Utilization of Parallel computing and Affinity Propagation Clustering in Identifying Sub-network Biomarker Genes of Cancer on web application

Thanyathorn Thanapattheerakul, Jaturong Kongmanee, Jonathan H. Chan
School of Information Technology (SIT), King Mongkut's University of Technology Thonburi
OVERVIEW

- Focus on improvement of the identification of genes related to Lung cancer
- Apply Affinity Propagation Clustering (APC) algorithm in order to accelerate the gene modules identification from co-expression network
- Apply Message Passing Interface (MPI) in order to accelerate the process and to reduce the space required simultaneously
- Implemented as Web application, www.gat.sit.kmutt.ac.th/apc
SYSTEM ARCHITECTURE

Fig. 2. System architecture
What is Affinity Propagation Clustering (APC)?
How does Affinity Propagation work?

Fig. 1. How Affinity Propagation works.
ADVANTAGES

- no need to specify or pre-determine the number of clusters such as K-means
- All data points are the exemplar of a cluster
DISADVANTAGE

- Consume huge memory and time
What is Message Passing Interface (MPI)?
CONCLUSION

- The system is the APC-based for analysis Subnetwork Identification Network on ‘gat.sit.kmutt.ac.th’
- The system uses Affinity Propagation Clustering (APC) in order to improve the identification of gene sub-networks
- The system applies Message Passing Interface (MPI) in order to accelerate the process and to reduce the space required
Q & A

Thank you for your time (:}