Structural sparsity in the real world

Erik Demaine[*], **Felix Reidl**, Peter Rossmanith, Fernando Sánchez Villaamil, Blair D. Sullivan[†] and Somnath Sikdar

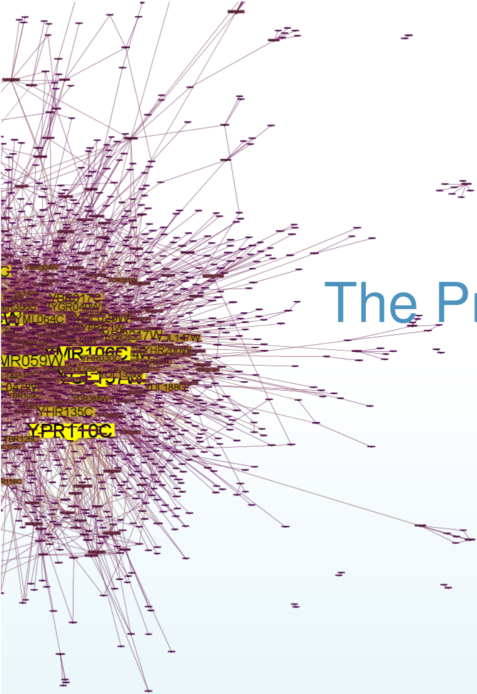
Theoretical Computer Science

RWTH AACHEN

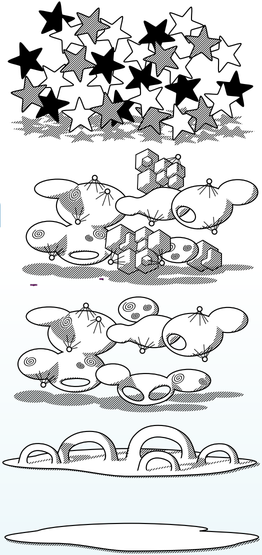
[*]MIT [†]NCSU

@Bergen 2015

Contents



The Program



Complex networks

Ubiquitous in real world

Empirical structure
- Small-world
- Heavy-tailed degree seq.
- Clustering

Algorithmic applications
- Disease spreading
- Attack resilience
- Fraud detection
- Drug discovery

Structural graph theory

Well-researched

Deep structural theorems
- WQO by minor relation
- Decomposition theorems
- Grid-theorem

Great algorithmic properties
- (E)PTAS
- Subexponential algorithms
- Linear kernels
- Model-checking

Can we bring these two fields together?

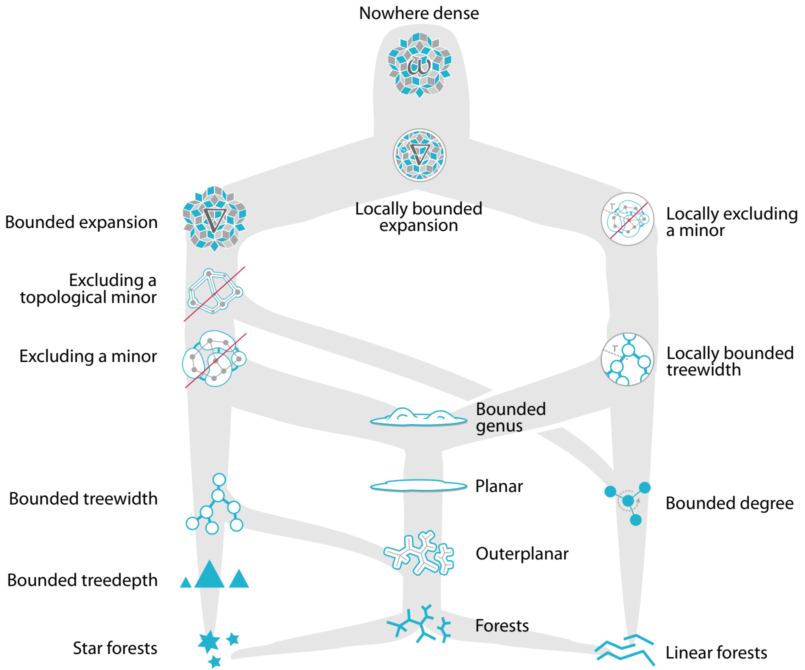
The idea

1. **Bridge the gap** by identifying a notion of sparseness that applies to complex networks.
2. **Develop** algorithmic tools for network related problems.
3. **Show experimentally** that the above is useful in practice.

The idea

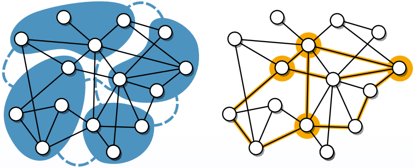
1. **Bridge the gap** by identifying a notion of sparseness that applies to complex networks.
   - Need general and stable notion of sparseness.
   - How to prove that it holds for complex networks?
2. **Develop** algorithmic tools for network related problems.
   - Unclear what problems are interesting.
3. **Show experimentally** that the above is useful in practice.
   - Show that structural sparseness appears in the real world.
   - Show that algorithms can compete with known approaches.

Structural Sparseness


Nowhere dense
Bounded expansion
Locally bounded expansion
Locally excluding a minor
Excluding a topological minor
Excluding a minor
Locally bounded treewidth
Bounded genus
Planar
Bounded treewidth
Bounded degree
Outerplanar
Bounded treedepth
Forests
Star forests
Linear forests

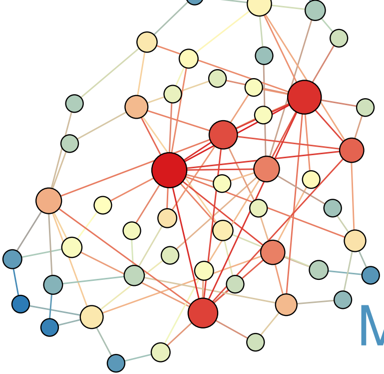
Bounded expansion

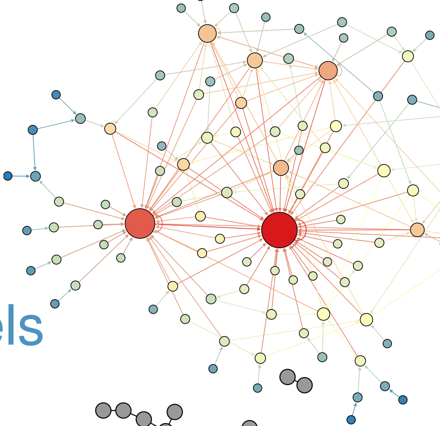
A graph class has bounded expansion if the density of its minors only depends on their depth.



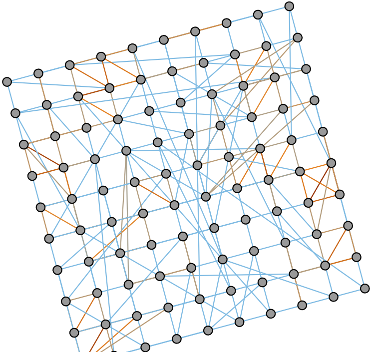
The following operations on a class of bounded expansion result again in a class of bounded expansion:

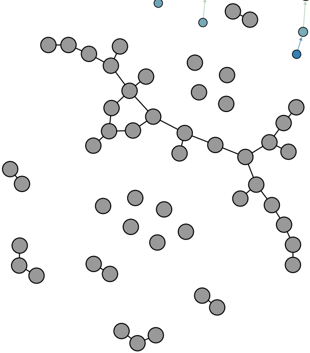
- Taking shallow minors/immersions (in particular subgraphs)
- Adding a universal vertex
- Replacing each vertex by a small clique (lexicographic product)

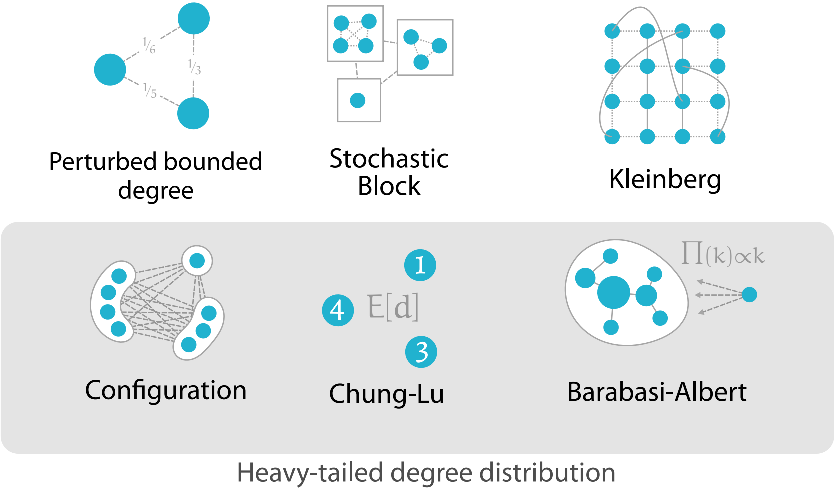




Models

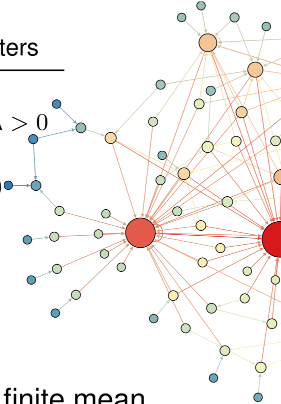





1/6
1/3
1/5
Perturbed bounded degree
Stochastic Block
Kleinberg
Configuration
1
4
E[d]
3
Chung-Lu
Π(k)∝k
Barabasi-Albert
Heavy-tailed degree distribution

The positive side

Name	Definition $f(d)$	Parameters
Power law	$d^{-\gamma}$	$\gamma > 2$
Power law w/ cutoff	$d^{-\gamma}e^{-\lambda d}$	$\gamma > 2$, $\lambda > 0$
Exponential	$e^{-\lambda d}$	$\lambda > 0$
Stretched exponential	$d^{\beta-1}e^{-\lambda d^{\beta}}$	$\lambda, \beta > 0$
Gaussian	$\exp(-\frac{(d-\mu)^2}{2\sigma^2})$	μ, σ
Log-normal	$d^{-1}\exp(-\frac{(\log d-\mu)^2}{2\sigma^2})$	μ, σ



Theorem

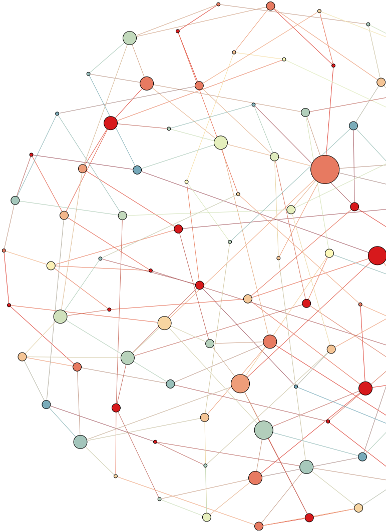
Let $\mathcal{D}$ be an asymptotic degree distribution with finite mean. Then random graphs generated by the Configuration Model or the Chung-Lu model with parameter $\mathcal{D}$ have bounded expansion with high probability.

The positive side

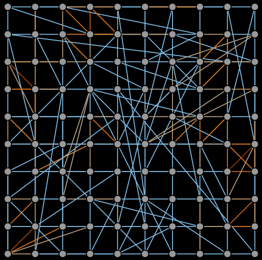
Theorem

The perturbed bounded degree model has bounded expansion with high probability.

Perturbing forests of $S_{\sqrt{n}}$ results in a somewhere dense class.

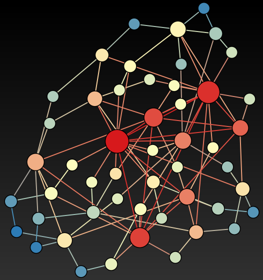


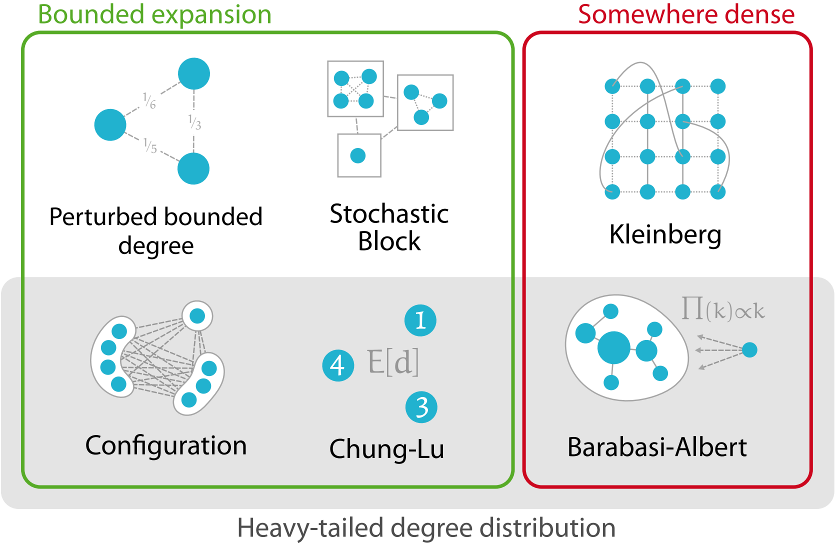
The negative side



Theorem
The Kleinberg Model is somewhere dense with high probability.

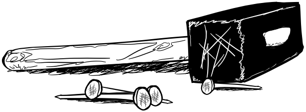
Theorem
The Barabási-Albert Model is somewhere dense with non-vanishing probability.





Bounded expansion
Somewhere dense
1/6
1/3
1/5
Perturbed bounded degree
Stochastic Block
Kleinberg
Configuration
1
4
E[d]
3
Chung-Lu
Π(k)∝k
Barabasi-Albert
Heavy-tailed degree distribution

Algorithms



Neighbourhood sizes

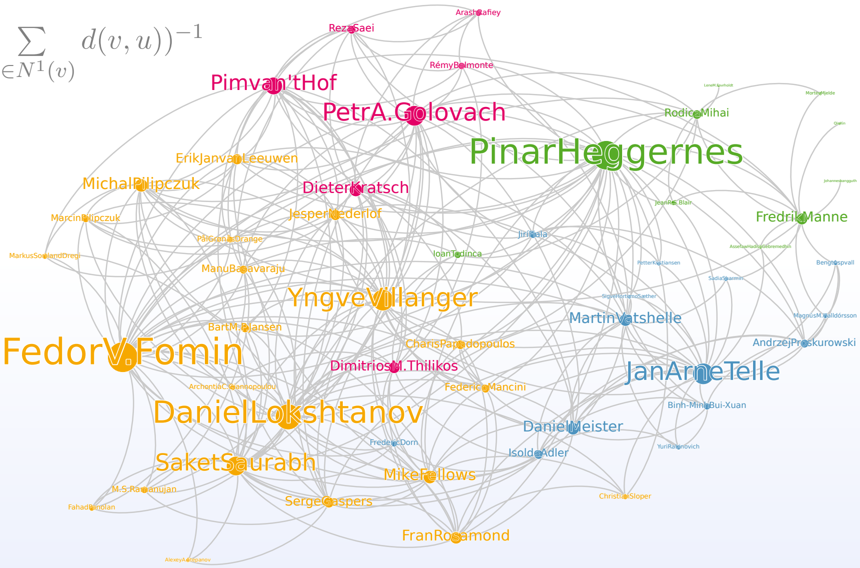
Measure	Definition	Localized
Closeness	$(\sum_{u \in V(G)} d(v,u))^{-1}$	$(\sum_{u \in N^r(v)} d(v,u))^{-1}$
Harmonic	$\sum_{u \in V(G)} d(v,u)^{-1}$	$\sum_{u \in N^r(v)} d(v,u)^{-1}$
Lin's index	$\frac{\lvert\{v \mid d(v,v) < \infty\}\rvert^2}{\sum_{u \in V(G): d(v,u) < \infty} d(v,u)}$	$\frac{\lvert N^r[v]\rvert^2}{\sum_{u \in N^r[v]} d(v,u)}$

Theorem

Let $\mathcal{G}$ be a graph class of bounded expansion. There is an algorithm that for every $r \in \mathbf{N}$ and $G \in \mathcal{G}$ computes the size of the i-th neighbourhood of every vertex of G, for all $i \leq r$, in linear time.

Closeness centrality

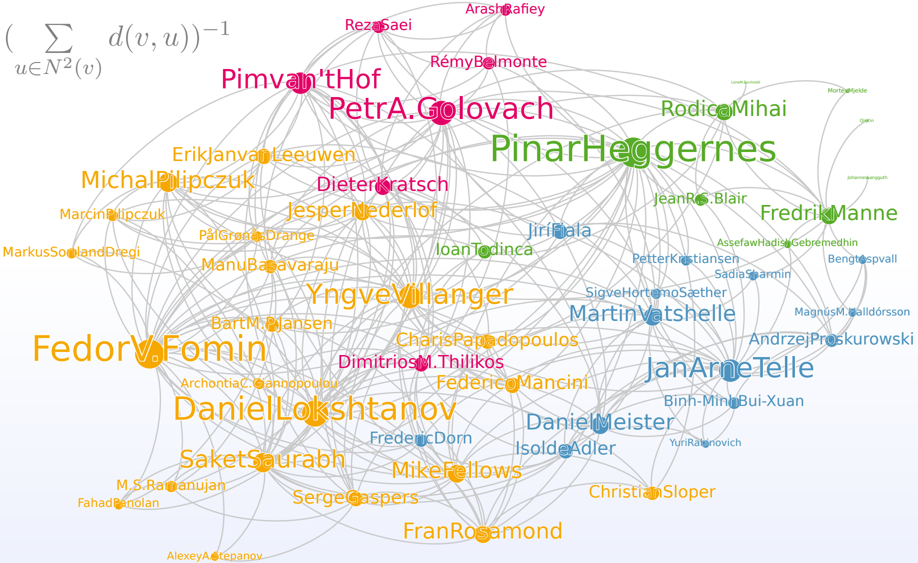
$$\left(\sum_{u\in N^1(v)} d(v,u)\right)^{-1}$$



Network provided by Pål

Closeness centrality

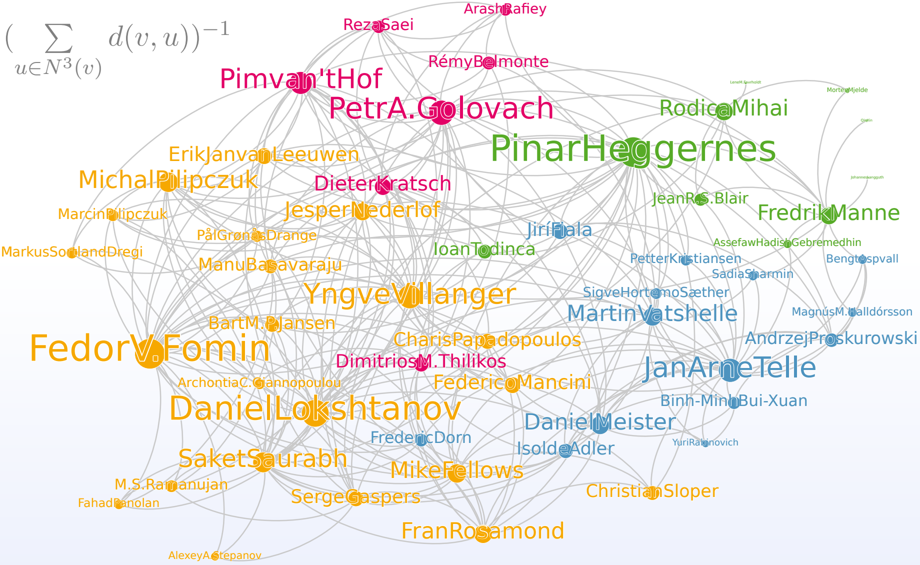
$$\left(\sum_{u \in N^2(v)} d(v,u)\right)^{-1}$$



Network provided by Pål

Closeness centrality

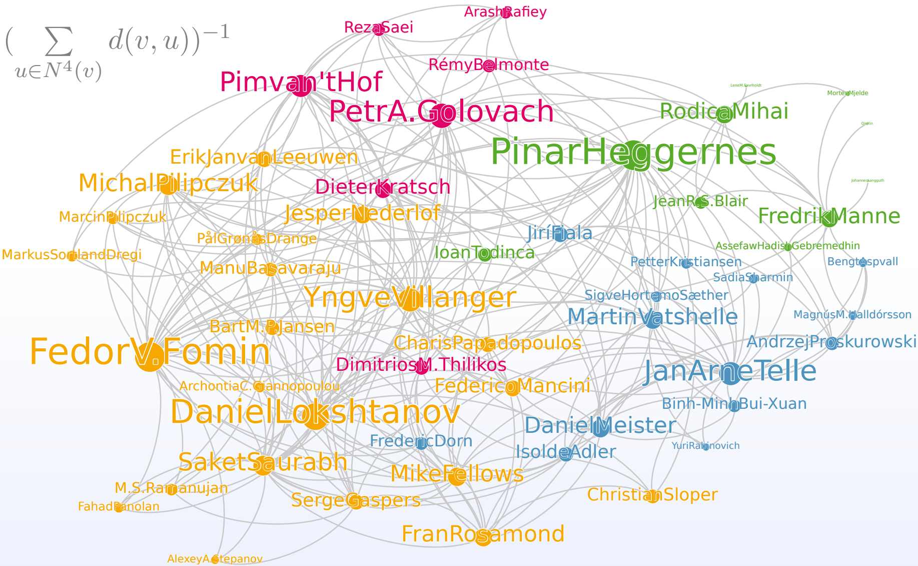
$$\left(\sum_{u \in N^3(v)} d(v,u)\right)^{-1}$$



Network provided by Pål

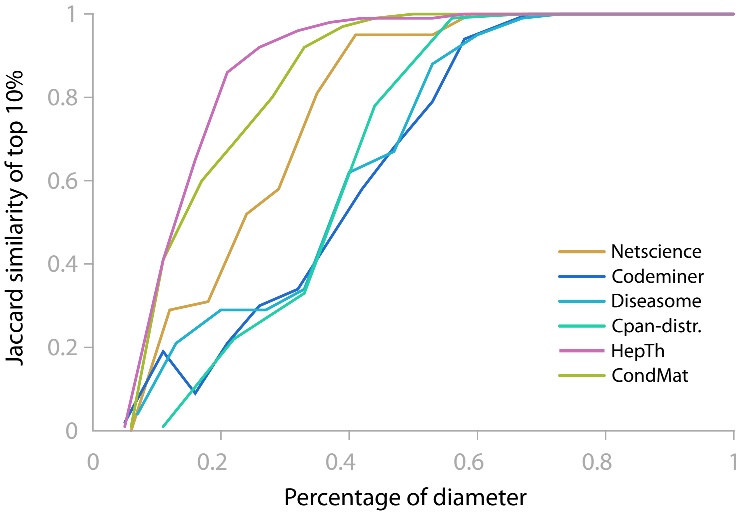
Closeness centrality

$$\Big(\sum_{u \in N^4(v)} d(v,u)\Big)^{-1}$$

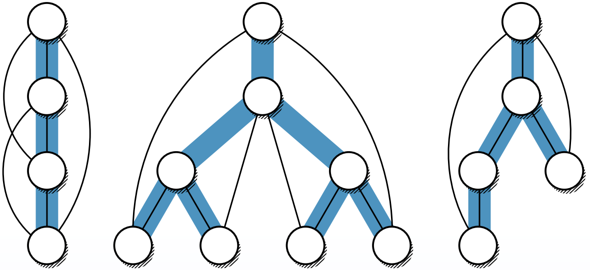


Network provided by Pål

Top-10% recovery



Counting substructures



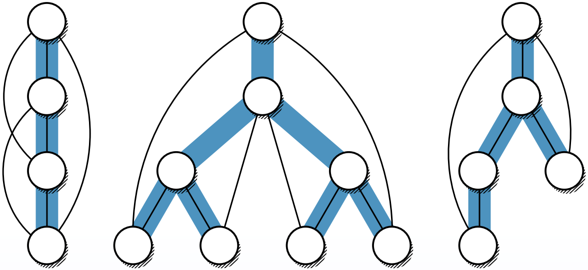
Theorem

Given a graph H on h vertices, a graph G on n vertices and a treedepth decomposition of G of height t, one can compute the

- number of isomorphisms from H to subgraphs of G,
- homomorphisms from H to subgraphs of G, or
- (induced) subgraphs of G isomorphic to H

in time $O(8^h \cdot t^h \cdot h^2 \cdot n)$ and space $O(4^h \cdot t^h \cdot ht \cdot \log n)$.

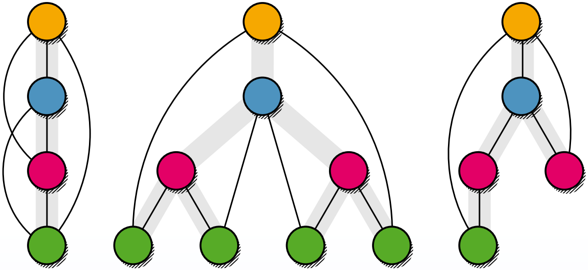
Counting substructures



Theorem (Nešetřil & Ossona de Mendez)

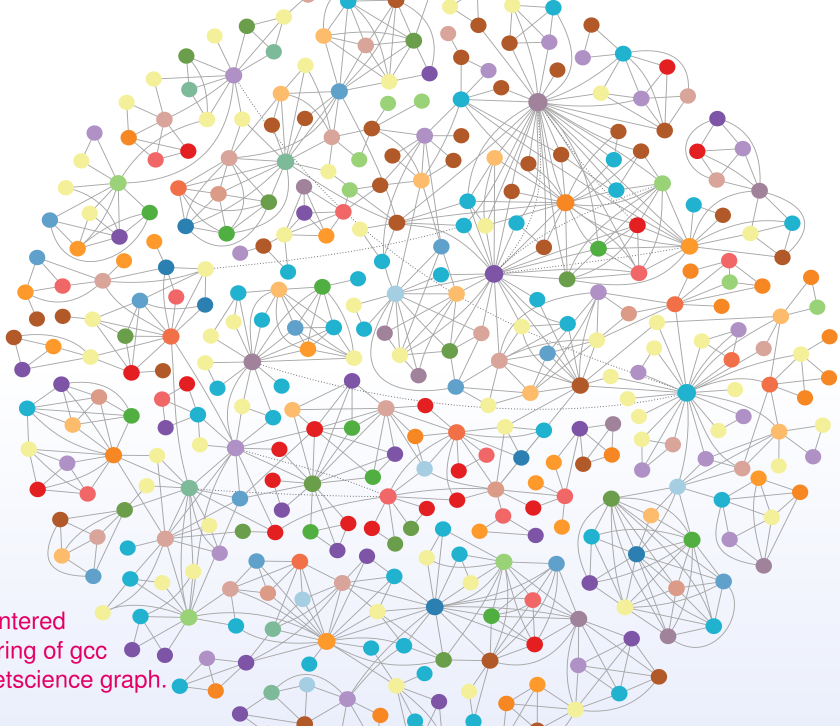
Let $\mathcal{G}$ be class of bounded expansion. There exists a function f such thaht for every p, every member of $\mathcal{G}$ has a p-centered coloring with at most $f(p)$ colors. Moreover, such a coloring can be computed in linear time.

Counting substructures

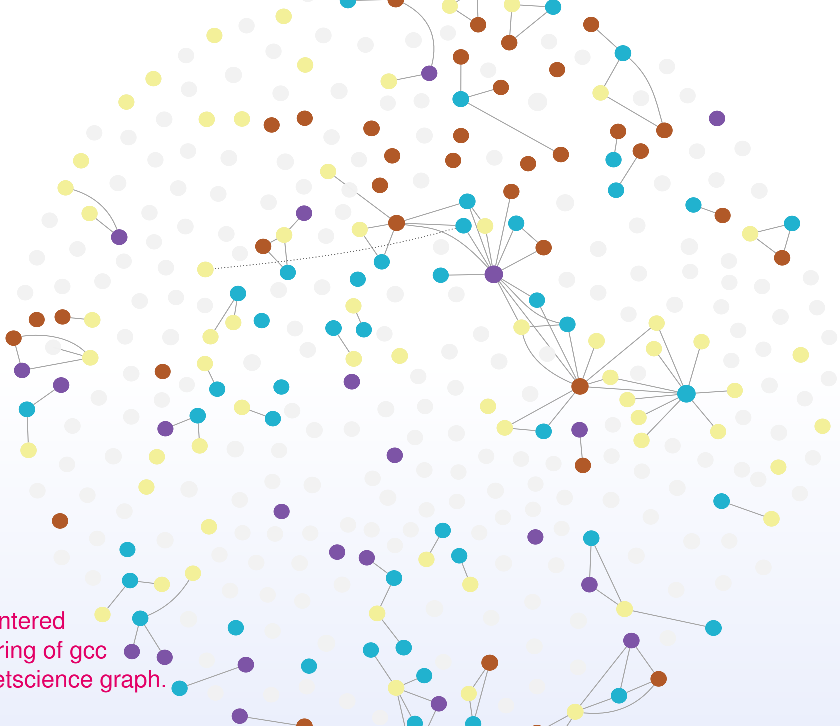


Theorem (Nešetřil & Ossona de Mendez)

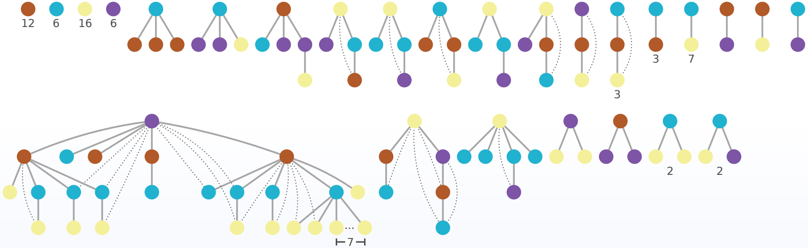
Let $\mathcal{G}$ be class of bounded expansion. There exists a function f such thaht for every p, every member of $\mathcal{G}$ has a p-centered coloring with at most $f(p)$ colors. Moreover, such a coloring can be computed in linear time.



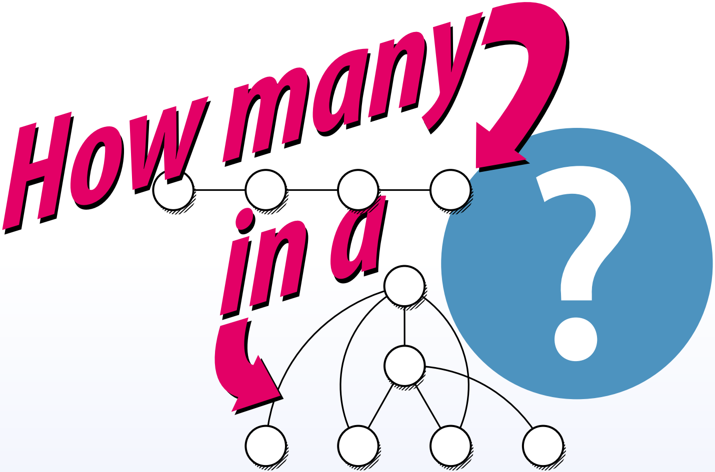
5-centered
coloring of gcc
of netscience graph.



5-centered
coloring of gcc
of netscience graph.

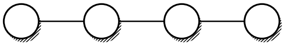
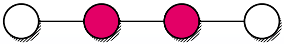
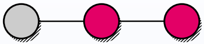


5-centered
coloring of gcc
of netscience graph.


How many
in a
?

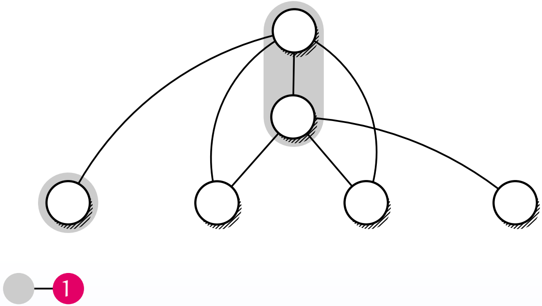
Example: Counting P_4s

Preprocessing: create k-Patterns (here: $k = 2$)

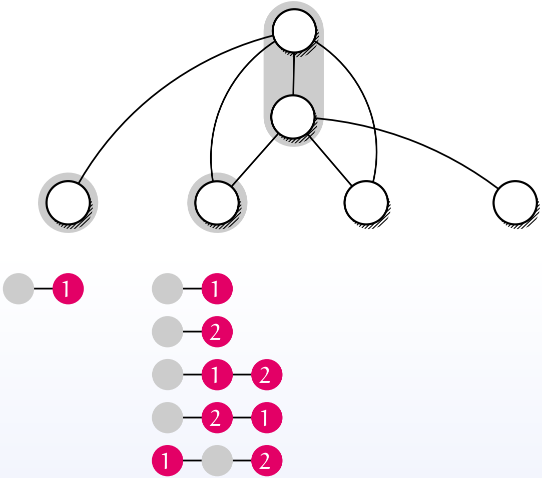
-  Take pattern graph P_4
-  Choose separator ■
-  Choose component ■
- 

Label separator

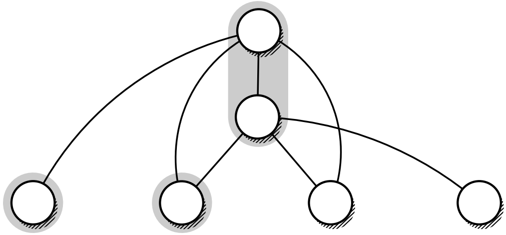
Example: Counting P_4s

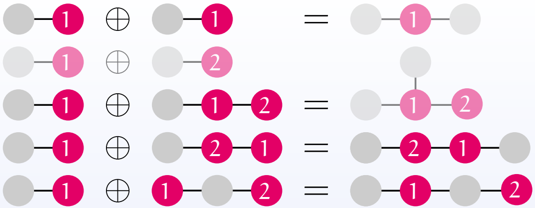


Example: Counting P_4s

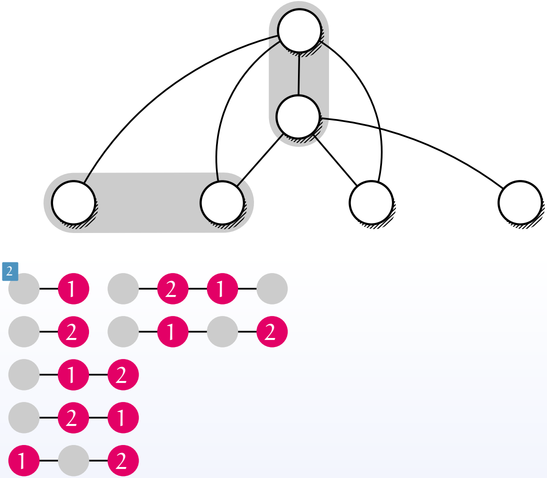


Example: Counting P_4s

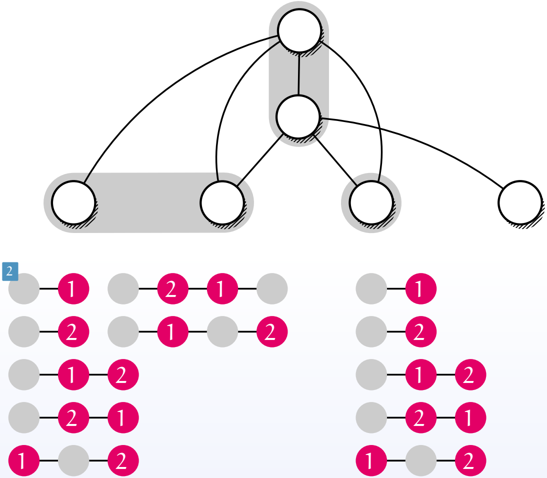




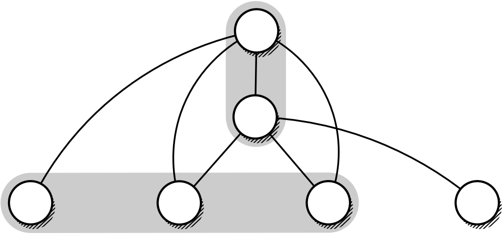
Example: Counting P_4s

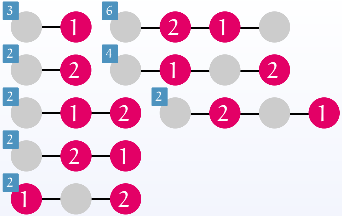


Example: Counting P_4s

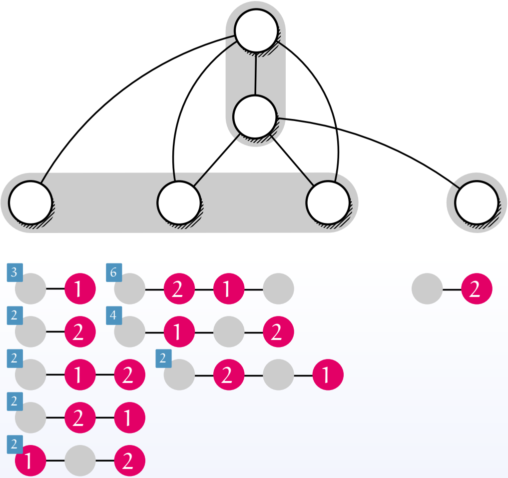


Example: Counting P_4s

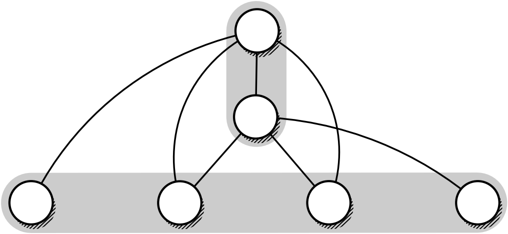


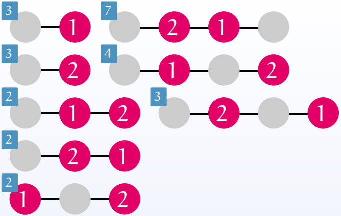


Example: Counting P_4s

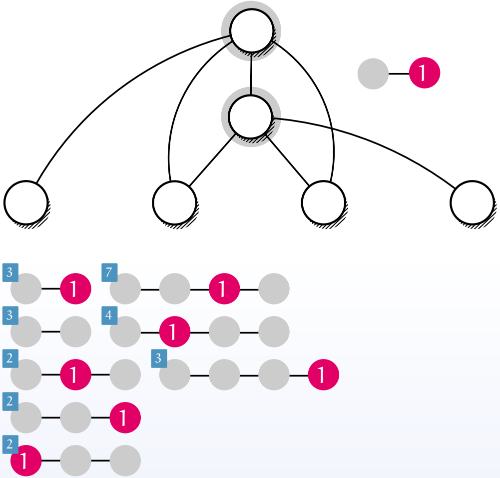


Example: Counting P_4s

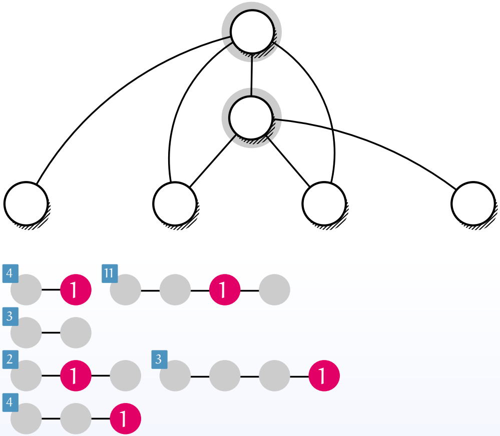




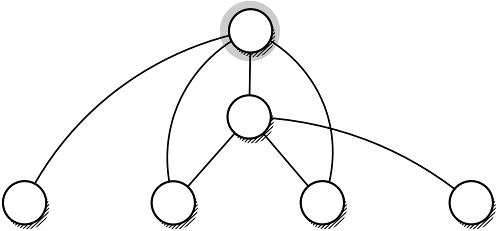
Example: Counting P_4s

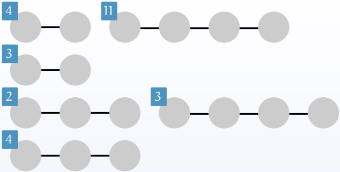


Example: Counting P_4s

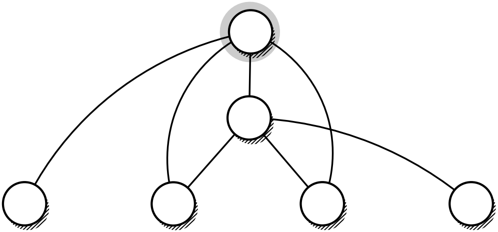


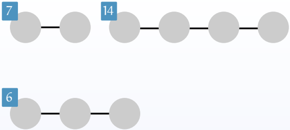
Example: Counting P_4s



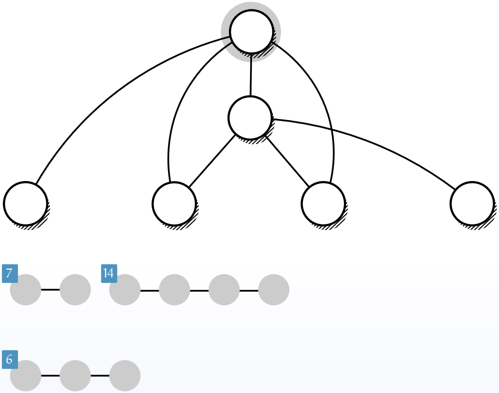


Example: Counting P_4s

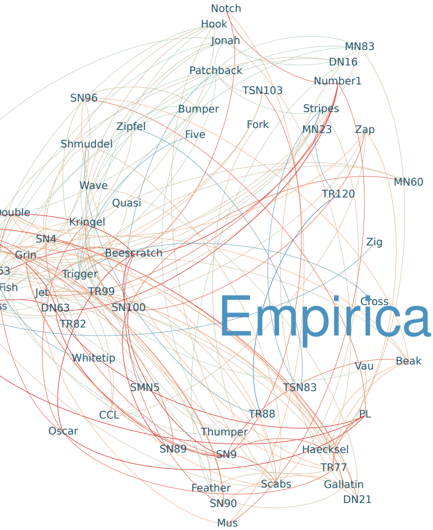




Example: Counting P_4s



There are seven P_4s in the target graph.



Empirical Sparseness

Closing the gap

In order to claim that our approach is useful in practice we cannot just rely on theory.

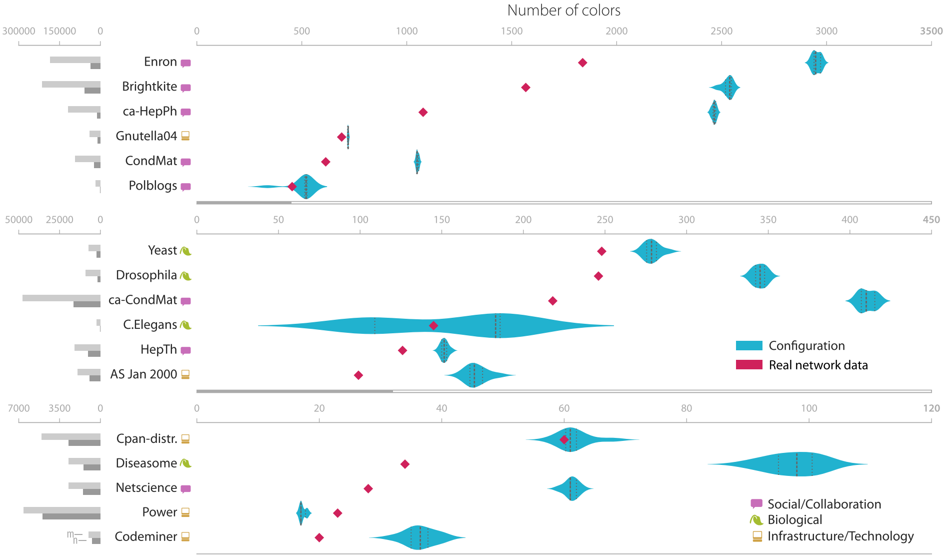
- Graph classes vs. concrete instances
- The bounds given by our proofs are enormous.
- Random graph models capture only some aspectes of complex networks.
- We prove asymptotic bounds.
  (although we show fast convergence)

Network	Vertices	Edges	p					
			2	3	4	5	6	∞
Airlines	235	1297	11	28	39	47	55	64
C.Elegans	306	2148	8	36	74	83	118	153
Codeminer	724	1017	5	10	15	17	23	51
Cpan-authors	839	2212	9	24	34	43	47	224
Diseasome	1419	2738	12	17	22	25	30	30
Polblogs	1491	16715	30	118	286	354	392	603
Netscience	1589	2742	20	20	28	28	28	20
Drosophila	1781	8911	12	65	137	188	263	395
Yeast	2284	6646	12	38	178	254	431	408
Cpan-distr.	2719	5016	5	14	32	42	56	224
Twittercrawl	3656	154824	89	561	1206	1285	1341	–
Power	4941	6594	6	12	20	21	34	95
AS Jan 2000	6474	13895	12	29	70	102	151	357
Hep-th	7610	15751	24	25	104	328	360	558
Gnutella04	10876	39994	8	43	626	–	–	–
ca-HepPh	12008	118489	239	296	1002	–	–	–
CondMat	16264	47594	18	47	255	1839	–	1310
ca-CondMat	23133	93497	26	89	665	–	–	–
Enron	36692	183831	27	214	1428	–	–	–
Brightkite	58228	214078	39	193	1421	–	–	–

Network	Vertices	Edges	p					
			2	3	4	5	6	∞
Airlines	235	1297	11	28	39	47	55	64
Power	4941	6594	6	12	20	21	34	95
AS Jan 2000	6474	13895	12	29	70	102	151	357
C.Elegans	306	2148	8	36	74	83	118	153
Diseasome	1419	2738	12	17	22	25	30	30
Drosophila	1781	8911	12	65	137	188	263	395
Yeast	2284	6646	12	38	178	254	431	408
Codeminer	724	1017	5	10	15	17	23	51
Gnutella04	10876	39994	8	43	626	–	–	–
Enron	36692	183831	27	214	1428	–	–	–
Brightkite	58228	214078	39	193	1421	–	–	–
Cpan-authors	839	2212	9	24	34	43	47	224
Polblogs	1491	16715	30	118	286	354	392	603
Netscience	1589	2742	20	20	28	28	28	20
Cpan-distr.	2719	5016	5	14	32	42	56	224
Twittercrawl	3656	154824	89	561	1206	1285	1341	–
Hep-th	7610	15751	24	25	104	328	360	558
ca-HepPh	12008	118489	239	296	1002	–	–	–
CondMat	16264	47594	18	47	255	1839	–	1310
ca-CondMat	23133	93497	26	89	665	–	–	–

			p					
Network	Vertices	Edges	2	3	4	5	6	∞
Airlines	235	1297	1.00	2.55	3.55	4.27	5.00	5.82
Power	4941	6594	1.00	2.00	3.33	3.50	5.67	15.83
AS Jan 2000	6474	13895	1.00	2.42	5.83	8.50	12.58	29.75
C.Elegans	306	2148	1.00	4.50	9.25	10.38	14.75	19.12
Diseasome	1419	2738	1.00	1.42	1.83	2.08	2.50	2.50
Drosophila	1781	8911	1.00	5.42	11.42	15.67	21.92	32.92
Yeast	2284	6646	1.00	3.17	14.83	21.17	35.92	34.00
Codeminer	724	1017	1.00	2.00	3.00	3.40	4.60	10.20
Gnutella04	10876	39994	1.00	5.38	78.25	–	–	–
Enron	36692	183831	1.00	7.93	52.89	–	–	–
Brightkite	58228	214078	1.00	4.95	36.44	–	–	–
Cpan-authors	839	2212	1.00	2.67	3.78	4.78	5.22	24.89
Polblogs	1491	16715	1.00	3.93	9.53	11.80	13.07	20.10
Netscience	1589	2742	1.00	1.00	1.40	1.40	1.40	1.00
Cpan-distr.	2719	5016	1.00	2.80	6.40	8.40	11.20	44.80
Twittercrawl	3656	154824	1.00	6.30	13.55	14.44	15.07	–
Hep-th	7610	15751	1.00	1.04	4.33	13.67	15.00	23.25
ca-HepPh	12008	118489	1.00	1.24	4.19	–	–	–
CondMat	16264	47594	1.00	2.61	14.17	102.17	–	72.78
ca-CondMat	23133	93497	1.00	3.42	25.58	–	–	–

Network structure



Conclusion

- We show that several important models of complex networks have bounded expansion.
- Besides the known algorithms (first-order model checking!) we show that relevant problems can be solved faster by using this fact.
- Our experiments demonstrate that many networks are structurally sparse.

Conclusion

- We show that several important models of complex networks have bounded expansion.
- Besides the known algorithms (first-order model checking!) we show that relevant problems can be solved faster by using this fact.
- Our experiments demonstrate that many networks are structurally sparse.

THANKS!

Questions?