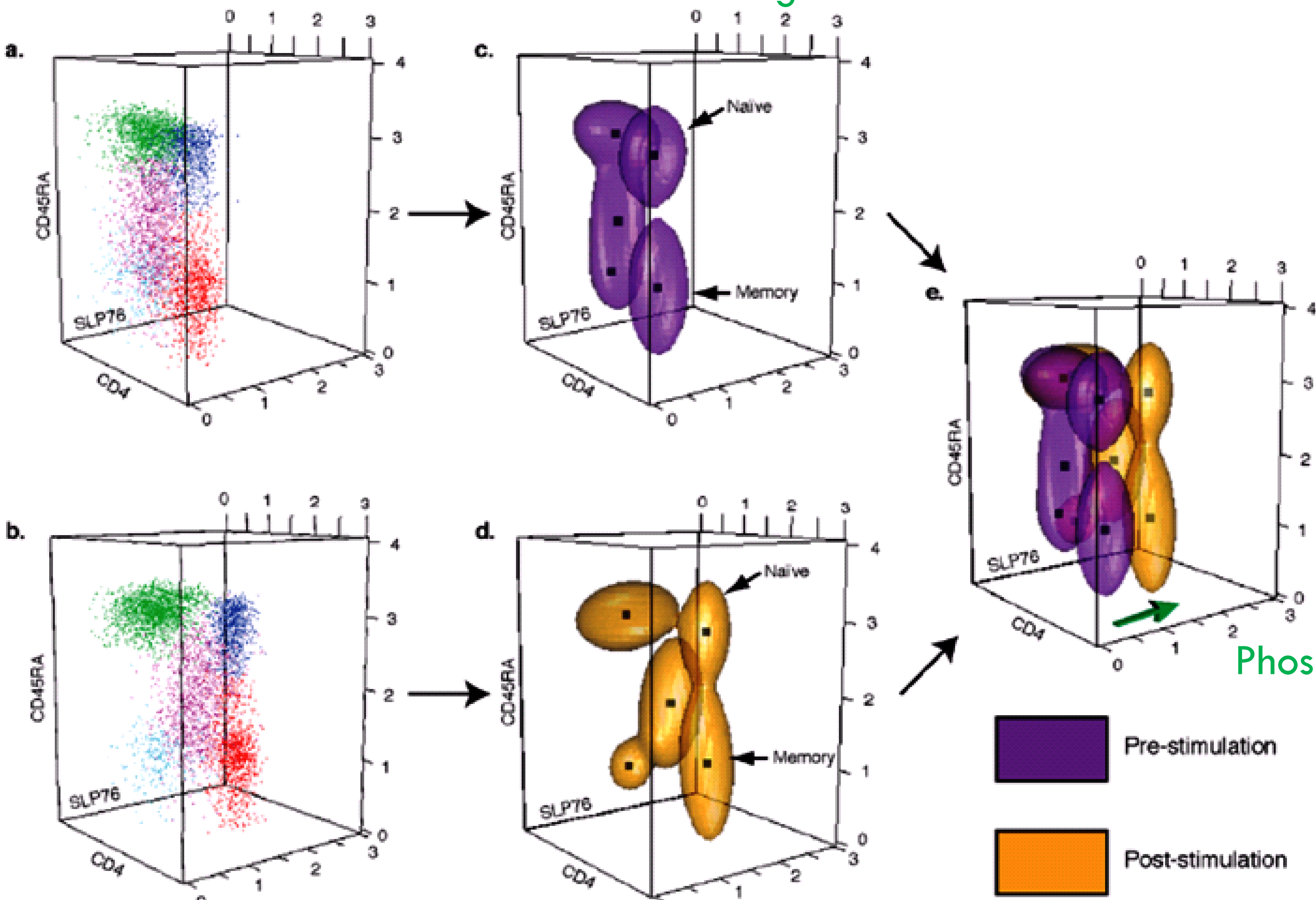
Results

- Detection of Phosphorylation effect in each sample
- Build template before and after phosphorylation
- Detection of Phosphorylation effect in meta-clusters.

Phosphorylation effect on each sample

Clustering samples

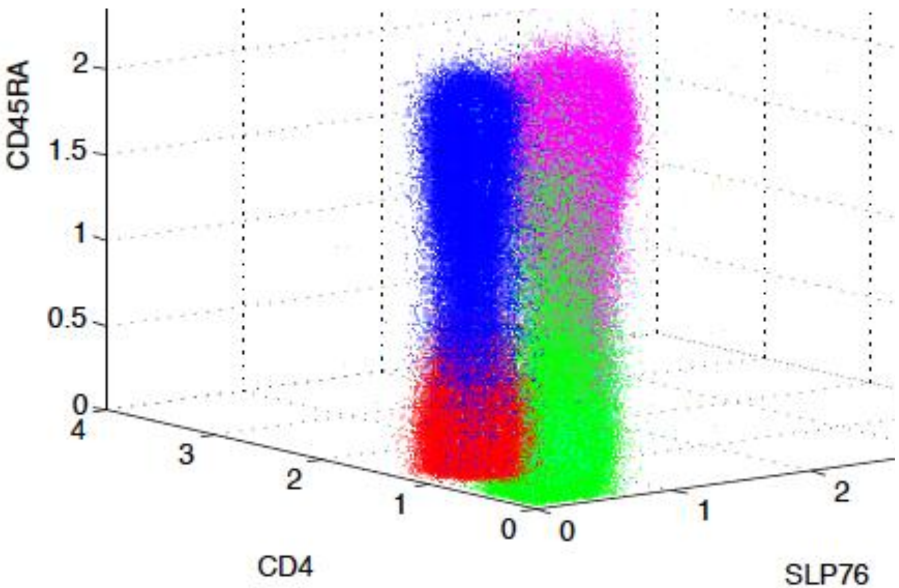
Parametric modeling and matching



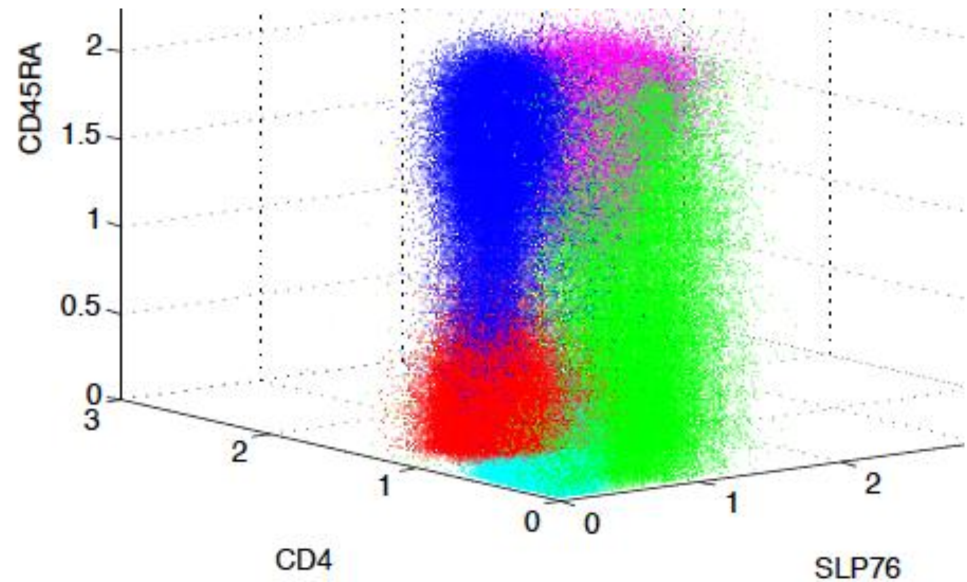
Increased Phosphorylation

Pyne et. al PNAS 2009

Templates

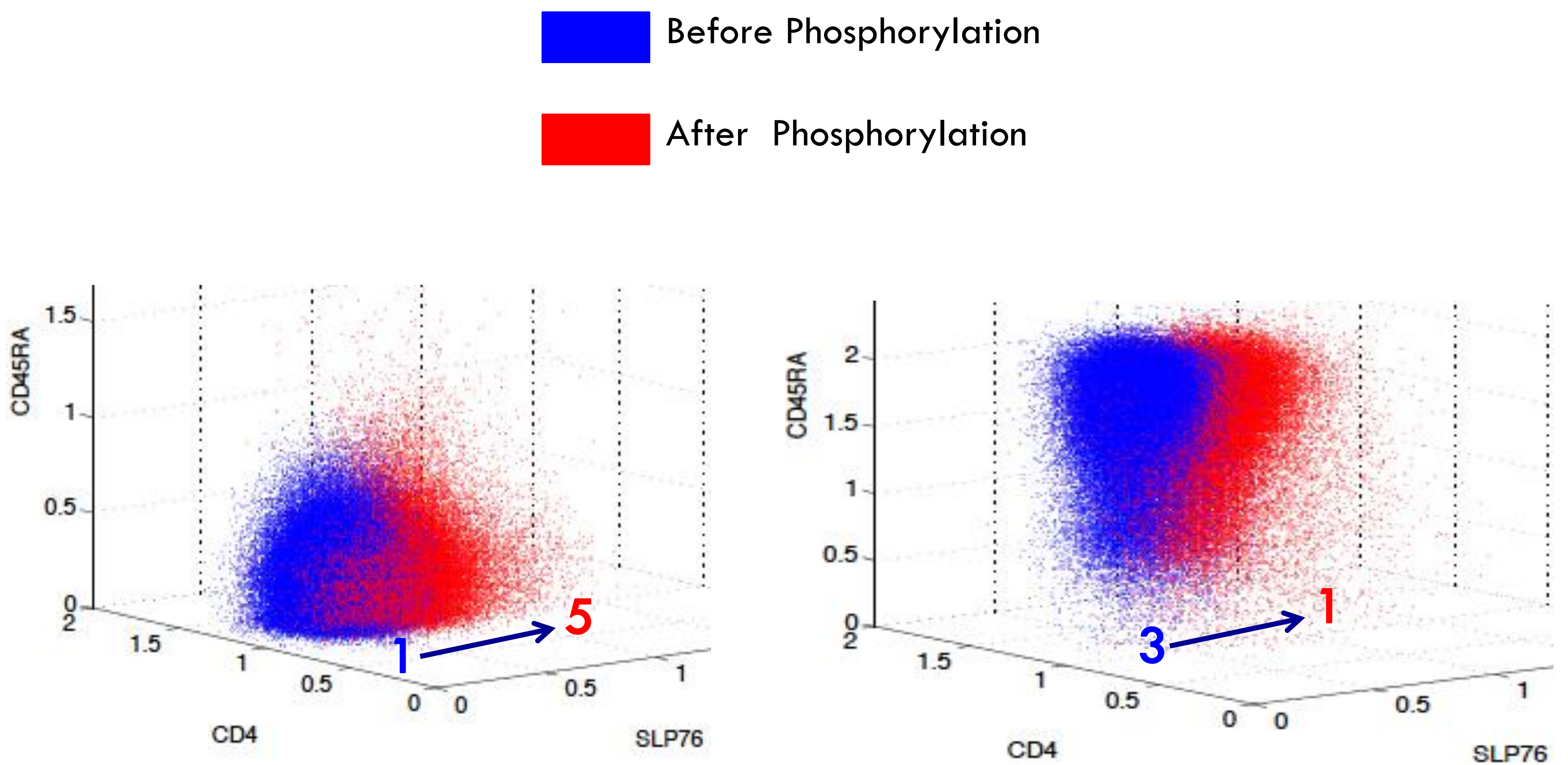


Pre-stimulation Template



Post-stimulation Template

Detection of increased Phosphorylation



Shown for one meta-cluster at a time

Discussion

- Can be considered as a general method of forming template from any non biological images.

Other challenging applications
(change of facial expression)

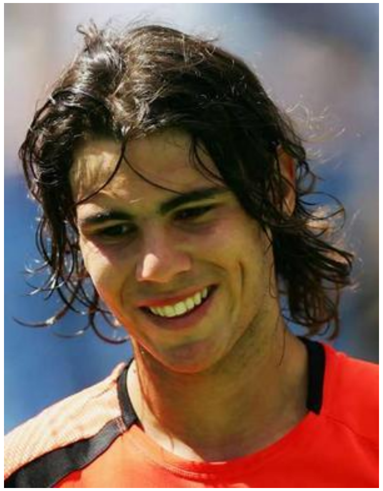
Thank You

- Supporting slides

Other applications

- Image template

+

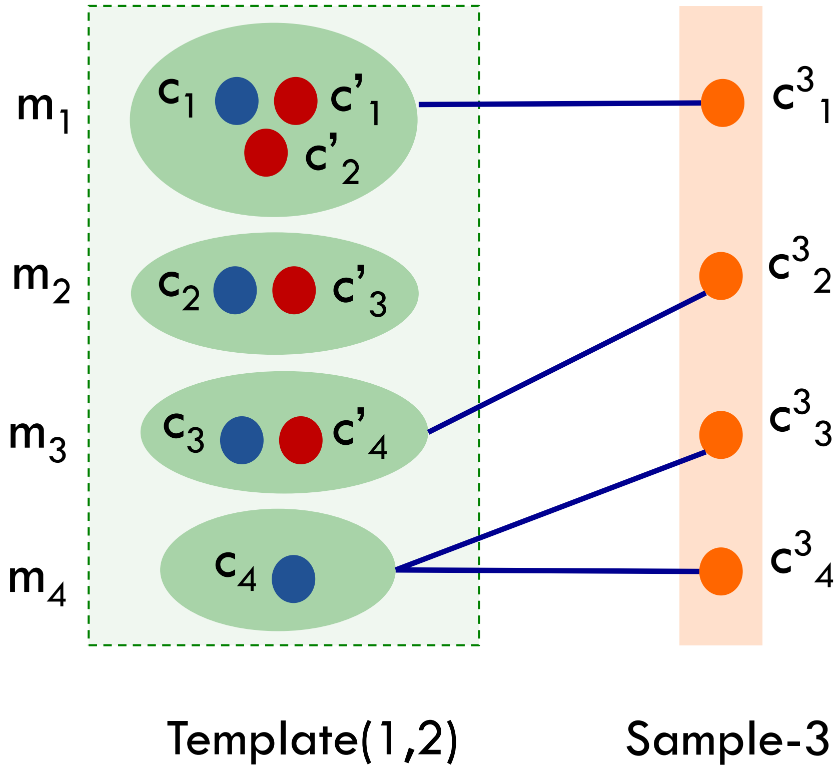


+

= ?

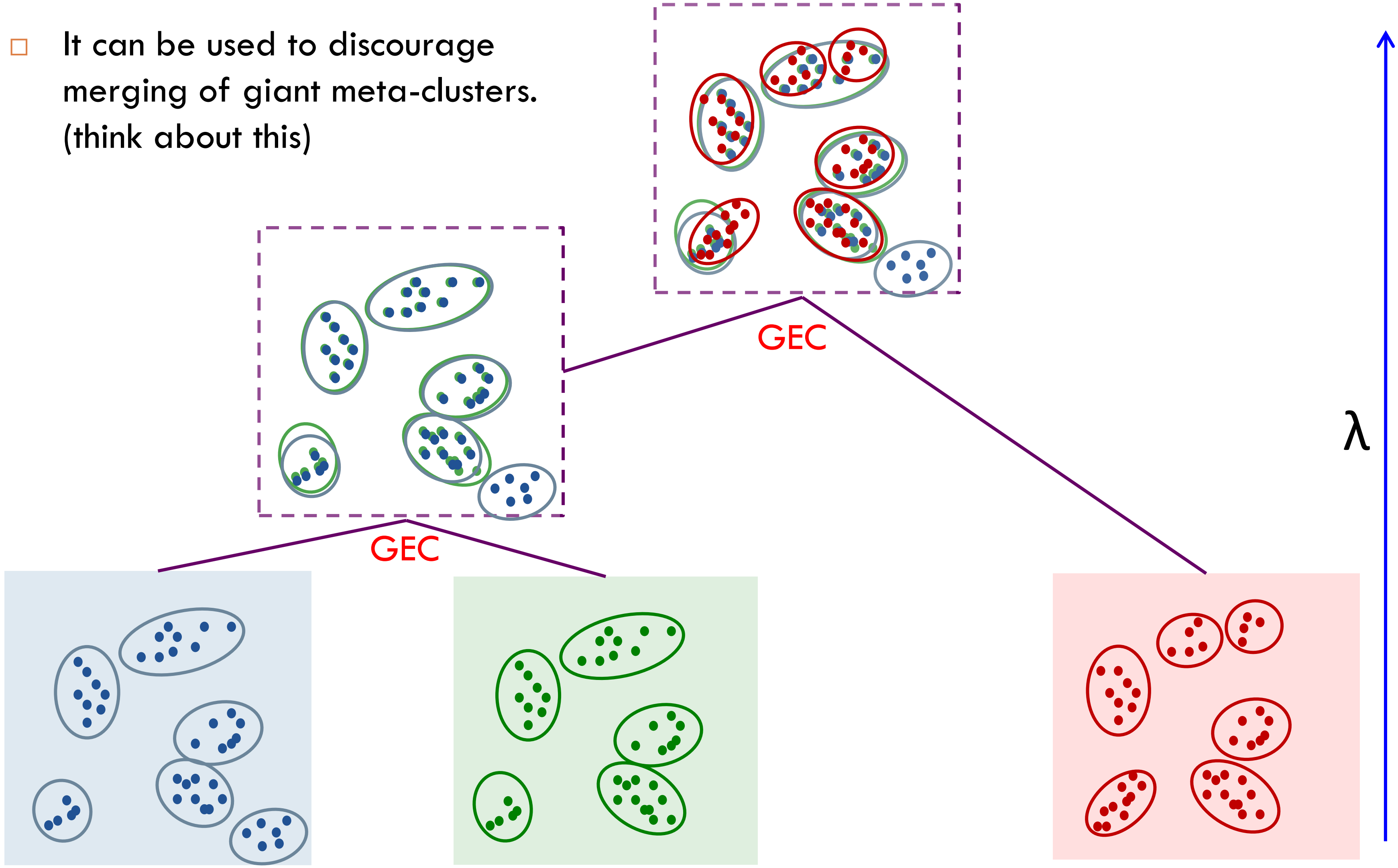
A closer look at GEC

- Number of original clusters present in a meta-cluster can be used as vertex-weight.

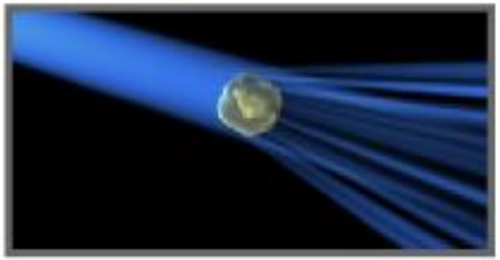


A closer look at (λ)

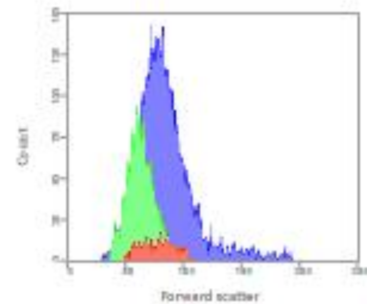
- It can be used to discourage merging of giant meta-clusters. (think about this)



Data collection in Flow cytometry

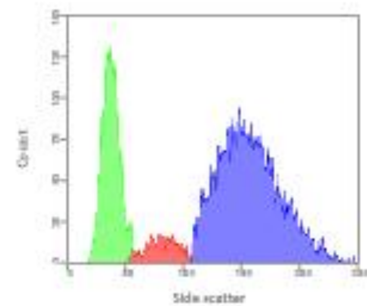


Forward scatter



Size

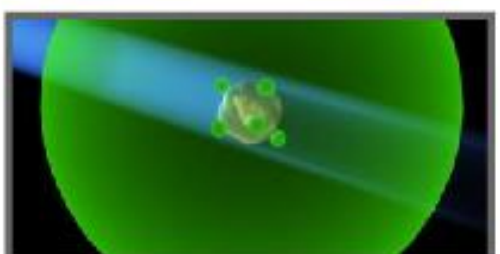
Side scatter



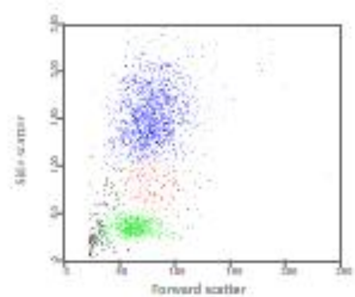
Complexity

Phenotype

Health



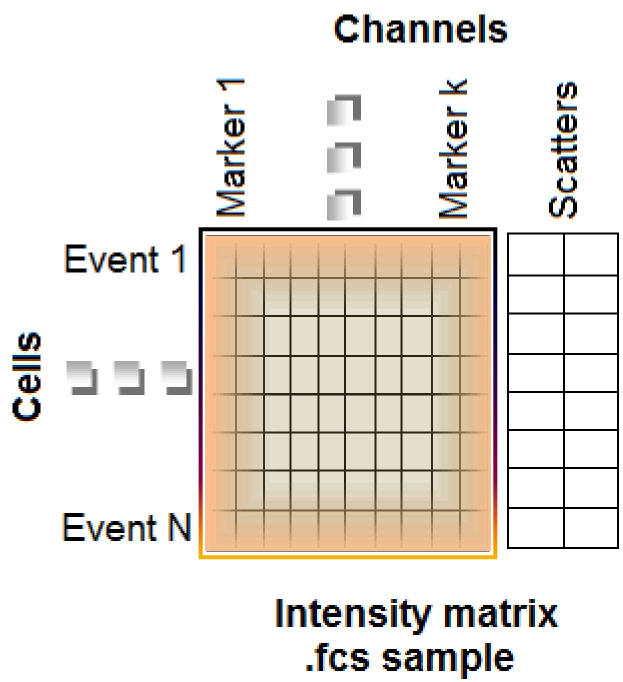
Fluorescence



[Source- http://probes.invitrogen.com/resources/education/tutorials/4Intro_Flow/player.html](http://probes.invitrogen.com/resources/education/tutorials/4Intro_Flow/player.html) (slide 8,9,10)

Output obtained from flow cytrometry

- We obtain multiple properties of a single cell from the experiment.
- Example

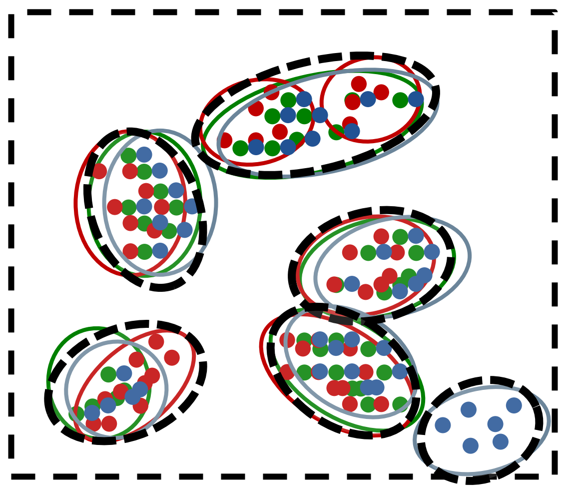


	FSC	SSC	CD4	CD8
Cell-1	2.1	2	3	1
Cell-2	4.5	2.3	12	1
Cell-3	2.3	3.4	3.3	4.4
Cell-4	2.2	2.1	3.1	3.3
…..	…	…	…	…

Two extreme solutions

1. Fine grained meta-clustering

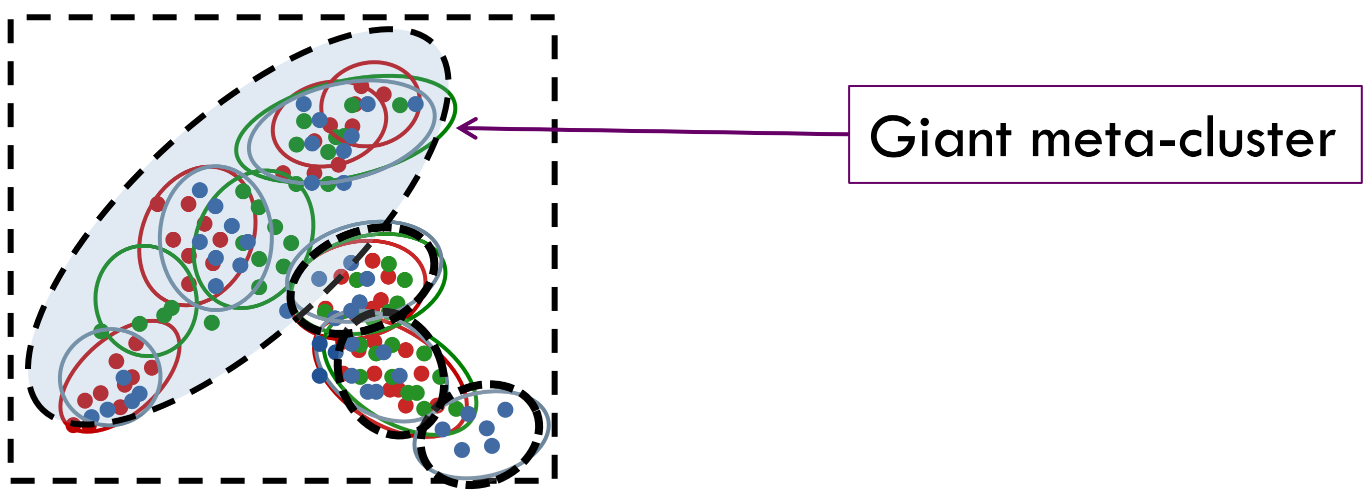
- Cluster all points from every sample together to form meta-cluster
- Assign cluster c^i_j to meta-cluster m_k if majority of the elements in c^i_j belongs to m_k .



Two extreme solutions

1. Fine grained meta-clustering

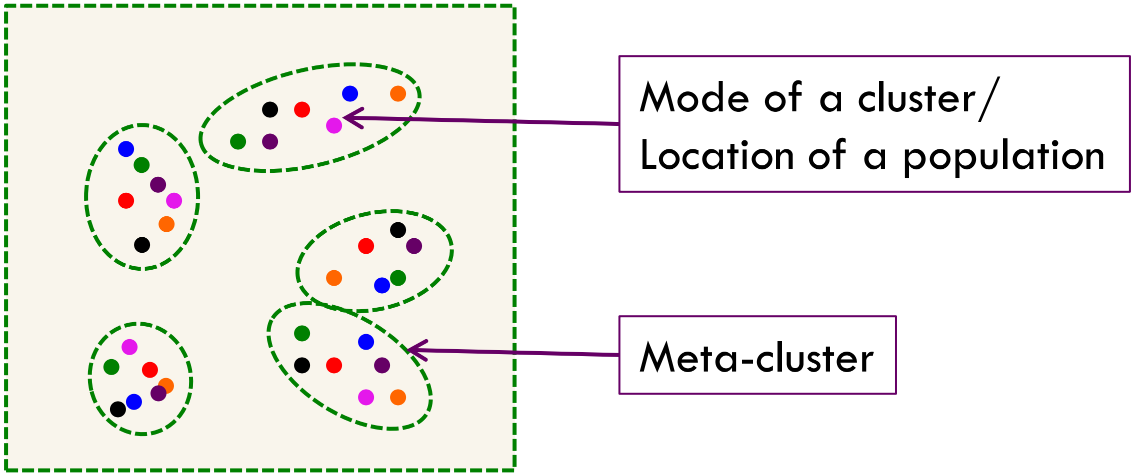
- Limitations:
  - For many samples computationally too expensive.
  - Giant Meta-clusters problem.
  - Single cluster may be spread over many meta-clusters



Two extreme solutions

2. coarse grained meta-clustering

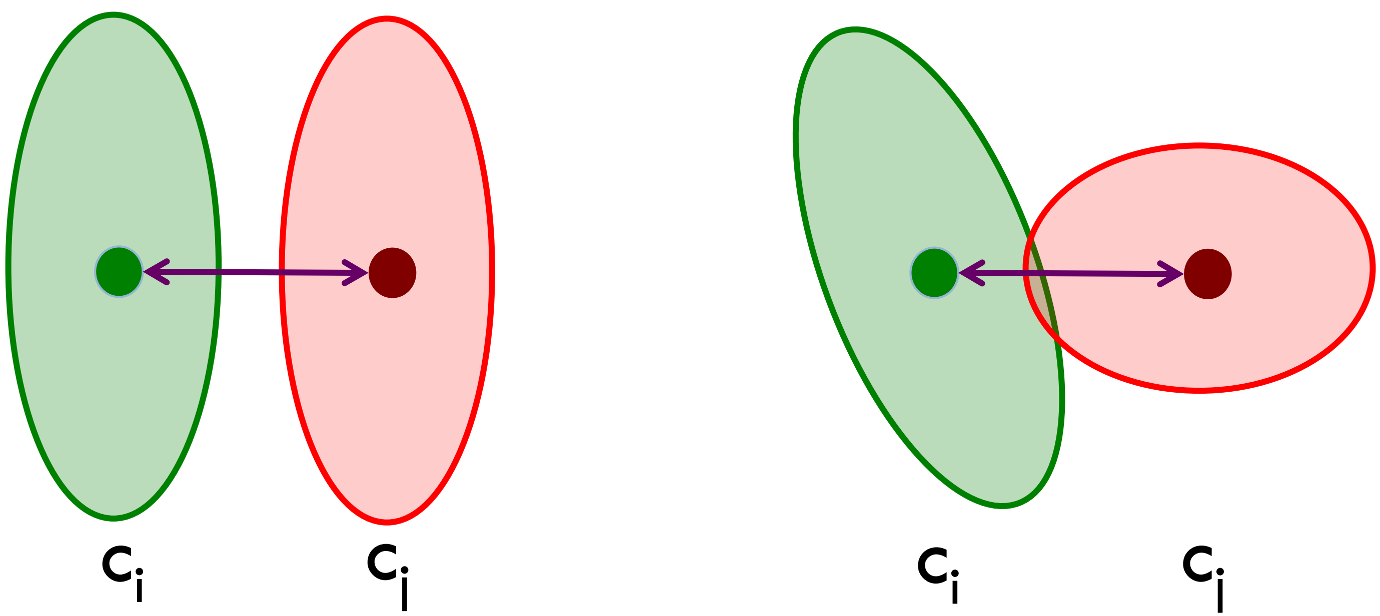
- Cluster centers/locations of all clusters from each sample to form meta-cluster.
- Assign cluster c^i_j to meta-cluster m_k if center of c^i_j belongs to m_k .



Template

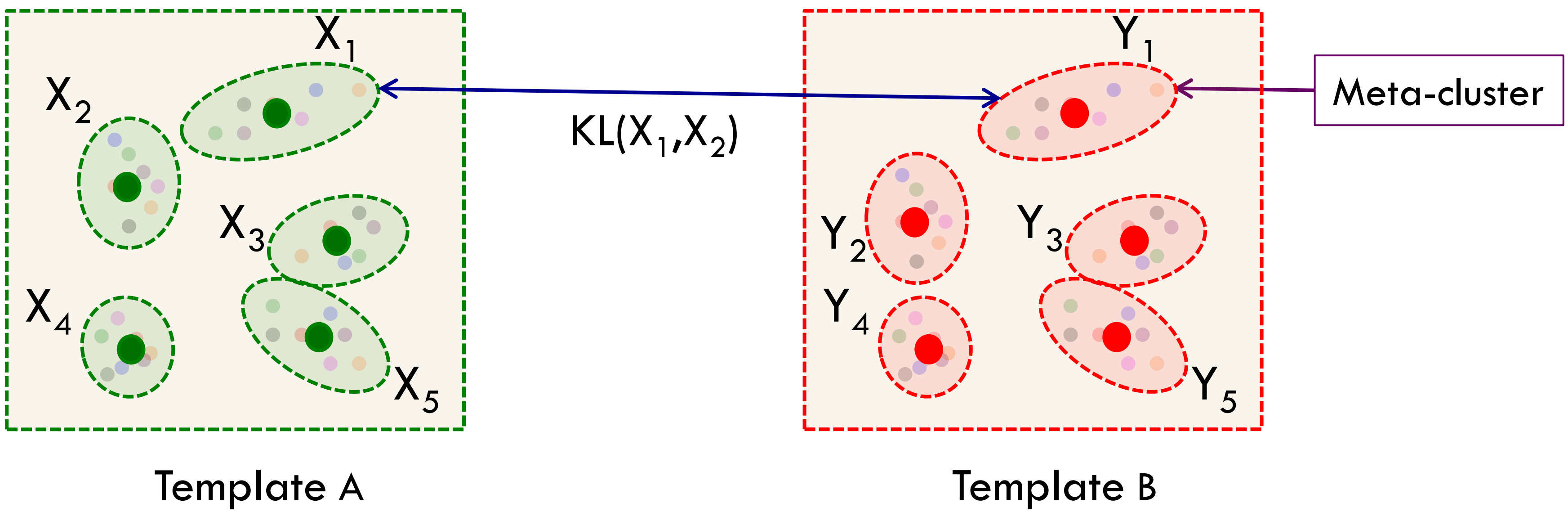
Dissimilarity / distance measure

- Consider the location and shape of every cluster during meta-clustering.
- Dissimilarity measure or distance should be different.
- Use KL-divergence or mutual information



Issues: distance between meta-clusters

2. KL divergence of the distribution of the meta-clusters (merging of distributions are required after matching)

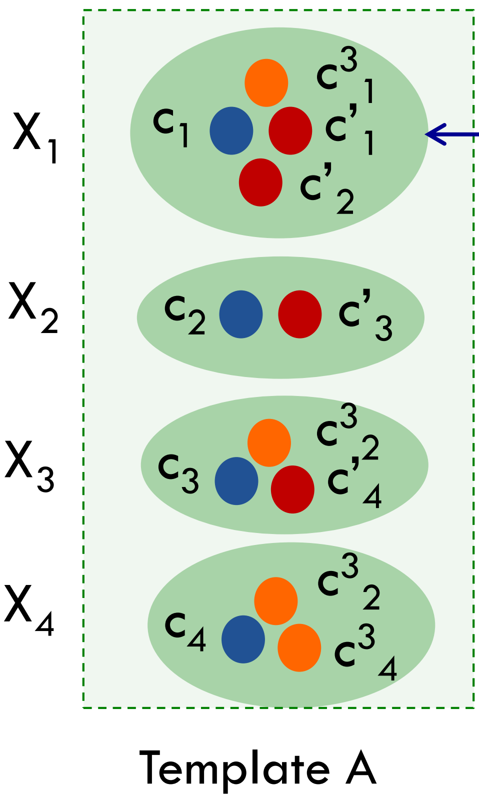


Hierarchical merging of templates

- We know how to merge templates
  - After merging what will be the distance from the new meta-clusters to (meta-)clusters in other templates/sample ?
  - What will be the order of merging ?

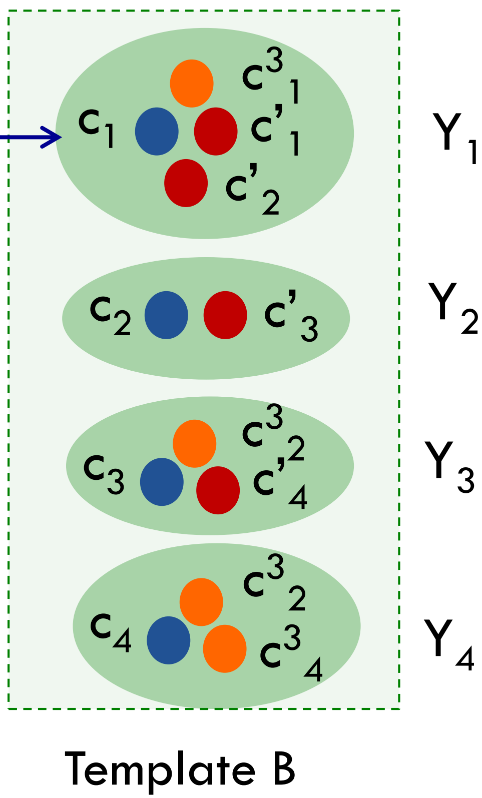
Issues: distance between meta-clusters

Average of the distances of the members



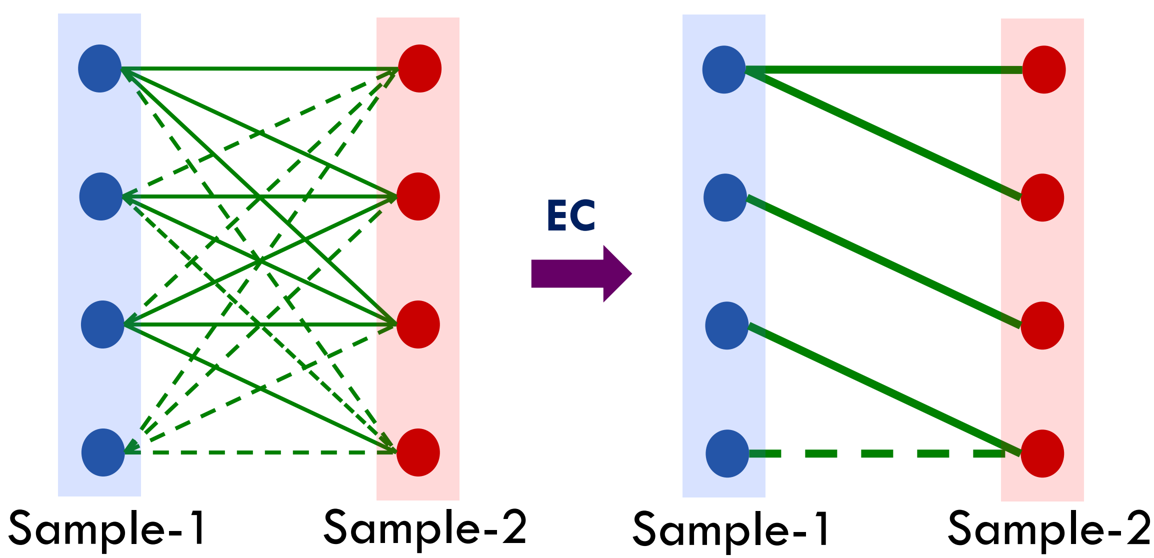
$$\frac{1}{|X_1|.|Y_1|}\sum_{x\in X_1}\sum_{y\in Y_1} d(x,y)$$

$d(x,y) = ½[KL(x,y)+KL(y,x)]$



Template from two samples: Generalized Edge Cover in a Bipartite Graph

- A minimum-weight edge cover (EC) is a subset of edges such that every vertex is incident to at least one edge of the set and summation of weight of edges in the set is minimum.
- Objective function: $\min \left(\sum_{(v_i, v_j) \in EC} c_{ij} \right)$



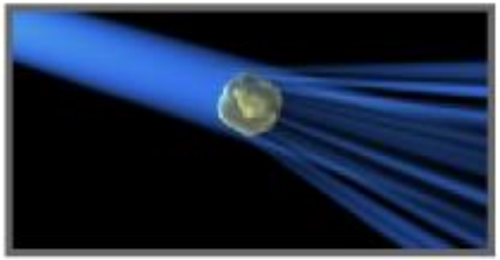
Issue-2: Order of template merging

- Use a measure of partition distance
  - Small if two partitions are similar, large if dissimilar
  - Partition distance as defined in Gasfield (2003) can not be used
- Covering distance: cost of Generalized edge cover.
  - Can be used as partition distance in this case
- Merge two templates with minimum covering distance.

Time complexity

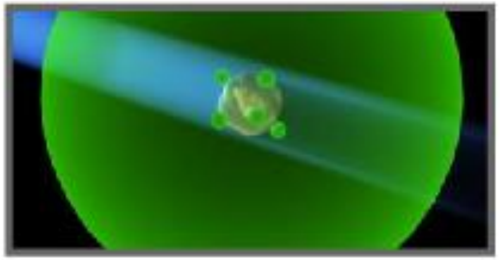
- Assume N samples.
- $O(N^2)$ initial GEC computation
- During merging of templates total number of GEC computation = $(N-1)+(N-2)+\ldots+1 = O(N^2)$
- Time to compute each GEC is $O(k^3\log(k))$ where k is the number of (meta-) clusters in a sample.
- Generally k is not large

Data collection in Flow cytometry



Forward scatter

Side scatter



Fluorescence

Source- http://probes.invitrogen.com/resources/education/tutorials/4Intro_Flow/player.html (slide 8,9,10)