On the Policy Iteration algorithm

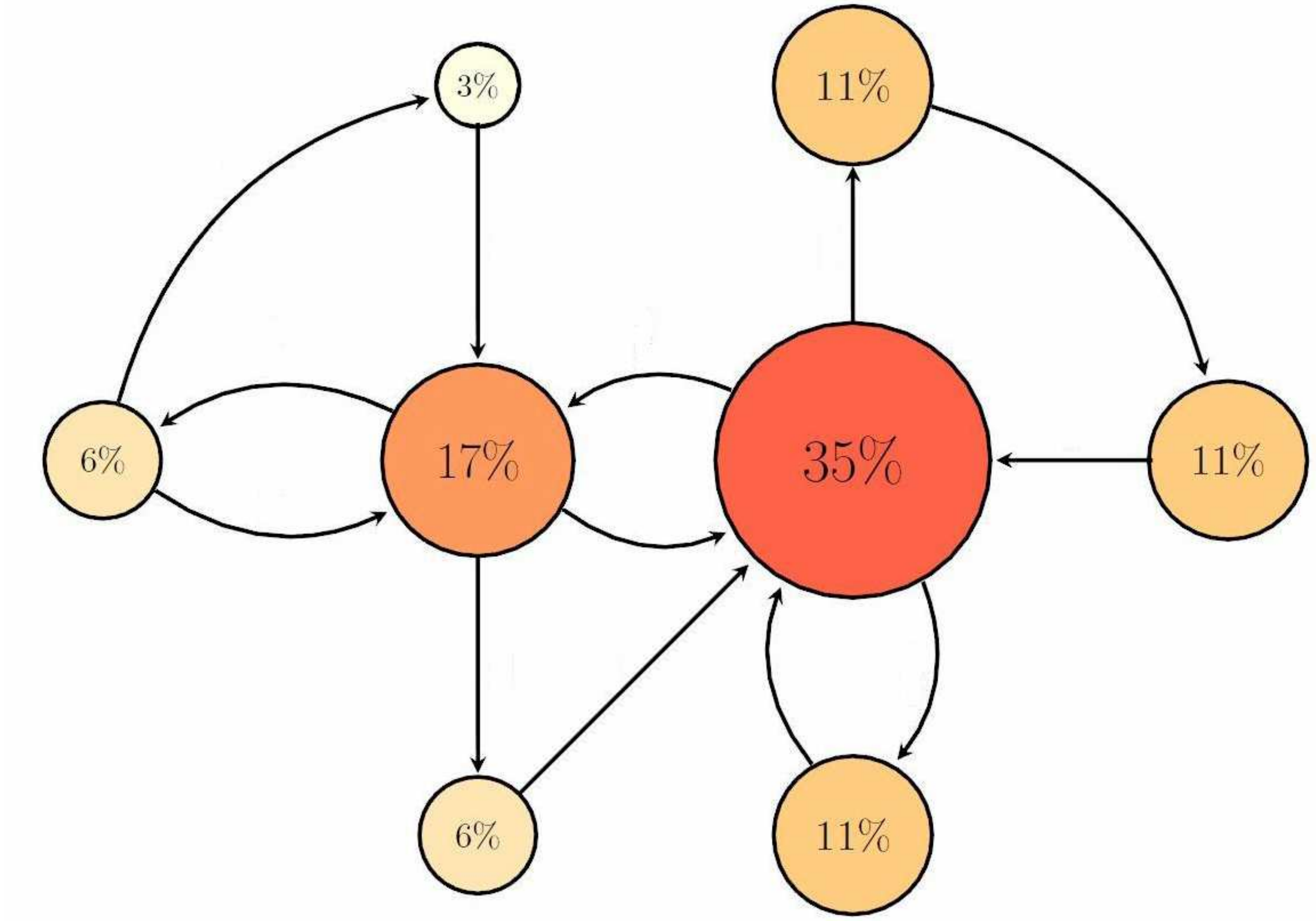
For PageRank optimization

Romain Hollanders UCLouvain / ICTEAM

Supervisors : Raphaël Jungers
Vincent Blondel
Jean-Charles Delvenne

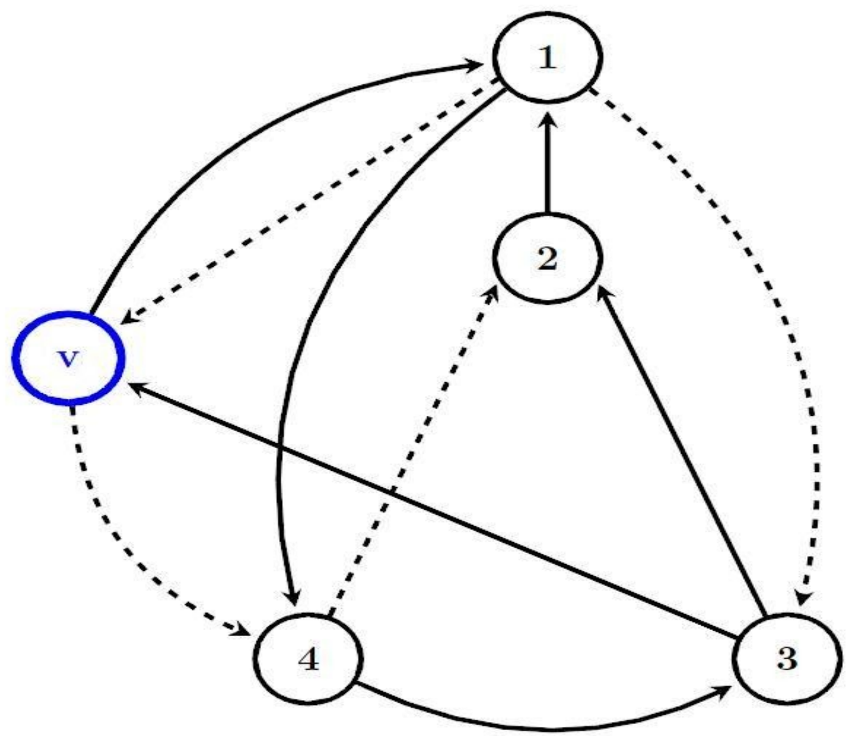
PageRank = average portion of time spent in a node

during an infinite random walk



A not so trivial task

PageRank Optimisation
by Edge selection ?



Several algorithms have been proposed

But which one should we use ?

1. Approximation of the optimal solution
   But running in polynomial time

2. Linear programming : runs in polynomial time
   But doesn't take all the problem's specificity into account

3. Iterative algorithm based on Policy Iteration : interesting behavior
   But no theoretical complexity results

Several algorithms have been proposed

But which one should we use ?

1. Approximation of the optimal solution
   But running in polynomial time

2. Linear programming : runs in polynomial time
   But doesn't take all the problem's specificity into account

3. Iterative algorithm based on Policy Iteration : interesting behavior
   But no theoretical complexity results

↳ Our main focus

Can we improve the existing complexity results
for our webmaster's problem ?

On the policy Iteration algorithm

for Page Rank Optimisation

1. **The Max-PageRank problem**

   Which problem do we want to solve?

2. **The PageRank Iteration algorithm**

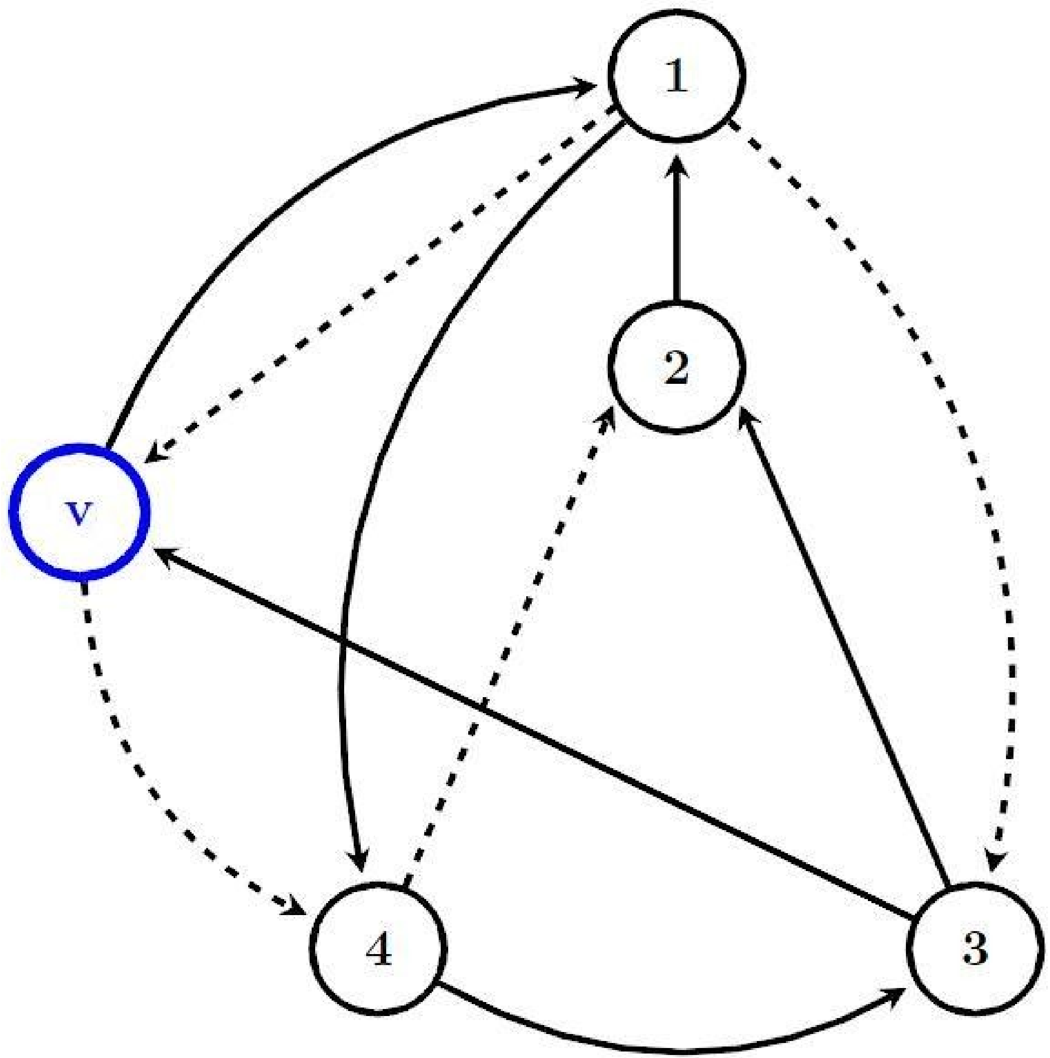
   How do we solve the problem?

3. **Our results**

   What did we find about the algorithm?

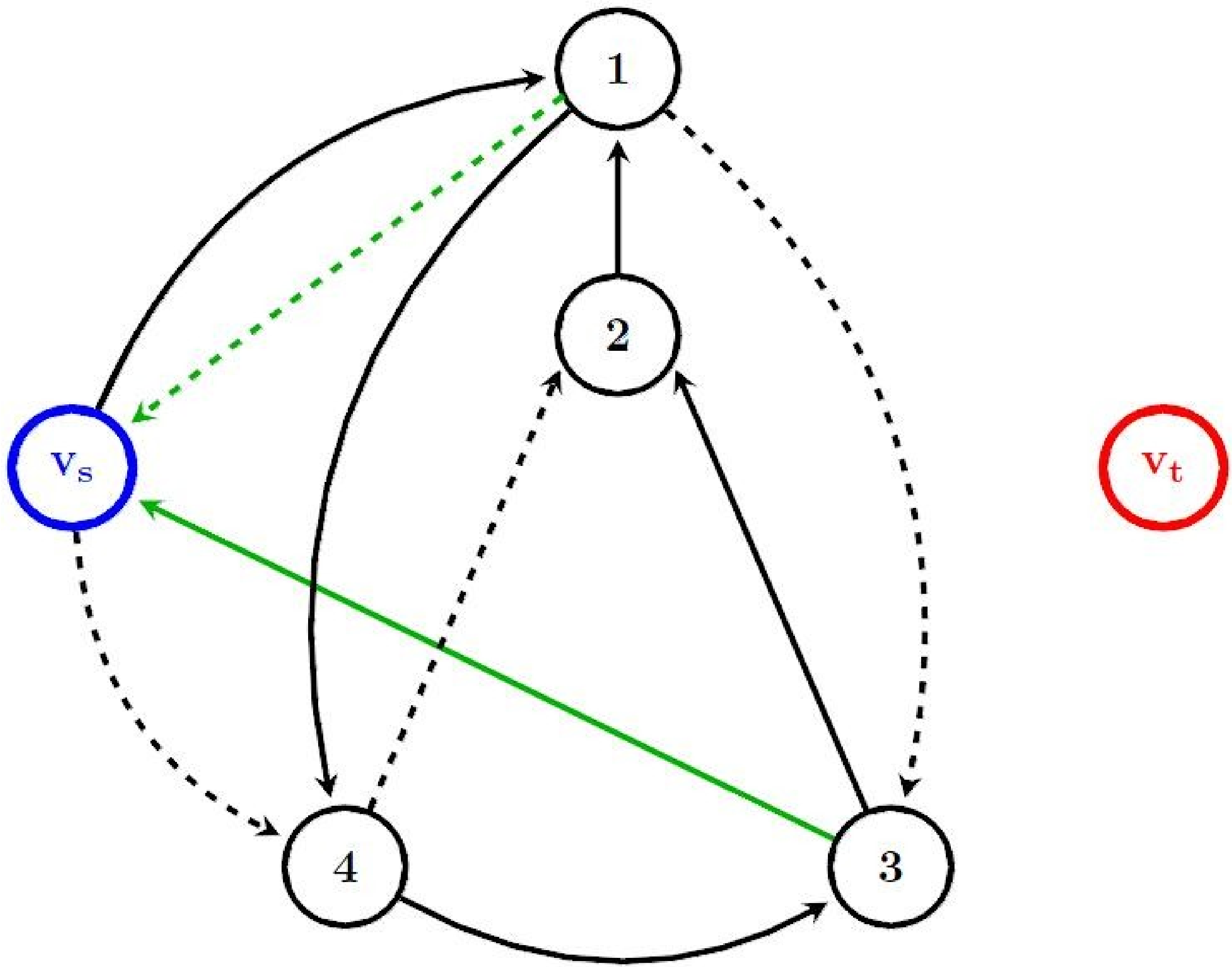
Which fragile edges should we activate ?

To maximize the PageRank of v or minimize its first return time



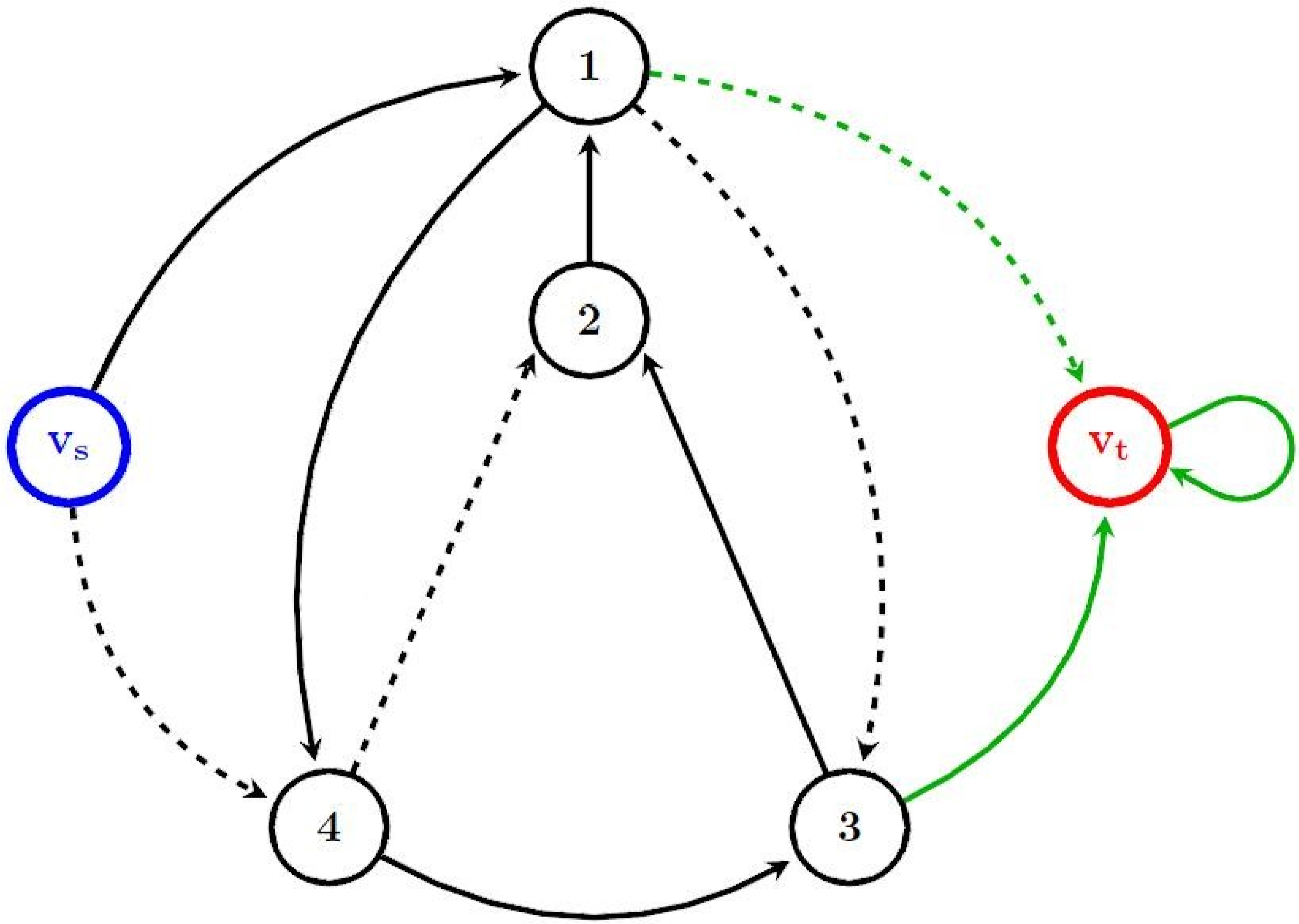
Which fragile edges should we activate ?

We formulate the problem as a Stochastic Shortest Path problem



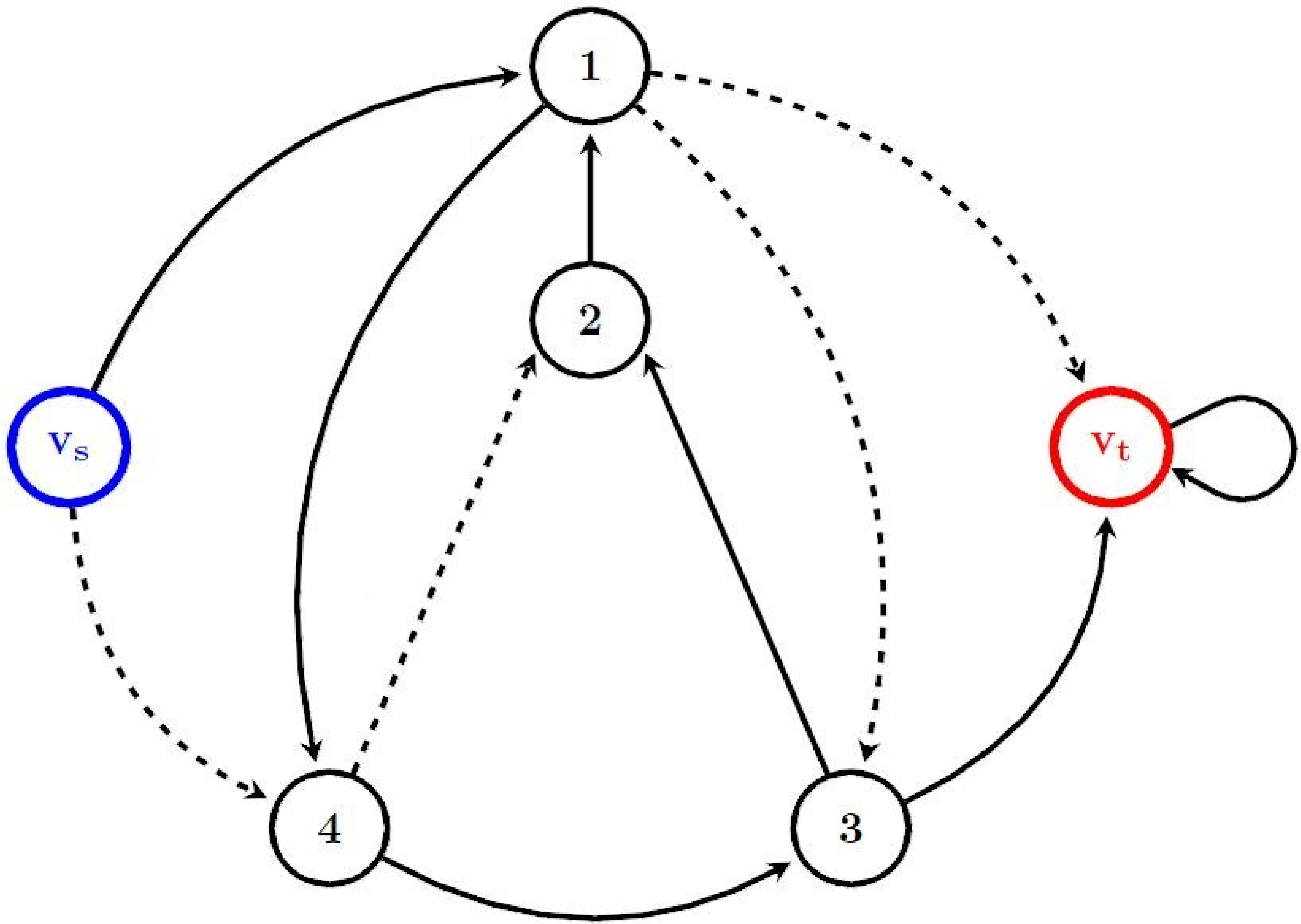
Which fragile edges should we activate ?

We formulate the problem as a Stochastic Shortest Path problem



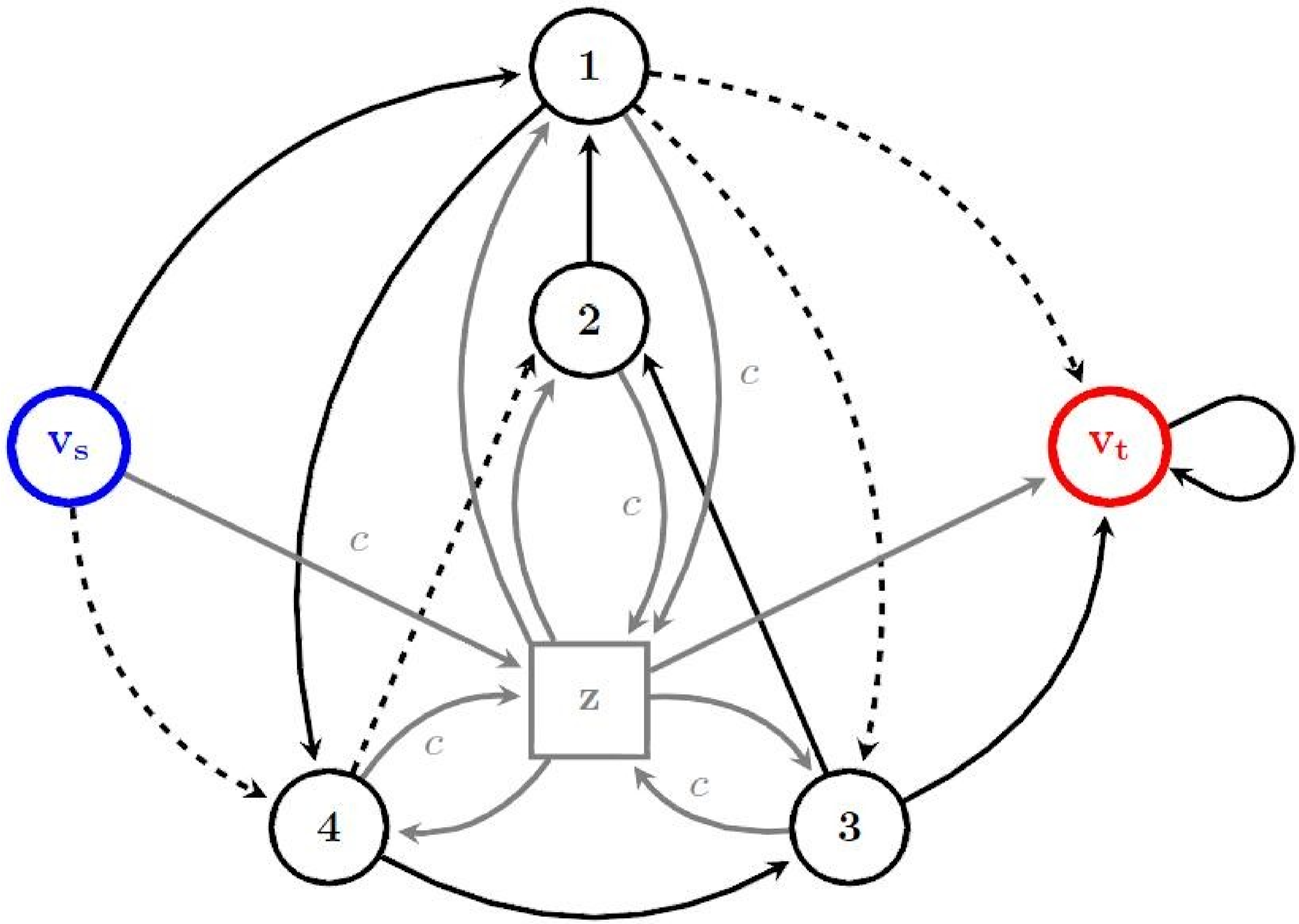
Which fragile edges should we activate ?

To maximize the PageRank of v_s or minimize the distance from v_s to v_t



We often add damping to the problem

The optimal solution may change but the problem is better conditioned



On the policy Iteration algorithm

for Page Rank Optimisation

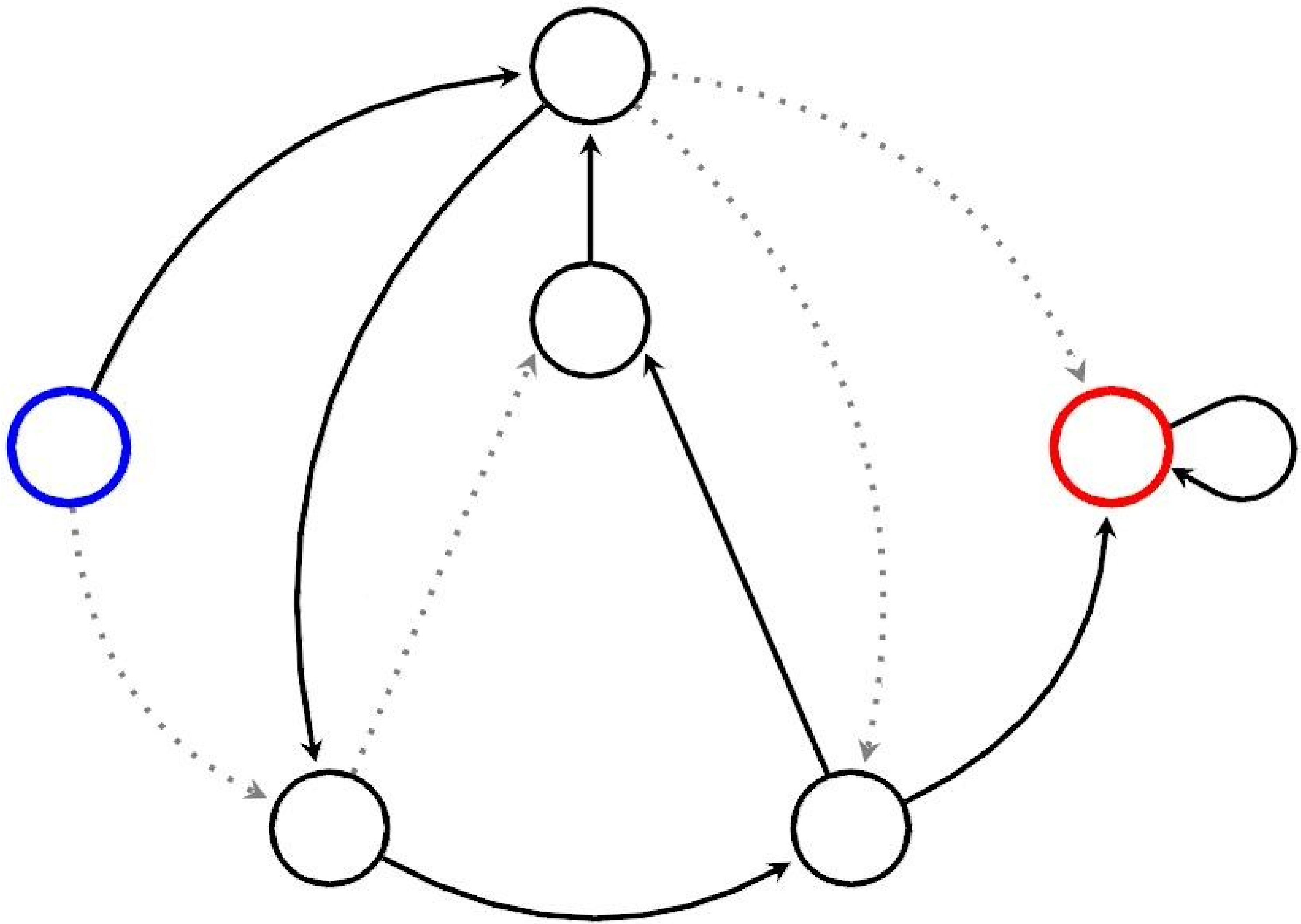
1. **The Max-PageRank problem**
   Which problem do we want to solve?

2. **The PageRank Iteration algorithm**
   How do we solve the problem?

3. **Our results**
   What did we find about the algorithm?

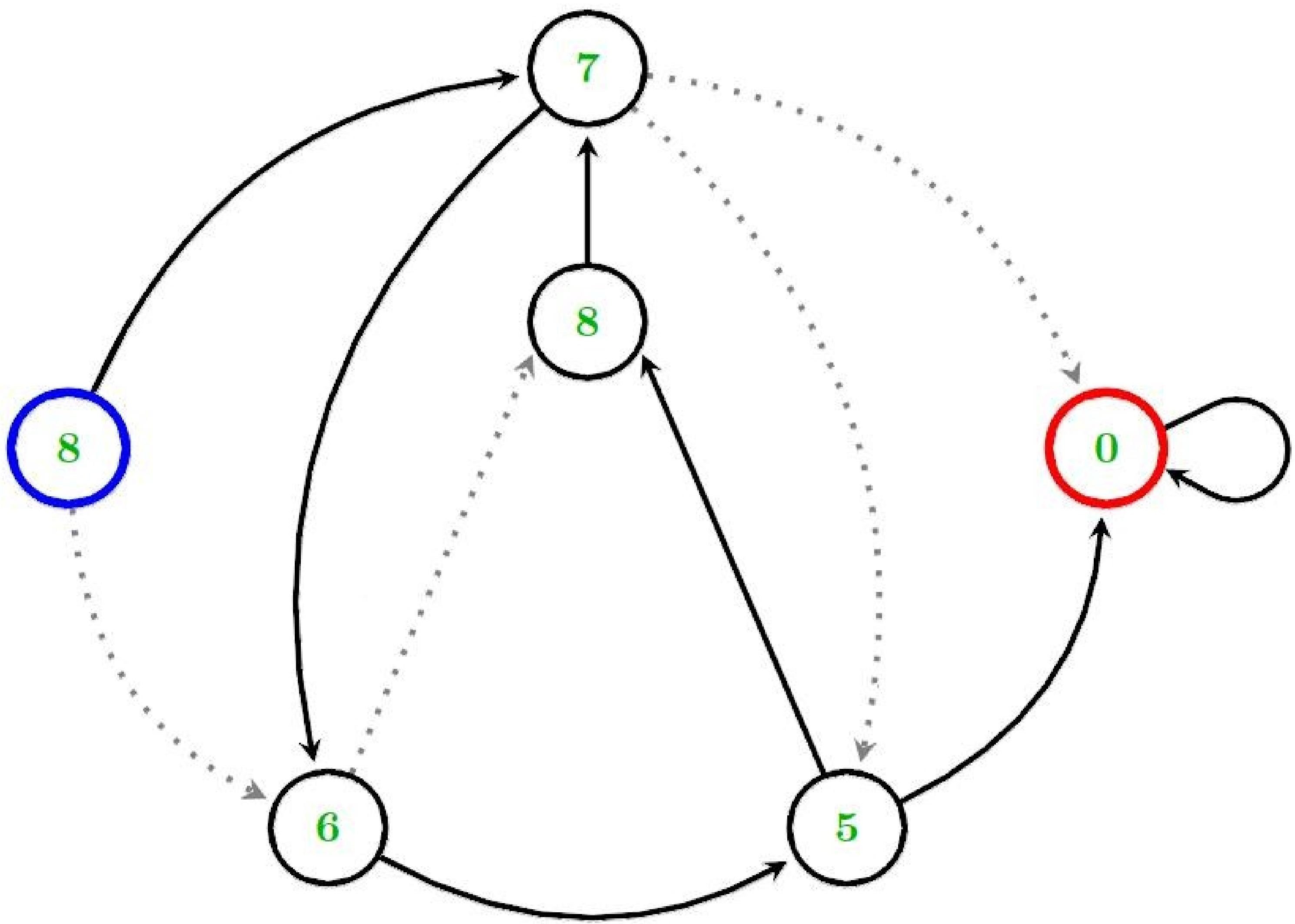
Iteration 1 : Evaluation step

Initial policy : all fragile edges are OFF



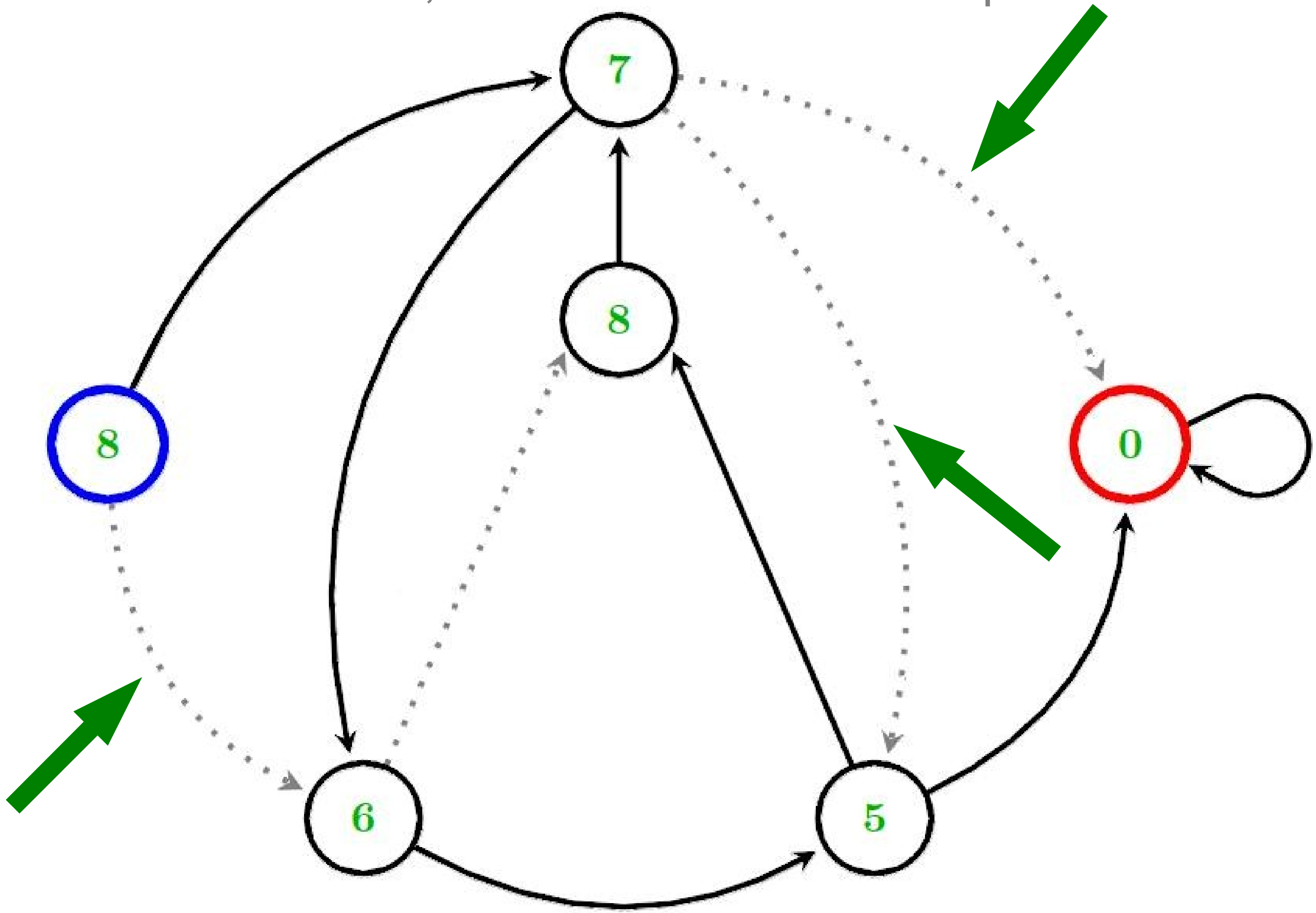
Iteration 1 : Evaluation step

Initial policy : all fragile edges are OFF

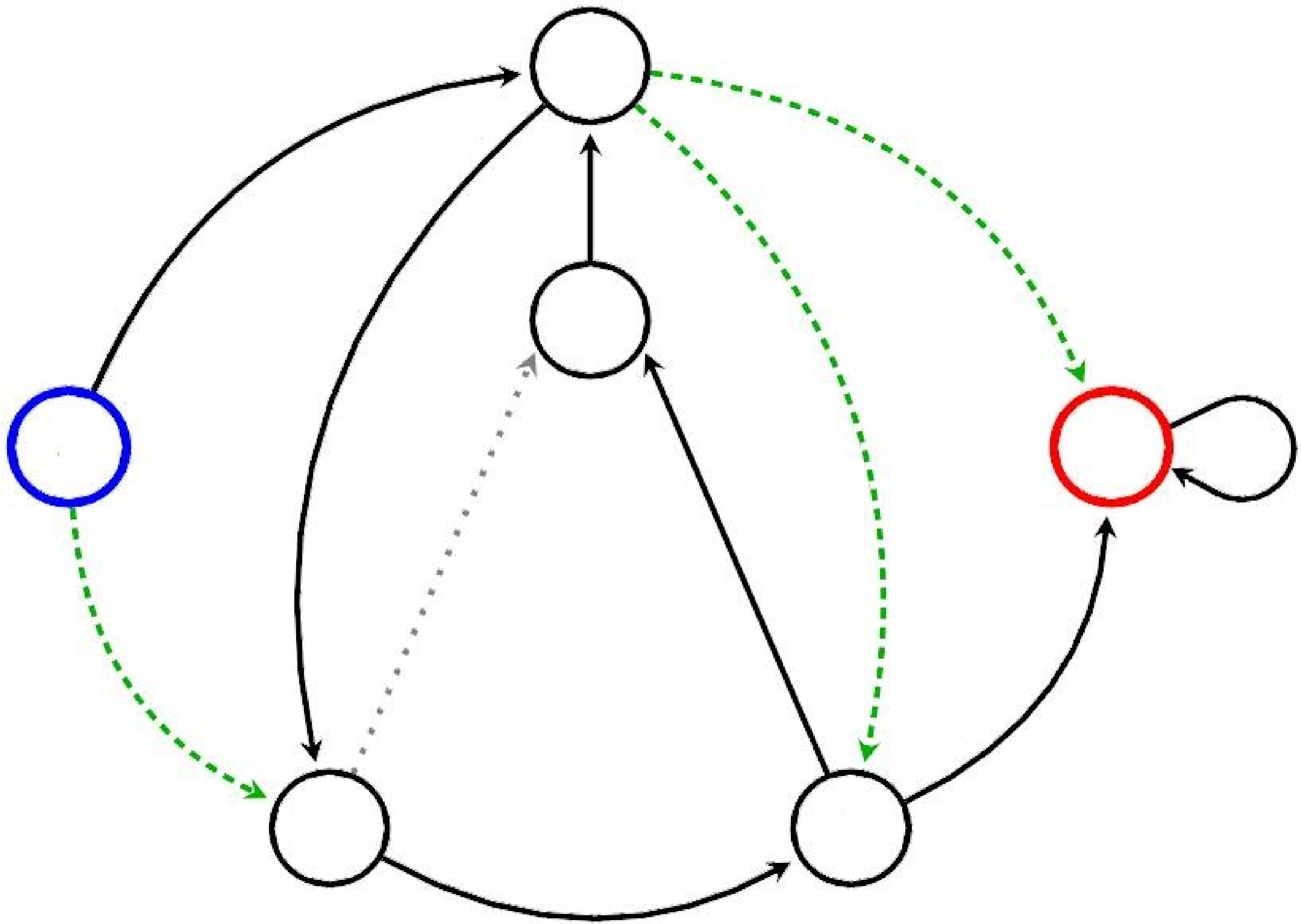


Iteration 1 : Improvement step

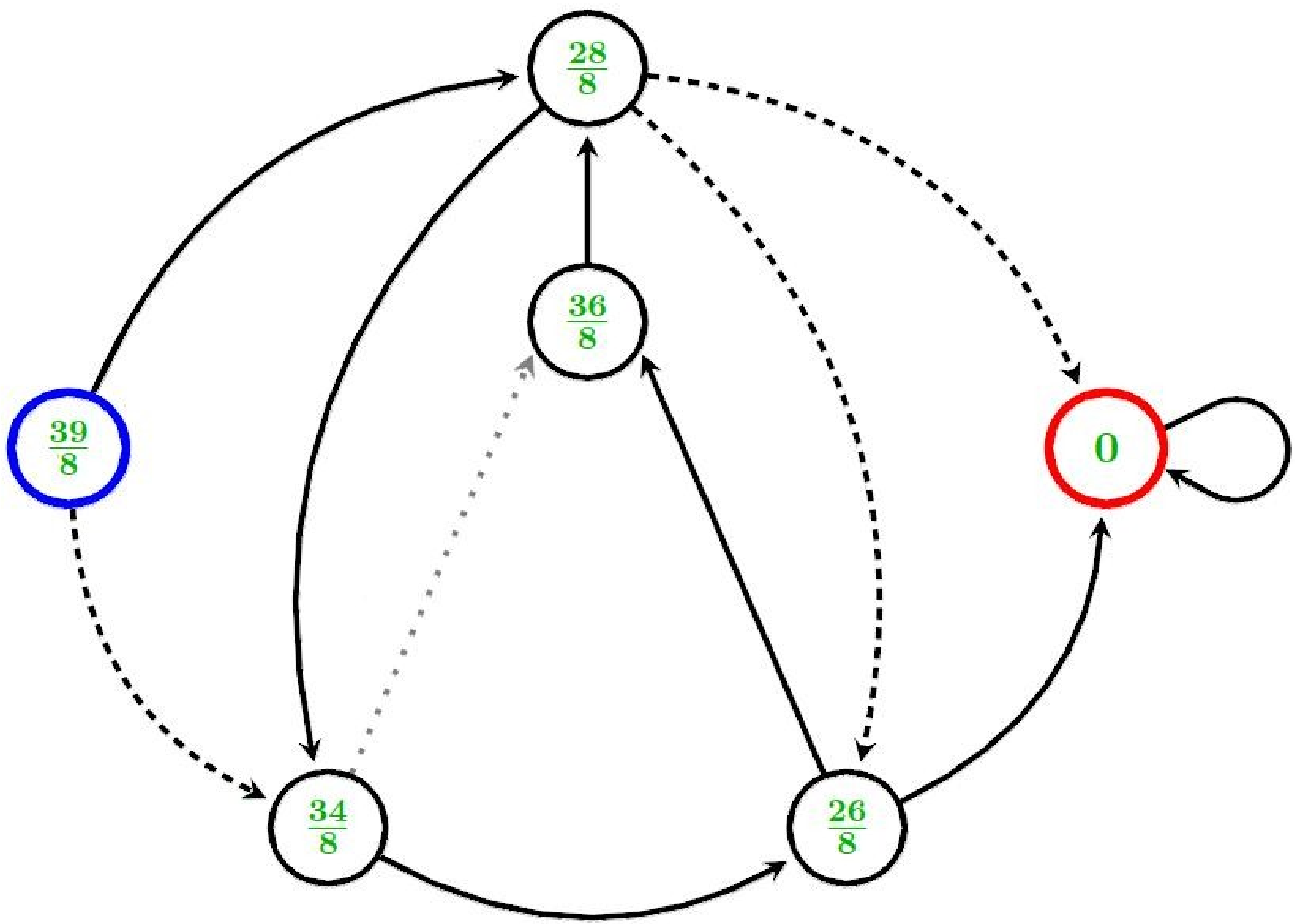
It is better to be at a distance d from the target node
than at distance > d+1, even at the cost of a displacement



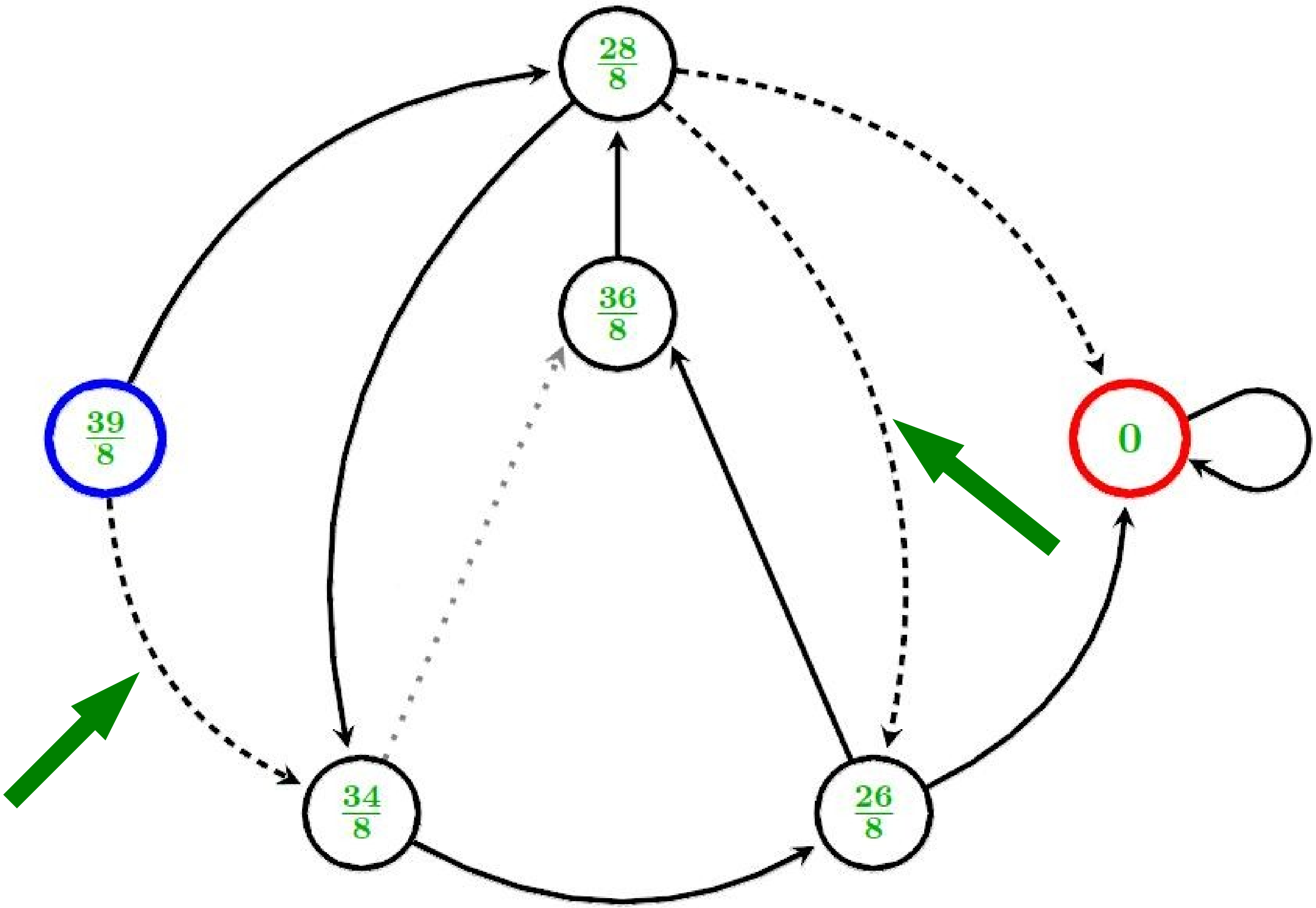
Iteration 1 : Improvement step



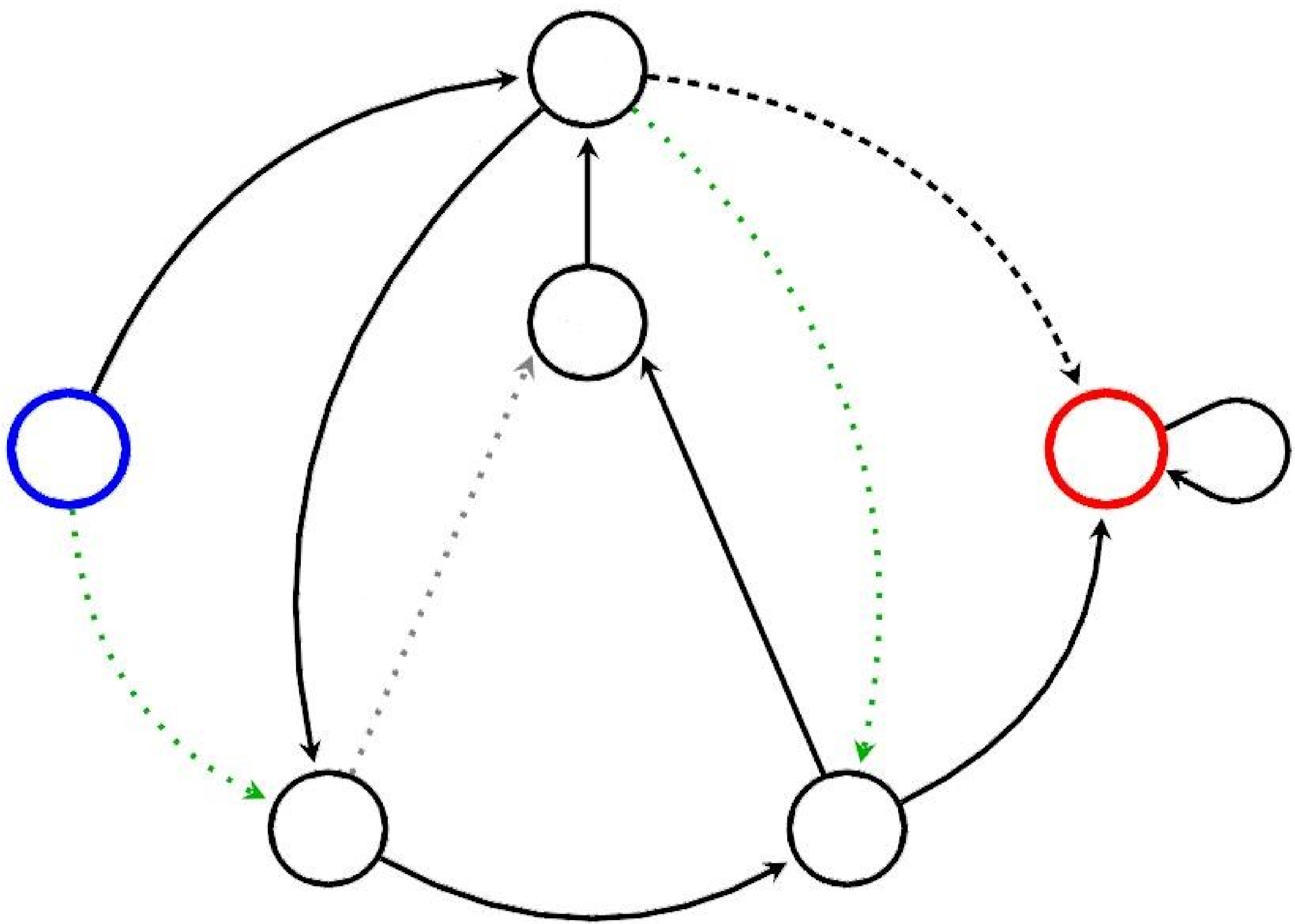
Iteration 2 : Evaluation step



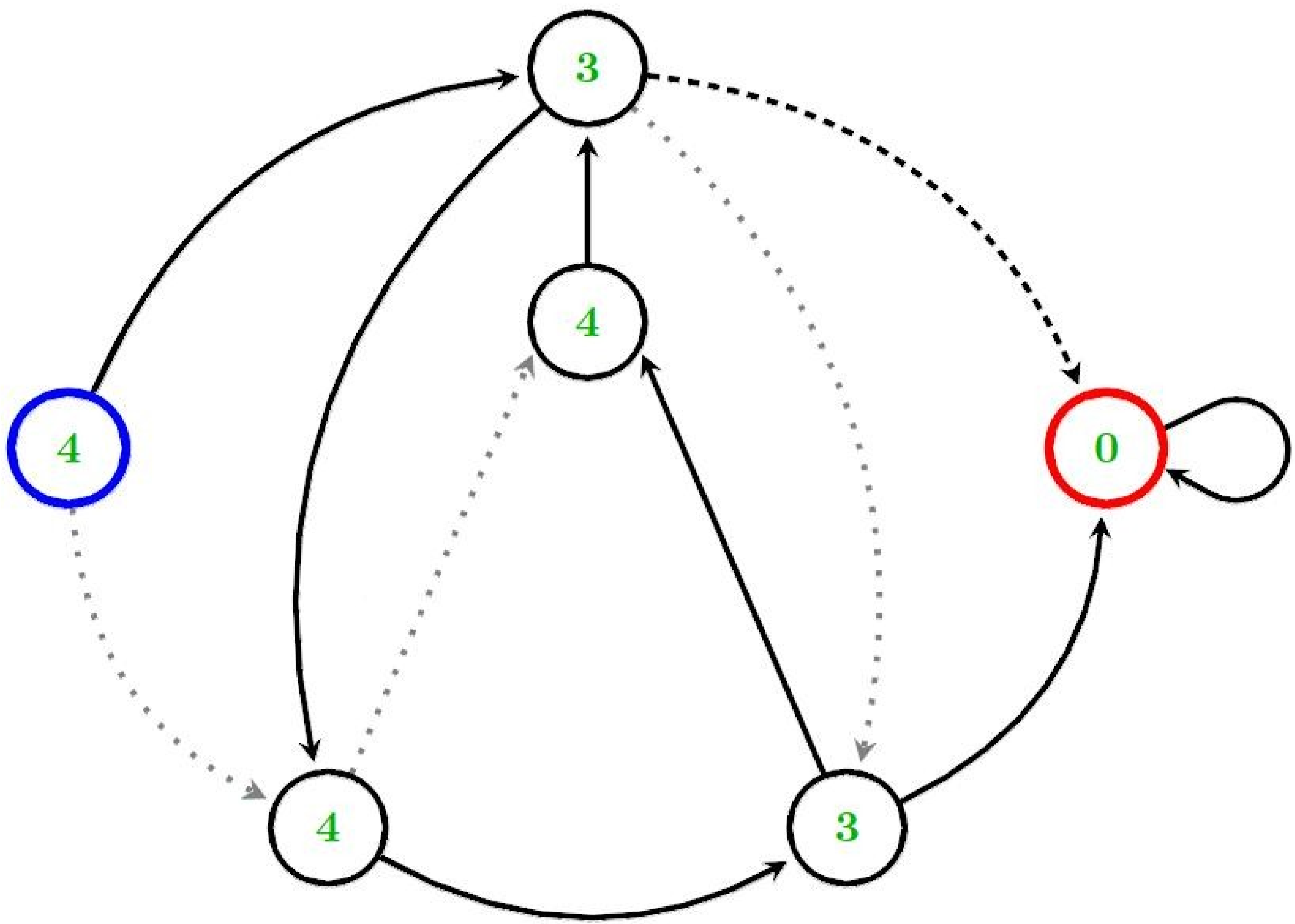
Iteration 2 : Improvement step



Iteration 2 : Improvement step

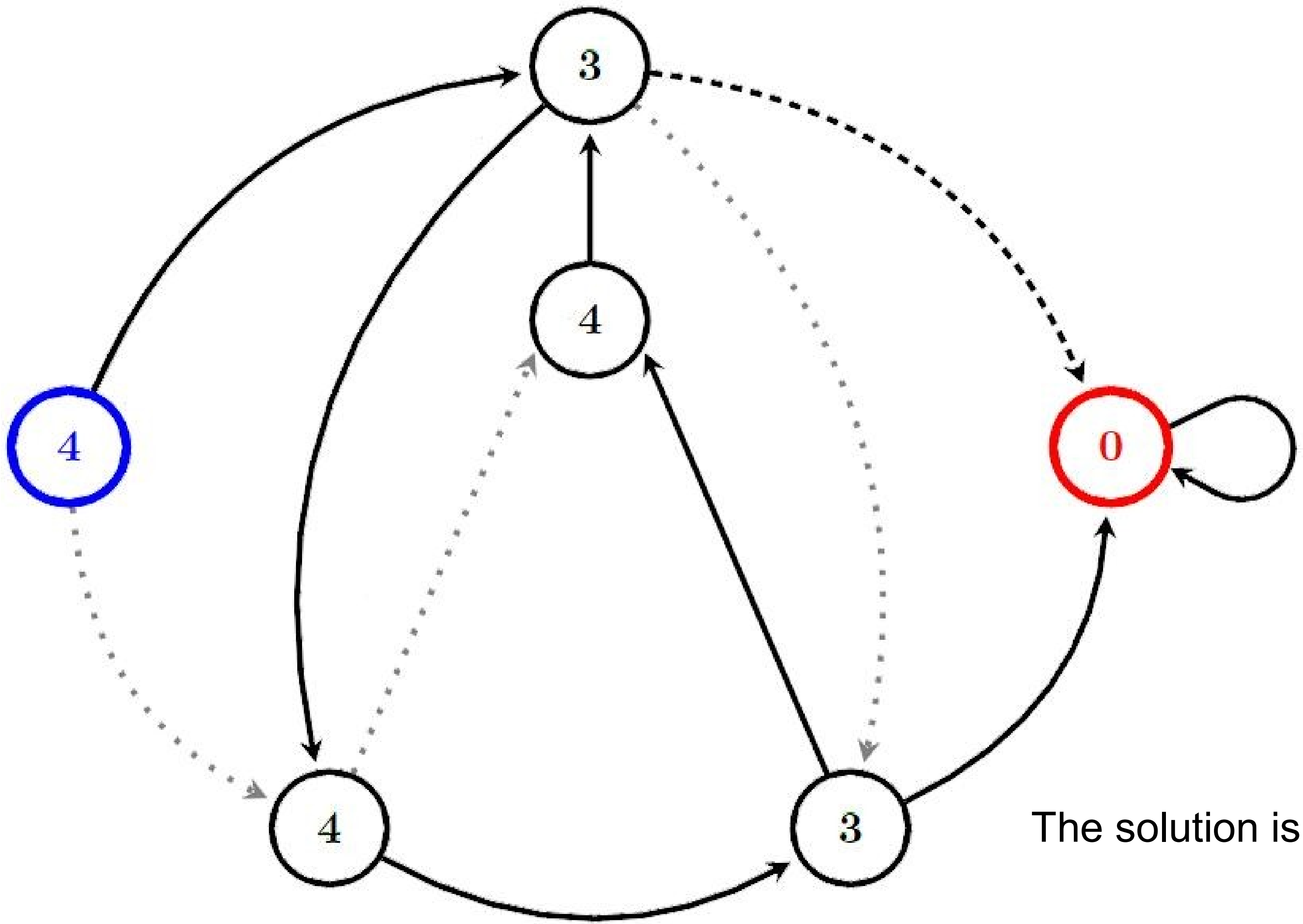


Iteration 3 : Evaluation step



Iteration 4 : convergence

Nothing changes



The solution is optimal !

On the policy Iteration algorithm

for Page Rank Optimisation

1. **The Max-PageRank problem**

   Which problem do we want to solve?

2. **The PageRank Iteration algorithm**

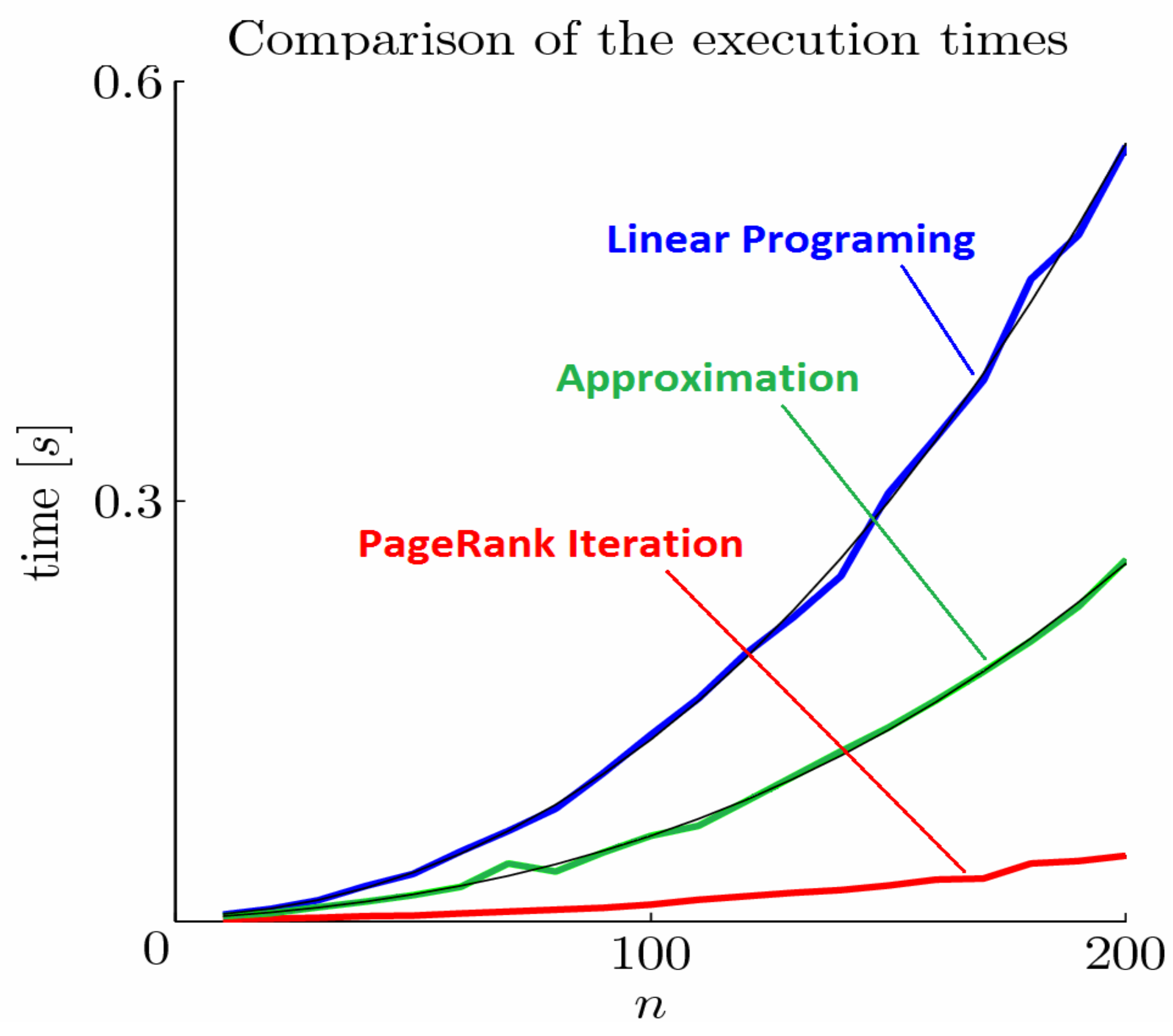
   How do we solve the problem?

3. **Our results**

   What did we find about the algorithm?

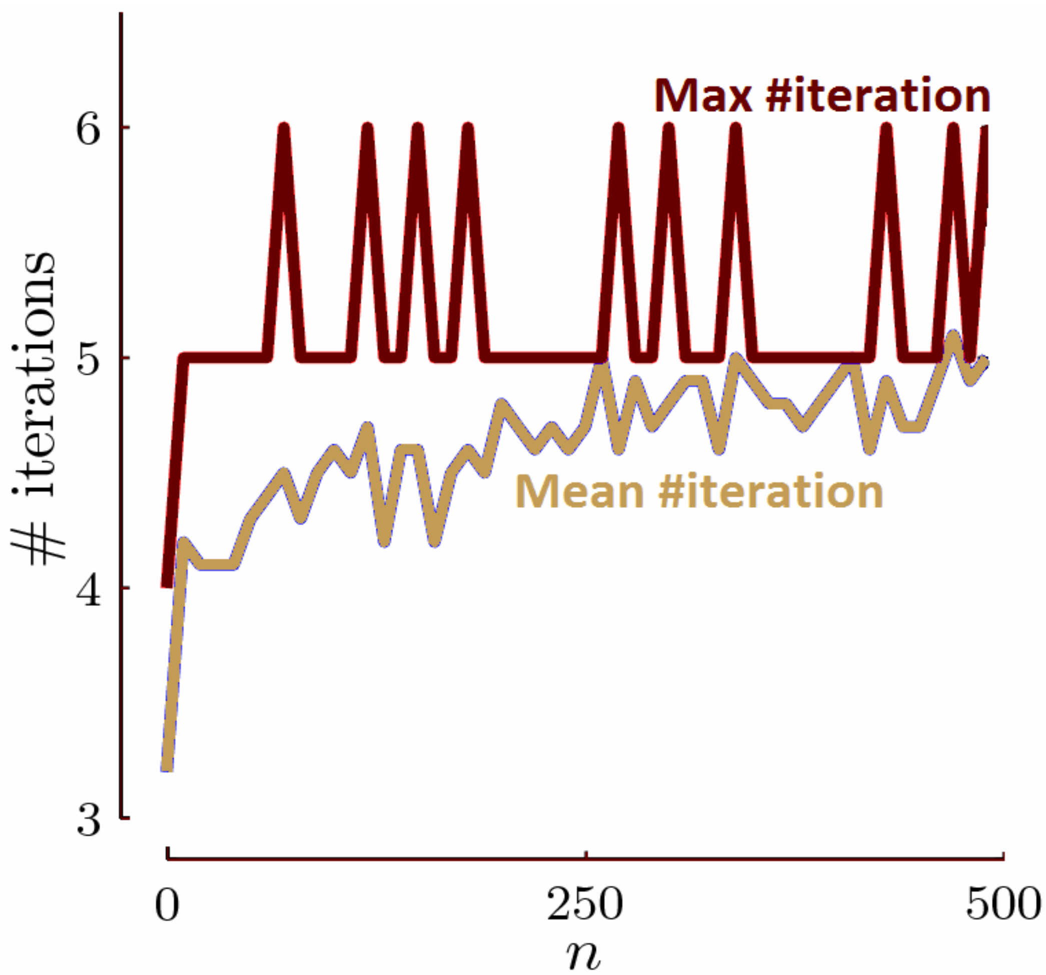
In practice, PRI is by far the best method

in terms of execution times



The number of iterations of PRI seems to grow at most linearly (or even logarithmically ?)

With respect to the problem size



No theoretical bounds on the complexity of PRI yet

This is where we step in !

- PRI is a particular case of Policy Iteration

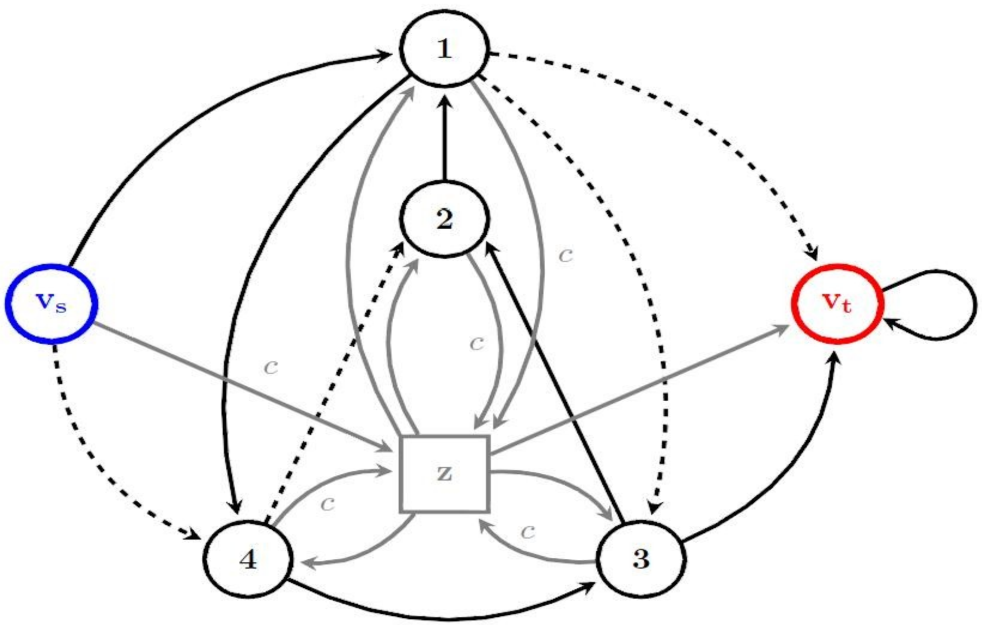
No theoretical bounds on the complexity of PRI yet

This is where we step in !

- PRI is a particular case of Policy Iteration
- We know that Policy Iteration converges in $O\left(p(L)\,\boxed{\eta^{2r}}\right)$ iterations where :
  - ➡ η is the smallest non-zero transition probability
  - ➡ r is the diameter of the graph

A polynomial bound for a particular case of PRI

This is where we step in !

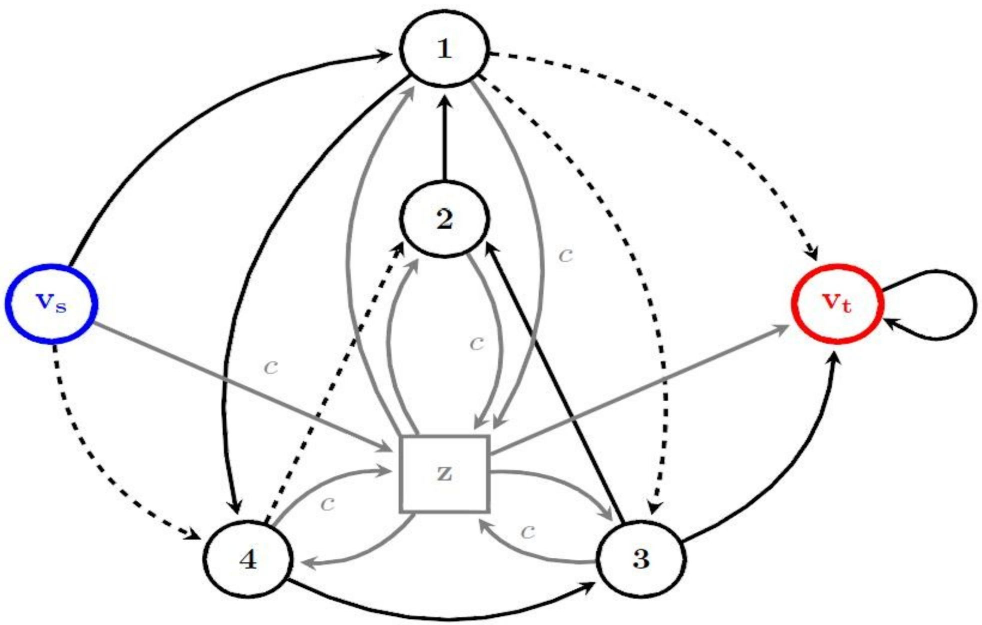


- In our case, with damping, we have :

➡ $$\eta = O\left(\min\left(\frac{1}{n}, c\right)\right)$$

A polynomial bound for a particular case of PRI

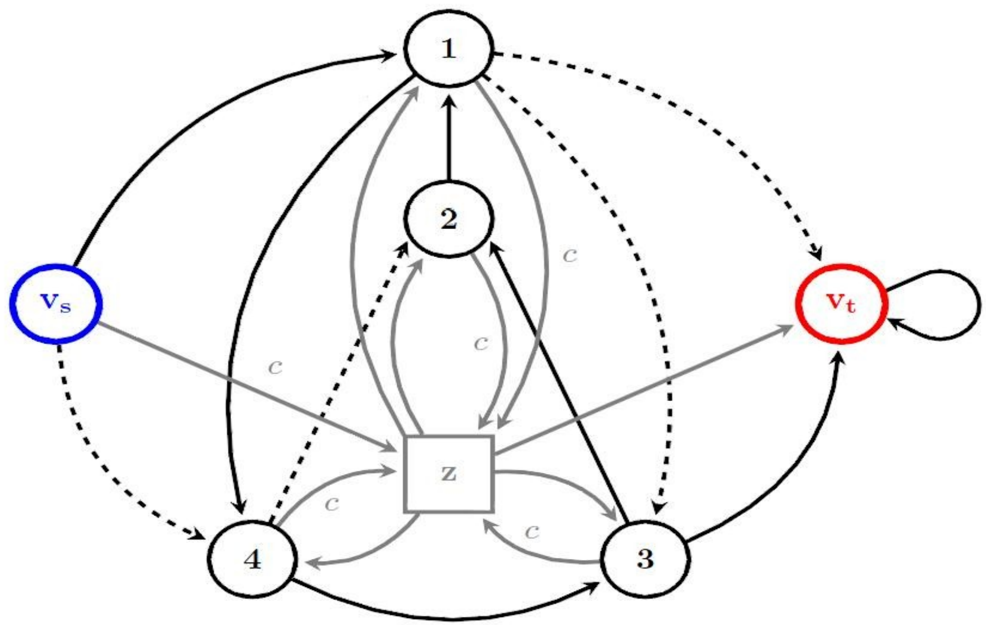
This is where we step in !



- In our case, with damping, we have :
  - ➡ $\eta = O\left(\min\left(\frac{1}{n}, c\right)\right)$
  - ➡ $r = 2$

A polynomial bound for a particular case of PRI

This is where we step in !



- In our case, with damping, we have :

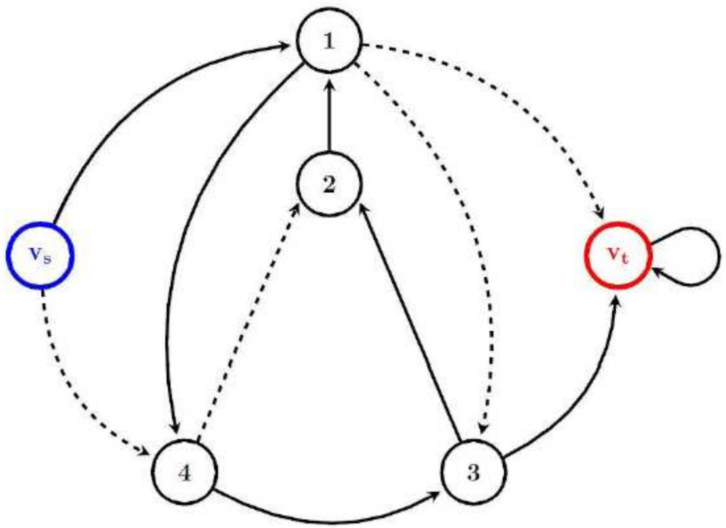
$$\left.\begin{array}{l} \eta = O\left(\min\left(\frac{1}{n}, c\right)\right) \\ r = 2 \end{array}\right\} \Longrightarrow$$

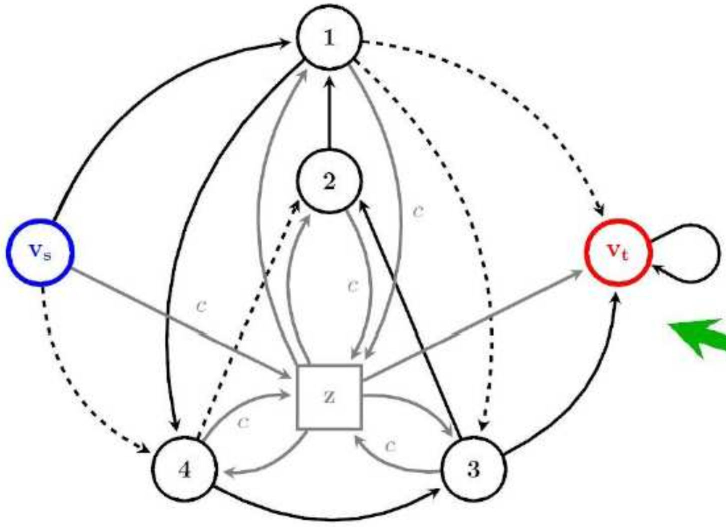
Weakly polynomial bound on the complexity of PRI

What can we do for the undamped case ?

Several ideas have been explored

1.





Can we find a sufficiently small c

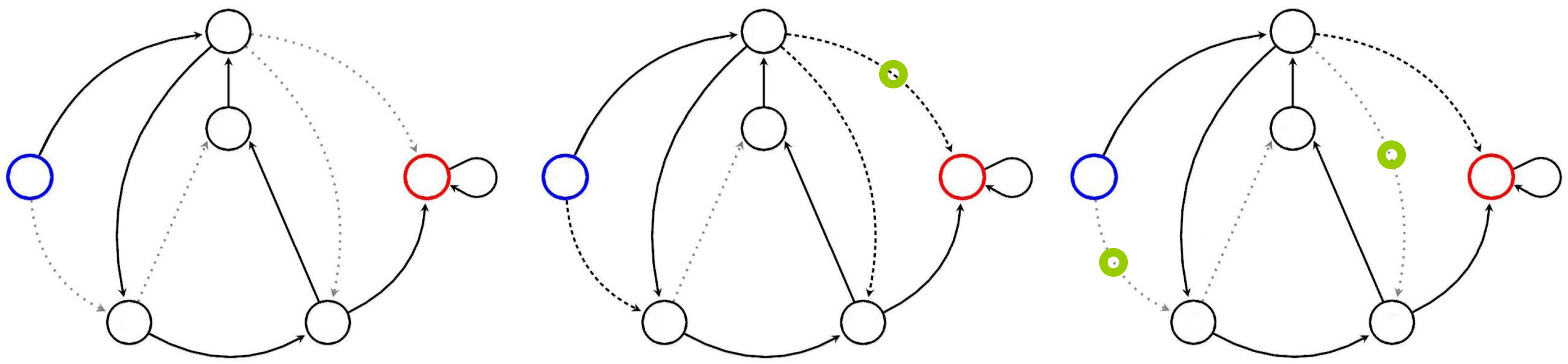
such that the solution of
the two problems is identical ?

What can we do for the undamped case ?

Several ideas have been explored

2. Can we show that PRI makes at least one definitive choice at each iteration ?

➡ If we do, then PRI takes at most [# fragile edges] iterations



Conclusions on the theoretical complexity of PRI

Few existing results

- Each iteration : asymptotic polynomial complexity
- Number of iterations :
  - ➡ New polynomial bound for the case with damping
    This is our main contribution
  - ➡ The undamped case is still open
    Several approaches have been explored

Conclusions on the theoretical complexity of PRI

Few existing results

- Each iteration : asymptotic polynomial complexity
- Number of iterations :
  - ➡ New polynomial bound for the case with damping
    This is our main contribution
  - ➡ The undamped case is still open
    Several approaches have been explored

Any ideas welcome !