A Numerical Evaluation of the Accuracy of Influence Maximization Algorithms

3rd International Workshop on MAISoN, Santa Clara

October 2, 2019

Hautahi Kingi
IMPAQ International

Li-An Daniel Wang
Sam Houston State University

Tom Shafer
Elder Research

Minh Huynh
IMPAQ International

Mike Trinh
IMPAQ International

Aaron Heuser
IMPAQ International

George Rochester
U.S. Food and Drug Administration

Antonio Paredes
U.S. Food and Drug Administration

Latest Paper Version: hautahi.com/work
Code: github.com/hautahi/IM-Evaluation

Agenda

1. Introduction to Influence Maximization

Agenda

1. Introduction to Influence Maximization
2. Approximate (1 - 1/e) Solutions
   - Greedy
   - Reverse Influence Sampling

Agenda

1. Introduction to Influence Maximization
2. Approximate (1 - 1/e) Solutions
   - Greedy
   - Reverse Influence Sampling
3. Exact Solution

Agenda

1. Introduction to Influence Maximization
2. Approximate (1 - 1/e) Solutions
    - Greedy
    - Reverse Influence Sampling
3. Exact Solution
4. Results

Influence Maximization

- Which nodes are most influential?

Influence Maximization

- Which nodes are most influential?
- Applications: Viral Marketing, Epidemiology, Fault Monitoring, Public Health

Influence Maximization

- Which nodes are most influential?
- Applications: Viral Marketing, Epidemiology, Fault Monitoring, Public Health
- Formal Primitives: Graph $G = (V, E)$, Spread Function $\sigma(S) \mapsto \mathbb{R}$

Influence Maximization

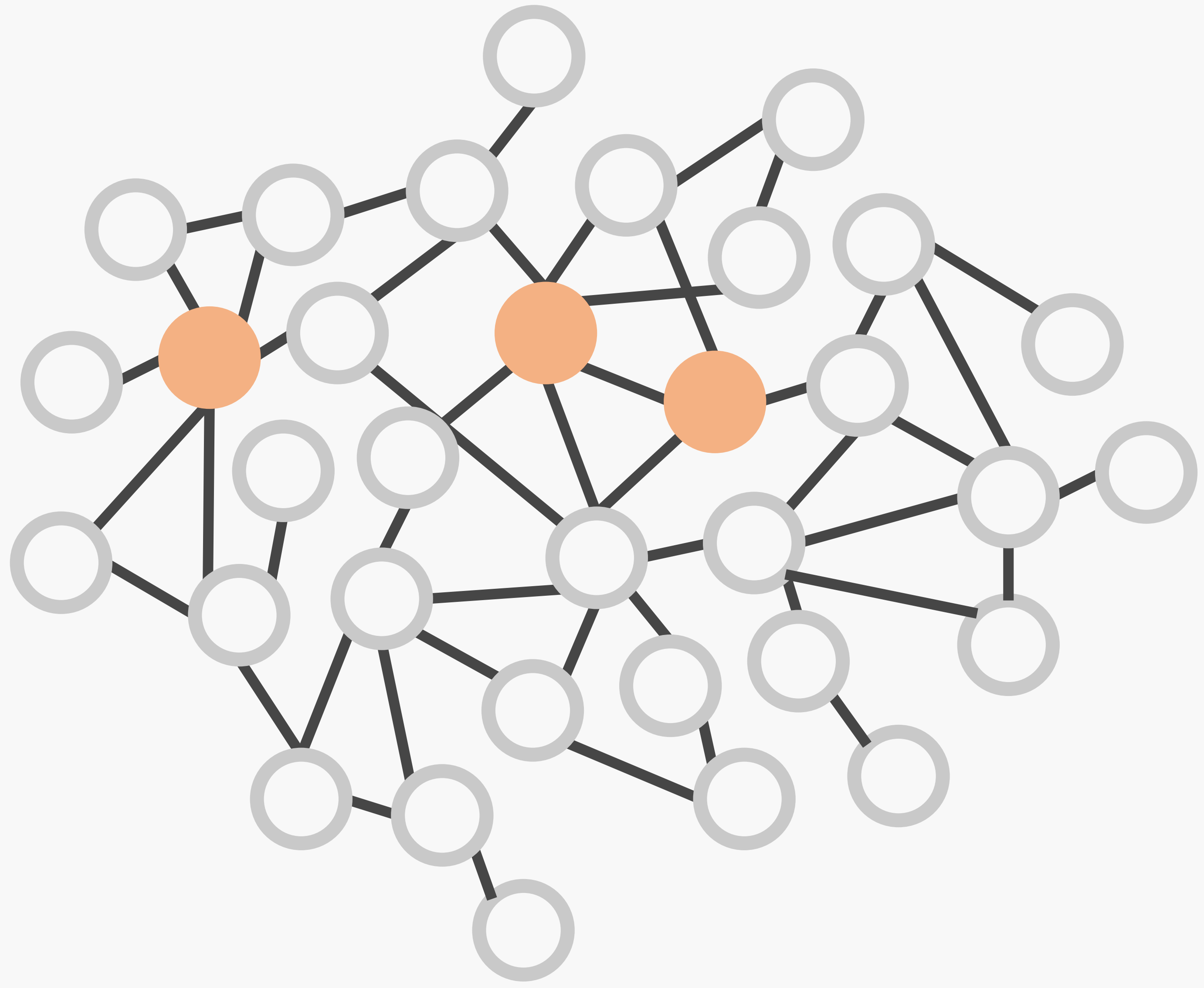
- Which nodes are most influential?
- Applications: Viral Marketing, Epidemiology, Fault Monitoring, Public Health
- Formal Primitives: Graph $G = (V, E)$, Spread Function $\sigma(S) \mapsto \mathbb{R}$
- Kempe et al. (2003) formulation:

$$\max_{S \subseteq V} \sigma(S) \quad \text{s.t.} \quad |S| \leq k$$

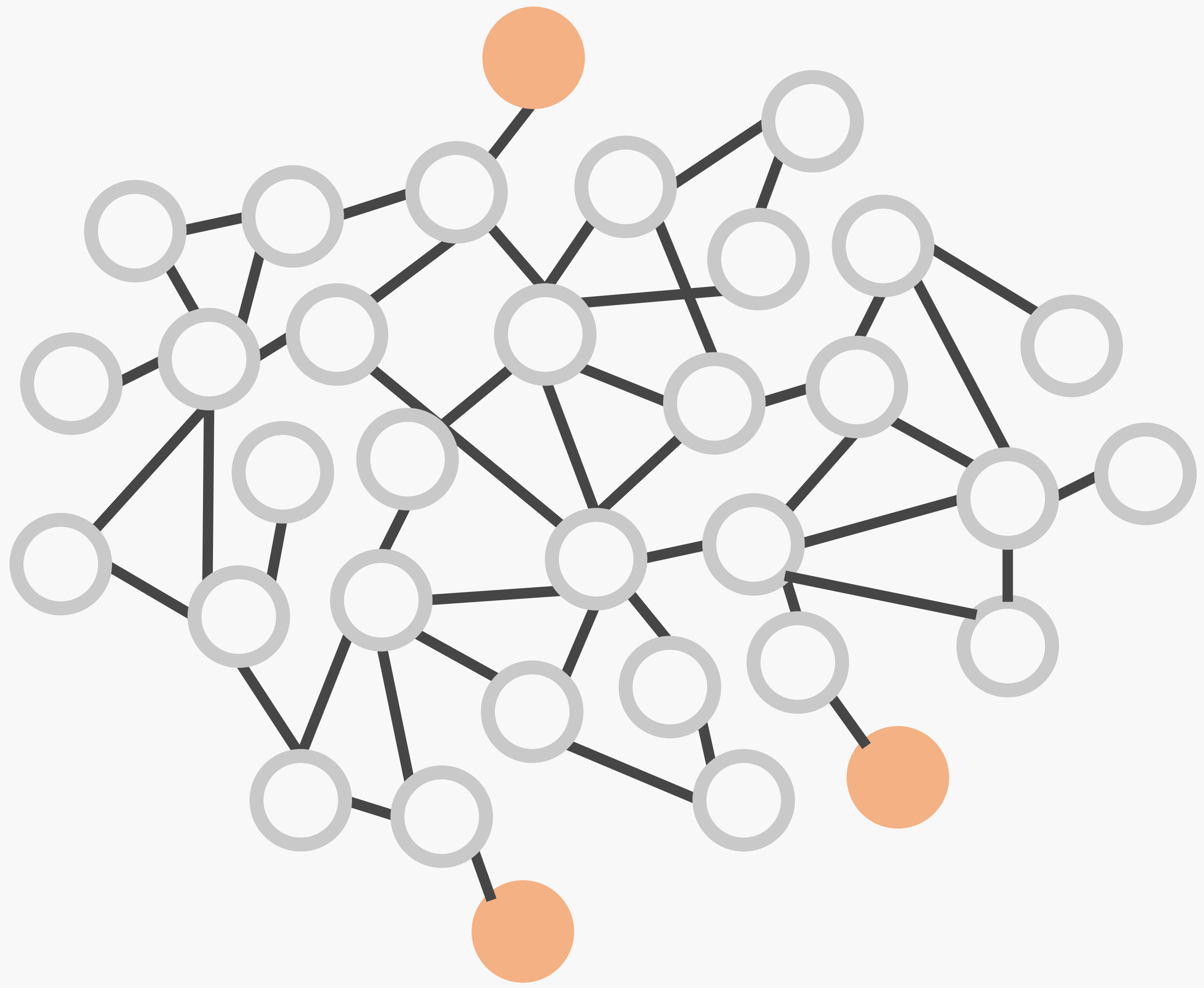
Influence Maximization

Influence Maximization

S_1

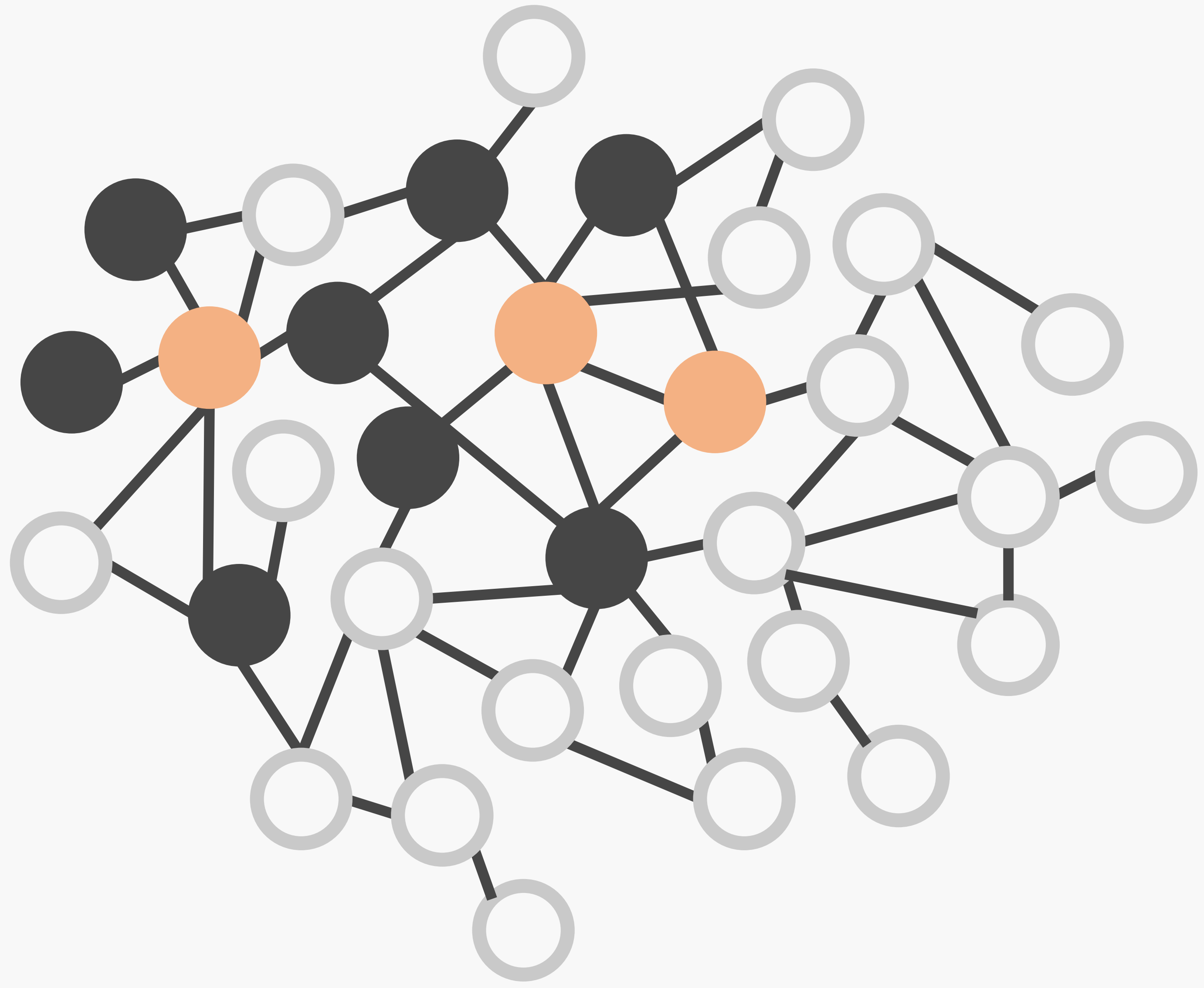


S_2

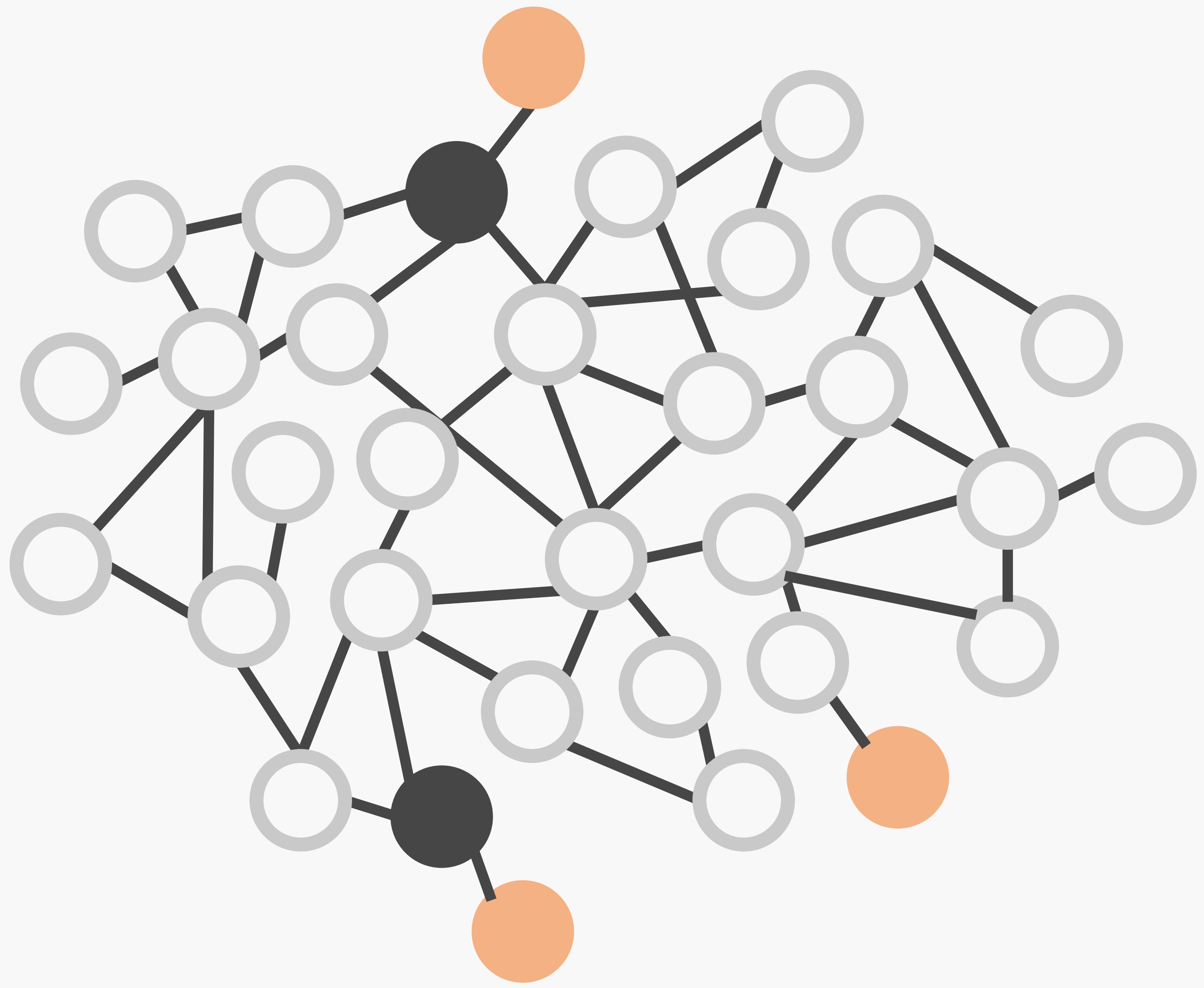


Influence Maximization

$\sigma(S_1)$

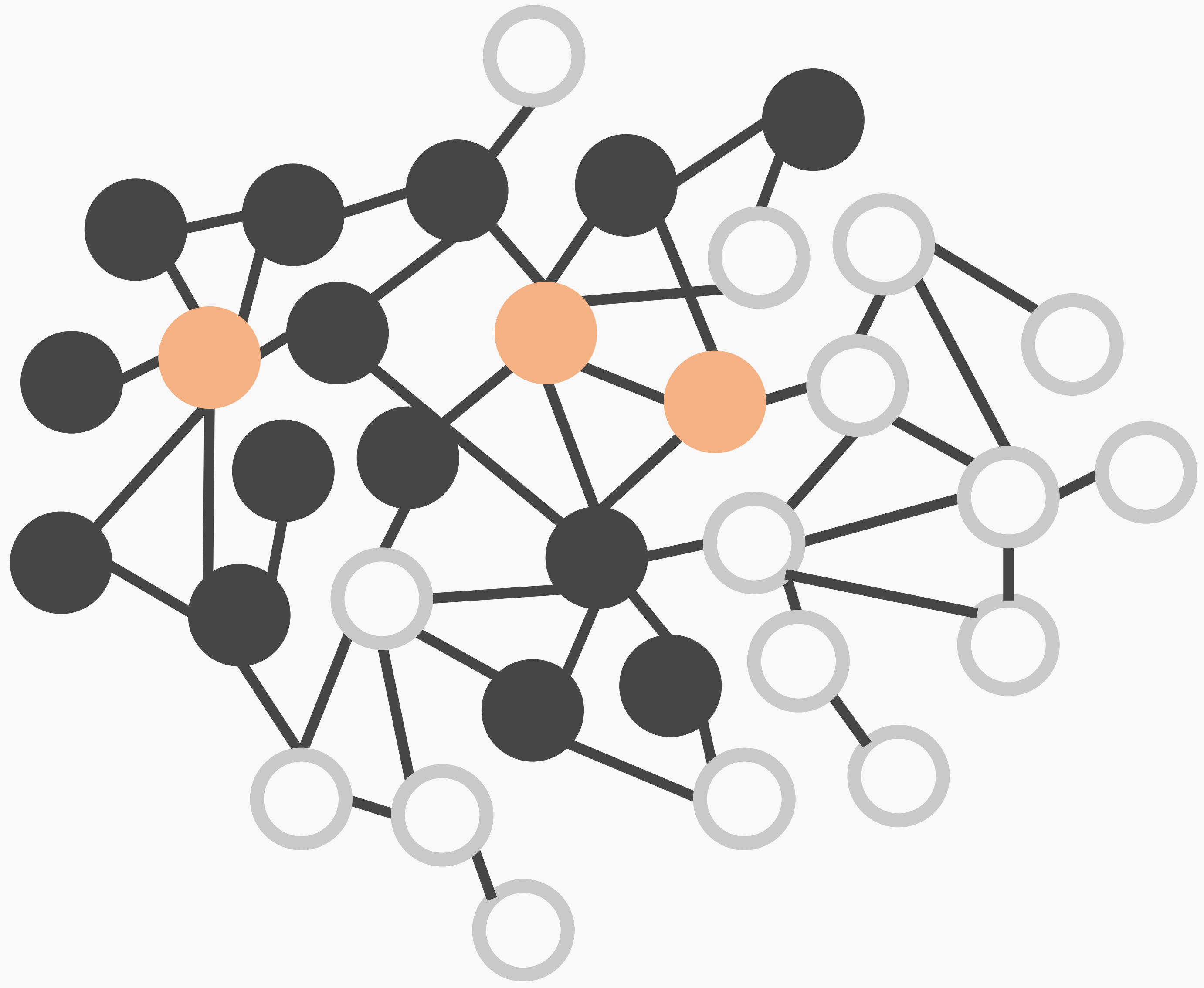


$\sigma(S_2)$

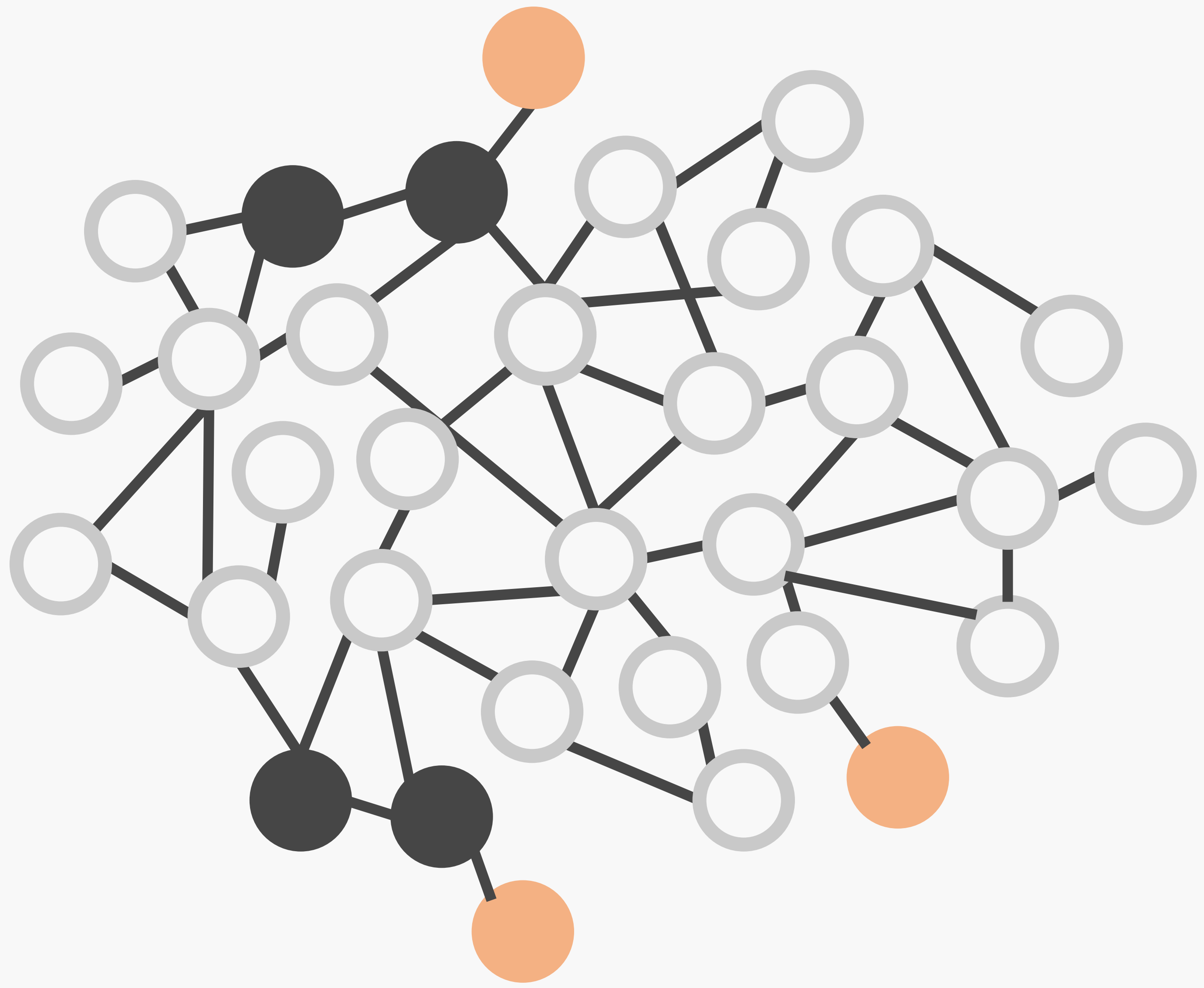


Influence Maximization

$\sigma(S_1)$

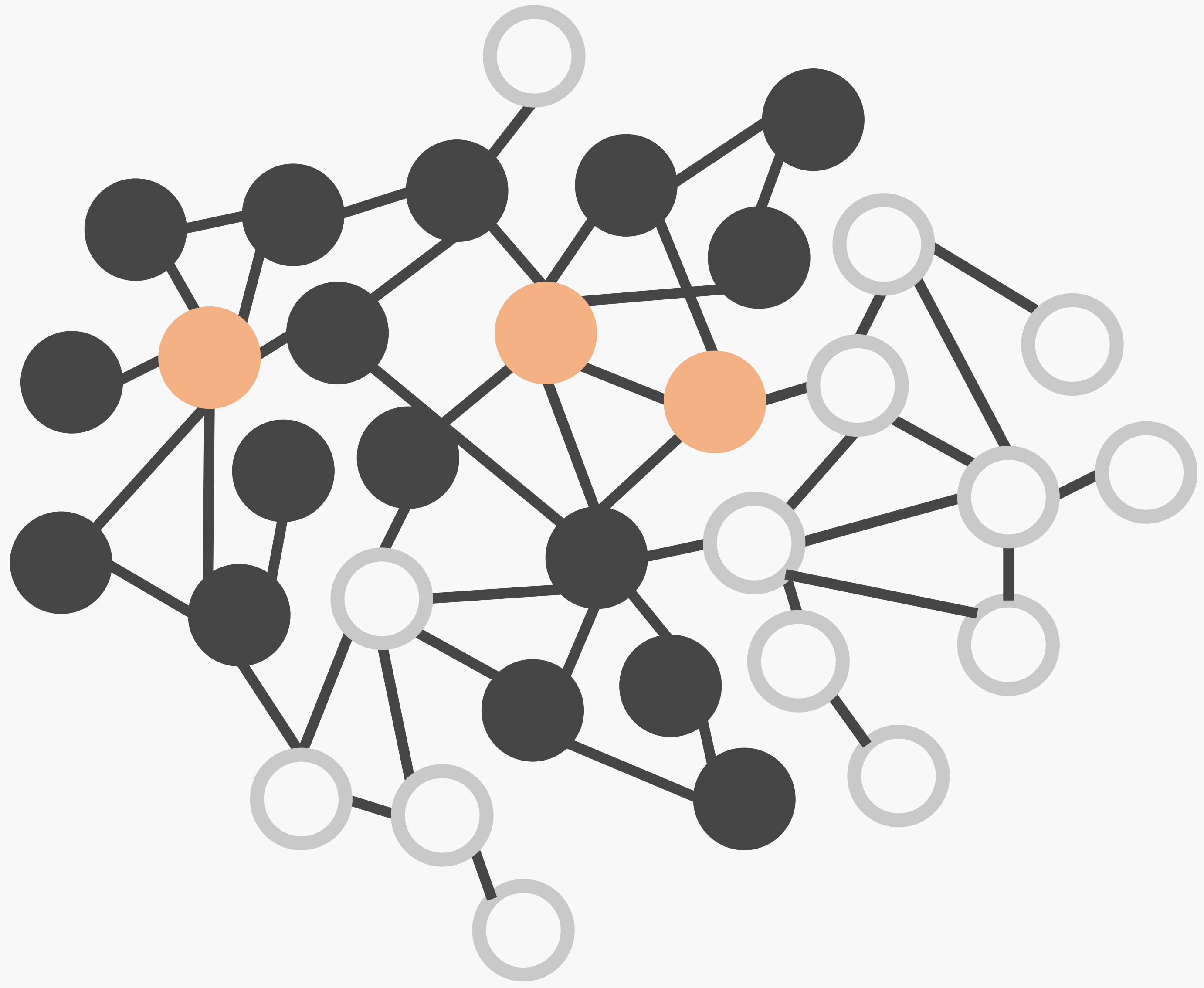


$\sigma(S_2)$

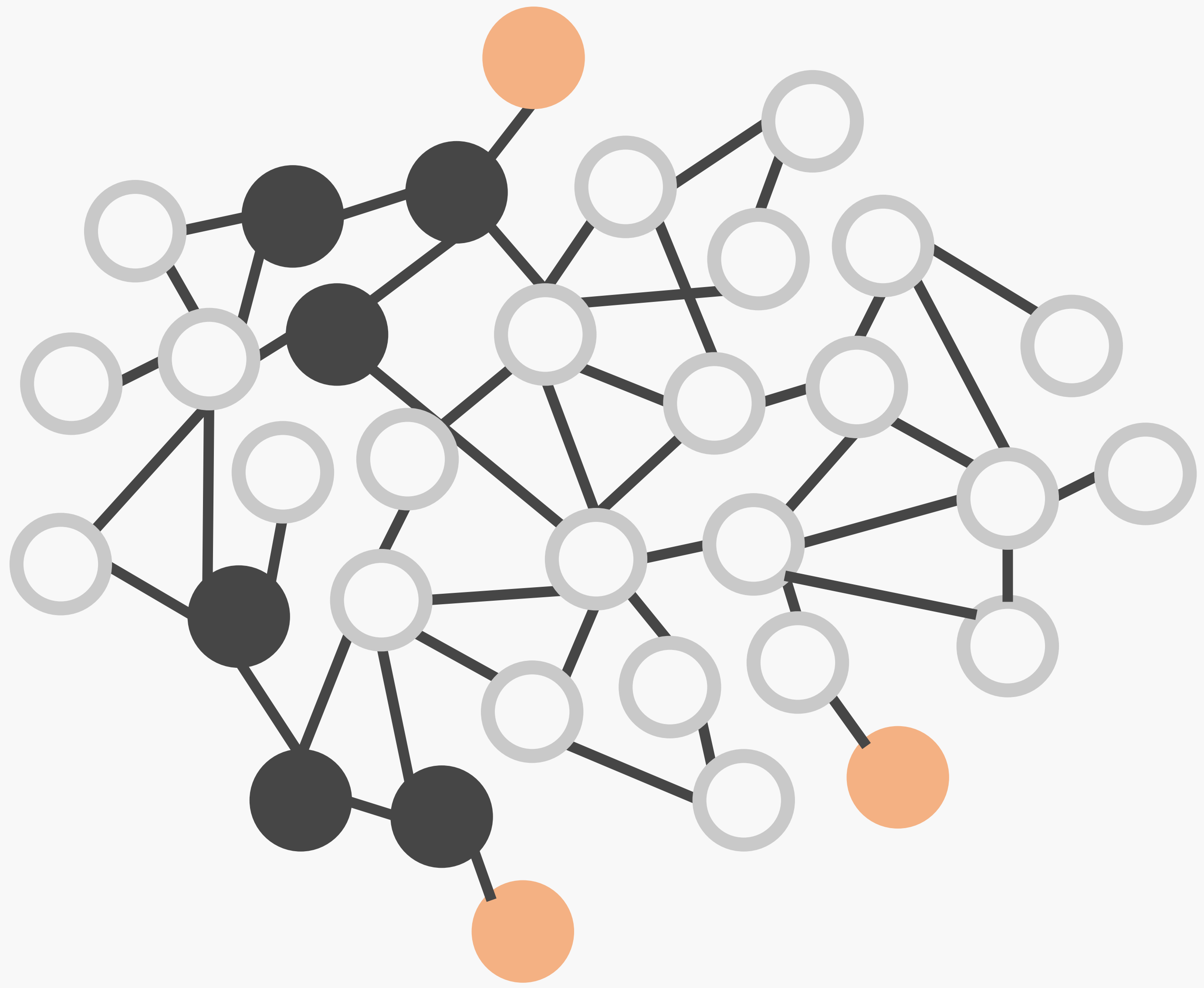


Influence Maximization

$\sigma(S_1)$

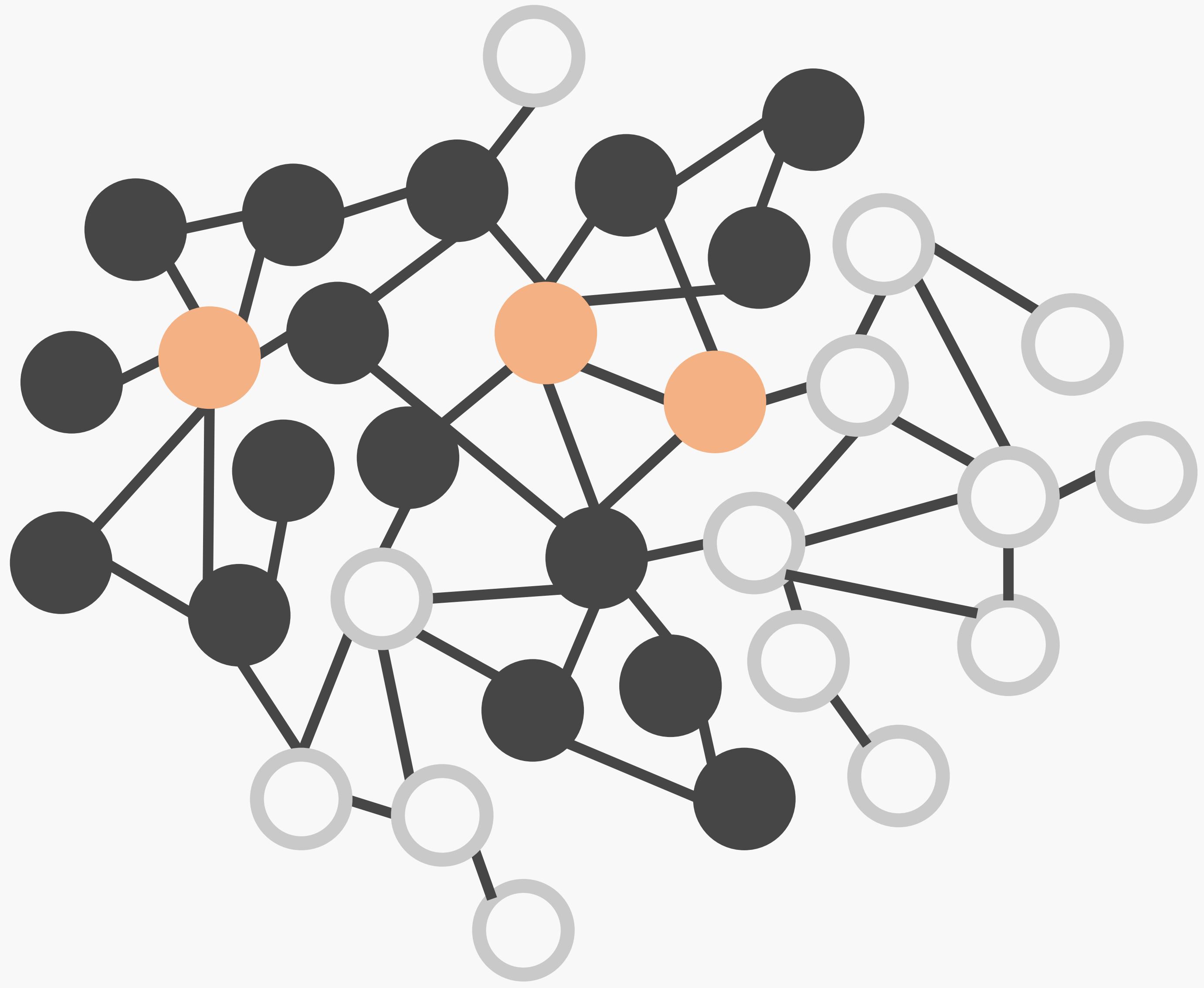


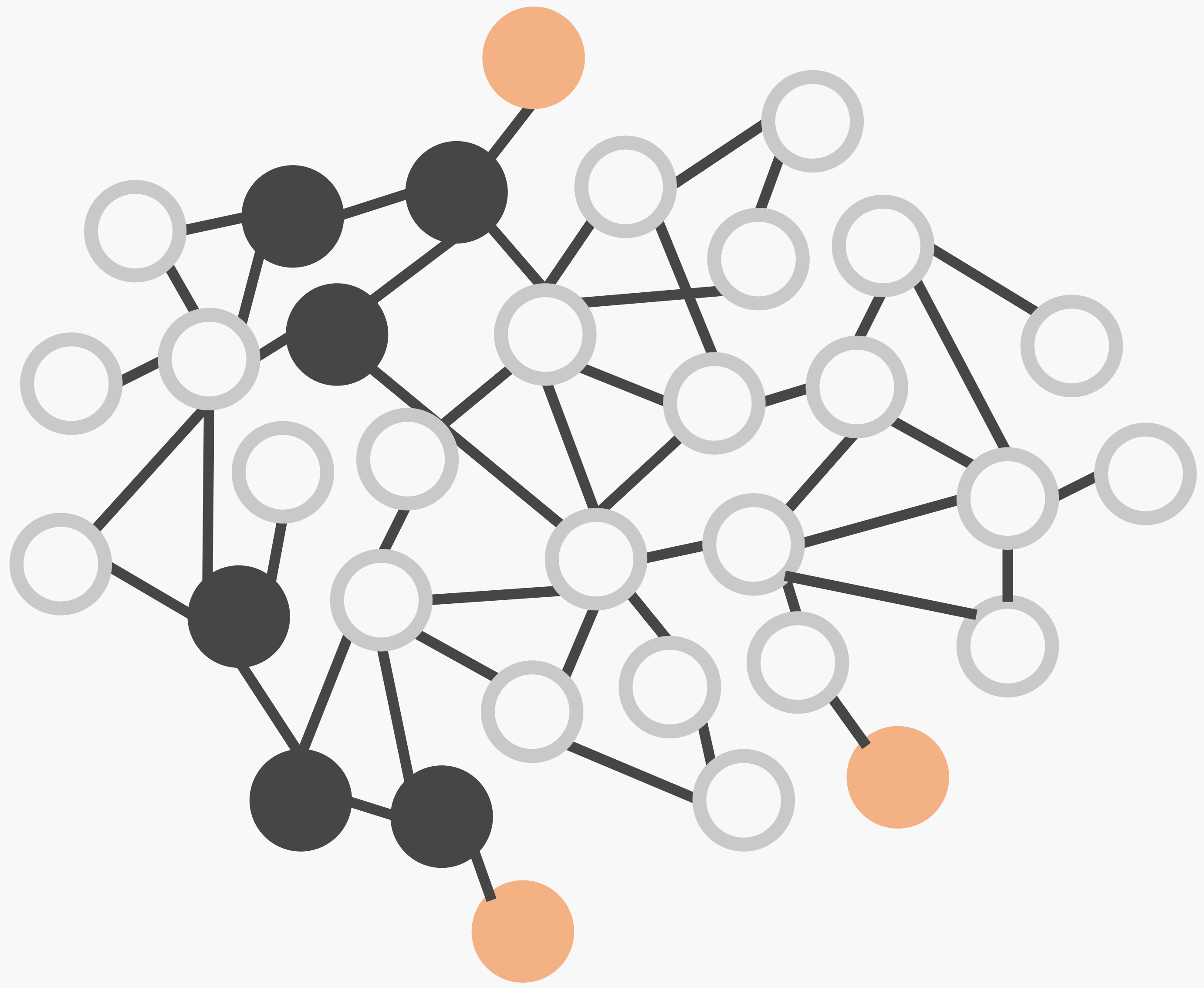
$\sigma(S_2)$



Influence Maximization

$$\sigma(S_1) > \sigma(S_2)$$

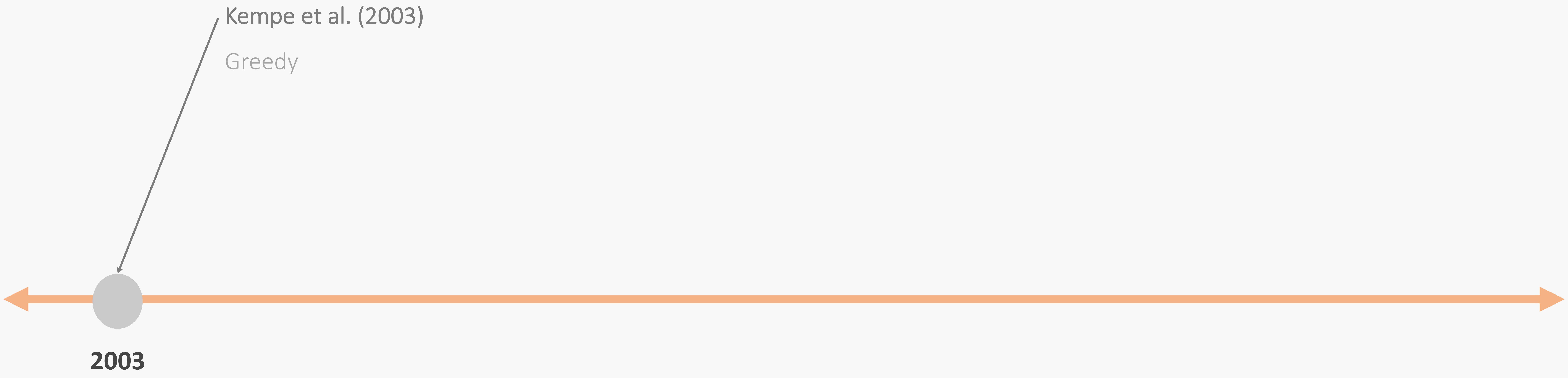




A Brief History

Computationally Challenging

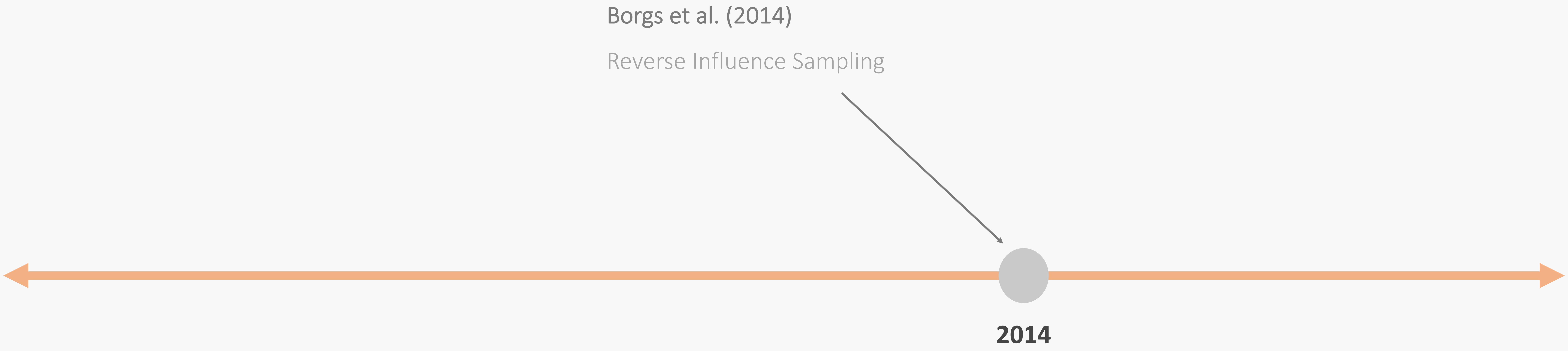
1. Size of domain is $\binom{\boldsymbol{n}}{\boldsymbol{k}}$
2. $\boldsymbol{\sigma}(\cdot)$ estimated via expensive Monte Carlo



A Brief History

Computationally Challenging

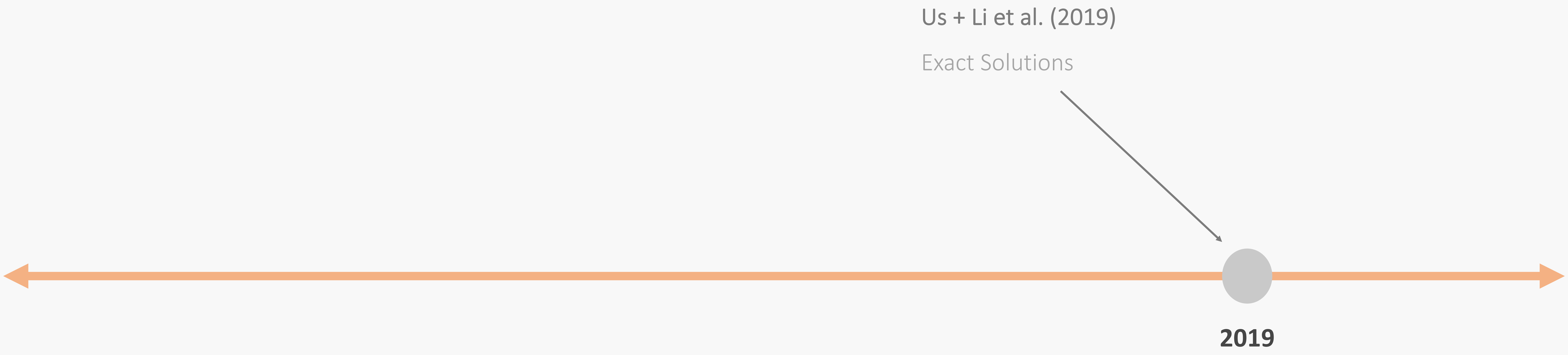
1. Size of domain is $\binom{n}{k}$
2. $\boldsymbol{\sigma}(\cdot)$ estimated via expensive Monte Carlo



A Brief History

Computationally Challenging

1. Size of domain is $\binom{n}{k}$
2. $\boldsymbol{\sigma}(\cdot)$ estimated via expensive Monte Carlo



Greedy Solution

Greedy Approximation

```
S, max = ∅, 0
for 1:k
    for v in V \ S
        if σ(S ∪ v) > max
            max, v* = σ(S ∪ v), v
        end if
    end for
    S = S ∪ v*
end for
```

Greedy Approximation

$\boldsymbol{S}, \mathbf{max} = \emptyset, \mathbf{0}$

for $\mathbf{1}: \boldsymbol{k}$

  for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

    if $\boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}) > \mathbf{max}$

      $\mathbf{max}, \mathbf{v}^* = \boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}), \boldsymbol{v}$

    end if

  end for

  $\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

$$\boldsymbol{O}(\boldsymbol{k} \cdot \boldsymbol{n} \cdot \boldsymbol{mc})$$

Greedy Approximation

$\boldsymbol{S}, \mathbf{max} = \emptyset, \mathbf{0}$

for $\mathbf{1}: \boldsymbol{k}$

    for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

        if $\boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}) > \mathbf{max}$

            $\mathbf{max}, \mathbf{v}^* = \boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}), \boldsymbol{v}$

        end if

    end for

    $\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

$$\boldsymbol{O}(\boldsymbol{k} \cdot \boldsymbol{n} \cdot \boldsymbol{mc})$$

$\boldsymbol{\sigma}(\cdot)$ sub-modular!

Greedy Approximation

$\boldsymbol{S}, \mathbf{max} = \emptyset, \mathbf{0}$

for $\mathbf{1}: \boldsymbol{k}$

  for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

    if $\boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}) > \mathbf{max}$

      $\mathbf{max}, \mathbf{v}^* = \boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}), \boldsymbol{v}$

    end if

  end for

  $\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

$$\boldsymbol{O}(\boldsymbol{k} \cdot \boldsymbol{n} \cdot \boldsymbol{mc})$$

$\boldsymbol{\sigma}(\cdot)$ sub-modular!

$$\sigma(S_{Greedy}) \geq \left(1 - \frac{1}{e} - \epsilon\right) \sigma(S_{OPT})$$

Greedy Approximation

$\boldsymbol{S}, \mathbf{max} = \emptyset, \mathbf{0}$

for $\mathbf{1}: \boldsymbol{k}$

  for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

    if $\boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}) > \mathbf{max}$

      $\mathbf{max}, \mathbf{v}^* = \boldsymbol{\sigma}(\boldsymbol{S} \cup \boldsymbol{v}), \boldsymbol{v}$

    end if

  end for

  $\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

$$\boldsymbol{O(k \cdot n \cdot mc)}$$

$\boldsymbol{\sigma}(\cdot)$ sub-modular!

$$\sigma(S_{Greedy}) \geq \left(1 - \frac{1}{e} - \epsilon\right) \sigma(S_{OPT})$$

submodularity

Greedy Approximation

```
S, max = ∅, 0
for 1: k
    for v in V \ S
        if σ(S ∪ v) > max
            max, v* = σ(S ∪ v), v
        end if
    end for
    S = S ∪ v*
end for
```

$$O(k \cdot n \cdot mc)$$

$\sigma(\cdot)$ sub-modular!

$$\sigma(S_{Greedy}) \geq \left(1 - \frac{1}{e} - \epsilon\right) \sigma(S_{OPT})$$

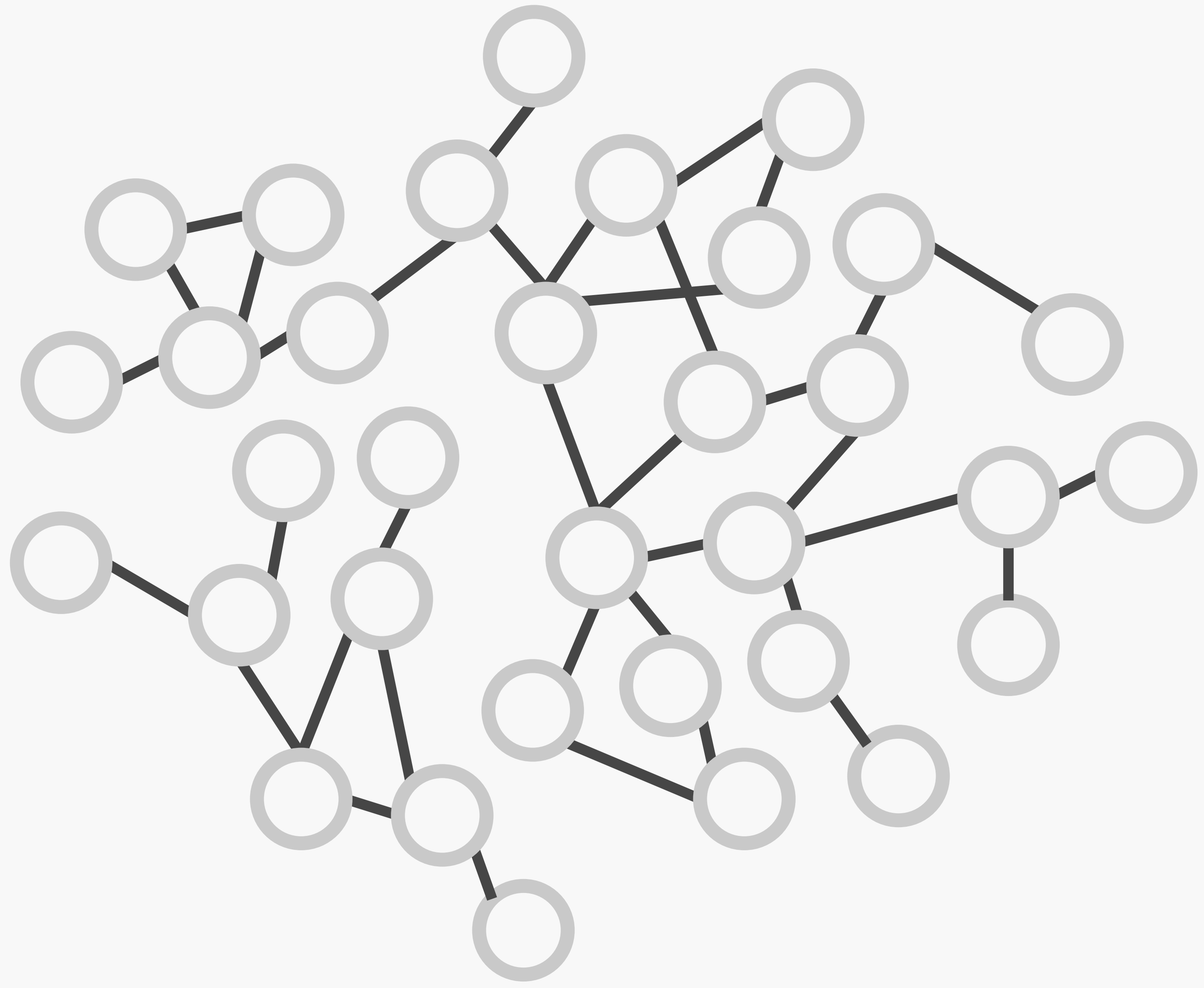
submodularity

$\sigma(\cdot)$ approximation

Reverse Influence Sampling

Random Reverse Reachable Sets

Random Reverse Reachable Sets



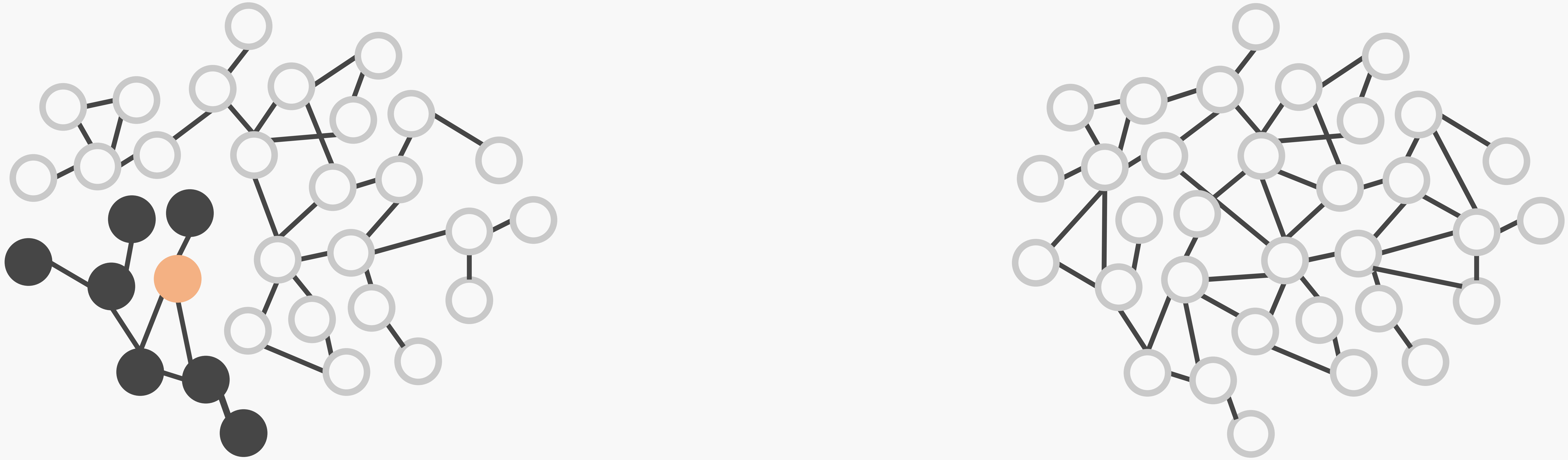
Random Reverse Reachable Sets

Random Reverse Reachable Sets

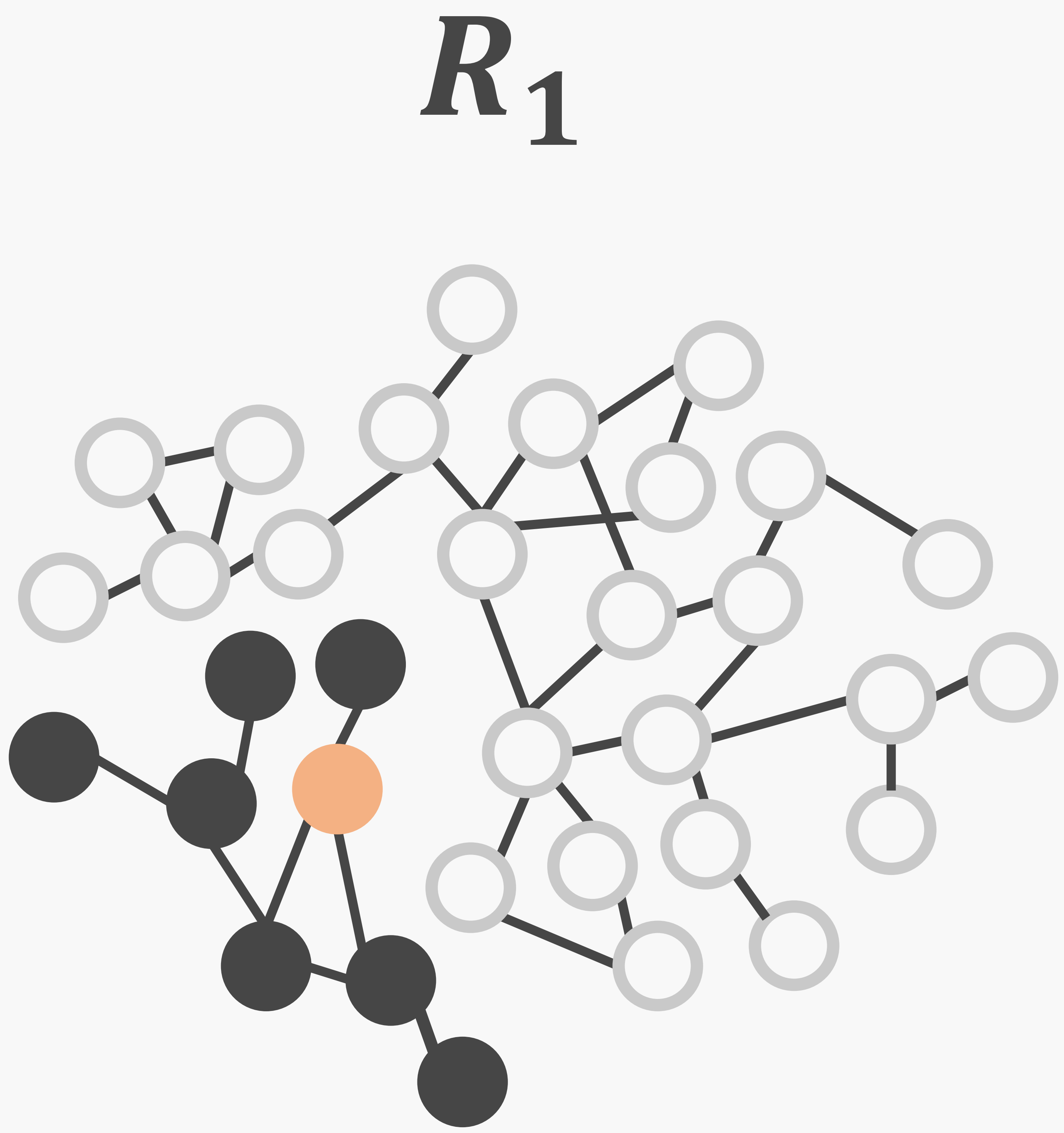
R_1

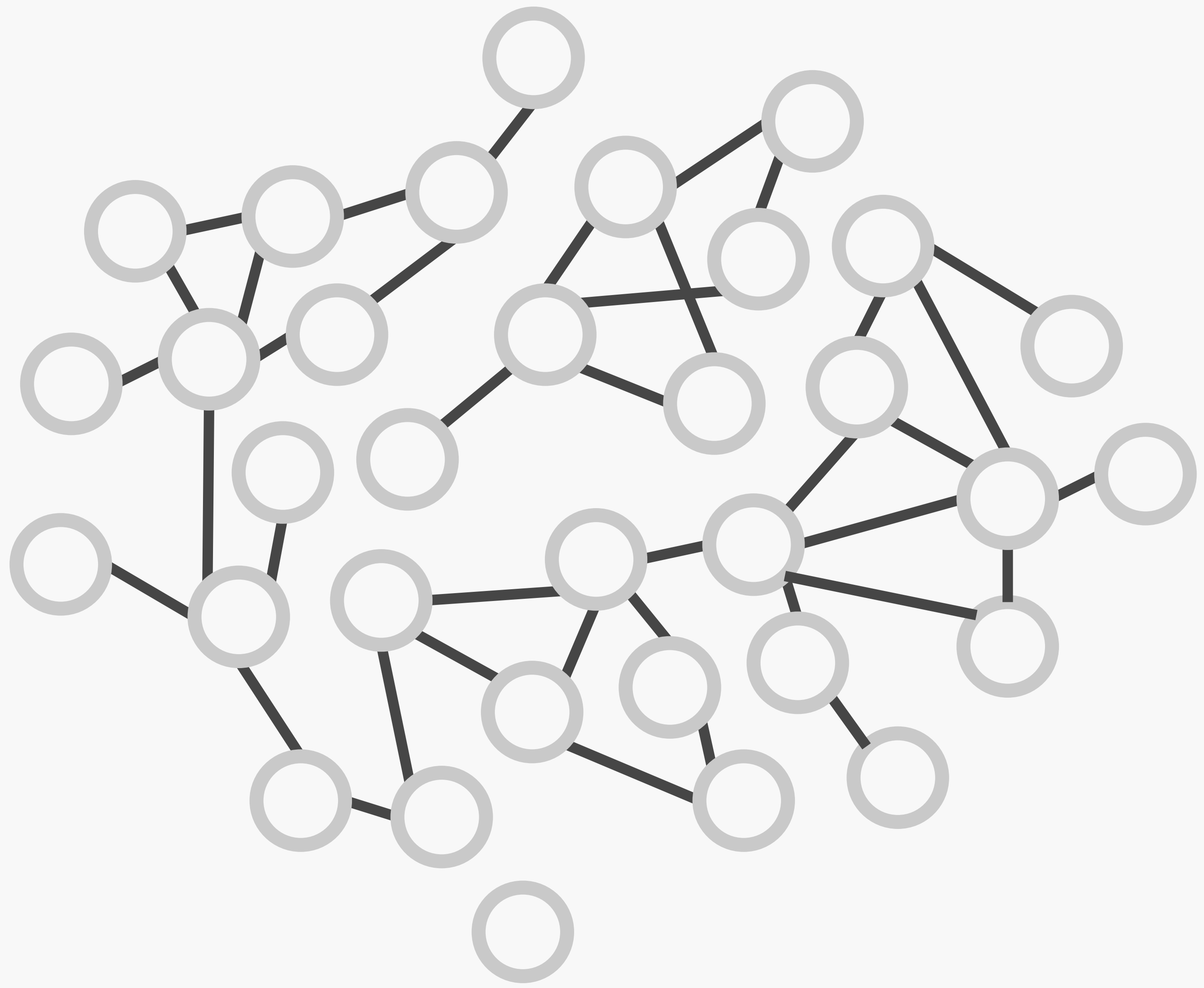
Random Reverse Reachable Sets

R_1



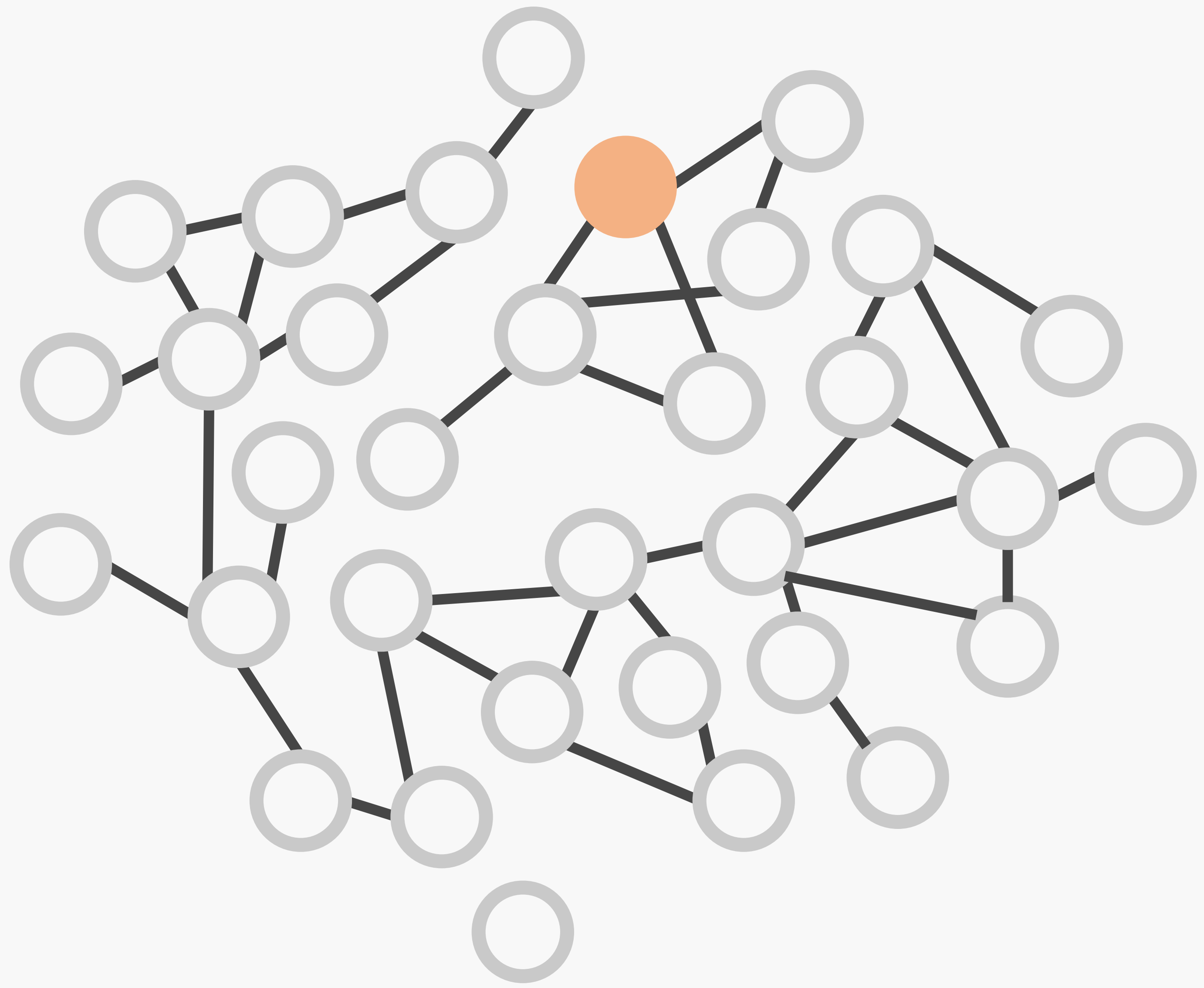
Random Reverse Reachable Sets





Random Reverse Reachable Sets

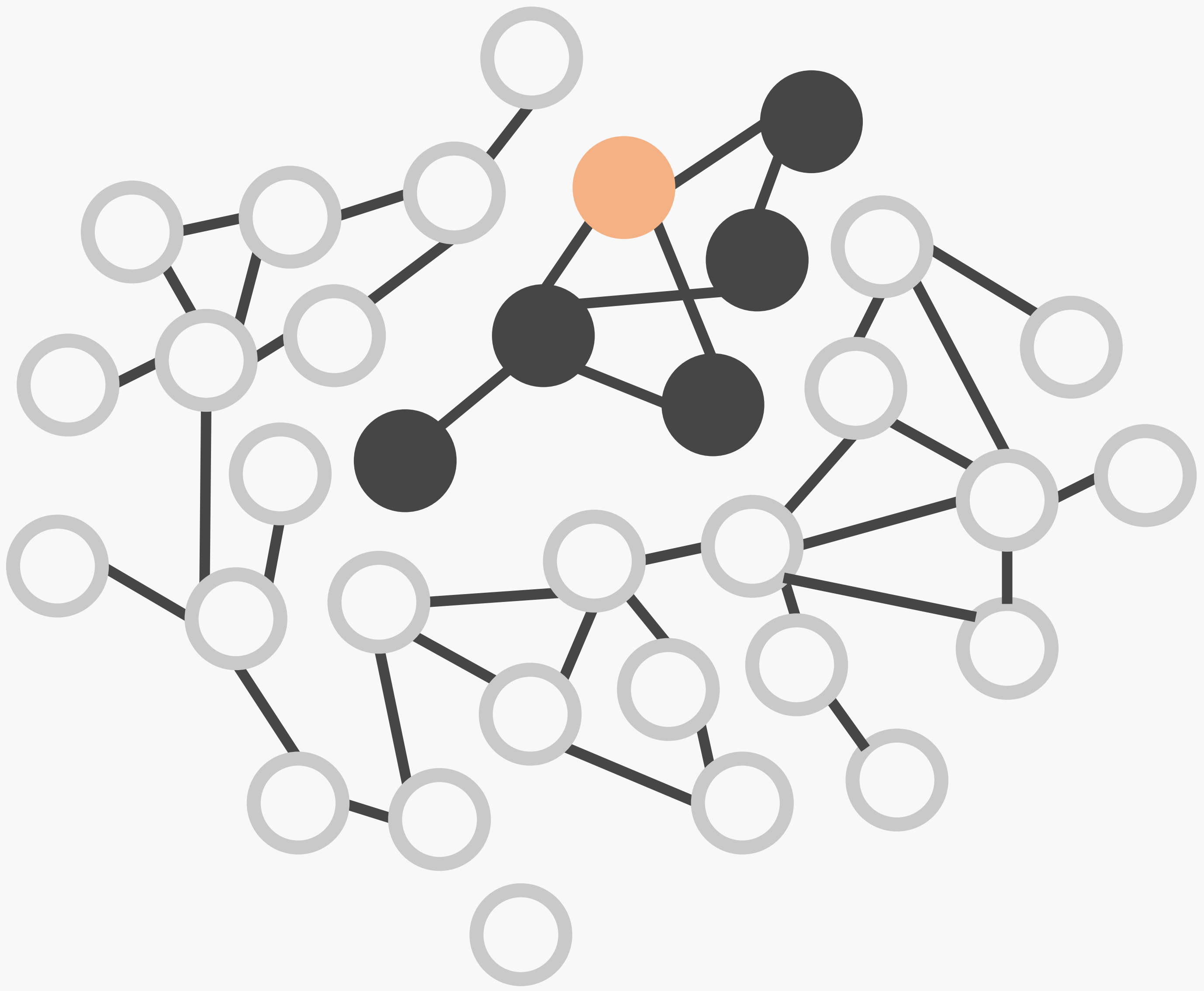
R_1



Random Reverse Reachable Sets

R_1

R_2



Reverse Influence Sampling

- Large Collection of RRR sets: $\mathfrak{R} = \{R_1, \dots, R_\theta\}$
- Fraction of RRR sets in $\mathfrak{R}$ covered by S: $\mathcal{F}_{\mathfrak{R}}(\mathrm{S})$
- **Lemma 1:**

$$\boldsymbol{\sigma}(\boldsymbol{S}) = \boldsymbol{n} \cdot \mathbb{E}(\mathcal{F}_{\mathfrak{R}}(\mathrm{S}))$$

Reverse Influence Sampling

- Large Collection of RRR sets: $\mathfrak{R} = \{R_1, \dots, R_\theta\}$
- Fraction of RRR sets in $\mathfrak{R}$ covered by S: $\mathcal{F}_{\mathfrak{R}}(\mathrm{S})$
- **Lemma 1:**

$$\boldsymbol{\sigma}(\boldsymbol{S}) = \boldsymbol{n} \cdot \mathbb{E}(\mathcal{F}_{\mathfrak{R}}(\mathrm{S}))$$

- **Lemma 2:**

If $\theta > (8 + 2\epsilon)n \frac{l \log n + \log\binom{n}{k} + \log 2}{OPT \cdot \epsilon^2}$ then

$$\left|\boldsymbol{\sigma}(\boldsymbol{S}) - \boldsymbol{n} \cdot \mathcal{F}_{\mathfrak{R}}(\mathrm{S})\right| < \frac{\epsilon}{2} OPT$$

RIS Approximation

Step 1: Generate many RRRSs

$$\mathfrak{R} = \{R_1, \dots, R_\theta\}$$

Step 2: Greedy Maximum Coverage

for $\mathbf{1}: \boldsymbol{k}$

for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

if $\mathcal{F}_{\mathfrak{R}}(\mathrm{S} \cup \boldsymbol{v}) > \mathbf{max}$

$\mathbf{max}, \boldsymbol{v}^* = \mathcal{F}_{\mathfrak{R}}(\mathrm{S} \cup \boldsymbol{v}), \boldsymbol{v}$

end if

end for

$\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

RIS Approximation

Step 1: Generate many RRRSs

$$\mathfrak{R} = \{R_1, \dots, R_\theta\}$$

Step 2: Greedy Maximum Coverage

for $\mathbf{1}:\boldsymbol{k}$

for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

if $\mathcal{F}_{\mathfrak{R}}(S \cup \boldsymbol{v}) > \mathbf{max}$

$\mathbf{max}, \boldsymbol{v}^* = \mathcal{F}_{\mathfrak{R}}(S \cup \boldsymbol{v}), \boldsymbol{v}$

end if

end for

$\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

$$\boldsymbol{O}(\boldsymbol{k} \cdot \boldsymbol{n} + \boldsymbol{\theta})$$

RIS Approximation

Step 1: Generate many RRRSs

$$\mathfrak{R} = \{R_1, \dots, R_\theta\}$$

Step 2: Greedy Maximum Coverage

for $\mathbf{1}: \boldsymbol{k}$

  for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

    if $\mathcal{F}_{\mathfrak{R}}(\mathrm{S} \cup \boldsymbol{v}) > \mathbf{max}$

      $\mathbf{max}, v^* = \mathcal{F}_{\mathfrak{R}}(\mathrm{S} \cup \boldsymbol{v}), \boldsymbol{v}$

    end if

  end for

  $\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

$$\boldsymbol{O}(\boldsymbol{k} \cdot \boldsymbol{n} + \boldsymbol{\theta})$$

$\mathcal{F}_{\mathfrak{R}}(\cdot)$ sub-modular!

RIS Approximation

Step 1: Generate many RRRSs

$$\mathfrak{R} = \{R_1, \ldots, R_\theta\}$$

Step 2: Greedy Maximum Coverage

for $\mathbf{1}: \boldsymbol{k}$

for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

if $\mathcal{F}_{\mathfrak{R}}(S \cup v) > \mathbf{max}$

$\mathbf{max}, v^* = \mathcal{F}_{\mathfrak{R}}(S \cup v), v$

end if

end for

$\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

$$\boldsymbol{O}(\boldsymbol{k} \cdot \boldsymbol{n} + \boldsymbol{\theta})$$

$\mathcal{F}_{\mathfrak{R}}(\cdot)$ sub-modular!

$$\sigma(S_{RIS}) \geq \left(1 - \frac{1}{e} - \epsilon\right) \sigma(S_{OPT})$$

RIS - Exact

RIS-Exact

Step 1: Generate many RRRSs

$$\mathfrak{R} = \{R_1, \dots, R_\theta\}$$

Step 2: Greedy Maximum Coverage

for $\mathbf{1}: \boldsymbol{k}$

for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

if $\mathcal{F}_{\mathfrak{R}}(\mathrm{S} \cup \boldsymbol{v}) > \mathbf{max}$

$\mathbf{max}, v^* = \mathcal{F}_{\mathfrak{R}}(\mathrm{S} \cup \boldsymbol{v}), \boldsymbol{v}$

end if

end for

$\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

RIS-Exact

Step 1: Generate many RRRSs

$$\Re = \{R_1, \ldots, R_\theta\}$$

Step 2: Greedy Maximum Coverage

for $\mathbf{1}:\boldsymbol{k}$

  for $\boldsymbol{v}$ in $\boldsymbol{V} \setminus \boldsymbol{S}$

    if $\mathcal{F}_{\Re}(\mathrm{S} \cup \boldsymbol{v}) > \mathbf{max}$

      $\mathbf{max}, \boldsymbol{v}^* = \mathcal{F}_{\Re}(\mathrm{S} \cup \boldsymbol{v}), \boldsymbol{v}$

    end if

  end for

  $\boldsymbol{S} = \boldsymbol{S} \cup \boldsymbol{v}^*$

end for

Step 1: Generate many RRRSs

$$\Re = \{R_1, \ldots, R_\theta\}$$

Step 2: Maximum Coverage

for $\boldsymbol{s}$ in $\binom{\boldsymbol{n}}{\boldsymbol{k}}$ candidates

  if $\mathcal{F}_{\Re}(\mathrm{s}) > \mathbf{max}$

    $\mathbf{max}, \boldsymbol{S} = \mathcal{F}_{\Re}(\mathrm{s}), \boldsymbol{v}$

  end if

end for

RIS-Exact Proof Sketch

$$\sigma(S_{RIS-E})$$

RIS-Exact Proof Sketch

$$\sigma(S_{RIS-E}) \geq n\,\mathcal{F}(S_{RIS-E}) - \frac{\epsilon}{2}\mathrm{OPT}$$

[Lemma]

RIS-Exact Proof Sketch

$$\sigma(S_{RIS-E}) \geq n\,\mathcal{F}(S_{RIS-E}) - \frac{\epsilon}{2}\mathrm{OPT}$$ [Lemma]

$$\geq n\,\mathcal{F}(S_{OPT}) - \frac{\epsilon}{2}\mathrm{OPT}$$ [definition of S_{RIS-E}]

RIS-Exact Proof Sketch

$$\sigma(S_{RIS-E}) \geq n\,\mathcal{F}(S_{RIS-E}) - \frac{\epsilon}{2}\mathrm{OPT}$$ [Lemma]

$$\geq n\,\mathcal{F}(S_{OPT}) - \frac{\epsilon}{2}\mathrm{OPT}$$ [definition of S_{RIS-E}]

$$\geq \boldsymbol{\sigma}(S_{OPT}) - \frac{\epsilon}{2}\mathrm{OPT} - \frac{\epsilon}{2}\mathrm{OPT}$$ [Lemma]

RIS-Exact Proof Sketch

$$
\begin{aligned}
\sigma(S_{RIS-E}) &\geq n\,\mathcal{F}(S_{RIS-E}) - \frac{\epsilon}{2}\mathrm{OPT} && \text{[Lemma]}\\
&\geq n\,\mathcal{F}(S_{OPT}) - \frac{\epsilon}{2}\mathrm{OPT} && \text{[definition of } S_{RIS-E}\text{]}\\
&\geq \boldsymbol{\sigma}(S_{OPT}) - \frac{\epsilon}{2}\mathrm{OPT} - \frac{\epsilon}{2}\mathrm{OPT} && \text{[Lemma]}\\
&\geq (1-\epsilon)OPT
\end{aligned}
$$

Experiments

GPU Implementation

- AWS EC2 *Deep Learning Base AMI Linux Version 19.1* Instance - Nividia Tesla K80

GPU Implementation

- AWS EC2 *Deep Learning Base AMI Linux Version 19.1* Instance - Nividia Tesla K80
- Python Numba

GPU Implementation

- AWS EC2 *Deep Learning Base AMI Linux Version 19.1* Instance - Nividia Tesla K80
- Python Numba
- Two distributed objects:

1. $\mathfrak{R} = \{R_1, \ldots, R_\theta\}$ coded as $\theta \times n$ array with $\mathfrak{R}_{ij} = \mathrm{TRUE}$ if node j is in R_i.
2. C coded as $1 \times \binom{n}{k}$ array with C_i representing the number of sets in $\mathfrak{R}$ covered by candidate seed set i.

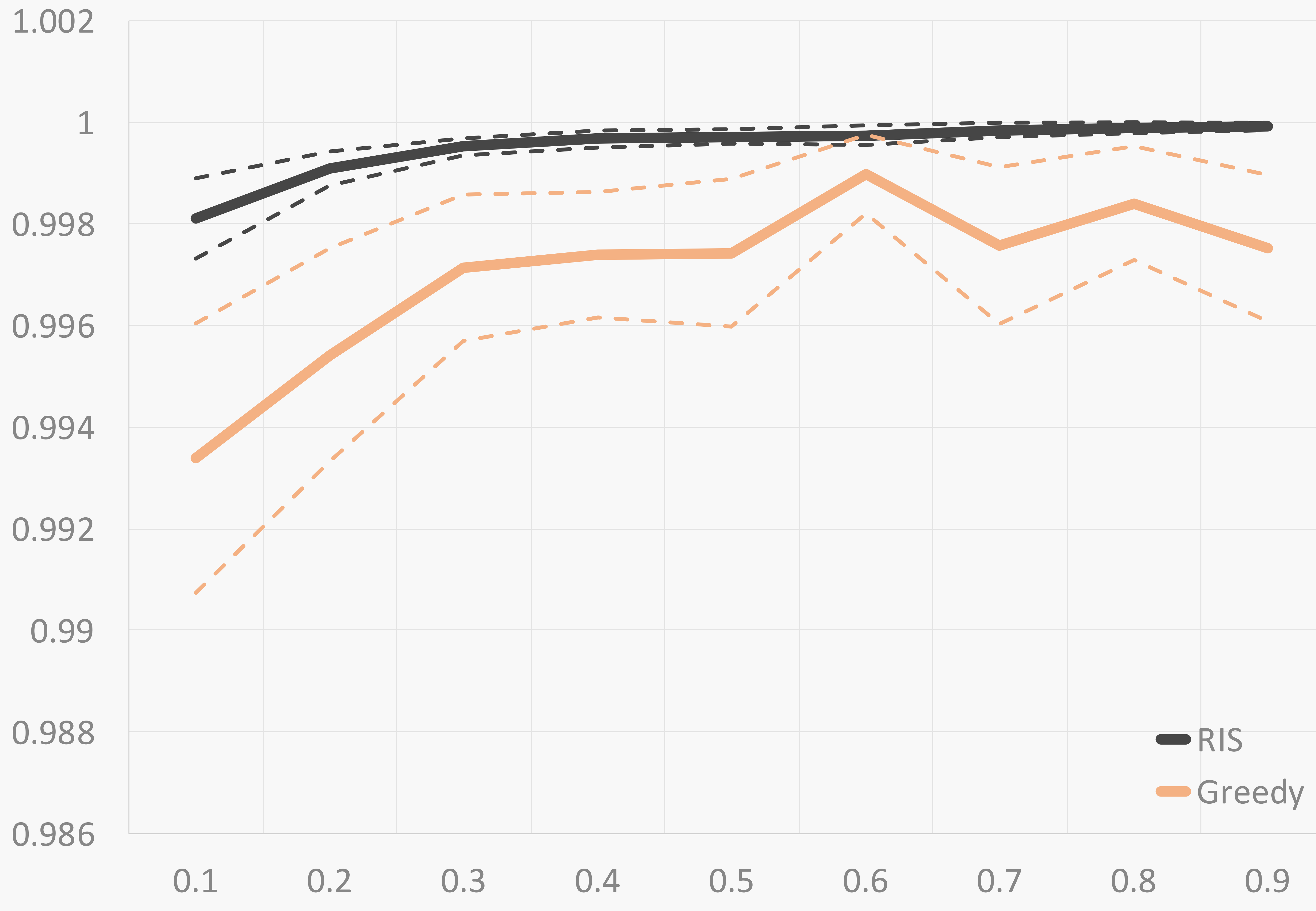
Experiment Parameters

- **3 Network Types**
  - Erdos-Renyi, Watts-Strogatz, Scale-Free
- **29 Parameter Configurations**
  - $q \in [0.1, 0.9]$, $\beta \in [0, 0.9]$, $\gamma \in [1.5, 4]$
- **10 Graph Instances**
- **6-7 Propagation Probabilities**
  - $p \in [0.01, 0.7]$
- **Network Size 100, Seed Set Size 4**
- **1,880 Total Simulations**
  - 180 seconds each
  - 4 days total

Approximations are near optimal

$$\delta = \frac{\sigma(S_{RIS})}{\sigma(S_{RIS-E})}$$

Weak positive relationship between network density and accuracy



Recap

1. RIS-Exact exploits Reverse Influence Sampling to make exact solutions feasible.
2. Top approximation algorithms are almost perfect.
3. Solution accuracy does not depend on network structure.

Thanks

Appendix

References

Kempe, Kleinberg & Tardos (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (p. 137-146). ACM.

Borgs, Brautbar, Chayes & Lucier (2014). Maximizing social influence in nearly optimal time. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms* (p. 946-957). Society for Industrial and Applied Mathematics.

Li, Smith, Dinh & Thai (2017). Why approximate when you can get the exact? optimal targeted viral marketing at scale. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 1 – 9

RIS Proof Sketch

$$\sigma(S_{RIS}) \geq n\,\boldsymbol{\mathcal{F}}(S_{RIS}) - \frac{\epsilon}{2}\mathrm{OPT}$$ [Lemma]

$$\geq \left(1 - \frac{1}{e}\right) n\,\boldsymbol{\mathcal{F}}(\boldsymbol{S}^{*}) - \frac{\epsilon}{2}\mathrm{OPT}$$ [Greedy submodular error]

$$\geq \left(1 - \frac{1}{e}\right) n\,\boldsymbol{\mathcal{F}}(S_{OPT}) - \frac{\epsilon}{2}\mathrm{OPT}$$ [definition of $\boldsymbol{S}^{*}$]

$$\geq \left(1 - \frac{1}{e}\right)\left(\boldsymbol{\sigma}(S_{OPT}) - \frac{\epsilon}{2}\mathrm{OPT}\right) - \frac{\epsilon}{2}\mathrm{OPT}$$ [Lemma]

$$\geq \left(1 - \frac{1}{e} - \epsilon\right) OPT$$