M2 Complex Systems – Complex Networks

Lecture 13

Complex networks as almost structured graphs

Autumn 2021 – ENS Lyon

Christophe Crespelle

christophe.crespelle@ens-lyon.fr

Modelling static networks

MODEL = RANDOM GENERATION OF SYNTHETIC NETWORKS

For simulating:

- phenomena
- algorithms
- protocols

In order to:

- design
- test
- predict
- better understand

Q: Do Internet protocols still work if Internet is 10 times larger ?

→ Generate a synthetic network and simulate

Modelling static networks

MODEL = RANDOM GENERATION OF SYNTHETIC NETWORKS

4 classic properties:

- Low global density
- Short distances
- Heterogeneous degrees
- High local density

Modelling static networks

MODEL = RANDOM GENERATION OF SYNTHETIC NETWORKS

4 classic properties:

- Low global density → parameter
- Short distances → induced by randomness

Erdös-Rényi 1960

- Heterogeneous degrees
- High local density

Modelling static networks

MODEL = RANDOM GENERATION OF SYNTHETIC NETWORKS

4 classic properties:

- Low global density → parameter
- Short distances → induced by randomness — Erdös-Rényi 1960
- Heterogeneous degrees → compatible with randomness — Molloy & Reed 1995
- High local density

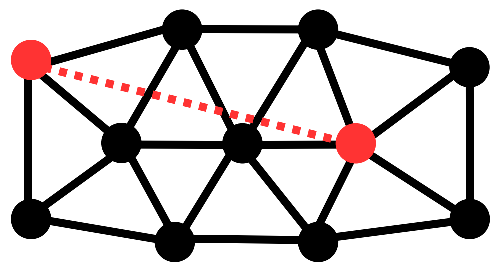
Modelling static networks

MODEL = RANDOM GENERATION OF SYNTHETIC NETWORKS

4 classic properties:

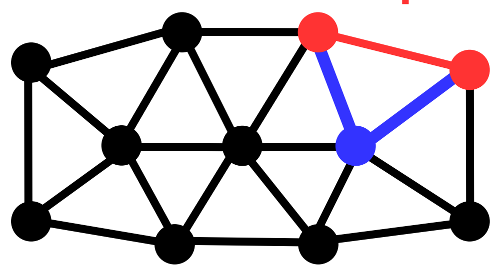
- Low global density → parameter
- Short distances → induced by randomness — Erdös-Rényi 1960
- Heterogeneous degrees → compatible with randomness — Molloy & Reed 1995
- High local density → **problem**

proba ???



global density

proba ???



local density

Modelling static networks

MODEL = RANDOM GENERATION OF SYNTHETIC NETWORKS

4 classic properties:

- Low global density → parameter
- Short distances → induced by randomness — Erdös-Rényi 1960
- Heterogeneous degrees → compatible with randomness — Molloy & Reed 1995
- High local density → **problem**

Big challenge: Generate networks having these 4 properties

low global density | short distances | heterogeneous degrees | high local density

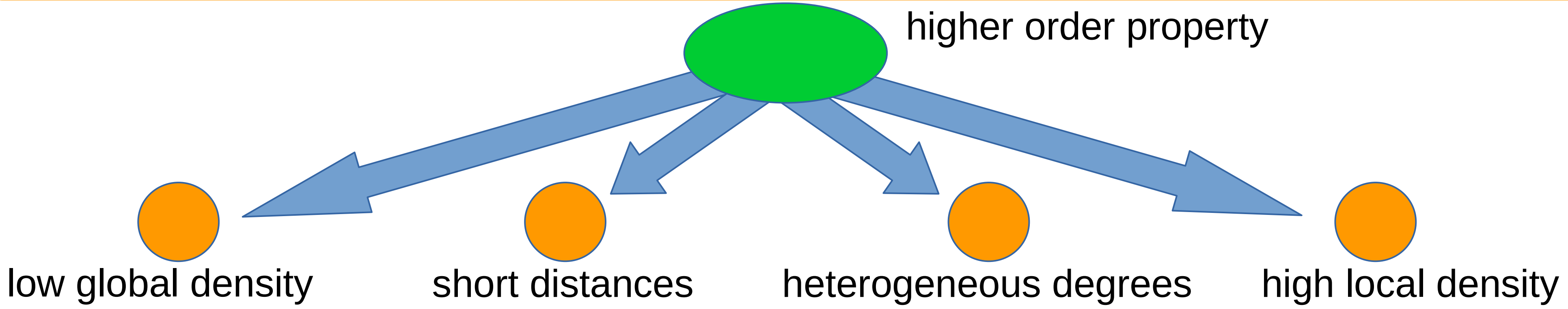
Modelling static networks

MODEL = RANDOM GENERATION OF SYNTHETIC NETWORKS

4 classic properties:

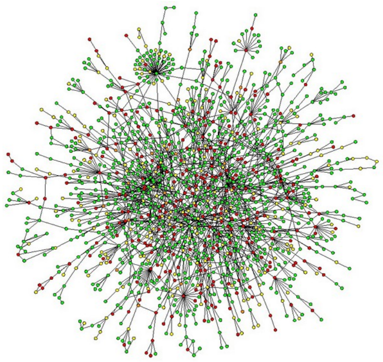
- Low global density → parameter
- Short distances → induced by randomness — Erdös-Rényi 1960
- Heterogeneous degrees → compatible with randomness — Molloy & Reed 1995
- High local density → **problem**

Big challenge: Generate networks having these 4 properties



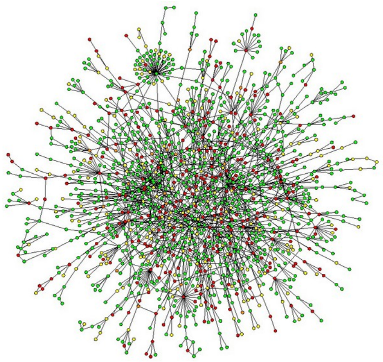
Idea: obtain these properties as a consequence of a higher order property

Almost structured graphs



- loosely constrained
  - → **randomness**
- strongly impacted by their context
  - → **structure**

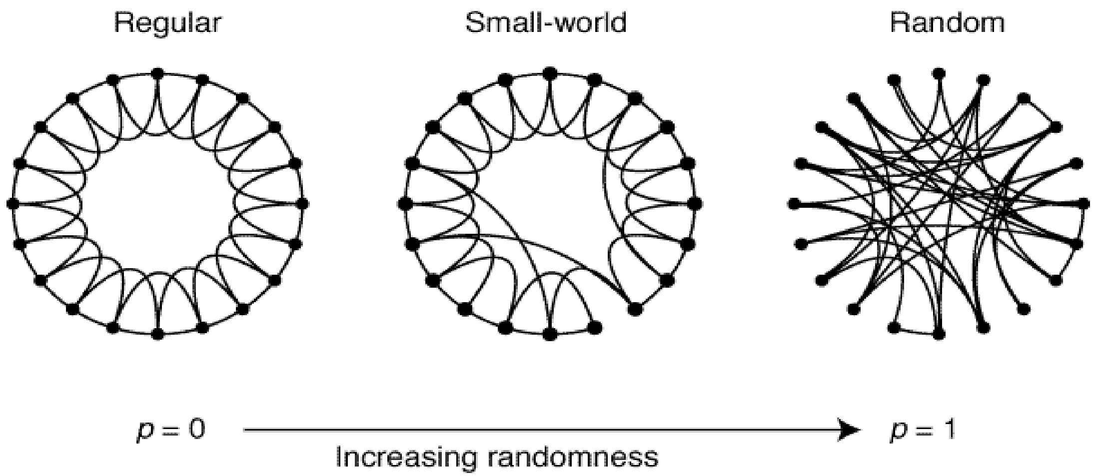
Almost structured graphs



- loosely constrained
  → **randomness**
- strongly impacted by their context
  → **structure**

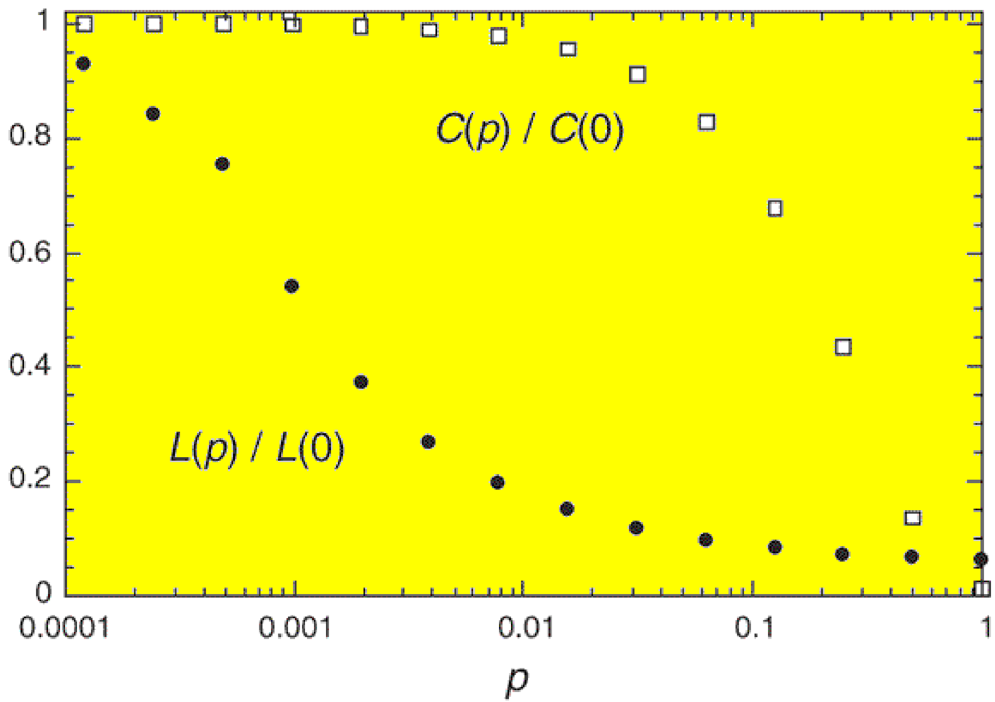
Complex networks	**=**	**structure**	**+**	**randomness**
[Watts & Strogatz 1998]		High local density		Short distances

Watts & Strogatz model



Regular lattice with n nodes
k^{th} power of the cycle, k<<n

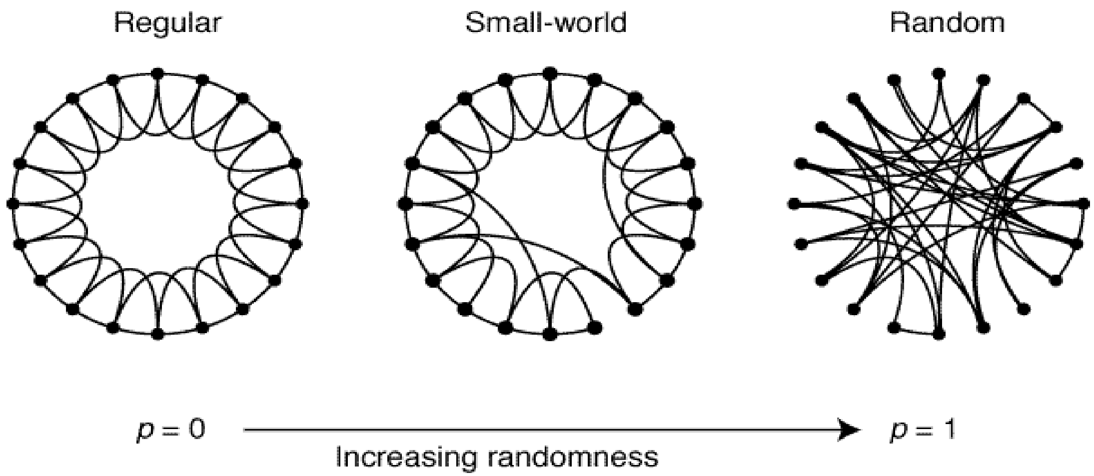
Second endpoint of each edge is rewired with probability p



Clustering C(p)
vs
average distance L(p)

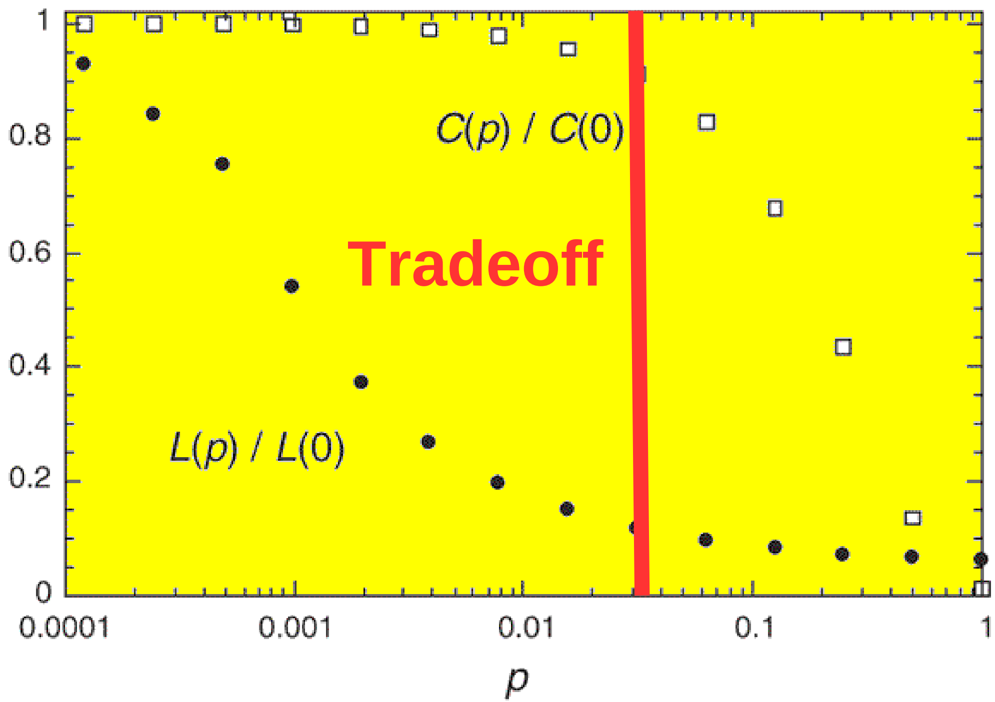
as p increases

Watts & Strogatz model



Regular lattice with n nodes
k[th] power of the cycle, k<<n

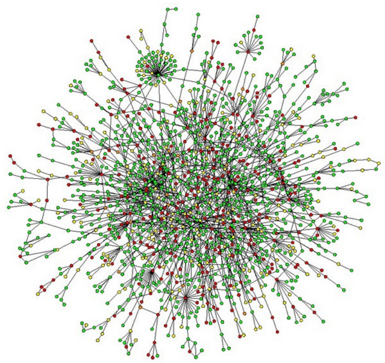
Second endpoint of each edge is rewired with probability p



Clustering C(p)
vs
average distance L(p)

as p increases

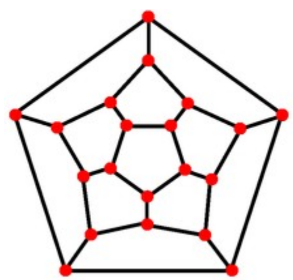
Almost structured graphs



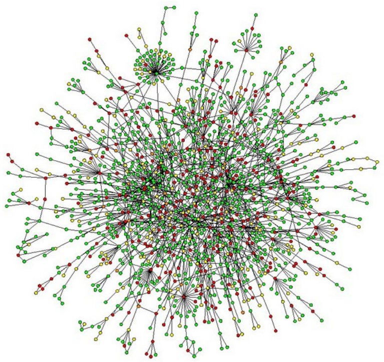
- loosely constrained
  → **randomness**
- strongly impacted by their context
  → **structure**

Complex networks = structure + randomness

1 strongly structured



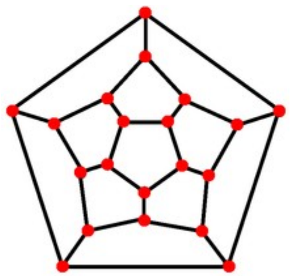
Almost structured graphs



- loosely constrained
  → **randomness**
- strongly impacted by their context
  → **structure**

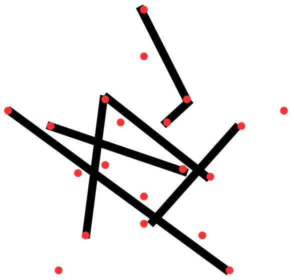
Complex networks	**=**	**structure**	**+**	**randomness**

1 strongly structured

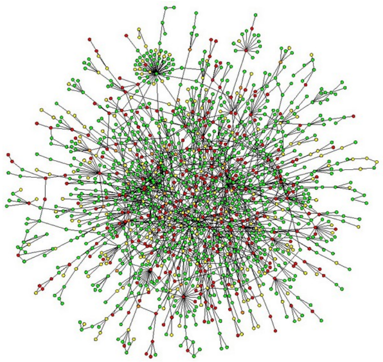


+

2 random modifications



Almost structured graphs

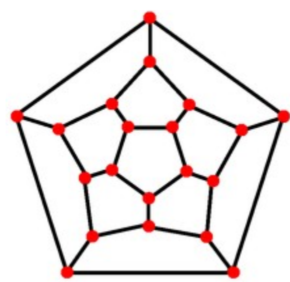


- loosely constrained
  → **randomness**
- strongly impacted by their context
  → **structure**

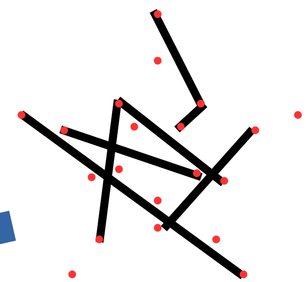
Complex networks = structure + randomness

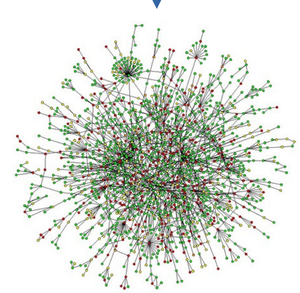
1 strongly structured

2 random modifications

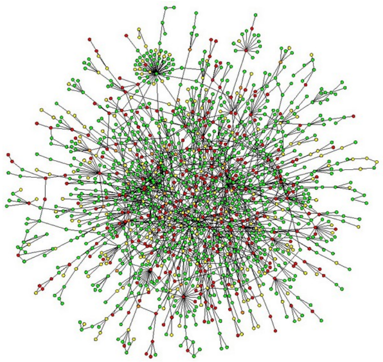


+





Almost structured graphs

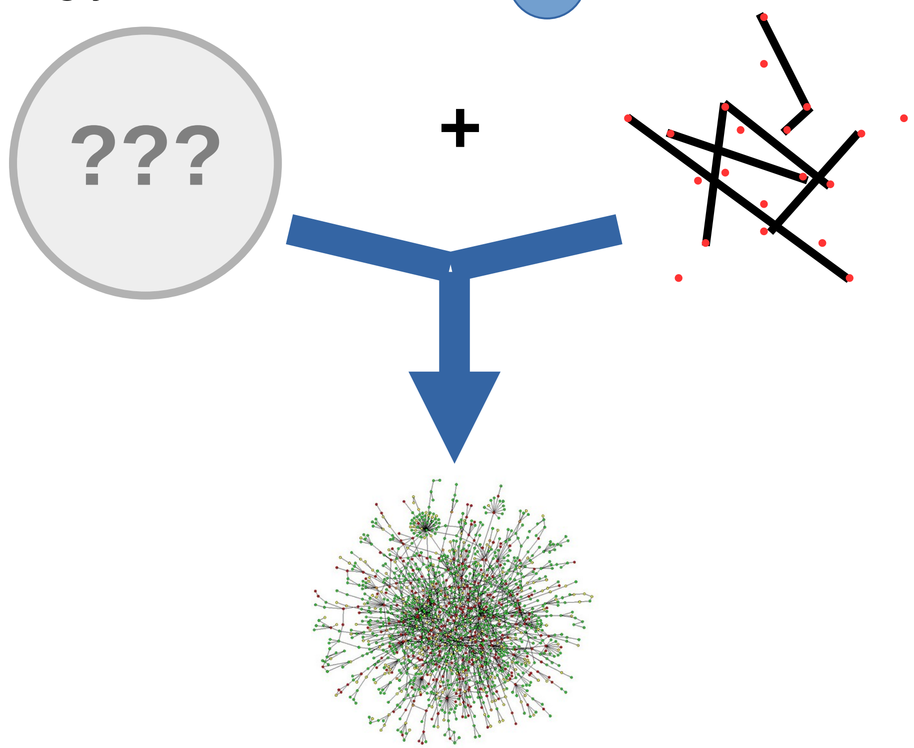


- loosely constrained
  → **randomness**
- strongly impacted by their context
  → **structure**

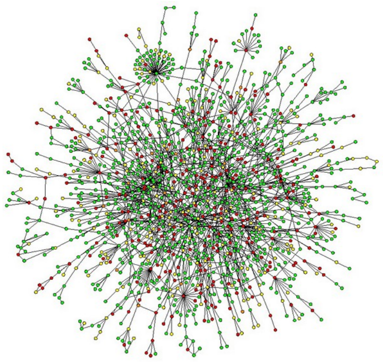
Complex networks = structure + randomness

1 strongly structured

2 random modifications



Almost structured graphs

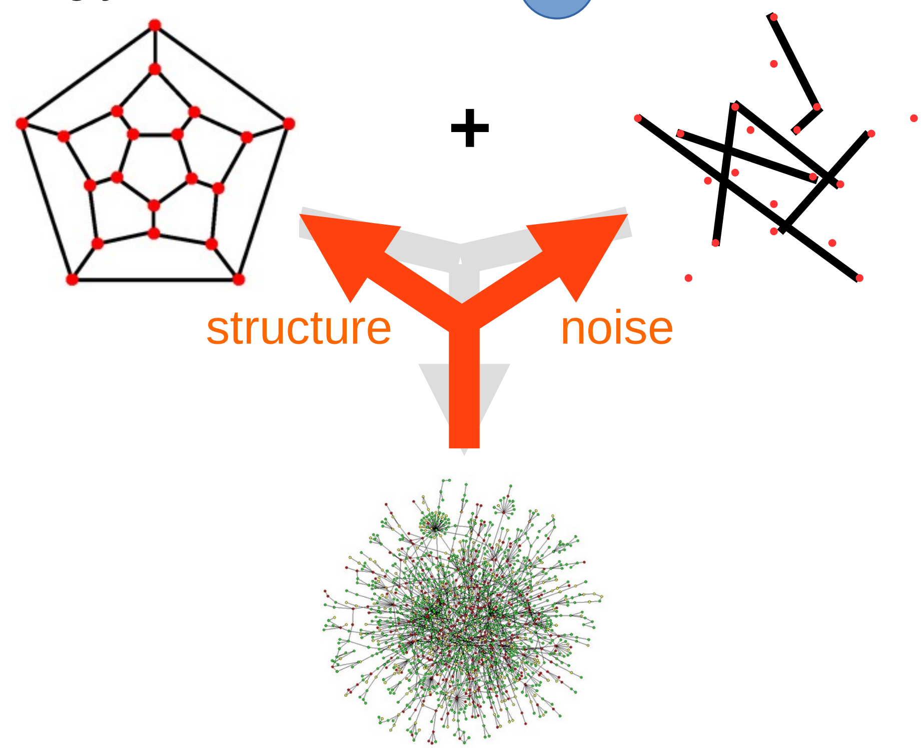


- loosely constrained
  → **randomness**
- strongly impacted by their context
  → **structure**

Complex networks = structure + randomness

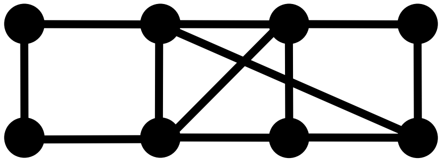
1 strongly structured

2 random modifications



Graph editing algorithms

INPUT

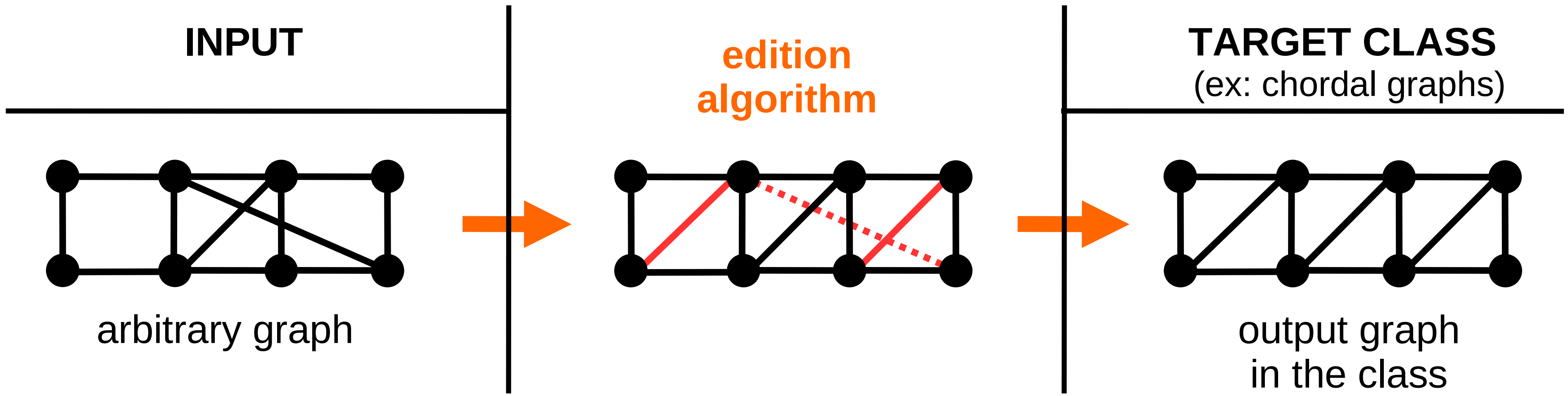


arbitrary graph

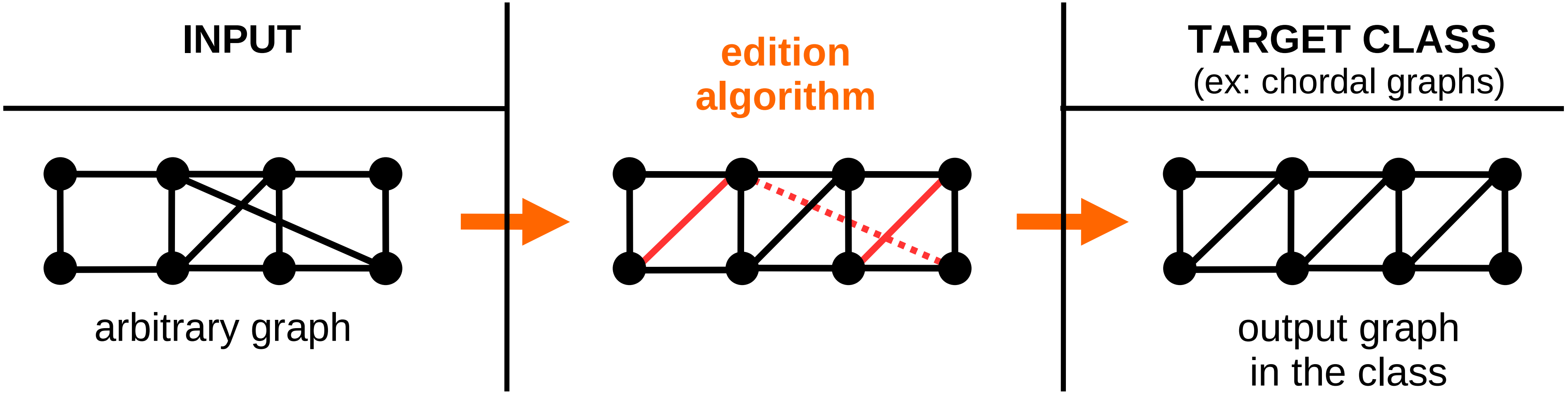
TARGET CLASS

(ex: chordal graphs)

Graph editing algorithms



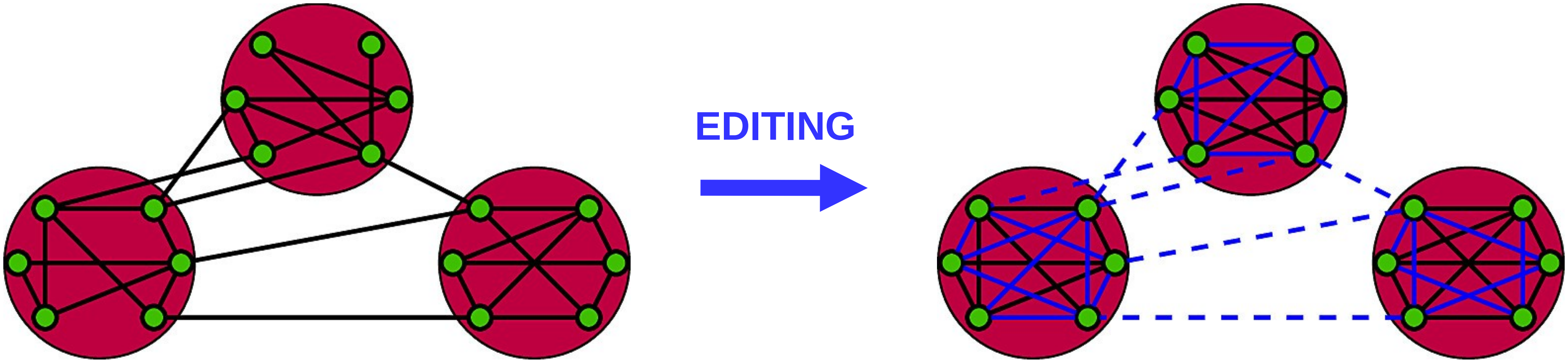
Graph editing algorithms



GOAL: perform as few modifications as possible

Graph editing algorithms

- Community detection

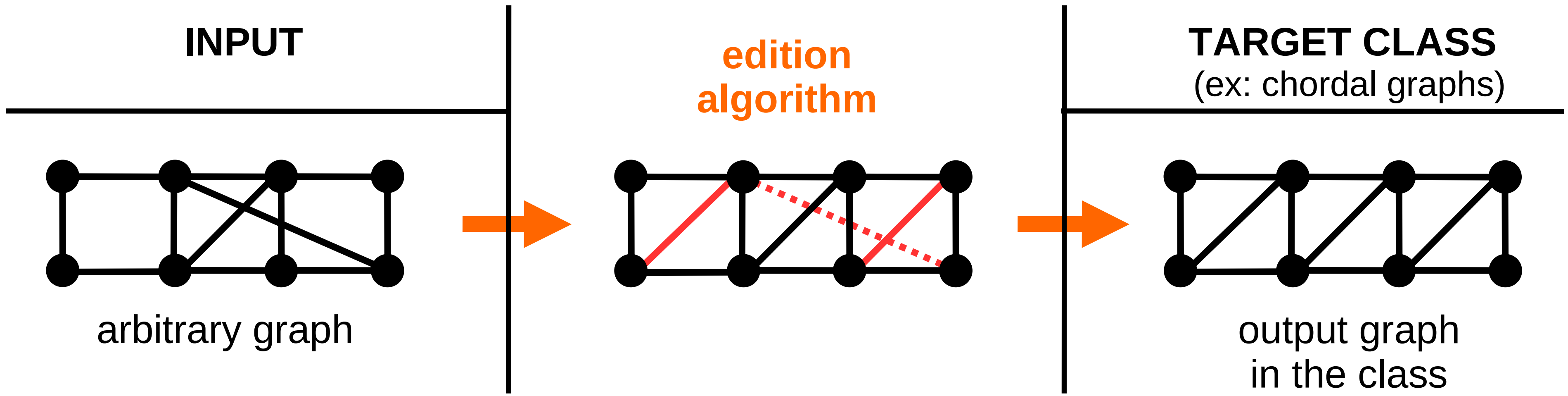


Original network

Resulting cluster graph

- Degree anonymization
  - Edit the graph so that all vertices have same degree

Graph editing algorithms



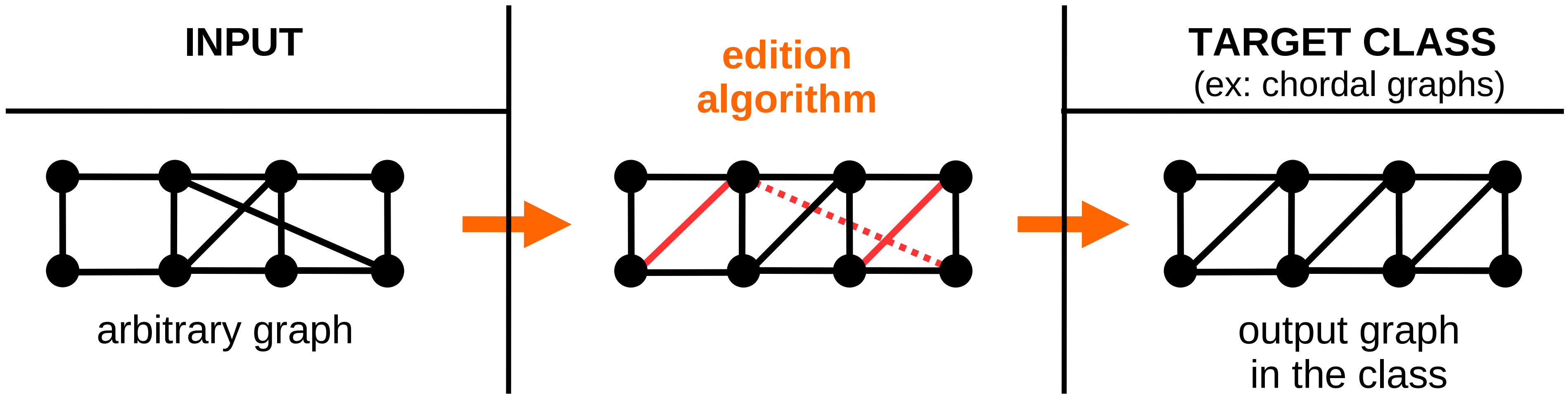
GOAL: perform as few modifications as possible

- Unfortunately: *minimum number* is ***NP-hard*** for most properties

  Even when only one type of modifications is allowed (eg. only additions)

- Different approaches:
  - Restricted inputs
  - Exact exponential algorithms
  - Parameterized algorithms
  - Approximation algorithms
  - *Inclusion minimal modification*

Graph editing algorithms



GOAL: perform as few modifications as possible

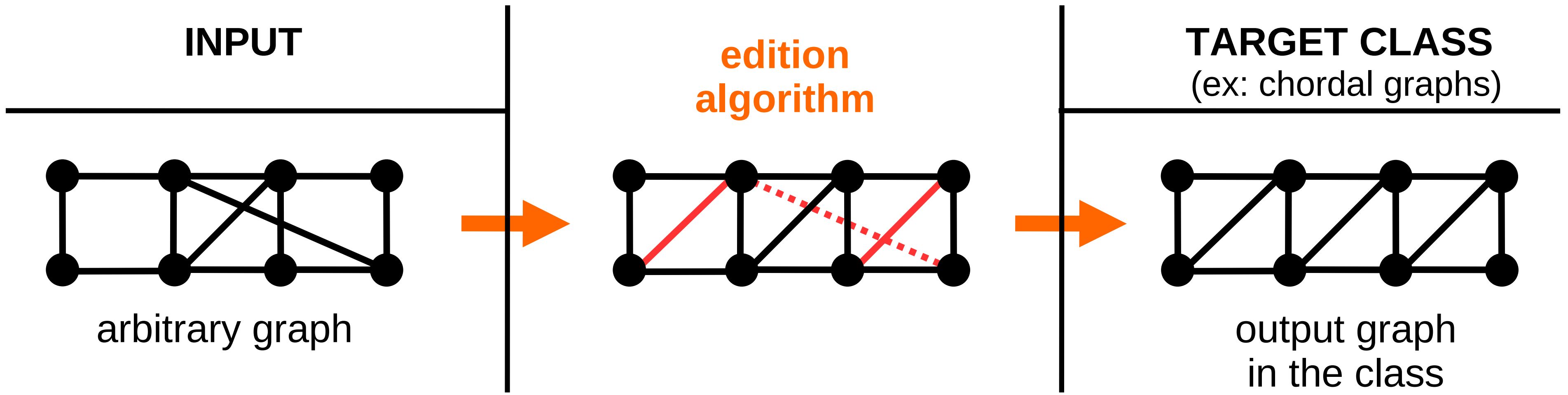
- Unfortunately: *minimum number* is ***NP-hard*** for most properties

  Even when only one type of modifications is allowed (eg. only additions)

- Relaxation of the problem:

  set of modifications *minimal for inclusion* → **polynomial time**

Graph editing algorithms



GOAL: perform as few modifications as possible

- Unfortunately: *minimum number* is ***NP-hard*** for most properties

  Even when only one type of modifications is allowed (eg. only additions)

- Relaxation of the problem:

  set of modifications *minimal for inclusion* → **polynomial time**

  ⚠ each target class needs a ***specific*** algorithm !

  Ex : interval graphs, permutation graphs, cographs

Results for some target classes

Completion:

- Interval completion : $O(n^2)$
  1981, 2005, 2013
- Chordal completion : O(nm)
  2006
- Trivially perfect completion : O(n+m')
  2008
- Comparability completion : $O(n^2m)$
  2008
- Split completion : O(n+m')
  2009
- Cograph completion : O(n+m')
  2010
- Permutation completion : $O(n^2)$
  2015

Deletion:

- Planar deletion : O(n+m)
  2006

Minimal cograph editing algorithms

Coedit : a tool for cograph editing

INPUT: an arbitrary graph

<u>Computes either:</u>

- a minimal cograph completion
- a minimal cograph deletion
- a minimal cograph editing

OUTPUT: the cotree of the cograph obtained

Input format:

# of vertices	n
degrees	u d°(u) v d°(v) ⋮
edges	u1 v1 u2 v2 ⋮

Output format:

# of nodes	n
Label of the root	l (=0 or 1)
# of children	u #child(u) v #child(v) ⋮
Edges of the tree	parent(u) u parent(v) v ⋮

- Written in C
- Sources available at https://www.ii.uib.no/~christophec/coedit/
- Under GNU GPL licence (can do whatever you want with it)

Algorithms

For completion

- An O(n+m') algorithm with ***minimum*** at each incremental step
  → improve heuristics
- An O(n+m log²n) algorithm
  → almost linear in the size of the ***input***

For editing

- An O(n+m) algorithm with ***minimum*** at each incremental step

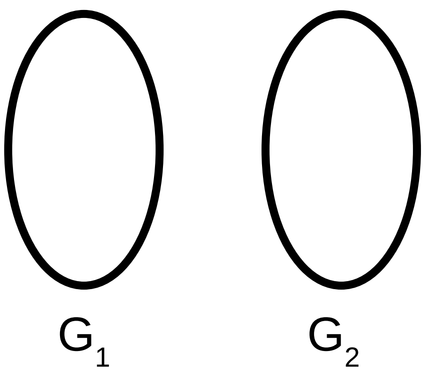
The vertex incremental approach : vertices are processed one by one

Cographs and incremental app.

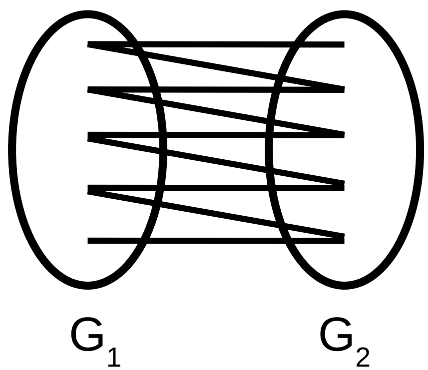
Obtained from single vertices by using 2 operations:

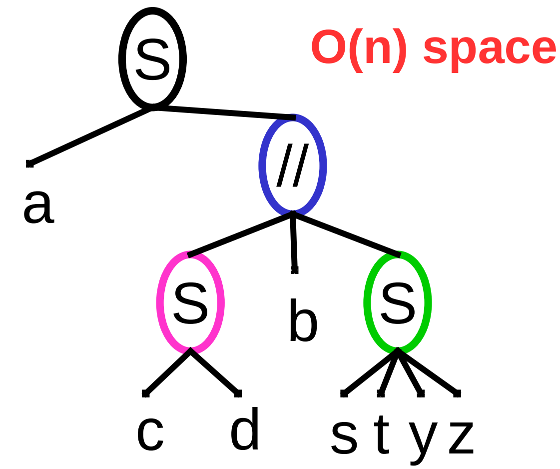
cotree

disjoint union
(//)



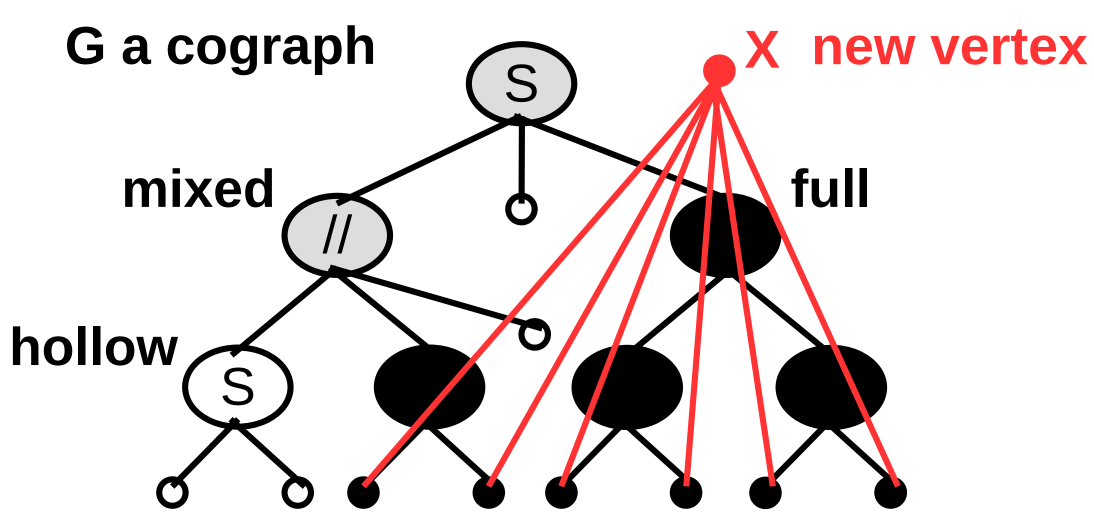
complete union
(S)





Incremental approach: a **cograph G** and **x** a new incoming vertex

G+x is not a cograph and we want to add (and/or delete) edges incident to x so that G+x become a cograph



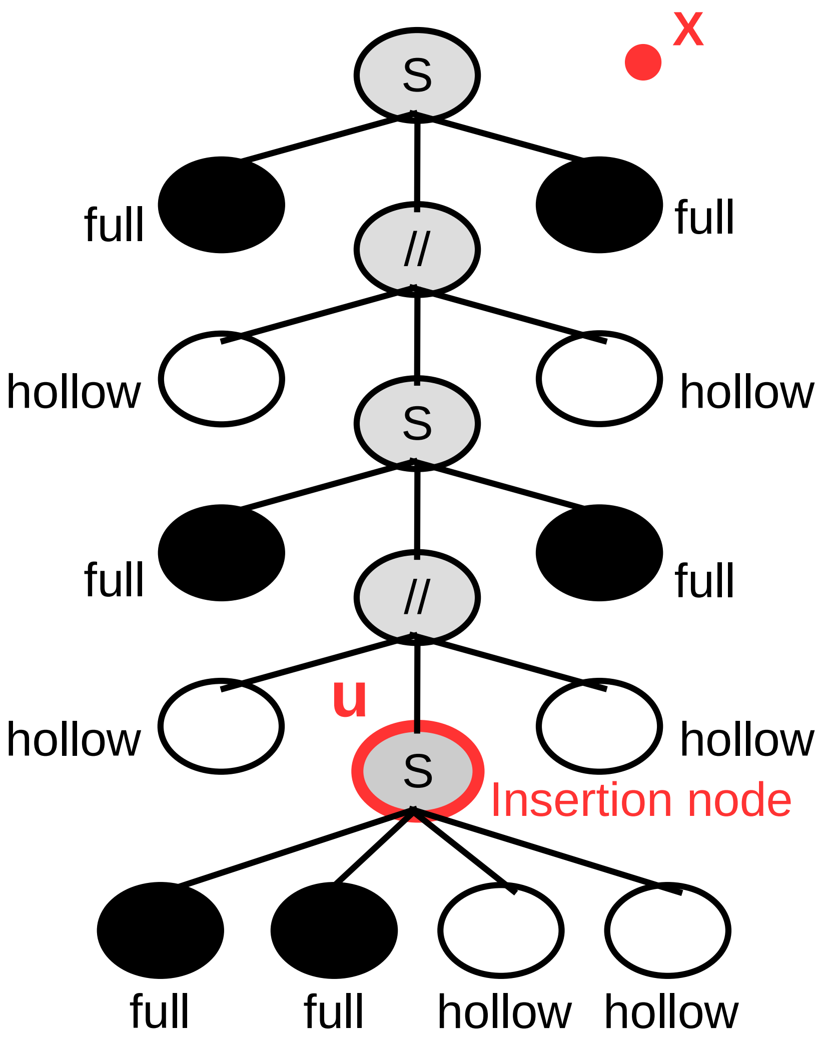
Completion algorithms

First algorithm: O(n+m')

A characterisation of cographs

[Corneil, Perl, Stewart 1981]

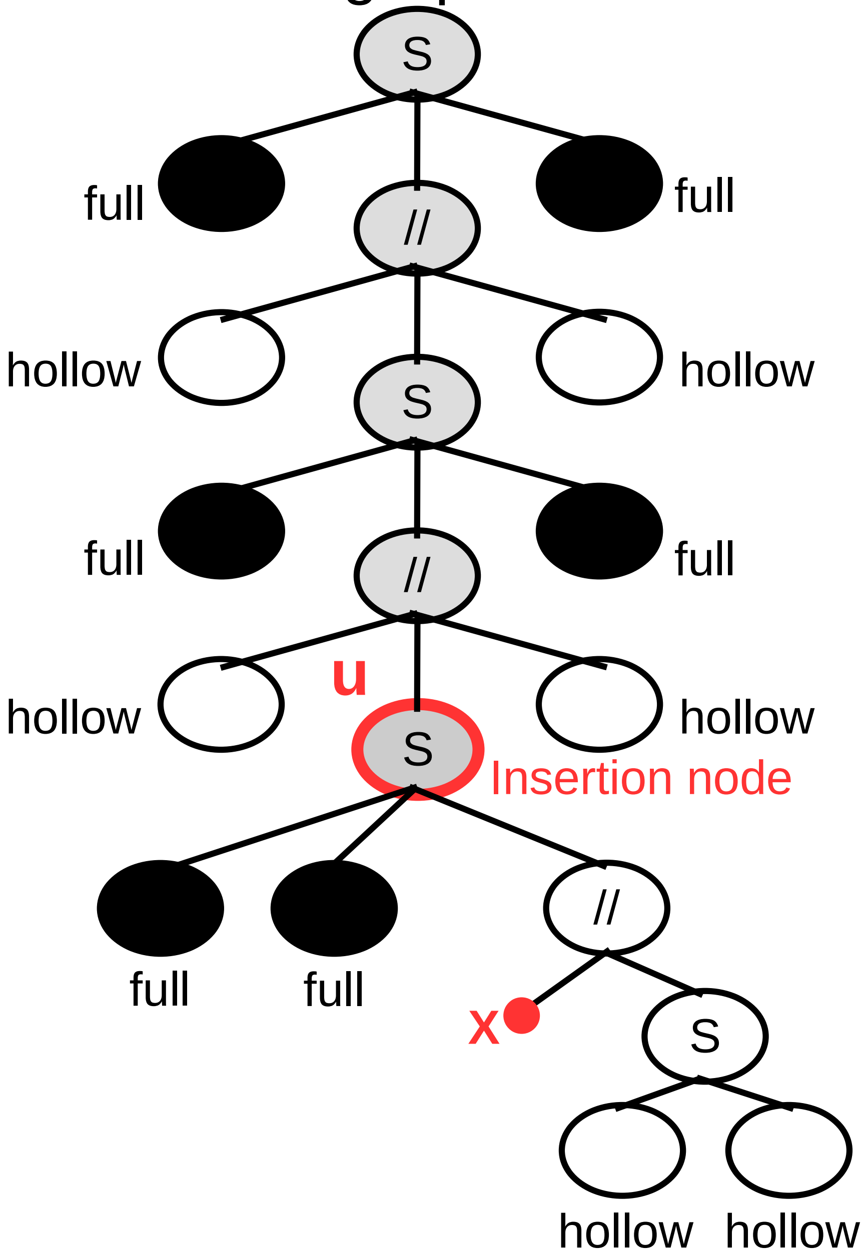
G+x is a cograph iff there exists a node u st.:



A characterisation of cographs

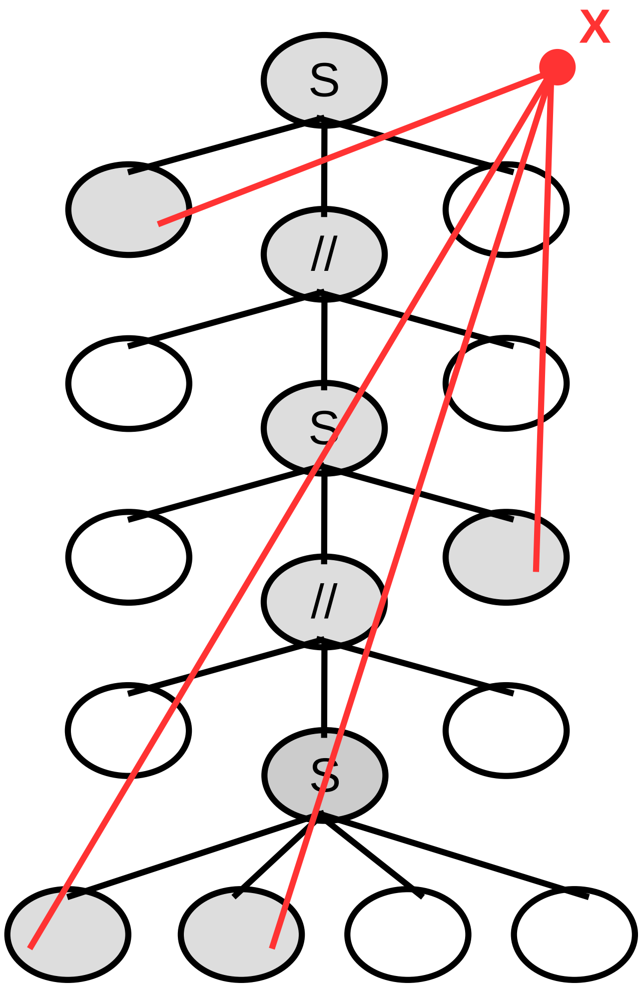
[Corneil, Perl, Stewart 1981]

G+x is a cograph iff there exists a node u st.:



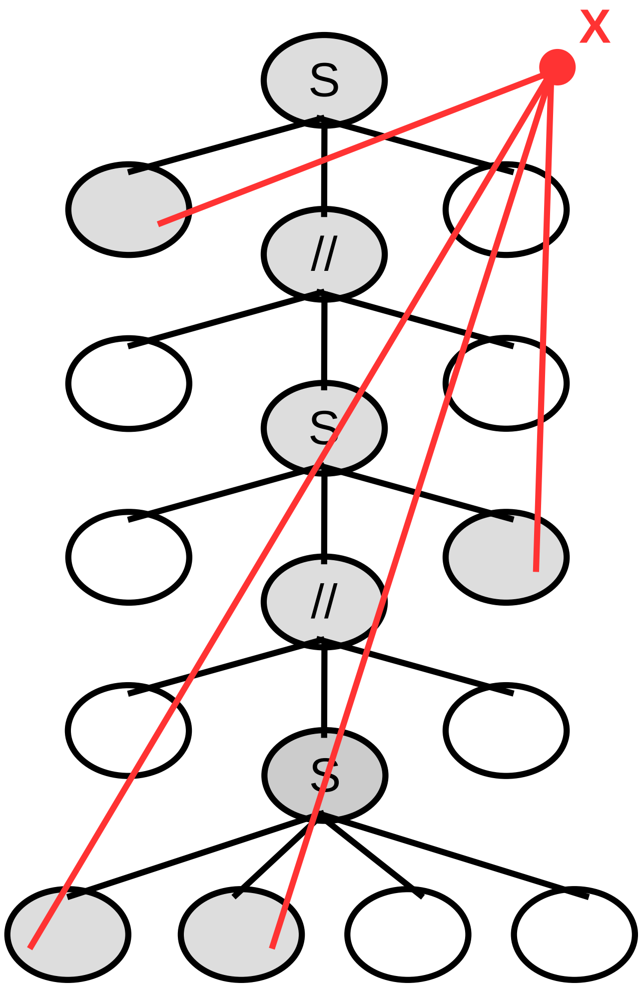
A characterisation of cographs

In our algorithm : G+x is not a cograph



A characterisation of cographs

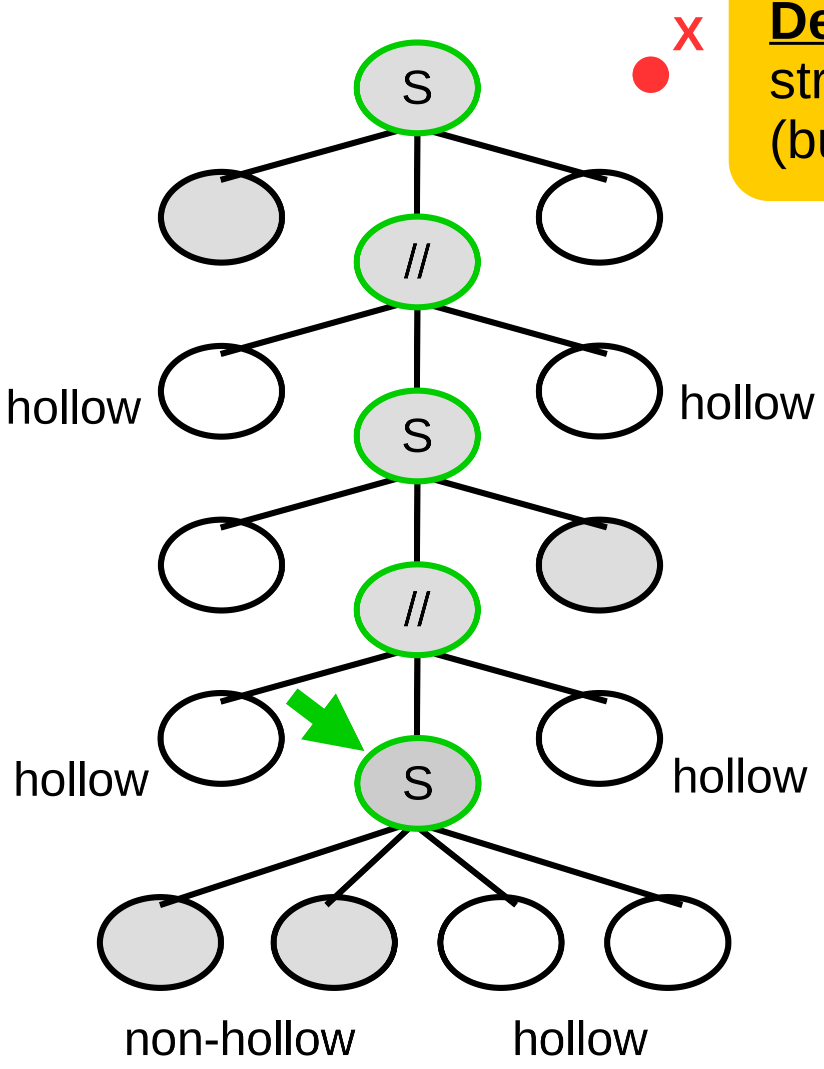
In our algorithm : G+x is not a cograph



Choose one node u for which you make the situation of the [CPS 81]'s theorem happen

Eligible nodes

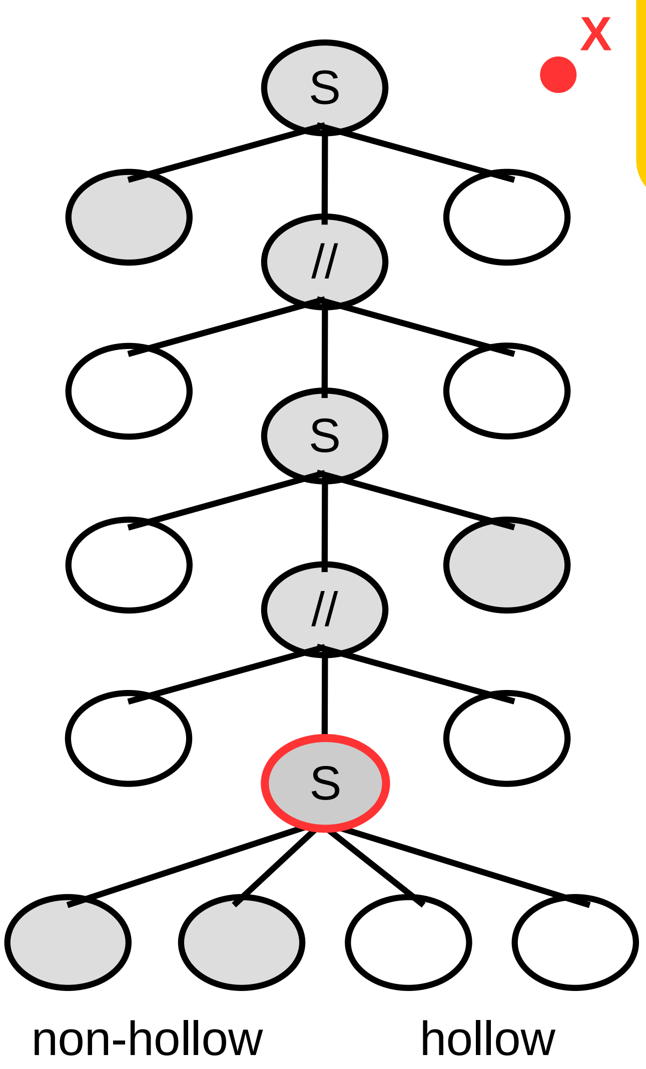
In our algorithm : G+x is not a cograph



Definition: u is an ***eligible node*** Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

Completion anchored at u

In our algorithm : G+x is not a cograph



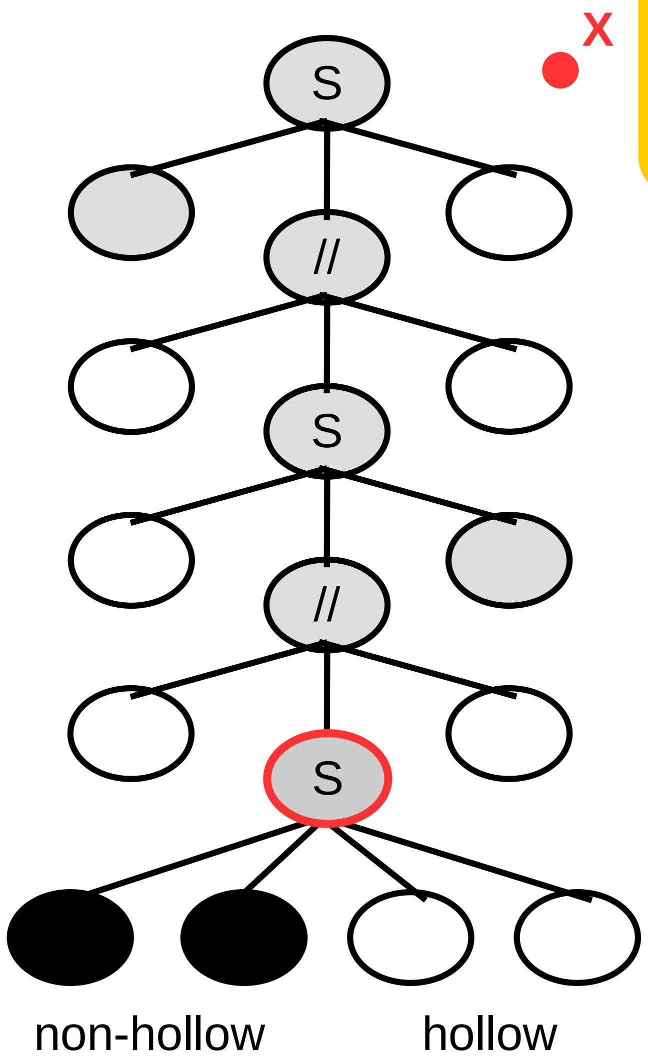
Definition: u is an ***eligible node*** Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

Proceed as follows :

1) choose one eligible node ***u***

Completion anchored at u

In our algorithm : G+x is not a cograph



Definition: u is an ***eligible node*** Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

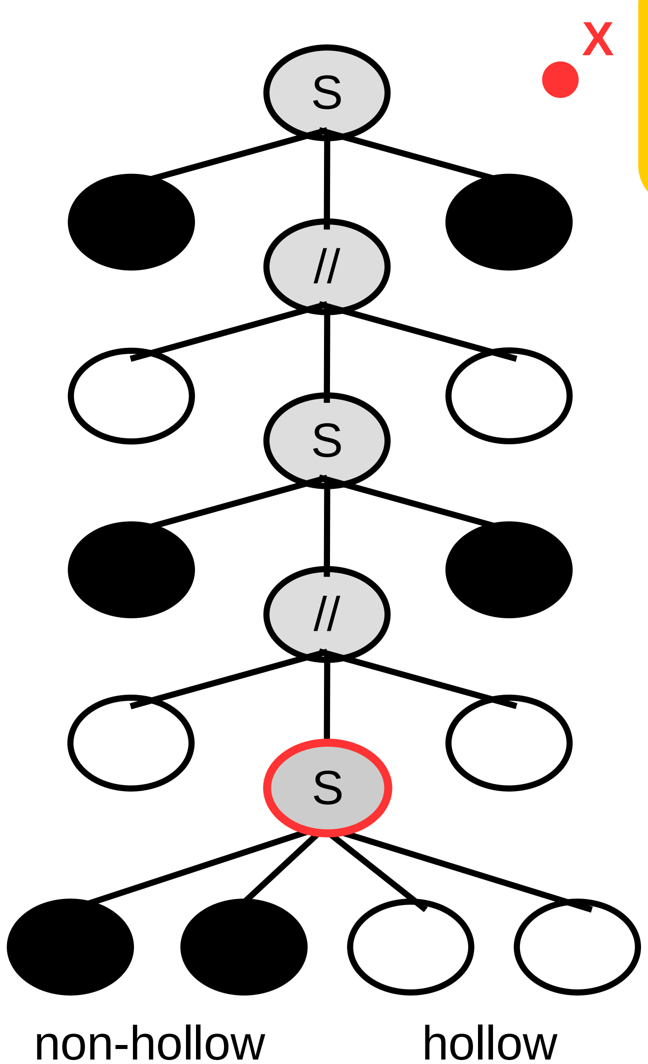
Proceed as follows :

1) choose one eligible node ***u***

2) make the non-hollow children of u become ***full*** (leave the others ***hollow***)

Completion anchored at u

In our algorithm : G+x is not a cograph



<u>Definition:</u> u is an ***eligible node*** Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

<u>Proceed as follows :</u>

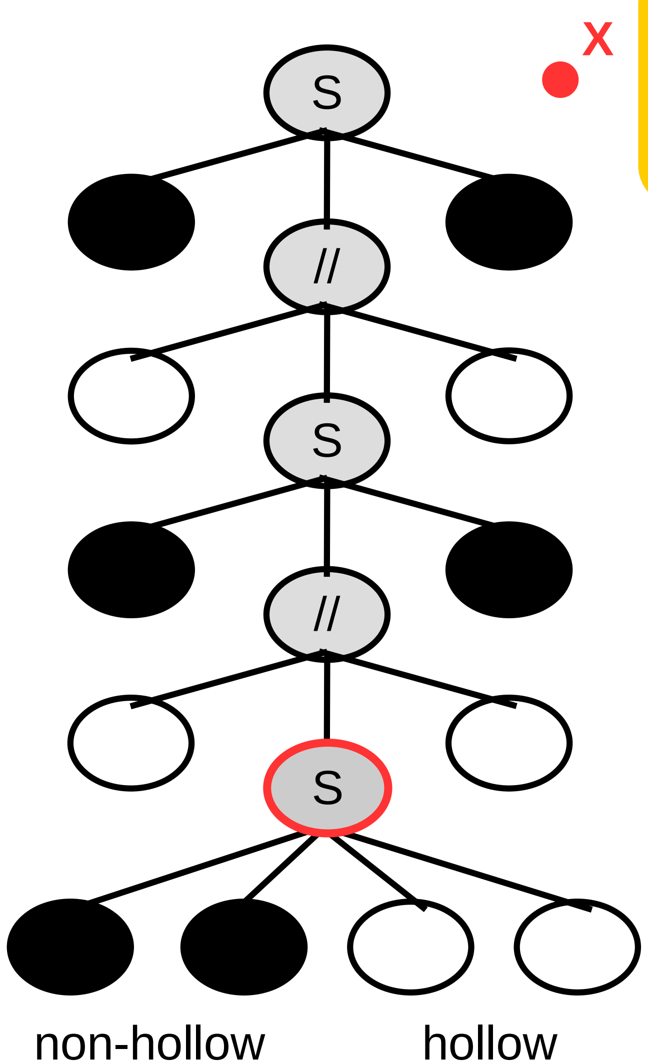
1) choose one eligible node ***u***

2) make the non-hollow children of u become ***full*** (leave the others ***hollow***)

3) for each ***series ancestor*** v of u, make all its children (but one) ***full***

Completion anchored at u

In our algorithm : G+x is not a cograph



<u>Definition:</u> u is an ***eligible node*** Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

<u>Proceed as follows :</u>

1) choose one eligible node ***u***

2) make the non-hollow children of u become ***full*** (leave the others ***hollow***)

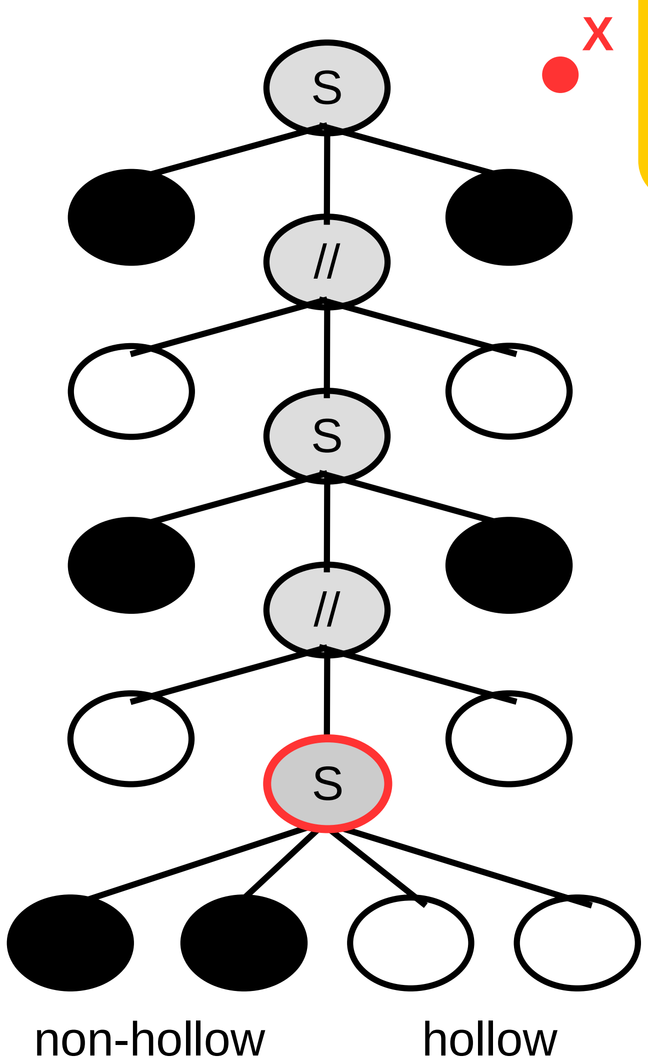
3) for each ***series ancestor*** v of u, make all its children (but one) ***full***

→ **you obtain a cograph completion of G+x**

called the ***completion anchored at u***

Completion anchored at u

In our algorithm : G+x is not a cograph



<u>Definition:</u> u is an ***eligible node*** Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

<u>Proceed as follows :</u>

1) choose one eligible node ***u***

2) make the non-hollow children of u become ***full*** (leave the others ***hollow***)

3) for each ***series ancestor*** v of u, make all its children (but one) ***full***

→ **you obtain a cograph completion of G+x**

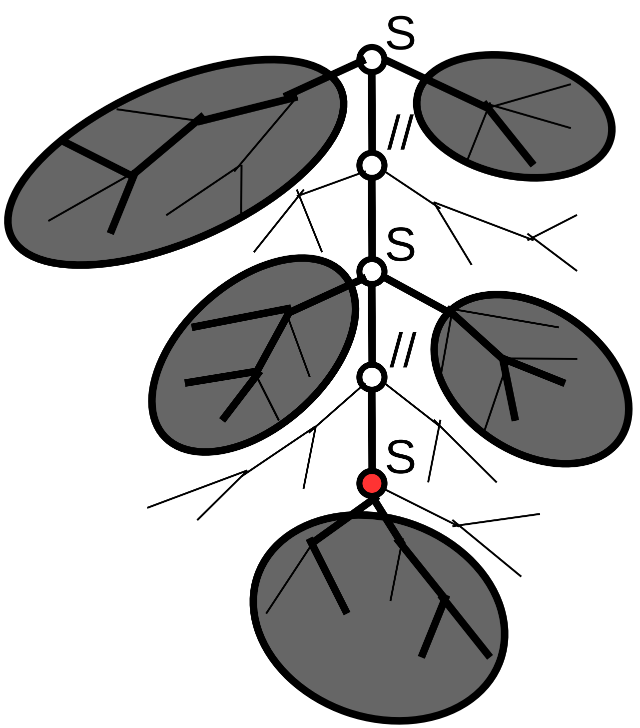
called the ***completion anchored at u***

<u>Question:</u> Is it minimal ?

→ **We have a characterization for this**

First algorithm : O(n+m')

- Search the tree bottom up from the leaves adjacent to x
- Find the eligible nodes that satisfy the characterization



Note : we search only non-hollow nodes

Complexity : O(d')

[LMP 10]

- Choose one u of ***minimum cost*** and update the data structure by running **[CPS 81]**'s algorithm.

Complexity : O(d') for one incremental step
O(n+m') for the whole algorithm

Completion algorithms

Second algorithm: $O(n + m \log^2 n)$

Why is O(n+m') not necessarily optimal?

- No reason to use adjacency lists to encode the output

→ there is an O(n) space representation of cographs

Why is O(n+m') not necessarily optimal?

- No reason to use adjacency lists to encode the output

→ there is an O(n) space representation of cographs

- What is the expected number of edges m' in a cograph completion?
  - If the input G has the vertex-expansion property, then G' has $O(n^2)$ edges
  - Random graphs with fixed average degree, **O(n) edges**, have the expansion property with high probability

→ **In practice, O(n+m') ~ $O(n^2)$**

→ **We achieve O(n+m $\log^2$n) time**

Why is O(n+m’) not necessarily optimal?

- No reason to use adjacency lists to encode the output

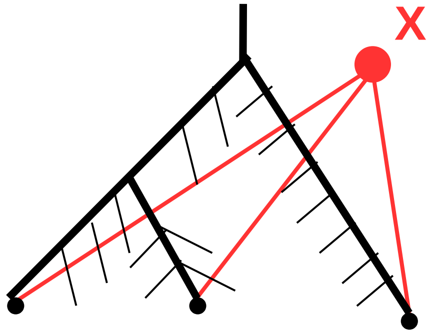
→ there is an O(n) space representation of cographs

- What is the expected number of edges m’ in a cograph completion?
  - If the input G has the vertex-expansion property, then G’ has $O(n^2)$ edges
  - Random graphs with fixed average degree, **O(n) edges**, have the expansion property with high probability

→ **In practice, $O(n+m') \sim O(n^2)$**

→ **We achieve $O(n+m \log^2 n)$ time**

- Where is the room for improvement of the complexity?



A ***constant*** number of neighbours of x can force to search an ***Ω(n)*** part of the co tree

Second algorithm : $O(n + m \log^2 n)$

- Note: we abandon the minimum incremental → ***only minimal***

- we use a dynamic data-structure for *lowest ancestor queries*
  [Sleator, Tarjan 1983]
  - In O(log n) time: w=lca(u,v) and w_u the child of w that is an ancestor of u
  - Update the structure in O(log n) time under elementary tree modifications

- we use *ordered lists*
  [Dietz, Sleator 1987]
  - In O(1) time: order between two elements in the list
  - Update the structure in O(1) time under deletion and insertion of an element

Second algorithm : O(n + m log²n)

<u>Our goal :</u> determine the ***lowest*** eligible, non-hollow and non-forced nodes

→ **minimal completion**

Second algorithm : $O(n + m \log^2 n)$

Our goal : determine the ***lowest*** eligible, non-hollow and non-forced nodes

→ **minimal completion**

- Lowest eligible nodes

  → highest parallel nodes with ≥2 non-hollow children

Second algorithm : $O(n + m \log^2 n)$

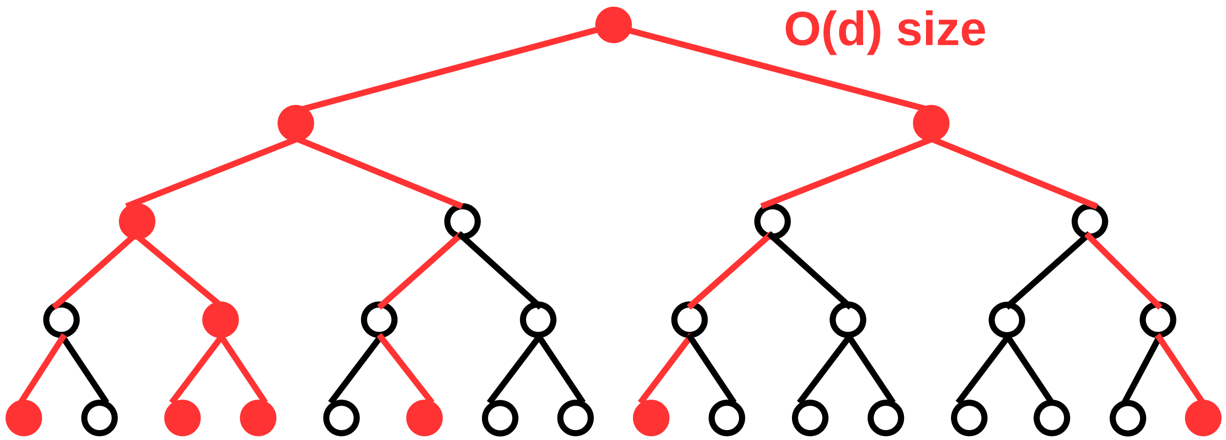
<u>Our goal :</u> determine the ***lowest*** eligible, non-hollow and non-forced nodes

➡ **minimal completion**

- Lowest eligible nodes

➡ highest parallel nodes with ≥2 non-hollow children

  - build T' : the subtree of lowest common ancestors of neighburs of x
  - Keep the highest parallel nodes in T'



Second algorithm : $O(n + m \log^2 n)$

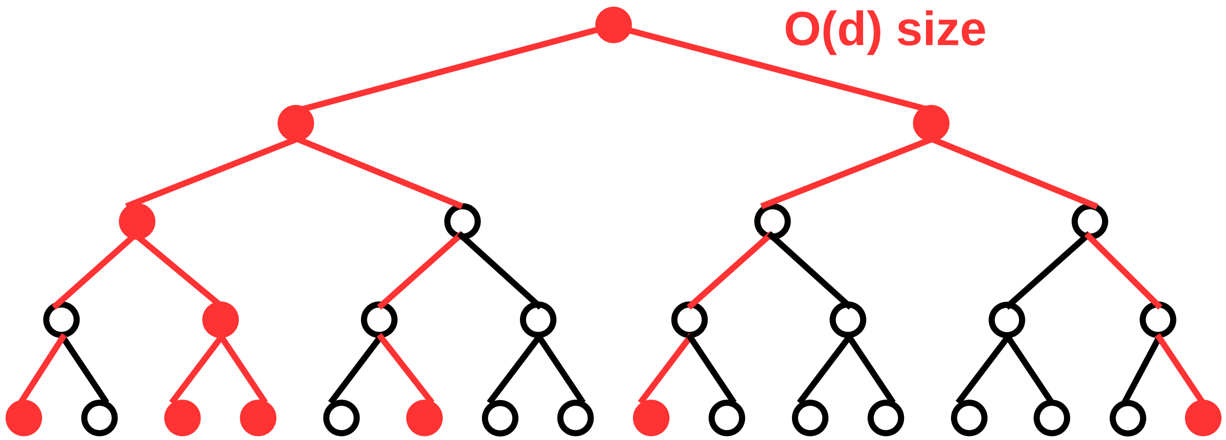
<u>Our goal :</u> determine the ***lowest*** eligible, non-hollow and non-forced nodes

→ **minimal completion**

- Lowest eligible nodes

  → highest parallel nodes with ≥2 non-hollow children

  - build T' : the subtree of lowest common ancestors of neighburs of x
  - Keep the highest parallel nodes in T'



1) sort neighbours of x from left to right : $O(d \log^2 n)$ time

Second algorithm : $O(n + m \log^2 n)$

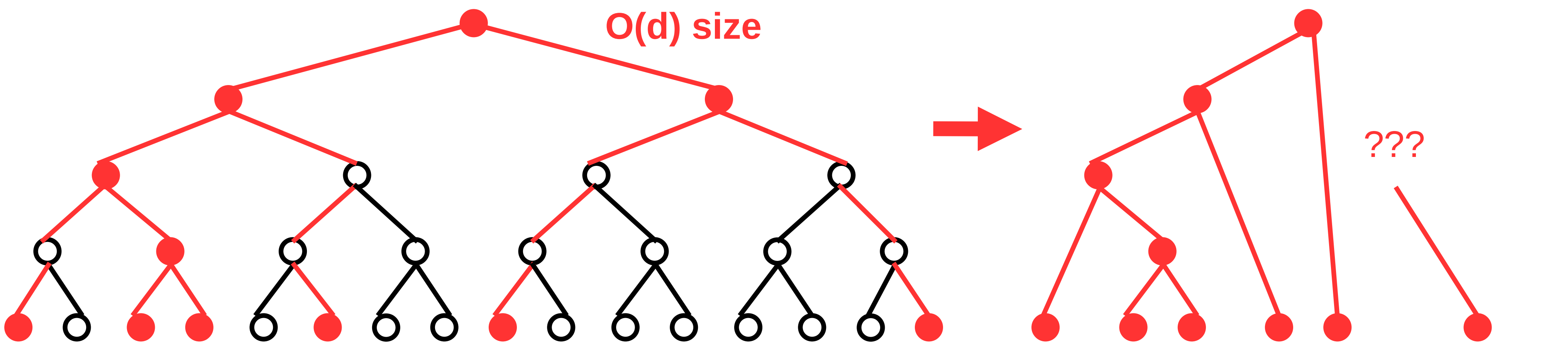
<u>Our goal :</u> determine the ***lowest*** eligible, non-hollow and non-forced nodes

→ **minimal completion**

- Lowest eligible nodes

  → highest parallel nodes with ≥2 non-hollow children

  - build T' : the subtree of lowest common ancestors of neighburs of x
  - Keep the highest parallel nodes in T'



1) sort neighbours of x from left to right : $O(d \log^2 n)$ time

2) insert neighbours one by one
Total :$O(d \log n)$ time

Second algorithm : O(n + m log²n)

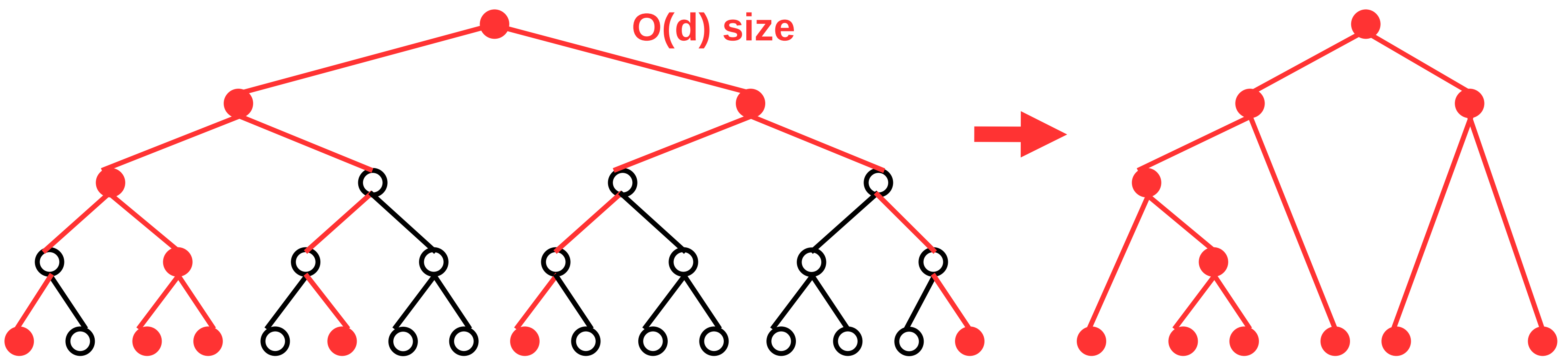
<u>Our goal :</u> determine the ***lowest*** eligible, non-hollow and non-forced nodes

→ **minimal completion**

- Lowest eligible nodes

→ highest parallel nodes with ≥2 non-hollow children

  - build T' : the subtree of lowest common ancestors of neighburs of x
  - Keep the highest parallel nodes in T'



1) sort neighbours of x from left to right : O(d $\log^2$n) time

2) insert neighbours one by one
Total :O(d log n) time

Second algorithm : $O(n + m \log^2 n)$

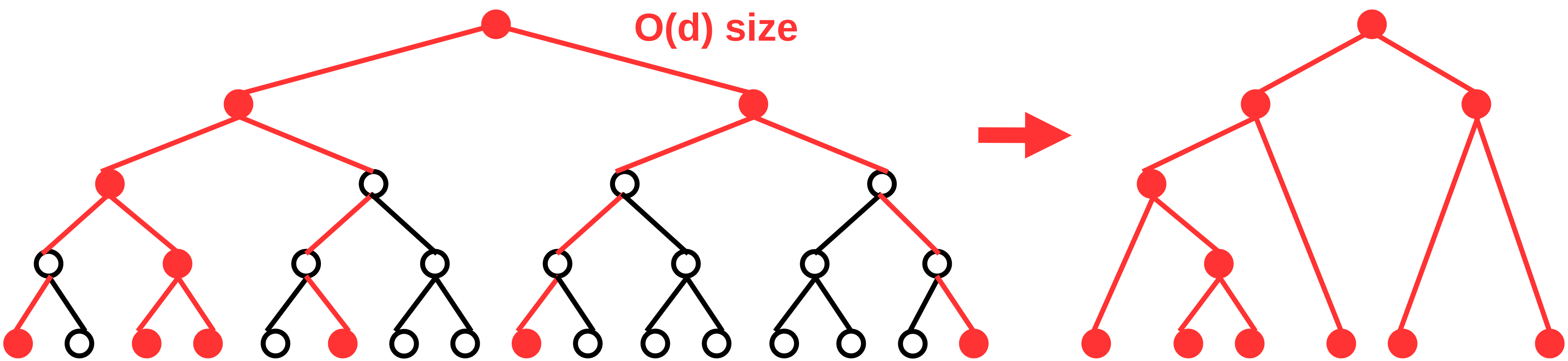
Our goal : determine the ***lowest*** eligible, non-hollow and non-forced nodes

→ **minimal completion**

- Lowest eligible nodes

  → highest parallel nodes with ≥2 non-hollow children

  - build T' : the subtree of lowest common ancestors of neighburs of x
  - Keep the highest parallel nodes in T'



1) sort neighbours of x from left to right : $O(d \log^2 n)$ time

2) insert neighbours one by one
Total :$O(d \log n)$ time

- Non-forced condition
  - Find the lowest non-forced node above each node of W (grand-parent)

Complexity : $O(d \log^2 n)$ for one incremental step
$O(n+m \log^2 n)$ for the whole algorithm

Editing algorithm

O(n + m) time

Algorithm for cograph editing

- Editing: use both additions and deletions of edges
- Minimal for inclusion
- **Linear time: O(n+m)**
- Additional feature: **minimum editing at each incremental step**
- **number of edits returned is ≤ m**

The local incremental approach

- Vertices are processed one by one
- Only edges ***incident to x*** are modified

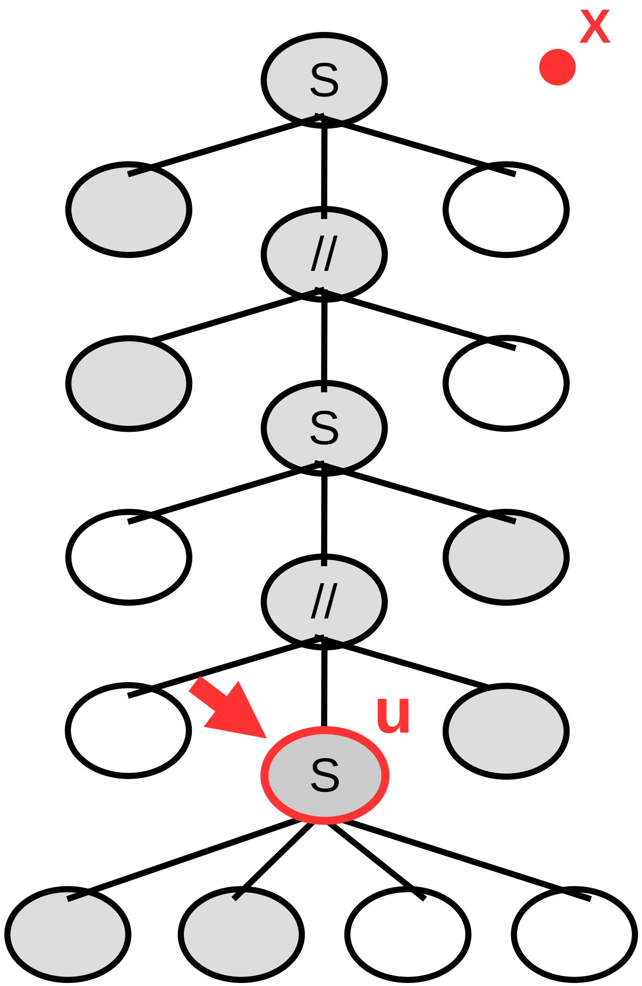
Always possible when:

- The class is ***hereditary***
- Contains no maximal element for induced subgraph relationship

Our goal : O(d) time complexity at each incremental step

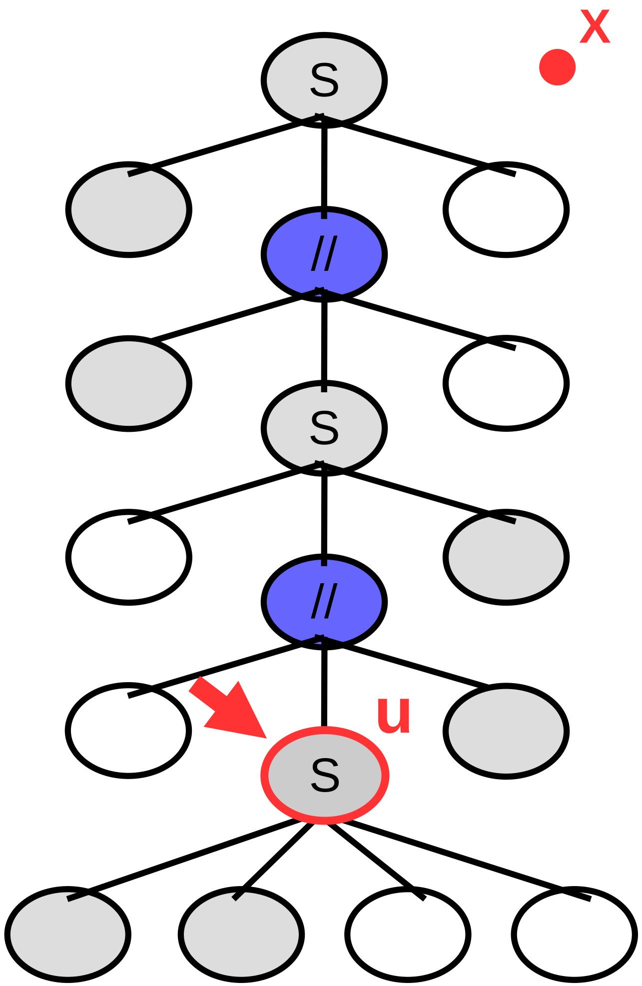
Editing anchored at u

In our algorithm : G+x is not a cograph



Editing anchored at u

In our algorithm : G+x is not a cograph

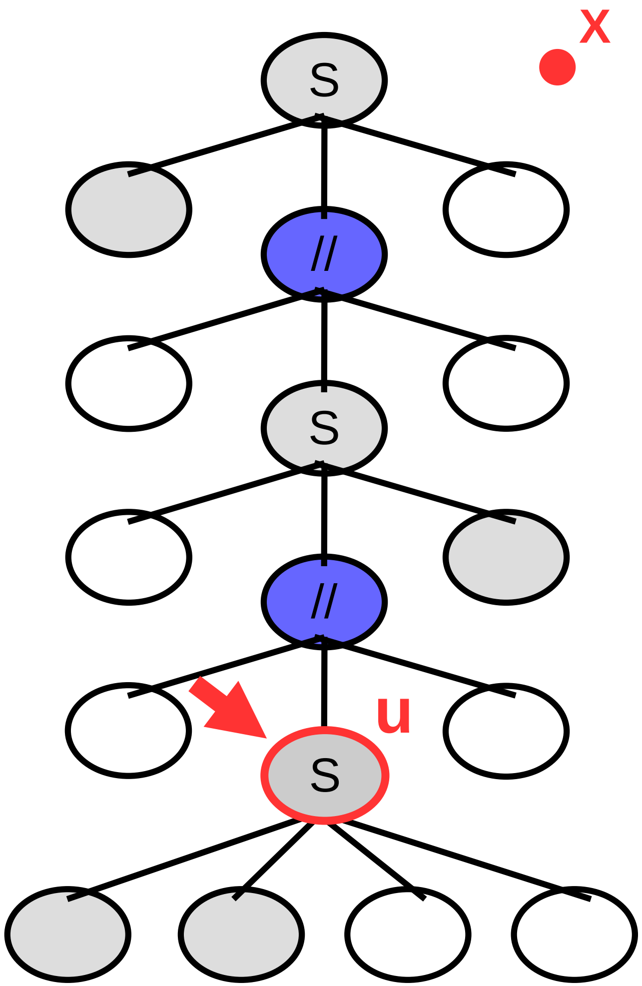


Proceed as follows :

1) for each ***parallel ancestor*** of u, make all its children (but one) ***hollow***

Editing anchored at u

In our algorithm : G+x is not a cograph

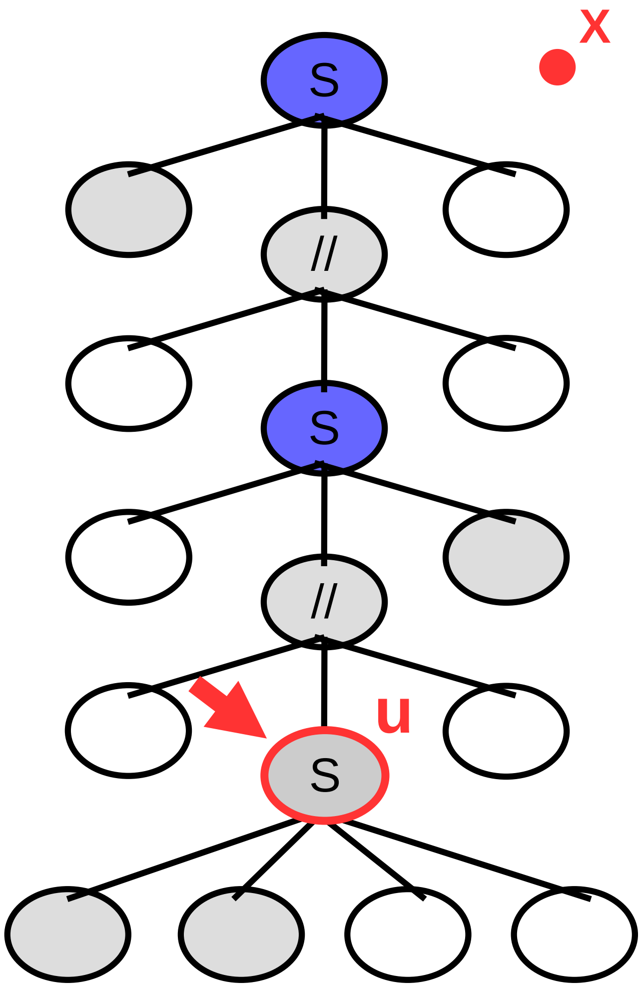


Proceed as follows :

1) for each ***parallel ancestor*** of u, make all its children (but one) ***hollow***

Editing anchored at u

In our algorithm : G+x is not a cograph



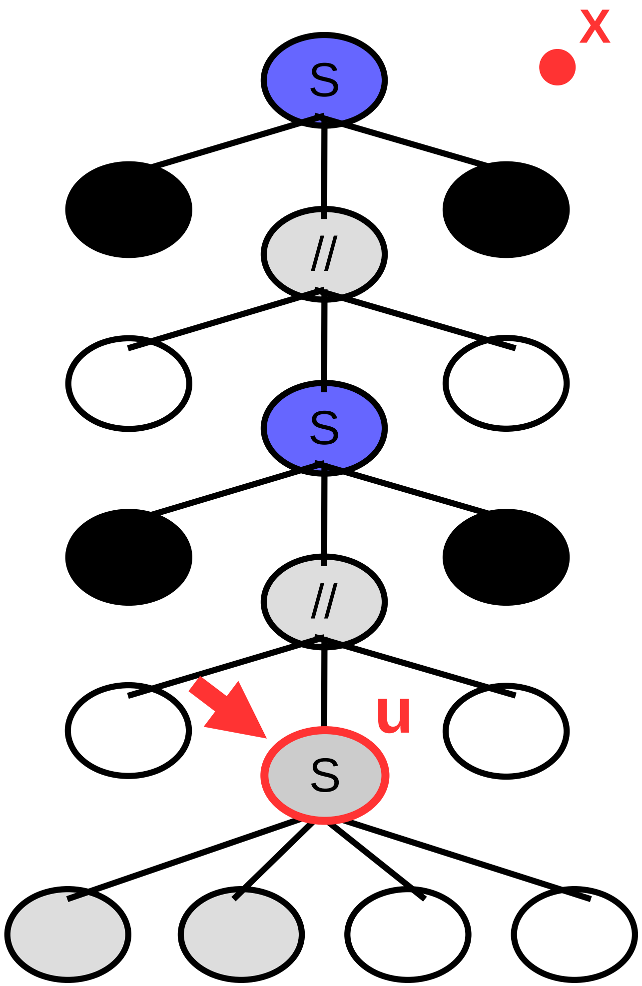
Proceed as follows :

1) for each ***parallel ancestor*** of u, make all its children (but one) ***hollow***

2) for each ***series ancestor*** of u, make all its children (but one) ***full***

Editing anchored at u

In our algorithm : G+x is not a cograph



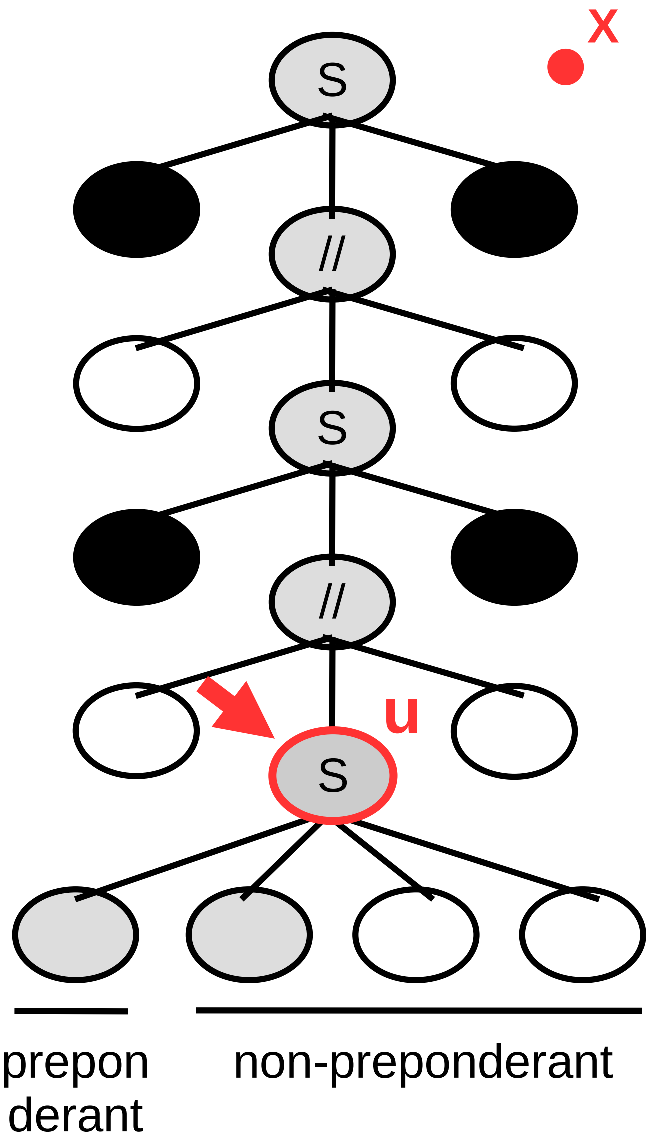
Proceed as follows :

1) for each ***parallel ancestor*** of u, make all its children (but one) ***hollow***

2) for each ***series ancestor*** of u, make all its children (but one) ***full***

Editing anchored at u

In our algorithm : G+x is not a cograph



Proceed as follows :

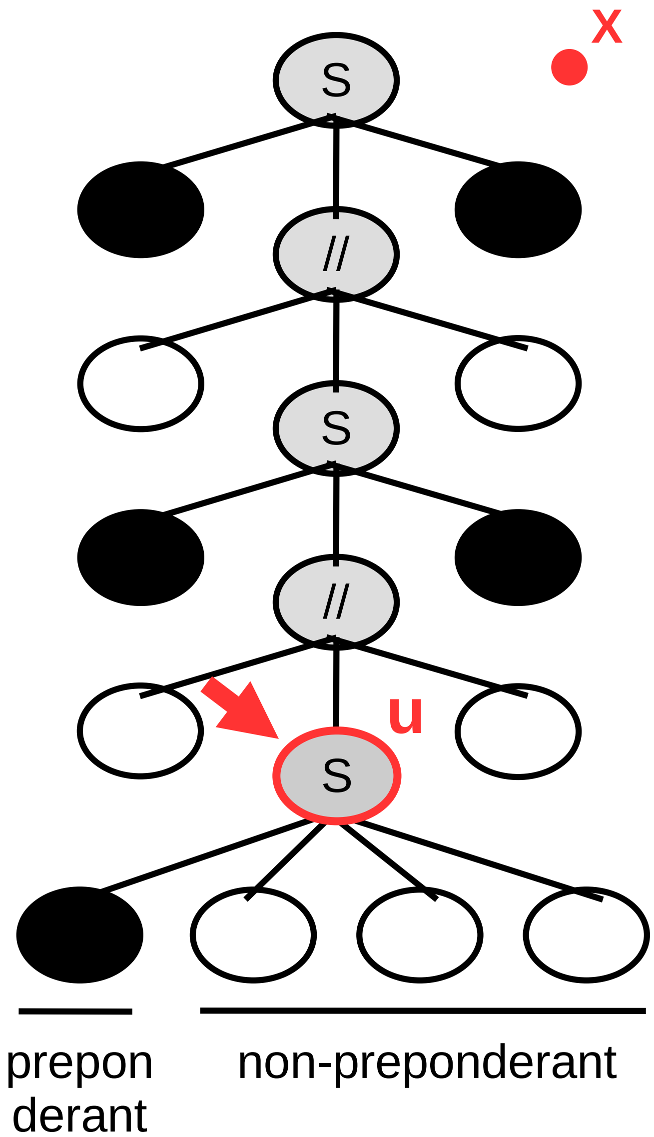
1) for each ***parallel ancestor*** of u, make all its children (but one) ***hollow***

2) for each ***series ancestor*** of u, make all its children (but one) ***full***

3) make the preponderant children of u become ***full*** and make the non-preponderant ones ***hollow***

Editing anchored at u

In our algorithm : G+x is not a cograph



Proceed as follows :

1) for each ***parallel ancestor*** of u, make all its children (but one) ***hollow***

2) for each ***series ancestor*** of u, make all its children (but one) ***full***

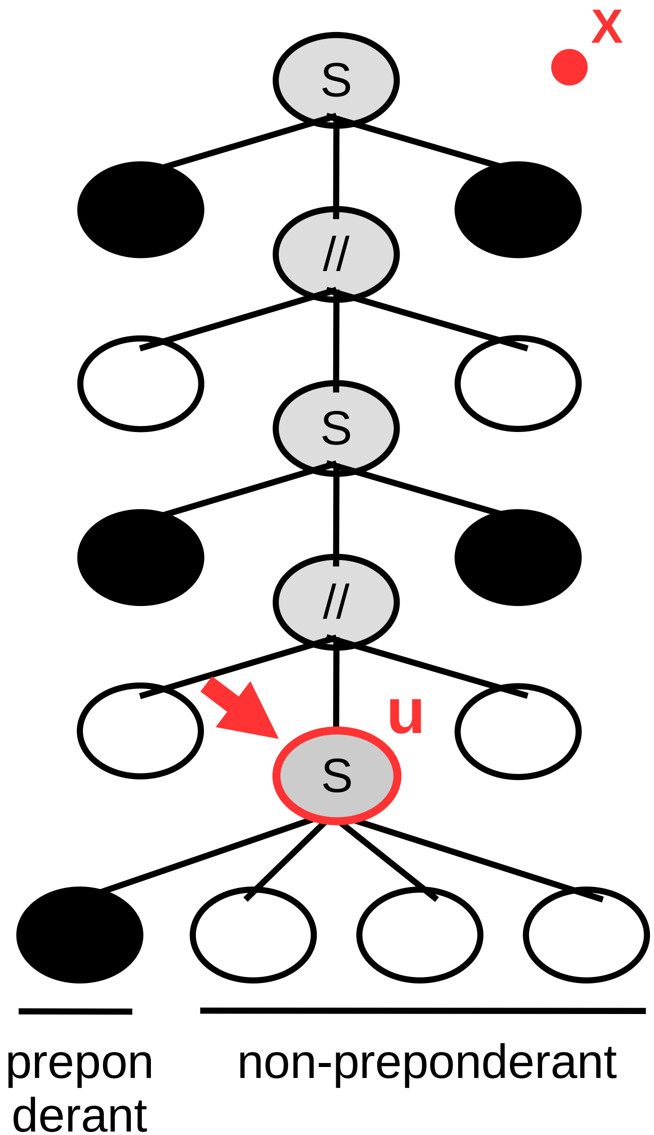
3) make the preponderant children of u become ***full*** and make the non-preponderant ones ***hollow***

→ **you obtain a cograph editing of G+x**

called the ***editing anchored at u***

Editing anchored at u

In our algorithm : G+x is not a cograph



Proceed as follows :

1) for each ***parallel ancestor*** of u, make all its children (but one) ***hollow***

2) for each ***series ancestor*** of u, make all its children (but one) ***full***

3) make the preponderant children of u become ***full*** and make the non-preponderant ones ***hollow***

→ **you obtain a cograph editing of G+x**

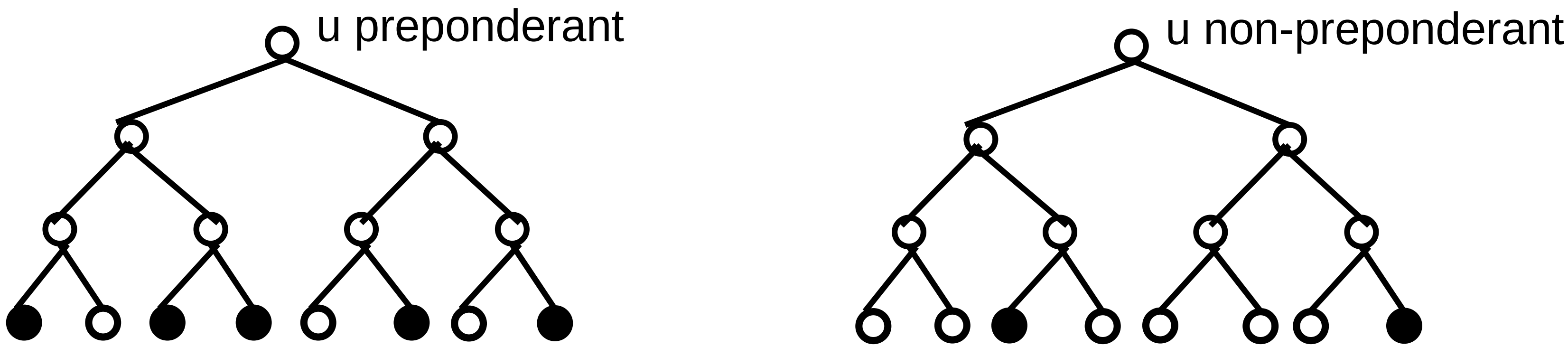
called the ***editing anchored at u***

Question: Is it minimal? minimum ?

→ **O(n) time algorithm trying all possible nodes of the cotree**

Maximal preponderant nodes

Def.: u is preponderant iff the subtree of u contains more neighbours of x than non-neighbours of x



Def.: u is maximal preponderant iff u is preponderant and no ancestor of u is.

Cor. [CPS81]: the insertion node of a minimum editing has a preponderant child

➡ **The insertion node is either in the subtree of some maximal preponderant node or is the parent of some maximal preponderant node**

Only O(d) candidates for the insertion node

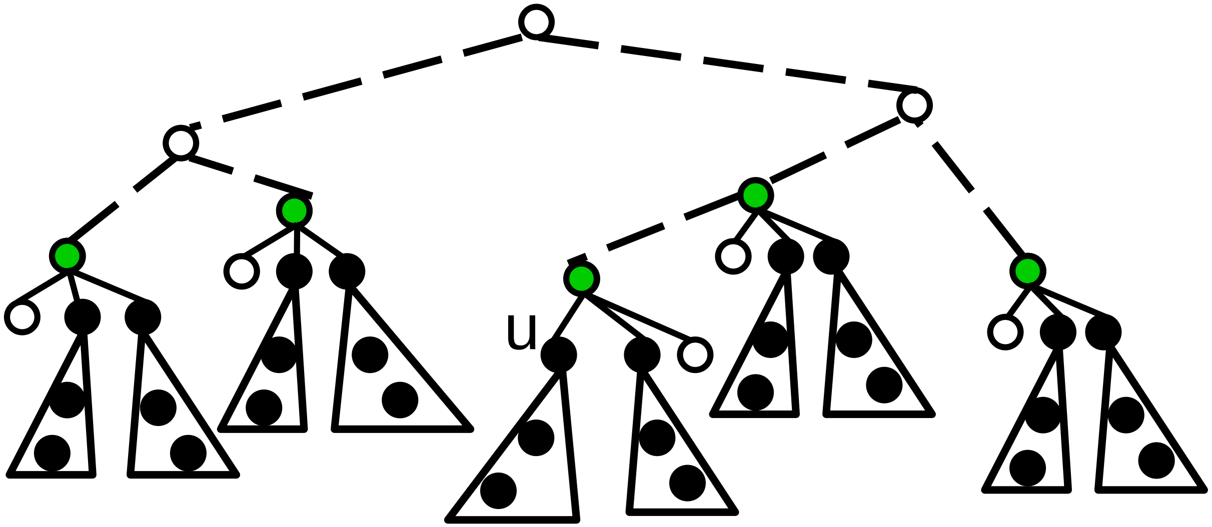
Outline of the algorithm

1) compute all maximal *preponderant* nodes (and their parents)

2) for each maximal preponderant node u, determine the minimum editing anchored in its subtree or at its parent

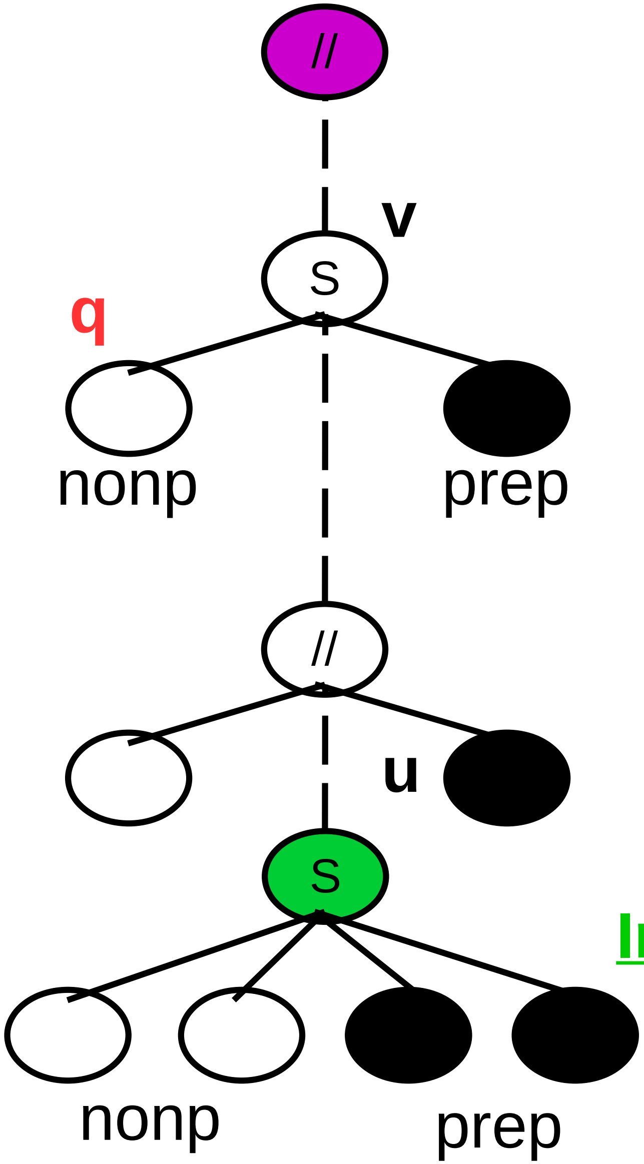
➡ O(n) algo applied on a subcotree where n=O(d)

3) keep the minimum editing among all the editings found for each maximal preponderant node u : need to compute ***cost-above(u)***



Principle of the bottom-up search

Obs.: we need the cost of the editing anchored at u only if it is less than the cost of the delete-all editing

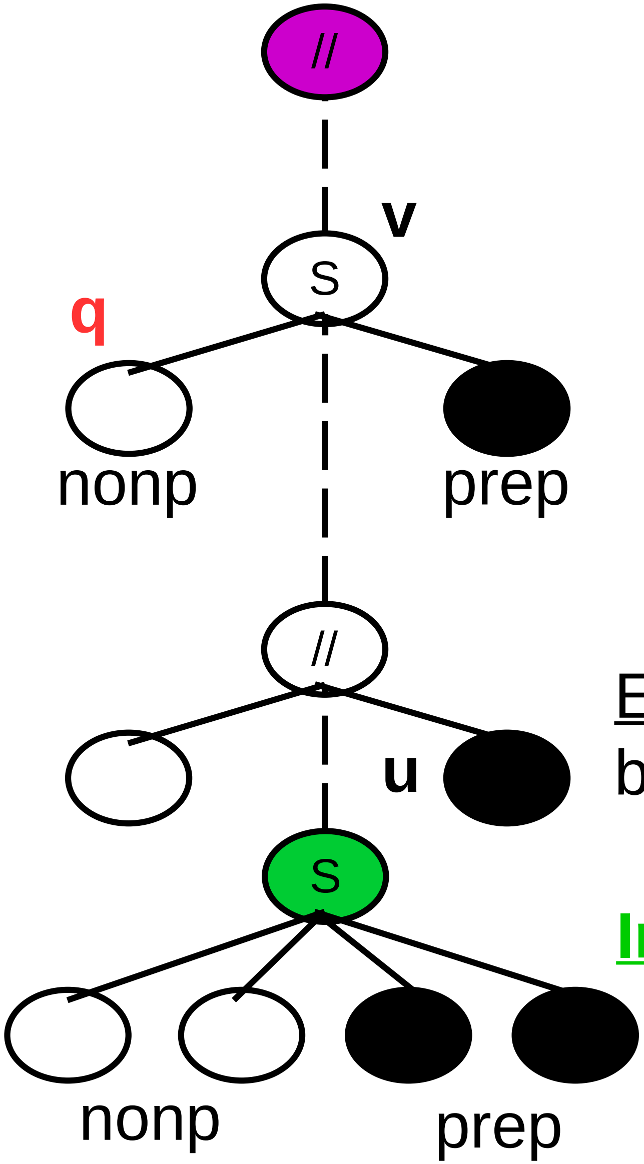


Initialisation : $bud(u)=B_{prep}(u)-W_{prep}(u)$

$= exc(u)$

Principle of the bottom-up search

Obs.: we need the cost of the editing anchored at u only if it is less than the cost of the delete-all editing



Encounter a parallel node:

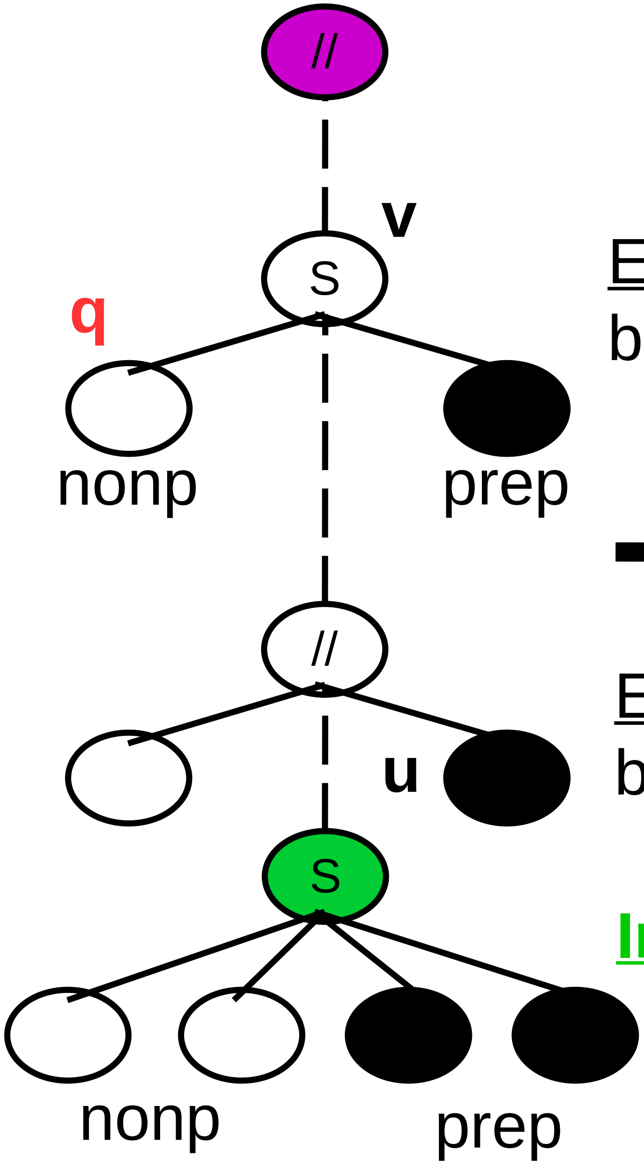
bud(u) unchanged

Initialisation : $bud(u)=B_{prep}(u)-W_{prep}(u)$

$= exc(u)$

Principle of the bottom-up search

Obs.: we need the cost of the editing anchored at u only if it is less than the cost of the delete-all editing



Encounter a series node v:

$$\text{bud}(u) \leftarrow \underbrace{\text{bud}(u) + B_{prep}(v) - W_{prep}(v)}_{= \textbf{bud}} + B_{nonp}(v) - W_{nonp}(v)$$

➡ Routine SearchTree(**q**,**bud**)

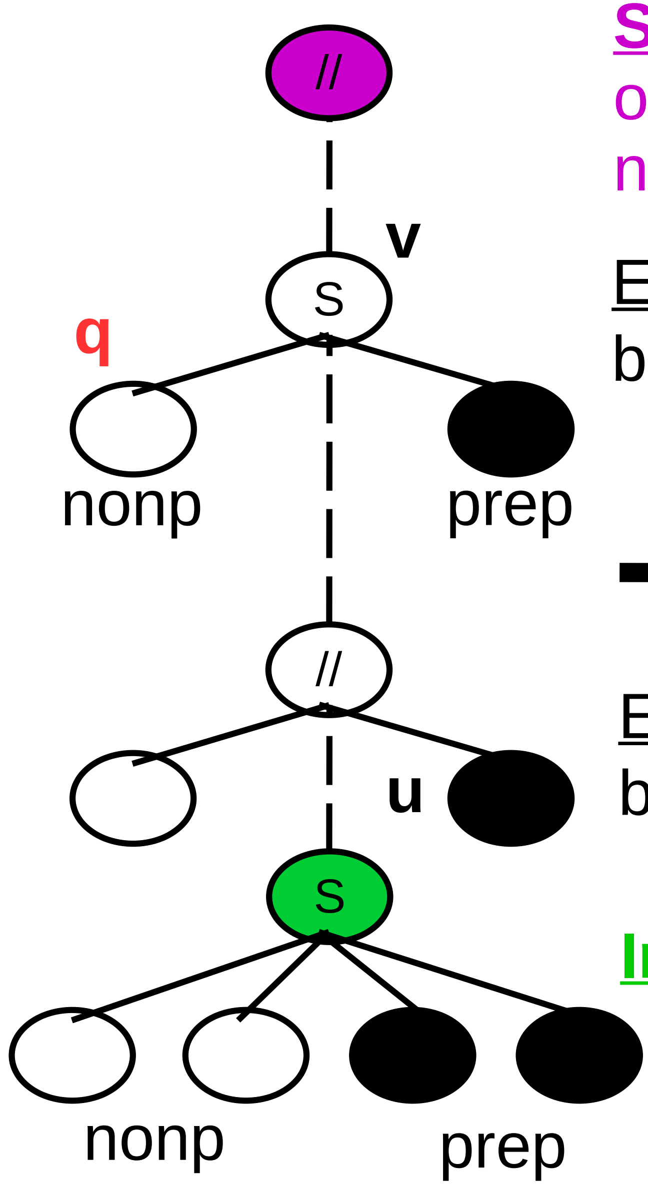
Encounter a parallel node:

bud(u) unchanged

Initialisation : $\text{bud}(u) = \underbrace{B_{prep}(u) - W_{prep}(u)}_{= \text{exc}(u)}$

Principle of the bottom-up search

Obs.: we need the cost of the editing anchored at u only if it is less than the cost of the delete-all editing



Stop when either the budget becomes negative or when the search reaches the root with non-negative budget → deduce cost-above(u)

Encounter a series node v:

$$bud(u) \leftarrow \underbrace{bud(u) + B_{prep}(v) - W_{prep}(v)}_{= \textbf{bud}} + B_{nonp}(v) - W_{nonp}(v)$$

→ Routine SearchTree(**q**,**bud**)

Encounter a parallel node:

bud(u) unchanged

Initialisation : $bud(u) = \underbrace{B_{prep}(u) - W_{prep}(u)}_{= exc(u)}$

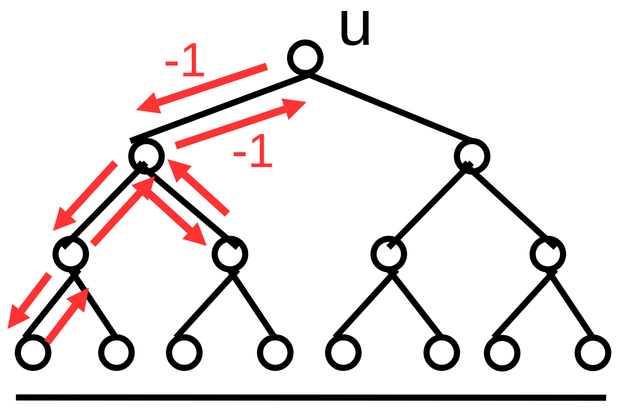
Routine SearchTree(u,s)

- Makes a **DFS limited by a *ttl*** and counts the difference between black and white leaves in *cpt*
  - Initially, *ttl* ← 2+5s and *cpt* ← s
  - *ttl* is decreased when an edge is traversed
  - DFS stops when *ttl*=-1

- Main property:

W(u)-B(u) ≤ s iff Search-tree(u, s) searches the entire subtree of u and ends with a value cpt ≥ 0.
Complexity : **O(min{s,W(u)-B(u)})**

why *ttl* ← 2+5s ?

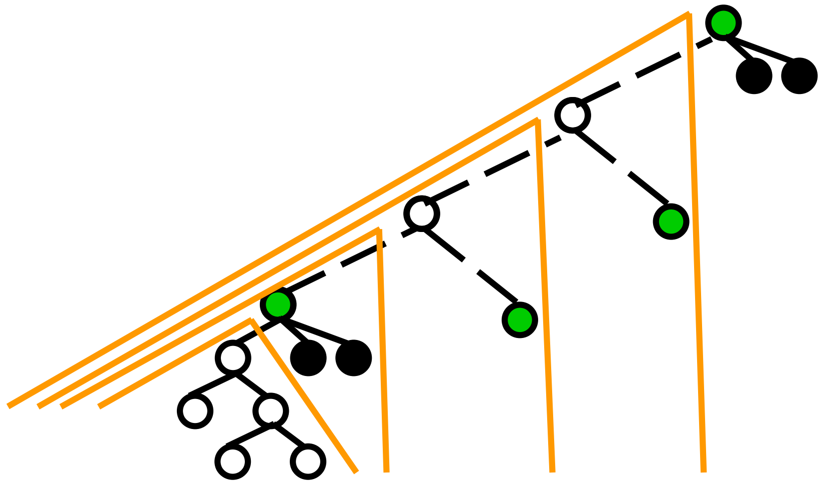


4s-2 edge traversals

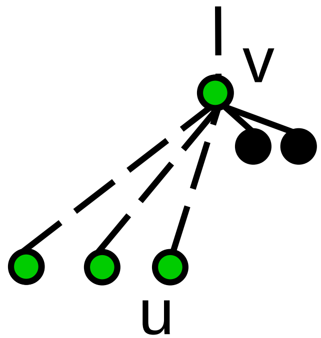
s white leaves

Two threats to the complexity

- Searching repeatedly the same part of the tree with the same budget



- Using repeatedly the same budget in the bottom-up search



$\mathrm{bud}(u) \leftarrow \mathrm{bud}(u)\ \mathbf{+B_{prep}(v)\text{-}W_{prep}(v)}\ +B_{nonp}(v)\text{-}W_{nonp}(v)$

Some open algorithmic questions

Inclusion-minimal cograph editing in linear time

- minimum at each incremental step
- at most m edits at the end

- Showing that minimal cograph completion is ***not*** solvable in linear time

  $O(n+m\log^2 n)$ from [Crespelle,Lokshtanov,Phan, Thierry 2020]

- Inclusion-minimal editing for other graph classes, in linear time?

Complex networks as almost cographs?

Cograph edition of real-world graphs

35 real-world graphs

+

8 random graphs

Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

Cograph edition of real-world graphs

RESULTS

35 real-world graphs

+

8 random graphs

Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

- Some networks are very close from cographs

Cograph edition of real-world graphs

Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

35 real-world graphs

+

8 random graphs

RESULTS

- Some networks are very close from cographs
- Random graphs are never

Cograph edition of real-world graphs

35 real-world graphs

+

8 random graphs

Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

RESULTS

- Some networks are very close from cographs
- Random graphs are never
- A wide range of proximity : 12% to 93%

Cograph edition of real-world graphs

35 real-world graphs

+

8 random graphs

Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

RESULTS

- Some networks are very close from cographs
- Random graphs are never
- A wide range of proximity : 12% to 93%
- The proximity with cographs highly depends on the real-world context

Cograph edition of real-world graphs

Close to cographs

- www
- software

Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

- The proximity with cographs highly depends on the real-world context

Cograph edition of real-world graphs

Not close not far

internet

road

Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

- The proximity with cographs highly depends on the real-world context

Cograph edition of real-world graphs

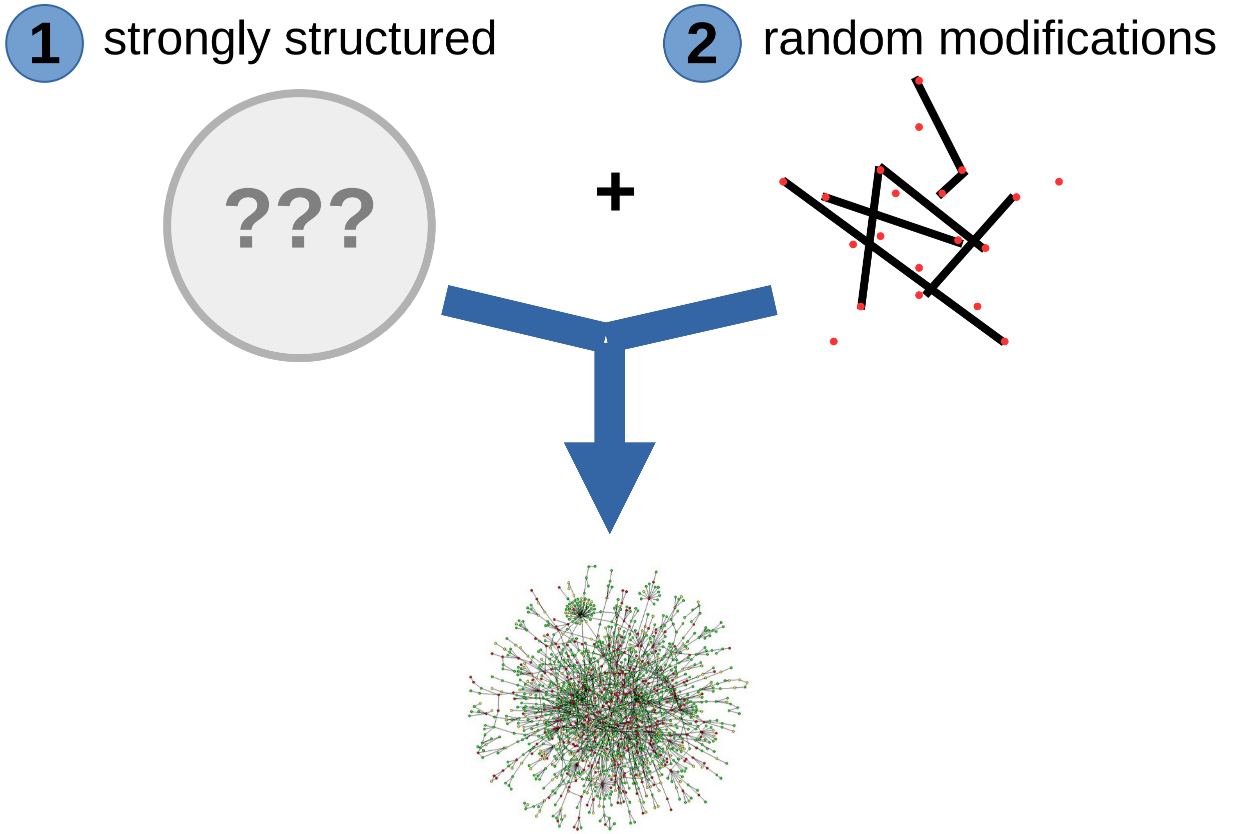
Context	Network	n	m	d°	%mod
WWW	in-2004	1 148 875	12 281 937	21.4	12 %
WWW	cnr-2000	227 058	2 187 201	19.3	19 %
PROTEIN	reactome	5 973	145 778	48.8	22 %
SOFTWARE	jdk	6 434	53 658	16.7	29 %
SOFTWARE	jung-j	6 120	50 290	16.4	29 %
WWW	eu-2005	835 044	15 718 784	37.7	29 %
CO-AUTHOR	ca-GrQc	4 158	13 422	6.5	34 %
CO-AUTHOR	ca-HepPh	11 204	117 619	21.0	34 %
SPECIES	foodweb	183	2 434	26.6	43 %
CO-AUTHOR	dblp	317 080	1 049 866	6.6	45 %
WORD-REL.	wordnet	145 145	656 230	9.0	48 %
COMMUNIC.	wiki-Talk	2 388 953	4 656 682	3.9	49 %
CO-SOLD	amazon	334 863	925 872	5.5	49 %
CO-AUTHOR	ca-CondMat	21 363	91 286	8.6	52 %
RANDOM	ER-Gnm_1M-2	796 208	958 827	2.4	52 %
CO-AUTHOR	ca-HepTh	8 638	24 806	5.7	54 %
INTERNET	as2000	6 474	12 572	3.9	54 %
ROAD	roadNet-TX	1 351 137	1 879 201	2.8	54 %
INTERNET	as-caida2007	26 475	53 381	4.0	55 %
CO-AUTHOR	ca-AstroPh	17 903	196 972	22.0	59 %
INTERNET	topology	34 761	107 720	6.2	61 %
RANDOM	ER-Gnm_1M-3	940 987	1 494 643	3.2	63 %
INTERNET	as-skitter	1 694 616	11 094 209	13.1	64 %
CO-OCCUR	bible-names	1 707	9 059	10.6	67 %
PROTEIN	figeys	2 217	6 418	5.8	67 %
CITATION-SCI.	cora	23 166	89 157	7.7	68 %
SOCIAL	youtube	1 134 890	2 987 624	5.3	69 %
CO-ACTOR	actor-col.	374 511	15 014 839	80.2	71 %
P2P-CONNECT.	p2p-Gnutella	62 561	147 878	4.7	71 %
RANDOM	ER-Gnm_1M-4	980 191	1 999 203	4.1	71 %
CITATION-SCI.	citeseer	365 154	1 721 981	9.4	75 %
CITATION-PAT.	cit-Patents	3 764 117	16 511 740	8.8	76 %
SOFTWARE	linux	30 817	213 208	13.8	77 %
SOCIAL	LiveJournal	3 997 962	34 681 189	17.4	78 %
CITATION-SCI.	cit-HepTh	27 400	352 021	25.7	79 %
RANDOM	ER-Gnm_1M-6	997 479	2 999 988	6.0	79 %
CITATION-SCI.	cit-HepPh	34 401	420 784	24.5	81 %
RANDOM	ER-Gnm_1M-8	999 684	3 999 999	8.0	84 %
RANDOM	ER-Gnm_1M-10	999 952	5 000 000	10.0	87 %
RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91 %
SOCIAL	orkut	3 072 441	117 185 083	76.3	91 %
RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	93 %
WORD-REL.	Thesaurus	23 132	297 094	25.7	93 %

Far from cographs

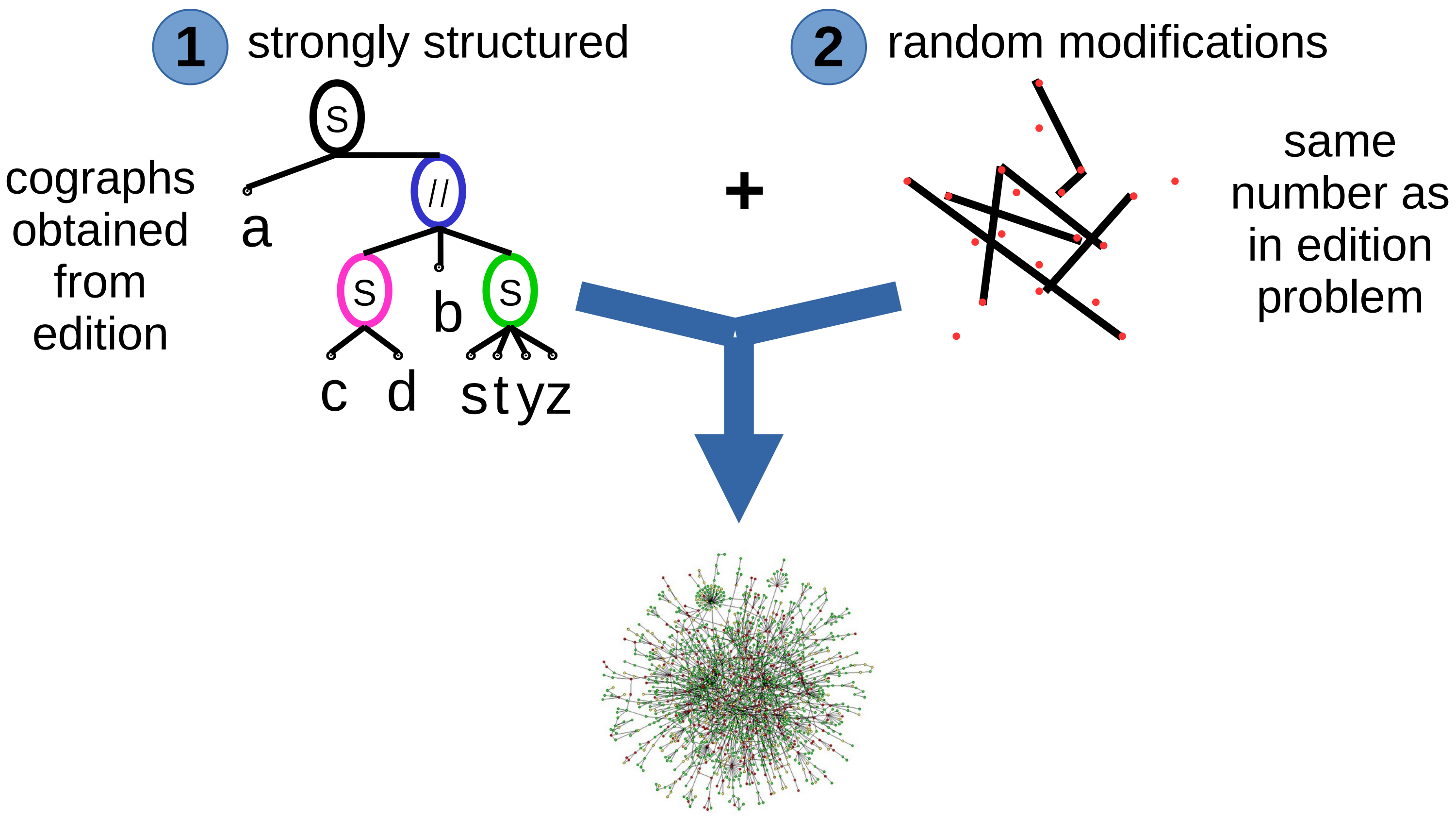
- citation
- social

- The proximity with cographs highly depends on the real-world context

Testing the modelling approach

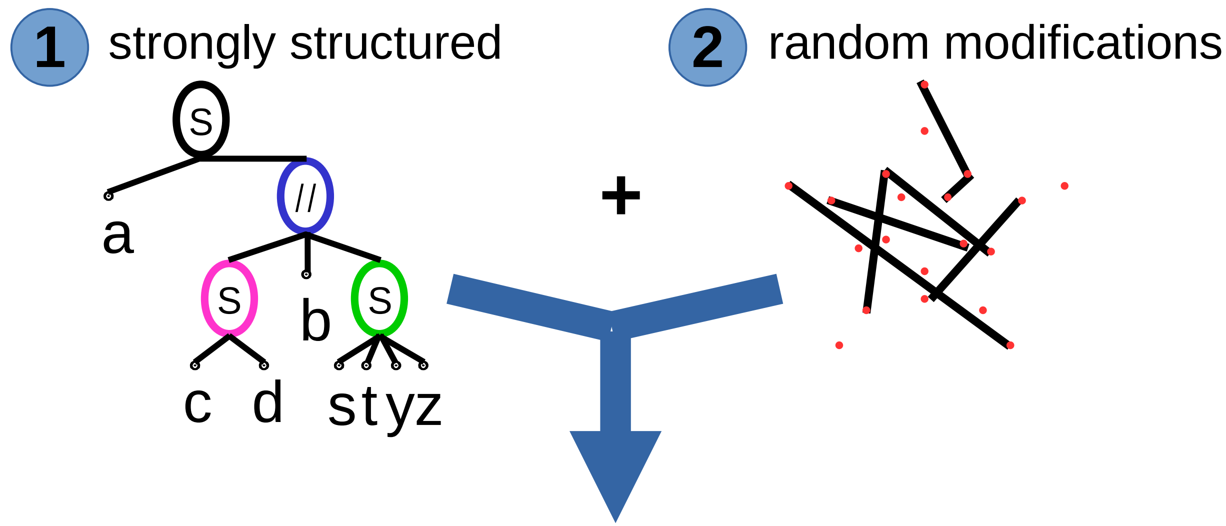


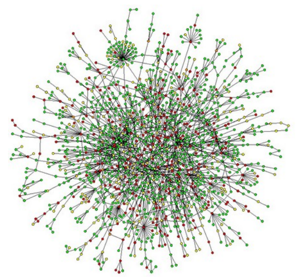
Testing the modelling approach



compare with the original
real-world network

Conclusion

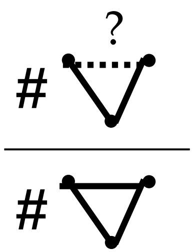


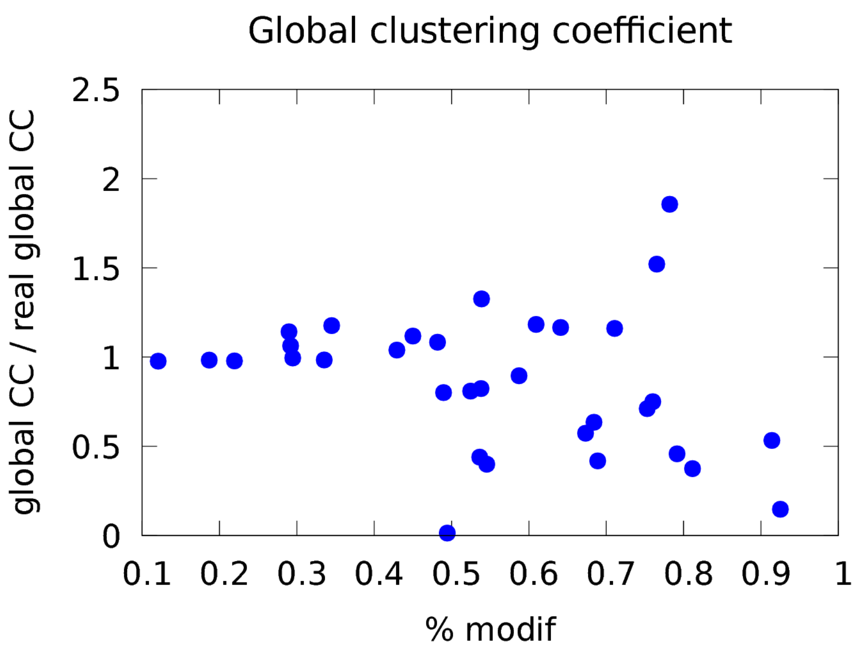


- ✔ global density
- ✔ distances
- ? degree distribution
- ? local density

Results of generation

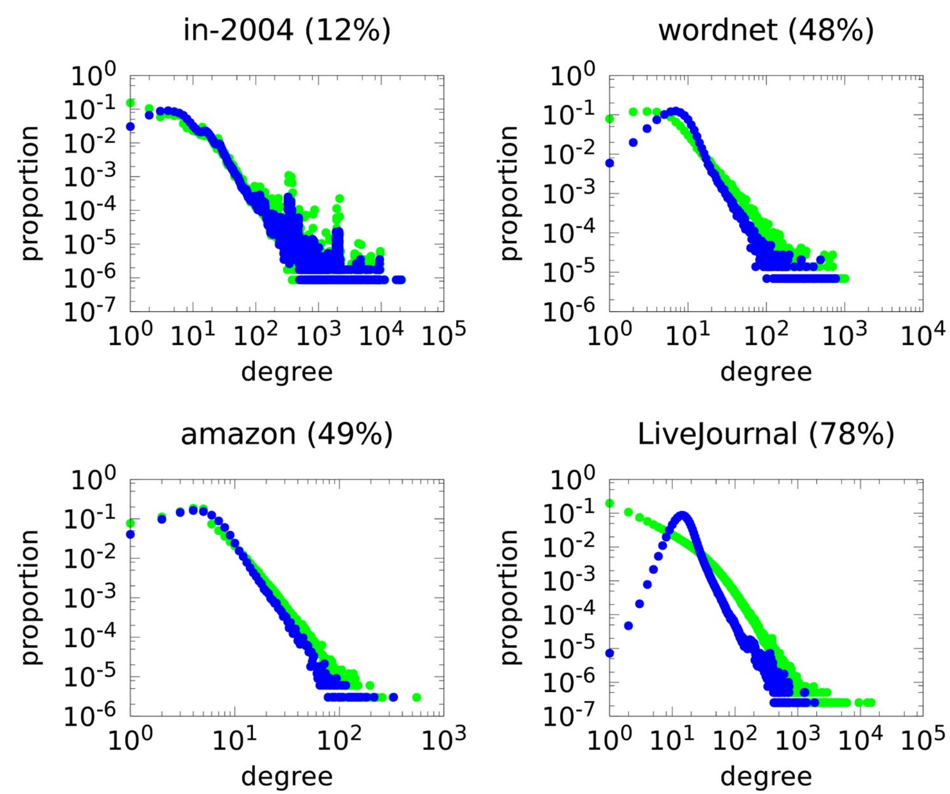
Local density

global CC = 

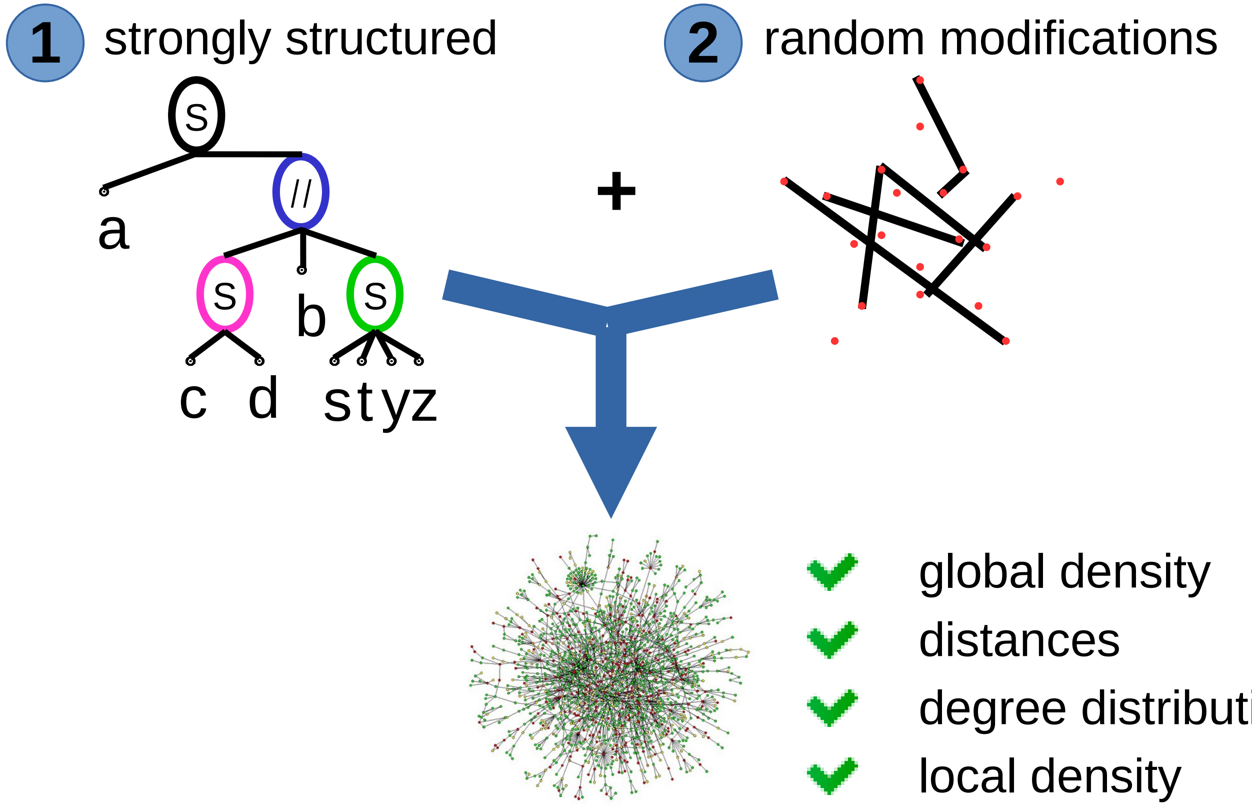


Degree distribution

Almost cograph model
Real distribution

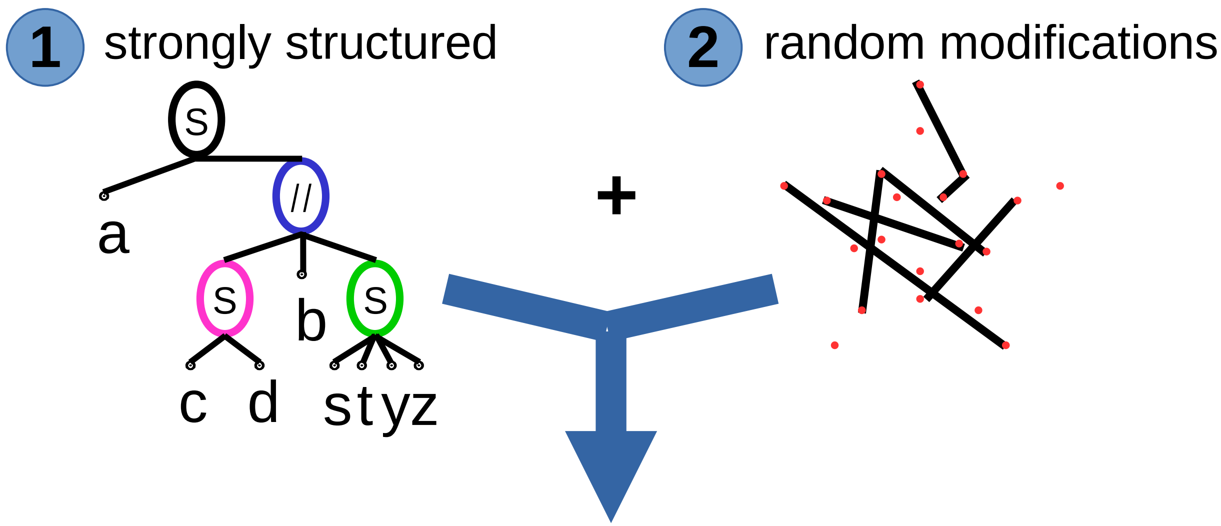


Conclusion

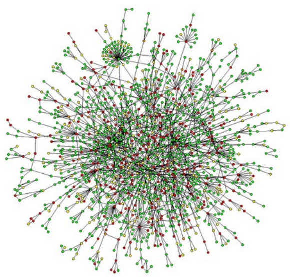


- global density
- distances
- degree distribution
- local density

Conclusion

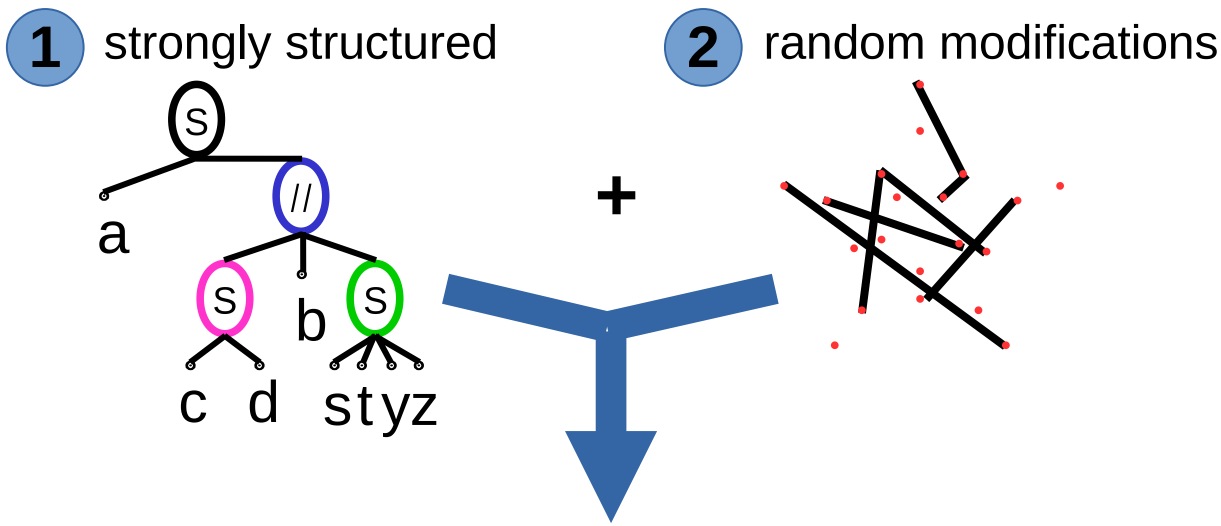


The cograph structure successfully captures these properties

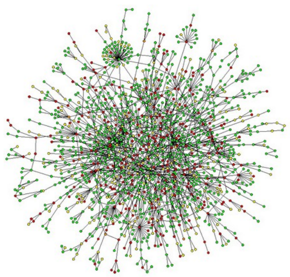


- global density
- distances
- degree distribution
- local density

Conclusion



The cograph structure successfully captures these properties



- global density
- distances
- degree distribution
- local density

<u>To complete the model</u>

- Edit a real-world graph into a cograph
- Generate a similar cotree
- Apply random modifications to the cograph

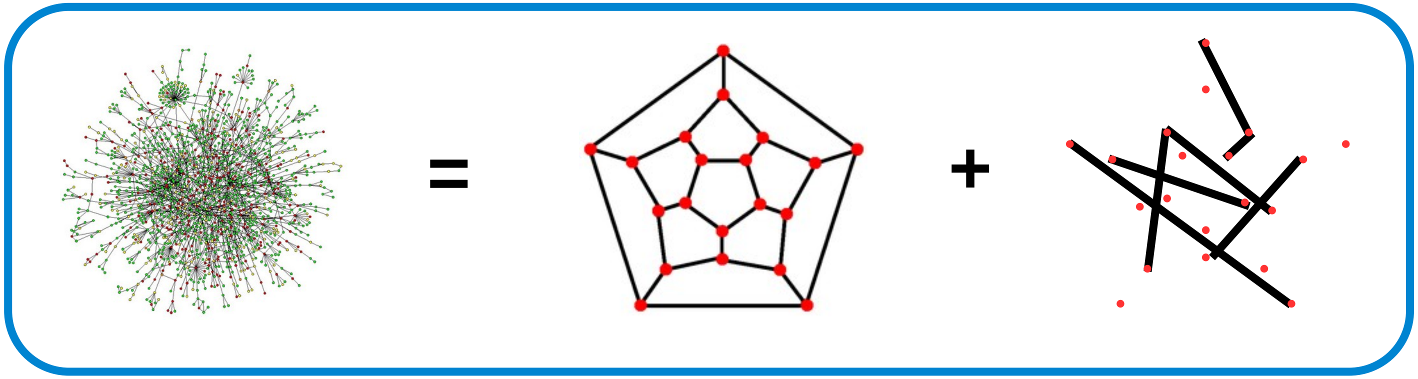
Perspectives

- Complete the modelling approach for cographs

- Consider other graph classes suitable for other kind of networks
  - Chordal graphs $\rightarrow$ social networks, citations
  - Related to planar graphs $\rightarrow$ internet, road networks

- Improve algorithms : complexity and quality
  - edition instead of completion
  - avoid incremental approach

Perspectives



- Modelling
- Efficient encoding : space + query time
- Analysis
  - Global organization
  - Specific roles
- Algorithmic theory of ***almost*** structured graphs

→ Take advantage of the proximity with a strongly structured graph