

# Community Detection from Genomic Dataset Across Human Cancers

Nandinee Haq ([nandinee@ece.ubc.ca](mailto:nandinee@ece.ubc.ca)),  
Z. Jane Wang ([zjanew@ece.ubc.ca](mailto:zjanew@ece.ubc.ca))

University of British Columbia, Vancouver, Canada

# Outline

- ❑ Motivation
- ❑ Community Detection Across Human Cancers
  - ❑ Dataset
  - ❑ Network Generation & Community Detection
  - ❑ Results
- ❑ Conclusion
- ❑ Future Works

# Motivation

- ❑ Cancer is driven by a combination of genes
- ❑ Gene combinations can vary across patients
- ❑ Tumor stratification for different tumor types is an active field of research where a population of tumors is divided into biologically meaningful subtypes.
- ❑ Most of these studies were carried out on tumors originating from the same organ

# Motivation

- ❑ Two important observations-
  - ❑ tumors originating from the same organ can significantly vary across subjects<sup>1</sup>
  - ❑ similar genomic alteration patterns can be observed across tumors originating from different tissues <sup>2</sup>

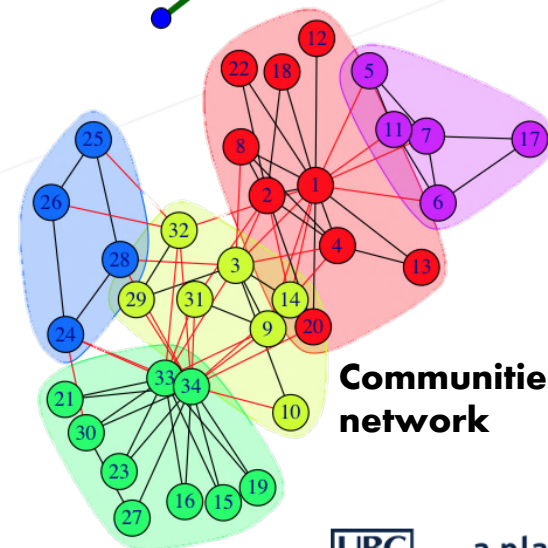
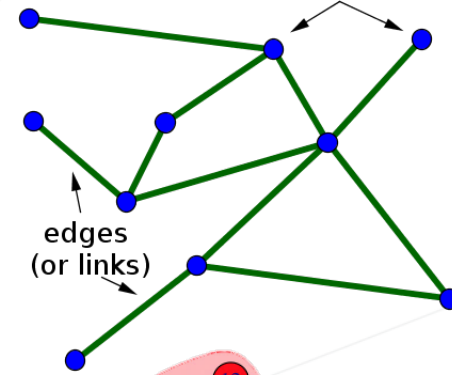
# Motivation

- To develop a data-driven technique for the cancer stratification problem by classifying cancers *independent of their origins* using the community detection approach

# Community Detection Across Human Cancers

- ❑ Tissue-independent tumor stratification is formulated as a weighted-graph clustering problem
- ❑ Network generated from subjects
- ❑ Community detection method applied for tumor clustering

**A simple network**  
nodes (or vertices)



# Genomic Dataset

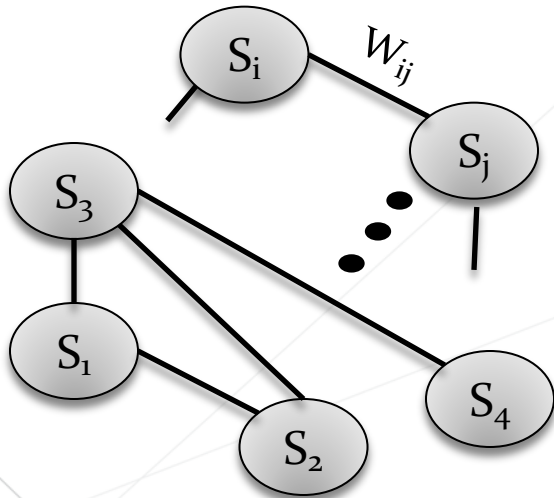
- Genomic features from around 3200 subjects, 12 cancer types from TCGA<sup>3</sup>
- 479 features<sup>4</sup>- 151 copy number losses, 116 copy number gains, 199 mutation and 13 methylation features
- Task is to find “*biologically meaningful*” clusters

Tumor Type	Number of cases
Bladder urothelial carcinoma (BLCA)	95
Breast invasive carcinoma (BRCA)	466
Colon and rectum adenocarcinoma (COREAD)	489
Glioblastoma multiformae (GBM)	216
Head and neck squamous cell carcinoma (HNSC)	299
Kidney renal clear-cell carcinoma (KIRC)	378
Acute myeloid leukemia (LAML)	164
Lung adenocarcinoma (LUAD)	224
Lung squamous cell carcinoma (LUSC)	182
Ovarian serous cystadenocarcinoma (OV)	445
Uterine corpus endometrioid carcinoma (UCEC)	241

# Network Generation

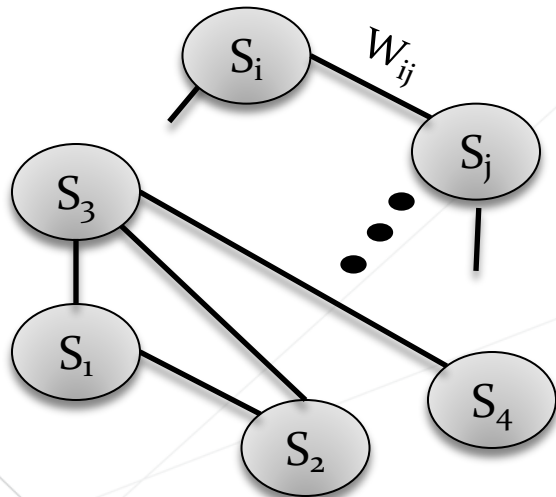
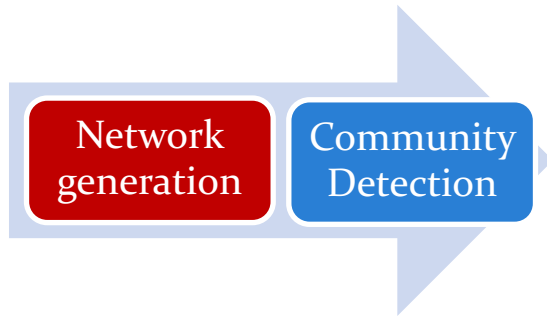
Network  
generation

Community  
Detection



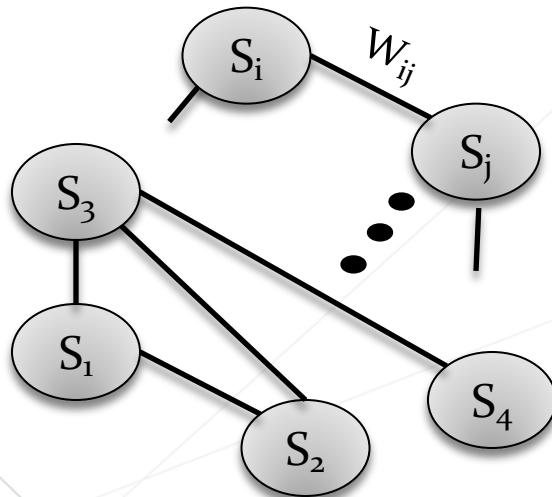
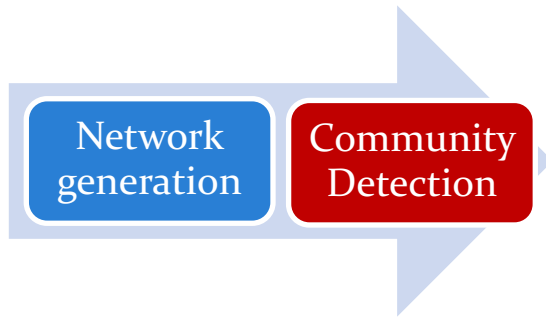


# Network Generation



- Subjects are considered as nodes
- Edge is drawn if two subjects have at least one common alteration.
- $W_{ij} = \sum_p C_{i,p} * C_{j,p}$   
 $W_{ij}$  = weight between subjects  $i$  and  $j$   
 $C_{i,p} \in [0,1]$  ;  
 $C_{i,p} = 1$  when feature  $- p$  is present in subject  $- i$
- Resulting graph has 3200 nodes and 1,851,740 edges.

# Community Detection

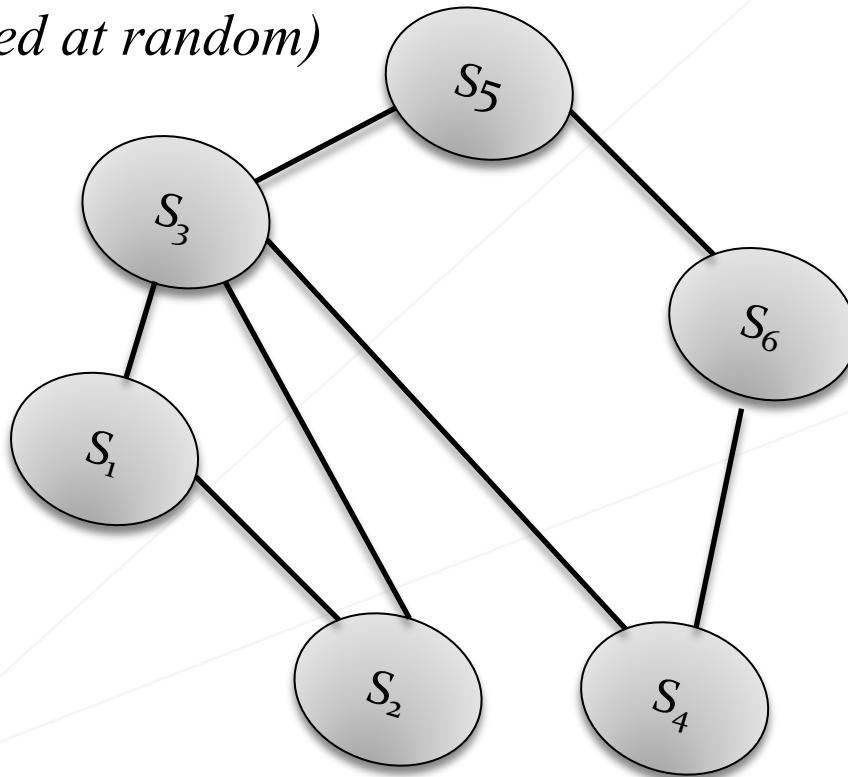


- ❑ Louvain Method<sup>5</sup> based on modularity maximization
- ❑ Does not require number of communities
- ❑ suitable for large graphs
- ❑ shown to outperform other modularity-based methods
- ❑ computationally less expensive

# Louvain Community Detection

Modularity : *(fraction of edges that fall within group) – (expected fraction if edges were distributed at random)*

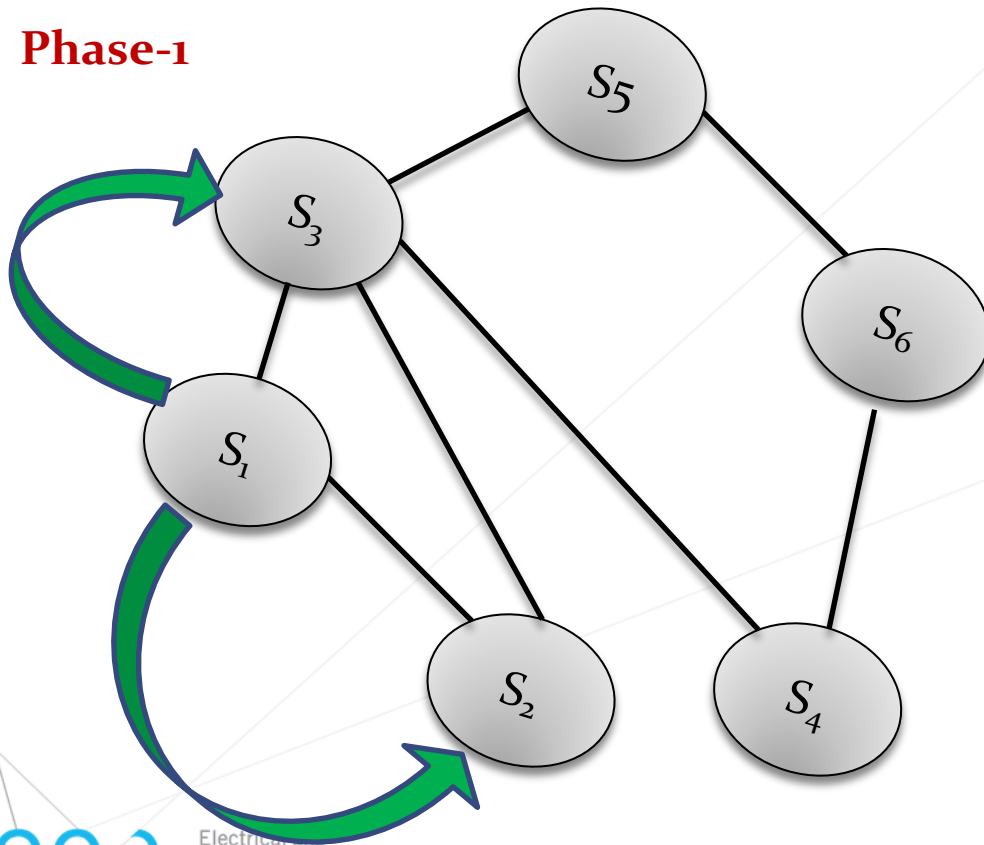
**Phase-1**



a different community is assigned to each node of the network.

# Louvain Community Detection

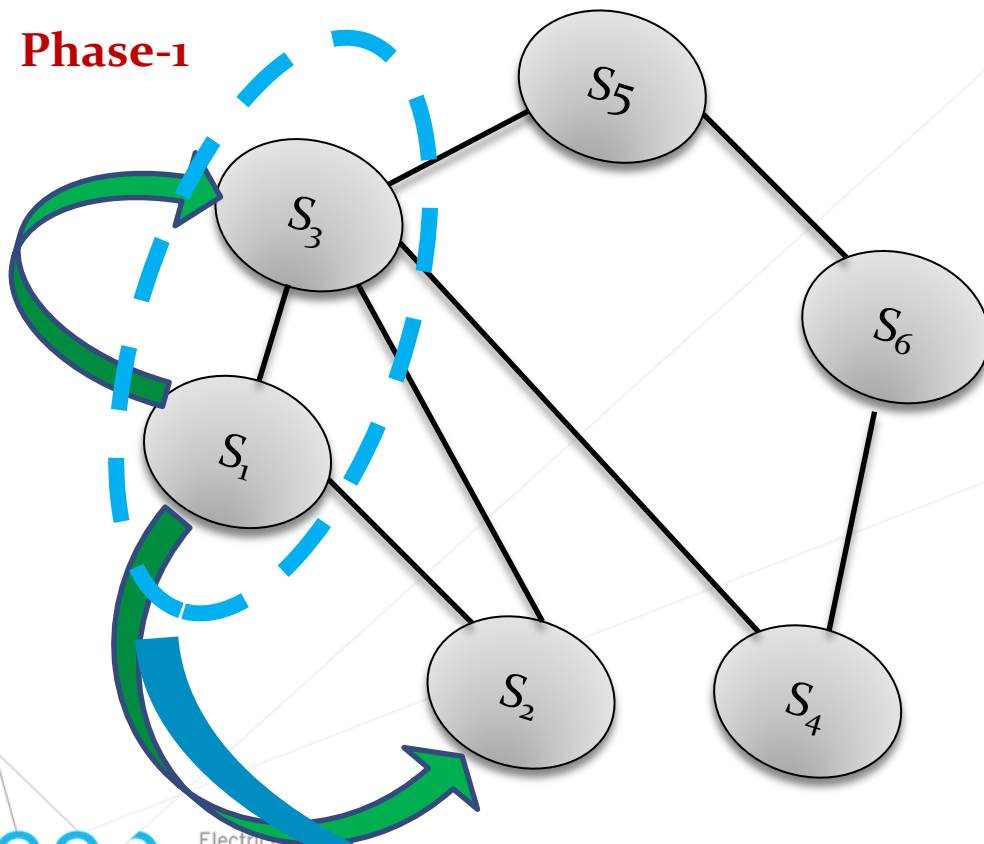
Phase-1



Calculate modularity gain if each node is placed in the community of each of its neighboring nodes

# Louvain Community Detection

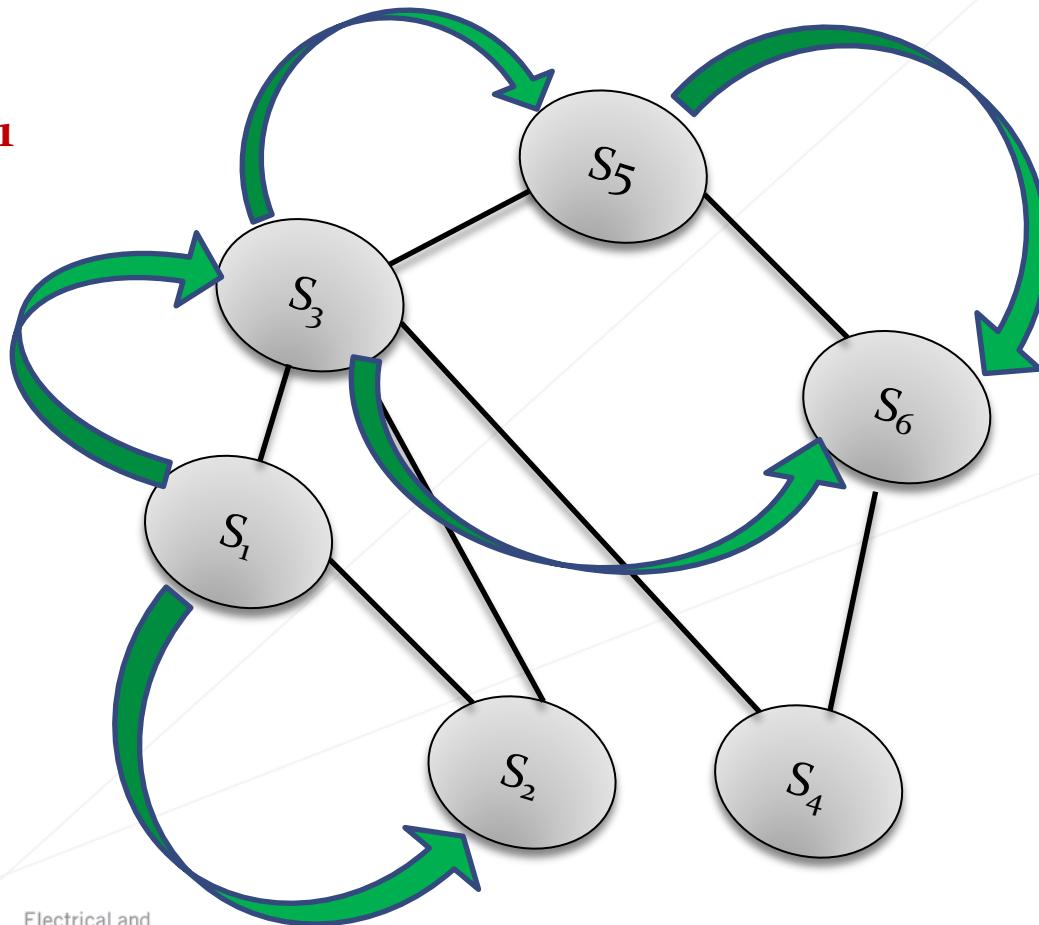
Phase-1



The node is finally placed to its neighbouring node for which maximum positive modularity gain is achieved.

# Louvain Community Detection

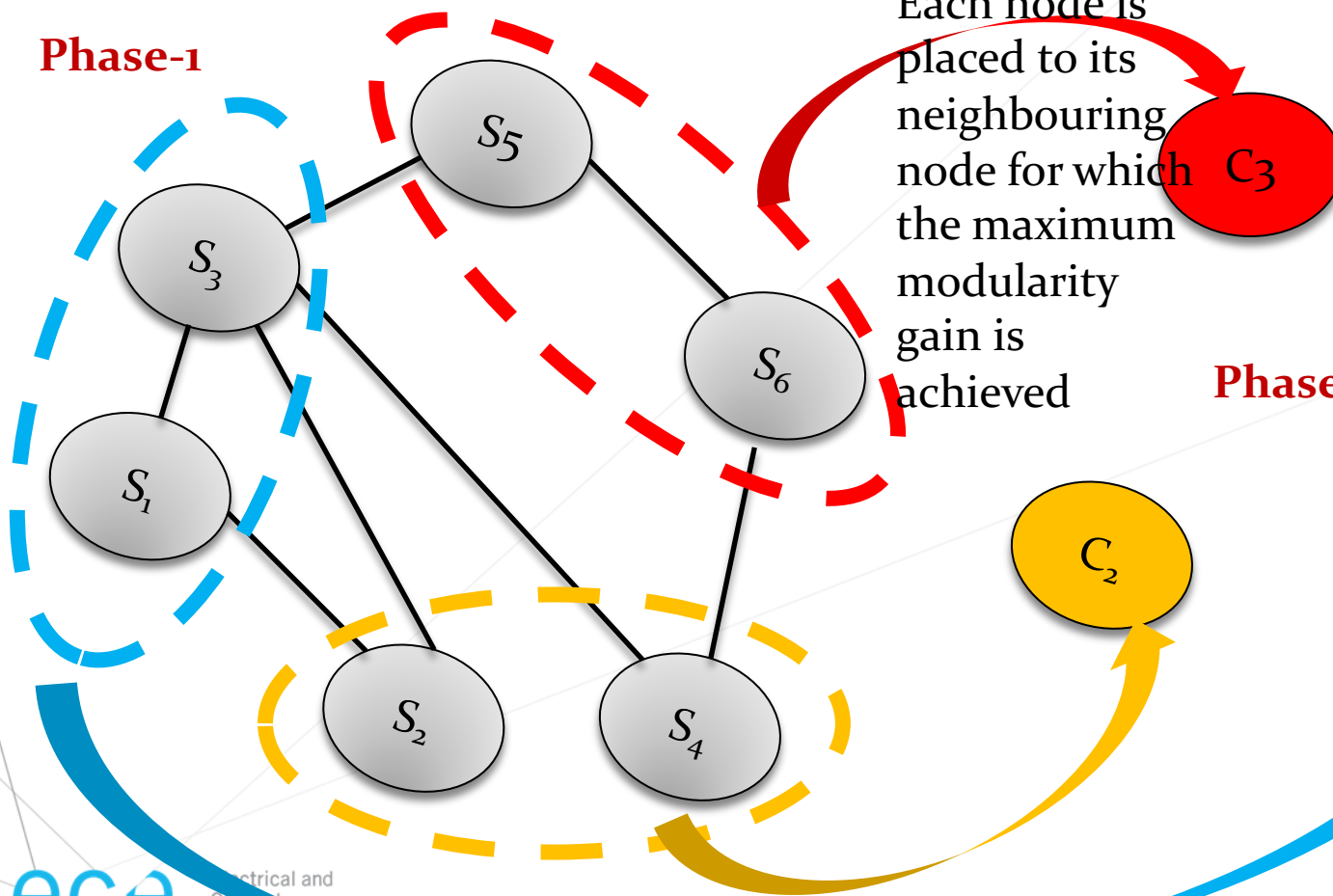
Phase-1



Calculate modularity gain if each node is placed in the community of each of its neighboring nodes

# Louvain Community Detection

Phase-1



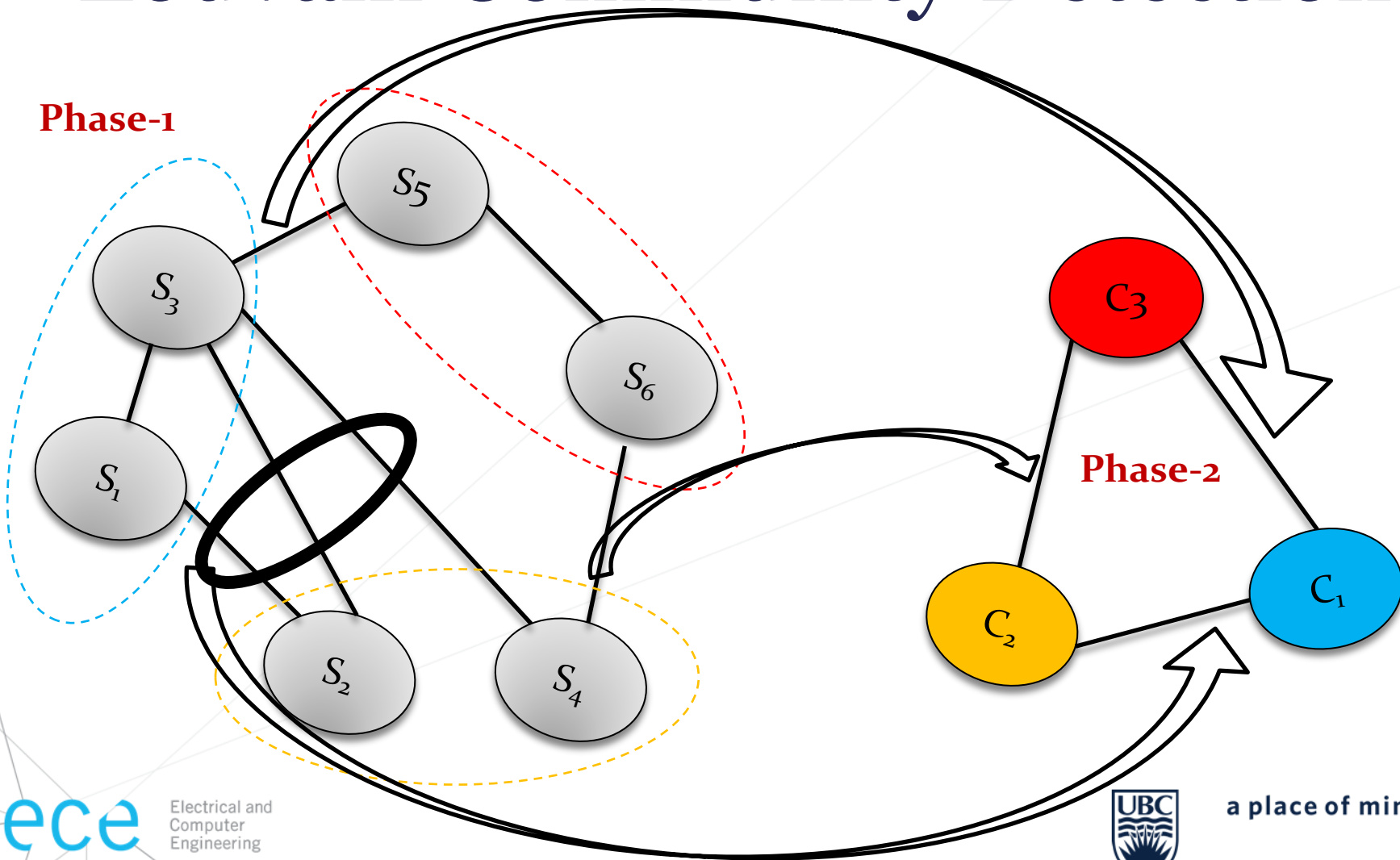
Each node is placed to its neighbouring node for which the maximum modularity gain is achieved

Phase-2

a new network is formed where each node now represents the community formed in the first step

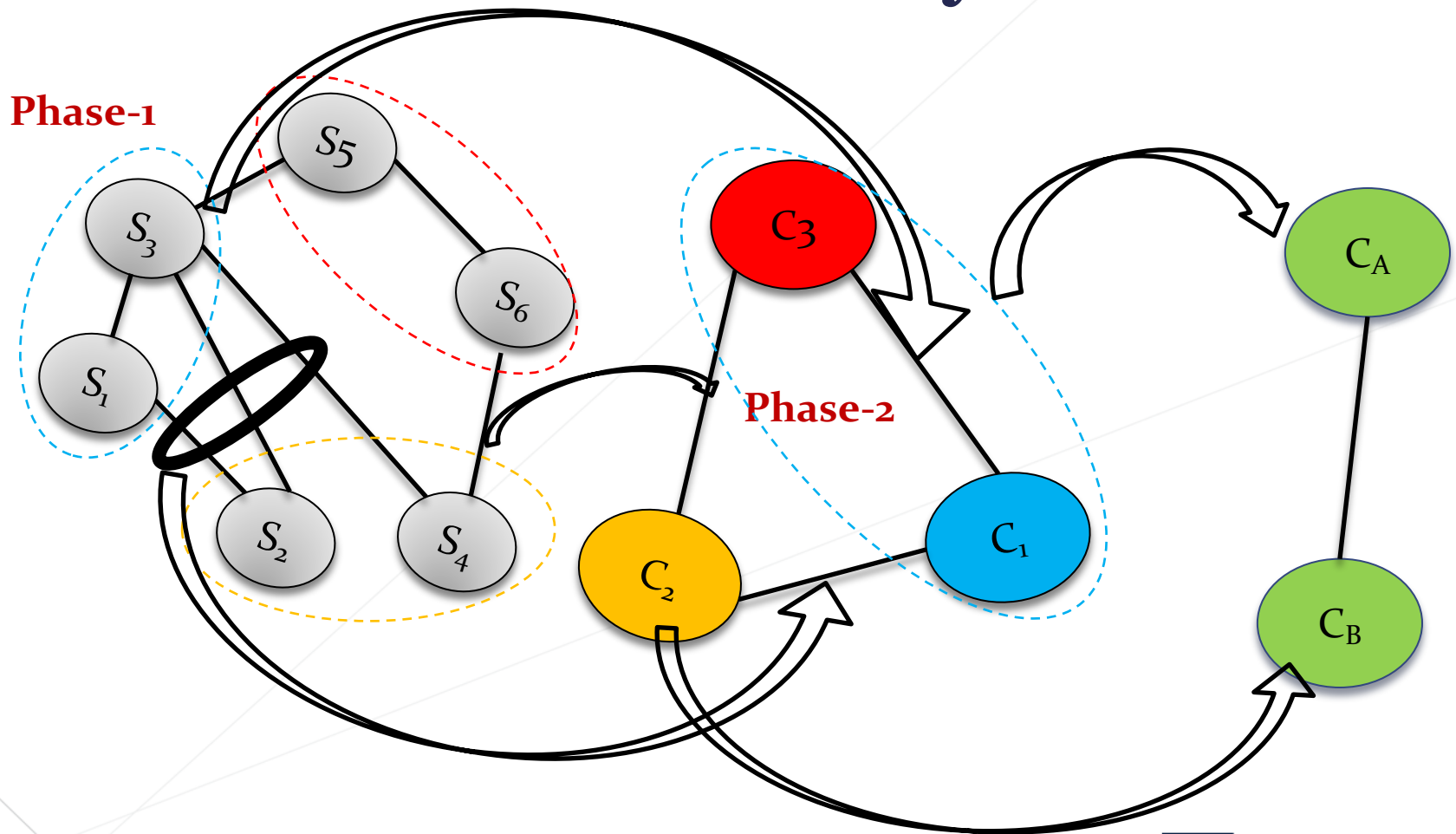
# Louvain Community Detection

Phase-1



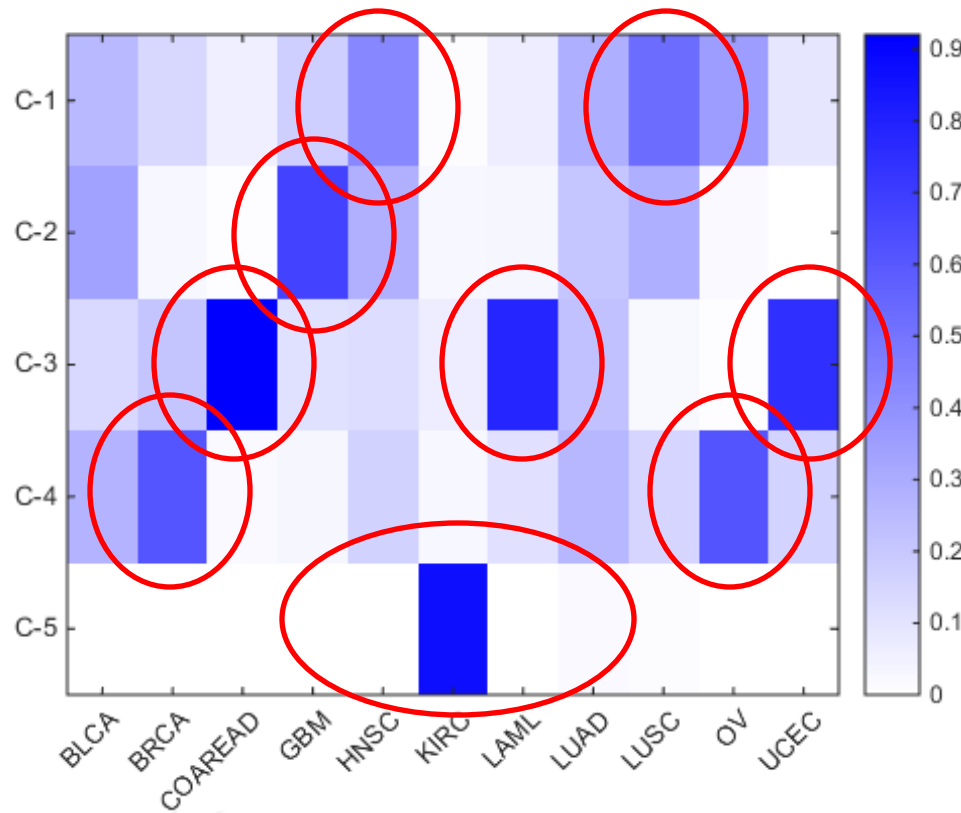


# Louvain Community Detection



# Community Detection Across Human Cancers

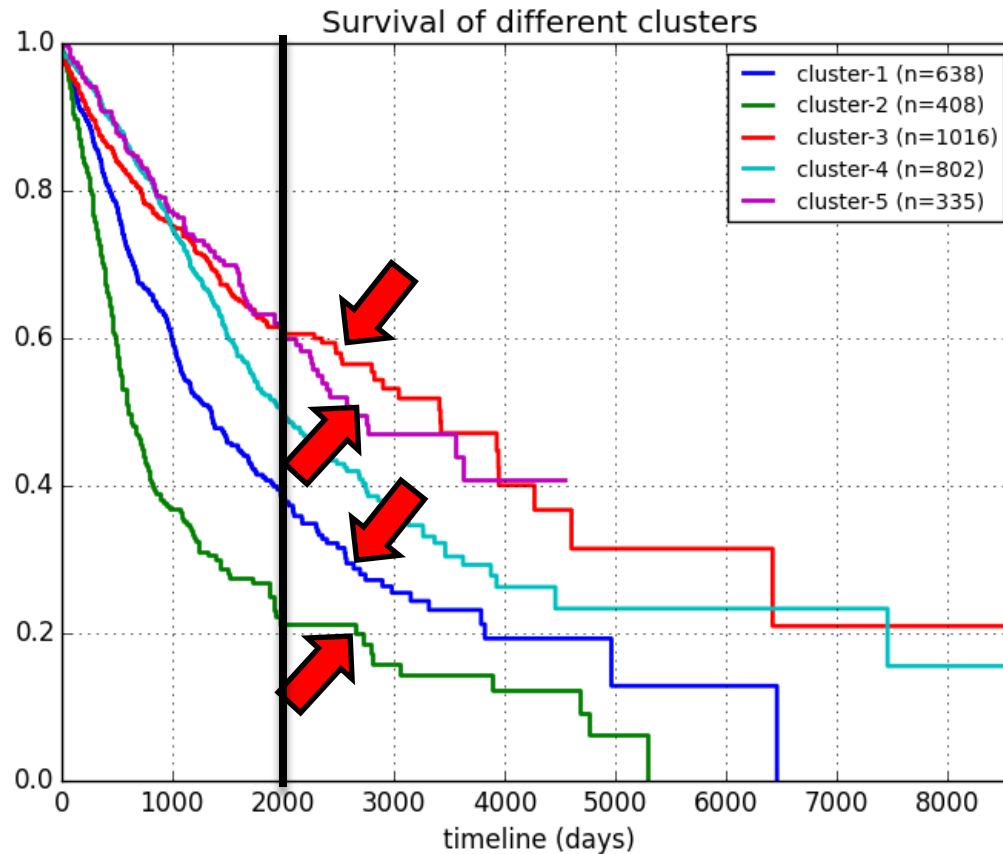
## Result



Community	Number of subjects (n)	Significant cancers
C-1	638	HNSC, LUSC
C-2	408	GBM
C-3	1016	COAREAD, LAML, UCEC
C-4	802	BRCA, OV
C-5	335	KIRC

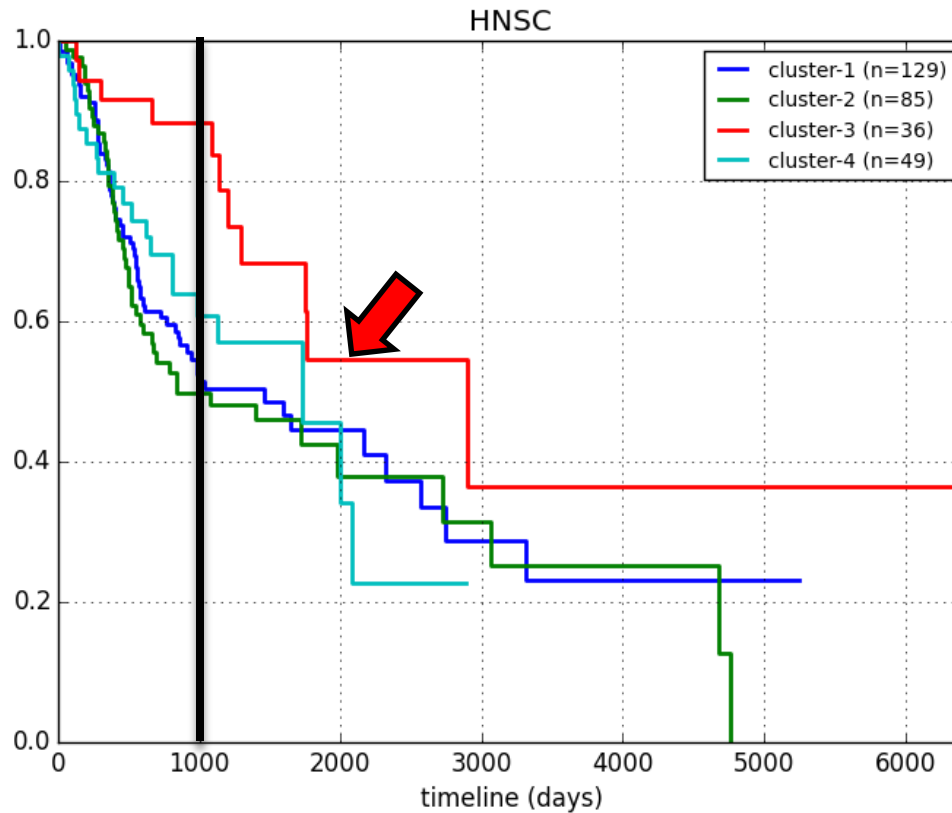
# Community Detection Across Human Cancers

## Result



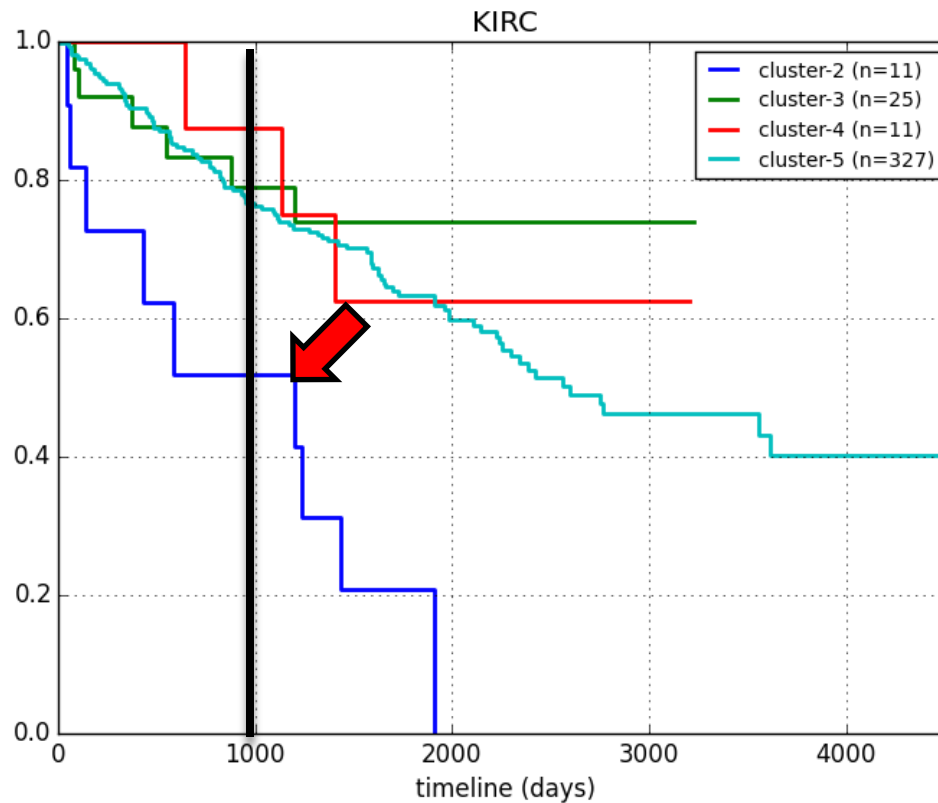
# Community Detection Across Human Cancers

## Result



# Community Detection Across Human Cancers

## Result



# Conclusion

- ❑ We tackled the cancer stratification problem from a data-driven clustering perspective
- ❑ Cancers samples are classified independent of their origins using the community detection approach
- ❑ We further investigated whether such data-driven clusters reveal different survival rate patterns

# Future Works

- ❑ Hierarchical community detection
- ❑ Analyze characteristic features associated with each tumor community
- ❑ Associate the characteristic features to their biological pathways and drug responses
- ❑ Relating biological pathways to tumor communities can provide important information for the design of personalized medicine

*Thank You!*

ece

Electrical and  
Computer  
Engineering



a place of mind