Optimal graphlet structures

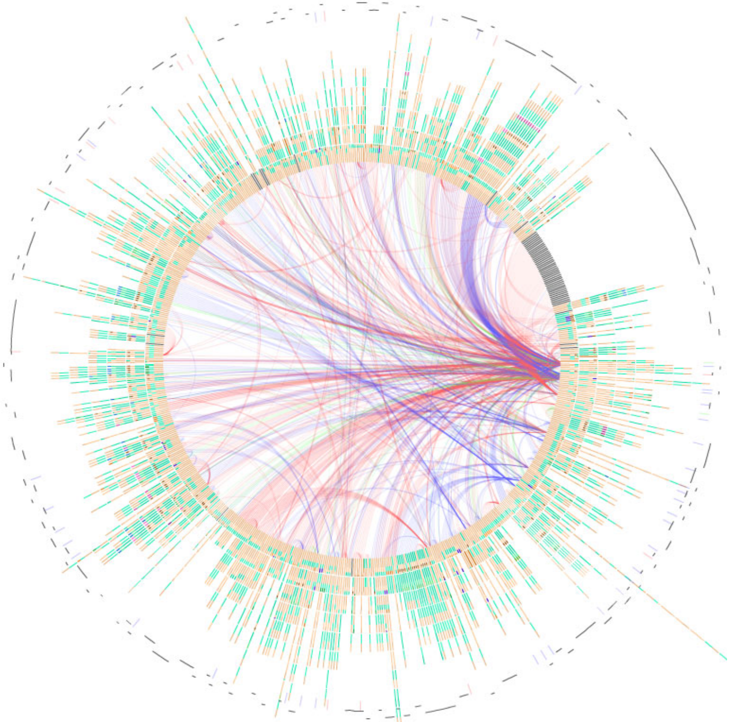
Clara Stegehuis

Eindhoven University of Technology

joint work with Remco van der Hofstad and Johan van Leeuwaarden

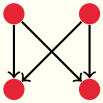
arXiv:1709.03466

May 9, 2018



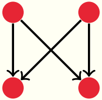
Frequently occurring subgraphs

Gene network

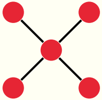


Frequently occurring subgraphs

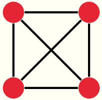
Gene network



Internet



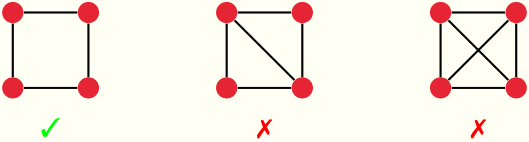
Collaborations



Graphlet counting

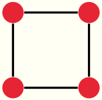
How many times does a connected graph H occur as *induced subgraph* in a random graph on n vertices?

Example:



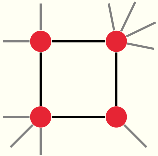
Graphlet counting

On which vertices does H typically occur as *induced subgraph* in a random graph on n vertices?



Graphlet counting

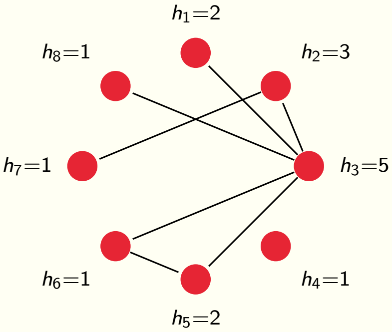
On which vertices does H typically occur as *induced subgraph* in a random graph on n vertices?



Inhomogeneous random graph

n vertices, weights h_i

$$p(h, h') = \min\left(\frac{hh'}{\mu n}, 1\right)$$



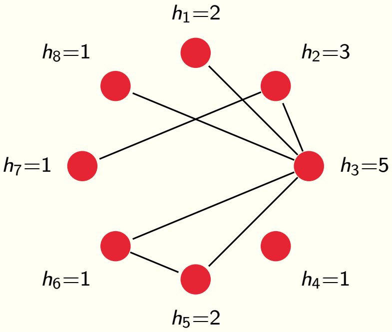
Inhomogeneous random graph

n vertices, weights h_i

$p(h, h') = \min\left(\frac{hh'}{\mu n}, 1\right)$

Degree of vertex $i \approx h_i$

$\mathbb{P}(h_i = x) \sim x^{-\tau} \quad \tau \in (2, 3)$



Example: Triangle counting

Number of triangles

$$N(\triangle) = \sum_{d_1,d_2,d_3} N(\triangle \text{ between vertices of degrees } d_1, d_2, d_3)$$

Number of triangles

$$N(\triangle) = \sum_{d_1,d_2,d_3} N(\triangle \text{ between vertices of degrees } d_1, d_2, d_3)$$
$$\approx N(\triangle \text{ between vertices of degrees } d_1^*, d_2^*, d_3^*)$$

Number of triangles

Choose $d_1 = \Theta(n^{\alpha_1})$, $d_2 = \Theta(n^{\alpha_2})$, $d_3 = \Theta(n^{\alpha_3})$.

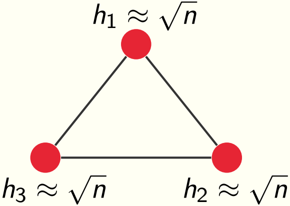
$$N(\triangle \text{ on degrees } n^{\alpha_1}, n^{\alpha_2}, n^{\alpha_3})$$
$$\approx (\# \text{ vertices of degrees } n^{\alpha_1}, n^{\alpha_2}, n^{\alpha_3})\mathbb{P}(\triangle \text{ on } n^{\alpha_1}, n^{\alpha_2}, n^{\alpha_3})$$

Number of triangles

$$N(\triangle \text{ on degrees } \Theta(n^{\alpha_1}), \Theta(n^{\alpha_2}), \Theta(n^{\alpha_3}))$$
$$\propto n^{(\alpha_1+\alpha_2+\alpha_3)(1-\tau)+1}$$
$$\times \min(n^{\alpha_1+\alpha_2-1}, 1) \min(n^{\alpha_1+\alpha_3-1}, 1) \min(n^{\alpha_2+\alpha_3-1}, 1)$$

Optimal triangle

$\alpha_1^* = \alpha_2^* = \alpha_3^* = \frac{1}{2}$



Counting other graphlets

Graphlet counting

$$N(H) = \sum_{d_1,\ldots,d_k} N(H \text{ on vertices of degrees } d_1,\ldots,d_k)$$
$$\approx N(H \text{ on vertices of degrees } d_1^*,\ldots,d_k^*)$$

Choose $d_1 = \Theta(n^{\alpha_1}),\ldots,d_k = \Theta(n^{\alpha_k})$, and find $\alpha_1^*,\ldots,\alpha_k^*$.

Graphlet counting

How many times occurs H as induced subgraph on vertices of degrees proportional to $(n^{\alpha_i})_{i=1,\dots,|V_H|}$?

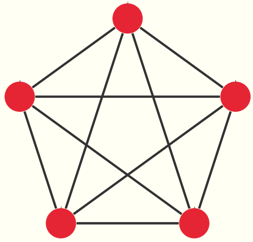
$$\{\# \text{ vertex sets with degrees } (n^{\alpha_i})_{i\in 1,\dots,|V_H|}\}$$
$$\times\, \mathbb{P}(H \text{ induced subgraph on these degrees})$$

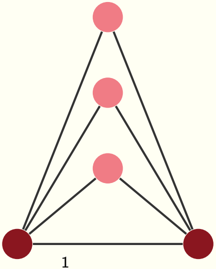
Graphlet counting

How many times occurs H as induced subgraph on vertices of degrees proportional to $(n^{\alpha_i})_{i=1,\dots,|V_H|}$?

$$\{\#\text{ vertex sets with degrees } (n^{\alpha_i})_{i\in 1,\dots,|V_H|}\}$$
$$\times\, \mathbb{P}(H \text{ induced subgraph on these degrees})$$
$$= \Theta(n^{|V_H|+\sum_i \alpha_i(1-\tau)})\Theta\left(\prod_{(i,j)\in E_H} \min(n^{\alpha_i+\alpha_j-1}, 1)\right)$$
$$\times\, \Theta\left(\prod_{(i,j)\notin E_H} \max(1-n^{\alpha_i+\alpha_j-1}, 0)\right)$$

Typical degrees of graphlets





Degrees in most likely configuration

$\Theta(1)$

$\Theta(n^{\frac{\tau-2}{\tau-1}})$

$\Theta(\sqrt{n})$

$\Theta(n^{\frac{1}{\tau-1}})$

Optimization problem

$$B(H) = \max_{\mathcal{P}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1,S_3} + E_{S_1,1} - E_{S_2,1}}{\tau - 1} \right],$$
$$\text{s.t.} \quad (u,v) \in E_H \quad \forall u \in S_2, v \in S_2 \cup S_3,$$

finds optimal partition into vertices with degrees $\Theta(n^{\frac{\tau-2}{\tau-1}})$, $\Theta(n^{\frac{1}{\tau-1}})$ and $\Theta(\sqrt{n})$.

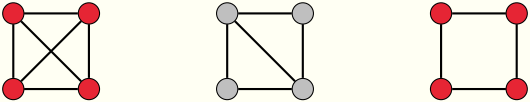
Optimization problem

$$B(H) = \max_{\mathcal{P}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1,S_3} + E_{S_1,1} - E_{S_2,1}}{\tau - 1} \right],$$

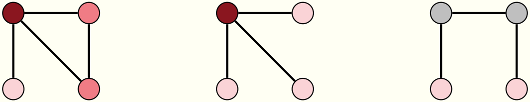
$$\text{s.t.} \quad (u, v) \in E_H \quad \forall u \in S_2, v \in S_2 \cup S_3,$$

finds optimal partition into vertices with degrees $\Theta(n^{\frac{\tau-2}{\tau-1}})$, $\Theta(n^{\frac{1}{\tau-1}})$ and $\Theta(\sqrt{n})$.
A function of $B(H)$ computes the graphlet count of H.

Atlas of graphlets



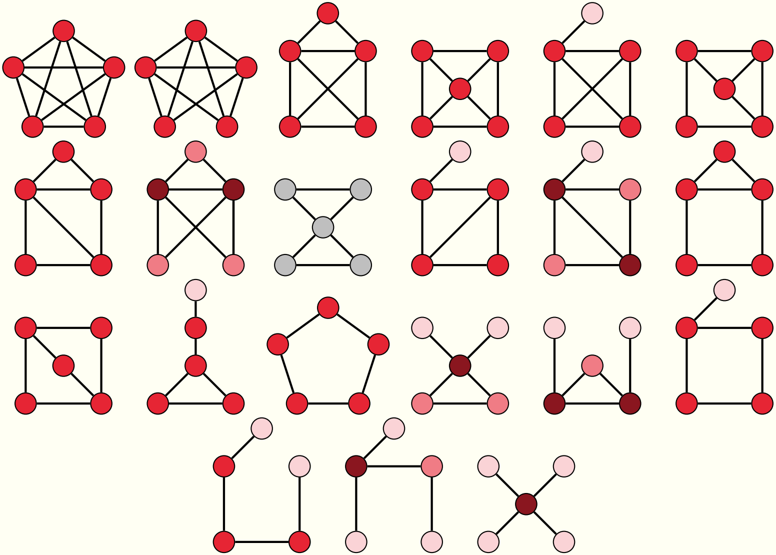
$n^{6-2\tau}$ | $n^{6-2\tau}\log(n)$ | $n^{6-2\tau}$



$n^{7-2\tau-\frac{1}{\tau-1}}$ | $n^{\frac{3}{\tau-1}}$ | $n^{4-\tau}\log(n)$

- : Vertices with degree $\Theta(1)$,
- : Vertices with degree $\Theta(n^{\frac{\tau-2}{\tau-1}})$,
- : Vertices with degree $\Theta(\sqrt{n})$,
- : Vertices with degree $\Theta(n^{\frac{1}{\tau-1}})$
- : Vertices where the optimizer is not unique.

Atlas of graphlets



Conclusion

- Graphlets occur on specific degrees
- These specific degrees take only four different orders of magnitude