

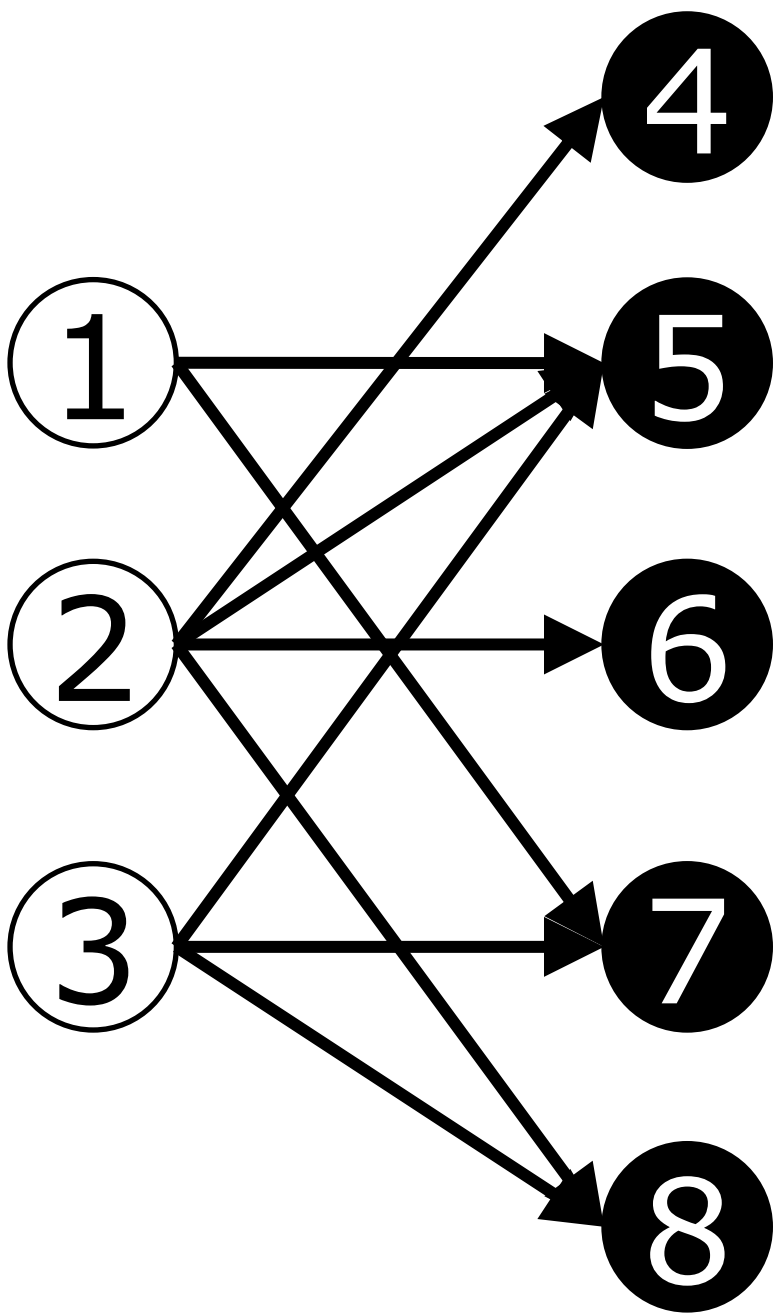
BiGraph: Bipartite-oriented Distributed Graph Partitioning for Big Learning

Rong Chen, Jiaxin Shi, Binyu Zang and Haibing Guan

What is the Bipartite Graph?

“A **bipartite** graph is a graph whose vertices can be divided into two **disjoint** sets U and V (that is, U and V are each independent sets) such that every edge connects a vertex in U to one in V ”

WIKIPEDIA
The Free Encyclopedia

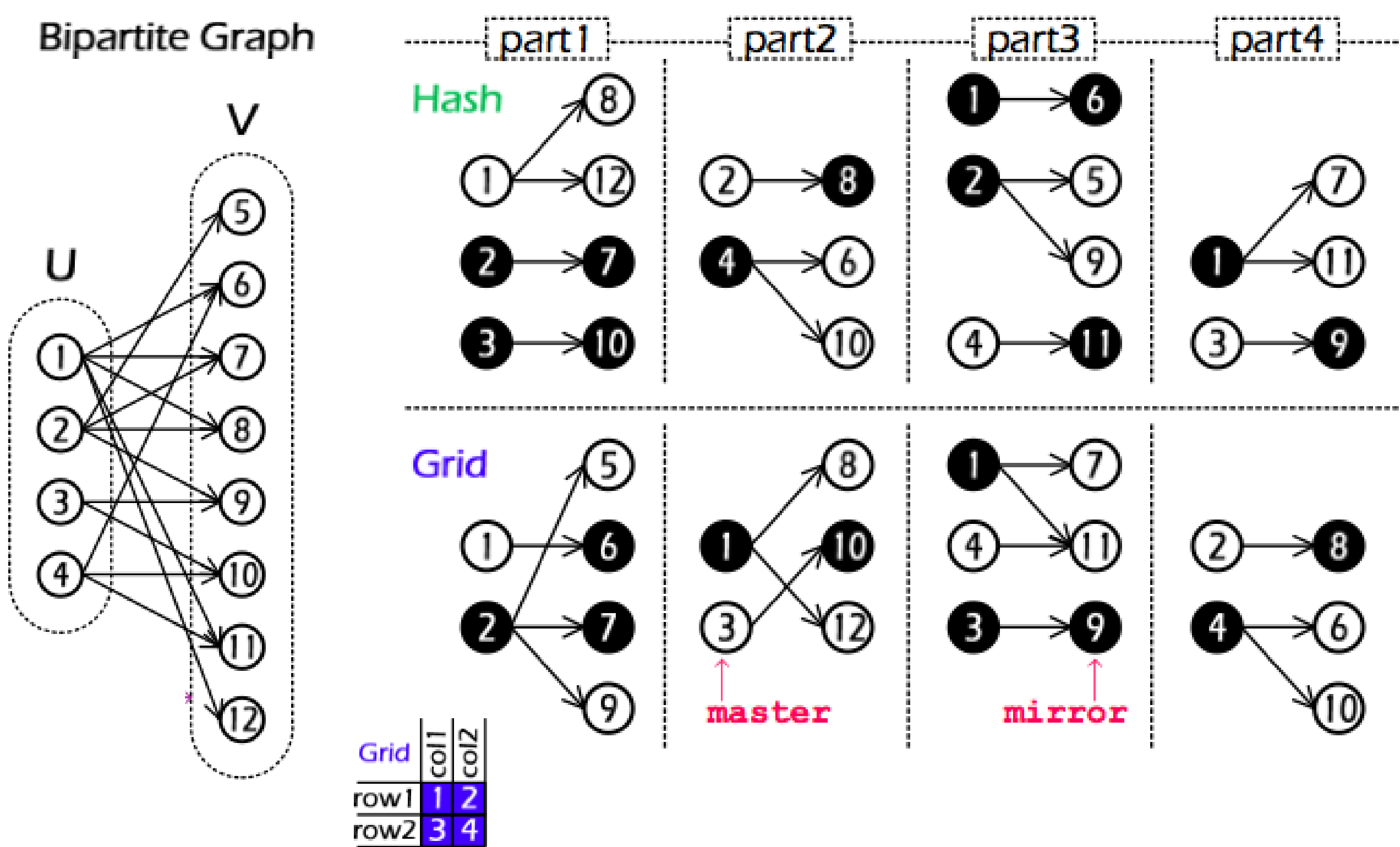


Many **Machine Learning** and **Data Mining** (MLDM) algorithms can be modeled as computing on **bipartite** graphs

→ e.g. Recommendation, Topic modeling, Medical diagnosis

Distributed Graph Partitioning

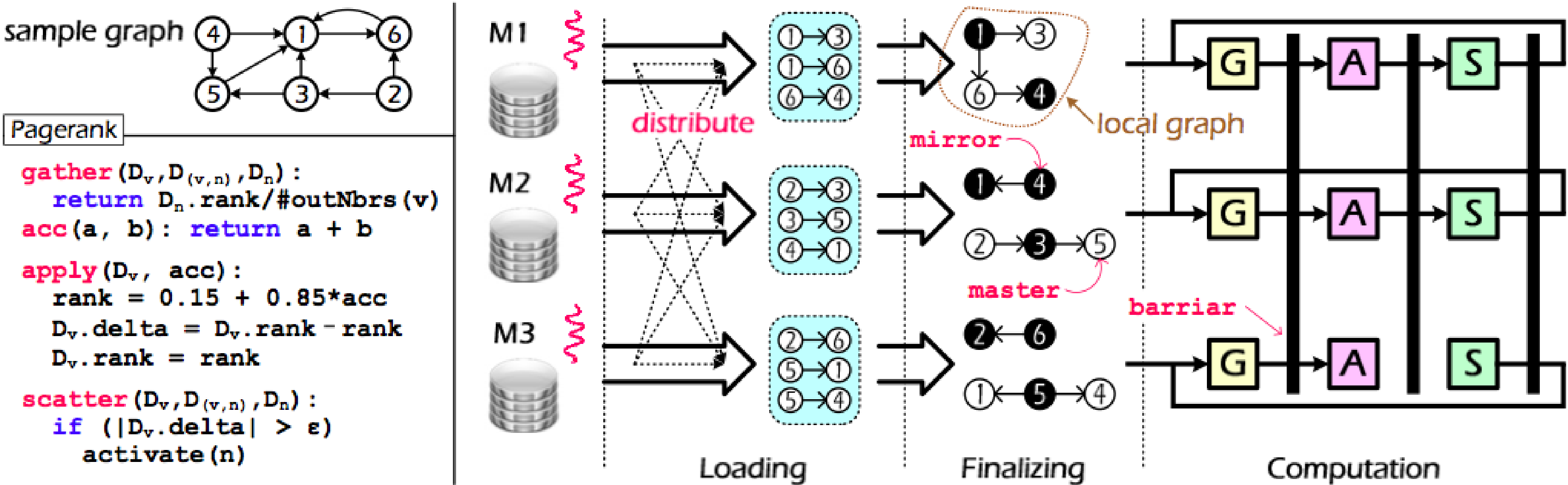
Existing Vertex-cut on **Bipartite** Graph



J. Gonzalez, et al. PowerGraph: Distributed graph-parallel computation on natural graphs. In OSDI, 2012.

N. Jain, et al. Graphbuilder: scalable graph ETL framework. In Workshop on Graph Data Management Experiences and Systems, 2013.

Graph-parallel Systems (e.g. GraphLab 2.2)



source code: <https://github.com/graphlab-code/graphlab>

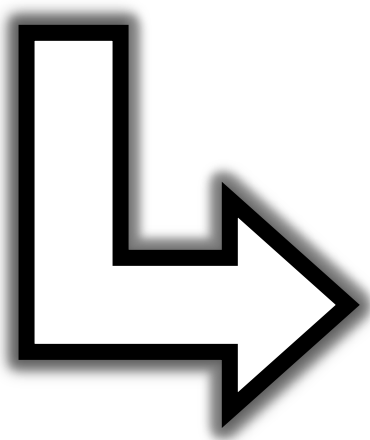
GraphLab

Challenges and Opportunities

1. Real-world bipartite graphs for MLDM are usually **imbalanced**
2. The comp. load of many MLDM algorithms may also be **skewed**
3. The size of data associated with vertices can be critical **skewed**

Bipartite-oriented Graph Partitioning

“Arbitrarily partitioning vertices belonging to the same subset may not introduce any replicas of vertices”

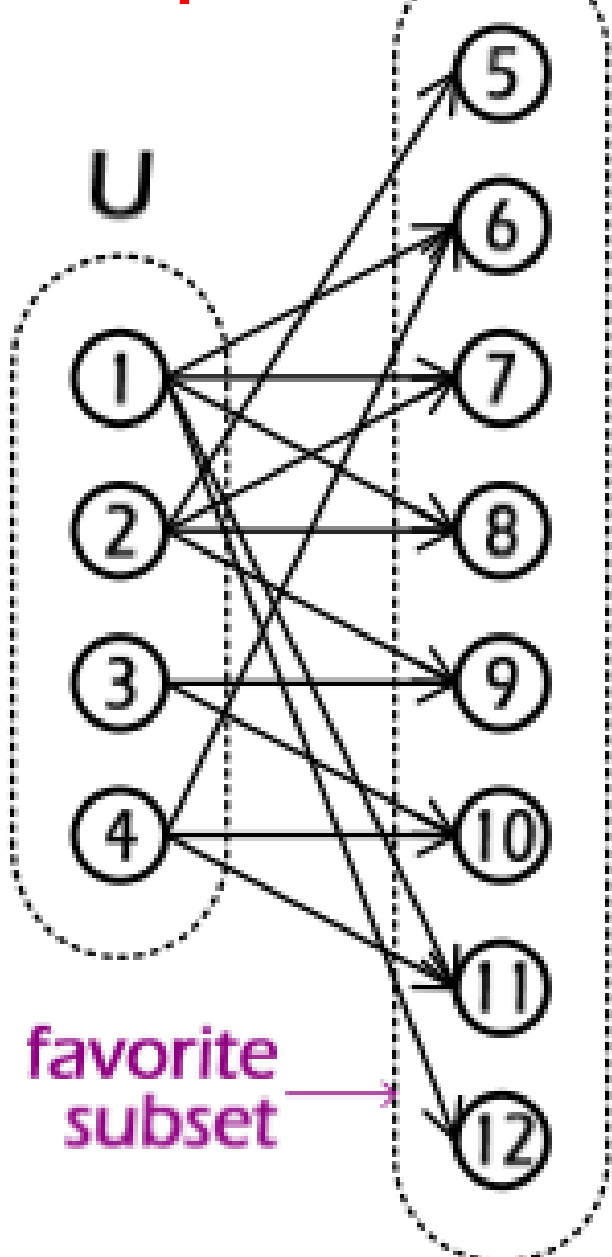


Bipartite-oriented Graph Partitioning

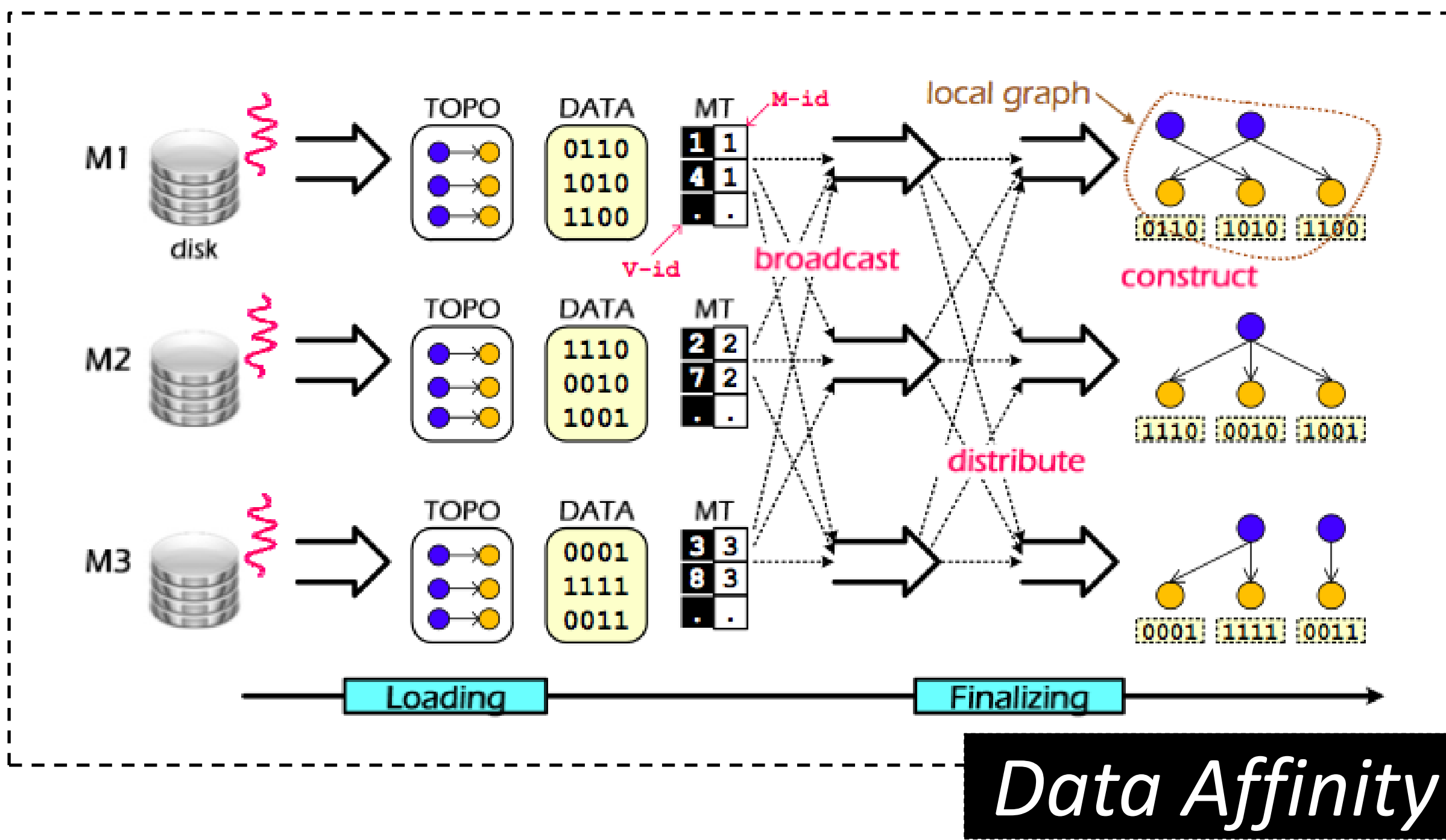
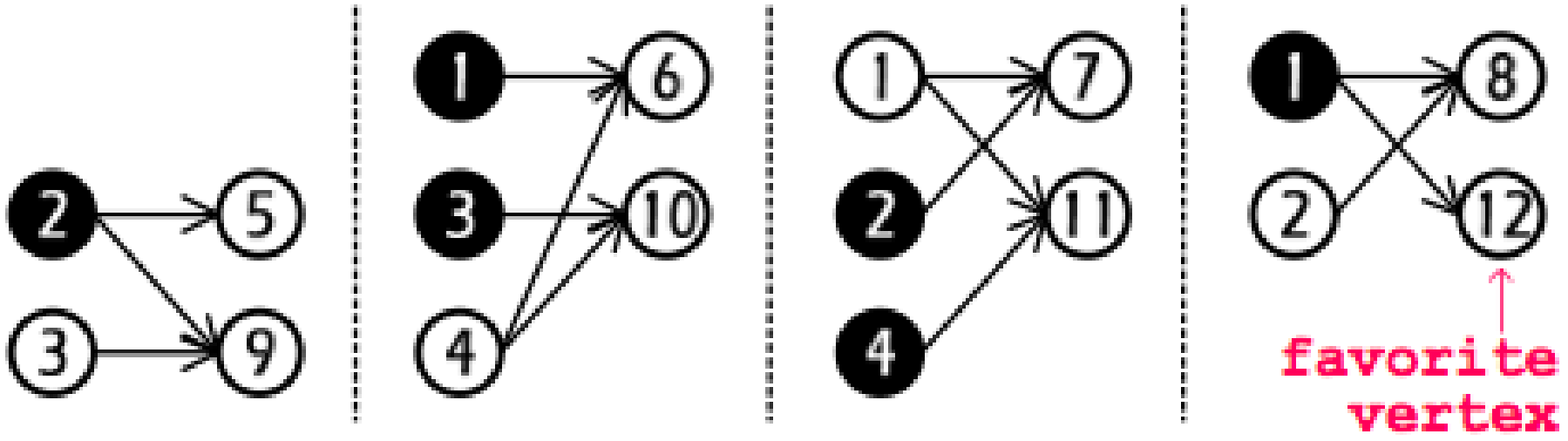
1. Bipartite-cut
2. Greedy heuristic
3. Data affinity

source : <http://ipads.se.sjtu.edu.cn/projects/powerlyra.html>

Bipartite
Graph



Bipartite-cut



Data Affinity

Experiment

