New Approximation Results for Resource Replication Problems

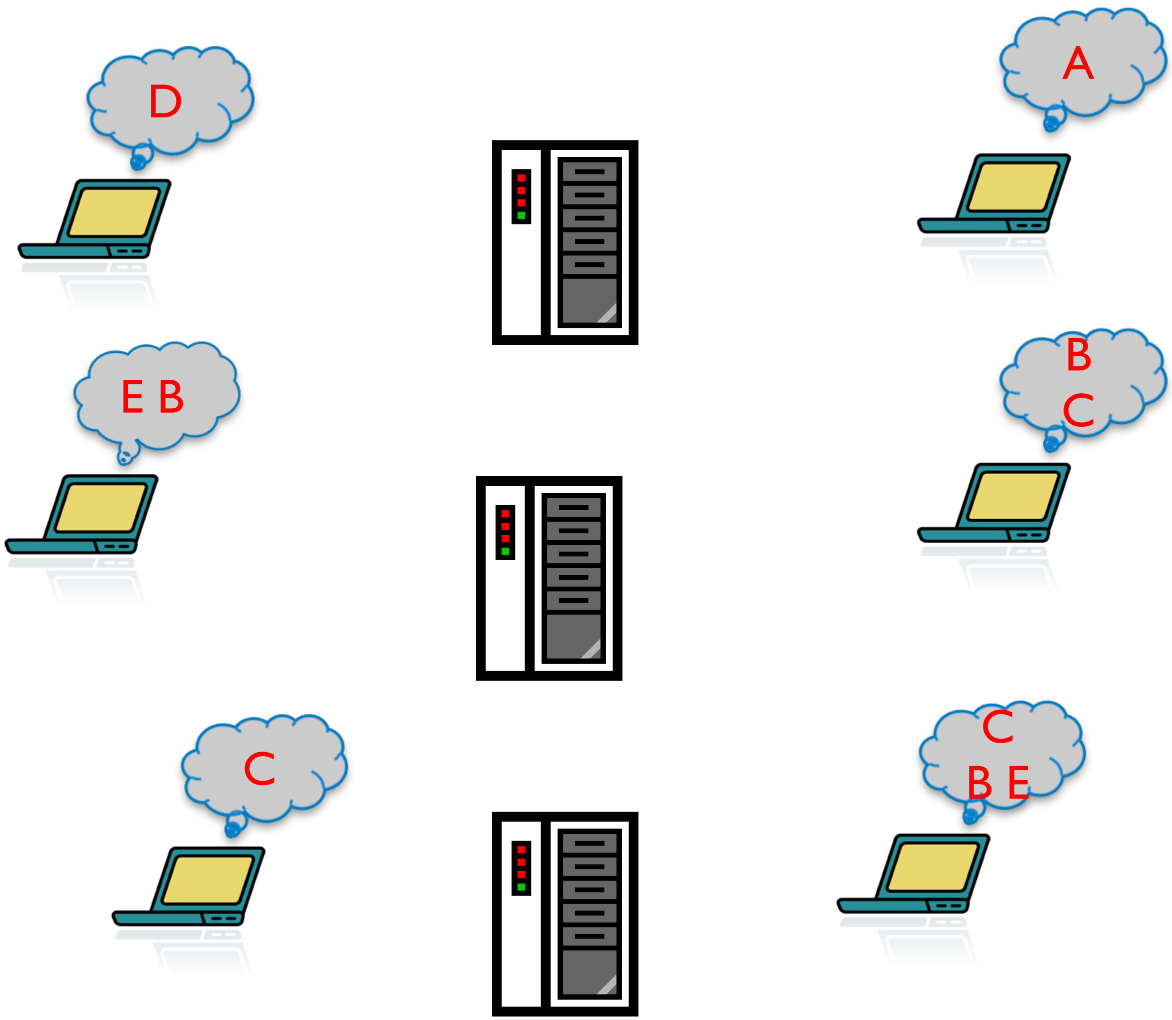
Samir Khuller[*], Barna Saha[$] and **Kanthi Sarpatwar**[*]

[*] University of Maryland College Park

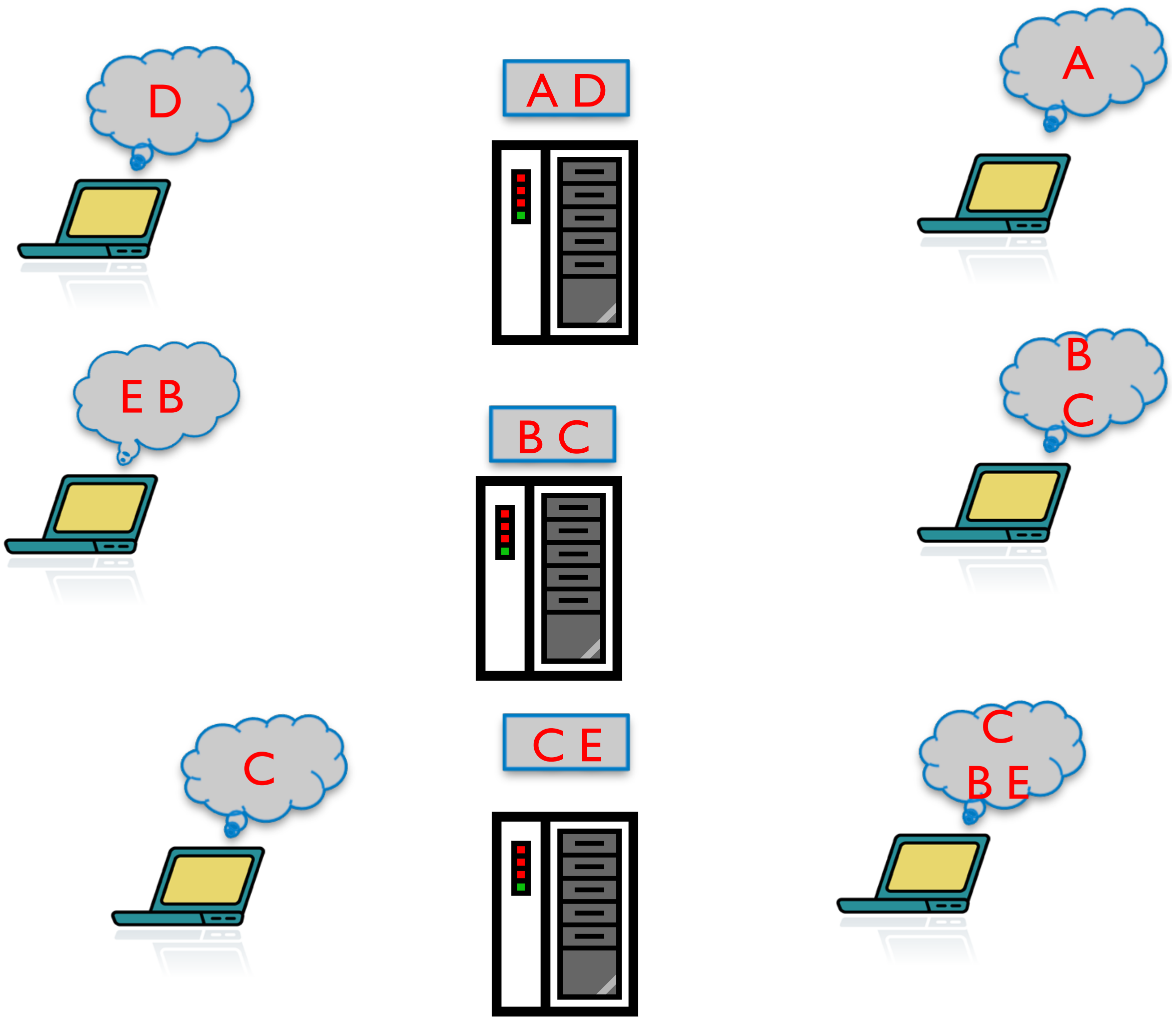
[$] AT&T Labs-Research

APPROX 2012

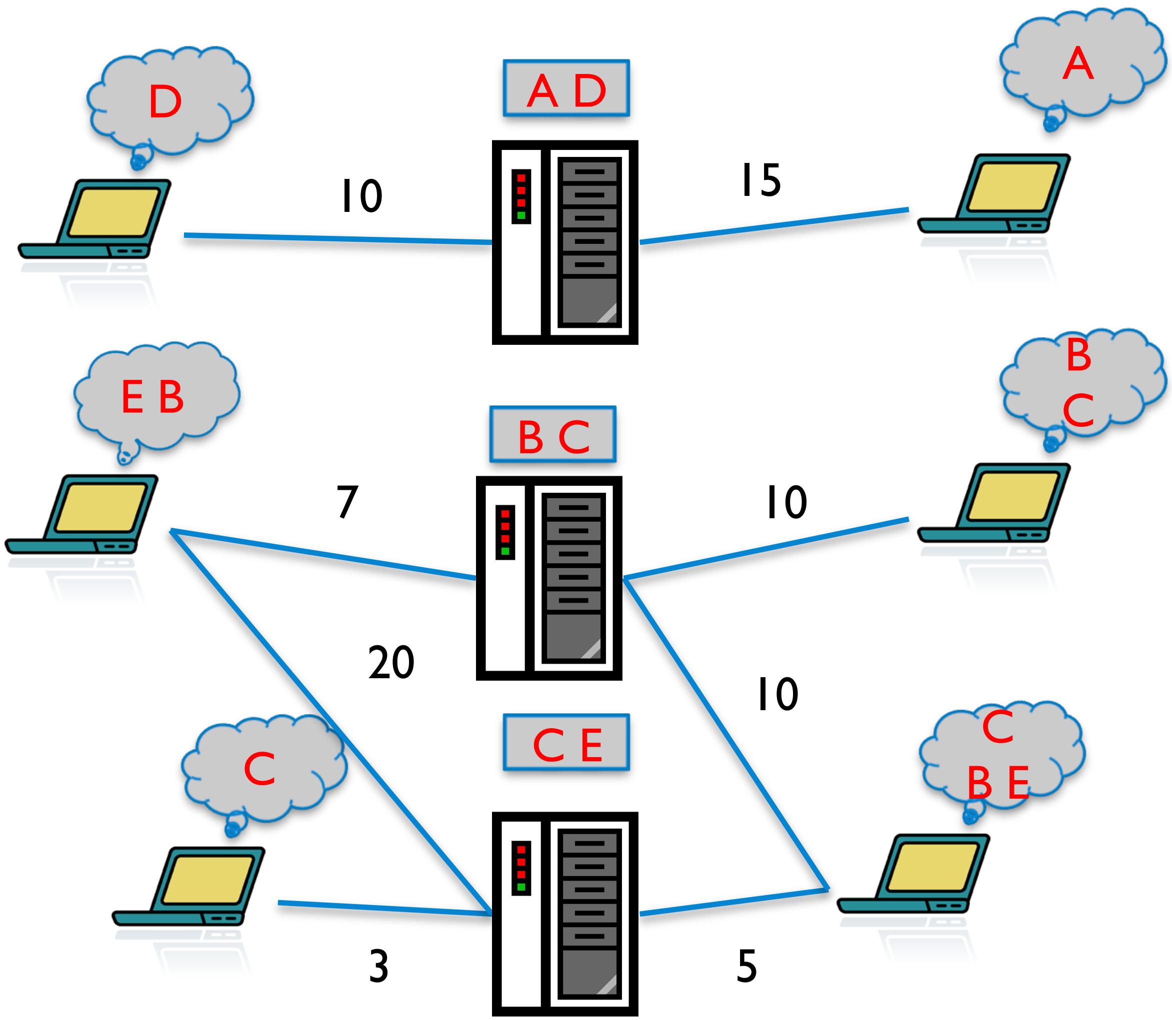
Resource Replication Problem



Resource Replication Problem

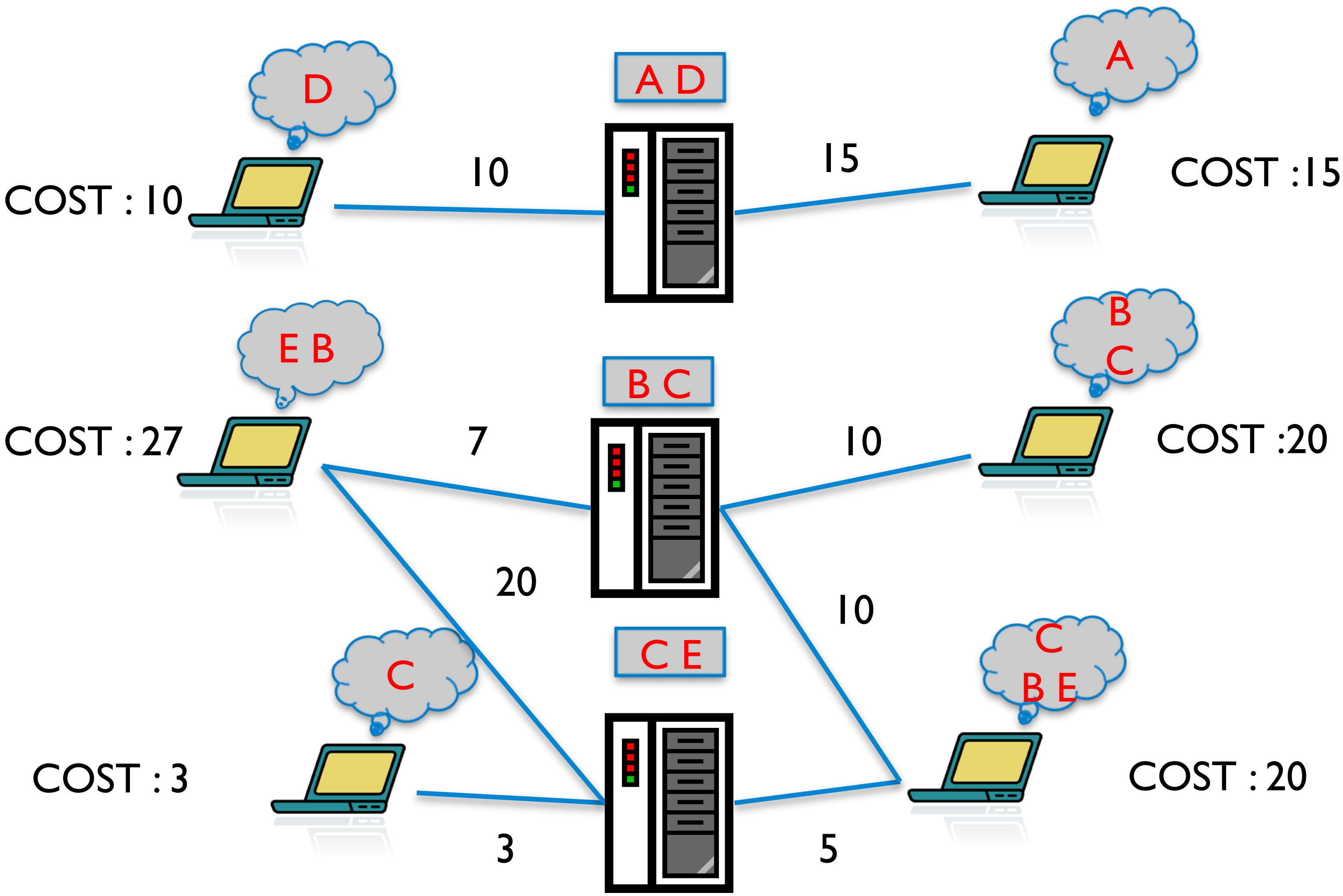


Resource Replication Problem



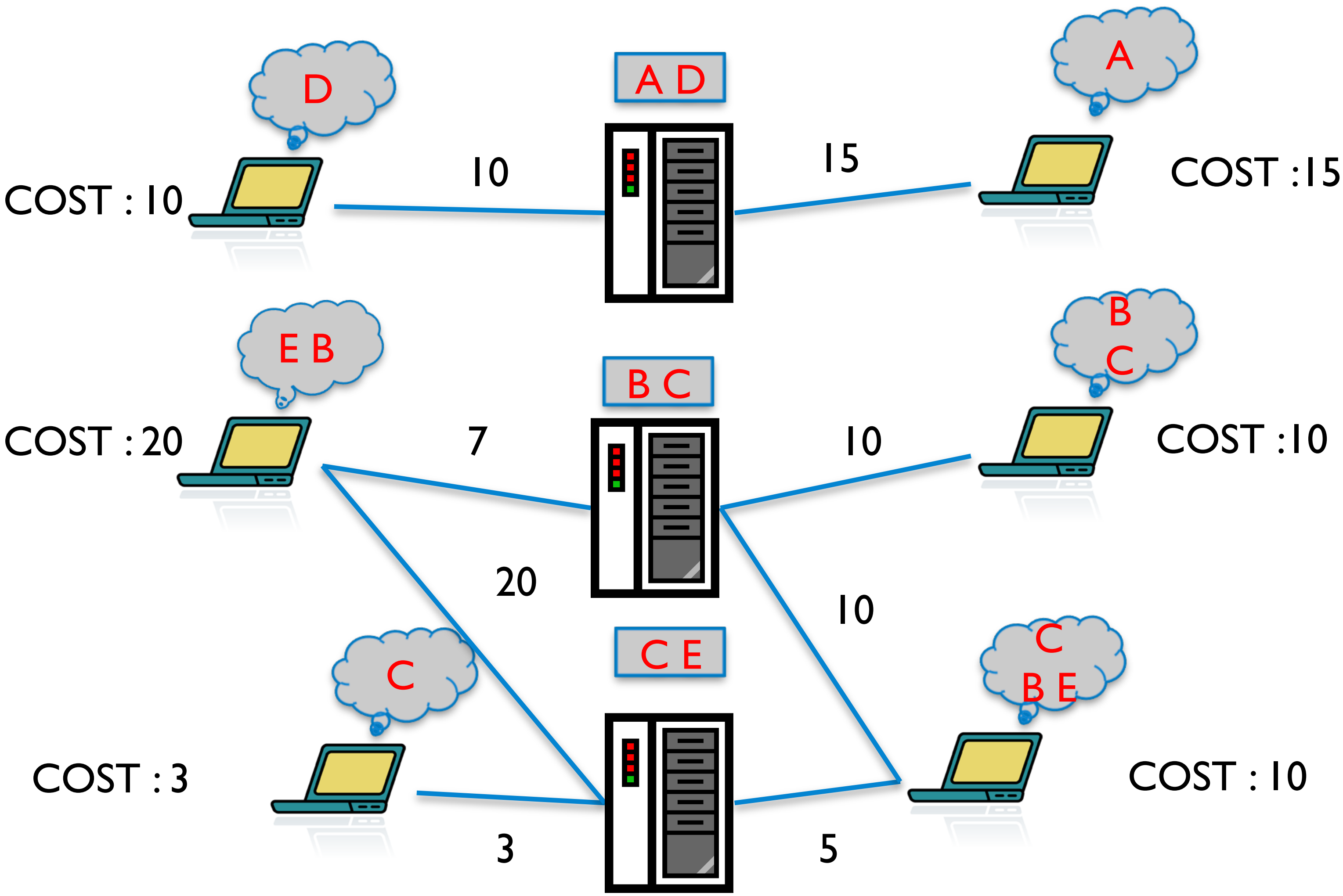
Min Sum Objective

MIN SUM COST: 95



Min Max Objective

MIN MAX
COST: 20



Objectives

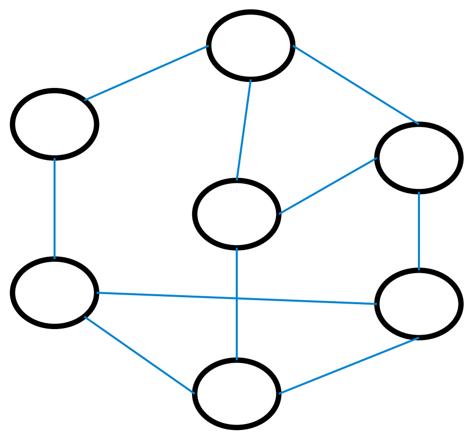
- Min Sum
  - An LP based 10 approximation by Baev, Rajaraman and Swamy[SODA 2001, SIAM J. Compt. 2008].
  - Not covered in this talk
- Min Max distance
  - Studied by Ko and Rubenstein[2003,2004].
  - This is the subject of current talk.

Motivation: Resource Replication

- Need of data replication arises in many contexts
  - File replica distribution
    - Reduce retrieval time: improves data locality
    - Fault tolerance
  - Distributed computation
    - Distribution of loads: improves latency
  - Channel allocation in wireless networks
    - Channel capacity can improve if nodes assigned to the same channel are far apart: reduce interference

Resource Replication

- Basic Problem:
  - Given K resources, N servers to host resources. Each client needs every resource.

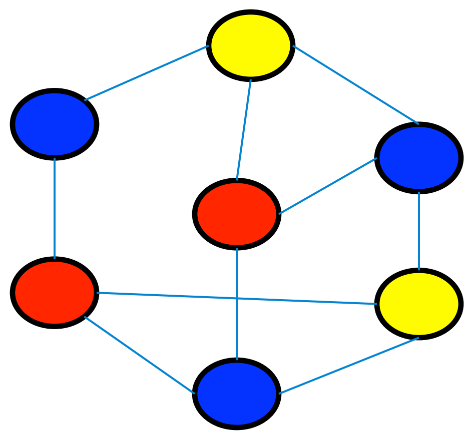


Edge denotes distance 1.
Shortest Path Metric.

- Goal: *minimize the maximum distance* each client needs to travel to get all the required resources.

Resource Replication Problem

- Basic Problem:
  - Given K resources, N servers to host resources. Each client needs every resource.



Edge denotes distance 1.
Shortest Path Metric.

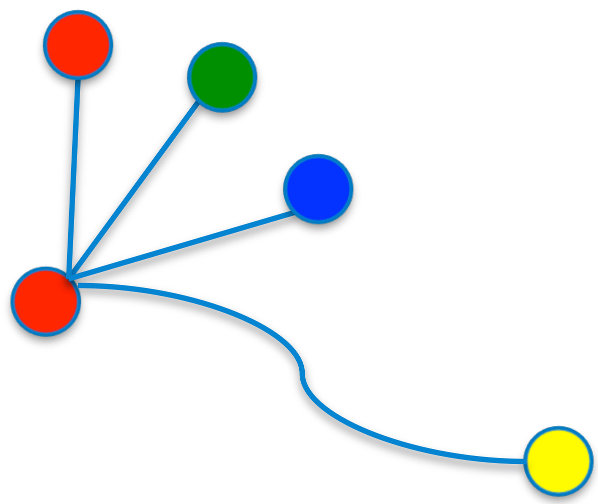
- Goal: *minimize the maximum distance* each client needs to travel to get all the required resources.

Basic Resource Replication Problem

- Studied by Ko and Rubenstein[ICNP 2003]
  - Distributed 3 approximation using local search
  - Local search may not converge in polynomial time
- We give a very simple 3 approximation which can be implemented in a distributed scenario as well.
  - Idea: Extending threshold graph technique for k center type algorithms

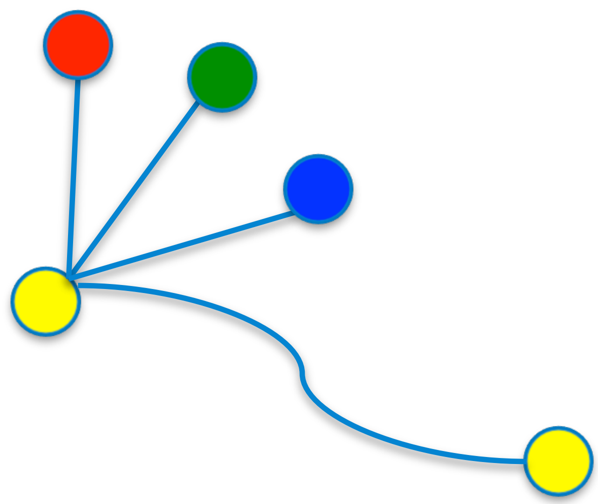
Ko and Rubenstein Algorithm

- Local Search Step-



Ko and Rubenstein Algorithm

- Local Search Step-



- Guaranteed to give a 3 approximation. Not polynomial time.

Our Results.

Problem Variant	Approx. Achieved	Hardness of Approximation	Description
Basic Resource Replication (BRR) Problem	3 approx.	No 2- ε approx. possible	Every node requires every resource. Each node can store exactly one resource type.
Subset Resource Replication (SRR) Problem	3 approx.	No 3- ε approx. possible	Nodes require arbitrary subset of resource types. Each node has designated capacity.
BRR with Outliers	3 approx.	No 2- ε approx. possible	Same as BRR. Need to satisfy at least M nodes.

Our Results (Contd.)

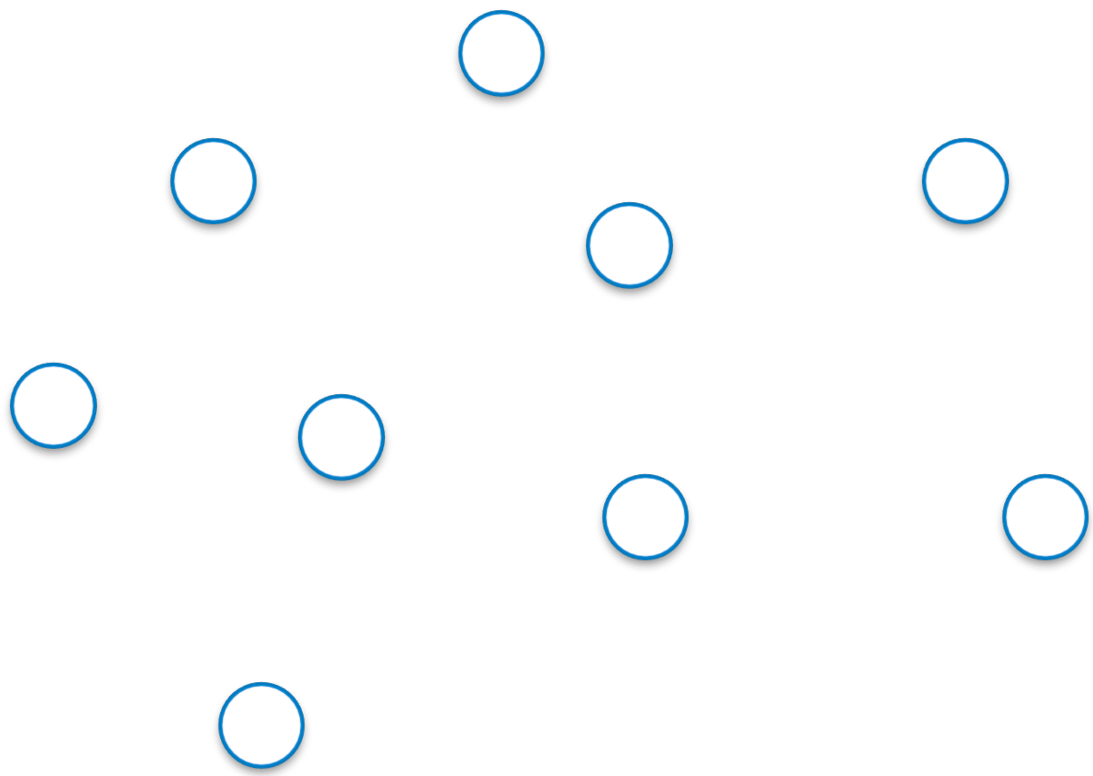
Problem Variant	Approx. Achieved	Hardness of Approximation	Description
SRR with Outliers	None	NP-Hard to approximate within any non-trivial approx..	Same as SRR. Need to satisfy at least M nodes.
BRR with Load Balancing	4 approx. with a violation of load capacity of factor 2.	No 2- ε arpprox.	Same as BRR. Each server can serve at most k clients.

Our Algorithm

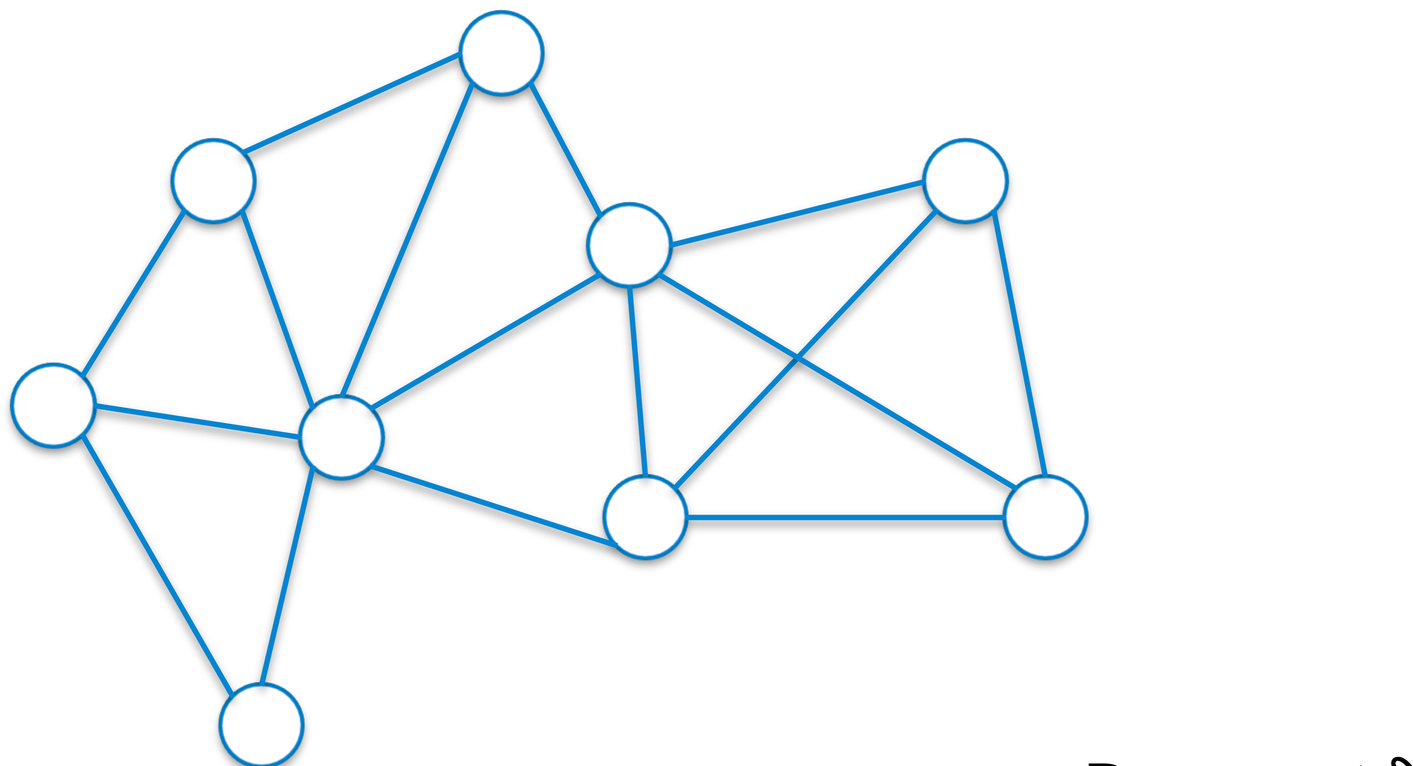
- Based on the threshold graph technique
  - Similar to the k-center.

- 3 approx. in polynomial time.
- Can be implemented in a distributed fashion.

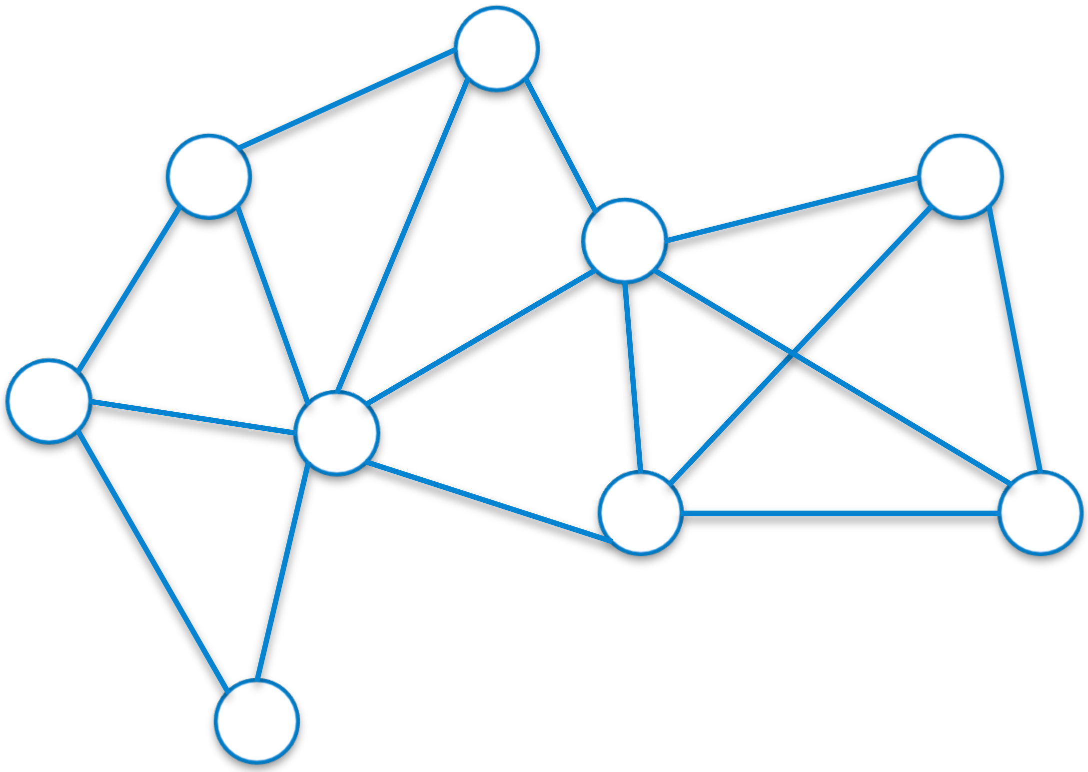
Graph G and resources {R, G, B}



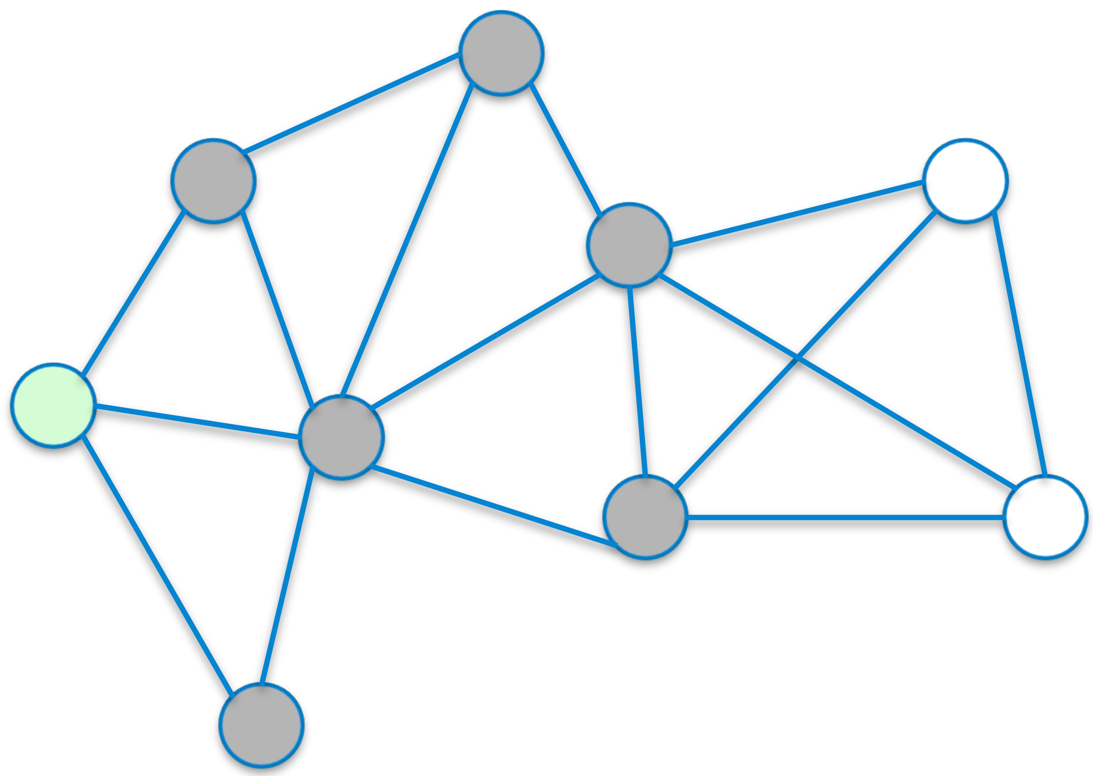
Guess the optimal distance δ and Construct **threshold graph.**



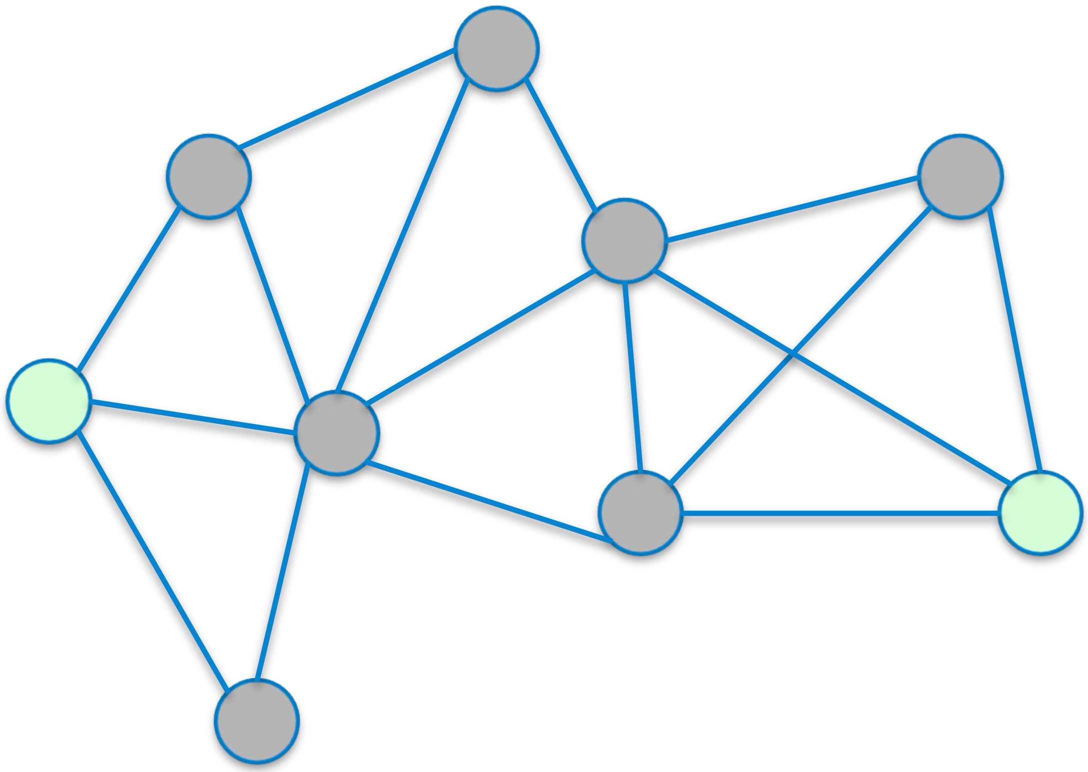
- Keep picking nodes and delete nodes within two hops.



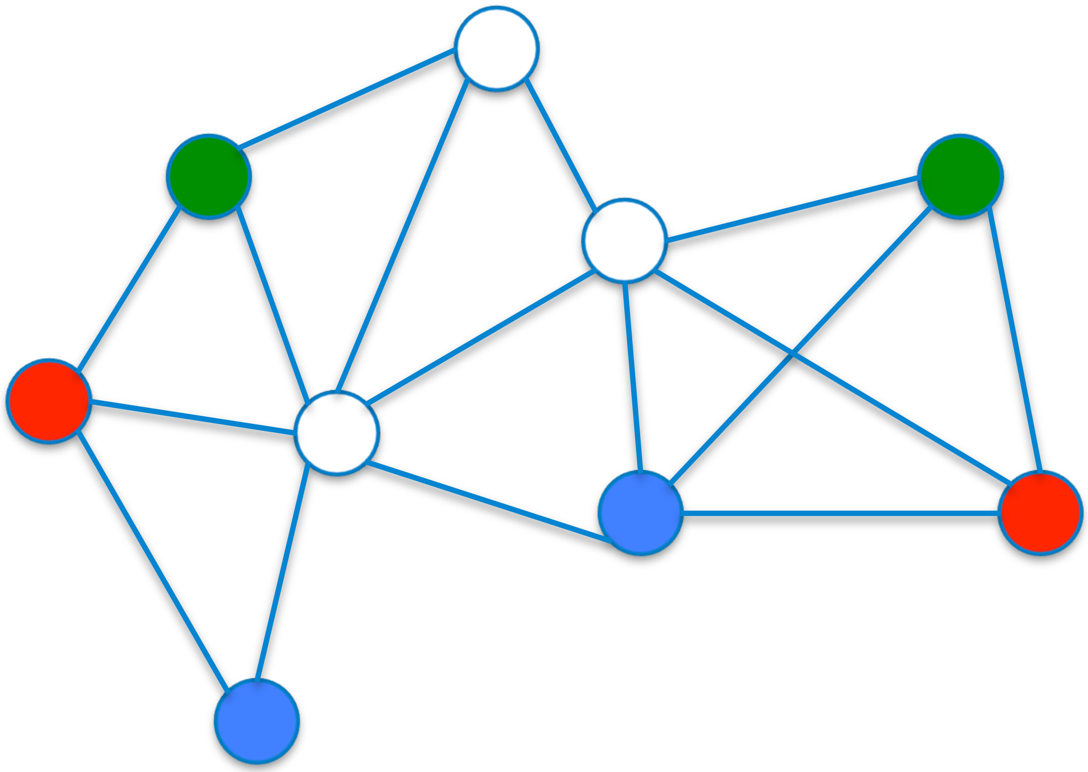
- Keep picking nodes and delete nodes within 2 hops. Call this set MIS.



- Keep picking nodes and delete nodes within two hops. Call this set MIS.



- Place k resources in each vertex of MIS's delta neighborhood.



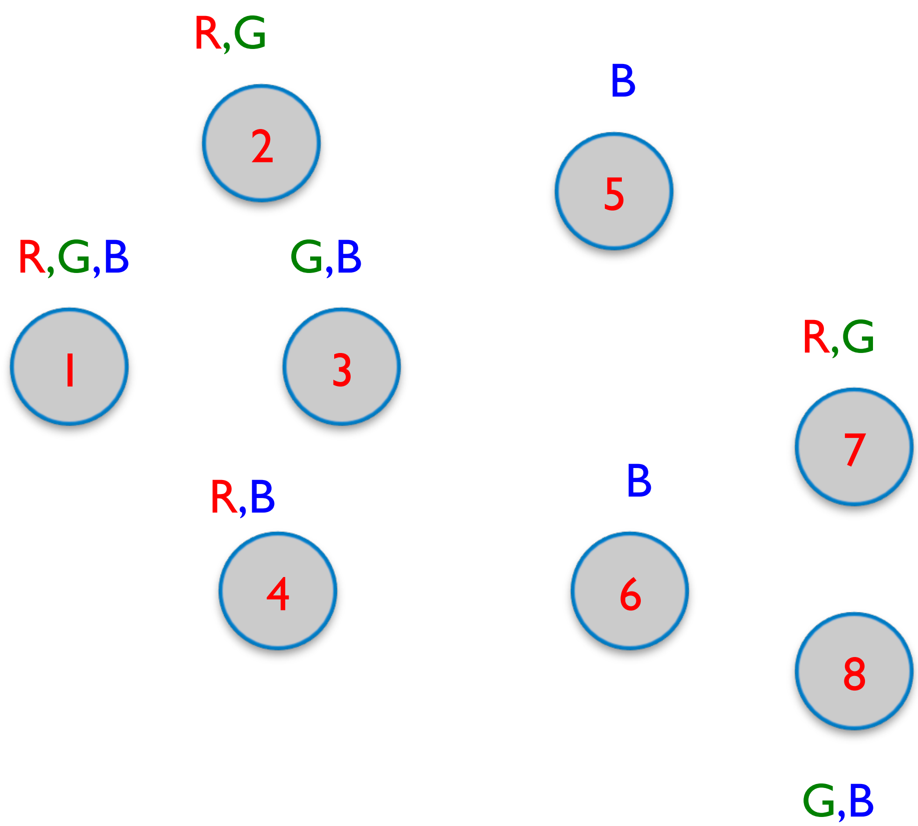
Why does it work ?

- Every vertex is within 2 hops of MIS.
- Each vertex in MIS has degree at least k-1.
- Placing colors on MIS and neighbors makes sure that every vertex has every color within 3 hops.
- Hence, we obtain a 3 approximation.

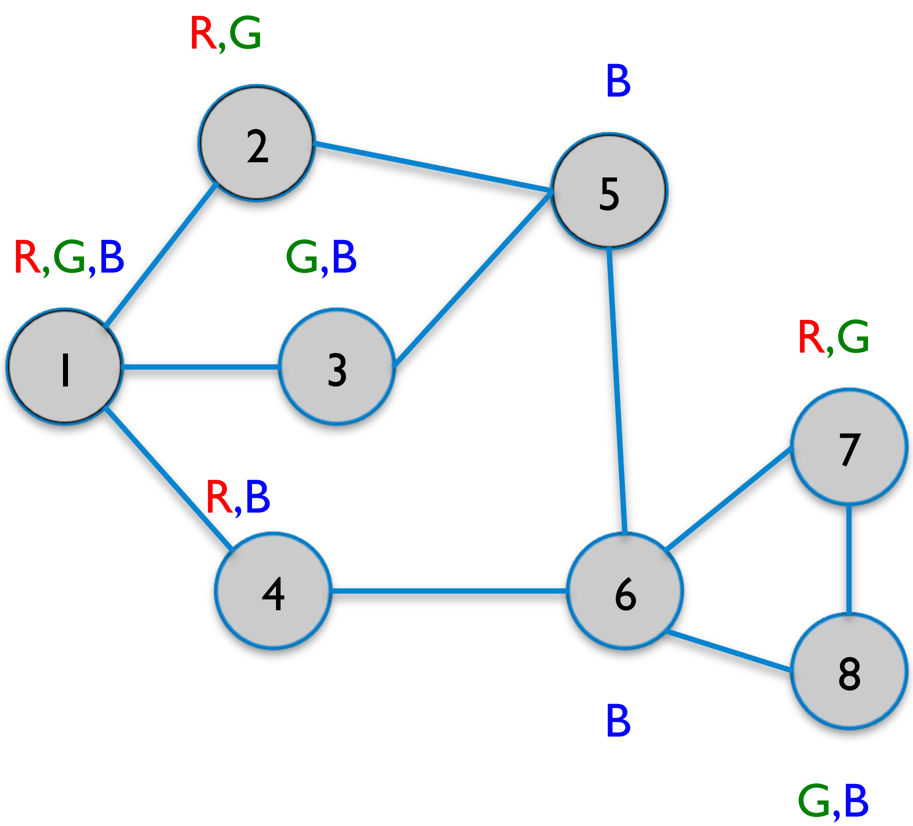
Subset Resource Replication Problem

- Want ~~all~~ SUBSET of colors.
- Can store ~~one~~ s_v colors per server
- Same Min Max objective.
- Also studied by Ko and Rubenstein[2004]
  - their heuristic does not provide any approximation guarantee.

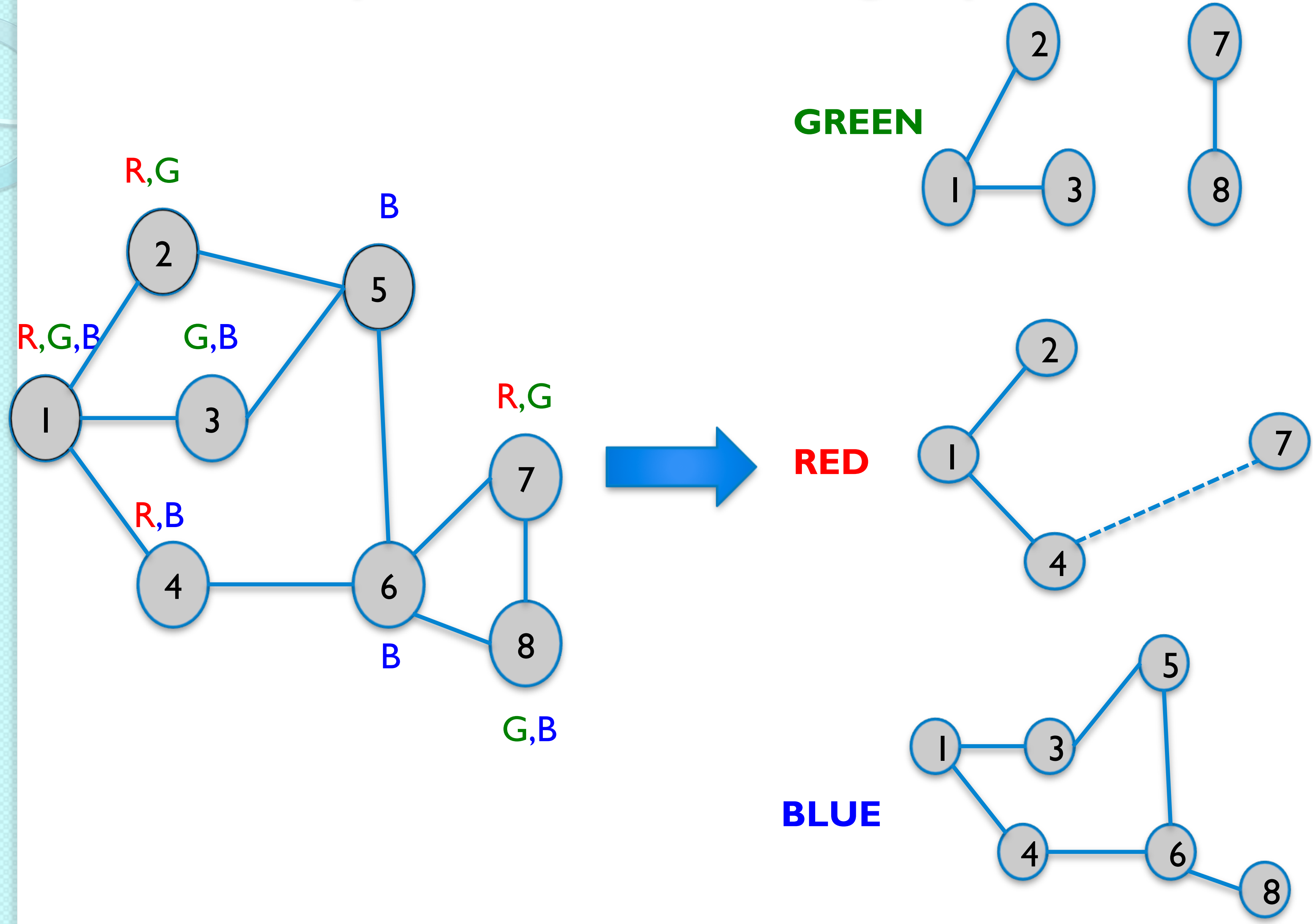
Given G and resources {R,G,B}



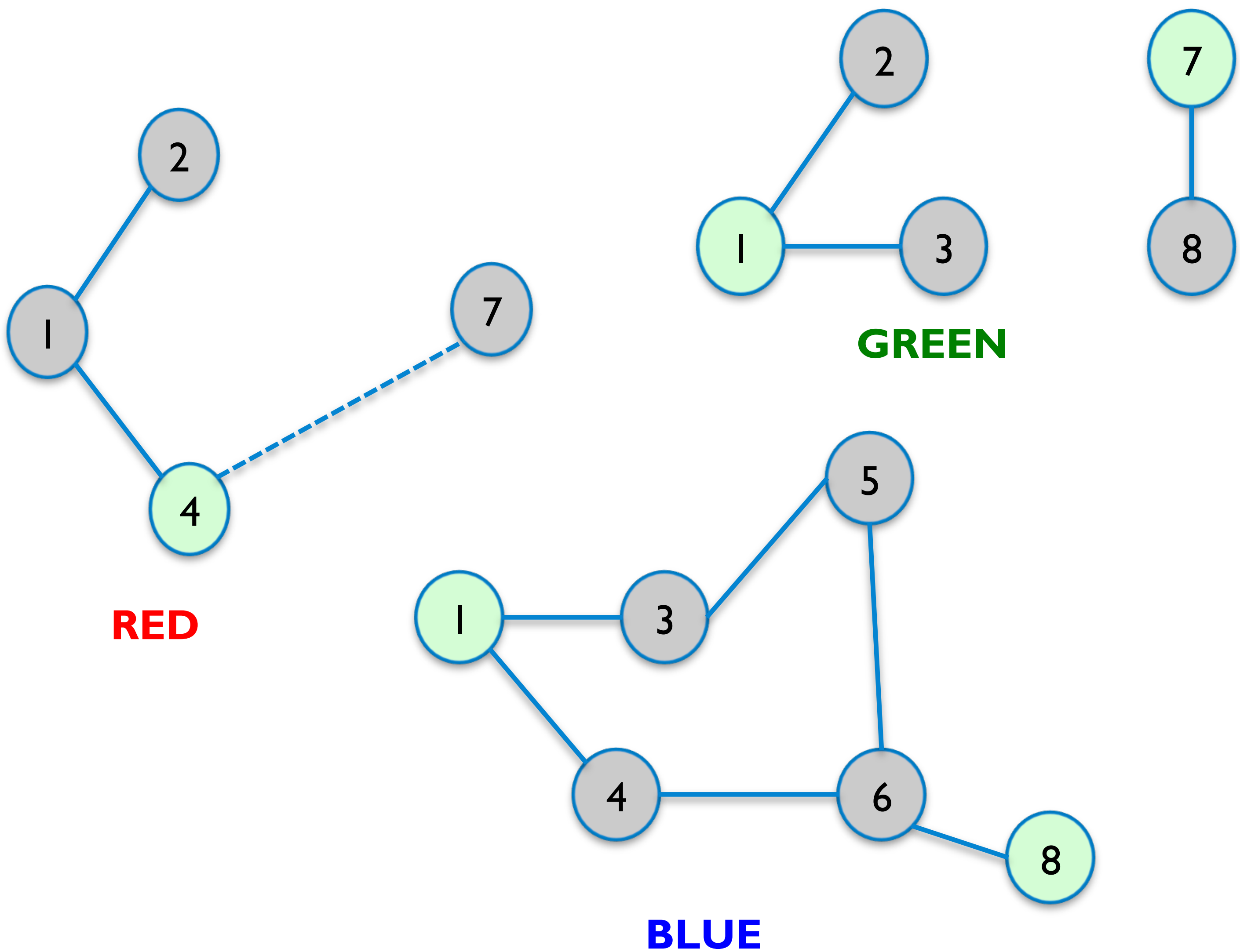
Guess δ ➔ threshold graph



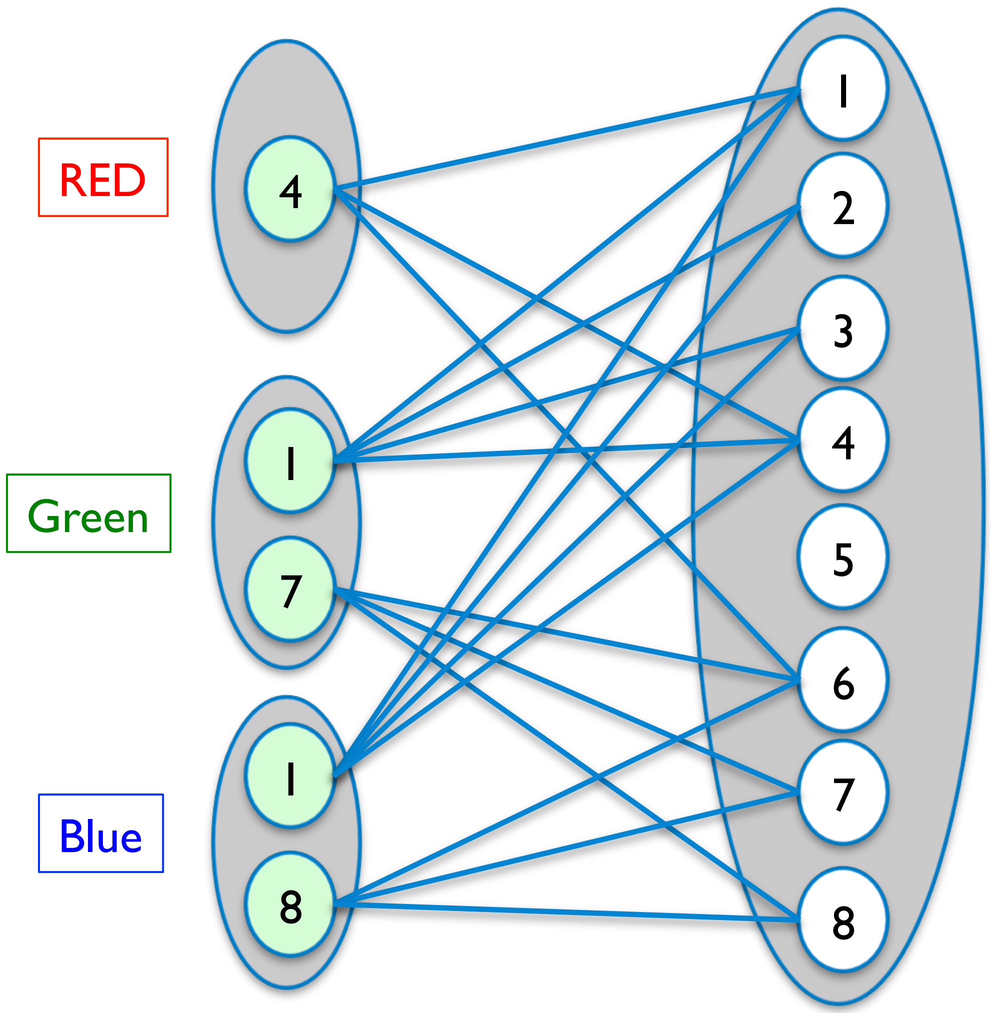
Decompose into sub-graphs.



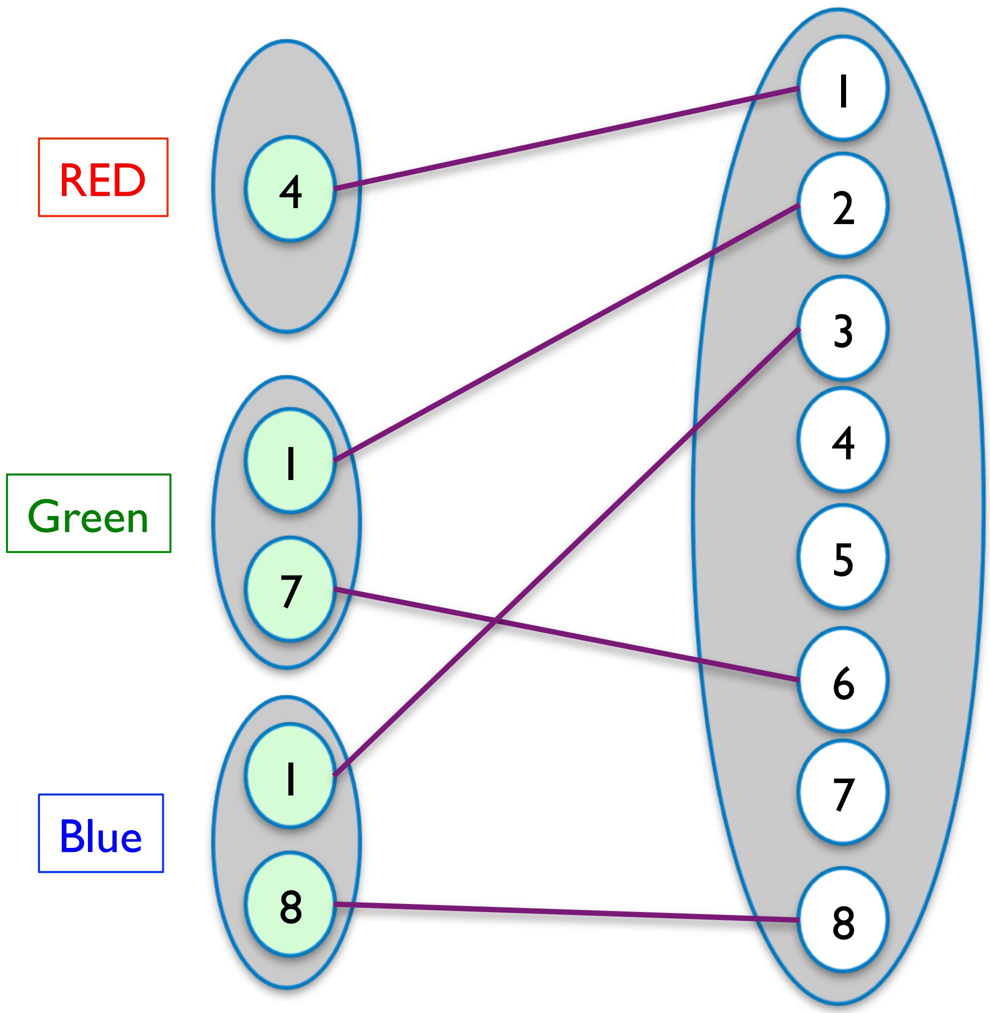
Compute **MIS** for each color graph



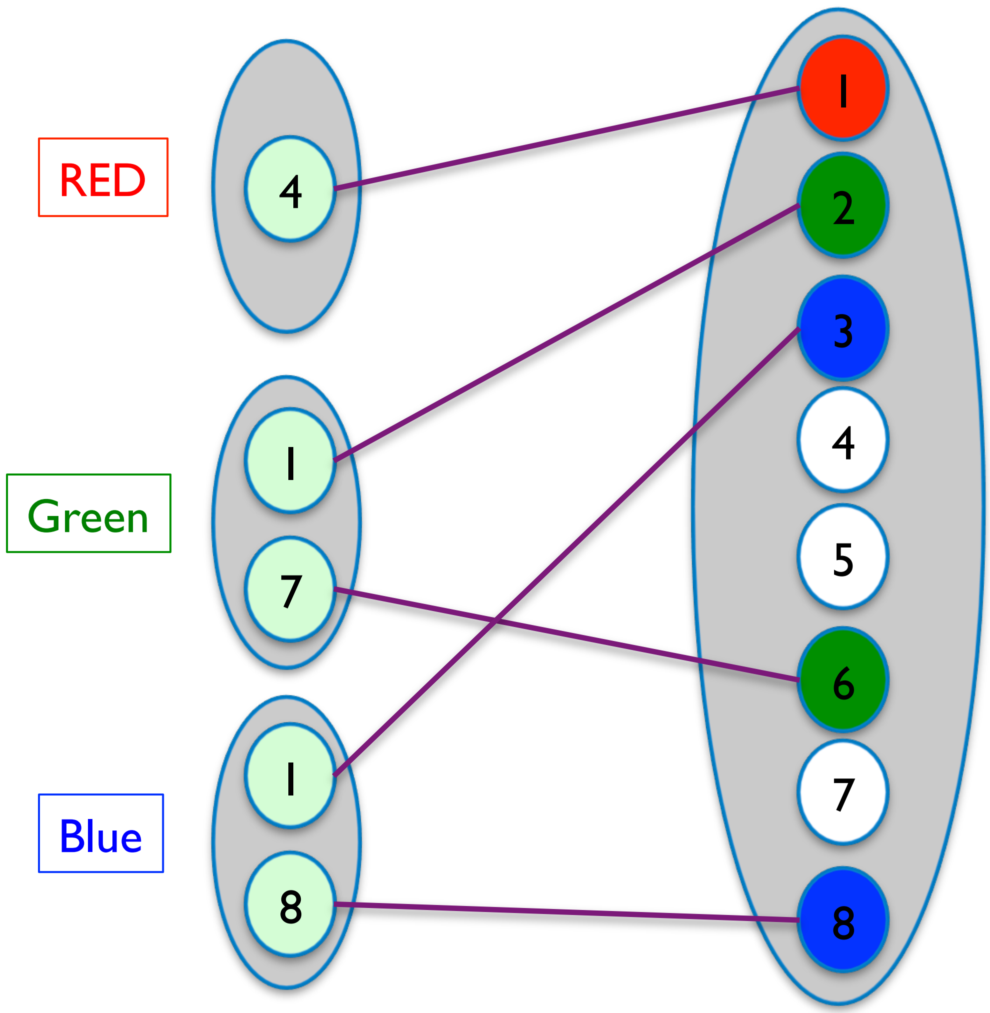
Compute a matching



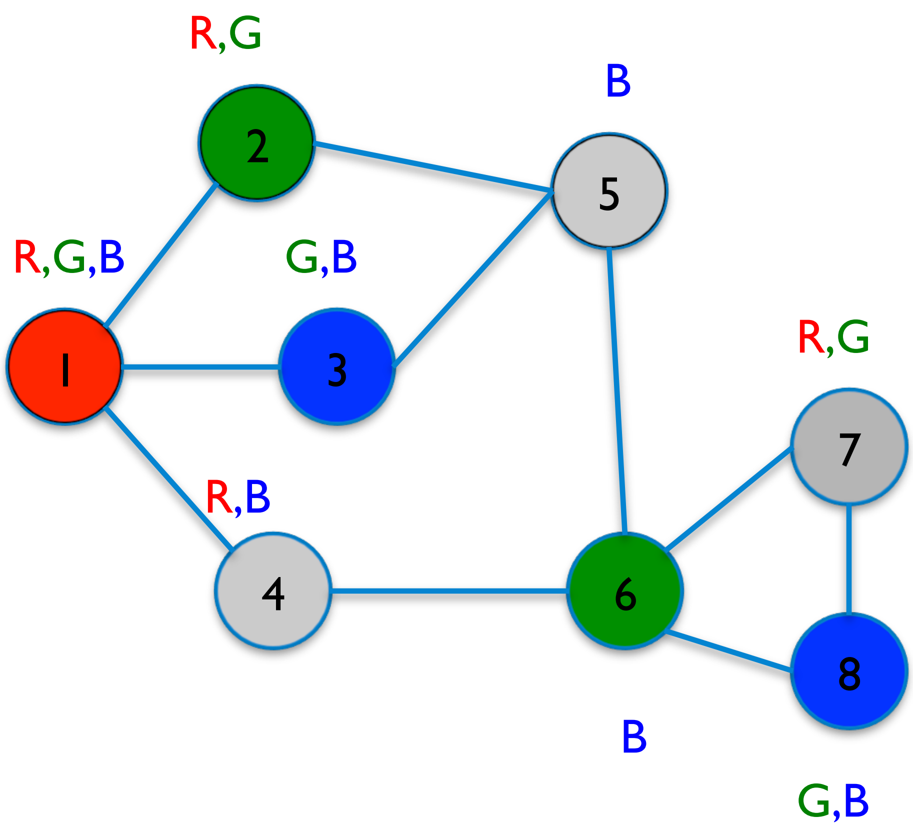
Compute a matching



Compute a matching



Assigning the colors.



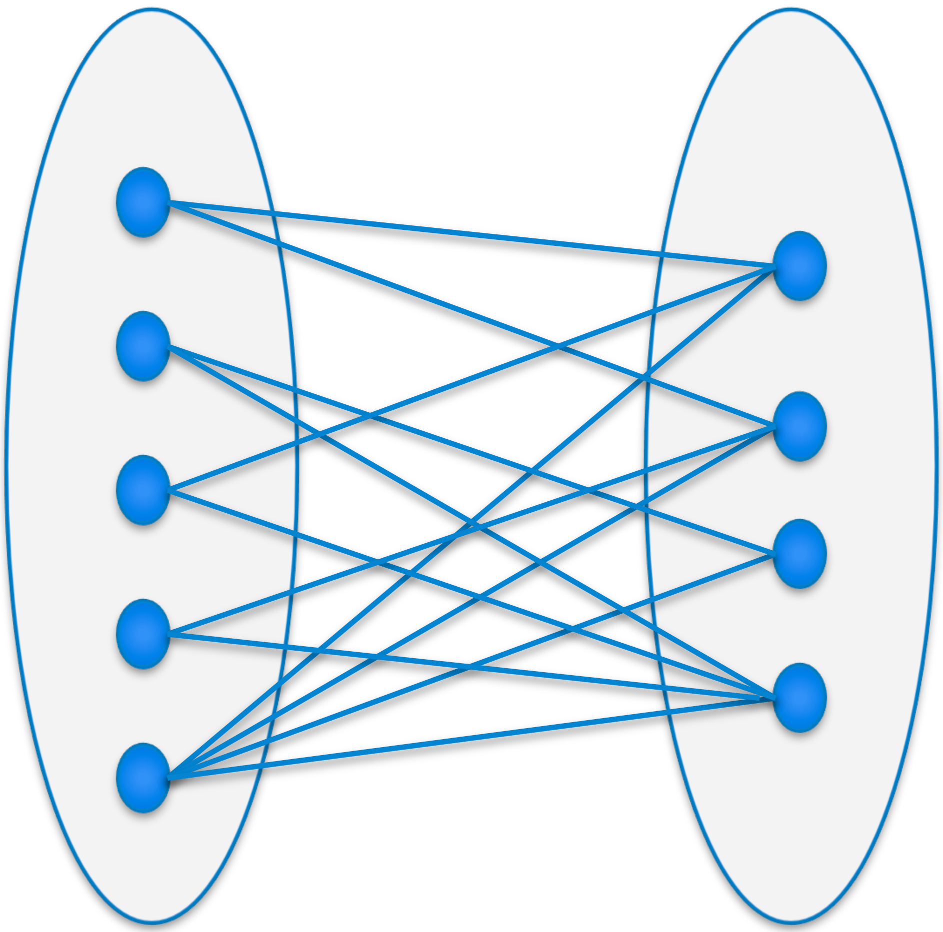
Subset Resource Replication Problem

- *Theorem 1: There exists a polynomial time 3-approximation algorithm for the Subset Resource Replication problem.*

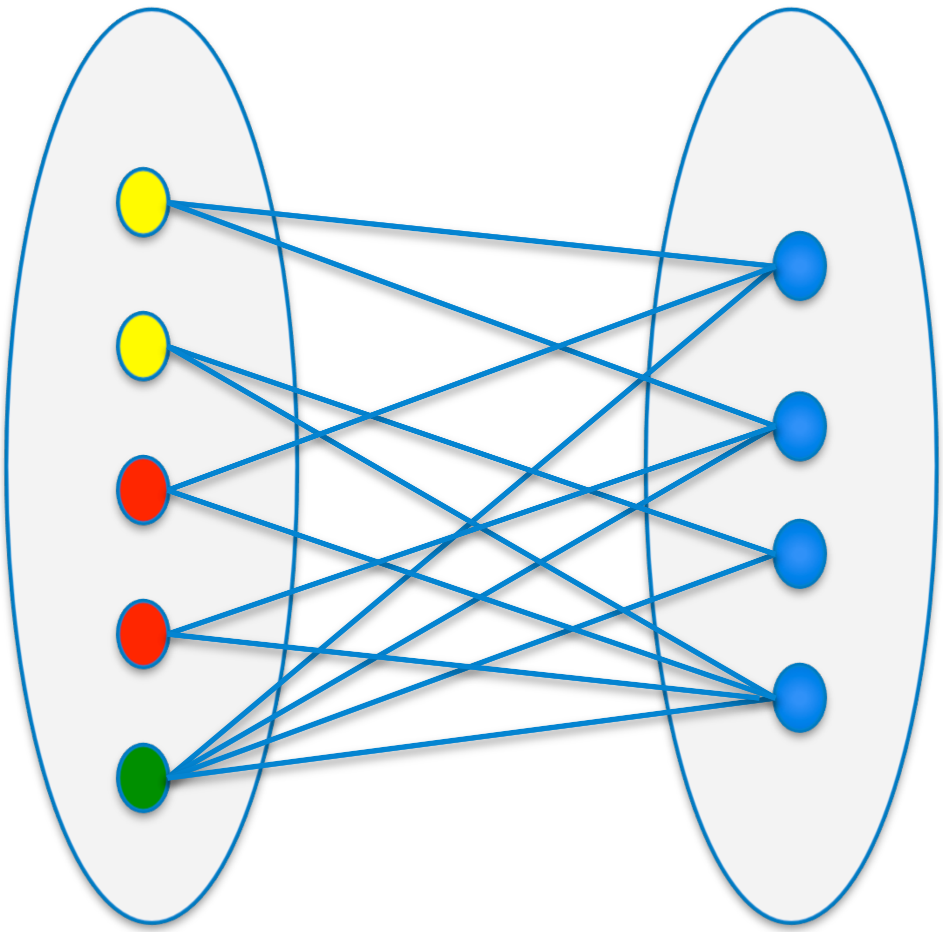
(3- ε) Hardness of Approximation

- Reduction from One Sided Domatic Number Problem (Feige, Halldorsson, Kortsarz and Srinivasan).
- Similar to the dominating set based reduction for the k-center problem.
- A similar reduction from Domatic Number Problem gives a (2- ε) hardness of approximation for Basic Resource Replication Problem.

One Sided Domatic Number.

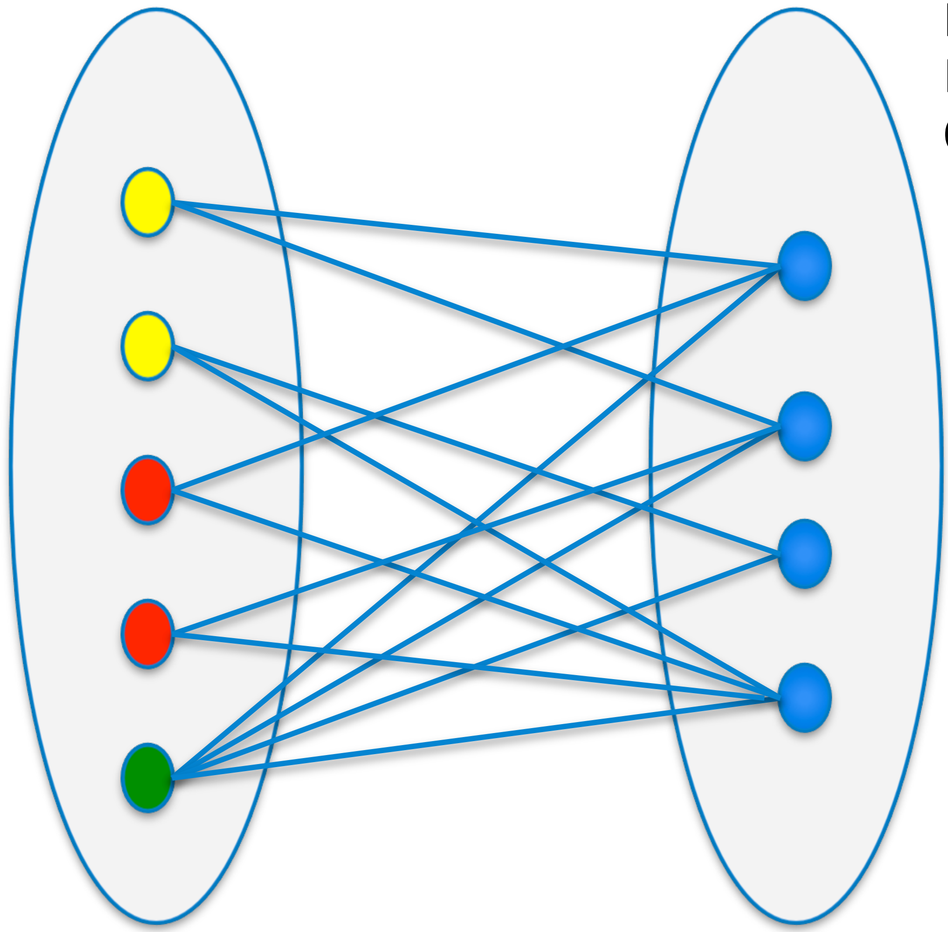


One Sided Domatic Number.



Domatic Number = 3

Reduction:



Blue Nodes:
Require all colors.
0 Capacity

Distance Metric:
Each edge – 1
Anti edge – 3

Other Vertices:
Unit capacity.
No resource required

Outlier Version

- Basic Resource Replication problem
  - 3-approximation
- Basic Resource Replication problem with bounds on number of replicas of each resource
  - 5-approximation
- Subset Resource Replication problem
  - No nontrivial approximation factor unless P=NP

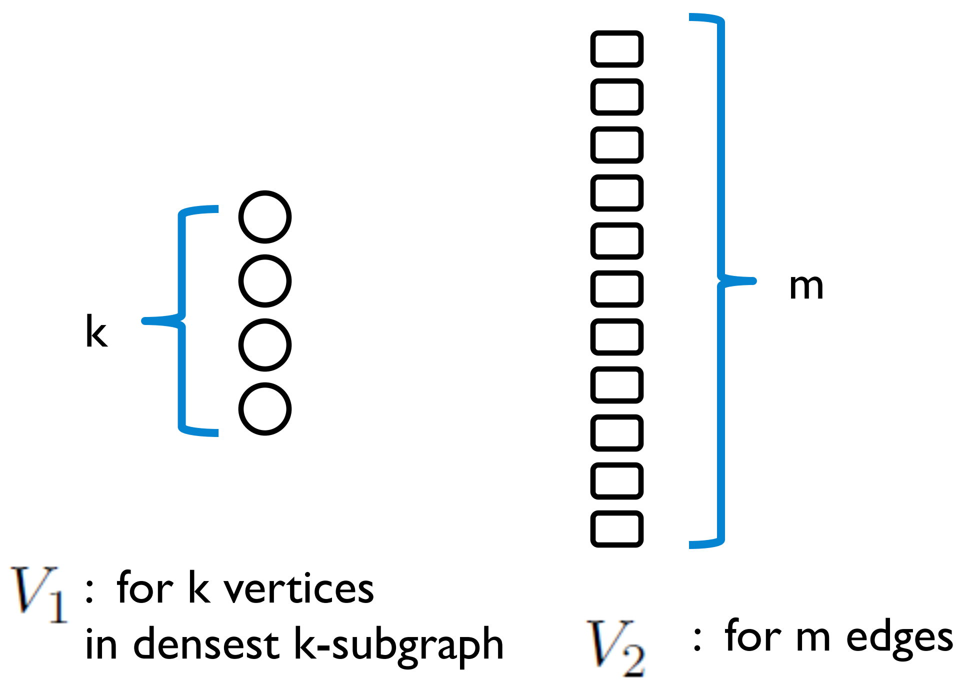
Outlier Version: Robust Subset Resource Replication Problem (RSRR)

- *Theorem : Assuming $P \neq NP$, there is no polynomial time algorithm which gives a positive approximation ratio for Robust Resource Replication Problem.*

- A polynomial time reduction of the densest k-sub-graph problem to the problem of deciding the "feasibility" of RSRR.

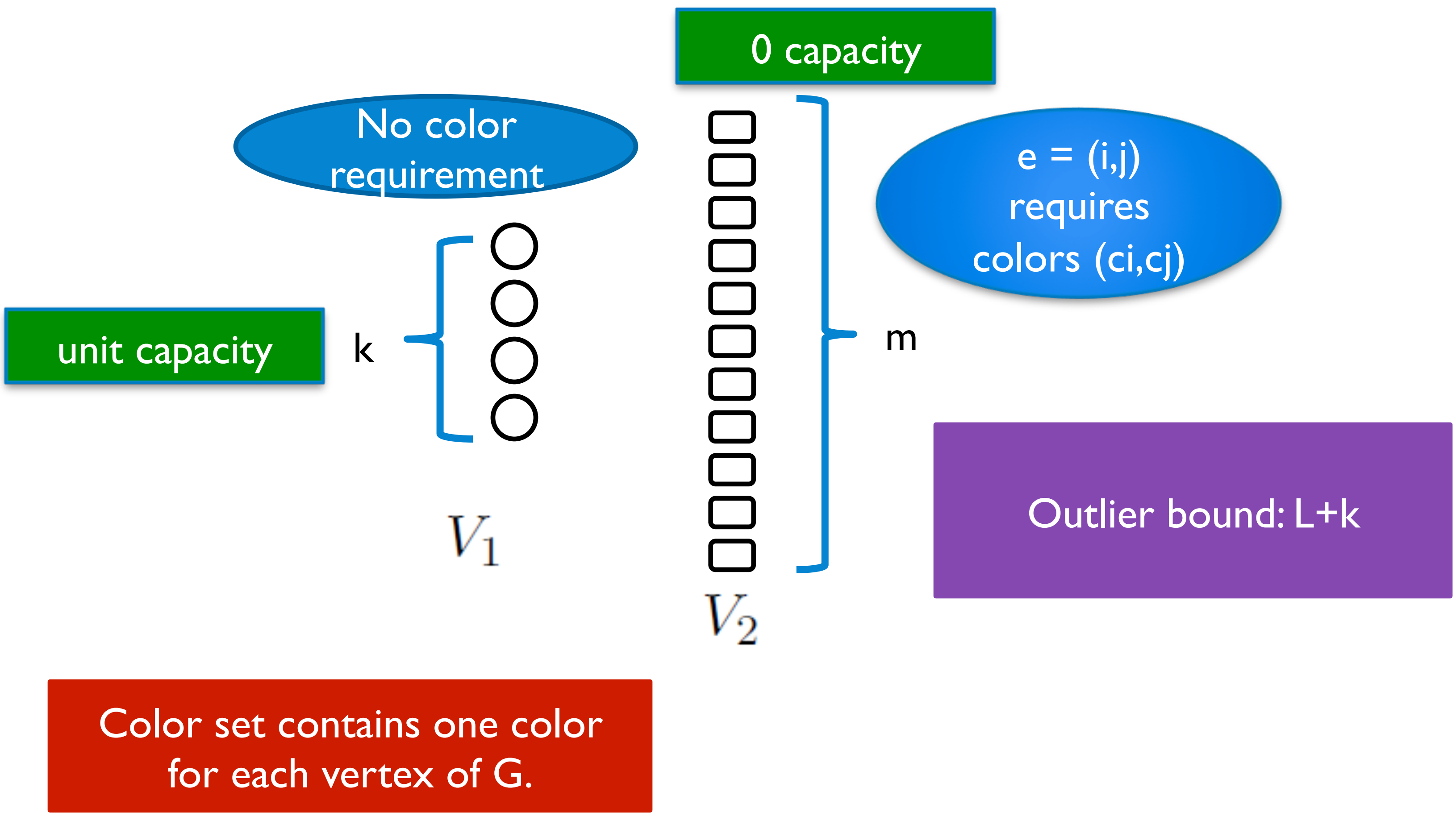
Robust Subset Resource Replication Problem (RSRR)

- Decision version of densest k-sub-graph problem:
  - Instance $\mathcal{I} = (G = (V, E), k, L),\ |V| = n,\ |E| = m$, decide if in G, there exists a sub-graph of size k containing L edges.
- Construct an instance of RSRR as follows:



Robust Subset Resource Replication Problem (RSRR)

- Construct an instance of RSRR as follows:



Future Direction

- Studying other objectives such as Min Sum with bound on number of replicas/ cost for replication
- Extending Load constraint to Subset Resource Replication problem/ Matching lower and upper bound for Basic Resource Replication Problem etc.

Questions ?