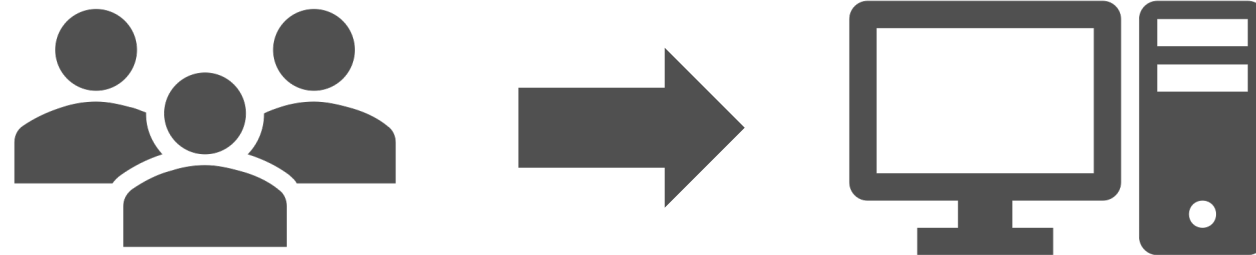


Jukka Suomela
Aalto University

**Can we
automate
our own work**
— or show that it is hard?

Computer science: ***what
can be automated?***



Computer science: ***what
can be automated?***

Today: ***can we automate
our own work?***

Focus: *theory of
distributed computing*

Consider a typical theory paper in e.g. PODC or DISC...

Abstract

The degree splitting problem requires coloring the edges of a graph so that each node has almost the same number of edges in each color. A directed variant of the problem requires orienting the edges so that each node has the same number of incoming and outgoing edges, again uniformly.

We present deterministic distributed algorithms for both problems that are faster than their counterparts presented by Ghaffari and Su [SODA'17] and faster, and have a much smaller discrepancy. This improves upon the best deterministic algorithm for $(2 + o(1))\Delta$ -edge-coloring, improving the discrepancy to $(2 + o(1))\Delta$.

1998 ACM Subject Classification C.2.2 Network Protocols

Keywords and phrases Distributed Graph Algorithms, Discrepancy

Digital Object Identifier 10.4230/LIPIcs.DISC.2017.1

1 Introduction and Related Work

In this work, we present improved distributed (LOCAL) algorithms for the *degree splitting problem*, and also use them to provide simple algorithms for the classic and well-studied problem of *edge coloring*.

LOCAL Model. In the standard LOCAL model of distributed computation, the input is abstracted as an n -node undirected graph $G = (V, E)$ where each node $v \in V$ has a unique $O(\log n)$ -bit identifier. Communication happens between nodes in the graph.

Consider a typical theory paper in
e.g. PODC or DISC...

***How much of the work is
done with computers?***

Abstract

The degree splitting problem requires coloring the edges of a graph so that each node has almost the same number of edges in each color. A directed variant of the problem requires orienting the edges so that each node has the same number of incoming and outgoing edges, again up to a small additive error.

We present deterministic distributed algorithms for both problems that are simpler and faster than their counterparts presented by Ghaffari and Su [SODA'17], and have a much smaller discrepancy. This is the first deterministic algorithm for $(2 + o(1))\Delta$ -edge-coloring, improving on the previous algorithm of [SODA'17].

1998 ACM Subject Classification C.2.2 Network Protocols

Keywords and phrases Distributed Graph Algorithms, Discrepancy

Digital Object Identifier 10.4230/LIPIcs.DISC.2017.1

1 Introduction and Related Work

In this work, we present improved distributed (LOCAL) algorithms for the *degree splitting problem*, and also use them to provide simpler algorithms for the classic and well-studied problem of *edge coloring*.

LOCAL Model. In the standard LOCAL model of distributed computation, a network is abstracted as an n -node undirected graph $G = (V, E)$ where each node $v \in V$ has a unique $O(\log n)$ -bit identifier. Communication happens between adjacent nodes.

Consider a typical theory paper in
e.g. PODC or DISC...

*How much of the work is
done with computers?*

*How much of it could be
done with computers?*

Abstract

The degree splitting problem requires coloring the edges of a graph so that each node has almost the same number of edges in each color. A directed variant of the problem requires orienting the edges so that each node has the same number of incoming and outgoing edges, again up to a small additive error.

We present deterministic distributed algorithms for both problems, which are simpler and faster than their counterparts presented by Ghaffari and Su [SODA'17], and have a much smaller discrepancy. This is the first deterministic algorithm for $(2 + o(1))\Delta$ -edge-coloring, improving on the previous bound of 3Δ .

1998 ACM Subject Classification C.2.2 Network Protocols

Keywords and phrases Distributed Graph Algorithms, Discrepancy

Digital Object Identifier 10.4230/LIPIcs.DISC.2017.1

1 Introduction and Related Work

In this work, we present improved distributed (LOCAL) algorithms for the *degree splitting problem*, and also use them to provide simple algorithms for the classic and well-studied problem of *edge coloring*.

LOCAL Model. In the standard LOCAL model of distributed computation, a network is abstracted as an n -node undirected graph $G = (V, E)$ where each node $v \in V$ has a unique $O(\log n)$ -bit identifier. Communication happens between adjacent nodes.

Standard process

- **Question:** is there an efficient distributed algorithm for solving task X in model M ?

Standard process

- **Question:** is there an efficient distributed algorithm for solving task X in model M ?
- **Approach:** find smart people, spend lots of time in front of a whiteboard ...

Standard process

- **Question:** is there an efficient distributed algorithm for solving task X in model M ?
- **Approach:** find smart people, spend lots of time in front of a whiteboard ...
- **End result:** algorithm, algorithm analysis, proof of correctness, lower bound proof ...

Standard process

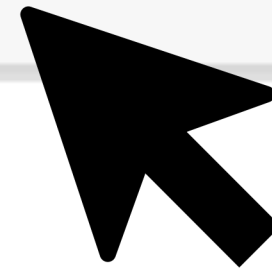
- **Question:** is there an efficient distributed algorithm for solving task X in model M ?
- **Approach:** find smart people, spend lots of time in front of a whiteboard ...
- **End result:** algorithm, algorithm analysis, proof of correctness, lower bound proof ...



Function V2

Automatic Lower Bound

Automatic Upper Bound



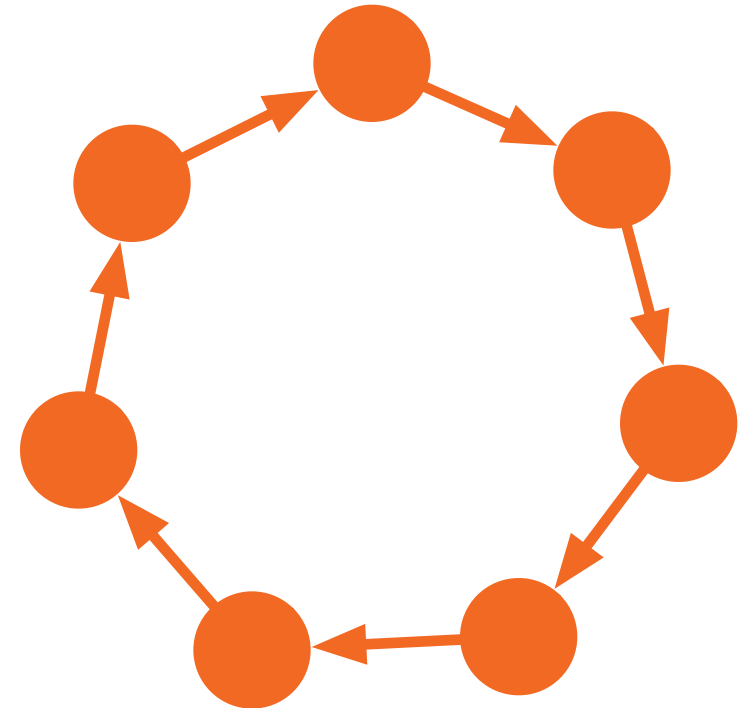
Lost sanity?

Toy example:

**Locally checkable
problems in cycles**

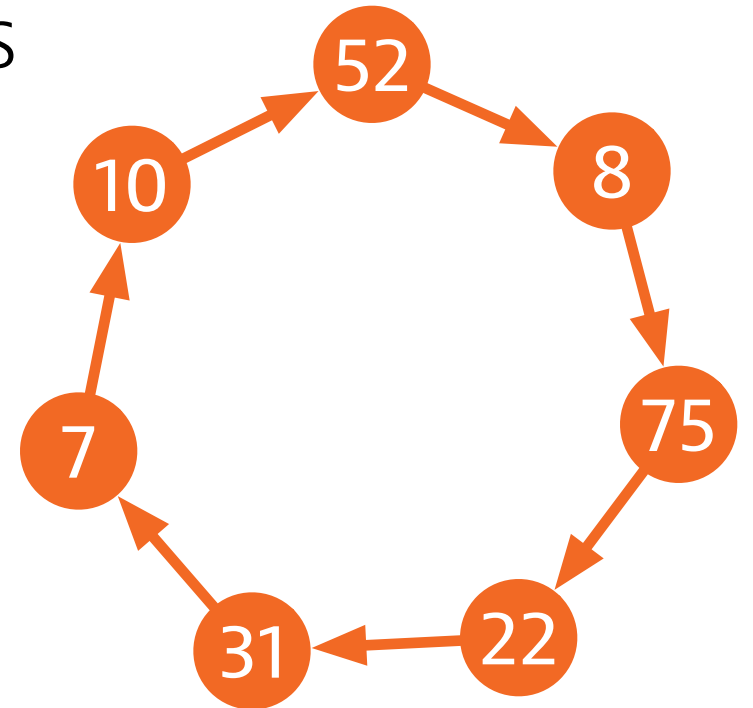
Setting

- **Computer network: cycle** of n computers
 - globally consistent orientation
 - each node has one "successor" and one "predecessor"



Setting

- **Computer network: cycle** of n computers
- **Model of computing: LOCAL model**
 - synchronous communication rounds
 - time = number of rounds until all nodes stop
 - unbounded message size
 - unlimited local computation
 - unique identifiers



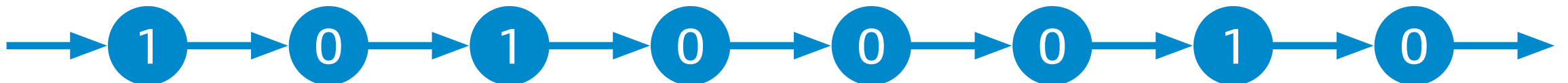
Setting

- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**
 - finite number of output labels
 - relation that tells which label sequences are valid

Setting

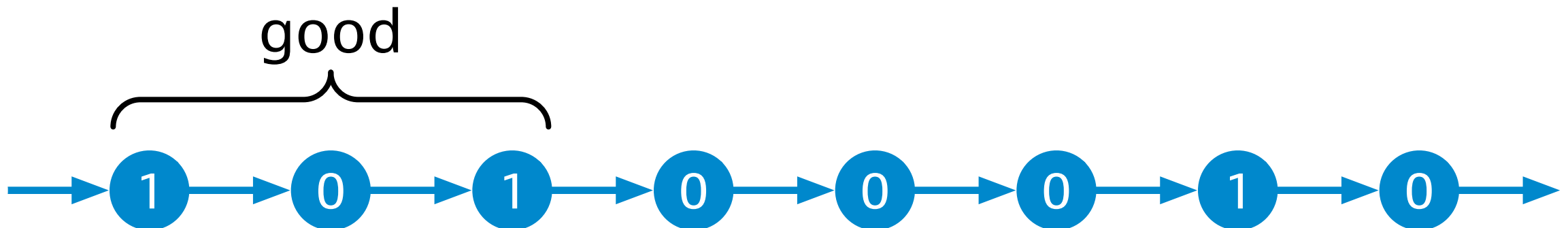
- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**

Example: maximal independent set



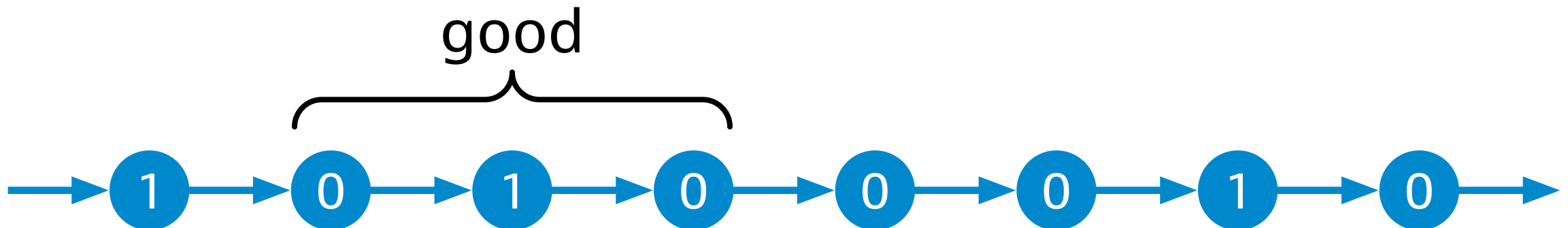
Setting

- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**



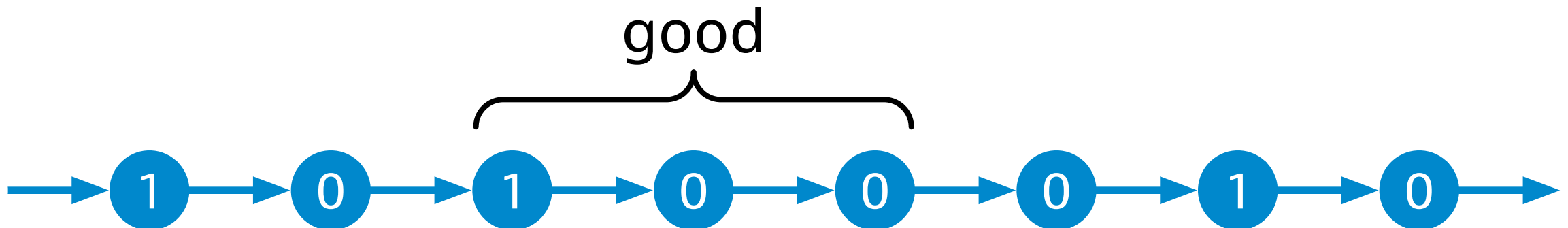
Setting

- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**



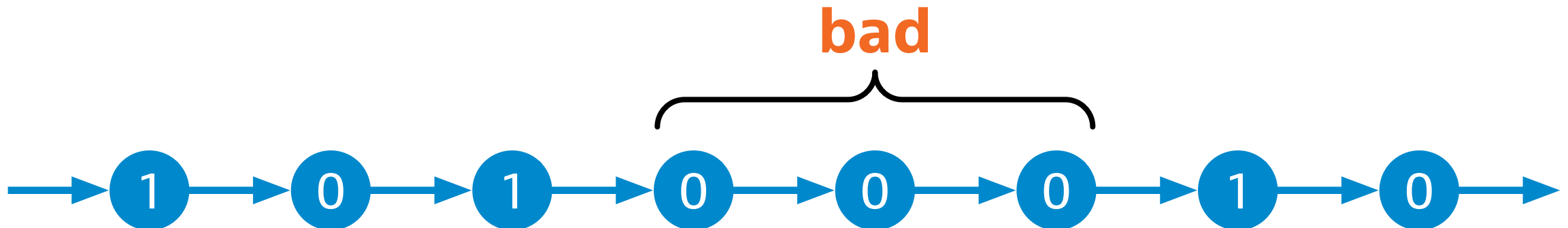
Setting

- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**



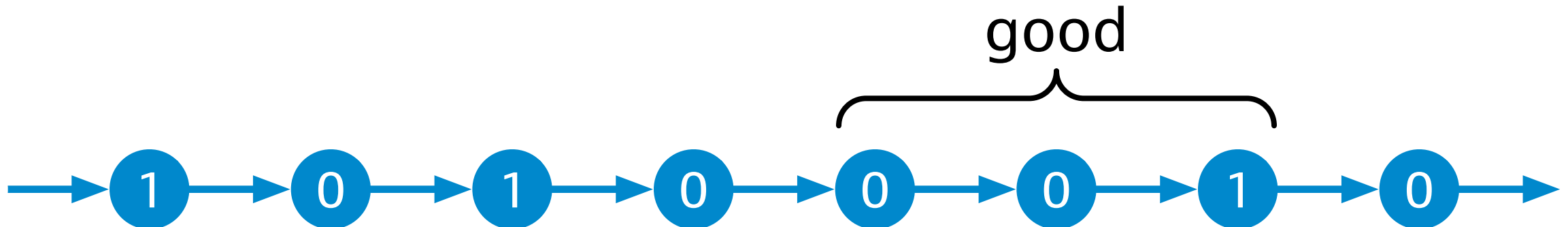
Setting

- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**



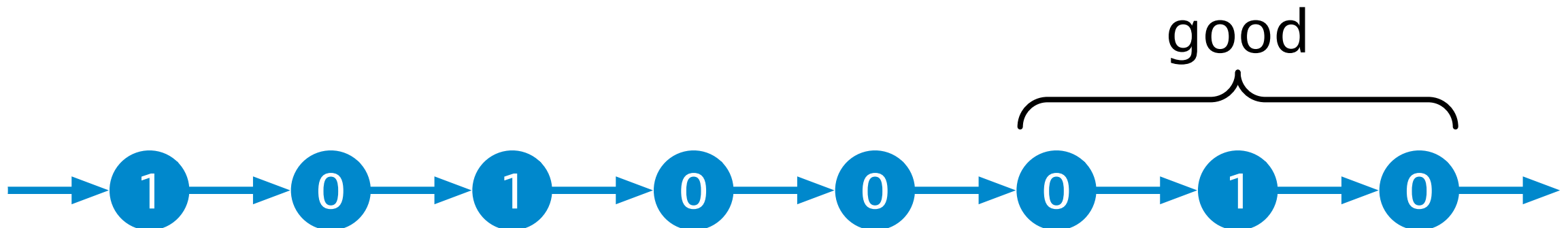
Setting

- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**



Setting

- *Computer network:* **cycle** of n computers
- *Model of computing:* **LOCAL model**
- *Problem:* any discrete problem you can define with **local constraints**



Valid label sequences

- *2-coloring*: **12, 21**
- *3-coloring*: **12, 21, 13, 31, 23, 32**
- *Independent set*: **01, 10, 00**
- *Maximal independent set*: **001, 010, 100, 101**
- *Distance-2 coloring with 3 colors*:
123, 132, 213, 231, 312, 321

Valid label sequences

- *2-coloring*: **12, 21**
- *3-coloring*: **12, 21, 13, 31, 23, 32**
- *Independent set*: **01, 10, 00**
- *Maximal independent set*: **001, 010, 100, 101**
- *Distance-2 coloring with 3 colors*:
123, 132, 213, 231, 312, 321

All possible
output labelings
in a window
of size k

Fully automatic

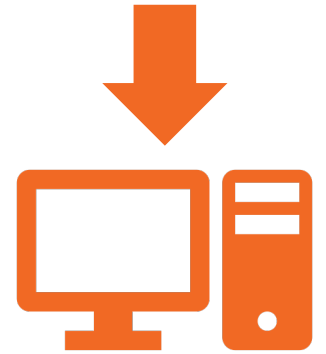
$$X = \{ 001, 010, 100, 101 \}$$

- Write down the specification of *any locally checkable problem* X

Fully automatic

- Write down the specification of *any locally checkable problem X*
- Then you can *find efficiently*
 - distributed round complexity of X
 - asymptotically optimal distributed algorithm for X

$X = \{ 001, 010, 100, 101 \}$



This algorithm solves X in time $O(\log^* n)$

Fully automatic

- Write down the specification of *any locally checkable problem X*
- Then you can *find efficiently*
 - distributed round complexity of X
 - asymptotically optimal distributed algorithm for X

**Polynomial time
(in the size
of problem
description)**

1 0 0

0 1 0

0 0 1

1 0 1

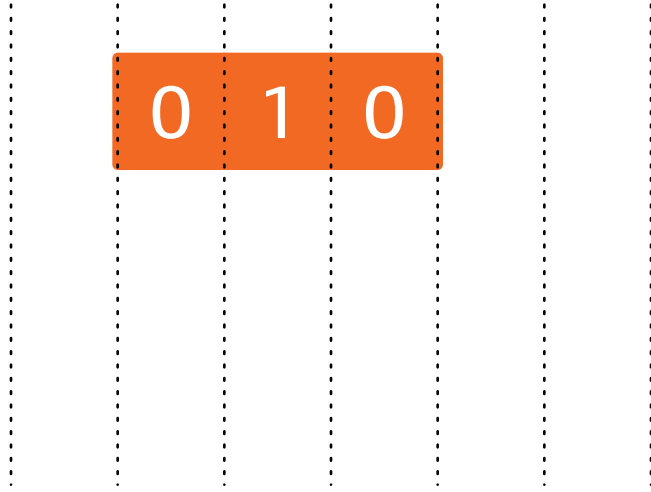
Example:
 X = maximal
independent
set problem

1 0 0

0 1 0

0 0 1

1 0 1

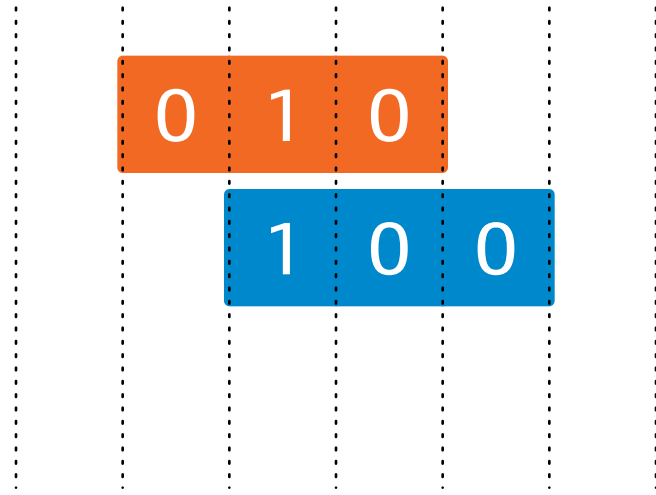


1 0 0

0 1 0

0 0 1

1 0 1



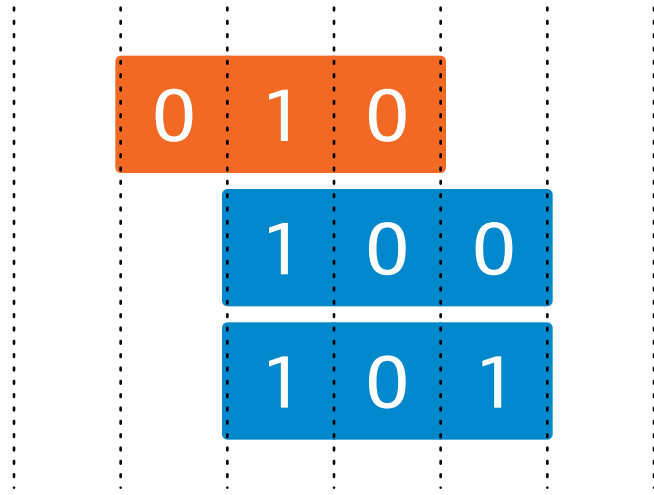
Compatible neighborhoods for adjacent nodes

1 0 0

0 1 0

0 0 1

1 0 1



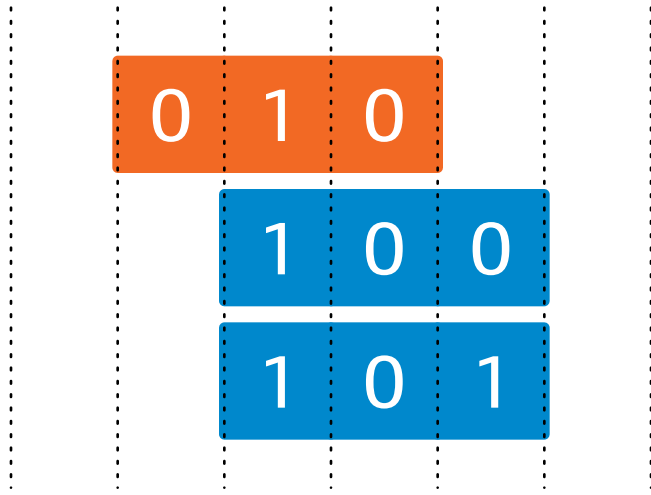
Compatible neighborhoods for adjacent nodes

1 0 0

0 1 0

0 0 1

1 0 1



1 0 0

0 1 0

0 0 1

1 0 1

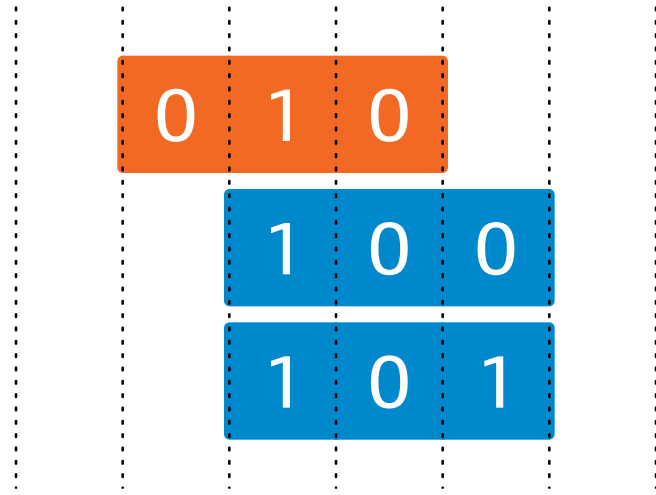


1 0 0

0 1 0

0 0 1

1 0 1

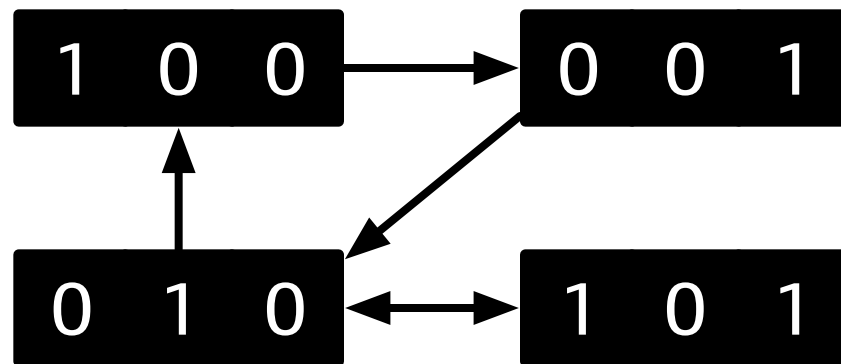
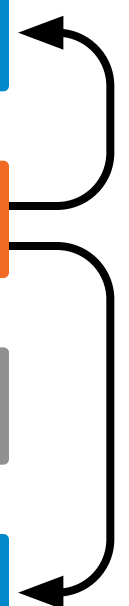


1 0 0

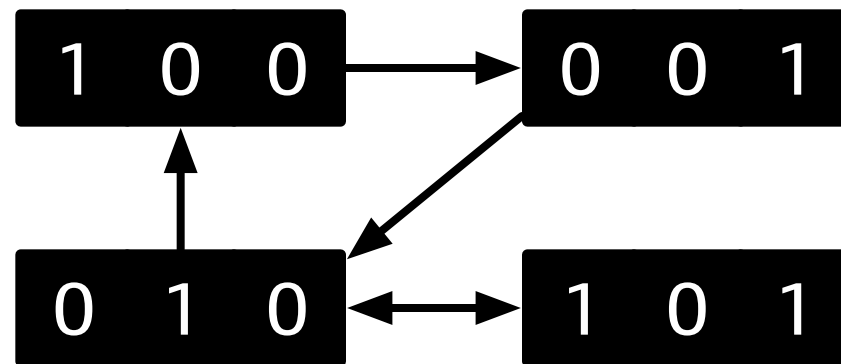
0 1 0

0 0 1

1 0 1



This graph
is all that
we need!

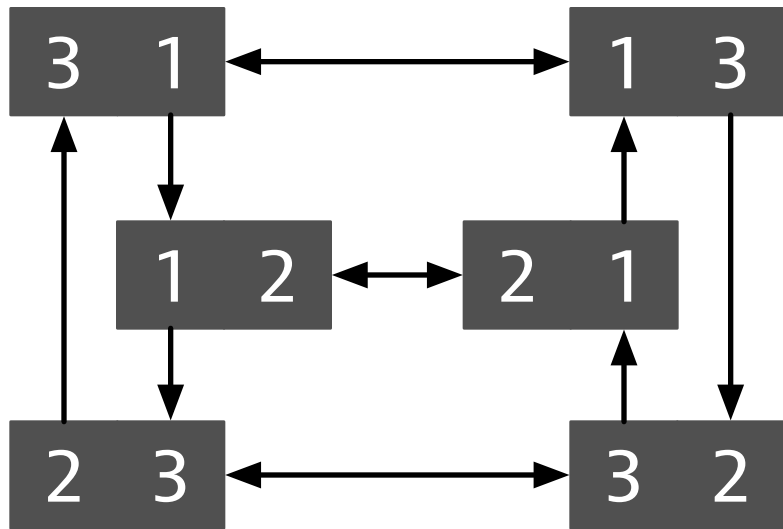


1 2

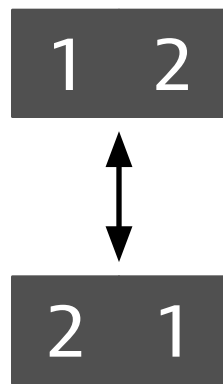


2 1

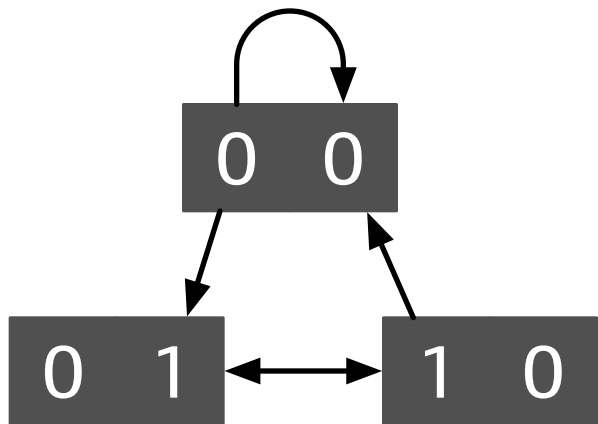
2-coloring



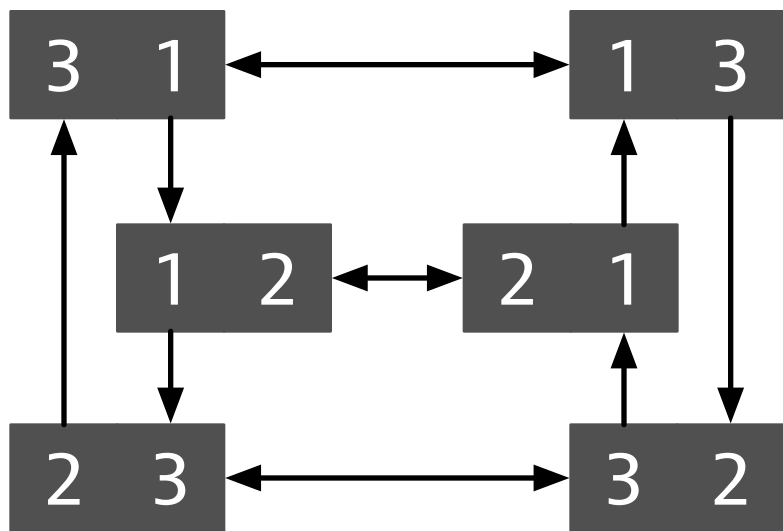
3-coloring



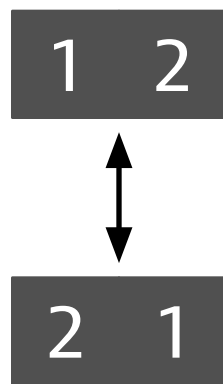
2-coloring



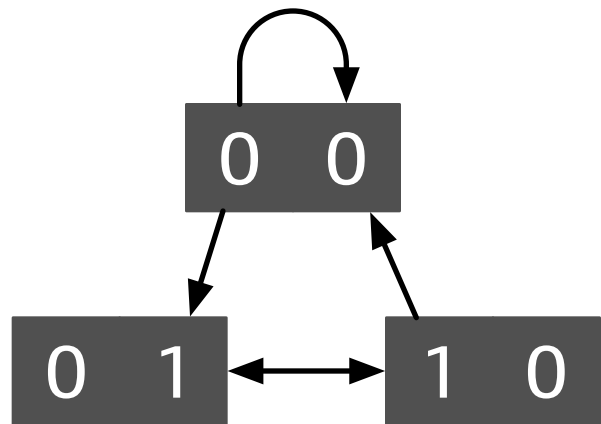
independent set



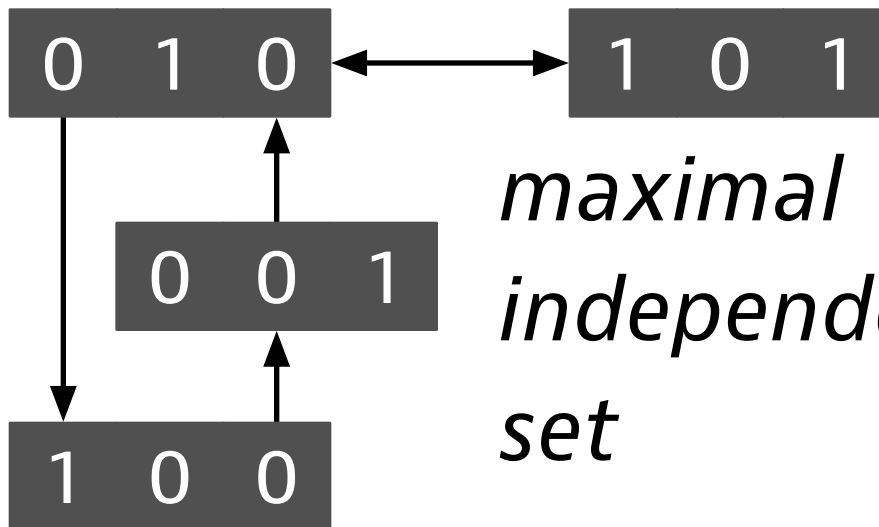
3-coloring



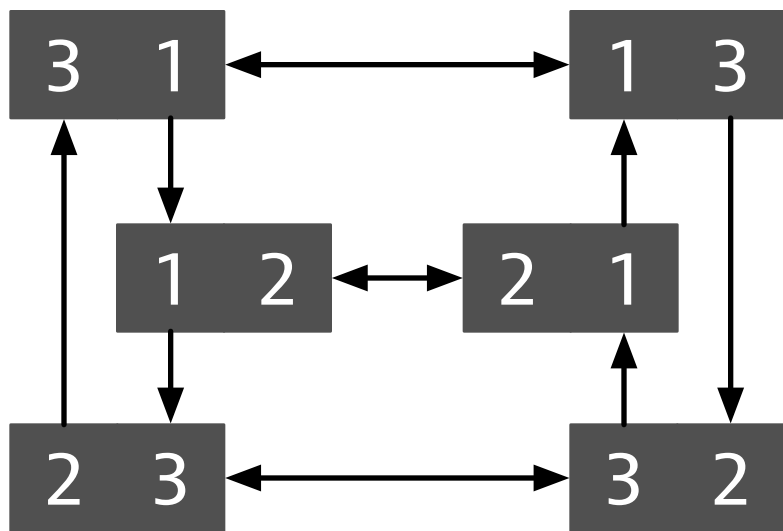
2-coloring



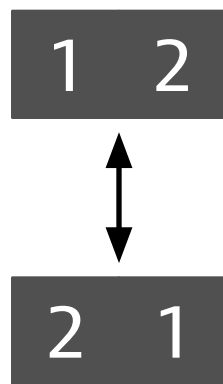
independent set



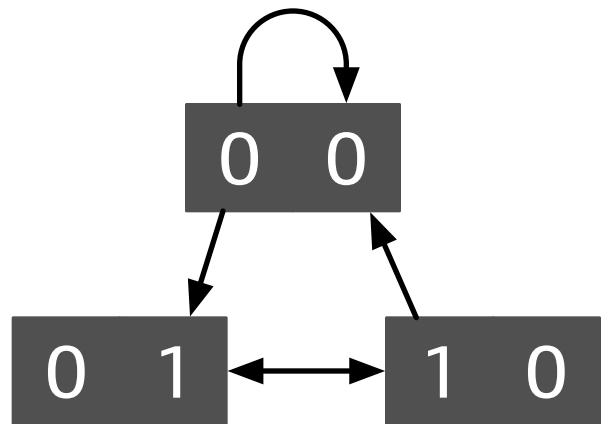
maximal independent set



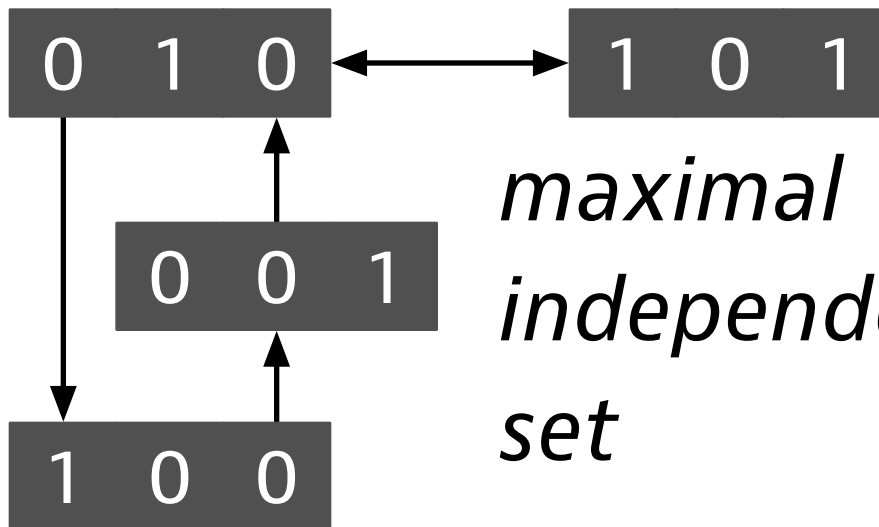
3-coloring



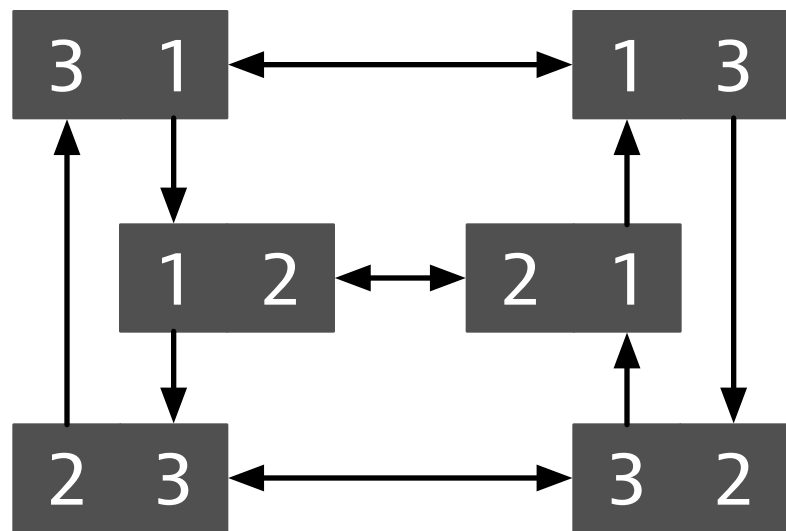
2-coloring



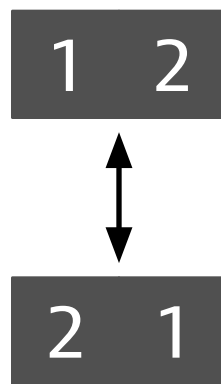
independent set



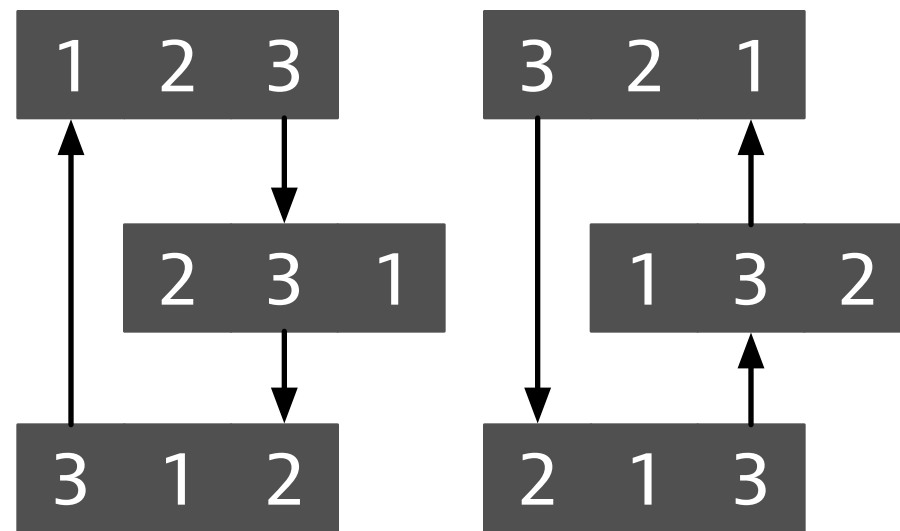
maximal independent set



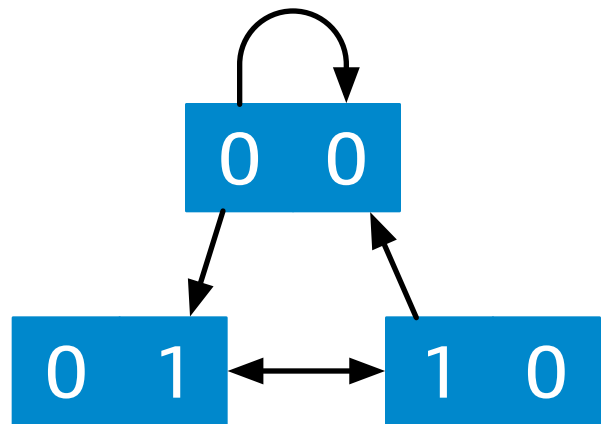
3-coloring



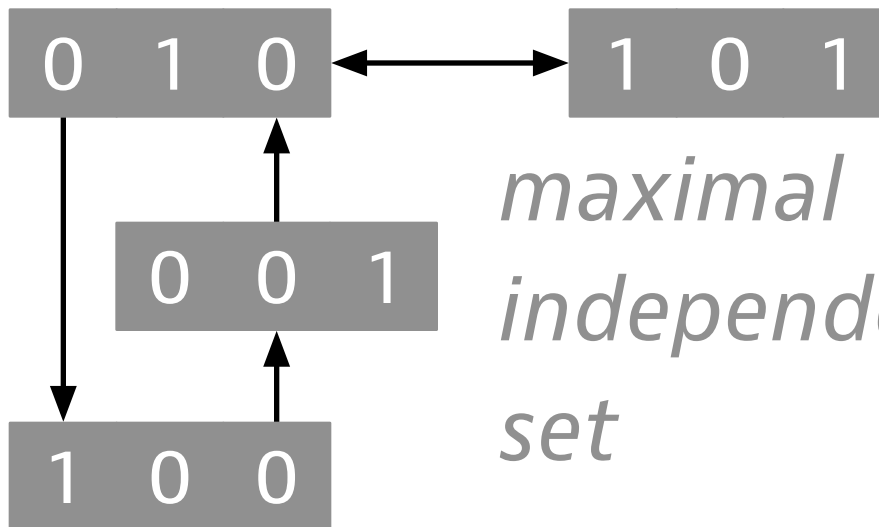
2-coloring



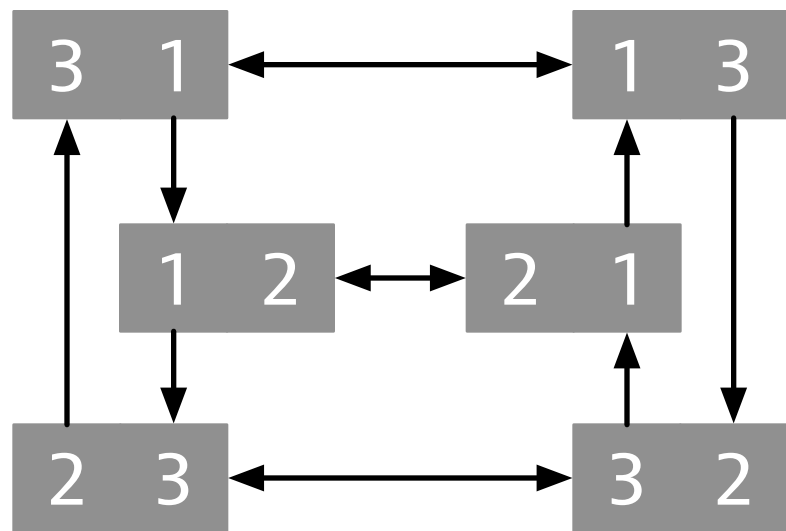
distance-2 coloring



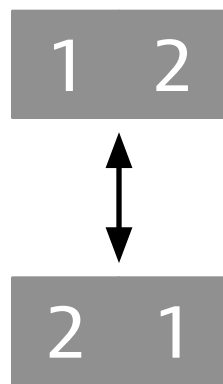
independent set



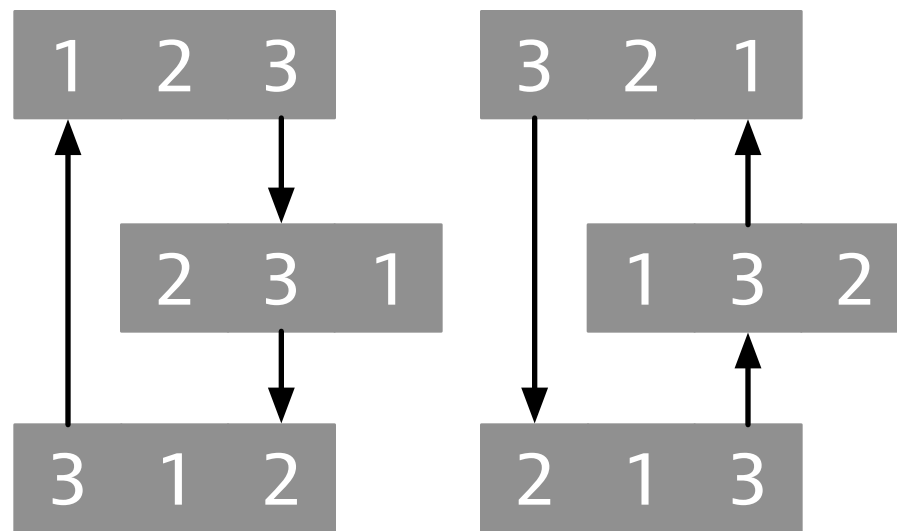
maximal independent set



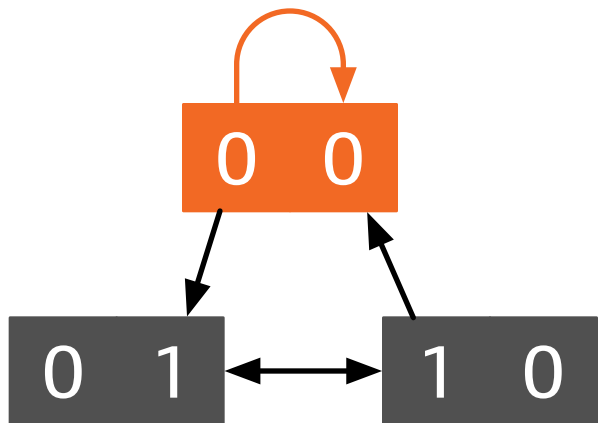
3-coloring



2-coloring

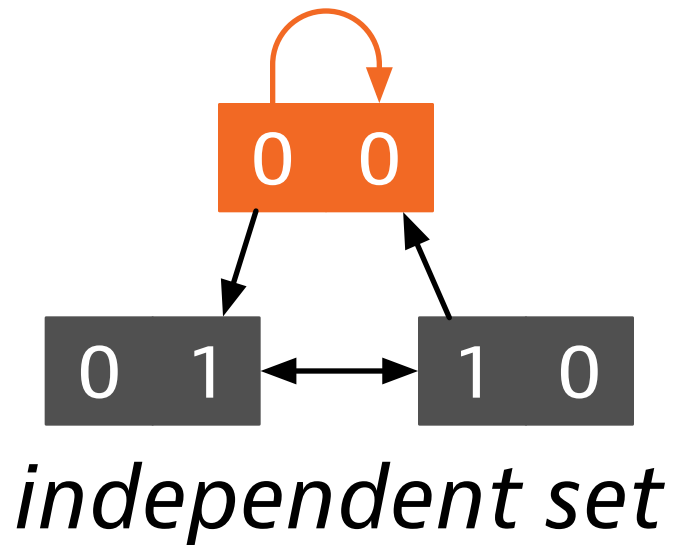


distance-2 coloring



independent set

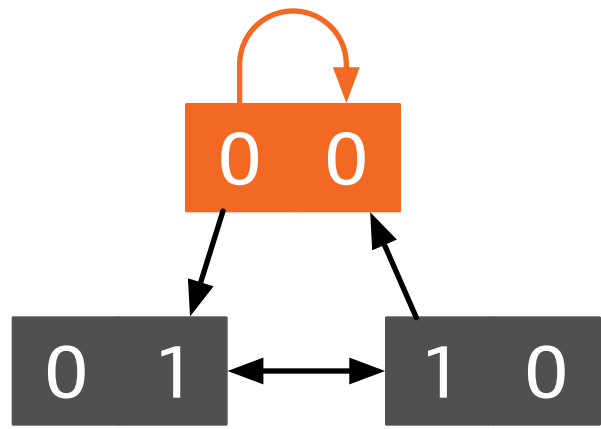
self-loop



self-loop
↓
solvable
in $O(1)$ rounds

Algorithm:

Constant output
(e.g. here all-0)



independent set

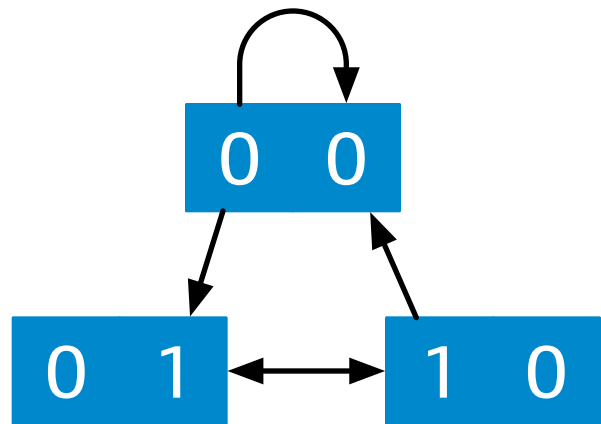
self-loop
↓ ↑
solvable
in $O(1)$ rounds

Proof: No self-loop

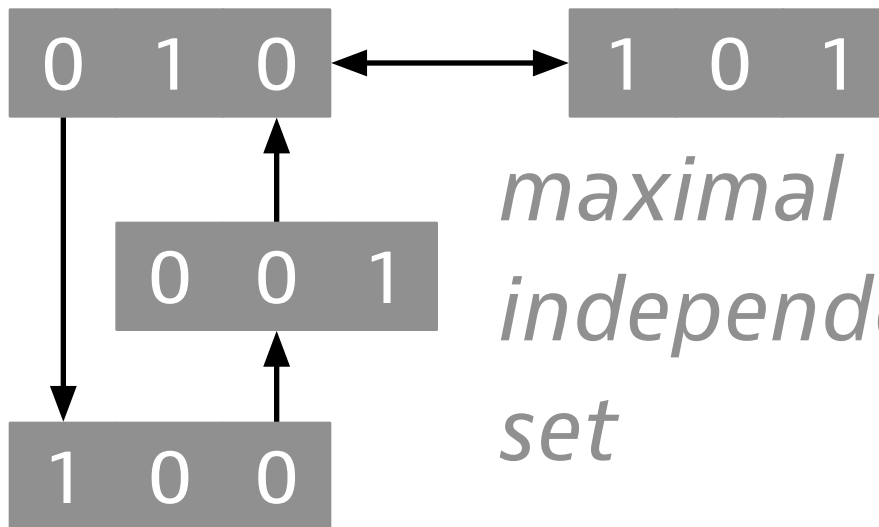
→ any solution breaks symmetry everywhere

→ can be used to find 3-coloring

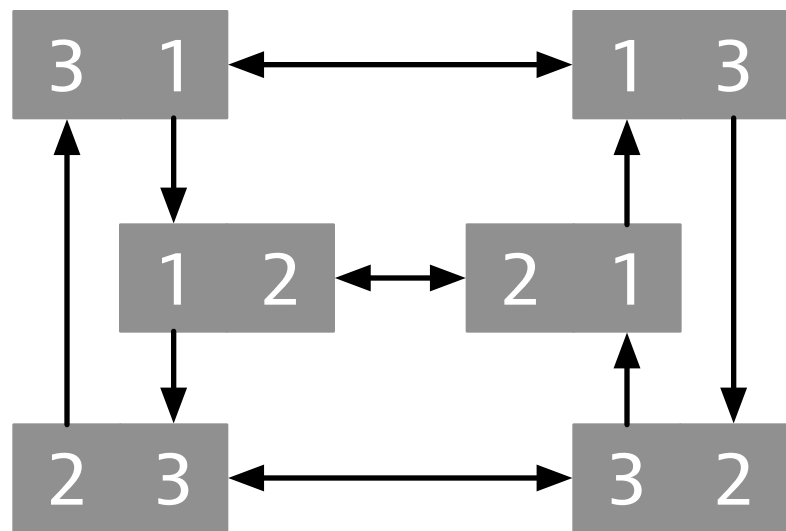
→ not possible in $\mathbf{o(\log^* n)}$ rounds



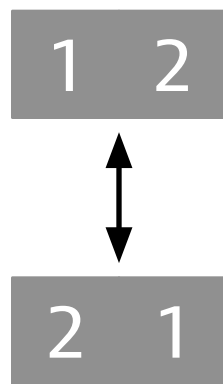
independent set



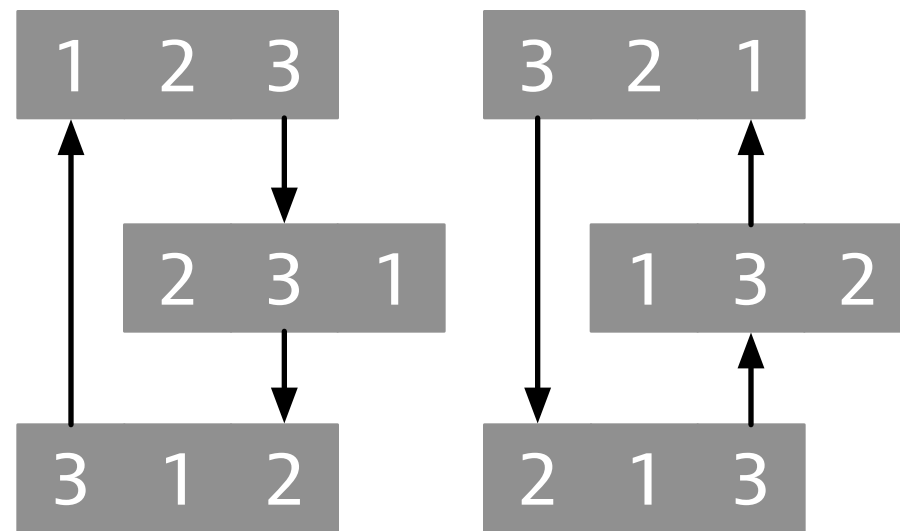
maximal independent set



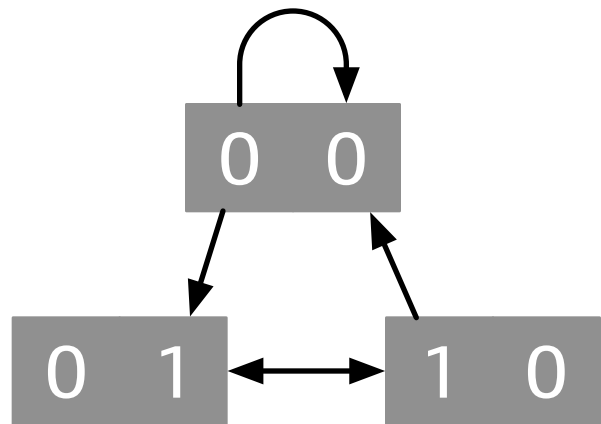
3-coloring



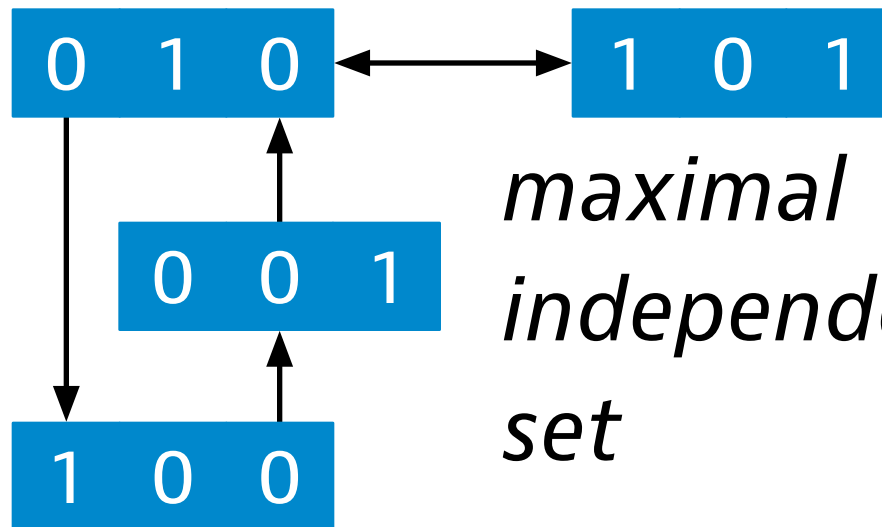
2-coloring



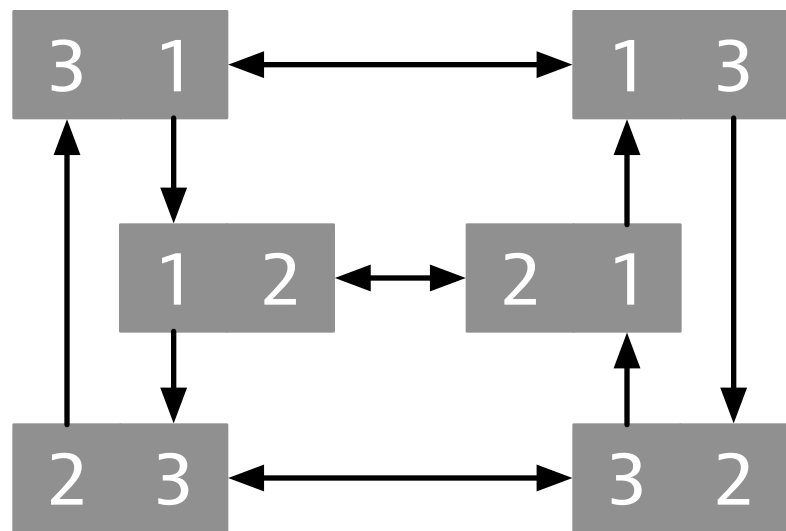
distance-2 coloring



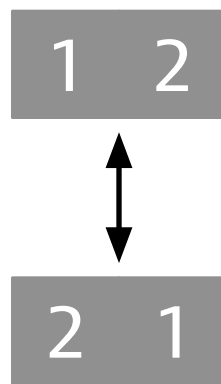
independent set



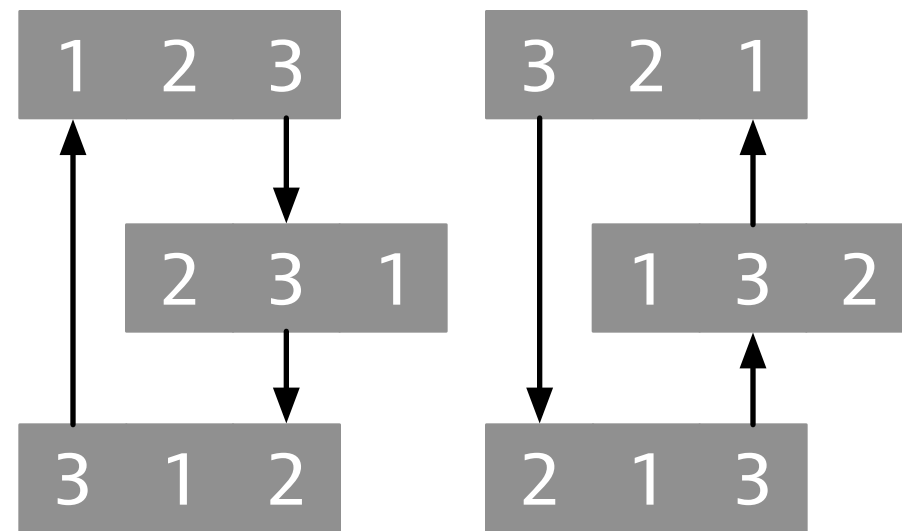
maximal independent set



3-coloring



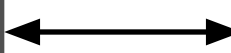
2-coloring



distance-2 coloring

0 1 0

1 0 1



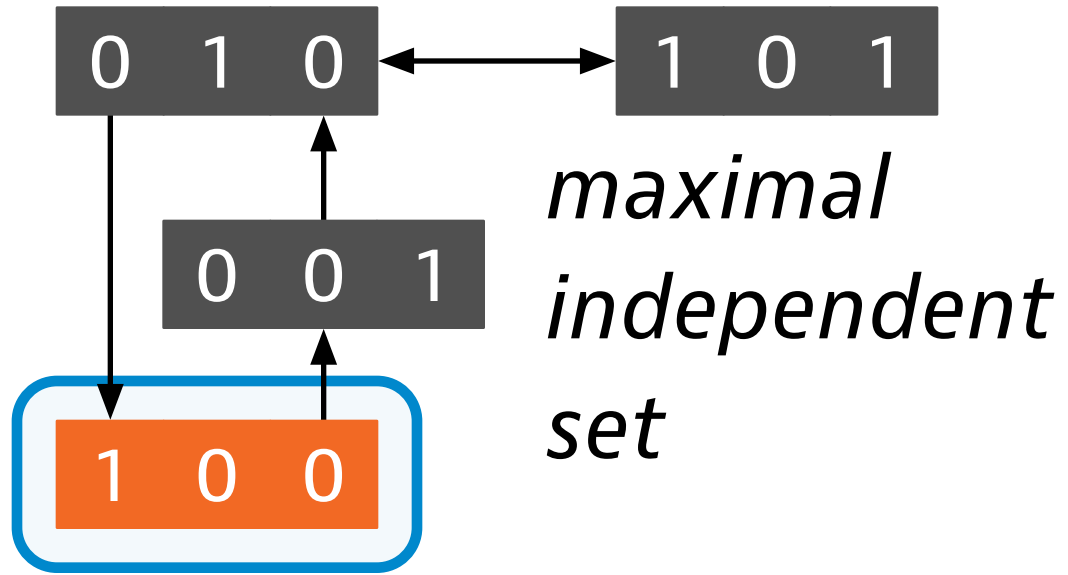
0 0 1

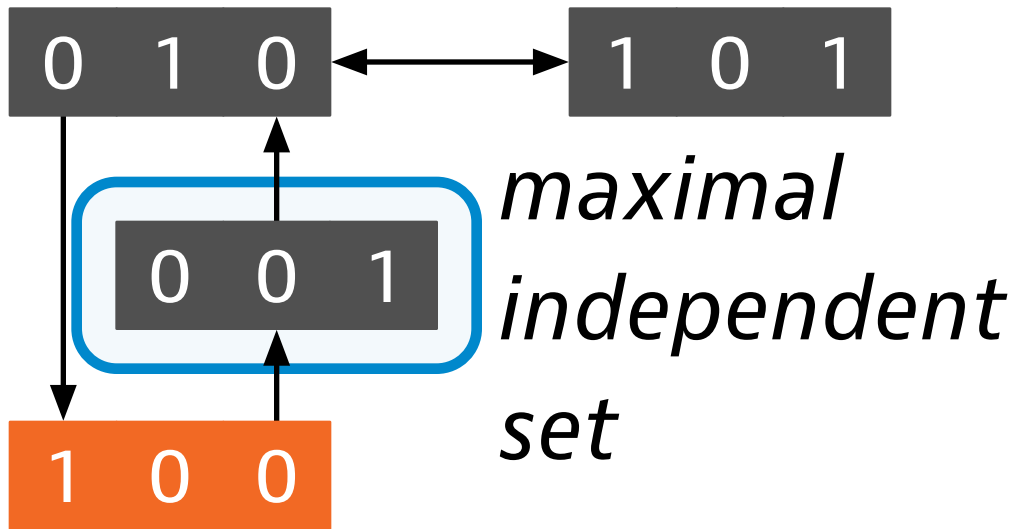
1 0 0



*maximal
independent
set*

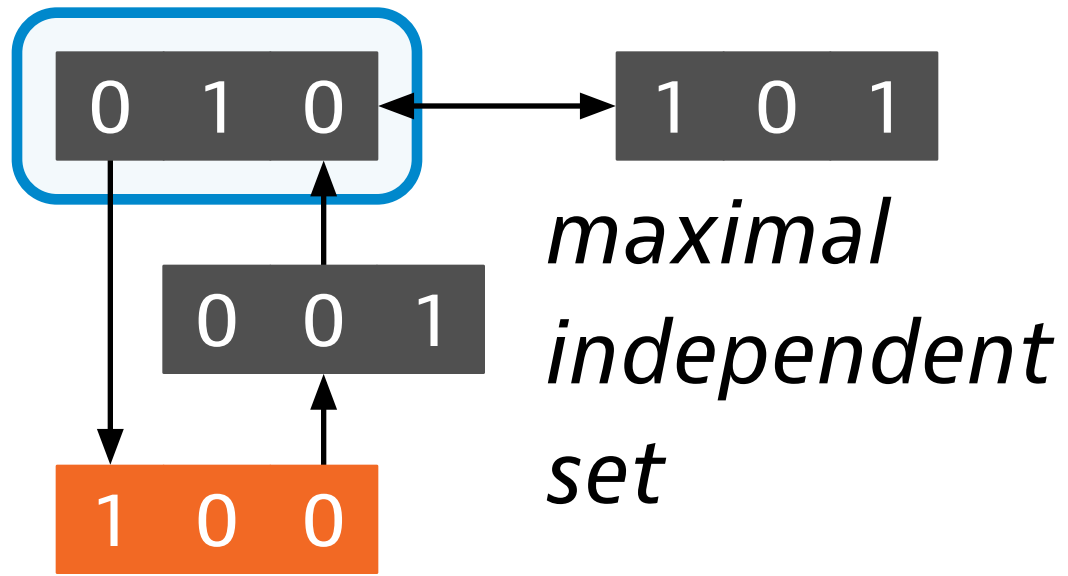
0



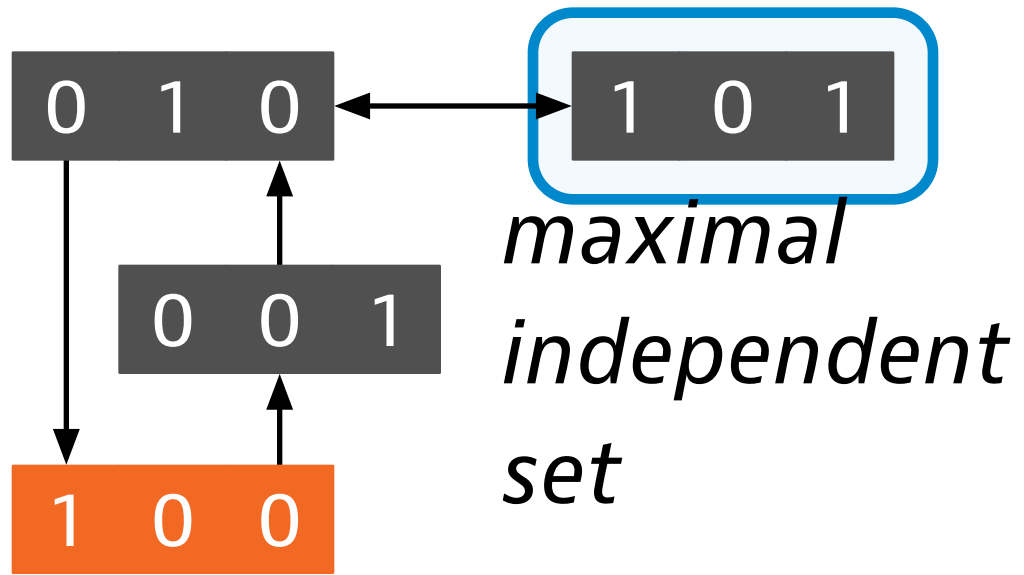


1

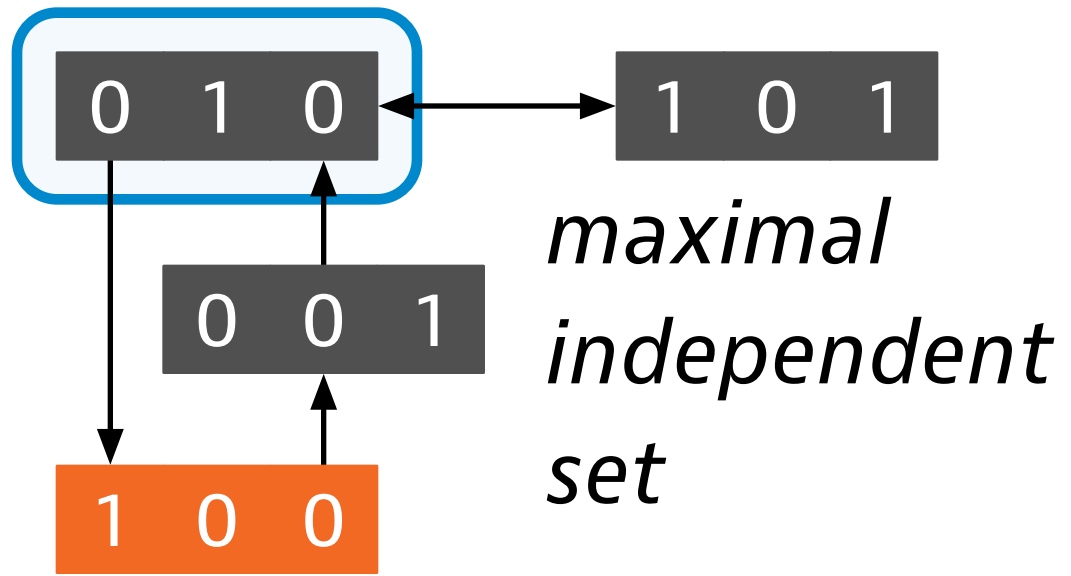
2



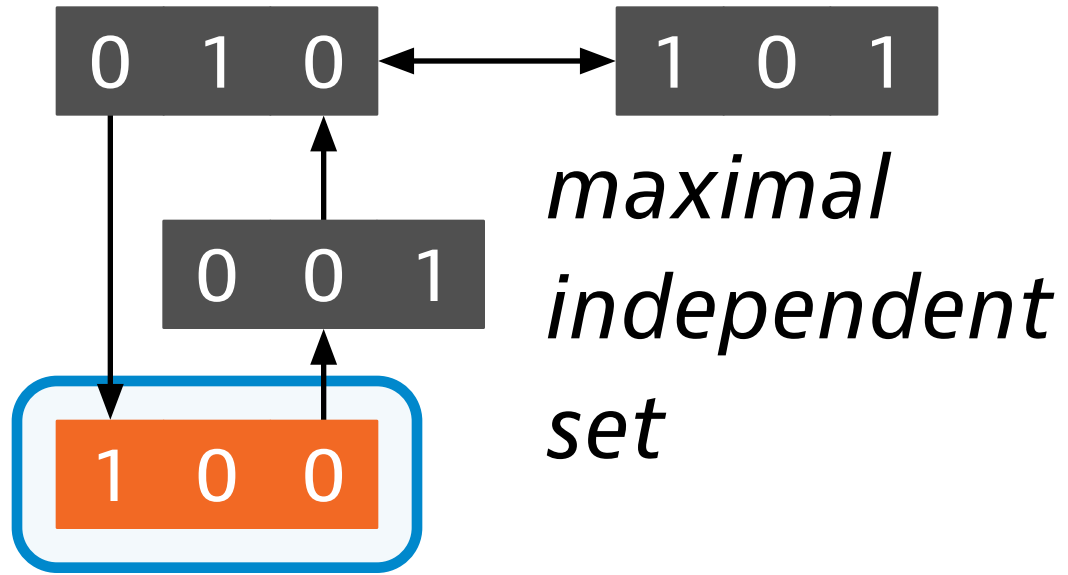
3



4

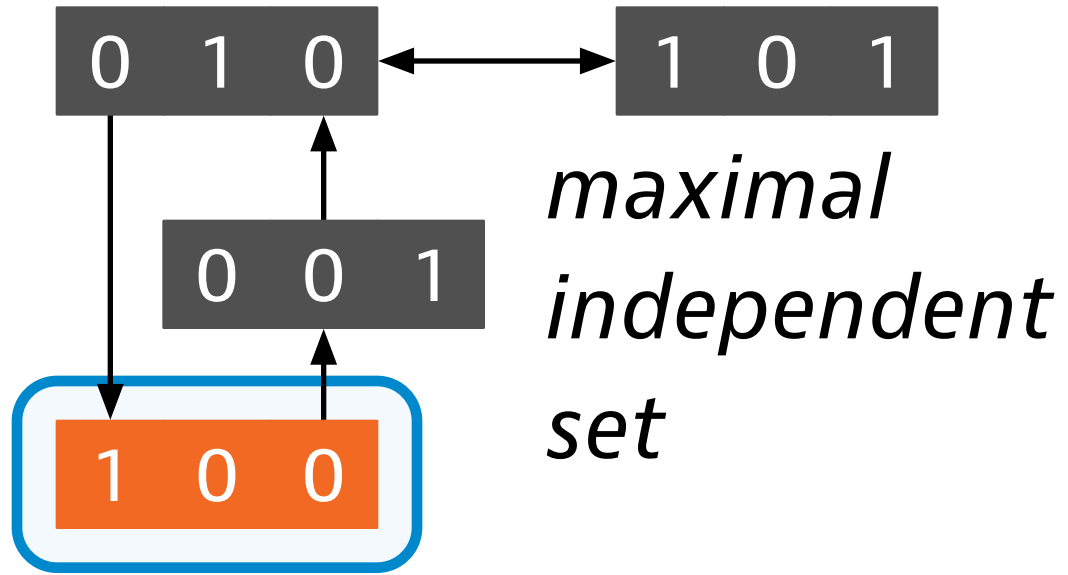


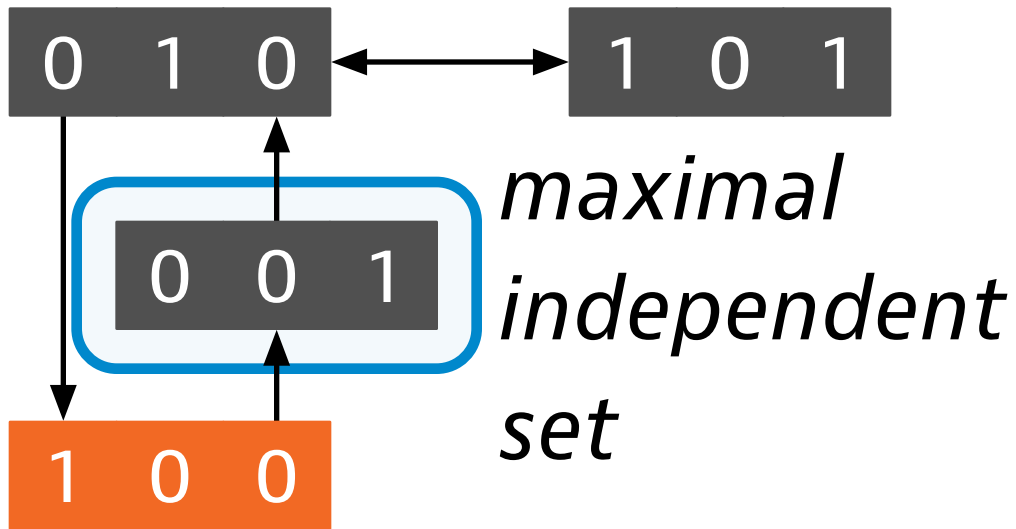
5



Self-returning walk of length 5

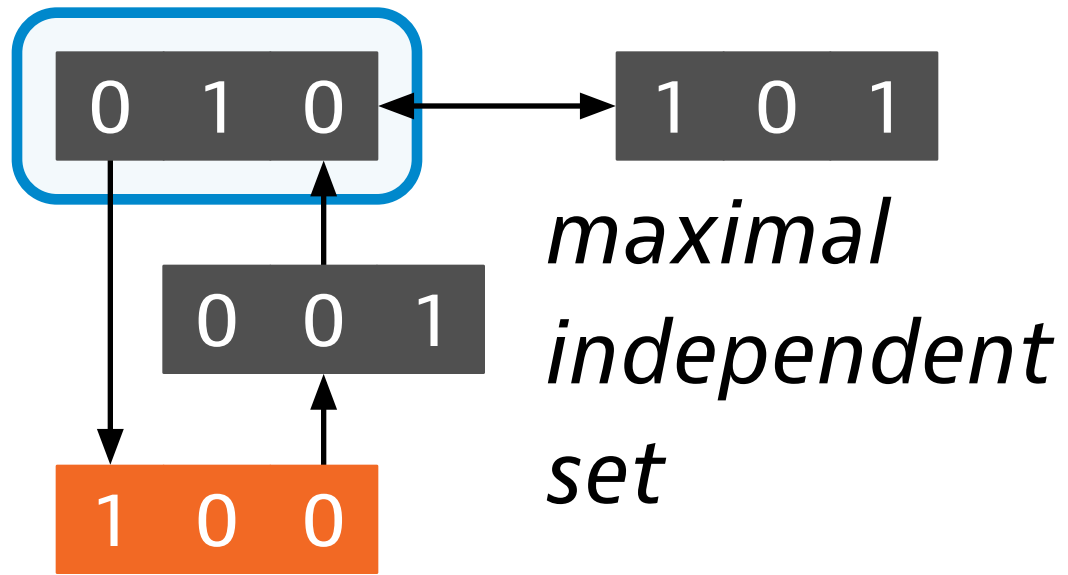
0



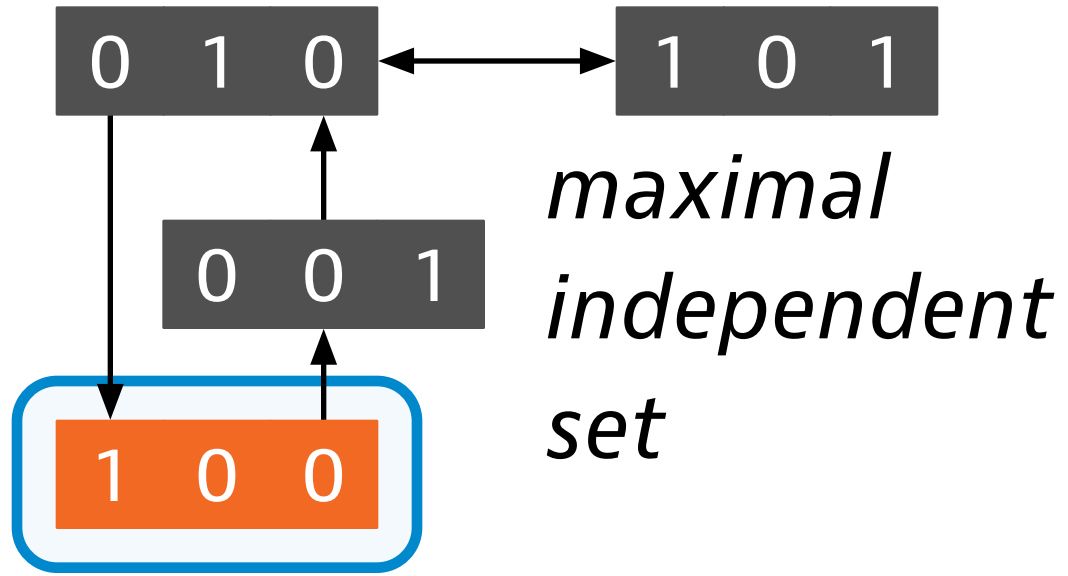


1

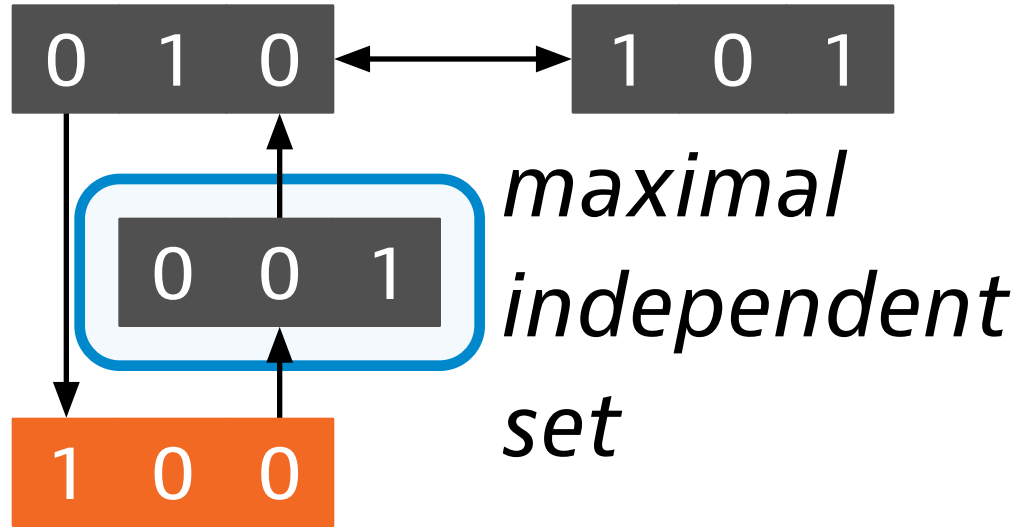
2



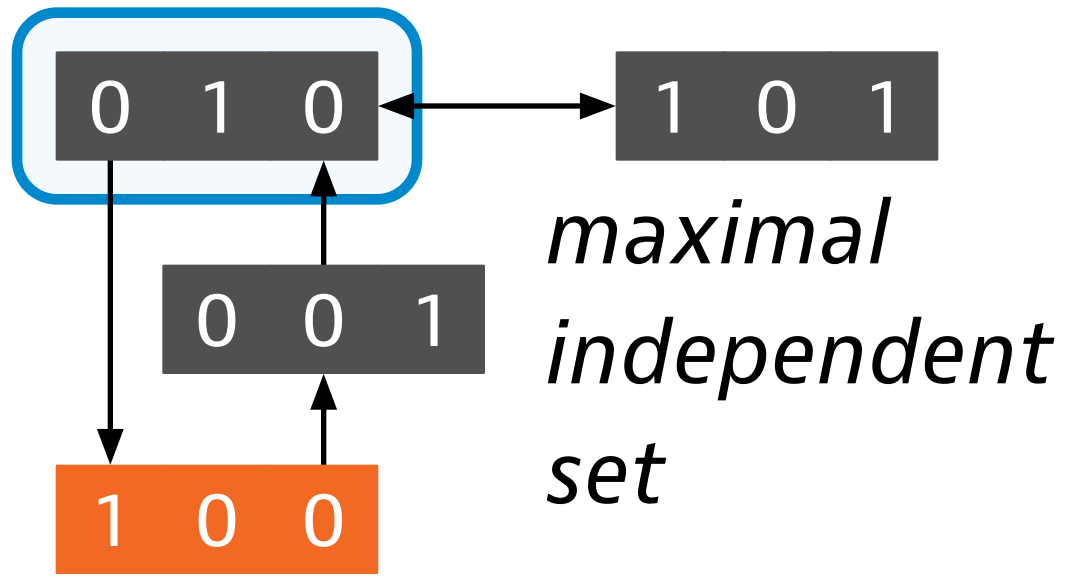
3



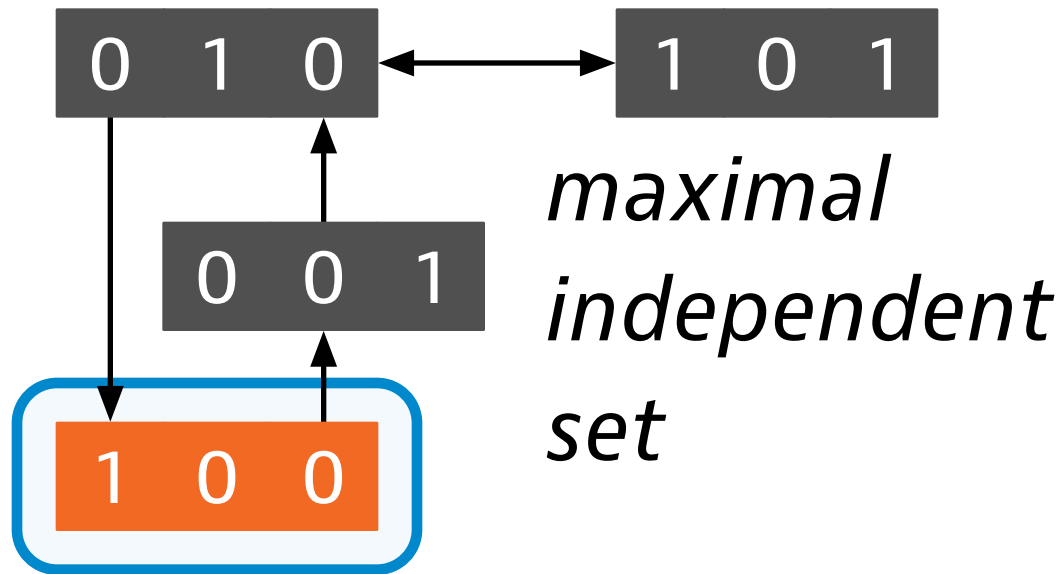
4



5

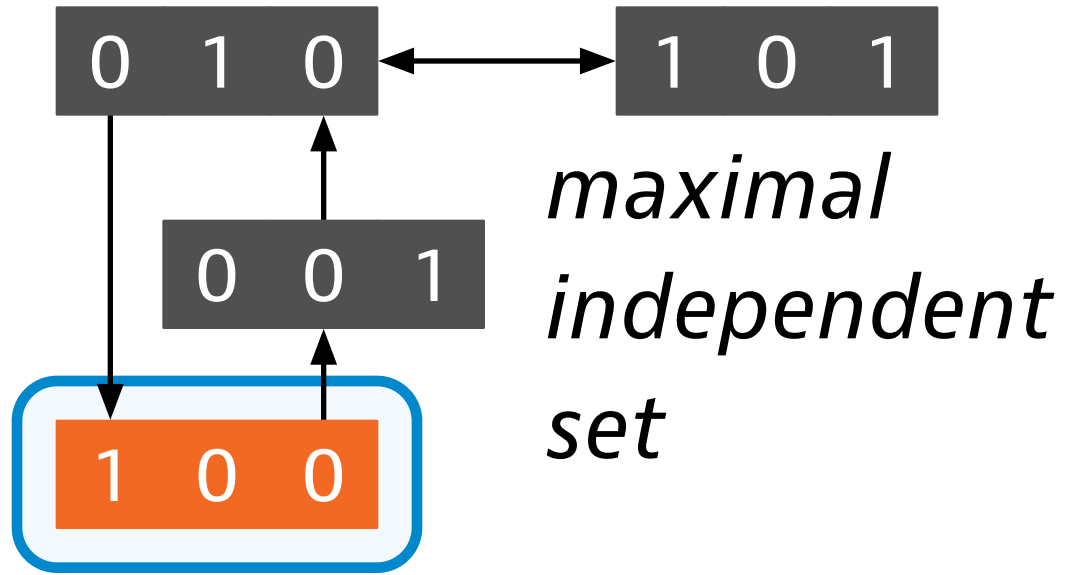


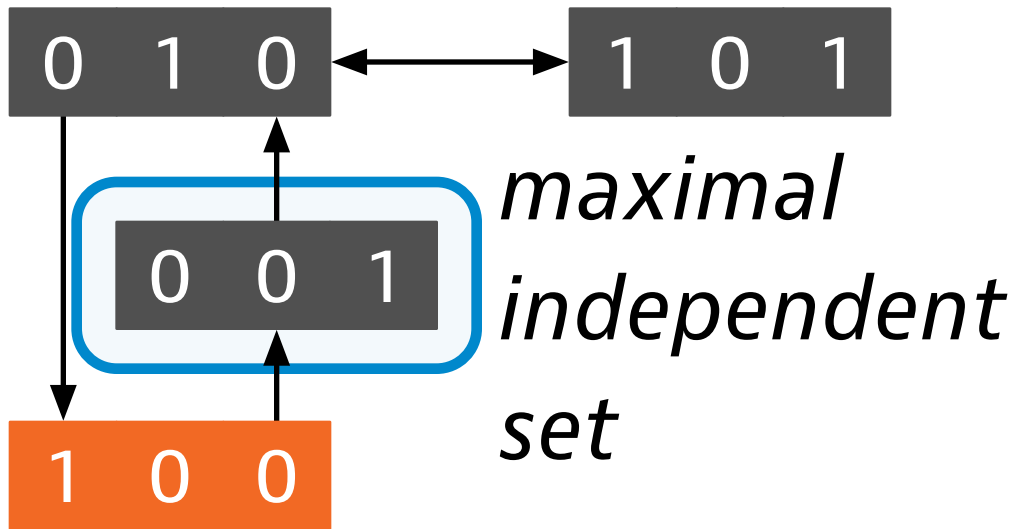
6



Self-returning walk of length 6

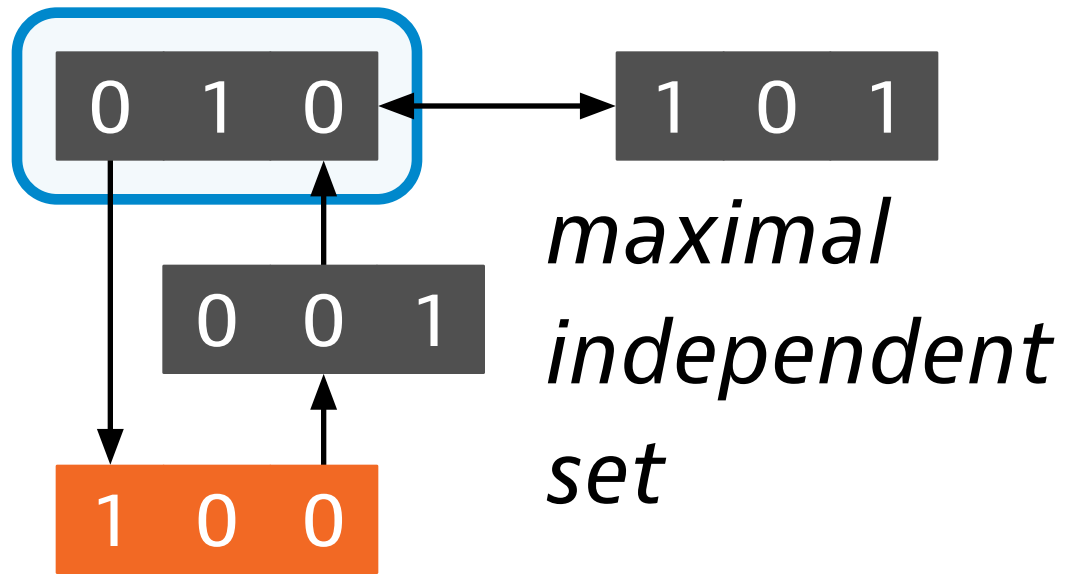
0



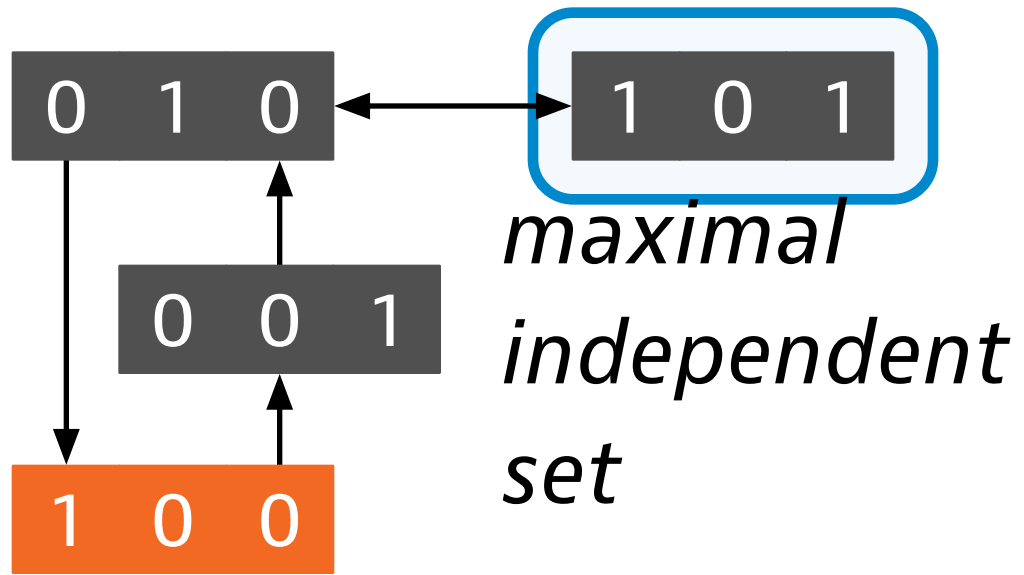


1

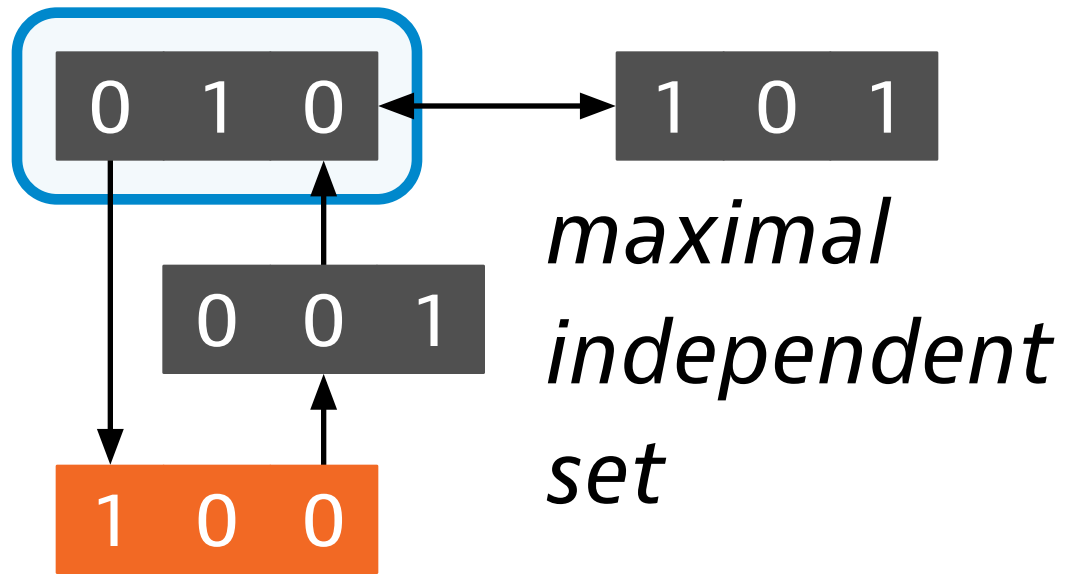
2



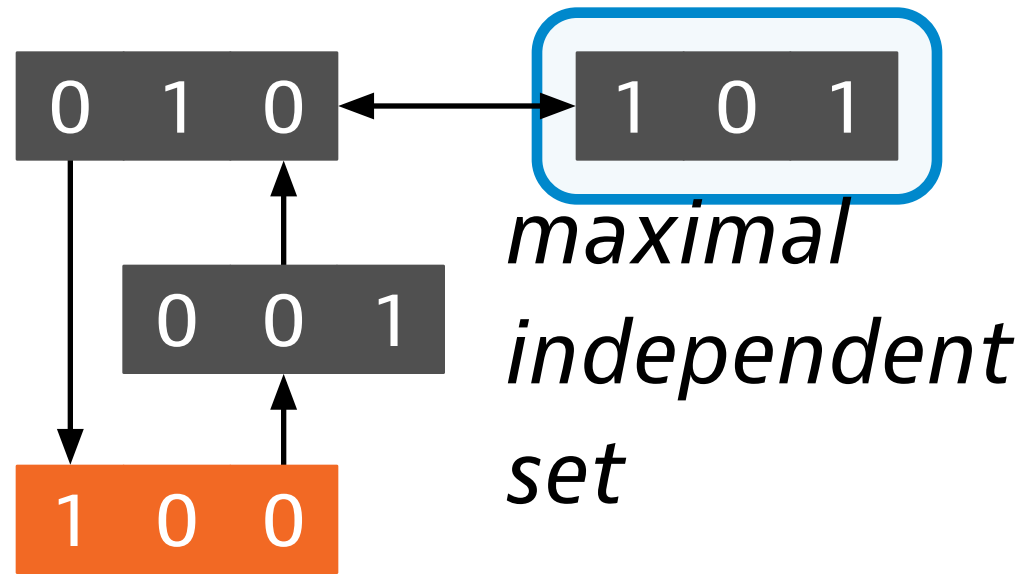
3



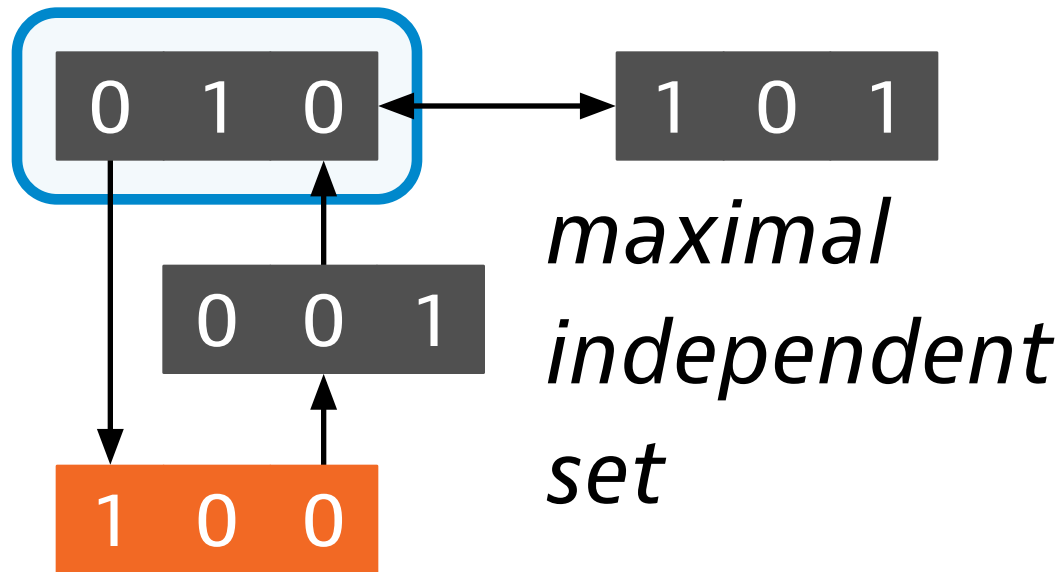
4

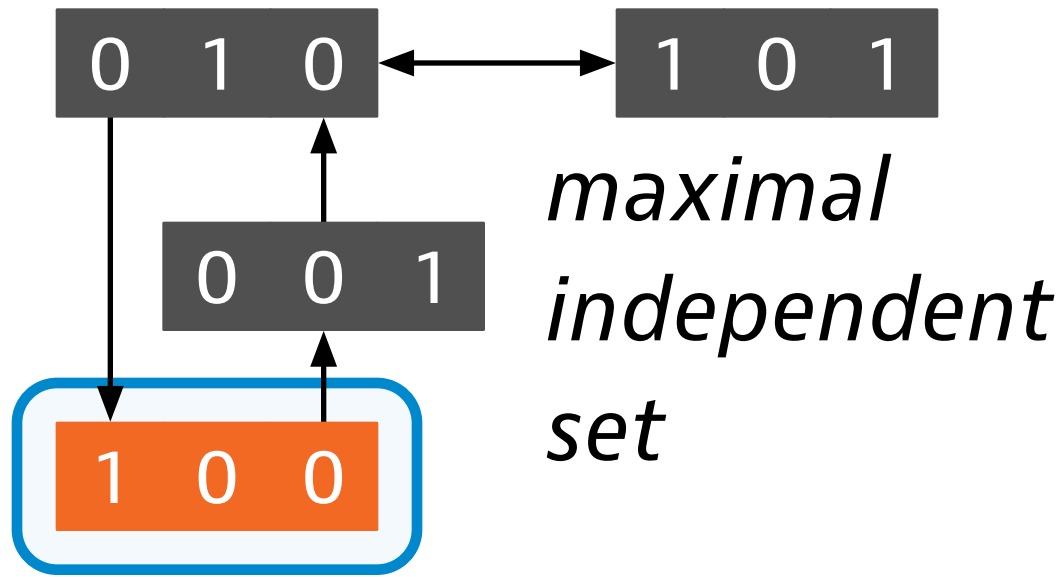


5



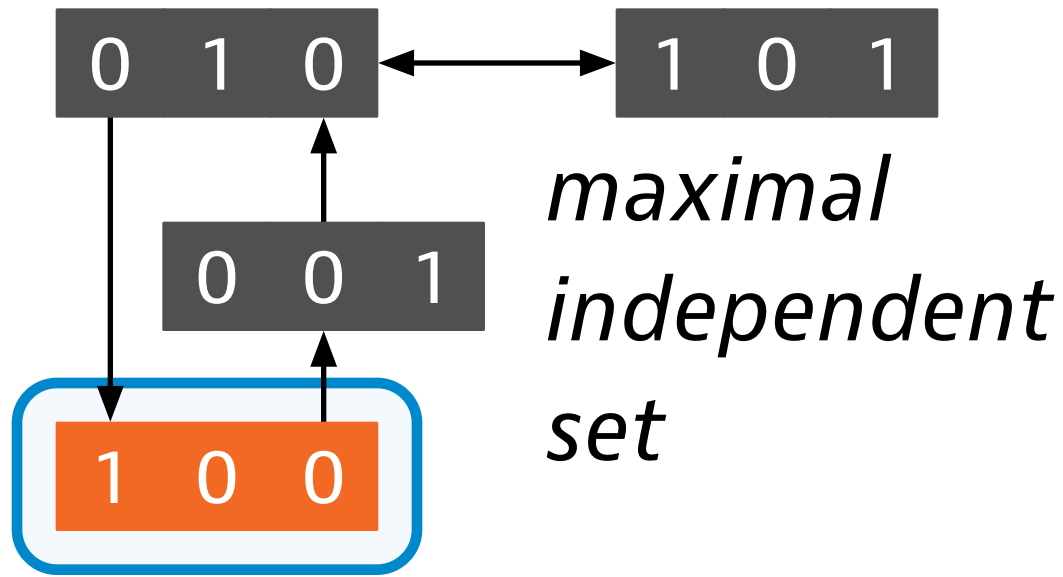
6





Self-returning
walk of length 7

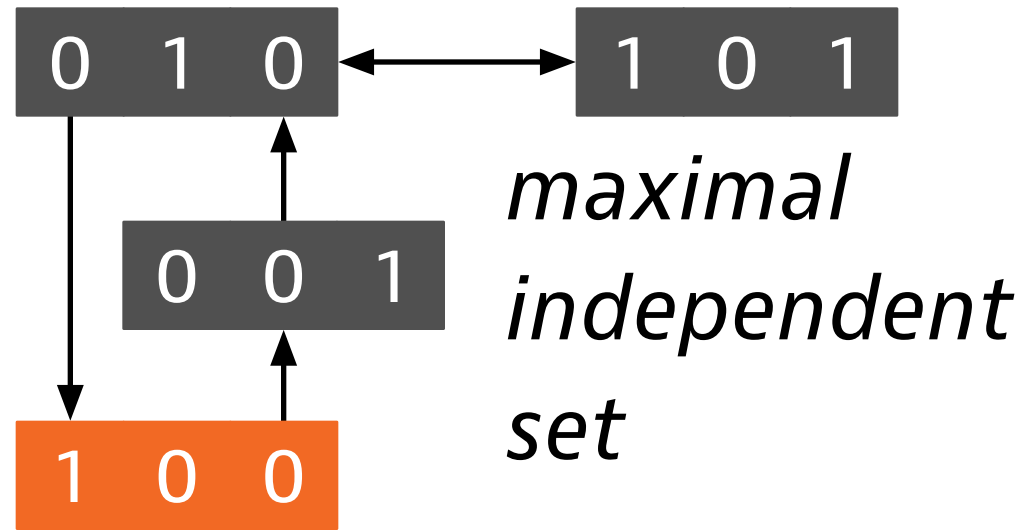
7



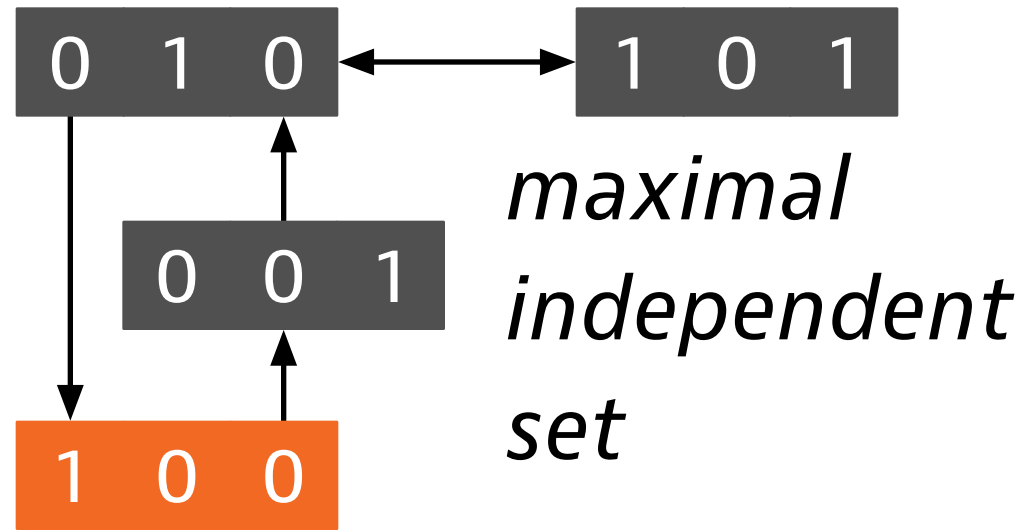
Self-returning walk of length k

$$k = 5, 6, 7, 8, \dots$$

“Flexible”:
for all $k \geq k_0$
there is a self-
returning walk
of length k



“Flexible”:
for all $k \geq k_0$
there is a self-
returning walk
of length k

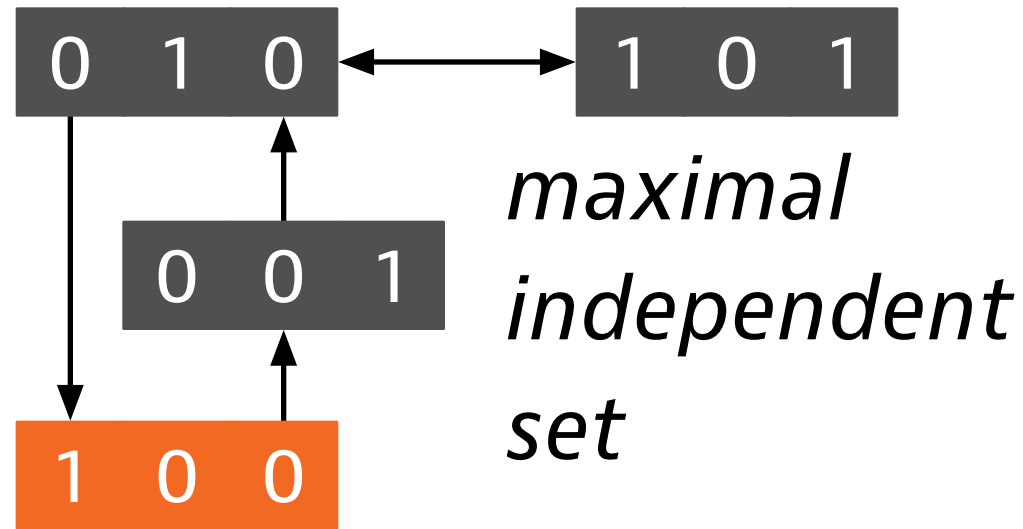


Decidable in
polynomial time

“Flexible”:
for all $k \geq k_0$
there is a self-
returning walk
of length k



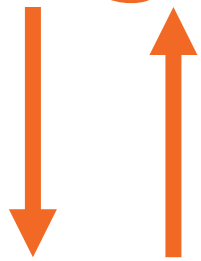
solvable in
 $O(\log^* n)$ rounds



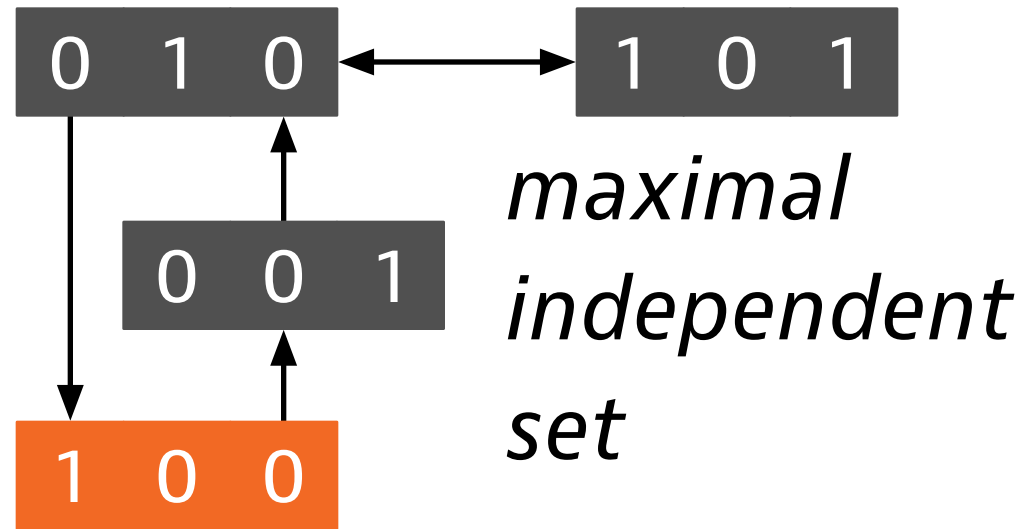
Algorithm:

- split in blocks of length $\geq k_0$
- use the flexible configuration at each block boundary
- fill in between boundaries by following a self-returning walk

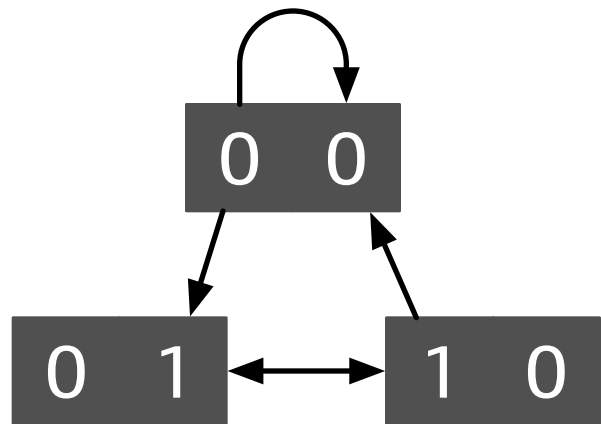
“Flexible”:
for all $k \geq k_0$
there is a self-
returning walk
of length k



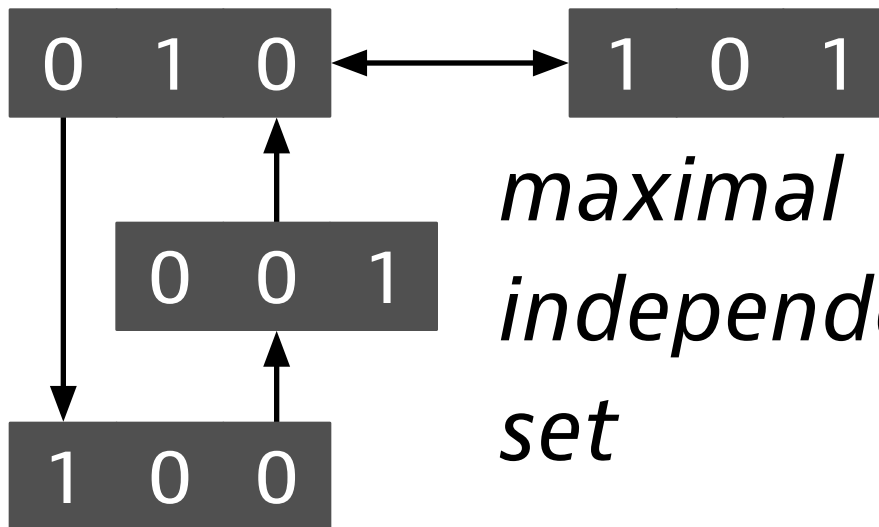
solvable in
 $O(\log^* n)$ rounds



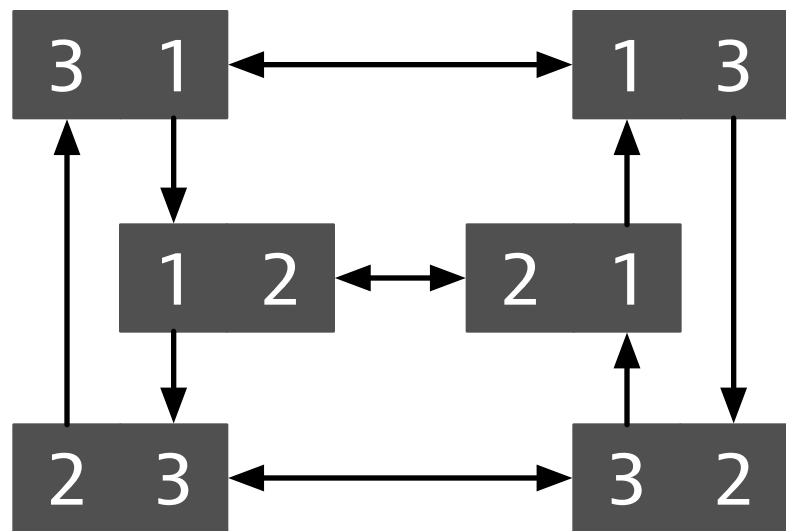
Proof: Not flexible \rightarrow must use
the same non-flexible configuration
at least twice far from each other;
not compatible for all distances
 \rightarrow global coordination needed
 \rightarrow not possible in $\mathbf{o(n)}$ rounds



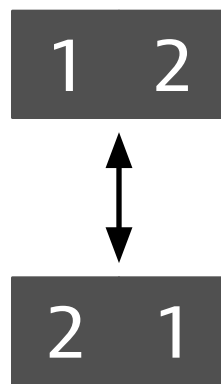
independent set



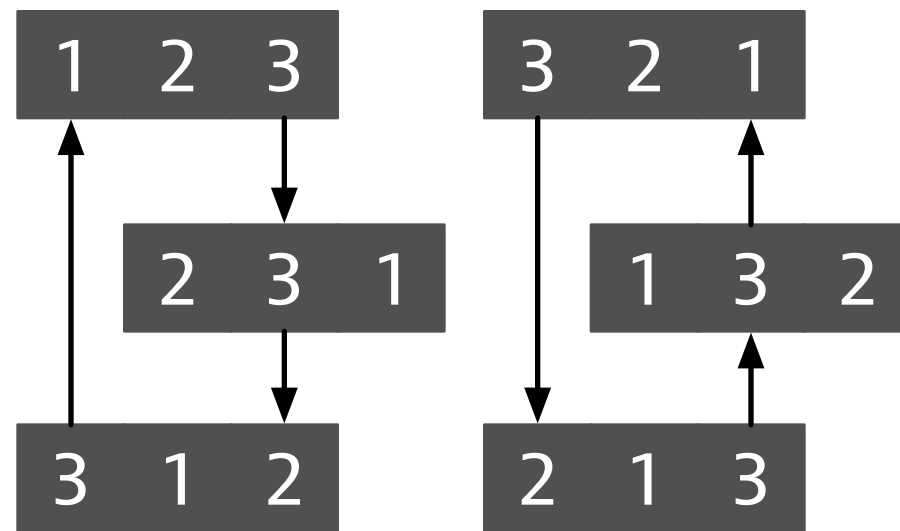
maximal independent set



3-coloring

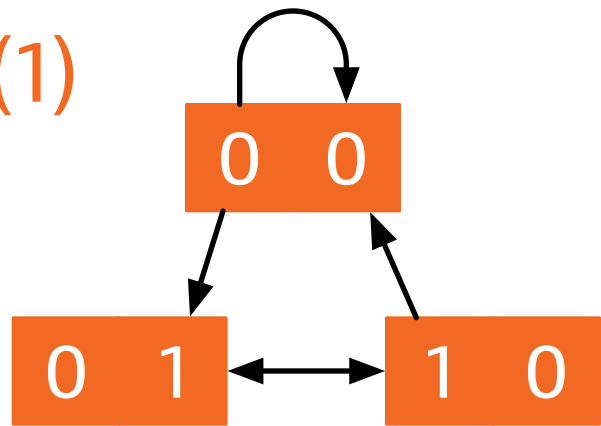


2-coloring

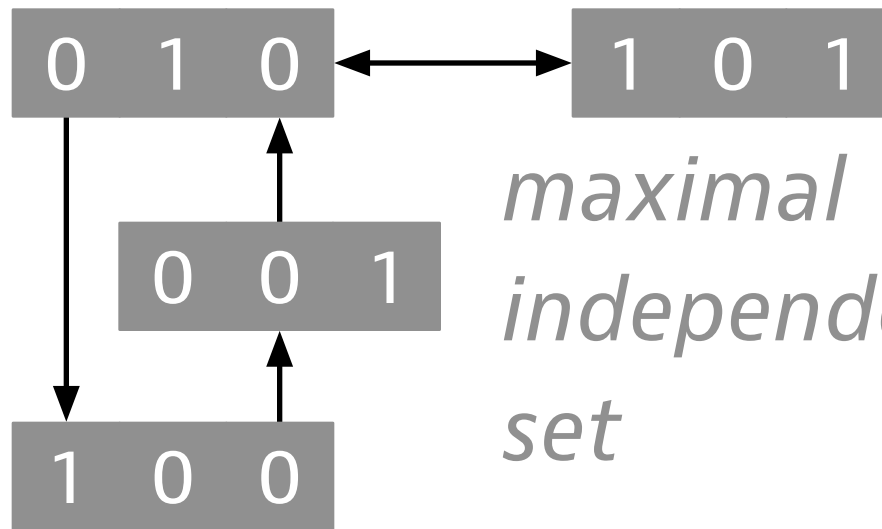


distance-2 coloring

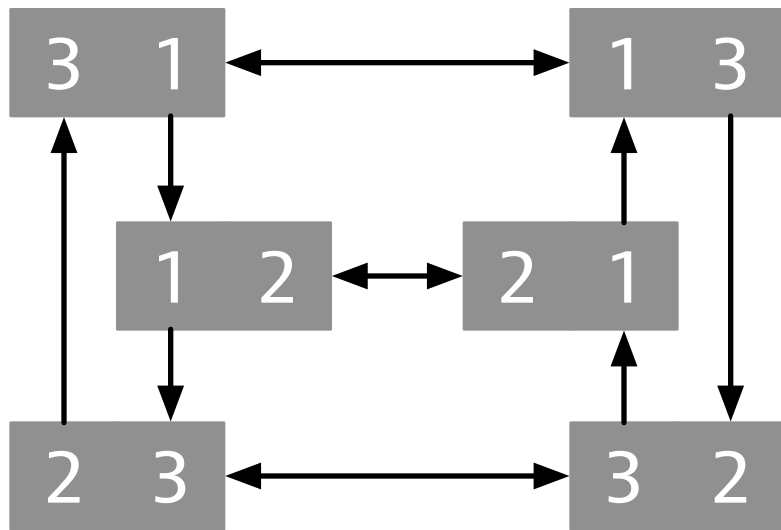
$O(1)$



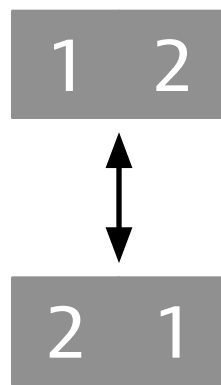
independent set



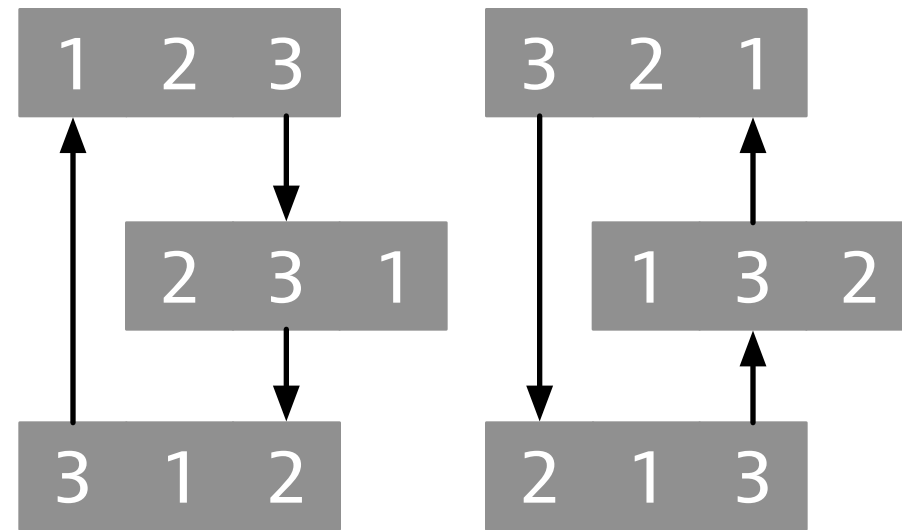
maximal independent set



3-coloring

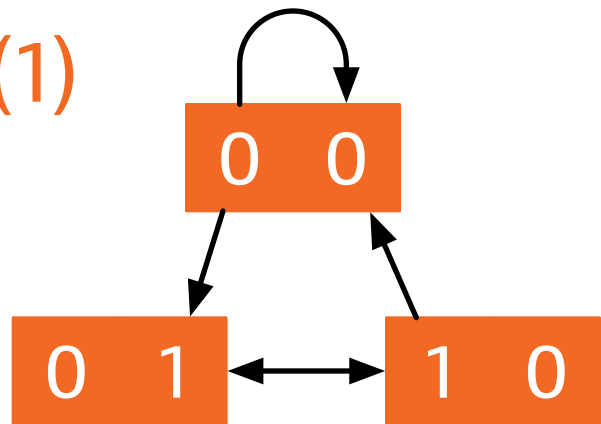


2-coloring

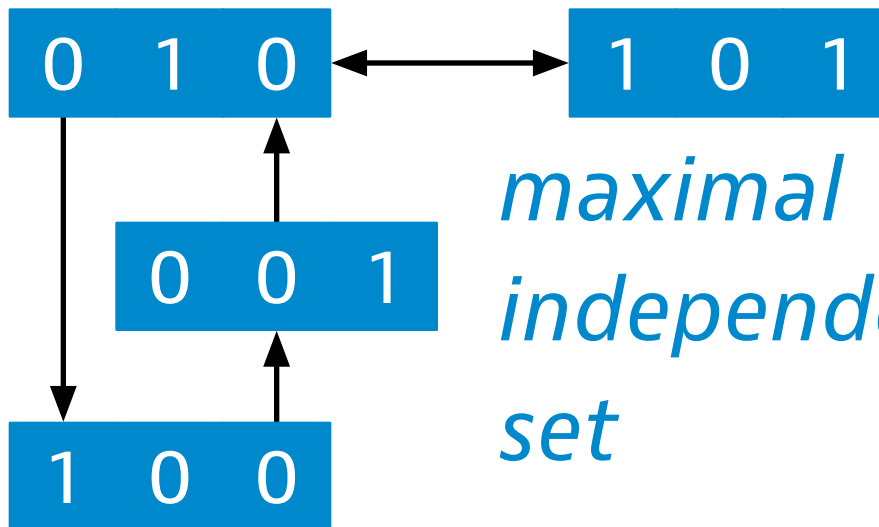


distance-2 coloring

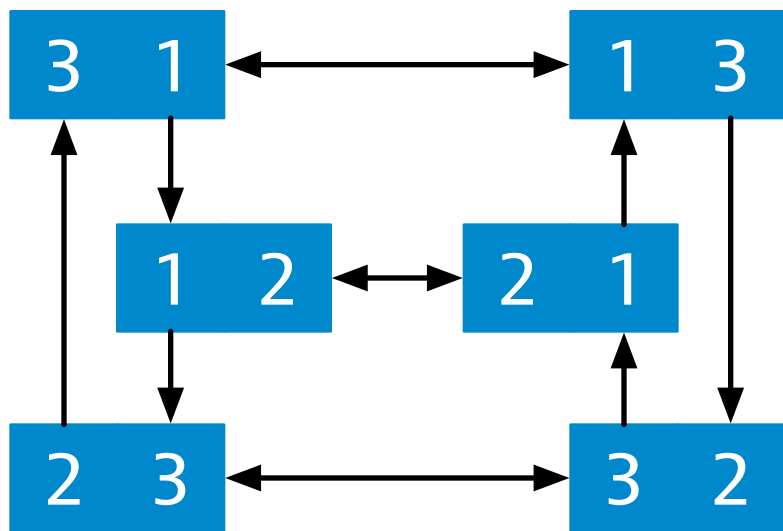
$O(1)$



independent set

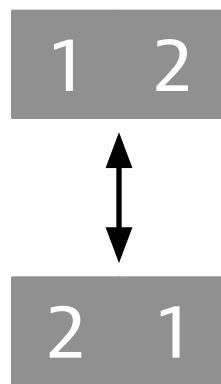


maximal independent set

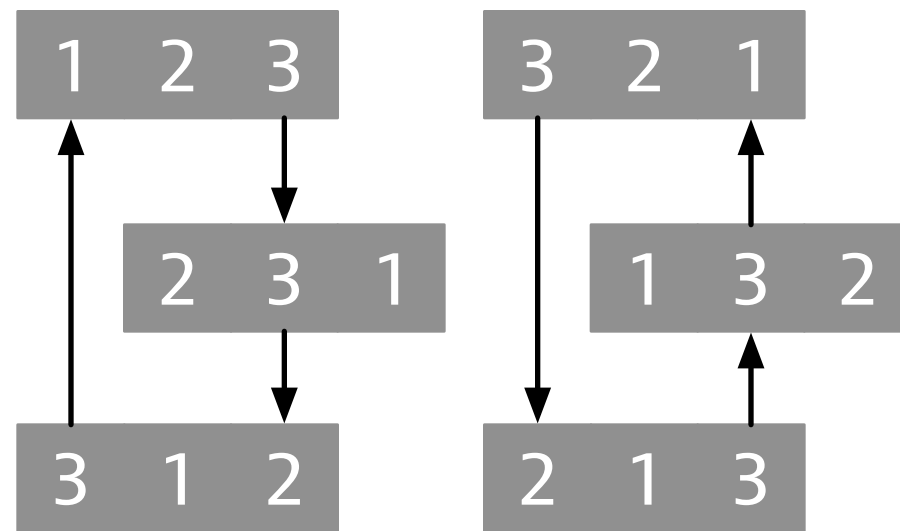


3-coloring

$O(\log^* n)$

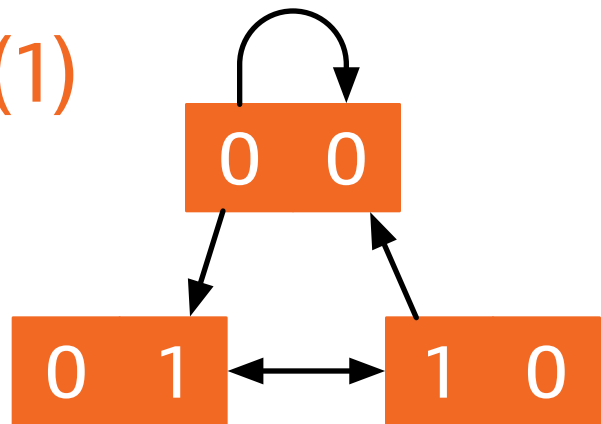


2-coloring

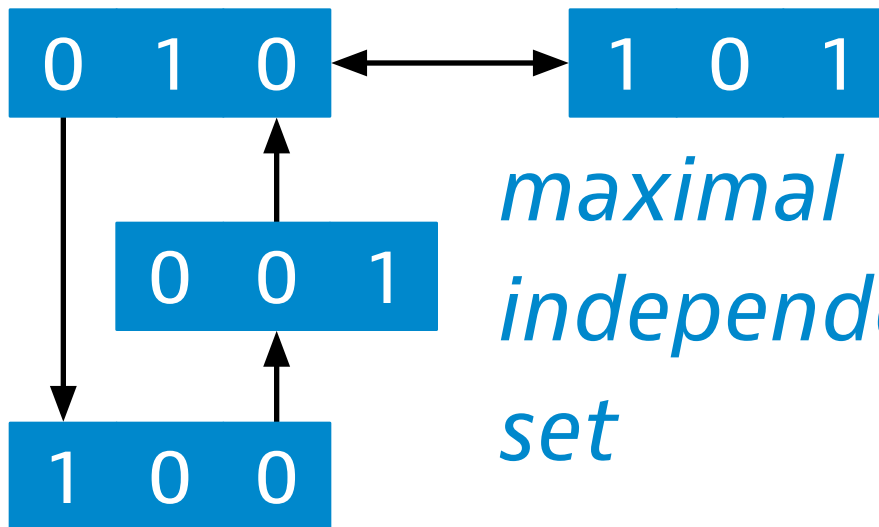


distance-2 coloring

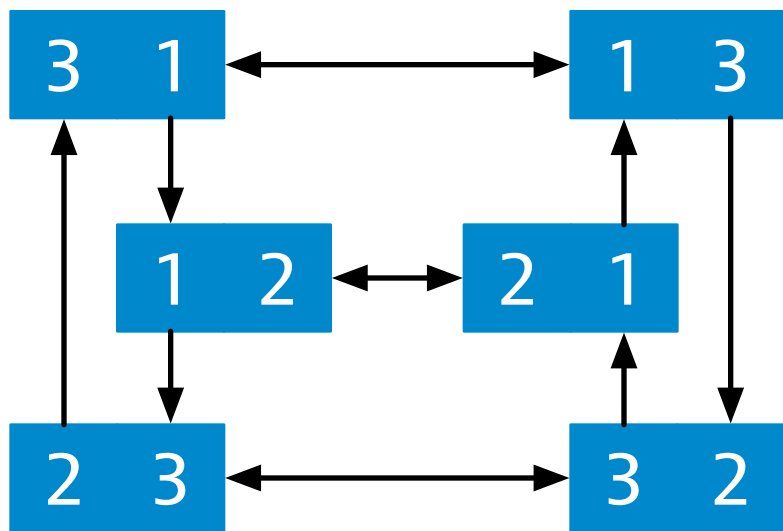
$O(1)$



independent set

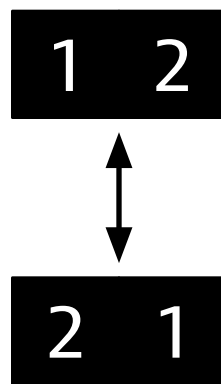


maximal independent set



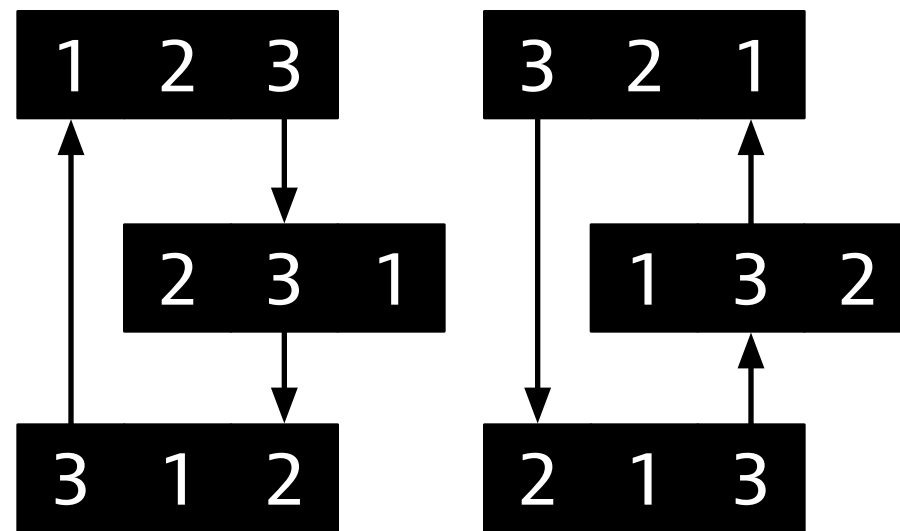
3-coloring

$O(\log^* n)$



2-coloring

$O(n)$

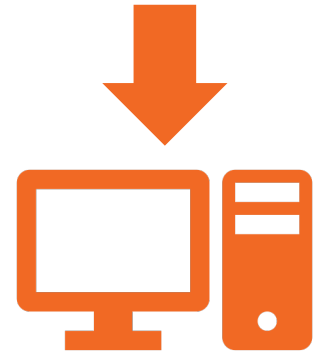


distance-2 coloring

Fully automatic

- Write down the specification of *any locally checkable problem X*
- Then you can *find efficiently*
 - distributed round complexity of X
 - asymptotically optimal distributed algorithm for X

$X = \{ 001, 010, 100, 101 \}$



This algorithm solves X in time $O(\log^* n)$

*“Oh but doing it
for **this** case is of
course trivial...”*

But what are
other cases in which
algorithm design &
lower-bound proofs
can be automated?

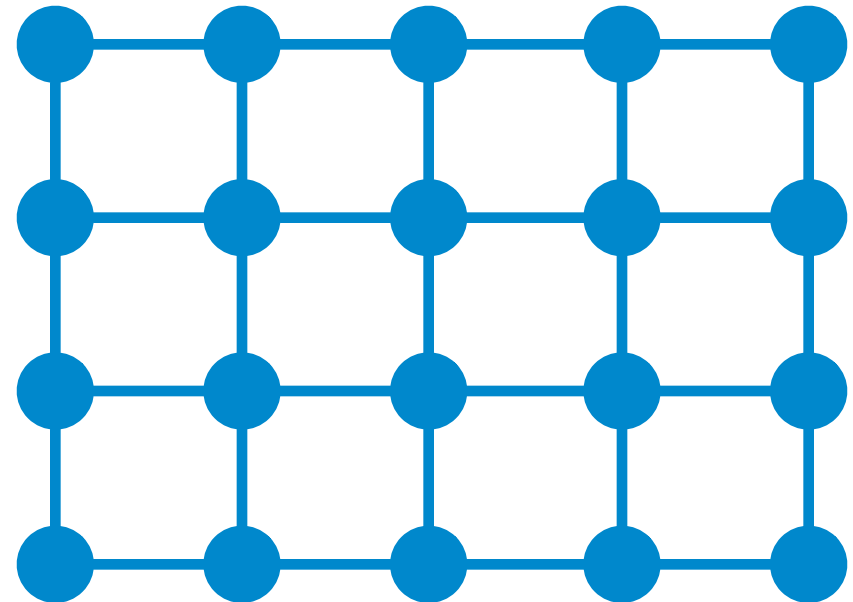
Cycles, paths



Cycles, paths



Grids

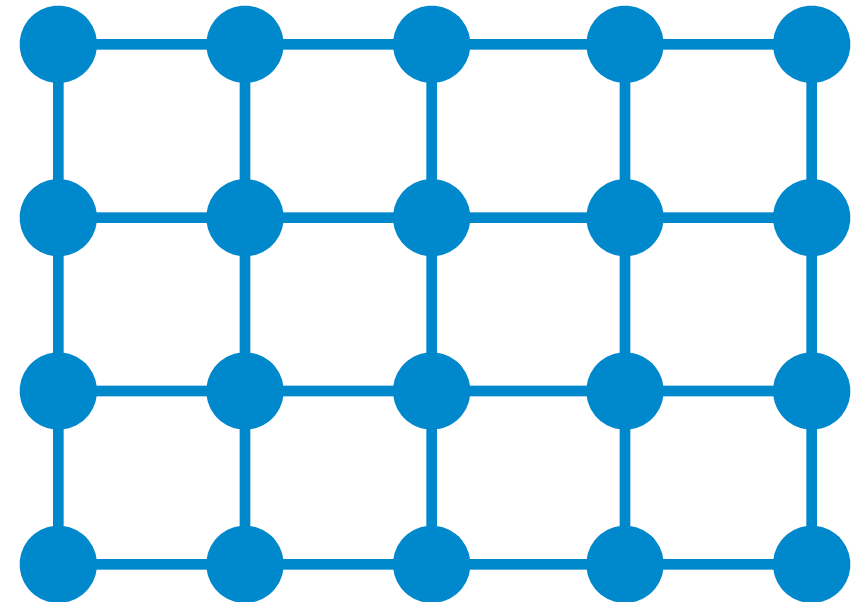


Cycles, paths

solution \approx
execution history of
a **finite automaton**



Grids



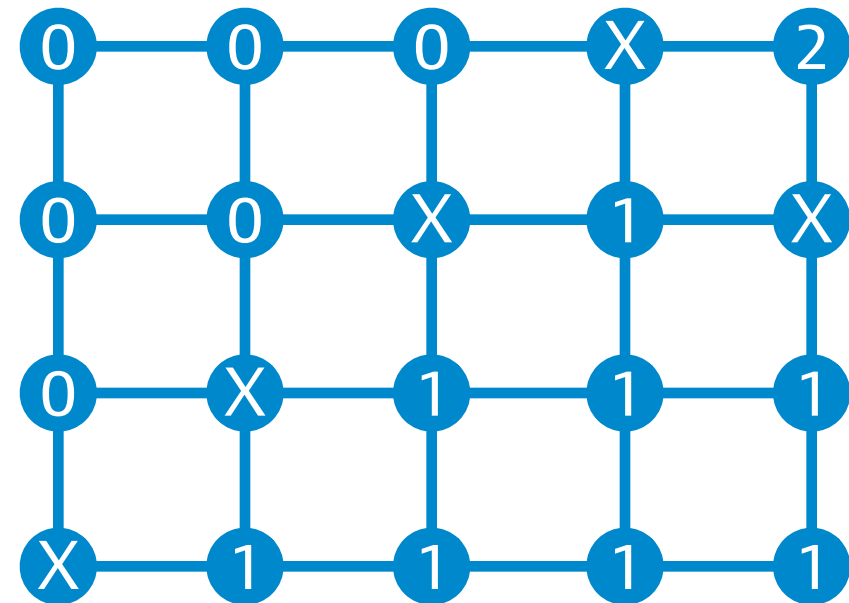
Cycles, paths

solution \approx
execution history of
a **finite automaton**



Grids

solution \approx
execution history of
a **Turing machine**



Cycles, paths

solution \approx
execution history of
a **finite automaton**

Many questions
(efficiently)
decidable

Grids

solution \approx
execution history of
a **Turing machine**

Cycles, paths

solution \approx
execution history of
a **finite automaton**

Many questions
(efficiently)
decidable

Grids

solution \approx
execution history of
a **Turing machine**

Many questions
undecidable

Undecidable

\neq

hopeless

Normal forms

Any algorithm \mathbf{A} that solves a locally checkable problem X fast can be written as $\mathbf{A} = \mathbf{B} \circ \mathbf{C}_k$

- \mathbf{C}_k = distance- k coloring
- \mathbf{B} = finite function that maps colored neighborhoods to local outputs

Normal forms

Any algorithm \mathbf{A} that solves a locally checkable problem X fast can be written as $\mathbf{A} = \mathbf{B} \circ \mathbf{C}_k$

- \mathbf{C}_k = distance- k coloring
- \mathbf{B} = finite function that maps colored neighborhoods to local outputs

"Fast" = e.g. $O(\log^* n)$

Normal forms

Any algorithm A that solves a locally checkable problem X fast can be written as $A = B \circ C_k$

- C_k = distance- k coloring
- B = finite function that maps colored neighborhoods to local outputs

Proof idea: Coloring \approx locally unique identifiers. If A fails with such fake identifiers, it also fails in some small graph with some real identifiers.

Normal forms

Any algorithm A that solves a locally checkable problem X fast can be written as $A = B \circ C_k$

- C_k = distance- k coloring
- B = finite function that maps colored neighborhoods to local outputs

For each $k = 1, 2, 3, \dots$:

- check all possible candidate functions B
- if any of them is good \rightarrow fast algorithm found!

Normal forms

Any algorithm A that solves a locally checkable problem X fast can be written as $A = B \circ C_k$

- C_k = distance- k coloring
- B = finite function that maps colored neighborhoods to local outputs

For each $k = 1, 2, 3, \dots$:

- check all possible candidate functions B
- if any of them is good \rightarrow fast algorithm found!

Normal forms

Any algorithm **A** that solves a locally checkable problem $\forall f$ can be written as $\mathbf{A} = \mathbf{B} \circ \mathbf{C}_k$

- $\mathbf{C}_k =$
- $\mathbf{B} =$

Finite computation for a given candidate B :
no worries about the halting problem

For each $k = 2, 3, \dots$

- check all possible candidate functions B
- if any of them is good \rightarrow fast algorithm found!

Normal forms

Undecidability:
*don't know when to stop if
fast algorithms don't exist*

gives a locally checkable
written as $A = B \circ C_k$

neighborhood function that maps colored
neighborhoods to local outputs

For each $k = 1, 2, 3, \dots$:

- check all possible candidate functions B
- if any of them is good \rightarrow fast algorithm found!

Normal forms

Any algorithm A that solves a locally checkable problem X fast can be written as $A = B \circ C_k$

- C_k = distance- k coloring
- B = finite function on k -neighborhoods

Computational complexity:
typically doubly-exponential in k

For each $k = 1, 2, 3, \dots$

- check all possible candidate functions B
- if any of them is good \rightarrow fast algorithm found!

Sometimes doable!

- Natural problems often solvable with a *small k*

Sometimes doable!

- Natural problems often solvable with a *small k*
- We can make it more feasible in practice:
 - more "*compact*" *normal forms*,
e.g. distance- k coloring \rightarrow ruling set

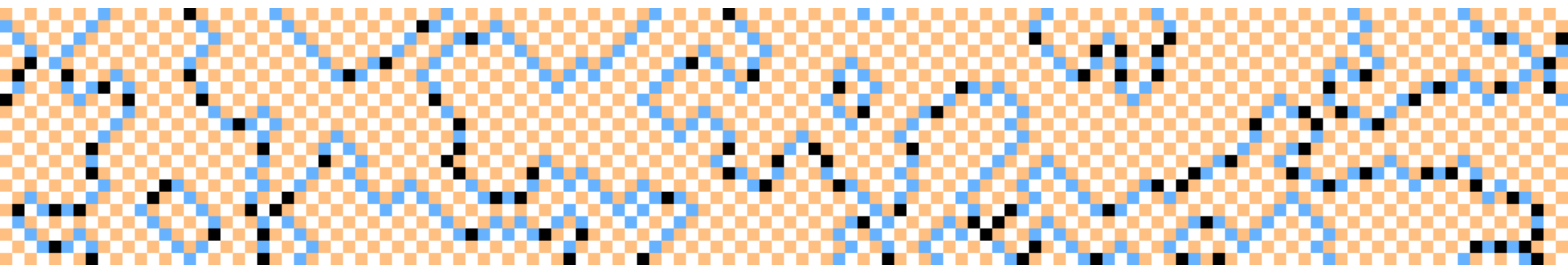
Sometimes doable!

- Natural problems often solvable with a *small k*
- We can make it more feasible in practice:
 - more "*compact normal forms*",
e.g. distance- k coloring \rightarrow ruling set
 - represent "*candidate B is good for this value of k* "
as a Boolean formula and use modern *SAT solvers*
to find such a B

Sometimes doable!

- Example: ***4-coloring in grids***
- Computers were much faster than human beings in figuring out that this is solvable in $O(\log^* n)$ rounds

[Brandt et al., PODC 2017]



Cycles, paths

solution \approx
execution history of
a **finite automaton**

Many questions
(efficiently)
decidable

Grids

solution \approx
execution history of
a **Turing machine**

Many questions
undecidable
(but there is hope!)

Cycles, paths

solution \approx
execution history of
a **finite automaton**

Grids + beyond

solution \approx
execution history of
a **Turing machine**

**Bad news apply to
any graph family that
contains large grids**

Cycles, paths

solution \approx
execution history of
a **finite automaton**

Grids + beyond

solution \approx
execution history of
a **Turing machine**

What is here
between paths
and grids?

Cycles, paths

solution \approx
execution history of
a **finite automaton**

Grids + beyond

solution \approx
execution history of
a **Turing machine**

Trees

Bounded treewidth

High girth

Cycles, paths

solution \approx
execution history of
a **finite automaton**

Grids + beyond

solution \approx
execution history of
a **Turing machine**

lots of open questions,
no known obstacles!

Trees
Bounded treewidth
High girth

Cycles, paths

solution \approx
execution history of
a **finite automaton**

Grids + beyond

solution \approx
execution history of
a **Turing machine**

some positive results
already known
(ask me, happy to tell more!)



Trees

Bounded treewidth
High girth

Big picture: **towards
meta-computational
research questions**

Meta questions

- **Traditional questions:** what is the best distributed algorithm for solving problem X ?
- **Meta-computational questions:** can we design an (efficient) *meta-algorithm* that finds the best distributed algorithm for *any problem X* in some problem family F ?

Classification

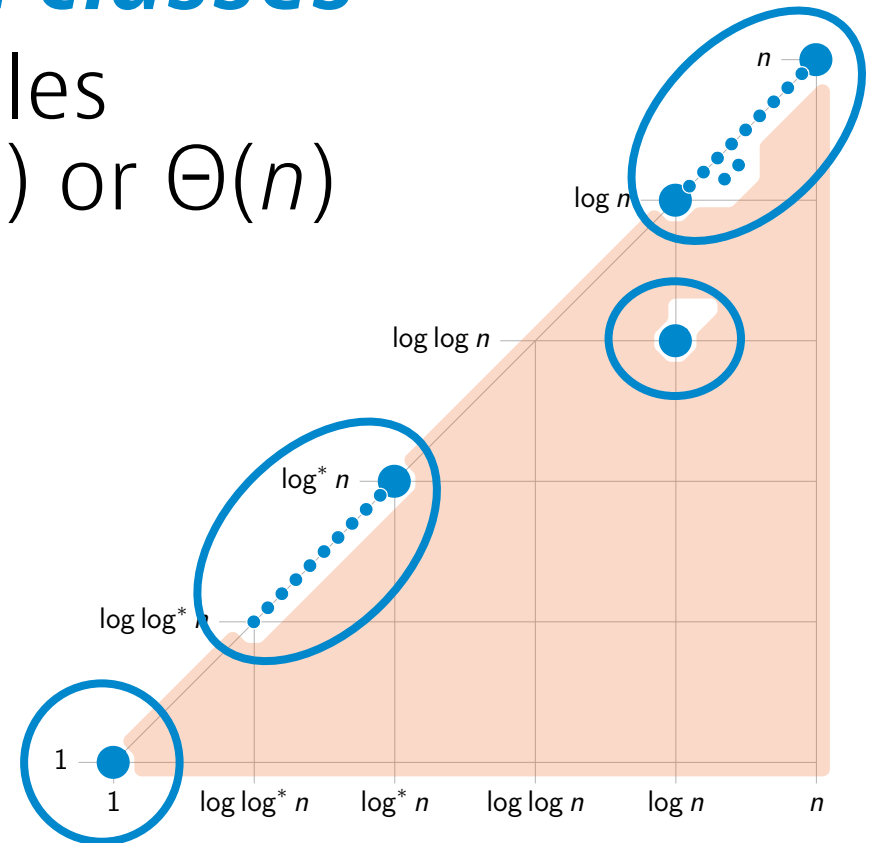
- **Classification:** *“Any problem in this family belongs to one of these classes”*

Classification

- **Classification:** *“Any problem in this family belongs to one of these classes”*
 - locally checkable problems in cycles have complexity $O(1)$ or $\Theta(\log^* n)$ or $\Theta(n)$

Classification

- **Classification:** *“Any problem in this family belongs to one of these classes”*
 - locally checkable problems in cycles have complexity $O(1)$ or $\Theta(\log^* n)$ or $\Theta(n)$
 - locally checkable problems in general graphs belong to one of four broad classes



Classification

- **Classification:** *“Any problem in this family belongs to one of these classes”*
 - locally checkable problems in cycles have complexity $O(1)$ or $\Theta(\log^* n)$ or $\Theta(n)$
- **Meta-algorithms:** *“Here is an efficient algorithm for determining the class of any given problem”*

Classification

- **Classification:** *“Any problem in this family belongs to one of these classes”*
 - locally checkable problems in cycles have complexity $O(1)$ or $\Theta(\log^* n)$ or $\Theta(n)$
- **Meta-algorithms:** *“Here is an efficient algorithm for determining the class of any given problem”*



Classification

- Computers can help with classification, too!

Classification

- Computers can help with classification, too!
- Classify a *finite sub-family of problems*, automate as much work as possible
 - e.g. bounded alphabet size, bounded degree

Classification

- Computers can help with classification, too!
- Classify a *finite sub-family of problems*, automate as much work as possible
 - e.g. bounded alphabet size, bounded degree
- Identify interesting nontrivial problems
 - e.g. where computers fail

Classification

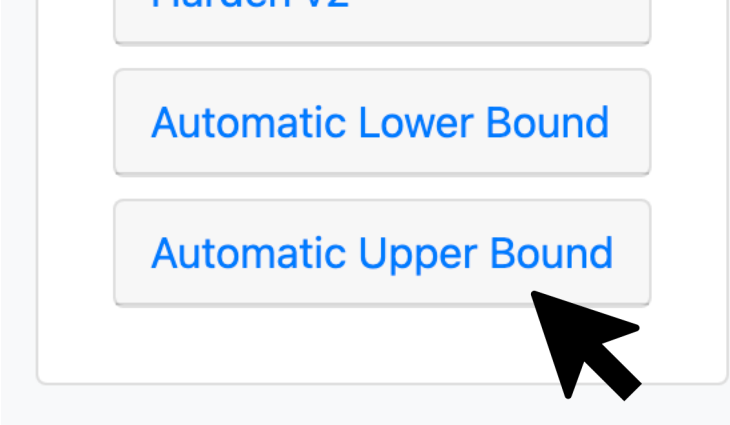
- Computers can help with classification, too!
- Classify a *finite sub-family of problems*, automate as much work as possible
 - e.g. bounded alphabet size, bounded degree
- Identify interesting nontrivial problems
 - e.g. where computers fail
- Detect patterns, generalize

Classification

- Computers can help with classification, too!
- Classify a ***finite sub-family of problems***,
automate as much work as possible
 - e.g. bounded alphabet size, bounded degree
- Identify interesting nontrivial problems
 - e.g. where con... fail
- Detect p...

How? Are there general techniques we can apply without much thinking?

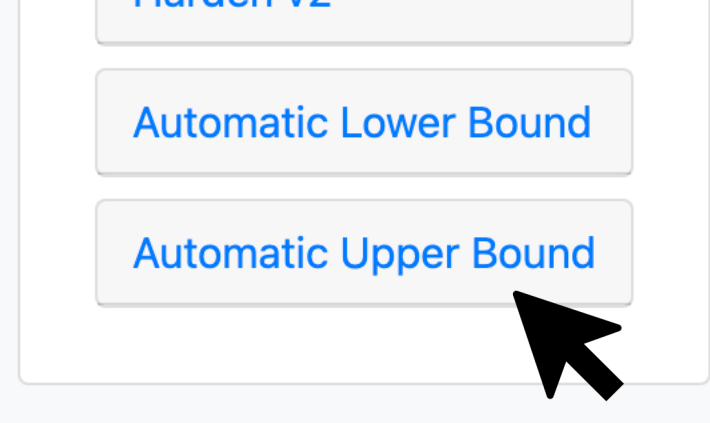
General techniques



- Example: *round elimination* technique
 - github.com/olidennis/round-eliminator
 - applicable to any locally checkable problem

[Brandt, PODC 2019]
[Olivetti, PODC 2020]

General techniques



- Example: *round elimination* technique
 - github.com/olidennis/round-eliminator
 - applicable to any locally checkable problem
- Does not always work — but when it works, you get algorithms and/or lower bound proofs for free!

[Brandt, PODC 2019]
[Olivetti, PODC 2020]

Success stories

- Lower bound for maximal matching and maximal independent set
- ***Six people and one computer program***
 - enabled rapid hypothesis testing and exploration of possible proof strategies

[Brandt et al., FOCS 2019]

Conclusions

Take-home messages

- *Can we automate our own work?*

Take-home messages

- *Can we automate our own work?*
 - **yes** — some questions on theory of distributed computing can be solved automatically!

Take-home messages

- *Can we automate our own work?*
 - **yes** — some questions on theory of distributed computing can be solved automatically!
 - known obstacles, tons of open questions

Take-home messages

- ***Can we automate our own work?***
 - **yes** — some questions on theory of distributed computing can be solved automatically!
 - known obstacles, tons of open questions
- ***Opportunities for human–computer collaboration!***

Take-home messages

- ***Can we automate our own work?***
 - **yes** — some questions on theory of distributed computing can be solved automatically!
 - known obstacles, tons of open questions
- ***Opportunities for human–computer collaboration!***
 - theory researchers who can write programs are going to have a competitive edge!