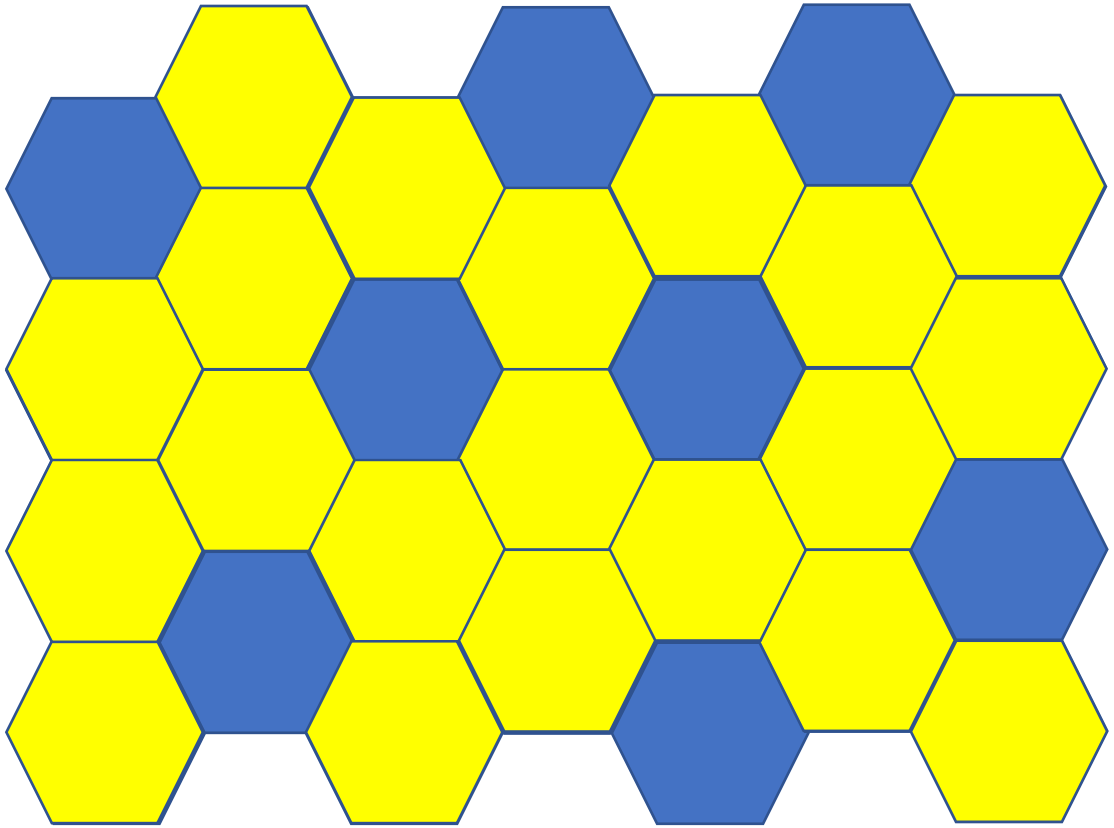
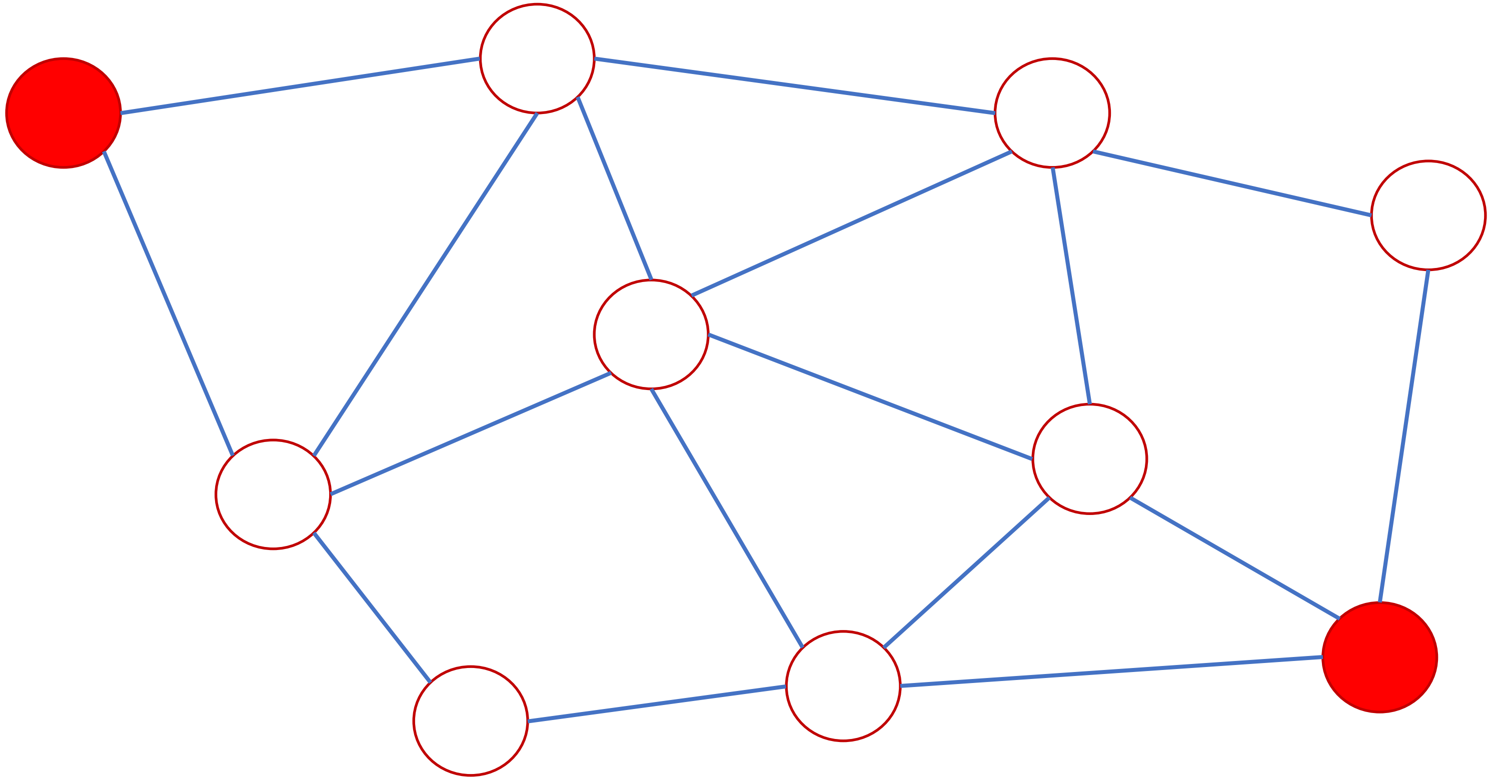
# Maximal Independent Set

Not Maximum… ☺



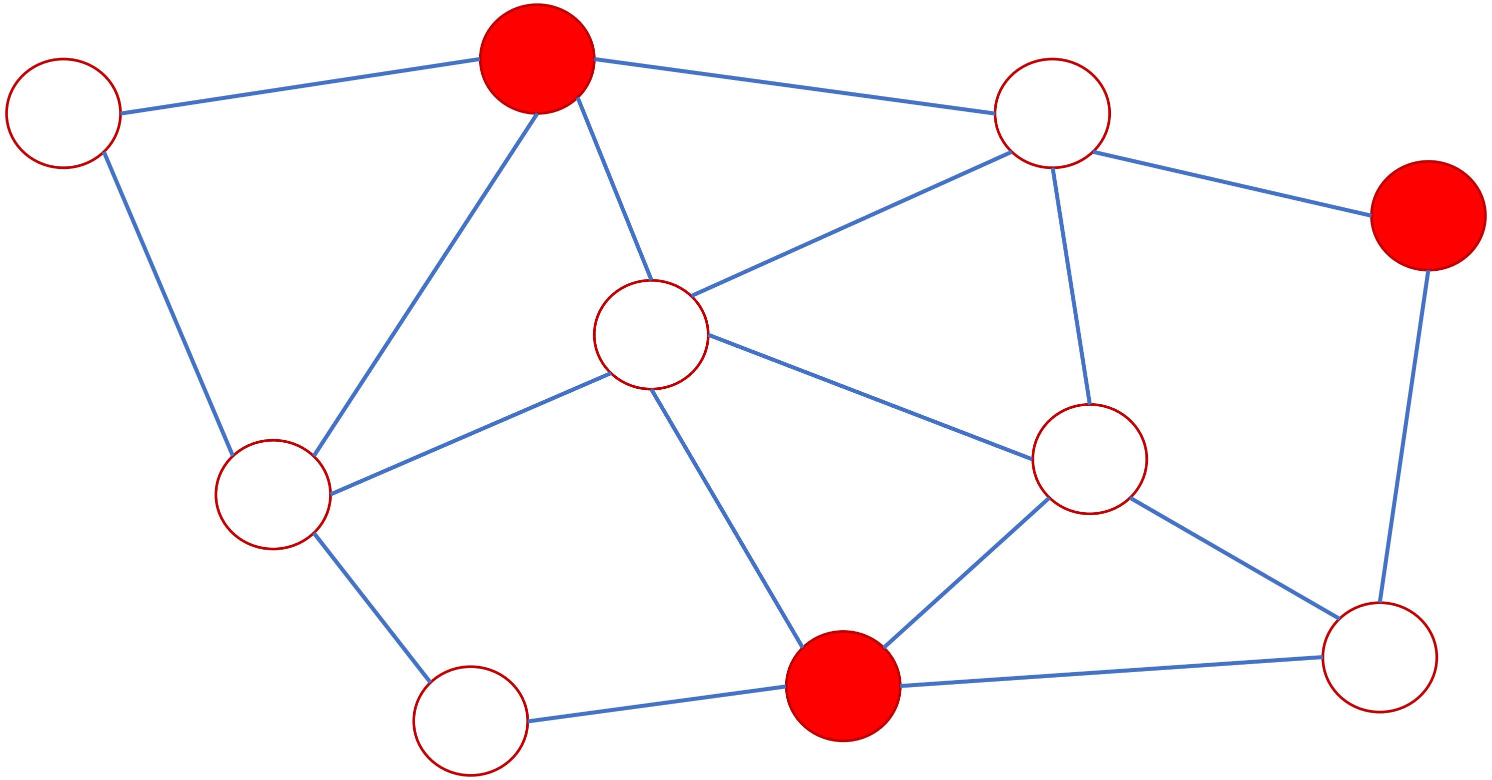
# Independent Set

Any set of nodes that are not adjacent



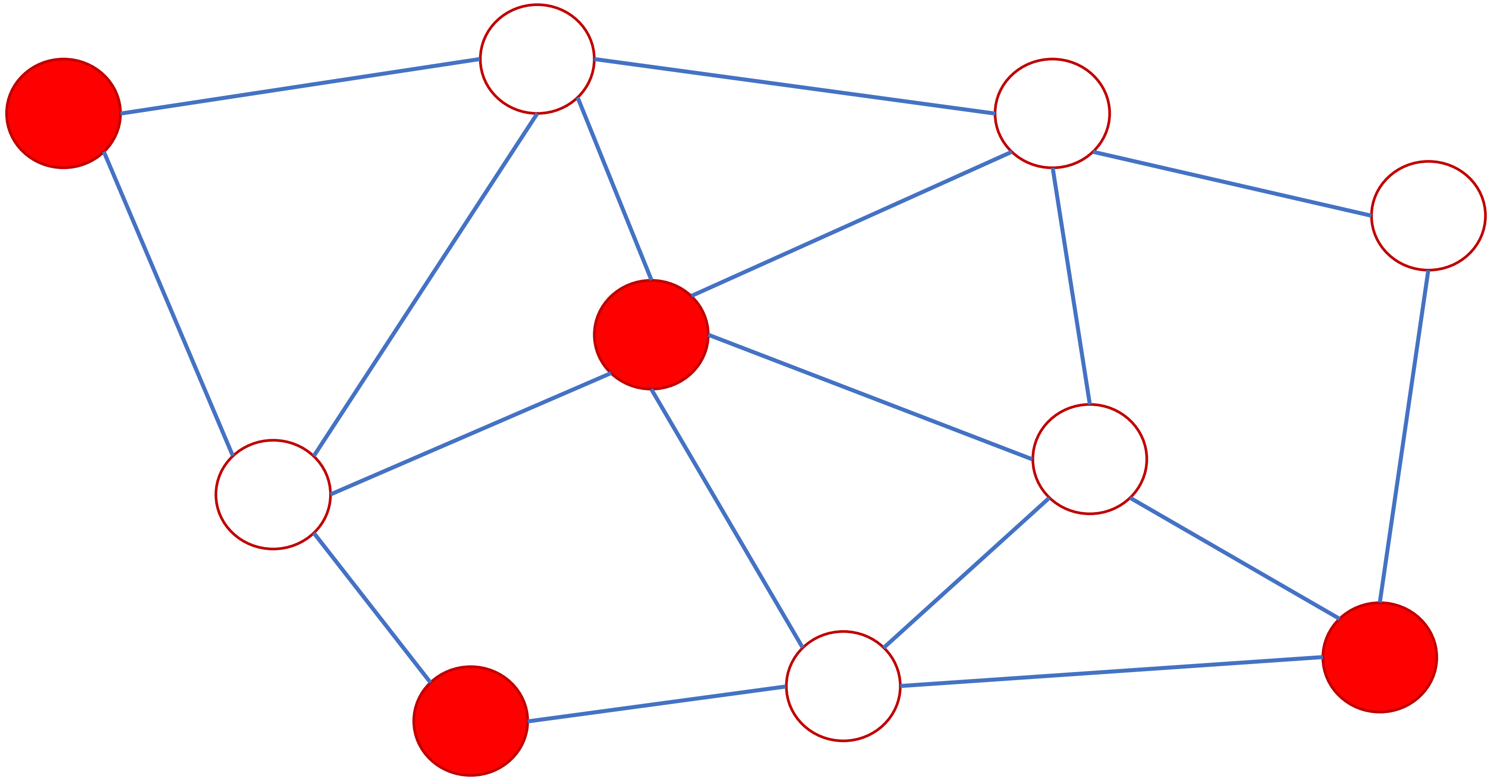
# Maximal Independent Set

An independent set that is no proper subset of another independent set



# Maximum Independent Set

This is an NP-Complete Problem



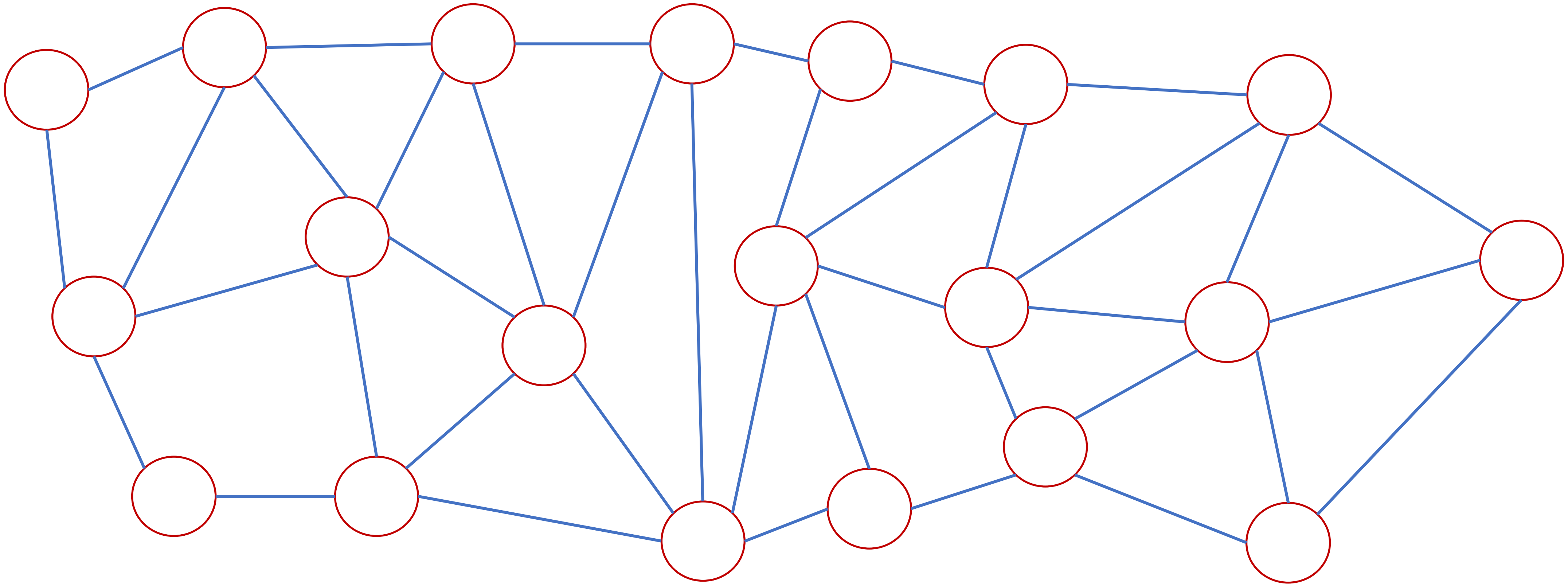
# Applications in Distributed Systems

- In a network graph consisting of nodes representing processors, a MIS defines a set of processors which can operate in parallel without interference
- For instance, in wireless ad hoc networks, to avoid interference, a conflict graph is built, and a MIS on that defines a clustering of the nodes enabling efficient routing
- (MAIN APPLICATION) Basic building block in many distributed algorithms like Leader Election

# A Sequential Greedy algorithm

Let $I$ contain the MIS. Initially $I = \emptyset$

Iteratively, pick a node and if possible, add it to $I$



# A Deterministic Distributed Algorithm For Computing MIS

Same as the sequential greedy algorithm, but  at each phase we may select any independent set

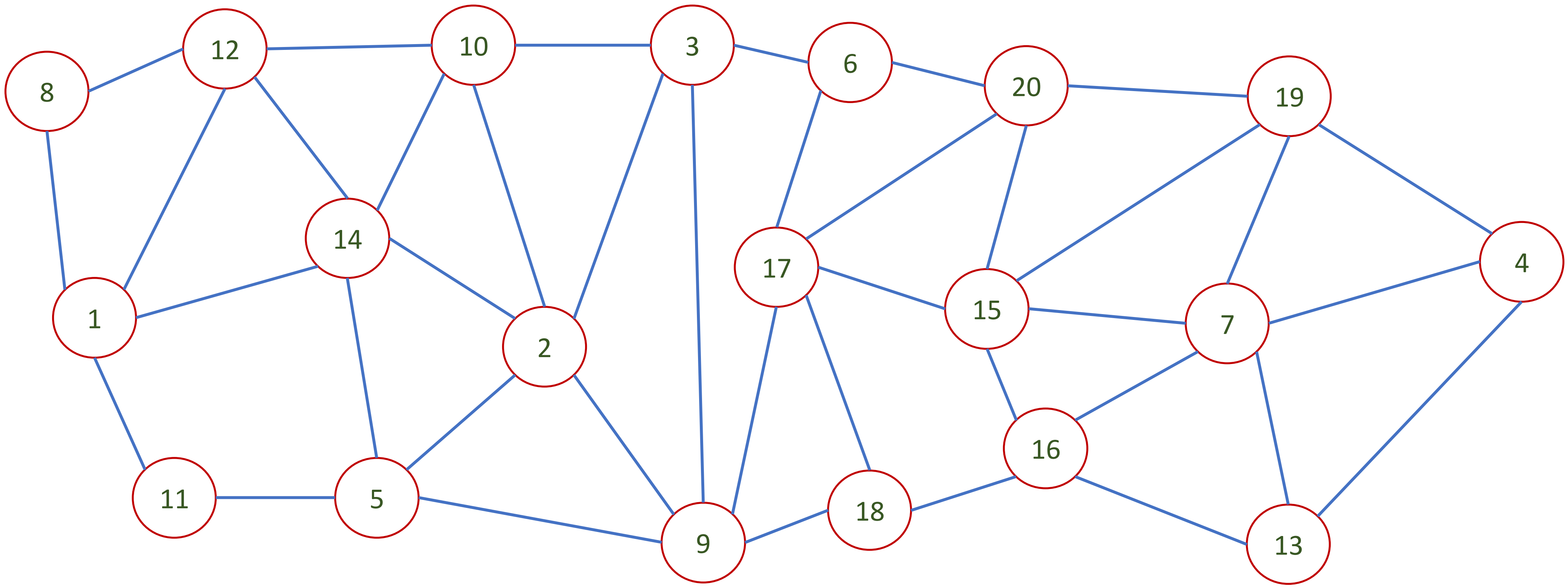
(instead of a single node)

How are we going to choose an independent set?

LOCAL model: use IDs

# Example

A node is getting into MIS if it has the highest ID among its current neighbors.



# Observations

The number of phases depends on the choice of independent set in each phase:

- The larger the independent set at each phase the faster the algorithm

When does the algorithm degenerates to the sequential algorithm?

# What if no IDs are Available?

There is no deterministic algorithm for computing an MIS on an anonymous ring network, assuming all nodes wake up simultaneously.

Proof Sketch:

1. No external input: All nodes start from the same state.
2. Inductively: at round $t$ all nodes send and receive the same messages
3. Due to the symmetry of the topology, all nodes will be at the same state at the end of round $t$.
4. If the algorithm terminates, then all nodes have the same state, and thus either all of them belong in the MIS or none.

# Connections to Coloring

- Each color class is an independent set
  - *But maybe not a maximal one*

Algorithm for MIS (assume we have a coloring $\{1,2,\ldots,c\}$ of the nodes):

1. First choose all nodes of color $i = 1$.
2. Iterate for each color $i = i + 1$:
   1. Add without conflict as many nodes as possible (in parallel) from color $i$

Given a coloring algorithm that needs $c$ colors and runs in $T$ rounds, we can construct an MIS in $c + T$ rounds.

# A Randomized Synchronous Distributed Algorithm

LOCAL model: Knowledge of the Max Degree

Same as before, with the exception that now the independent set in each phase is chosen probabilistically.

# The Algorithm

**Assumption:**

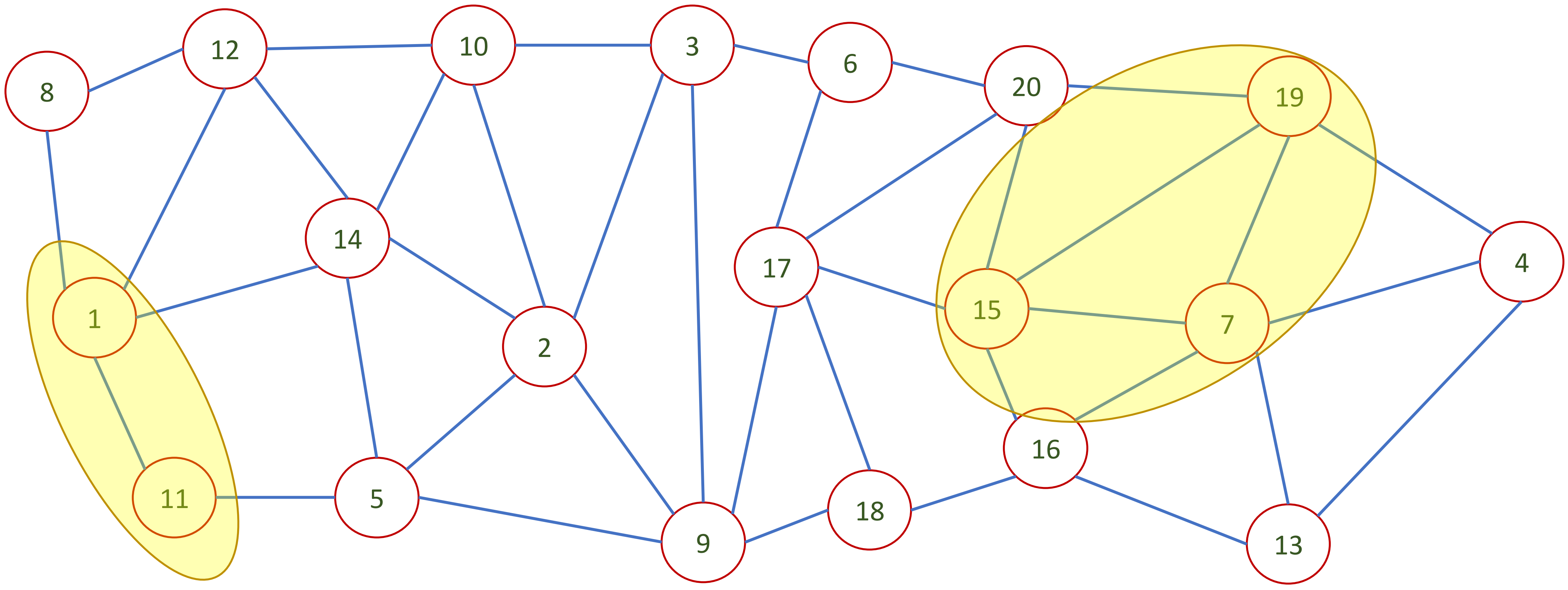
Let $\Delta$ be the max node degree in the graph. All nodes know $\Delta$.

At round $k$:

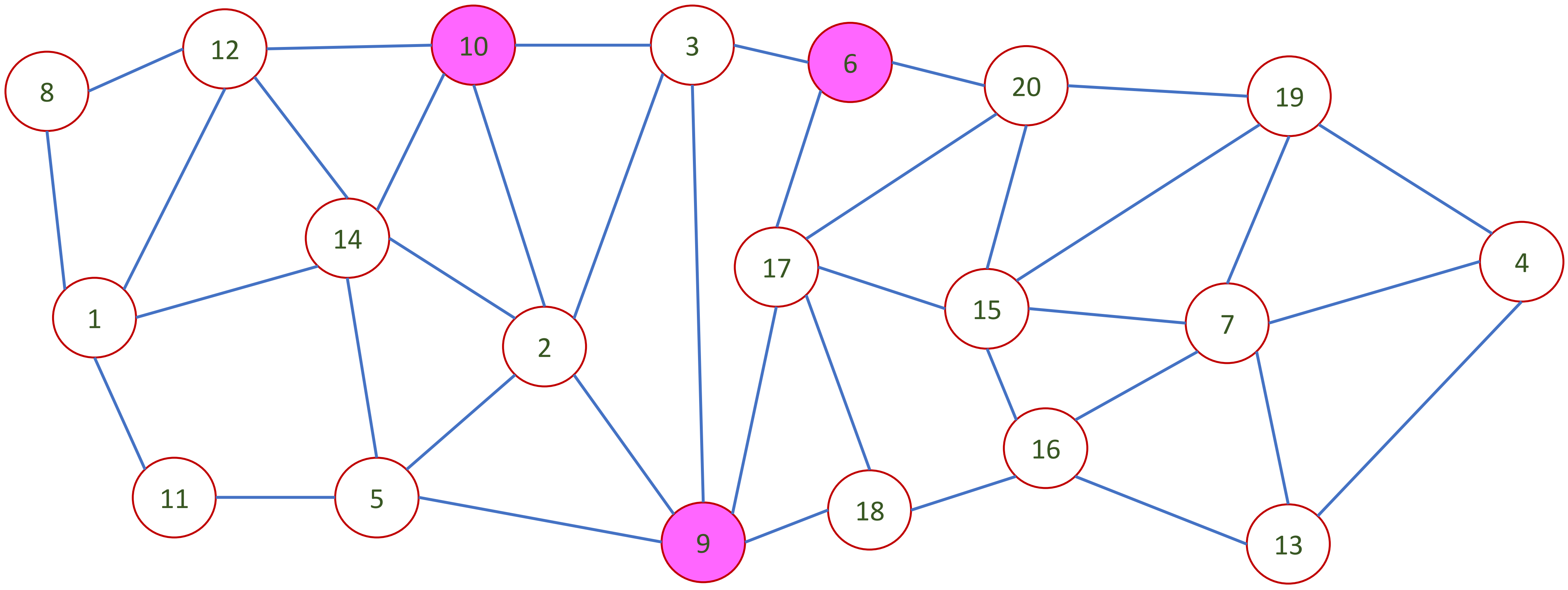
1. Each node in the graph elects itself as a candidate for MIS with probability $p = \frac{1}{\Delta}$.
2. The set of elected nodes that are adjacent, are un-elected.
3. All remaining elected nodes are added to the MIS and the graph $G$ is updated.

# Example

Unelect the conflicting nodes.

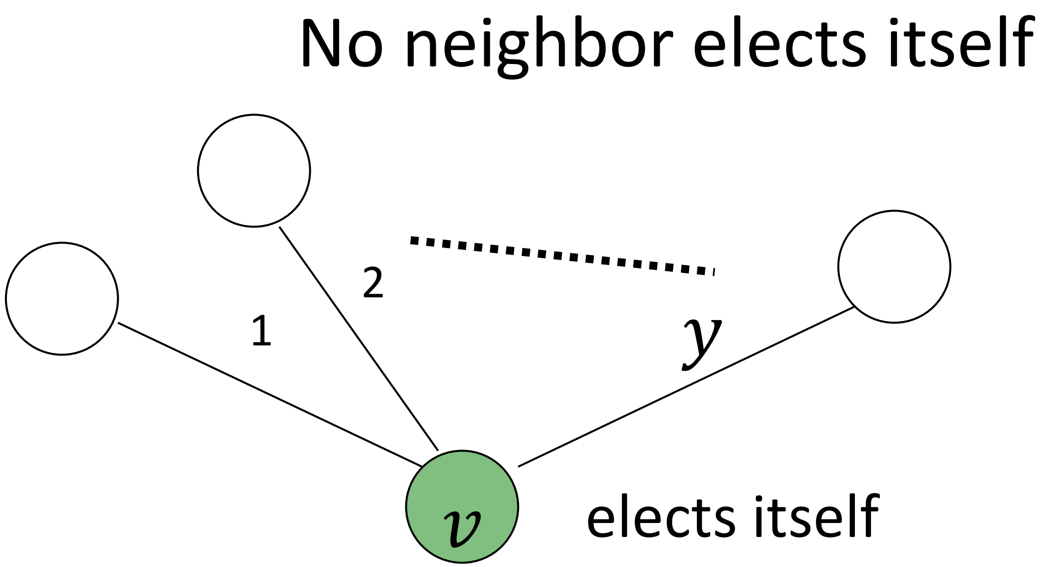


# Example



# Analysis ☹

Success for a node $v$ in round $k$: $v$ becomes elected (in the IS) or un-elected (neighbor of an elected)




# Analysis – Fundamental Inequalities

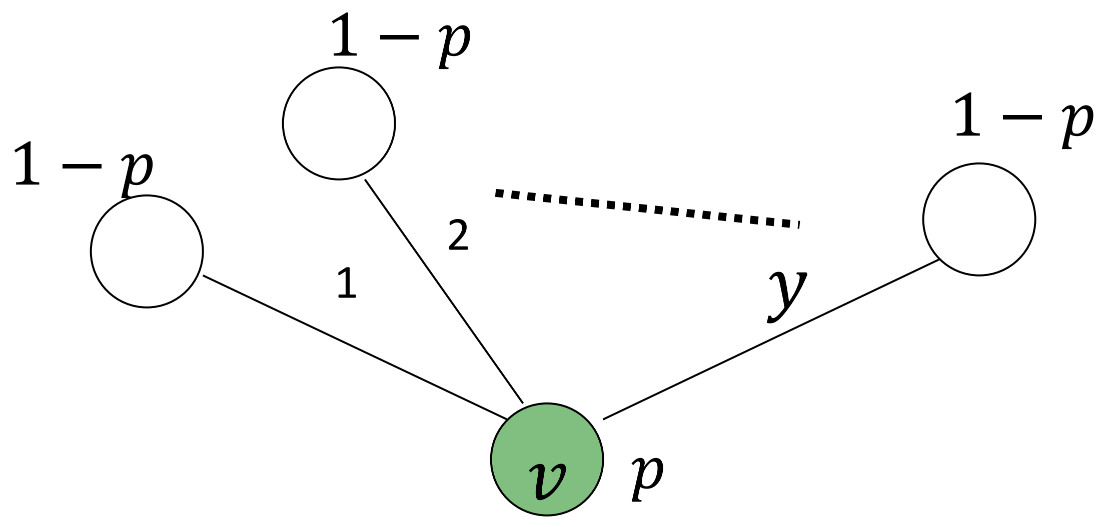
$$For\ n \geq 1, |t| \leq n: \quad e^t\left(1-\frac{t^2}{n}\right) \leq \left(1+\frac{t}{n}\right)^n \leq e^t$$

$$For\ 0 < p \leq 1, k \geq 1: \quad 1-\mathrm{p} \leq \left(1-\frac{p}{k}\right)^k$$

# Probability of Success in a Round

At least $p(1-p)^y \geq p(1-p)^\Delta = \frac{1}{\Delta}\left(1-\frac{1}{\Delta}\right)^\Delta$

By first inequality: $\frac{1}{\Delta}\left(1-\frac{1}{\Delta}\right)^\Delta \geq \frac{1}{e\Delta}\left(1-\frac{1}{\Delta}\right) \geq \frac{1}{2e\Delta}$, for $\Delta \geq 2$




$For\ n \geq 1, |t| \leq n$: $e^t\left(1-\frac{t^2}{n}\right) \leq \left(1+\frac{t}{n}\right)^n \leq e^t$

# An Expected Bound

What is the expected number of rounds for $v$ to get into the IS?

$$\leq \frac{1}{probability\ of\ success\ in\ round} = 2e\Delta\ rounds$$

# Bad Event for One Node in $G$

After $4e\Delta \ln n$ rounds, node $v$ did not become inactive (elected or neighbor to an elected)

$$\left(1 - \frac{1}{2e\Delta}\right)^{4e\Delta \ln n} \leq \frac{1}{e^{2 \ln n}} = \frac{1}{n^2}$$

$For\ n \geq 1, |t| \leq n$: $\quad e^t\left(1 - \frac{t^2}{n}\right) \leq \left(1 + \frac{t}{n}\right)^n \leq e^t$

# Bad Event for Any Node in $G$

After $4e\Delta \ln n$ rounds, at least one node did not become inactive (elected or neighbor to an elected)

$$\Pr(bad\ event\ for\ any\ node) \leq \sum_{x \in G} \Pr(bad\ event\ for\ x)$$

$$\leq n\frac{1}{n^2} = \frac{1}{n}$$

(Use of Union Bound in the first inequality)

# Good Event for All Nodes in $G$

Within $4e\Delta \ln n$ rounds, all nodes become inactive (elected or neighbor to an elected)

$$1 - \Pr(bad\ event) \geq 1 - \frac{1}{n}$$

It means that it holds with high probability

# Concluding

Total number of rounds:

$$4e\Delta \ln n = O(\Delta \log n) \textit{ with high probability}$$

Time duration of each round:

$$O(1)$$

## Total Time:

$$O(\Delta \log n)$$

# Luby's MIS Distributed Algorithm

## A Considerable Improvement

At round $k$:

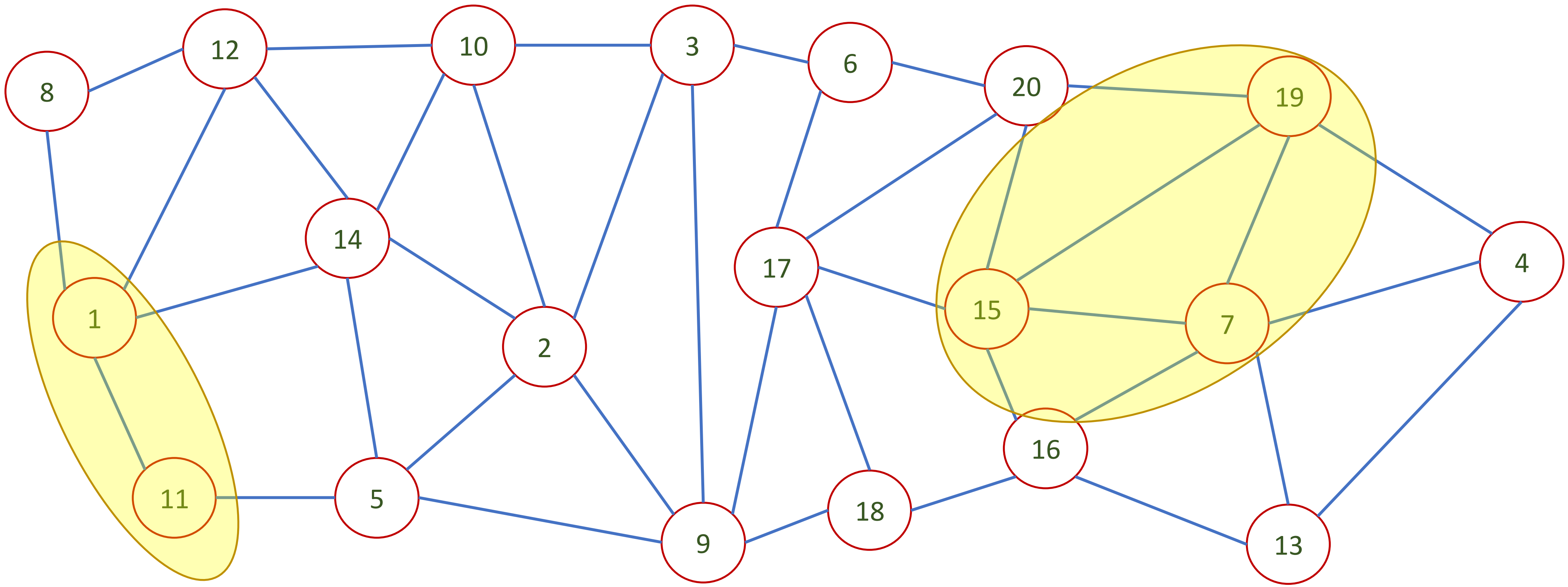
1. Each node $v$ in the graph elects itself as a candidate for MIS with probability $p_v = \frac{1}{2d(v)}$, where $d(v)$ is the current degree of $v$.

2. If two or more elected nodes are adjacent, then the one with the highest degree wins. If many nodes have the same degree, break the ties arbitrarily.

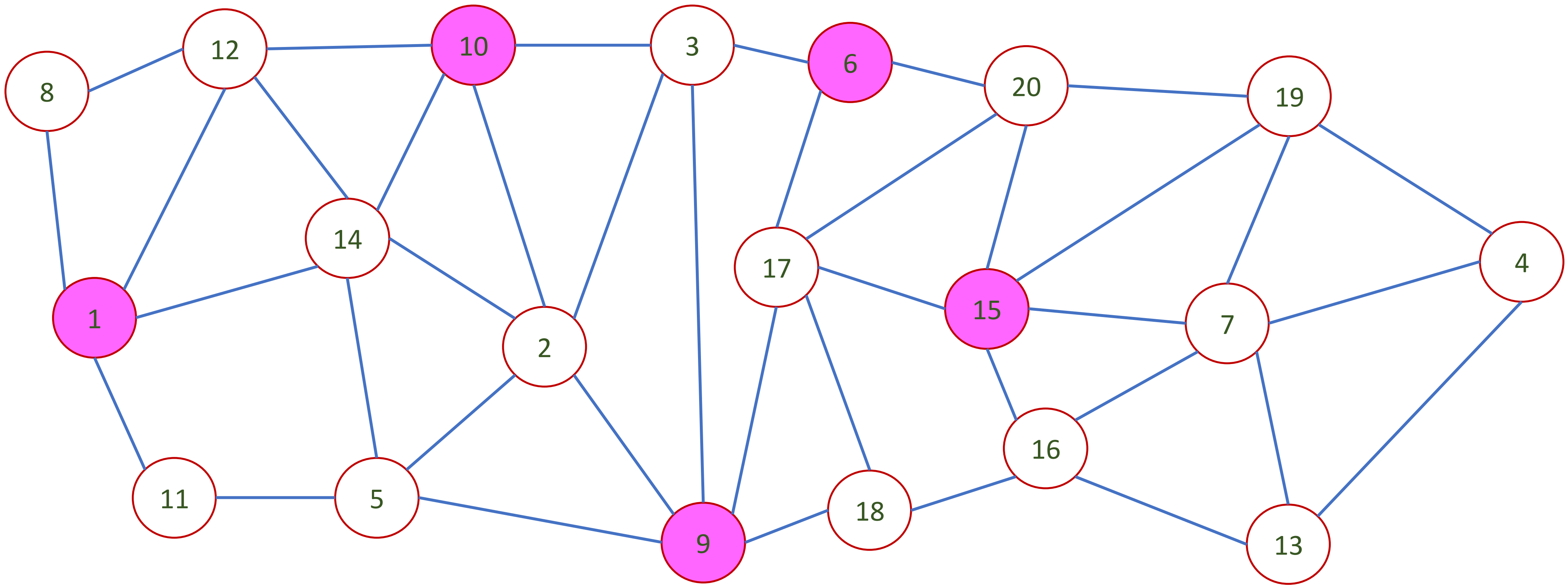
3. All remaining elected nodes are added to the MIS and the graph is updated.

# Example

Resolve conflicts



# Example



# No Analysis ☺

The expected number of rounds is $O(\log n)$

The number of rounds is $O(\log \Delta \log n)$ with high probability, where $\Delta$ is the maximum degree.

# Some Algorithms

| | Knowledge | Time (expected) | Message Size (# of bits) | Bit Complexity (per channel) |
|---|---|---|---|---|
| Simple Randomized | Max degree of the graph | $O(\Delta \log n)$ | 1 | $O(\Delta \log n)$ |
| Luby (Lynch) | Size of the graph | $O(\log n)$ | $\log n$ | $O(\log^2 n)$ |
| Luby (Peleg) | Max Degree in 2-Neighbourhood | $O(\log^2 n)$ | $\log n$ | $O(\log^3 n)$ |
| Luby (Wattenhofer) | Max of Neighbours degrees | $O(\log n)$ | $\log n$ | $O(\log^2 n)$ |
| Alon, Babai and Itai | Max of Neighbours degrees | $O(\log n)$ | $\log n$ | $O(\log^2 n)$ |
| **Follows** | **Size of Graph, Max degree** | $O(\log^2 n)$ | **1** | $O(1)$ |
| **Follows (2)** | **None** | $O(\log n)$ | **1** | $O(1)$ |

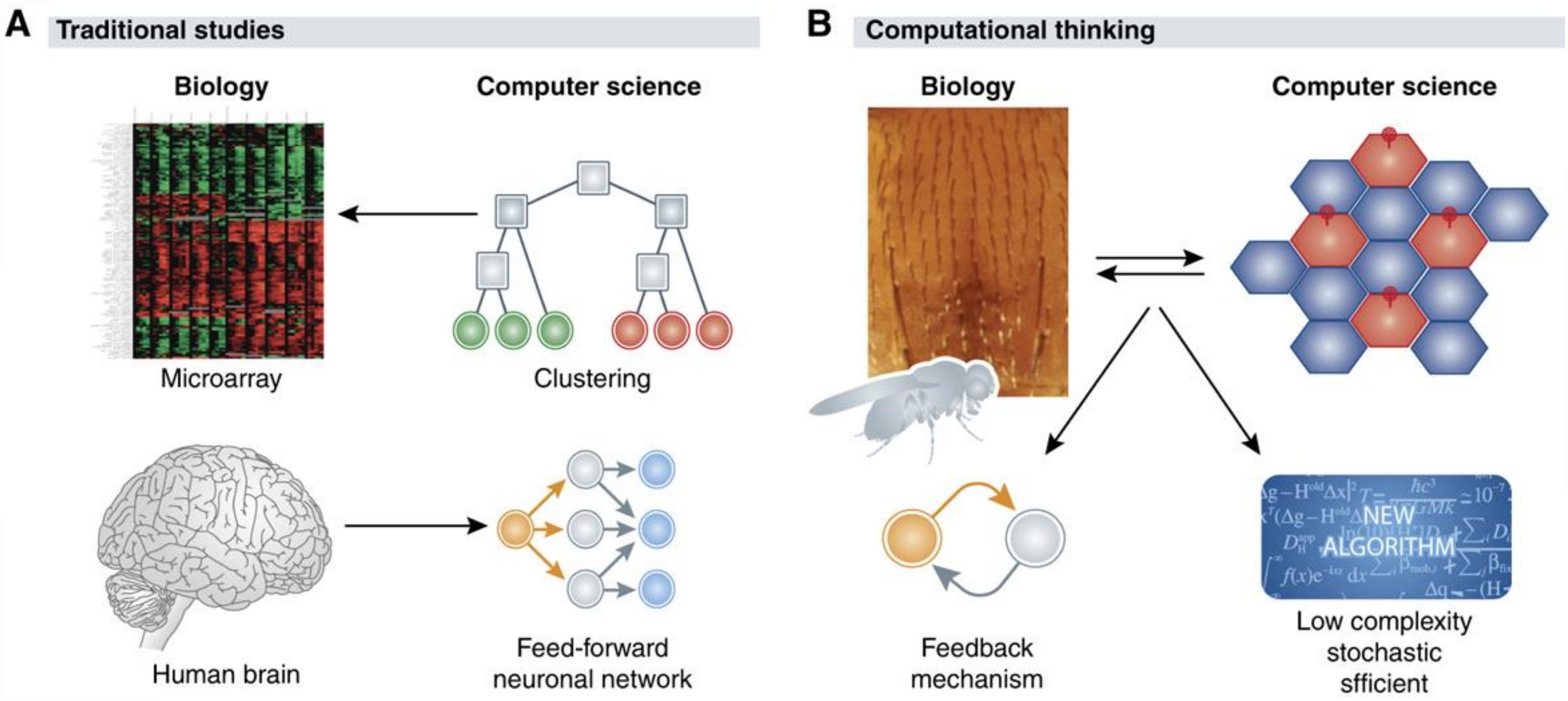



Image taken from http://www.algorithmsinnature.org/

Nature ⟶ Algorithms

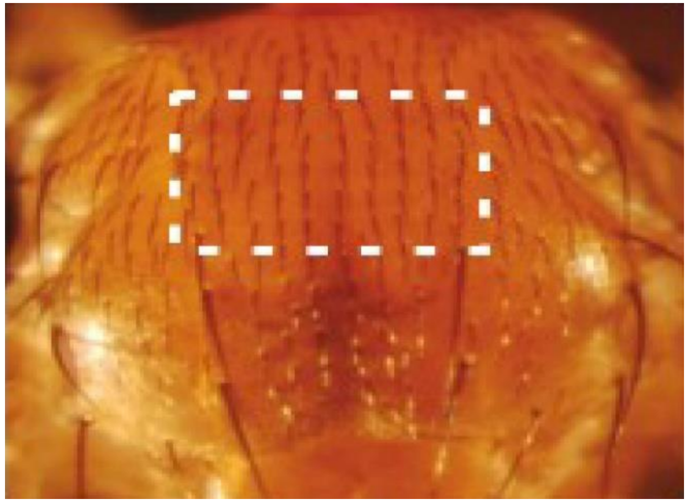
Nature ⟵ Algorithms

# Inspiring Flies…

# A Biological Approach to Distributed MIS(Science 2011)

During development of a fly the Sensory Organ Precursor (SOP) cells are chosen such that:

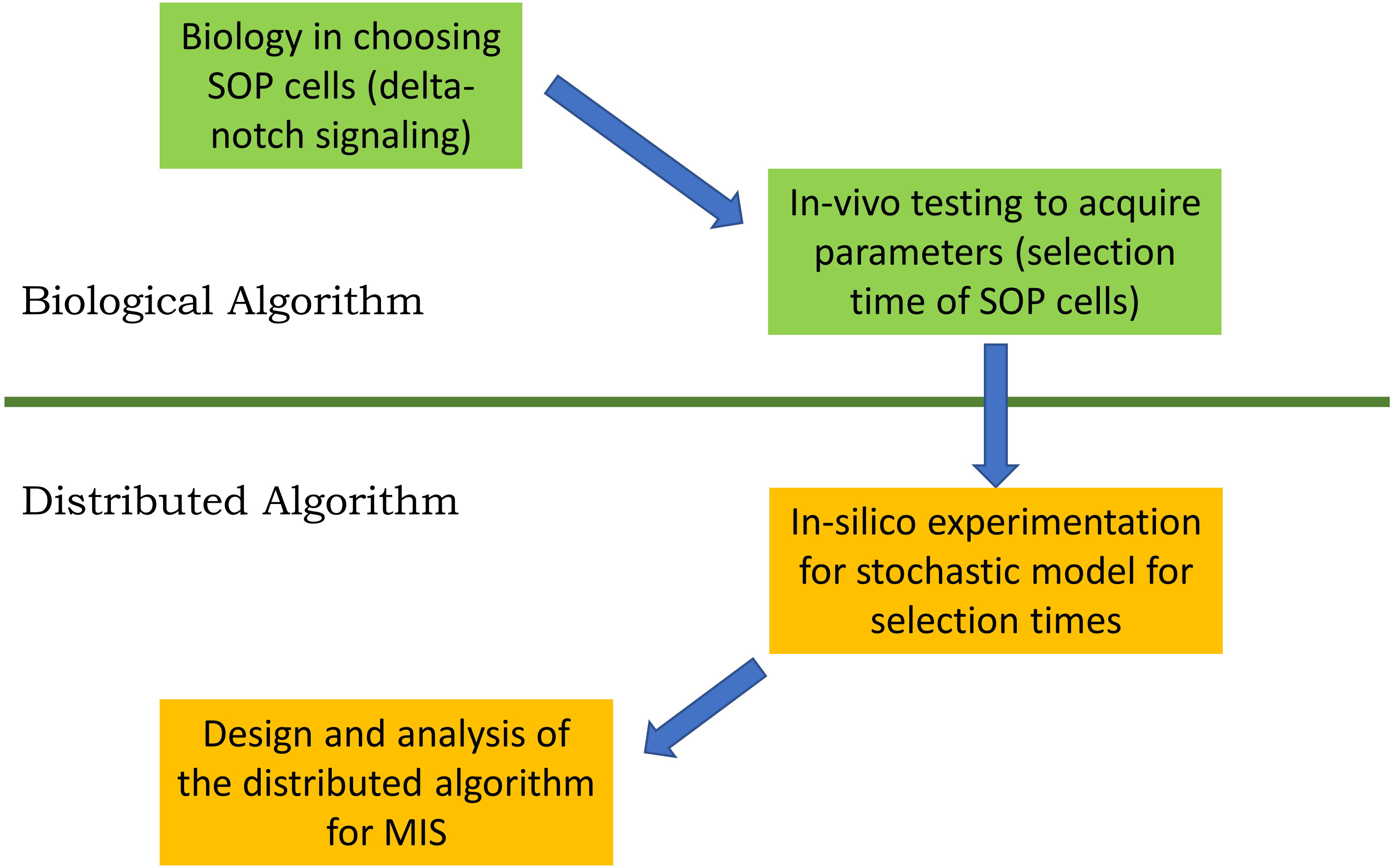
1. All cells are either a SOP or inhibited by a SOP
2. No two SOP cells are adjacent



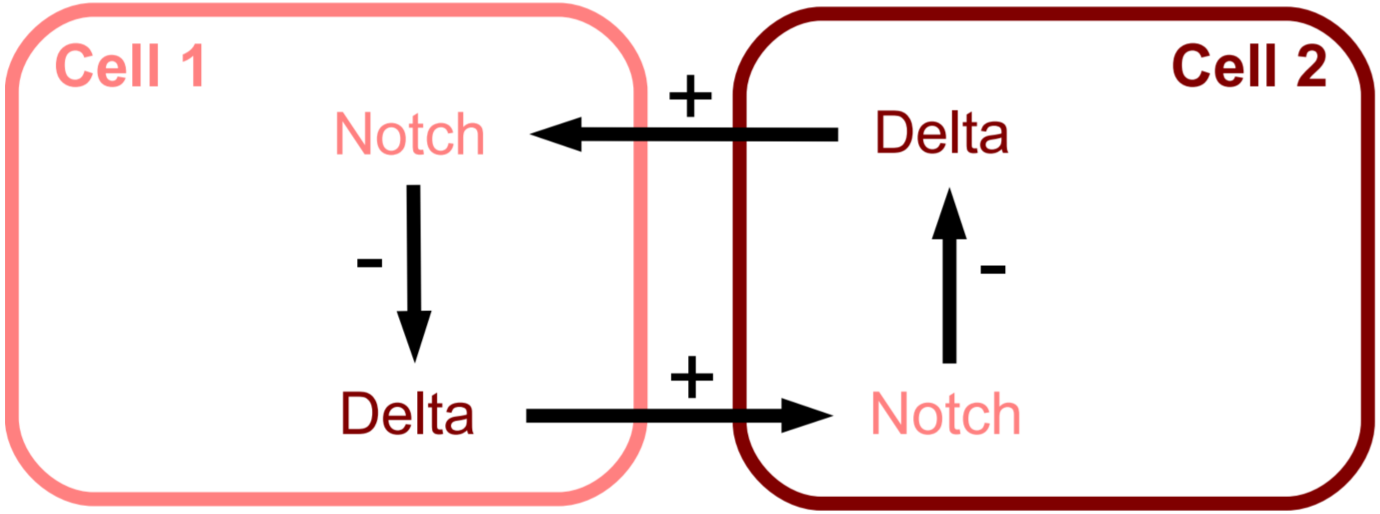
Maximal Independent Set:

1. All nodes are either in the set or connected to a node in the set
2. No two nodes in the set are connected

# What did they do?



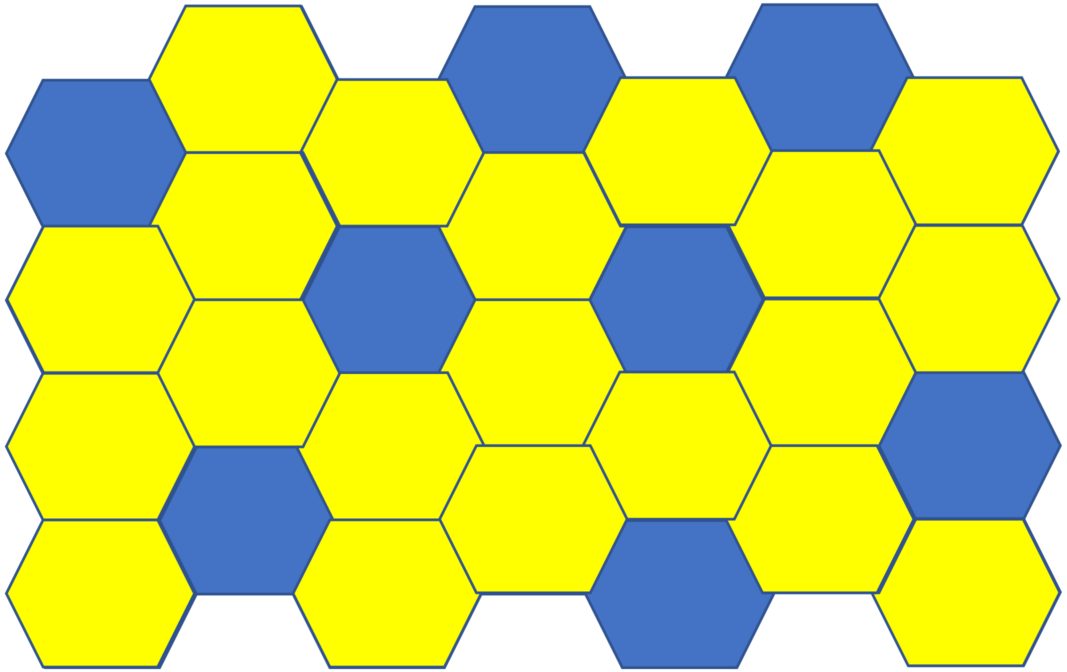
# The Delta-Notch Mechanism




- Provides a communication channel between adjacent cells during development
  - Delta in one cell can bind to, and transactivate, the transmembrane protein Notch in its adjacent cells; Delta and Notch in the same cell mutually inactivate each other
- Generates a switch between two signaling states:
  - Sending (high Delta/low Notch)
  - Receiving (high Notch/low Delta)

# Properties of the Biological System: Topology

Cells cannot distinguish between inhibited cells from cells that have not decided whether they are SOP or not. Similarly, in this model no node can distinguish anything about its neighborhood (active vs inactive, etc.)



Number of neighbors of each cell is bounded

# Properties of the Biological System: Stochasticity

SOP selection is a stochastic process:

- Delta Firing
- Sites that become SOPs

# Properties of the Biological System: Communication

1-bit communication:

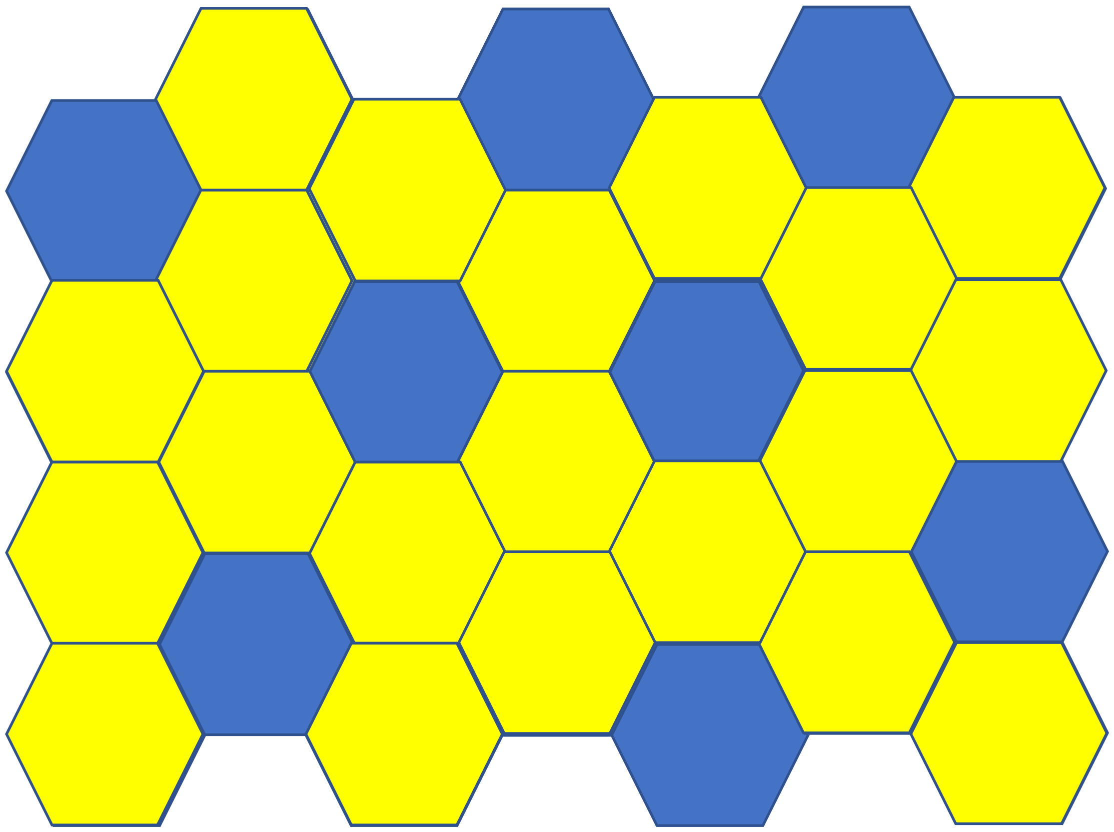
Although the precise biological communication is not fully resolved, many models of this communication are based on a threshold decision and a binary communication process.

# Properties of the Biological System: Synchronicity

The biological system seems to be asynchronous. However, there are some characteristics that justify the use of a synchronous model.

1. Cells start the SOP process based on an external biochemical signal that reaches all cells at around the same time.
2. Messaging through the Delta-Notch signaling has a predictive delay.

# The Core Idea of the Algorithm



# Distributed Biologically-Inspired Synchronous Algorithm for MIS

Algorithm: $MIS(n, \Delta)$ at node $v$

1. For $i = 0: \log \Delta$
    1. For $j = 0: M \log n$

    Exchange 1

    2. $v = 0$
    3. With probability $\frac{1}{2^{\log \Delta - i}}$ broadcast $B$ to neighbors and set $v = 1$
    4. If received message from neighbor, then $v = 0$

    Exchange 2

    5. If $v = 1$ then

        Broadcast $B$; Join $MIS$; exit
    7. else

        If received message $B$ in this exchange, then mark node $v$ inactive; exit the algorithm

+ No knowledge of degree
+ Single Bit Messages
+ Optimal Message Complexity
– $O(\log^2 n)$ rounds
– Knowledge of $n$ (number of nodes) and $\Delta$ (max degree)

# Analysis

## Correcteness

- No two nodes in $MIS$ are connected to each other
- If $w$ exits without joining $MIS$, then it is connected to a node in $MIS$
- If $w$ is in $MIS$, then all neighbors of $w$ become inactive

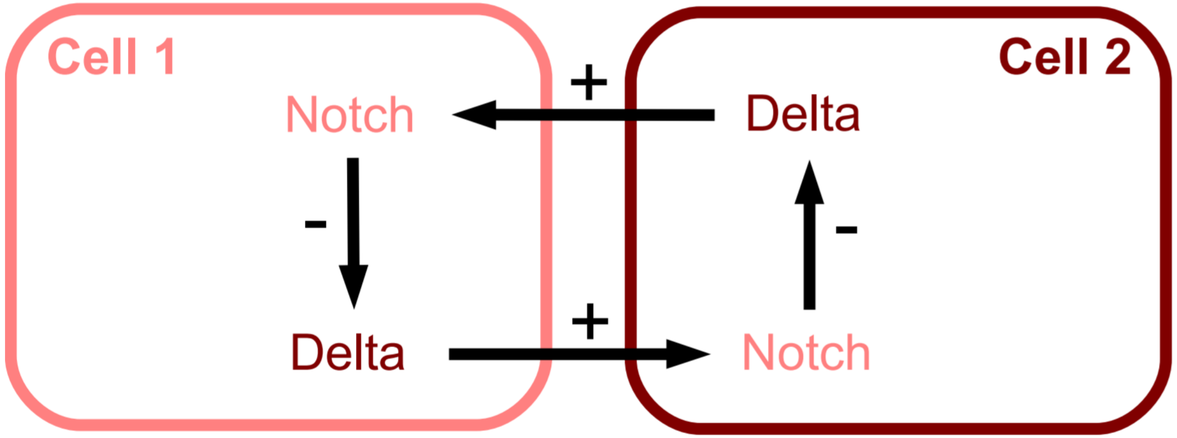
## Runtime

- With probability at least $1 - \frac{i}{n^2}$ there are no nodes with degree $> \frac{\Delta}{2^i}$ at the end of phase $i$
- With probability $1 - \frac{\log \Delta}{n^2} > 1 - \frac{\log n}{n^2}$ all nodes are either in $MIS$ or connected to a node in $MIS$ by the end of the algorithm

# Improvement

Whatever the probabilities, no better than $O(\log^2 n)$ can be achieved

***Idea:*** Adapt the probabilities used at each node based on local feedback from adjacent nodes



# The Algorithm

Algorithm: $MIS$ at node $v$

1. Set $p_v = \frac{1}{2}$
2. While active do

   Exchange 1

   1. With probability $p_v$ signal to all adjacent nodes
   2. If any neighbor is signaling then

      Stop signaling and set $p_v = \frac{p_v}{2}$
   3. else

      $p_v = \min\left\{2p_v, \frac{1}{2}\right\}$

   Exchange 2

   4. If signaling then

      Joins the $MIS$; Terminate (inactive)
   5. else if any neighbor is signaling then

      Terminate (inactive)

+ No knowledge
+ Single Bit Messages
+ Optimal Message Complexity
+ $O(\log n)$ rounds

# References


1. D. Peleg. Distributed Computing: A Locality-Sensitive Approach. Chapter 8: Maximal Independent Sets (MIS), 2000. https://doi.org/10.1137/1.9780898719772
2. S. Schmid and P.S. Mandal. Distributed Network Algorithms. Lecture Notes for GIAN Course, Chapter 8: Maximal Independent Set, 2016.
3. Y. Afek, N. Alon, O. Barad, E. Hornstein, N. Barkai and Z. Bar-Joseph. A Biological Solution to a Fundamental Distributed Computing Problem. Science 2011.
4. A. Scott, P. Jeavons and L. Xu. Feedback from nature: an optimal distributed algorithm for maximal independent set selection. PODC 2013.