Finding induced subgraphs

Clara Stegehuis

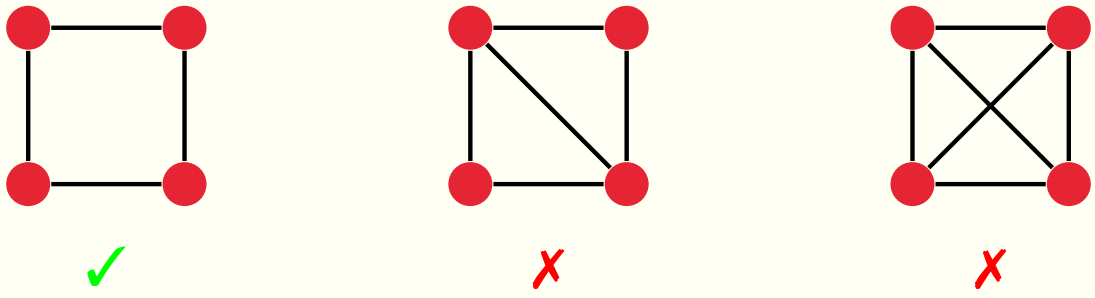
Eindhoven University of Technology
joint work with Ellen Cardinaels and Johan van Leeuwaarden

May 17, 2018

Induced subgraph problem

Does a connected graph H occur as *induced subgraph* in graph G on n vertices?

Example:



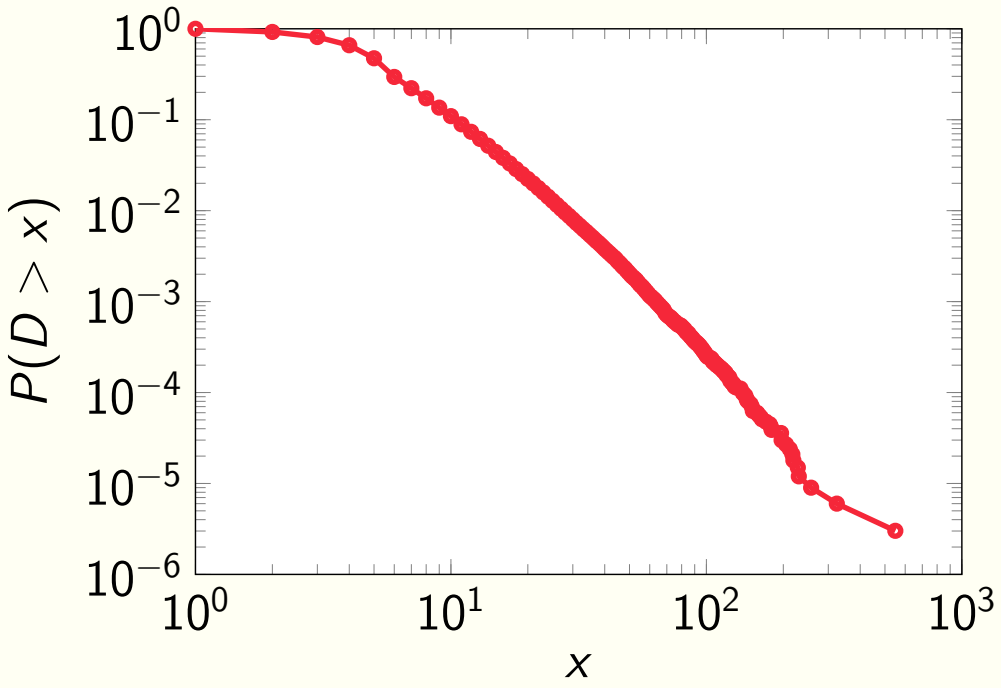
Induced subgraph problem

- Difficult in general: MAX CLIQUE is NP hard on general networks
- For fixed k, can be solved in $O(n^k)$ time.
- For $k = 4$ can be solved in $O(n^{1.51})$ time whp.

Induced subgraph problem

How can we find an induced subgraph in a random graph?

Scale-free networks

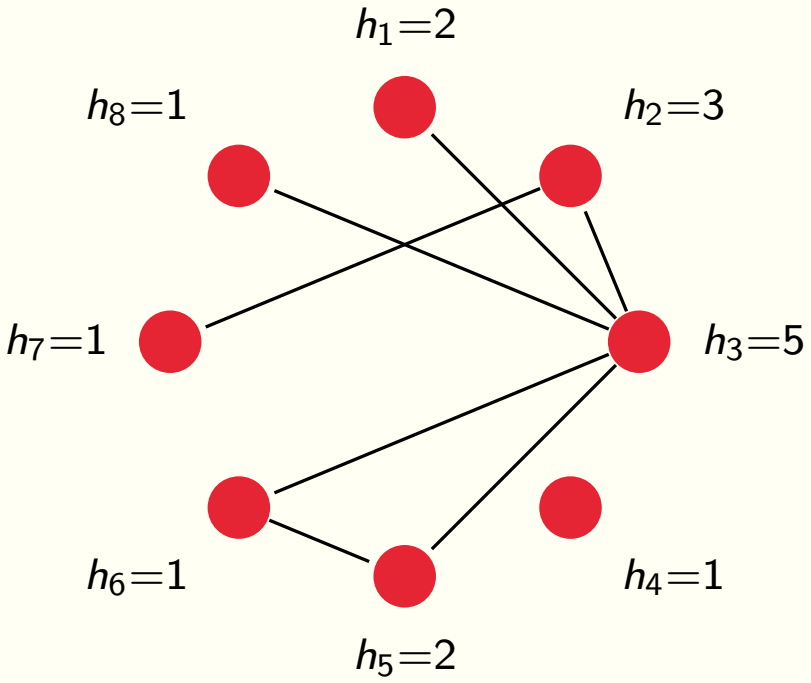


$$P(D = k) \approx k^{-\tau}, \quad \tau \in (2, 3)$$

Inhomogeneous random graph

n vertices, weights h_i

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



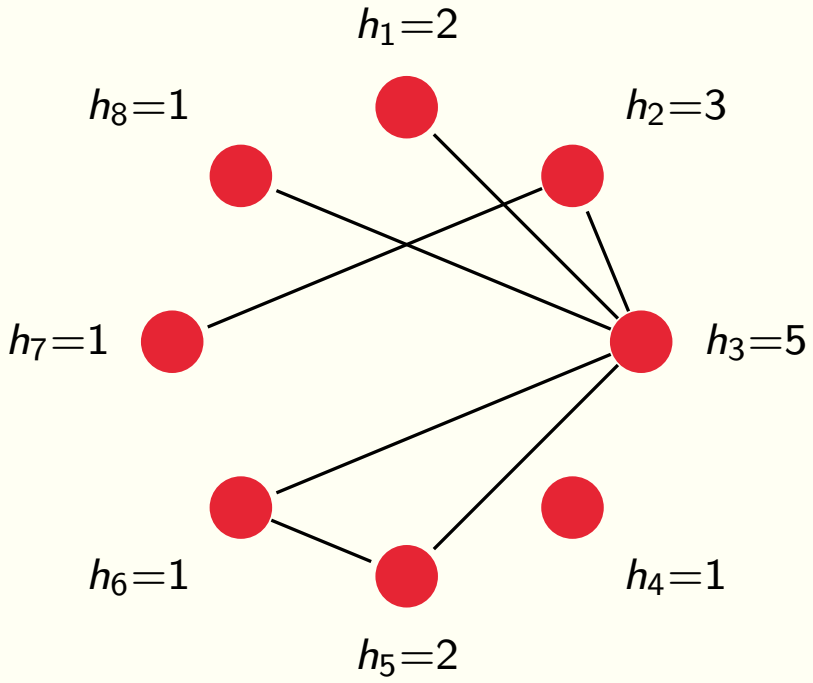
Inhomogeneous random graph

n vertices, weights h_i

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

$d_i \approx h_i$

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



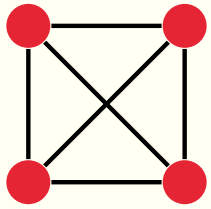
Finding cliques [Friedrich 2015]

Algorithm to find clique of size k:

1. Delete vertices with $D_i < \sqrt{n}/\log(n)$.
   Repeat:
2. Select k randomly chosen vertices
3. Check if they form a clique.

Large degrees

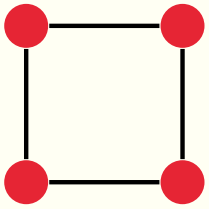
When $D_i, D_j \gg \sqrt{n}$:
$p(D_i, D_j) = \min(D_i D_j / \mu n, 1) = 1.$



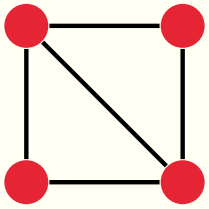
✓

Large degrees

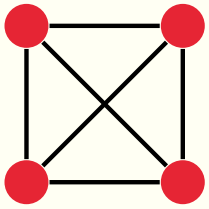
When $D_i, D_j \gg \sqrt{n}$:
$p(D_i, D_j) = \min(D_i D_j / \mu n, 1) = 1.$



✗



✗



✓

$\sqrt{n}$ degrees

Choose a, b such that $ab < \mu$.

$$p(a\sqrt{n}, b\sqrt{n}) = \min\left(\frac{ab}{\mu}, 1\right) = \frac{ab}{\mu} < 1$$

$$P(H \text{ induced on } \sqrt{n} \text{ degrees}) = C_H$$

On average, find H in $1/C_H$ searches.

Finding induced subgraphs

Algorithm to find H on k vertices:

1. Delete vertices with $D_i \notin [a\sqrt{n}, b\sqrt{n}]$
   Repeat:
2. Select k randomly chosen vertices
3. Check if H is present

Finding subgraphs

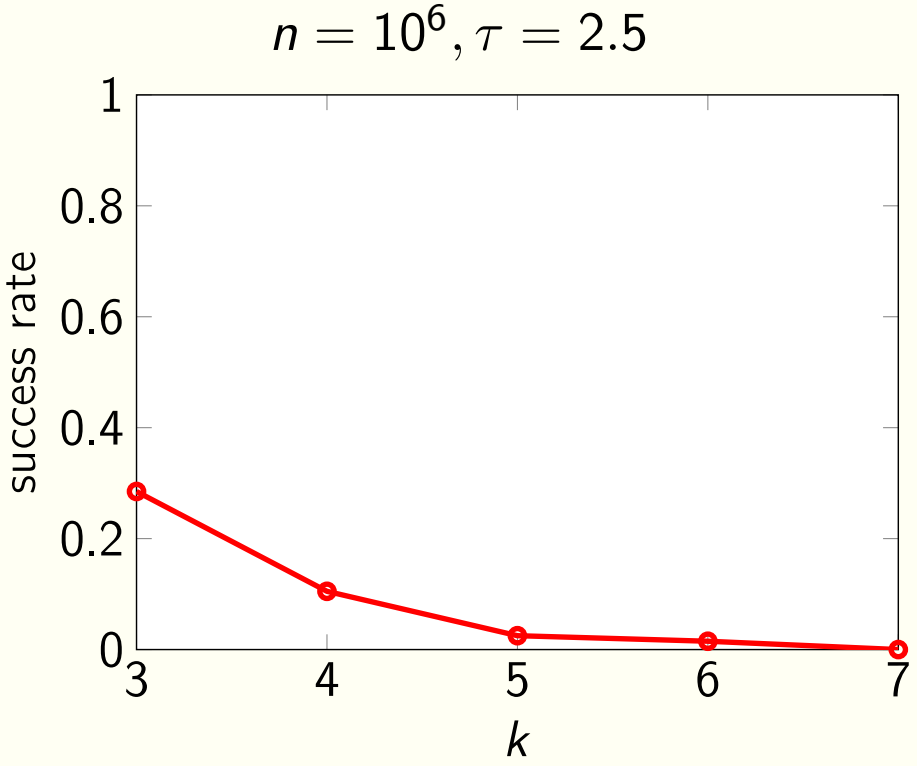
Theorem

For any connected graph H of fixed size k, Algorithm 1 finds H as induced subgraph in at most nk edge checks with high probability.

Analysis

1. Show that weights and degrees are close
2. Show that polynomially many vertices remain
3. In remaining graph $P(H \text{ is induced subgraph}) > C_H > 0$.

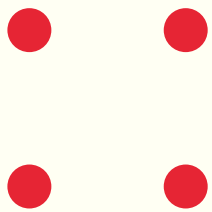
Results: Cycles of length k



Choosing k vertices

Option 1

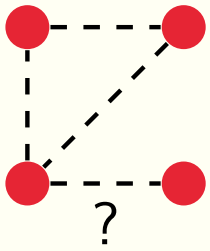
Choose uniformly at random



Choosing k vertices

Option 1

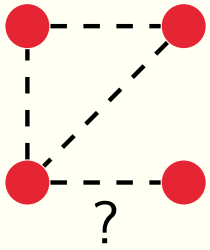
Choose uniformly at random



Choosing k vertices

Option 1

Choose uniformly at random



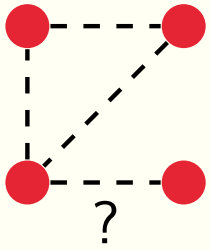
Option 2

Choose first vertex uniformly, then choose neighbors

Choosing k vertices

Option 1

Choose uniformly at random



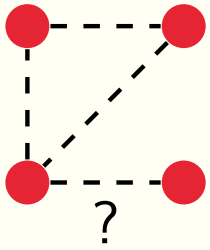
Option 2

Choose first vertex uniformly, then choose neighbors

Choosing k vertices

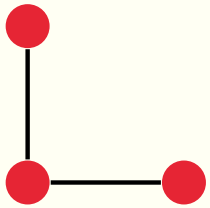
Option 1

Choose uniformly at random



Option 2

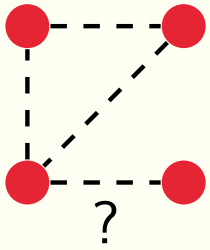
Choose first vertex uniformly, then choose neighbors



Choosing k vertices

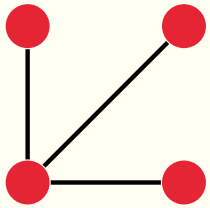
Option 1

Choose uniformly at random



Option 2

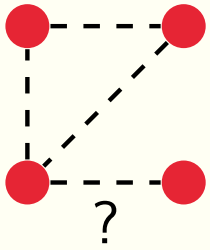
Choose first vertex uniformly, then choose neighbors



Choosing k vertices

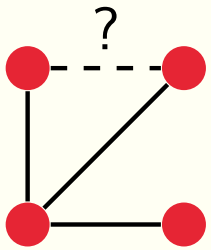
Option 1

Choose uniformly at random

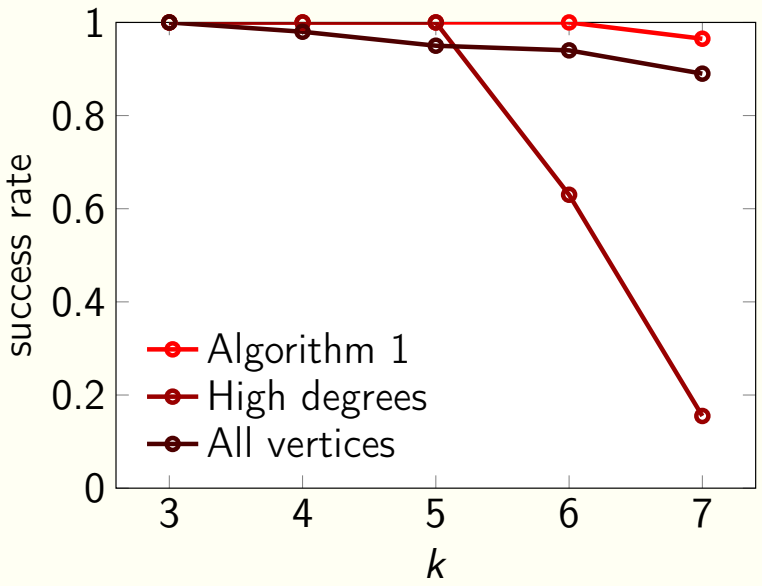


Option 2

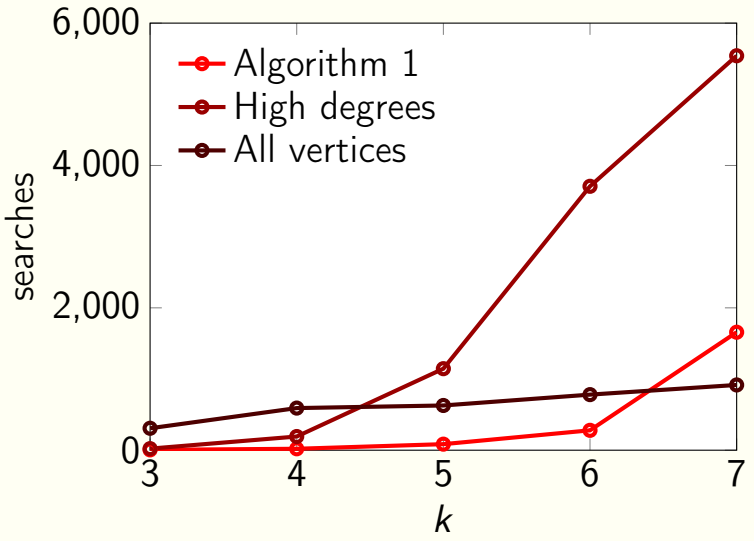
Choose first vertex uniformly, then choose neighbors



Performance: Cycles of length k

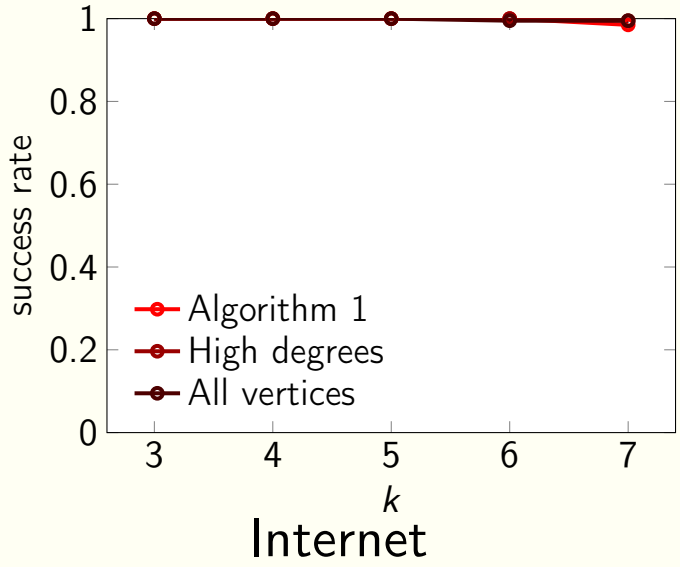


Success rate

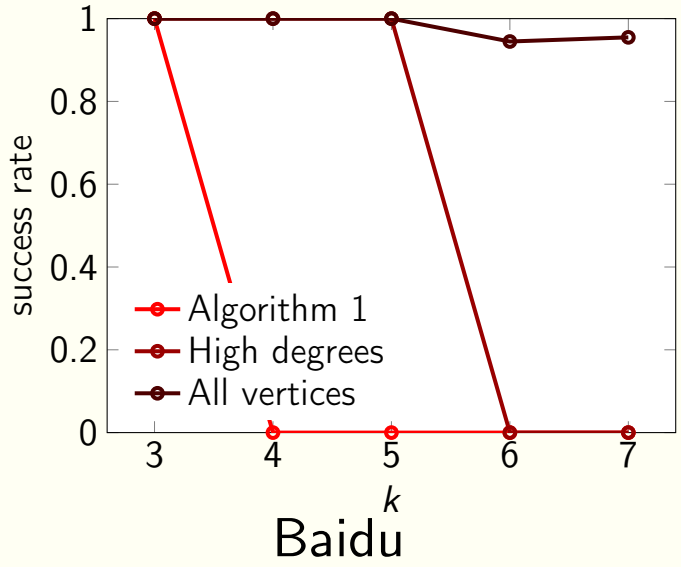


Number of searches

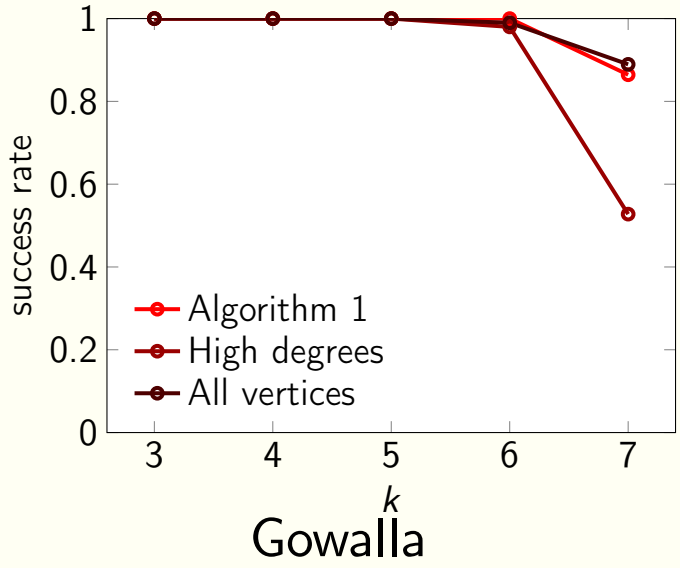
Real-world data: Success rate



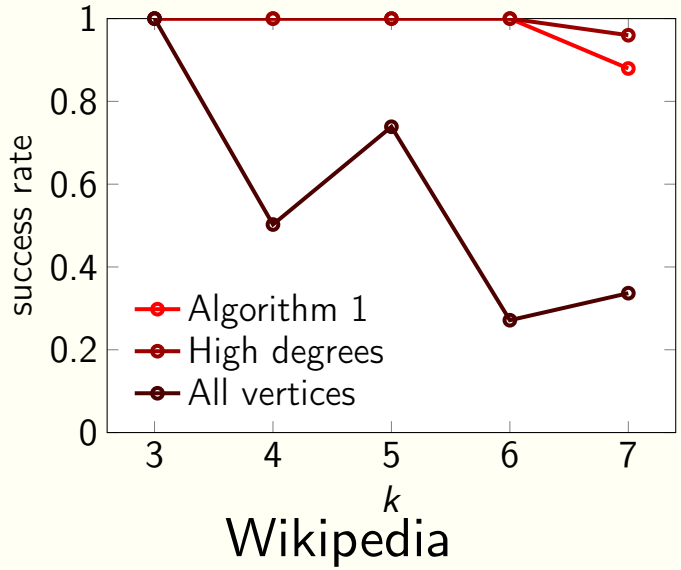
Internet



Baidu

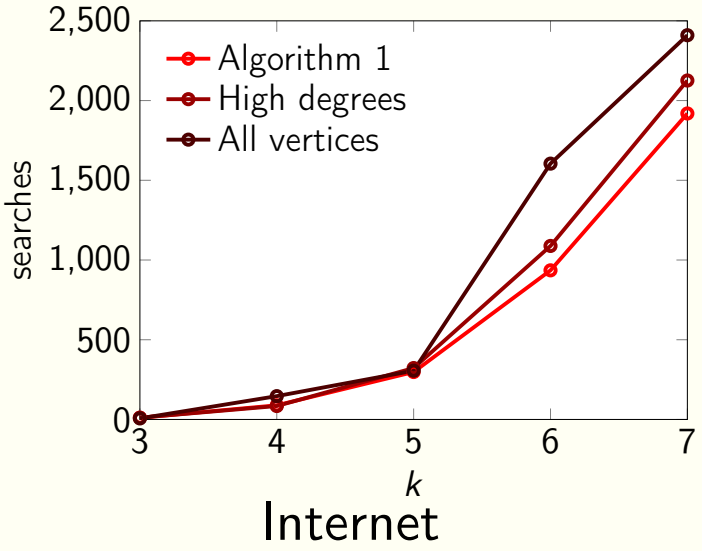


Gowalla

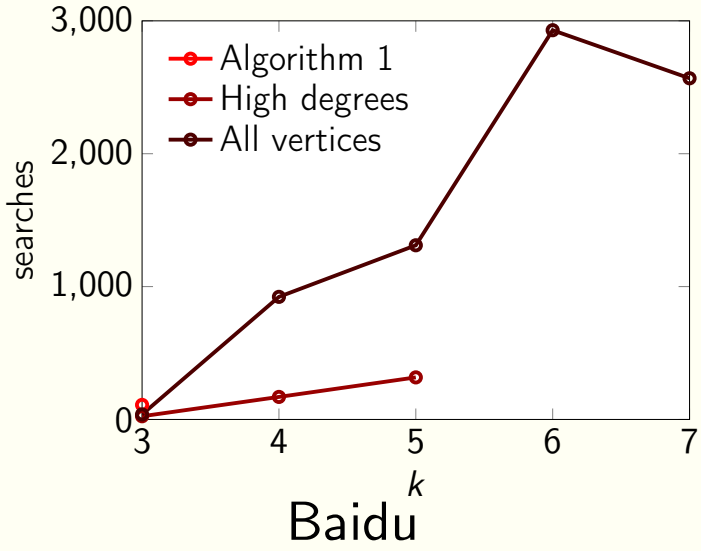


Wikipedia

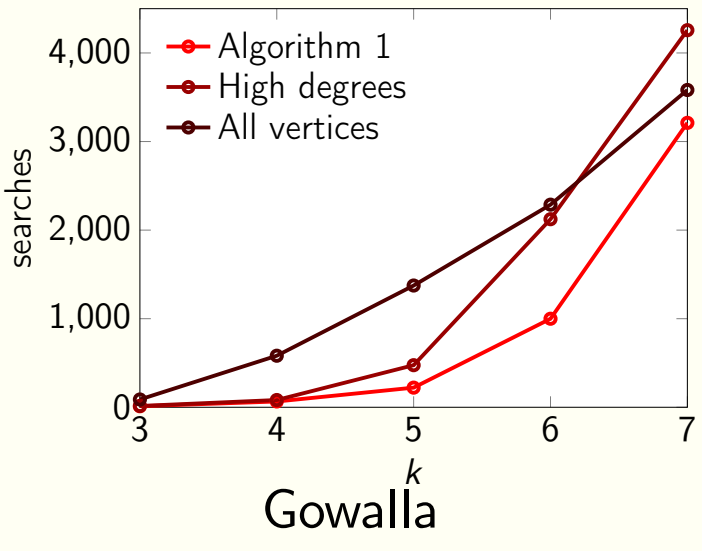
Real-world data: Searches



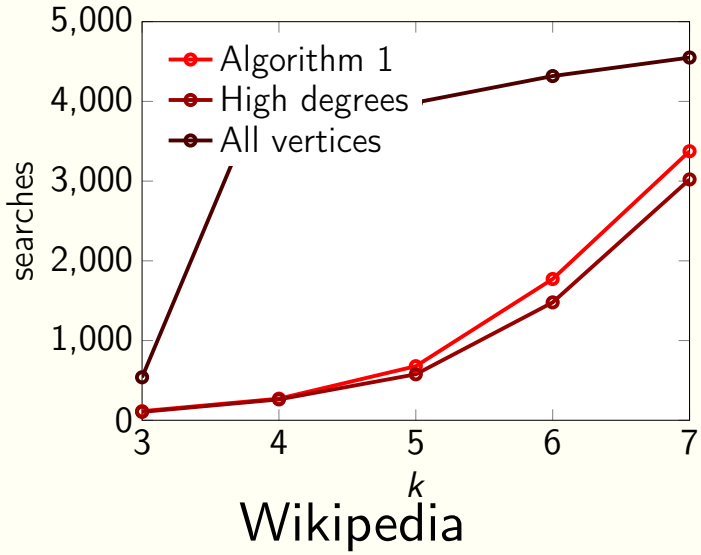
Internet



Baidu



Gowalla

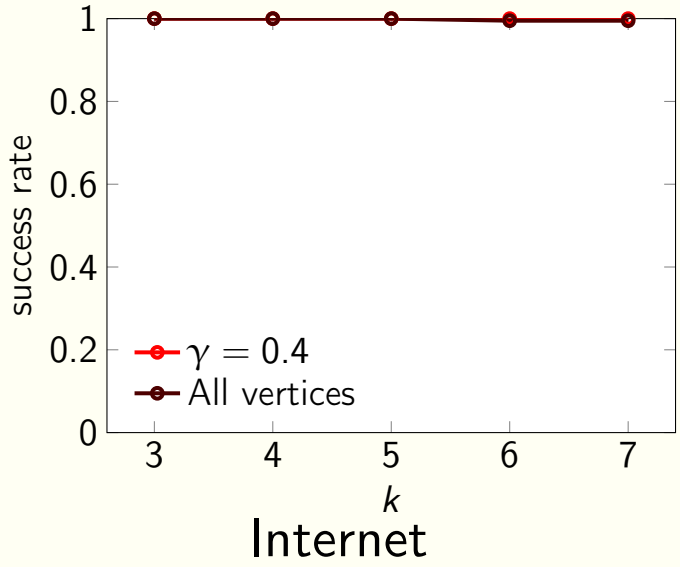


Wikipedia

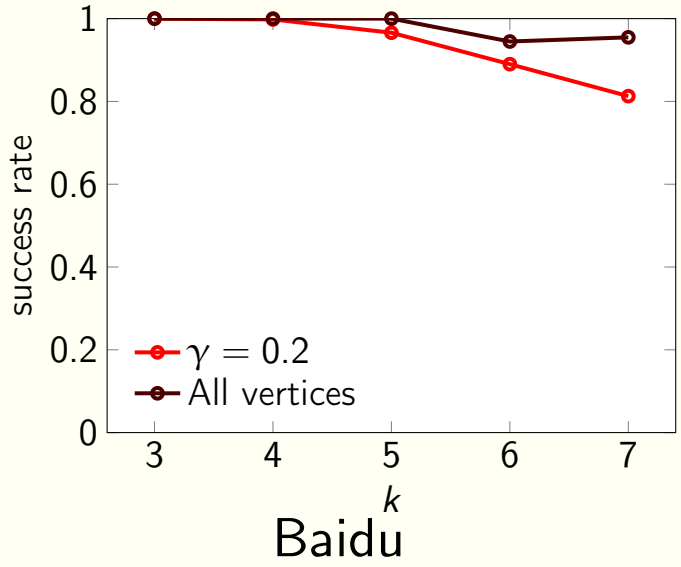
Real-world data

- In real-world networks, searching on $\sqrt{n}$ degrees is not efficient
- Search on degrees proportional to n^{γ} instead

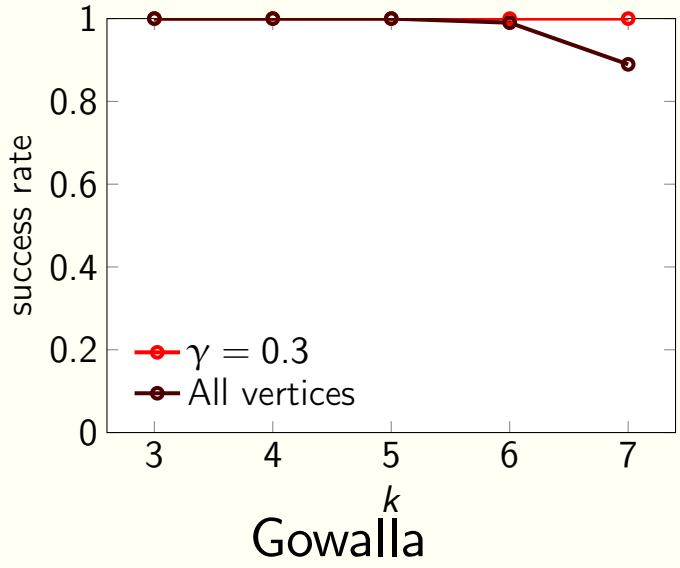
Real-world data: Success rate



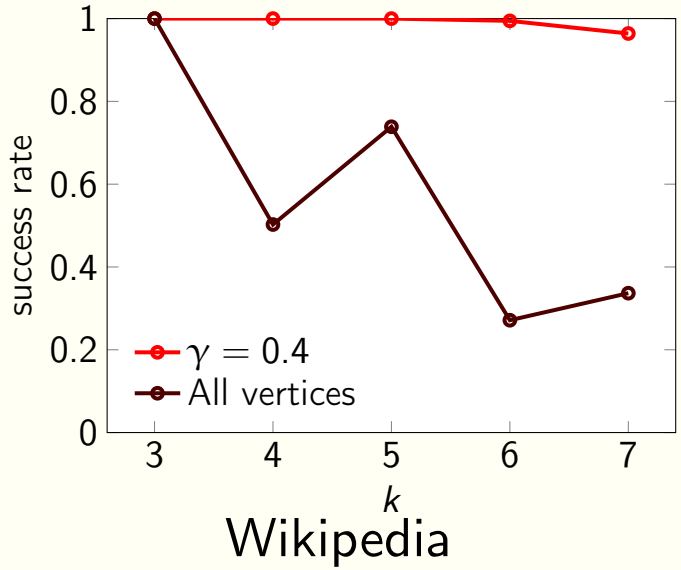
Internet



Baidu

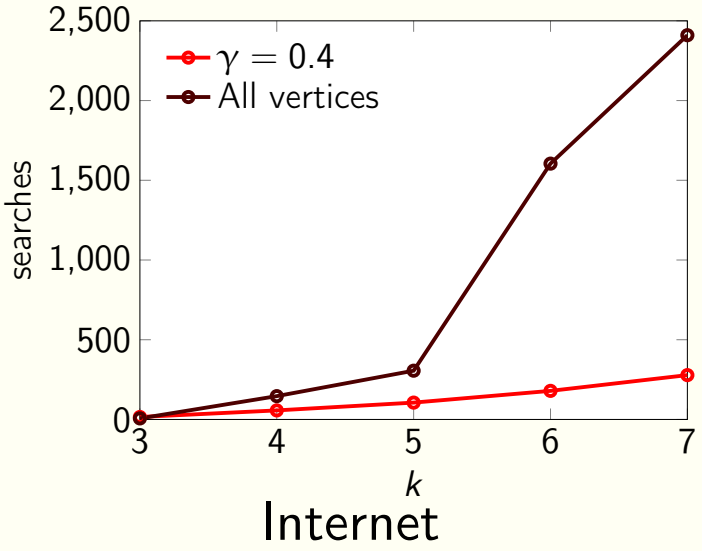


Gowalla

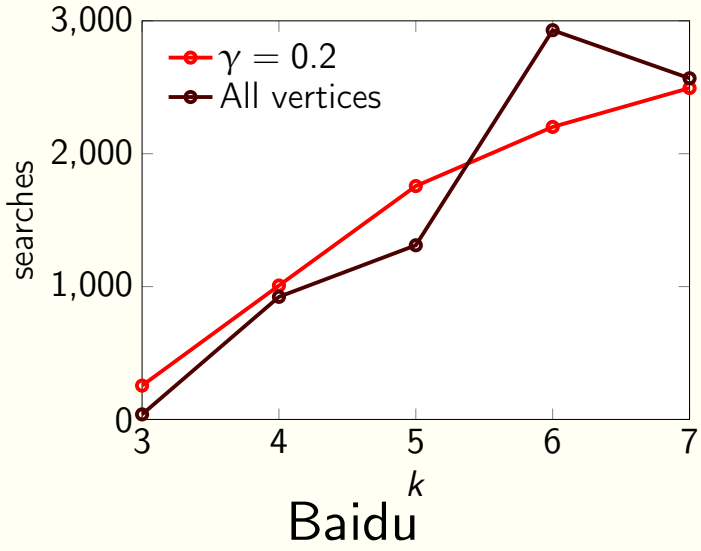


Wikipedia

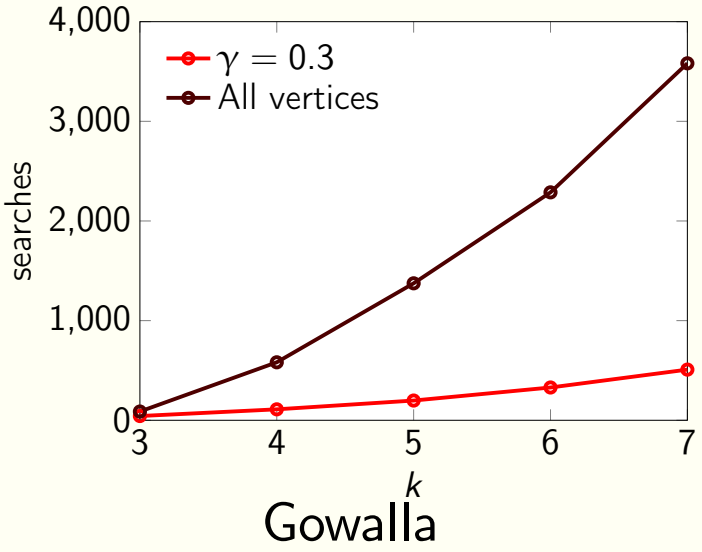
Real-world data: Searches



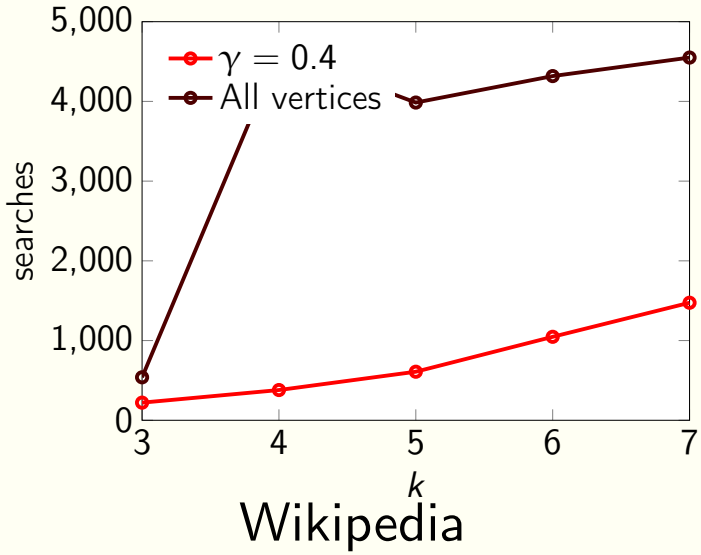
Internet



Baidu



Gowalla



Wikipedia

Conclusion

- By selecting $\sqrt{n}$ degrees, we can efficiently find induced subgraphs on inhomogeneous random graphs
- In real-world networks induced subgraphs occur on other degree magnitudes

Possible extensions

- What about other random graph models?
- How can we find the optimal induced subgraph scale on real-world data?