

A DYNAMIC PROGRAMMING ALGORITHM FOR A MAXIMUM s -CLIQUE SET ON TREES

*16th International Conference on Reachability Problems
(RP'22)*

18-October-2022.

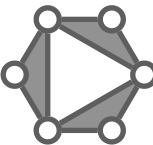
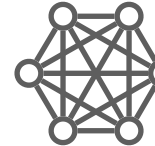



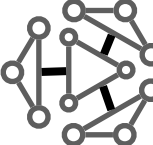

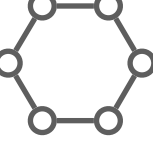


ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

José Alberto Fernández-Zepeda, **Alejandro Flores Lamas**, Matthew Hague, Joel Antonio Trejo-Sánchez

AGENDA



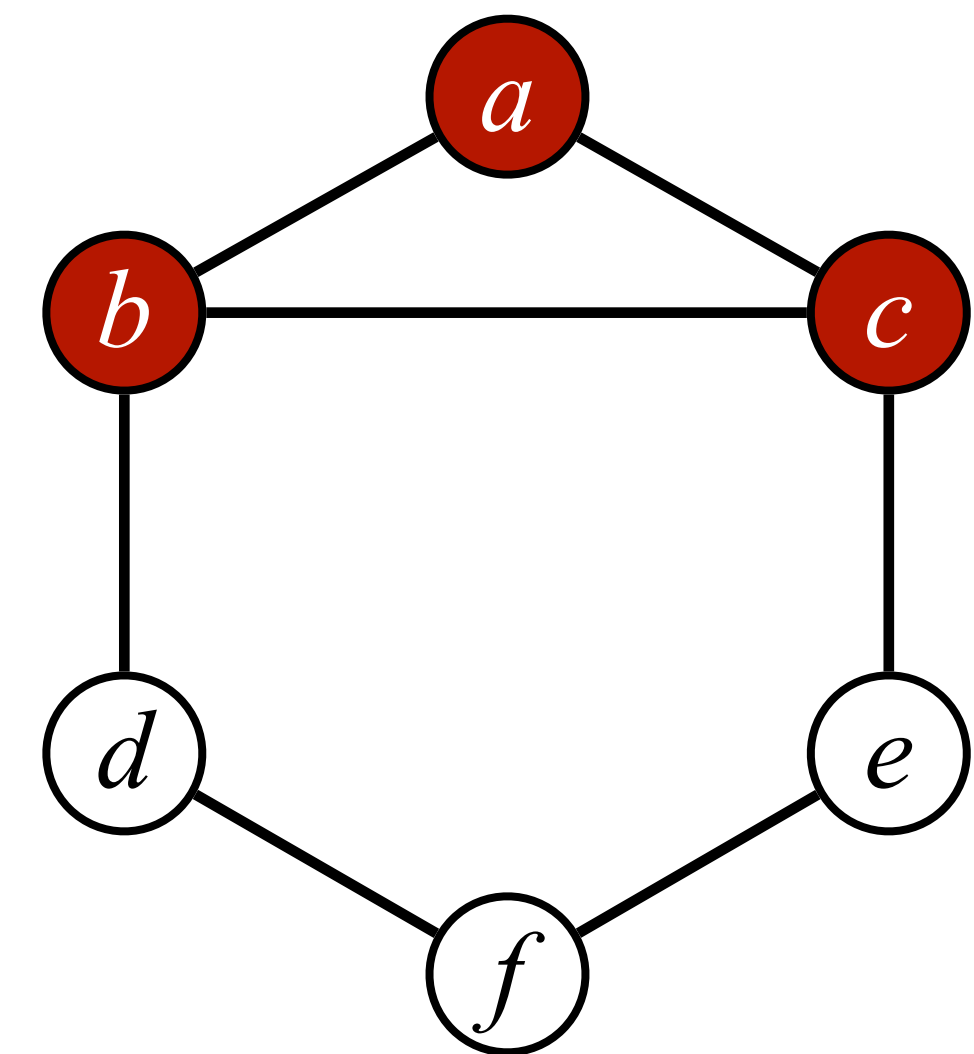
-  INTRODUCTION
 -  CLIQUE
 -  CLIQUE RELAXATIONS
 -  LANDSCAPE AND TAXONOMIC FRAMEWORK
-  SOLVING THE MAXIMUM s -CLIQUE PROBLEM ON TREES
-  OTHER METHODOLOGIES FOR THE s -CLIQUE PROBLEM
-  EXPERIMENTS
-  FINAL NOTES

CLIQUE



Given a graph $G = (V_G, E_G)$ a **clique** is a subset of vertices $S \subseteq V_G$ such that

- Each pair of distinct vertices $u, v \in S$ are adjacent.
- In other words, a clique $G' = (S, E'_G)$ is a complete subgraph of G .
- The size of a clique is the number of vertices it contains.



 \in set

$dist(\text{red circle}, \text{red circle}) = 1$

DURING THE 50s



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

PSYCHOMETRIKA—VOL. 15, NO. 2
JUNE, 1950

Journal of Mathematical Sociology
1973 Vol. 3, pp 113-126

© Gordon and Breach Science Publisher:
Printed in Birkenhead, England

CONNECTIVITY AND GENERALIZED CLIQUES IN SOCIOMETRIC GROUP STRUCTURE

R. DUNCAN LUCE

GRADUATE STUDENT, DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

By using the concepts of antimetry and n -chain it is possible to define and to investigate some properties of connectivity in a sociometric group. It is shown that the number of elements in a group, the number of antimetries, and the degree of connectivity must satisfy certain inequalities. Using the ideas of connectivity, a generalized concept of clique, called an n -clique, is introduced. n -cliques are shown to have a very close relationship to the existence of cliques in an artificial structure defined on the same set of elements, thus permitting the determination of n -cliques by means of the same simple matrix procedures used to obtain the clique structures. The presence of two or more m -cliques, where m is the number of elements in the group, is proved to mean an almost complete splitting of the group.

1. Introduction

This paper is devoted to extending the theoretical and practical mathematical results presented in an earlier paper (3). In that paper it was shown that certain elementary matrix operations permit, to some extent, an analysis of the simple type of psychological group structure which is often expressed by a sociometric diagram. Specifically, we envisage a finite set of m (≥ 2) elements i, j, k, \dots having a structure defined on them as follows: For any two elements i and j of the set there either exists or does not exist some *one type* of “directional” relationship *from i to j* . This one type of relationship on the group may assume various forms such as communication or friendship; and so to acquire generality in the discussion we shall speak of an *antimetry* from i to j . Then communication is merely a special type of antimetry. It is arbitrarily assumed that no antimetry can exist from an element i to itself. By “directional” is meant that knowing of the existence or non-existence of an antimetry from i to j does not give any information about the existence or non-existence of an antimetry from j to i . For example, if the form of the antimetry is “ i communicates to j ,” then the specific knowledge that a can communicate to b does not tell us, in general, whether b can communicate to a .

169

A GRAPH-THEORETIC DEFINITION OF A SOCIOMETRIC CLIQUE *

RICHARD D. ALBA

Columbia University

The intent of this paper is to provide a definition of a sociometric clique in the language of graph theory. This problem is viewed from two perspectives: maintaining fidelity to the intuitive notion of a clique; and providing adequate computational mechanics for large bodies of data. Luce's (1950) concept of an K -clique is used, but further qualifications are added. Two statistics or measures with associated probability distributions are defined for testing the adequacy of a subgraph which qualifies according to the definition.

1. INTRODUCTION

One of the factors which has most confused the discussion of sociometric clique identification in large bodies of data is the absence of a formal definition of a clique. Luce and Perry (1949) and Luce (1950) are the only well-known sources for such a formalization, and their attempts have not been followed out in the literature. Rather, more recent attempts to satisfactorily identify cliques have employed *ad hoc* clustering or taxonomic procedures which allow the data to suggest natural groupings. Notable among these are the approaches of MacRae (1960), Coleman and MacRae (1960), and Hubbel (1965).

One difficulty in the approach of Luce and Perry (1949) and Luce (1950) was that an adequate computational procedure to locate subsets of the data which satisfied their criteria was lacking. Recent computational literature, such as Bonner (1963) and especially Auguston and Minker (1970), contains algorithms which can be used to identify these subsets in reasonably large bodies of data (whose sizes are on the order of 100 to 1000 individuals).

Other difficulties, however, remain. The definition in Luce and Perry (1949), in which a clique is defined as a maximal complete subgraph, is too stringent for most purposes. The concept of n -clique, as presented in Luce (1950), may provide a suitable basis for a formalization but its properties must be explored further before any judgment can be made.

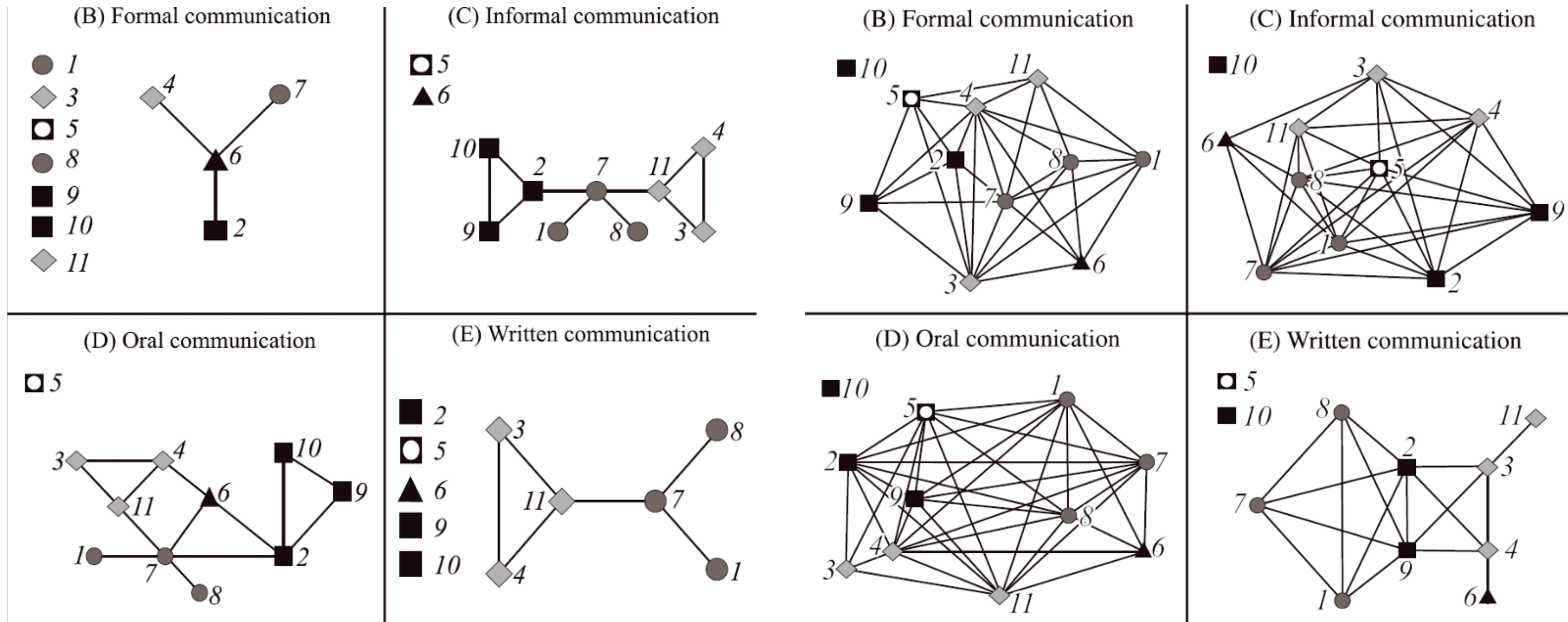
The intent of this paper is to provide a suitable definition of a sociometric clique.

*A version of this paper was read at the 1972 meetings of the American Sociological Association in New Orleans. This paper was prepared under National Science Foundation Grant GS 3231 to Professor Charles Kadushin of Teachers College, Columbia University. I am grateful for the comments of Professor Kadushin, as well as those of Dr. Kenneth Land of the Russell Sage Foundation and Professor Seymour Spileman of the University of Wisconsin to a draft of this paper. I particularly want to thank Dr. Tom Louis of Columbia University for his help in formulating the applications of probability theory in this paper.

113

- Luce, R.D., 1950. Connectivity and generalized cliques in sociometric group structure. Psychometrika, 15(2), pp.169-190.
- Alba, R.D., 1973. A graph-theoretic definition of a sociometric clique. Journal of Mathematical Sociology, 3(1), pp.113-126.

APPLICATION EXAMPLE



CLIQUE RELAXATIONS



CONNECTIVITY AND GENERALIZED CLIQUES IN SOCIOMETRIC GROUP STRUCTURE

R. DUNCAN LUCE
GRADUATE STUDENT, DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

By using the concepts of antimetry and n -chain it is possible to define and to investigate some properties of connectivity in a sociometric group. It is shown that the number of elements in a group, the number of antimetries, and the degree of connectivity must satisfy certain inequalities. Using the idea of n -chains, a generalized concept of clique, called an n -clique, is shown to have a very close relationship to the concept of cliques in an artificial structure defined by elements, thus permitting the determination of the same simple matrix procedures used to determine the presence of two or more m -cliques. The presence of two or more m -cliques in a group, is proved to be equivalent to the splitting of the group.

1. Introduction

This paper is devoted to extending the mathematical results presented in an earlier paper it was shown that certain elementary properties, to some extent, an analysis of the simple structure which is often expressed by a sociometric diagram. Specifically, we envisage a finite set of m (≥ 2) elements i, j, k, \dots having a structure defined on them as follows: For any two elements i and j of the set there either exists or does not exist some *one type* of “directional” relationship *from i to j* . This one type of relationship on the group may assume various forms such as communication or friendship; and so to acquire generality in the discussion we shall speak of an *antimetry* from i to j . Then communication is merely a special type of antimetry. It is arbitrarily assumed that no antimetry can exist from an element i to itself. By “directional” is meant that knowing of the existence or non-existence of an antimetry from i to j does not give any information about the existence or non-existence of an antimetry from j to i . For example, if the form of the antimetry is “ i communicates to j ,” then the specific knowledge that a can communicate to b does not tell us, in general, whether b can communicate to a .

A GRAPH-THEORETIC DEFINITION OF A SOCIOMETRIC CLIQUE *

RICHARD D. ALBA

Columbia University

The intent of this paper is to provide a definition of a sociometric clique in the language of graph theory. This problem is viewed from two perspectives: maintaining fidelity to the intuitive notion of a clique; and providing adequate computational mechanics for large bodies of data. Luce's (1950) concept of an K -clique is used, but further qualifications are added. Two statistics or measures with associated probability distributions are defined for testing the adequacy of a subgraph which qualifies

INTRODUCTION

Most confused the discussion of sociometric clique data is the absence of a formal definition of a clique. Luce (1950) are the only well-known sources for such a definition have not been followed out in the literature. Rather, those who identify cliques have employed *ad hoc* clustering procedures to follow the data to suggest natural groupings. Notable examples are of MacRae (1960), Coleman and MacRae (1960),

that of Luce and Perry (1949) and Luce (1950) was that an adequate computational procedure to locate subsets of the data which satisfied their criteria was lacking. Recent computational literature, such as Bonner (1963) and especially Auguston and Minker (1970), contains algorithms which can be used to identify these subsets in reasonably large bodies of data (whose sizes are on the order of 100 to 1000 individuals).

Other difficulties, however, remain. The definition in Luce and Perry (1949), in which a clique is defined as a maximal complete subgraph, is too stringent for most purposes. The concept of n -clique, as presented in Luce (1950), may provide a suitable basis for a formalization but its properties must be explored further before any judgment can be made.

The intent of this paper is to provide a suitable definition of a sociometric clique.

*A version of this paper was read at the 1972 meetings of the American Sociological Association in New Orleans. This paper was prepared under National Science Foundation Grant GS 3231 to Professor Charles Kadushin of Teachers College, Columbia University. I am grateful for the comments of Professor Kadushin, as well as those of Dr. Kenneth Land of the Russell Sage Foundation and Professor Seymour Spilerman of the University of Wisconsin to a draft of this paper. I particularly want to thank Dr. Tom Louis of Columbia University for his help in formulating the applications of probability theory in this paper.

Clique problem
(maximum, maximal,
enumeration...)

- Luce, R.D., 1950. Connectivity and generalized cliques in sociometric group structure. Psychometrika, 15(2), pp.169-190.
- Alba, R.D., 1973. A graph-theoretic definition of a sociometric clique. Journal of Mathematical Sociology, 3(1), pp.113-126.

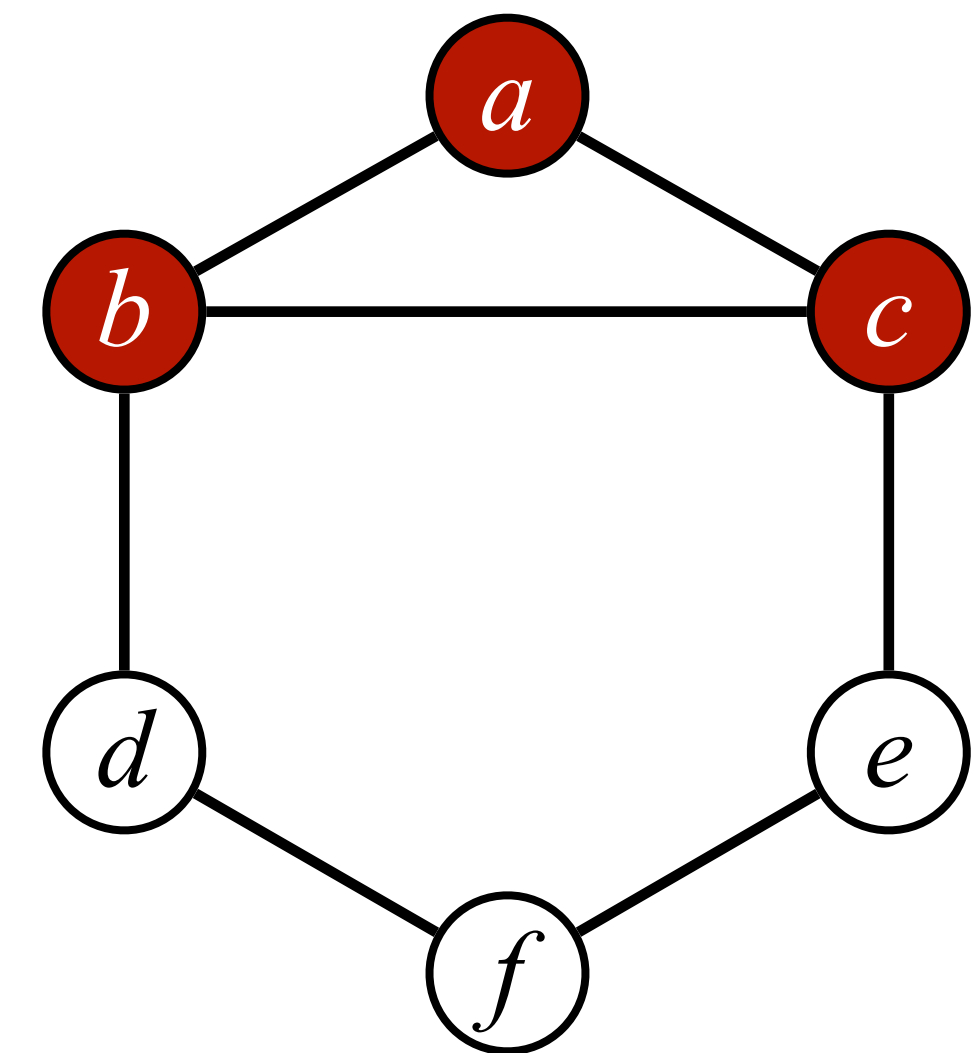
s -CLIQUE



Given a graph $G = (V_G, E_G)$, an **s -clique** is a subset of vertices $S \subseteq V_G$ such that:

- The distance between each pair of distinct vertices $u, v \in S$ is at most s edges long.
 - $\text{dist}(u, v) \leq s$

Clique as an 1-clique



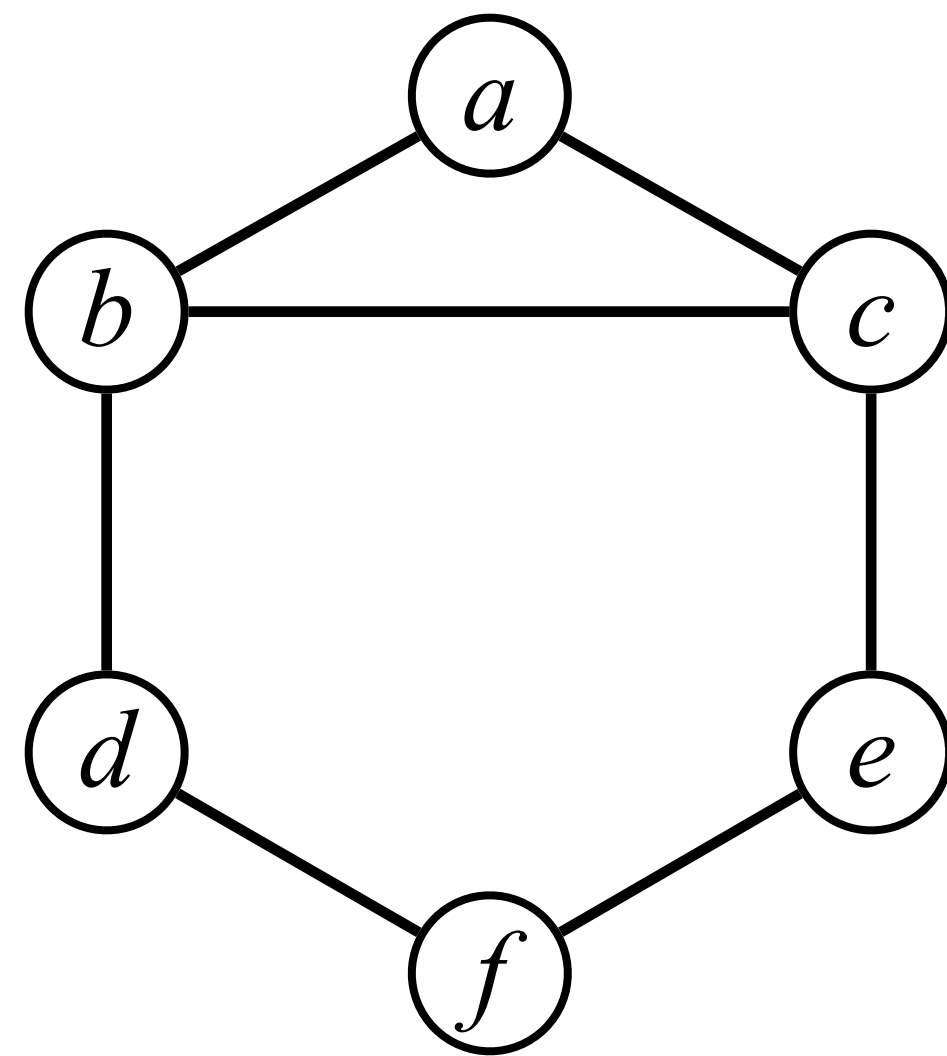
$$\text{dist}(u, v) \leq 1$$



2-CLIQUE

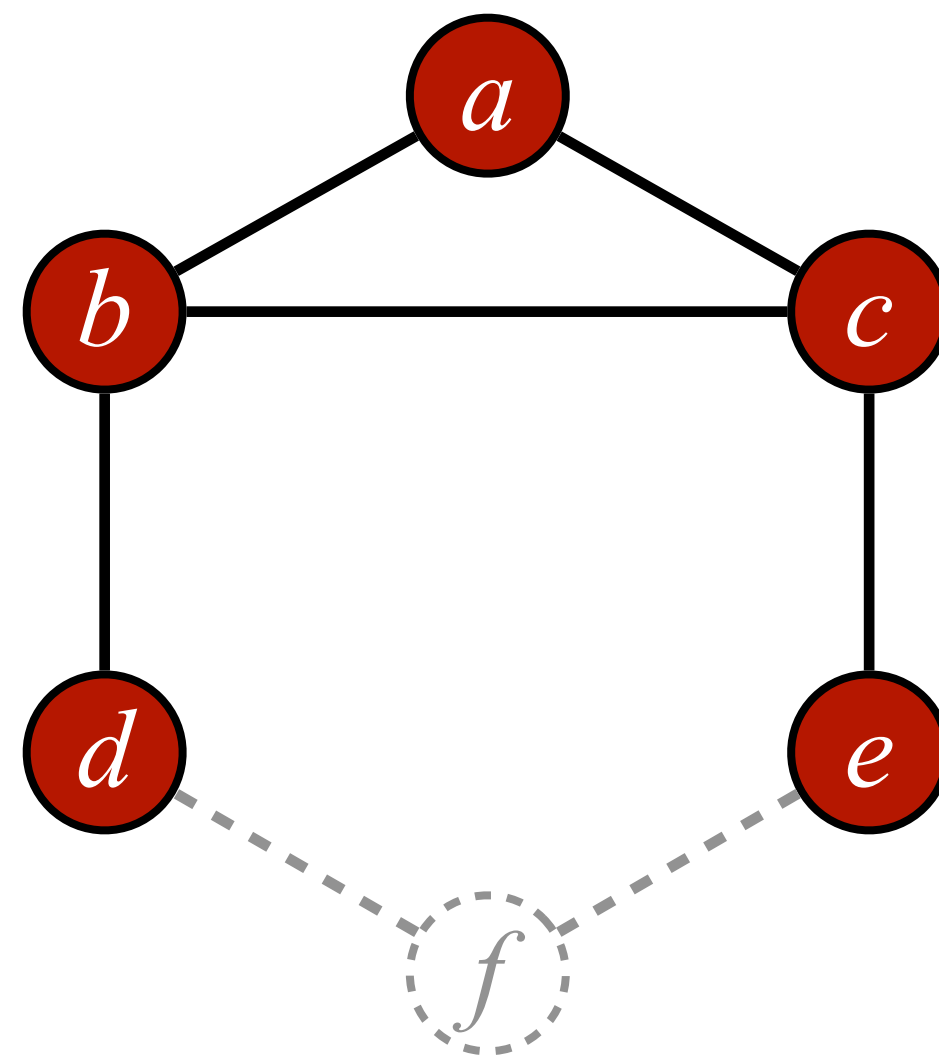


Input graph:
 $G = (V_G, E_G)$



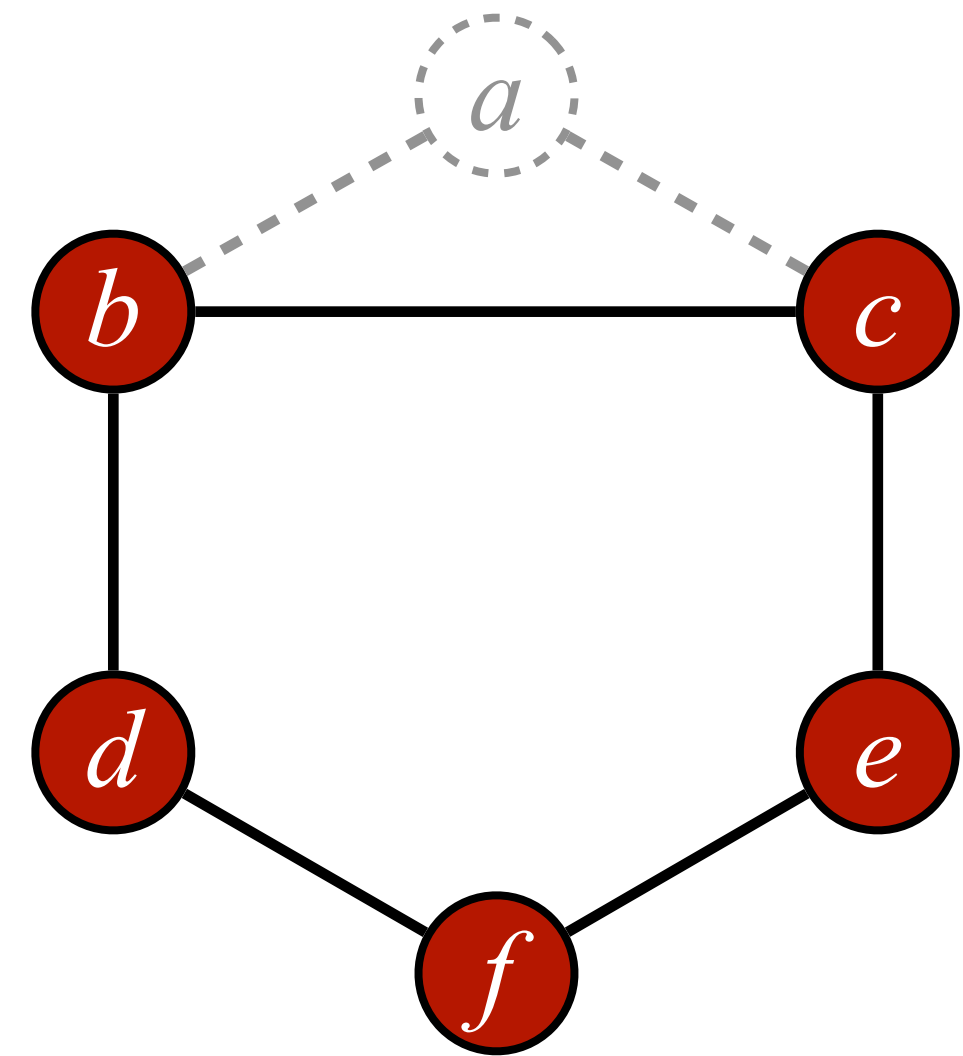
 \in set

$S^1 = \{a, b, c, d, e\}$

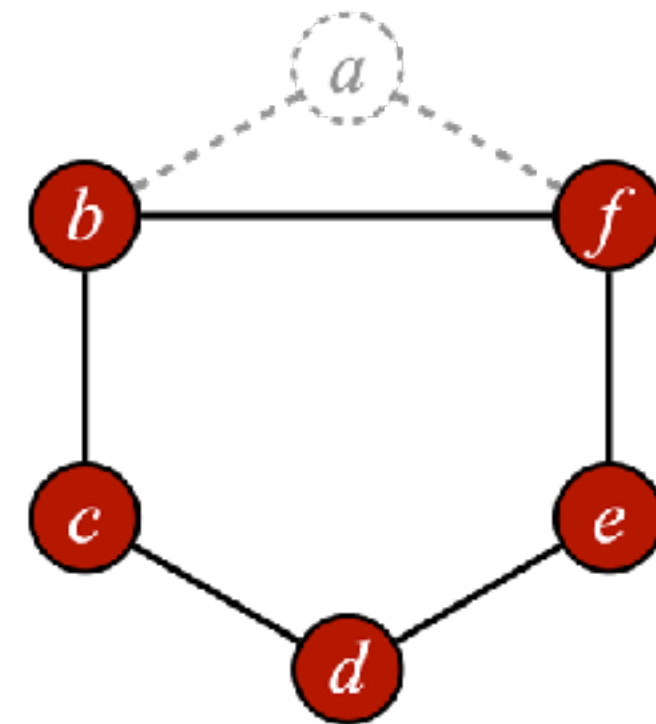
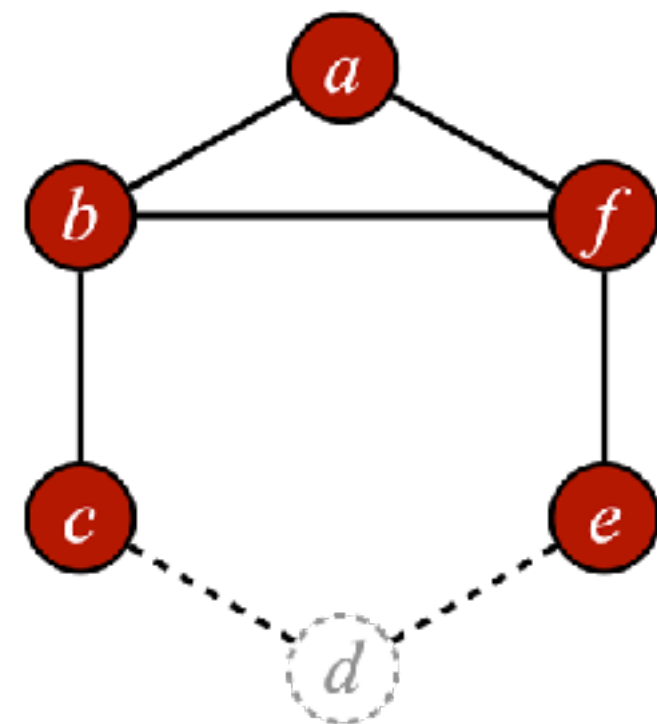


$\text{dist}(d, e)$ uses edges
 $\{(d, f), (f, e)\}$

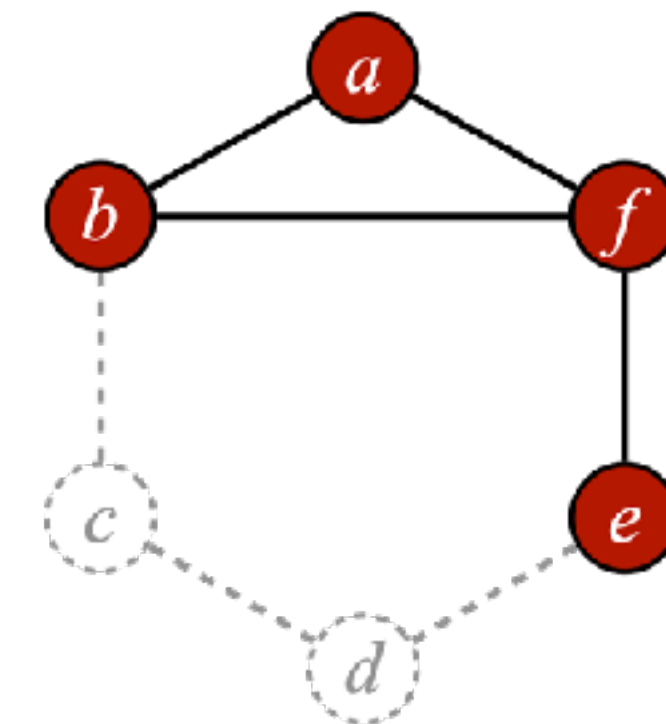
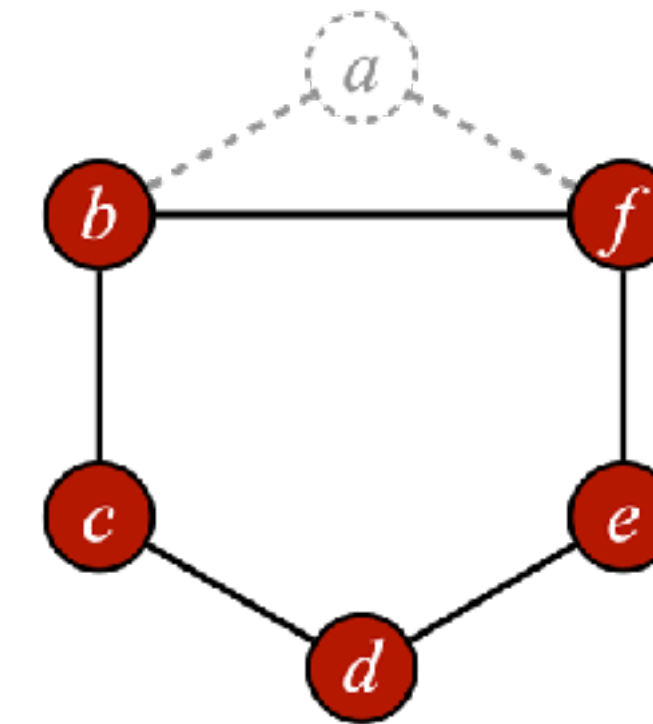
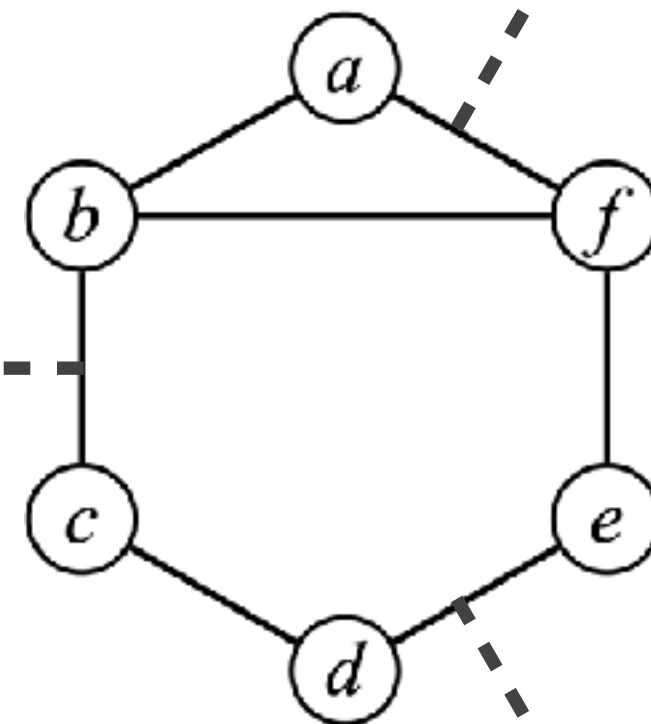
$S^2 = \{b, c, d, e, f\}$



CLIQUE, CLUBS, CLANS

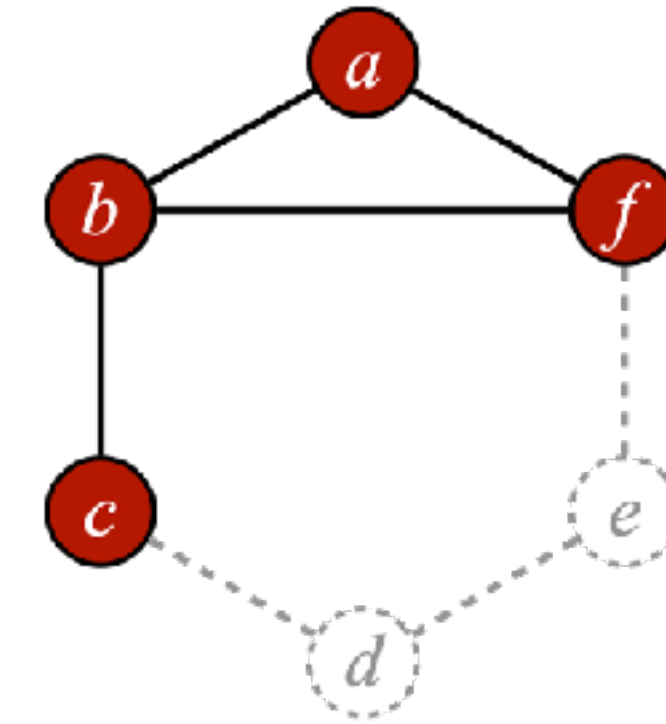
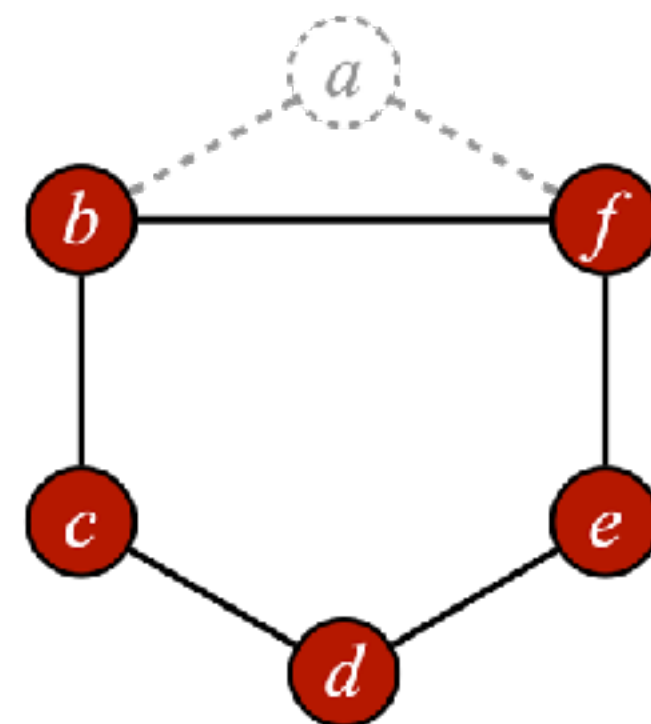


2-clique

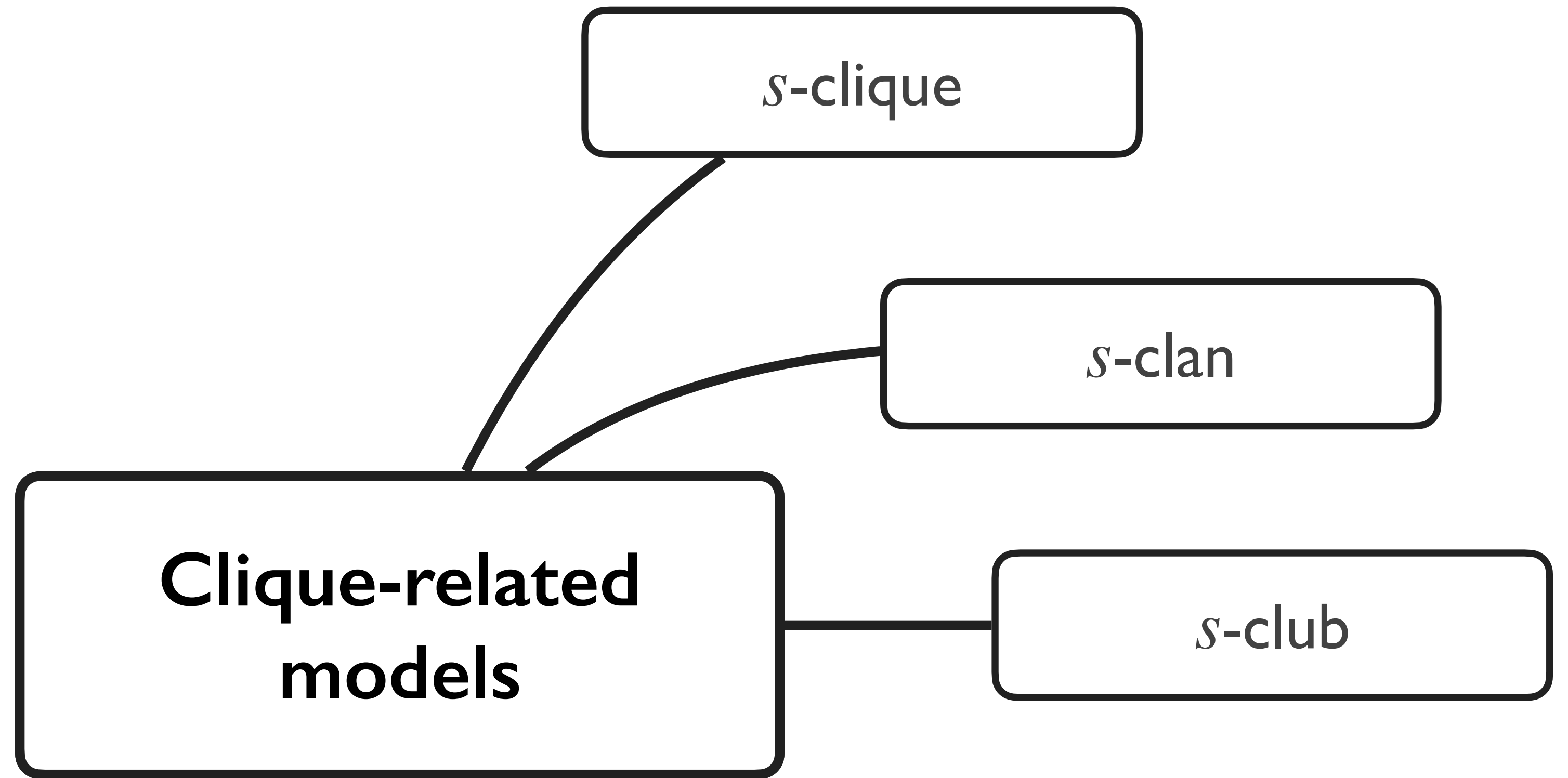


2-club

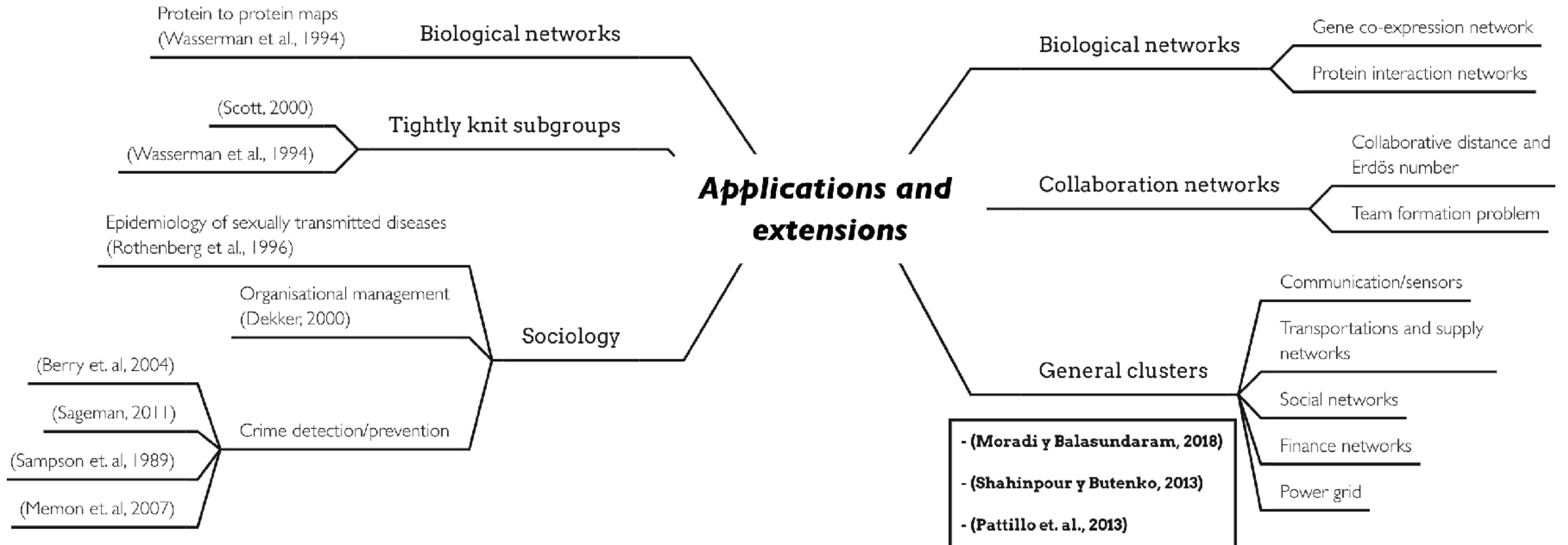
2-clan

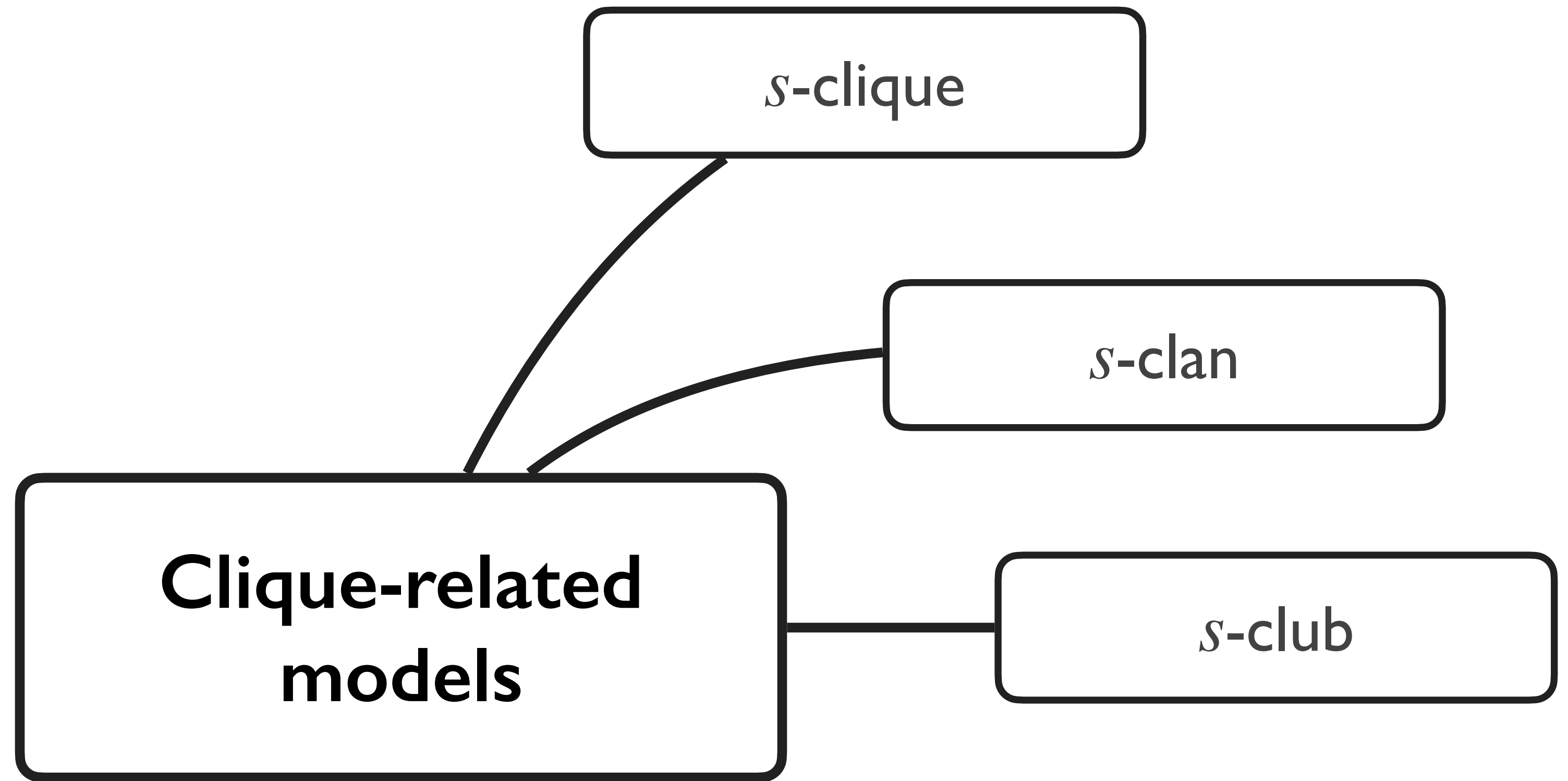


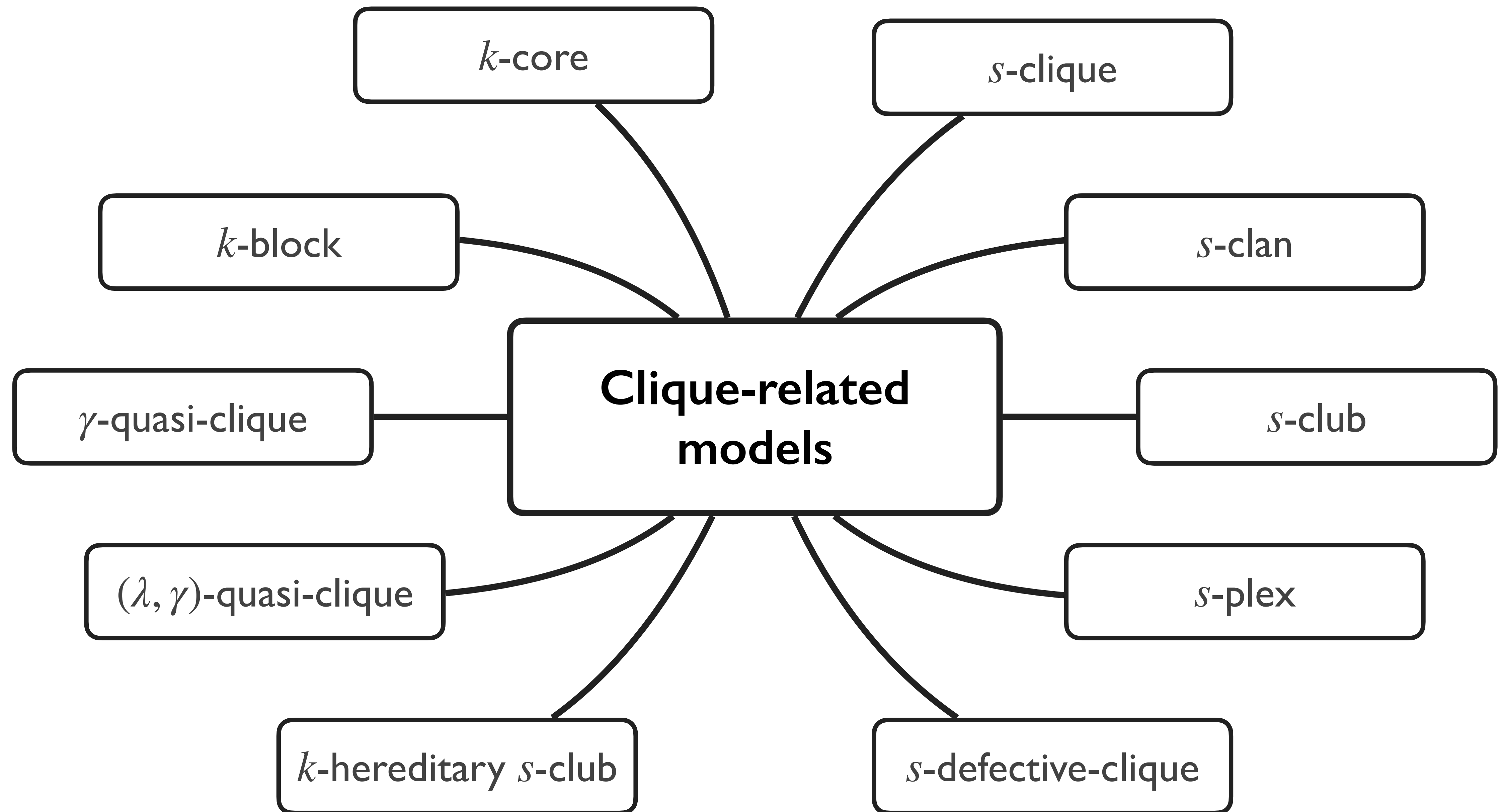
**Clique problem
(maximum, maximal,
enumeration...)**



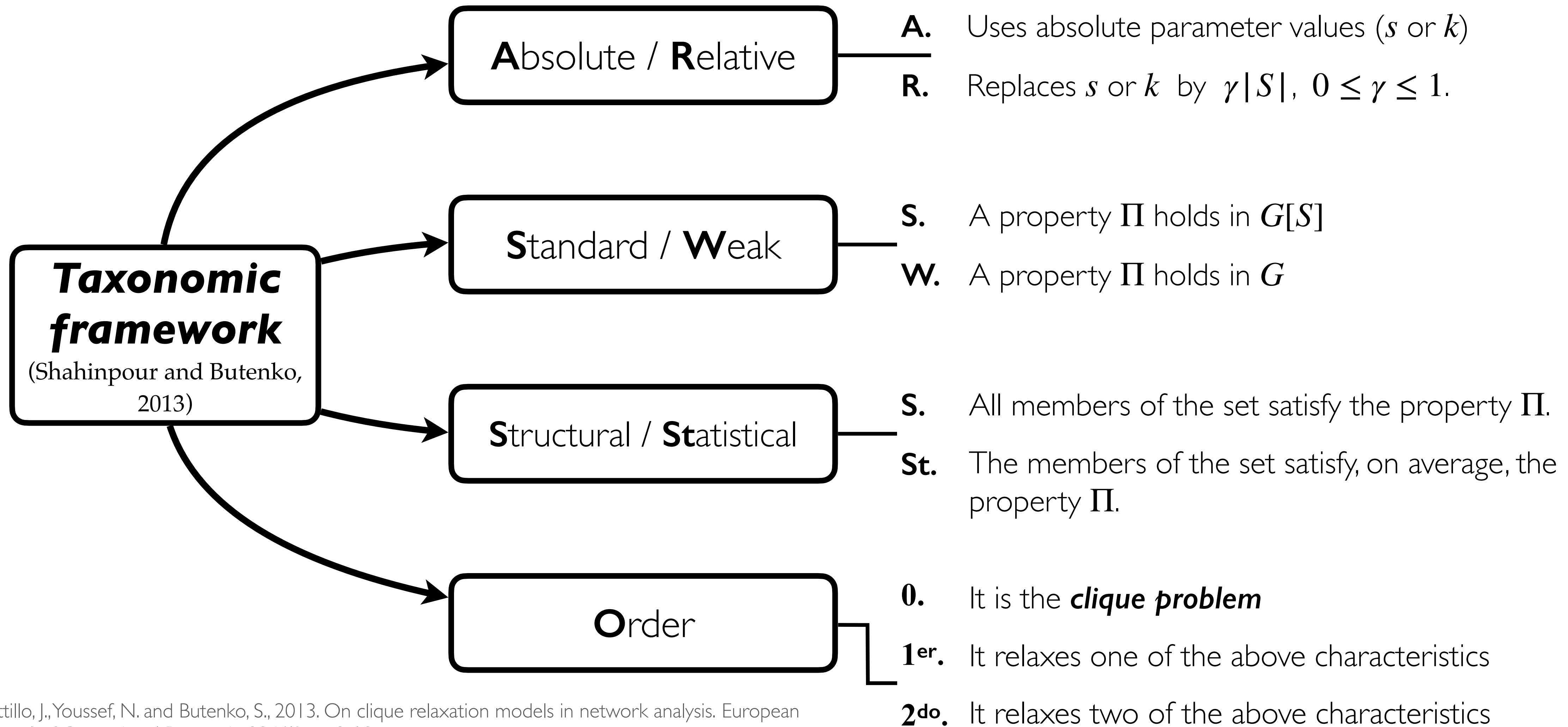
APPLICATIONS AND EXTENSIONS







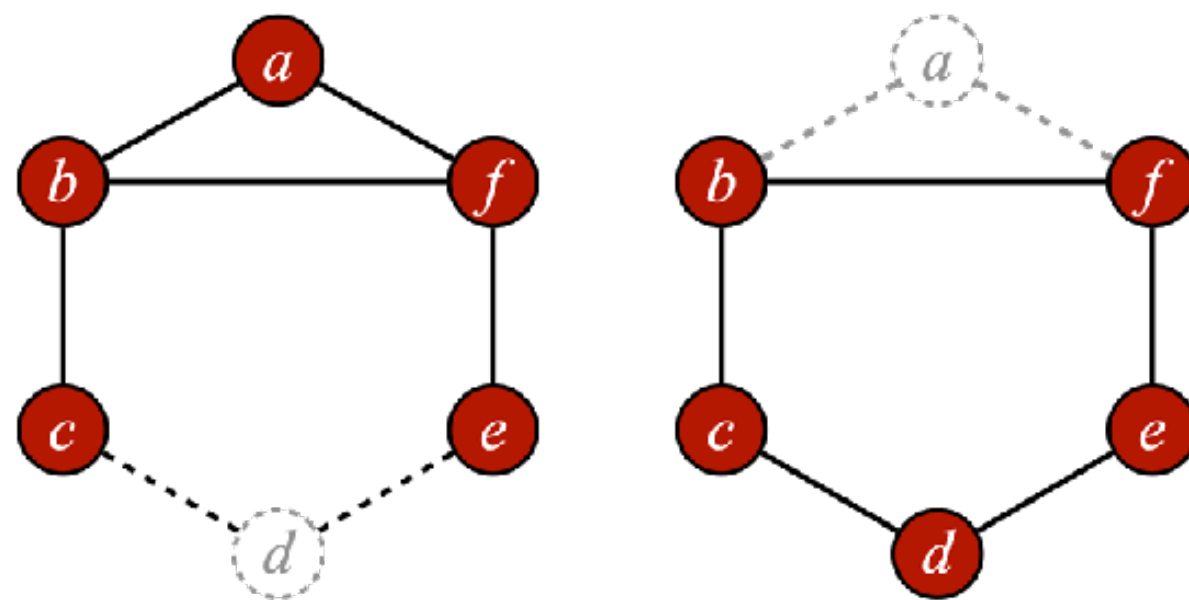
FRAMEWORK



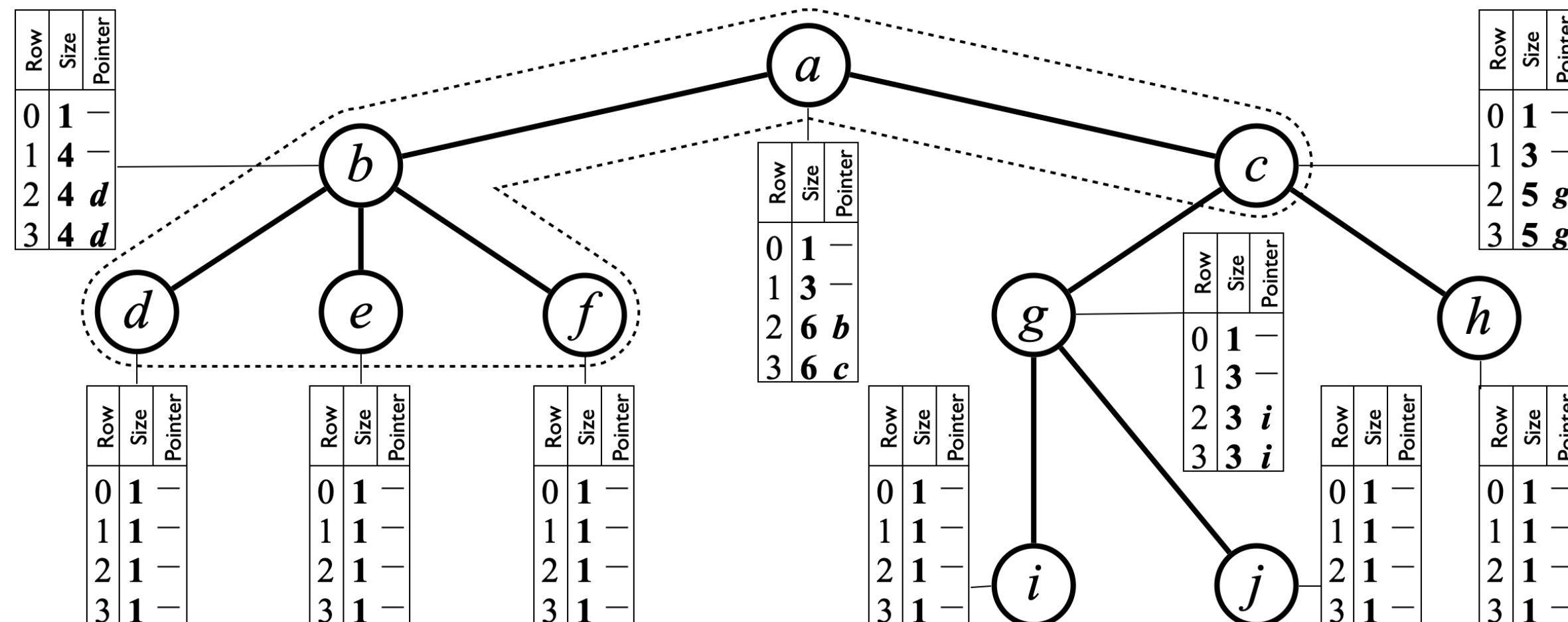
THE MAXIMUM s -CLIQUE PROBLEM ON TREES (THIS TALK)



s -clique problem



Dynamic programming



Taxonomic framework
(Shahinpour and Butenko, 2013)

Absolute / Relative

- A.** Uses absolute parameter values (s or k)
- R.** Replaces s or k by $\gamma|S|$, $0 \leq \gamma \leq 1$.

Standard / Weak

- S.** A property Π holds in $G[S]$
- W.** A property Π holds in G

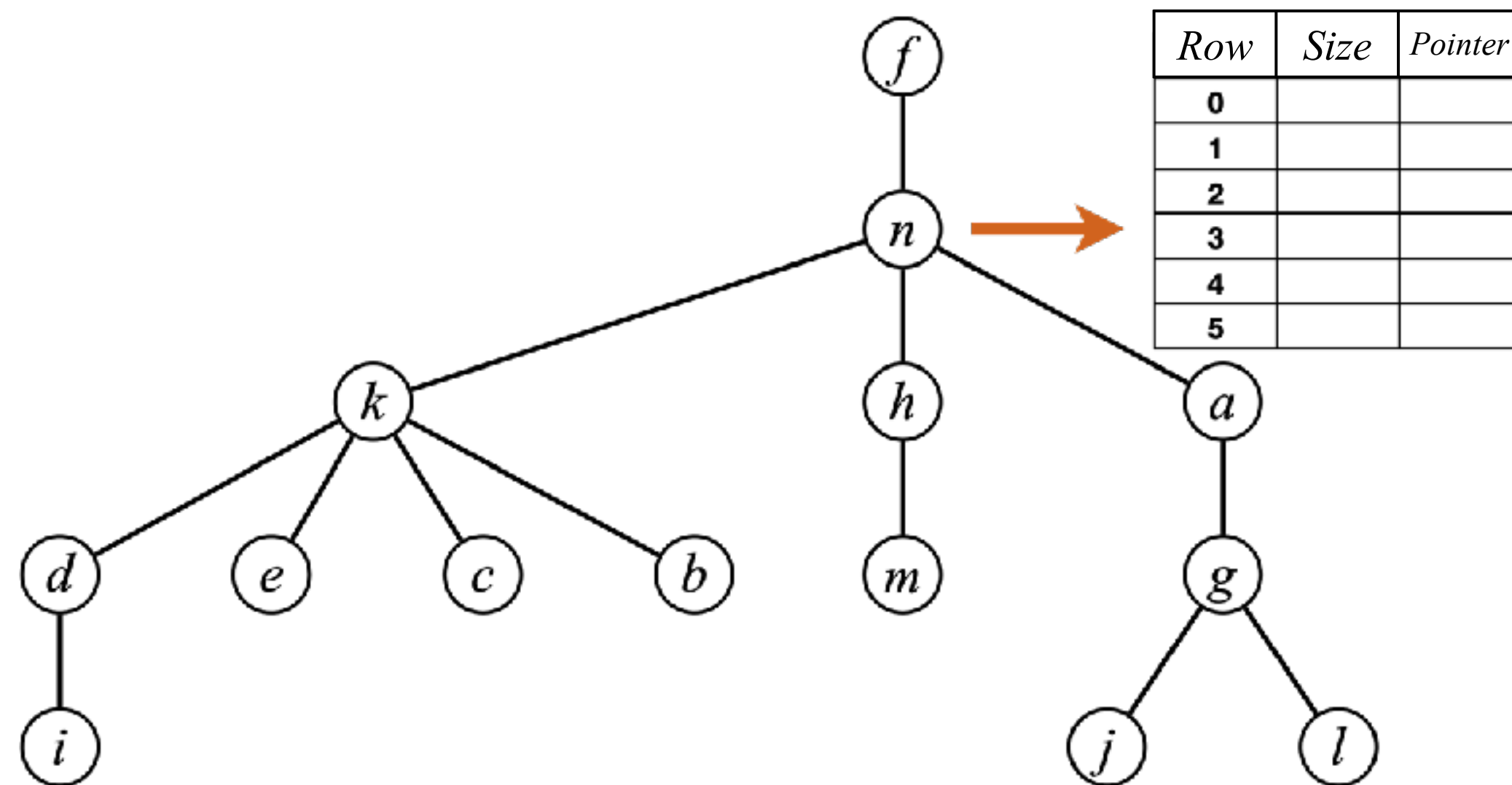
Structural / Statistical

- S.** All members of the set satisfy the property Π .
- St.** The members of the set satisfy on average the property Π .

Order

- 0.** This is the **clique problem**
- 1^{er}.** It relaxes one of the above characteristics
- 2^{do}.** It relaxes two of the above characteristics
- ...

STRATEGY



Let $T = (V_T, E_T)$ be a tree graph,
 $v \in V_T$ and s an integer.

For this example, consider $s = 5$

Each vertex v has an associated table with the following:

- $s + 1$ rows
- 2 columns: **Size** and **Pointer**

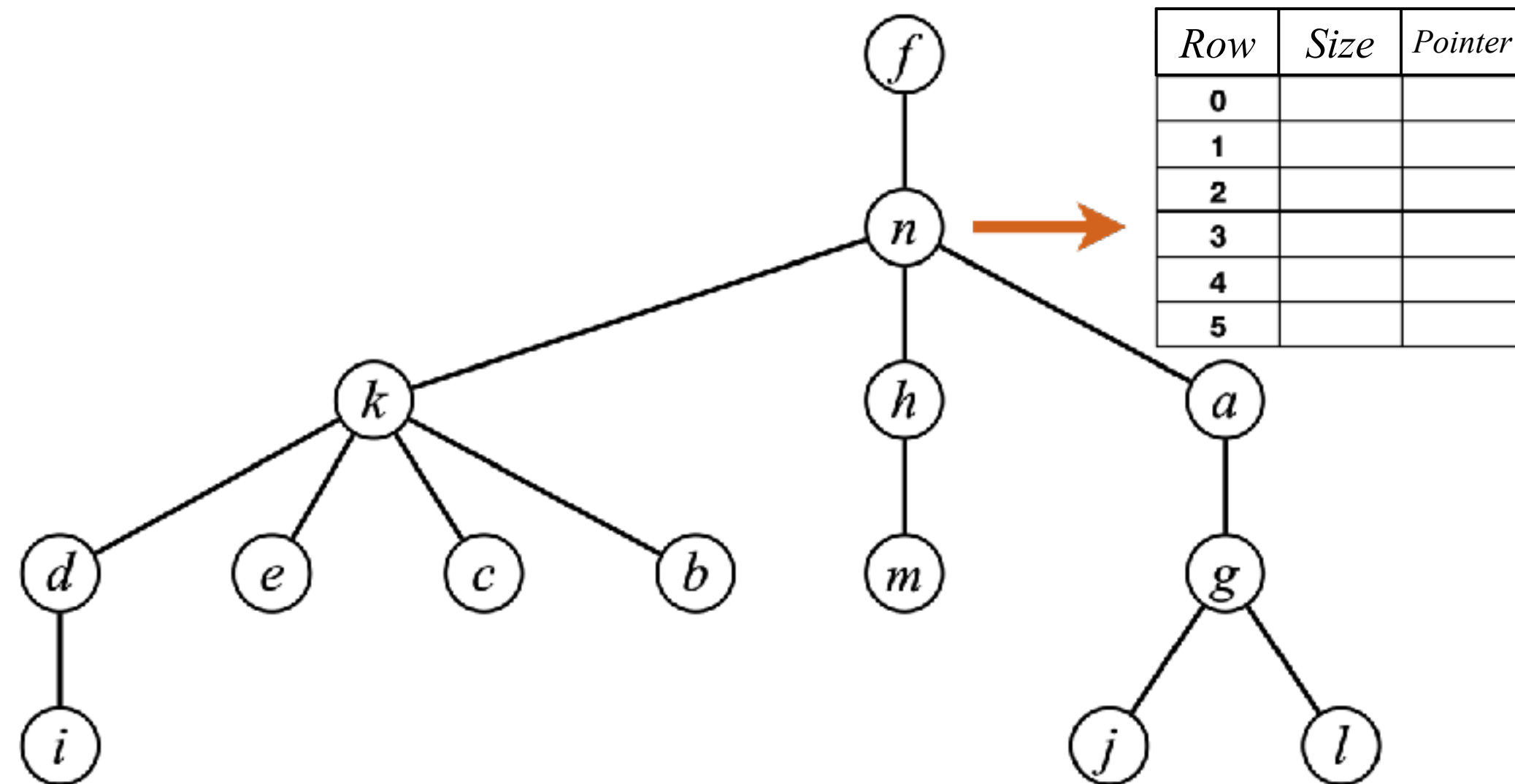
If v is a leaf vertex, trivial case:

$$t.table[row].size \leftarrow 1, \forall row \in [s]$$

Otherwise, there are **four cases**:

- $row = 0$
- $1 \leq row \leq \left\lfloor \frac{s}{2} \right\rfloor$
- $\left\lfloor \frac{s}{2} \right\rfloor < row \leq s - 1$
- $row = s$

STRATEGY



Let $T = (V_T, E_T)$ be a tree graph,
 $v \in V_T$ and s an integer.

For this example, consider $s = 5$

Each vertex v has an associated table with the following:

- $s + 1$ rows
- 2 columns: **Size** and **Pointer**

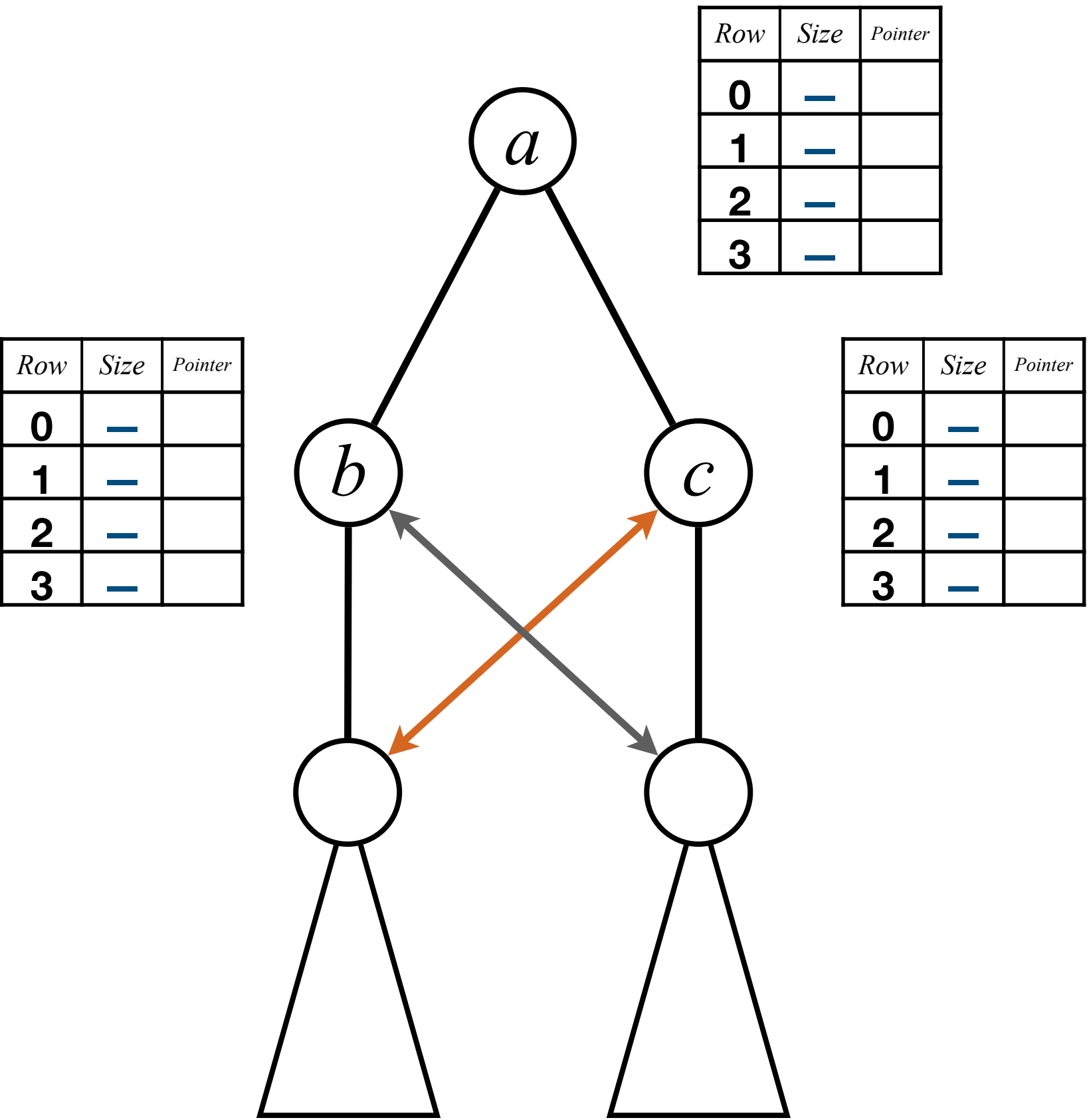
If v is a leaf vertex, trivial case:

$$t.table[row].size \leftarrow 1, \forall row \in [s]$$

Otherwise, there are **four cases**:

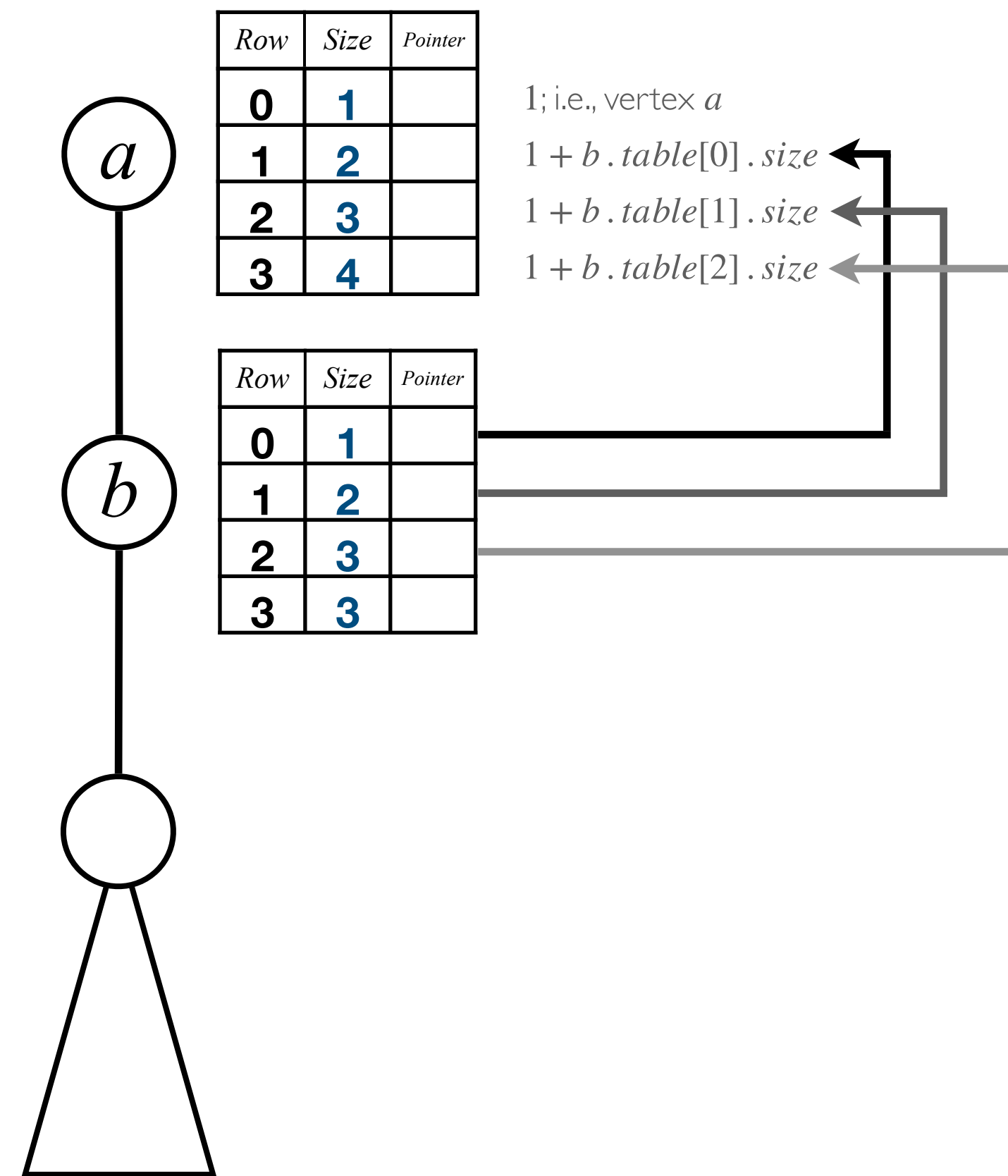
- $row = 0$
- $1 \leq row \leq \left\lfloor \frac{s}{2} \right\rfloor$
- $\left\lfloor \frac{s}{2} \right\rfloor < row \leq s - 1$
- $row = s$

STRATEGY



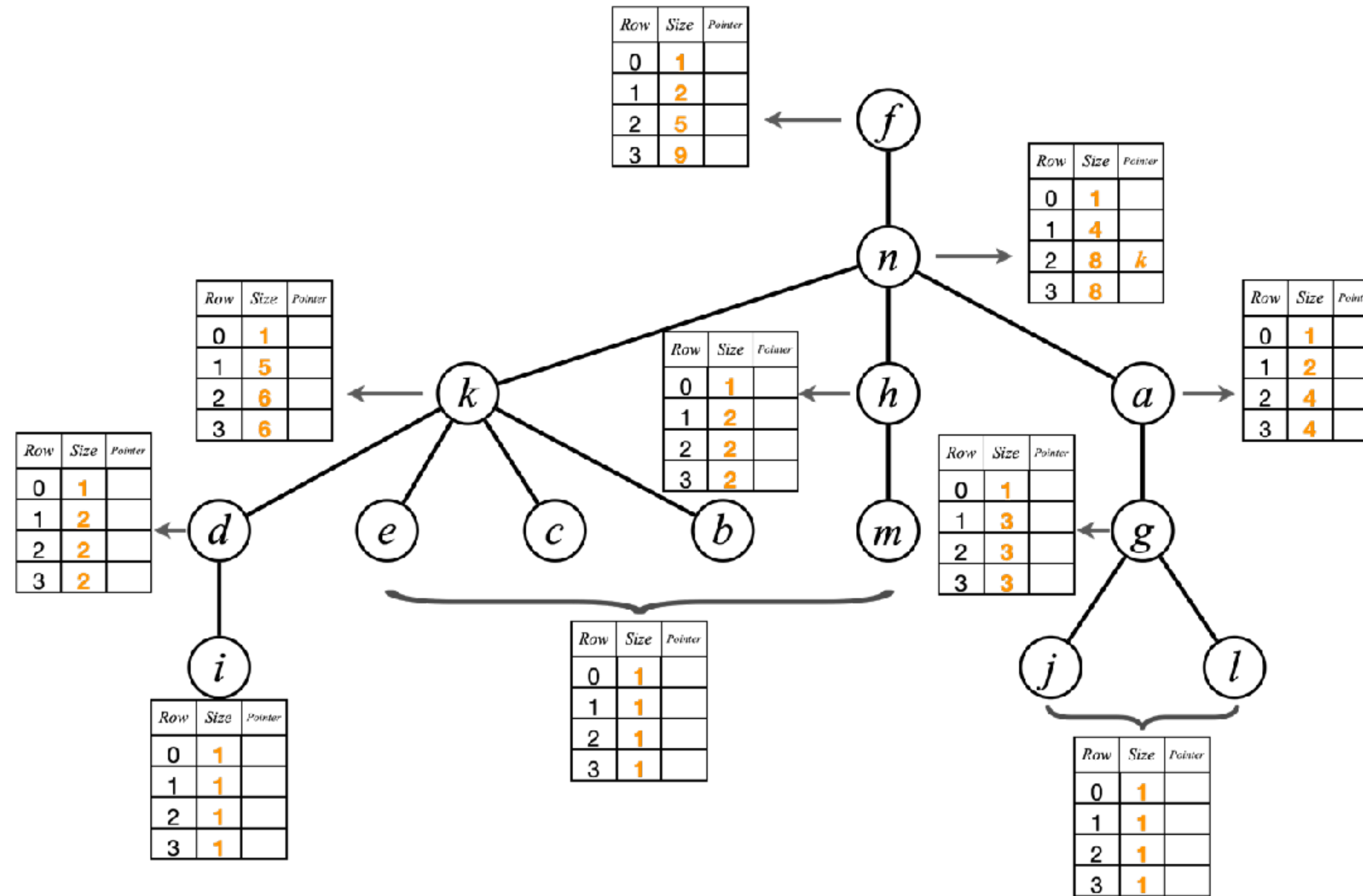
Case	Action
1) $row = 0$	$a.table[0].size \leftarrow 1$
2) $1 \leq row \leq \left\lfloor \frac{s}{2} \right\rfloor$	$a.table[row].size \leftarrow 1 + \sum_{w \in C_a} w.table[row-1].size$
3) $\left\lfloor \frac{s}{2} \right\rfloor < row \leq s-1$	$total \leftarrow \sum_{w \in C_a} w.table[s-row-1]$ $a.table[row].size \leftarrow 1 + \max_{w \in C_a} \left\{ total - (w.table[s-row-1].size + w.table[row-1].size) \right\}$
4) $row = s$	$a.table[s].size \leftarrow 1 + \max_{w \in C_a} \{ w.table[s-1].size \}$
$a.table[row].size \leftarrow \max(a.table[row].size, a.table[row-1].size), \text{ for } 1 \leq row \leq s$ Update $a.table[row].pointer \leftarrow a.table[row-1].pointer$, accordingly	

STRATEGY

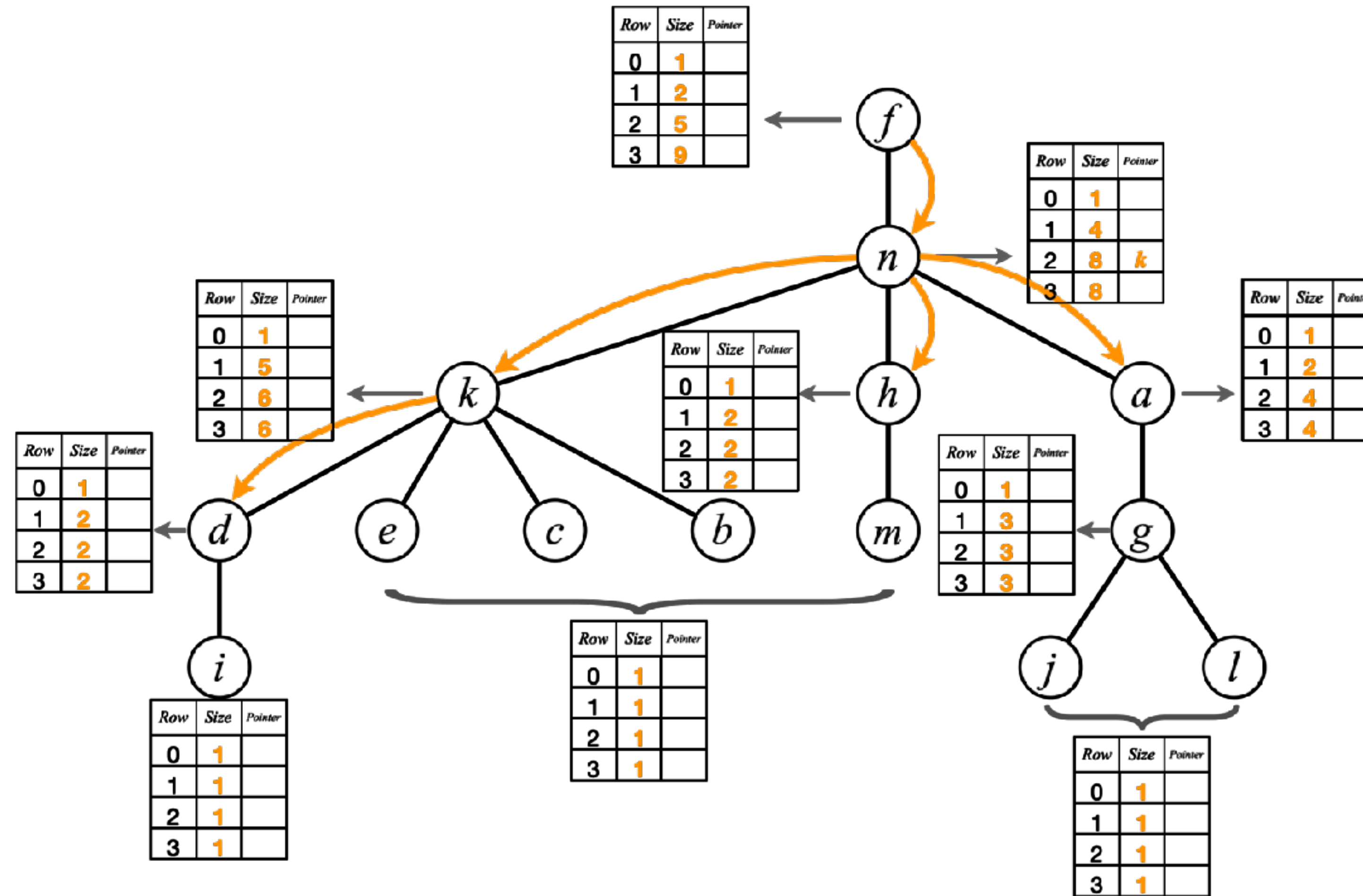


Case	Action
1) $row = 0$	$a.table[0].size \leftarrow 1$
2) $1 \leq row \leq \left\lfloor \frac{s}{2} \right\rfloor$	$a.table[row].size \leftarrow 1 + \sum_{w \in C_a} w.table[row-1].size$
3) $\left\lfloor \frac{s}{2} \right\rfloor < row \leq s-1$	$total \leftarrow \sum_{w \in C_a} w.table[s-row-1]$ $a.table[row].size \leftarrow 1 + \max_{w \in C_a} \left\{ total - (w.table[s-row-1].size + w.table[row-1].size) \right\}$
4) $row = s$	$a.table[s].size \leftarrow 1 + \max_{w \in C_a} \{ w.table[s-1].size \}$
$a.table[row].size \leftarrow \max(a.table[row].size, a.table[row-1].size), \text{ for } 1 \leq row \leq s$ Update $a.table[row].pointer \leftarrow a.table[row-1].pointer$, accordingly	

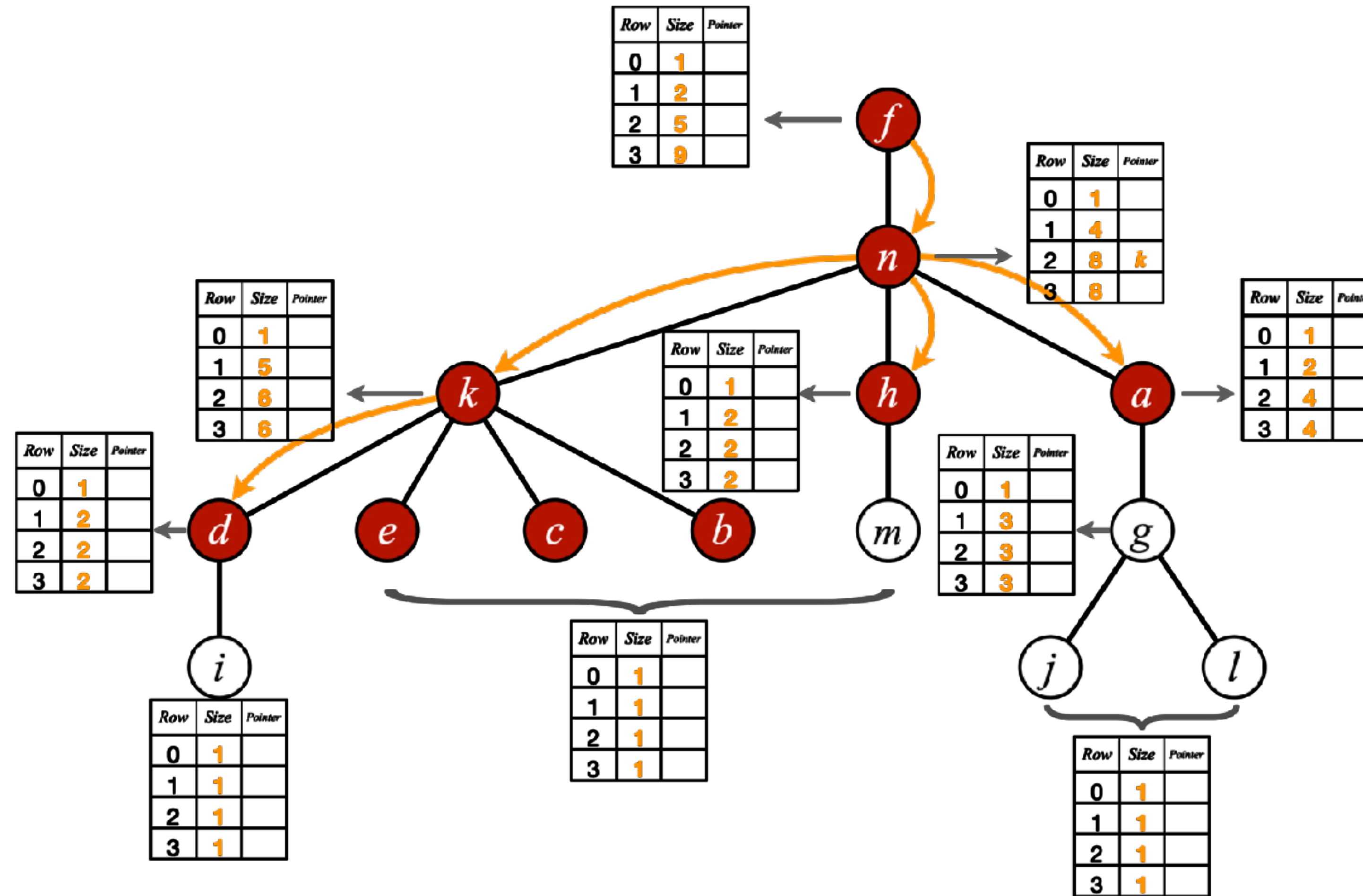
STRATEGY



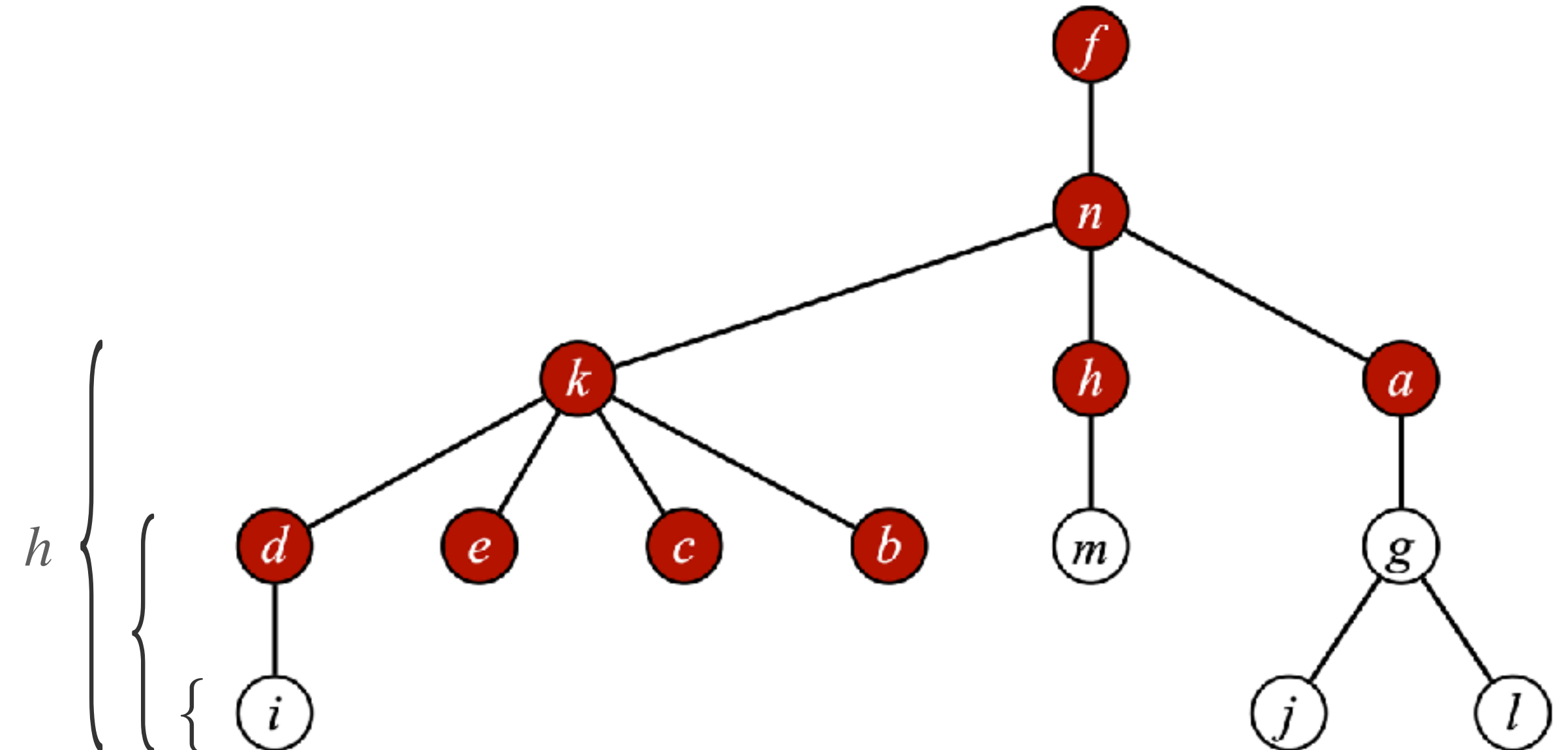
STRATEGY



STRATEGY



CORRECTNESS AND ANALYSIS



- Filling in the table of each vertex v : $\longrightarrow O(s \cdot |C_v|)$ u.t.
- Traversals on the tree: $\longrightarrow O(n)$ u.t.
- Total execution time: $\longrightarrow O(s \cdot n)$ u.t.

OTHER METHODOLOGIES



Methodologies for the s -clique

Exact

<u>Bron & Kerbosh, 1973</u>	$\left\{ \begin{array}{l} \text{Clique, arbitrary graph} \\ \underline{O(3^{\frac{n}{3}}) \text{ u.t.}} \end{array} \right.$
<u>Schäfer, 2009</u>	$\left\{ \begin{array}{l} \text{\textit{s}-club, trees} \\ \underline{O(s^2 \cdot n) \text{ u.t.}} \end{array} \right.$
<u>This work</u>	$\left\{ \begin{array}{l} \text{\textit{s}-clique, trees} \\ \underline{O(s \cdot n) \text{ u.t.}} \end{array} \right.$
<u>Power of a graph</u>	$\left\{ \begin{array}{l} \text{\textit{s}-clique, arbitrary graph} \\ \underline{\text{Exponential}} \end{array} \right.$

Heuristics

<u>Edachery, et, al., 1999</u>	$\left\{ \begin{array}{l} \text{\textit{s}-clique (clusters), arbitrary graph} \\ \underline{O(n^3) \text{ u.t.}} \end{array} \right.$
<u>Behar & Cohen, 2018</u>	$\left\{ \begin{array}{l} \underline{\sim \text{Bron y Kerbosh}} \end{array} \right.$

Mathematical formulation

$\left\{ \begin{array}{l} \underline{\text{Balasundaram, et, al., 2005*}} \end{array} \right.$	$\left\{ \begin{array}{l} \underline{\text{Arbitrary graph}} \\ \underline{\text{CPLEX}} \\ \underline{\text{Small values for } s} \end{array} \right.$
--	---

SCHÄFER'S ALGORITHM & OUR PROPOSAL



Schäfer's

The maximum s -club
problem on trees

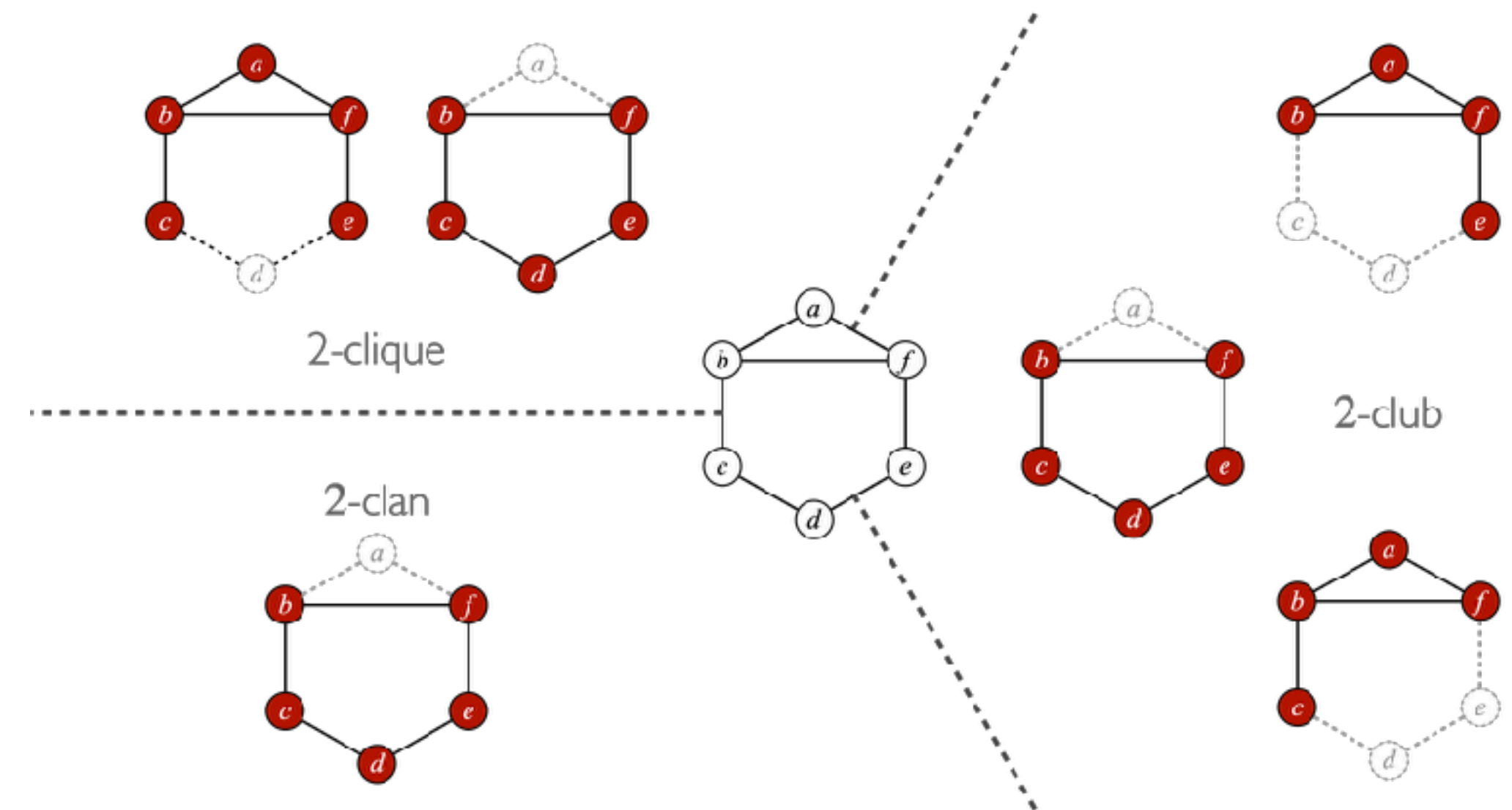
$$T(n) = O(s^2 \cdot n)$$

Our proposal

The maximum s -clique
problem on trees

$$T(n) = O(s \cdot n)$$

On trees: the s -clan, s -club,
and s -clique problems are
equivalent



EXPERIMENTS



Data set

- 1 $|\mathcal{T}_{22_16}| = 12,761$, $|V_T| = 22$, Prof. Brendan McKay.
- 2 T_D , a doubly logarithmic tree of height 5 ($|V_T| = 119,041$)
- 3 T_L a linear tree ($|V_T| = 10,000$)
- 4 T_B a full balanced binary tree with 2^{16} leaves
- 5 $|\mathcal{T}_{Ph}| = 5,722$, $|V_T| = \text{up to } 297$, PhylomeDB
- 6 $T_{\eta e6}$ for $\eta \in \{0.3, 0.5, 0.75, 1\}$ and $e6 = 10^6$ nodes

$$s \in \left\{ \begin{array}{l} 4, 5, 10, 15, 20, \\ 25, 30, 100, 500, \\ 1\,000, 10\,000 \end{array} \right\}$$

where appropriate

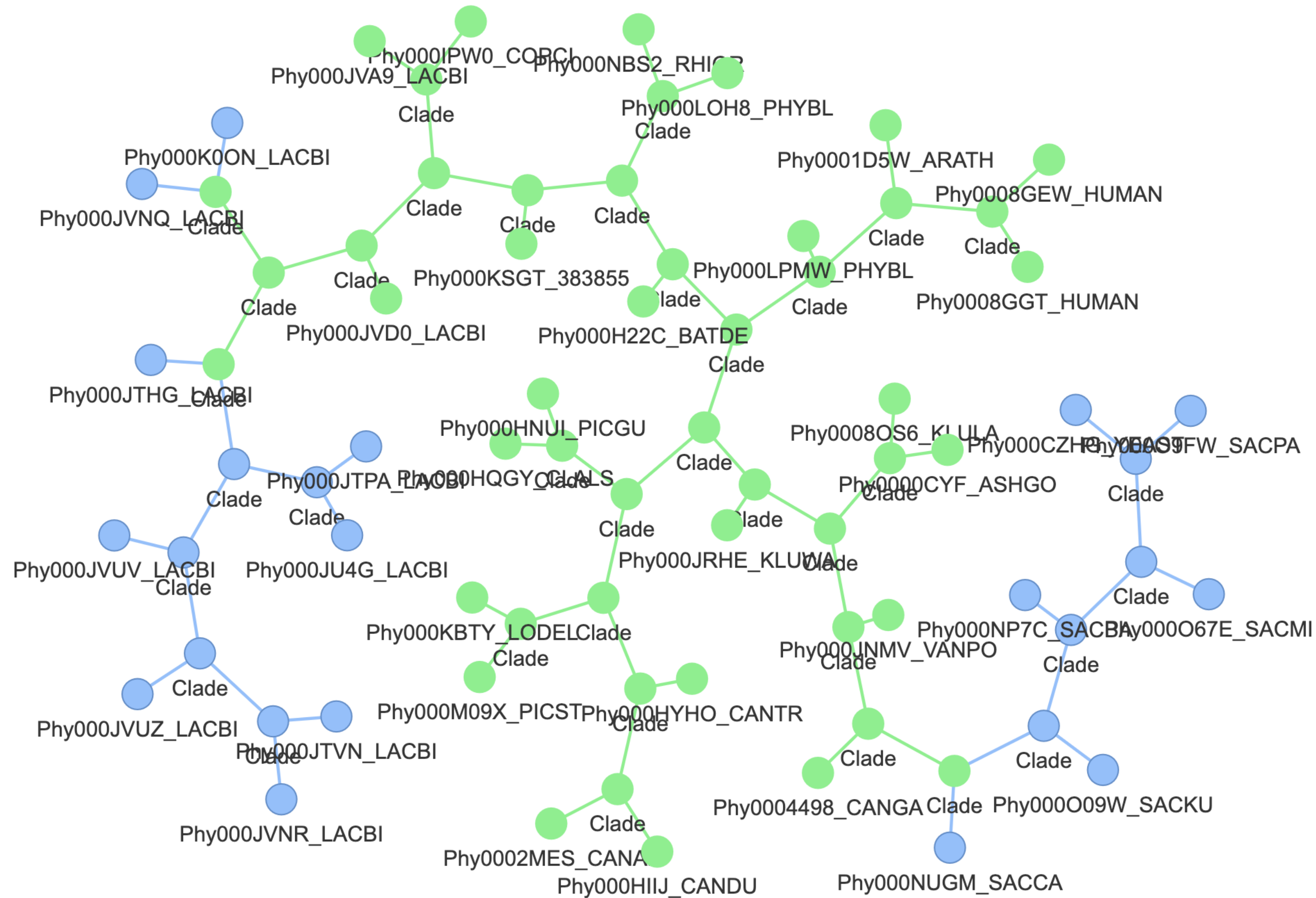
RESULTS



Table 2: Wall-clock running time of Schäfer's implementation and MAX-DsT on the six cases of studies. We denote 'timeouts' and 'not applicable' by '—' and 'n/a', respectively.

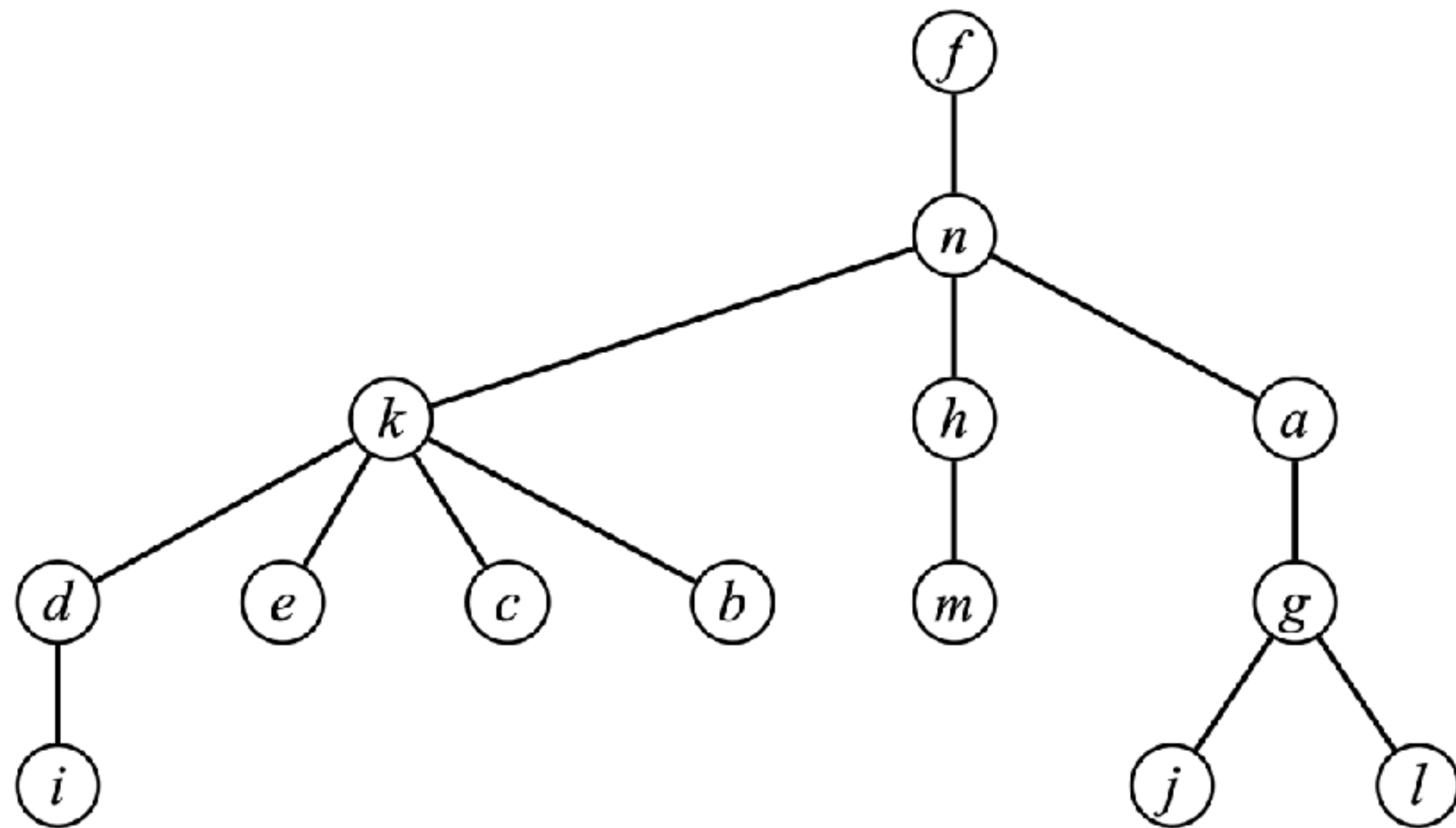
	Row	Graph	Running time in seconds		$\frac{t_{\text{Schäfer}}}{t_{\text{MAX-DsT}}}$	Row	Graph	Running time in seconds		$\frac{t_{\text{Schäfer}}}{t_{\text{MAX-DsT}}}$	
			$t_{\text{Schäfer}}$	$t_{\text{MAX-DsT}}$				$t_{\text{Schäfer}}$	$t_{\text{MAX-DsT}}$		
1 { 2 { 3 {	1	\mathcal{T}_{22_16}	132.4492	60.8545	2.1	14	\mathcal{T}_{Ph}	719.4773	185.4435	3.8	5 {
	2	$T_D, s = 4$	109.9832	1.8577	59.2	15	$T_{0.3e6}, s = 10$	1 022.6843	9.4978	107.7	
	3	$T_D, s = 5$	116.5945	1.5734	74.1	16	$T_{0.3e6}, s = 100$	1 439.42	17.3046	83.1	
	4	$T_L, s = 10$	1.5289	0.5584	2.7	17	$T_{0.3e6}, s = 500$	5 989.8246	74.7689	80.1	
3 { 4 {	5	$T_L, s = 100$	12.6688	4.8969	2.5	18	$T_{0.5e6}, s = 10$	2 945.4381	15.6059	188.7	6 {
	6	$T_L, s = 1 000$	602.9626	45.2492	13.3	19	$T_{0.5e6}, s = 100$	3 654.6010	44.1117	82.8	
	7	$T_L, s = 10 000$	19 859.2017	241.7339	82.1	20	$T_{0.5e6}, s = 500$	—	2 661.9666	n/a	
	8	$T_B, s = 5$	117.9299	2.0445	57.6	21	$T_{0.75e6}, s = 10$	6 436.4398	22.7611	282.7	
4 {	9	$T_B, s = 10$	121.7708	1.9870	61.2	22	$T_{0.75e6}, s = 100$	7 807.3684	94.7391	82.4	
	10	$T_B, s = 15$	125.8261	1.9811	63.5	23	$T_{0.75e6}, s = 500$	—	5 657.6981	n/a	
	11	$T_B, s = 20$	125.4565	1.9592	64.0	24	$T_{1e6}, s = 10$	11 529.7250	31.8786	361.6	
	12	$T_B, s = 25$	126.7537	1.9809	63.9	25	$T_{1e6}, s = 100$	14 167.638	198.9311	71.2	
	13	$T_B, s = 30$	148.4265	1.9831	74.8	26	$T_{1e6}, s = 500$	—	11 422.2738	n/a	

A PHYLOGENETIC TREE



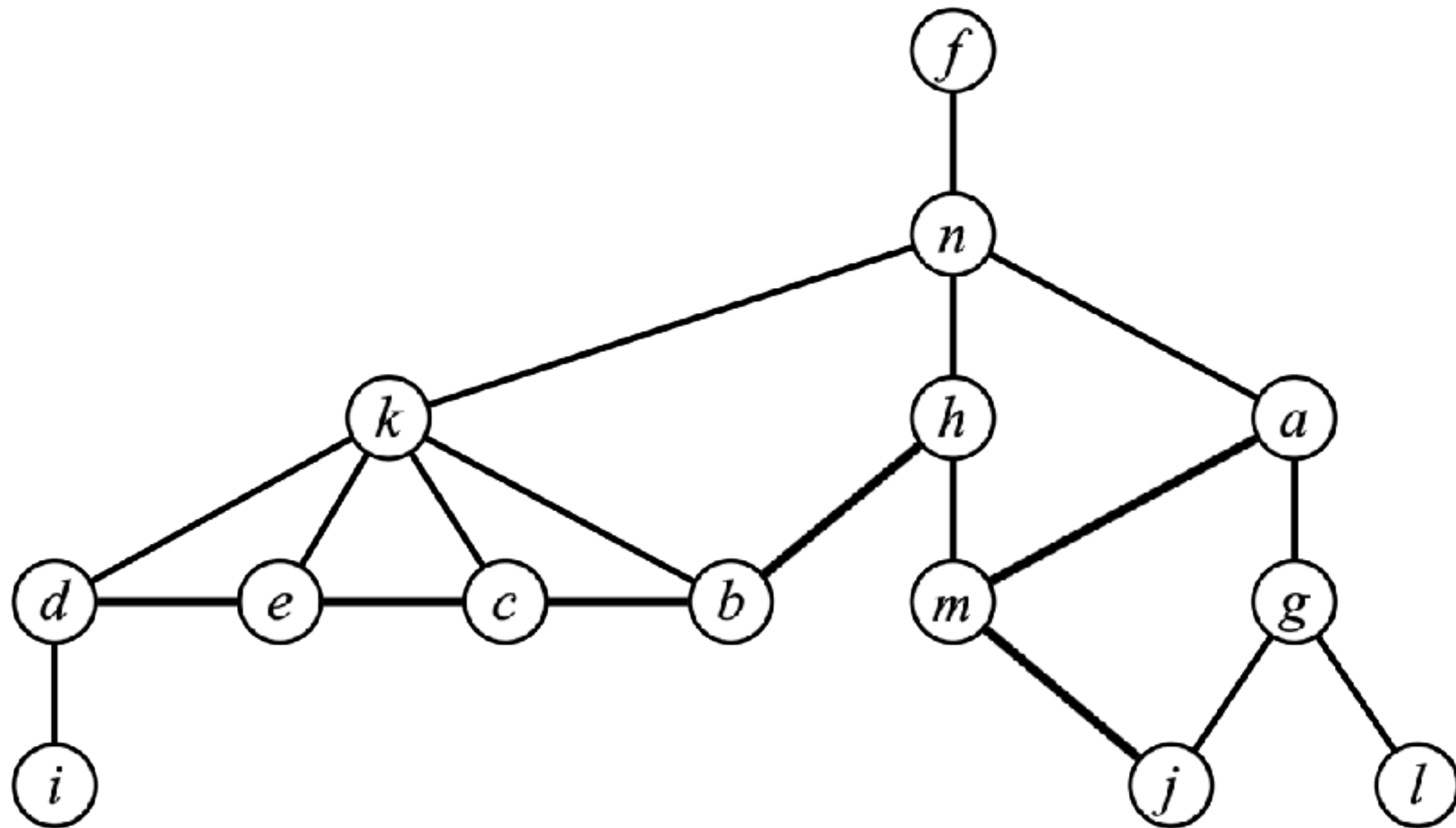
It can potentially help us
find relationships that we
didn't see before!

FUTURE WORK



- Extend this dynamic programming approach for a broader graph family
- Preferably one with an underlying structure close to a tree
 - cordal graphs
 - outerplanar graphs
 - bounded treewidth

FUTURE WORK

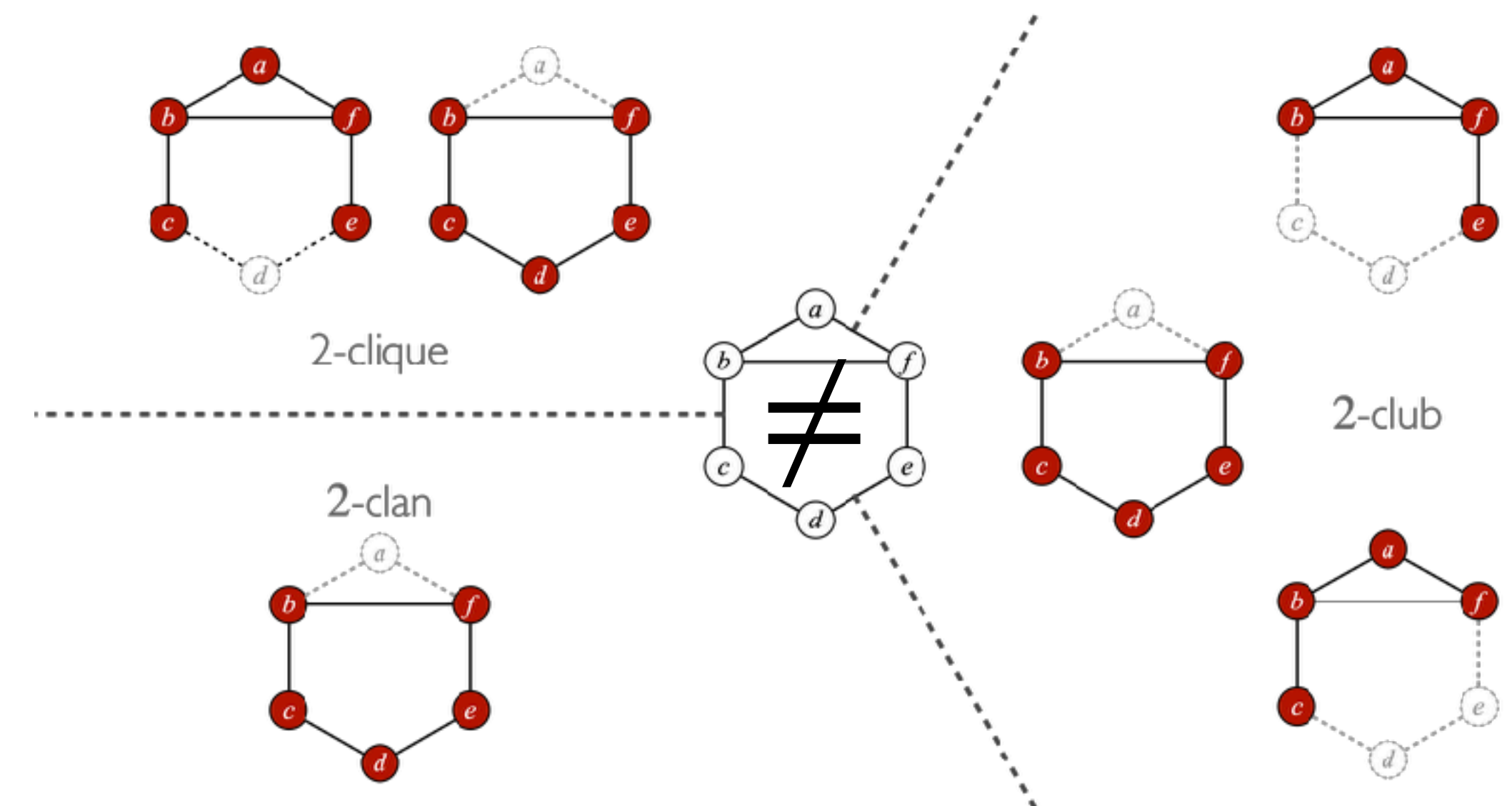


- Extend this dynamic programming approach for a broader graph family
- Preferably one with an underlying structure close to a tree
 - chordal graphs
 - outerplanar graphs
 - bounded treewidth

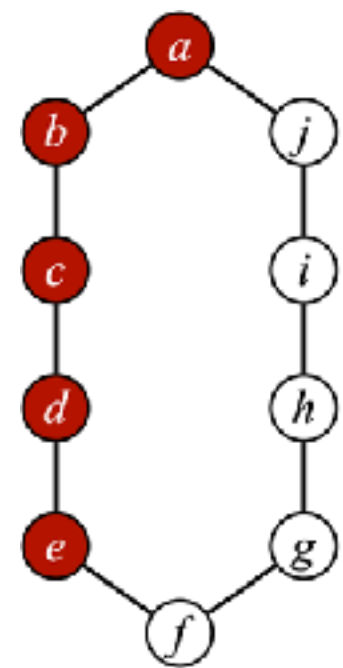
CHALLENGES + ADDITIONAL INFORMATION



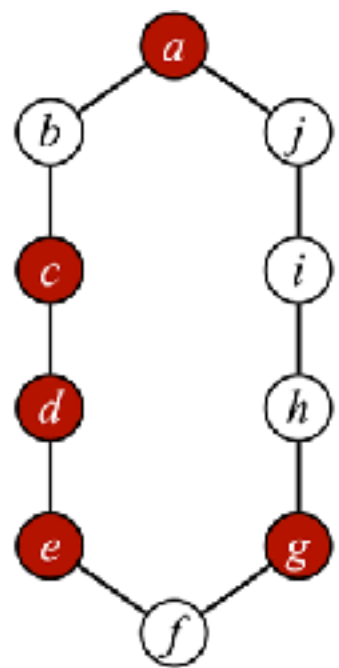
- There could exist more than one path between some pair of vertices
- Cycle-breaking strategy or some hierarchy-like structure to avoid overlapping computations
- On other graphs, the s -clique, s -clan, and s -club might not reflect equivalent problems.



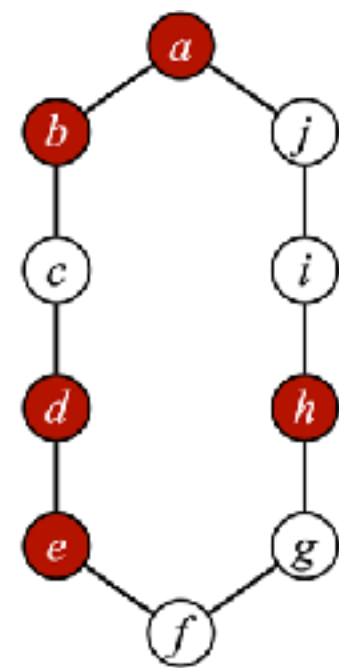
EXAMPLE



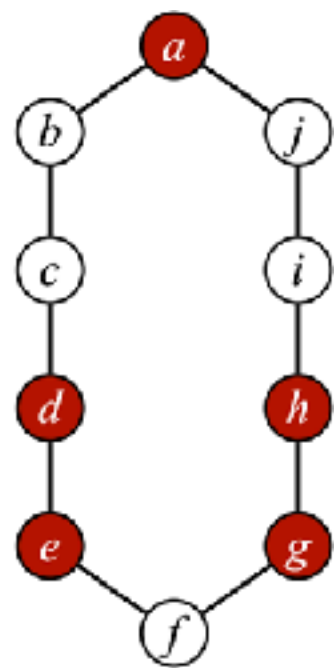
a)



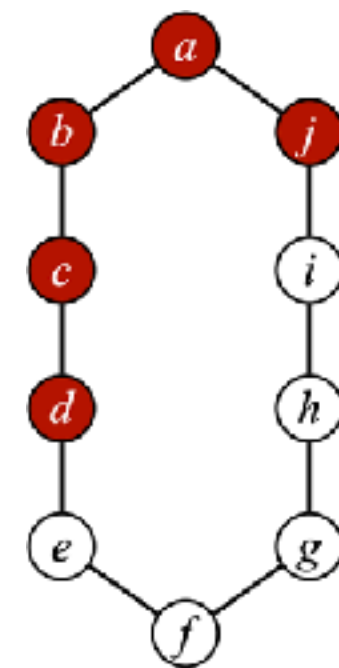
b)



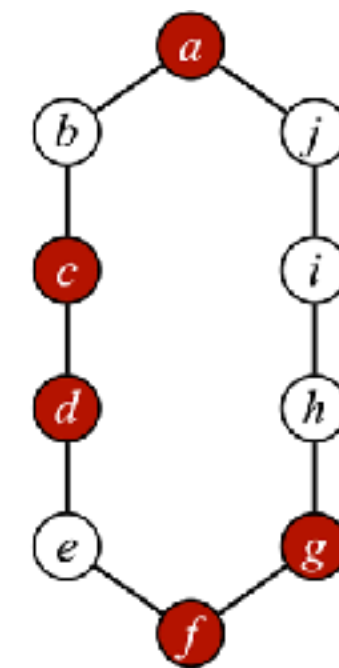
c)



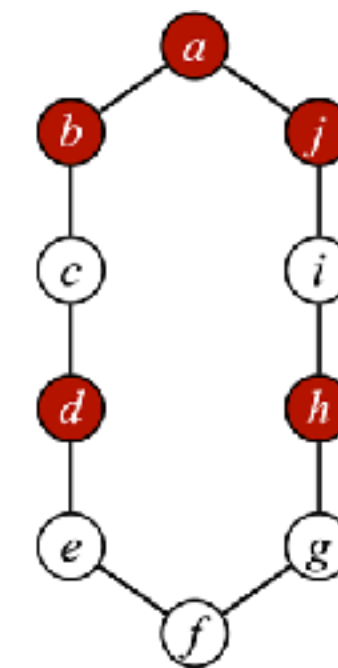
d)



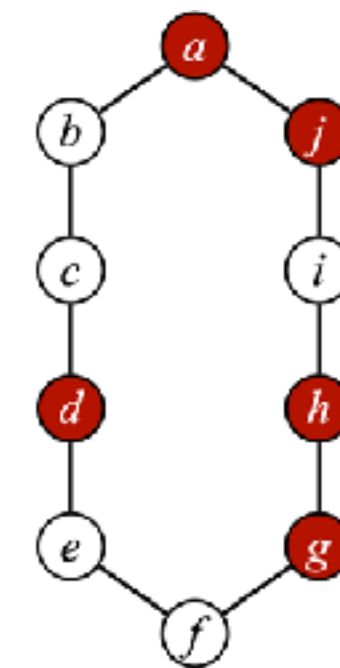
i)



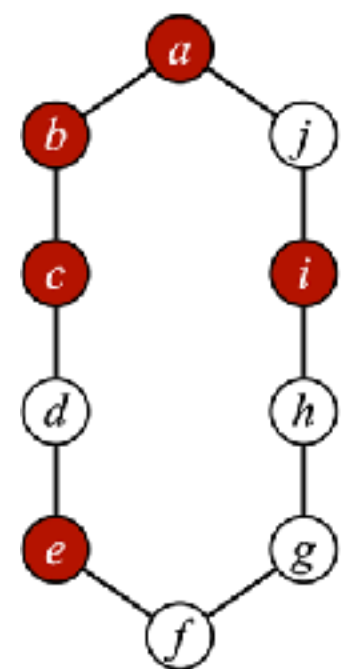
j)



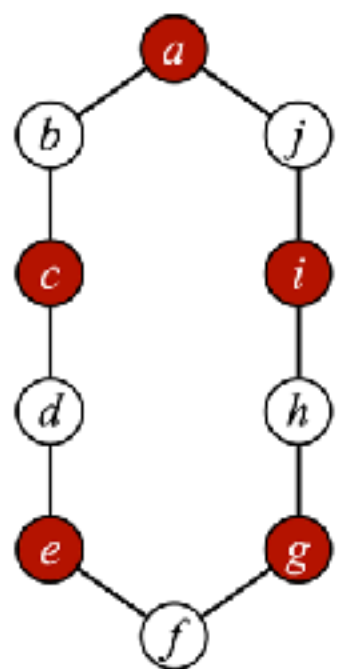
k)



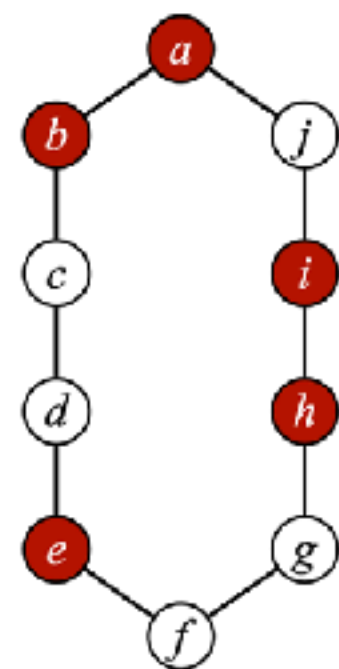
l)



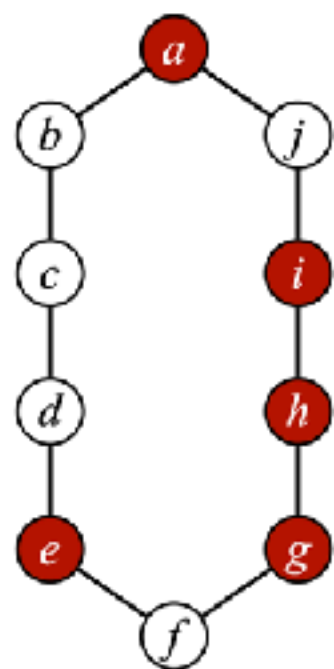
e)



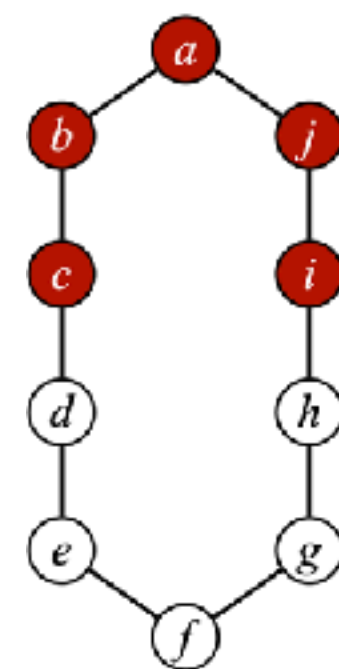
f)



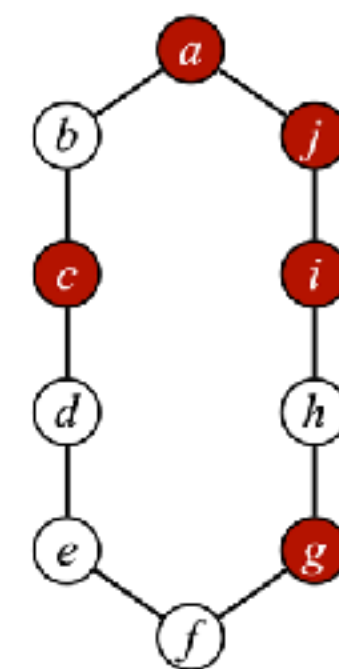
g)



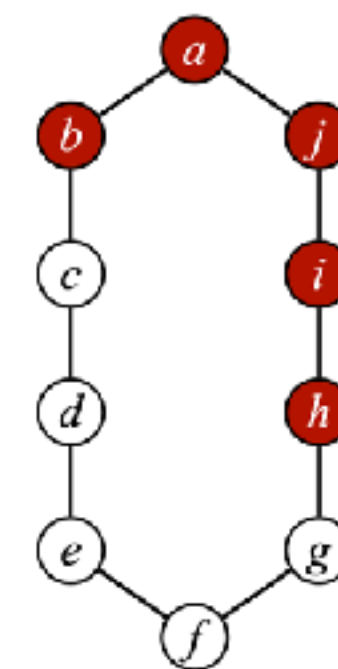
h)



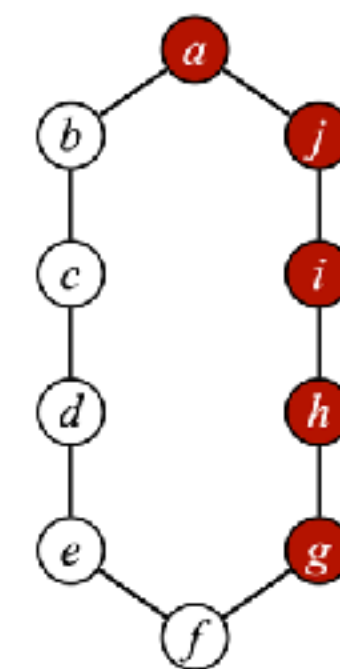
m)



n)



o)



p)

 $\in \text{set}$

$$\text{dist}(\bullet, \bullet) = 4$$



THANK YOU



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

COMMENTS?

fernand@cicese.mx 

joel.trejo@cimat.mx 



Matthew.Hague@rhul.ac.uk 

Alejandro.Flores-Lamas@rhul.ac.uk 

THANK YOU



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON