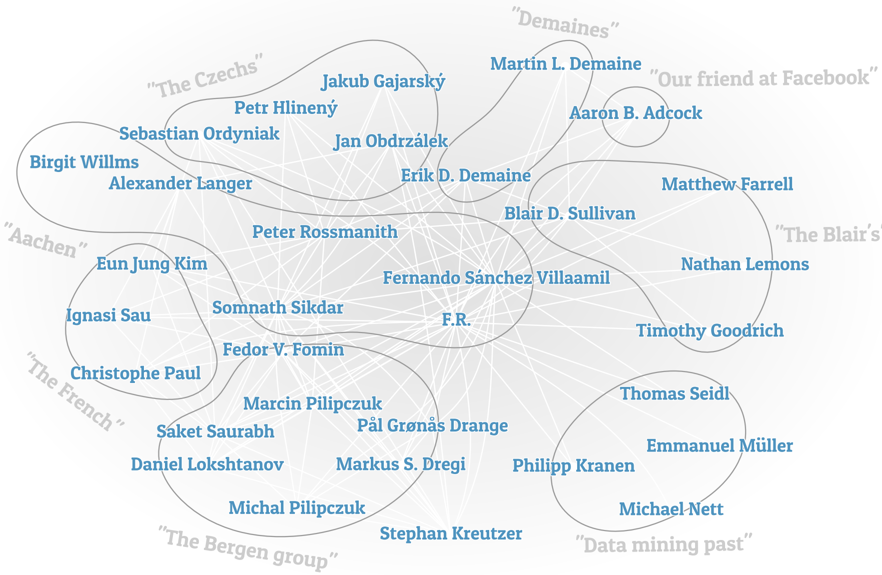
Structural Sparseness and Complex Networks

Felix Reidl

Dec 4th 2015

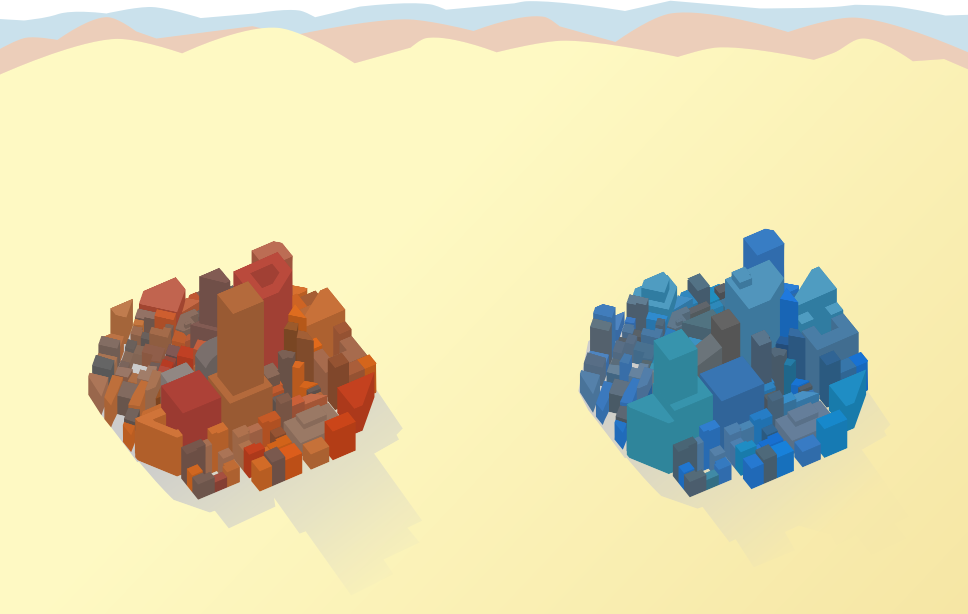

"The Czechs"
"Demaines"
"Our friend at Facebook"
"Aachen"
"The Blair's"
"The French"
"The Bergen group"
"Data mining past"
Jakub Gajarský
Petr Hlinený
Sebastian Ordyniak
Jan Obdrzálek
Martin L. Demaine
Aaron B. Adcock
Erik D. Demaine
Birgit Willms
Alexander Langer
Matthew Farrell
Blair D. Sullivan
Peter Rossmanith
Eun Jung Kim
Nathan Lemons
Fernando Sánchez Villaamil
Somnath Sikdar
Ignasi Sau
F.R.
Timothy Goodrich
Fedor V. Fomin
Christophe Paul
Thomas Seidl
Marcin Pilipczuk
Pål Grønås Drange
Saket Saurabh
Emmanuel Müller
Daniel Lokshtanov
Markus S. Dregi
Philipp Kranen
Michal Pilipczuk
Michael Nett
Stephan Kreutzer

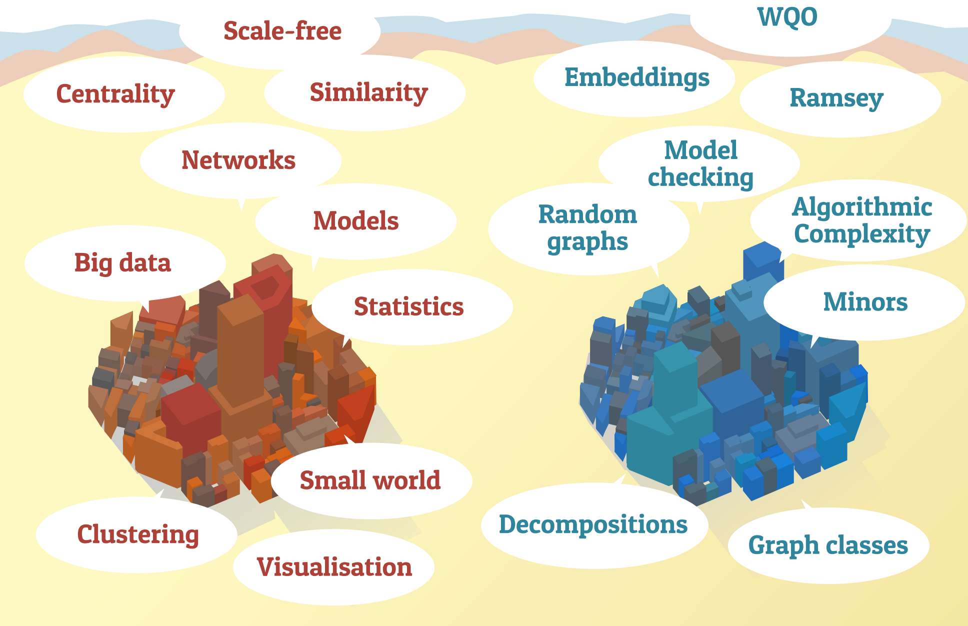
The City & the City

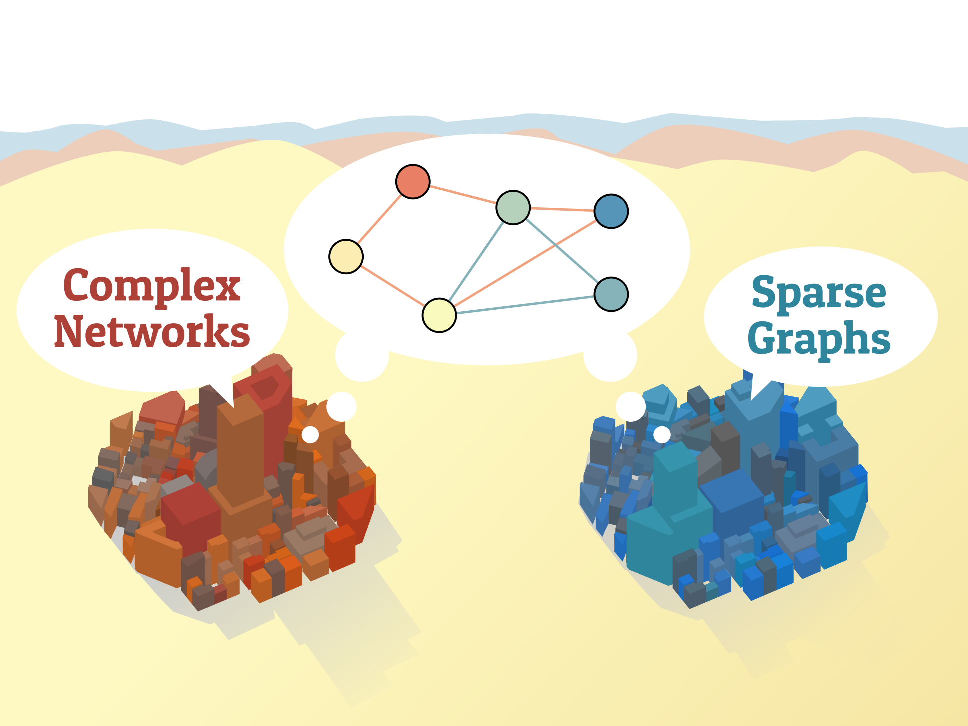
by China Miéville

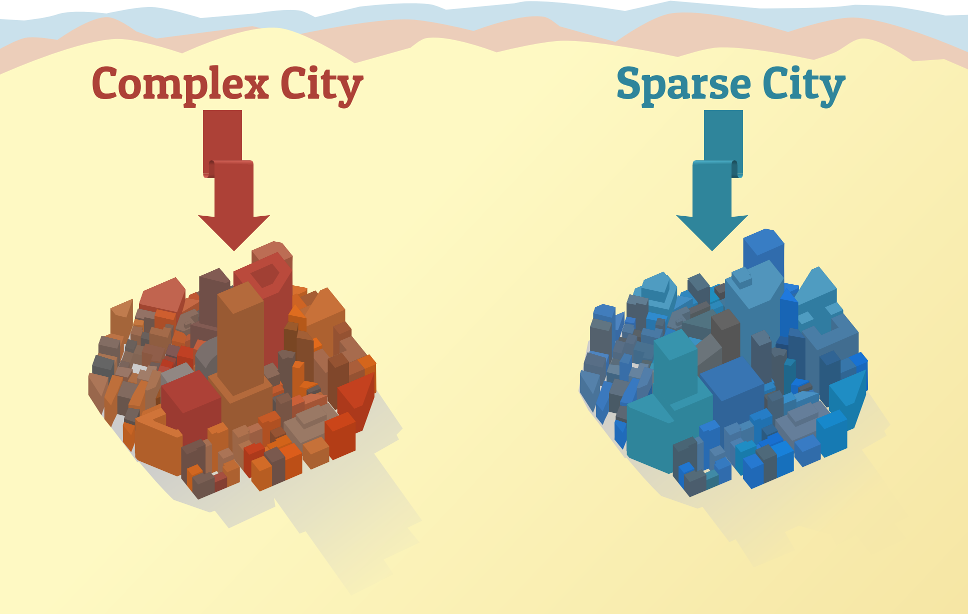
The City & the City

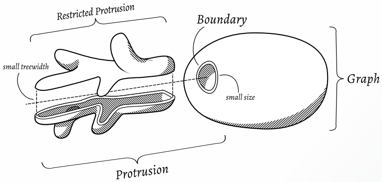
The City & the City

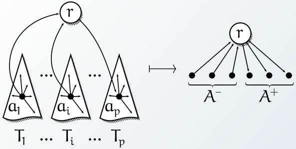



Scale-free
WQO
Centrality
Similarity
Embeddings
Ramsey
Networks
Model checking
Models
Random graphs
Algorithmic Complexity
Big data
Statistics
Minors
Small world
Clustering
Visualisation
Decompositions
Graph classes

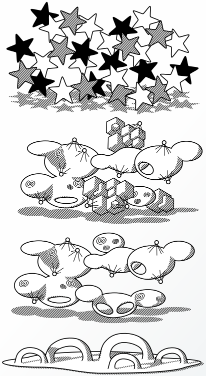

Complex Networks
Sparse Graphs

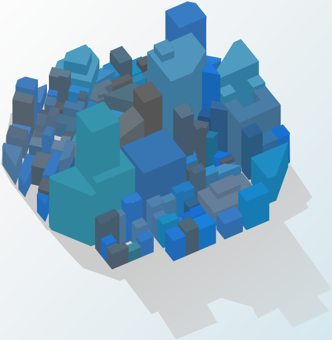

Complex City
Sparse City

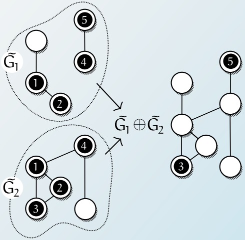

Restricted Protrusion
Boundary
small treewidth
small size
Graph
Protrusion

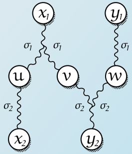

r
a1 ai ap
T1 … Ti … Tp
A− A+

Sparse City






G̃1
G̃2
G̃1 ⊕ G̃2
1
2
3
4
5


x1
y1
u
v
w
x2
y2
σ1
σ2

What sparse city cares about

How many colours do I need
to colour a graph?

How dense can a graph get until
a complete subgraph on k vertices
appears?

Can I efficiently answer questions
framed in a certain logic?

How often does this small graph
fit into this large graph?

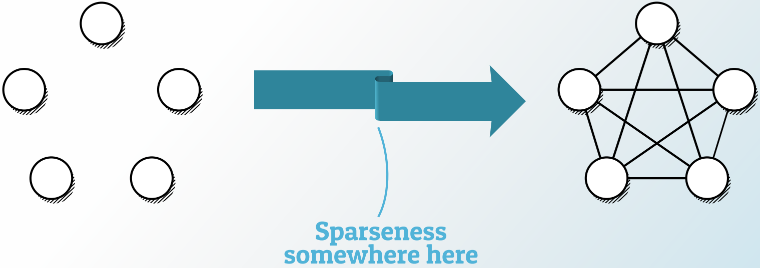
What problems are efficiently
solvable?

Can my graph be decomposed in
a nice manner?

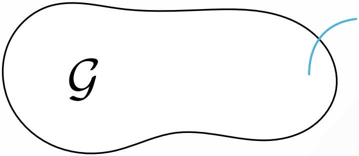
Sparse City

Sparseness

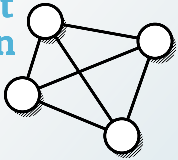
"Not too many edges"



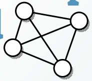
Forbidden substructures

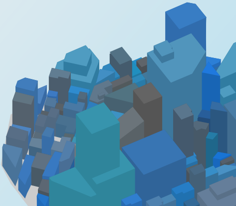


All graphs
that do not
contain



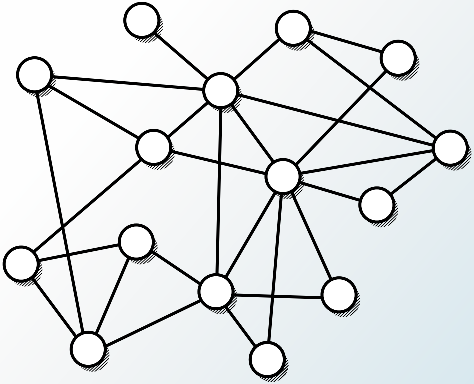
as a 'substructure'.

- Very nice for proofs:
  if we find , we have a contradiction.
- Structural sparseness!?

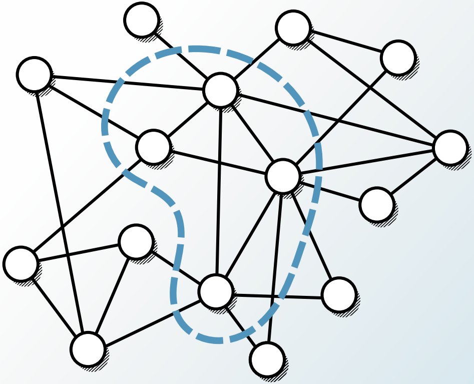


Sparse City

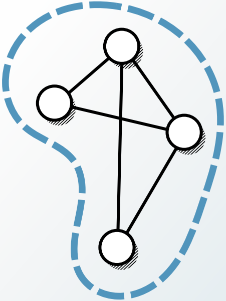
Substructures



Substructures

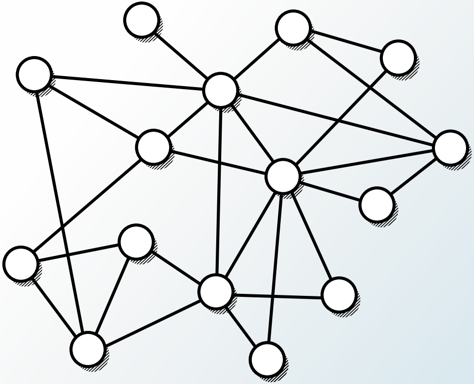


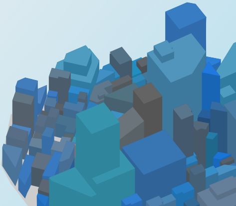
Substructures



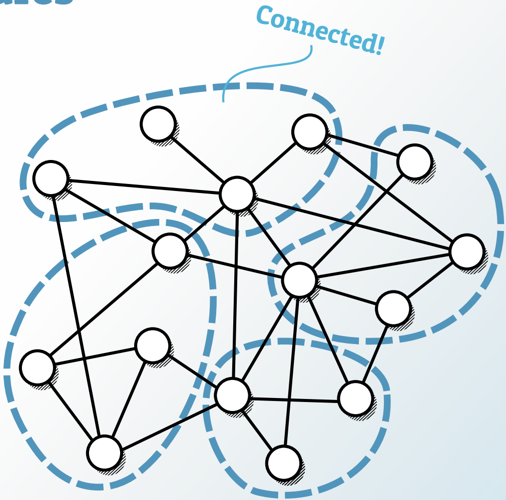
(Induced) Subgraph

Substructures

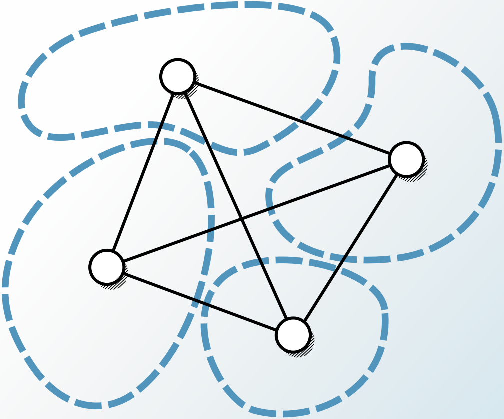




Substructures

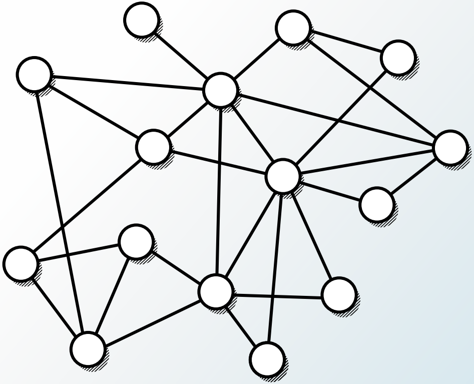


Substructures

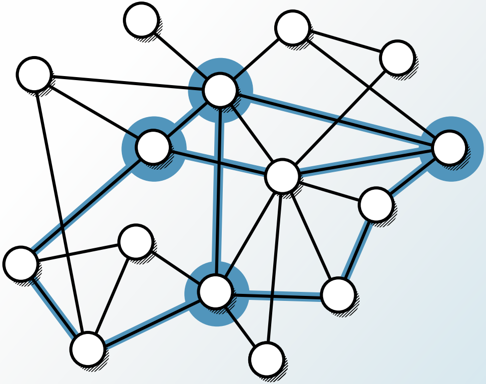


Minor

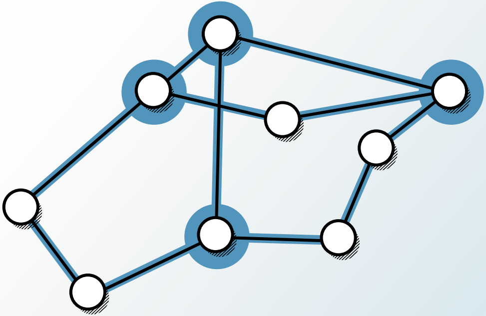
Substructures



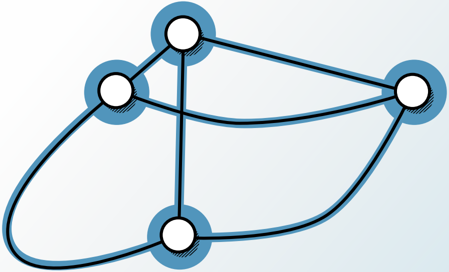
Substructures



Substructures

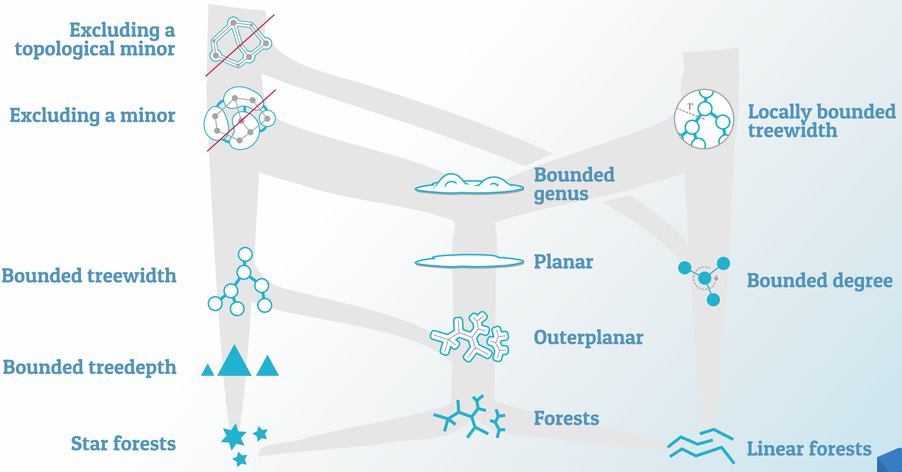


Substructures



Topological minor

Hierarchy (Part I)

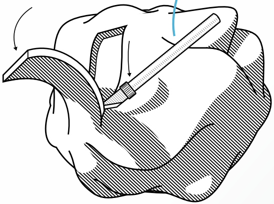


Sparse City

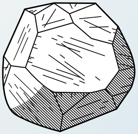
Kernelisation

Polynomial-time scalpel

The hard part



Apply exhaustively



Input size n

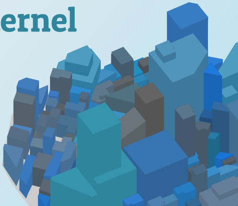
Parameter k

Instance of parametrised problem

Input size $f(k)$

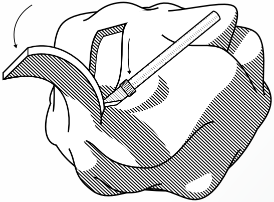
Parameter k'

Kernel

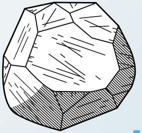


Sparse City

Kernelisation



Apply exhaustively



Kernel size

Input size n

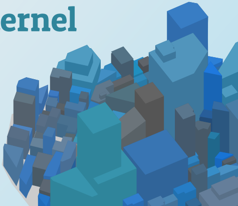
Parameter k

Instance of parametrised problem

Input size $f(k)$

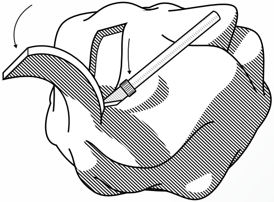
Parameter k'

Kernel

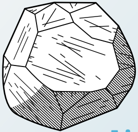


Sparse City

Kernelisation



Apply exhaustively



Linear kernel

Input size n

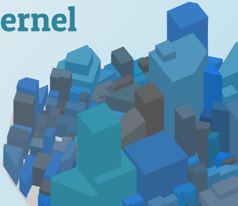
Parameter k

Instance of parametrised problem

Input size $c \cdot k$

Parameter $k' \leq k$

Kernel



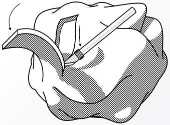
Sparse City

Meta-Kernelisation

PROBLEM	
Input:	xx xxx x x x x x xxxxx
Parameter:	xx x x
Problem:	xx xxx xx x xx xxxxx

Suitable problem

+

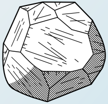


Reduction rule

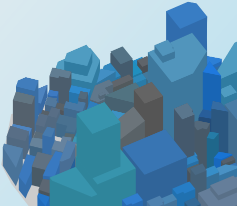
+

Sparse class

=

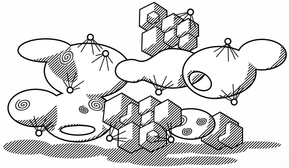


Kernel



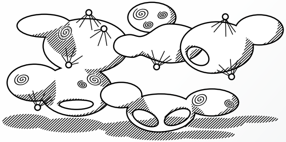
Sparse City

Pushing it up



H-Topological-Minor-Free — Treewidth-bounding

∪

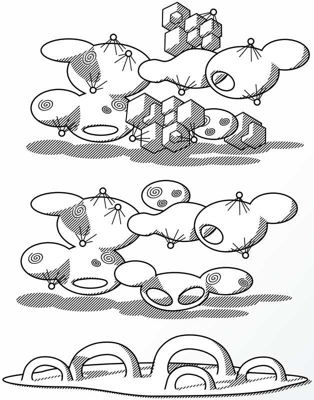


H-Minor-Free — Contraction-bidimensional

∪

Bounded Genus — Quasi-Coverable

Running out of problems



Few problems

H-Topological-Minor-Free

Treewidth-bounding

⋃

H-Minor-Free

Contraction-bidimensional

⋃

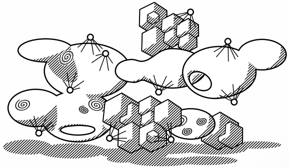
Bounded Genus

Quasi-Coverable

Many problems

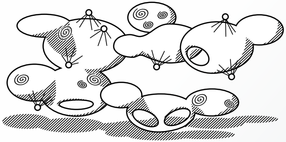
Sparse City

Structural parametrisation



H-Topological-Minor-Free | Treewidth-t modulator

∪

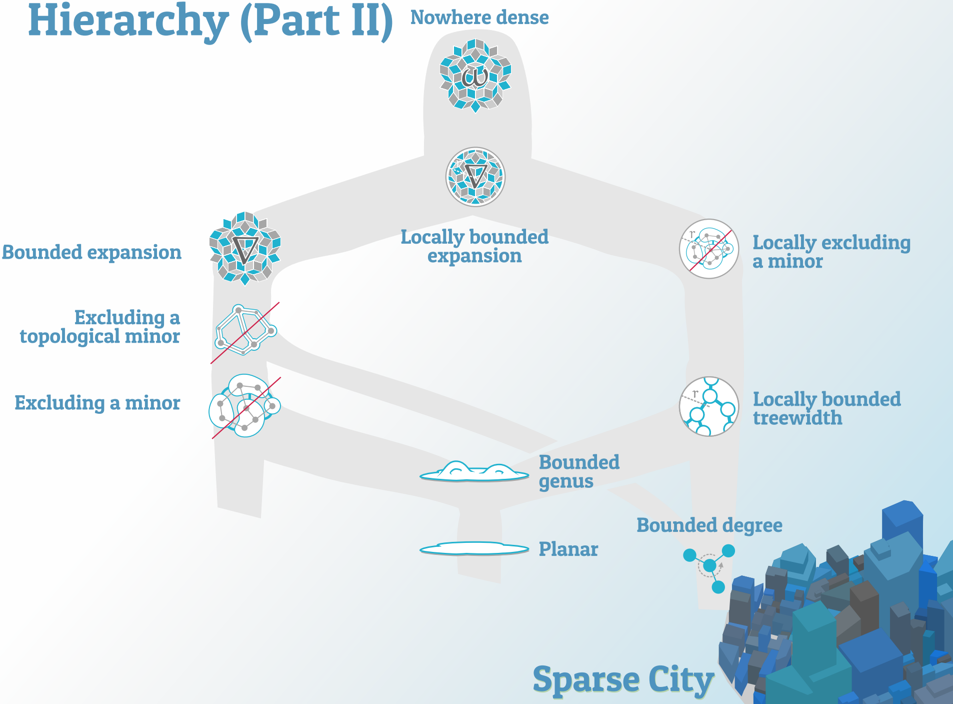


H-Minor-Free | Treewidth-t modulator

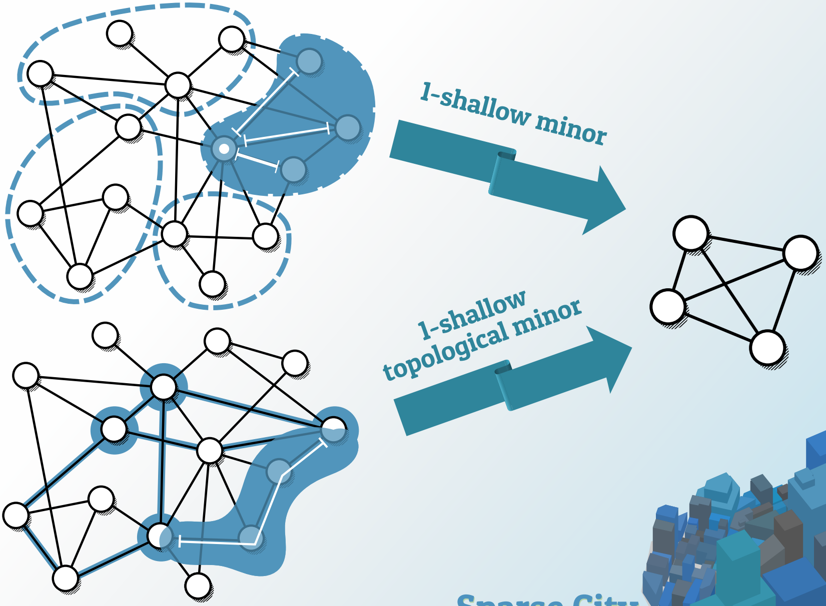
∪

Bounded Genus | Treewidth-t modulator

Hierarchy (Part II)

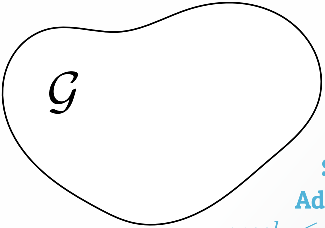


Shallow minors



Sparse City

Bounded expansion classes



All r-shallow minors appearing in this class have density at most f(r).

Same for r-shallow topological minors.

Same for r-shallow immersions.

Admits low-treedepth colourings.

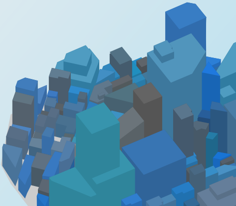
$wcol_r \leq f(r)$

$col_r \leq f(r)$

$f(r)$-quasi-wide

- 'Depth-dependent sparseness'
- Robust!

We can do interesting things without increasing the expansion by too much.

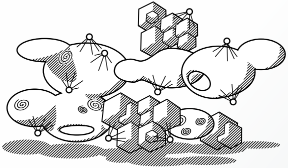


Sparse City

Structural parametrisation

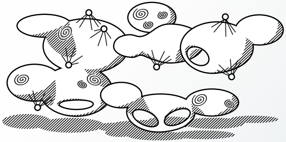
Bounded expansion / nowhere dense — Treedepth-t modulator

∪



H-Topological-Minor-Free — Treewidth-t modulator

∪



H-Minor-Free — Treewidth-t modulator

∪

Bounded Genus

Sparse City

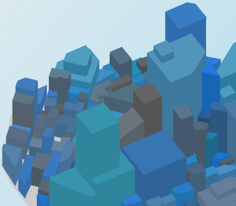
Summary

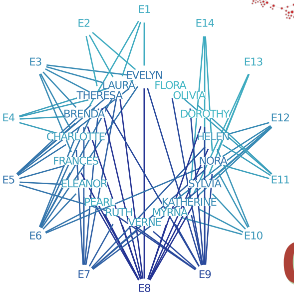
- **Bounded expansion/nowhere dense classes have extremely good algorithmic properties!**
- **Extension of meta-kernelisation**
- **Linear kernel for Dominating Set**

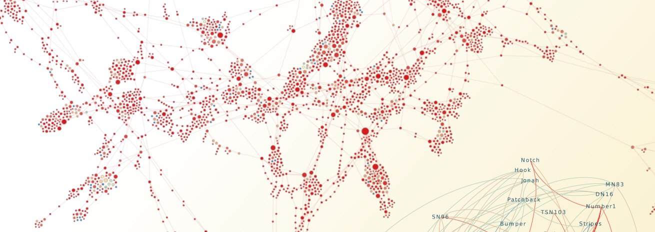
  } Thesis

- **First-Order model checking in linear time**
  Dvořák, Král, Thomas / Grohe, Kreutzer, Siebertz
- **Many nice structural properties**
  Some new ones in my thesis...

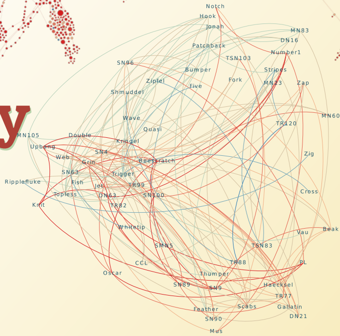
Sparse City

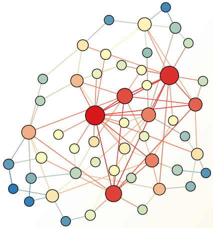


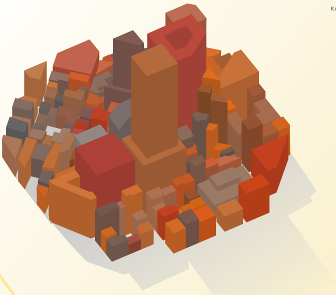


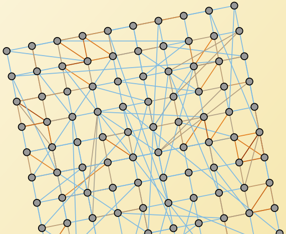


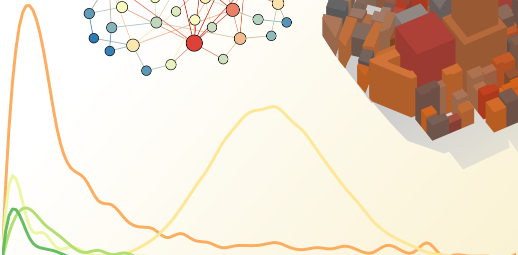
Complex City







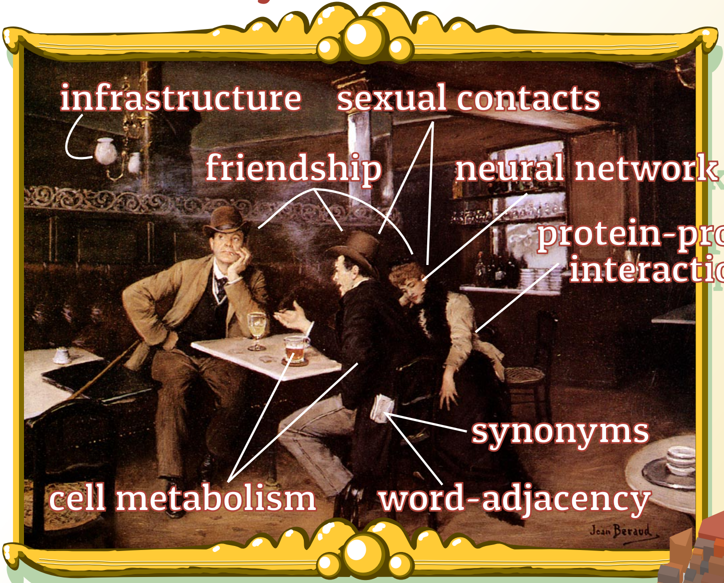




Complex networks

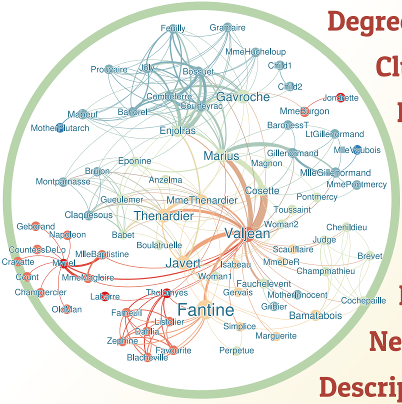
- Social networks since ~1920
- Networks are everywhere
- Very easy to collect nowadays!
- Commonalities?

Networks everywhere



Complex City

What complex city cares about



Density

Degree distribution

Clustering coefficient

Diameter

Centrality indices

Core decompositon

Community structure

Frequent subgraphs

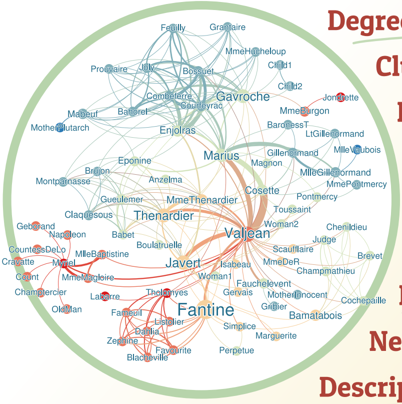
Network similarity

Descriptive modelling

Generative modelling

Complex City

What complex city cares about



Density

Degree distribution

Clustering coefficient

Diameter

Centrality indices

Core decompositon

Community structure

Frequent subgraphs

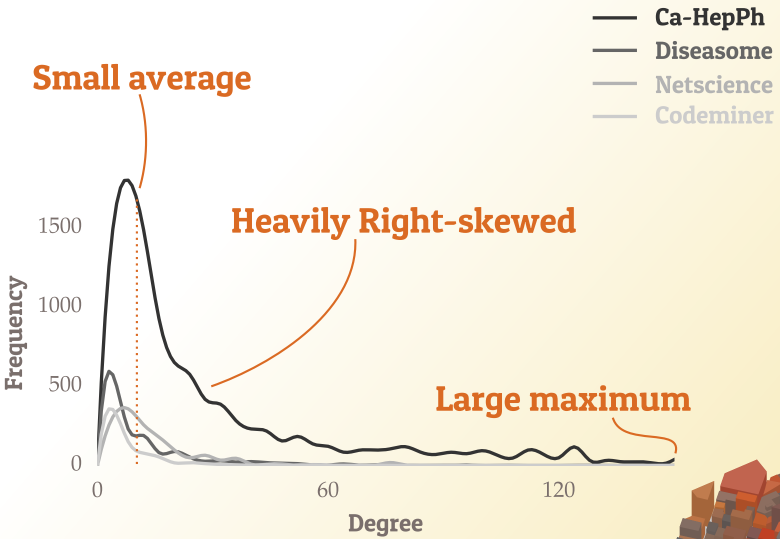
Network similarity

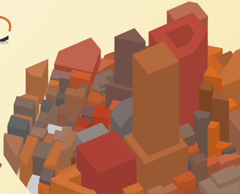
Descriptive modelling

Generative modelling

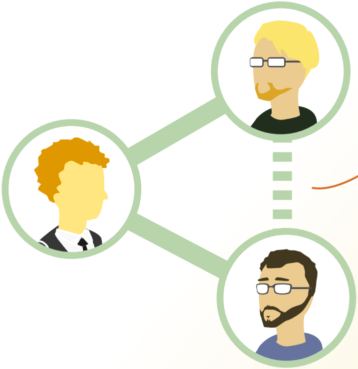
Complex City

Degree distribution



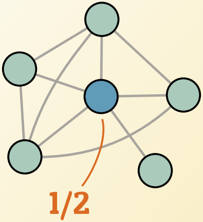


Clustering coefficient

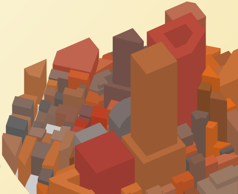


Are friends of mine
more likely to be friends?

$$\frac{\text{\#edges between friends}}{\text{pairs of friends}}$$



- **Most networks exhibit high clustering**

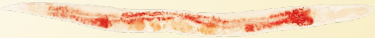


Complex City

Centrality indices

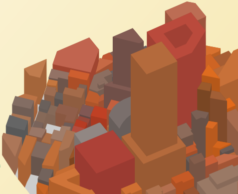
Which node in here
is important?

- Only use network!
- Many different measures,
  no consensus



- Several measures based
  around neighbourhood sizes, e.g.

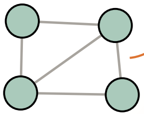
$$c_H(v) = \sum_{u \in G} d(v,u)^{-1}$$



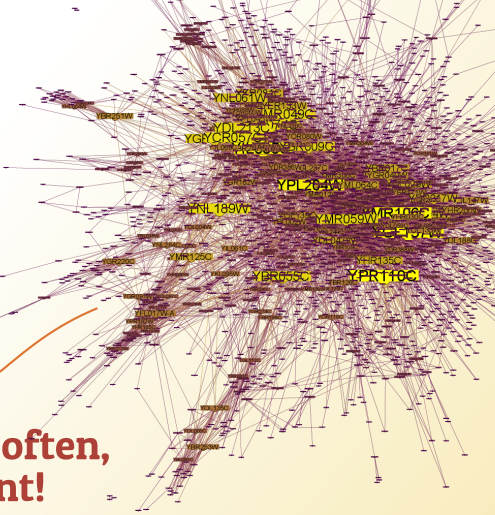
Complex City

Frequent subgraphs

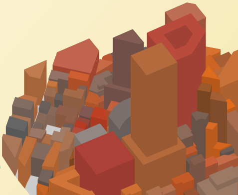
How often is this



contained in this?



- If a subgraph appears often, it probably is important!
- Network similarity measure
- Expensive to compute!

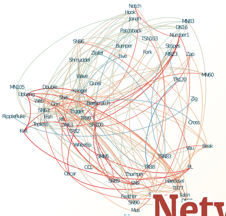


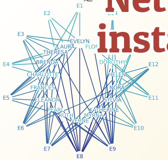
Descriptive modelling

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

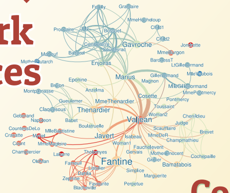
Mathematical Theory

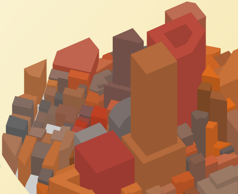
???





Network instances





Complex City

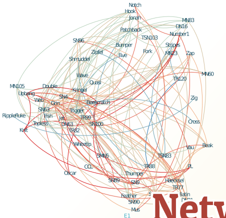
Descriptive modelling

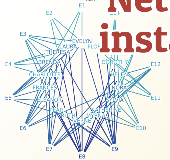
$$Pr[\|G\| \geq \xi k] \leq \left(\frac{e\beta D^2}{2n\xi k e^{D^2/2n}}\right)^{\xi k}$$

Mathematical Theory

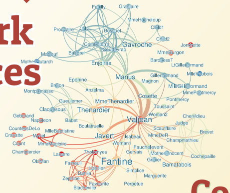
Network model

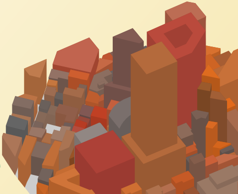
- Random network
- Tunable parameters
- Replicates some statistics





Network instances

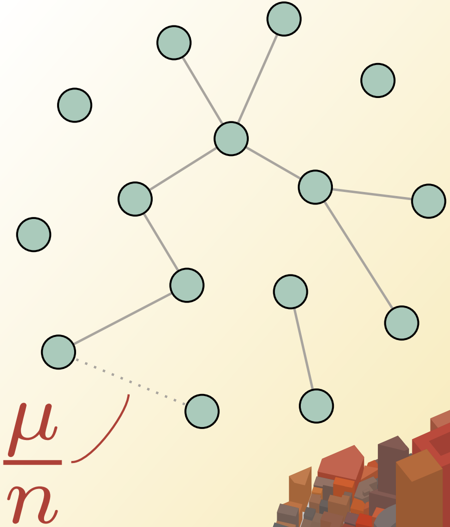




Complex City

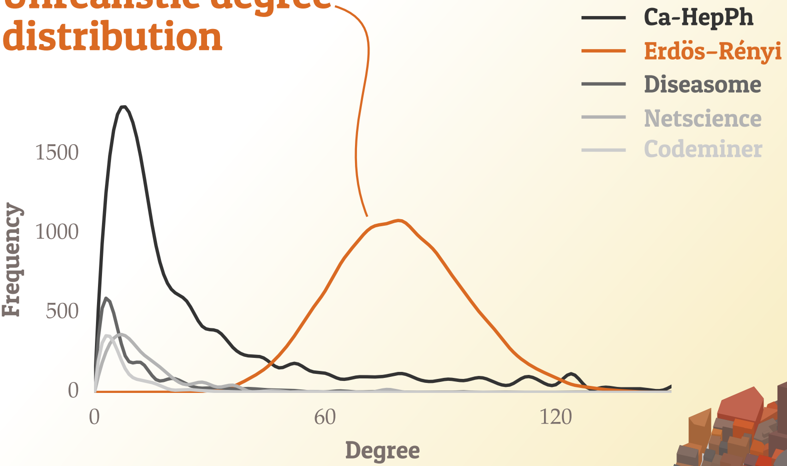
The Erdös–Rényi model

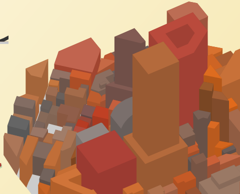
- Well-understood
- Nice properties
- No clustering
- Unrealistic degree distribution



The Erdös–Rényi model

- Unrealistic degree distribution

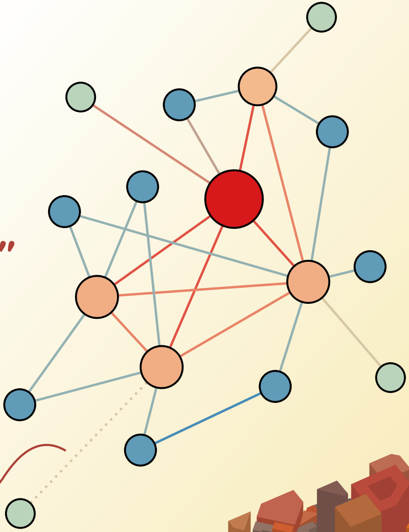




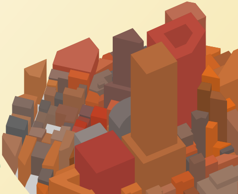
Complex City

The Chung-Lu model

- Close to E.-R.
- Prescribe degree distribution
- No "further assumptions"
- No clustering



$$\frac{w_{\bullet} w_{\bullet}}{2m}$$

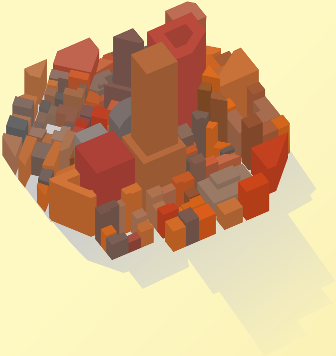


Summary

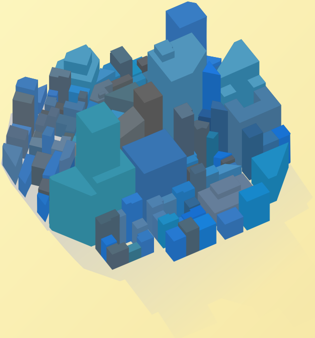
- Very large data sets (millions of nodes)
- From very different areas
- A lot of algorithmic questions
- Need very efficient algorithms!

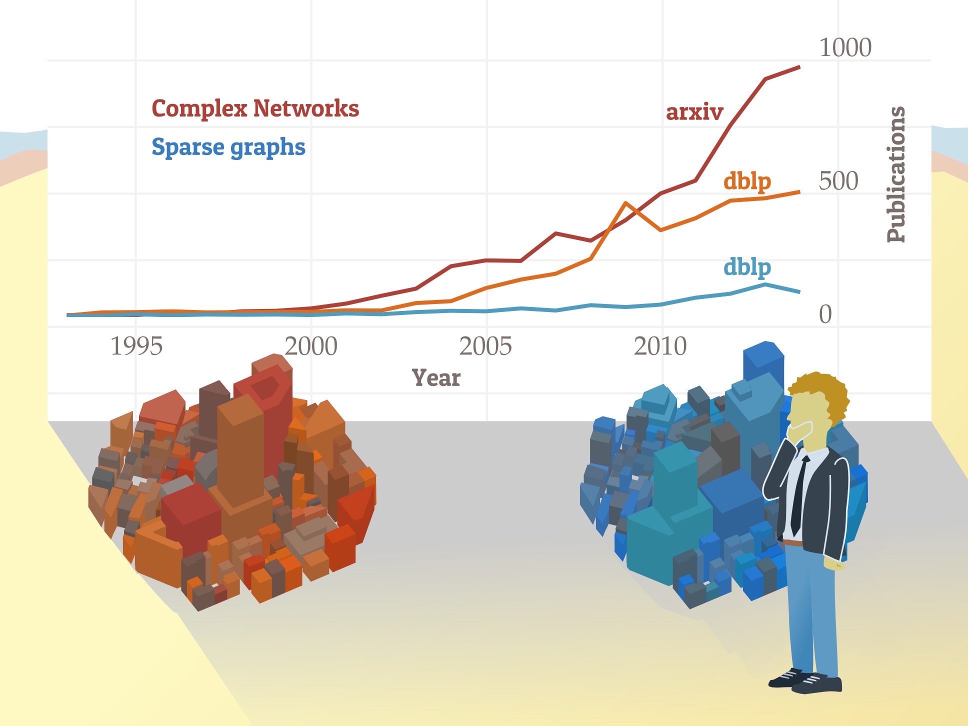
Complex City

- Work with sparse graphs
- Need efficient algorithms
- Have interesting problems

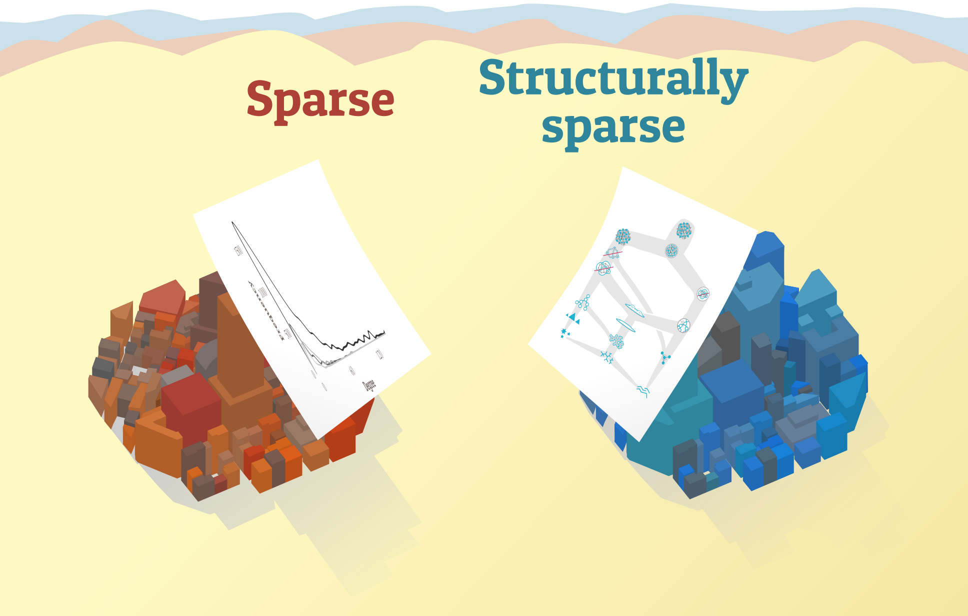


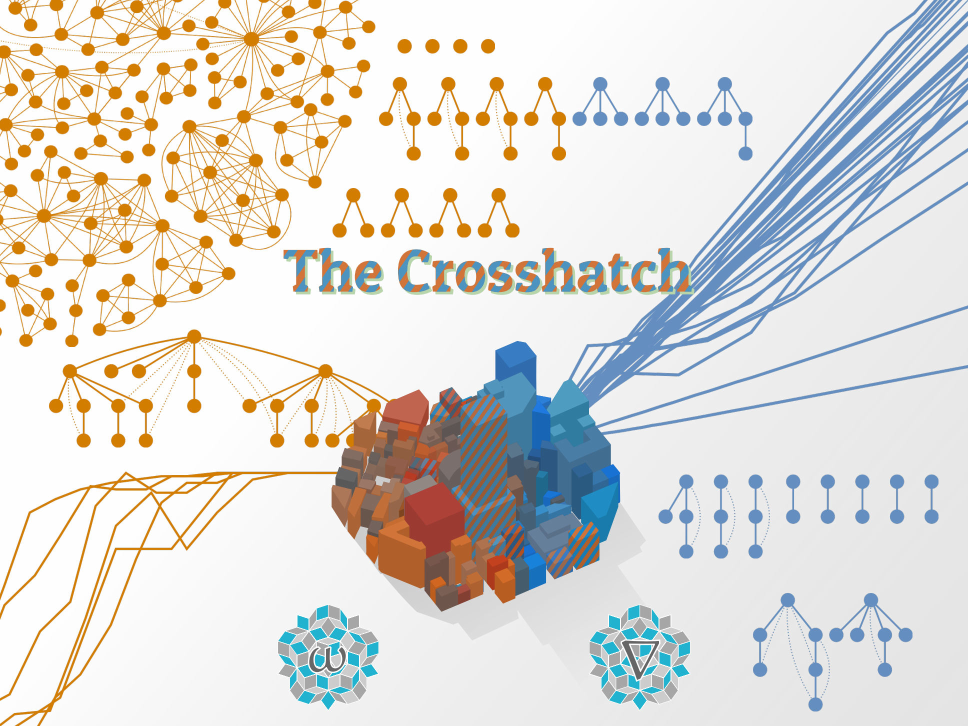
- Work with sparse graphs
- Have efficient algorithms
- Like interesting problems





Complex Networks
Sparse graphs
arxiv
dblp
dblp
1000
500
0
Publications
1995
2000
2005
2010
Year


Sparse
Structurally sparse
Degree


The Crosshatch

A small overlap

Sparse Erdös–Rényi

- $G(n, \mu/n)$ **has bounded expansion a.a.s.**
  Nešetřil, Ossona de Mendez, Wood

- **Maybe Chung-Lu does, too?**

What about other models?

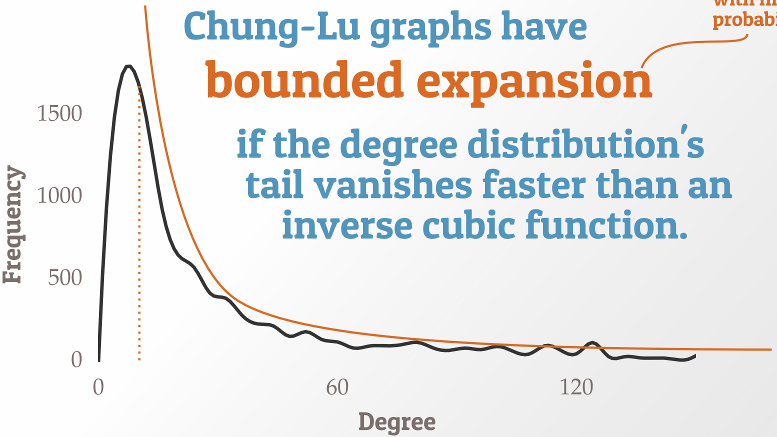
For which degree distributions?

Do those appear in the real world?

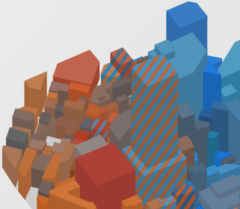
Can we even test that?

A tale of tails

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.



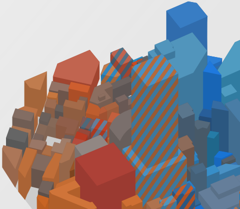
The Crosshatch

Phase transition of Chung-Lu

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

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

  - 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



The Crosshatch

Tail-bounds for real networks?

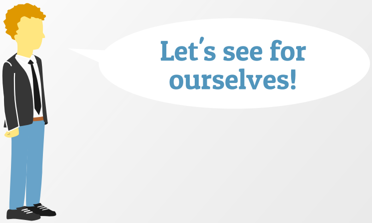
- **Famous claim: real degree distributions follow power law** $\sim \frac{1}{d^\gamma}$

  small gamma: tail does not vanish fast enough

- **Rigorous statistical tests: almost never pure power law, but exponential cut-off**

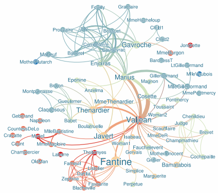
  Clauset, Shalizi, Newman

  tail vanishes quick enough



The Crosshatch

Real structural sparseness



Statistical test of degree distribution

Tail looks subcubic → Cannot decide

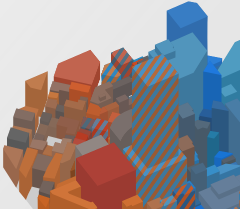
Tail looks supercubic ↓

Compare density to Chung-Lu w/ same degree distribution

Denser than Chung-Lu → Cannot decide

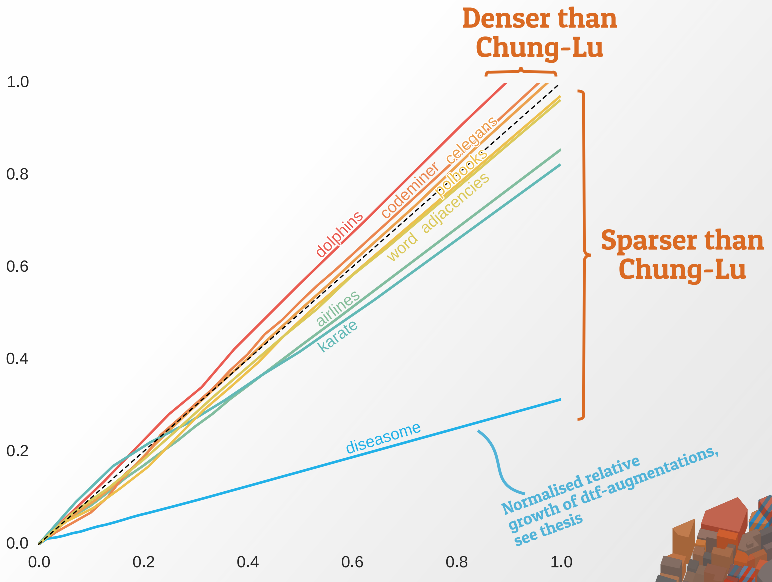
Sparser than Chung-Lu ↓

Structurally sparse!



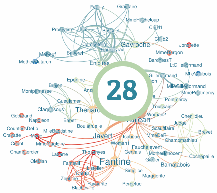
The Crosshatch

Chung-Lu vs. real network



The Crosshatch

Real structural sparseness



28

Statistical test of degree distribution

Tail looks subcubic → 6 Cannot decide

Tail looks supercubic ↓

22

Compare density to Chung-Lu w/ same degree distribution

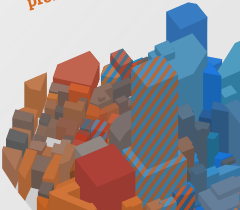
Denser than Chung-Lu → Cannot decide 8

7 of those are probably sparse

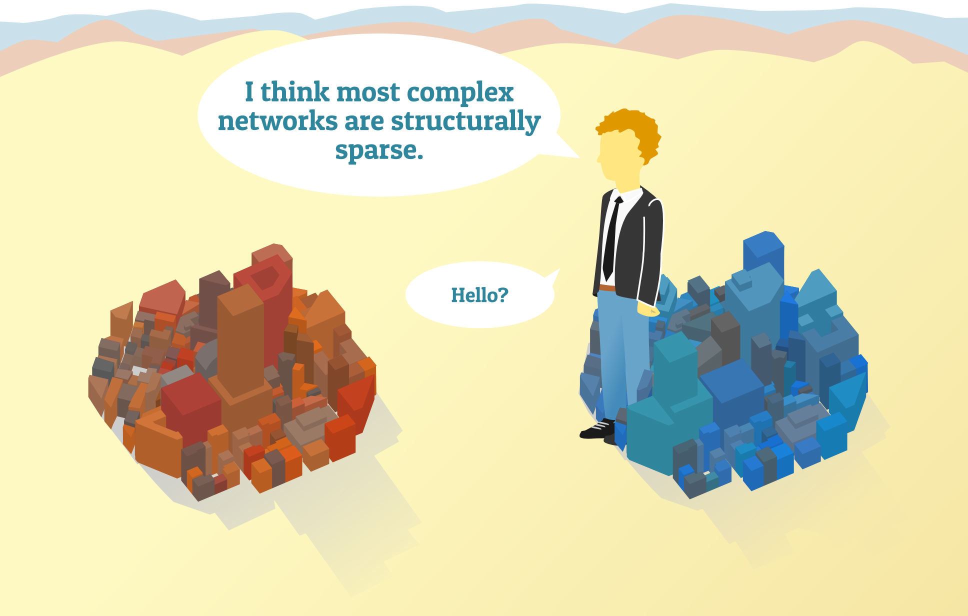
Sparser than Chung-Lu ↓

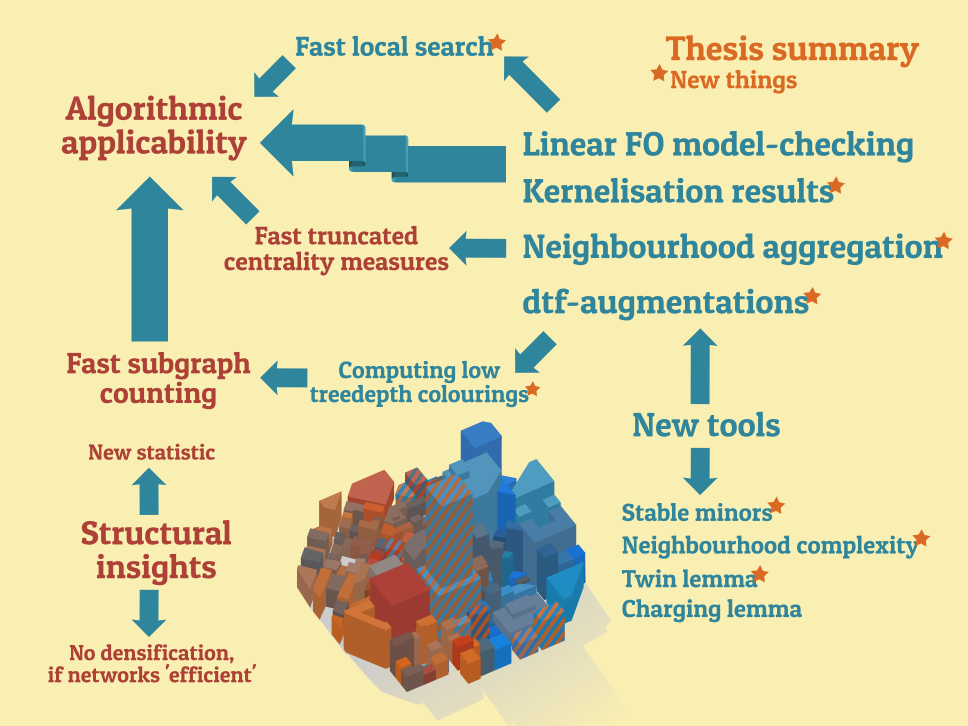
14

Structurally sparse!



The Crosshatch


I think most complex networks are structurally sparse.
Hello?



Thesis summary
★ New things
Fast local search ★
Algorithmic applicability
Linear FO model-checking
Kernelisation results ★
Neighbourhood aggregation ★
dtf-augmentations ★
Fast truncated centrality measures
Fast subgraph counting
Computing low treedepth colourings ★
New tools
Stable minors ★
Neighbourhood complexity ★
Twin lemma ★
Charging lemma
New statistic
Structural insights
No densification, if networks 'efficient'

Open questions and future work

Apply low treedepth colourings to practical problem.

What about other network models?

Fix attachment models.

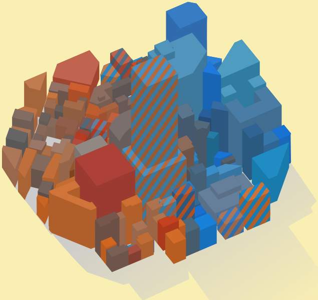
Collect interesting network problems!

Kernel for r-Dominating Set in nowhere dense classes?

Kernel for other problems under natural parametrisation.

Neighbourhood complexity of nowhere dense classes?

Better bounds for dtf-augmentations.



THANKS!

Questions?