Structural Sparseness & Complex Networks

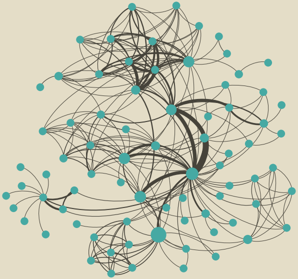
Felix Reidl, NCSU
felix.reidl@gmail.com

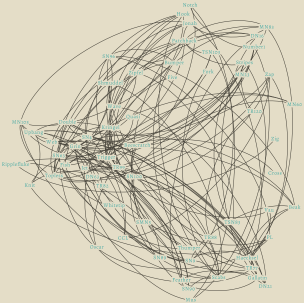
SIAM DM '16

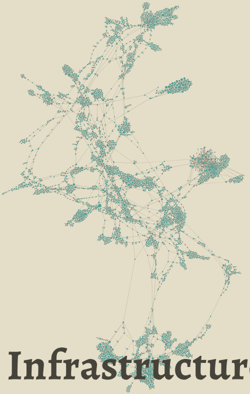
Complex Networks

= Real world graphs

(+ a lot of annotations)







Sociology

Friendships,
Collaborations,
Communication,
...

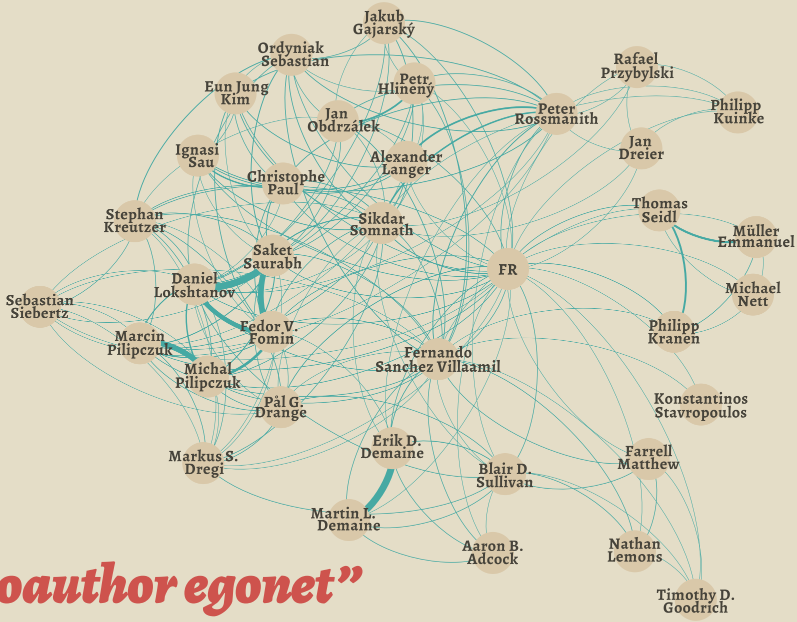
Biology

Gene-gene interactions,
Protein-protein inter.,
Neural networks,
...

Infrastructure

Road networks,
Power grids,
Computer networks,
....

Speaking of networks

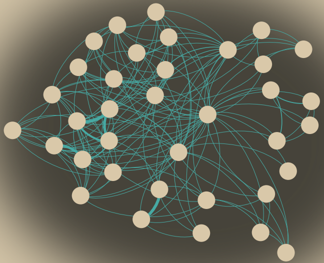


"Coauthor egonet"

Properties of Real Graphs

Giant connected component

Maximum degree around n^{ε}, $\varepsilon < 1$



Clustered

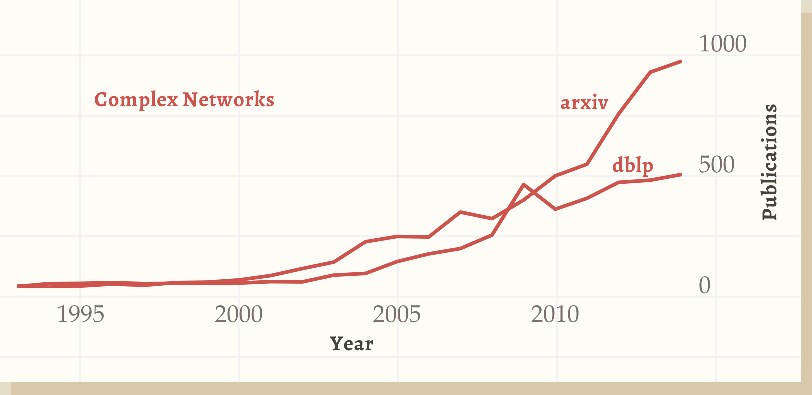
Low diameter

Random (but not uniform)

Sparse

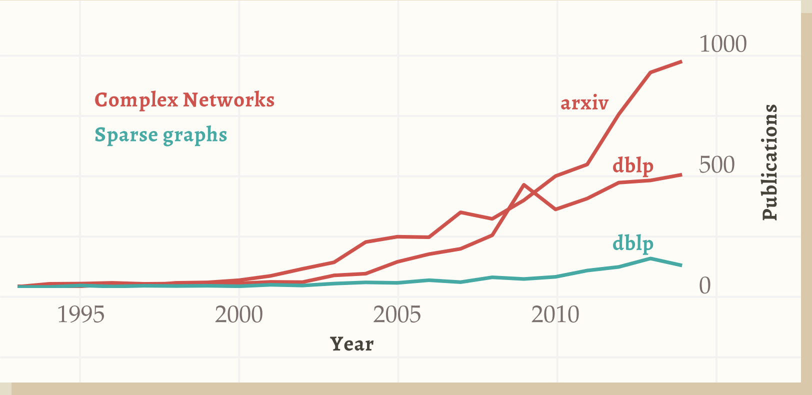
A Booming Field

1) We collect a *lot* of **network data**.
2) We need to **compute** things on them.
3) Sparse graphs have **nice algorithms**.



A Booming Field

1) We collect a *lot* of network data.
2) We need to compute things on them.
3) Sparse graphs have nice algorithms.



Goldilock's sparseness principle

Consider class $\mathcal{G}$ where

1) all *subgraphs* of $\mathcal{G}$ are sparse.
   Degenerate class: not very tractable.

Goldilock's sparseness principle

Consider class $\mathcal{G}$ where

1) all *subgraphs* of $\mathcal{G}$ are sparse.
   Degenerate class: not very tractable.

3) all *minors* of $\mathcal{G}$ are sparse.
   Excluded minor class:
   to restrictive for networks.

Goldilock's sparseness principle

Consider class $\mathcal{G}$ where

1) all *subgraphs* of $\mathcal{G}$ are sparse.
   Degenerate class: not very tractable.

2) all *shallow minors* of $\mathcal{G}$ are sparse.
   Bounded expansion class:
   exactly what we need!

3) all *minors* of $\mathcal{G}$ are sparse.
   Excluded minor class:
   to restrictive for networks.

Sparse Hierarchy

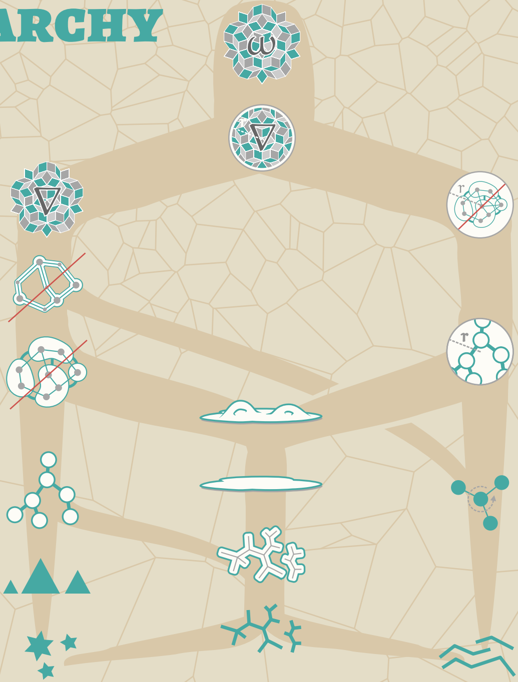
Applicability
to real world

Less

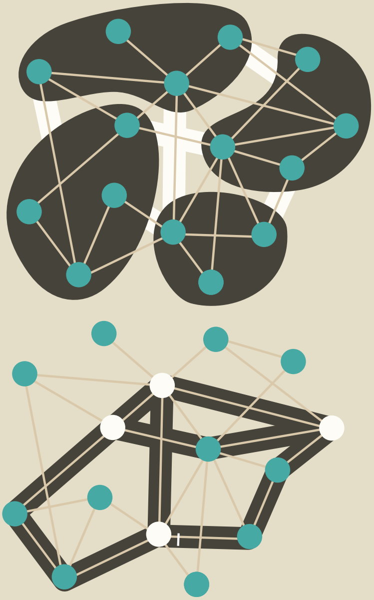
Structure

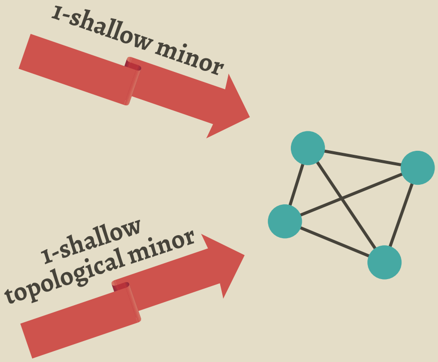
More

Algorithmic
tractability



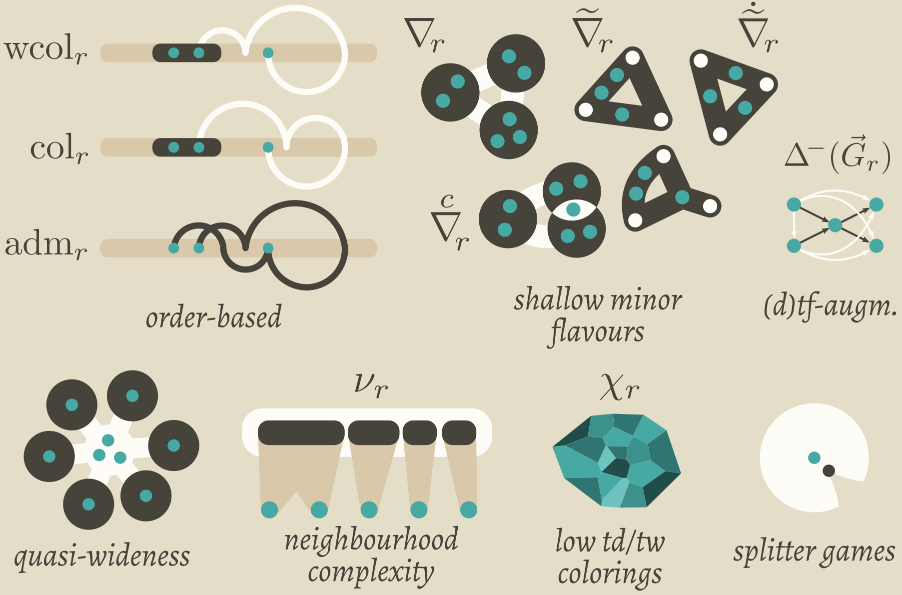
Shallow minors





Bounded expansion:
density bounded by a function of depth

Choose your poison

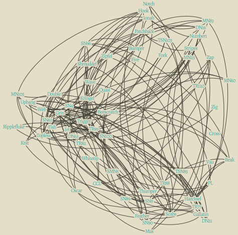


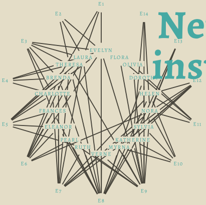
From data to theory

$$Pr[\|G\| \geq \xi k] \leq \left(\frac{e\beta D^2}{2n\xi k e^{D^2/2n}}\right)^{\xi k}$$

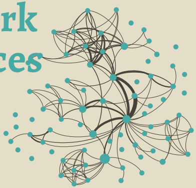
Mathematical theory

???





Network instances



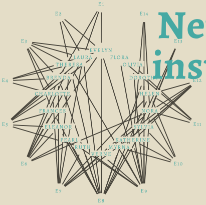
From data to theory

$$Pr[\|G\| \geq \xi k] \leq \left(\frac{e\beta D^2}{2n\xi k e^{D^2/2n}}\right)^{\xi k}$$

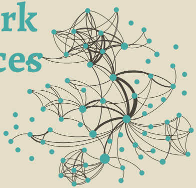
Mathematical theory

Network model

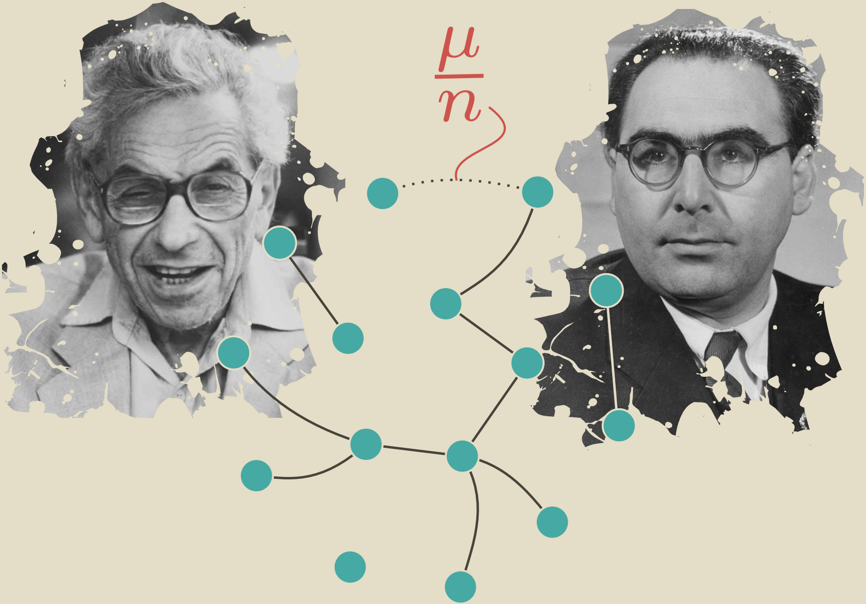
- Random network
- Tunable parameters
- Replicates some statistics



Network instances

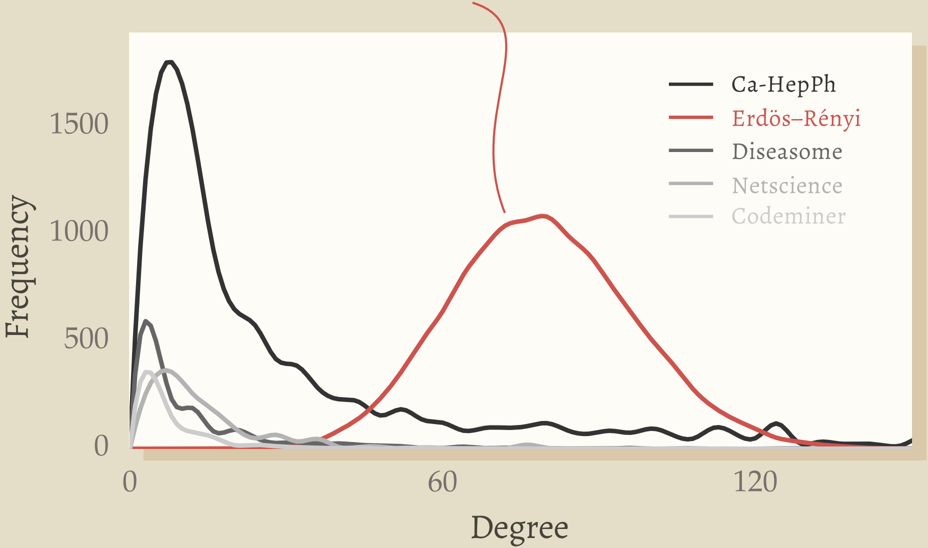


Erdős–Rényi: Structurally Sparse

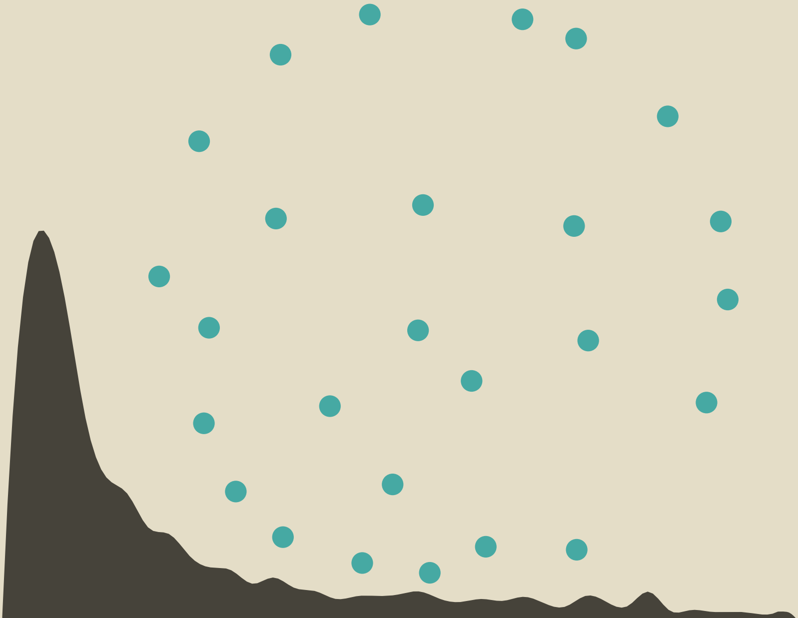


Deficiency of ER

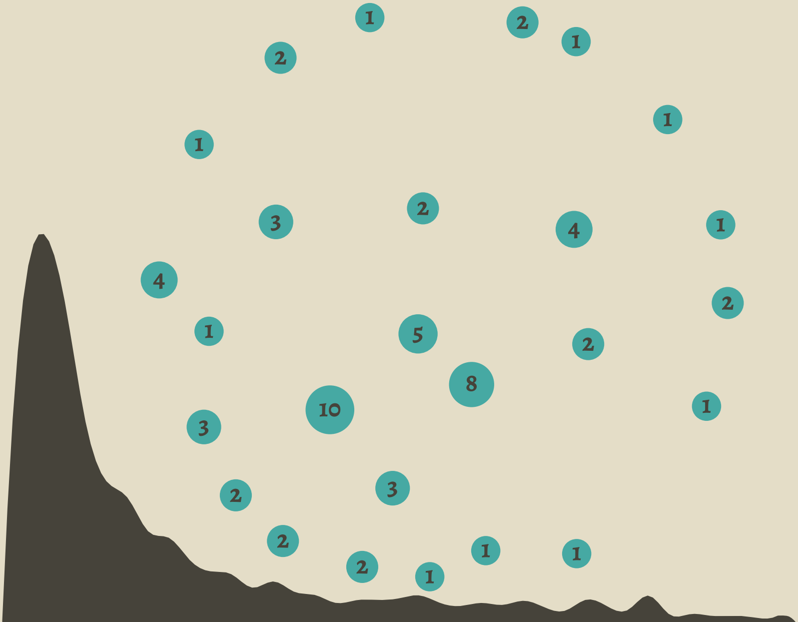
Unrealistic degree distribution



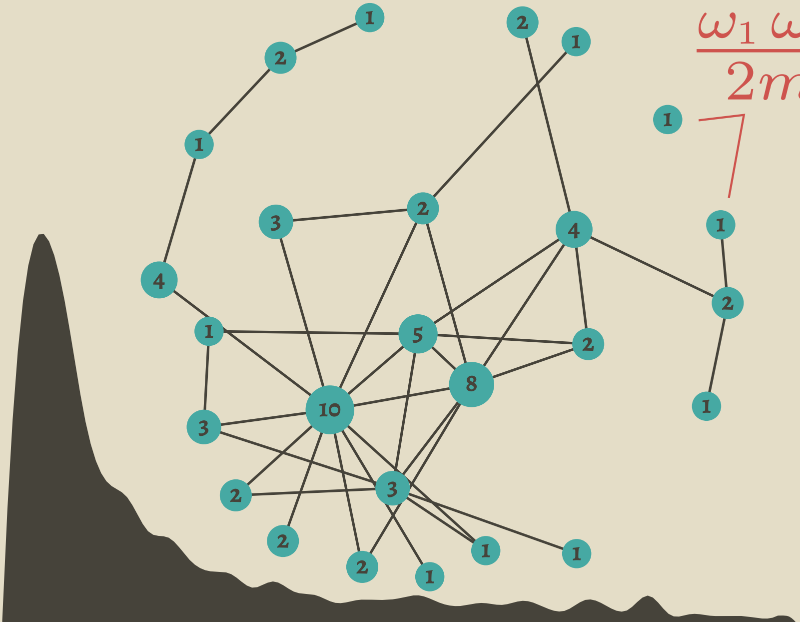
Chung–Lu: Better by a degree.



Chung–Lu: Better by a degree.



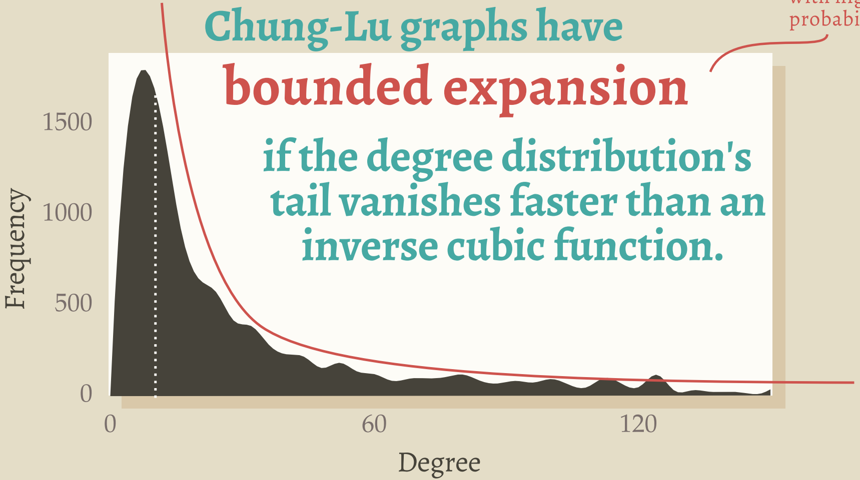
Chung–Lu: Better by a degree.



$$\frac{\omega_1 \, \omega_2}{2m}$$

The devil is in the D-tail

Chung-Lu graphs have **bounded expansion** (with high probability) if the degree distribution's tail vanishes faster than an inverse cubic function.



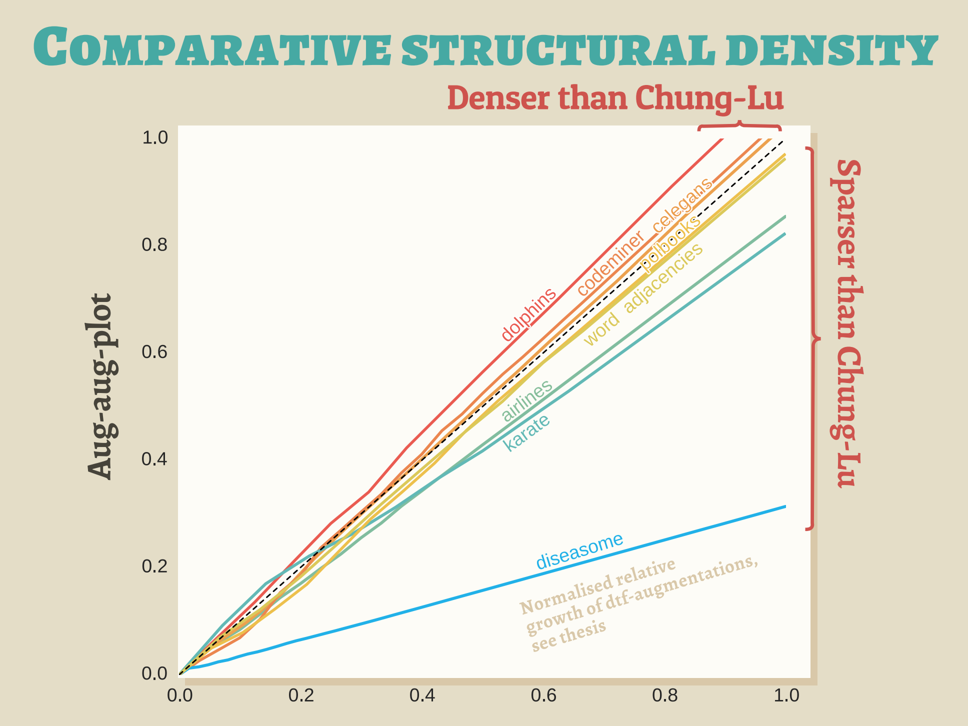
(Proof idea: couple occurences of shallow top. minors to subgraphs in a different Chung-Lu graph, bound probability of dense subgraph in that graph.)

Structural phase transition

- **Degree distribution with tail-bound $\frac{1}{h(d)}$:**

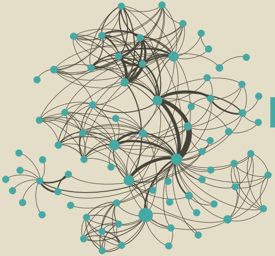
$$h(d) = \begin{cases} \Omega(d^{3+\epsilon}) & \text{bounded expansion} \\ \Theta(d^{3+o(1)}) & \text{nowhere dense} \\ O(d^{3-\epsilon}) & \text{somewhere dense} \end{cases}$$

- **Proof idea for lower bounds: more coupling.**
- **The same works for the so-called 'configuration model'**
- **Also works for similar models with non-vanishing clustering**



Comparative structural density
Denser than Chung-Lu
Sparser than Chung-Lu
Aug-aug-plot
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
adjacencies
word
airlines
karate
diseasome
Normalised relative
growth of dtf-augmentations,
see thesis

Real Structural Sparseness



Statistical test of degree distribution

Tail looks subcubic → Cannot decide

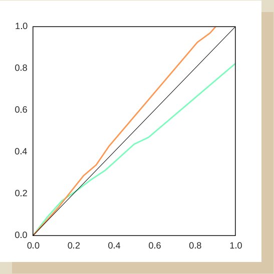
Tail looks supercubic ↓

Compare density to Chung-Lu w/ same degree distribution

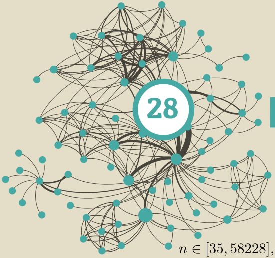
Denser than Chung-Lu → Cannot decide

Sparser than Chung-Lu ↘

Structurally sparse!



Real Structural Sparseness



28

$n \in [35, 58228]$,
$m \in [78, 214078]$

Statistical test of degree distribution

Tail looks subcubic → 6 Cannot decide

Tail looks supercubic ↓

22

Compare density to Chung-Lu w/ same degree distribution

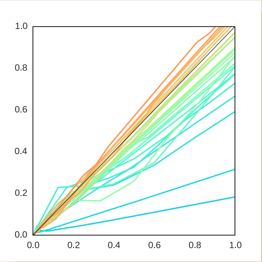
Denser than Chung-Lu → 8 Cannot decide

7 of those are probably sparse

Sparser than Chung-Lu ↘

14

Structurally sparse!



Four reasons to care

Tractability

- Linear FO-Model checking
- Linear r-Domset kernel
- r-Domset approximation
- Linear kernels for td-modulators
- r-neighbourhood aggregation

Applications

- Fast local search
- Centrality estimates
- Motif counting

Elegance

- "Sparseness + depth"
- Sparseness tailored to fit
- Stability under class operations

Practical

- Predicted by models
- (almost) linear time
- Parallelizable

Dawar
Demaine
Drange
Dregi
Dvořák
Fomin
Gajarský
Grohe
Hliněný
Král
Kreutzer
Lokshtanov
Nešetřil
Obdržálek
Ordyniak
Ossona de Mendez
Pilipczuk
Pilipczuk
Reidl
Rossmanith
Sánchez Villaamil
Saurabh
Siebertz
Sikdar
Sullivan
Thomas
Wood

THANKS!

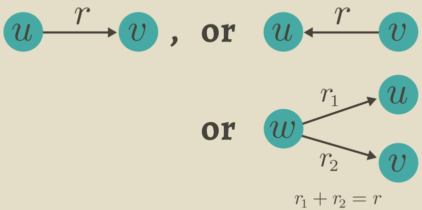
Questions?

DTF-AUGMENTATIONS

Theorem: Let $G \in \mathcal{G}$ from a bounded expansion class. There exists a sequence $\vec{G}_1, \vec{G}_2, \ldots$ of edge-weighted digraphs such that

a) $\Delta^-(\vec{G}_r) \leqslant f(r)$

b) For all u, v with $\mathrm{dist}_G(u,v) \leqslant r$ either



is present in $\vec{G}_r$.

Moreover, for fixed r this sequence is computable in linear time.