# Coedit: a tool for minimal cograph edge modification

#### **Christophe Crespelle**

University of Bergen

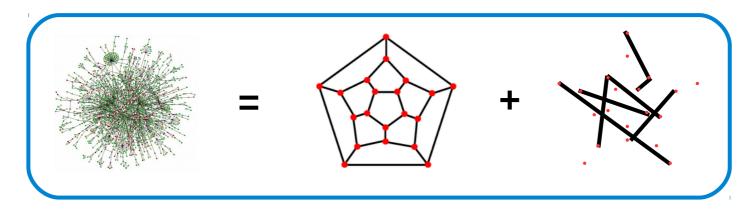
with Daniel Lokshtanov, Thi Ha Duong Phan and Eric Thierry



### **Goal of PROXNET project**

## Representing real-world complex networks as *almost* structured graphs

Complex network = structured graph + noise



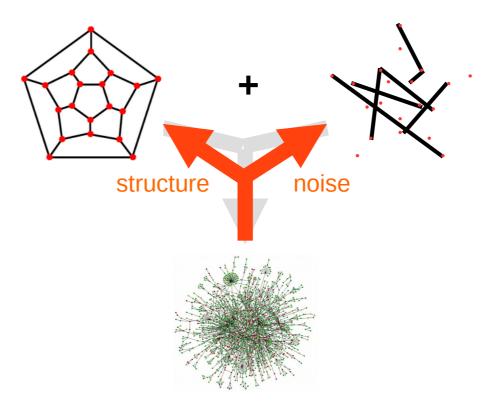
Modelling

- Efficient encoding : space + query time
- Understand their structure (global organisation, specific roles)
- Algorithmic theory of *almost* structured graphs

Take advantage of the proximity with a strongly structured graph

## **Goal of PROXNET project**

Representing real-world complex networks as *almost* structured graphs



Edge modification problems (editing, completion, deletion)

Polynomial-time algorithms: set of modifications minimal for inclusion

### Coedit

#### **INPUT: an arbitrary graph**

#### Computes either:

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

#### **OUTPUT: the cotree of the cograph obtained**

# of vertices degrees	∫u d°(u) ∖ v d°(v) i	# of nodes Label of the root # of children	l (=0 or 1)
edges	u2 v2		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 log<sup>2</sup>n) algorithm

almost linear in the size of the *input* 

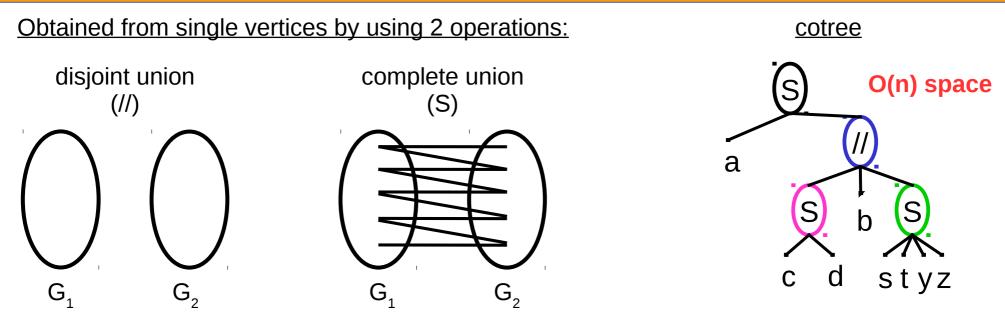
For editing

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

The vertex incremental approach : vertices are processed one by one

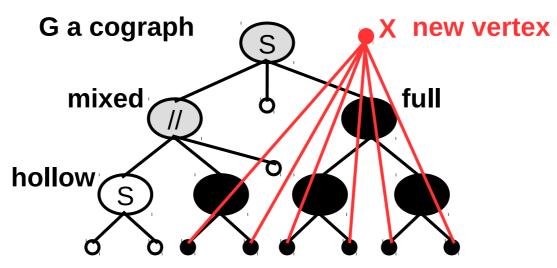


## **Cographs and incremental app.**



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

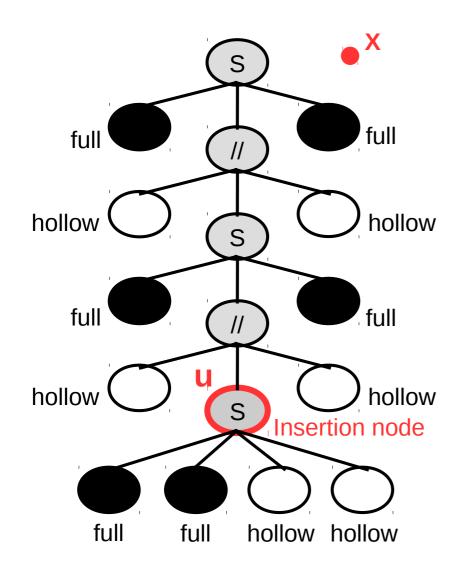


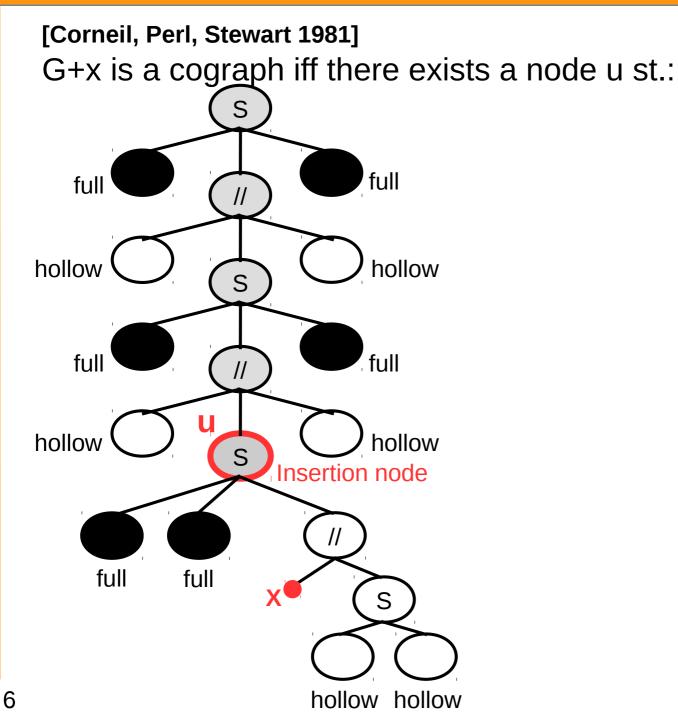
#### **Completion algorithms**

First algorithm: O(n+m')

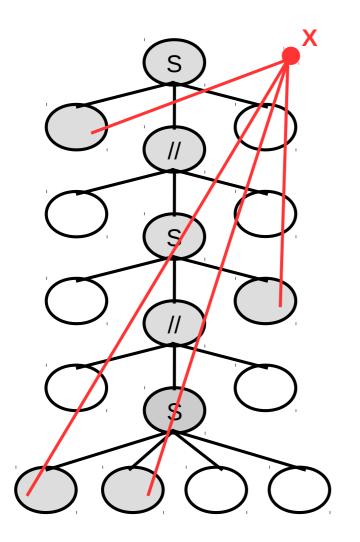
#### [Corneil, Perl, Stewart 1981]

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

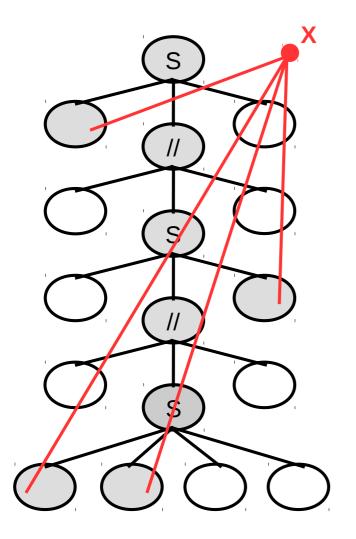




In our algorithm : G+x is not a cograph



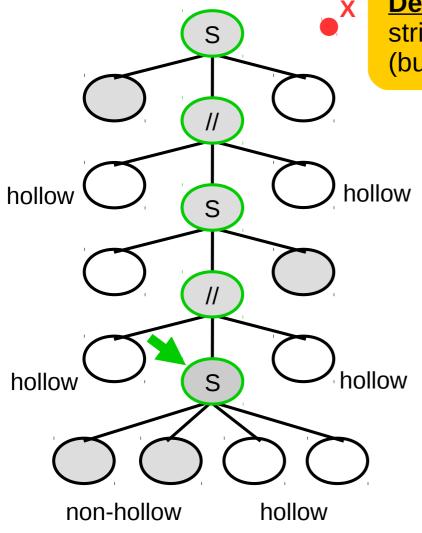
In our algorithm : G+x is not a cograph



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

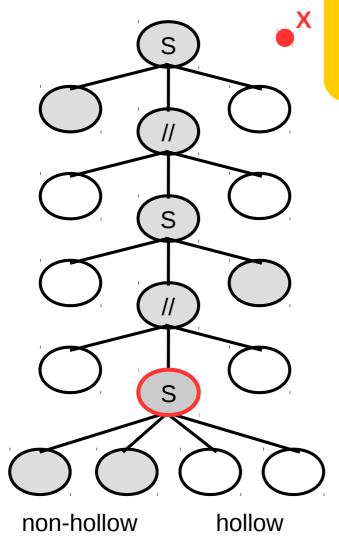
### **Eligible nodes**

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

In our algorithm : G+x is not a cograph

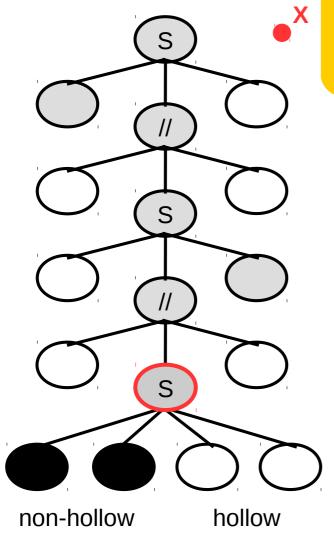


<u>**Definition:**</u> u is an *eligible node* Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

Proceed as follows :

1) choose one eligible node *u* 

In our algorithm : G+x is not a cograph



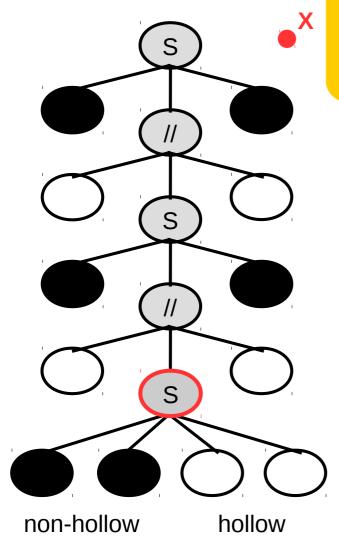
<u>**Definition:**</u> u is an *eligible node* Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

Proceed as follows :

1) choose one eligible node *u* 

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

In our algorithm : G+x is not a cograph



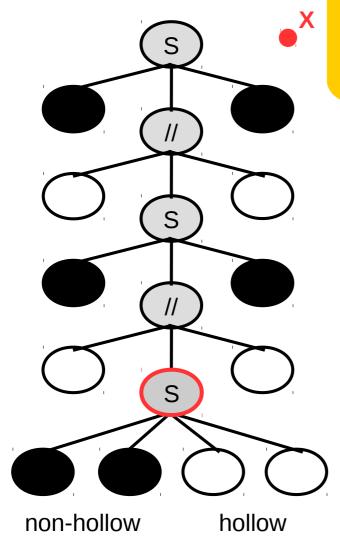
<u>**Definition:**</u> u is an *eligible node* Iff all parallel strict ancestors of u are such that all their children (but one) are hollow

Proceed as follows :

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**

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

Proceed as follows :

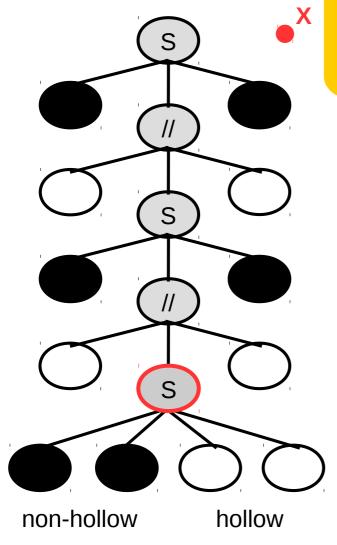
1) choose one eligible node *u* 

- 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* 

#### 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

Proceed as follows :

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 ?

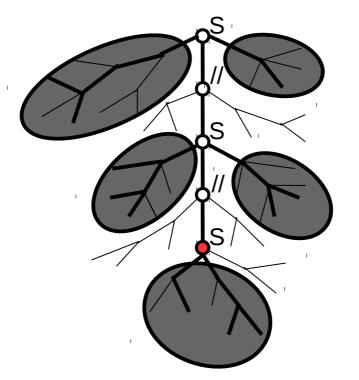


We have a characterization for this

## First algorithm : O(n+m')

Search the tree bottom up from the leaves adjacent to x

Find the eligible nodes that satisfy the characterization



<u>Note :</u> we search only non-hollow nodes

<u>Complexity</u> : O(d') [LMP 10]

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

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

#### **Completion algorithms**

#### Second algorithm: O(n + m log<sup>2</sup>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<sup>2</sup>) edges
- Random graphs with fixed average degree, O(n) edges, have the expansion property with high probability
- In practice, O(n+m') ~ O(n<sup>2</sup>)

11

We achieve O(n+m log<sup>2</sup>n) time

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

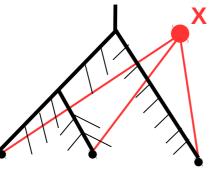
No reason to use adjacency lists to encode the output

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<sup>2</sup>) edges
- Random graphs with fixed average degree, O(n) edges, have the expansion property with high probability
- In practice, O(n+m') ~ O(n<sup>2</sup>)
- We achieve O(n+m log²n) time

Where is the room for improvement of the complexity?



11

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

## Second algorithm : O(n + m log<sup>2</sup>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

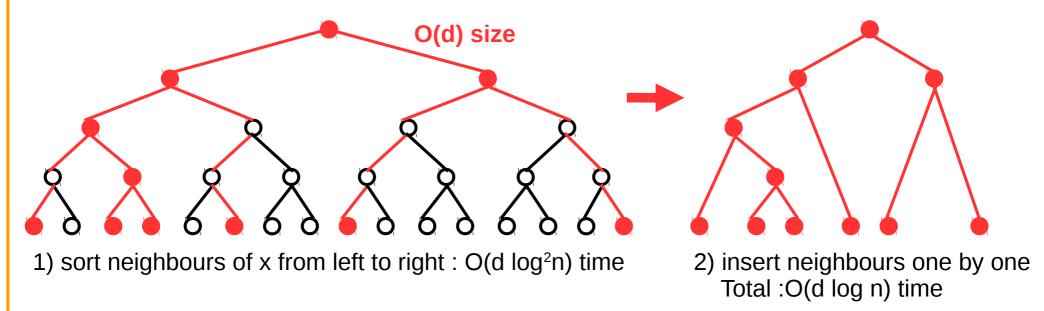
## Second algorithm : O(n + m log<sup>2</sup>n)

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

Lowest eligible nodes

▶ highest parallel nodes with  $\geq 2$  non-hollow children

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



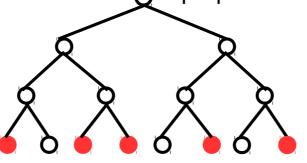
<u>Complexity :</u> O(d log<sup>2</sup>n) for one incremental step O(n+m log<sup>2</sup>n) for the whole algorithm

#### **Editing algorithm** O(n + m) time

## **Cograph editing**

Use both addition and deletion of edges
Find a minimum cardinality modification at each incremental step
Complexity O(n+m) time, O(d) time at each incremental step
Obs.: a minimum editing is not worse than deleting all edges incident to x
1) compute all maximal *preponderant* nodes and their budget

 $\mathbf{Q}$  u preponderant, budget is 5-3=2



2) for each parent u of some preponderant node, climb in the tree and try to fill what must be by using only the budgets of the children of u reach the root : success, otherwise : failure

ensures an O(d) time complexity

#### **Coedit : use case**

	Context	Network	n	m	$d^{\circ}$	%mod
	WWW	in-2004	1148875	12281937	21.4	12%
	WWW	cnr-2000	227058	2187201	19.3	19%
	PROTEIN	reactome	5973	145778	48.8	22%
	SOFTWARE	jdk	6434	53658	16.7	29%
	SOFTWARE	jung-j	6120	50290	16.4	29%
	WWW	eu-2005	835044	15718784	37.7	29%
	CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%
	CO-AUTHOR	ca-HepPh	11204	117619	21.0	34%
	SPECIES	foodweb	183	2434	26.6	43%
	CO-AUTHOR	dblp	317080	1049866	6.6	45%
	WORD-REL.	wordnet	145145	656230	9.0	48%
	COMMUNIC.	wiki-Talk	2388953	4656682	3.9	49%
	CO-SOLD	amazon	334863	925872	5.5	49%
	CO-AUTHOR	ca-CondMat	21363	91286	8.6	52%
	RANDOM	ER-Gnm_1M-2	796208	958827	2.4	52%
real-world	CO-AUTHOR	ca-HepTh	8638	24806	5.7	54%
_	INTERNET	as2000	6474	12572	3.9	54%
graphs	ROAD	roadNet-TX	1351137	1879201	2.8	54%
9	INTERNET	as-caida2007	26475	53381	4.0	55%
	CO-AUTHOR	ca-AstroPh	17903	196972	22.0	59%
	INTERNET	topology	34761	107720	6.2	61%
+	RANDOM	ER-Gnm_1M-3	940987	1494643	3.2	63%
	INTERNET	as-skitter	1694616	11094209	13.1	64%
	CO-OCCUR	bible-names	1 707	9059	10.6	67%
random	PROTEIN	figeys	2217	6418	5.8	67%
ranuom	CITATION-SCI.	cora	23166	89157	7.7	68%
graphs	SOCIAL	youtube	1134890	2987624	5.3	69%
graphs	CO-ACTOR	actor-col.	374511	15014839	80.2	71%
	P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	71%
	RANDOM	ER-Gnm_1M-4	980191	1999203	4.1	71%
	CITATION-SCI.	citeseer	365154	1721981	9.4	75%
	CITATION-PAT.	cit-Patents	3764117	16511740	8.8	76%
	SOFTWARE	linux	30817	213 208	13.8	77%
	SOCIAL	LiveJournal	3997962	34 681 189	17.4	78%
	CITATION-SCI.	cit-HepTh	27 400	352021	25.7	79%
	RANDOM	ER-Gnm_1M-6	997479	2 999 988	6.0	79%
	CITATION-SCI.	cit-HepPh	34 401	420784	24.5	81%
	RANDOM	ER-Gnm_1M-8	999684	3 999 999	8.0	84%
	RANDOM	ER-Gnm_1M-10	999952	5 000 000	10.0	87%
	RANDOM	ER-Gnm_1M-15	1000000	7 500 000	15.0	91 %
	SOCIAL	orkut	3072441	117 185 083	76.3	91 %
	RANDOM	ER-Gnm_1M-20	1000000	10 000 000	20.0	93%
	WORD-REL.	Thesaurus	23132	297094	25.7	93%

	Context	Network	n	m	d°	%mod
	WWW	in-2004	1148875	12281937	21.4	12%
	WWW	cnr-2000	227058	2187201	19.3	19%
	PROTEIN	reactome	5973	145778	48.8	22%
	SOFTWARE	jdk	6434	53658	16.7	29%
	SOFTWARE	jung-j	6120	50290	16.4	29%
	WWW	eu-2005	835044	15718784	37.7	29%
	CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%
	CO-AUTHOR	ca-HepPh	11204	117619	21.0	34%
	SPECIES	foodweb	183	2434	26.6	43%
	CO-AUTHOR	dblp	317080	1049866	6.6	45%
	WORD-REL.	wordnet	145145	656230	9.0	48%
	COMMUNIC.	wiki-Talk	2388953	4656682	3.9	49%
	CO-SOLD	amazon	334863	925872	5.5	49%
	CO-AUTHOR	ca-CondMat	21363	91286	8.6	52%
	RANDOM	ER-Gnm_1M-2	796208	958827	2.4	52%
35 real-world	CO-AUTHOR	ca-HepTh	8638	24806	5.7	54%
	INTERNET	as2000	6474	12572	3.9	54%
graphs	ROAD	roadNet-TX	1351137	1879201	2.8	54%
9	INTERNET	as-caida2007	26475	53381	4.0	55%
	CO-AUTHOR	ca-AstroPh	17903	196972	22.0	59%
	INTERNET	topology	34761	107720	6.2	61%
+	RANDOM	ER-Gnm_1M-3	940987	1494643	3.2	63%
	INTERNET	as-skitter	1694616	11094209	13.1	64%
	CO-OCCUR	bible-names	1707	9059	10.6	67%
8 random	PROTEIN	figeys	2217	6418	5.8	67%
014110011	CITATION-SCI.	cora	23166	89157	7.7	68%
graphs	SOCIAL	youtube	1134890	2987624	5.3	69%
graphs	CO-ACTOR	actor-col.	374511	15014839	80.2	71%
	P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	71%
	RANDOM	ER-Gnm_1M-4	980191	1999203	4.1	71%
	CITATION-SCI.	citeseer	365154	1721981	9.4	75%
	CITATION-PAT.	cit-Patents	3764117	16511740	8.8	76%
	SOFTWARE	linux	30817	213208	13.8	77%
	SOCIAL	LiveJournal	3997962	34 681 189	17.4	78%
	CITATION-SCI.	cit-HepTh	27 400	352021	25.7	79%
	RANDOM	ER-Gnm_1M-6	997479	2 999 988	6.0	79%
	CITATION-SCI.	cit-HepPh	34 401	420784	24.5	81%
	RANDOM	ER-Gnm_1M-8	999684	3 999 999	8.0	84%
	RANDOM	ER-Gnm_1M-10	999952	5000000	10.0	87%
	RANDOM	ER-Gnm_1M-15	1 000 000	7 500 000	15.0	91%
	SOCIAL	orkut	3072441	117 185 083	76.3	91%
	RANDOM	ER-Gnm_1M-20	1000000	10 000 000	20.0	93%
	WORD-REL.	Thesaurus	23132	297094	25.7	93%



Some networks are very close from cographs

$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		Context	Network	n	m	d°	%mod
S5 real-world graphs         WWW PROTEIN         cnr-2000 reactome         227058         2187201         19.3         19% PROTEIN           505 TWARE         jung-j         6120         50290         16.4         29% SOFTWARE         29% WWW         eu-2005         835041         1571874         37.7         29% SOFTWARE           WWW         eu-2005         835041         1571874         37.7         29% SOFTWARE         20% WWW         20.4         117619         21.0         34% SOFTWARE           SOFTWARE         jung-j         6120         50290         16.4         29% CO-AUTHOR         ca-HepPh         11204         117619         21.0         34% SOFTWARE           SPECIES         foodweb         188         2434         26.6         43% SOFTWARE         317080         1049866         6.6         45% WORD-REL.         wordnet         145145         656230         9.0         48% CO-AUTHOR         ca-CondMat         21363         91286         8.6         52% CO-AUTHOR         ca-CondMat							
PROTEIN         reactome         5973         145778         48.8         22%           SOFTWARE         juk         6434         53658         16.7         29%           SOFTWARE         jung-j         6120         50200         16.4         29%           WWW         eu-2005         835044         15718784         37.7         29%           CO-AUTHOR         ca-GrQc         4158         13422         65         34%           CO-AUTHOR         ca-GrQc         4158         13422         65         43%           SPECIES         foodweb         183         2434         26.6         43%           CO-AUTHOR         dblp         317080         1049866         66         45%           COMUNIC.         wiki-Talk         2388953         465623         9.0         48%           CO-AUTHOR         ca-CondMat         21363         91286         86         52%           RANDOM         ER-Gm.IM-2         796208         95827         2.5         49%           CO-AUTHOR         ca-HepTh         8838         24806         5.7         54%           NTERNET         assciata2007         26475         53381         40         55%<		WWW			2187201	19.3	
35 real-world graphs       jdk       6434       53658       16.7       29%         SOFTWARE       jung-j       6120       50290       16.4       29%         WWW       eu-2005       835044       15718784       37.7       29%         CO-AUTHOR       ca-GrQc       4158       13422       6.5       34%         CO-AUTHOR       ca-HepPh       11204       1171619       21.0       34%         SPECIES       foodweb       183       2434       26.6       43%         CO-AUTHOR       dblp       317080       1049866       6.6       45%         WORD-REL       wordnet       145145       656230       9.0       48%         CO-AUTHOR       ca-CondMat       2388953       4656682       39       49%         CO-AUTHOR       ca-CondMat       21363       91286       8.6       52%         RANDOM       ER-Gmm.IM-2       796208       995827       2.4       52%         CO-AUTHOR       ca-Hep1h       8638       24806       5.7       54%         INTERNET       as-caida2007       26475       53381       4.0       55%         CO-AUTHOR       ca-AstroPh       17901       2.8		PROTEIN	reactome		145778	48.8	
WWW         eu-2005         835044         15718784         37.7         29%           CO-AUTHOR         ca-GrQc         4158         13422         6.5         34%           SPECIES         foodweb         183         2434         26.6         43%           CO-AUTHOR         dblp         31708         1049866         6.6         45%           WORD-REL         wordnet         145145         656230         9.0         48%           CO-AUTHOR         dblp         31708         1049866         6.6         45%           WORD-REL         wordnet         145145         656230         9.0         48%           CO-SOLD         amazon         334863         925872         5.5         49%           CO-AUTHOR         ca-HepTh         8638         24806         5.7         54%           INTERNET         as2000         6474         12572         3.9         54%           INTERNET         ascaida2007         26475         53381         4.0         55%           CO-AUTHOR         ca-HepTh         858         2406         5.7         54%           INTERNET         ascaida2007         26475         53381         4.0         55		SOFTWARE	jdk	6434	53658	16.7	
$ \begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		SOFTWARE	jung-j	6120	50290	16.4	29%
Species         Co-AUTHOR SPECIES         ca-HepPh foodweb         11204         117619         21.0         34% SPECIES           CO-AUTHOR         dblp         317080         1049866         6.6         43% WORD-REL           WORD-REL         wordnet         115114         656230         9.0         48% COMMUNIC           CO-SOLD         amazon         334863         992862         3.9         49% CO-SOLD           CO-AUTHOR         ca-HepTh         8638         24806         5.7         54% SPECIES           RANDOM         ER-Gnm_IM-2         796208         958827         2.4         52% SPECIES           CO-AUTHOR         ca-HepTh         8638         24806         5.7         54% SPECIES           INTERNET         ascaida2007         26475         53381         4.0         55% CO-AUTHOR           INTERNET         ascaida2007         26475         53381         4.0         55% CO-AUTHOR         55% CO-AUTHOR         55% CO-AUTHOR         64764         11094209         1.1         64% CO           NTERNET         asscaide2007         26475         53381         4.0         55% CO-AUTHOR         6486         5.8         67% CITATION-SCI.         6764         58         67% CITATION-SCI.         <		WWW	eu-2005	835044	15718784	37.7	29%
SPECIES         foodweb         183         2434         26.6         43%           CO-AUTHOR         dblp         317080         1049866         6.6         45%           WORD-REL         wordnet         145145         656230         9.0         48%           COMMUNIC         wiki-Talk         2388953         4656682         3.9         49%           CO-SOLD         amazon         334863         925872         5.5         49%           CO-AUTHOR         ca-CondMat         21363         91286         8.6         52%           RANDOM         ER-Gmm.1M-2         796208         958827         2.4         52%           CO-AUTHOR         ca-Hep1h         8638         24806         5.7         54%           INTERNET         as2000         6474         12572         3.9         54%           INTERNET         as-caida2007         26475         53381         4.0         55%           CO-AUTHOR         ca-AstroPh         17903         196972         2.0         59%           RANDOM         ER-Gmm.1M-3         940987         1494643         3.2         63%           INTERNET         as-skitter         1694616         11094209		CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%
Solution         Co-AUTHOR WORD-REL.         dblp wordnet         317080         1049866         6.6         45% WORD-REL.           WORD-REL.         wordnet         145145         656230         9.0         48% CO-SOLD           CO-SOLD         mazon         334863         925872         5.5         49% CO-SOLD           CO-AUTHOR         ca-CondMat         21363         91286         8.6         52%           RANDOM         ER-Gm.1M-2         796208         958827         2.4         52%           RANDOM         CO-AUTHOR         ca-CondMat         21363         91286         8.6         52%           RANDOM         ER-Gm.1M-2         796208         958827         2.4         52%           RANDOM         CO-AUTHOR         ca-CondMat         21363         91286         8.6         52%           RANDOM         Co-AUTHOR         ca-CondMat         1351137         1879201         2.8         54%           NTERNET         ascaida2007         26475         53381         4.0         55%           CO-AUTHOR         ca-AstroPh         17903         16972         22.0         59%           INTERNET         asskitter         1694616         11094209         13		CO-AUTHOR	ca-HepPh	11204	117619		
MORD-REL.         wordnet         145145         656230         9.0         48%           COMMUNIC.         wiki-Talk         2388953         4656682         3.9         49%           CO-SOLD         amazon         334863         925872         5.5         44%           CO-AUTHOR         ca-CondMat         21363         91286         8.6         52%           RANDOM         ER-Gmm.1M-2         796208         958827         2.4         52%           RANDOM         ER-Gmm.1M-2         796208         958827         2.4         52%           CO-AUTHOR         ca-CondMat         1351137         1879201         2.8         54%           ROAD         roadNet-TX         1351137         1879201         2.8         54%           INTERNET         ascaida2007         26475         53381         4.0         55%           CO-AUTHOR         ca-AstroPh         17903         196972         22.0         59%           INTERNET         asskitter         1694616         11094209         13.1         64%           CO-OCCUR         bible-names         1707         9059         10.6         67%           SOCIAL         youtube         134890			foodweb	183		26.6	
35 real-world graphs         CO-MUUNIC. (Co-SOLD         wiki-Talk         238953         4656682         3.9         49%           CO-SOLD         amazon         334863         925872         5.5         49%           CO-AUTHOR         ca-CondMat         21363         91286         8.6         52%           RANDOM         ER-Gm.1M-2         706208         958827         2.4         52%           CO-AUTHOR         ca-HepTh         8638         24806         5.7         54%           INTERNET         ascoid2007         26475         53381         4.0         55%           CO-AUTHOR         ca-AstroPh         17903         196972         22.0         59%           INTERNET         topology         34761         107720         6.2         61%           RANDOM         ER-Gm.1M-3         940987         1494643         3.2         66%           INTERNET         topology         34761         107720         6.2         61%           RANDOM         ER-Gm.1M-3         940987         1494643         3.2         66%           CO-ACTOR         actor-col.         374511         1501483         80.2         71%           CO-ACTOR         actor-col.<			dblp	317080			
Solution         CO-SOLD CO-AUTHOR         amazon ca-CondMat         334863         925872         5.5         49% 52%           35 real-world graphs         RANDOM         ER-Gm.IM-2         796208         958827         2.4         52%           CO-AUTHOR         ca-HepTh         8638         24806         5.7         54%           MODM         ER-Gm.IM-2         1351137         1879201         2.8         54%           INTERNET         as2000         6474         12572         3.9         54%           ROAD         roadNet-TX         1351137         1879201         2.8         54%           INTERNET         ascaida2007         26475         53381         4.0         55%           CO-AUTHOR         ca-AstroPh         17003         196972         2.0         59%           INTERNET         topology         34761         100720         6.2         61%           RANDOM         ER-Gnm.IM-3         940987         1494643         3.2         63%           INTERNET         asskitter         1694616         11094209         13.1         64%           CO-ACUR         bible-names         1707         9059         10.6         67%           CO-A							
35 real-world graphs         CO-AUTHOR         ca-CondMat         21363         91286         8.6         52%           RANDOM         ER-Gm.1M-2         796208         958827         2.4         52%           CO-AUTHOR         ca-HepTh         8638         24806         5.7         54%           INTERNET         as2000         6474         12572         3.9         54%           ROAD         roadNet-TX         1351137         1879201         2.8         54%           INTERNET         as-caida2007         26475         53381         4.0         55%           CO-AUTHOR         ca-AstroPh         17903         196972         22.0         59%           INTERNET         topology         34761         100720         6.2         61%           RANDOM         ER-Gm.1M-3         940987         1494643         3.2         63%           INTERNET         as-skitter         1694616         11094209         1.1         64%           CO-OCCUR         bible-names         1707         9059         10.6         67%           POTEIN         figeys         2217         6418         5.8         60%           CO-ACTOR         actor-col.         374			wiki-Talk				
35 real-world graphs       RANDOM       ER-Gnm_1M-2       796208       958827       2.4       52%         CO-AUTHOR       ca-Hep1h       8638       24806       5.7       54%         INTERNET       as2000       6474       12572       3.9       54%         NOAD       roadke-TX       1351137       1879201       2.8       54%         INTERNET       as-caida2007       26475       53381       4.0       55%         CO-AUTHOR       ca-AstroPh       17903       196972       22.0       59%         INTERNET       topology       34761       107720       6.2       61%         RANDOM       ER-Gnm_1M-3       940987       1494643       3.2       63%         INTERNET       topology       34761       107720       6.2       61%         RANDOM       ER-Gnm_1M-3       940987       1494643       3.2       63%         CO-OCCUR       bible-names       1707       9059       10.6       67%         CTATION-SCL       cora       23166       89157       7.7       68%         SOCIAL       youtube       1134890       2987624       5.3       69%         CO-ACTOR       caceer							
35 real-world graphs       CO-AUTHOR ca-HepTh as2000 6474 12572 3.9 54%         NTERNET as2000 6474 12572 3.9 54%         ROAD roadNet-TX 1351137 1879201 2.8 54%         INTERNET as-caida2007 26475 53381 4.0 55%         CO-AUTHOR ca-AstroPh 17903 196972 22.0 59%         INTERNET topology 34761 107720 6.2 61%         RANDOM ER-Gnm.1M-3 940987 1494643 3.2 63%         INTERNET as-skitter 1694616 11094209 13.1 64%         CO-OCCUR bible-names 1707 9059 10.6 67%         PROTEIN figeys 2217 6418 5.8 67%         CO-ACTOR actor-col. 374511 15014839 80.2 71%         POP-CONNECT. p2p-Gnutella 62561 147878 4.7 71%         RANDOM ER-Gnm.1M-4 980191 1999203 4.1 71%         CITATION-SCI. cit-Recr       365154 1721981 9.4 75%         CITATION-PAT. cit-Patents 3764117 16511740 8.8 76%         SOCIAL LiveJournal 3997962 34681189 17.4 78%         SOCIAL LiveJournal 3997962 34							
graphs       INTERNET       as2000       6474       12572       3.9       54%         ROAD       roadNet-TX       1351137       1879201       2.8       54%         NTERNET       as-caida2007       26475       53381       4.0       55%         CO-AUTHOR       ca-AstroPh       17003       196972       22.0       59%         INTERNET       topology       34761       1007720       6.2       61%         RANDOM       ER-Gnm.1M-3       940987       1494643       3.2       63%         INTERNET       as-skitter       1694616       11094209       13.1       64%         CO-OCCUR       bible-names       1707       9059       10.6       67%         PROTEIN       figeys       2217       6418       5.8       69%         CO-ACTOR       actor-col.       374511       15014839       80.2       71%         P2P-CONNECT.       p2p-Gnutella       62561       147878       4.7       71%         CITATION-SCI.       citseer       365154       1721981       9.4       75%         SOFTWARE       linux       30817       213208       13.8       77%         SOCIAL       LiveJournal	o = 1 1 1						
graphs       INTERNET       as2000       6474       12572       3.9       54%         ROAD       roadNet-TX       1351137       1879201       2.8       54%         NTERNET       as-caida2007       26475       53381       4.0       55%         CO-AUTHOR       ca-AstroPh       17003       196972       22.0       59%         INTERNET       topology       34761       1007720       6.2       61%         RANDOM       ER-Gnm.1M-3       940987       1494643       3.2       63%         INTERNET       as-skitter       1694616       11094209       13.1       64%         CO-OCCUR       bible-names       1707       9059       10.6       67%         PROTEIN       figeys       2217       6418       5.8       69%         CO-ACTOR       actor-col.       374511       15014839       80.2       71%         P2P-CONNECT.       p2p-Gnutella       62561       147878       4.7       71%         CITATION-SCI.       citseer       365154       1721981       9.4       75%         SOFTWARE       linux       30817       213208       13.8       77%         SOCIAL       LiveJournal	35 real-world						
H       INTERNET       as-caida2007       26 475       53 381       4.0       55%         CO-AUTHOR       ca-AstroPh       17903       196972       22.0       59%         INTERNET       topology       34 761       107720       6.2       61%         RANDOM       ER-Gnm.1M-3       940987       1494 643       3.2       63%         INTERNET       as-skitter       1694 616       11094 209       13.1       64%         CO-OCCUR       bible-names       1707       9059       10.6       67%         PROTEIN       figeys       2217       6418       5.8       67%         CITATION-SCI.       cora       23166       89157       7.7       68%         SOCIAL       youtube       1134 890       2987 624       5.3       69%         CO-ACTOR       actor-col.       374 511       15014 839       80.2       71%         P2P-CONNECT.       p2p-Gnutella       62 561       147 878       4.7       71%         RANDOM       ER-Gnm.1M-4       980 191       1999 203       4.1       71%         CO-ACTOR       citeseer       365 154       1721 981       9.4       75%         CITATION-SCI. <t< th=""><th></th><td></td><td></td><td></td><td></td><td></td><td></td></t<>							
H       INTERNET       as-caida2007       26 475       53 381       4.0       55%         CO-AUTHOR       ca-AstroPh       17903       196972       22.0       59%         INTERNET       topology       34 761       107720       6.2       61%         RANDOM       ER-Gnm.1M-3       940987       1494 643       3.2       63%         INTERNET       as-skitter       1694 616       11094 209       13.1       64%         CO-OCCUR       bible-names       1707       9059       10.6       67%         PROTEIN       figeys       2217       6418       5.8       67%         CITATION-SCI.       cora       23166       89157       7.7       68%         SOCIAL       youtube       1134 890       2987 624       5.3       69%         CO-ACTOR       actor-col.       374 511       15014 839       80.2       71%         P2P-CONNECT.       p2p-Gnutella       62 561       147 878       4.7       71%         RANDOM       ER-Gnm.1M-4       980 191       1999 203       4.1       71%         CO-ACTOR       citeseer       365 154       1721 981       9.4       75%         CITATION-SCI. <t< th=""><th>grapns</th><td></td><td></td><td></td><td></td><td></td><td></td></t<>	grapns						
+         INTERNET         topology         34761         107720         6.2         61%           RANDOM         ER-Gnm.1M-3         940987         1494643         3.2         63%           INTERNET         as-skitter         1694616         11094209         13.1         64%           CO-OCCUR         bible-names         1707         9059         10.6         67%           PROTEIN         figeys         2217         6418         5.8         67%           CTATION-SCI.         cora         23166         89157         7.7         68%           SOCIAL         youtube         1134890         2987624         5.3         69%           CO-ACTOR         actor-col.         374511         15014839         80.2         71%           P2P-CONNECT.         p2p-Gnutella         62561         147878         4.7         71%           RANDOM         ER-Gnm.1M-4         980191         1999203         4.1         71%           CITATION-SCI.         citeseer         365154         1721981         9.4         75%           CITATION-SCI.         cite-Patents         3764117         16511740         8.8         76%           SOCIAL         linux	8 1						
+       RANDOM       ER-Gnm.1M-3       940987       1494643       3.2       63%         INTERNET       as-skitter       1694616       11094209       13.1       64%         CO-OCCUR       bible-names       1707       9059       10.6       67%         PROTEIN       figeys       2217       6418       5.8       67%         CITATION-SCI.       cora       23166       89157       7.7       68%         SOCIAL       youtube       1134890       2987624       5.3       69%         CO-ACTOR       actor-col.       374511       15014839       80.2       71%         P2P-CONNECT.       p2p-Gnutella       62561       147878       4.7       71%         RANDOM       ER-Gnm.1M-4       980191       1999203       4.1       71%         CITATION-SCI.       citeseer       365154       1721981       9.4       75%         CITATION-SCI.       citeseer       365154       1721981       9.4       75%         SOCIAL       LiveJournal       3997962       34681189       17.4       78%         CITATION-SCI.       cit-HepTh       27400       352021       25.7       79%         CITATION-SCI.							
B random graphs         INTERNET CO-OCCUR         as-skitter bible-names         1694616         11094209         13.1         64%           CO-OCCUR         bible-names         1707         9059         10.6         67%           PROTEIN         figeys         2217         6418         5.8         67%           CITATION-SCI.         cora         23166         89157         7.7         68%           SOCIAL         youtube         1134890         2987624         5.3         69%           CO-ACTOR         actor-col.         374511         15014839         80.2         71%           P2P-CONNECT.         p2p-Gnutella         62561         147878         4.7         71%           RANDOM         ER-Gnm.1M-4         980191         1999203         4.1         71%           CITATION-SCI.         citeseer         365154         1721981         9.4         75%           CITATION-PAT.         cit-Patents         3764117         16511740         8.8         76%           SOCIAL         LiveJournal         3997962         34 681189         17.4         78%           CITATION-SCI.         cit-HepTh         27400         352021         25.7         79%           <	Т		1 00				
Sandong graphs         CO-OCCUR         bible-names         1707         9059         10.6         67%           PROTEIN         figeys         2217         6418         5.8         67%           CITATION-SCI.         cora         23166         89157         7.7         68%           SOCIAL         youtube         1134890         2987624         5.3         69%           CO-ACTOR         actor-col.         374511         15014839         80.2         71%           P2P-CONNECT.         p2p-Gnutella         62561         147878         4.7         71%           RANDOM         ER-Gnm.1M-4         980191         1999203         4.1         71%           CITATION-SCI.         citeseer         365154         1721981         9.4         75%           CITATION-PAT.         cit-Patents         3764117         16511740         8.8         76%           SOFTWARE         linux         30817         213208         13.8         77%           SOCIAL         LiveJournal         3997962         34681189         17.4         78%           CITATION-SCI.         cit-HepTh         27400         352021         25.7         79%           RANDOM         ER-Gn	Ŧ						
B random graphs         PROTEIN         figeys         2217         6418         5.8         67%           CITATION-SCI.         cora         23166         89157         7.7         668%           SOCIAL         youtube         1134890         2987624         5.3         69%           CO-ACTOR         actor-col.         374511         15014839         80.2         71%           P2P-CONNECT.         p2p-Gnutella         62561         147878         4.7         71%           RANDOM         ER-Gnm.1M-4         980191         1999203         4.1         71%           CITATION-SCI.         citeseer         365154         1721981         9.4         75%           CITATION-PAT.         cit-Patents         3764117         16511740         8.8         76%           SOFTWARE         linux         30817         213208         13.8         77%           SOCIAL         LiveJournal         3997962         34681189         17.4         78%           CITATION-SCI.         cit-HepTh         27400         352021         25.7         79%           RANDOM         ER-Gnm.1M-6         997479         2999988         6.0         79%           RANDOM         E							
8 random       CITATION-SCI.       cora       23166       89157       7.7       68%         graphs       SOCIAL       youtube       1134890       2987624       5.3       69%         CO-ACTOR       actor-col.       374511       15014839       80.2       71%         P2P-CONNECT.       p2p-Gnutella       62561       147878       4.7       71%         RANDOM       ER-Gnm_1M-4       980191       1999203       4.1       71%         CITATION-SCI.       citeseer       365154       1721981       9.4       75%         CITATION-PAT.       cit-Patents       3764117       16511740       8.8       76%         SOCIAL       LiveJournal       3997962       34681189       17.4       78%         CITATION-SCI.       cit-HepTh       27400       352021       25.7       79%         SOCIAL       LiveJournal       3997962       34681189       17.4       78%         CITATION-SCI.       cit-HepTh       27400       352021       25.7       79%         RANDOM       ER-Gnm_1M-6       997479       2999988       6.0       79%         CITATION-SCI.       cit-HepPh       34401       420784       24.5       81%							
Graphs         Initial Solution of the solutis of the solutis of the solution of the solution of the solution	8 random						
<b>Graphs</b> CO-ACTOR P2P-CONNECT.actor-col. p2p-Gnutella3745111501483980.271%P2P-CONNECT.p2p-Gnutella625611478784.771%RANDOMER-Gnm.1M-498019119992034.171%CITATION-SCI.citeseer36515417219819.475%CITATION-PAT.cit-Patents3764117165117408.876%SOFTWARElinux3081721320813.877%SOCIALLiveJournal39979623468118917.478%CITATION-SCI.cit-HepTh2740035202125.779%RANDOMER-Gnm.1M-699747929999886.079%RANDOMER-Gnm.1M-899968439999998.084%RANDOMER-Gnm.1M-10999952500000010.087%RANDOMER-Gnm.1M-151000000750000015.091%							
P2P-CONNECT.p2p-Gnutella62 561147 8784.771 %RANDOMER-Gnm.1M-4980 1911999 2034.171 %CITATION-SCI.citeseer365 1541721 9819.475 %CITATION-PAT.cit-Patents3764 11716 511 7408.876 %SOFTWARElinux30 817213 20813.877 %SOCIALLiveJournal3997 96234 681 18917.478 %CITATION-SCI.cit-HepTh27 400352 02125.779 %RANDOMER-Gnm.1M-6997 4792999 9886.079 %RANDOMER-Gnm.1M-8999 6843999 9998.084 %RANDOMER-Gnm.1M-10999 9525000 00010.087 %RANDOMER-Gnm.1M-151000 0007 500 00015.091 %	graphs		·				
RANDOMER-Gnm_1M-4980 1911 999 2034.171%CITATION-SCI.citeseer365 1541 721 9819.475%CITATION-PAT.cit-Patents3764 11716 511 7408.876%SOFTWARElinux30 817213 20813.877%SOCIALLiveJournal3 997 96234 681 18917.478%CITATION-SCI.cit-HepTh27 400352 02125.779%RANDOMER-Gnm_1M-6997 4792 999 9886.079%CITATION-SCI.cit-HepPh34 401420 78424.581%RANDOMER-Gnm_1M-8999 6843 999 9998.084%RANDOMER-Gnm_1M-10999 9525 000 00010.087%RANDOMER-Gnm_1M-151 000 0007 500 00015.091%	9						
CITATION-SCI.citeseer36515417219819.475%CITATION-PAT.cit-Patents3764117165117408.876%SOFTWARElinux3081721320813.877%SOCIALLiveJournal39979623468118917.478%CITATION-SCI.cit-HepTh2740035202125.779%RANDOMER-Gnm.1M-699747929999886.079%CITATION-SCI.cit-HepPh3440142078424.581%RANDOMER-Gnm.1M-899968439999998.084%RANDOMER-Gnm.1M-10999952500000010.087%RANDOMER-Gnm.1M-151000000750000015.091%							
CITATION-PAT.cit-Patents3764117165117408.876%SOFTWARElinux3081721320813.877%SOCIALLiveJournal39979623468118917.478%CITATION-SCI.cit-HepTh2740035202125.779%RANDOMER-Gnm.1M-699747929999886.079%CITATION-SCI.cit-HepPh3440142078424.581%RANDOMER-Gnm.1M-899968439999998.084%RANDOMER-Gnm.1M-10999952500000010.087%RANDOMER-Gnm.1M-151000000750000015.091%							
SOFTWARE       linux       30817       213208       13.8       77%         SOCIAL       LiveJournal       3997962       34681189       17.4       78%         CITATION-SCI.       cit-HepTh       27400       352021       25.7       79%         RANDOM       ER-Gnm.1M-6       997479       2999988       6.0       79%         CITATION-SCI.       cit-HepPh       34401       420784       24.5       81%         RANDOM       ER-Gnm.1M-8       999684       3999999       8.0       84%         RANDOM       ER-Gnm.1M-10       999952       5000000       10.0       87%         RANDOM       ER-Gnm.1M-15       1000000       7500000       15.0       91%							
SOCIAL       LiveJournal       3997962       34681189       17.4       78%         CITATION-SCI.       cit-HepTh       27400       352021       25.7       79%         RANDOM       ER-Gnm.1M-6       997479       2999988       6.0       79%         CITATION-SCI.       cit-HepPh       34401       420784       24.5       81%         RANDOM       ER-Gnm.1M-8       999684       3999999       8.0       84%         RANDOM       ER-Gnm.1M-10       999952       5000000       10.0       87%         RANDOM       ER-Gnm.1M-15       1000000       7500000       15.0       91%							
CITATION-SCI.cit-HepTh2740035202125.779%RANDOMER-Gnm.1M-699747929999886.079%CITATION-SCI.cit-HepPh3440142078424.581%RANDOMER-Gnm.1M-899968439999998.084%RANDOMER-Gnm.1M-10999952500000010.087%RANDOMER-Gnm.1M-151000000750000015.091%							
RANDOMER-Gnm-1M-699747929999886.079%CITATION-SCI.cit-HepPh3440142078424.581%RANDOMER-Gnm-1M-899968439999998.084%RANDOMER-Gnm-1M-10999952500000010.087%RANDOMER-Gnm-1M-151000000750000015.091%							
CITATION-SCI.cit-HepPh3440142078424.581%RANDOMER-Gnm_1M-899968439999998.084%RANDOMER-Gnm_1M-10999952500000010.087%RANDOMER-Gnm_1M-151000000750000015.091%			-				
RANDOMER-Gnm_1M-8999 6843 999 9998.084 %RANDOMER-Gnm_1M-10999 9525 000 00010.087 %RANDOMER-Gnm_1M-151 000 0007 500 00015.091 %							
RANDOM RANDOMER-Gnm_1M-10999 9525 000 00010.087 %RANDOMER-Gnm_1M-151 000 0007 500 00015.091 %							
RANDOM         ER-Gnm_1M-15         1 000 000         7 500 000         15.0         91 %							
10001 AD $101 M B 101 A 441 111 100 000 1 10.0 1 91 70$		SOCIAL	orkut	3072441	117185083	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

	Context	Network	n	m	$d^{\circ}$	%mod	
	WWW	in-2004	1148875	12281937	21.4	12%	
	WWW	cnr-2000	227058	2187201	19.3	19%	•
	PROTEIN	reactome	5973	145778	48.8	22%	
	SOFTWARE	jdk	6434	53658	16.7	29%	
	SOFTWARE	jung-j	6120	50290	16.4	29%	
	WWW	eu-2005	835044	15718784	37.7	29%	_
	CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%	
	CO-AUTHOR	ca-HepPh	11204	117619	21.0	34%	
	SPECIES	foodweb	183	2434	26.6	43%	
	CO-AUTHOR	dblp	317080	1049866	6.6	45%	
	WORD-REL.	wordnet	145145	656230	9.0	48%	
	COMMUNIC.	wiki-Talk	2388953	4656682	3.9	49%	
	CO-SOLD	amazon	334863	925872	5.5	49%	
	CO-AUTHOR	ca-CondMat	21363	91286	8.6	52%	
	RANDOM	ER-Gnm_1M-2	796208	958827	2.4	52%	
35 real-world	CO-AUTHOR	ca-HepTh	8 6 3 8	24806	5.7	54%	
	INTERNET	as2000	6474	12572	3.9	54%	
graphs	ROAD	roadNet-TX	1351137	1879201	2.8	54%	
9	INTERNET	as-caida2007	26475	53381	4.0	55%	
	CO-AUTHOR	ca-AstroPh	17903	196972	22.0	59%	
	INTERNET	topology	34761	107720	6.2	61%	
+	RANDOM	ER-Gnm_1M-3	940987	1494643	3.2	63%	
	INTERNET	as-skitter	1694616	11094209	13.1	64%	
	CO-OCCUR	bible-names	1 707	9059	10.6	67%	
8 random	PROTEIN	figeys	2217	6418	5.8	67%	
0 14110011	CITATION-SCI.	cora	23166	89157	7.7	68%	
graphs	SOCIAL	youtube	1134890	2987624	5.3	69%	
graphs	CO-ACTOR	actor-col.	374511	15014839	80.2	71%	
	P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	71%	
	RANDOM	ER-Gnm_1M-4	980191	1999203	4.1	71%	
	CITATION-SCI.	citeseer	365154	1721981	9.4	75%	
	CITATION-PAT.	cit-Patents	3764117	16511740	8.8	76%	
	SOFTWARE	linux	30817	213208	13.8	77%	
	SOCIAL	LiveJournal	3997962	34681189	17.4	78%	
	CITATION-SCI.	cit-HepTh	27400	352021	25.7	79%	
	RANDOM	ER-Gnm_1M-6	997479	2999988	6.0	79%	
	CITATION-SCI.	cit-HepPh	34401	420784	24.5	81%	
	RANDOM	ER-Gnm_1M-8	999684	3999999	8.0	84 %	
	RANDOM	ER-Gnm_1M-10	999952	5000000	10.0	87 %	
	RANDOM	ER-Gnm_1M-15	1 000 000	7500000	15.0	91%	
	SOCIAL	orkut	3072441	117185083	76.3	91%	
	RANDOM	ER-Gnm_1M-20	1000000	10000000	20.0	93%	
	WORD-REL.	Thesaurus	23132	297094	25.7	93%	

#### **RESULTS**

Some networks are very close from cographs

Random graphs are never

A wide range of proximity :

12% to 93%

	Context	Network	n	m	$\mathbf{d}^{\circ}$	%mod	
	WWW	in-2004	1148875	12281937	21.4	12%	
	WWW	cnr-2000	227058	2187201	19.3	19%	
	PROTEIN	reactome	5973	145778	48.8	22%	
	SOFTWARE	jdk	6434	53658	16.7	29%	
	SOFTWARE	jung-j	6120	50290	16.4	29%	
	WWW	eu-2005	835044	15718784	37.7	29%	
	CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%	📕 Som
	CO-AUTHOR	ca-HepPh	11204	117619	21.0	34%	
	SPECIES	foodweb	183	2434	26.6	43%	مامدر
	CO-AUTHOR	dblp	317080	1049866	6.6	45%	close
	WORD-REL.	wordnet	145145	656230	9.0	48%	
	COMMUNIC.	wiki-Talk	2388953	4656682	3.9	49%	
	CO-SOLD	amazon	334863	925872	5.5	49%	Done
	CO-AUTHOR	ca-CondMat	21363	91286	8.6	52%	Rano
	RANDOM	ER-Gnm_1M-2	796208	958827	2.4	52%	
35 real-world	CO-AUTHOR	ca-HepTh	8 6 3 8	24806	5.7	54%	
	INTERNET	as2000	6474	12572	3.9	54%	
graphs	ROAD	roadNet-TX	1351137	1879201	2.8	54%	
9.49.10	INTERNET	as-caida2007	26475	53381	4.0	55%	
	CO-AUTHOR	ca-AstroPh	17903	196972	22.0	59%	
_	INTERNET	topology	34761	107720	6.2	61%	
+	RANDOM	ER-Gnm_1M-3	940987	1494643	3.2	63%	
	INTERNET	as-skitter	1694616	11094209	13.1	64%	
	CO-OCCUR	bible-names	1707	9059	10.6	67%	A wi
9 random	PROTEIN	figeys	2217	6418	5.8	67%	
8 random	CITATION-SCI.	cora	23166	89157	7.7	68%	12%
graphs	SOCIAL	youtube	1134890	2987624	5.3	69%	<b>TZ</b> /0
yiapiis	CO-ACTOR	actor-col.	374511	15014839	80.2	71%	
	P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	71%	
	RANDOM	ER-Gnm_1M-4	980191	1999203	4.1	71%	The
	CITATION-SCI.	citeseer	365154	1721981	9.4	75%	
	CITATION-PAT.	cit-Patents	3764117	16511740	8.8	76%	
	SOFTWARE	linux	30817	213208	13.8	77%	highl
	SOCIAL	LiveJournal	3997962	34681189	17.4	78%	ingin
	CITATION-SCI.	$\operatorname{cit-HepTh}$	27400	352021	25.7	79%	
	RANDOM	ER-Gnm_1M-6	997479	2999988	6.0	79%	real-
	CITATION-SCI.	cit-HepPh	34401	420784	24.5	81%	
	RANDOM	ER-Gnm_1M-8	999684	3999999	8.0	84%	
	RANDOM	ER-Gnm_1M-10	999952	5000000	10.0	87%	
	RANDOM	ER-Gnm_1M-15	1000000	7500000	15.0	91%	
	SOCIAL	orkut	3072441	117185083	76.3	91%	
	RANDOM	ER-Gnm_1M-20	1000000	10000000	20.0	93%	
	WORD-REL.	Thesaurus	23132	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

35

	Context	Network	n	m	$d^{\circ}$	%mod
	WWW	in-2004	1148875	12281937	21.4	12%
	WWW	cnr-2000	227058	2187201	19.3	19%
	PROTEIN	reactome	5973	145778	48.8	22%
lose to cographs	SOFTWARE	jdk	6434	53658	16.7	29%
iuse lu cugraphis	SOFTWARE	jung-j	6120	50290	16.4	29%
	WWW	eu-2005	835044	15718784	37.7	29%
WWW	CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%
	CO-AUTHOR	ca-HepPh	11204	117619	21.0	34%
	SPECIES	foodweb	183	2434	26.6	43%
software	CO-AUTHOR	dblp	317080	1049866	6.6	45%
	WORD-REL.	wordnet	145145	656230	9.0	48%
	COMMUNIC.	wiki-Talk	2388953	4656682	3.9	49%
	CO-SOLD	amazon	334863	925872	5.5	49%
	CO-AUTHOR	ca-CondMat	21363	91286	8.6	52%
	RANDOM	ER-Gnm_1M-2	796 208	958827	2.4	52%
	CO-AUTHOR	ca-HepTh	8 6 3 8	24 806	5.7	54%
	INTERNET	as2000	6474	12572	3.9	54%
	ROAD	roadNet-TX	1351137	1879201	2.8	54%
	INTERNET	as-caida2007	26475	53381	4.0	55%
	CO-AUTHOR	ca-AstroPh	17903	196972	22.0	59%
	INTERNET	topology	34761	107720	6.2	61%
	RANDOM	ER-Gnm_1M-3	940 987	1494643	3.2	63 %
	INTERNET	as-skitter	1694616	11 094 209	13.1	64%
	CO-OCCUR	bible-names	1 707	9 0 5 9	10.6	67 %
	PROTEIN	figeys	2 217	6 4 1 8	5.8	67 %
	CITATION-SCI.	cora	23166	89157	7.7	68 %
	SOCIAL	youtube	1134890	2987624	5.3	69%
	CO-ACTOR	actor-col.	374511	15014839	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	365154	1 721 981	9.4	75%
	CITATION-PAT.	cit-Patents	3764117	16511740	8.8	76%
	SOFTWARE	linux	30817	213 208	13.8	77%
	SOCIAL	LiveJournal	3997962	34 681 189	17.4	78%
	CITATION-SCI.	cit-HepTh	27400	352 021	25.7	79%
	RANDOM	ER-Gnm_1M-6	997479	2 999 988	6.0	79%
	CITATION-SCI.	cit-HepPh	34 401	420784	24.5	81 %
	RANDOM	ER-Gnm_1M-8	999684	3 999 999	8.0	84 %
	RANDOM	ER-Gnm_1M-10	999952	5 000 000	10.0	87%
	RANDOM	ER-Gnm_1M-15	1 000 000	7500000	10.0 15.0	91%
	SOCIAL	orkut	3072441	117 185 083	76.3	91%
	RANDOM	ER-Gnm_1M-20	1 000 000	10 000 000	20.0	91%
	WORD-REL.	Thesaurus	23 132	297 094	20.0 25.7	93%
	WORD-REL.	rnesaurus	23 132	297 094	20.1	93 70

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

<u>C</u>

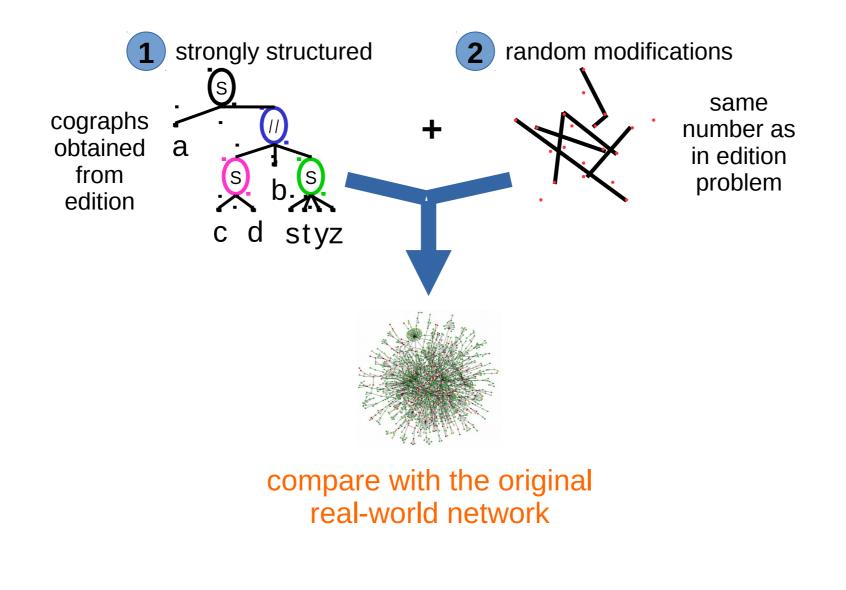
	Context	Network	n	m	d°	%mod
	WWW	in-2004	1148875	12281937	21.4	12%
	WWW	cnr-2000	227058	2187201	19.3	19%
	PROTEIN	reactome	5973	145778	48.8	22%
	SOFTWARE	jdk	6434	53658	16.7	29%
	SOFTWARE	jung-j	6120	50290	16.4	29%
	WWW	eu-2005	835044	15718784	37.7	29%
	CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%
	CO-AUTHOR	ca-HepPh	11204	117619	21.0	34%
	SPECIES	foodweb	183	2434	26.6	43%
	CO-AUTHOR	dblp	317080	1049866	6.6	45%
	WORD-REL.	wordnet	145145	656230	9.0	48%
	COMMUNIC.	wiki-Talk	2388953	4656682	3.9	49%
	CO-SOLD	amazon	334863	925872	5.5	49%
	CO-AUTHOR	ca-CondMat	21363	91286	8.6	52%
	RANDOM	ER-Gnm_1M-2	796208	958827	2.4	52%
	CO-AUTHOR	ca-HepTh	8 6 3 8	24806	5.7	54%
	INTERNET	as2000	6474	12572	3.9	54%
	ROAD	roadNet-TX	1351137	1879201	2.8	54%
<u>Not close not far</u>	INTERNET	as-caida2007	26475	53381	4.0	55%
	CO-AUTHOR	ca-AstroPh	17903	196972	22.0	59%
internet	INTERNET	topology	34761	107720	6.2	61%
	RANDOM	ER-Gnm_1M-3	940987	1494643	3.2	63%
	INTERNET	as-skitter	1694616	11094209	13.1	64%
road	CO-OCCUR	bible-names	1707	9059	10.6	67%
	PROTEIN	figeys	2217	6418	5.8	67%
	CITATION-SCI.	cora	23166	89157	7.7	68%
	SOCIAL	youtube	1134890	2987624	5.3	69%
	CO-ACTOR	actor-col.	374511	15014839	80.2	71%
	P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	71%
	RANDOM	ER-Gnm_1M-4	980191	1999203	4.1	71%
	CITATION-SCI.	citeseer	365154	1721981	9.4	75%
	CITATION-PAT.	cit-Patents	3764117	16511740	8.8	76%
	SOFTWARE	linux	30817	213208	13.8	77%
	SOCIAL	LiveJournal	3997962	34681189	17.4	78%
	CITATION-SCI.	cit-HepTh	27400	352021	25.7	79%
	RANDOM	ER-Gnm_1M-6	997479	2999988	6.0	79%
	CITATION-SCI.	cit-HepPh	34401	420784	24.5	81%
	RANDOM	ER-Gnm_1M-8	999684	3999999	8.0	84%
	RANDOM	ER-Gnm_1M-10	999952	5000000	10.0	87%
	RANDOM	ER-Gnm_1M-15	1000000	7500000	15.0	91%
	SOCIAL	orkut	3072441	117185083	76.3	91%
	RANDOM	ER-Gnm_1M-20	1000000	10000000	20.0	93%
15	WORD-REL.	Thesaurus	23132	297094	25.7	93%

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

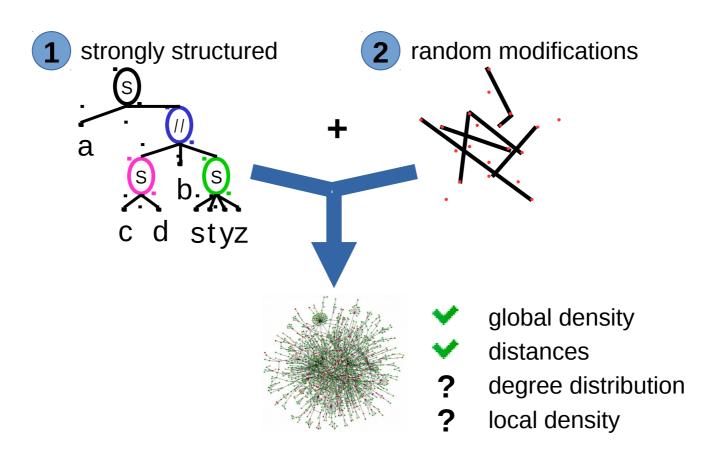
	Context	Network	n	m	$\mathbf{d}^{\circ}$	%mod
	WWW	in-2004	1148875	12281937	21.4	12%
	WWW	cnr-2000	227058	2187201	19.3	19%
	PROTEIN	reactome	5973	145778	48.8	22%
	SOFTWARE	jdk	6434	53658	16.7	29%
	SOFTWARE	jung-j	6120	50290	16.4	29%
	WWW	eu-2005	835044	15718784	37.7	29%
	CO-AUTHOR	ca-GrQc	4158	13422	6.5	34%
	CO-AUTHOR	ca-HepPh	11204	117619	21.0	34%
	SPECIES	foodweb	183	2 4 3 4	26.6	43%
	CO-AUTHOR	dblp	317080	1049866	6.6	45%
	WORD-REL.	wordnet	145145	656230	9.0	48%
	COMMUNIC.	wiki-Talk	2388953	4656682	3.9	49%
	CO-SOLD	amazon	334863	925872	5.5	49%
	CO-AUTHOR	ca-CondMat	21363	91286	8.6	52%
	RANDOM	ER-Gnm_1M-2	796208	958827	2.4	52%
	CO-AUTHOR	ca-HepTh	8 6 3 8	24806	5.7	54%
	INTERNET	as2000	6474	12572	3.9	54%
	ROAD	roadNet-TX	1351137	1879201	2.8	54%
	INTERNET	as-caida2007	26475	53381	4.0	55%
	CO-AUTHOR	ca-AstroPh	17903	196972	22.0	59%
	INTERNET	topology	34761	107 720	6.2	61%
	RANDOM	ER-Gnm_1M-3	940987	1494643	3.2	63%
	INTERNET	as-skitter	1694616	11094209	13.1	64%
	CO-OCCUR	bible-names	1 707	9 0 5 9	10.6	67%
	PROTEIN	figeys	2217	6 4 1 8	5.8	67%
	CITATION-SCI.	cora	23166	89157	7.7	68%
	SOCIAL	youtube	1134890	2987624	5.3	69%
	CO-ACTOR	actor-col.	374511	15014839	80.2	71%
	P2P-CONNECT.	p2p-Gnutella	62561	147 878	4.7	71%
	RANDOM	ER-Gnm_1M-4	980191	1999203	4.1	71%
Far from cographs	CITATION-SCI.	citeseer	365154	1 721 981	9.4	75%
U	CITATION-PAT.	cit-Patents	3764117	16511740	8.8	76%
aitation	SOFTWARE	linux	30817	213208	13.8	77%
citation	SOCIAL	LiveJournal	3997962	34681189	17.4	78%
	CITATION-SCI.	cit-HepTh	27400	352021	25.7	79%
social	RANDOM	ER-Gnm_1M-6	997479	2999988	6.0	79%
300lui	CITATION-SCI.	cit-HepPh	34401	420784	24.5	81 %
	RANDOM	ER-Gnm_1M-8	999684	3999999	8.0	84%
	RANDOM	ER-Gnm_1M-10	999952	5000000	10.0	87%
	RANDOM	ER-Gnm_1M-15	1000000	7500000	15.0	91%
	SOCIAL	orkut	3072441	117185083	76.3	91 %
•	RANDOM	ER-Gnm_1M-20	1000000	10000000	20.0	93%
15	WORD-REL.	Thesaurus	23132	297094	25.7	93%

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

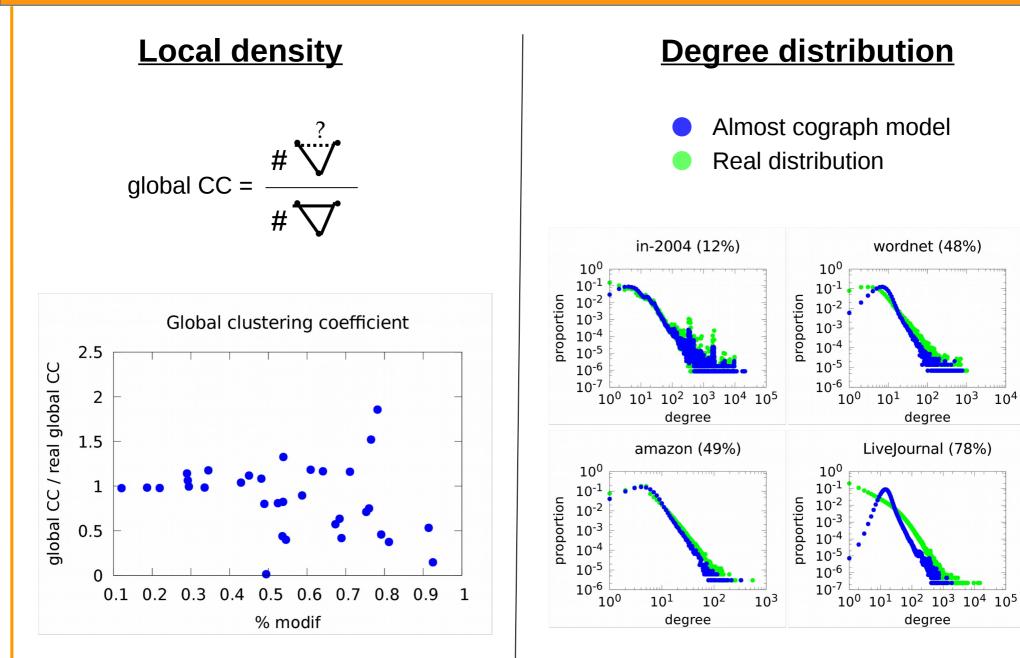
### **Testing the modelling approach**



### Conclusion

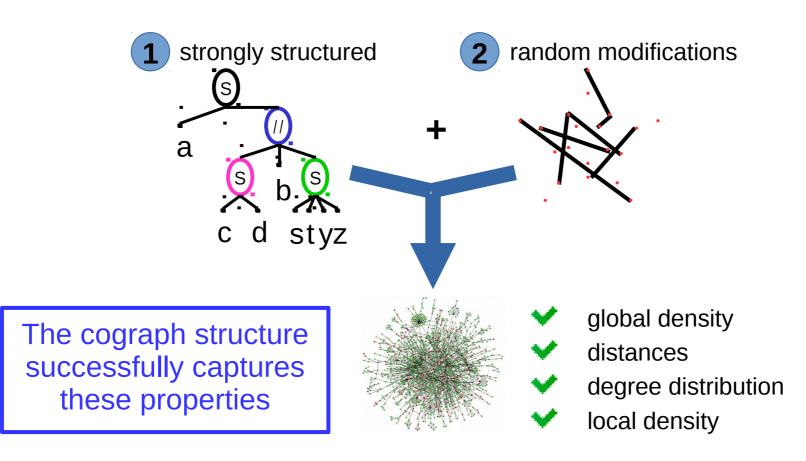


### **Results of generation**

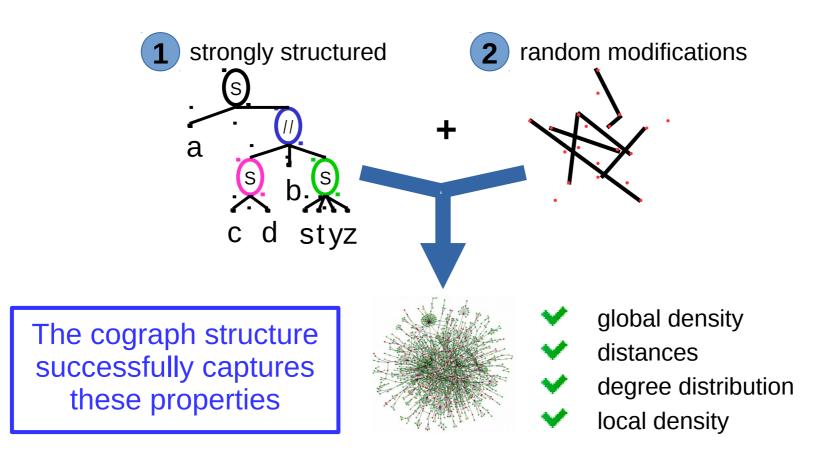


 $10^{4}$ 

### Conclusion



## Conclusion



#### To complete the model

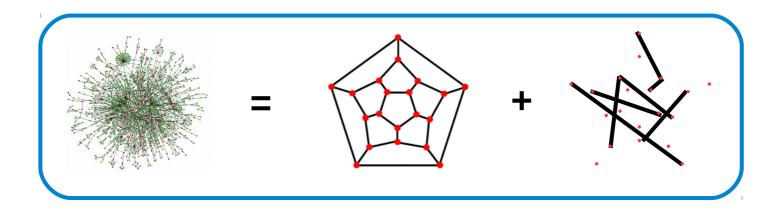
- Edit a real-world graph into a cograph
- Generate a similar cotree
  - Apply random modifications to the cograph

### **Perspectives**

Assess the quality of the set of modifications obtained from the inclusion-minimal approach

Consider other graph classes suitable for other kinds of networks

- Chordal graphs  $\rightarrow$  social networks, citations
- Related to planar graphs  $\rightarrow$  internet, road networks

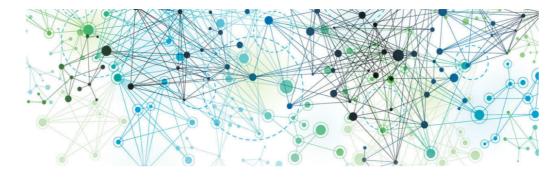


Other possibilities of this representation

- Efficient encoding
- Algorithmics of almost structured graphs

#### PROXNET - Modelling Complex Networks Through Graph Editing Problems

Marie Sklodowska-Curie Actions of the European Union



#### About PROXNET

PROXNET is a project funded by the MSCA program of the European Union. It is hosted at the University of Bergen, with principal researcher Christophe Crespelle and supervised by Pinar Heggernes.

The goal of the PROXNET project is to open a new way for analysing, modelling and managing complex networks, through graph editing problems. The reason why these networks are said to be complex is that they are loosely structured, due to the part of uncertainty and randomness they contain. On the other hand, the real-world context where they come from strongly constrains their organisation and gives them some specific structure. The difficulty in retrieving this structure is that it is altered by the noise resulting from the uncertainty and randomness that these networks contain. In the PROXNET project, we retrieve the hidden structures of complex networks thanks to graph editing problems, which consist in changing some adjacencies of the graph in order to obtain a desired property. We develop the algorithms necessary to solve graph editing problems on huge instances of graphs, we apply them to real-world datasets and use the results obtained in order to design new models of complex networks.

#### **Contact information**

e-mail christophe dot crespelle at uib dot no (dot = ".", at = "@")
Postal address
Postboks 7803
5020 Bergen, Norway
Location Høyteknologisenteret
Thormøhlens Gate 55, Bergen

#### News

01/23/2020	Workshop on Graph Modification: Algorithms, Experiments and New Problems in Bergen, Norway.
06/03/2019	Workshop on Kernelization in Bergen, Norway.
03/04/2019	Conference on Algorithms, Optimization and Learning in Dynamics Environments in Hanoi, Vietnam
11/15/2018	Graph Theory and Applications Workshop in Hanoi, Vietnam.
09/17/2018	Operation Research + Parameterized Complexity Workshop in Solstrand, Norway.
08/09/2018	China-Norway FPT workshop in Bergen, Norway.
03/21/2018	16th Annual Winter School in Algorithms, Graph Theory and Combinatorics in Geilo, Norway.

#### Software

Coedit Minimal completion, deletion and editing of an arbitrary graph into a cograph. Released January 2020. sources

# Coedit: a tool for minimal cograph edge modification

#### **Christophe Crespelle**

University of Bergen

with Daniel Lokshtanov, Thi Ha Duong Phan and Eric Thierry

