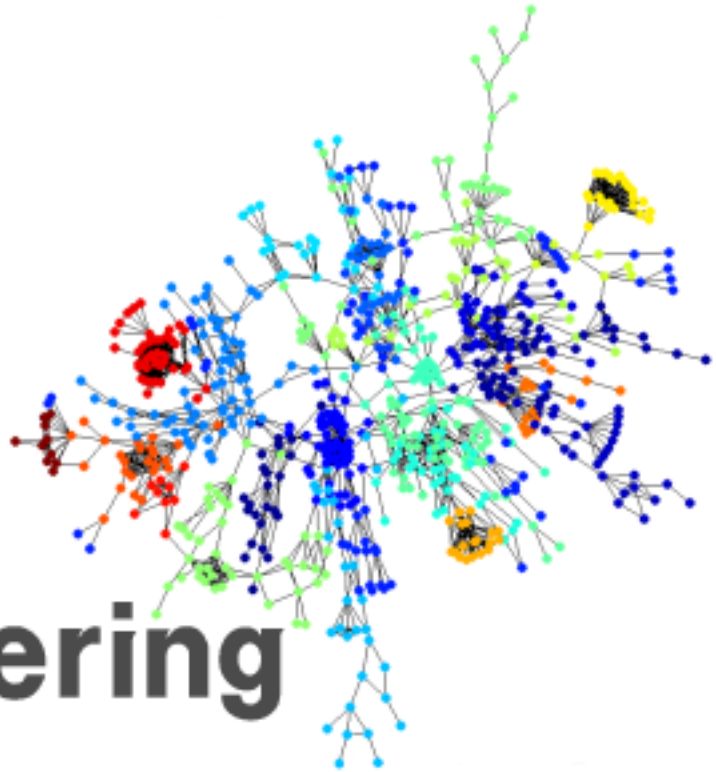


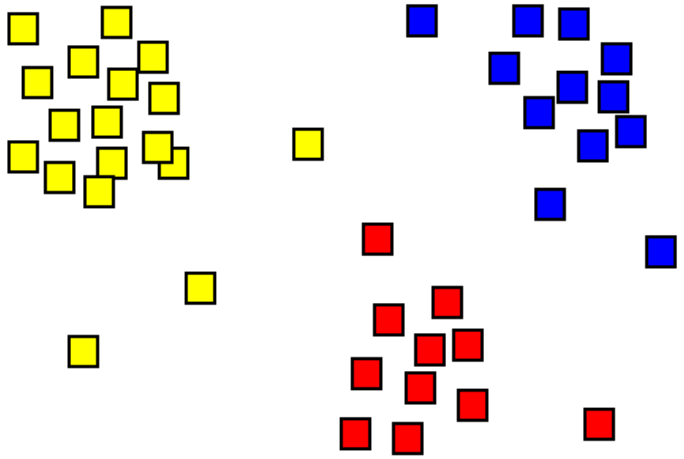
Community Clustering

중간발표2

CLUSTERING?

the task of assigning a set of objects into groups

: called clusters



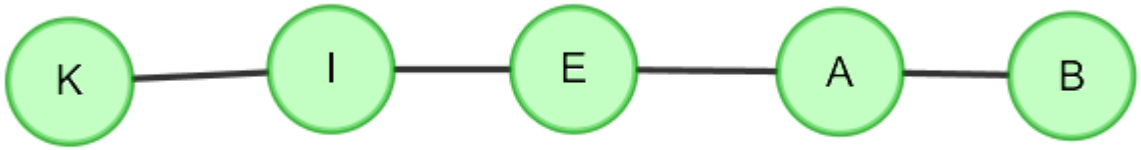
CLUSTERING?

useful method to analyze BigData
-SNS analysis
-Gene Sequence analysis
-Business and Marketing

NEED VARIOUS APPROACHES!

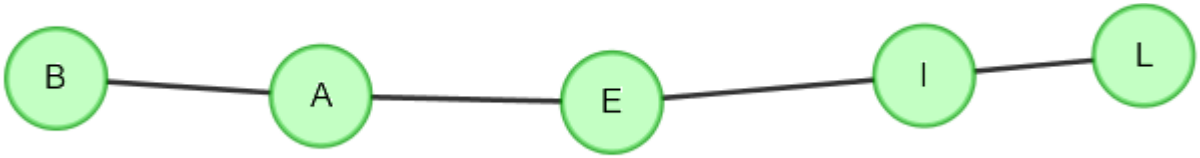
DIFFERENCES

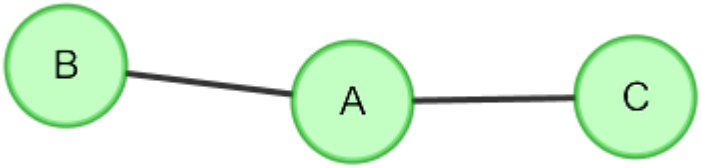
- Fresh approach that has not been tried before
- Allowing overlapping and hierarchical clustering
- Improvement of performance



ROUP

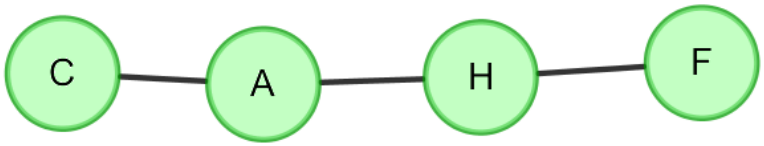
QUENCY

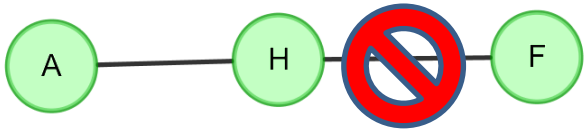


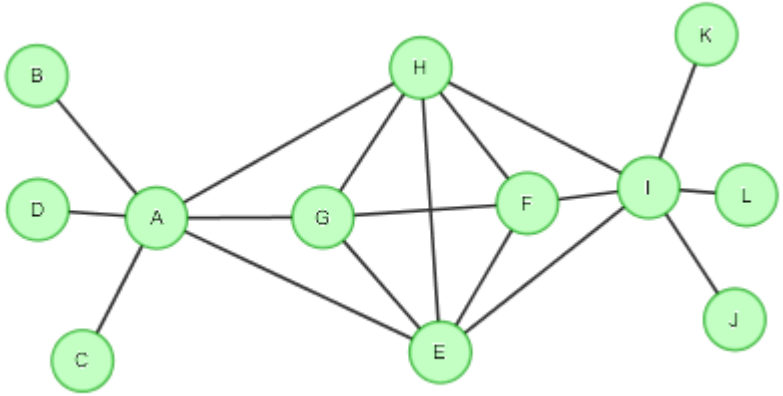


Why

cy?



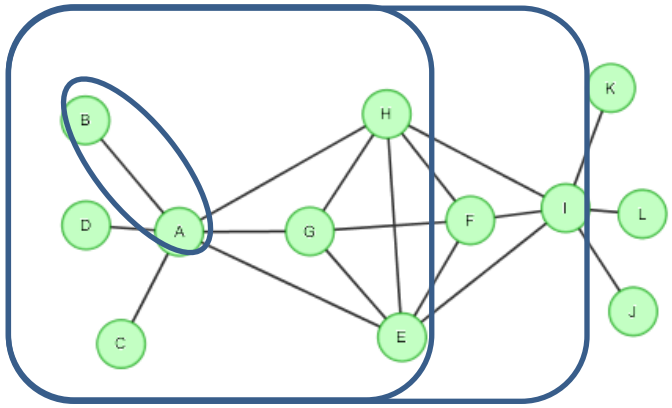


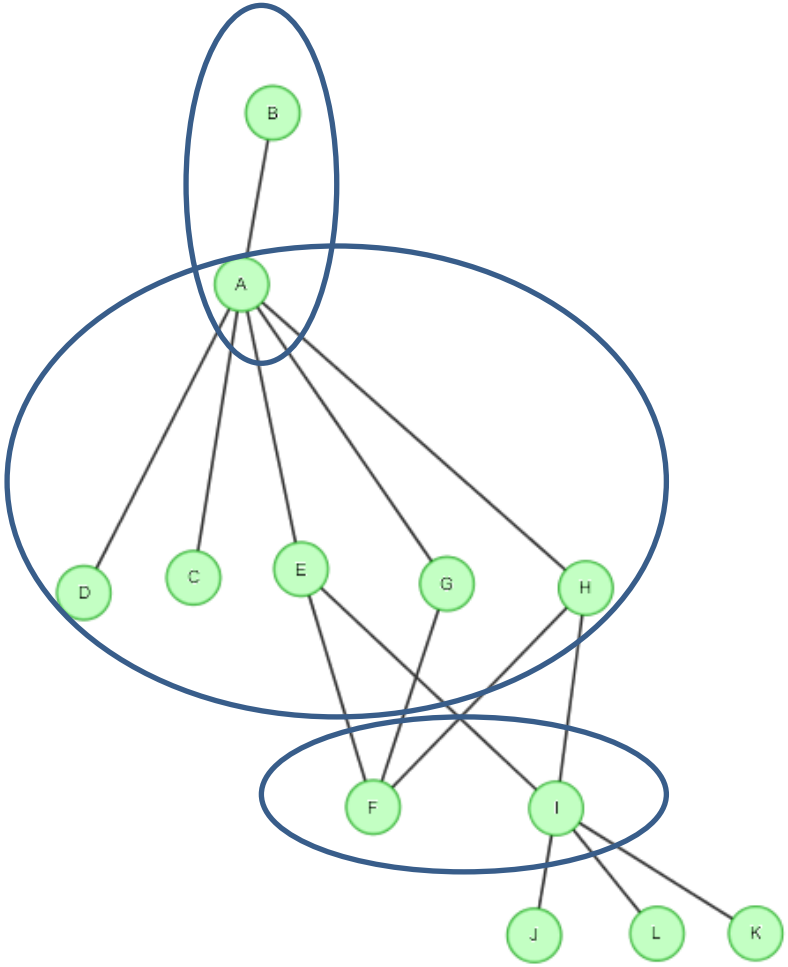


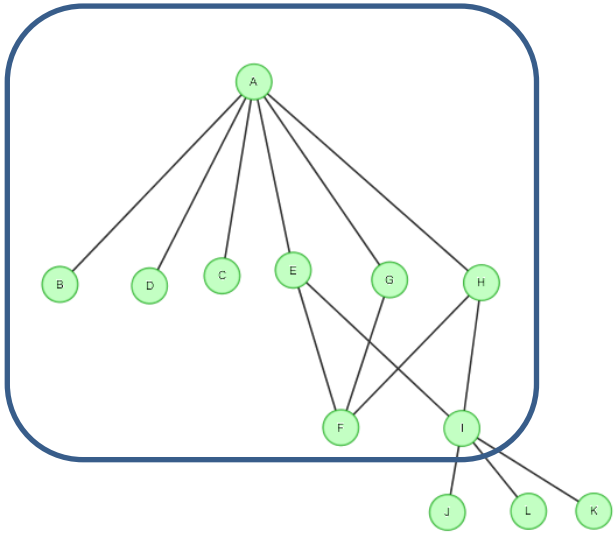
GRAPH CHARACTERISTIC

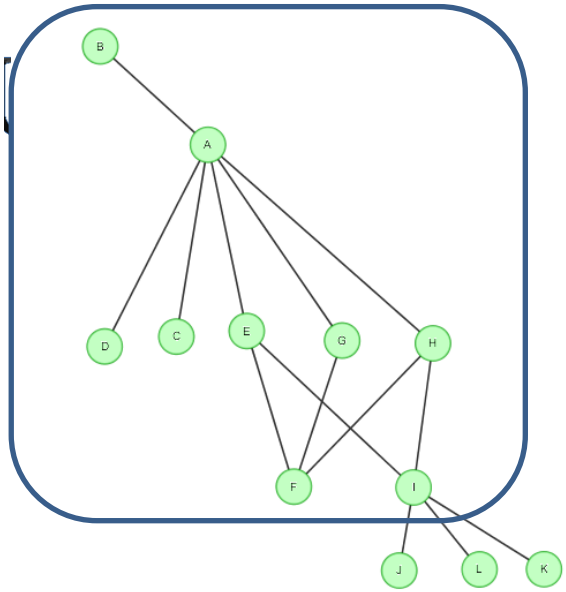
Go-Along
Propagation
Conversion

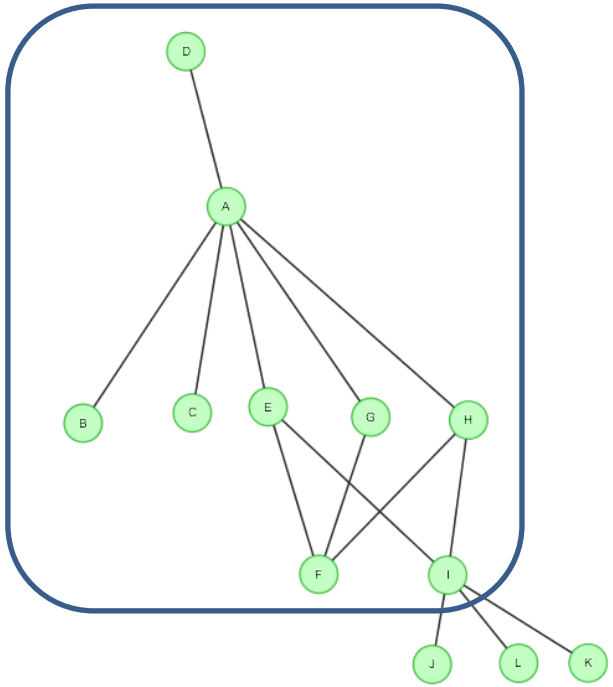
Meaningful Edge?

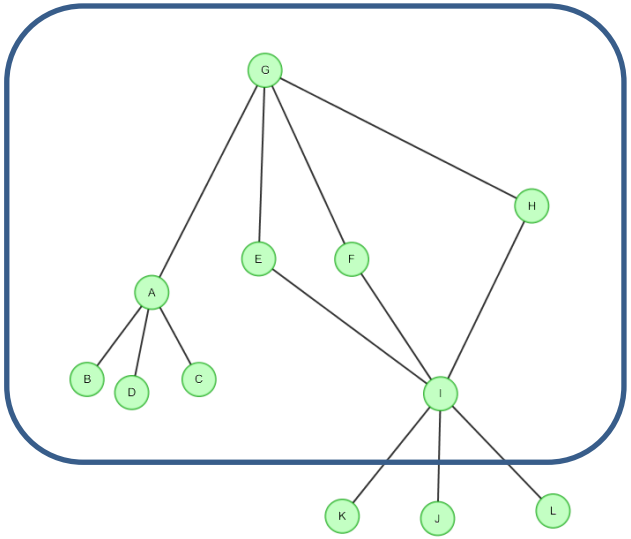


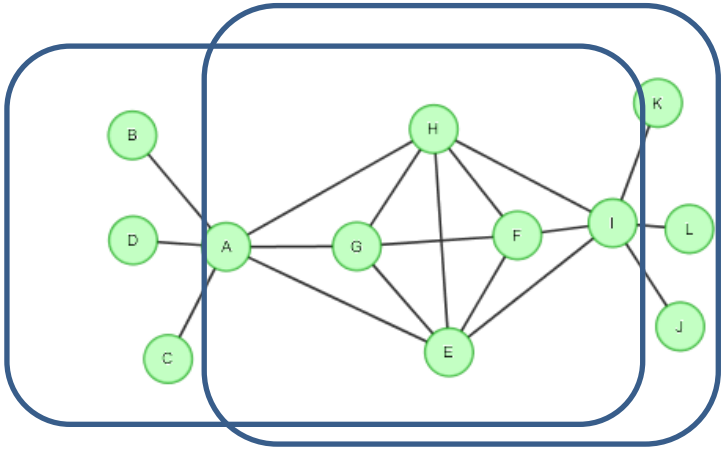



A
B
D
C
E
G
H
F
I
J
L
K


B
A
D
C
E
G
H
F
I
J
L
K

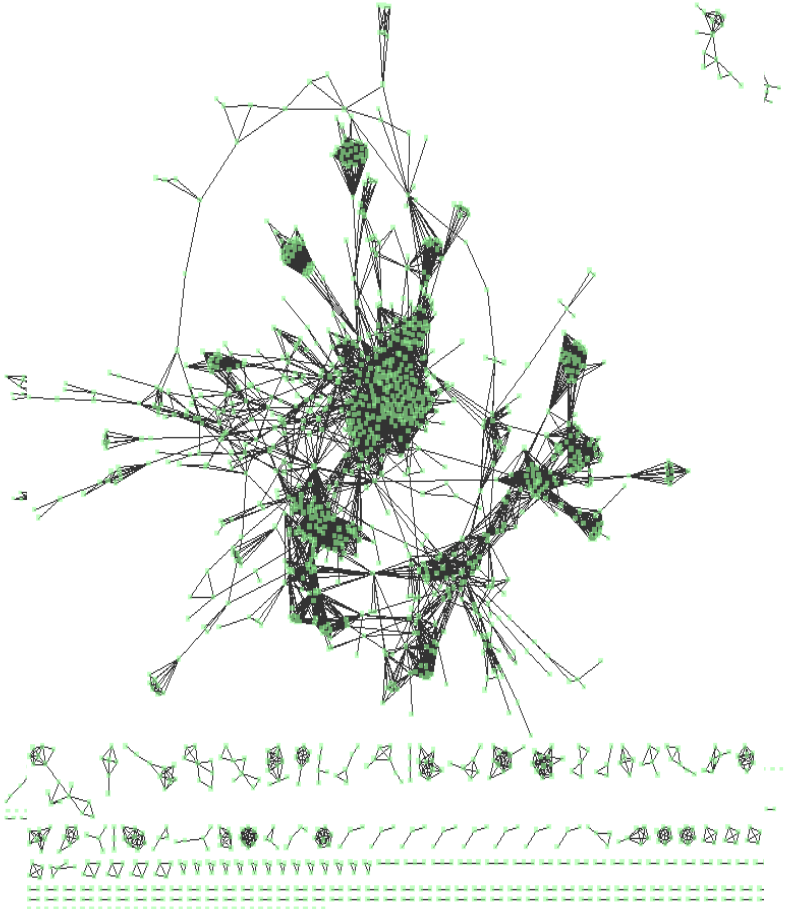

D
A
B
C
E
G
H
F
I
J
L
K


G
H
E
F
A
B
D
C
I
K
J
L

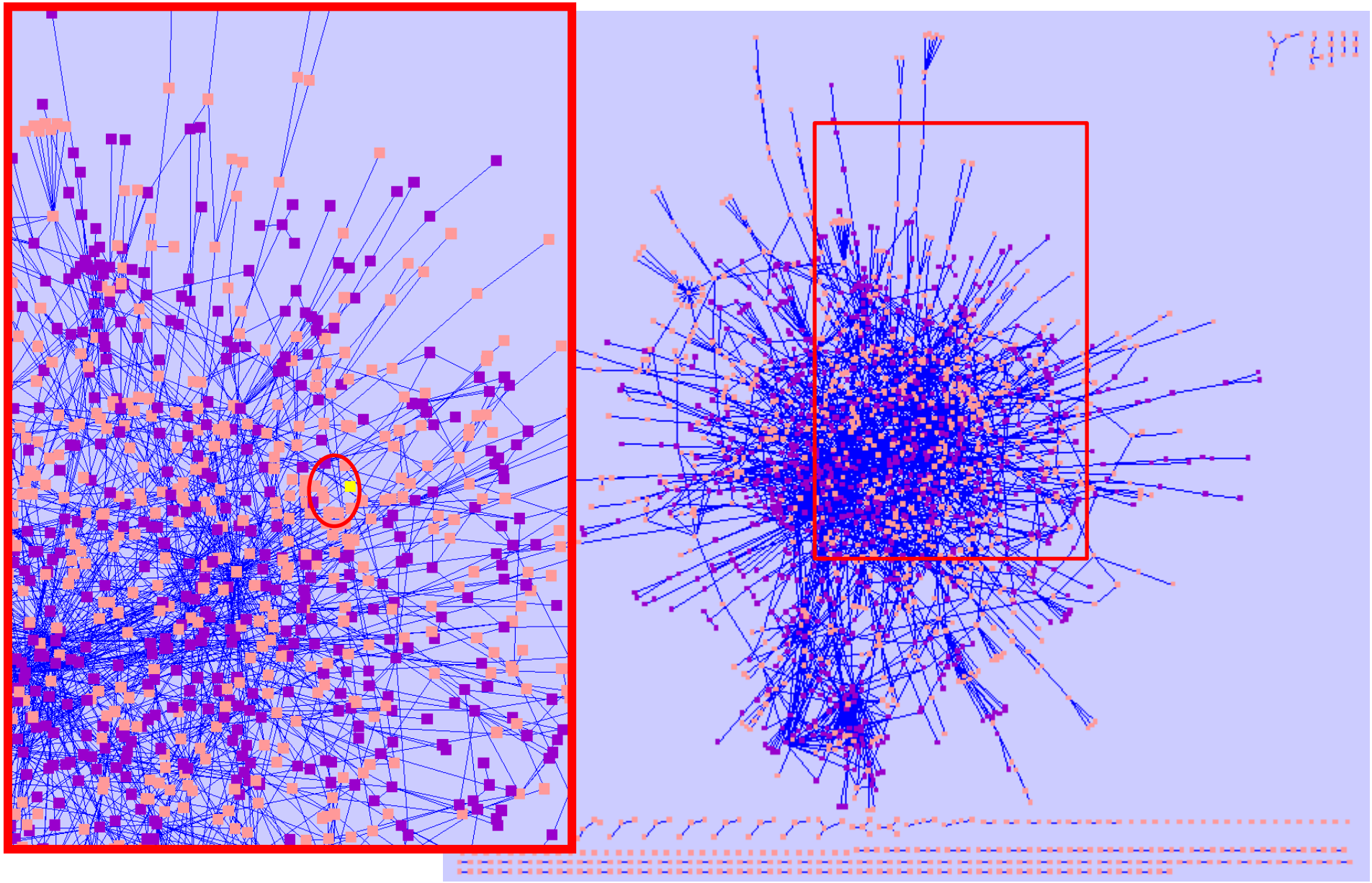

K
H
B
D
A
G
F
I
L
J
C
E

CANDIDATE DATASET

- Y2H_Union
- LC_multiple
- APMS_Collins



APMS_Collins



VALIDATION

N	X	LOD	P	P_adj	Gene-Ontology-ID	Gene-Ontology-Attribute
8	8	1.974	3.067e-7	<0.001	GO:0030127	COPII vesicle coat
7	9	1.220	0.00005413	0.04300	GO:0005979	regulation of glycogen biosynthetic process
11	20	0.8273	0.00004755	0.03700	GO:0070847	core mediator complex
18	35	0.7709	6.722e-7	<0.001	GO:0000086	G2/M transition of mitotic cell cycle
19	46	0.5971	0.00001986	0.01600	GO:0019207	kinase regulator activity
18	44	0.5900	0.00003832	0.03300	GO:0019887	protein kinase regulator activity
20	49	0.5887	0.00001489	0.01500	GO:0046983	protein dimerization activity
21	56	0.5289	0.00004208	0.03300	GO:0005643	nuclear pore
28	81	0.4749	0.00001413	0.01500	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine a...
27	80	0.4589	0.00003277	0.03000	GO:0000398	nuclear mRNA splicing, via spliceosome
28	83	0.4588	0.00002373	0.01900	GO:0000956	nuclear-transcribed mRNA catabolic process
28	84	0.4510	0.00003047	0.02600	GO:0006402	mRNA catabolic process

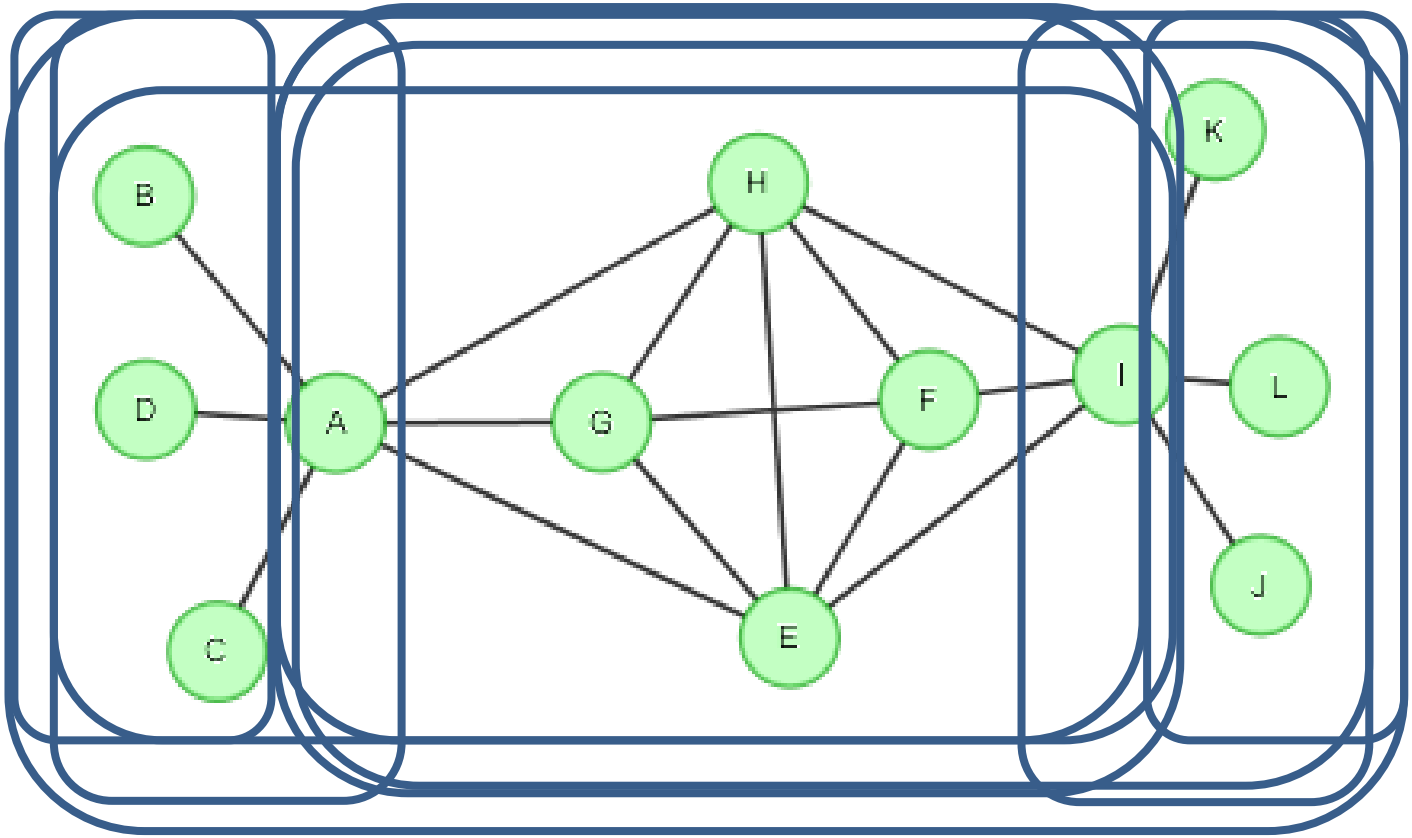
N : Number of entities in query
X : Total Number of entites
P & P_adj : null hypothesis, less than 0.05 or 0.01

73	278	0.3104	0.000001196	<0.001	GO:0016071	mRNA metabolic process
84	318	0.2935	0.000001173	<0.001	GO:0007049	cell cycle
54	211	0.2910	0.00006461	0.04900	GO:0051128	regulation of cellular component organization
78	308	0.2894	0.000002379	0.004000	GO:0006886	intracellular protein transport
107	438	0.2748	2.602e-7	<0.001	GO:0022402	cell cycle process
120	496	0.2707	6.065e-8	<0.001	GO:0015031	protein transport

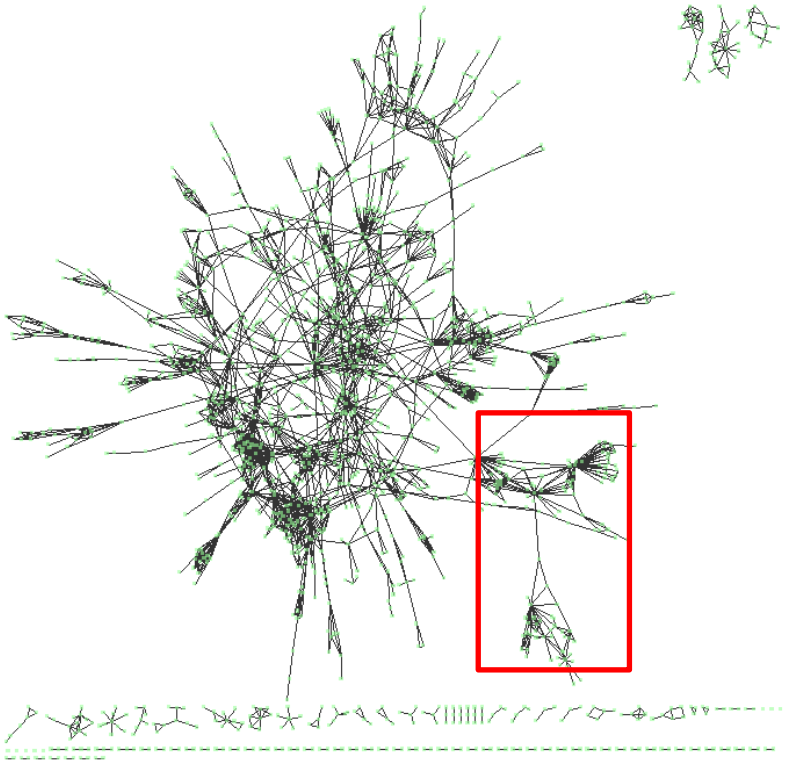
PROBLEM

Cluster is too big to be meaningful!
-GO(Gene Ontology)

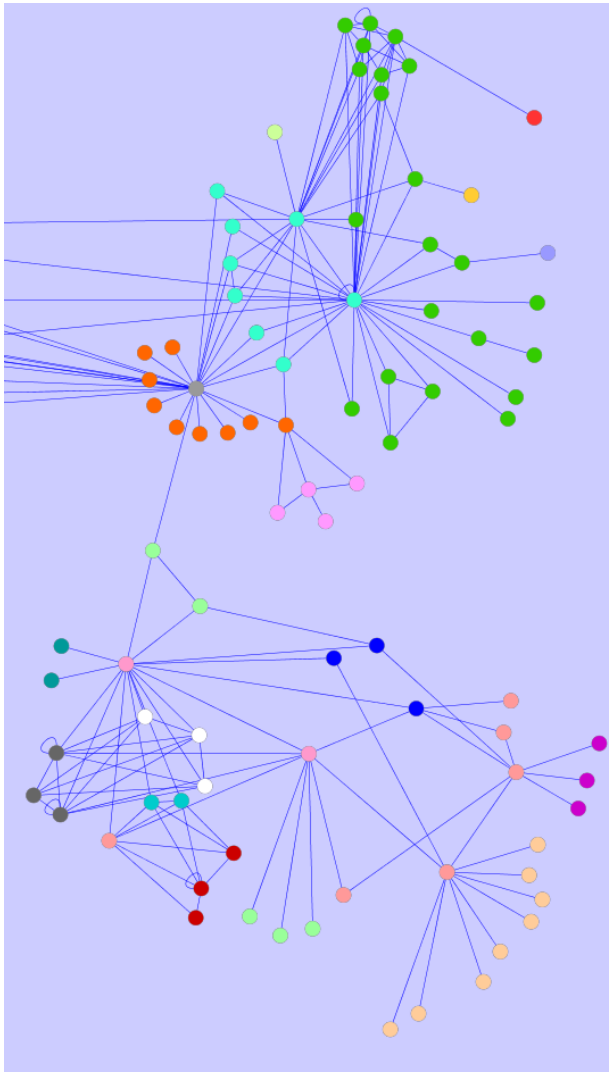
DIVIDE SET



VISUALIZATION



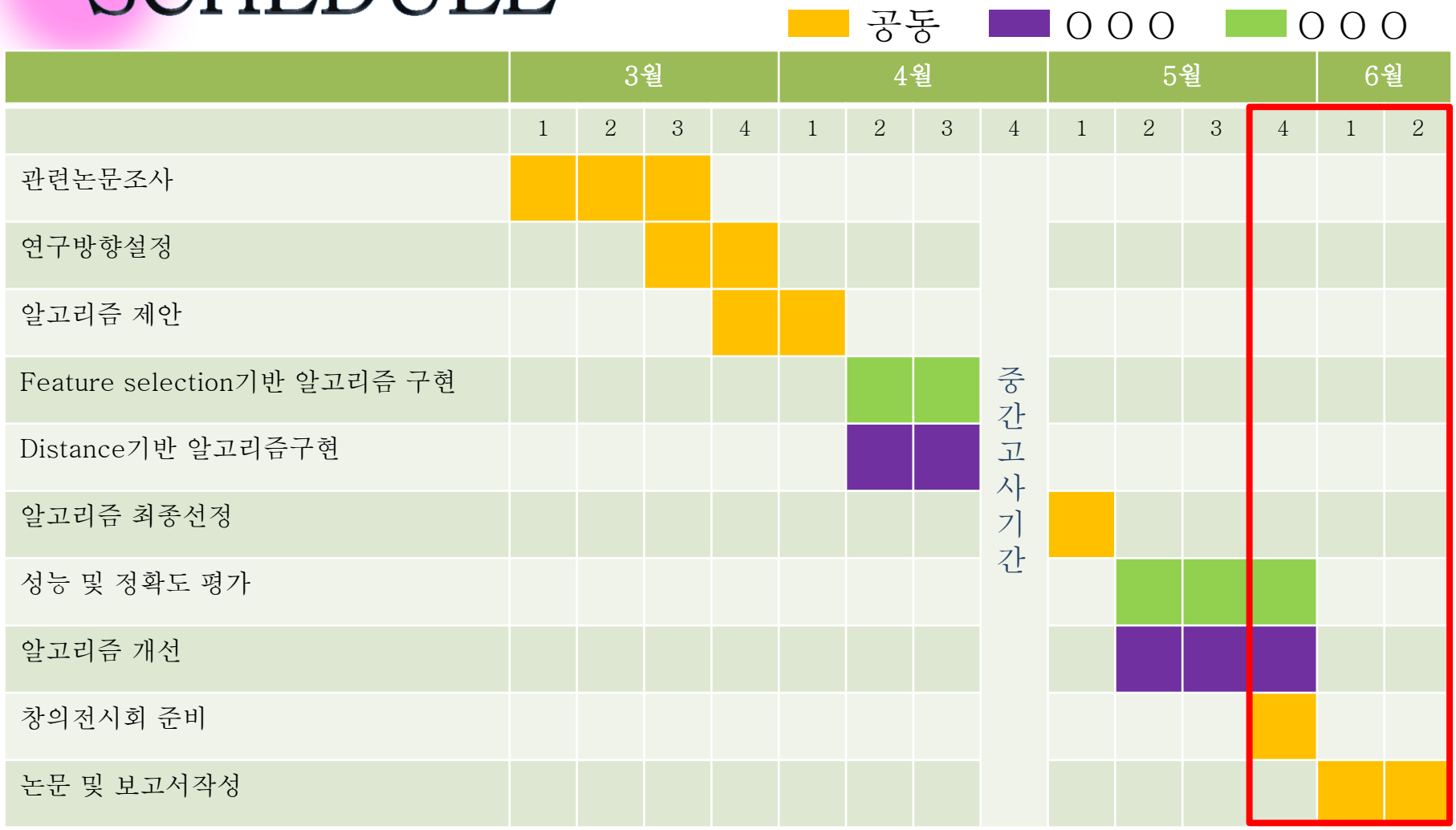
LC_Multiple



WHAT IS LEFT?

- Decide overlapping node
- Output small, overlapping group
- Validate with GO
- Compare with known algorithm

SCHEDULE


공동
O O O
O O O
3월
4월
5월
6월
1 2 3 4 1 2 3 4 1 2 3 4 1 2
관련논문조사
연구방향설정
알고리즘 제안
Feature selection기반 알고리즘 구현
Distance기반 알고리즘구현
알고리즘 최종선정
성능 및 정확도 평가
알고리즘 개선
창의전시회 준비
논문 및 보고서작성
중간고사기간

Question?