On the impact of small-world on local search

Andrea Roli

andrea.roli@unibo.it

DEIS

Università degli Studi di Bologna

Campus of Cesena

Motivation

The impact of *structure* – whatever it is – on search algorithms is dramatically relevant

- Identify most difficult instances (for a given technique)
- Understand *why* an instance is difficult
- Exploit this bit of information to choose the best solver, or a combination of solvers
- Evaluate the quality of benchmarks

Goal

Previous work [Walsh, 1999]

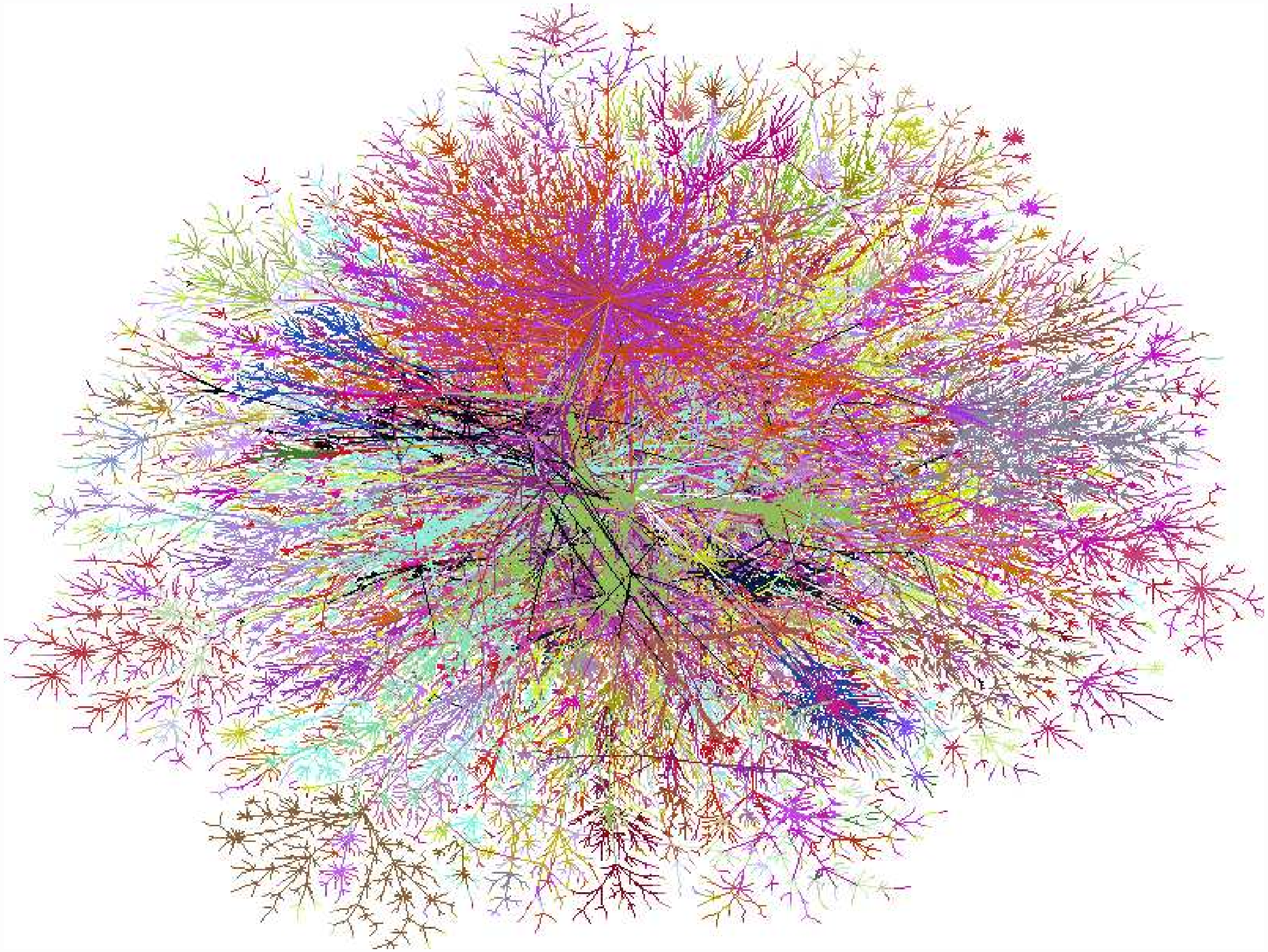
- CSP instances defined over ‘small-world’ graphs are harder to solve for complete algorithms

Question: What about local search behavior on small-world instances?

Outline

- Background: Complex networks
- *Structure* in CSPs
- Small-world SAT instances
- Experimental results
- Discussion

Complex networks



Complex networks

- System topology is crucial for understanding its dynamics
- Graph theory provides useful models
- Complex networks: emerging research field

Graphs as structure abstraction

- Entities represented as graph nodes
- relations $\leftrightarrow$ arcs
- Node: either one entity or an entire subsystem

Main characteristics

- Node degree (distribution, average, etc.)
- Diameter, characteristic path length et similia
- Clustering (i.e., cliquishness tendency)

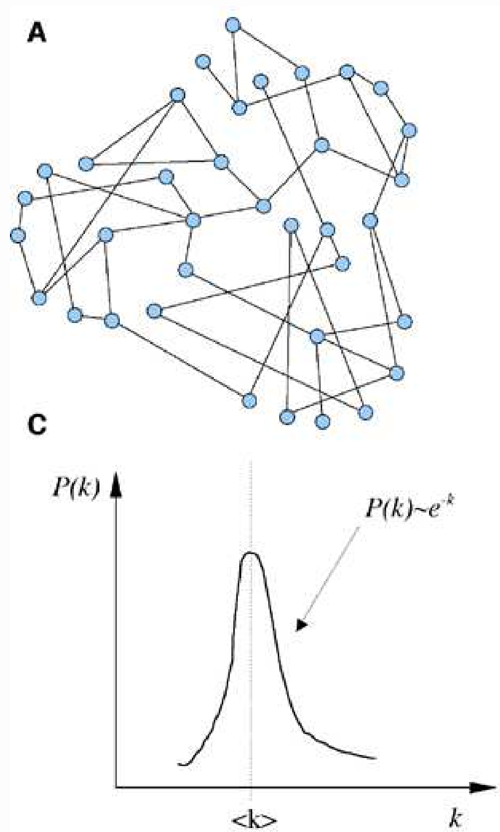
Random graphs

- First developed model for system structure
- Several important applications
- Random graphs fail to represent social and biological systems

Random graphs

- Node degree distribution: Poissonian (approx Normal)
- Characteristic path length: low
- Clustering: low

Random graphs



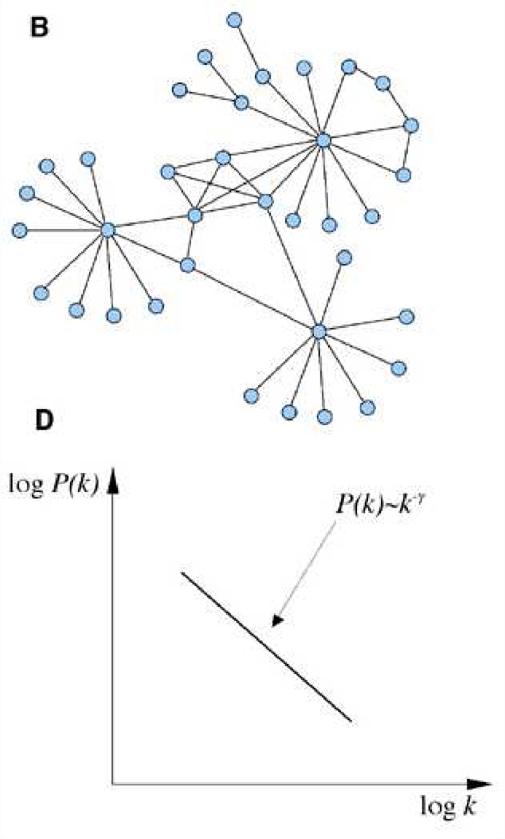
Scale-free networks

- Relations among individuals in a society (e.g., scientific collaborations)
- Web pages structure
- Internet structure
- . . .

Scale-free networks

- Node degree distribution:
  nodes with degree $k \sim k^{-\gamma}$ (γ parameter)
- Very few *hubs* (but not negligible) and many nodes with few connections
- Robust wrt random failures
- Sensitive to attacks

Scale-free networks



Scale-free networks formation

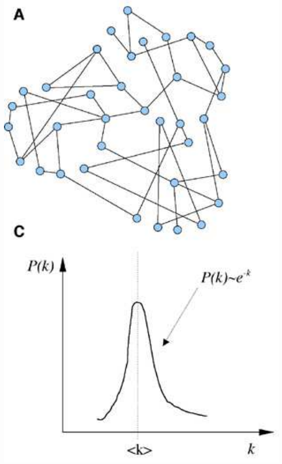
- *Growth*: older nodes has on average a higher number of connections
- *Preferential attachment*: new nodes are more likely to connect to nodes with higher degree (probability proportional to the degree)
- Model variants that take into account also the *fitness*

Small-world

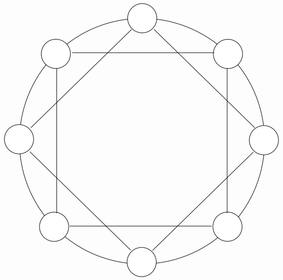
- Any pair of nodes connected by few hops (short characteristic path length)
- High degree of cliquishness (high clustering coefficient)
- Examples:
  - Social networks
  - World Wide Web
  - Scientific collaboration network
  - *C.Elegans* worm neural network

Characteristic length

Informally: average path length between any pair of nodes.

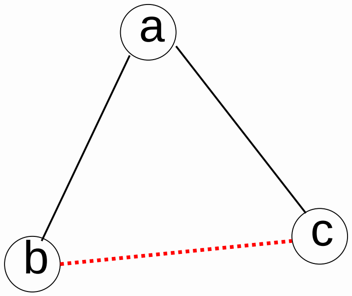


Random graphs → short

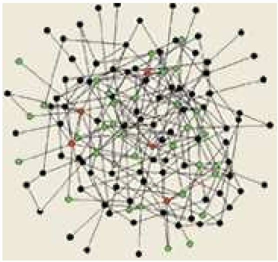


Grid graphs → long

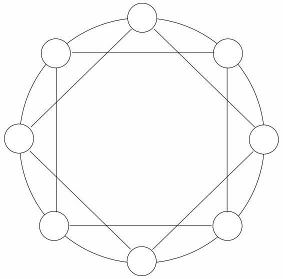
Clustering



Informally: it quantifies the probability that, given node a connected to b and c, there is an edge between b and c



Random graphs $\longrightarrow$ low



Grid graphs $\longrightarrow$ high

Structure

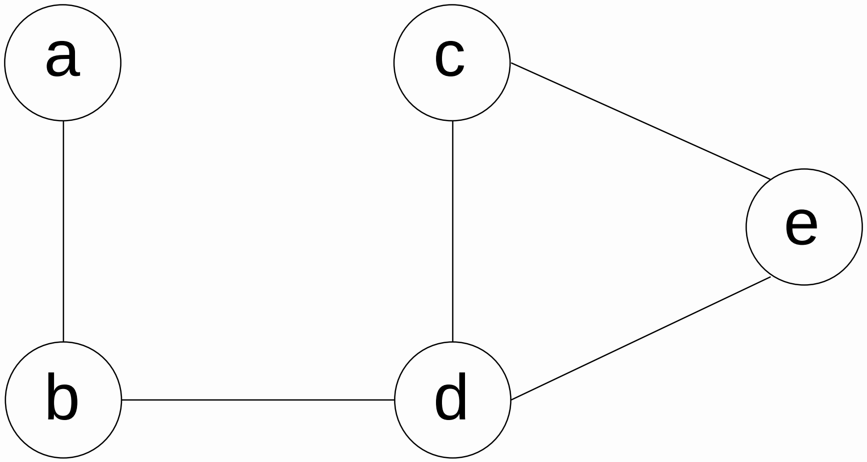
- Diverse meanings
- *Structure* vs. *random*
- Usually *real world* problems are said to be structured
- Attempts to define quantitative measures (entropy, compression ratio, etc.)

► Graph representation of relations among problem entities

SATgraphs

$$(a \vee \neg b) \wedge (b \vee d) \wedge (c \vee \neg d \vee \neg e)$$

$$\Downarrow$$



Remember the initial goal.

Previous work [Walsh, 1999]

- CSP instances defined over ‘small-world’ graphs are harder to solve for complete algorithms

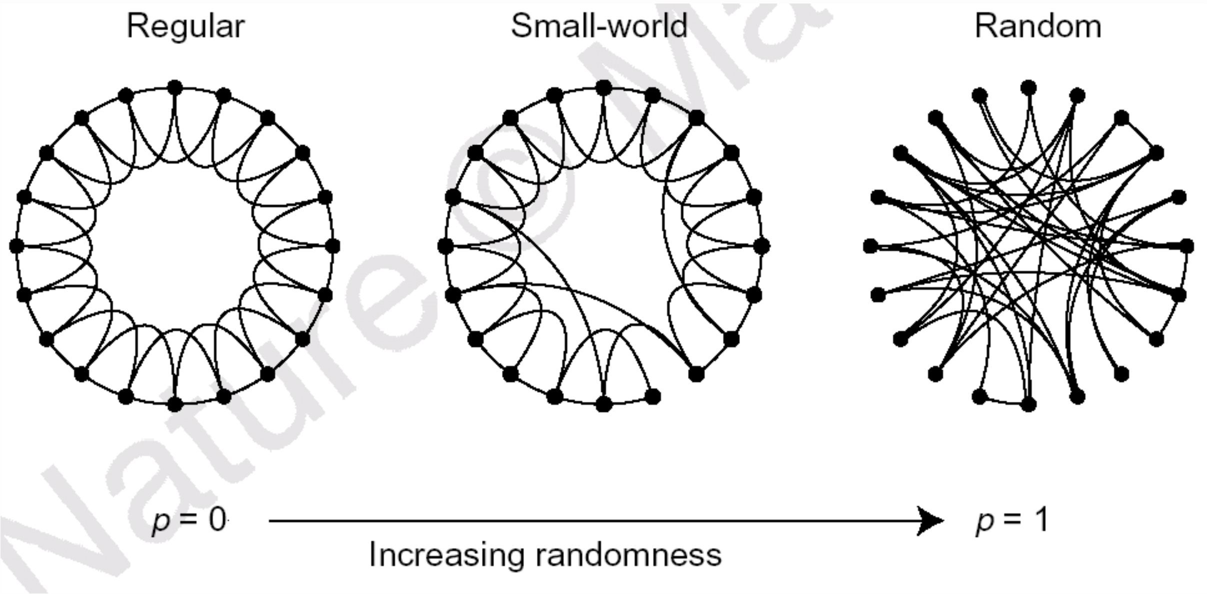
Question: What about local search behavior on small-world instances?

Experimental issues

- Small-world SAT instances
  - Procedure to generate instances
  - Measuring ‘small-world’ property
- Attacking the benchmark with local search algorithms
  - GSAT
  - WalkSAT
  - ILS-SAT

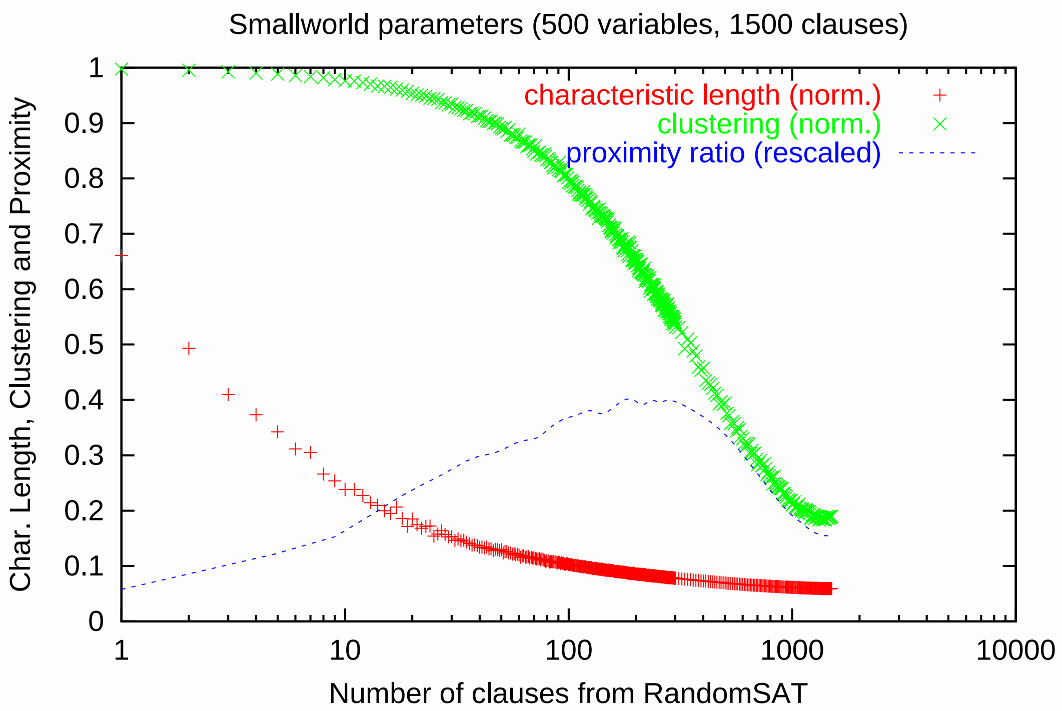
Small-world SAT

► Morphing between a lattice SAT instance and a random SAT instance. [Gent et al., 1999]

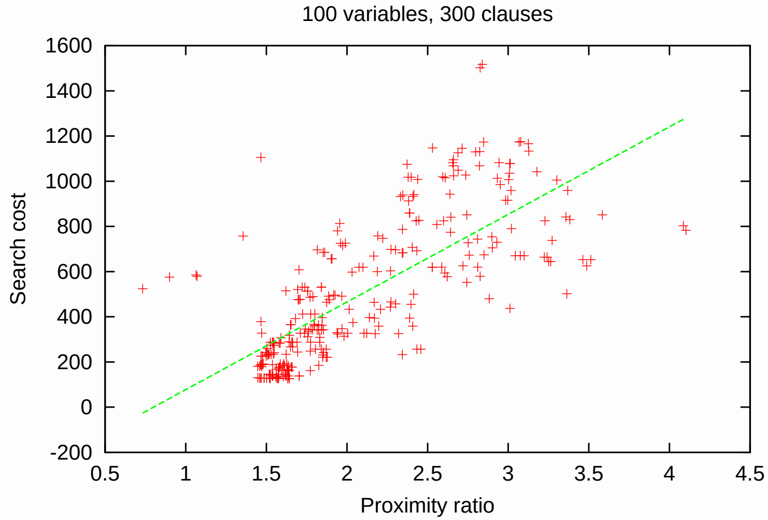


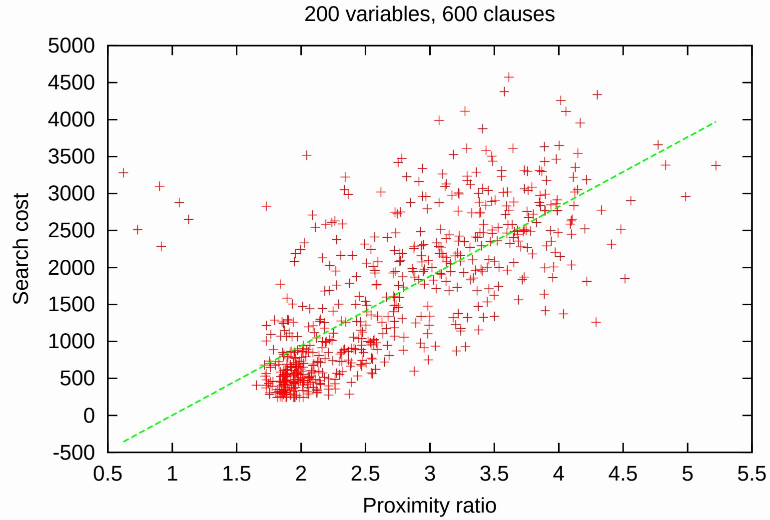
Small-world SAT

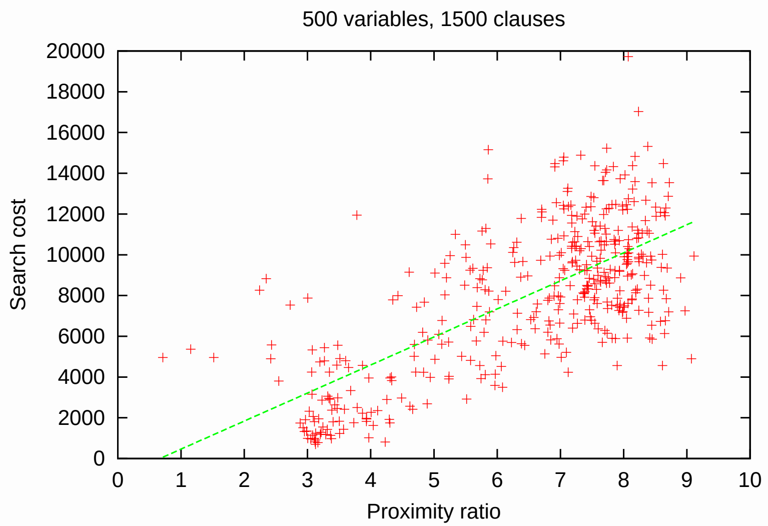
Length, clustering and proximity ratio (normalized ratio clustering/length)

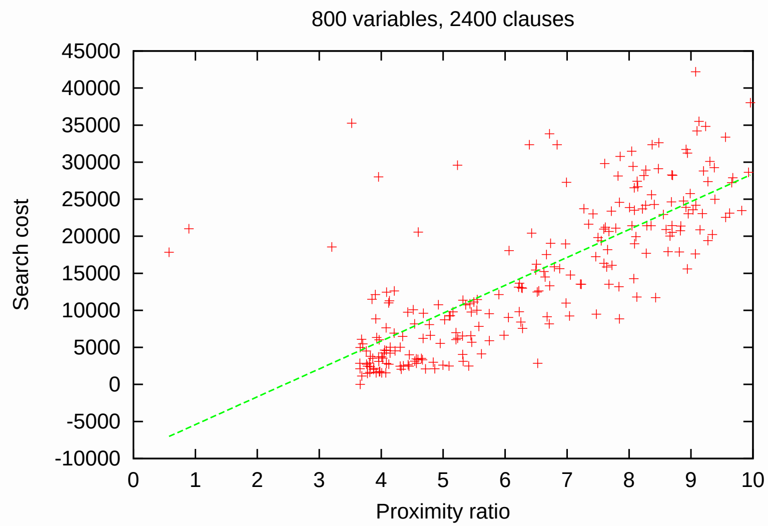


Complete algorithm





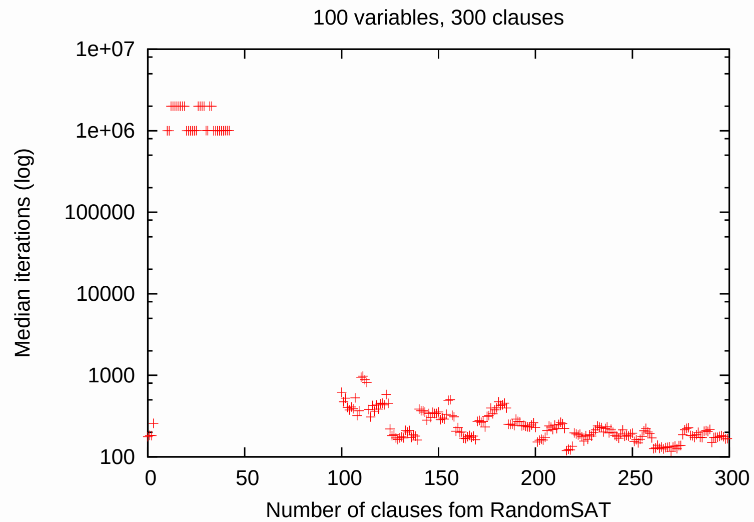


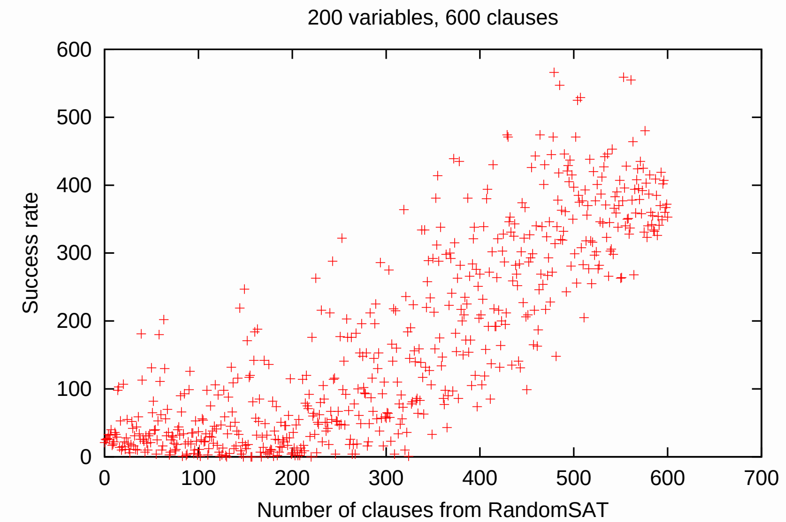


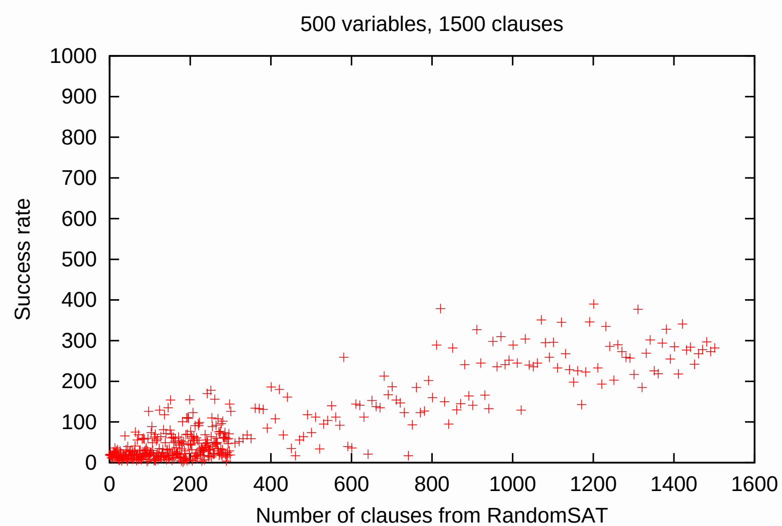
Outline of the results

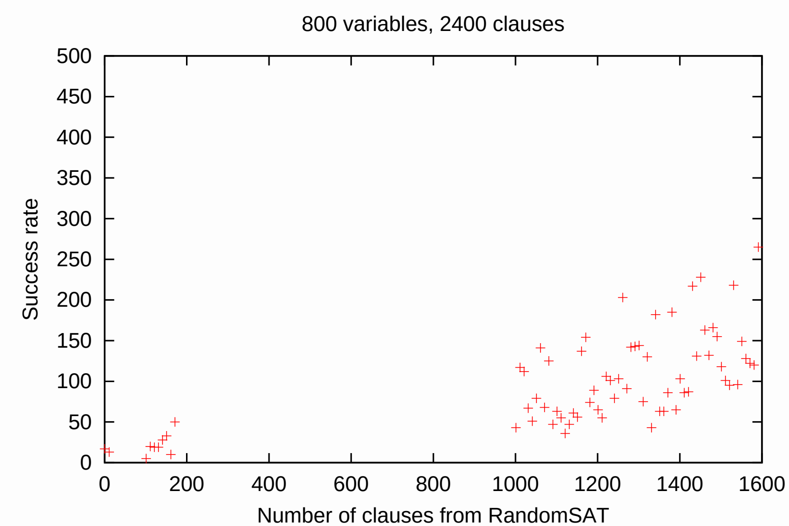
- No common behavior across different algorithms
- ‘Mild’ tendency of small-world and hardness correlation

GSAT

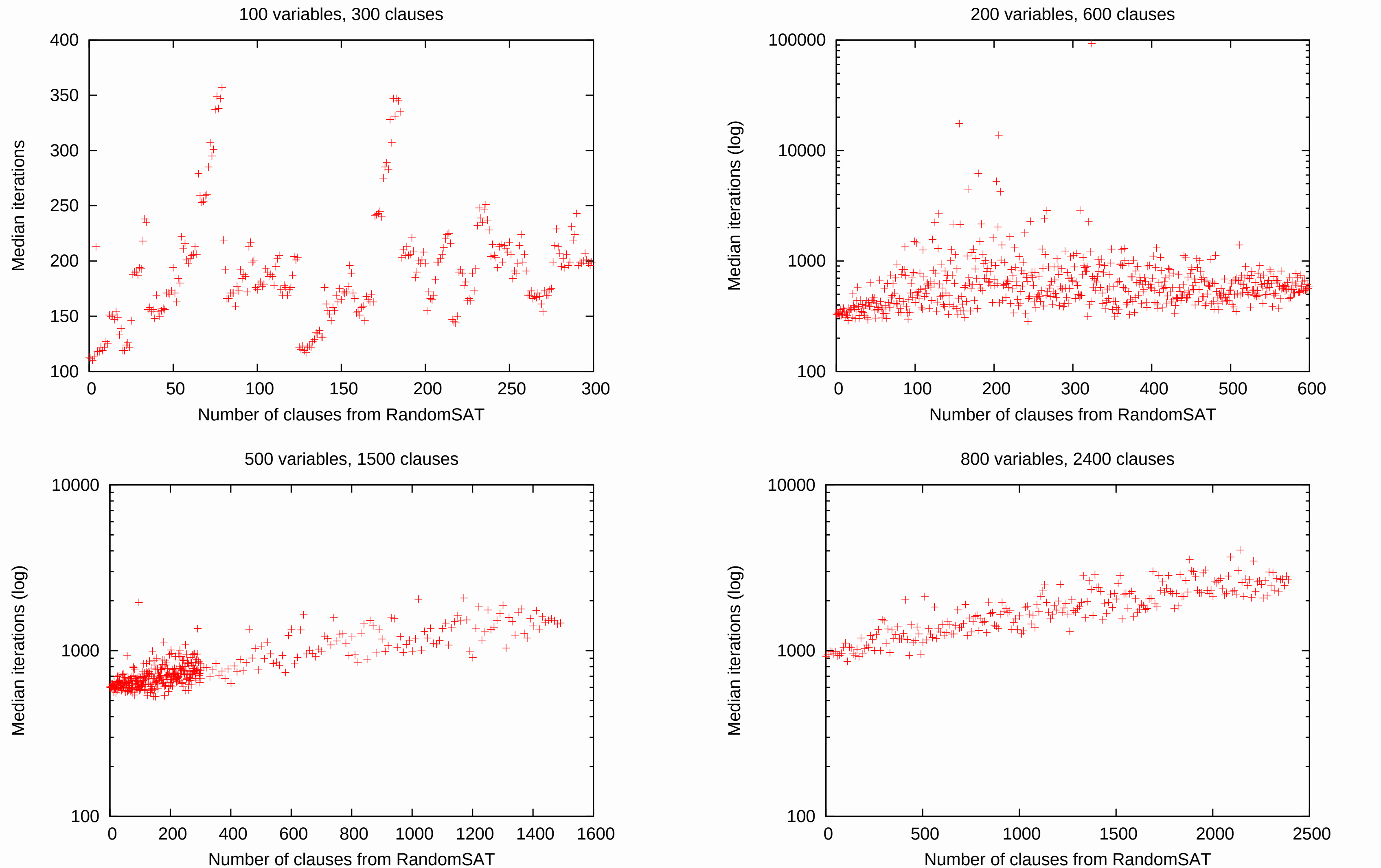

100 variables, 300 clauses
Median iterations (log)
1e+07
1e+06
100000
10000
1000
100
0
50
100
150
200
250
300
Number of clauses fom RandomSAT


200 variables, 600 clauses
Success rate
600
500
400
300
200
100
0
0
100
200
300
400
500
600
700
Number of clauses from RandomSAT

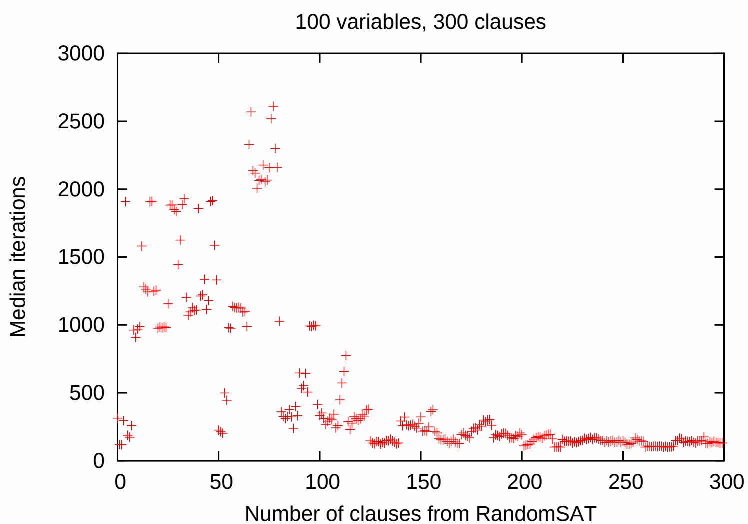

500 variables, 1500 clauses
Success rate
1000
900
800
700
600
500
400
300
200
100
0
0
200
400
600
800
1000
1200
1400
1600
Number of clauses from RandomSAT

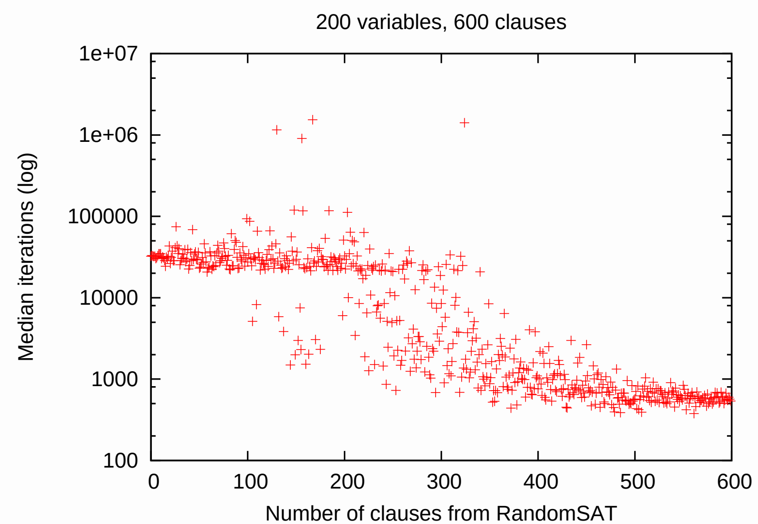

800 variables, 2400 clauses
Success rate
500
450
400
350
300
250
200
150
100
50
0
0
200
400
600
800
1000
1200
1400
1600
Number of clauses from RandomSAT

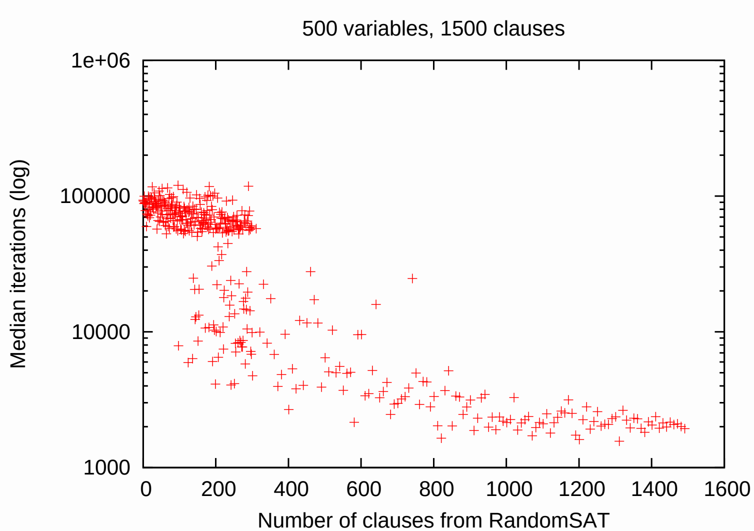
WalkSAT

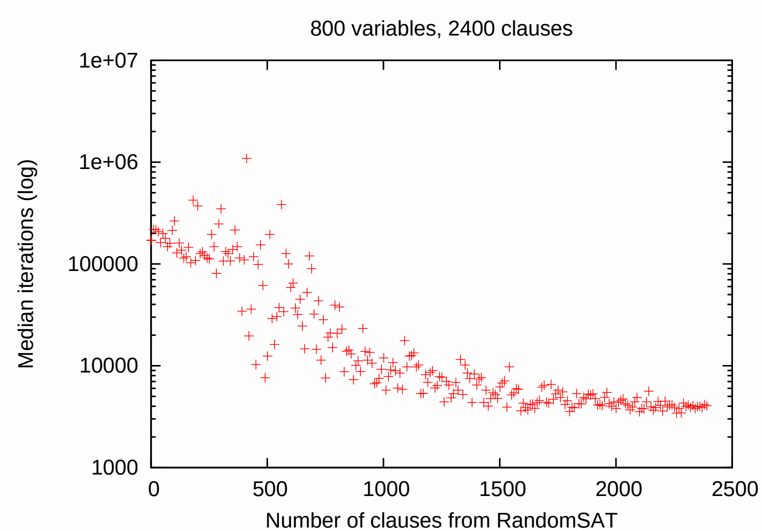

100 variables, 300 clauses
Median iterations
400
350
300
250
200
150
100
0
50
100
150
200
250
300
Number of clauses from RandomSAT
200 variables, 600 clauses
Median iterations (log)
100000
10000
1000
100
0
100
200
300
400
500
600
Number of clauses from RandomSAT
500 variables, 1500 clauses
Median iterations (log)
10000
1000
100
0
200
400
600
800
1000
1200
1400
1600
Number of clauses from RandomSAT
800 variables, 2400 clauses
Median iterations (log)
10000
1000
100
0
500
1000
1500
2000
2500
Number of clauses from RandomSAT

ILS-SAT


100 variables, 300 clauses
Median iterations
3000
2500
2000
1500
1000
500
0
0
50
100
150
200
250
300
Number of clauses from RandomSAT


200 variables, 600 clauses
Median iterations (log)
1e+07
1e+06
100000
10000
1000
100
0
100
200
300
400
500
600
Number of clauses from RandomSAT


500 variables, 1500 clauses
Median iterations (log)
1e+06
100000
10000
1000
0
200
400
600
800
1000
1200
1400
1600
Number of clauses from RandomSAT


800 variables, 2400 clauses
Median iterations (log)
1e+07
1e+06
100000
10000
1000
0
500
1000
1500
2000
2500
Number of clauses from RandomSAT

Discussion

- Many small-world/lattice SAT instances are harder for GSAT and ILS-SAT
- WalkSAT exhibits a peculiar behavior

► The relation between SATgraph and search landscape plays a very important role

Future work

- **Connections between constraint graph properties and search space characteristics**
- Exploring strengths and weaknesses of the heuristics w.r.t. constraint graph properties
- Relation between problem encoding and graph properties
- Alternative formulations to study the structure of a problem can be used (e.g., weighted graphs)