COMPLEX NETWORKS MEET SPARSITY SOLVING PROBLEMS DEFINABLE IN FIRST-ORDER LOGIC

Jan Dreier, Philipp Kuinke, Peter Rossmanith

May 25 2021

RWTH Aachen University, TU Wien

1

1

> MSO on treewidth



> MSO on treewidth

FO on sparse graphs





1



Network Science







Motivation













FO Model-Checking

[Grohe, Kreutzer, Sieberz 2011]

First-order formulas φ can be evaluated on sparse graphs in time $f(|\varphi|)n^{1+\varepsilon}.$

FO Model-Checking

[Grohe, Kreutzer, Sieberz 2011]

First-order formulas φ can be evaluated on sparse graphs in time $f(|\varphi|)n^{1+\varepsilon}.$

 \bigcirc dominating set of size k:

$$\exists x_1 \dots \exists x_k \, \forall y \, \bigvee_i y \sim x_i \lor y = x_i$$

FO Model-Checking

[Grohe, Kreutzer, Sieberz 2011]

First-order formulas φ can be evaluated on sparse graphs in time $f(|\varphi|)n^{1+\varepsilon}.$

 \bigcirc dominating set of size k:

$$\exists x_1 \dots \exists x_k \, \forall y \, \bigvee_i y \sim x_i \lor y = x_i$$

 \bigcirc independent set of size k:

$$\exists x_1 \dots \exists x_k \bigwedge_{i,j} x_i \not\sim x_j \land x_i \neq x_j$$

FO Model-Checking

[Grohe, Kreutzer, Sieberz 2011]

First-order formulas φ can be evaluated on sparse graphs in time $f(|\varphi|)n^{1+\varepsilon}.$

 \bigcirc dominating set of size k:

$$\exists x_1 \dots \exists x_k \, \forall y \, \bigvee_i y \sim x_i \lor y = x_i$$

 \bigcirc independent set of size k:

$$\exists x_1 \dots \exists x_k \bigwedge_{i,j} x_i \not\sim x_j \land x_i \neq x_j$$

best algorithms on general graphs: $n^{O(k)}$

FO Model-Checking

[Grohe, Kreutzer, Sieberz 2011]

First-order formulas φ can be evaluated on sparse graphs in time $f(|\varphi|)n^{1+\varepsilon}.$

 \bigcirc dominating set of size k:

$$\exists x_1 \dots \exists x_k \, \forall y \, \bigvee_i y \sim x_i \lor y = x_i$$

 \bigcirc independent set of size k:

$$\exists x_1 \dots \exists x_k \bigwedge_{i,j} x_i \not\sim x_j \land x_i \neq x_j$$

best algorithms on general graphs: $n^{O(k)}$

on sparse graphs: $f(k)n^{1+\varepsilon}$

First-Order Logic

First-Order Logic

can be expressed:

 \bigcirc dominating set of size k

- \bigcirc dominating set of size k
- \bigcirc independent set of size k

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k
- fundamentals of SQL database queries

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k
- fundamentals of SQL database queries

on general graphs: $n^{O(k)}$, on sparse graphs: $f(k)n^{1+\varepsilon}$

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k
- fundamentals of SQL database queries

on general graphs: $n^{O(k)}$, on sparse graphs: $f(k)n^{1+\varepsilon}$

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k
- fundamentals of SQL database queries

on general graphs: $n^{O(k)}$, on sparse graphs: $f(k)n^{1+\varepsilon}$

cannot be expressed:

o are two vertices in the same connected component?

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k
- fundamentals of SQL database queries

on general graphs: $n^{O(k)}$, on sparse graphs: $f(k)n^{1+\varepsilon}$

- o are two vertices in the same connected component?
- \bigcirc do two vertices have distance at most $\log(n)$?

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k
- fundamentals of SQL database queries

on general graphs: $n^{O(k)}$, on sparse graphs: $f(k)n^{1+\varepsilon}$

- o are two vertices in the same connected component?
- \bigcirc do two vertices have distance at most $\log(n)$?
- is something acyclic?

- \bigcirc dominating set of size k
- \bigcirc independent set of size k
- \bigcirc motif counting of size k
- fundamentals of SQL database queries

on general graphs: $n^{O(k)}$, on sparse graphs: $f(k)n^{1+\varepsilon}$

- o are two vertices in the same connected component?
- \bigcirc do two vertices have distance at most $\log(n)$?
- is something acyclic?

Sparse Graph Classes





 $\frac{|E|}{1} \le c$

for every graph in the graph class.



 $\frac{|E|}{2} \le 2$

for every graph in the graph class.









$$\frac{|E|}{|V|} \le f(r)$$

for every r-shallow
minor of every graph
in the graph class.
What Are Sparse Graphs?



$$\frac{|E|}{|V|} \leq f(r)$$
 for every r-shallow minor of every graph in the graph class.

real world: for small *r* (such as 1,2,3) sparse (w.r.t. gen. col. numbers) [Nadara, Pilipczuk, Rabinovich, Reidl, Siebertz 2019]







Some central properties:



Some central properties:

• Skewed degree distribution Fraction of vertices with degree kproportional to $k^{-\alpha}$ with $2 \le \alpha \le 3$?



Some central properties:

○ Skewed degree distribution Fraction of vertices with degree kproportional to $k^{-\alpha}$ with $2 \le \alpha \le 3$?

Clustered

If we have a common friend we are likely friends as well



Some central properties:

○ Skewed degree distribution Fraction of vertices with degree kproportional to $k^{-\alpha}$ with $2 \le \alpha \le 3$?

Clustered

If we have a common friend we are likely friends as well

Small-world property Everyone is close to everyone





































Sparsity of Random Graph Models



[Grohe 2001], [Farrell et. al. 2015], [Demaine et. al. 2019], [Dreier et. al. 2020]

RESULTS

[Dreier, Kuinke, Rossmanith 2020]

○ fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$

○ fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$

unclustered

real networks: typically clustered

 \bigcirc fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$

unclustered

real networks: typically clustered

Theorem

Given a first-order sentence φ and a graph G sampled from a 3-power-law-bounded model, one can decide whether φ is true on G in expected time $f(|\varphi|)n^{1+\varepsilon}$ for every $\varepsilon > 0$.

[Dreier, Kuinke, Rossmanith 2020]

A more direct way to get a desirable degree distribution.



14

A random graph model with vertices $1, \ldots, n$ is 3-power-law-bounded if the probability that some subset of edges $E \subseteq \binom{\{1,\ldots,n\}}{2}$ is present is at most

A random graph model with vertices $1, \ldots, n$ is 3-power-law-bounded if the probability that some subset of edges $E \subseteq {\binom{\{1,\ldots,n\}}{2}}$ is present is at most

$$\prod_{ij\in E} \frac{1}{\sqrt{i\cdot j}}$$

A random graph model with vertices $1, \ldots, n$ is 3-power-law-bounded if the probability that some subset of edges $E \subseteq \binom{\{1,\ldots,n\}}{2}$ is present is at most

$$\log(n)^{O(|E|^2)} \prod_{ij \in E} \frac{1}{\sqrt{i \cdot j}}.$$

A random graph model with vertices $1, \ldots, n$ is 3-power-law-bounded if the probability that some subset of edges $E \subseteq \binom{\{1,\ldots,n\}}{2}$ is present is at most

$$\log(n)^{O(|E|^2)} \prod_{ij \in E} \frac{1}{\sqrt{i \cdot j}}.$$

- Preferential attachment model
- O Chung–Lu model
- O Erdös–Rényi model
- Configuration model

X

- Hyperbolic random graph model
- random intersection model
- Watts–Strogatz model
- Kleinberg model

Asymptotic Structure of 3-power-law-bounded models

Asymptotic Structure of 3-power-law-bounded models



Asymptotic Structure of 3-power-law-bounded models


Asymptotic Structure of 3-power-law-bounded models



Asymptotic Structure of 3-power-law-bounded models



Asymptotic Structure of 3-power-law-bounded models



Input: graph sampled from 3-power-law-bounded model



Algorithm





Algorithm



prune trees





prune protrusions





use brute force on core



repeat for every neighborhood



Done!

Practical Considerations I



○ our run time:

Practical Considerations II

○ our run time:



○ under worst-case complexity this is optimal



- under worst-case complexity this is optimal
- open question: can we do better in the average-case?

A random graph model is *3-power-law-bounded* if (roughly speaking):

A random graph model is *3-power-law-bounded* if (roughly speaking):

○ fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$

A random graph model is *3-power-law-bounded* if (roughly speaking):

- fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$
- unclustered

real networks: typically clustered

A random graph model is *3-power-law-bounded* if (roughly speaking):

- fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$
- unclustered

real networks: typically clustered

Theorem

Given a first-order sentence φ and a graph G sampled from a 3-power-law-bounded model, one can decide whether φ is true on G in expected time $f(|\varphi|)n^{1+\varepsilon}$ for every $\varepsilon > 0$.

A random graph model is *3-power-law-bounded* if (roughly speaking):

○ fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$



unclustered

real networks: typically clustered

Theorem

Given a first-order sentence φ and a graph G sampled from a 3-power-law-bounded model, one can decide whether φ is true on G in expected time $f(|\varphi|)n^{1+\varepsilon}$ for every $\varepsilon > 0$.

A random graph model is *3-power-law-bounded* if (roughly speaking):

- \bigcirc fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$
- unclustered

real networks: typically clustered

Theorem

Given a first-order sentence φ and a graph G sampled from a 3-power-law-bounded model, one can decide whether φ is true on G in expected time $f(|\varphi|)n^{1+\varepsilon}$ for every $\varepsilon > 0$.



optimal

A random graph model is *3-power-law-bounded* if (roughly speaking):

- \bigcirc fraction of vertices with degree k is $O(k^{-3})$ real networks: typically $k^{-\alpha}$ with $2 \le \alpha \le 3$
- unclustered

real networks: typically clustered

Theorem

Given a first-order sentence φ and a graph G sampled from a 3-power-law-bounded model, one can decide whether φ is true on G in expected time $f(|\varphi|)n^{1+\varepsilon}$ for every $\varepsilon > 0$.

Big Question: model-checking on clustered models?





Can we solve the problem on Hyperbolic Random Graphs?



LOWER BOUNDS

○ worst-case: P and NP

- worst-case: P and NP
- average-case: avgP and (NP, PComp)?

- worst-case: P and NP
- average-case: avgP and (NP, PComp)?
- expected time example:

$$\sum_{x \in \mathbf{N}} \mathbf{P}[X = x] \cdot \mathsf{time}(x) = O(1)$$









[Dreier, Rossmanith 2019]



Without Colors


Without Colors



Zero-One Laws

1/2 random graph



Zero-One Laws



Zero-One Laws







• W[1]-complete:

 $\exists x_1 \dots \exists x_k$

• W[2]-complete:

 $\exists x_1 \dots \exists x_k \forall y$

• W[3]-complete:

 $\exists x_1 \dots \exists x_k \forall y \exists z$

• W[4]-complete:

 $\exists x_1 \dots \exists x_k \forall y \exists z \forall w$

• W[4]-complete:

 $\exists x_1 \dots \exists x_k \forall y \exists z \forall w$

average-case on 1/2 random graph

• W[4]-complete:

 $\exists x_1 \dots \exists x_k \forall y \exists z \forall w$

\bigcirc average-case on 1/2 random graph \circ FPT (i.e., $f(|\varphi|)n$ time):

 $\exists x_1 \ldots \exists x_k$

• W[4]-complete:

$$\exists x_1 \dots \exists x_k \forall y \exists z \forall w$$

o average-case on 1/2 random graph o FPT (i.e., f(|φ|)n time):

 $\exists x_1 \ldots \exists x_k$

• equally hard:

 $\exists x_1 \dots \exists x_k \forall y$

• W[4]-complete:

$$\exists x_1 \dots \exists x_k \forall y \exists z \forall w$$

o average-case on 1/2 random graph o FPT (i.e., f(|φ|)n time):

 $\exists x_1 \ldots \exists x_k$

• equally hard:

 $\exists x_1 \dots \exists x_k \forall y \exists z$

• W[4]-complete:

$$\exists x_1 \dots \exists x_k \forall y \exists z \forall w$$

○ average-case on 1/2 random graph ○ FPT (i.e., f(|φ|)n time):

 $\exists x_1 \ldots \exists x_k$

• equally hard:

 $\exists x_1 \dots \exists x_k \forall y \exists z \forall w$

• W[4]-complete:

$$\exists x_1 \dots \exists x_k \forall y \exists z \forall w$$

o average-case on 1/2 random graph o FPT (i.e., f(|φ|)n time):

 $\exists x_1 \ldots \exists x_k$

• equally hard:

 $\exists x_1 \dots \exists x_k \forall y \exists z \forall w \exists u$

• W[4]-complete:

$$\exists x_1 \dots \exists x_k \forall y \exists z \forall w$$

○ average-case on 1/2 random graph ○ FPT (i.e., f(|φ|)n time):

 $\exists x_1 \ldots \exists x_k$

• equally hard:

 $\exists x_1 \dots \exists x_k \forall y \exists z \forall w \exists u \dots$

 Solve all problems definable in FO logic on unclustered graphs that are not too skewed.

- Solve all problems definable in FO logic on unclustered graphs that are not too skewed.
- O How about clustered models?

- Solve all problems definable in FO logic on unclustered graphs that are not too skewed.
- How about clustered models?
- O How about about (uncolored) lower bounds?

Thanks!

dreier@ac.tuwien.ac.at