Structural sparsity in the real world

Felix Reidl

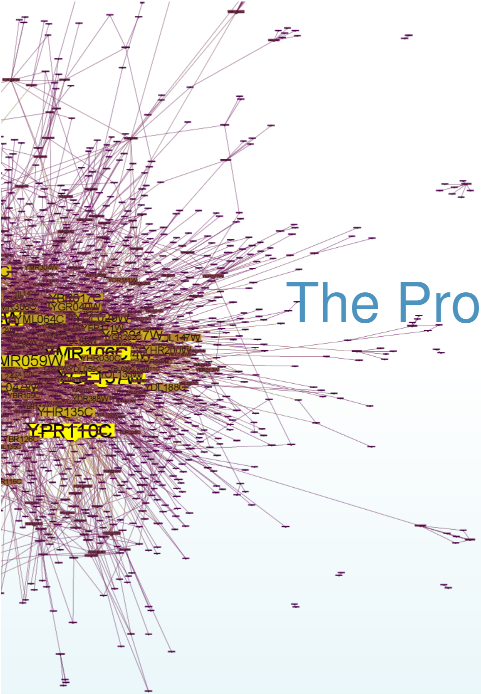
Theoretical Computer Science

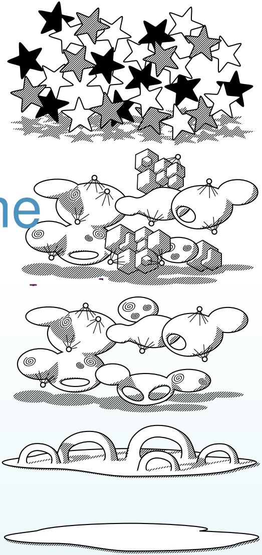
RWTH AACHEN

@abc-Workshop 2015

Contents

The Programme





Preface

The following contains results from the following papers, hence the respective co-authors deserve credit:

- Structural sparsity of complex networks: Bounded expansion in random models and real-world graphs.
  Erik D. Demaine, FR, P. Rossmanith, F. Sánchez Villaamil, S. Sikdar, and B. D. Sullivan.
- Hyperbolicity, degeneracy, and expansion of random intersectiongraphs.
  M. Farrel, T. D. Goodrich, N. Lemons, FR, F. Sánchez Villaamil, and B. D. Sullivan.
- Kernelization using structural parameters on sparse graph classes.
  J. Gajarský, P. Hliněný, J. Obdržálek, S. Ordyniak, FR, P. Rossmanith, F. Sánchez Villaamil, and S. Sikdar.
- Kernelization and sparseness: the case of dominating set.
  P. G. Drange, M. S. Dregi, F. V. Fomin, S. Kreutzer, D. Lokshtanov, M. Pilipczuk, M. Pilipczuk, FR, S. Saurabh, F. Sánchez Villaamil, and S. Sikdar.

The whole story can (soon) be found in my thesis :)

My motivation

- We have huge amounts of network data from various fields
    - Friendships, collaborations, face-to-face interaction,...
    - Protein-protein interaction, food webs, brain networks,...
    - Communication patterns, transportation, ...

My motivation

- We have huge amounts of network data from various fields
  - Friendships, collaborations, face-to-face interaction,...
  - Protein-protein interaction, food webs, brain networks,...
  - Communication patterns, transportation, ...
- We have a lot of algorithmic questions regarding such data

My motivation

- We have huge amounts of network data from various fields
  - Friendships, collaborations, face-to-face interaction,...
  - Protein-protein interaction, food webs, brain networks,...
  - Communication patterns, transportation, ...
- We have a lot of algorithmic questions regarding such data, e.g., motif discovery, centrality of members, propagation of information or diseases.

My motivation

- We have huge amounts of network data from various fields
  - Friendships, collaborations, face-to-face interaction,...
  - Protein-protein interaction, food webs, brain networks,...
  - Communication patterns, transportation, ...
- We have a lot of algorithmic questions regarding such data, e.g., motif discovery, centrality of members, propagation of information or diseases.
- We know—empirically—that these networks are sparse...

My motivation

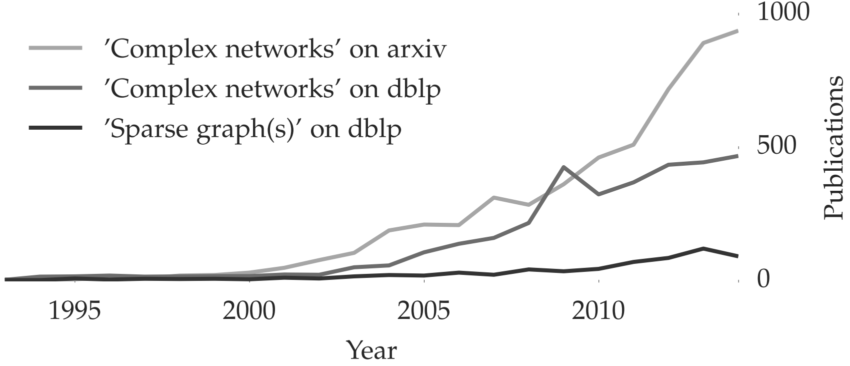
- We have huge amounts of network data from various fields
  - Friendships, collaborations, face-to-face interaction,...
  - Protein-protein interaction, food webs, brain networks,...
  - Communication patterns, transportation, ...
- We have a lot of algorithmic questions regarding such data, e.g., motif discovery, centrality of members, propagation of information or diseases.
- We know—empirically—that these networks are sparse...
  ...and sparse graphs have good algorithmic properties!

My motivation

- We have huge amounts of network data from various fields
  - Friendships, collaborations, face-to-face interaction,...
  - Protein-protein interaction, food webs, brain networks,...
  - Communication patterns, transportation, ...
- We have a lot of algorithmic questions regarding such data, e.g., motif discovery, centrality of members, propagation of information or diseases.
- We know—empirically—that these networks are sparse...
  ...and sparse graphs have good algorithmic properties!

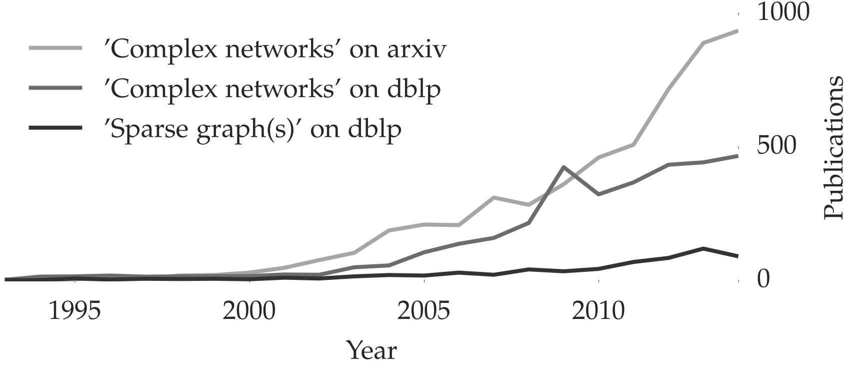
The perfect playground for sparse graph theory!

...Why?



'Complex networks' on arxiv
'Complex networks' on dblp
'Sparse graph(s)' on dblp
1000
500
0
Publications
1995
2000
2005
2010
Year

...Why?



Sparseness $\neq$ structural sparseness!

The Programme

1. **Bridge the gap** by identifying a notion of structural 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 Programme

1. **Bridge the gap** by identifying a notion of structural sparseness that applies to complex networks.
   - Many notions of sparseness (e.g. planar) too strict!
   - How to *prove* sparseness for complex networks?
2. **Develop** algorithmic tools for network related problems.
3. **Show experimentally** that the above is useful in practice.

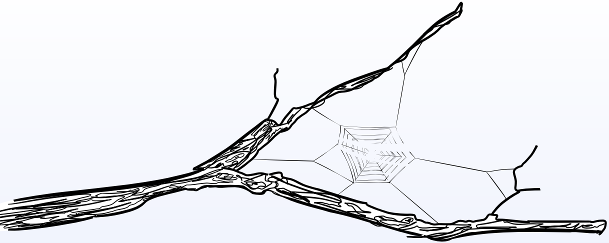
The Programme

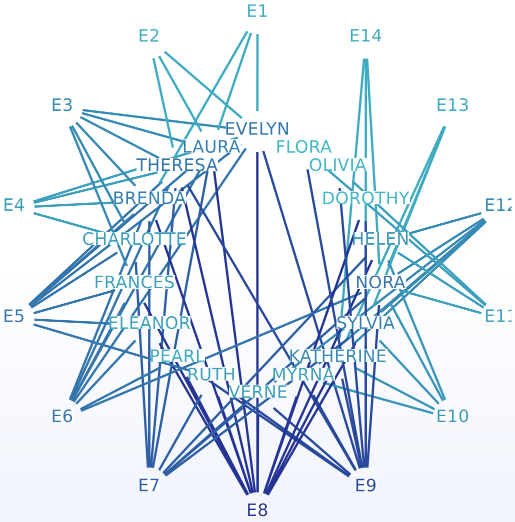
1. **Bridge the gap** by identifying a notion of structural sparseness that applies to complex networks.
   - Many notions of sparseness (e.g. planar) too strict!
   - How to *prove* sparseness for complex networks?
2. **Develop** algorithmic tools for network related problems.
   - Unclear what problems are interesting.
3. **Show experimentally** that the above is useful in practice.

The Programme

1. **Bridge the gap** by identifying a notion of structural sparseness that applies to complex networks.
   - Many notions of sparseness (e.g. planar) too strict!
   - How to *prove* sparseness 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.

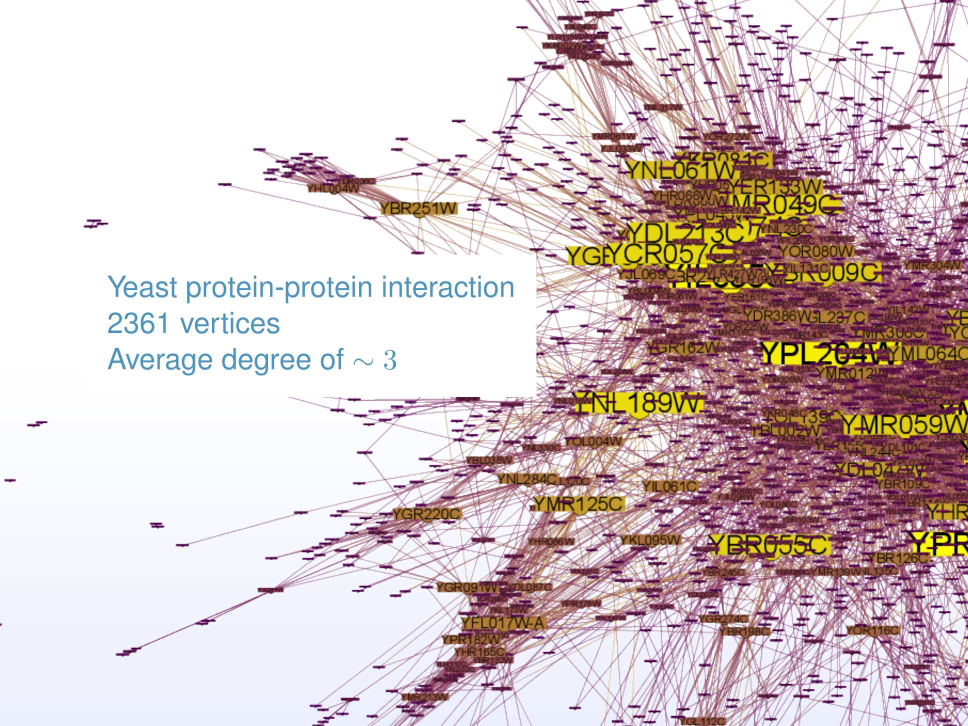
Complex Networks: Examples

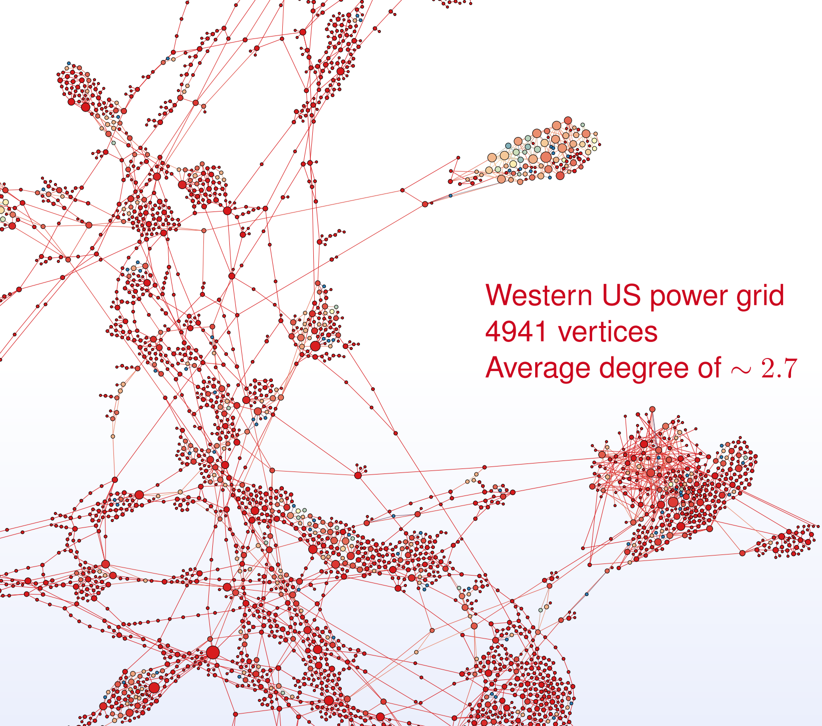




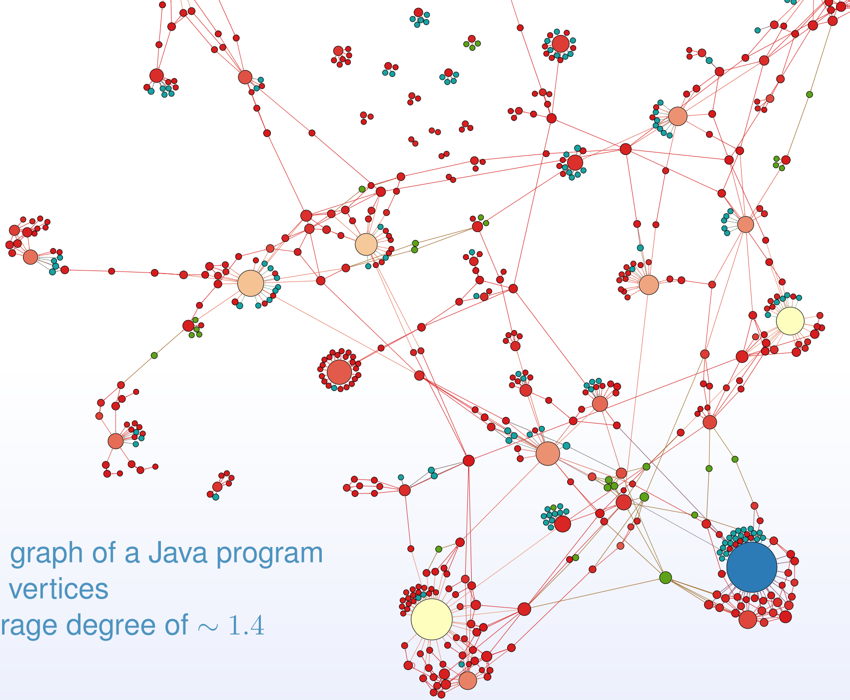
Southern Women
Davis et al., 1930
18 women
14 events over 9 month

Yeast protein-protein interaction
2361 vertices
Average degree of ~ 3

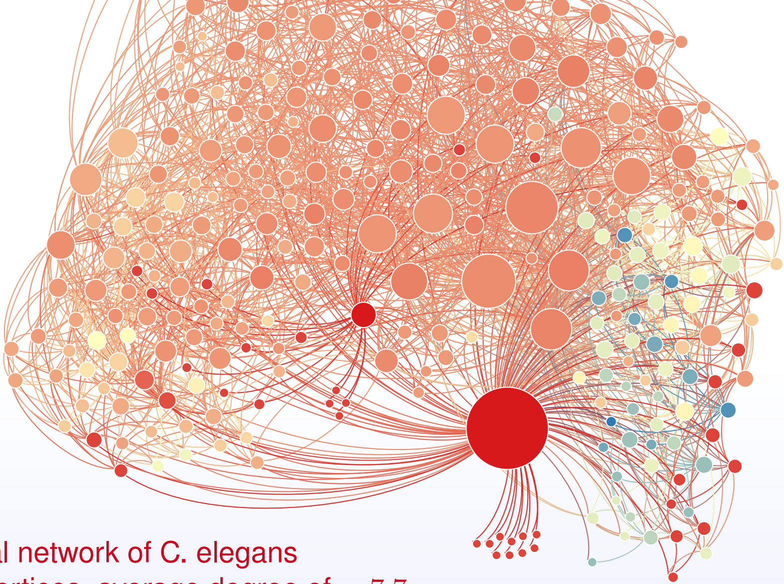




Western US power grid
4941 vertices
Average degree of ~ 2.7

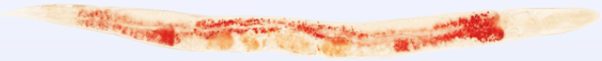


Call graph of a Java program
724 vertices
Average degree of ~ 1.4

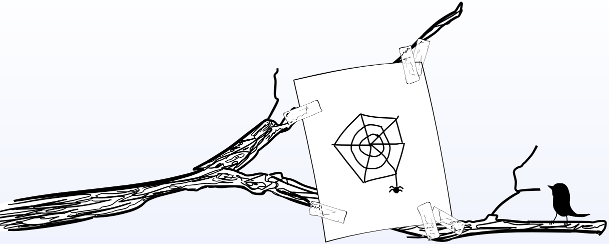


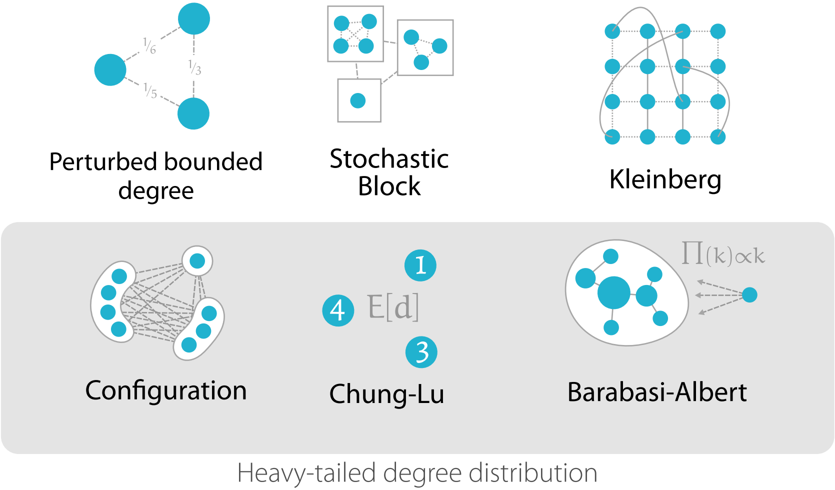
Neural network of C. elegans

297 vertices, average degree of ~ 7.7



Network models



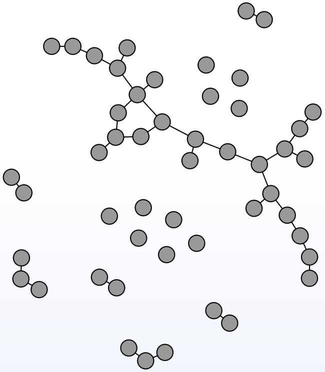


1/6
1/3
1/5
Perturbed bounded degree
Stochastic Block
Kleinberg
Configuration
1
4
E[d]
3
Chung-Lu
Π(k)∝k
Barabasi-Albert
Heavy-tailed degree distribution

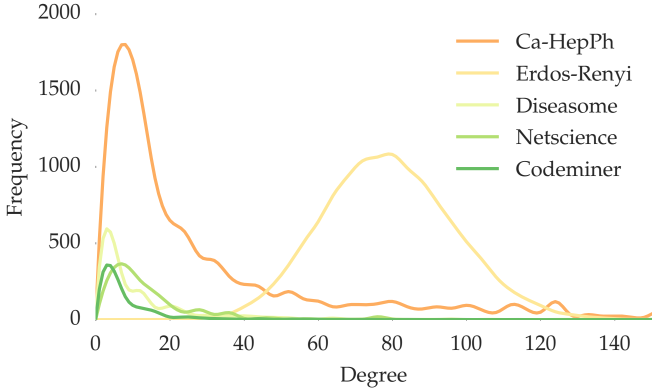
Erdős-Rényi

$G(n,p)$: n-vertex graph in which every edge is present with probability p. For sparse graphs, we want $np = O(1)$.

- Well-understood
- Simple model
- Clustering $\sim p$
- Degree distribution too symmetric and concentrated



Degree distributions

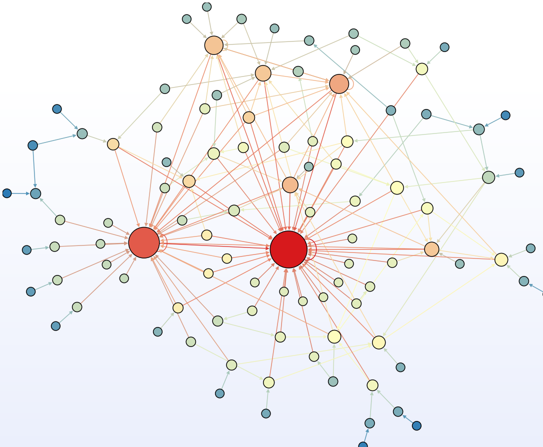


Power law	$d^{-\gamma}$	Power law w/ cutoff	$d^{-\gamma} e^{-\lambda d}$
Exponential	$e^{-\lambda d}$	Stretched exponential	$d^{\beta-1} e^{-\lambda d^{\beta}}$
Gaussian	$\exp(-\frac{(d-\mu)^2}{2\sigma^2})$	Log-normal	$d^{-1} \exp(-\frac{(\log d-\mu)^2}{2\sigma^2})$

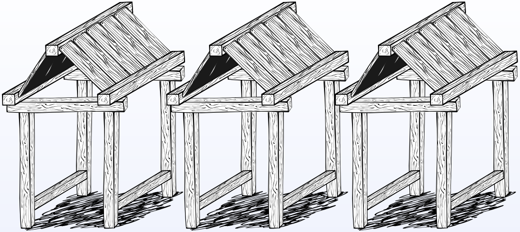
Chung-Lu / Configuration model

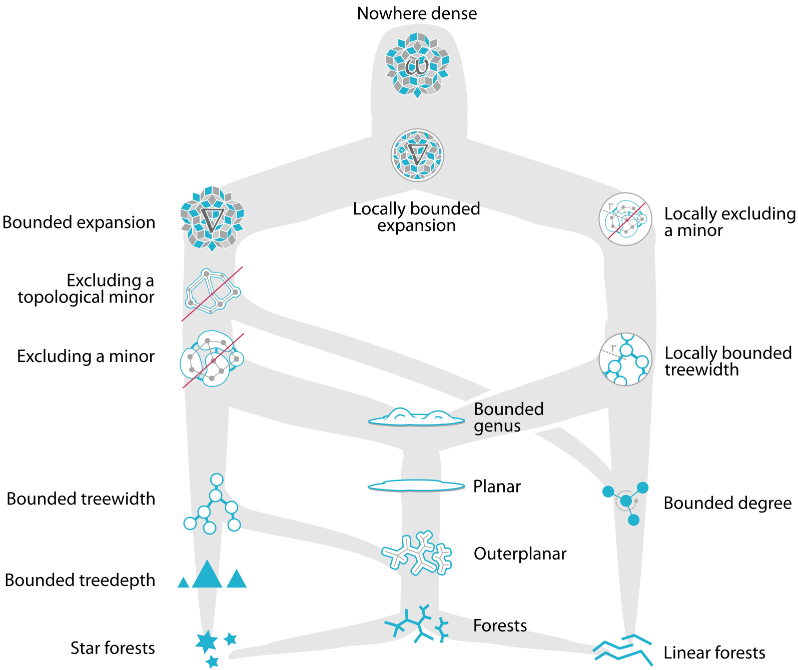
Fix a degree-distribution. Create a degree sequence $d_1, \ldots, d_n$ for n vertices. Now connect each pair of vertices u, v with probability $d_u d_v / \sum_i d_i$ independently at random.
(Configuration model slightly different)

- Simple model
- Very flexible
- Clustering depends on distribution (can vanish)



Structural sparseness

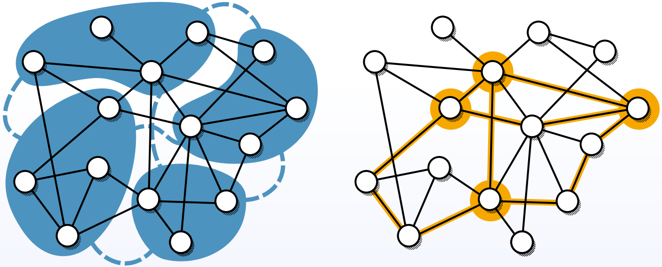




Nowhere dense
Bounded expansion
Locally bounded expansion
Locally excluding a minor
Excluding a topological minor
Excluding a minor
Locally bounded treewidth
Bounded genus
Planar
Bounded degree
Bounded treewidth
Outerplanar
Bounded treedepth
Forests
Star forests
Linear forests

Bounded expansion

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



Bounded expansion:
Robustness

Classes of bounded expansion are closed* under

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

Bounded expansion:
Robustness

Classes of bounded expansion are closed* under

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

Many other equivalent characterisations besides density of shallow minors: shallow immersions, weakly linked colourings, low treedepth colourings, neighbour complexity,...

Bounded expansion:
Usefulness

Theorem (Dvořák, Král, and Thomas)

First-order model-checking is possible in linear time.

Theorem

Dominating Set and r-Dominating Set admit linear kernels.

Theorem (Nešetřil, Ossona de Mendez)

Compute short-distance oracle in linear time.

Theorem

Compute oracle for the size of r-neighbourhoods in linear time.

Theorem (Nešetřil, Ossona de Mendez)

Find out how often fixed graph H occurs as a subgraph/homomorphism in linear time.

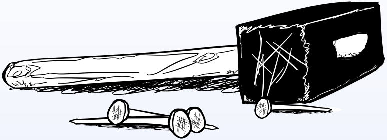
Bounded expansion:
Applicable!

Theorem

Let (D_n) be a sparse degree distribution sequence with tail $h(d)$. Both the configuration model and the Chung–Lu model, with high probability,

- have bounded expansion for $h(d) = \Omega(d^{3+\epsilon})$,
- are nowhere dense (with unbounded expansion) for $h(d) = \Theta(d^{3+o(1)})$,
- and are somewhere dense for $h(d) = O(d^{3-\epsilon})$.

Empirical Sparseness



Closing the gap

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

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 proves are enormous.
- Random graph models capture only some aspectes of complex networks.
- We prove asymptotic bounds.
  (although we show fast convergence)

Distribution tails, aug-aug plots

From theory: if degree distribution has a supercubic tail-bound, then Chung–Lu/Configuration model is structurally sparse.

Distribution tails, aug-aug plots

From theory: if degree distribution has a supercubic tail-bound, then Chung–Lu/Configuration model is structurally sparse.

1. Fit the degree distribution to plausible distributions and then decide whether the tail has a supercubic bound.

Distribution tails, aug-aug plots

From theory: if degree distribution has a supercubic tail-bound, then Chung–Lu/Configuration model is structurally sparse.

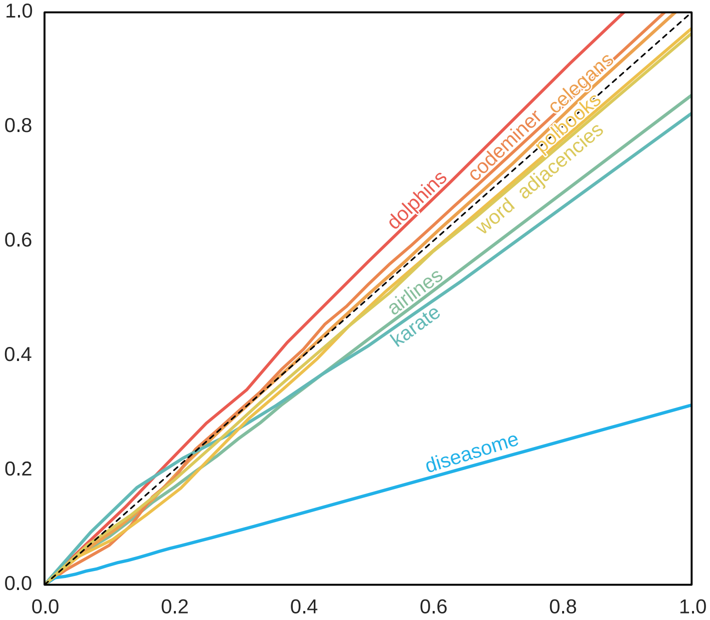
1. Fit the degree distribution to plausible distributions and then decide whether the tail has a supercubic bound.
2. Plot structural sparseness of the network against that of a random graph* with the same degree distribution.

Distribution tails, aug-aug plots

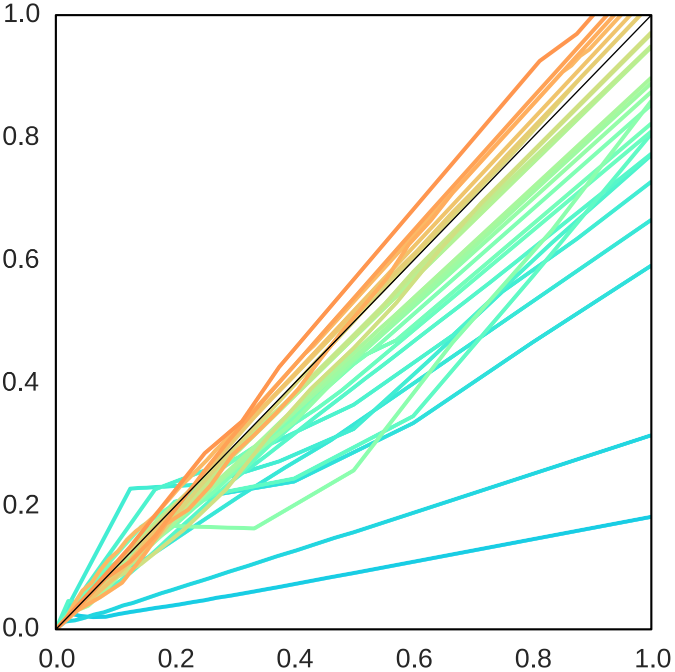
From theory: if degree distribution has a supercubic tail-bound, then Chung–Lu/Configuration model is structurally sparse.

1. Fit the degree distribution to plausible distributions and then decide whether the tail has a supercubic bound.
2. Plot structural sparseness of the network against that of a random graph* with the same degree distribution.

Crucial: we have sparseness measure for different depths.



1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
dolphins
codeminer
celegans
polbooks
word adjacencies
airlines
karate
diseasome



1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Conclusion

- We show that 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 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?