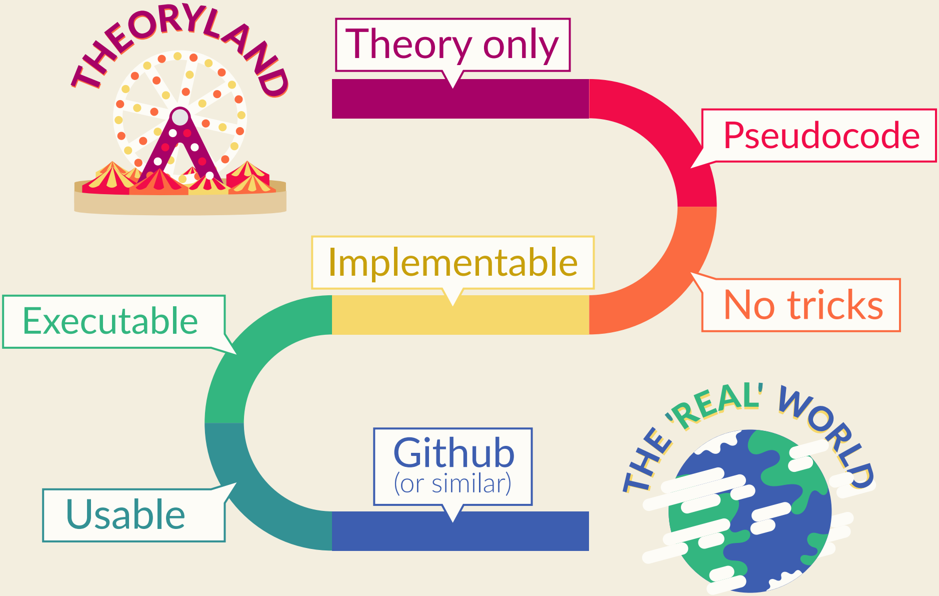
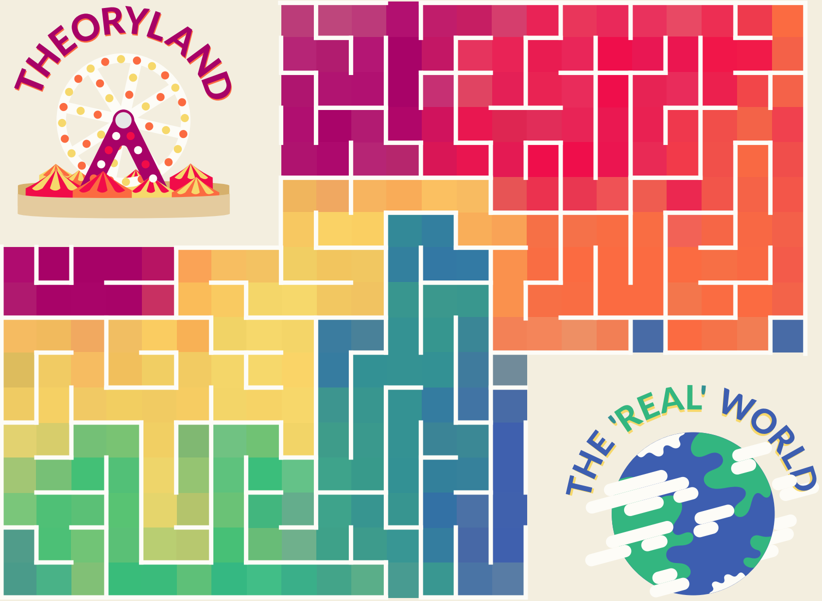
Complex networks & sparsity
Part IV: Implementation

Felix Reidl
Blair D. Sullivan
DOCCOURSE '18

The long and winding road

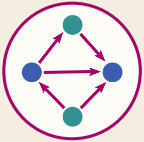


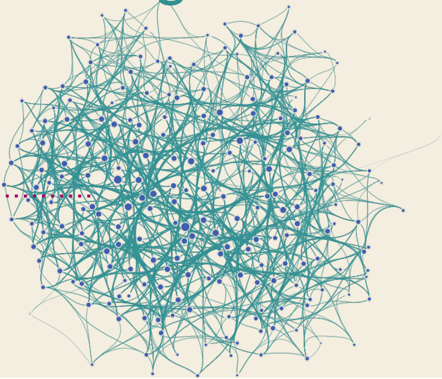
The long and winding ~~road~~



Motif-counting

We want to count the number of times a given **motif graph**





appears in a larger host graph (network).

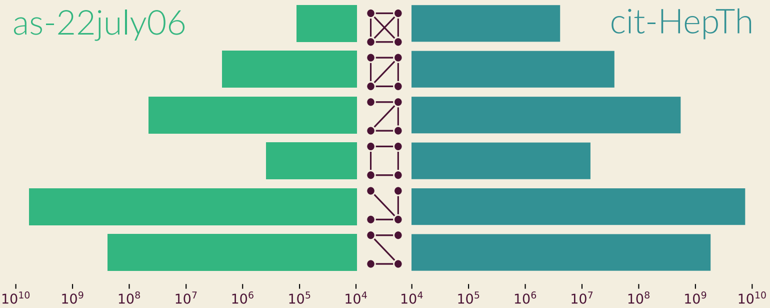
Motifs that appear more often **than expected** *might* play an important **function** in the network.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. **Network motifs: simple building blocks of complex networks**. Science. 2002 Oct 25;298(5594):824-7.

Ribeiro P, Silva F, Kaiser M. **Strategies for network motifs discovery**. InE-Science, 2009. e-Science'09. Fifth IEEE International Conference on 2009 Dec 9 (pp. 80-87). IEEE.

Graphlets

We want to count all (connected) induced subgraphs up to a given size.



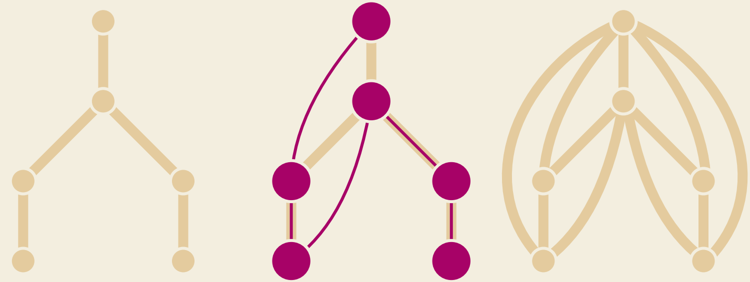
The **graphlet degree distribution** or the **graphlet degree** can be used to compare networks.

Pržulj N, Corneil DG, Jurisica I.
Modeling interactome: scale-free or geometric?.
Bioinformatics. 2004 Jul 29;20(18):3508-15

Pržulj N. B**iological network comparison using graphlet degree distribution.**
Bioinformatics. 2007 Jan 15;23(2):e177-83.

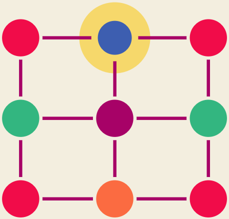
Treedepth

Def. A graph has *treedepth* d if it is the subgraph of the *closure* of a tree of height d.



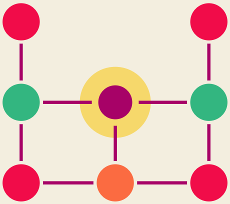
Treedepth: centered colourings

A vertex colouring is *centered* if every connected subgraph H contains a vertex whose color is unique in H.



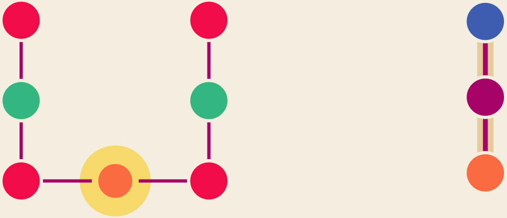
Treedepth: centered colourings

A vertex colouring is *centered* if every connected subgraph H contains a vertex whose color is unique in H.



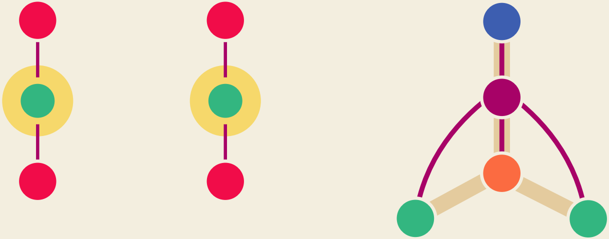
Treedepth: centered colourings

A vertex colouring is *centered* if every connected subgraph H contains a vertex whose color is unique in H.



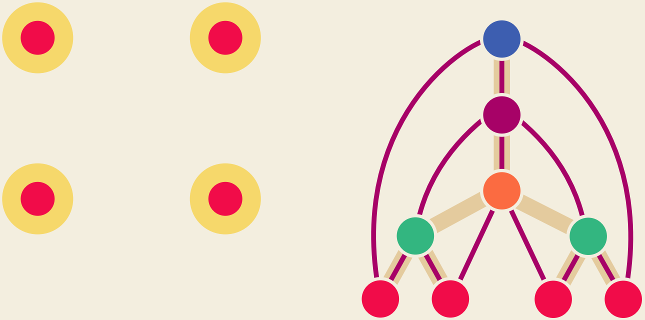
Treedepth: centered colourings

A vertex colouring is *centered* if every connected subgraph H contains a vertex whose color is unique in H.



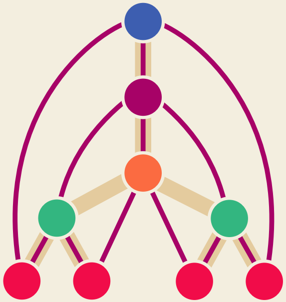
Treedepth: centered colourings

A vertex colouring is *centered* if every connected subgraph H contains a vertex whose color is unique in H.



Treedepth: centered colourings

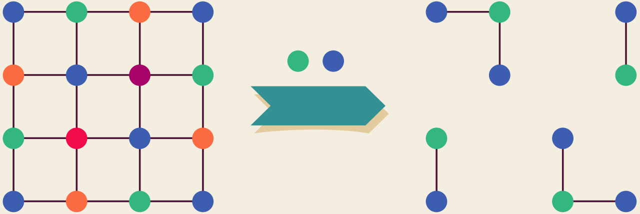
A vertex colouring is *centered* if every connected subgraph H contains a vertex whose color is unique in H.



The centered colouring number $\chi_{\mathrm{cen}}(G)$ is equal to the treedepth of G.

Low treedepth colourings

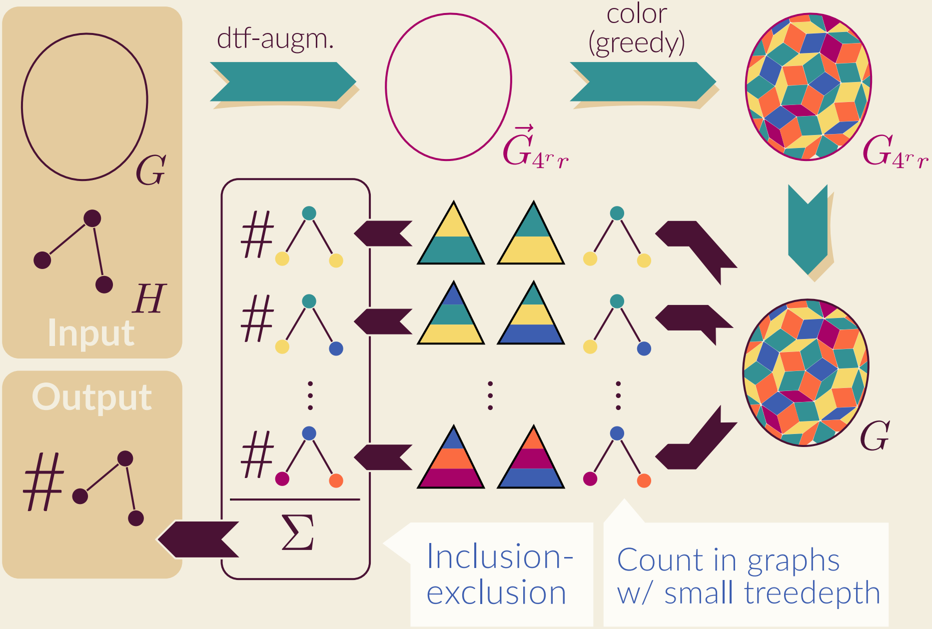
A vertex colouring is a **r-treedepth colouring** if every set of $i < r$ colours induce a subgraph of treedepth i.



$\chi_r(G) :=$ min #colours needed in an r-treedepth colouring of G

A graph class has bounded expansion iff it is χ_r-bounded.

Motif counting using χ_r

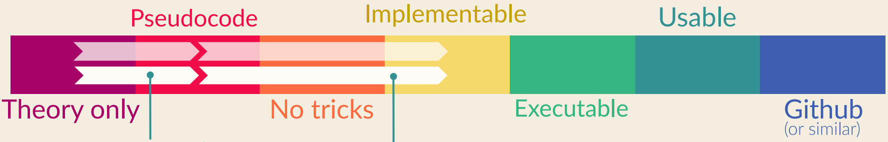


CONCUSS

Engineering motif-counting

$$f(h) \cdot 2^{O(h^2)} n$$

$$f(h) \cdot h^{O(h)} n$$

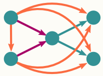


Nešetřil J, De Mendez PO.
Sparsity: graphs, structures, and algorithms.
Springer Science & Business Media; 2012 Apr 24.

Demaine ED, Reidl F, Rossmanith P,
Sánchez Villaamil F, Sikdar S, Sullivan BD.
Structural sparsity of complex networks:
Bounded expansion in random models and real-world graphs.
arXiv preprint arXiv:1406.2587. 2014 Jun 10.

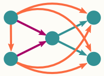
Engineering motif-counting

tf-augmentations



Absolutely impractical

dtf-augmentations



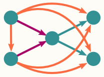
Test colouring after each step

Nešetřil J, De Mendez PO. **Sparsity: graphs, structures, and algorithms**. Springer Science & Business Media; 2012 Apr 24.

Reidl F. **Structural sparseness and complex networks**. (Doctoral dissertation, Dissertation, Aachen, Techn. Hochsch., 2015).

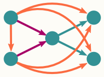
Engineering motif-counting

tf-augmentations



Absolutely impractical

dtf-augmentations



Test colouring after each step

Good engineering & heuristical improvements

Nešetřil J, De Mendez PO.
Sparsity: graphs, structures, and algorithms.
Springer Science & Business Media; 2012 Apr 24.

Reidl F. **Structural sparseness and complex networks**.
(Doctoral dissertation, Dissertation, Aachen, Techn. Hochsch., 2015).

Theory in Practice Group (NCSU)
with great help from students
Clayton G. Hobbs & Brandon Mork
https://github.com/TheoryInPractice/CONCUSS

Is it *practical*?

The current implementation is vastly outperformed by other algorithms (vf2) on practical instances.

Number of colours

Size of motif

$$\binom{f(h)}{h} h^{O(h)} n$$

There are artificial graph classes in which the algorithm performs better.

O'Brien MP, Sullivan BD.
Experimental evaluation of counting subgraph isomorphisms in classes of bounded expansion.
arXiv preprint arXiv:1712.06690. 2017 Dec 18.

Is it *practical*?

The current implementation is vastly outperformed by other algorithms (vf2) on practical instances.

Number of colours — $f(h)$

Size of motif — h

$$\binom{f(h)}{h} h^{O(h)} n$$

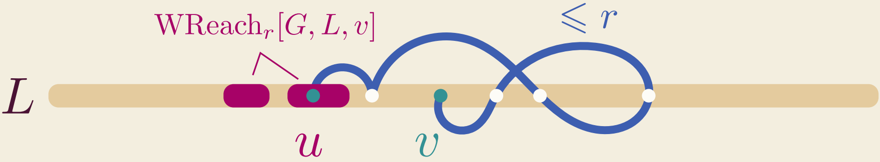
There are artificial graph classes in which the algorithm performs better.

O'Brien MP, Sullivan BD.
Experimental evaluation of counting subgraph isomorphisms in classes of bounded expansion.
arXiv preprint arXiv:1712.06690. 2017 Dec 18.

1) Improve colouring algorithm
2) Don't use low-treedepth colourings

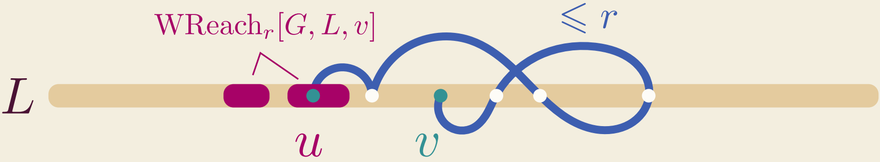
Mandoline

Weak coloring & bounded expansion



u is **weakly r-reachable** from v if there exists a path from v to u of length at most r such that u is the path's leftmost vertex.

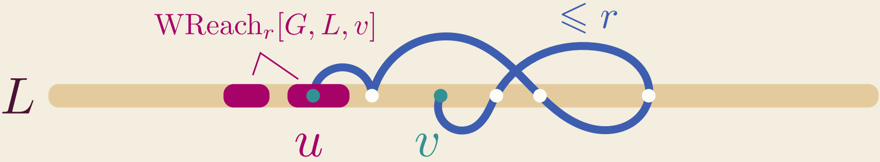
Weak coloring & bounded expansion



u is **weakly r-reachable** from v if there exists a path from v to u of length at most r such that u is the path's leftmost vertex.

$$\mathrm{wcol}_r(G) := \min_{L \in \Pi(G)} \max_{v \in G} |\,\mathrm{WReach}_r[G, L, v]|$$

Weak coloring & bounded expansion



u is **weakly r-reachable** from v if there exists a path from v to u of length at most r such that u is the path's leftmost vertex.

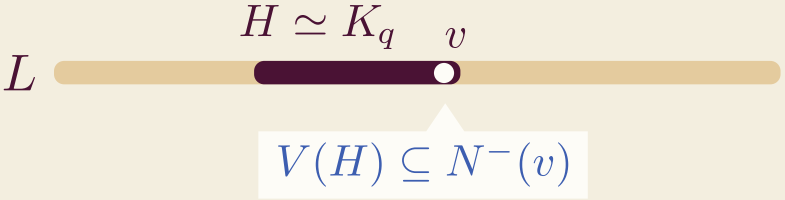
$$\mathrm{wcol}_r(G) := \min_{L \in \Pi(G)} \max_{v \in G} |\,\mathrm{WReach}_r[G, L, v]|$$

A graph class has bounded expansion iff it is wcol_r-bounded.

Let's start with something easy!

We count cliques in a d-degenerate graph.

Observation: every clique is contained in the left-neighbourhood of its *last* vertex.



Let's start with something easy!

We count cliques in a d-degenerate graph.

Observation: every clique is contained in the left-neighbourhood of its *last* vertex.

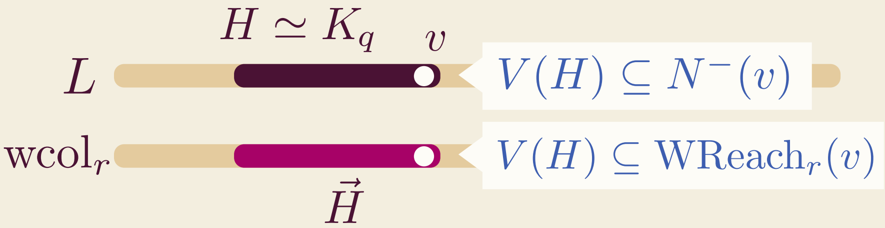
$V(H) \subseteq N^-(v)$

Therefore we can enumerate all cliques by enumerating all cliques in $N^-(v)$ for all $v \in G$!

$O(2^d n)$ time!

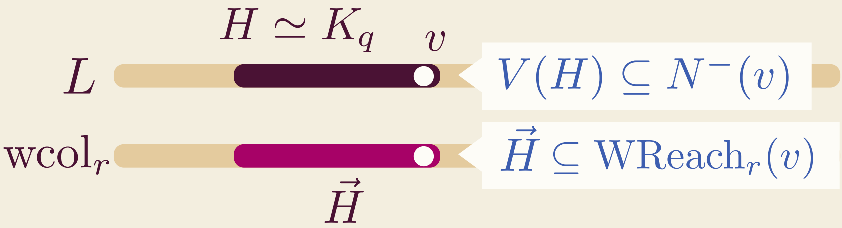
Does it blend?

Can we 'lift' this algorithm to wcol?



Does it blend?

Can we 'lift' this algorithm to wcol?

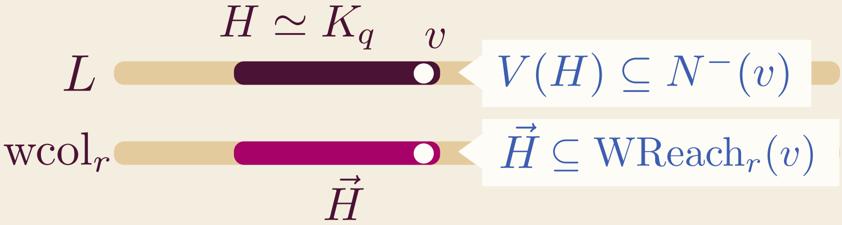


1. What is the 'last' vertex of H?

   Enumerate all orderings $\vec{H}$ of H.

Does it blend?

Can we 'lift' this algorithm to wcol?



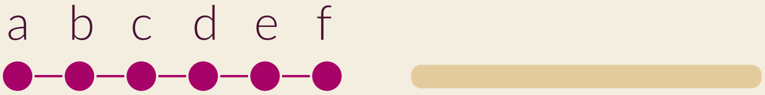
1. What is the 'last' vertex of H?

   Enumerate all orderings $\vec{H}$ of H.

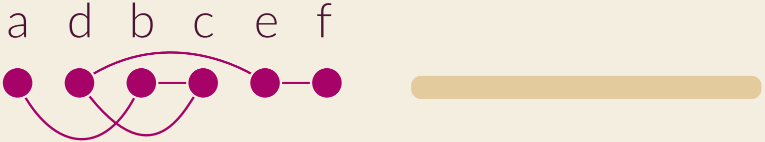
2. Does $\vec{H} \subseteq \mathrm{WReach}_r(v)$ actually hold?

   Only sometimes!

Two ways to order a P_6



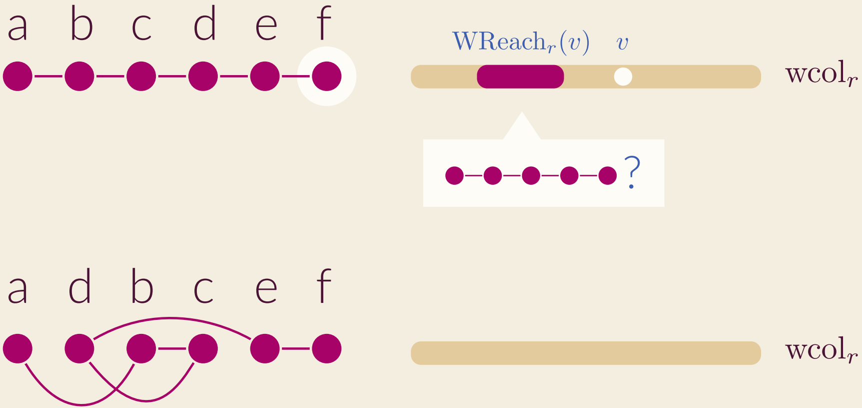
wcol_r



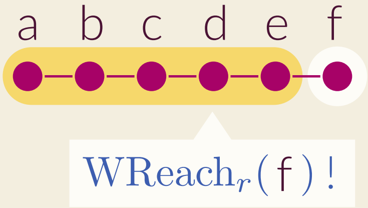
wcol_r

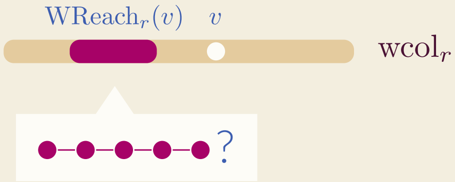
Two ways to order a P_6

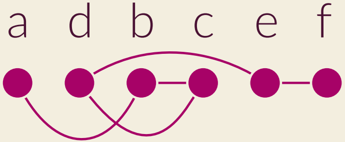
Two ways to order a P_6

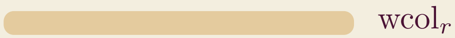


Two ways to order a P_6

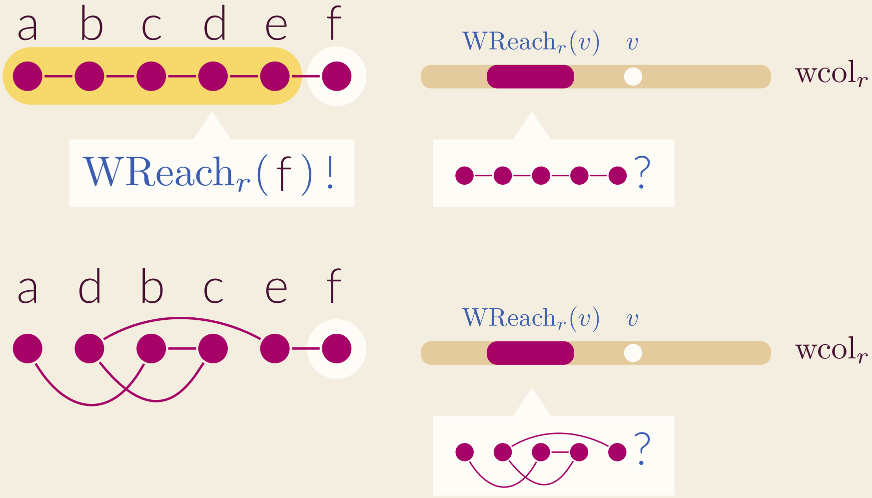




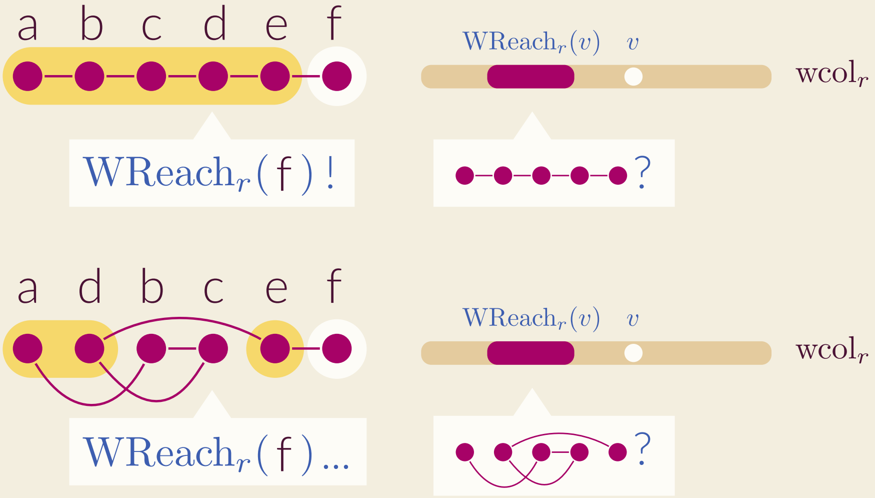




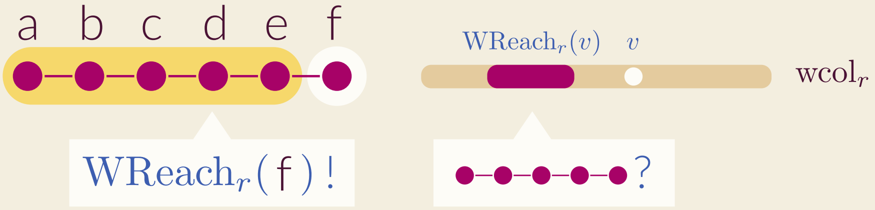
Two ways to order a P_6

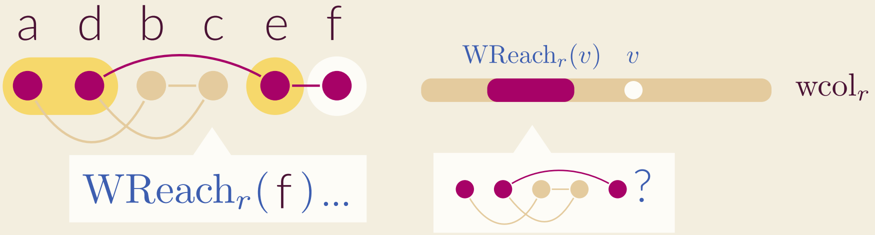


Two ways to order a P_6

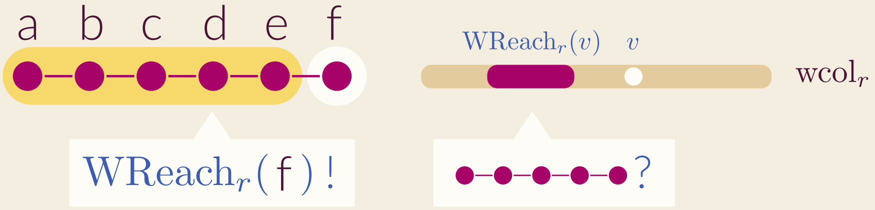


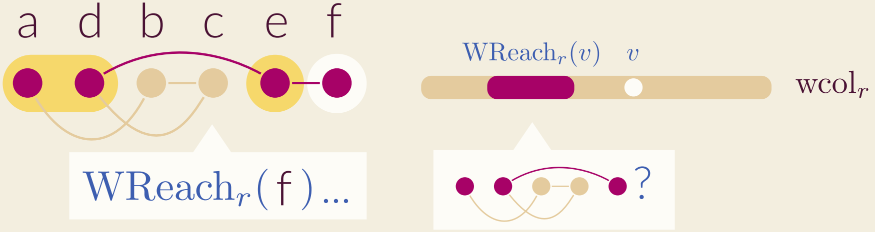
Two ways to order a P_6





Two ways to order a P_6





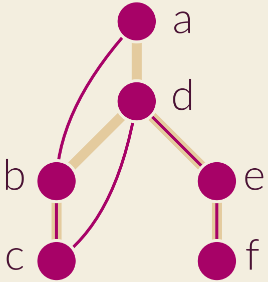
Is there a nice formalization of this property?

Treedepth: elimination orderings

Given an ordering $\prec$ of $V(G)$, we compute a treedepth decomposition as follows:

For every connected component of G, remove the minimum vertex and add it as the current root, then recurse on the resulting components.

adbcef adefbc
adbecf adebfc
adbefc adebcf

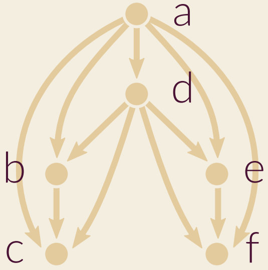


Treedepth: elimination orderings

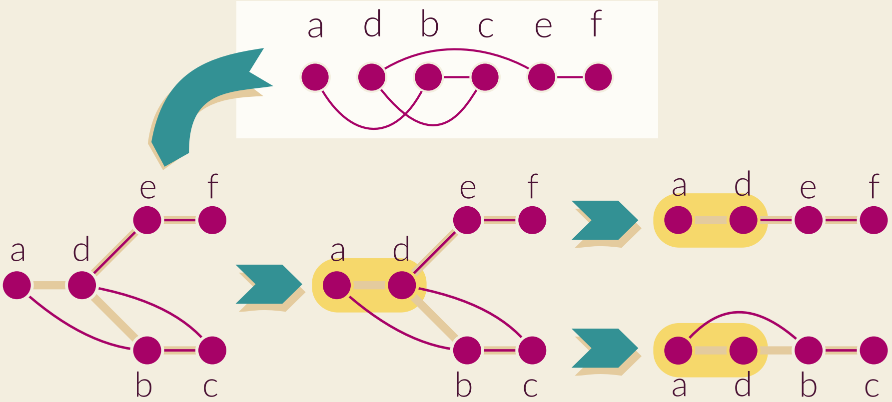
Given an ordering $\prec$ of $V(G)$, we compute a treedepth decomposition as follows:

For every connected component of G, remove the minimum vertex and add it as the current root, then recurse on the resulting components.

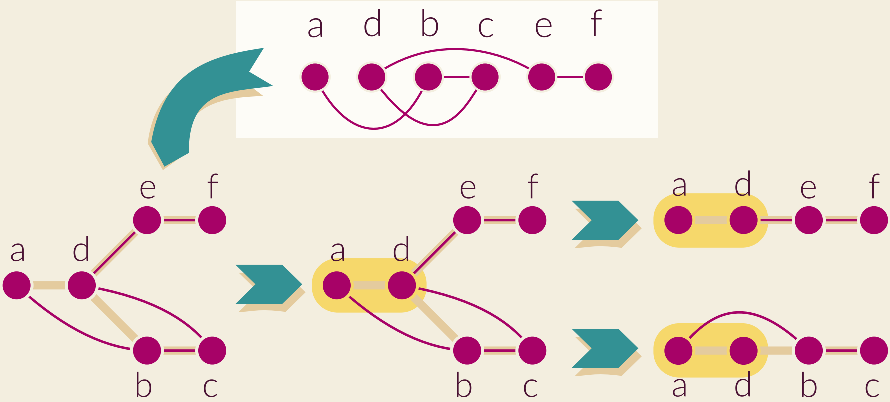
adbcef adefbc
adbecf adebfc
adbefc adebcf



Decomposition!



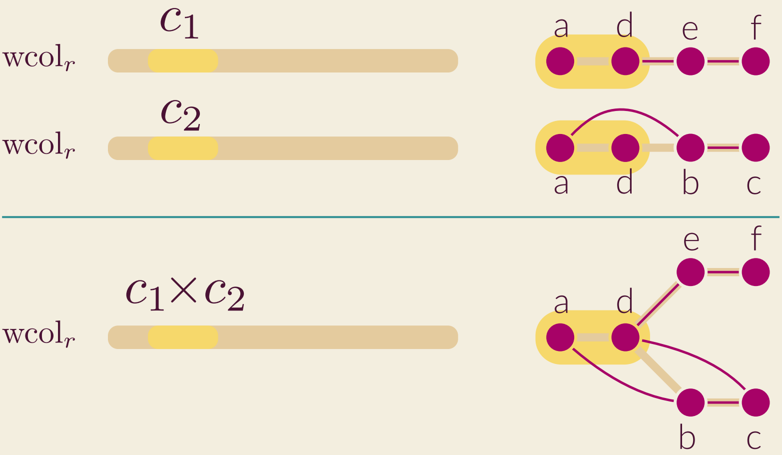
Decomposition!



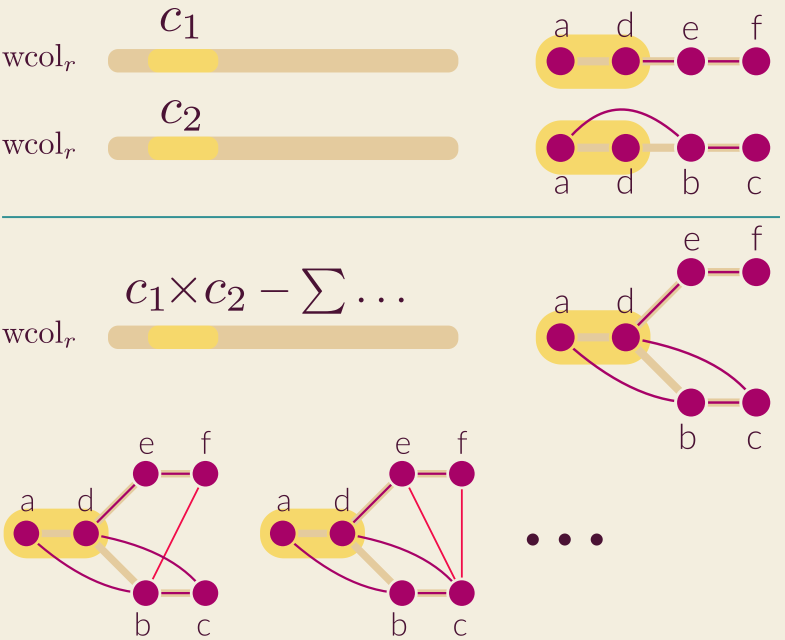
We can count linear pieces!

Progress! These pieces are linear!

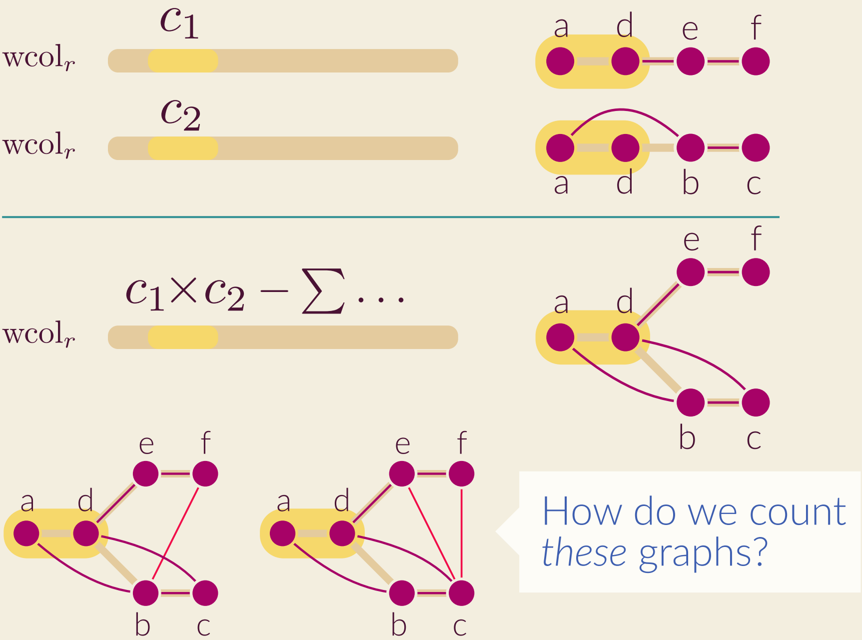
Count & combine!



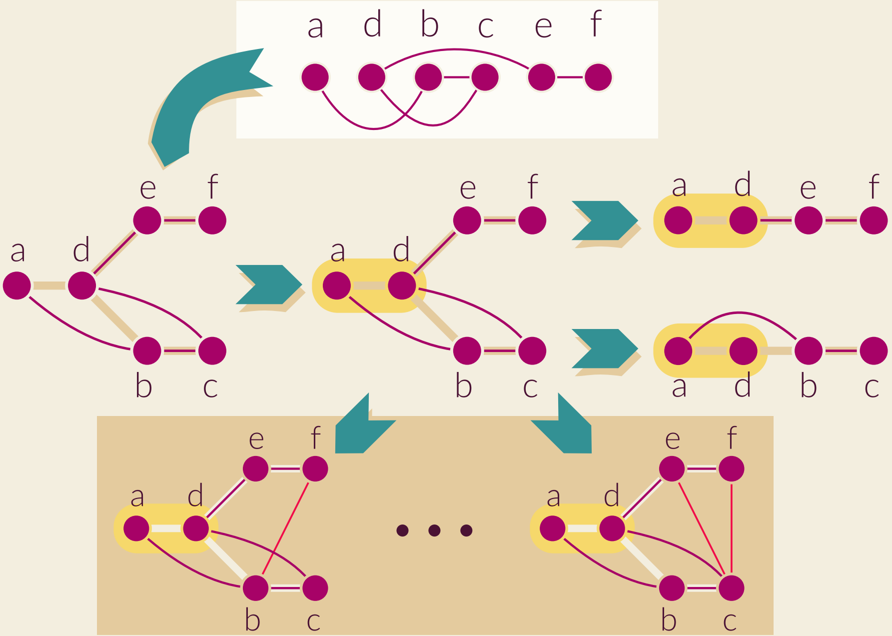
Count & combine!



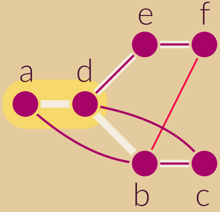
Count & combine!



Decomposition!

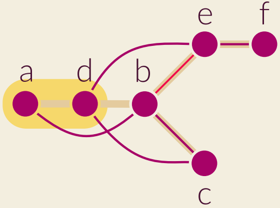


Decomposition!

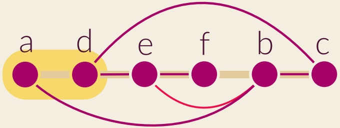


adbcef
adbecf
adbefc
adefbc
adebfc
adebcf

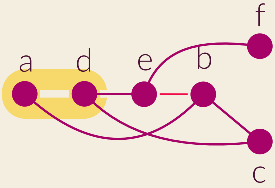
adbecf
adbefc
adbcef



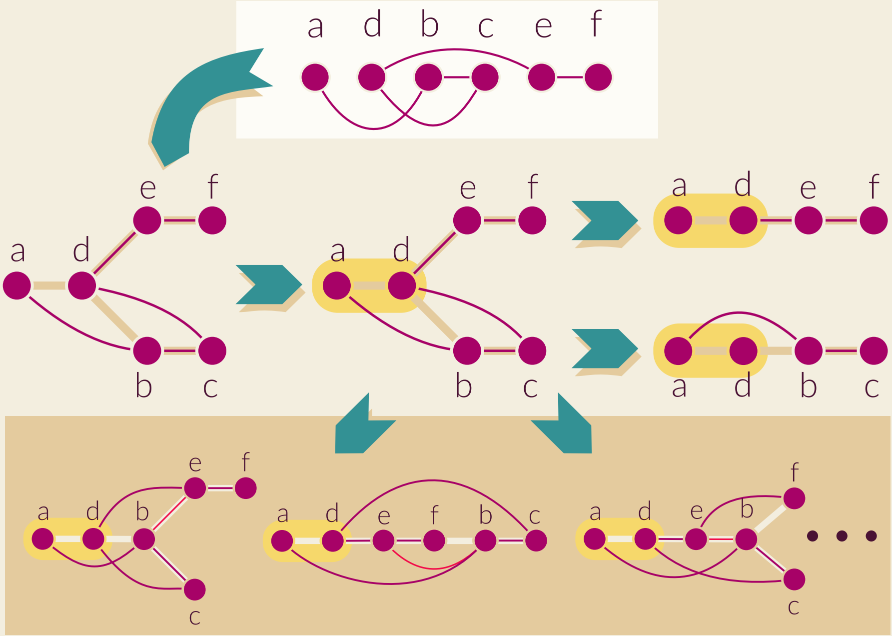
adefbc



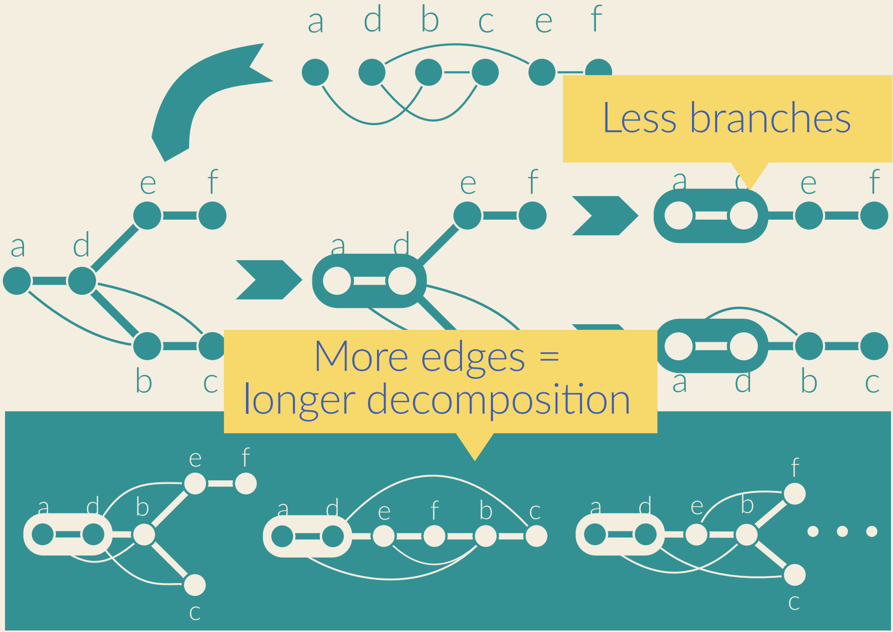
adebfc
adebcf



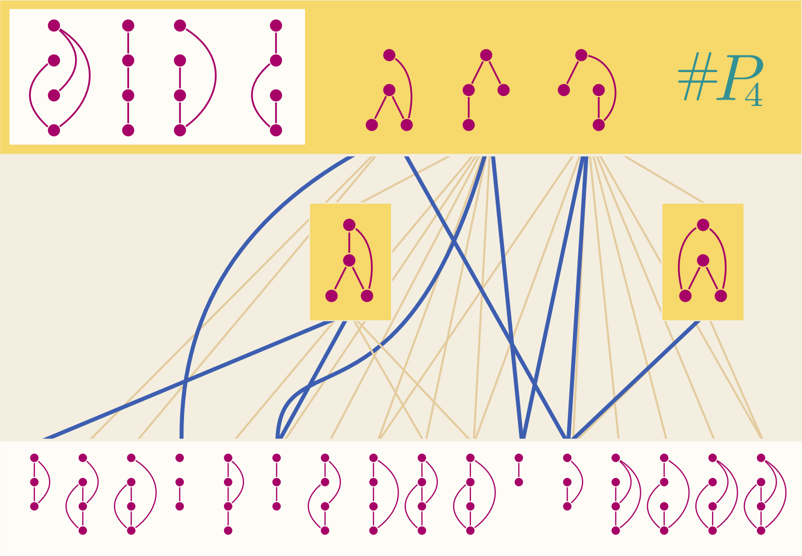
Decomposition!



Decomposition!

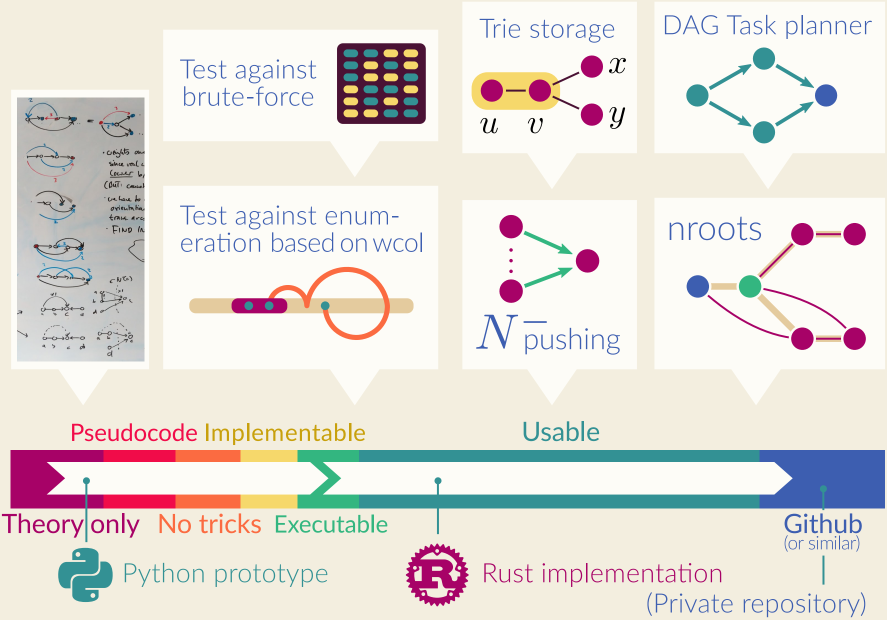


Motif counting using wcol_r

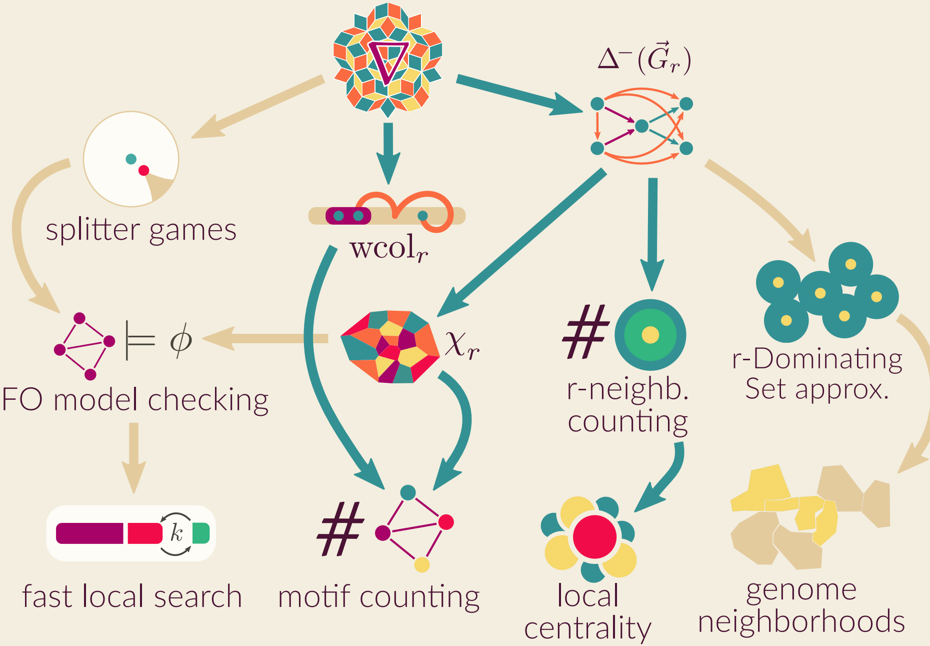


Composition

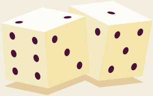
Mandoline progress



Applications & Algorithms



Open questions & future work



Classification of more models!

How do we classify models like preferential attachment?

Can we 'measure' bnd. exp. in practice?

Approximate r-nbhd counting without exponential dependence?

r-nbhd counting in time $O(2^{\omega_r(G)} n)$?

Make existing algorithms *practical*!

wcol/dtf specifically for networks!

Long term: 'sparsity' programming library

Find (interdisciplinary) collaborations

THANKS!

Demaine ED, Reidl F, Rossmanith P, Villaamil FS, Sikdar S, Sullivan BD. **Structural sparsity of complex networks: Random graph models and linear algorithms**. CoRR, abs/1406.2587. 2014 Jun 10.

Sullivan BD, Farrell M, Villaamil FS, Reidl F, Lemons N, Goodrich T. **Hyperbolicity, degeneracy, and expansion of random intersection graphs**. Internet Mathematics. 2017 Feb 9;9062017(1):1278.

Chin AJ, Goodrich TD, O'Brien MP, Reidl F, Sullivan BD, van der Poel A. **Asymptotic Analysis of Equivalences and Core-Structures in Kronecker-Style Graph Models**. InData Mining (ICDM), 2016 IEEE 16th International Conference on 2016 Dec 12 (pp. 829-834). IEEE.

O'Brien MP, Sullivan BD. **Experimental evaluation of counting subgraph isomorphisms in classes of bounded expansion**. arXiv preprint arXiv:1712.06690. 2017 Dec 18.

Gutin G, Mertzios GB, Reidl F. **Lower and Upper Bound for Computing the Size of All Second Neighbourhoods**. arXiv preprint arXiv:1805.01684. 2018 May 4.

GORDON AND BETTY MOORE FOUNDATION