E. GALIMBERTI, F. BONCHI, F. GULLO

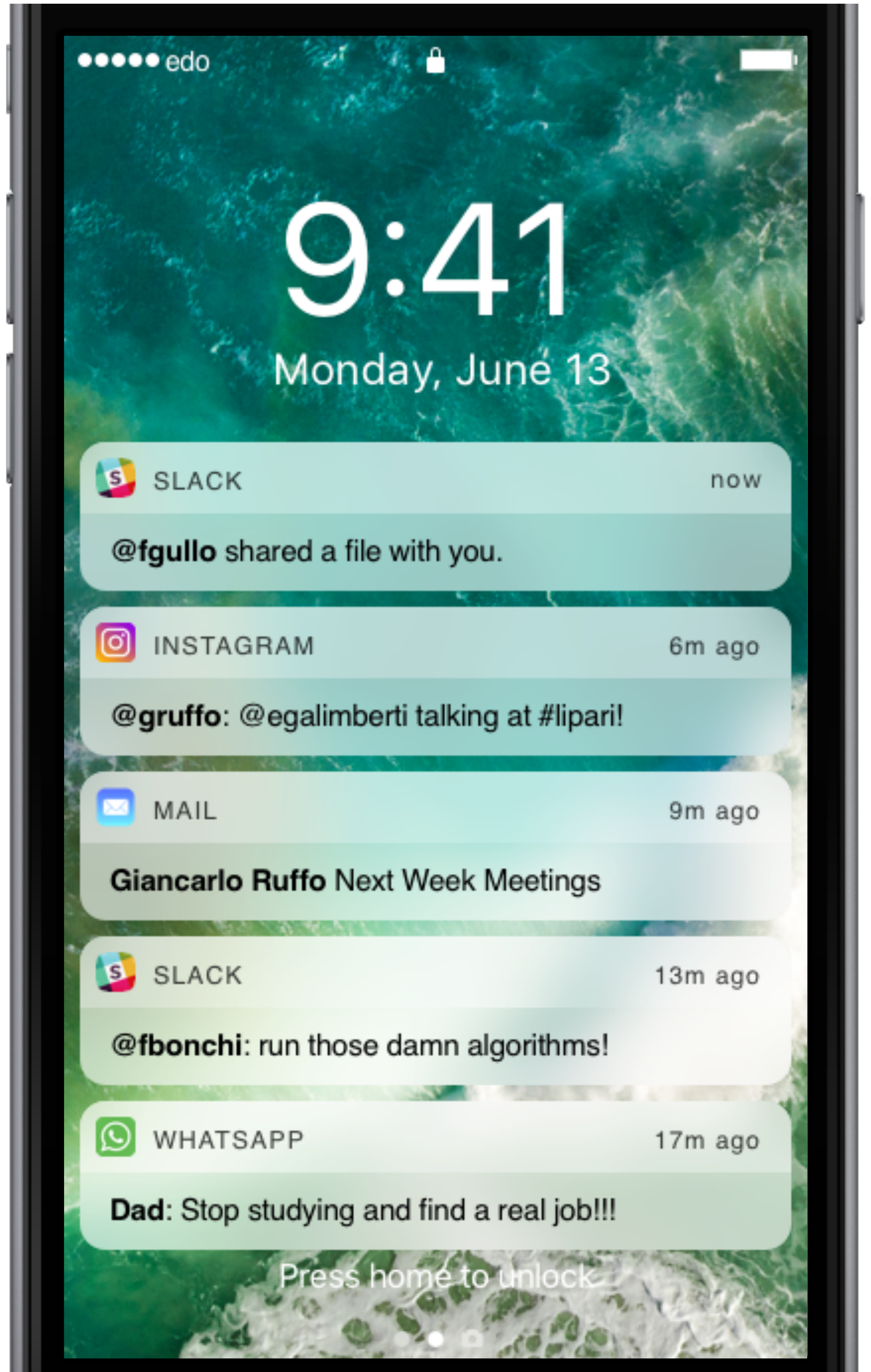
CORE DECOMPOSITION AND DENSEST SUBGRAPH IN MULTILAYER NETWORKS

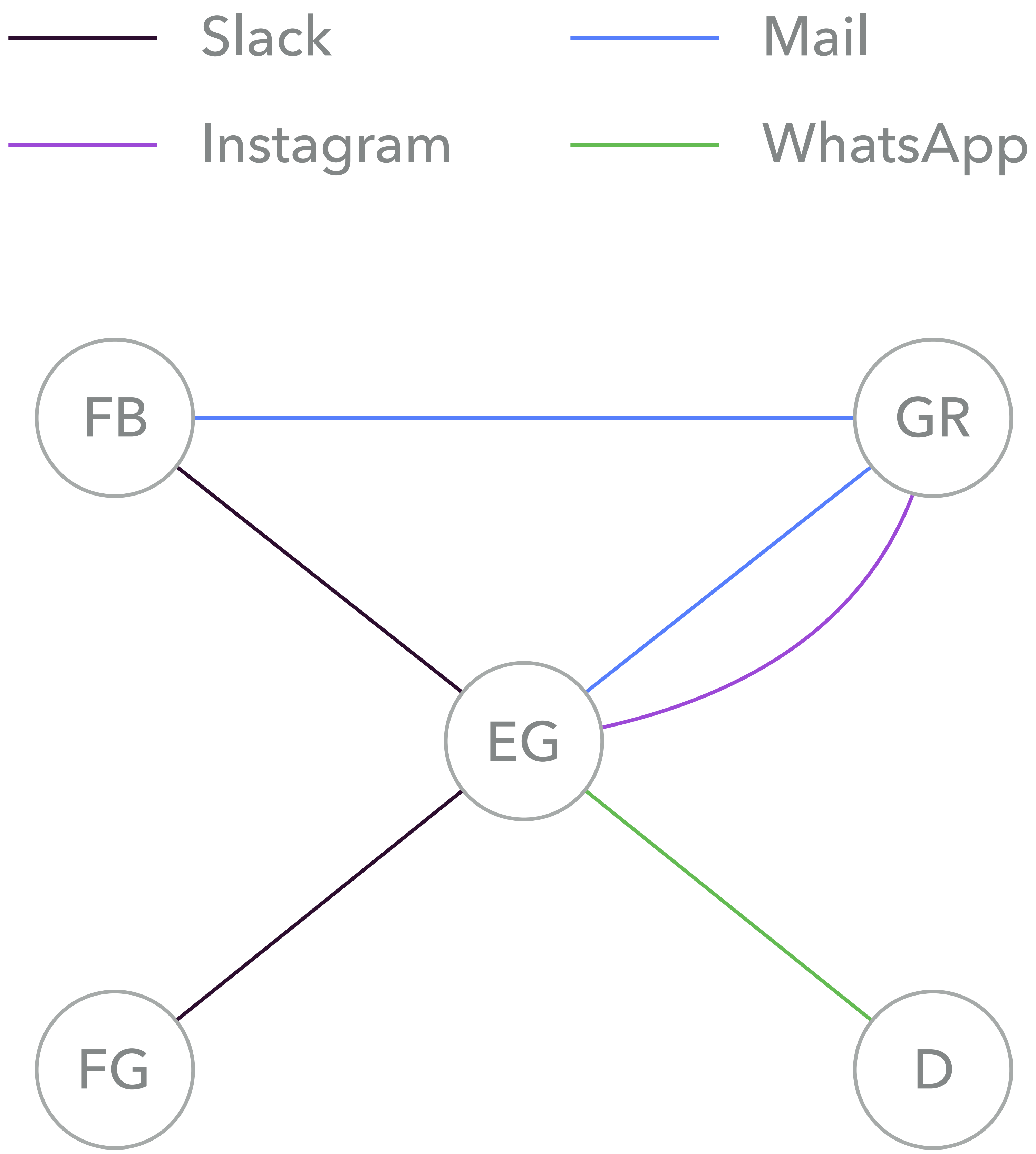
AGENDA

- Multilayer Networks
- Core Decomposition and Densest Subgraph
- Multilayer Core Decomposition
- Experiments
- Multilayer Densest Subgraph

MULTILAYER NETWORKS

MULTILAYER NETWORKS





MULTILAYER NETWORKS

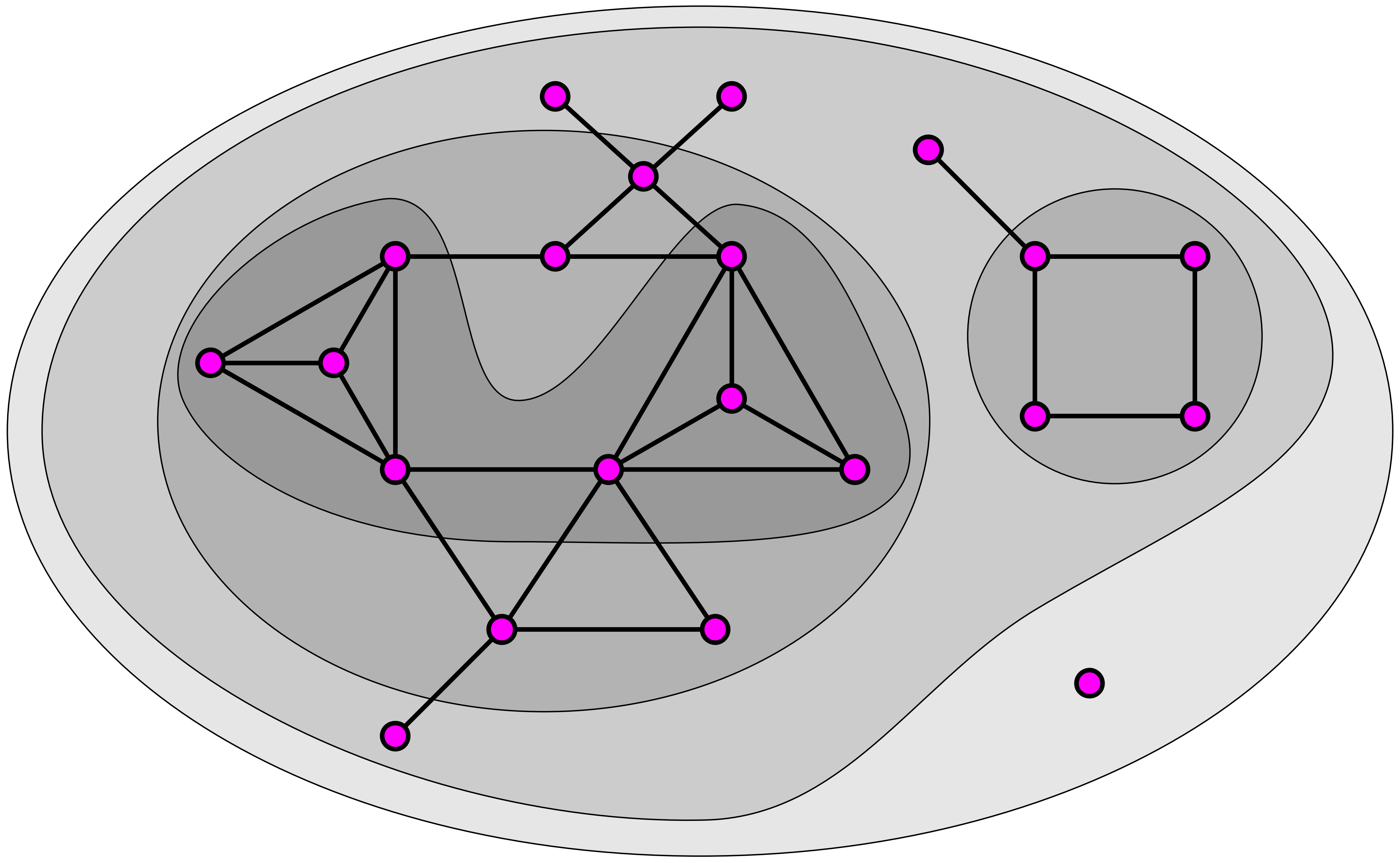
- Many real-world applications:
  - social media
  - biology
  - finance
  - transportation systems
  - critical infrastructures

- Represented by **multilayer graphs** $G=(V,E,L)$ where
  - V is a set of **vertices**
  - L is a set of **layers**
  - $E \subseteq V \times V \times L$ is a set of **labeled edges**

CORE DECOMPOSITION AND DENSEST SUBGRAPH

CORE DECOMPOSITION

Given a simple, single-layer, graph G.
The **k-core** (or core of order k) of G is a maximal subgraph $G[C_k]$ such that every vertex u in C_k has degree at least k.
The set of all k-cores forms the **core decomposition** of G.



CORE DECOMPOSITION

- It can be computed in linear time
- It has been studied for various types of graph
  - uncertain
  - directed
  - weighted
- Azimi-Tafreshi *et al.* study the **core-percolation problem** on multilayer networks from a physics standpoint, **without providing any algorithm**

DENSEST SUBGRAPH

Given a simple, single-layer, graph G.
The **densest subgraph** is the subgraph of G maximizing the average-degree density.

- Exact polynomial time algorithm
- Linear-time 1/2-approximation algorithm
- Jethava *et al.* formulate the **densest common subgraph** problem, i.e., find a subgraph maximizing the minimum average degree over **all layers**

MULTILAYER CORE DECOMPOSITION

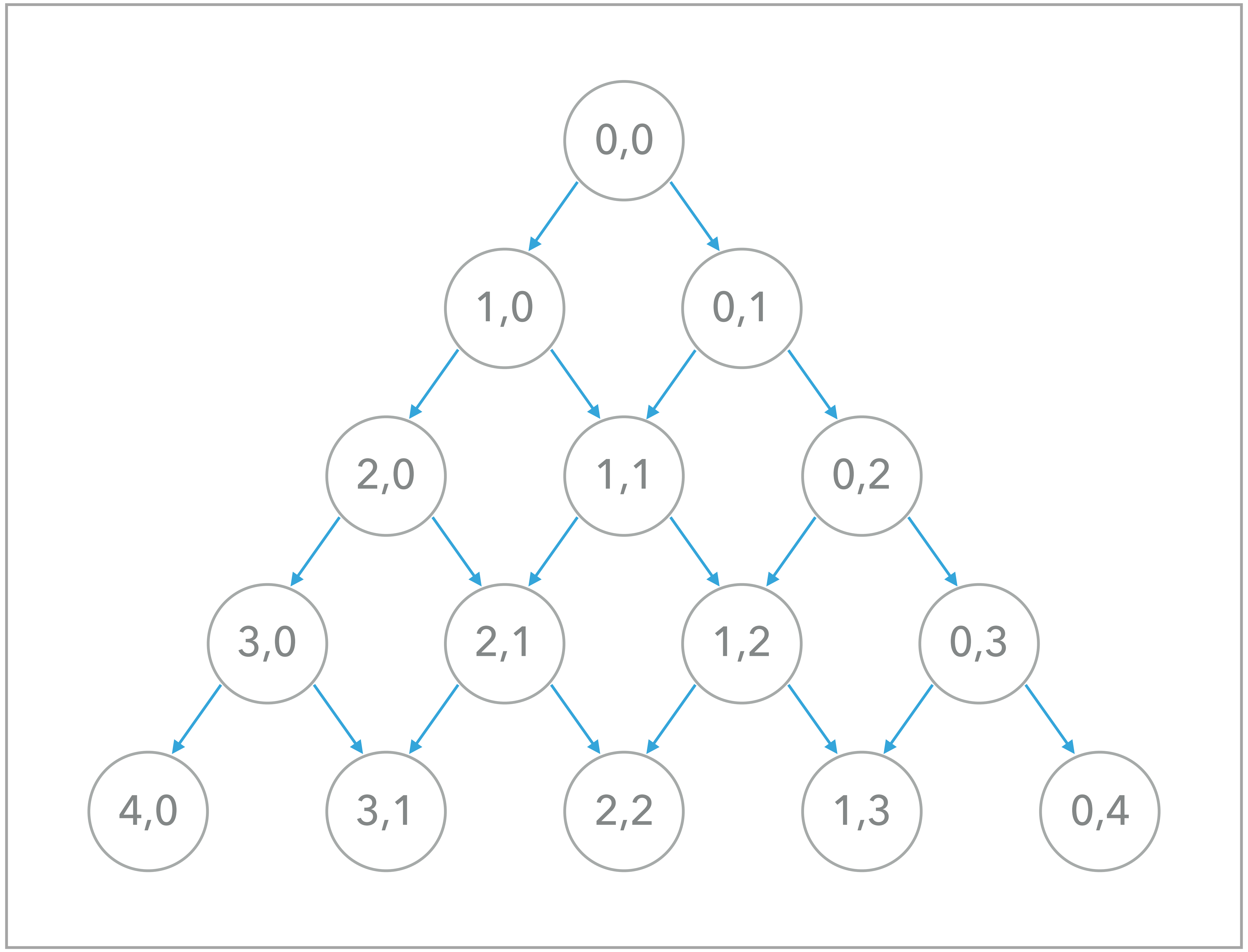
MULTILAYER CORE DECOMPOSITION

Let G=(V,E,L) be a multilayer graph and an |L|-dimensional integer vector $\mathbf{k}=[k_l]$.
The **multilayer k-core** of G is a maximal subgraph $G[C_\mathbf{k}]$ whose vertices have at least degree k_l in $C_\mathbf{k}$, for all layers l in L.

Given a multilayer graph G=(V,E,L), find the set of all **non-empty** and **distinct** multilayer cores of G.
Such a set constitutes the **multilayer core decomposition** of G.

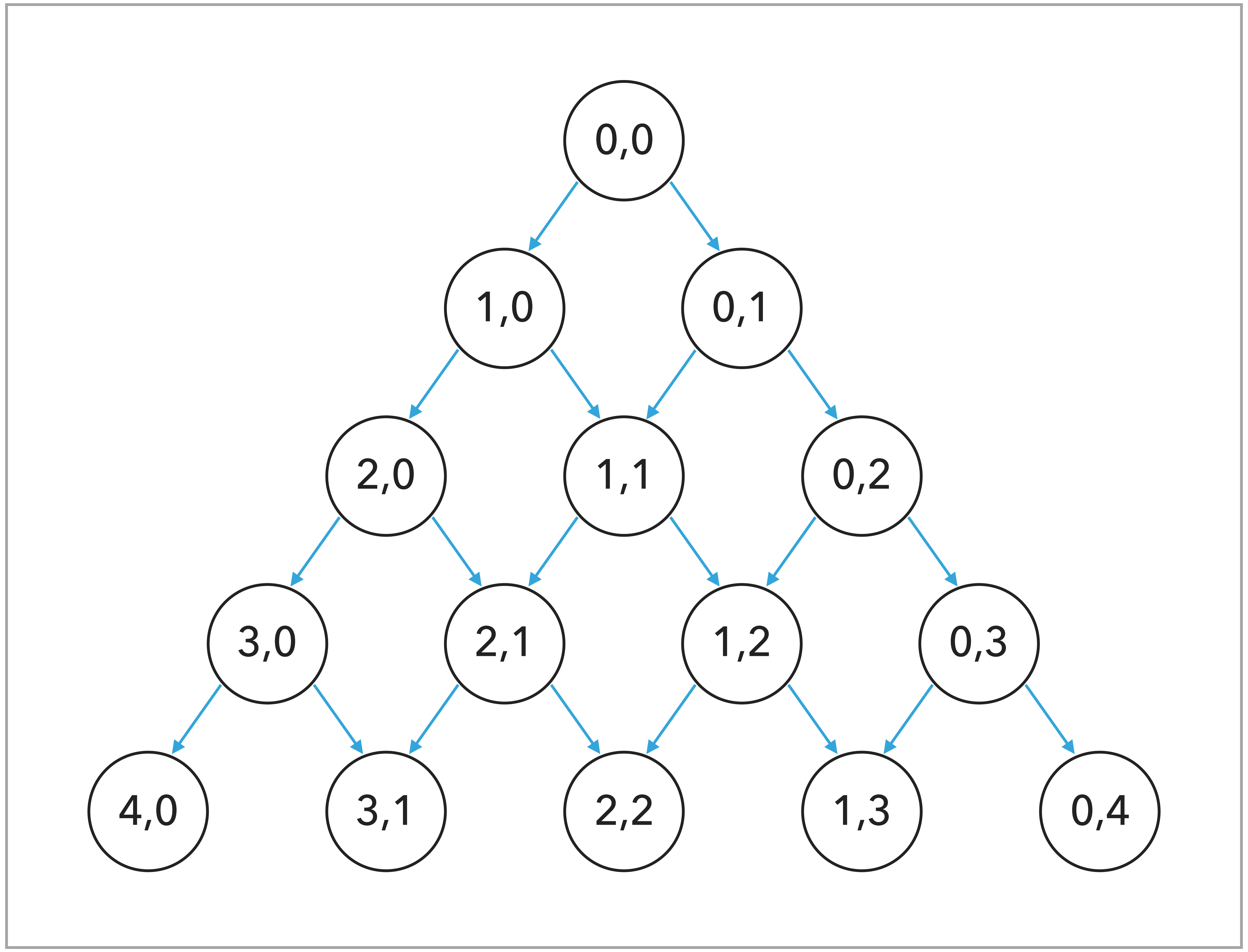
- The number of multilayer cores to be output may be exponential in the number of layers
- **No polynomial-time algorithm can exist**

SEARCH SPACE: CORE LATTICE



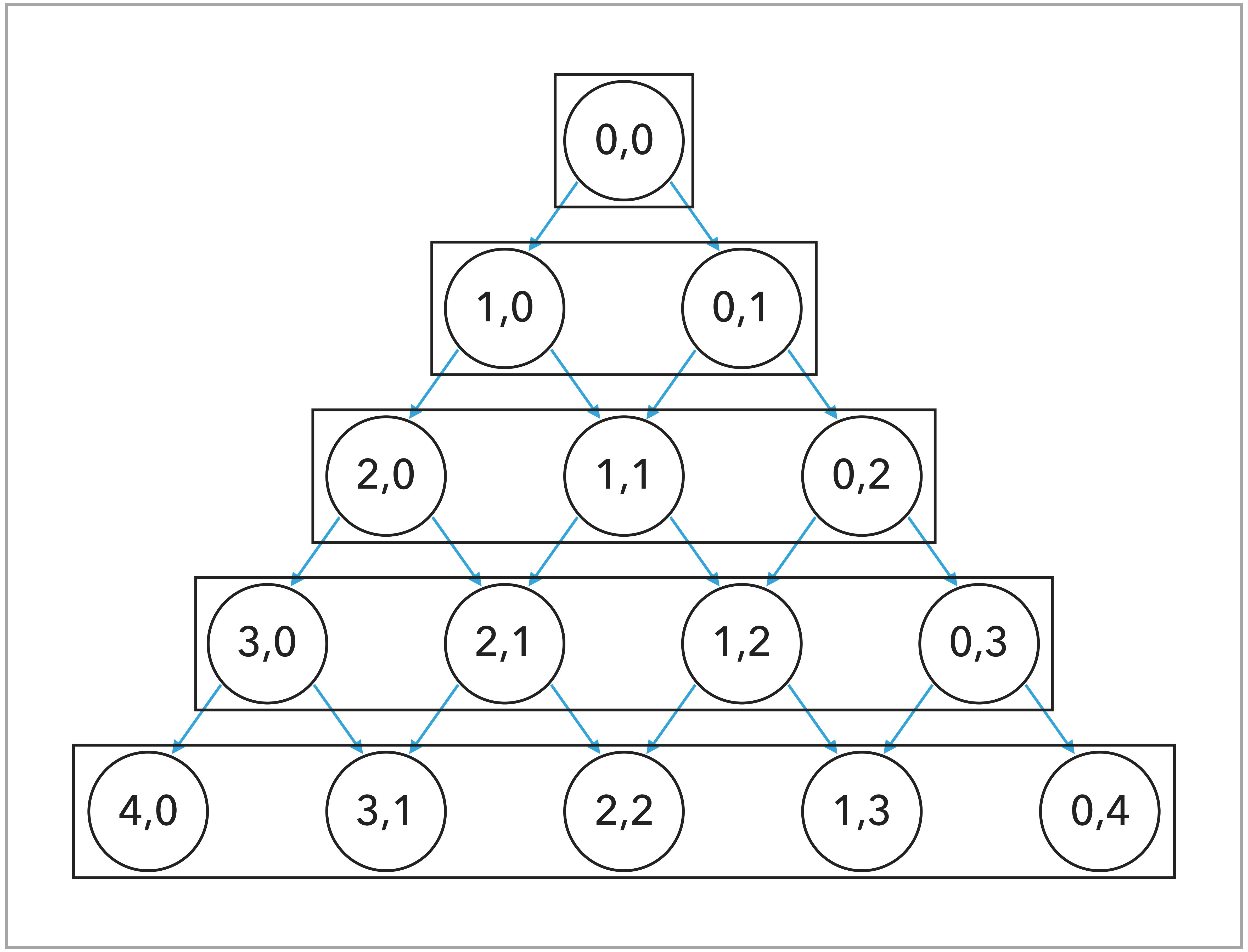
- A **k**-core with coreness vector **k**=[k_l] is **contained** into any **k'**-core described by a coreness vector **k'**=[$k_{l'}$] whose components $k_{l'}$ are all no more than components k_l

NAIVE ALGORITHM



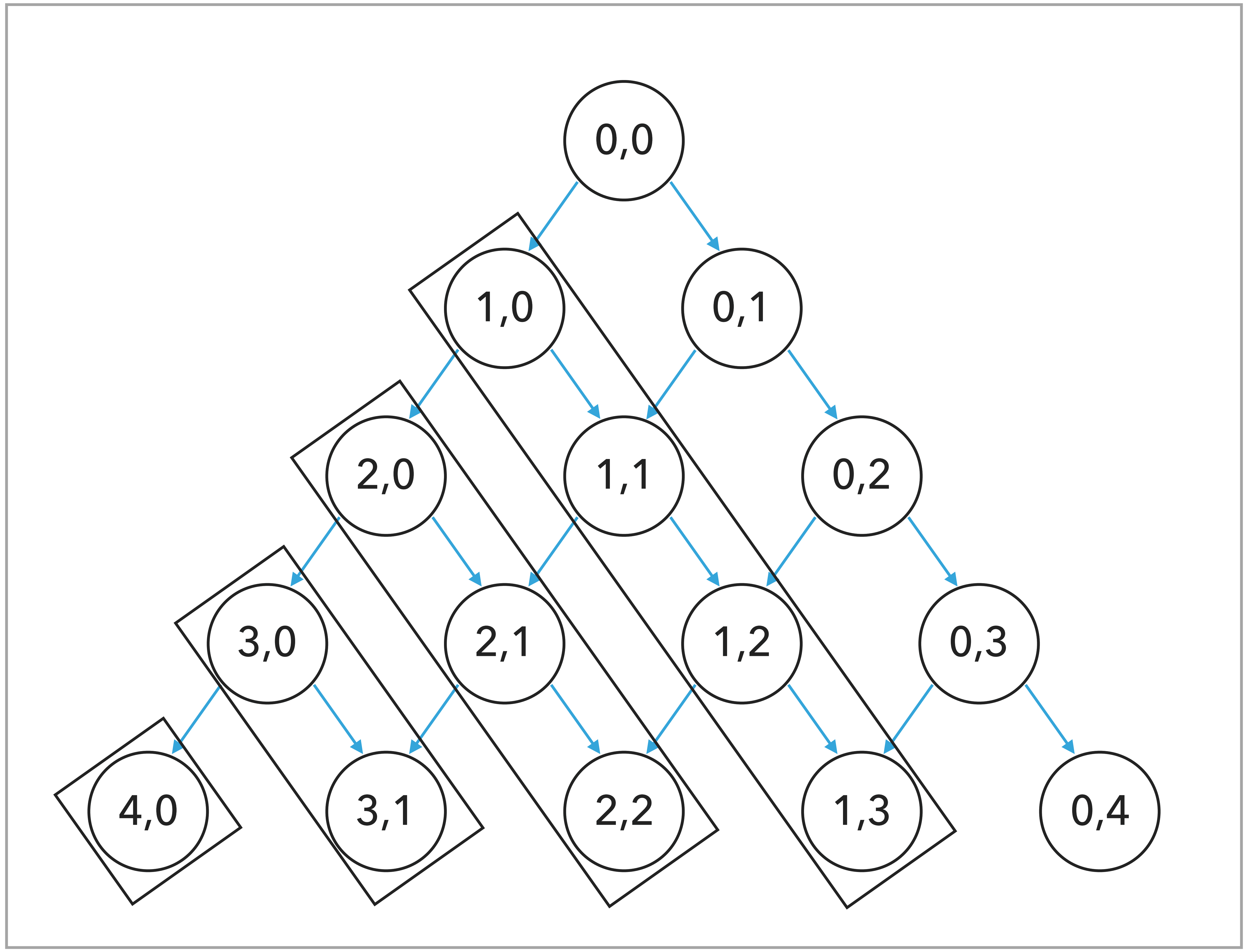
- Every possible core is computed separately and without a specific ordering
- Weaknesses:
  - each core is computed starting from the whole input graph
  - a lot of non-distinct and/or empty (thus, unnecessary) cores may be computed

BREADTH-FIRST ALGORITHM



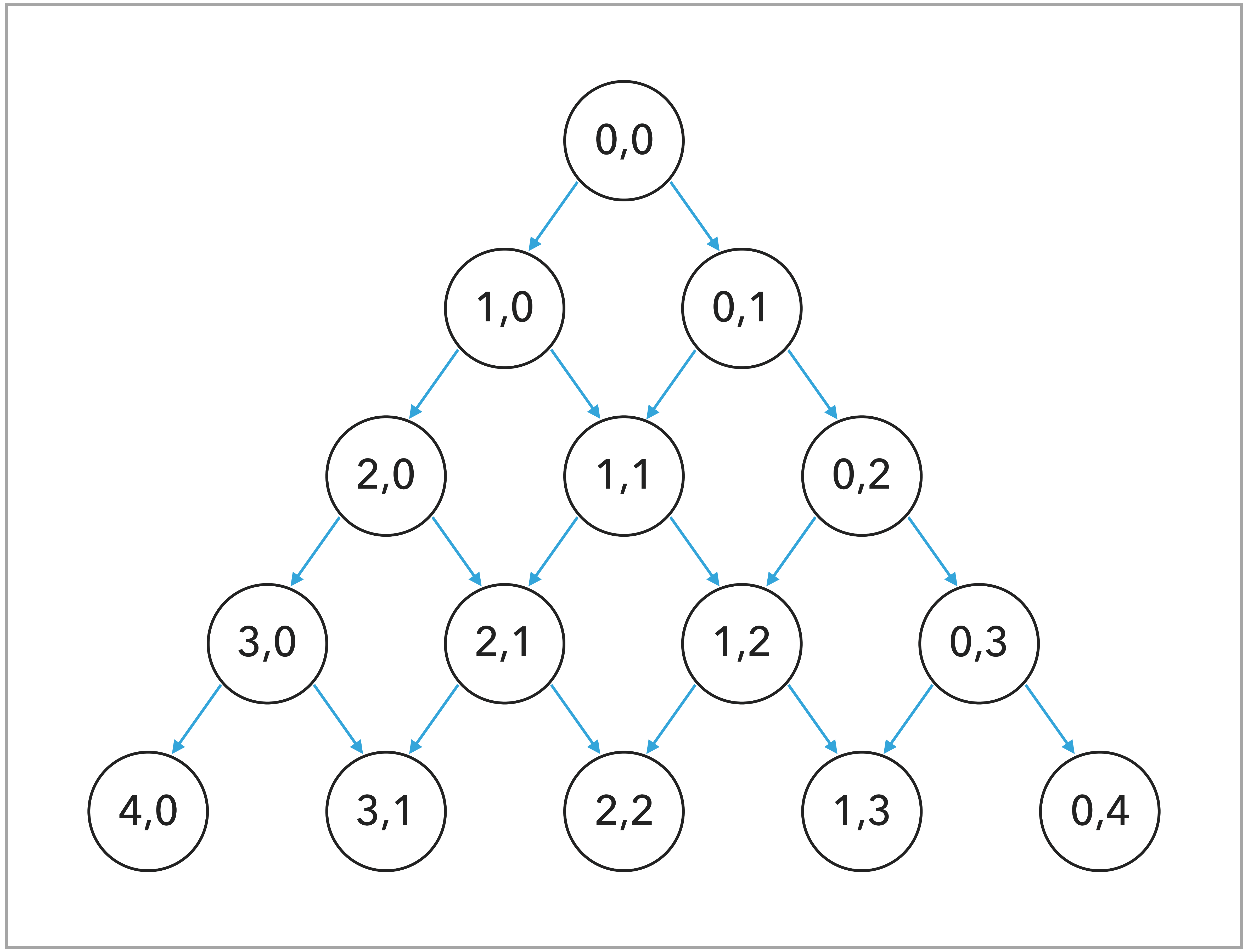
- The core lattice is explored **level by level**
- Cores are computed from the intersection of all their fathers
- Cores having less fathers then the number of non-zero components of its coreness vector **k** are not visited
- Weaknesses:
  - the computation of the cores within a straight path can be performed more efficiently
  - non-distinct cores are computed

DEPTH-FIRST ALGORITHM



- The core lattice is explored **path by path**, resembling a depth-first search
- The algorithm iteratively picks a non-leaf core **k**=[k_l] and computes all cores in the path varying a component of k
- Not all paths have to be explored to visit the whole core lattice
- Weaknesses:
  - cores may be computed multiple times
  - cores are computed starting from larger subgraphs
  - non-distinct cores are still computed

HYBRID ALGORITHM



- The algorithm starts with a single-layer core decomposition for each layer
- Then it performs a breadth-first search equipped with a **"look-ahead" mechanism**
- All cores are computed once and non-distinct cores are skipped

EXPERIMENTS

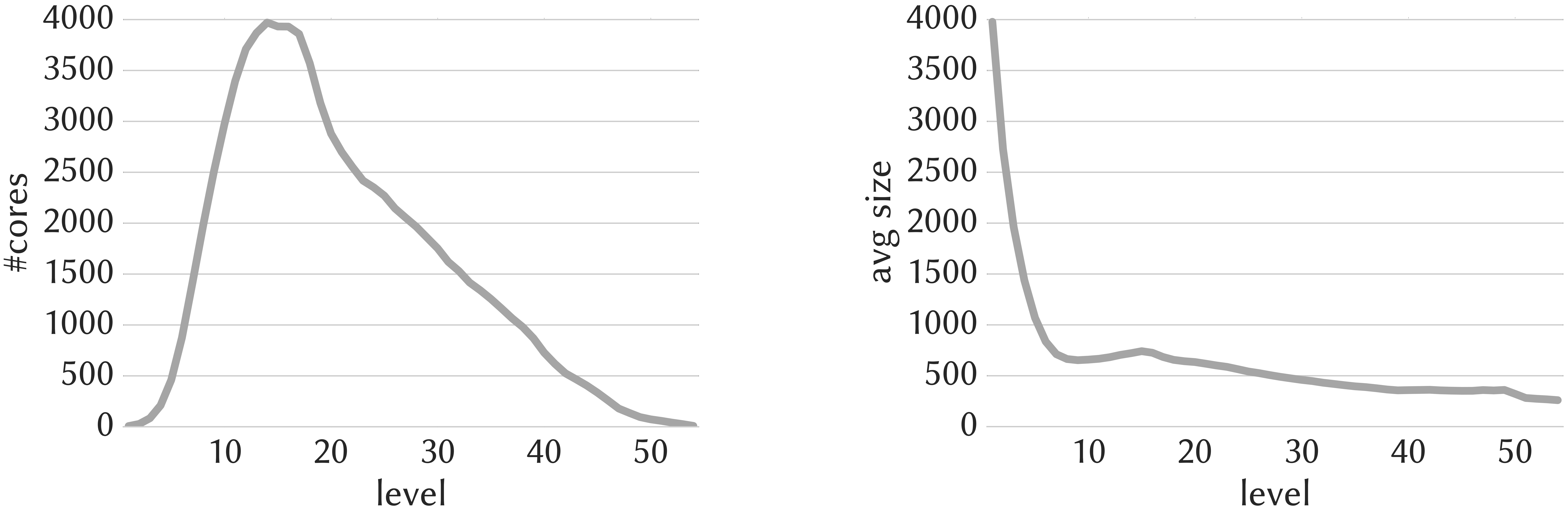
DATASETS

dataset	$\|V\|$	$\|E\|$	$\|L\|$	domain
Homo	18k	153k	7	genetic
SacchCere	6.5k	247k	7	genetic
DBLP	513k	1.0	10	co-authorship
ObamaInIsrael	2.2M	3.8M	3	social
Amazon	410k	8.1M	4	co-purchasing
FriendfeedTwitter	155k	13M	2	social
Higgs	456k	13M	4	social
Friendfeed	510k	18M	3	social

EFFICIENCY

dataset	#output cores	method	time (s)	#computed cores
SacchCere	74,426	N	19,282	278,402
		BFS	802	89,883
		DFS	2,117	223,643
		H	819	83,978
DBLP	3,346	N	104,361	34,572
		BFS	66	6,184
		DFS	219	38,887
		H	26	5,037
Amazon	1,164	BFS	2,349	1,354
		DFS	3,809	2,459
		H	2,464	1,334
Friendfeed	365,666	BFS	45,568	546,631
		DFS	12,211	568,107
		H	37,495	389,323

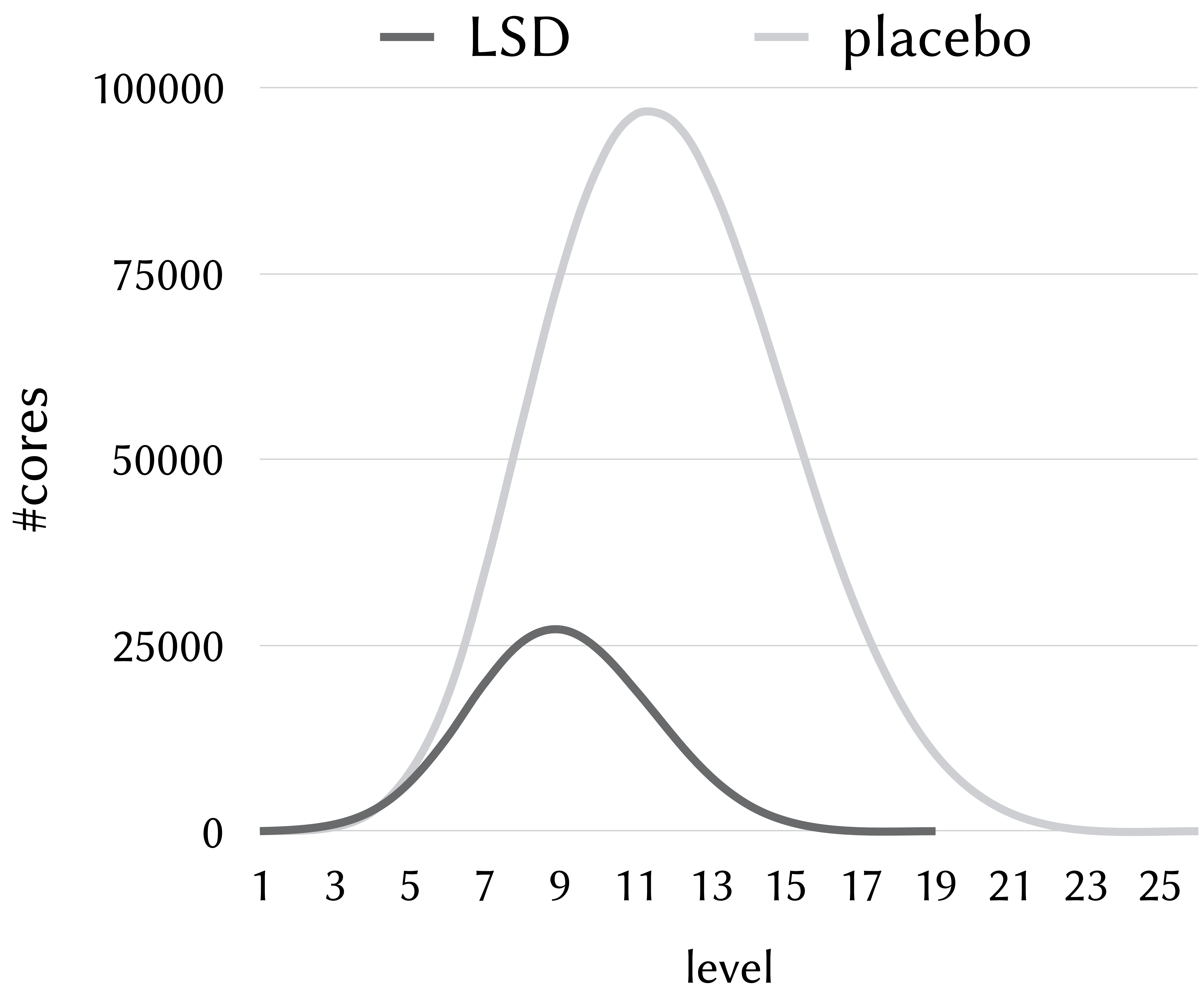
RESULTS



SacchCere

CASE STUDY: BRAIN

- Dataset to study the effect of LSD on the human brain:
  - 3 neuroimaging techniques
  - 15 individuals
  - 2 states
- 6 multilayer networks:
  - 165 vertices
  - 15 layers



MULTILAYER DENSEST SUBGRAPH

MULTILAYER DENSEST SUBGRAPH

Given a multilayer graph G=(V,E,L), a positive real number β, and a real-valued function

$$\delta(S) = \max_{\hat{L} \subseteq L} \min_{l \in \hat{L}} \frac{|E_l[S]|}{|S|} |\hat{L}|^{\beta}$$

find a subset S* of V that maximizes function δ.

- β controls the importance of the two ingredients of the objective function δ
- Solving the problem allows for automatically finding a set of layers of interest for the densest subgraph S*

APPROXIMATION ALGORITHM

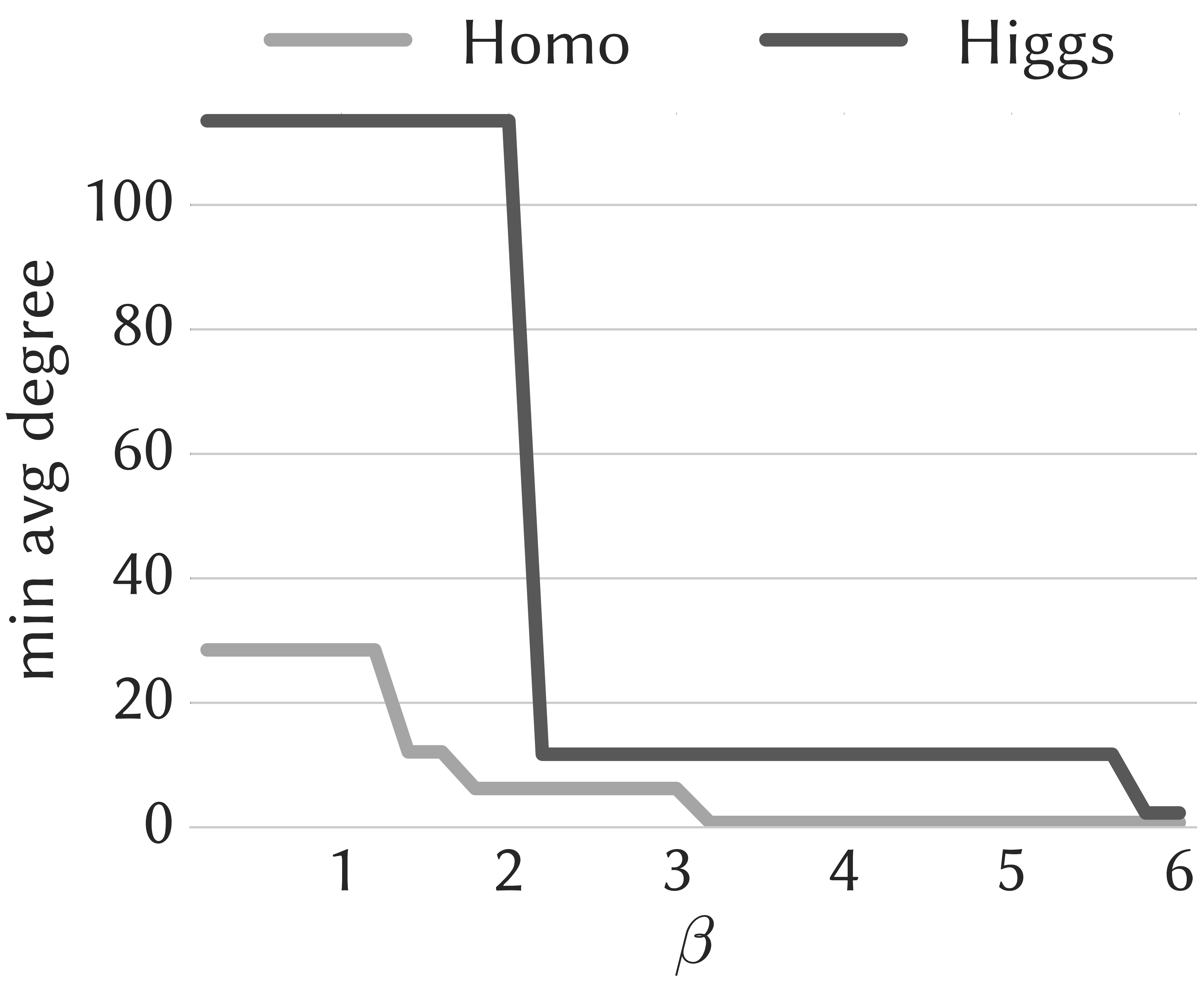
- Compute the multilayer core decomposition of the input graph
- Among all cores, take the one maximizing the objective function δ as the output densest subgraph

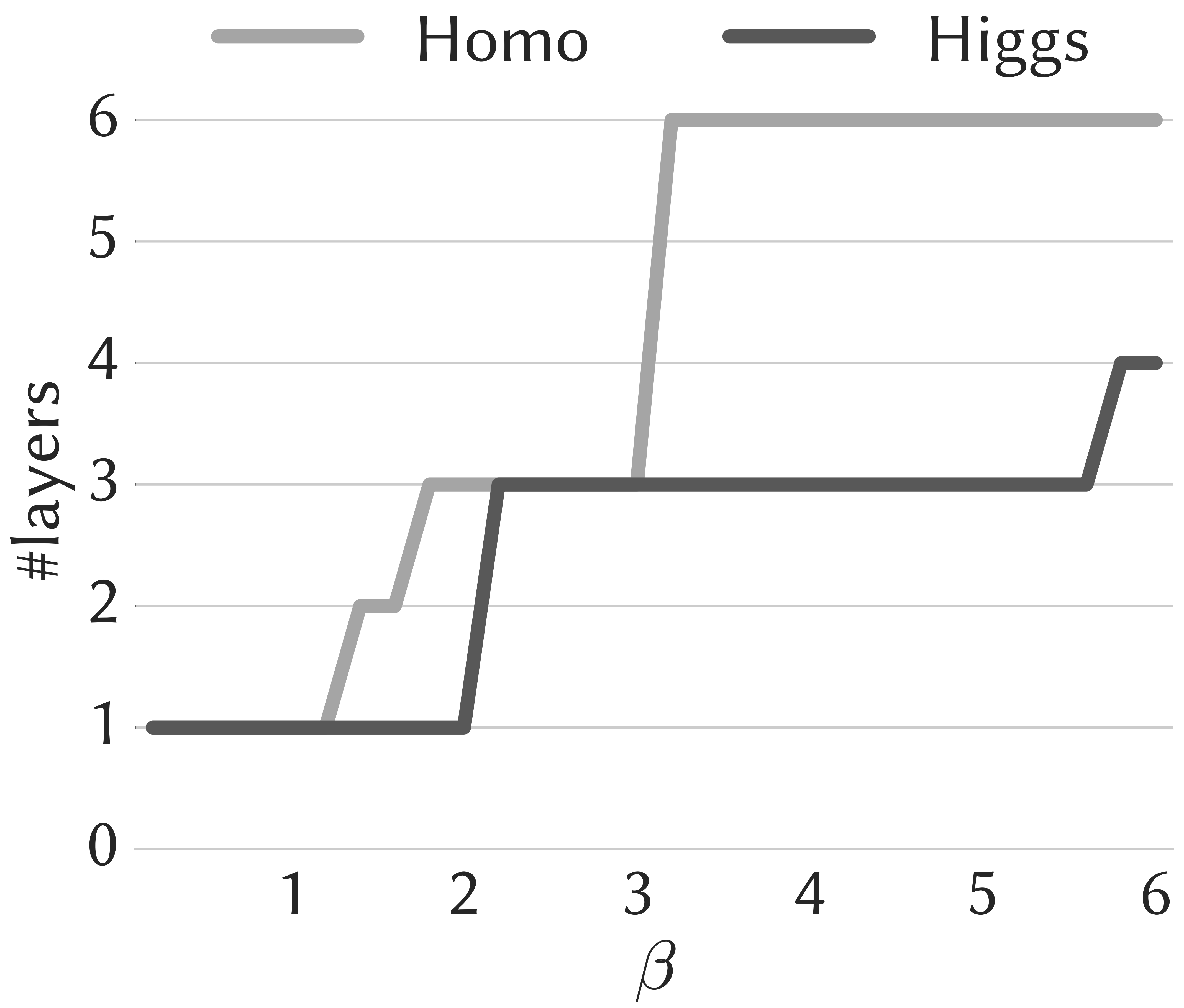
Let C* denote the core maximizing the density function δ, then

$$\delta(C^*) \geq \frac{1}{2|L|^{\beta}} \delta(S^*),$$

i.e., the algorithm achieves **$1/2|L|^{\beta}$ approximation guarantees**.

RESULTS


Homo
Higgs
min avg degree
100
80
60
40
20
0
1
2
3
4
5
6
β


Homo
Higgs
#layers
6
5
4
3
2
1
0
1
2
3
4
5
6
β

E. GALIMBERTI, F. BONCHI, F. GULLO

CORE DECOMPOSITION AND DENSEST SUBGRAPH IN MULTILAYER NETWORKS

ANECDOTAL EVIDENCE: DBLP