Rumor Spreading and Conductance

Flavio Chierichetti *Silvio Lattanzi* Alessandro Panconesi

Sapienza
University of Rome

Why is rumor spreading fast in social networks?

- How to answer this question?
- How to define rumor spreading?
- What are social networks?

Rumor spreading

Rumor spreading

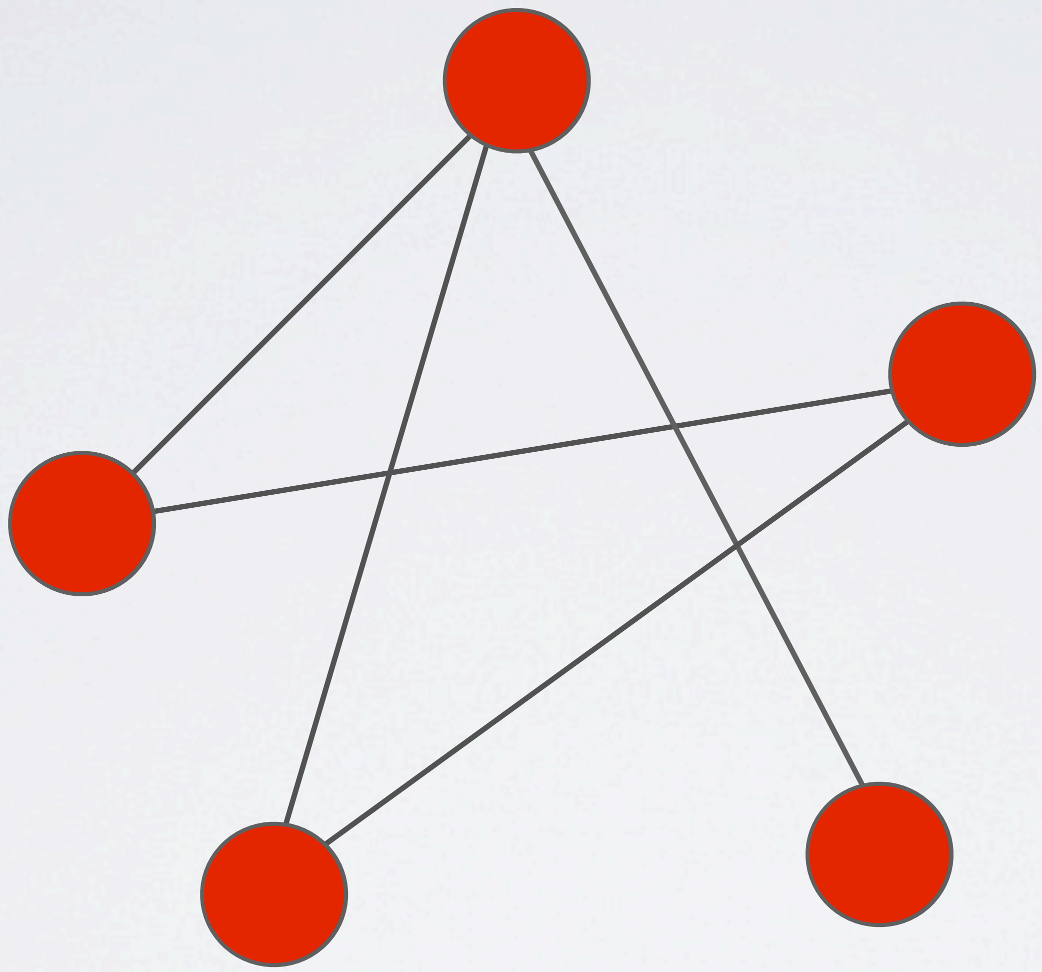
- Push, Pull and Push-Pull

  Introduced in the contest of distributed database.

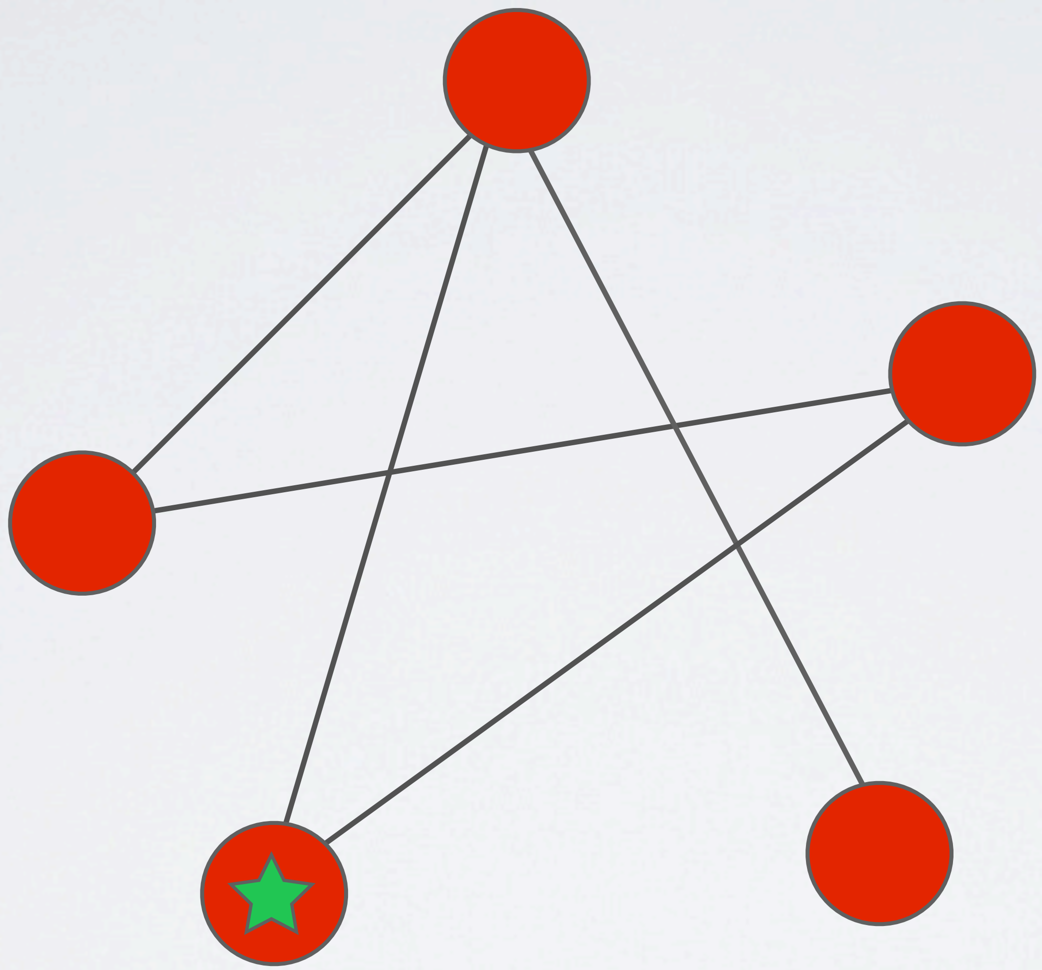
  Demers, Greene, Hauser, Irish, Larson, Shenker, Sturgis, Swinehart, Terry, PODC 1987

- Basic mechanisms for information dissemination in networks.

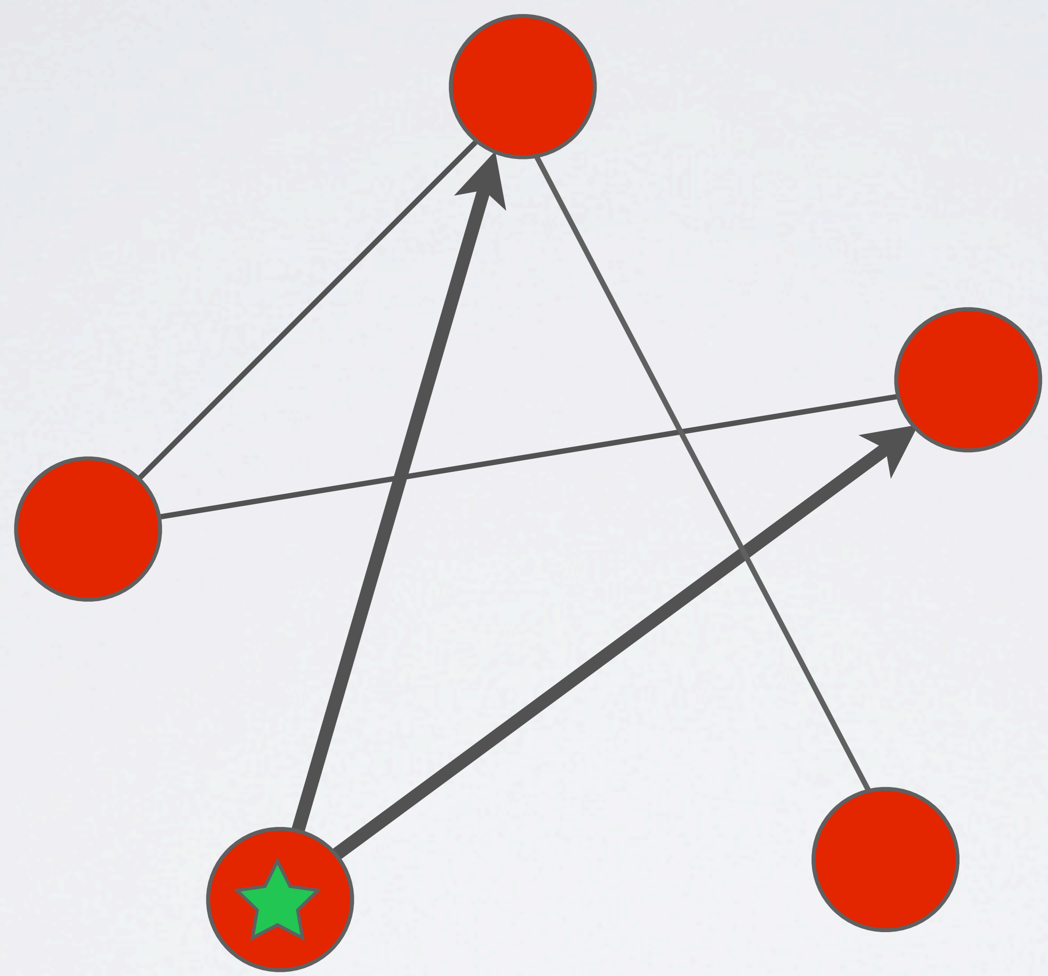
Push



Push

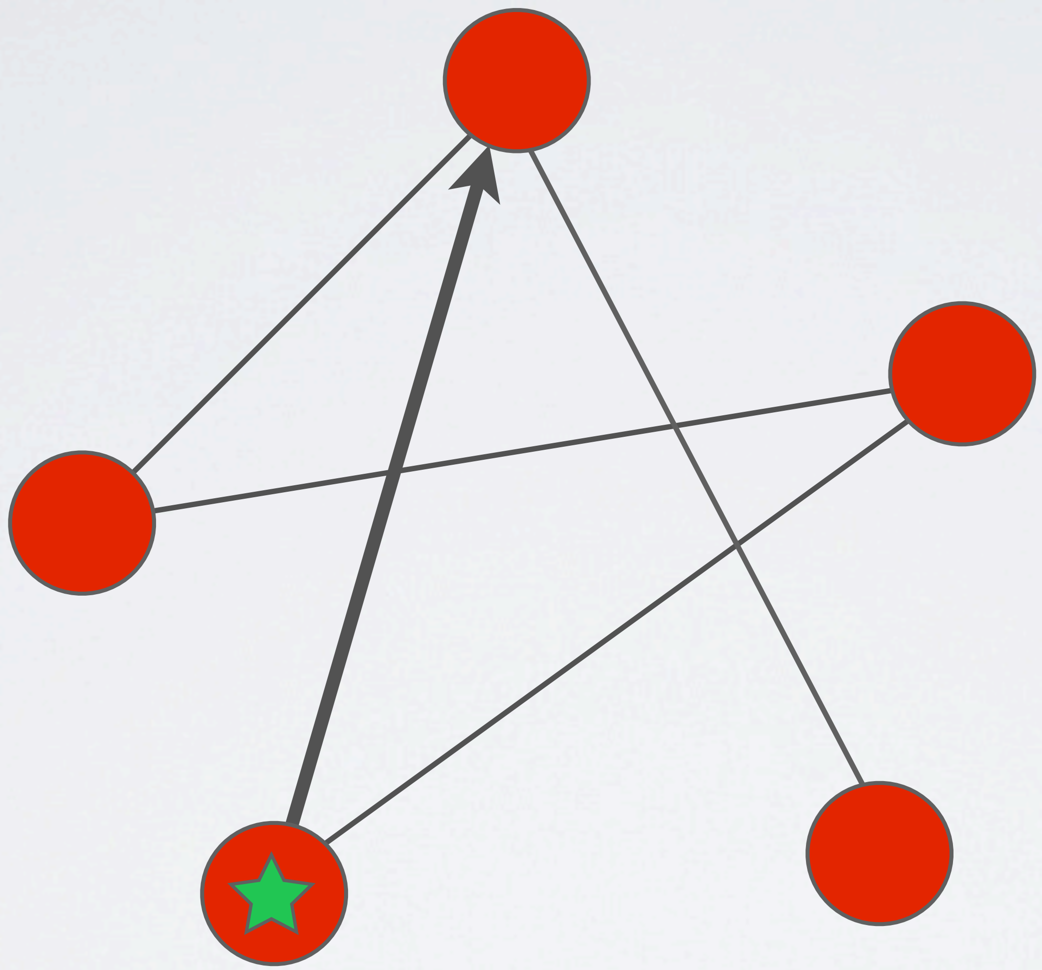


Push



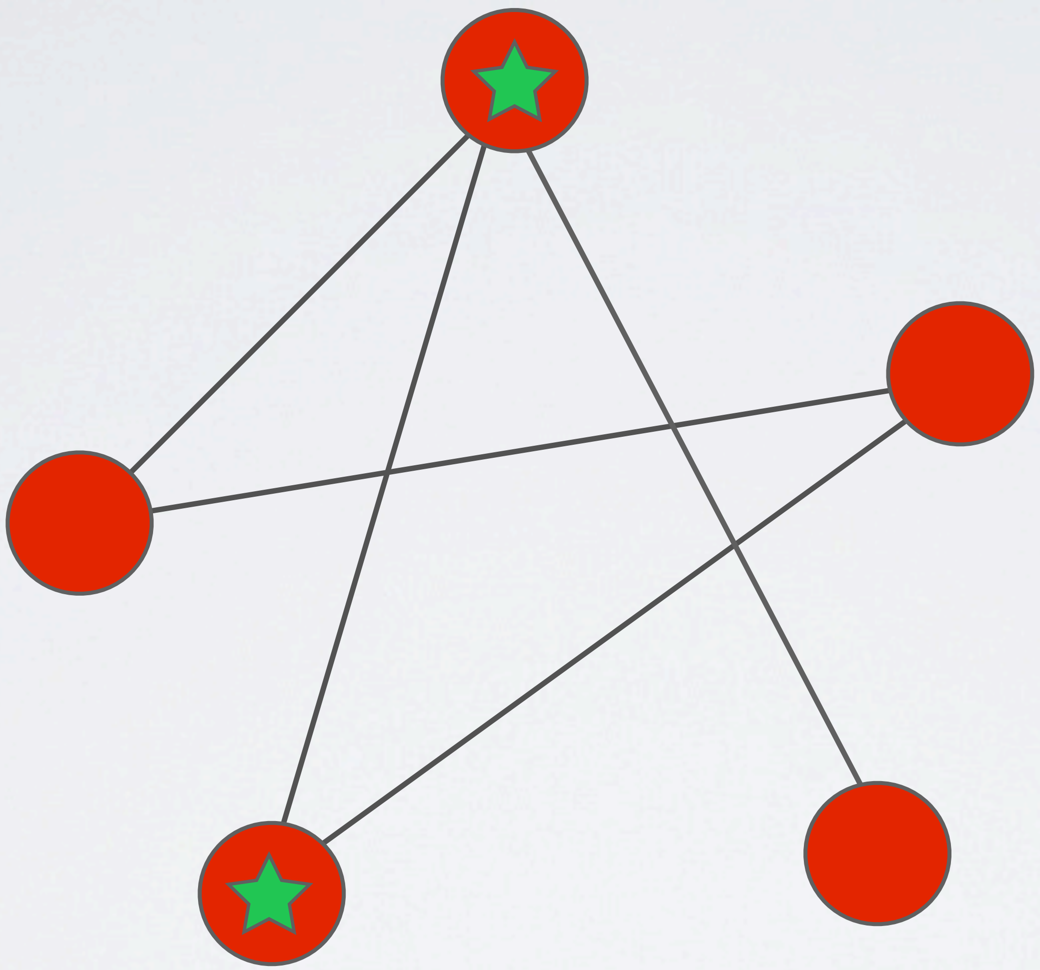
In each round, each informed node "pushes" the information to a neighbor chosen UAR.

Push



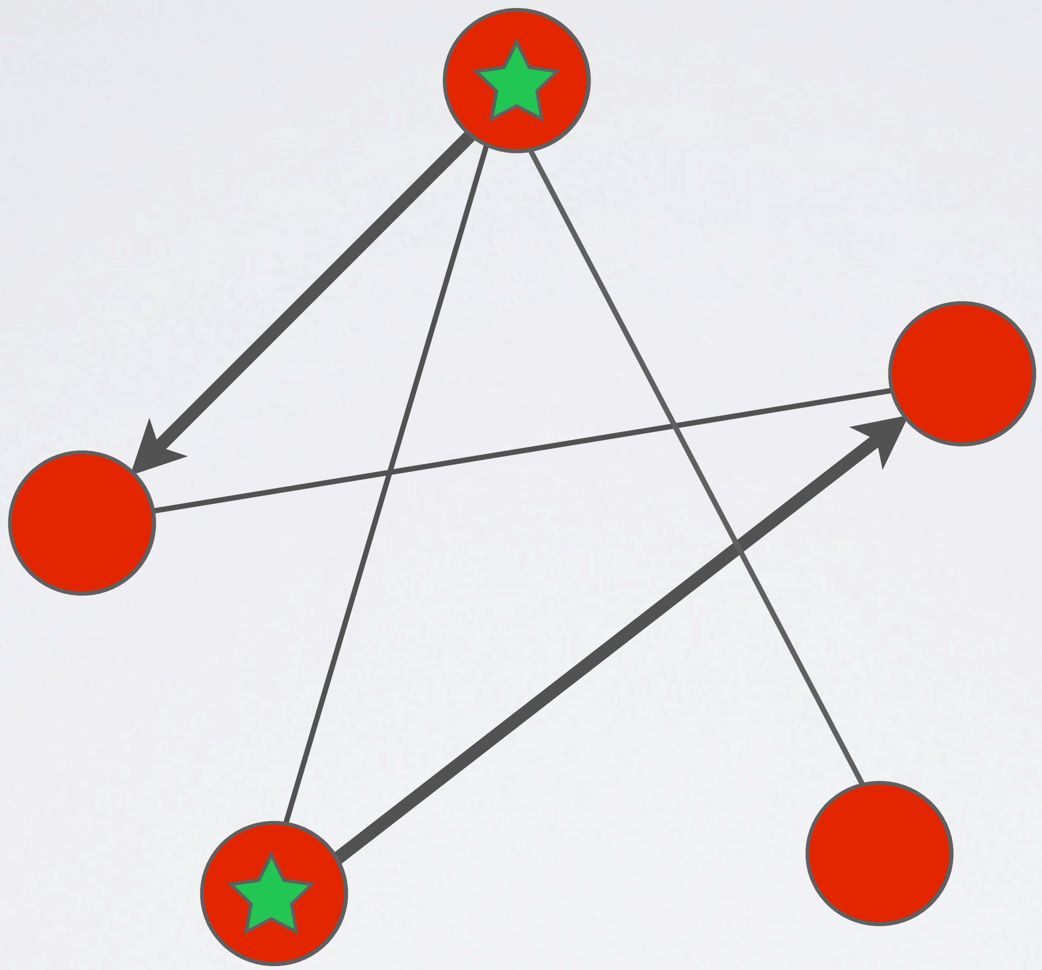
In each round, each informed node “pushes” the information to a neighbor chosen UAR.

Push



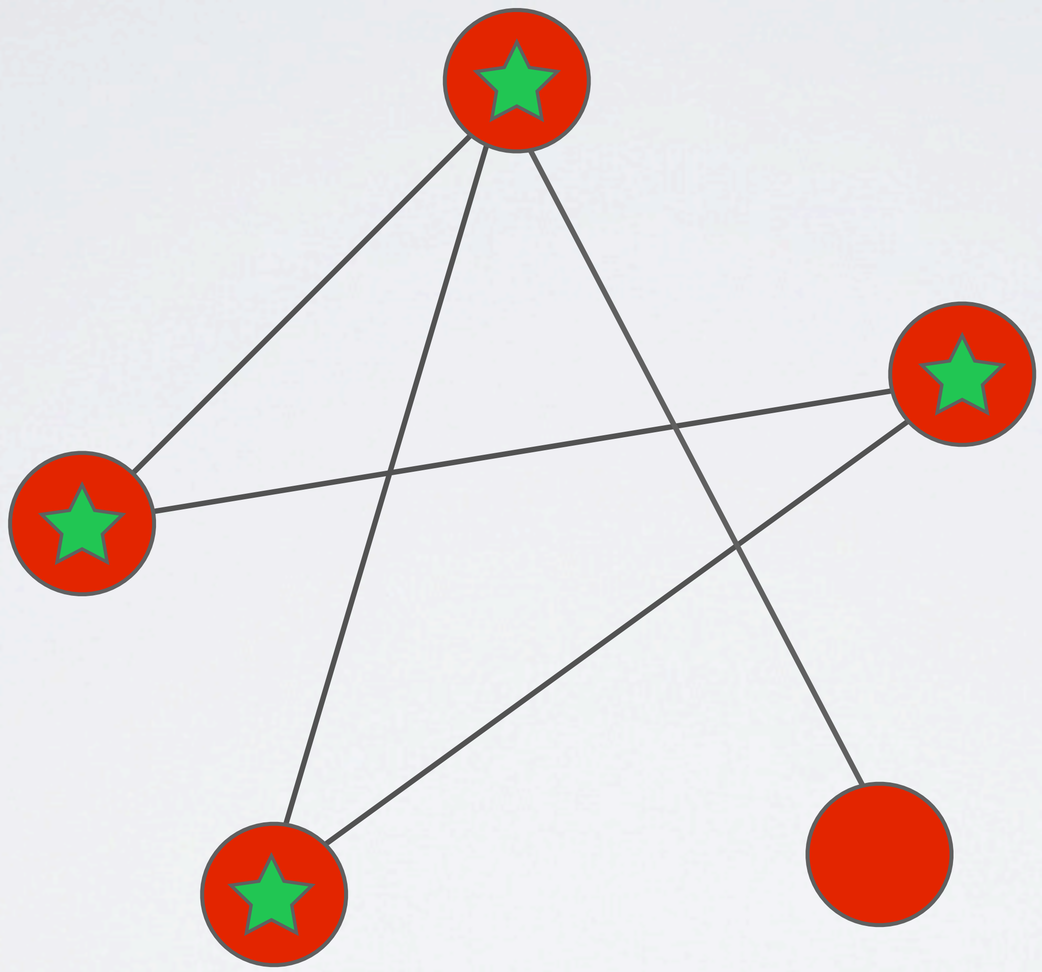
In each round, each informed node "pushes" the information to a neighbor chosen UAR.

Push



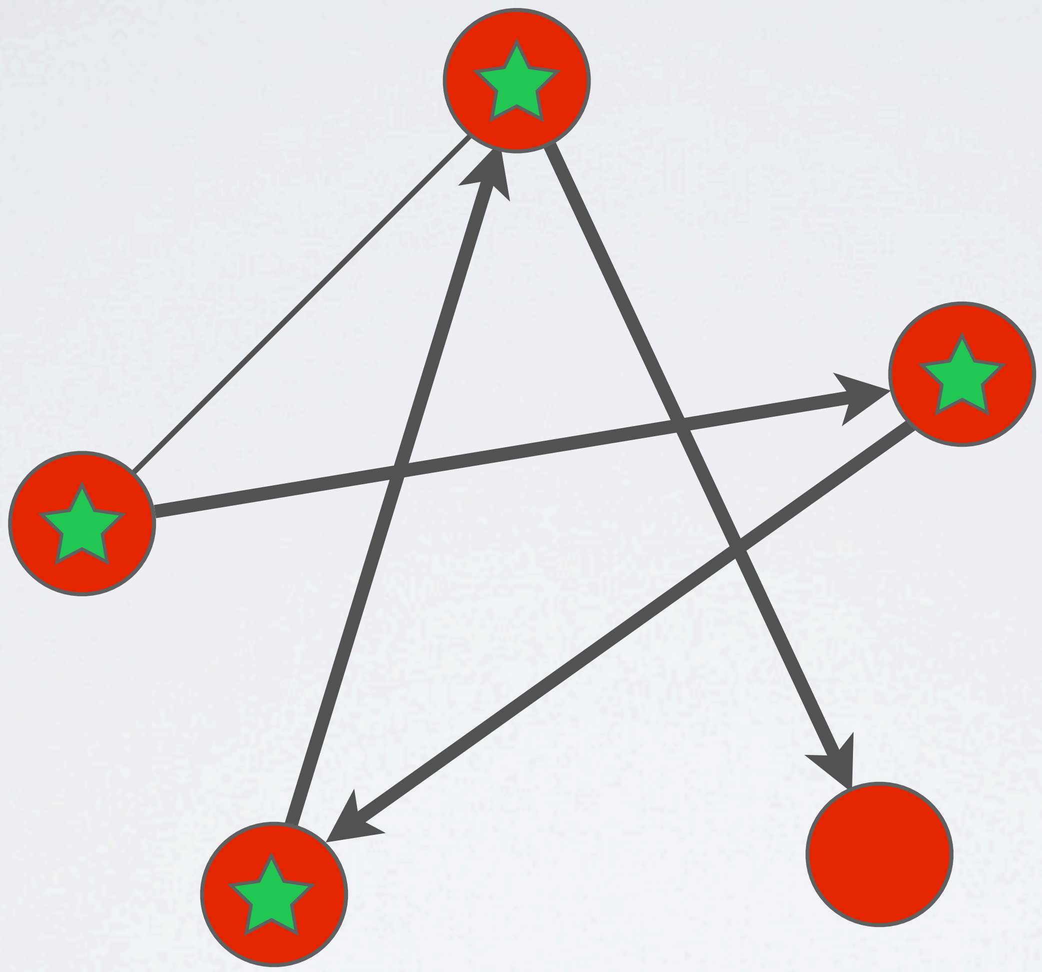
In each round, each informed node “pushes” the information to a neighbor chosen UAR.

Push



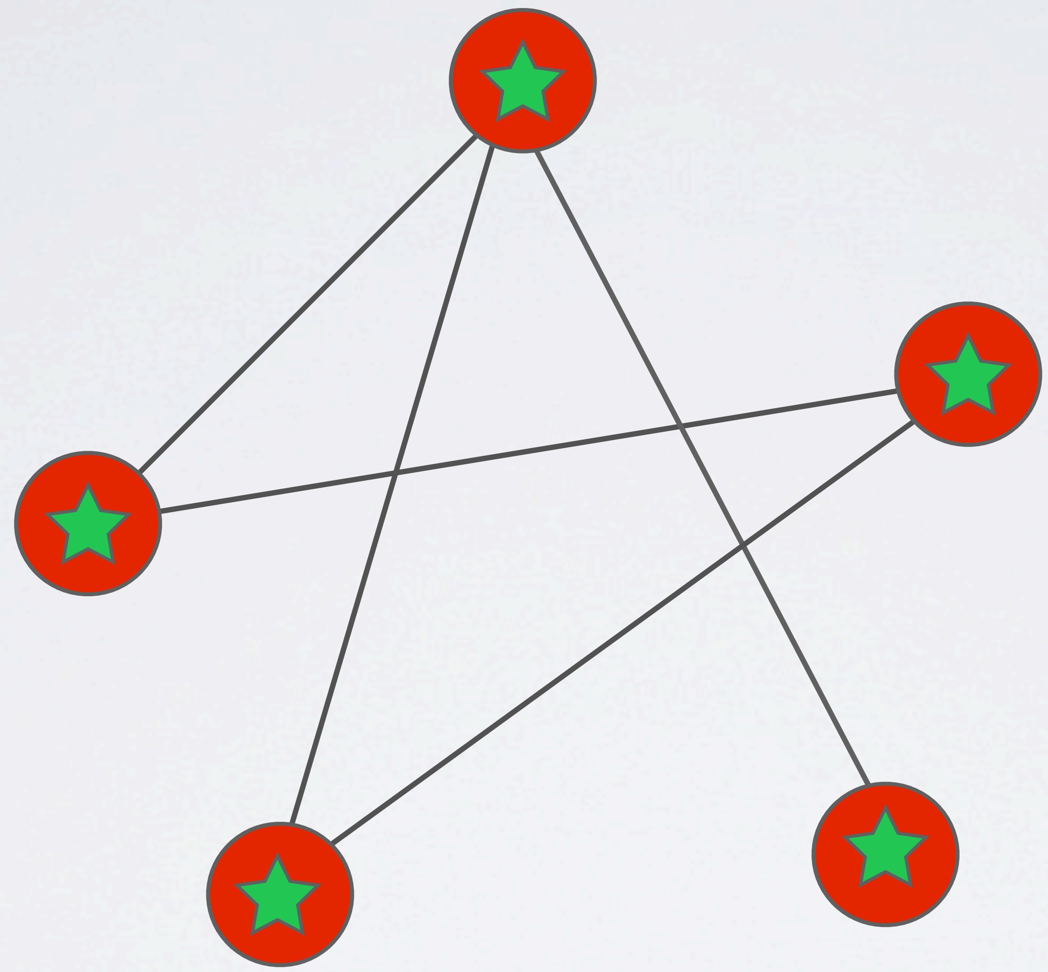
In each round, each informed node "pushes" the information to a neighbor chosen UAR.

Push



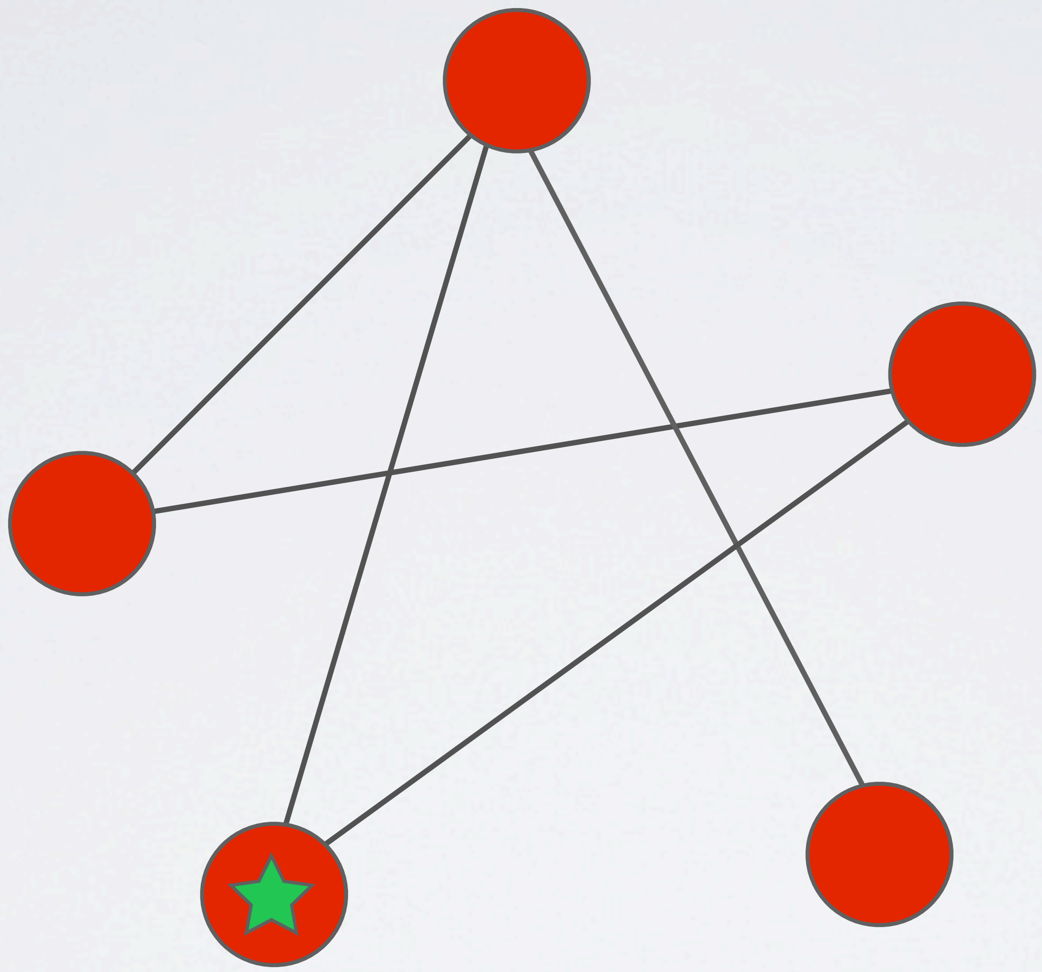
In each round, each informed node “pushes” the information to a neighbor chosen UAR.

Push

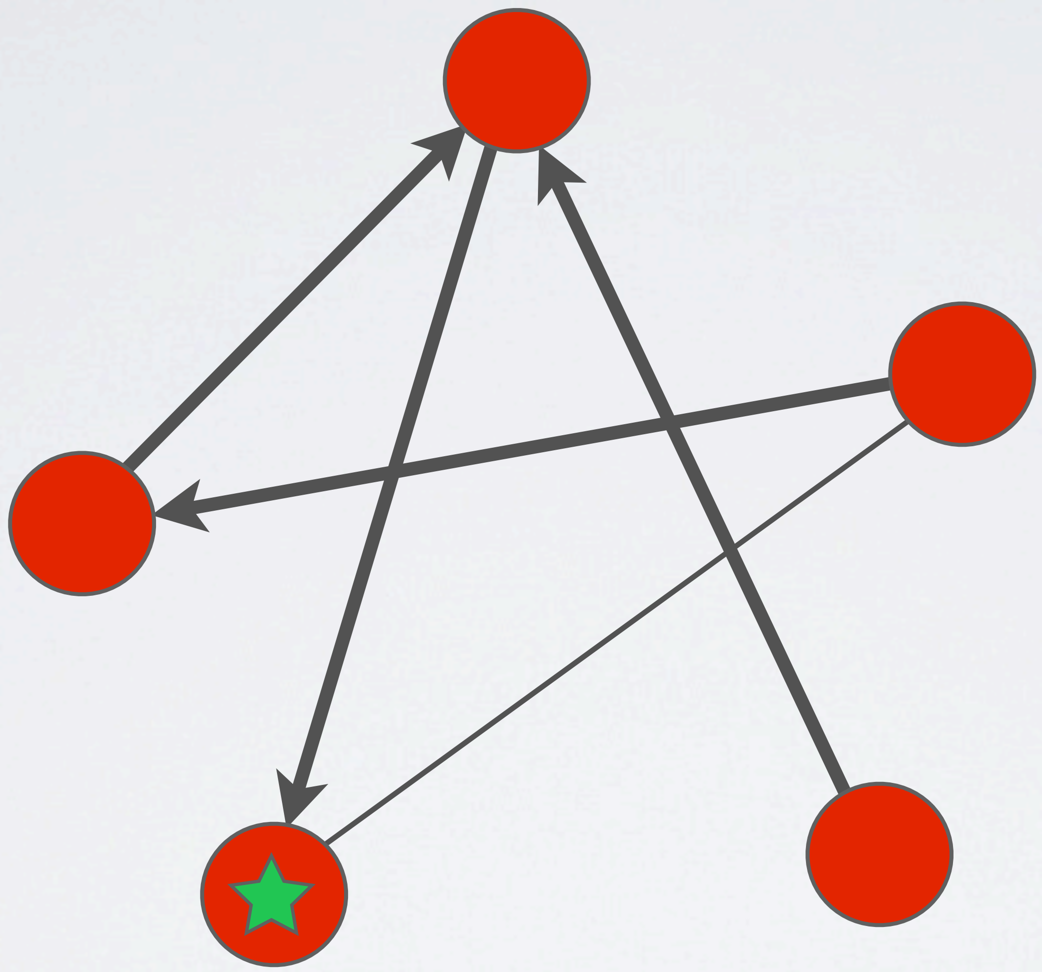


In each round, each informed node “pushes” the information to a neighbor chosen UAR.

Pull

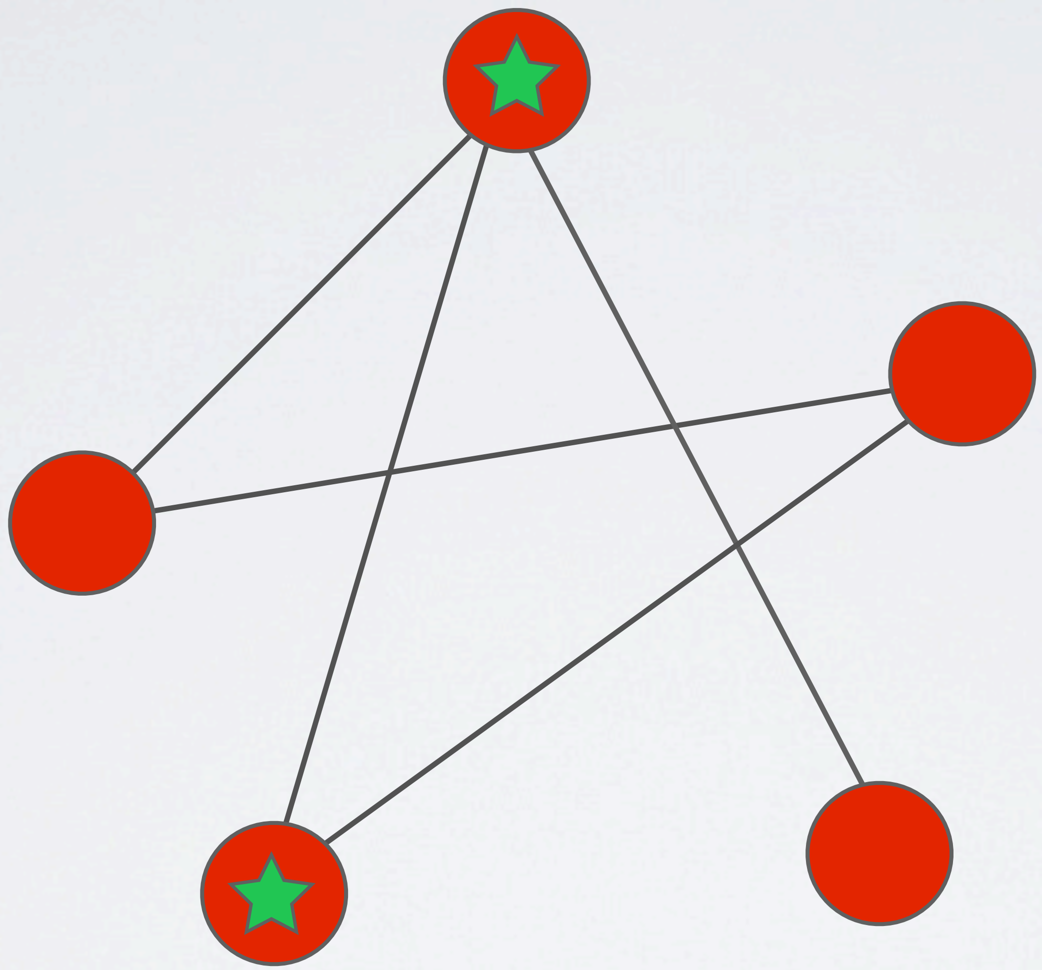


Pull



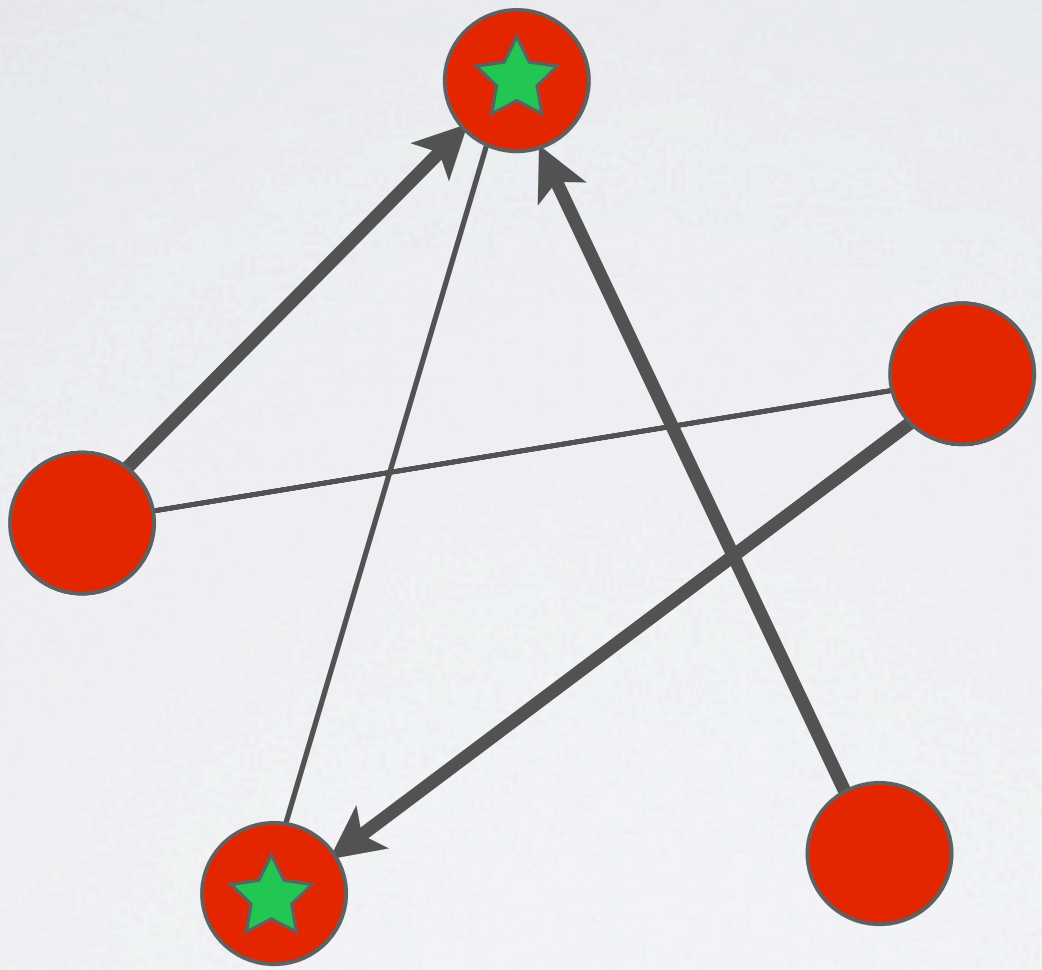
In each round, each uninformed node tries to “pull” the information from a neighbor chosen UAR.

Pull



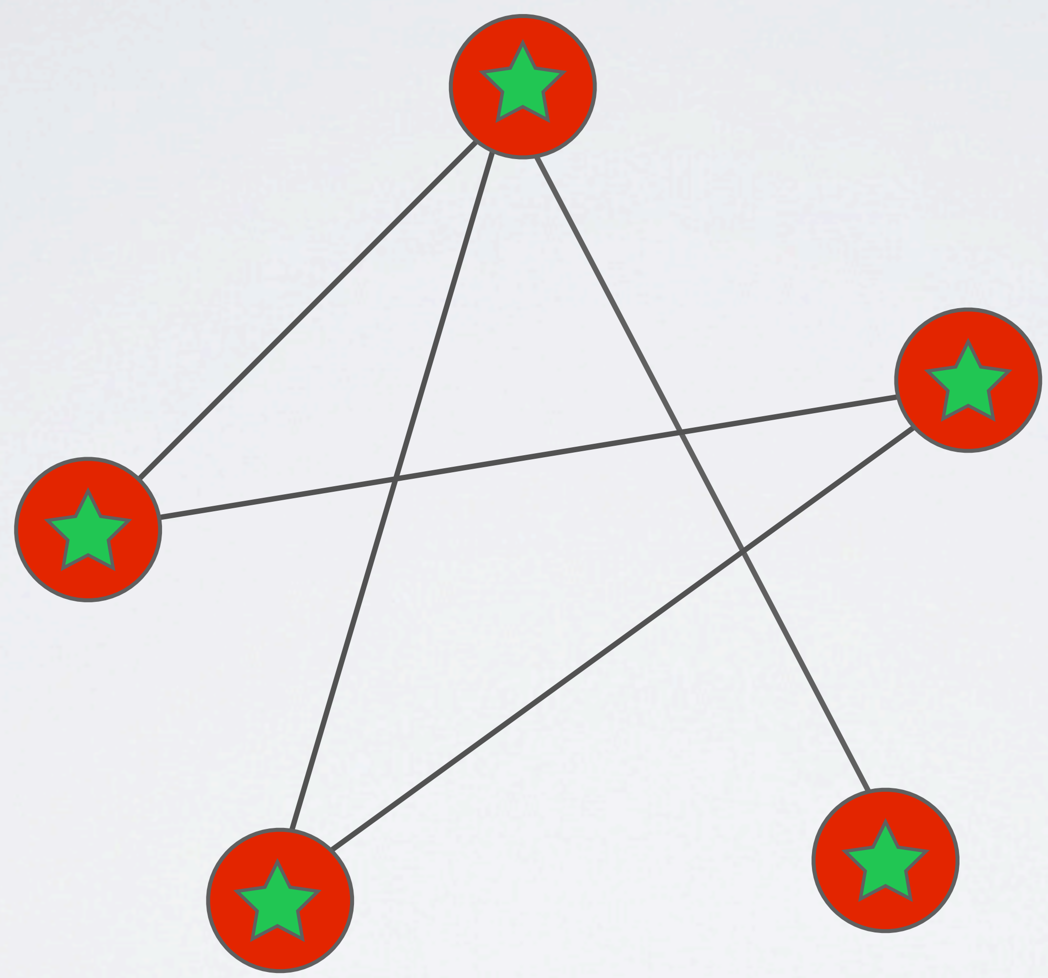
In each round, each uninformed node tries to “pull” the information from a neighbor chosen UAR.

Pull



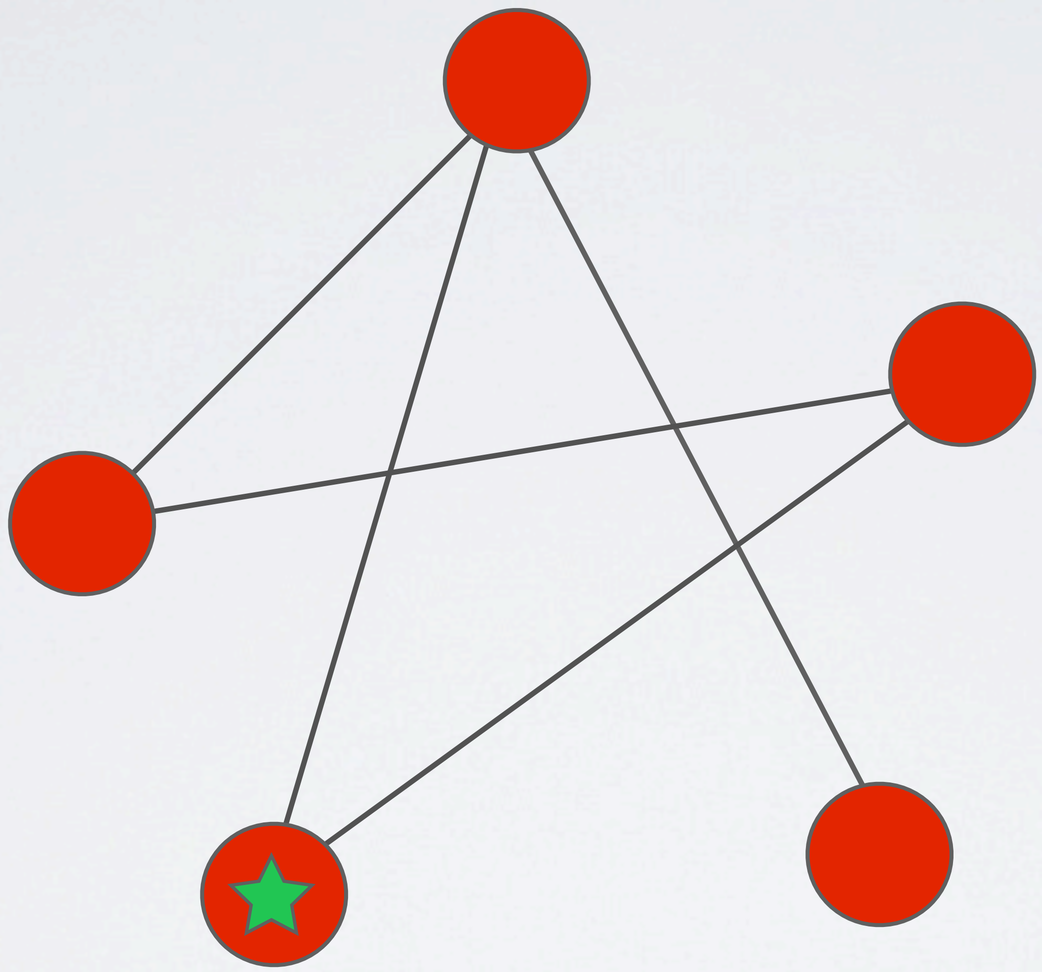
In each round, each uninformed node tries to “pull” the information from a neighbor chosen UAR.

Pull

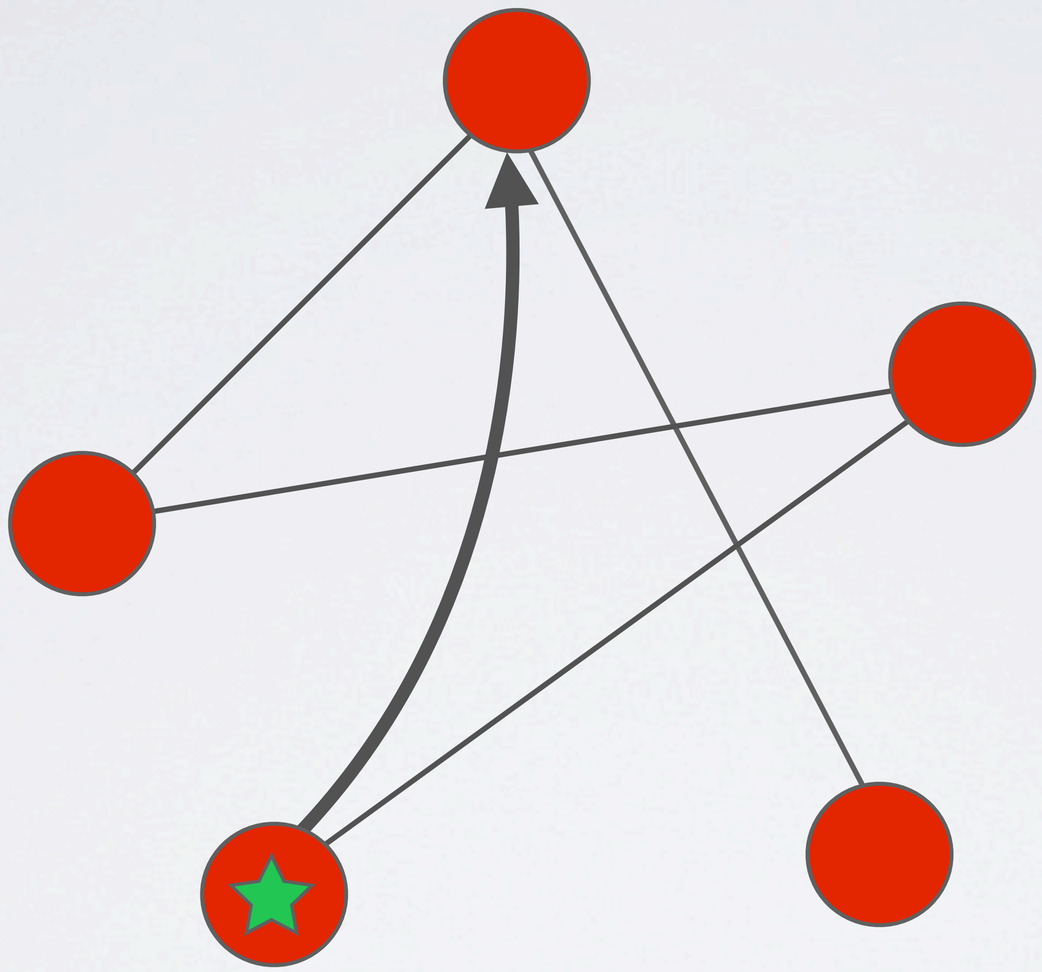


In each round, each uninformed node tries to "pull" the information from a neighbor chosen UAR.

Push-Pull

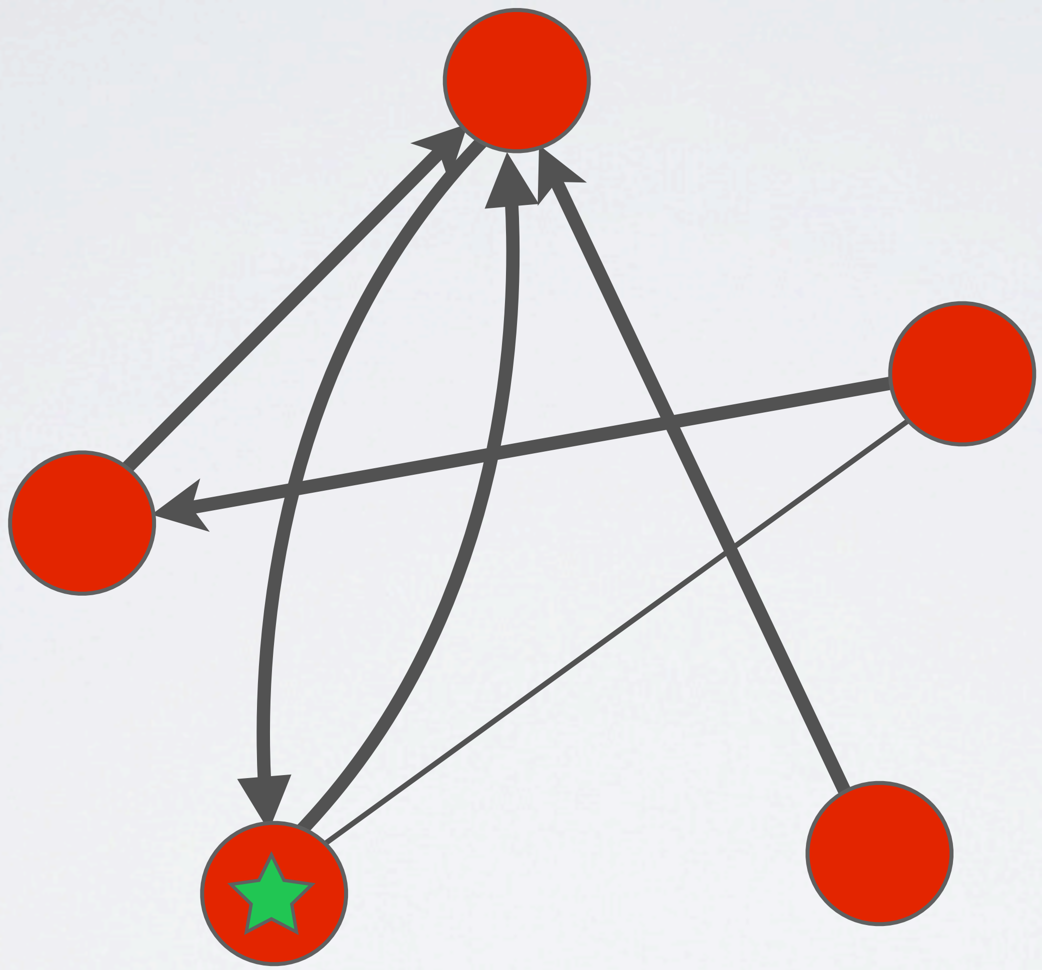


Push-Pull



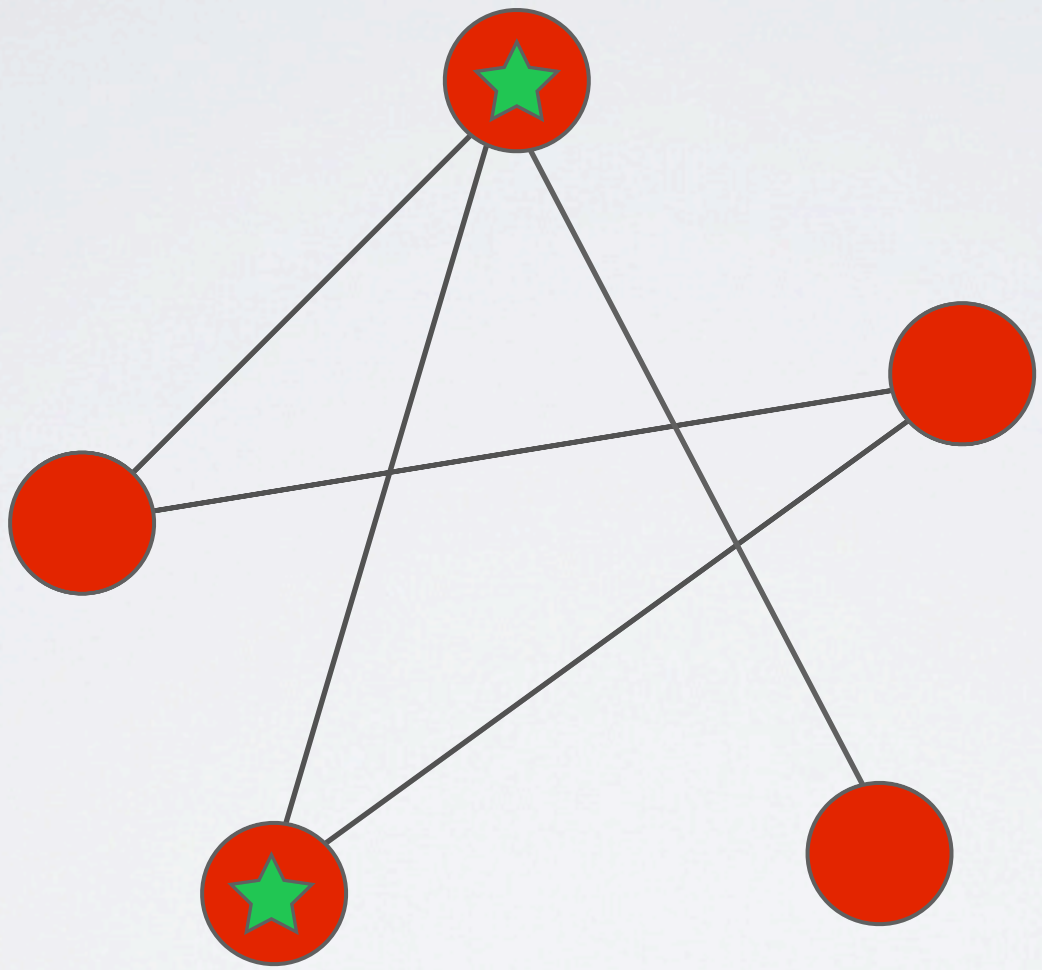
Each node performs both actions on neighbor chosen UAR.

Push-Pull



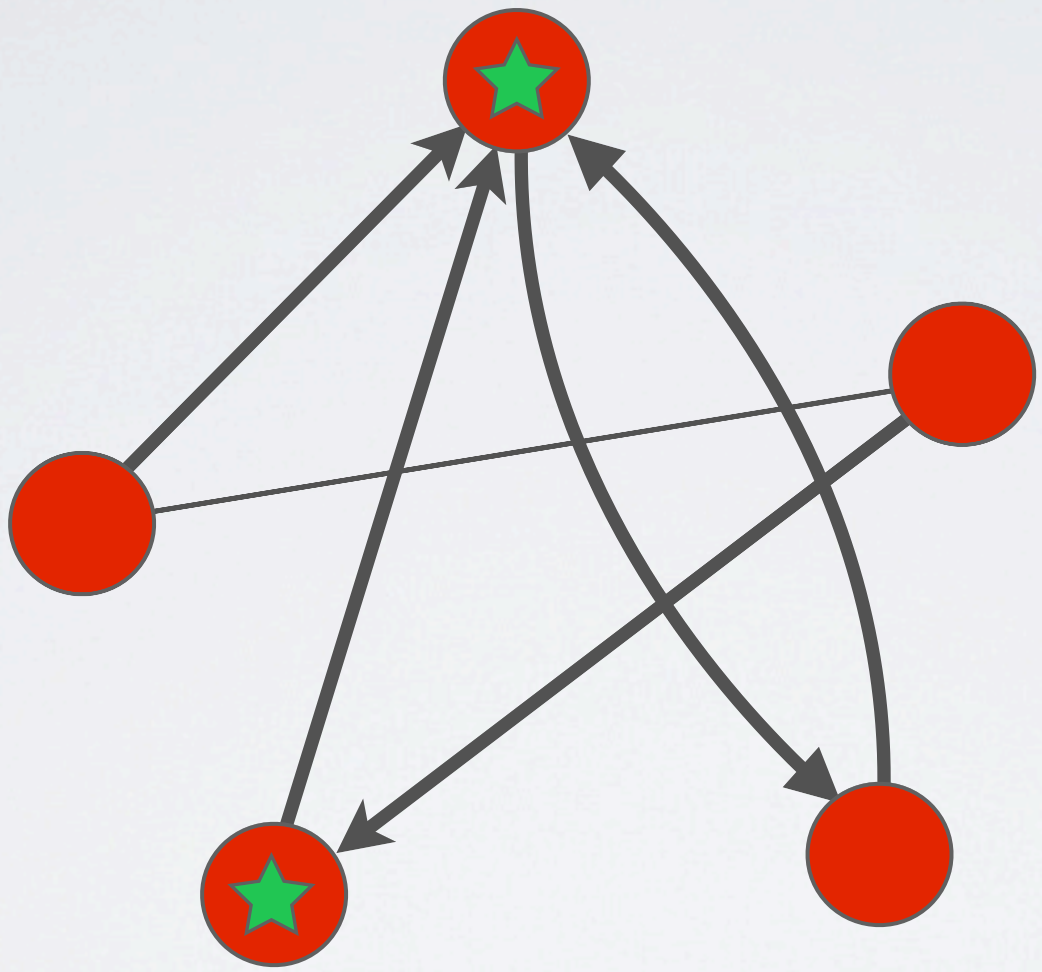
Each node performs both actions on neighbor chosen UAR.

Push-Pull



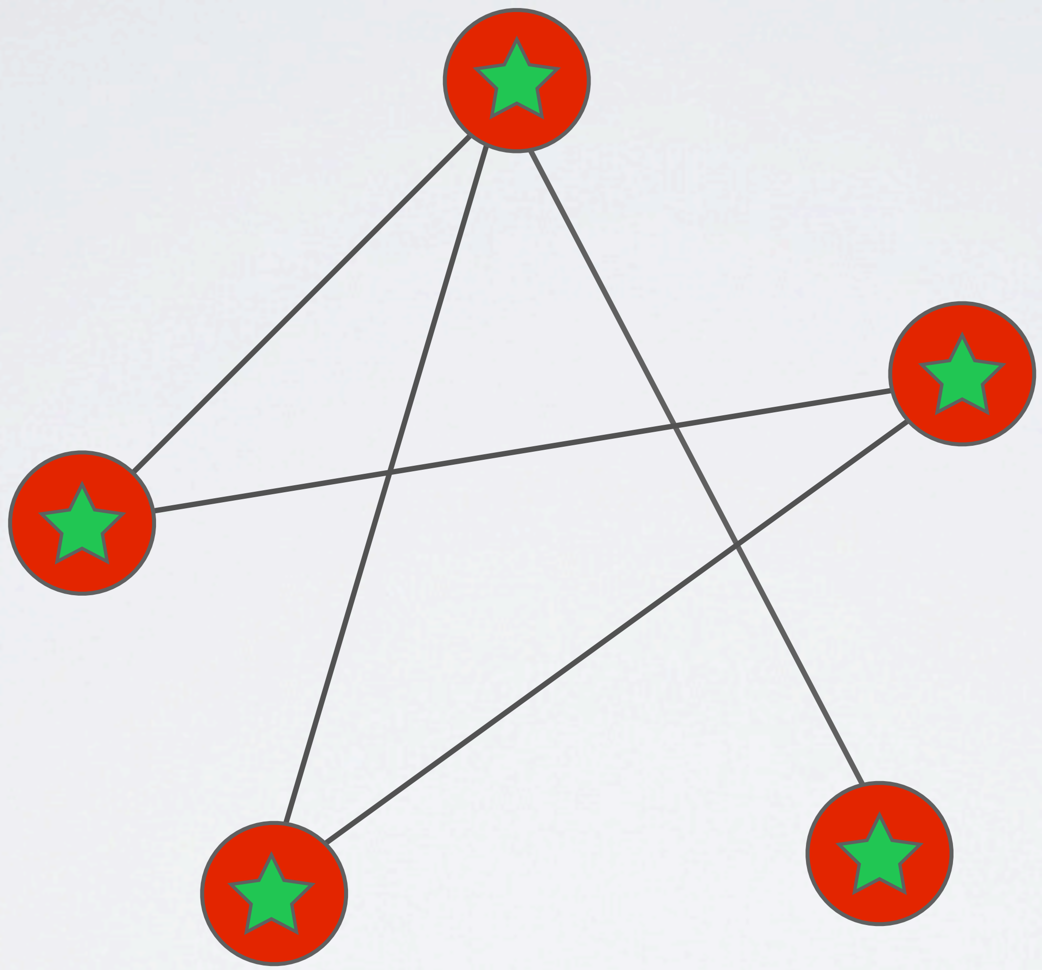
Each node performs both actions on neighbor chosen UAR.

Push-Pull



Each node performs both actions on neighbor chosen UAR.

Push-Pull



Each node performs both actions on neighbor chosen UAR.

Push

Pull

Performance

- What are the completion times T_{PUSH}, T_{PULL}, $T_{PUSH\text{-}PULL}$?

- How many rounds will it take for each node to know the information with probability $1 - o(1)$, assuming a *worst-case* source?

Previous work

- $\boldsymbol{T_{PUSH}}, \boldsymbol{T_{PULL}} = \Theta(\log n)$ if $G = K_n$

  Frieze, Grimmet, *Algorithms* 1985

Previous work

- T_{PUSH}, T_{PULL} = $\Theta(\log n)$ if $G = K_n$

  Frieze, Grimmet, *Algorithms* 1985

- $T_{PUSH} \leq O(n \log n)$

  $T_{PUSH} \leq O(\Delta(G)\ (\ diam(G) + \log n\ ))$

  $T_{PUSH} \leq O(\log n)$ in Hypercubes and $G(n,p)$ Graphs

  Feige, Peleg, Raghavan, Upfal, *Algorithms* 1990

Previous work

- T_{PUSH}, T_{PULL} = $\Theta(\log n)$ if $G = K_n$

  Frieze, Grimmet, *Algorithms* 1985

Huge in Social Networks!

- $T_{PUSH} \leq O(n \log n)$

  $T_{PUSH} \leq O(\Delta(G) (\text{ diam}(G) + \log n))$

  $T_{PUSH} \leq O(\log n)$ in Hypercubes and $G(n,p)$ Graphs

  Feige, Peleg, Raghavan, Upfal, *Algorithms* 1990

Previous work

- T_{PUSH} O(log n) in "quasi-regular" expanders

  Berenbrink, Elsässer, Friedetzky, *PODC* 2008
  Doerr, Friedrich, Sauerwald, *ICALP* 2009
  Sauerwald, *ISAAC* 2007

- $T_{PUSH\text{-}PULL}$ $O(\log^2 n)$ in PA graphs

  $T_{PUSH\text{-}PULL}$ $O(\text{poly}(\Phi^{-1} \log n))$ if conductance = Φ

  Chierichetti, Lattanzi, Panconesi, *ICALP* 2009, *SODA* 2010

- Non uniform rumor spreading and conductance

  Boyd, Ghosh, Prabhakar, Shah, *IEEE Transaction on Information Theory* 2006
  Mosk-Aoyama, Shah, *IEEE Transaction on Information Theory* 2008

Previous work

- T_{PUSH} O(log n) in "quasi-regular" expanders
  Berenbrink, Elsässer, Friedetzky, *PODC* 2008
  Doerr, Friedrich, Sauerwald, *ICALP* 2009
  Sauerwald, *ISAAC* 2007

Social Networks are highly irregular!

- $T_{PUSH\text{-}PULL}$ $O(\log^2 n)$ in PA graphs
  $T_{PUSH\text{-}PULL}$ $O(\mathrm{poly}(\Phi^{-1} \log n))$ if conductance = Φ
  Chierichetti, Lattanzi, Panconesi, *ICALP* 2009, *SODA* 2010

- Non uniform rumor spreading and conductance
  Boyd, Ghosh, Prabhakar, Shah, *IEEE Transaction on Information Theory* 2006
  Mosk-Aoyama, Shah, *IEEE Transaction on Information Theory* 2008

Previous work

- T_{PUSH} O(log n) in "quasi-regular" expanders

  Berenbrink, Elsässer, Friedetzky, *PODC* 2008
  Doerr, Friedrich, Sauerwald, *ICALP* 2009
  Sauerwald, *ISAAC* 2007

Connections to Spielman-Teng sparsification theory

- $T_{PUSH\text{-}PULL}$ $O(\log^2 n)$ in PA graphs

  $T_{PUSH\text{-}PULL}$ $O(\text{poly}(\Phi^{-1} \log n))$ if conductance = Φ

  Chierichetti, Lattanzi, Panconesi, *ICALP* 2009, *SODA* 2010

- Non uniform rumor spreading and conductance

  Boyd, Ghosh, Prabhakar, Shah, *IEEE Transaction on Information Theory* 2006
  Mosk-Aoyama, Shah, *IEEE Transaction on Information Theory* 2008

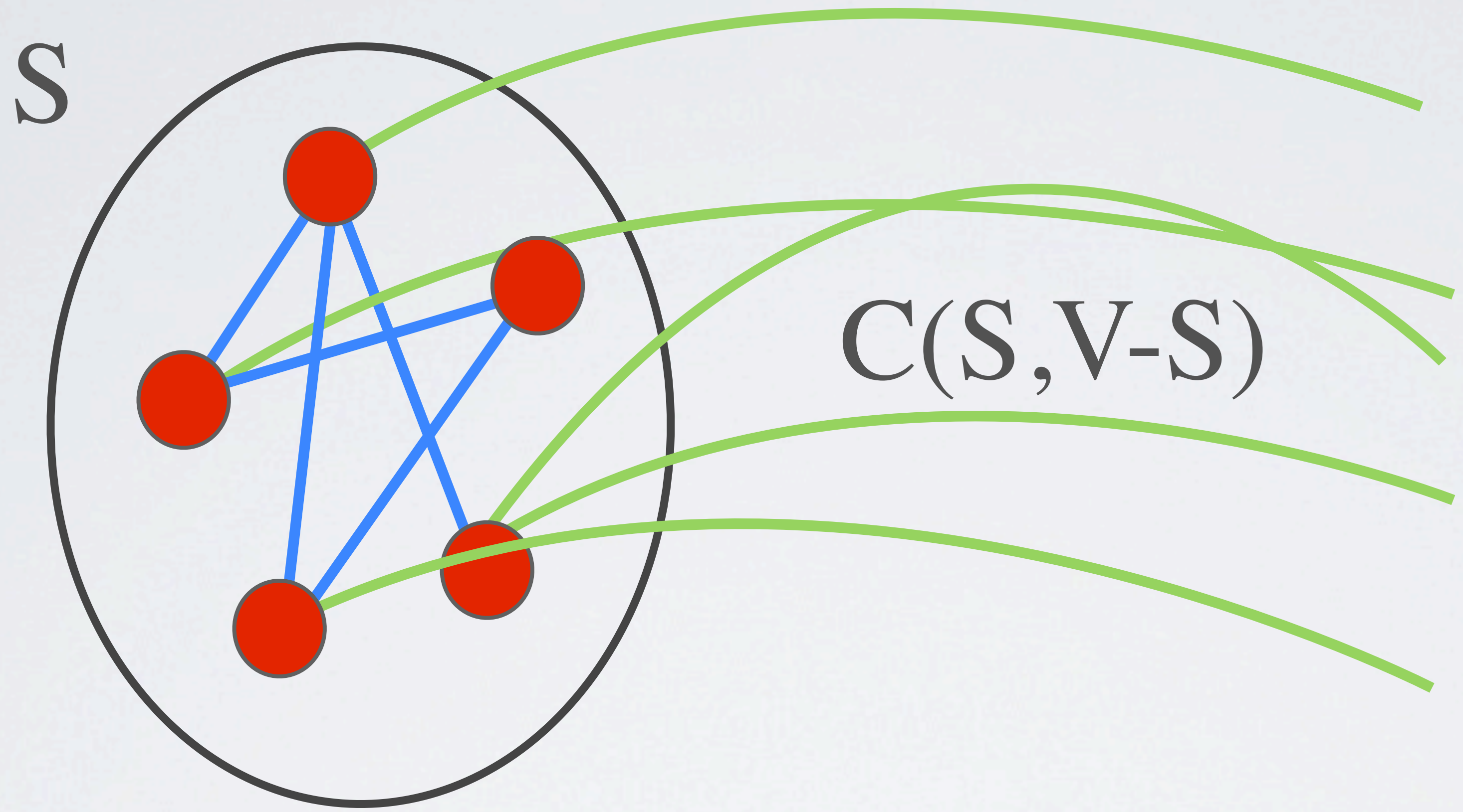
Social networks

Empirical evidence

- Leskovec, Lang, Dasgupta and Mahoney give empirical evidence that social networks have conductance $\Omega\left(\frac{1}{\log n}\right)$

- Can we relate the performance of rumor spreading algorithms with the conductance of the graph?

Conductance



$$\phi(G) = \min_{\substack{S \subseteq V \\ \mathrm{vol}(S) \leq |E|}} \frac{|C(S, V - S)|}{\min(\mathrm{vol}(S), \mathrm{vol}(V - S))}$$

Social networks

↓

high conductance

Social networks

↓

<u>high conductance</u>

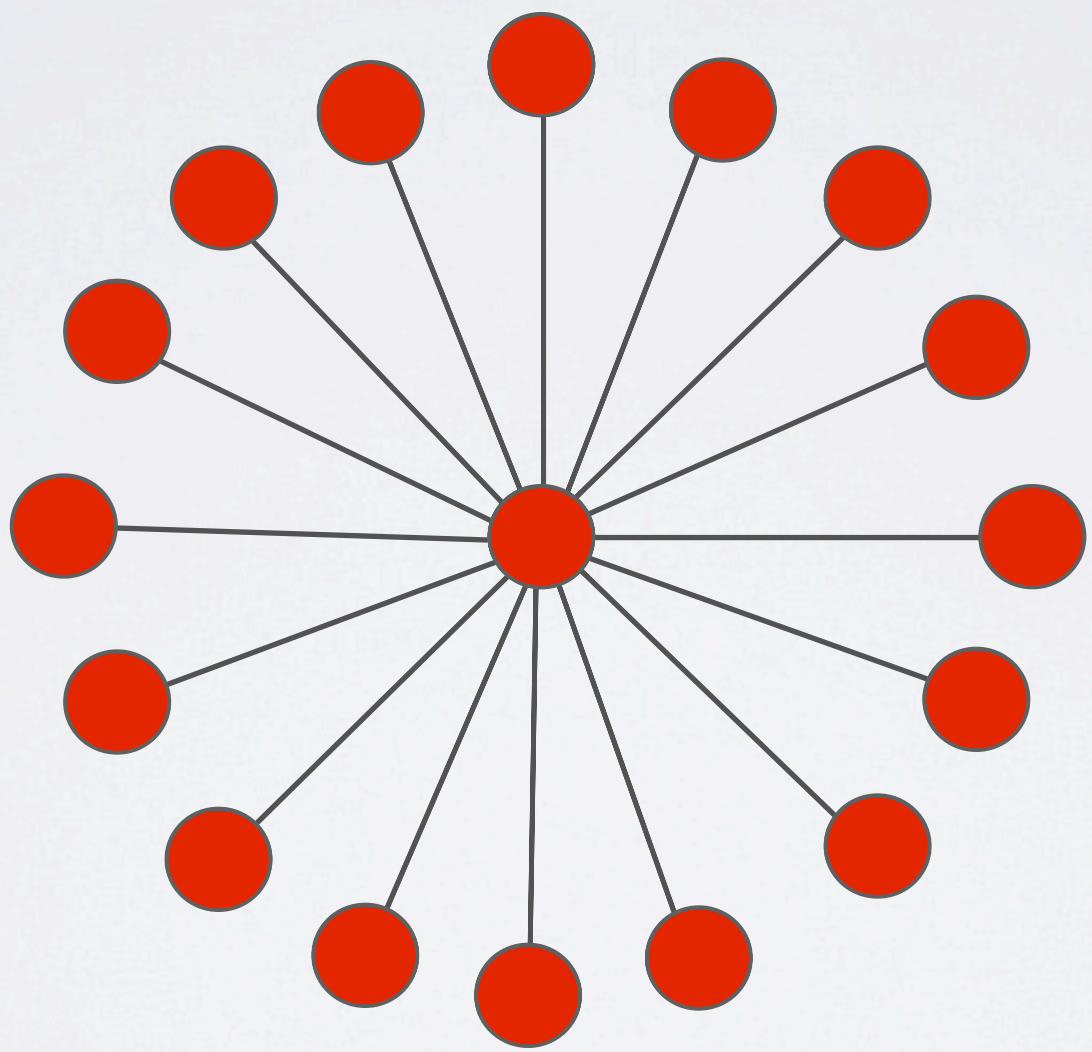
Is rumor spreading fast on high conductance graphs?

T_{PUSH}

Is the *PUSH* strategy fast?

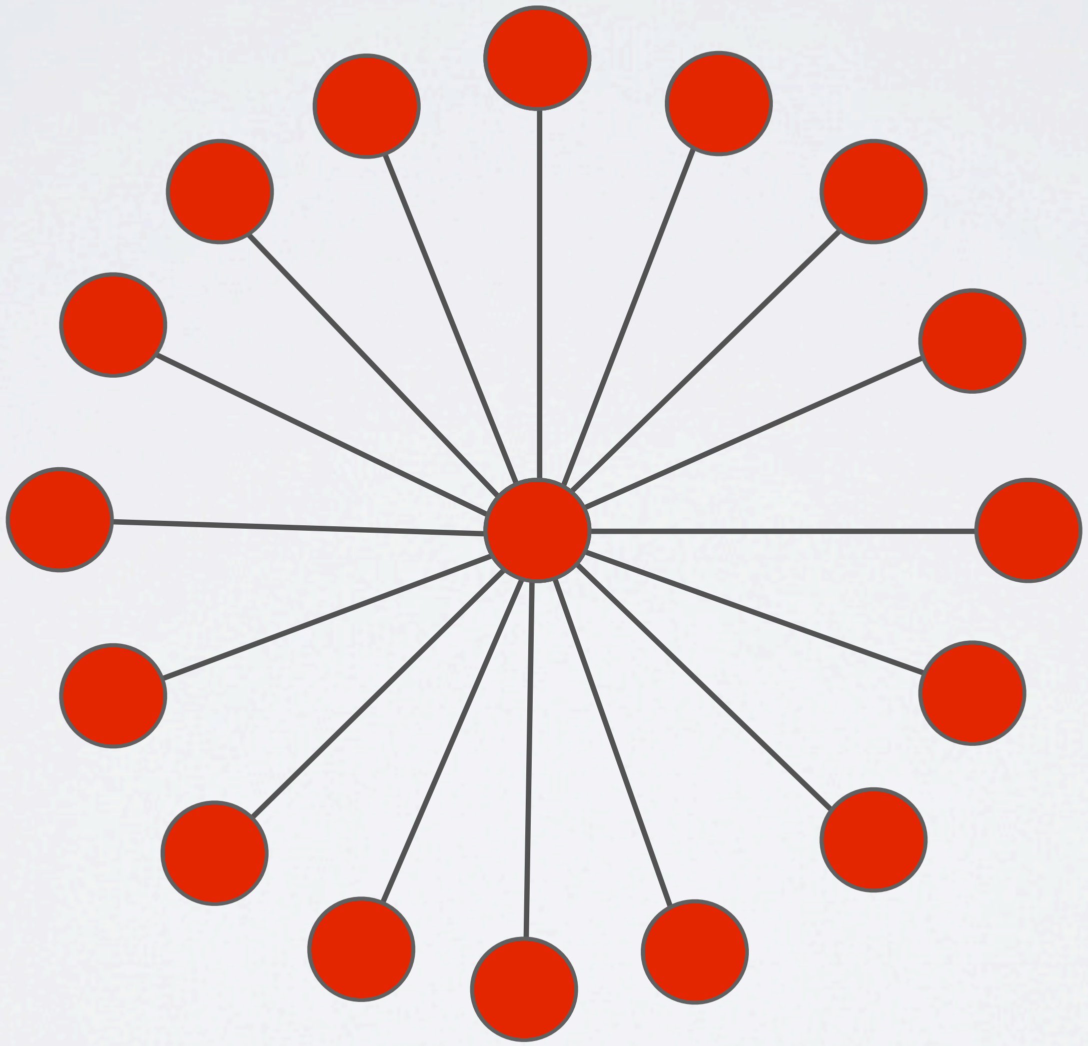
T_{PUSH}

Is the *PUSH* strategy fast?



T_{PUSH}

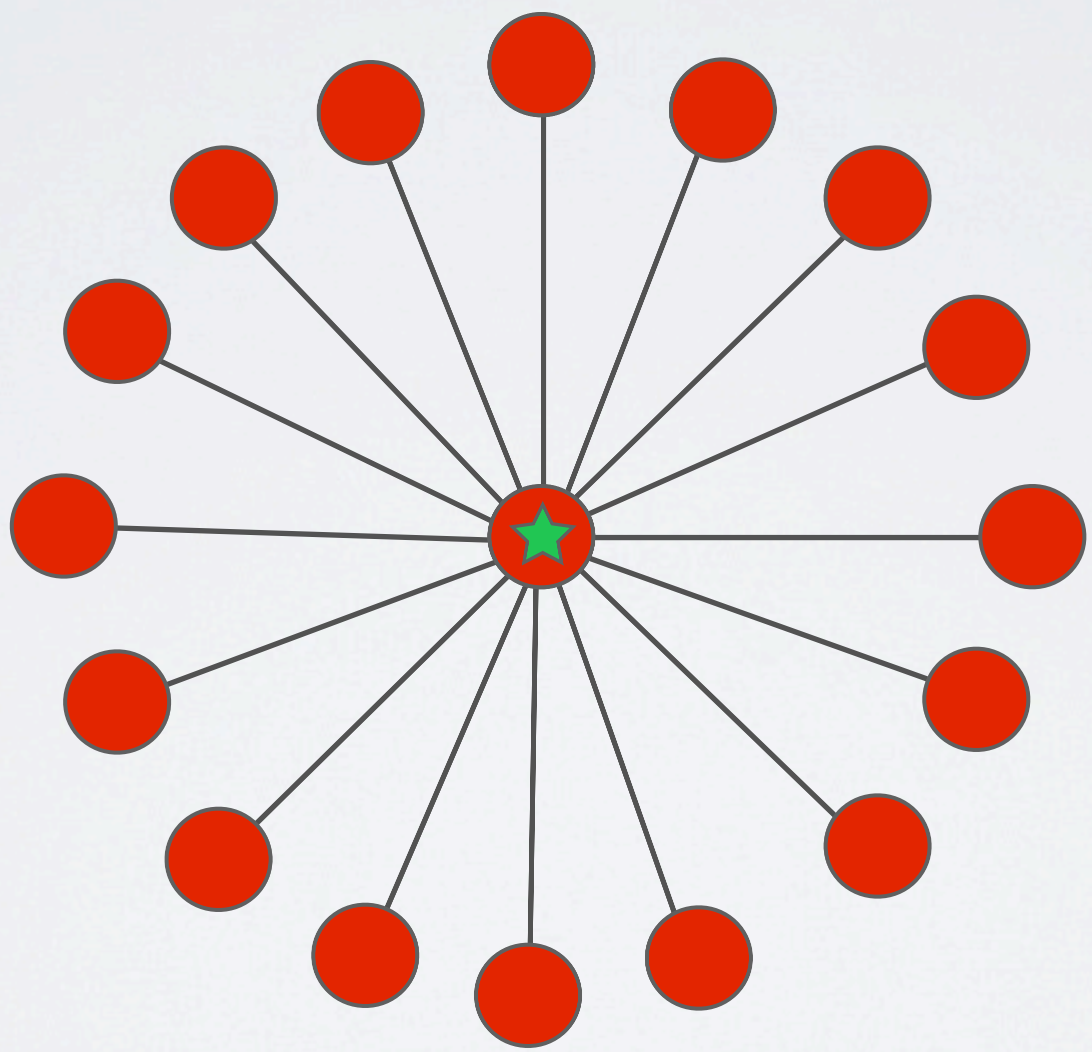
Is the *PUSH* strategy fast?



The star has constant conductance.

T_{PUSH}

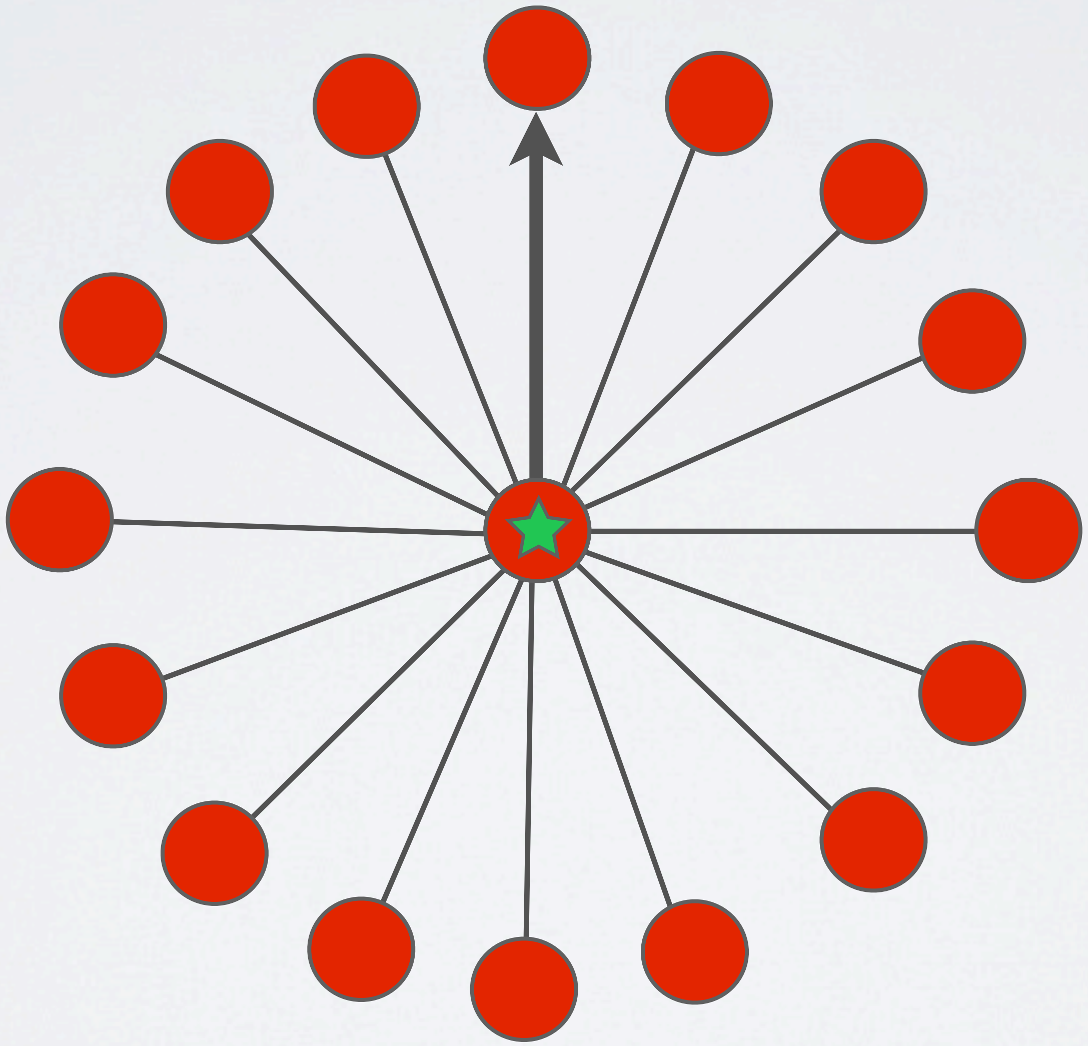
Is the *PUSH* strategy fast?



The star has constant conductance.

*T**PUSH*

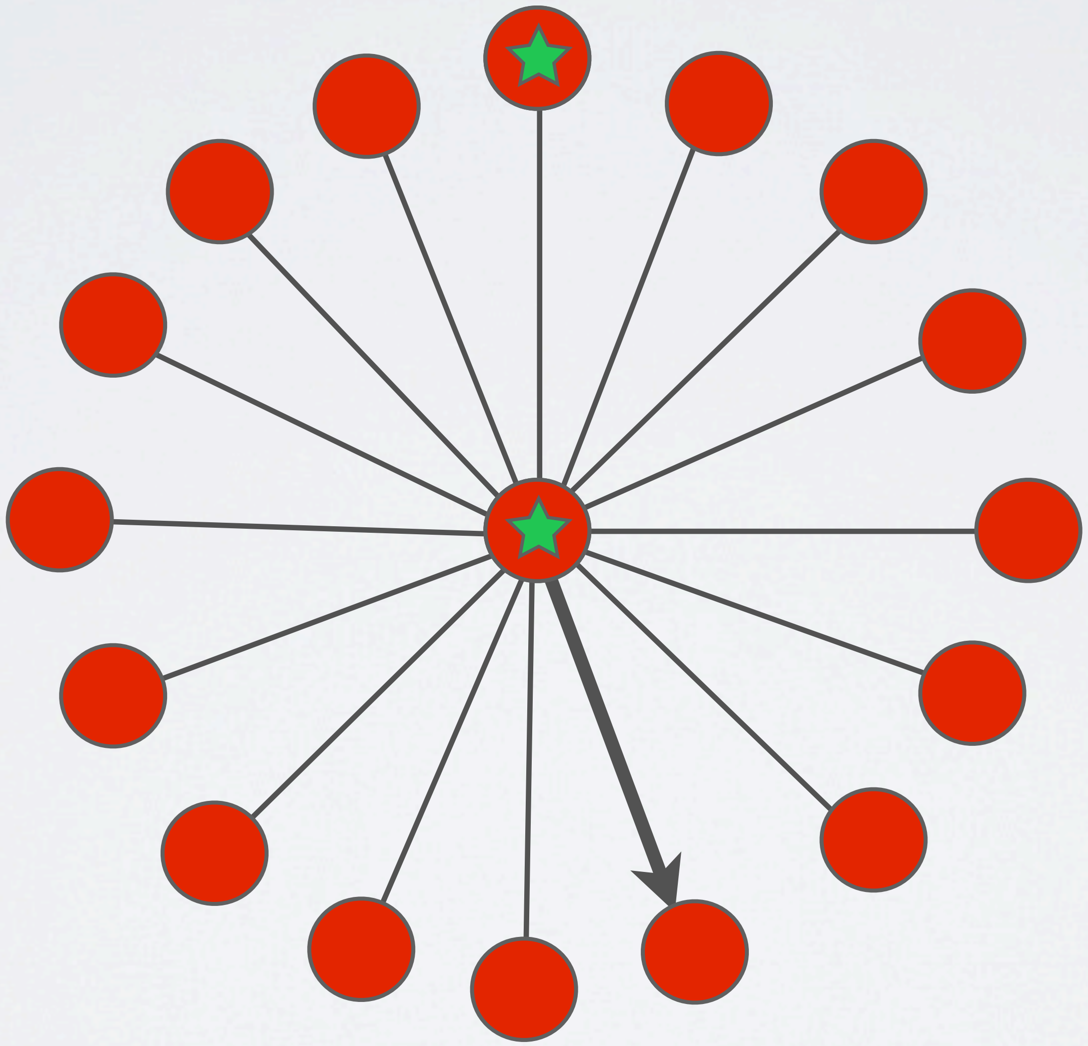
Is the *PUSH* strategy fast?



The star has constant conductance.

T_{PUSH}

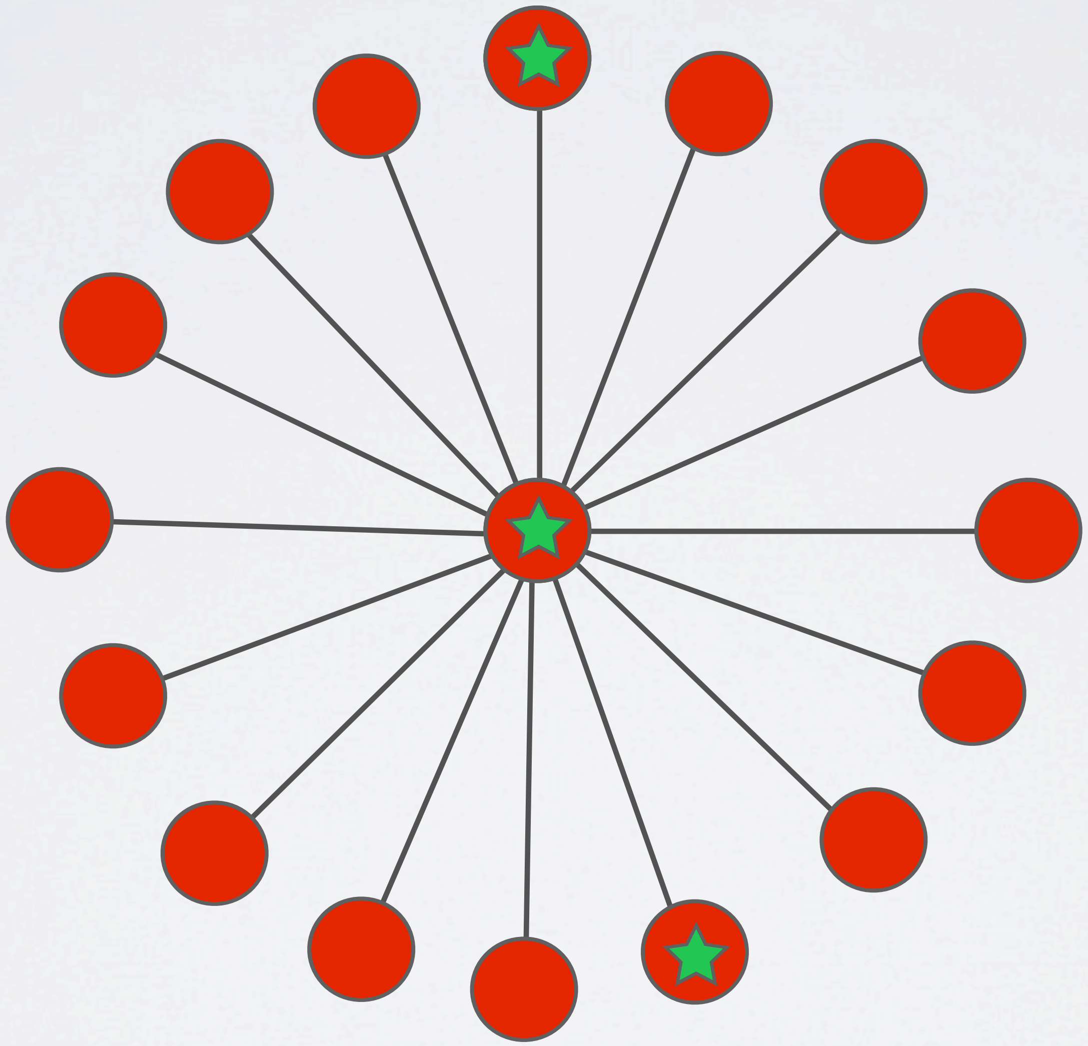
Is the *PUSH* strategy fast?



The star has constant conductance.

*T*PUSH

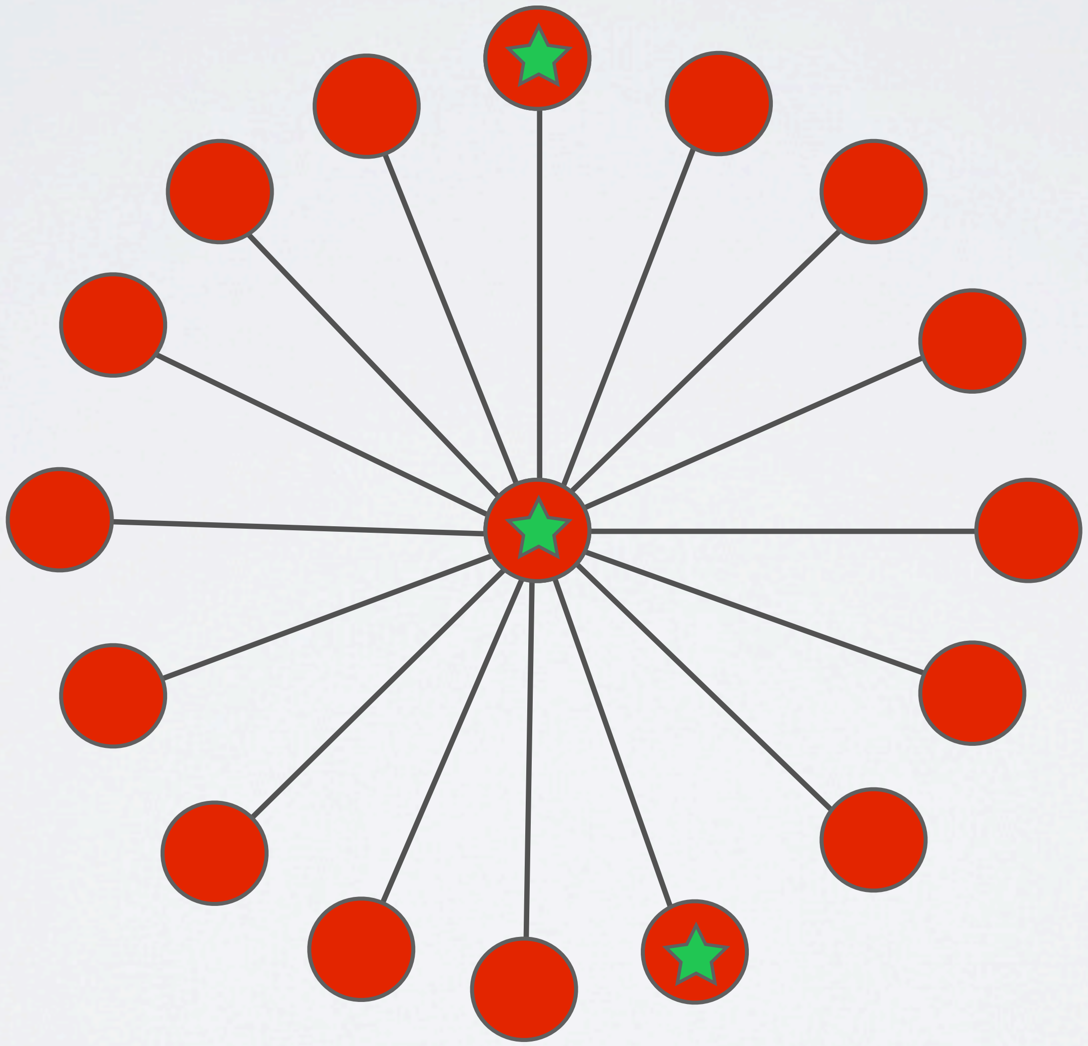
Is the *PUSH* strategy fast?



Coupon Collector!

T$_{PUSH}$

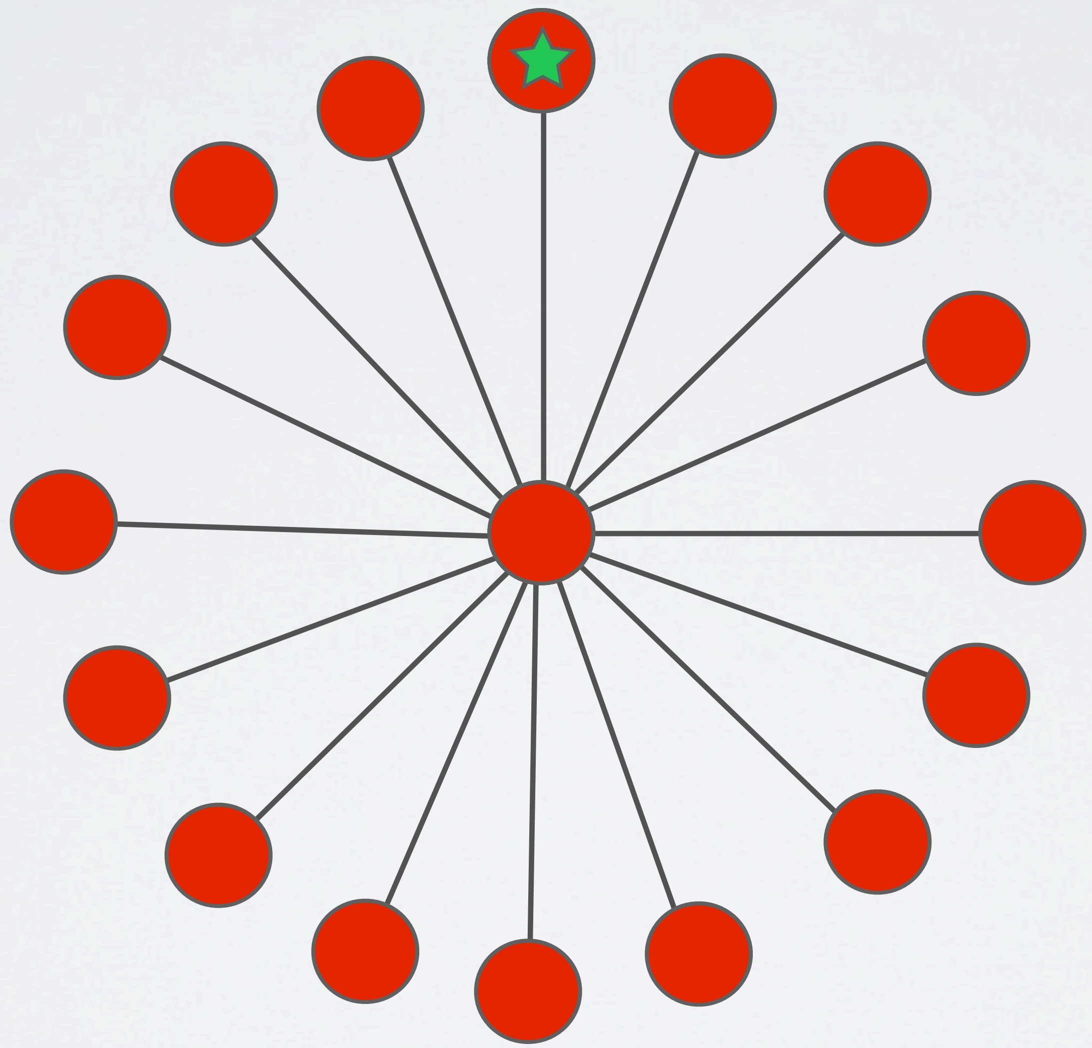
Is the *PUSH* strategy fast? **NO**



Coupon Collector!

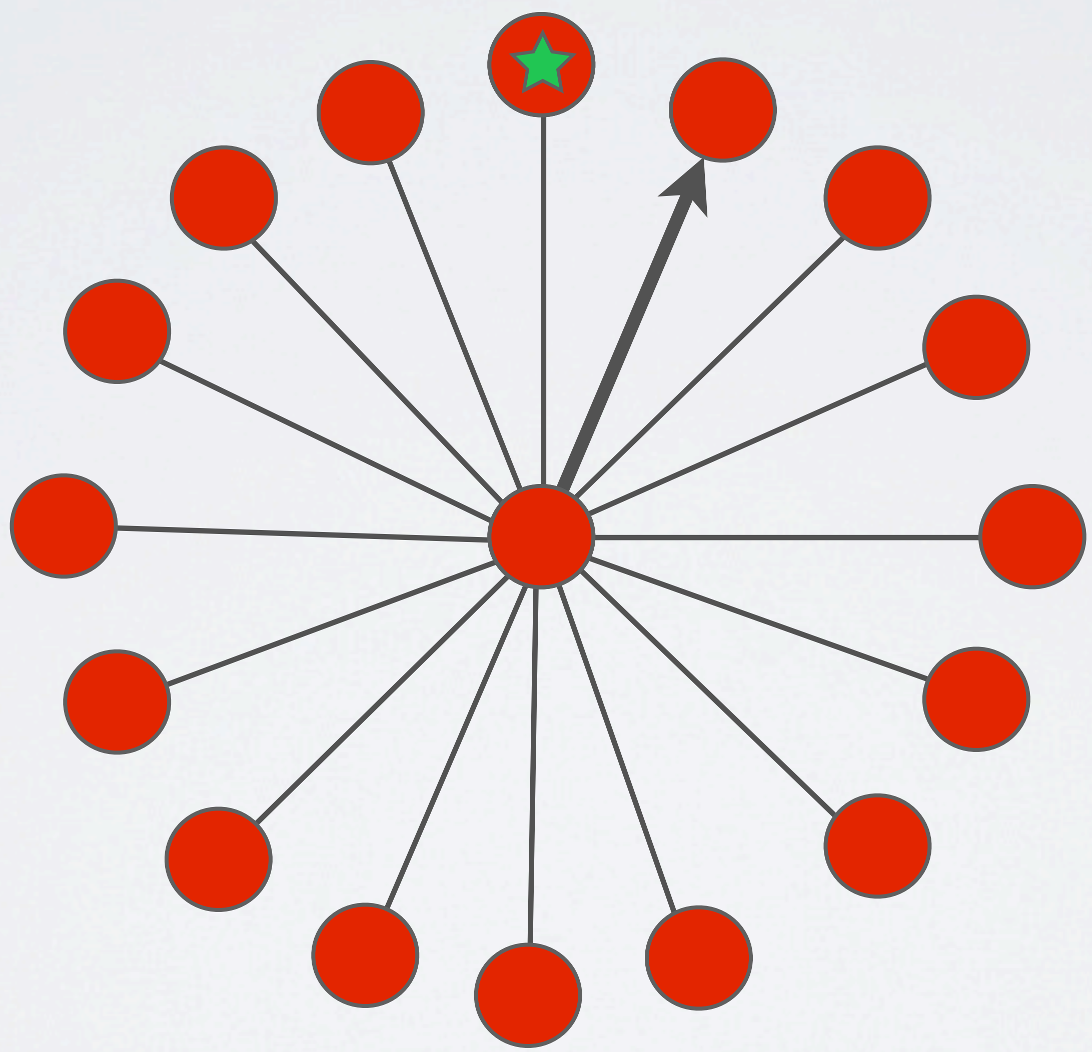
TPULL

Is the *PULL* strategy fast?



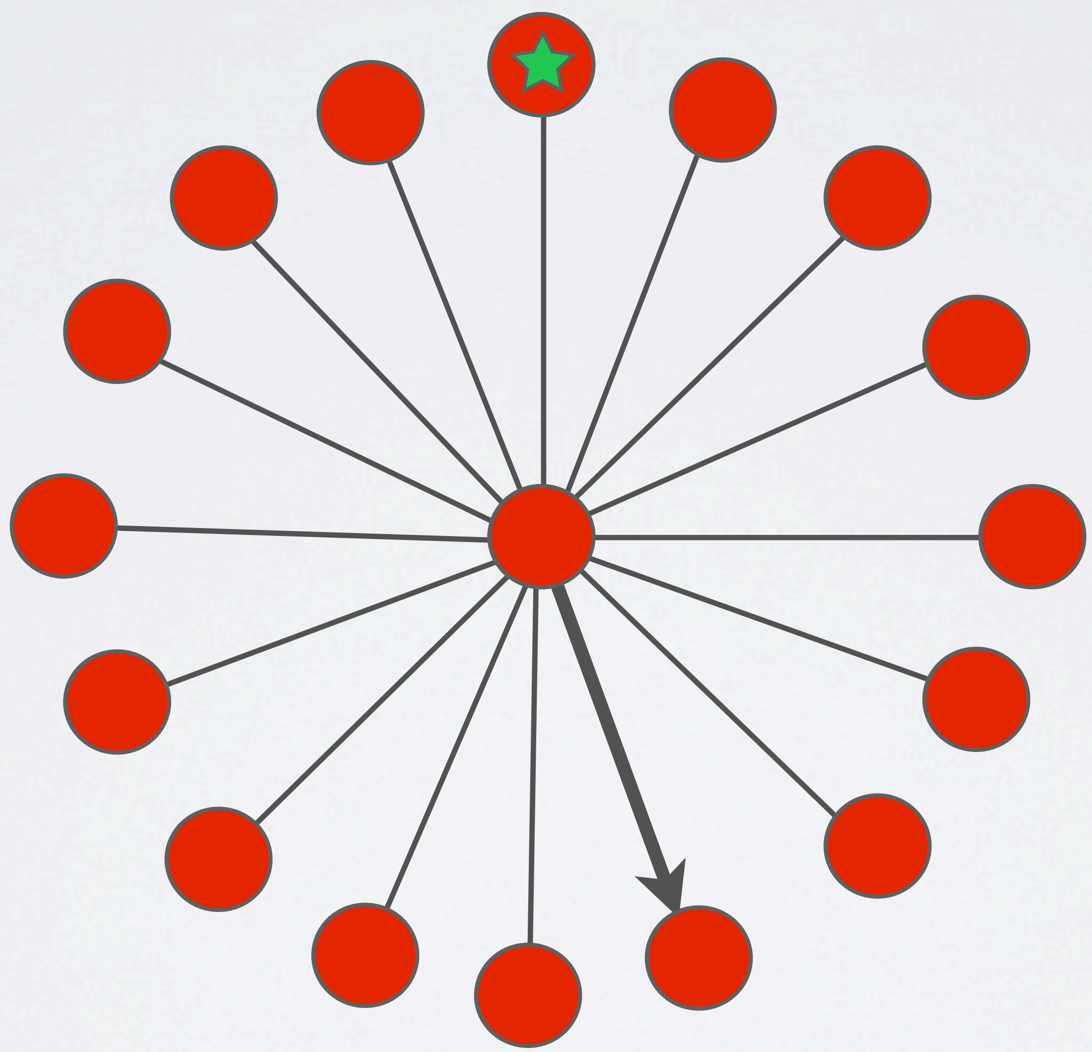
T PULL

Is the *PULL* strategy fast?



T

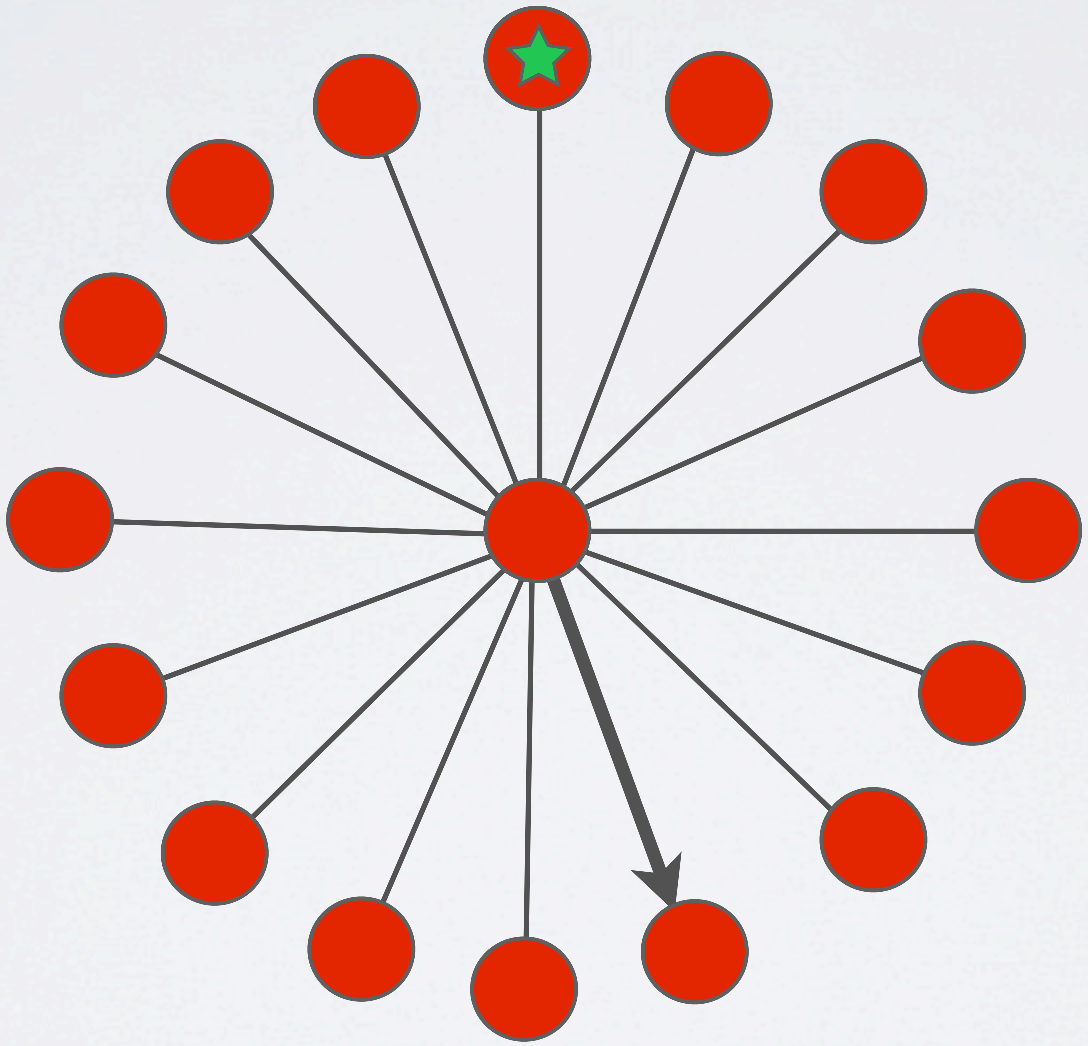
Is the *PULL* strategy fast?



The central node has to *PULL* the information from the right node.

T PULL

Is the *PULL* strategy fast? **NO**



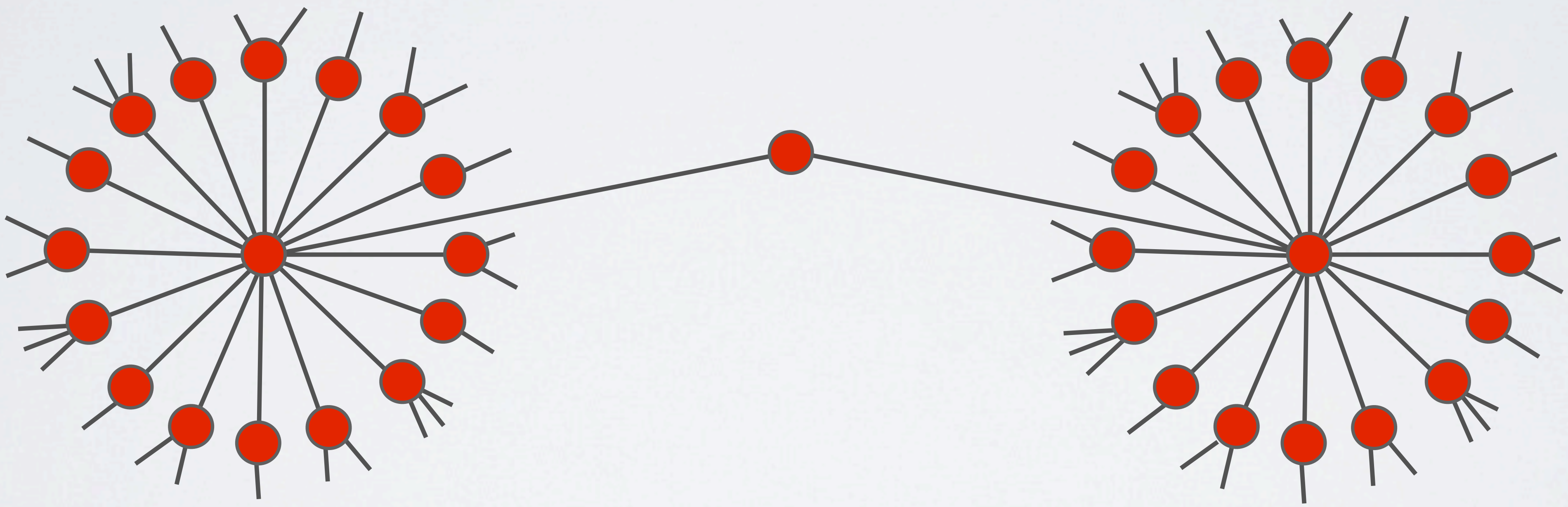
The central node has to *PULL* the information from the right node.

T_{PUSH} and T_{PULL} in social networks

Social networks do not look like a star...

T_{PUSH} and T_{PULL} in social networks

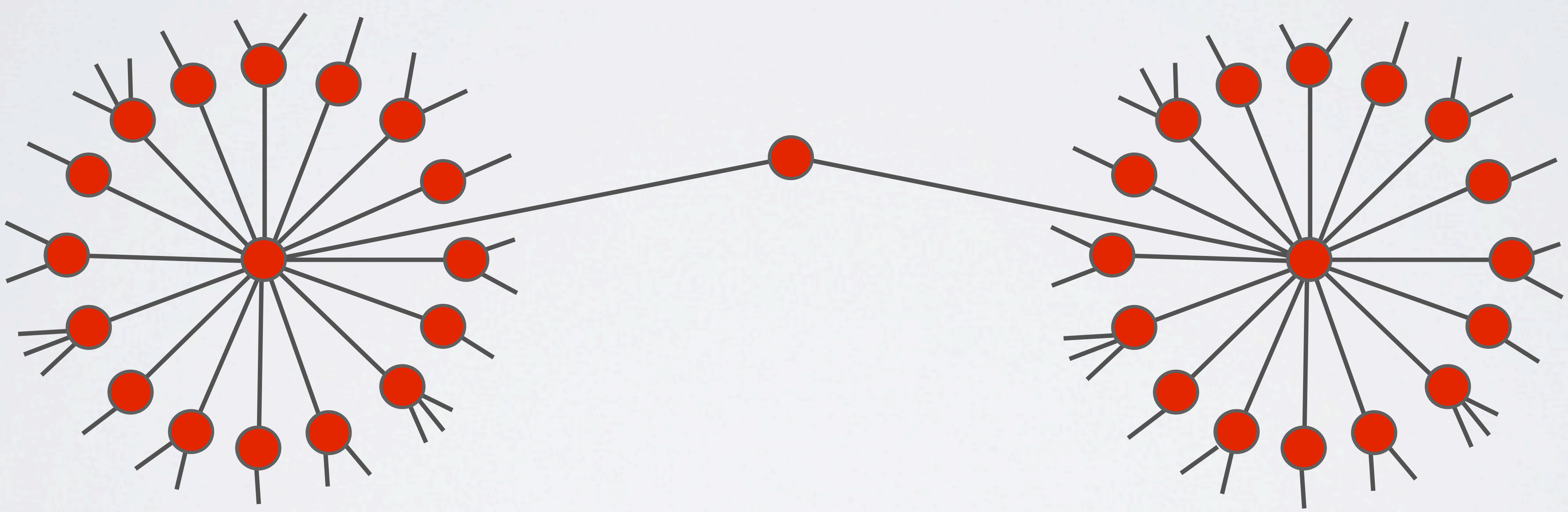
Social networks do not look like a star...



... but there are low degree nodes connected only to high degree nodes.

T_{PUSH} and T_{PULL} in social networks

Social networks do not look like a star...



... but there are low degree nodes connected only to high degree nodes.

SAME ISSUES!!!

$T_{\text{PUSH-PULL}}$?

Theorem

Let G be a graph with conductance Φ, then w.h.p.

$$T_{PUSH-PULL} = O\left(\frac{\log n}{\Phi}\left(\log\frac{1}{\Phi}\right)^2\right)$$

Theorem

Let G be a graph with conductance Φ, then w.h.p.

$$T_{PUSH-PULL} = O\left(\frac{\log n}{\Phi}\left(\log\frac{1}{\Phi}\right)^2\right)$$

$$T_{PUSH-PULL} = \Omega\left(\frac{\log n}{\Phi}\right)$$

Lower bound

- Take any 3-regular graph of constant vertex expansion of order $\Theta(n\Phi)$ and diameter $\Theta(\log n)$

Lower bound

- Take any 3-regular graph of constant vertex expansion of order $\Theta(n\Phi)$ and diameter $\Theta(\log n)$

- Replace each edge with a path of length $\Theta(\Phi^{-1})$

Lower bound

- Take any 3-regular graph of constant vertex expansion of order $\Theta(n\Phi)$ and diameter $\Theta(\log n)$
- Replace each edge with a path of length $\Theta(\Phi^{-1})$
- The resulting graph will have order $\Theta(n)$, diameter $\Theta(\Phi^{-1} \log n)$ and conductance $\Theta(\Phi)$.

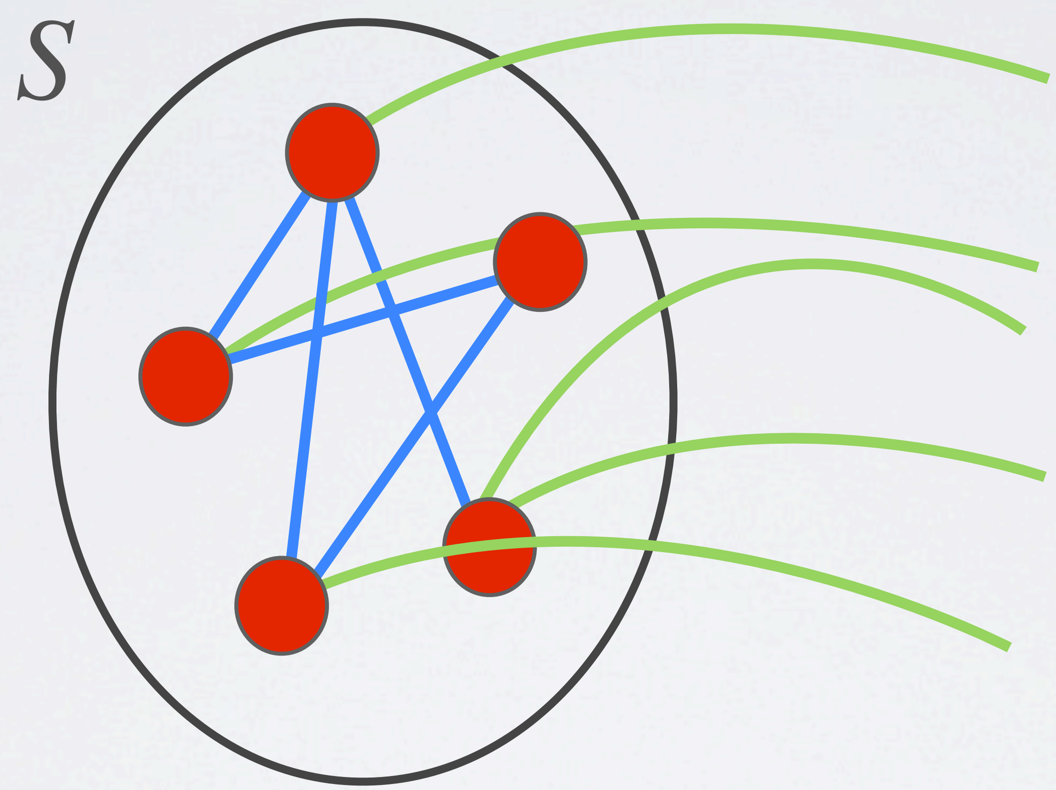
Upper bound

Let G be a graph with conductance Φ, then w.h.p.

$$T_{PUSH-PULL} = O\left(\frac{\log n}{\Phi^2}\right)$$

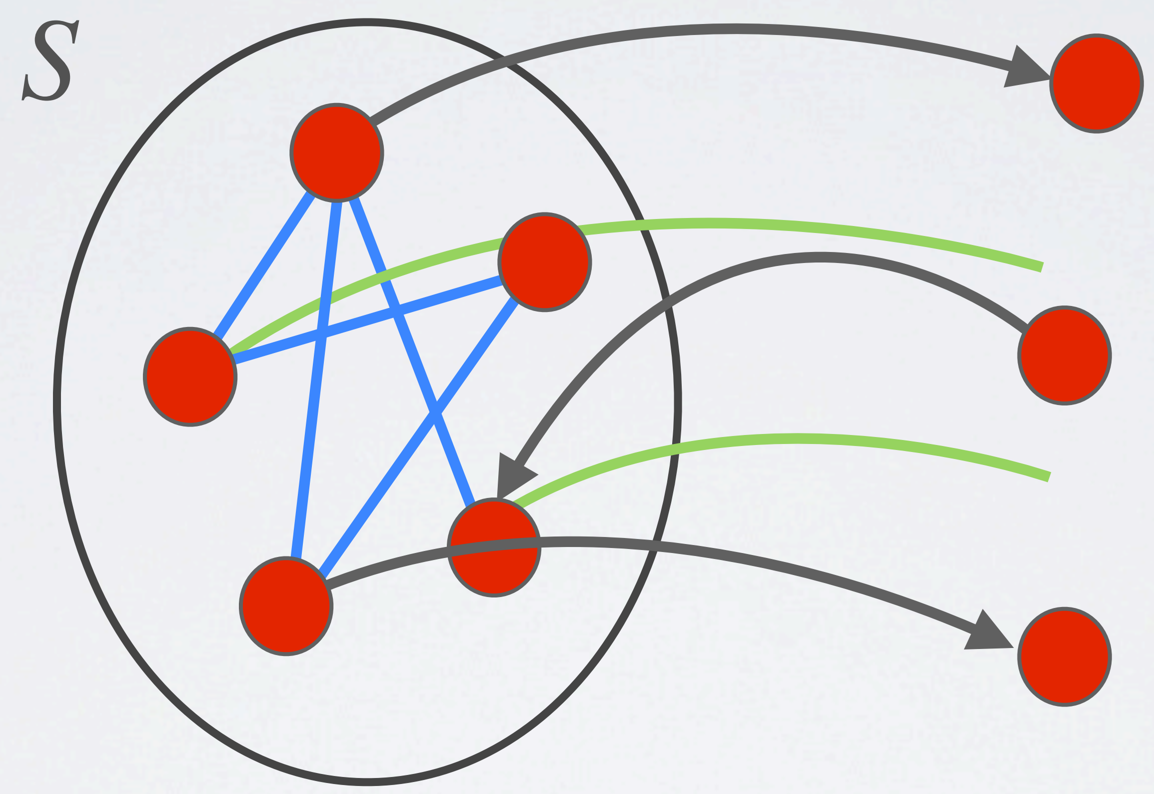
Proof strategy

Key lemma



Proof strategy

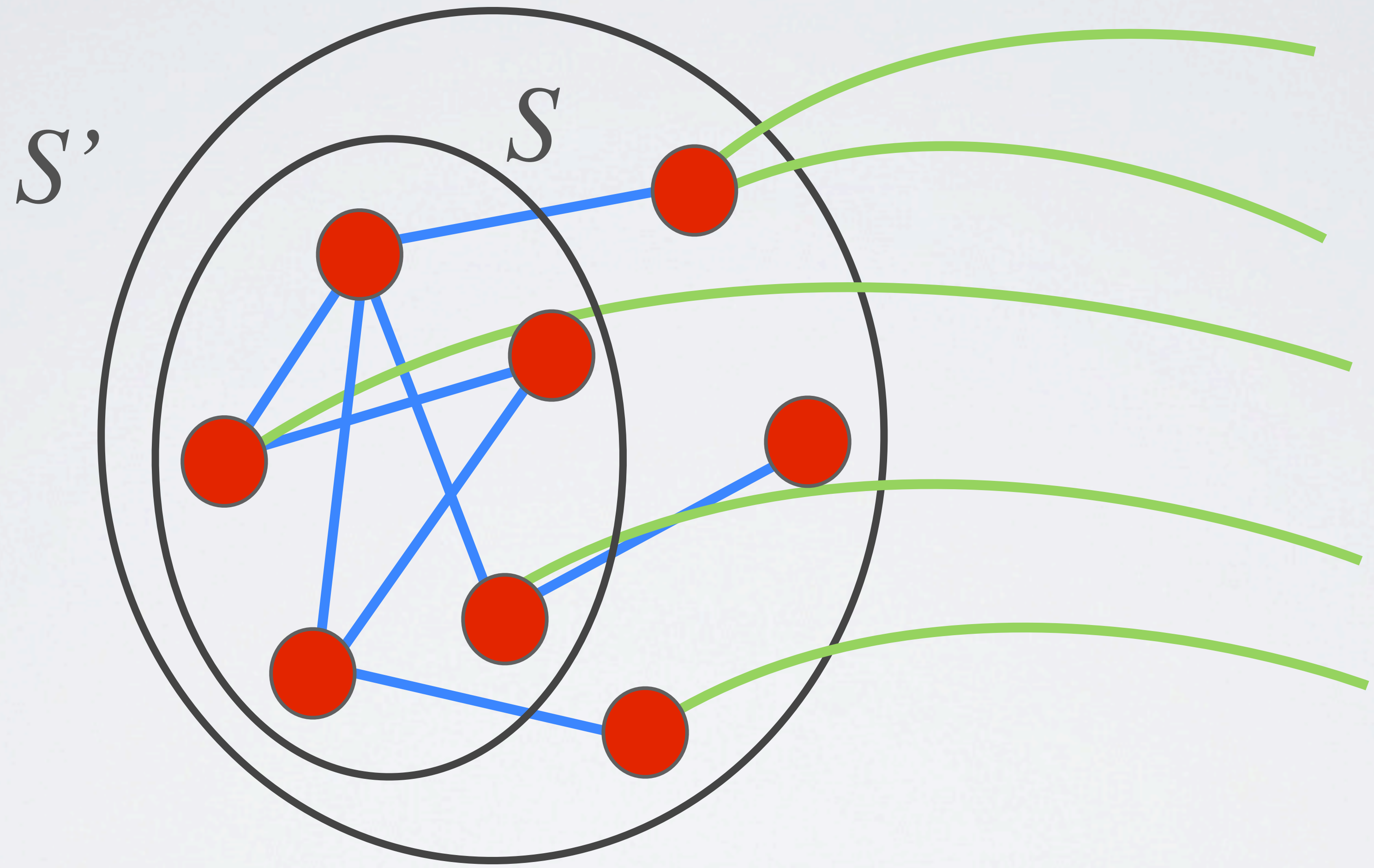
Key lemma



We consider the process for $O(\Phi^{-1})$ steps.

Proof strategy

Key lemma

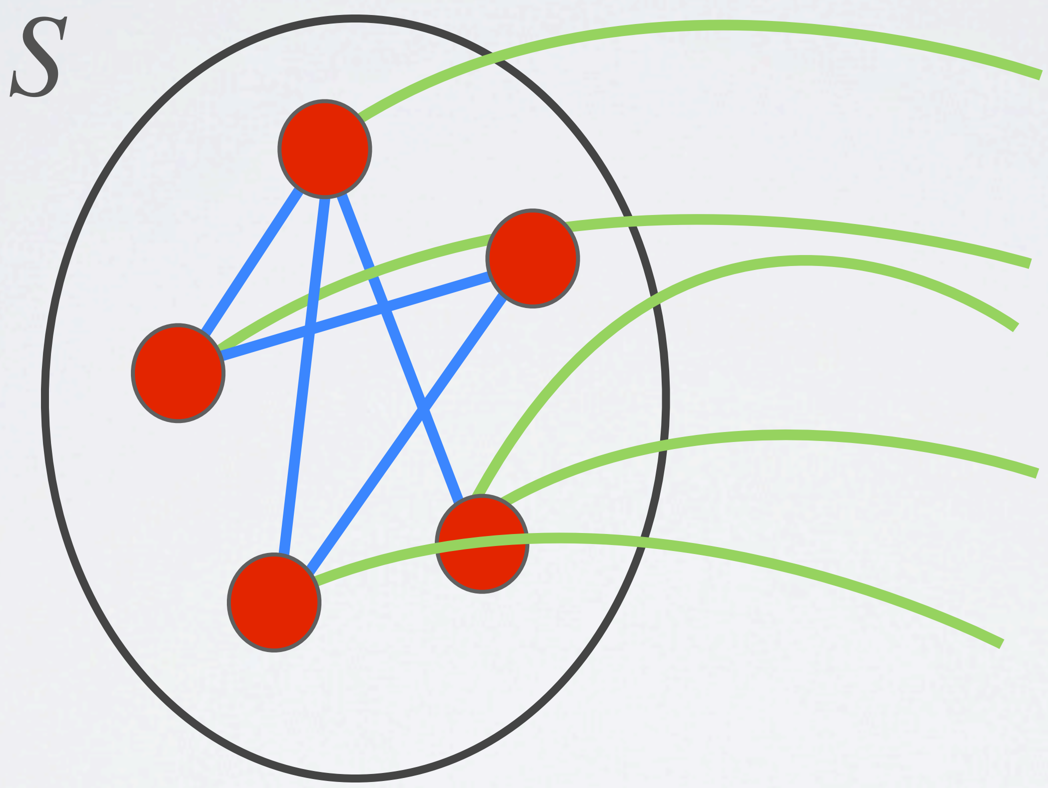


After $\Theta(\Phi^{-1})$ steps with $\Theta(1)$ probability, if S' is the new set of informed nodes, then

$$\mathrm{Vol}(S') \geq (1 + \Omega(\Phi))\,\mathrm{Vol}(S)$$

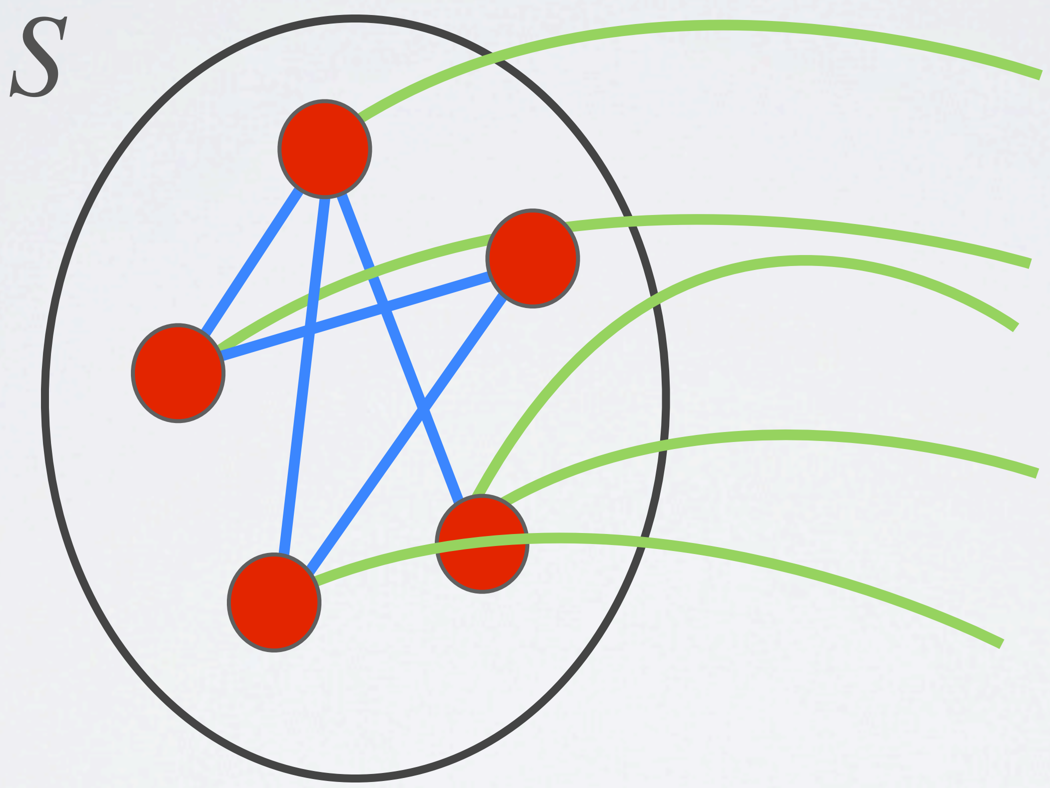
Proof strategy

We consider macro-phases composed by $O(\Phi^{-1})$ steps



Proof strategy

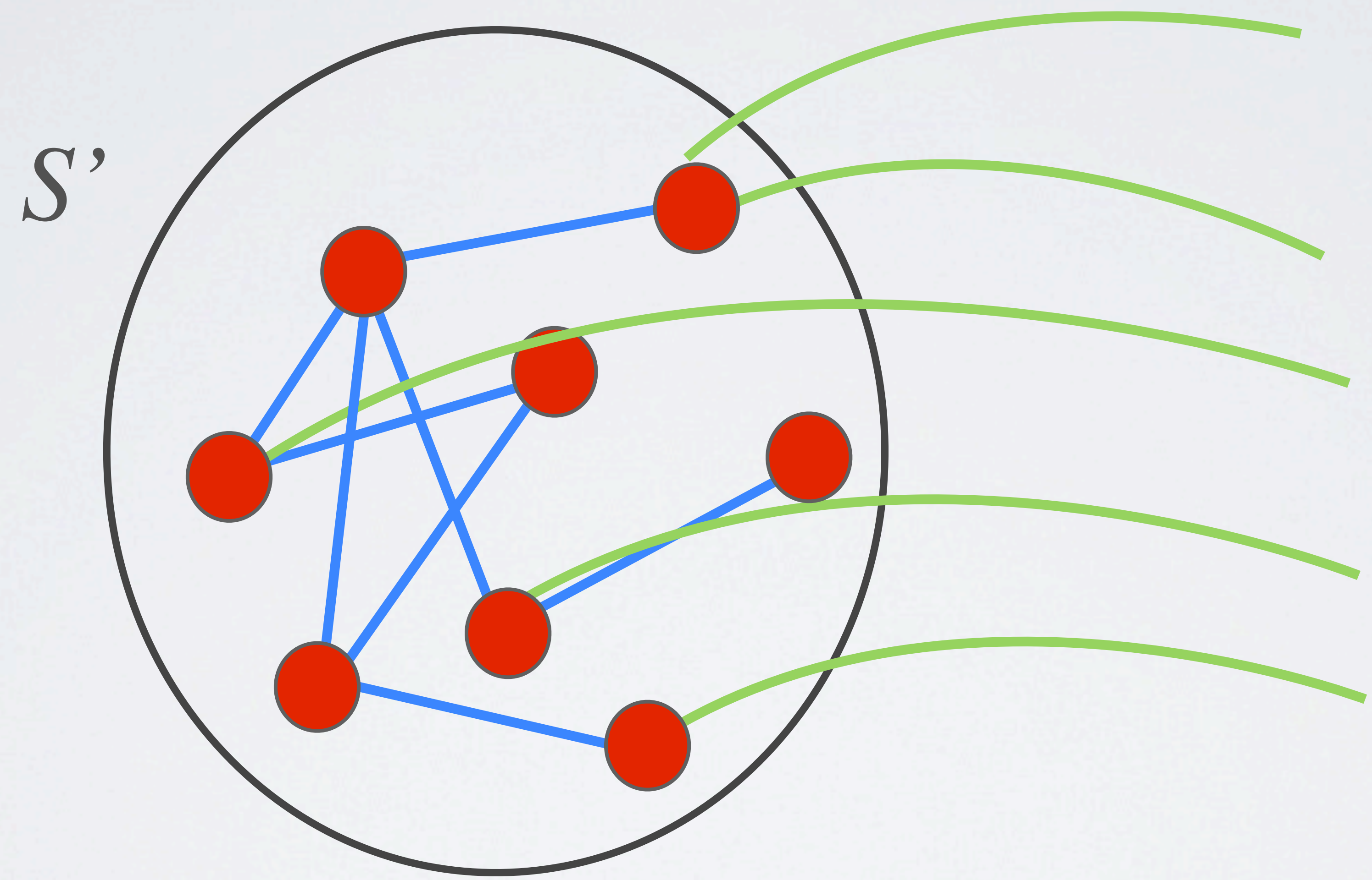
We consider macro-phases composed by $O(\Phi^{-1})$ steps



We say that a macro-phase is successful if the volume increases by a factor of $(1+ \Phi)$.

Proof strategy

We consider macro-phases composed by $O(\Phi^{-1})$ steps

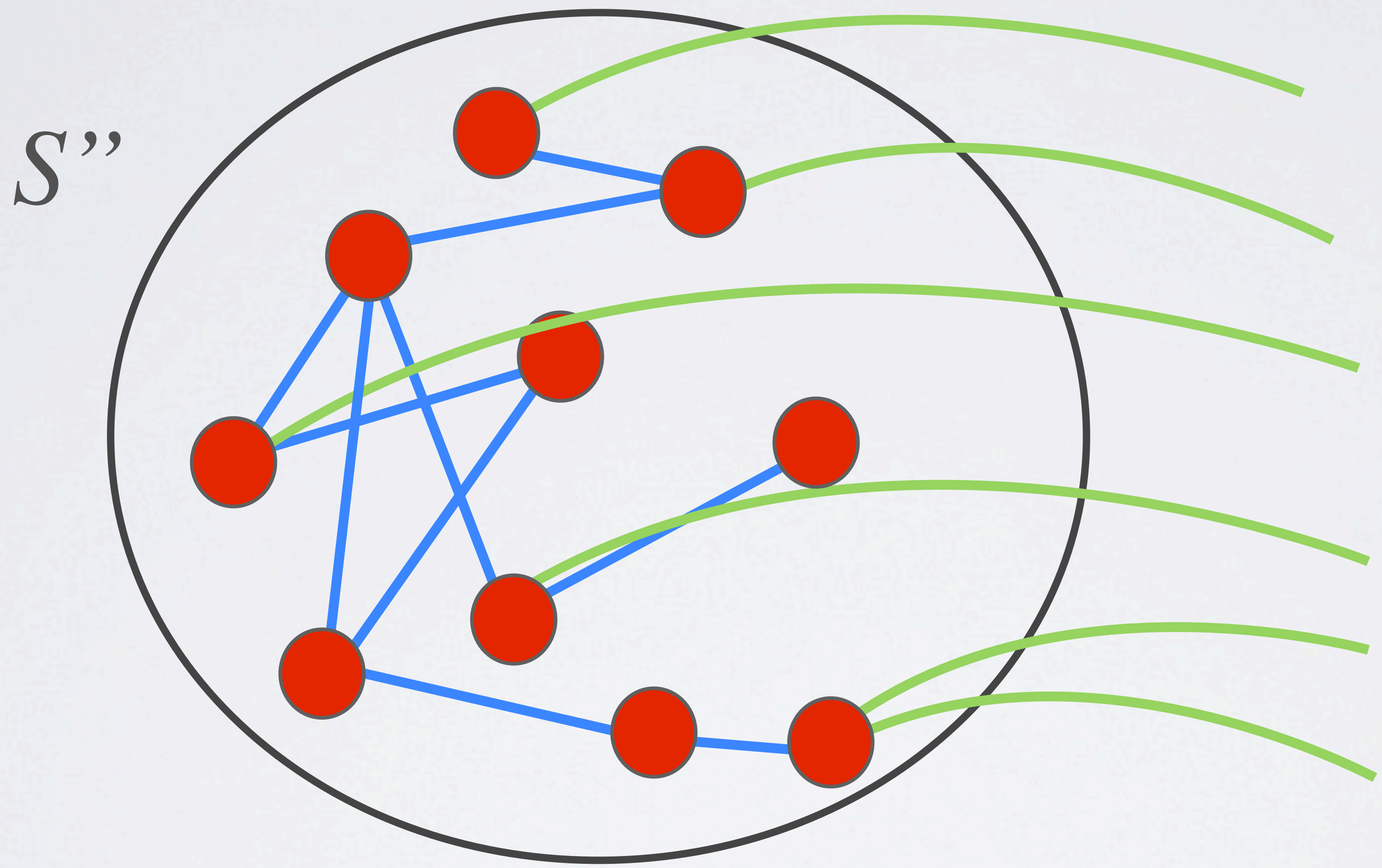


After 1 successful macro-phase, we have:

$$\mathrm{Vol}(S') \geq (1 + \Omega(\Phi))\ \mathrm{Vol}(S)$$

Proof strategy

We consider macro-phases composed by $O(\Phi^{-1})$ steps

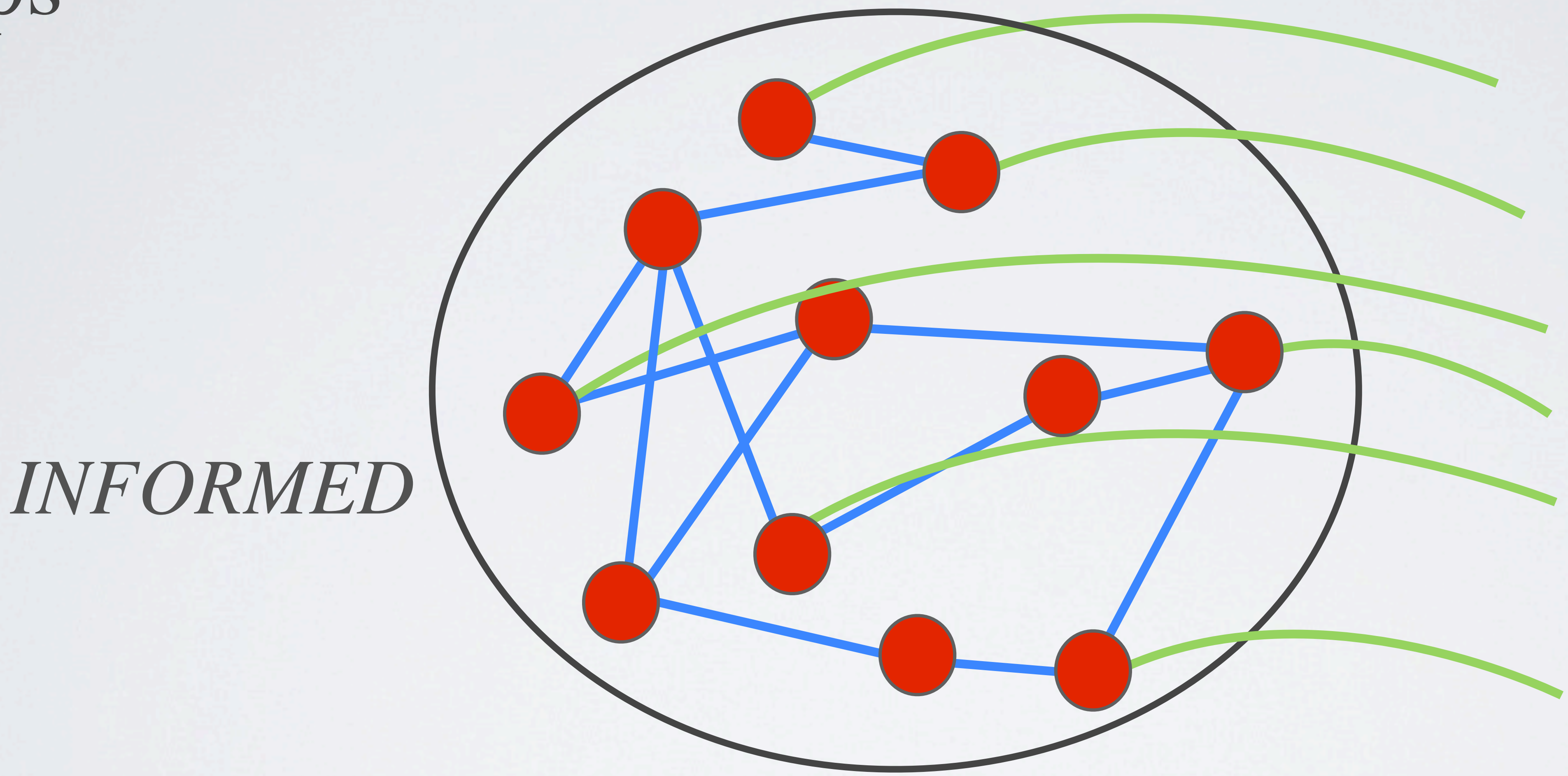


After 2 successful macro-phases, we have:

$$\mathrm{Vol}(S'') \geq (1+\Omega(\Phi))^2\ \mathrm{Vol}(S)$$

Proof strategy

We consider macro-phases composed by $O(\Phi^{-1})$ steps



After $\Theta(\Phi^{-1} \log n)$ successful macro-phases, we have:

$$\mathrm{Vol}(\mathit{INFORMED}) > \mathrm{Vol}(G)/2$$

Proof strategy

- A macro-phase is successful with constant probability.

Proof strategy

- A macro-phase is successful with constant probability.

- After $O(\Phi^{-1} \log n)$ successful macro-phases, we have $\mathrm{Vol}(INFORMED) \geq \mathrm{Vol}(G)/2$

Proof strategy

- A macro-phase is successful with constant probability.
- After $O(\Phi^{-1} \log n)$ successful macro-phases, we have Vol$(INFORMED) \geq$ Vol$(G)/2$
- Using the Chernoff bound after $O(\Phi^{-1} \log n)$ macro-phases, we have $O(\Phi^{-1} \log n)$ successful macro-phases.

Proof strategy

After $O(\Phi^{-2} \log n)$ steps we have

$\mathrm{Vol}(\mathit{INFORMED}) > \mathrm{Vol}(G)/2$ w.h.p.

Proof strategy

After $O(\Phi^{-2} \log n)$ steps we have

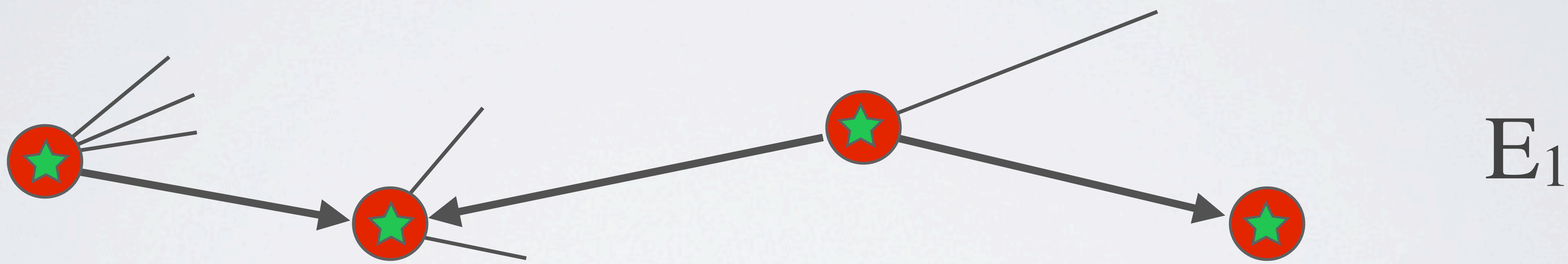
$\text{Vol}(\mathit{INFORMED}) > \text{Vol}(G)/2$ w.h.p.

E_1

Proof strategy

After $O(\Phi^{-2} \log n)$ steps we have

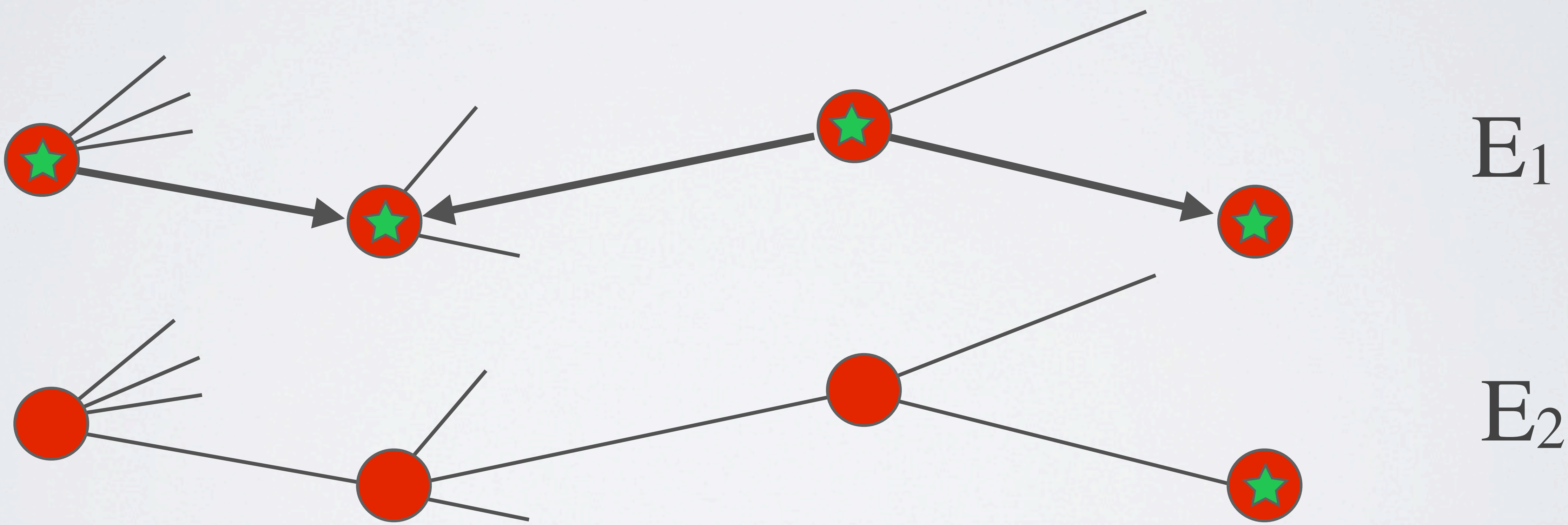
$$\mathrm{Vol}(\mathit{INFORMED}) > \mathrm{Vol}(G)/2 \text{ w.h.p.}$$



Proof strategy

After $O(\Phi^{-2} \log n)$ steps we have

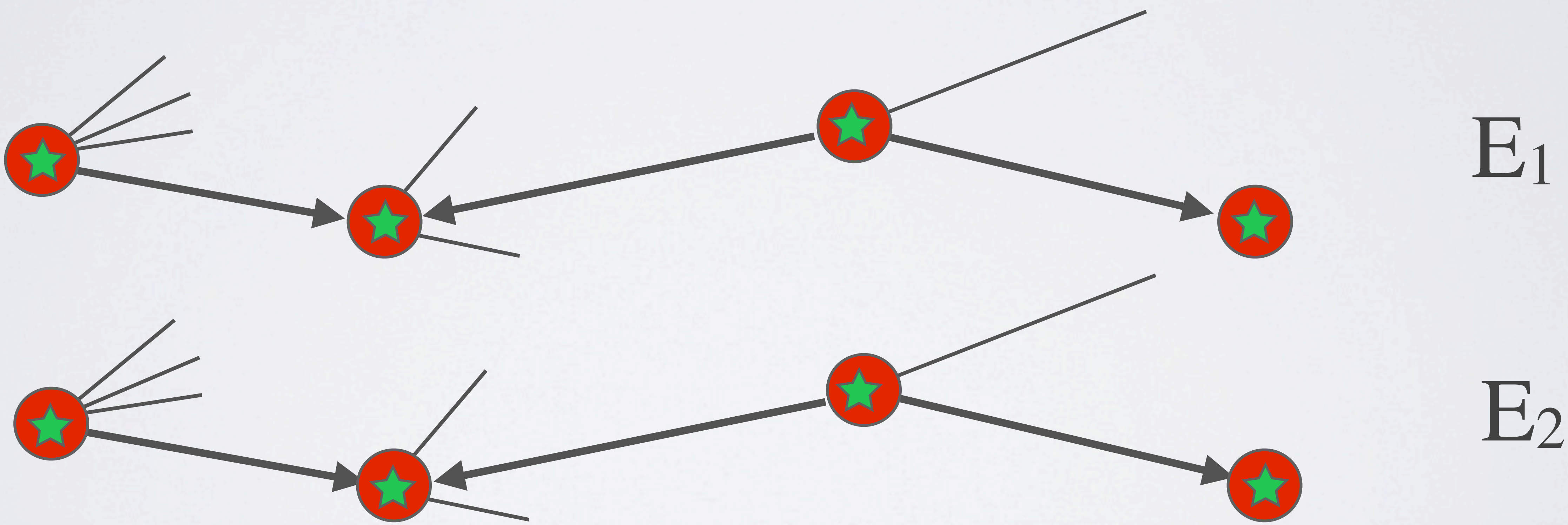
$$\mathrm{Vol}(INFORMED) > \mathrm{Vol}(G)/2 \text{ w.h.p.}$$



Proof strategy

After $O(\Phi^{-2} \log n)$ steps we have

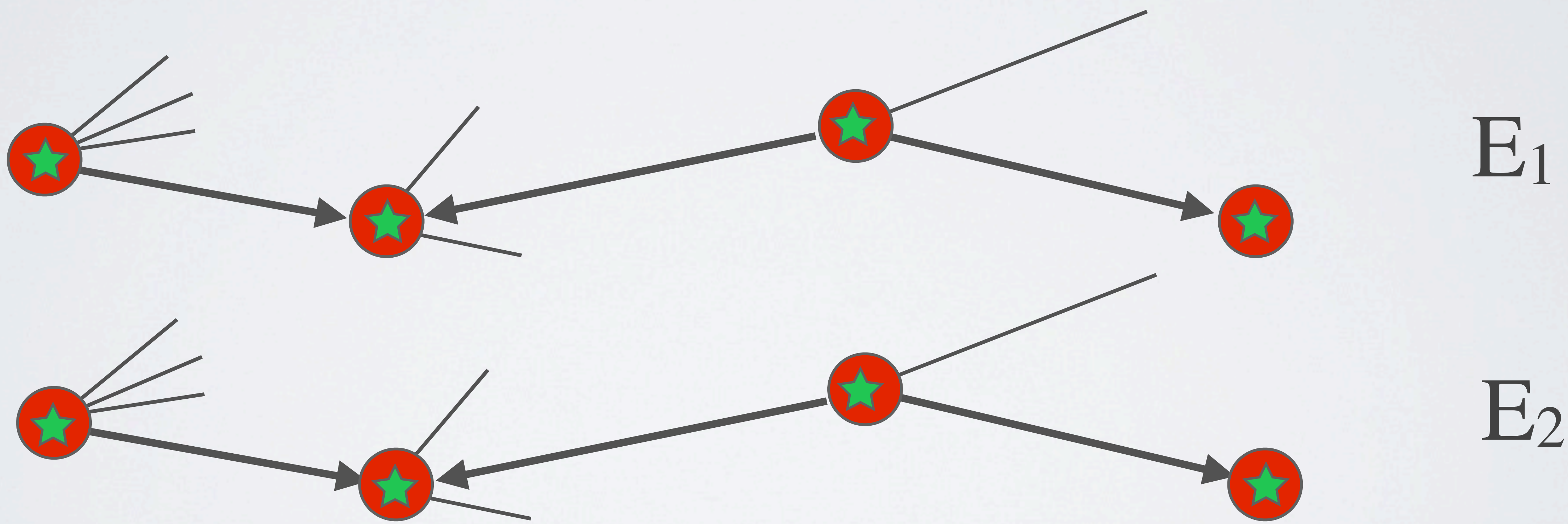
$$\mathrm{Vol}(INFORMED) > \mathrm{Vol}(G)/2 \text{ w.h.p.}$$



Proof strategy

After $O(\Phi^{-2} \log n)$ steps we have

$\mathrm{Vol}(INFORMED) > \mathrm{Vol}(G)/2$ w.h.p.



$P(E_1) = P(E_2)$

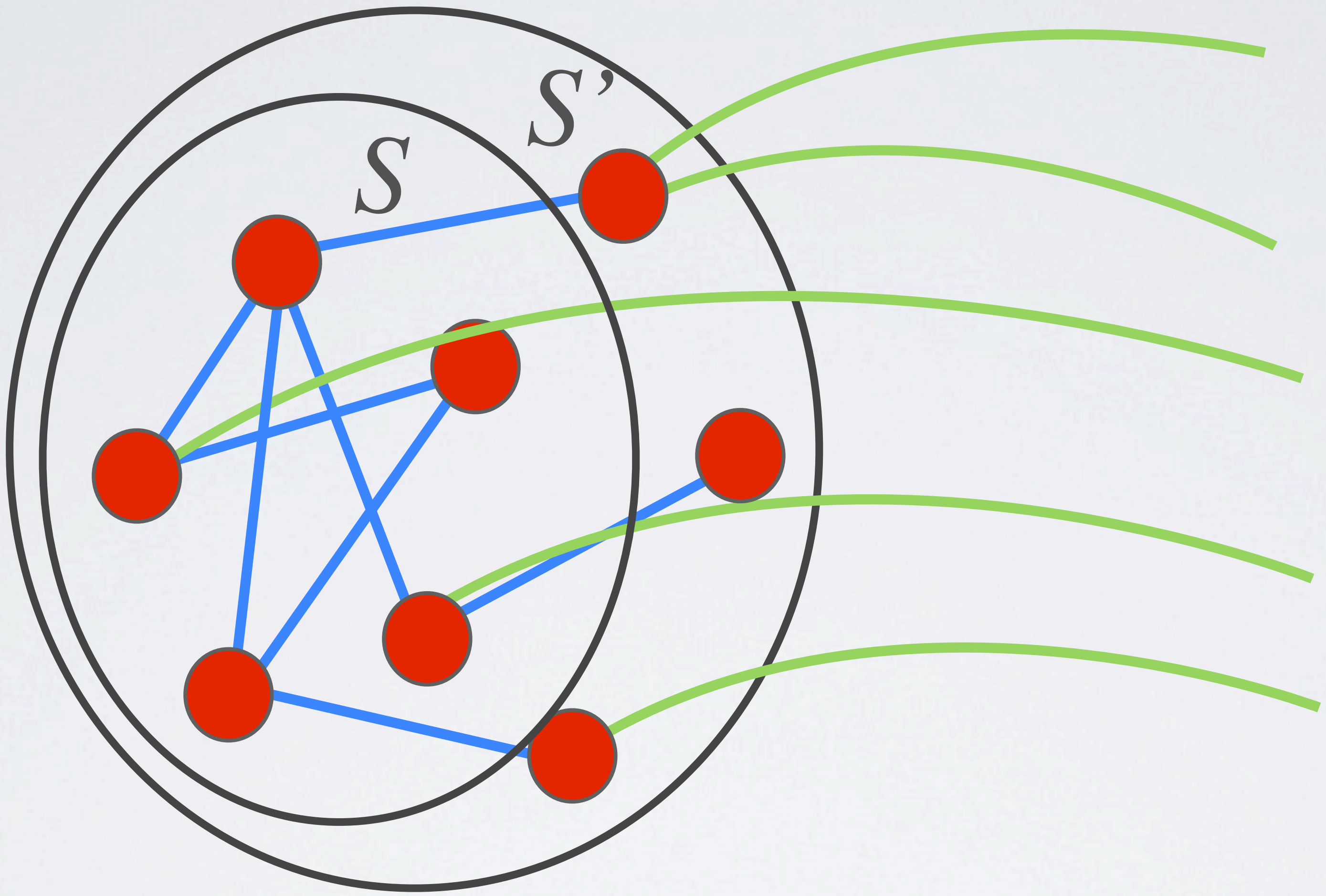
Proof strategy

- After $O(\Phi^{-2} \log n)$ steps we have
  $\mathrm{Vol}(INFORMED) > \mathrm{Vol}(G)/2$ w.h.p.

- After $O(\Phi^{-2} \log n)$ steps each node pulls the information from a set of nodes of $\mathrm{Vol}(G)/2$ w.h.p.

- After $O(\Phi^{-2} \log n)$ steps all the nodes have the info w.h.p.

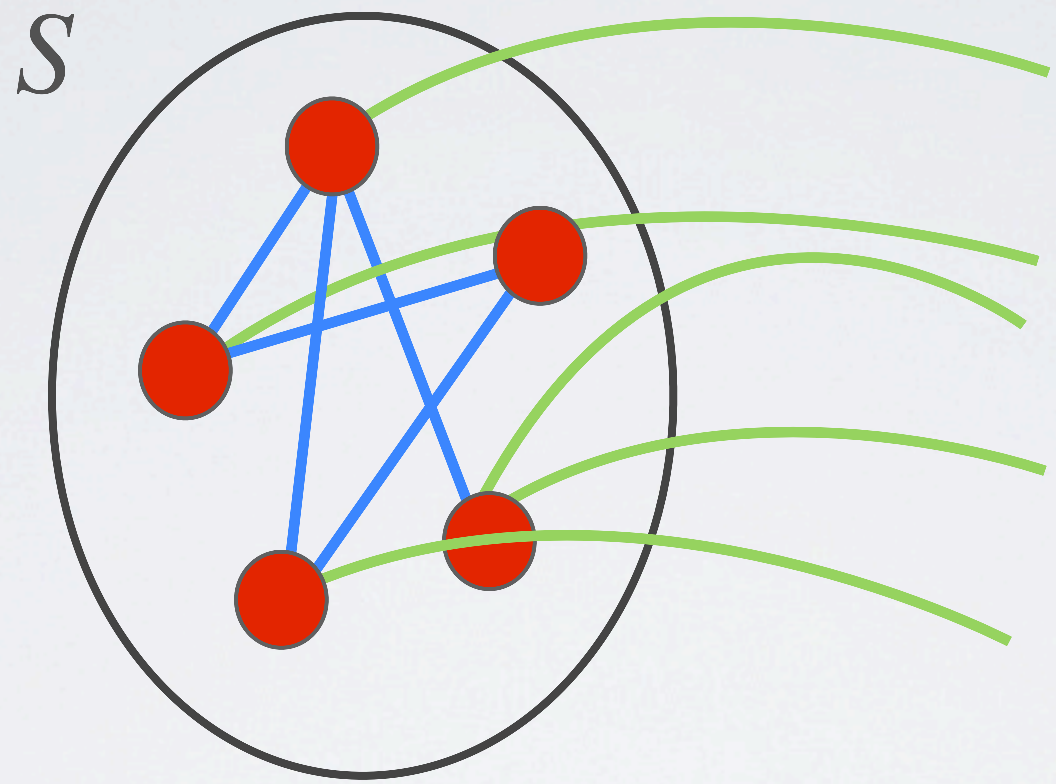
Key lemma



After $O(\Phi^{-1})$ steps with constant probability, we have that for the new set of informed nodes S'

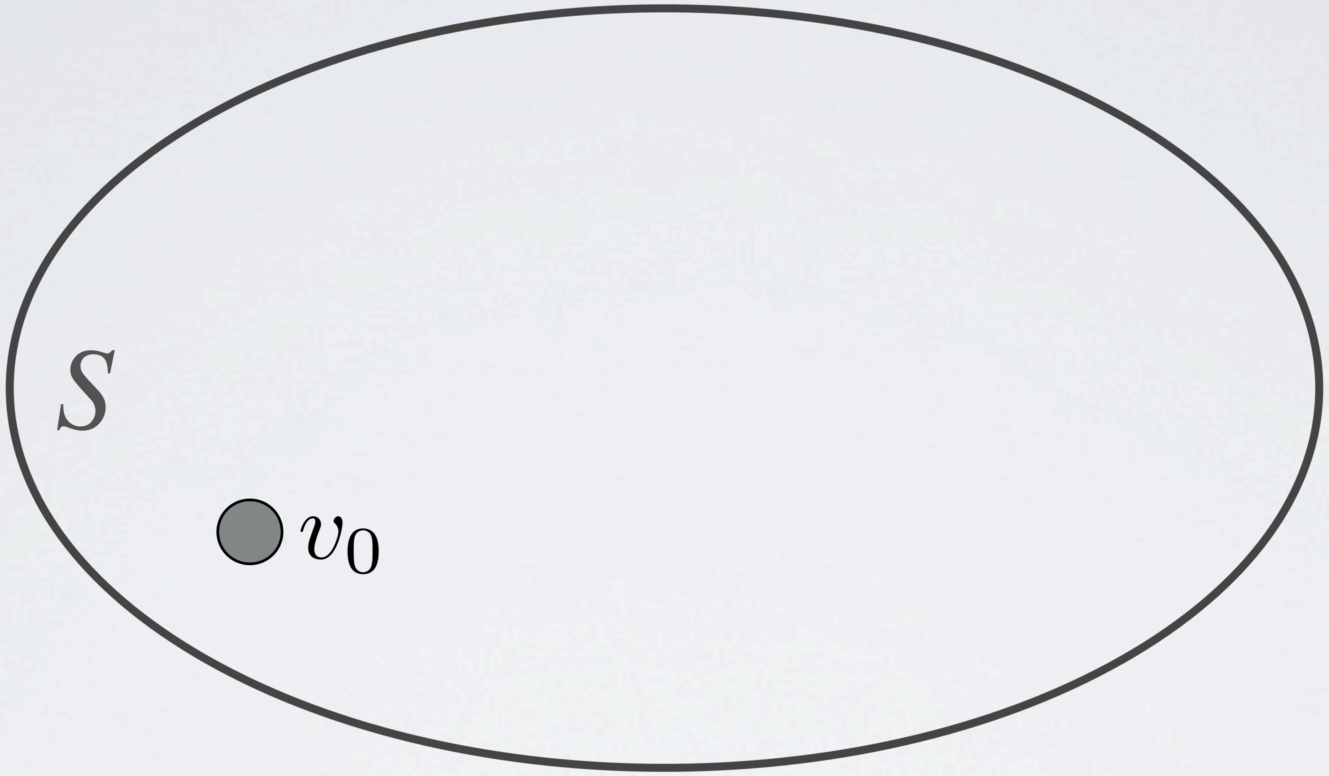
$$\mathrm{Vol}(S') \geq (1+\Omega(\Phi))\,\mathrm{Vol}(S)$$

Sketch of the proof

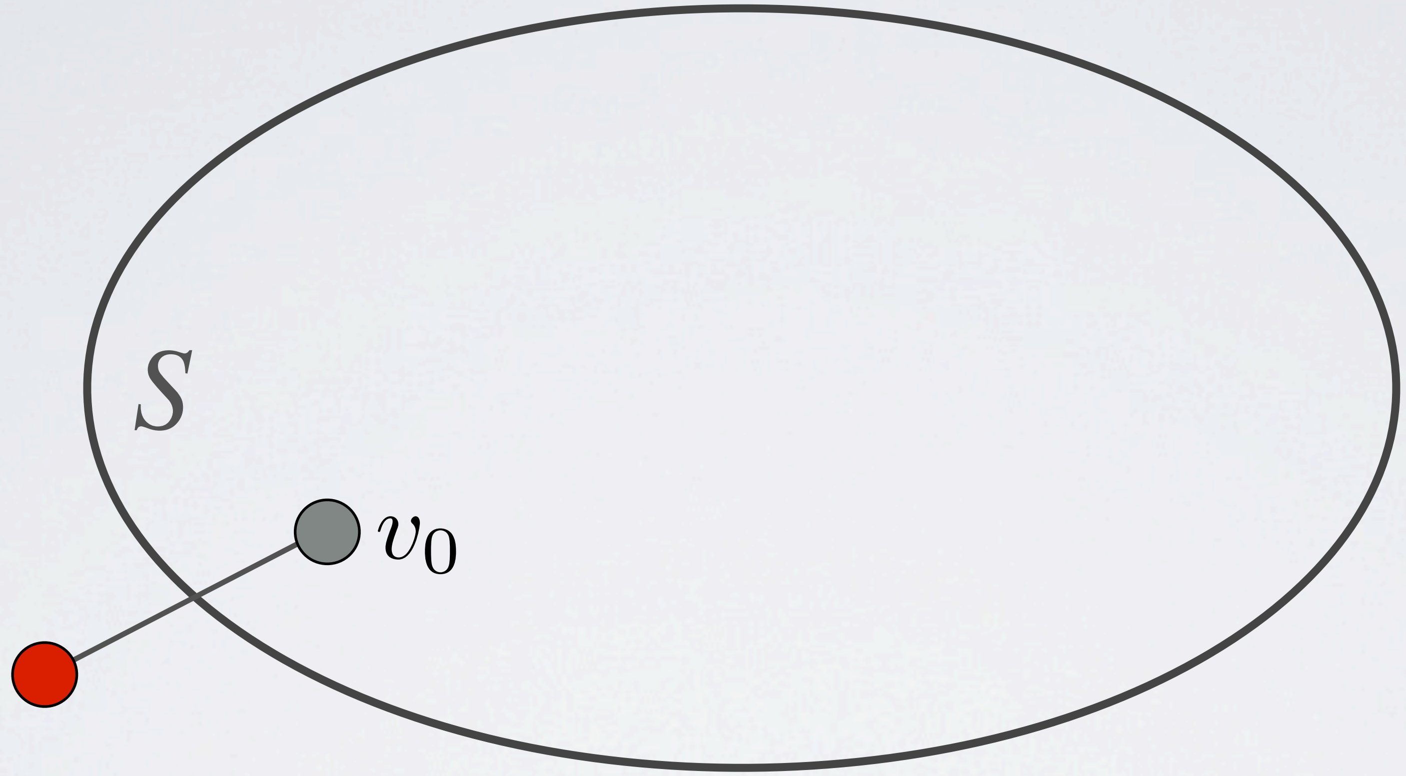


Idea: analyze what happens to each node in S in a macro-phase

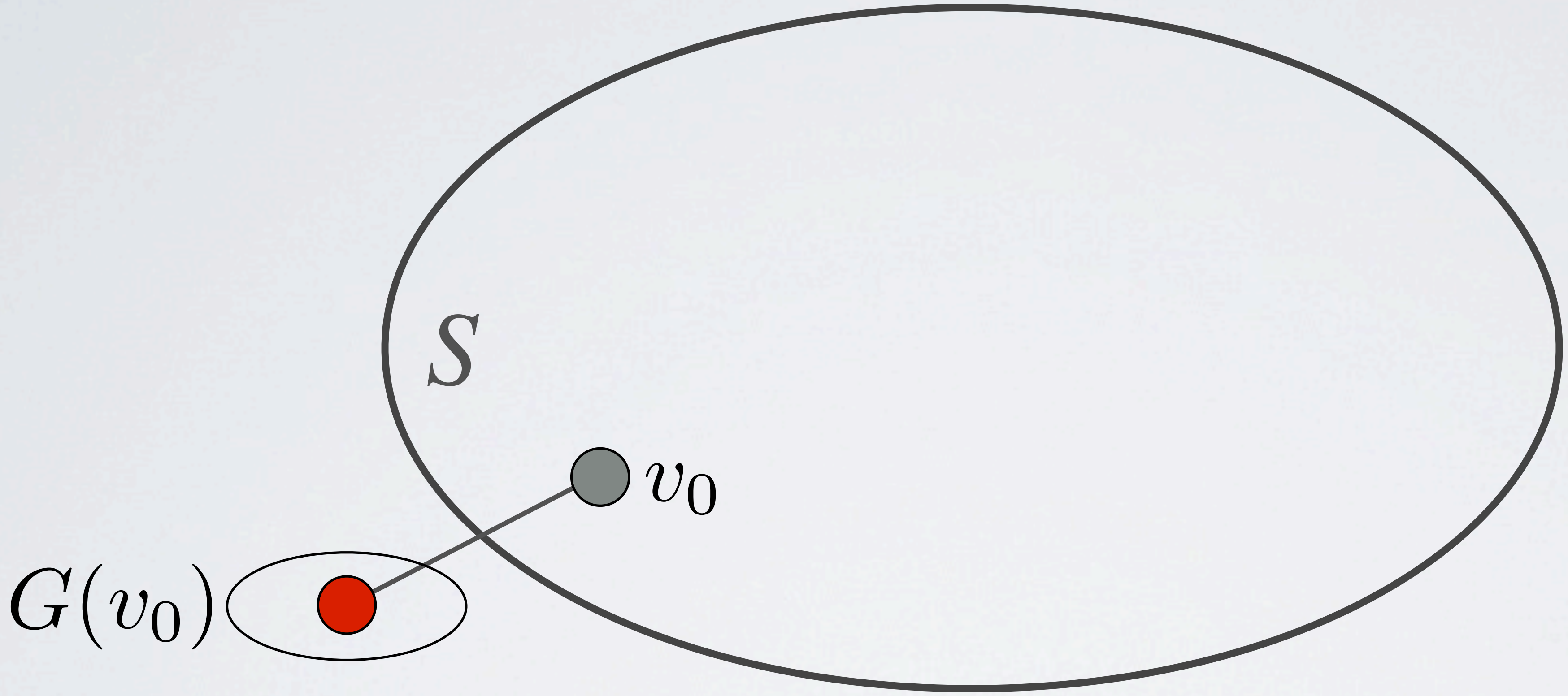
Sketch of the proof



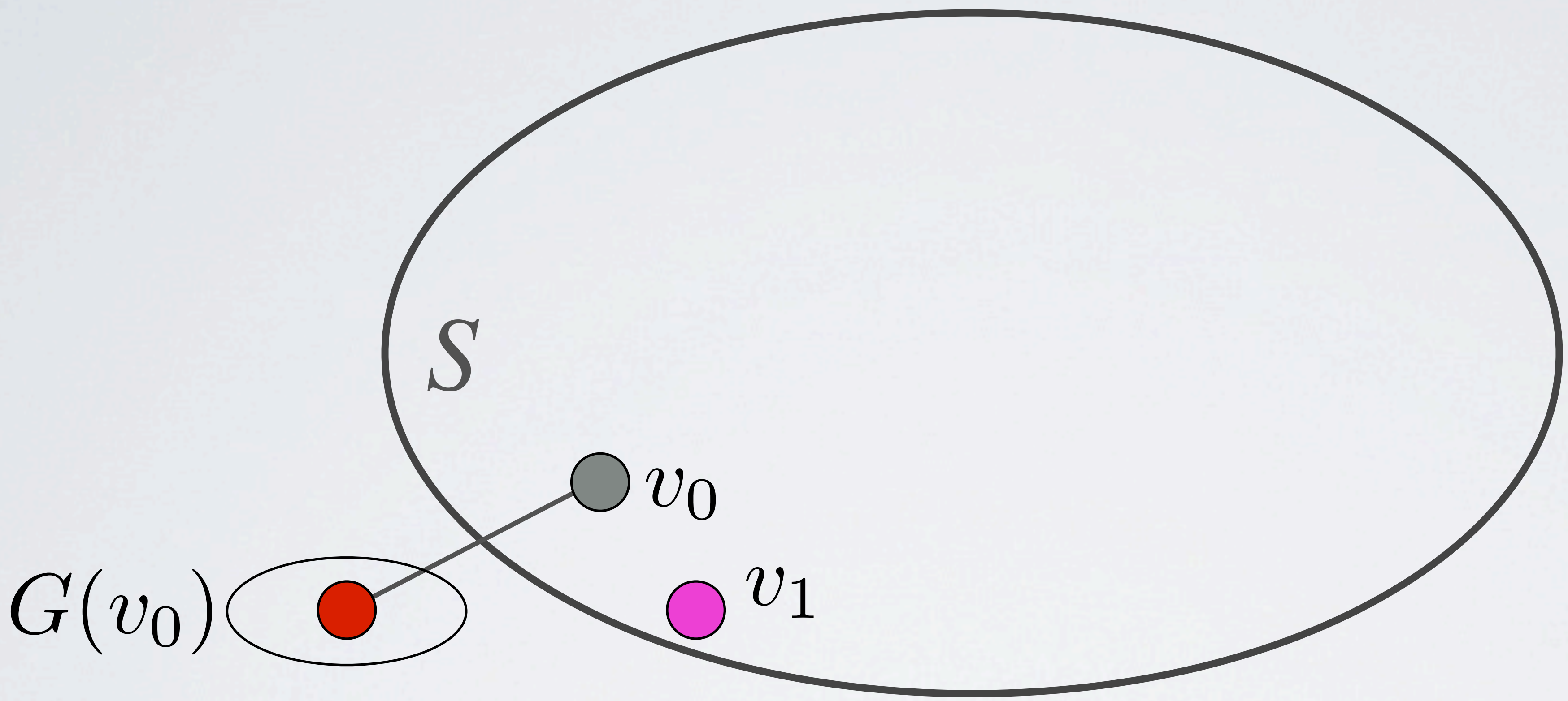
Sketch of the proof



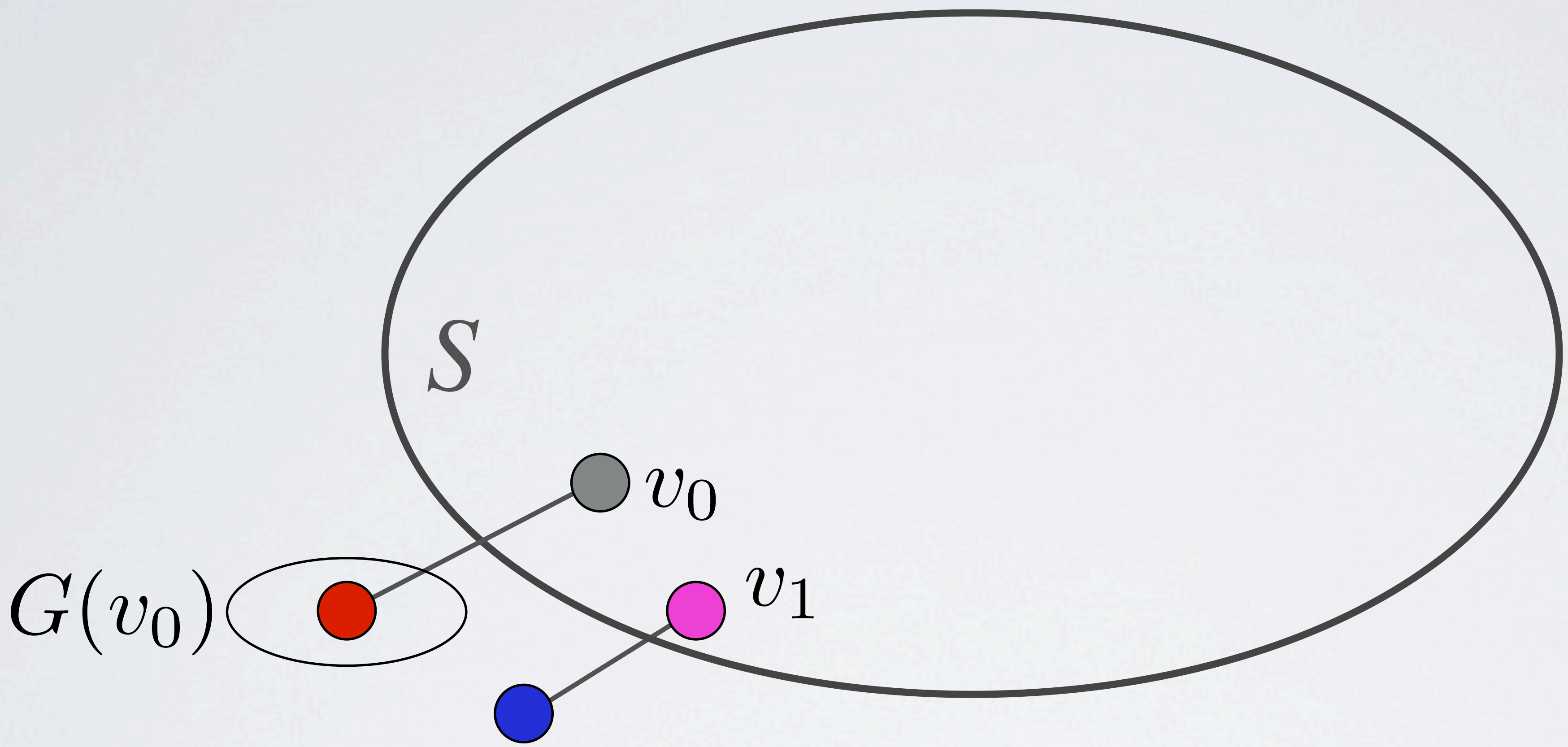
Sketch of the proof



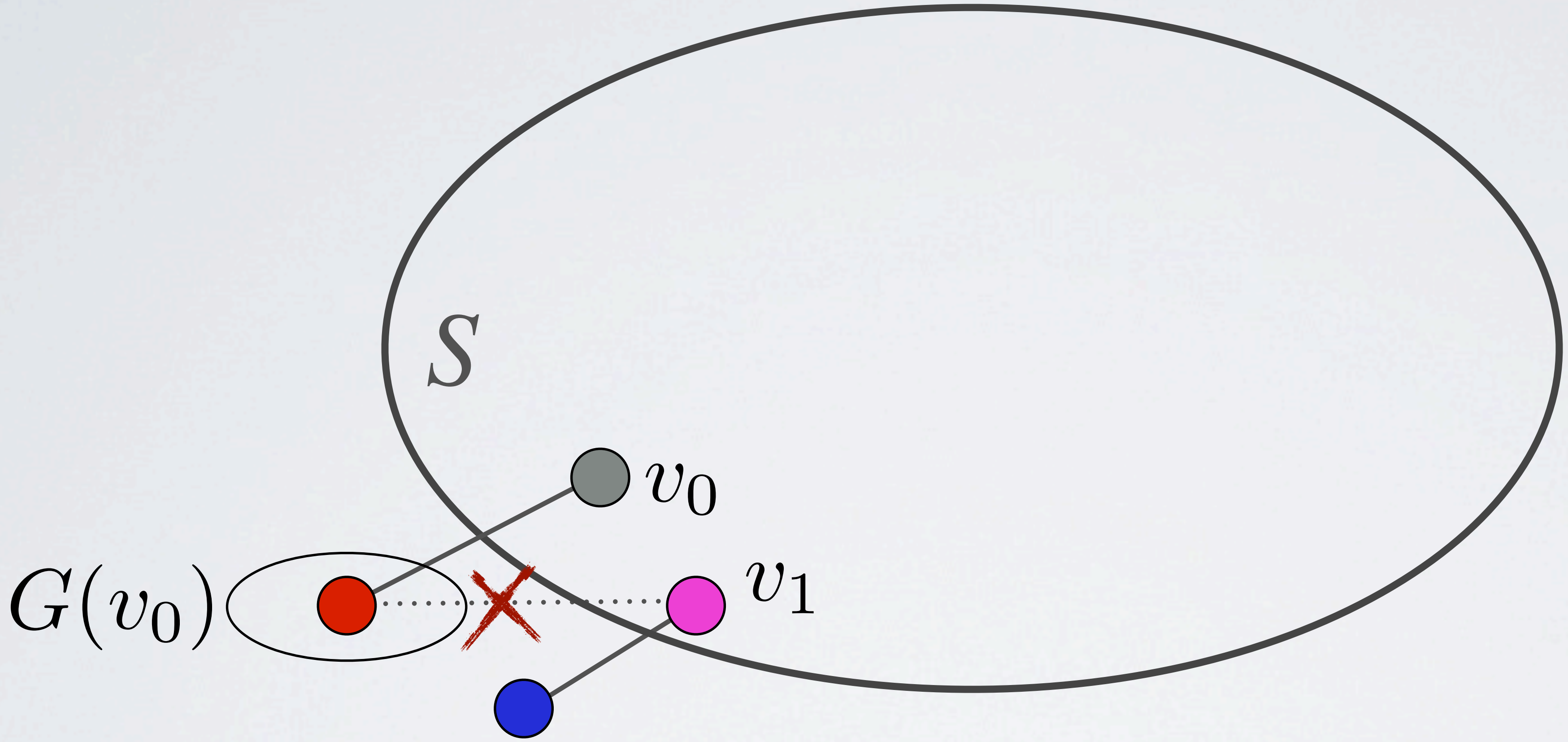
Sketch of the proof



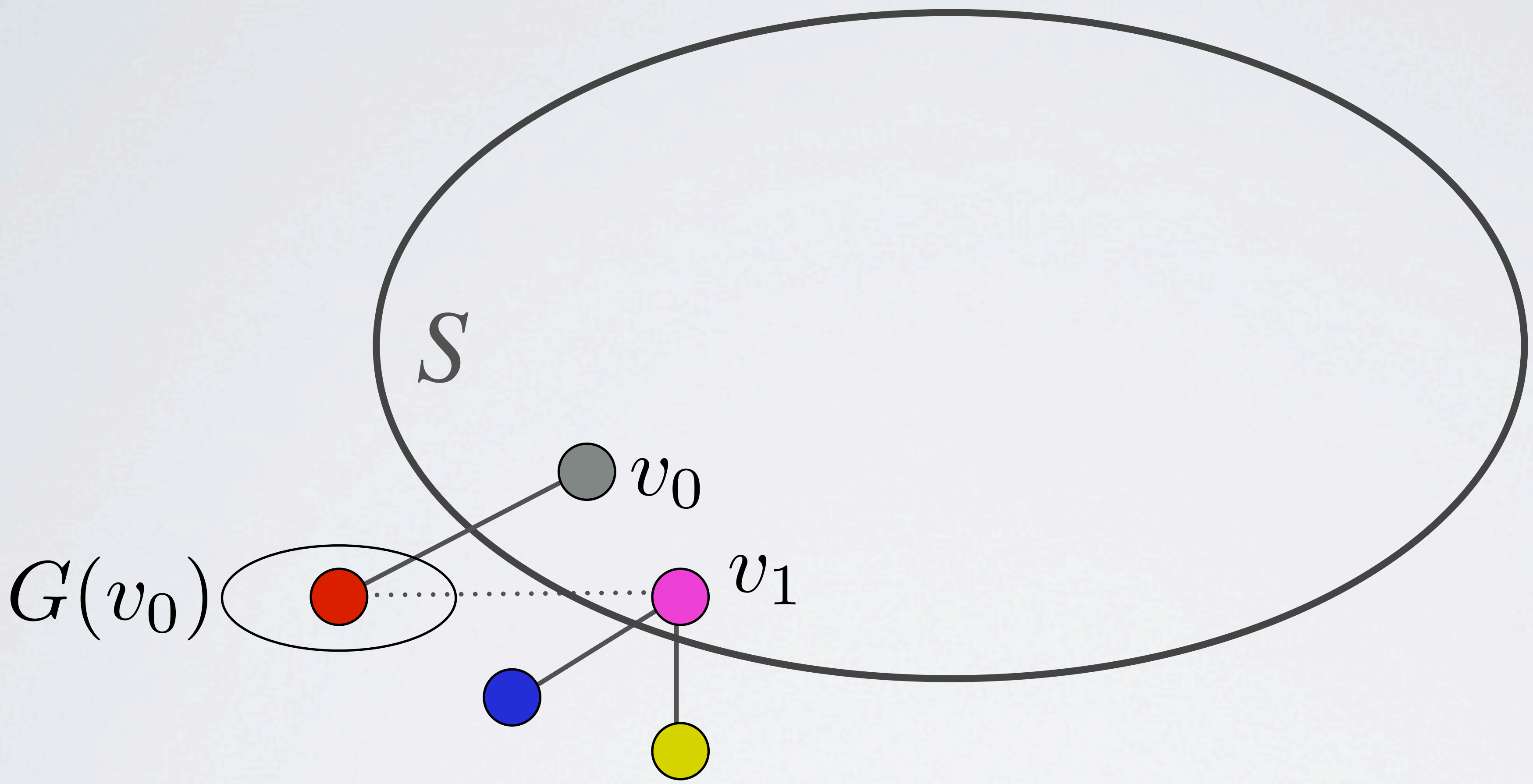
Sketch of the proof



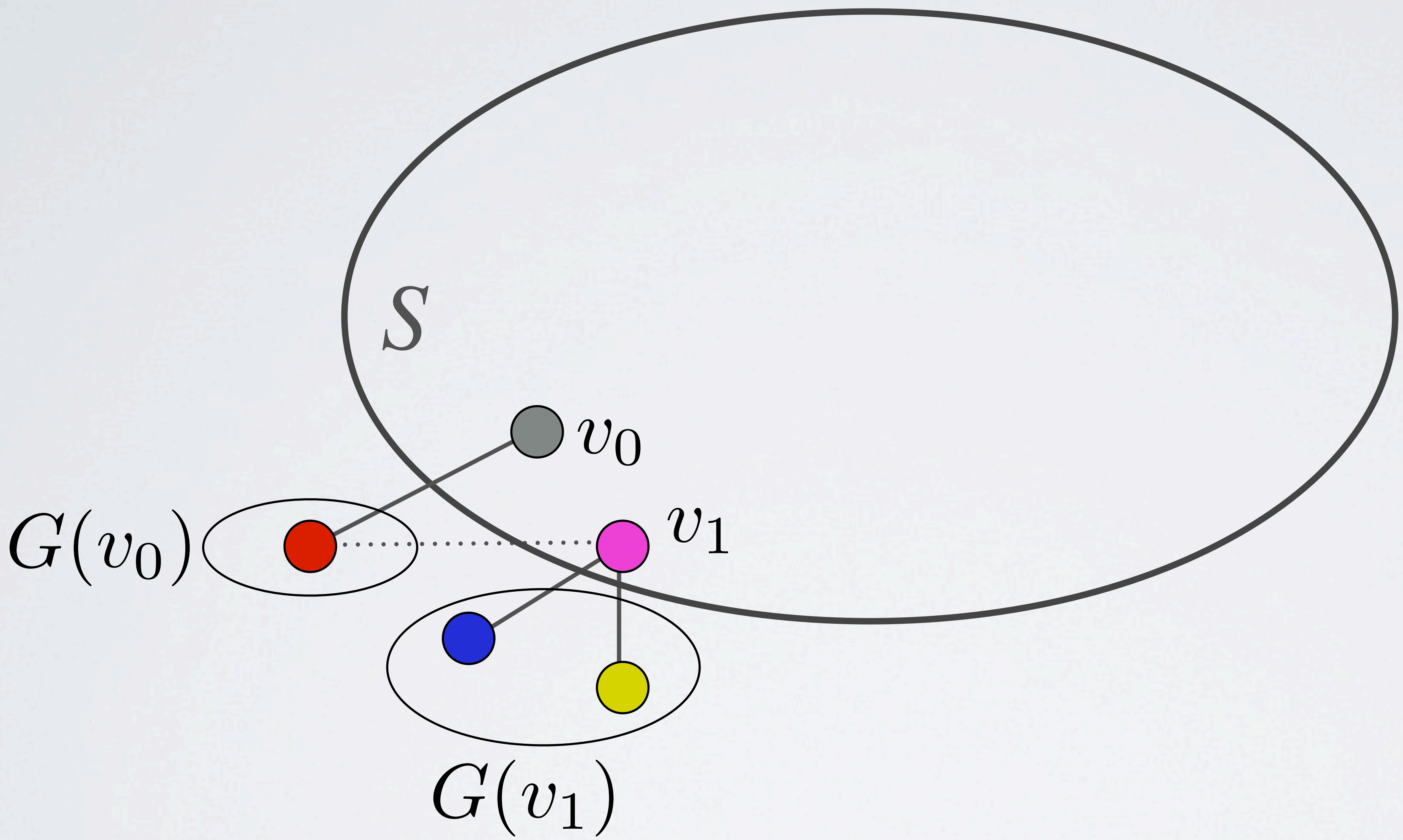
Sketch of the proof



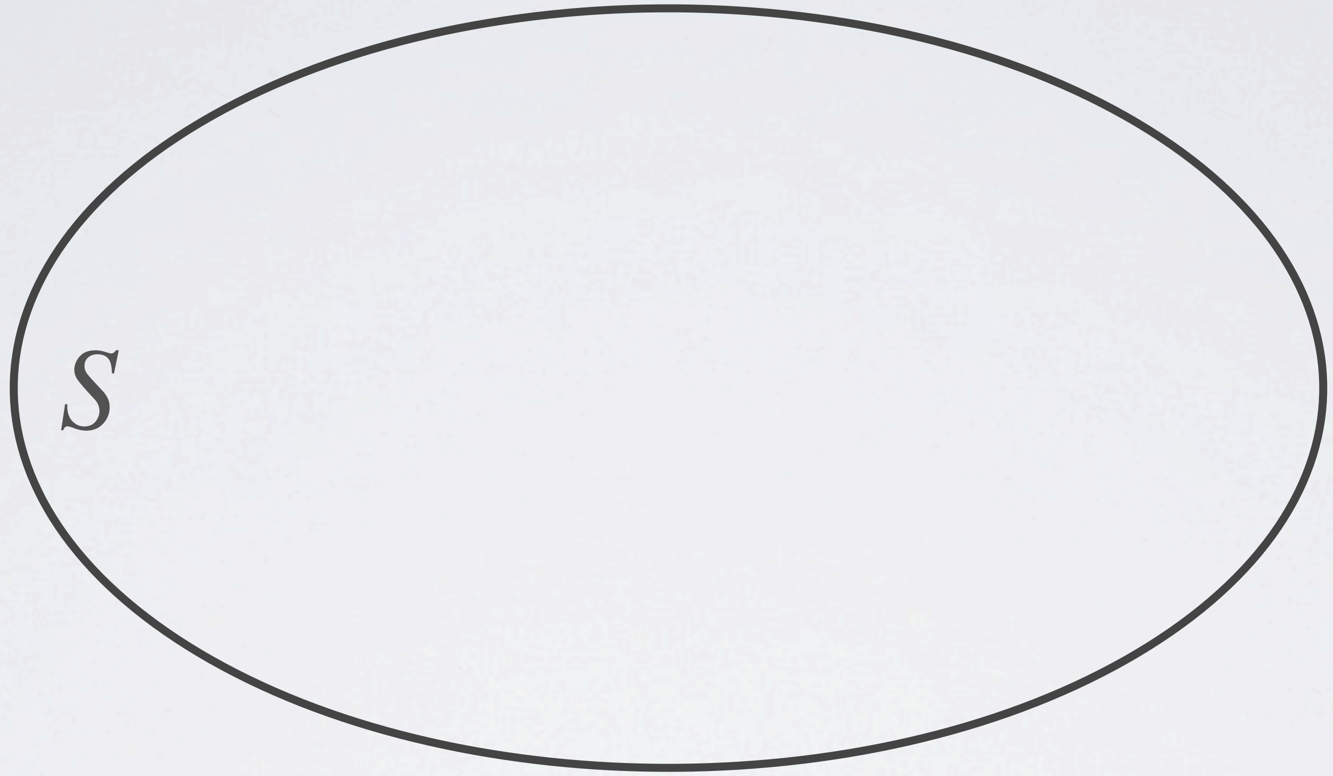
Sketch of the proof



Sketch of the proof

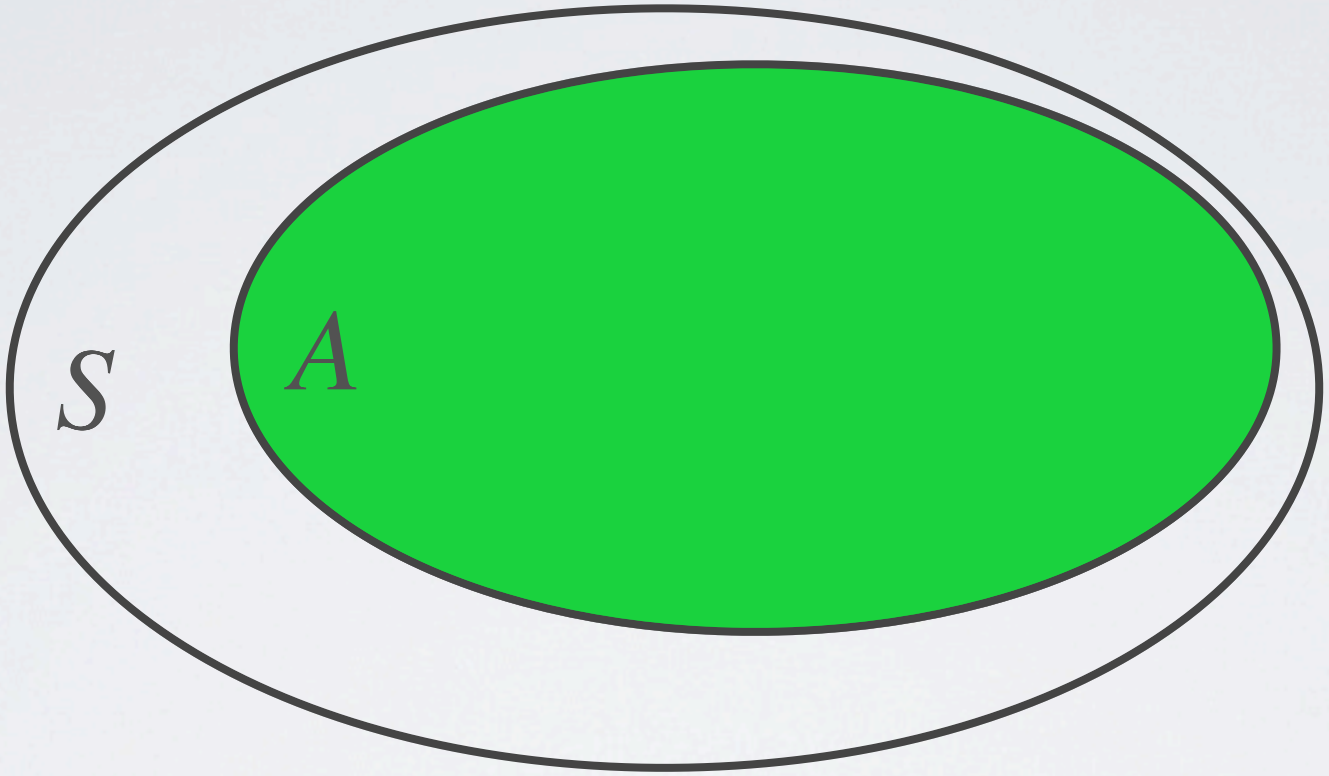


Sketch of the proof



We define the following sets:

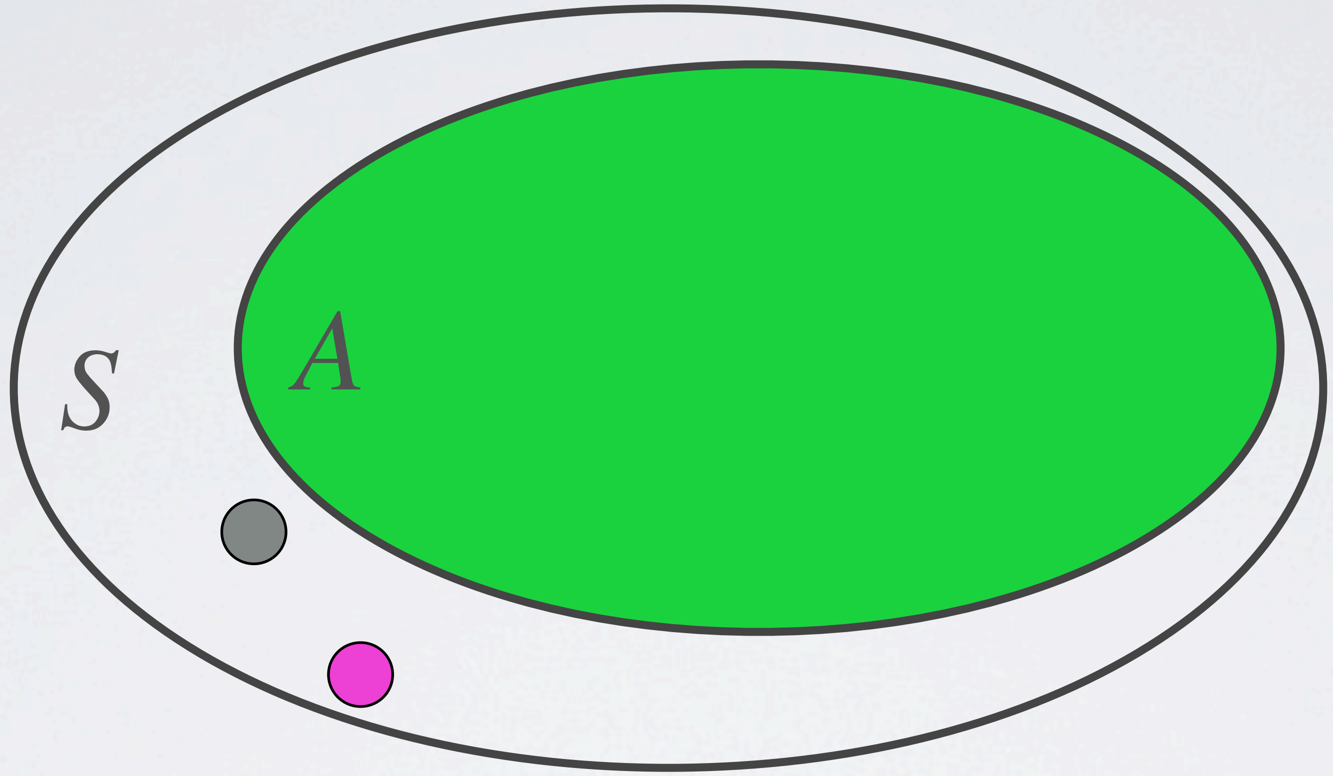
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider

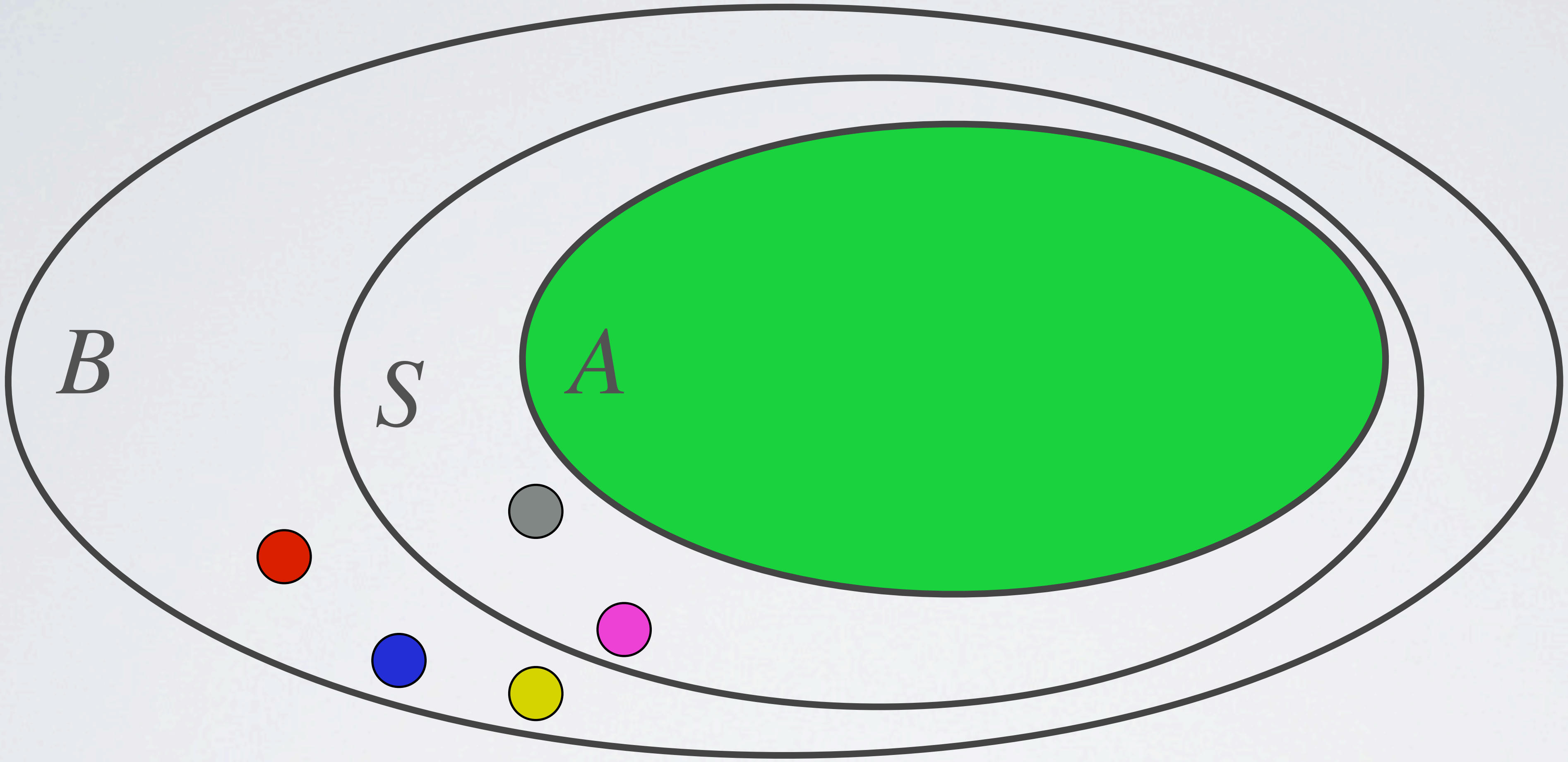
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider

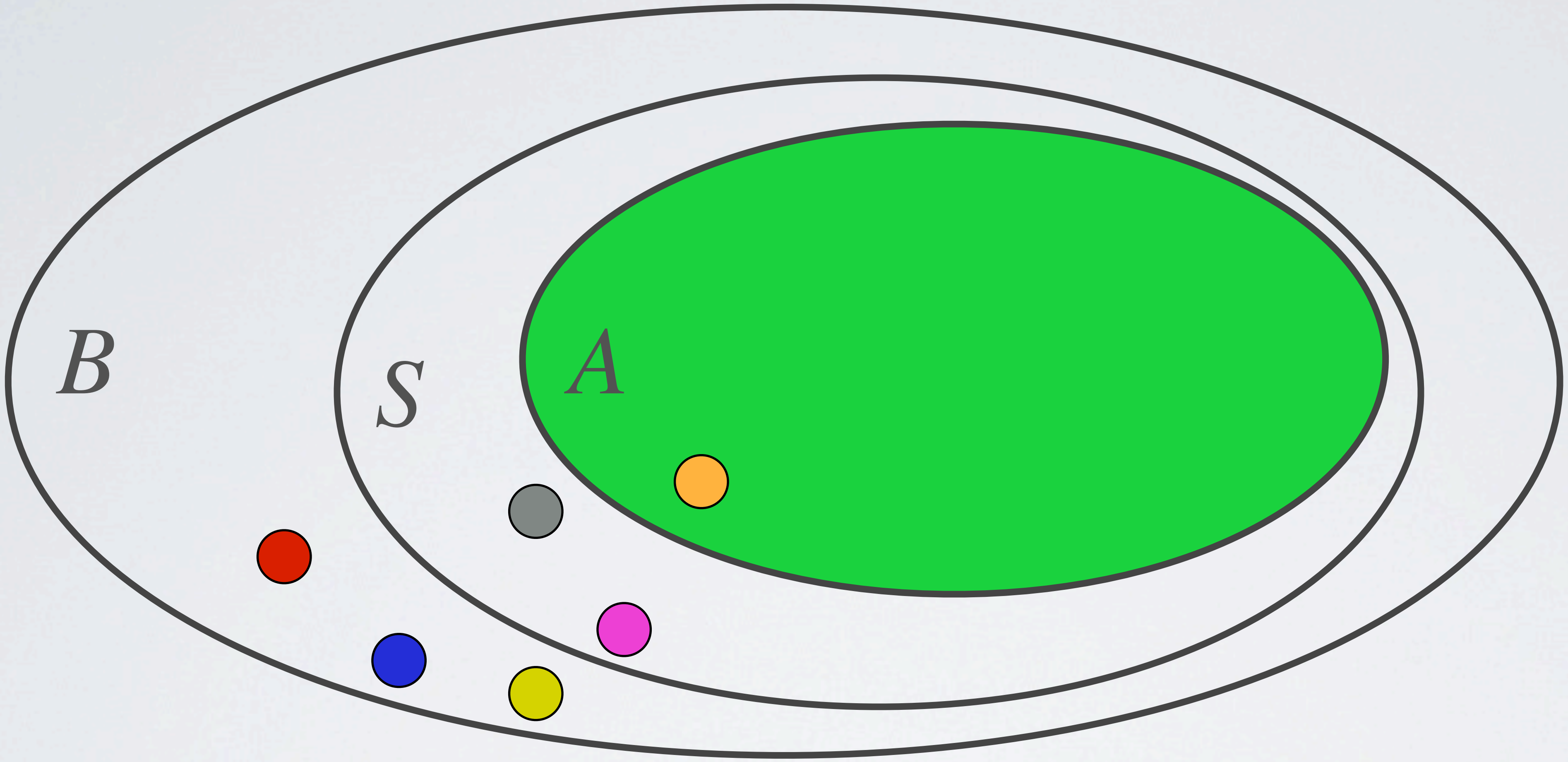
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider
- $B \supseteq S$, informed nodes at the current phase

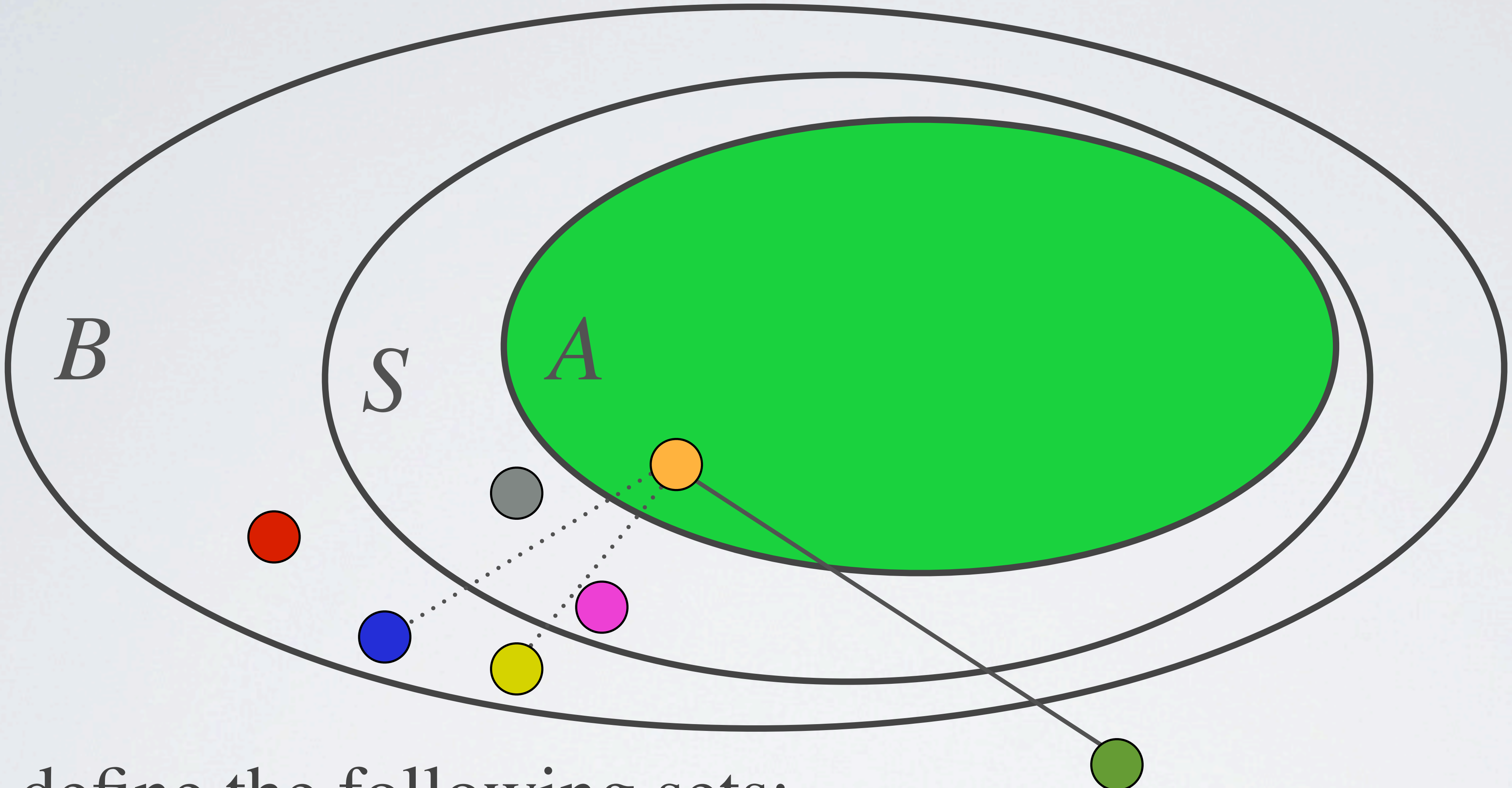
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider
- $B \supseteq S$, informed nodes at the current phase

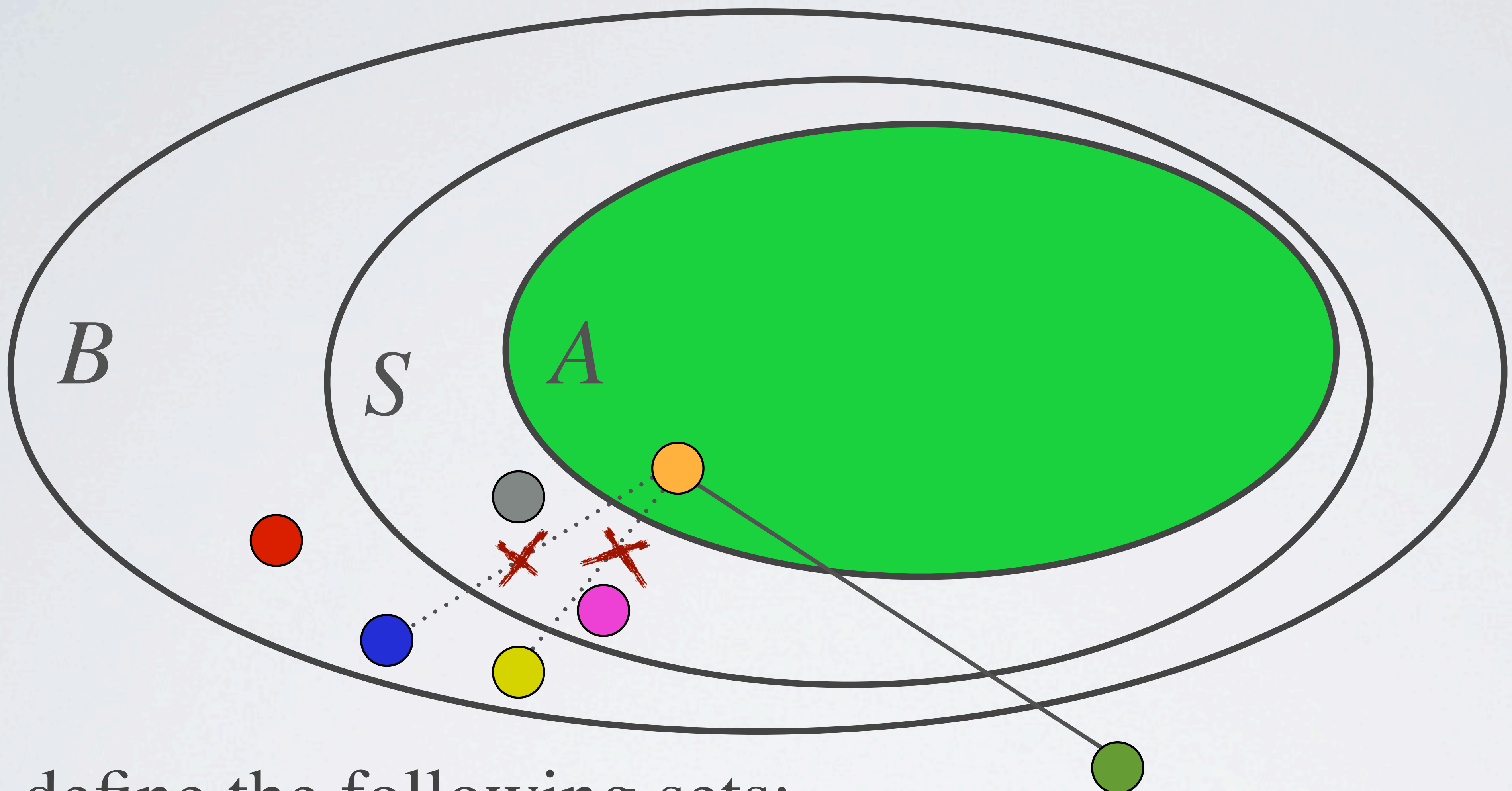
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider
- $B \supseteq S$, informed nodes at the current phase

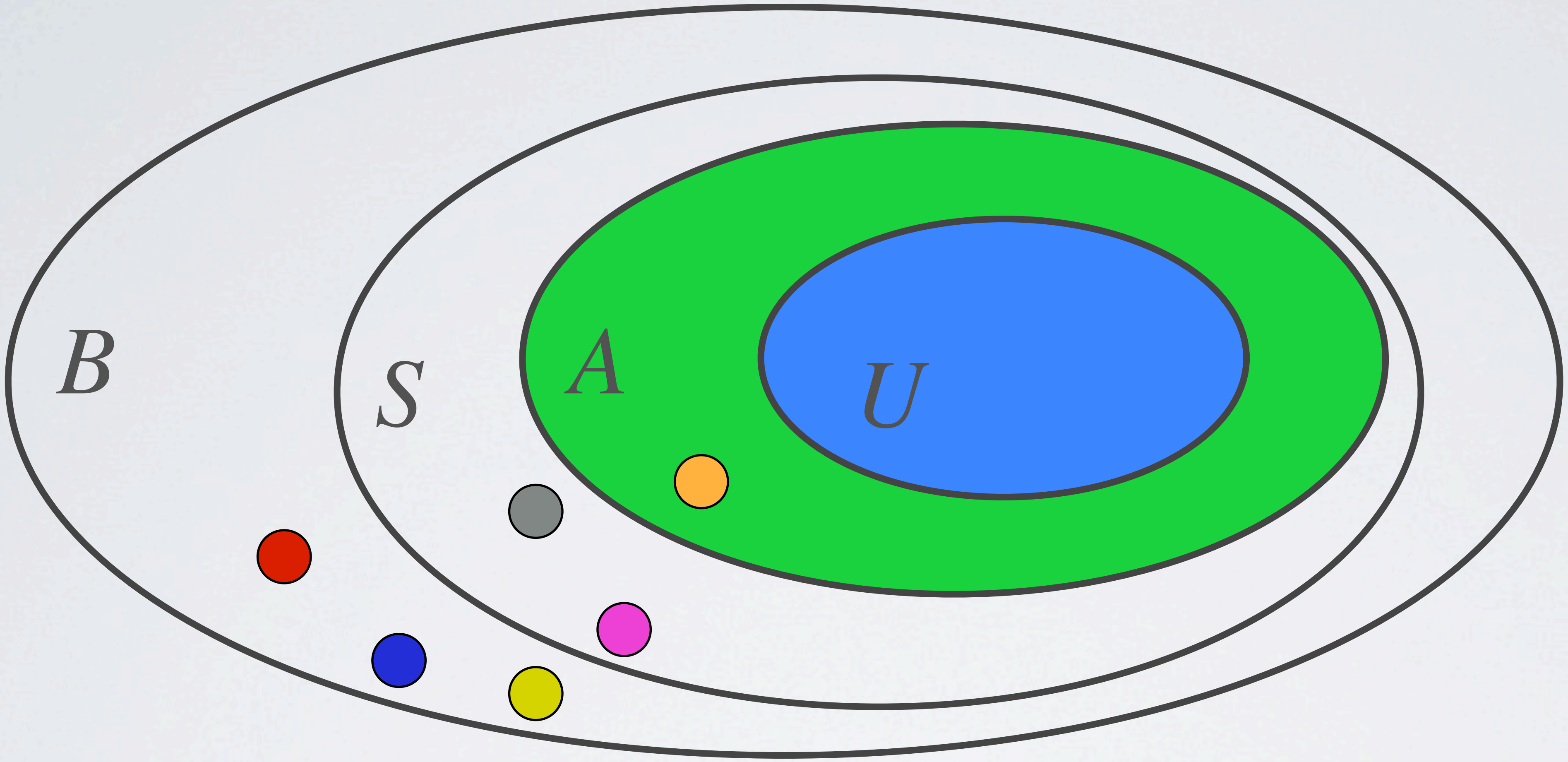
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider
- $B \supseteq S$, informed nodes at the current phase

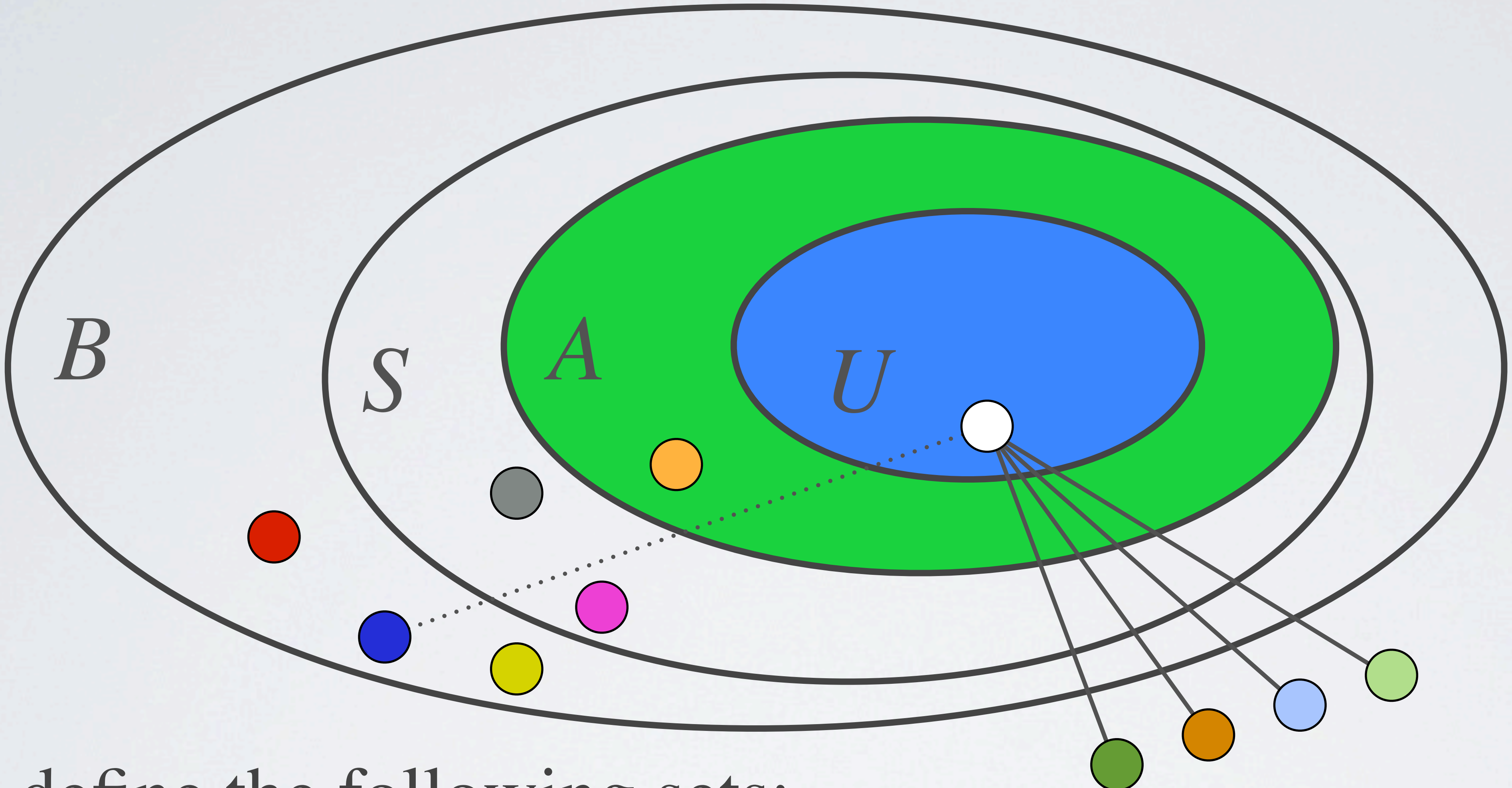
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider
- $B \supseteq S$, informed nodes at the current phase
- U useful nodes in A

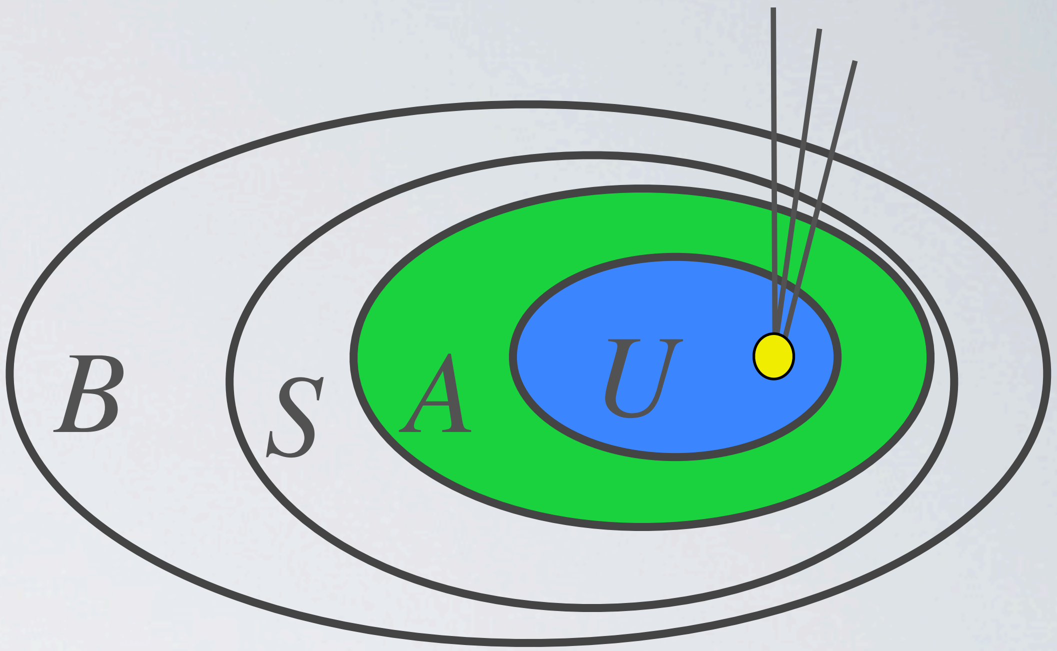
Sketch of the proof



We define the following sets:

- $A \subseteq S$, informed nodes that we still have to consider
- $B \supseteq S$, informed nodes at the current phase
- U useful nodes in A

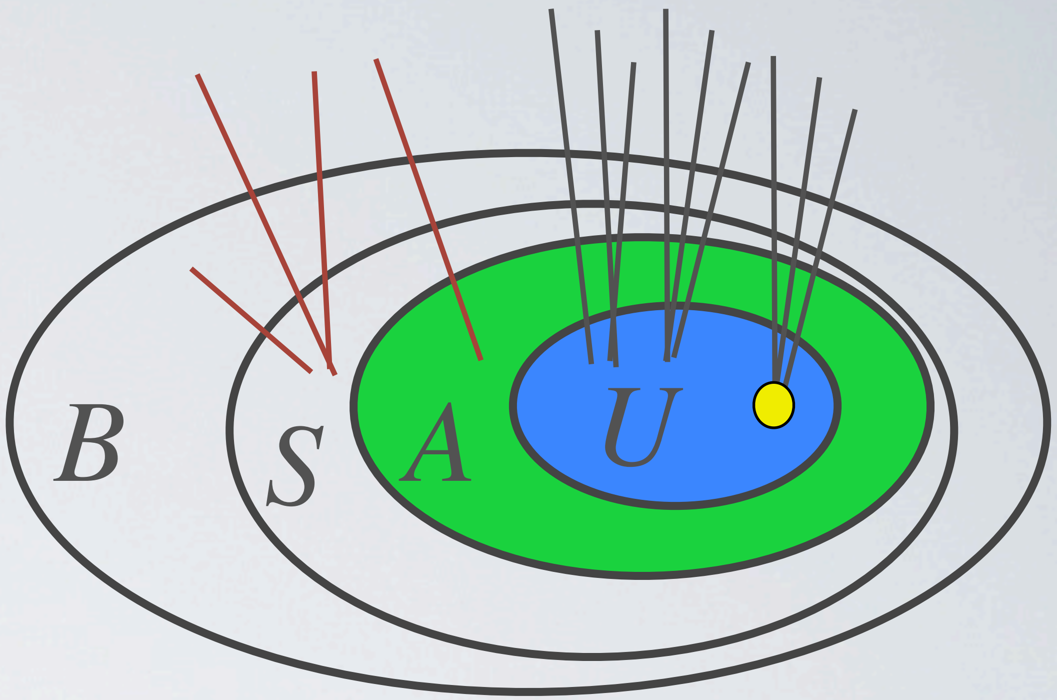
Sketch of the proof



The set of useful nodes U is

$$U = U_B(A) = \left\{ v \in A \mid \frac{\deg_B^+(v)}{\deg(v)} \geq \frac{\phi}{2} \right\}$$

Sketch of the proof

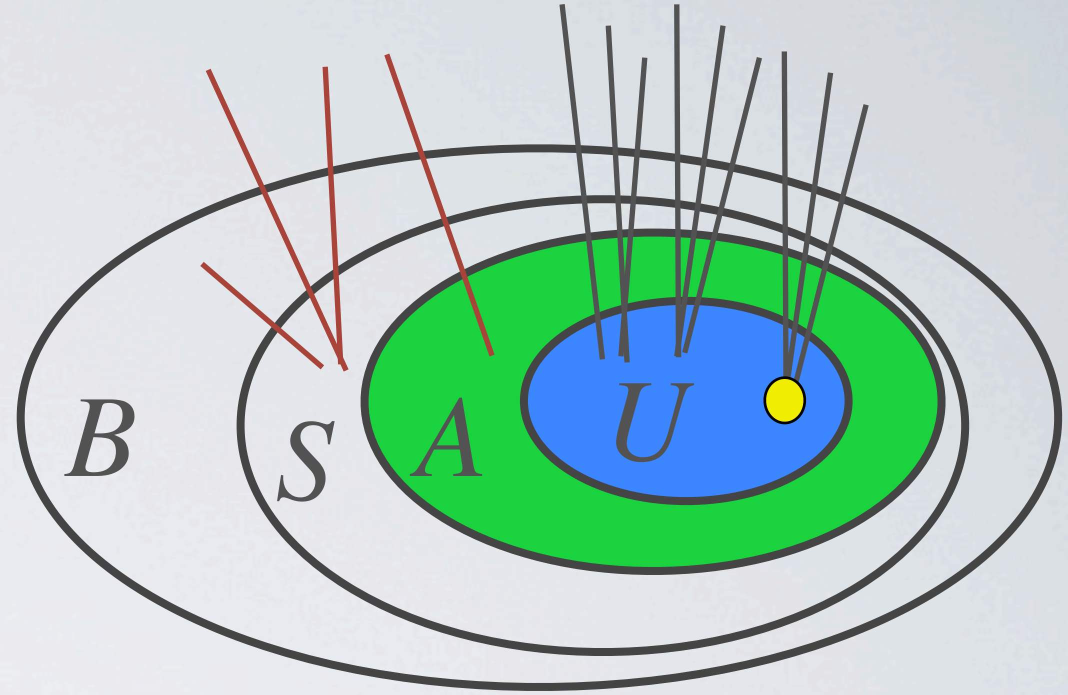


The set of useful nodes U is

$$U = U_B(A) = \left\{ v \in A \mid \frac{\deg_B^+(v)}{\deg(v)} \geq \frac{\phi}{2} \right\}$$

1. The cut (U, V - B) is a large part of the cut (S, V - S), which has size at least Φ Vol(S).

Sketch of the proof



The set of useful nodes *U* is

$$U = U_B(A) = \left\{ v \in A \mid \frac{\deg_B^+(v)}{\deg(v)} \geq \frac{\phi}{2} \right\}$$

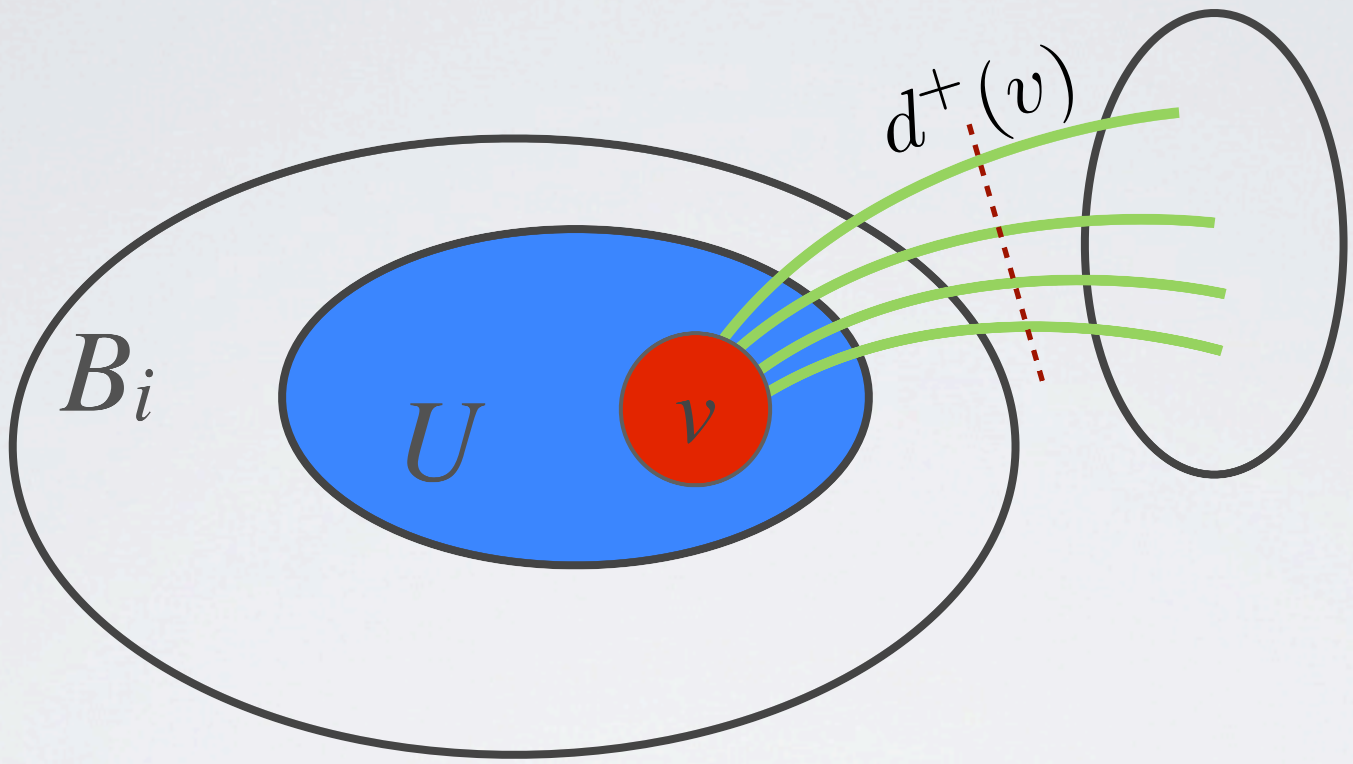
1. The cut (U, V - B) is a large part of the cut (S, V - S), which has size at least Φ Vol(S).
2. And, furthermore, each node in U will have constant probability of gaining a constant fraction of its edges in the cut.

Sketch of the proof

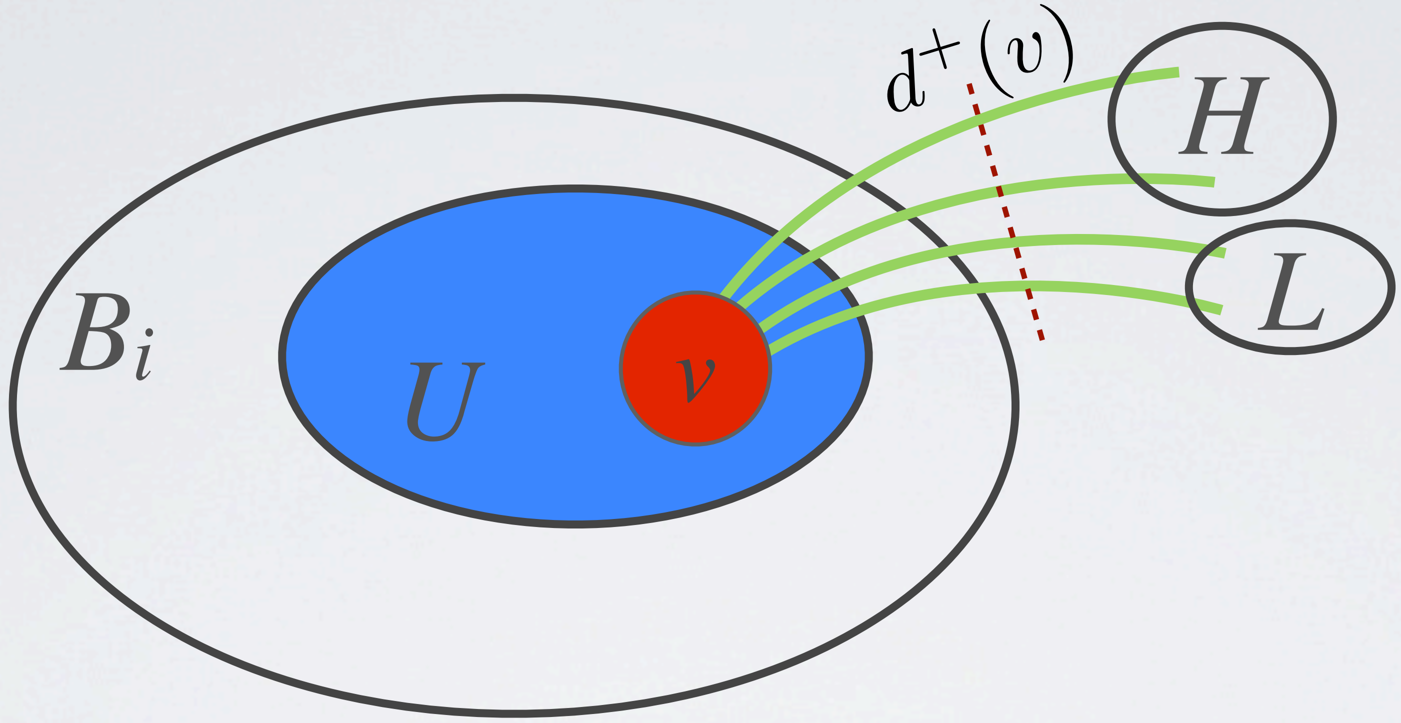
In order to get the key lemma we prove that for every macro-phase, and every *v* in *U*

$$\Pr\left[G(v) \geq \frac{1}{20} \cdot \deg_B^+(v)\right] \geq 1 - e^{-1}$$

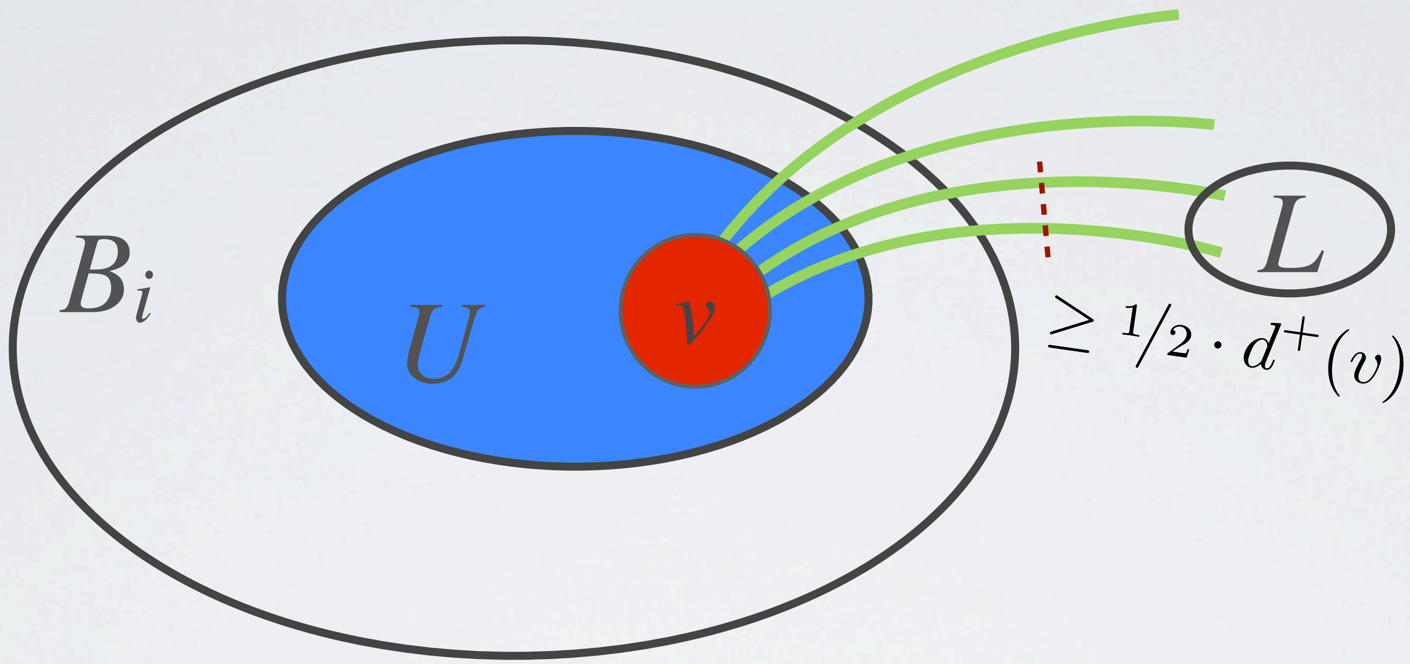
Sketch of the proof



Sketch of the proof



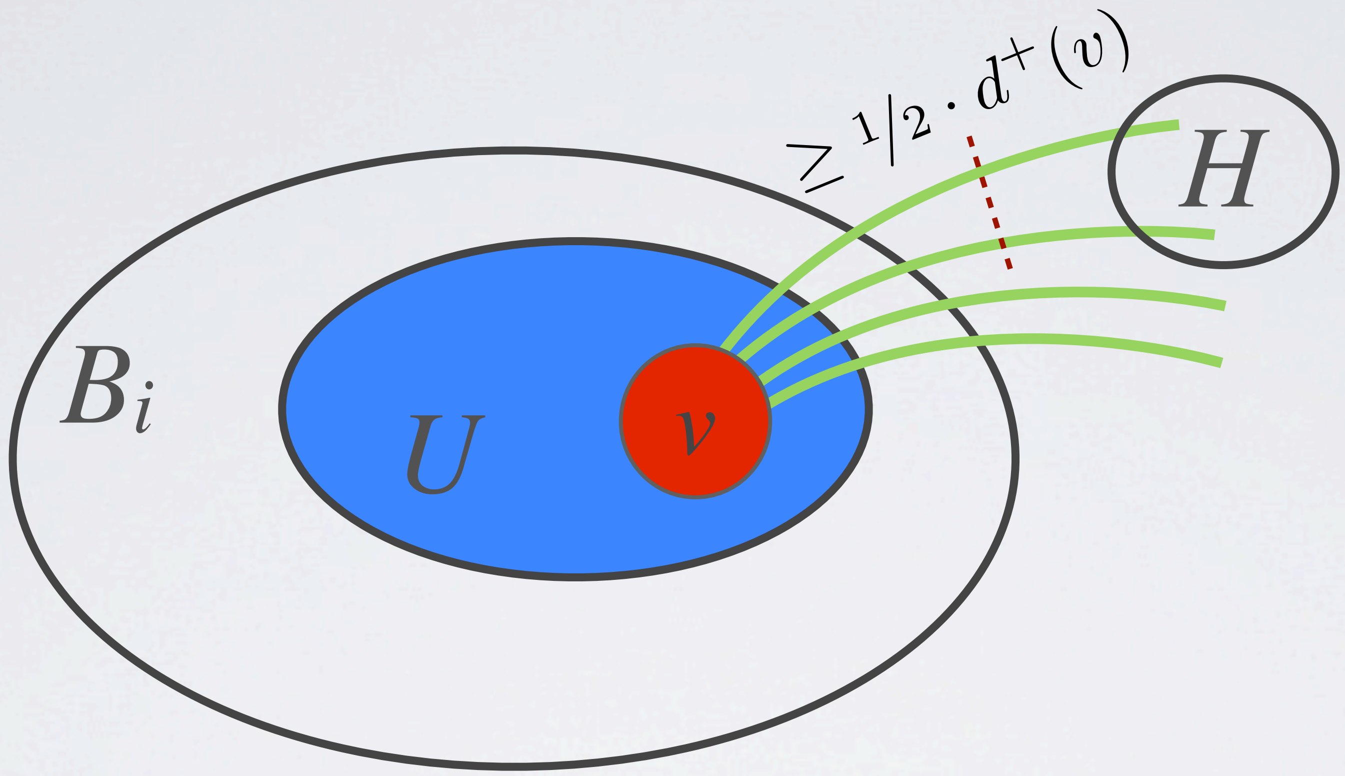
Sketch of the proof



By applying Chebyshev inequality with some arithmetic manipulation, we get that, in the PULL regime:

$$\Pr\left[g(v) > \frac{1}{20} \cdot d^+(v)\right] \geq \frac{1}{10}$$

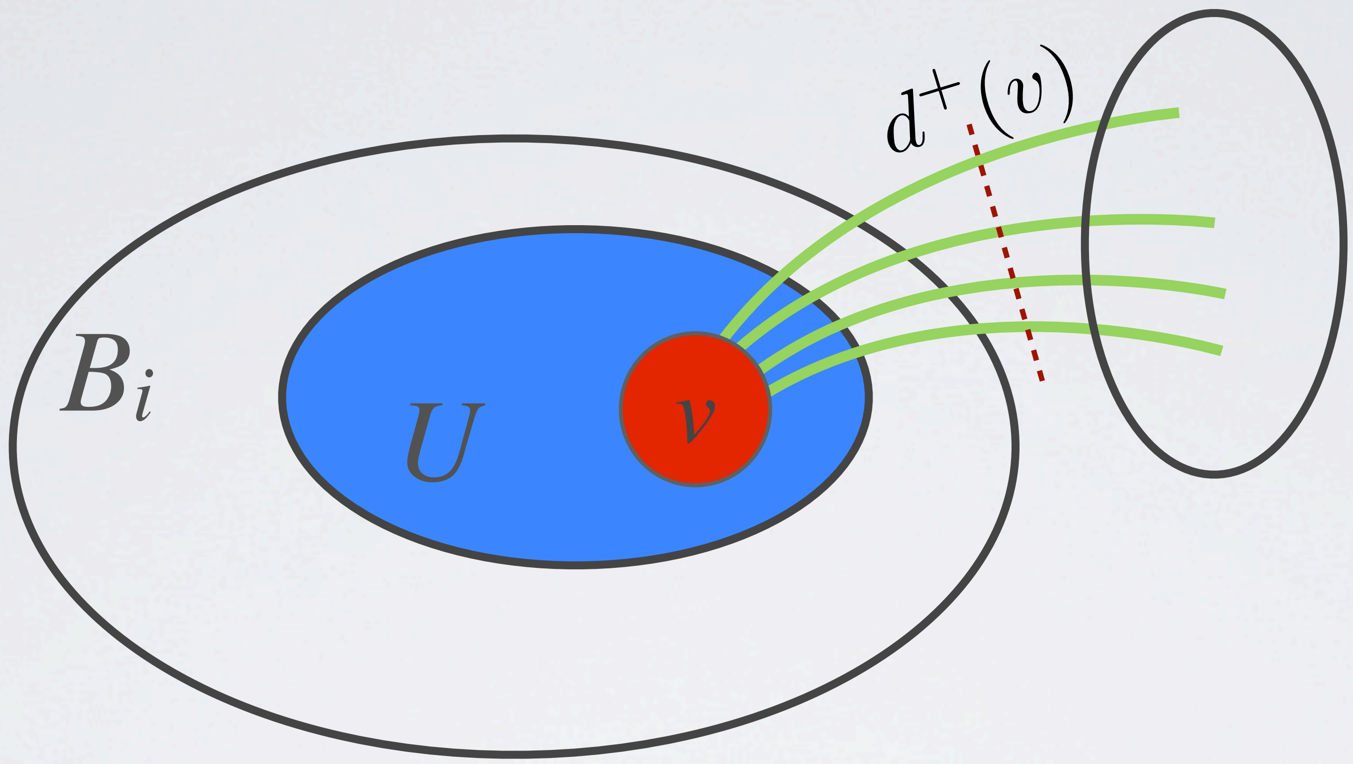
Sketch of the proof



In the PUSH regime, we had:

$$\Pr\left[g(v) > \frac{1}{20} \cdot d^+(v)\right] \geq \frac{\phi}{10}$$

Sketch of the proof



So, in general,

$$\Pr\left[g(v) > \frac{1}{20} \cdot d^+(v)\right] \geq \frac{\phi}{10}$$

Sketch of the proof

Since we go on for Φ^{-1} steps,

$$\Pr\left[G(v) \geq \frac{1}{20} \cdot \deg_B^+(v)\right] \geq 1 - e^{-1}$$

Upper bound

Let G be a graph with conductance Φ, then w.h.p.

$$T_{PUSH-PULL} = O\left(\frac{\log n}{\Phi^2}\right)$$

The tighter bound

Let G be a graph with conductance Φ, then w.h.p.

$$T_{PUSH-PULL} = O\left(\frac{\log n}{\Phi}\left(\log\frac{1}{\Phi}\right)^2\right)$$

Can the key lemma be improved?

After $\Theta(\Phi^{-1})$ steps with $\Theta(1)$ probability, if S' is the new set of informed nodes, then

$$\mathrm{Vol}(S') \geq (1 + \Omega(\Phi))\, \mathrm{Vol}(S)$$

Can the key lemma be improved?

After $\Theta(\Phi^{-1})$ steps with $\Theta(1)$ probability, if S' is the new set of informed nodes, then

$$\mathrm{Vol}(S') \geq (1 + \Omega(\Phi))\ \mathrm{Vol}(S)$$

Can the key lemma be improved?

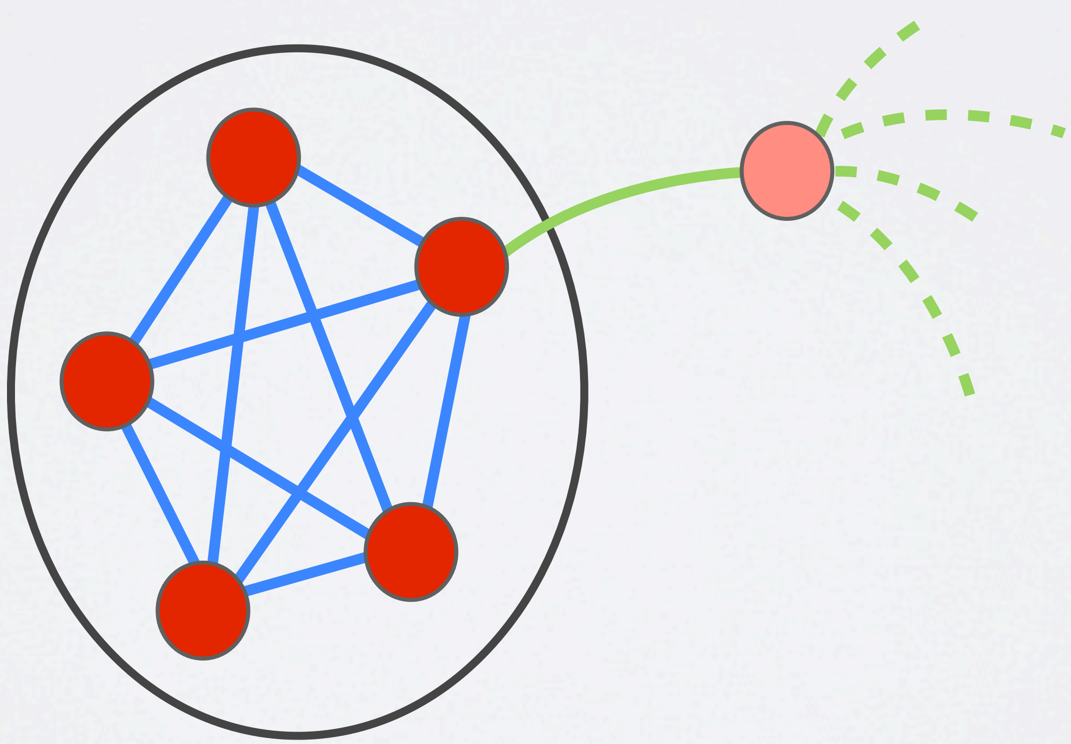
After $\Theta(\Phi^{-1})$ steps with $\Theta(1)$ probability, if S' is the new set of informed nodes, then

$$\text{Vol}(S') \geq (1 + \Omega(\Phi))\, \text{Vol}(S)$$

Can the key lemma be improved?

After $\Theta(\Phi^{-1})$ steps with $\Theta(1)$ probability, if S' is the new set of informed nodes, then

$$\mathrm{Vol}(S') \geq (1 + \Omega(\Phi))\ \mathrm{Vol}(S)$$

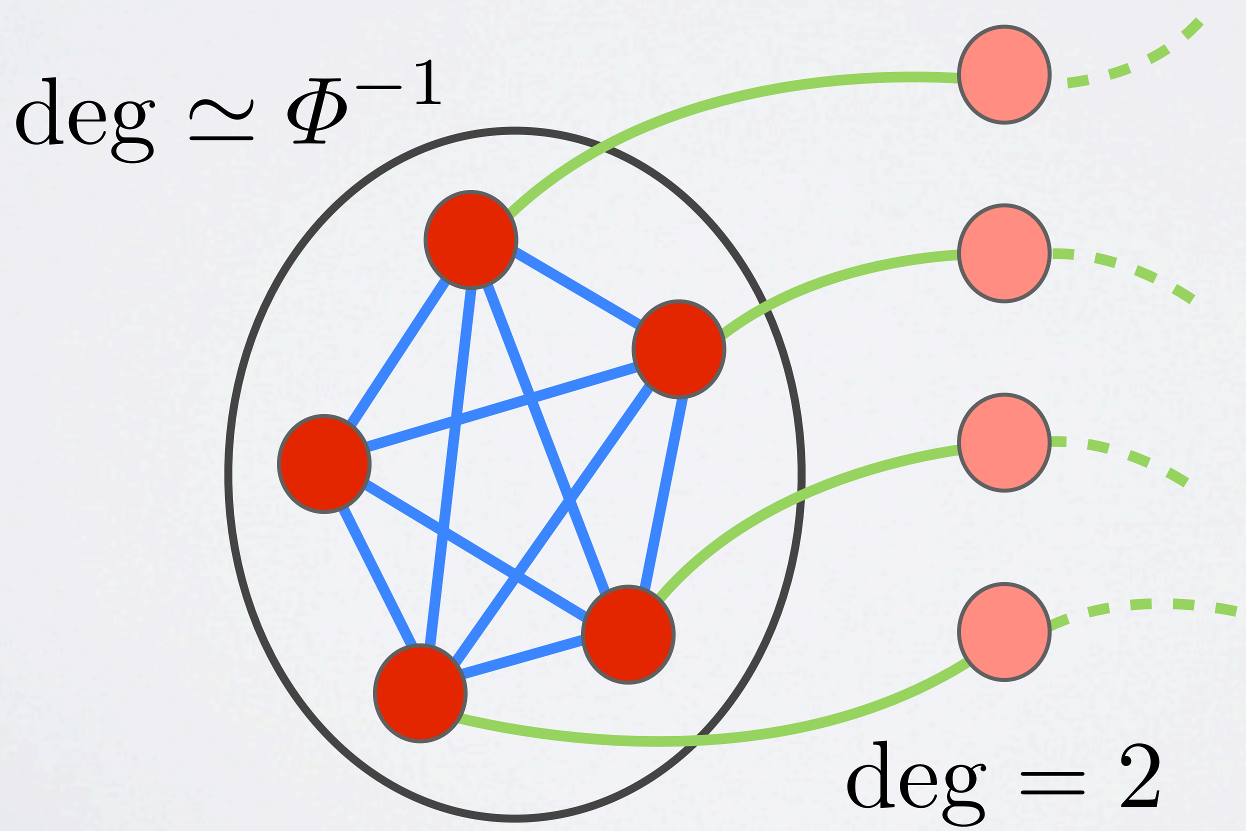


$$\forall v \ \ \deg(v) \simeq \Phi^{-1}$$

Can the key lemma be improved?

After $\Theta(\Phi^{-1})$ steps with $\Theta(1)$ probability, if S' is the new set of informed nodes, then

$$\mathrm{Vol}(S') \geq (1 + \Omega(\Phi))\, \mathrm{Vol}(S)$$



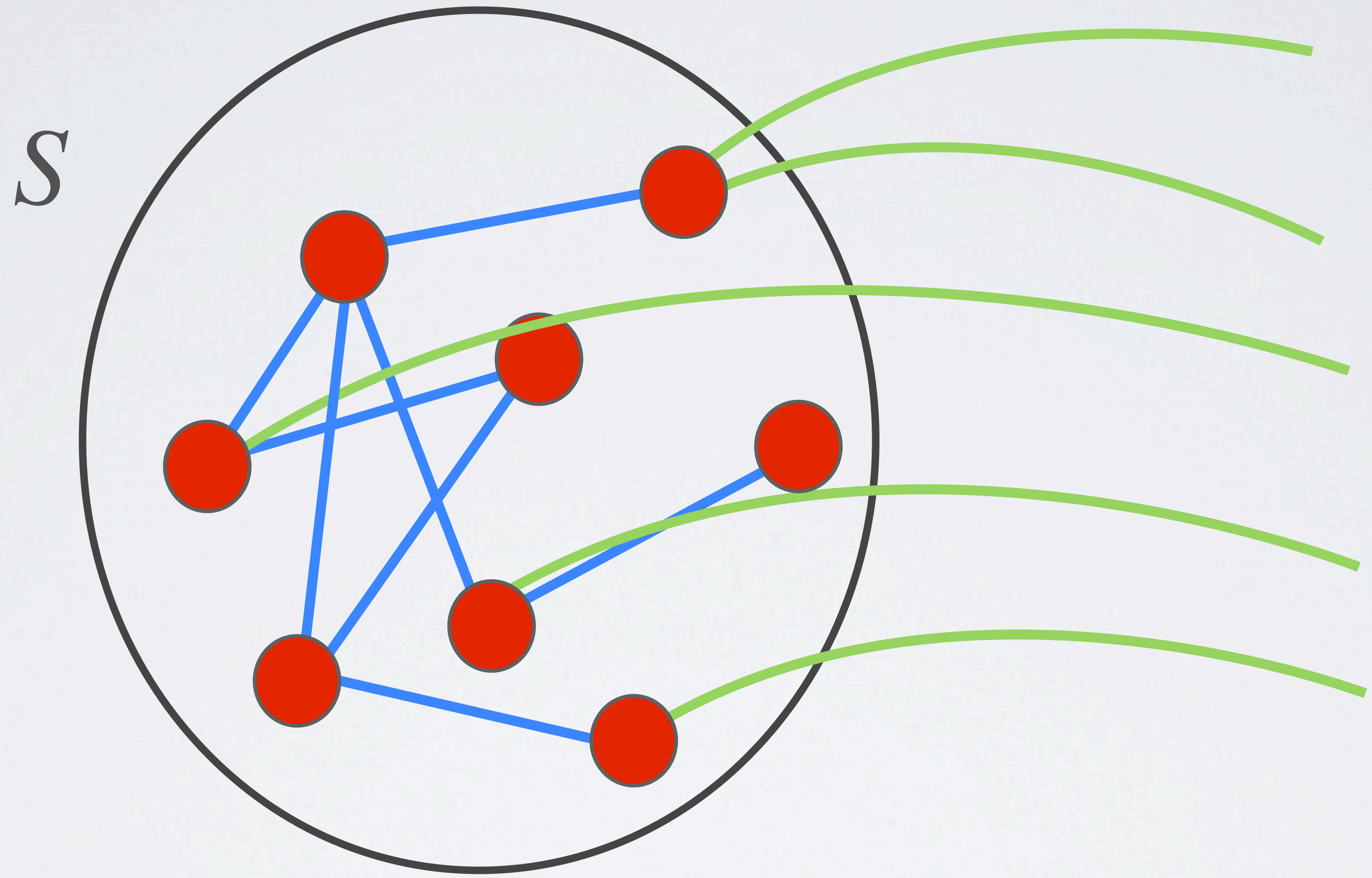
Can the key lemma be improved?

After $\Theta(\Phi^{-1})$ steps with $\Theta(1)$ probability, if S' is the new set of informed nodes, then

$$\mathrm{Vol}(S') \geq (1 + \Omega(\Phi))\ \mathrm{Vol}(S)$$

No?

Stronger key lemma



For some $p \geq \Phi$, after $O(1/p)$ steps with constant probability, we have that for the new set of informed nodes S'

$$\mathrm{vol}(S') \geq \left(1 + \Omega\left(\frac{\phi}{p \log^2 \phi^{-1}}\right)\right) \cdot \mathrm{vol}(S)$$

Conclusion

- We studied the rumor spreading problem in graph of conductance Φ.
- We showed that the *PUSH* and the *PULL* strategies are not fast,
- and that the *PUSH-PULL* strategy is fast, and we gave an almost tight bound for its performance.

Conclusion

- We studied the rumor spreading problem in graph of conductance Φ.
- We showed that the *PUSH* and the *PULL* strategies are not fast, (fast if some kind of "degree uniformity" exists)
- and that the *PUSH-PULL* strategy is fast, and we gave an almost tight bound for its performance.

Open problems

- Find a tight bound for the *PUSH-PULL* strategy.
- Study the relationship between rumor spreading and vertex expansion.
- Can the *PUSH* strategy inform efficiently a large part of a social network?

Thank you!
Questions?