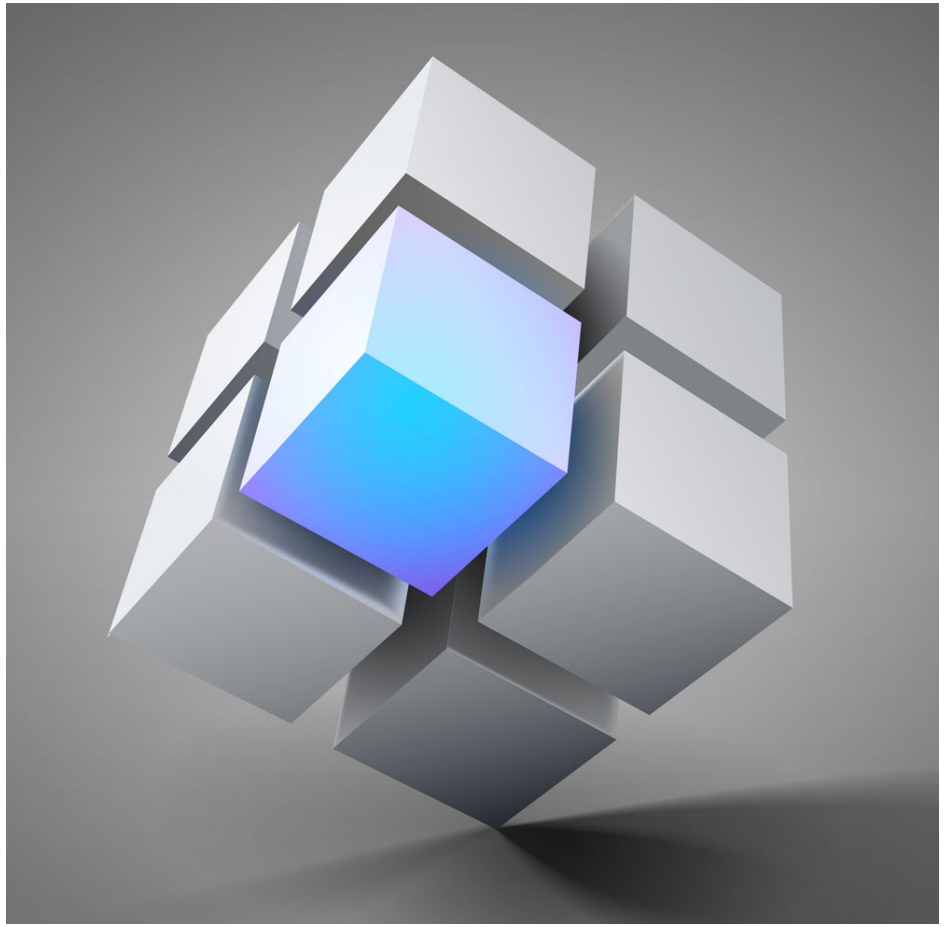
Multi-Round Influence Maximization

Lichao Sun, Weiran Huang, Philip S. Yu, Wei Chen
University of Illinois at Chicago, Microsoft Research, Tsinghua University

Roadmap



1 **Problem**

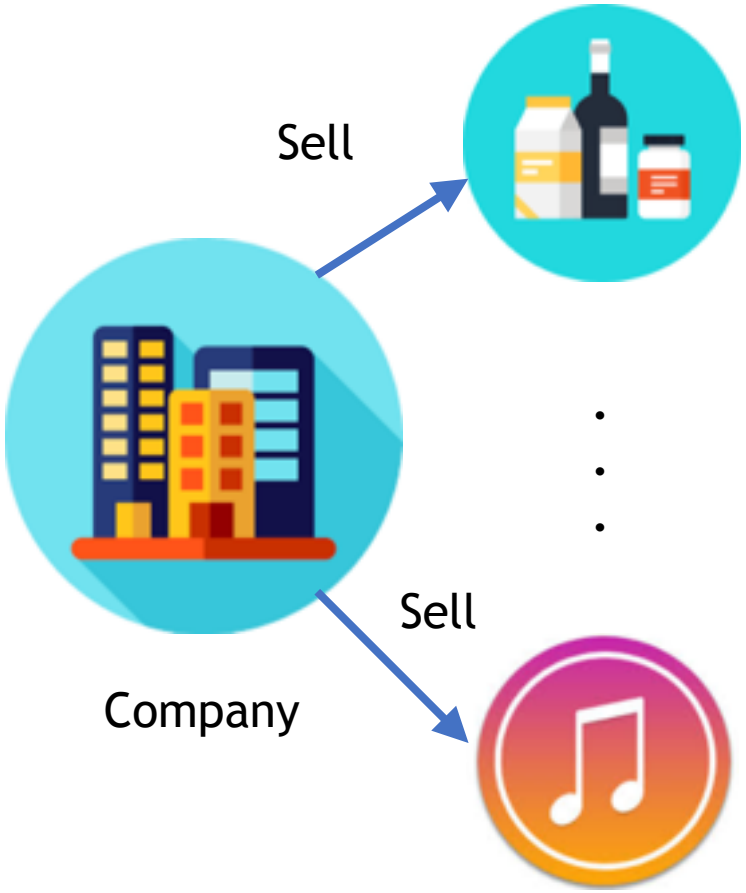
2 Methodology

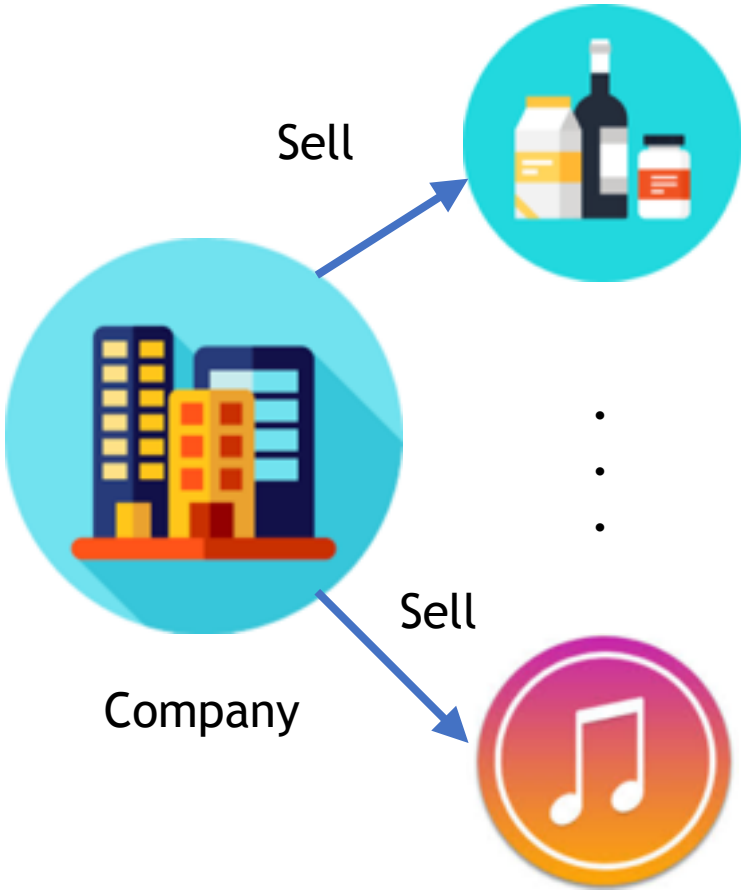
3 Experiments

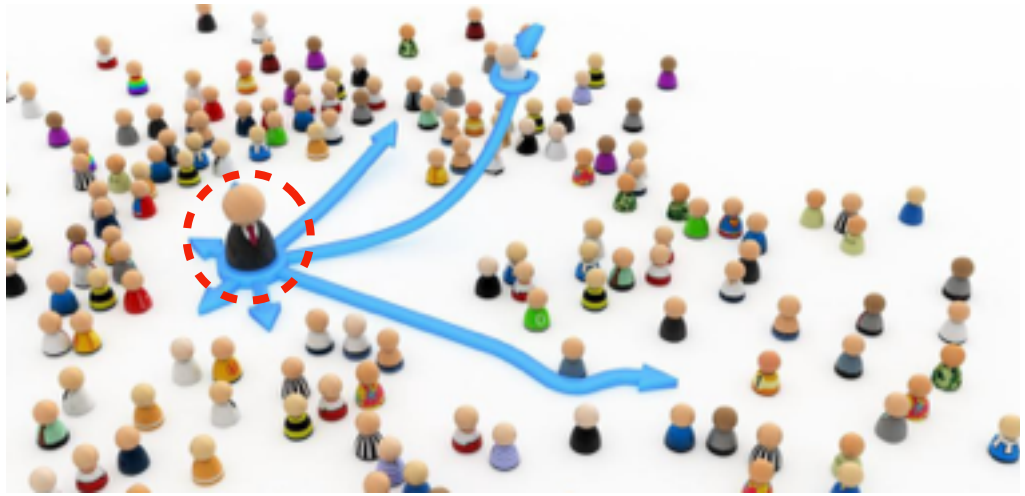
4 Conclusions

Problem Statement

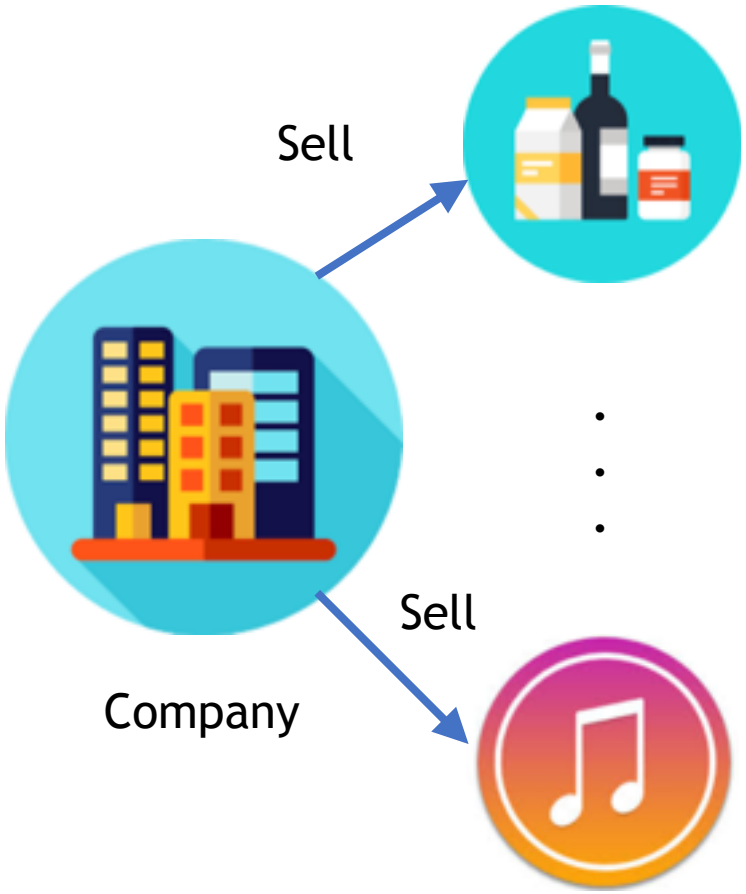
Company

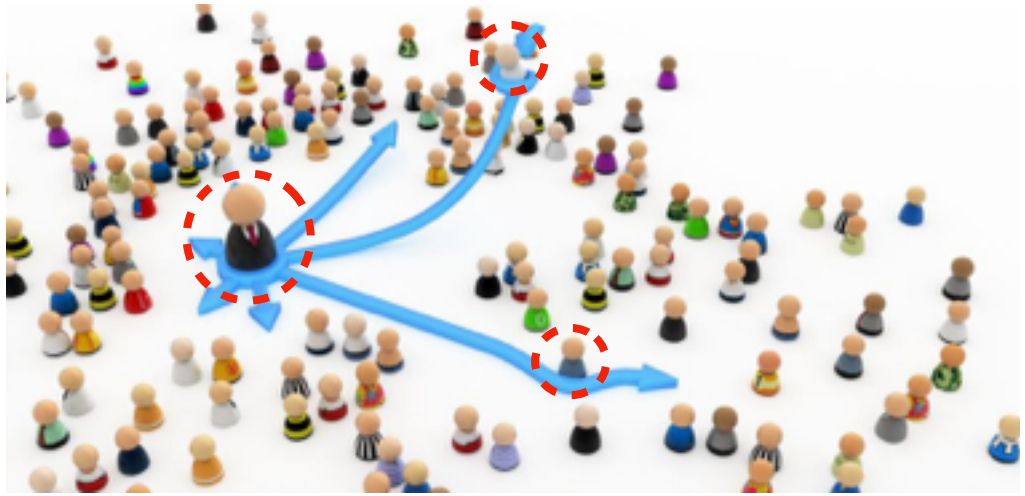

Sell
Sell
Company



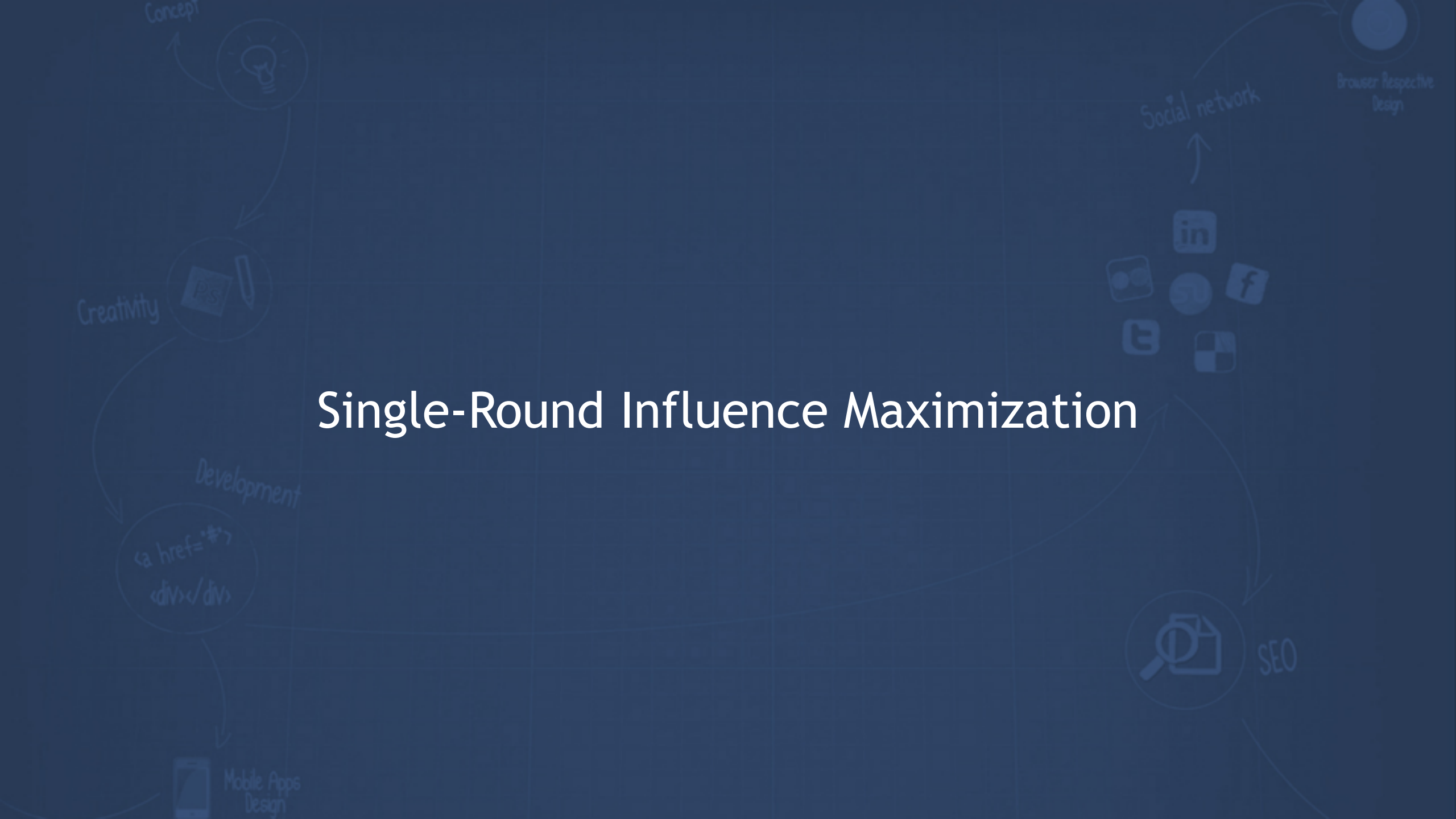


Find Influential member





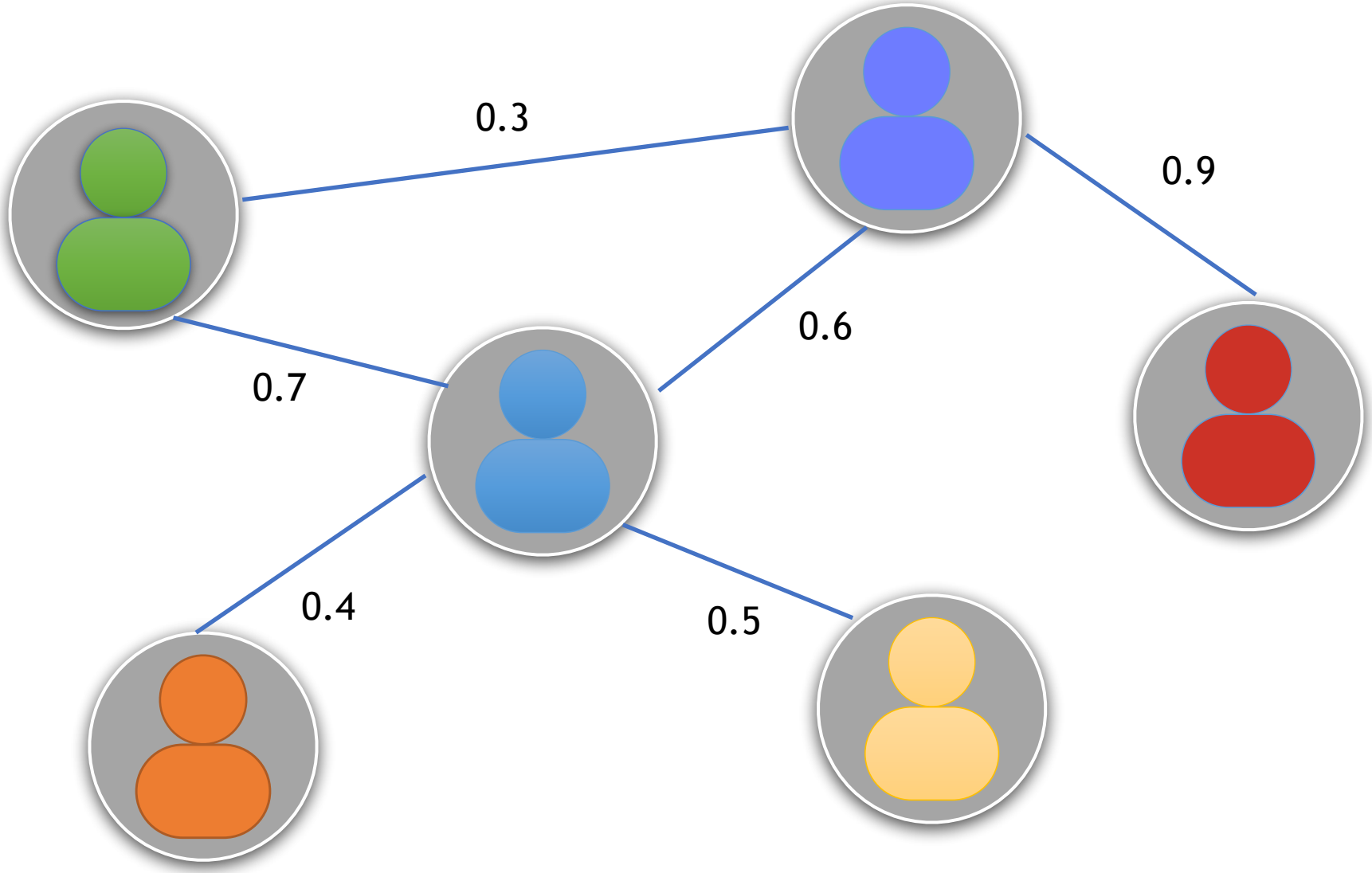
Find a set of Influential members

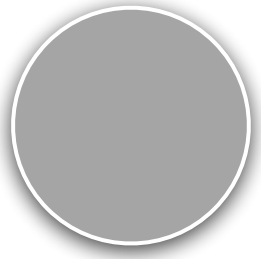


Single-Round Influence Maximization

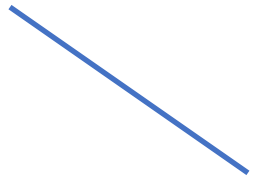
- Influence maximization is the problem of finding a small set of seed nodes in a social network that maximizes the spread of influence under certain influence cascade models. (Kempe, and Kleinberg, 2003)
  - Influence cascade models: (1) Independent cascade model (2) Linear-threshold model (3) Triggering Model
  - Activation probability is known

- Find the optimal solution is NP-hard question (Set cover problem). (Kempe, Kleinberg, and Tardos 2003)
  - Greedy
  - Monto-Carlo simulation


0.3
0.9
0.6
0.7
0.4
0.5



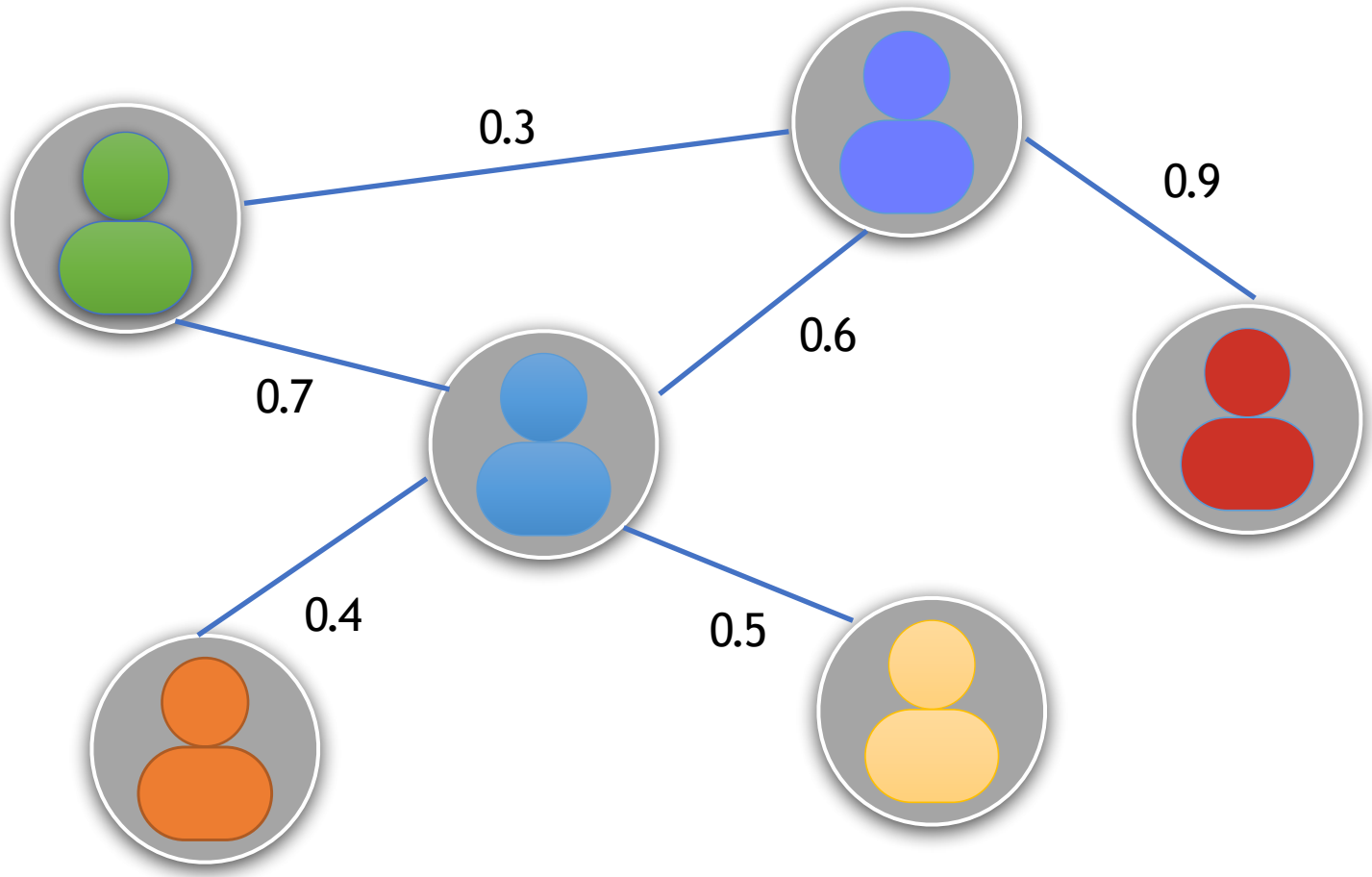
Node

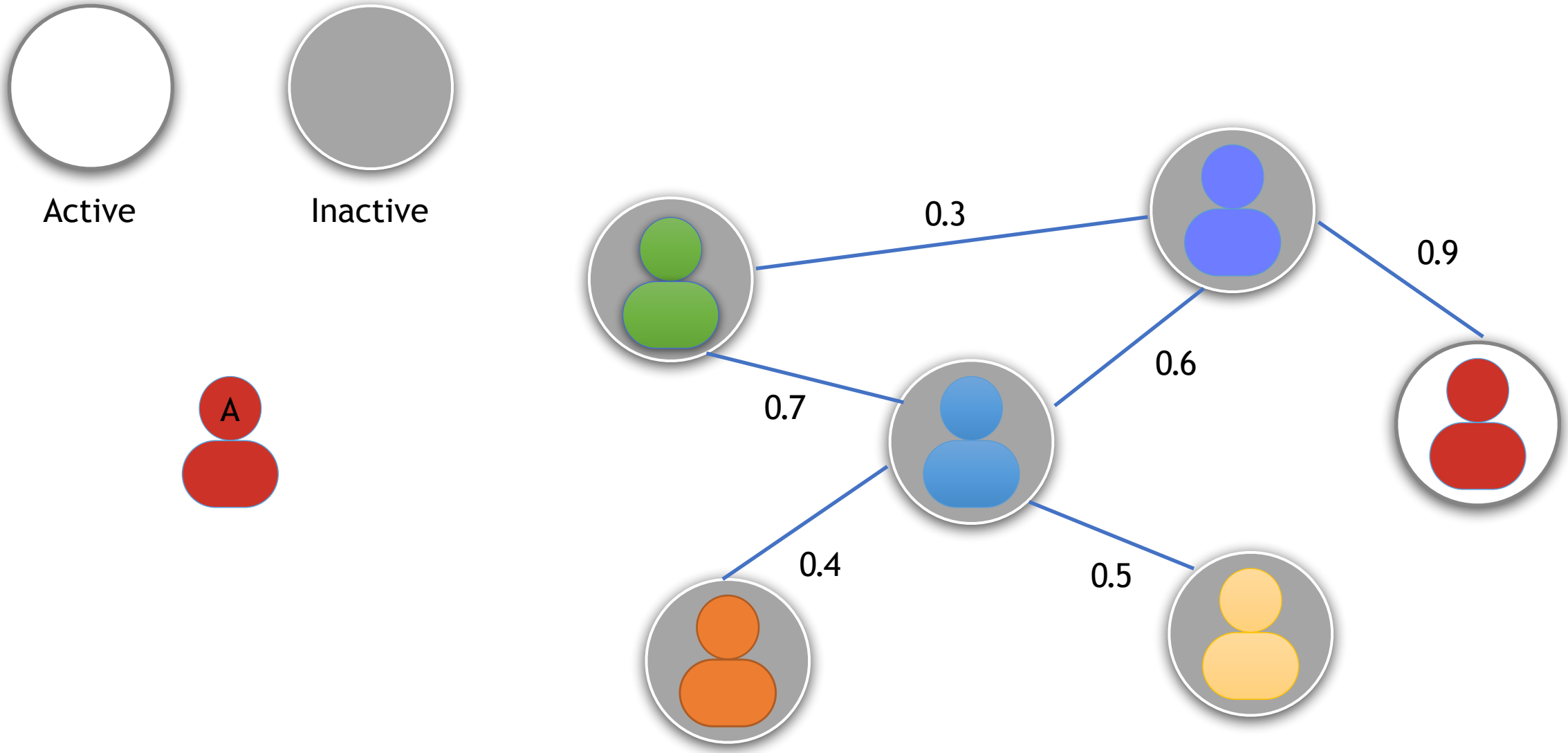


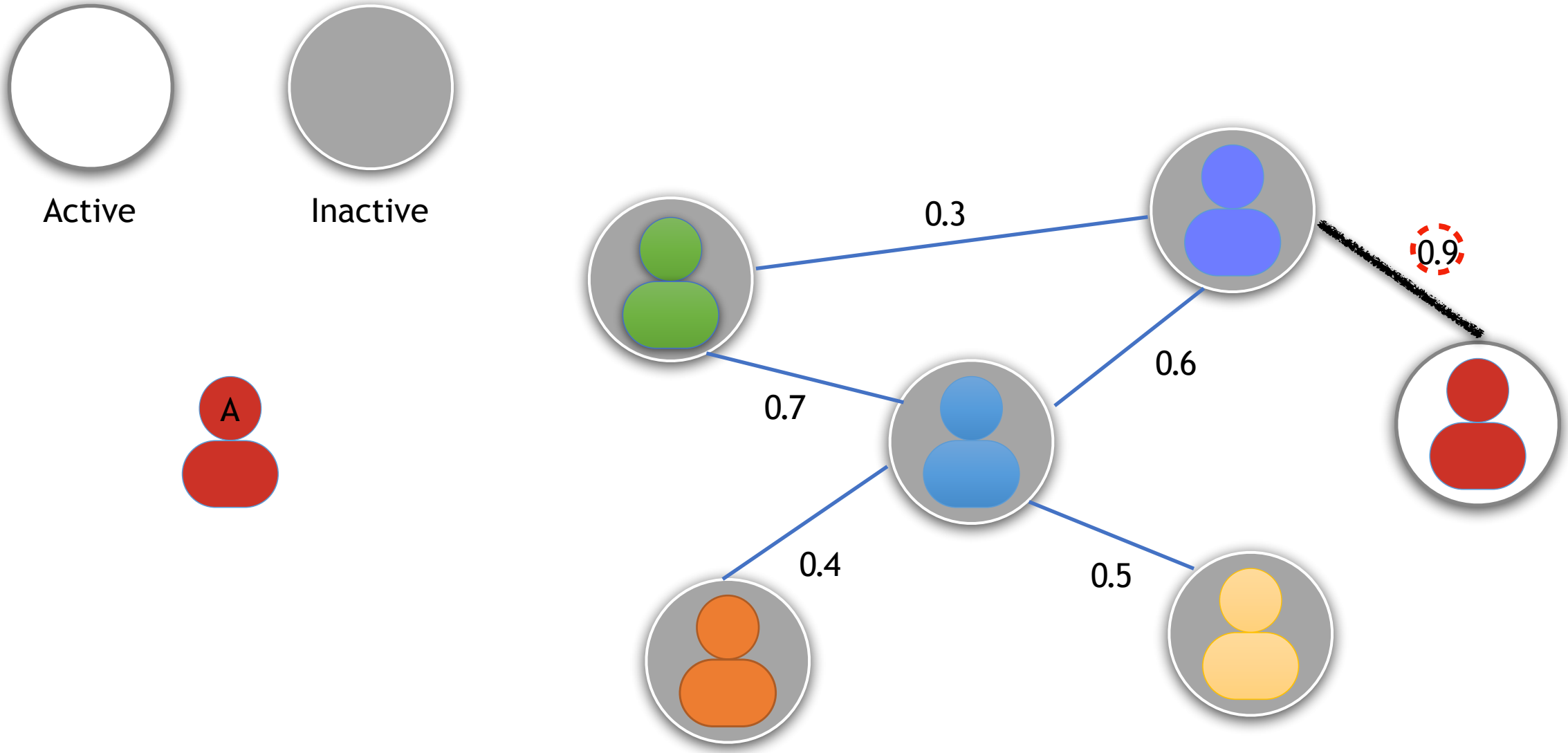
Edge

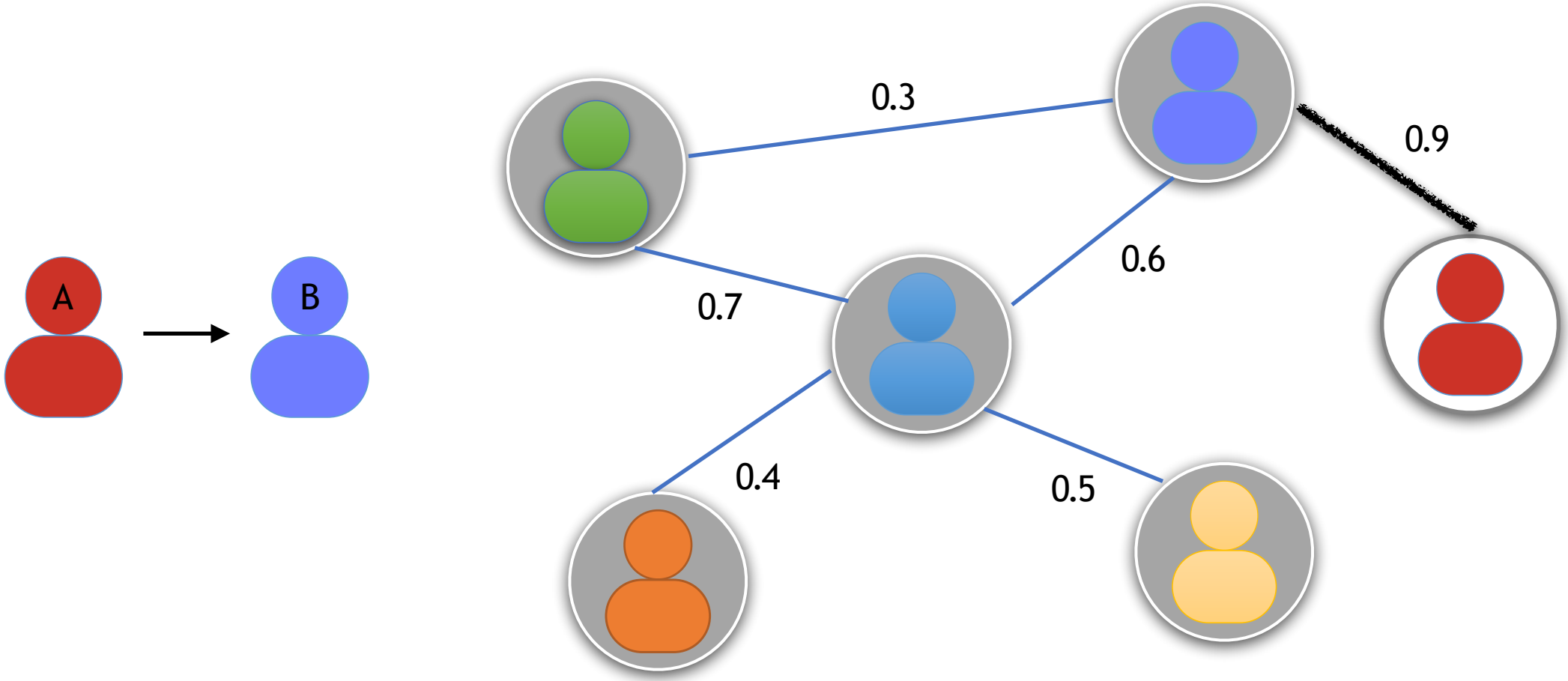
0.9

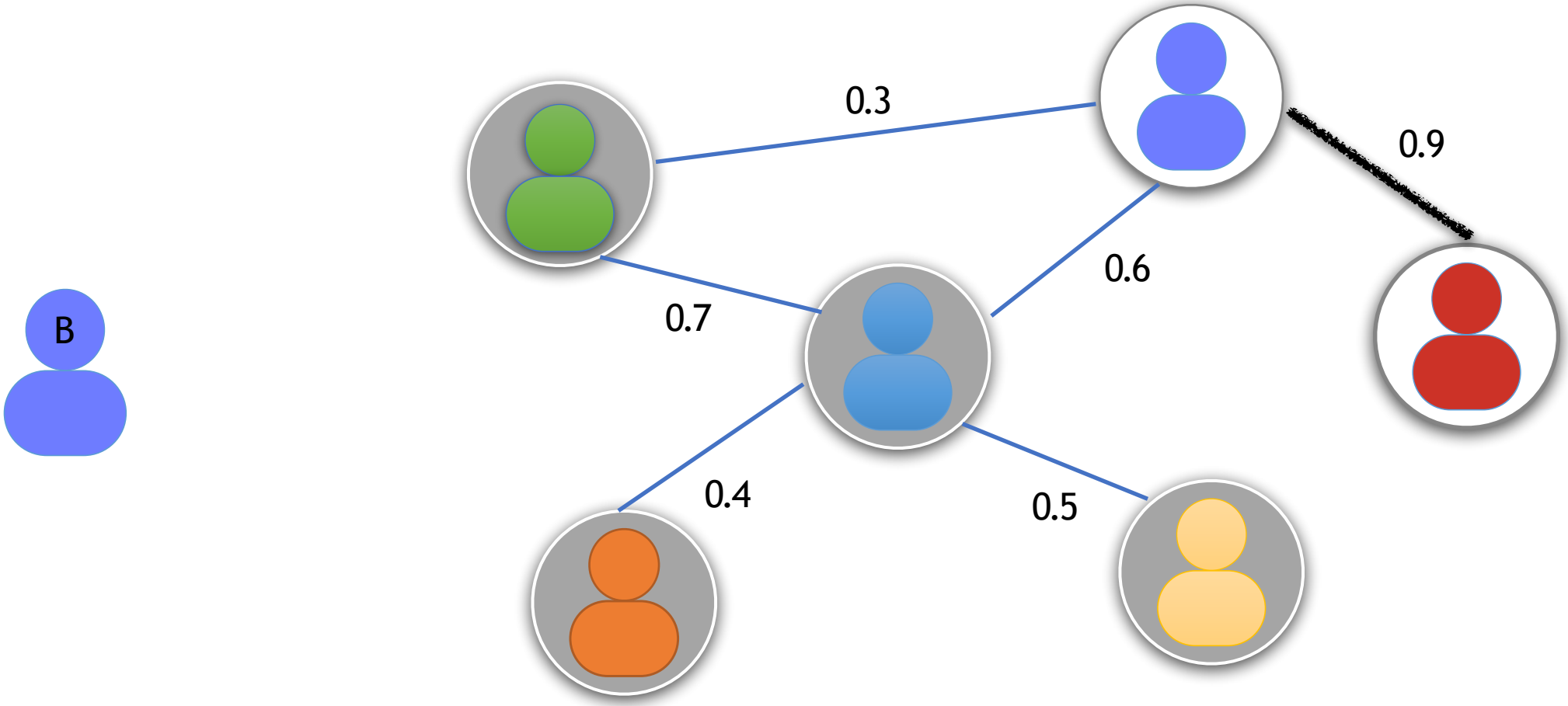
Influential Probability

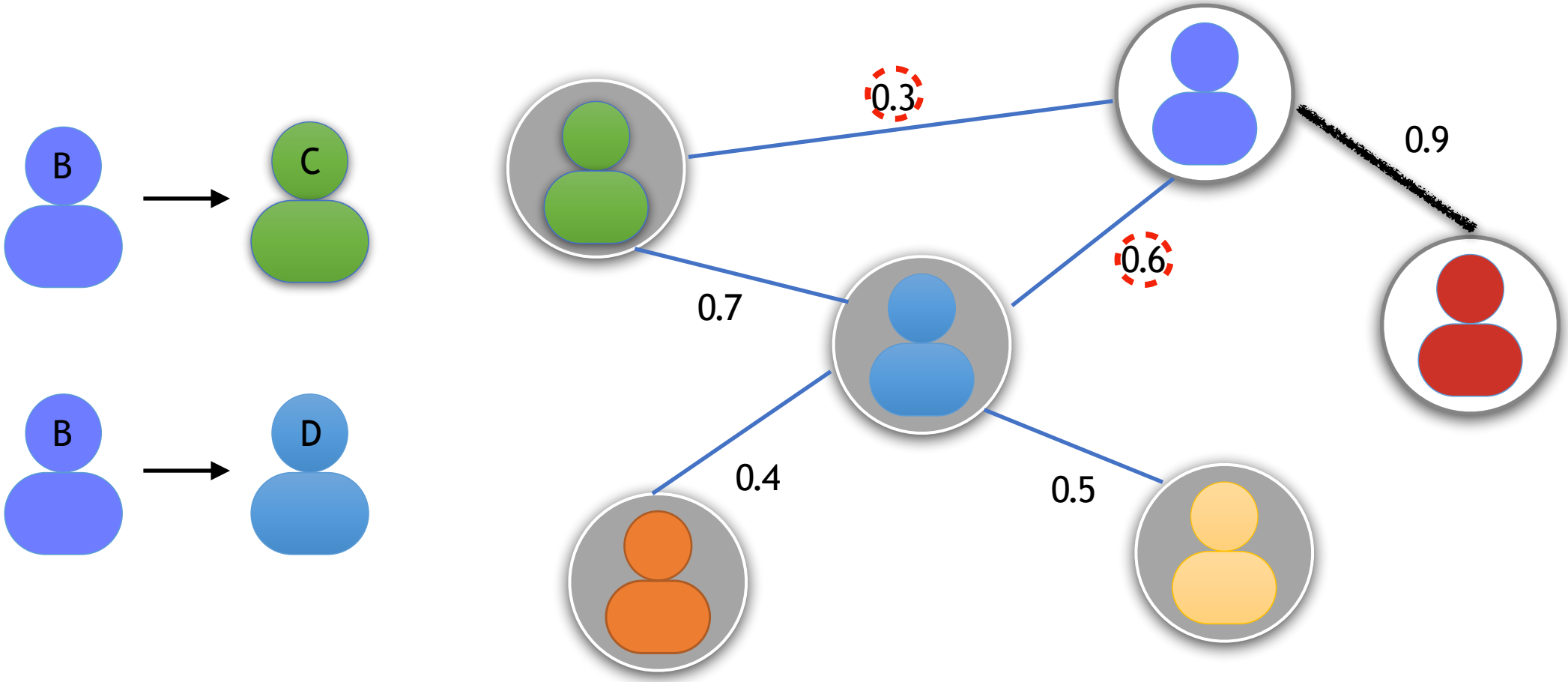


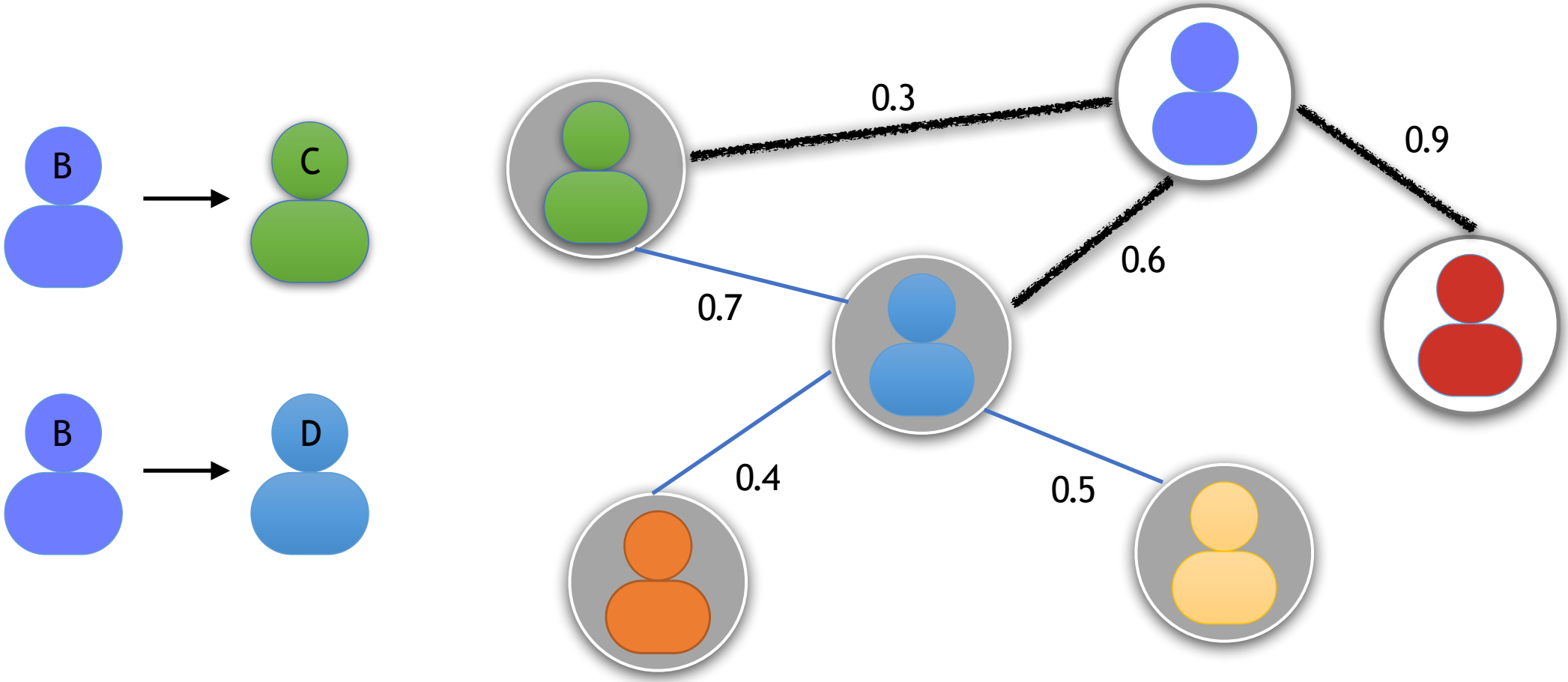

Active
Inactive
A
0.3
0.9
0.6
0.7
0.4
0.5

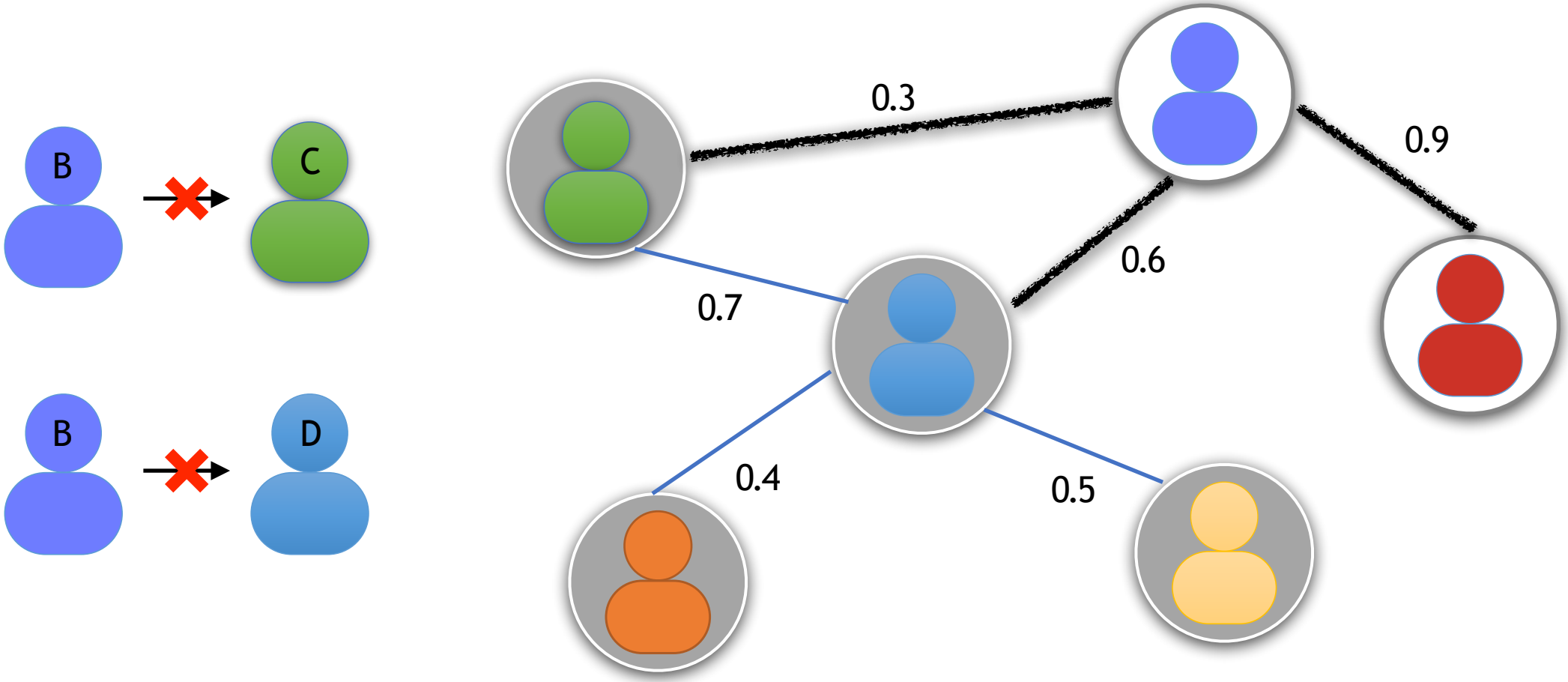

Active
Inactive
A
0.3
0.9
0.6
0.7
0.4
0.5

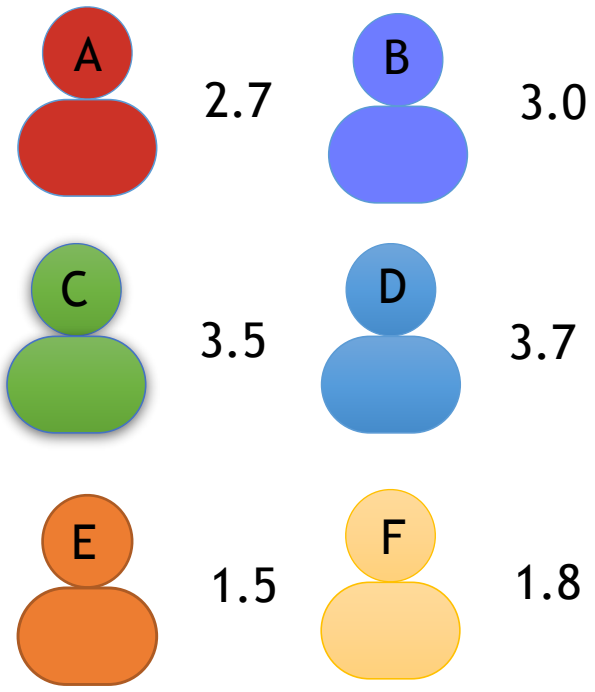

A
B
0.3
0.9
0.7
0.6
0.4
0.5

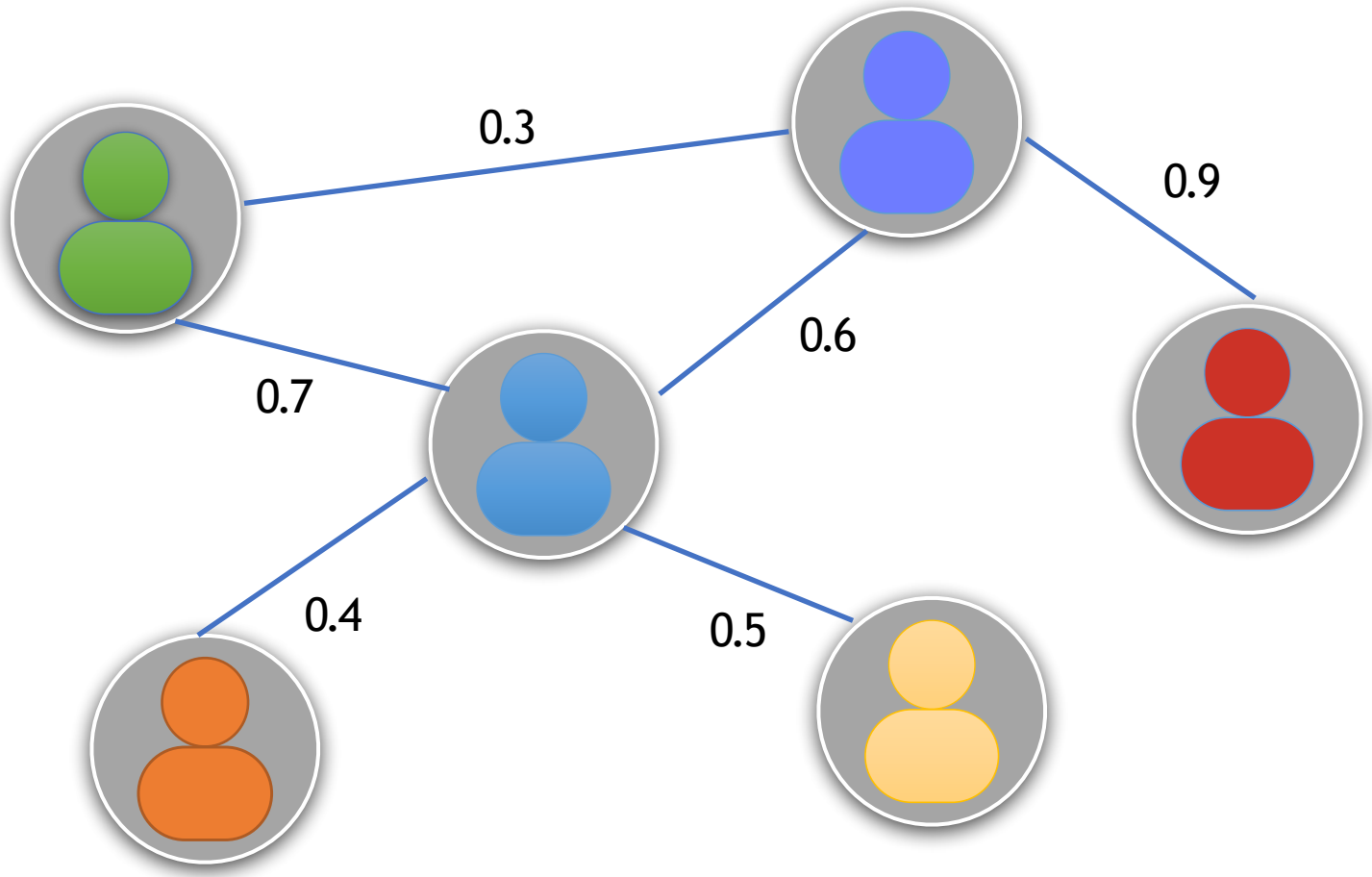

B
0.3
0.9
0.6
0.7
0.4
0.5

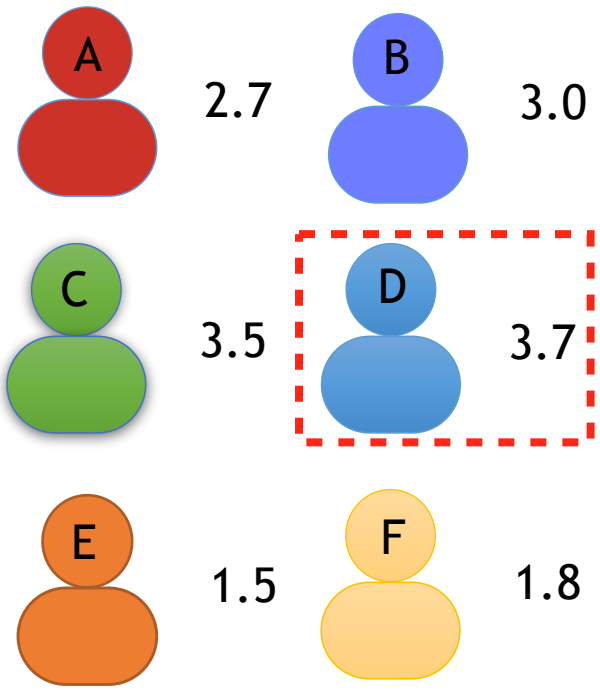

B
C
B
D
0.3
0.9
0.6
0.7
0.4
0.5

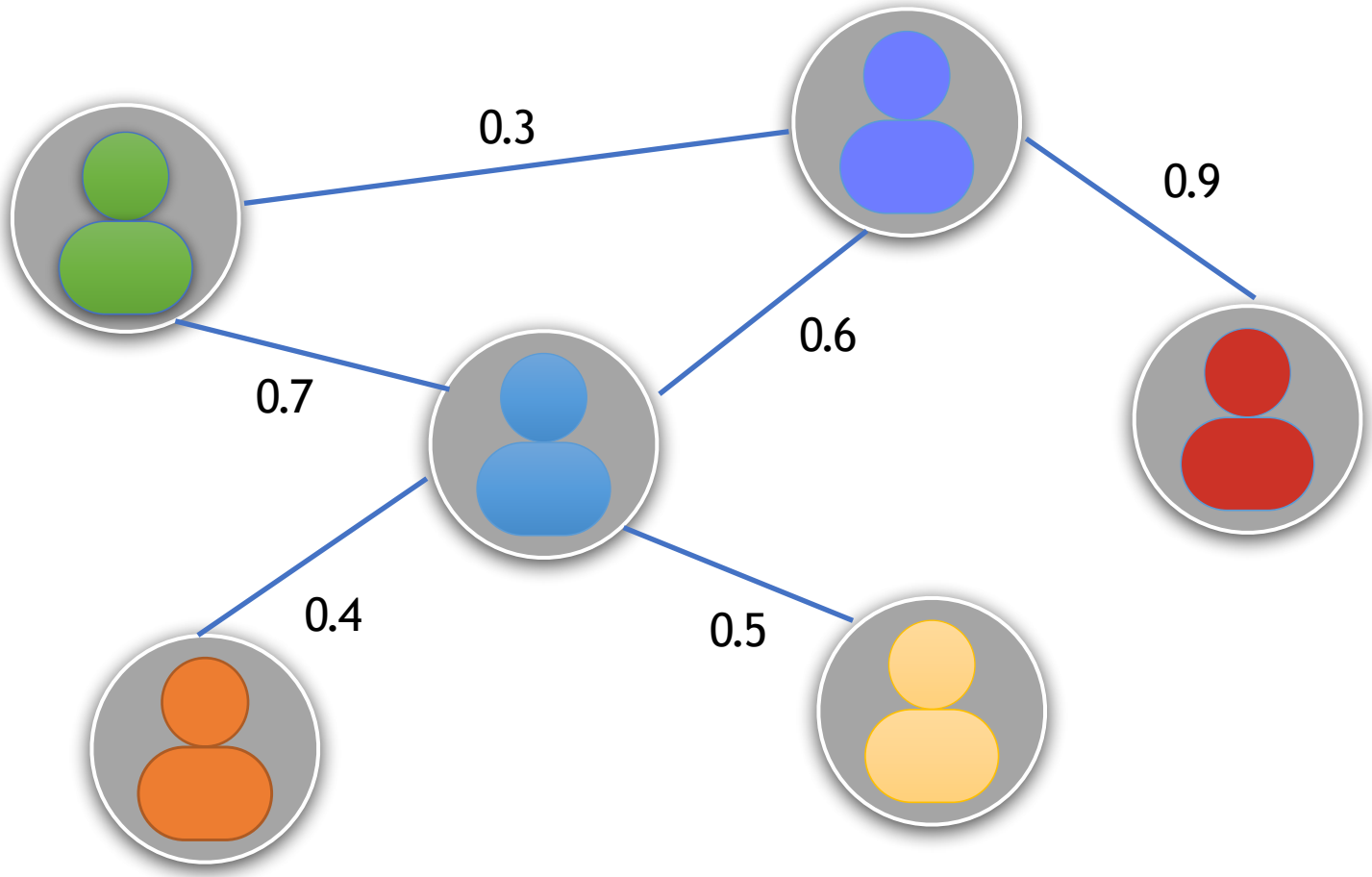

B
C
B
D
0.3
0.9
0.6
0.7
0.4
0.5


B
C
B
D
0.3
0.9
0.6
0.7
0.4
0.5

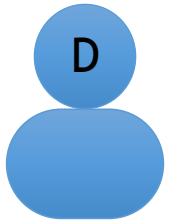

A
2.7
B
3.0
C
3.5
D
3.7
E
1.5
F
1.8

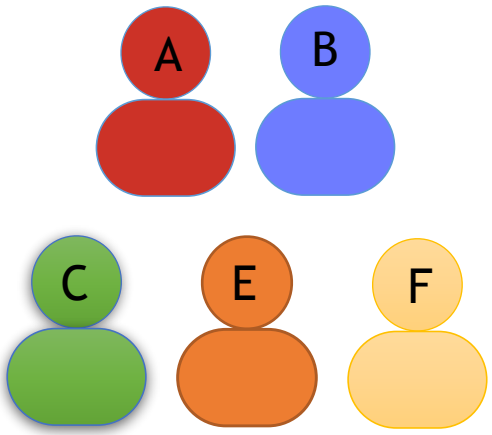

0.3
0.9
0.6
0.7
0.4
0.5


A
2.7
B
3.0
C
3.5
D
3.7
E
1.5
F
1.8

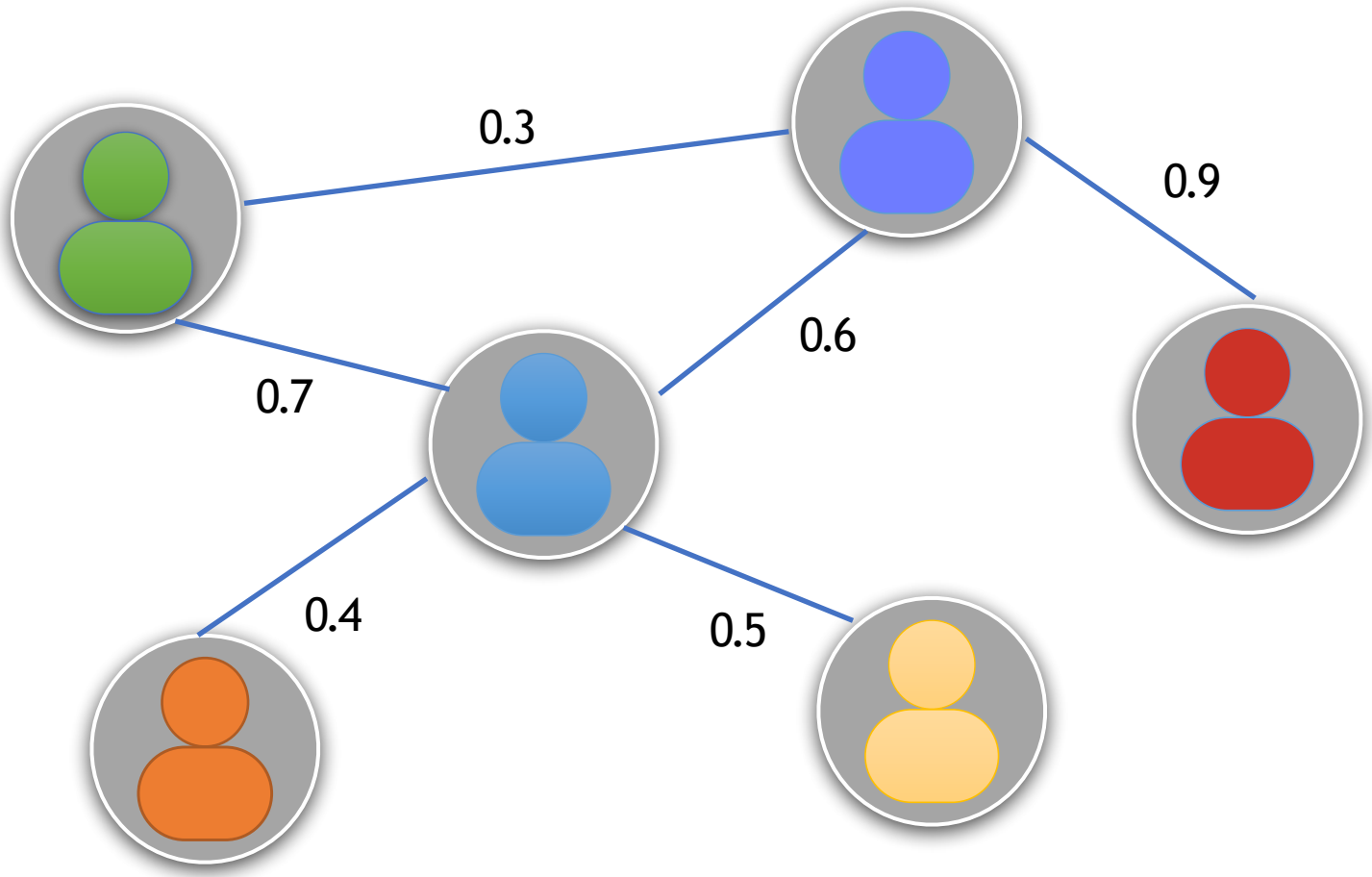

0.3
0.9
0.6
0.7
0.4
0.5

seed set:





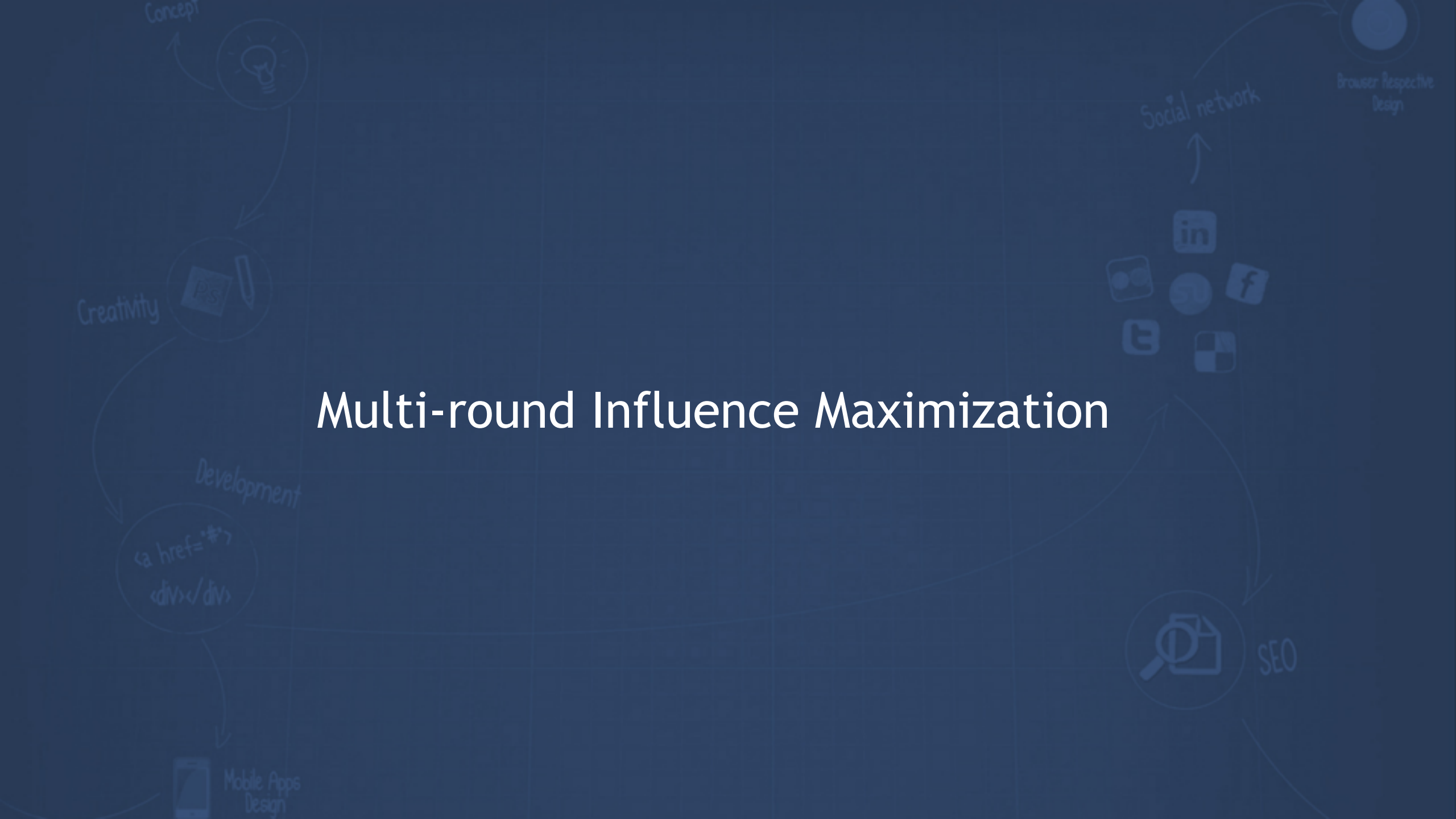
Find the best one which can produce largest network impact with seed set



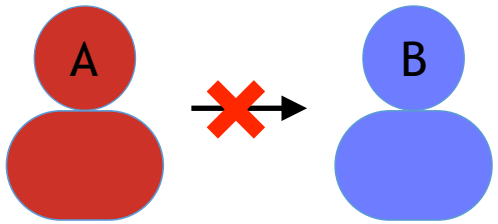
Summary: Single-Round Influence Maximization

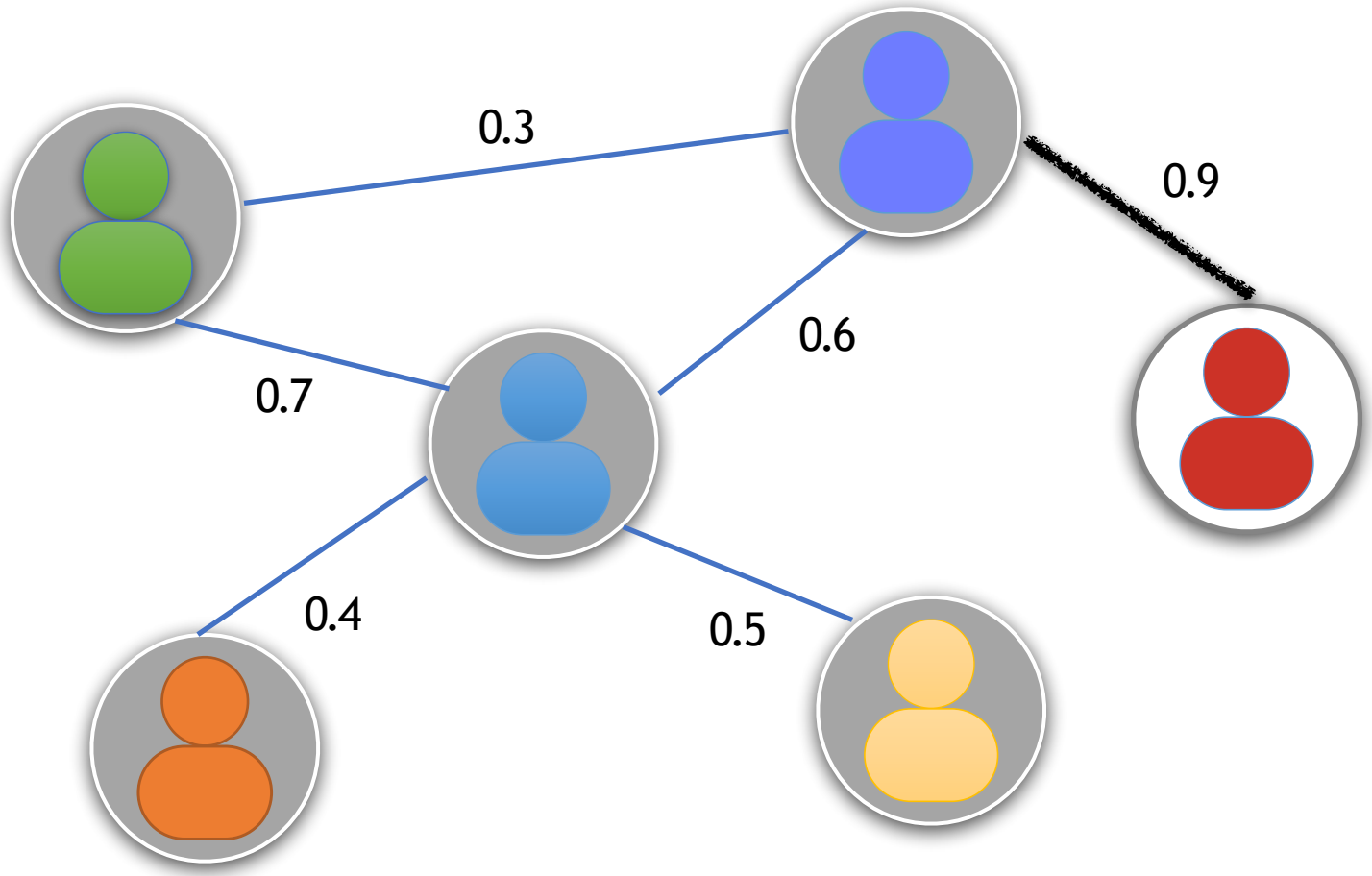
- Given a probability Graph G and budget k
- Greedy algorithm to pick k nodes as a seed set
- Every node can be activated at most once
- Every node can not be duplicated in the seed set

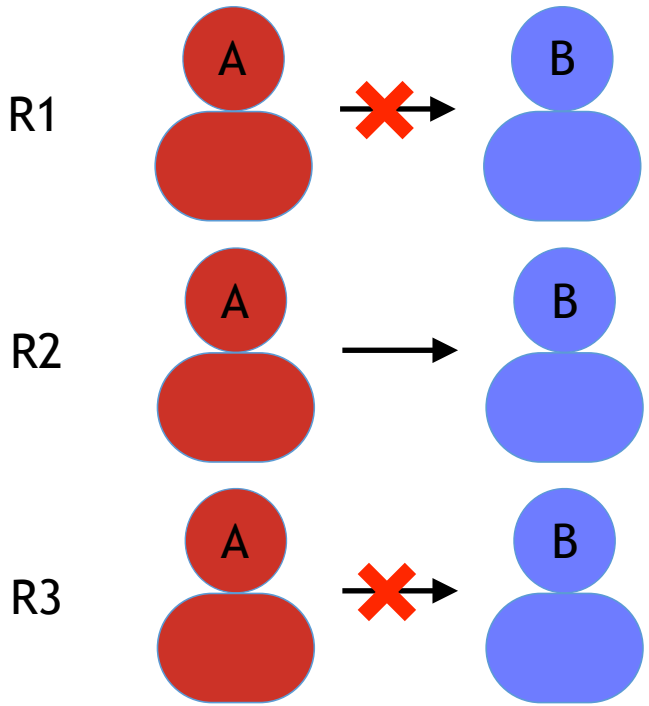
If company wants to do advertisements twice with total 2k budget, how to design the node selection strategy?

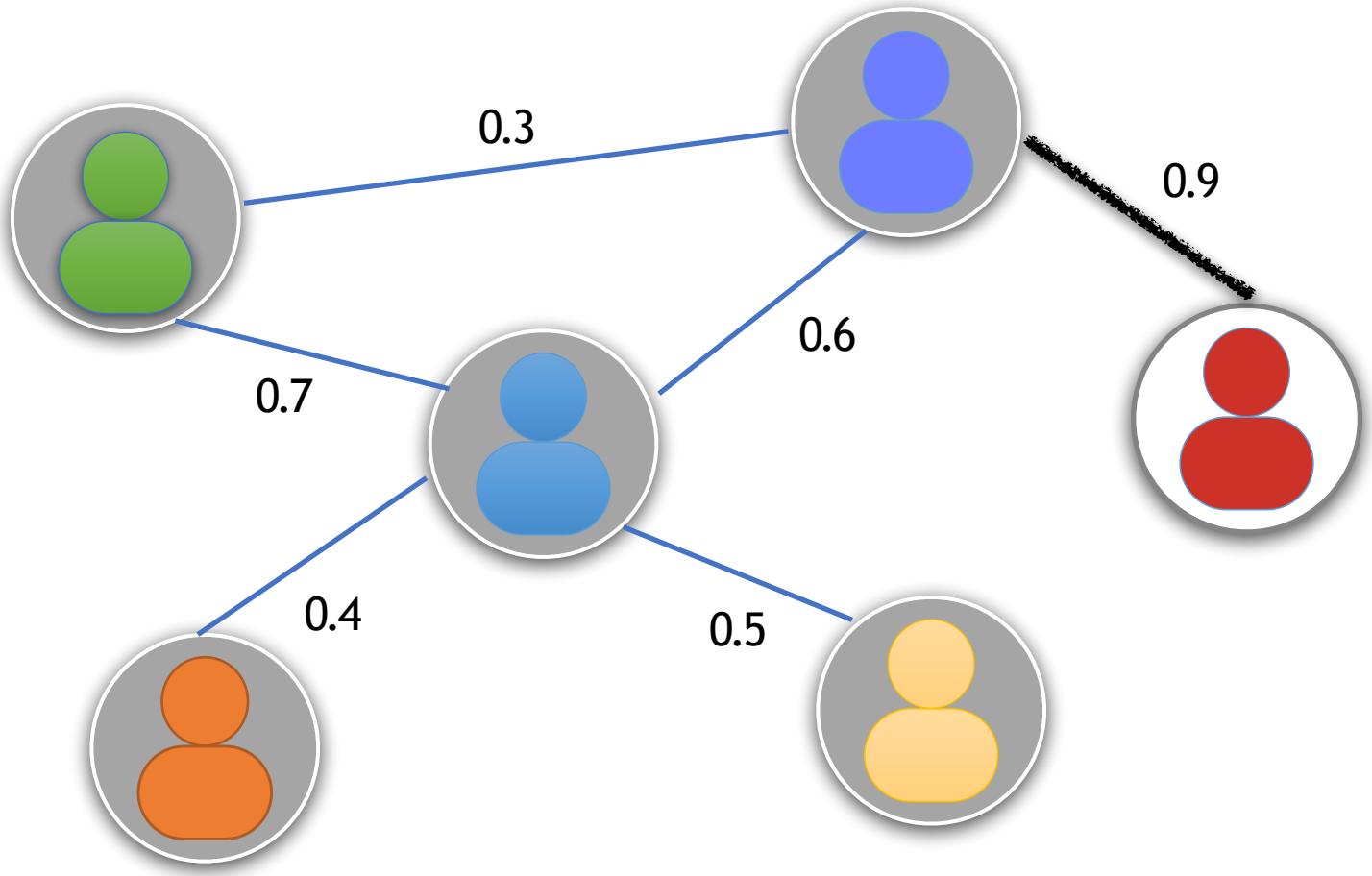


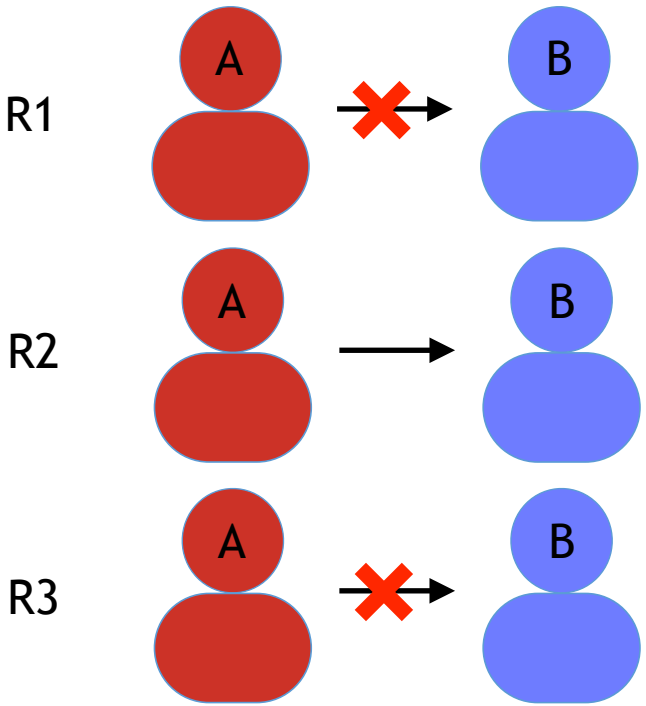
Multi-round Influence Maximization

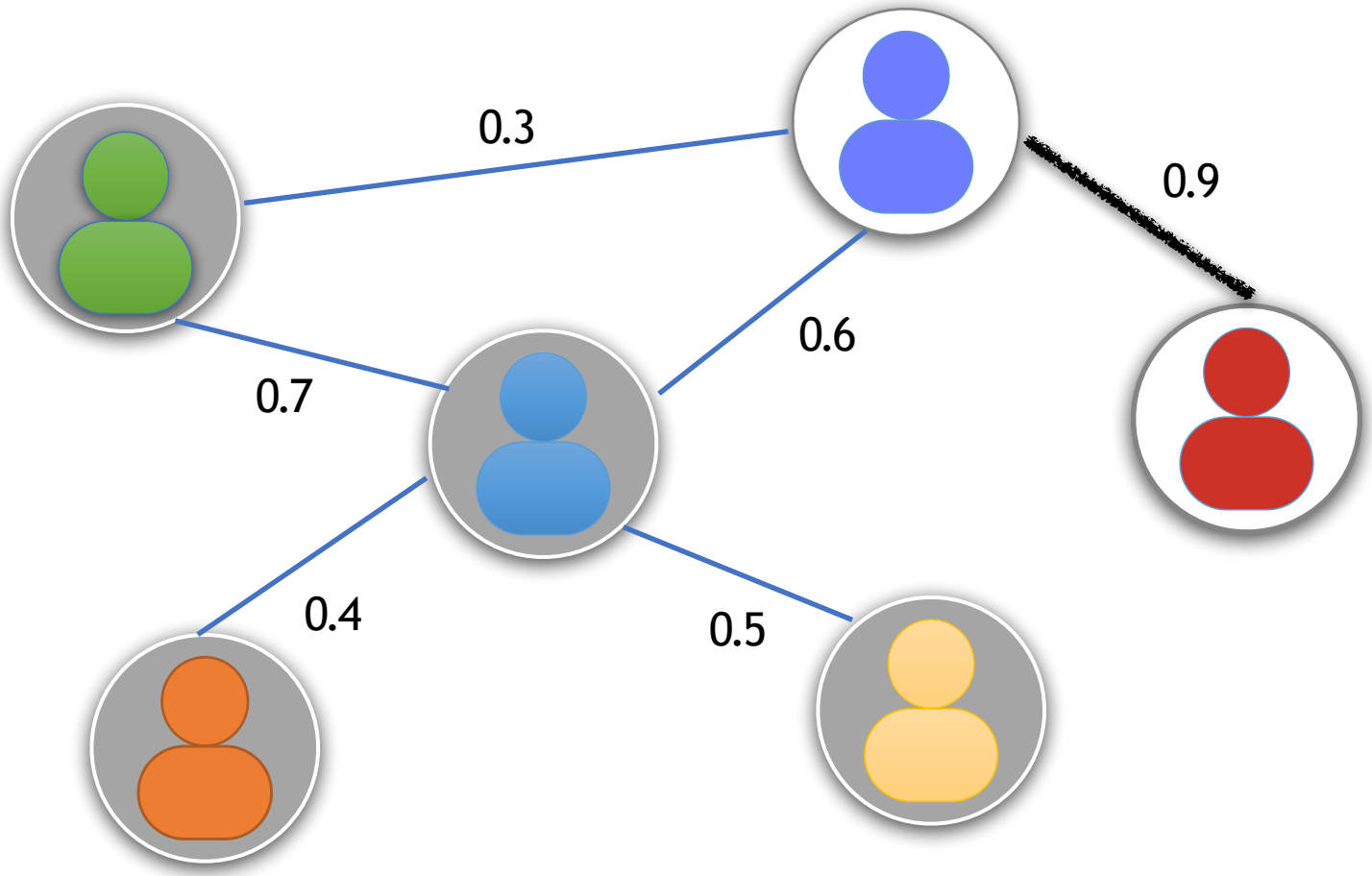

A
B


0.3
0.9
0.6
0.7
0.4
0.5


R1
A
B
R2
A
B
R3
A
B


0.3
0.9
0.6
0.7
0.4
0.5


R1
A
B
R2
A
B
R3
A
B

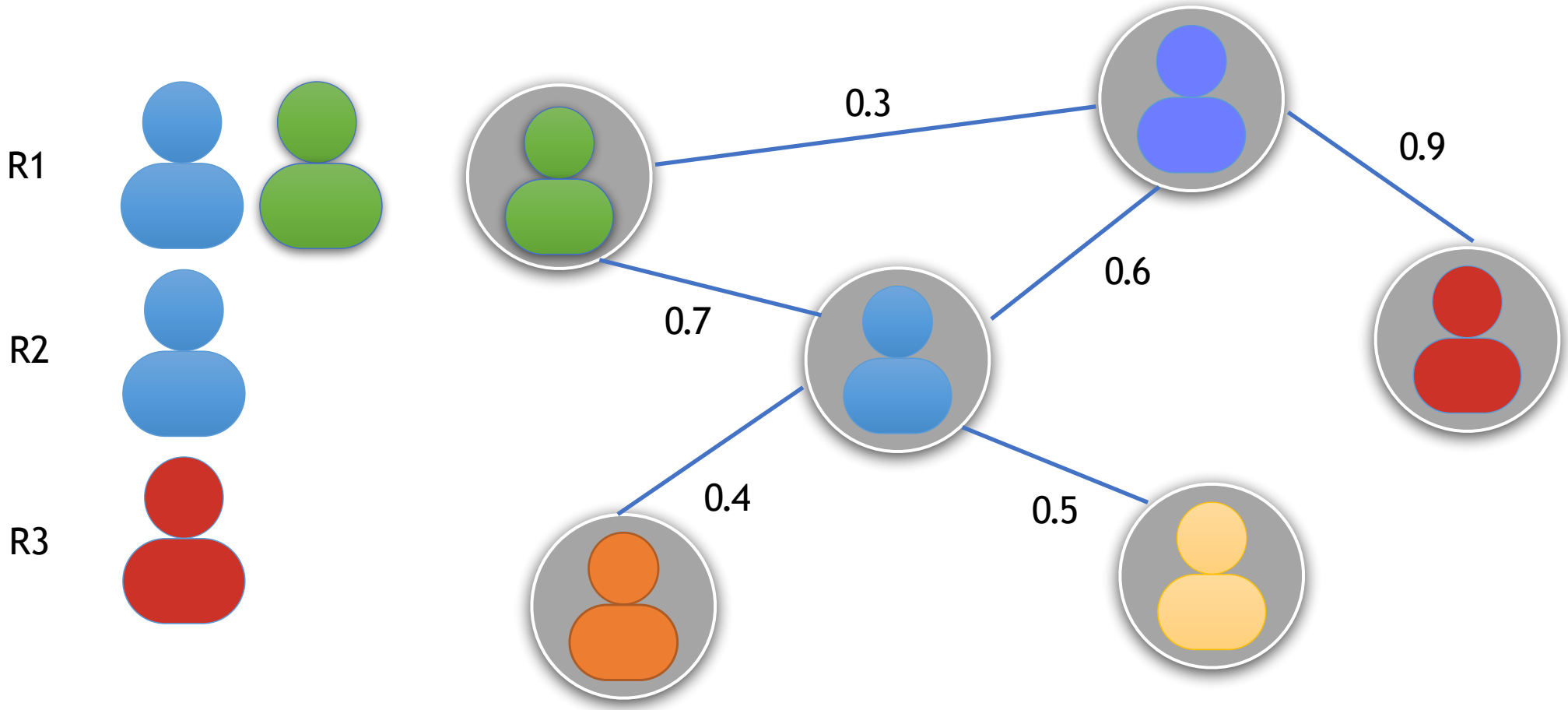

0.3
0.9
0.6
0.7
0.4
0.5

Summary: Multi-Round Influence Maximization

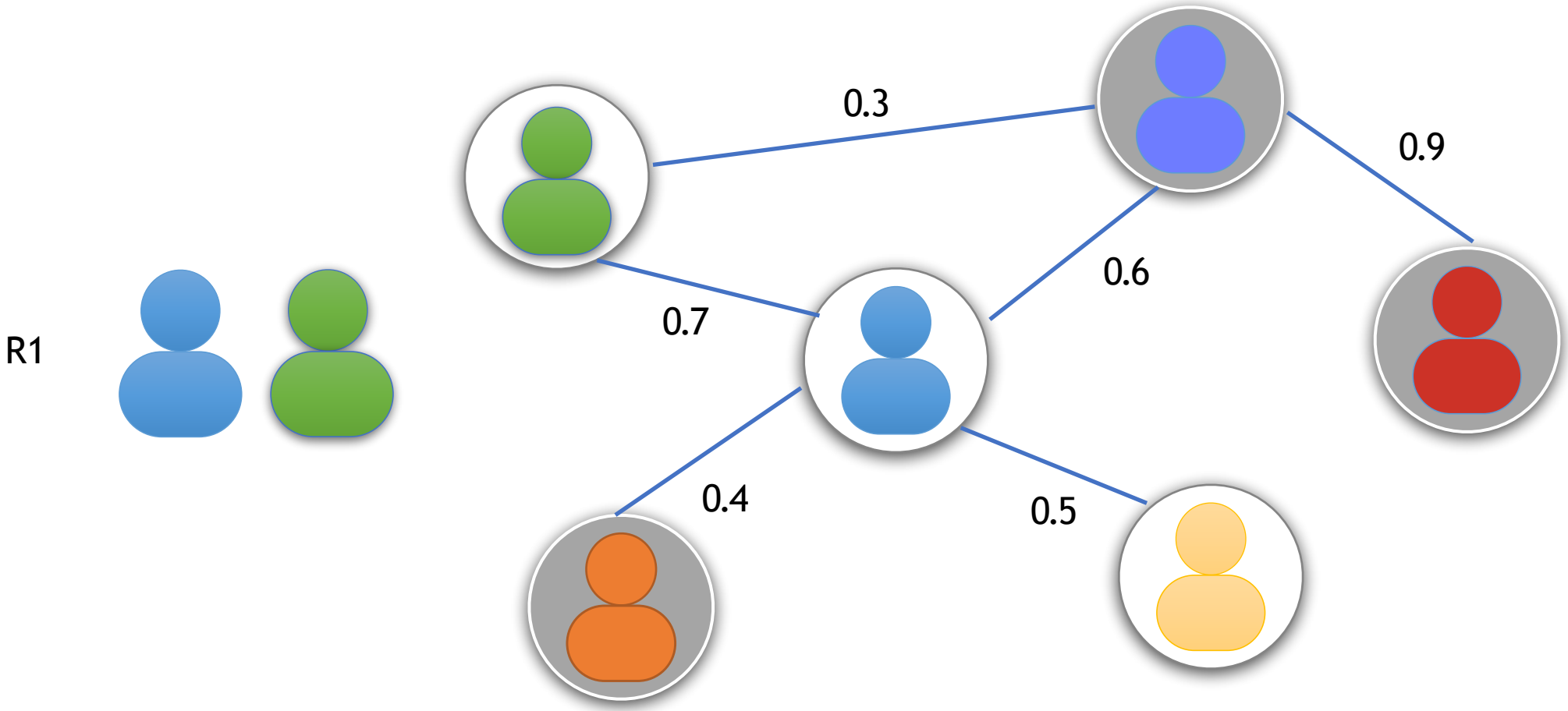
- Given a probability Graph G, budget k and round T
- Greed pick Tk nodes as a seed set
- Every node can be activated at most once in each round
- Every node can not be duplicated in the seed set in each round

- Every node can be activated more than once in multiple round (no double count score but can share information)
- Every node can be duplicated in the seed set in multiple round

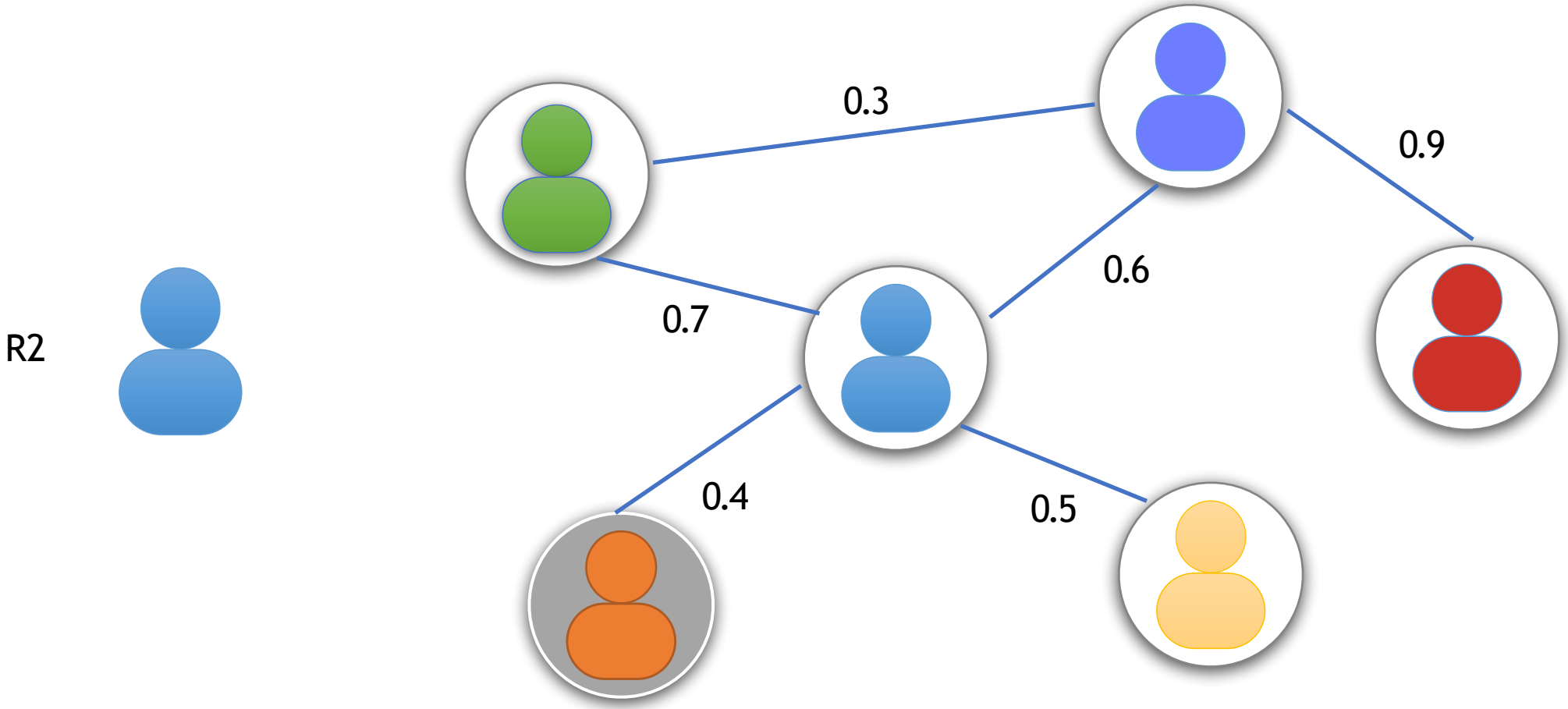
Non-Adaptive Multi-Round Influence Maximization



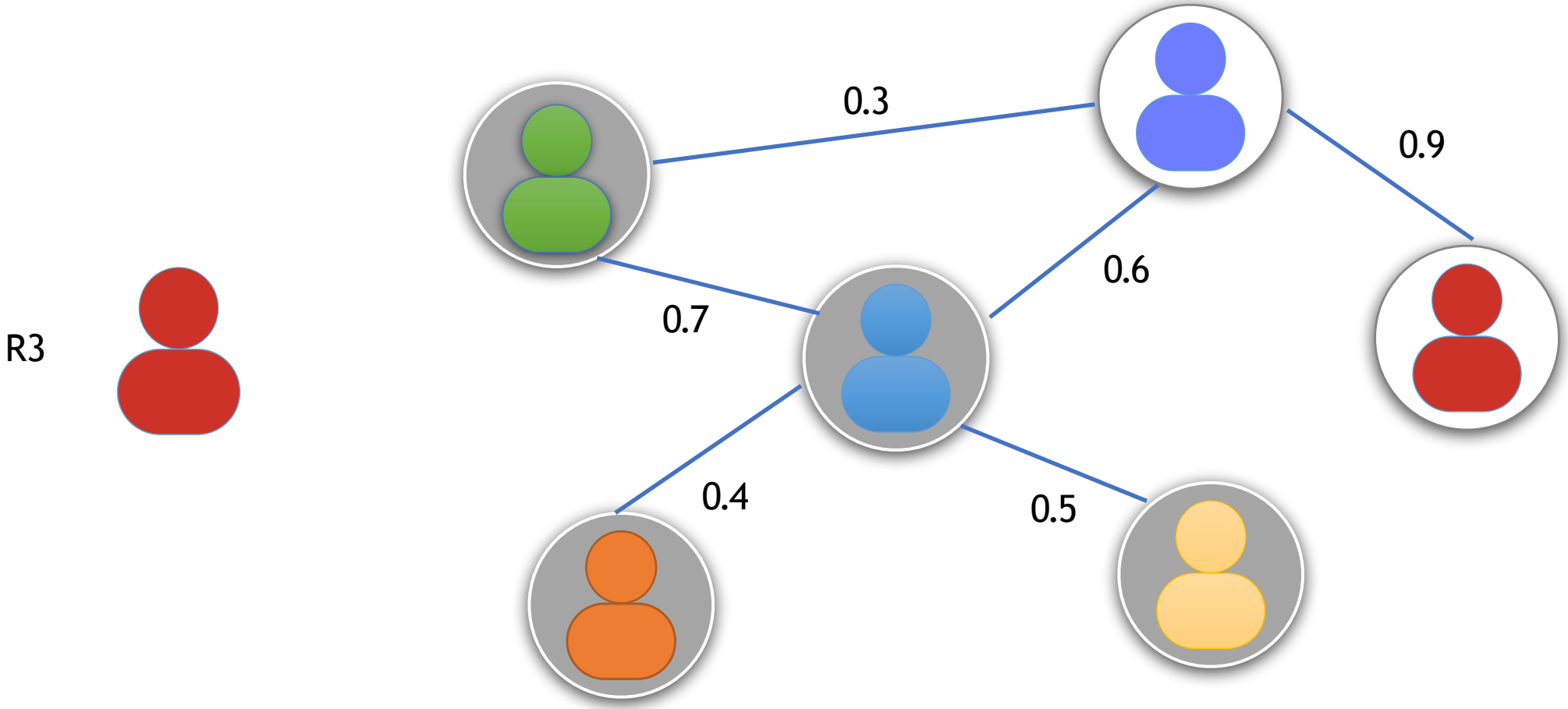
Non-Adaptive Multi-Round Influence Maximization



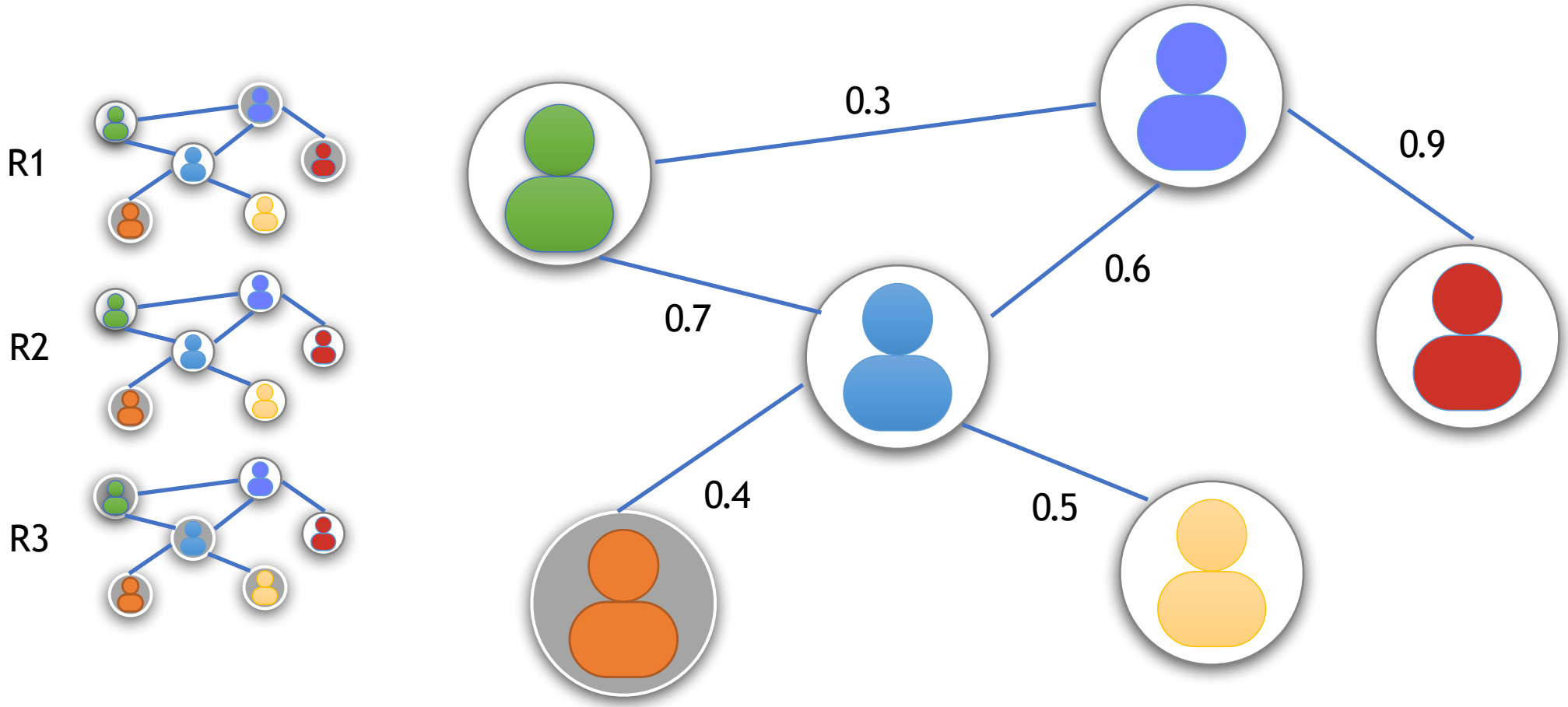
Non-Adaptive Multi-Round Influence Maximization



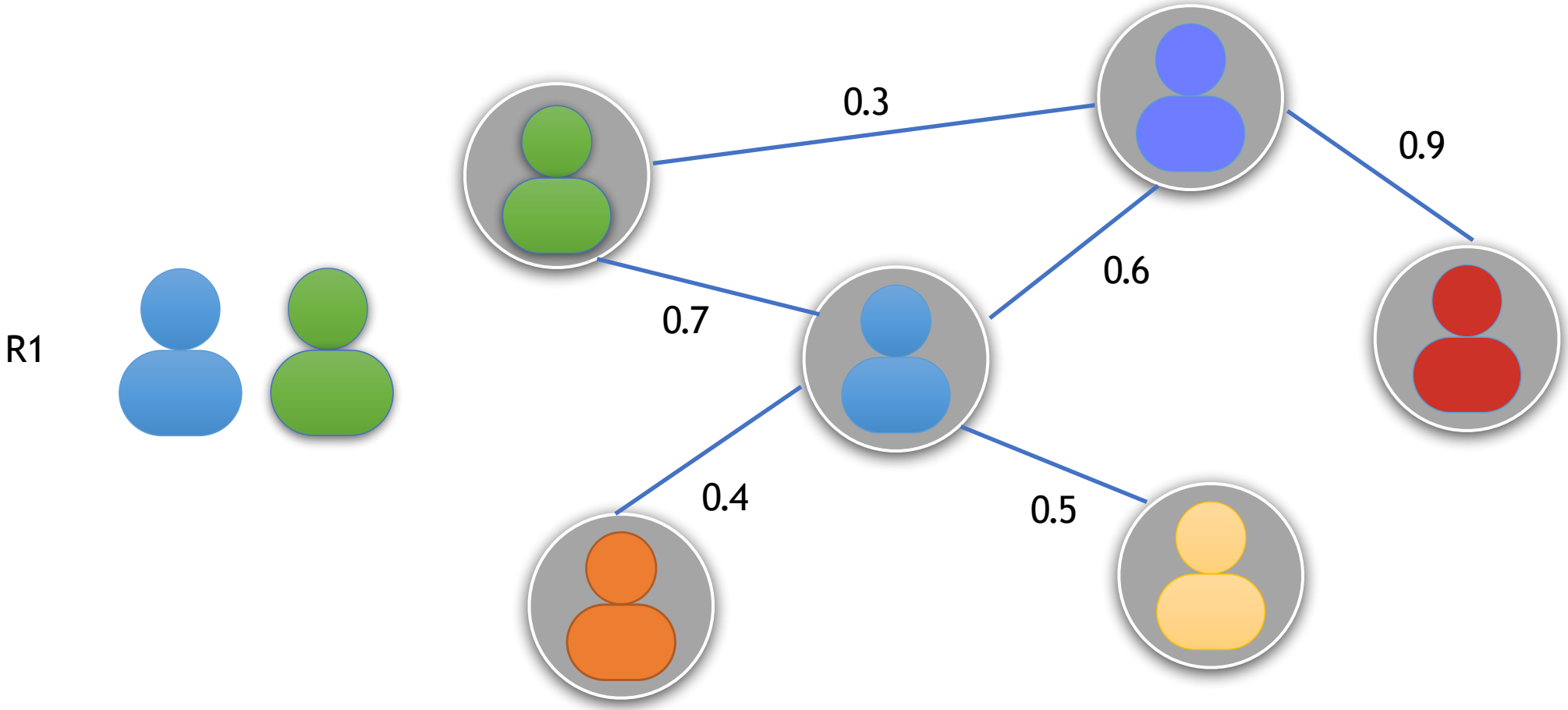
Non-Adaptive Multi-Round Influence Maximization



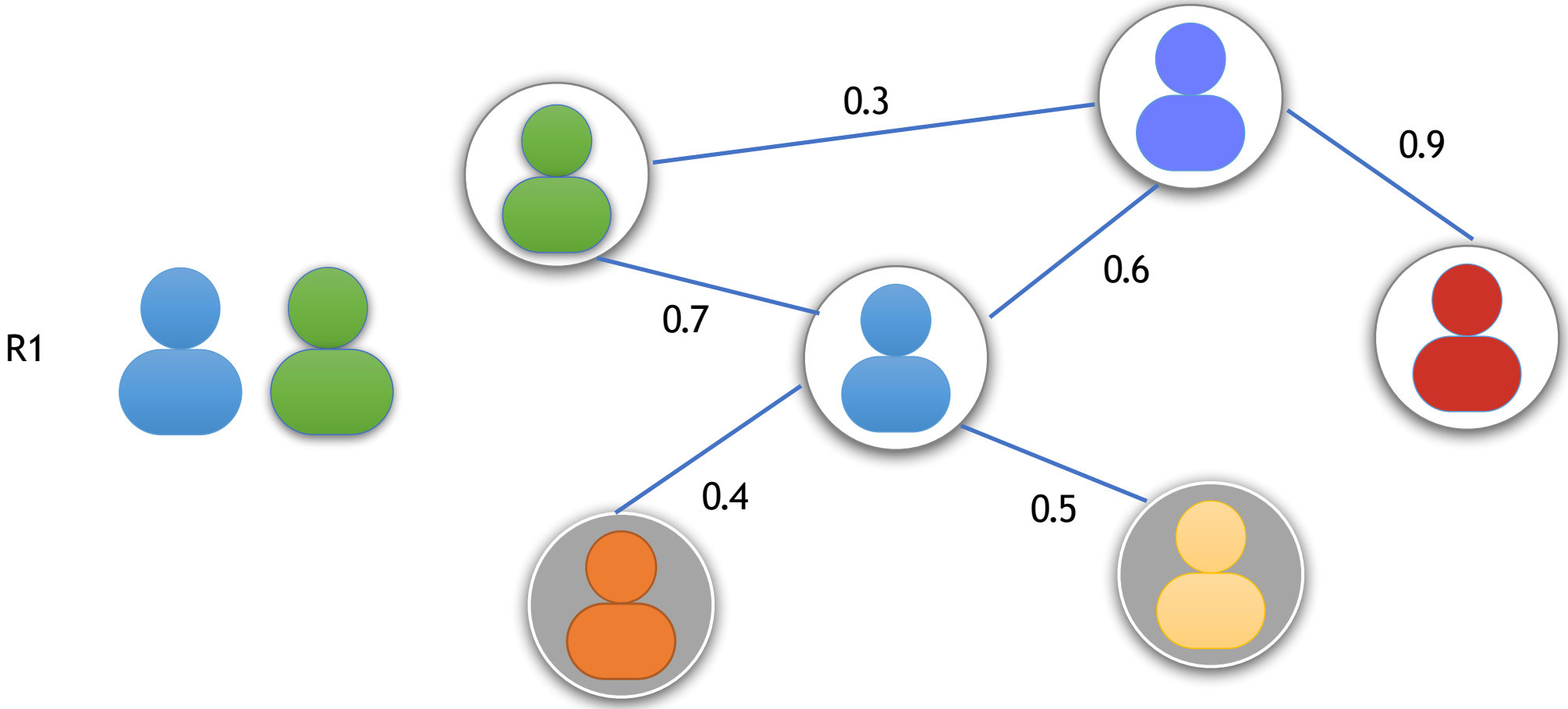
Non-Adaptive Multi-Round Influence Maximization



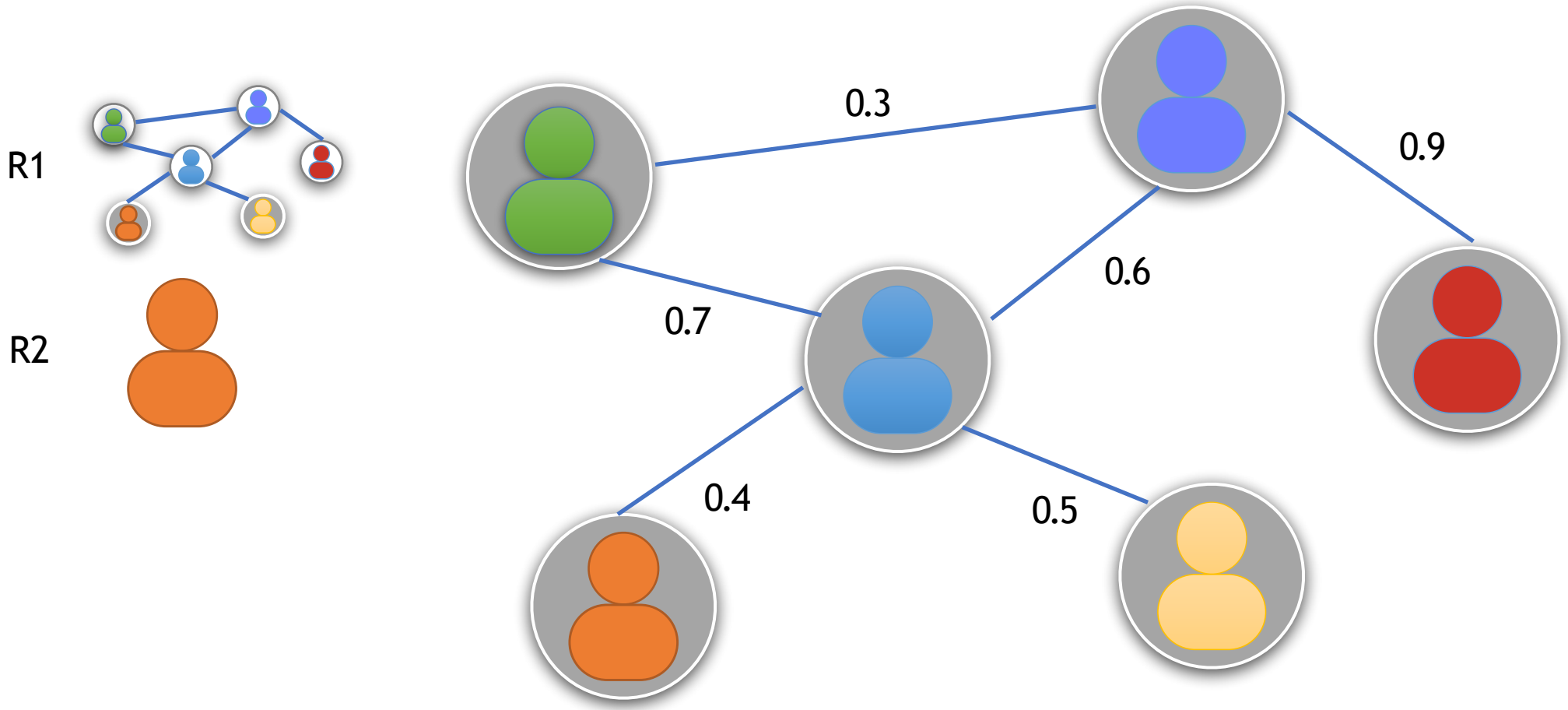
Adaptive Multi-Round Influence Maximization



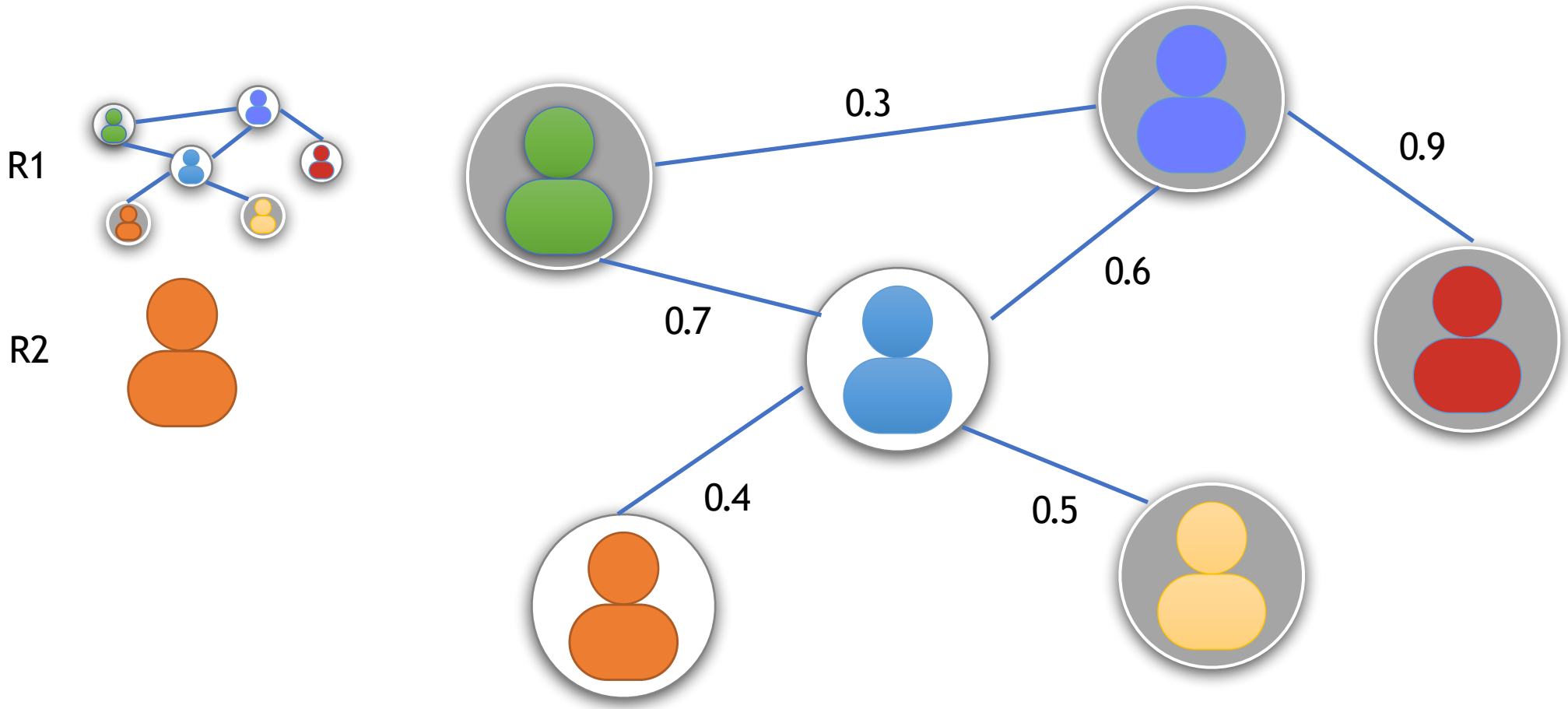
Adaptive Multi-Round Influence Maximization



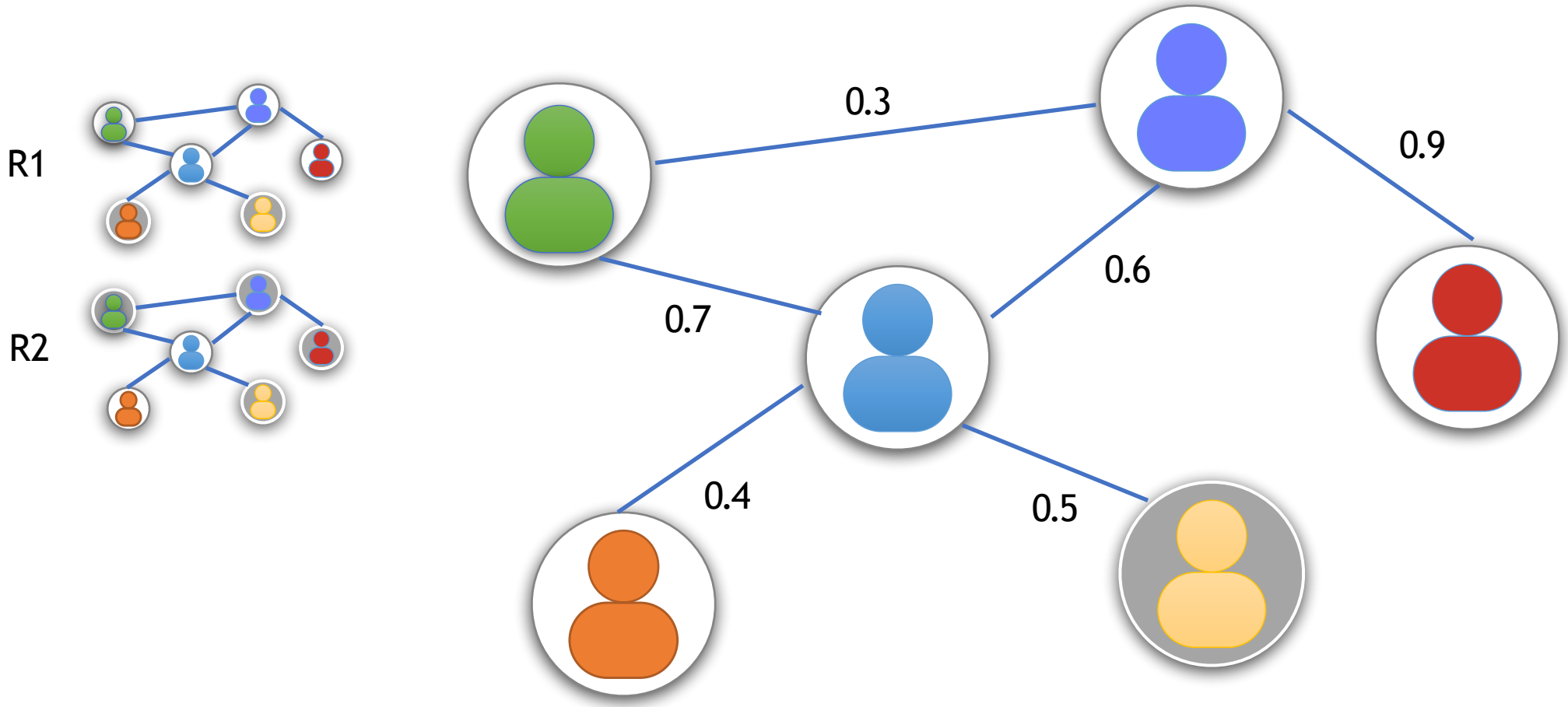
Adaptive Multi-Round Influence Maximization



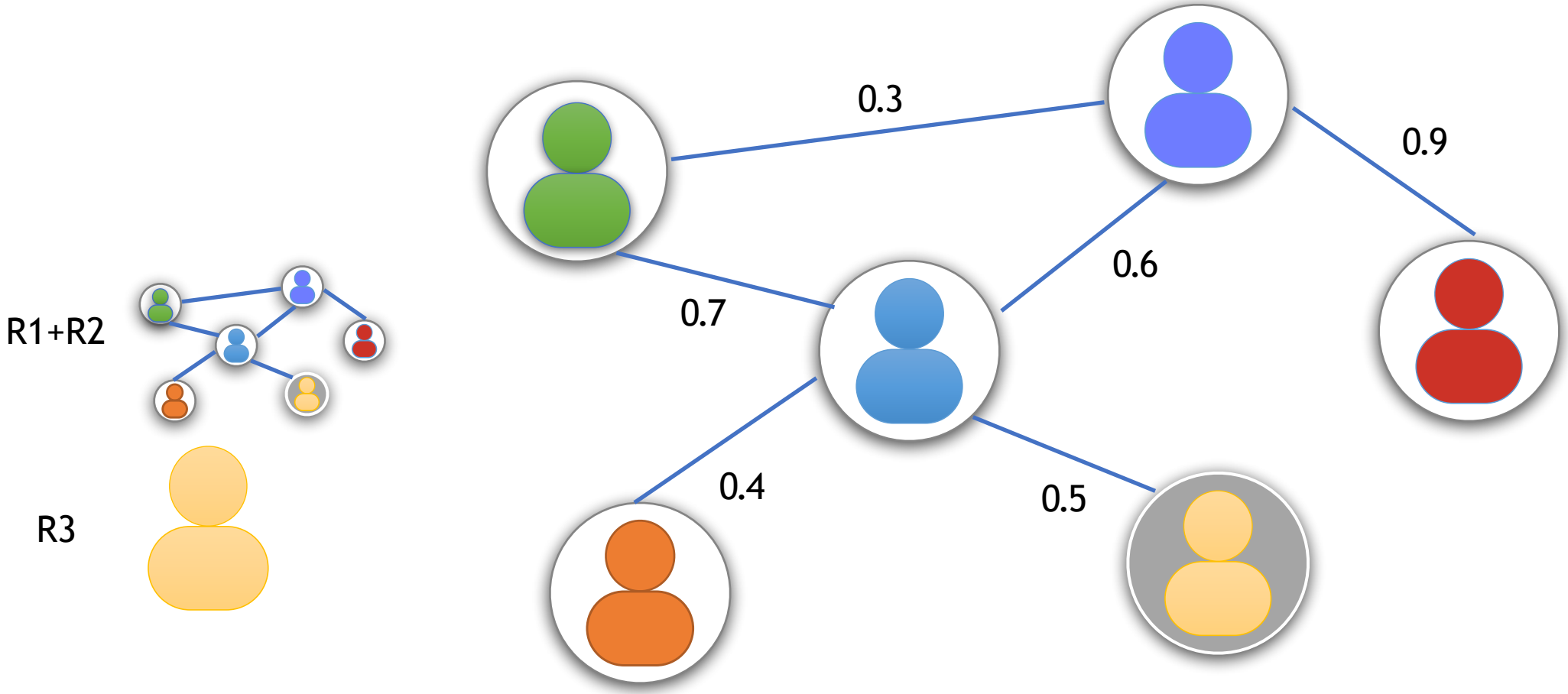
Adaptive Multi-Round Influence Maximization



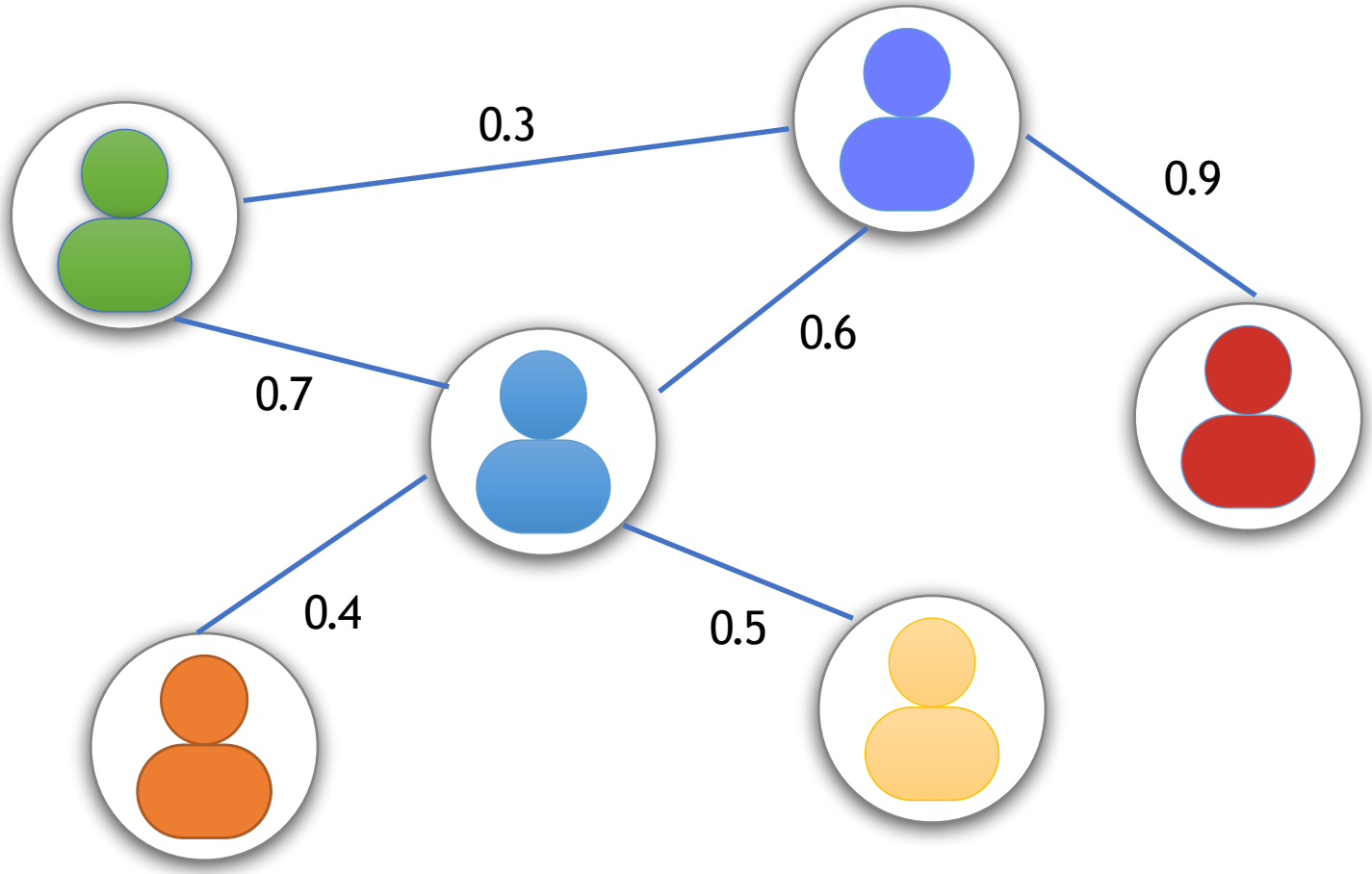
Adaptive Multi-Round Influence Maximization



Adaptive Multi-Round Influence Maximization



Adaptive Multi-Round Influence Maximization

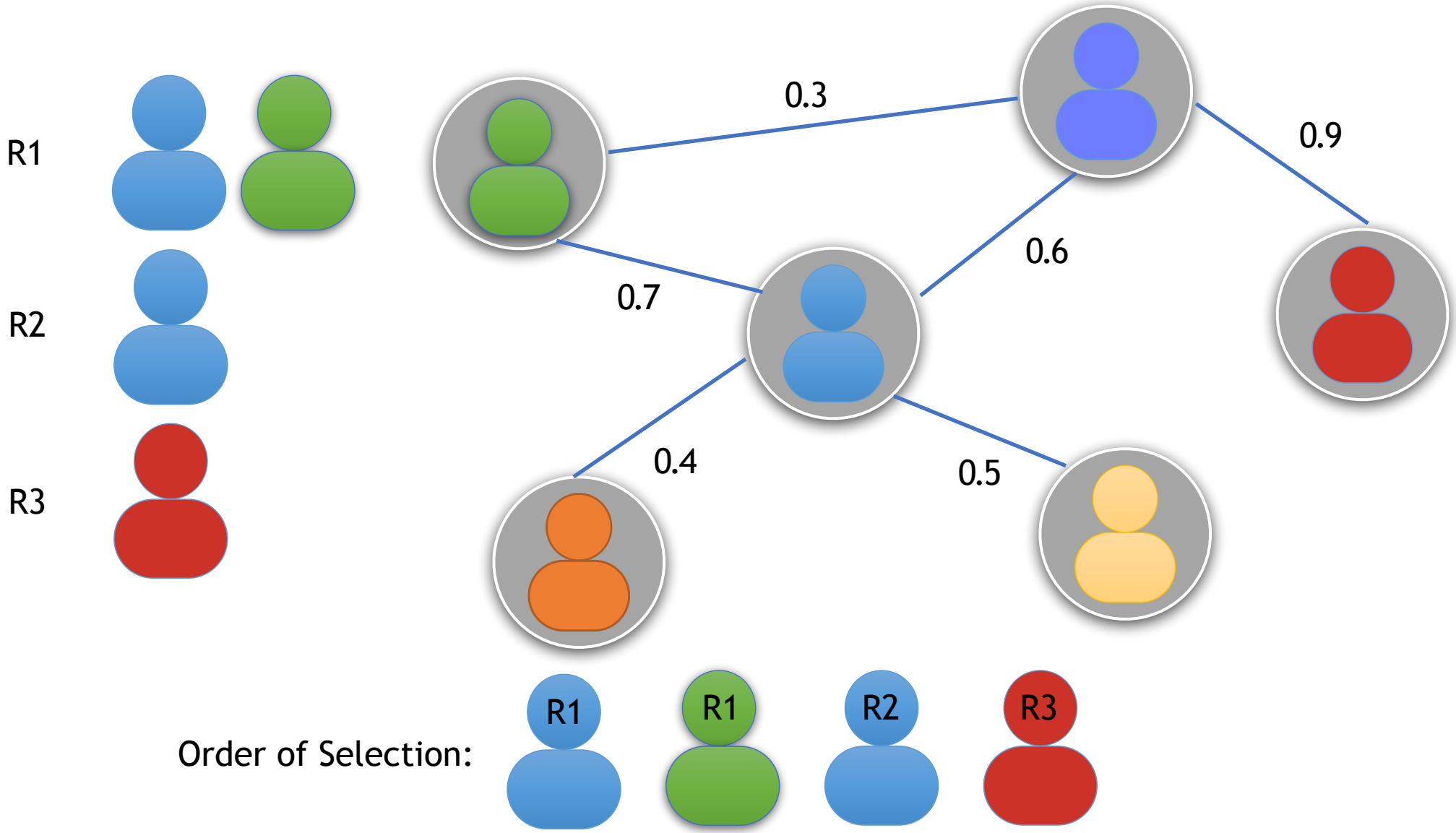


Summary: Adaptive versus Non-Adaptive

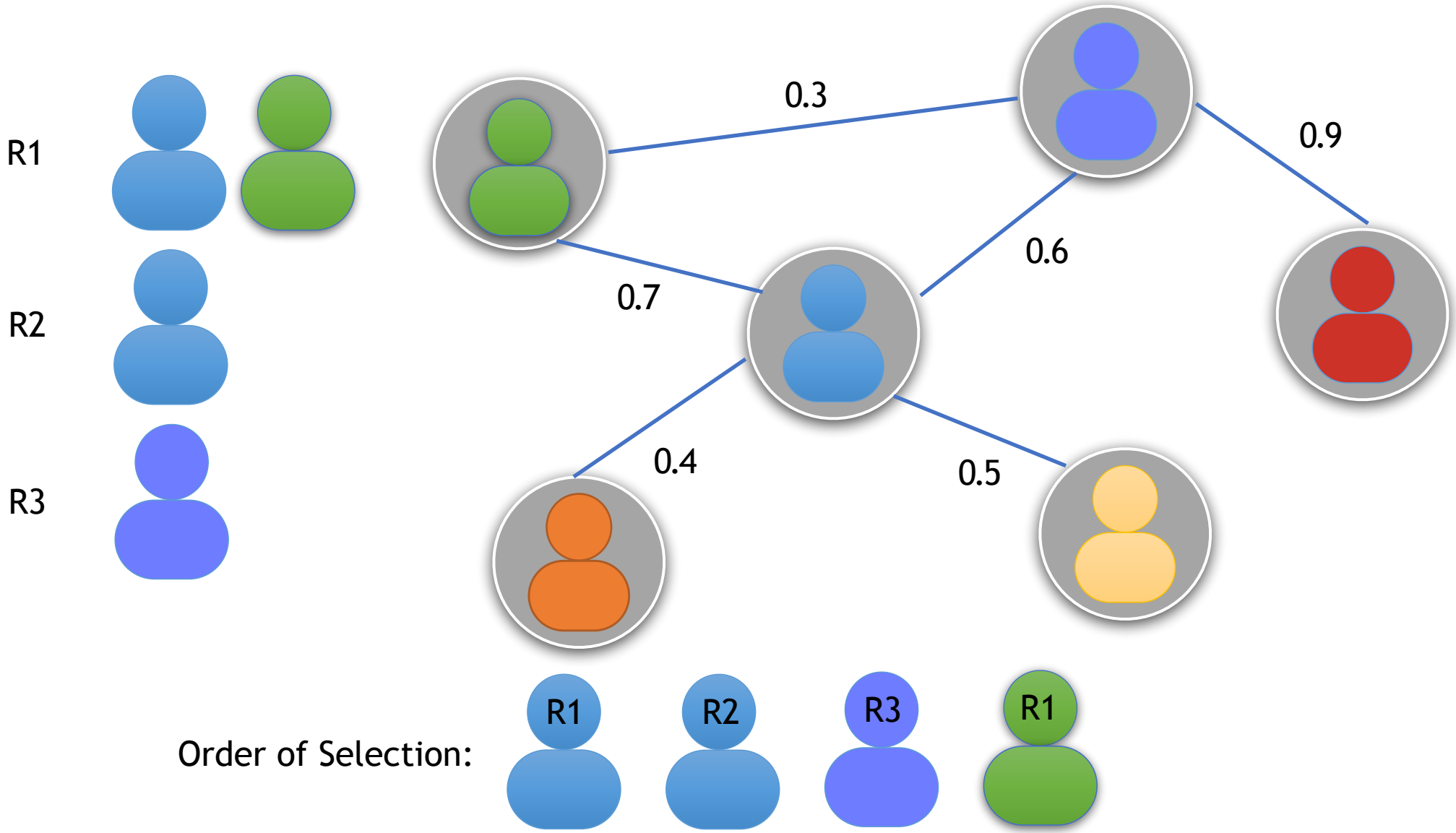
- Non-adaptive algorithm:
  - select all seeds at the beginning
  - do not need to wait for the information feedbacks from previous rounds

- Adaptive algorithm
  - use all influence information in previous rounds for next node selection
  - generally better performance than non-adaptive algorithm

Non-Adaptive Cross-Round Node Selection



Non-Adaptive Cross-Round Node Selection



Summary: Within-round versus Cross-round

- Within-round setting
  - adaptive and non-adaptive
  - faster due to small search space

- Cross-round setting
  - non-adaptive only
  - better performance due to global greedy (large search space)

Summary: Multi-Round Influence Maximization

- Non-adaptive model
  - cross-round seed selection (good performance)
  - within-round seed selection (fast)

- Adaptive model
  - within-round seed selection (fast, good performance, feedbacks)

THEOREM 3.2. *Let $\mathcal{S}^*$ be the optimal solution of the non-adaptive MRIM under cross-round setting. For every $\varepsilon > 0$ and $\ell > 0$, with probability at least $1 - \frac{1}{n^\ell}$, the output $\mathcal{S}^o$ of* CR-Greedy *satisfies*

$$\rho(\mathcal{S}^o) \geq \left(\frac{1}{2} - \varepsilon\right)\rho(\mathcal{S}^*),$$

if CR-Greedy *uses $R = \lceil 31k^2T^2n\log(2kn^{\ell+1})/\varepsilon^2\rceil$ as input. In this case, the total running time is $O(k^3\ell T^4n^2m\log(n)/\varepsilon^2)$, assuming each simulation finishes in $O(m)$ time.*

THEOREM 3.3. *Let $\mathcal{S}^*$ be the optimal solution of the non-adaptive MRIM under within-round setting. For every $\varepsilon > 0$ and $\ell > 0$, with probability at least $1 - \frac{1}{n^\ell}$, the output $\mathcal{S}^o$ of* WR-Greedy *satisfies*

$$\rho(\mathcal{S}^o) \geq \left(1 - e^{-(1-\frac{1}{e})} - \varepsilon\right)\rho(\mathcal{S}^*),$$

if WR-Greedy *uses $R = \lceil 31k^2n\log(2kn^{\ell+1}T)/\varepsilon^2\rceil$ as input. In this case, the total running time is $O(k^3\ell Tn^2m\log(nT)/\varepsilon^2)$, assuming each simulation finishes in $O(m)$ time.*

THEOREM 4.3. *Let π^{ag} represents the policy corresponding to the* AdaGreedy *algorithm. For any $\varepsilon > 0$ and $\ell > 0$, if we use $R = \lceil 31k^2n\log(2kn^{\ell+1}T)/\varepsilon^2\rceil$ simulations for each influence spread estimation, then with probability at least $1 - \frac{1}{n^\ell}$,*

$$f_{\text{avg}}(\pi^{ag}) \geq \left(1 - e^{-(1-\frac{1}{e})} - \varepsilon\right) f_{\text{avg}}(\pi^*).$$

In this case, the total running time for T-round AdaGreedy *is $O(k^3\ell Tn^2m\log(nT)/\varepsilon^2)$.*

Time: Non-adaptive CR > Adaptive WR = Non-adaptive WR

Performance: Non-adaptive CR >= Non-adaptive WR
Adaptive WR >= Non-adaptive WR

THEOREM 3.2. *Let $\mathcal{S}^*$ be the optimal solution of the non-adaptive MRIM under cross-round setting. For every $\varepsilon > 0$ and $\ell > 0$, with probability at least $1 - \frac{1}{n^\ell}$, the output $\mathcal{S}^o$ of* CR-Greedy *satisfies*

$$\rho(\mathcal{S}^o) \geq \left(\frac{1}{2} - \varepsilon\right)\rho(\mathcal{S}^*),$$

if CR-Greedy *uses $R = \lceil 31k^2T^2n\log(2kn^{\ell+1})/\varepsilon^2\rceil$ as input. In this case, the total running time is $O(k^3\ell T^4n^2m\log(n)/\varepsilon^2)$, assuming each simulation finishes in $O(m)$ time.*

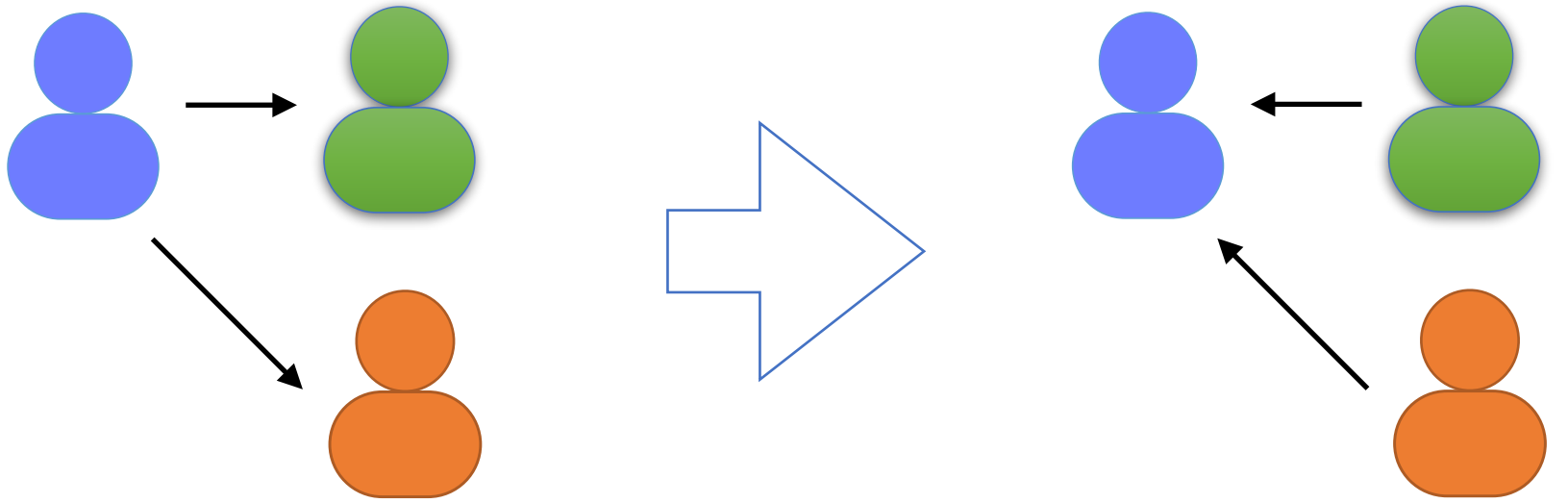
THEOREM 4.3. *Let π^{ag} represents the policy corresponding to the* AdaGreedy *algorithm. For any $\varepsilon > 0$ and $\ell > 0$, if we use $R = \lceil 31k^2n\log(2kn^{\ell+1}T)/\varepsilon^2\rceil$ simulations for each influence spread estimation, then with probability at least $1 - \frac{1}{n^\ell}$,*

$$f_{\text{avg}}(\pi^{\text{ag}}) \geq \left(1 - e^{-(1-\frac{1}{e})} - \varepsilon\right) f_{\text{avg}}(\pi^*).$$

In this case, the total running time for T-round AdaGreedy *is $O(k^3\ell Tn^2m\log(nT)/\varepsilon^2)$.*

Which is better??? Non-adaptive CR, Adaptive WR ???

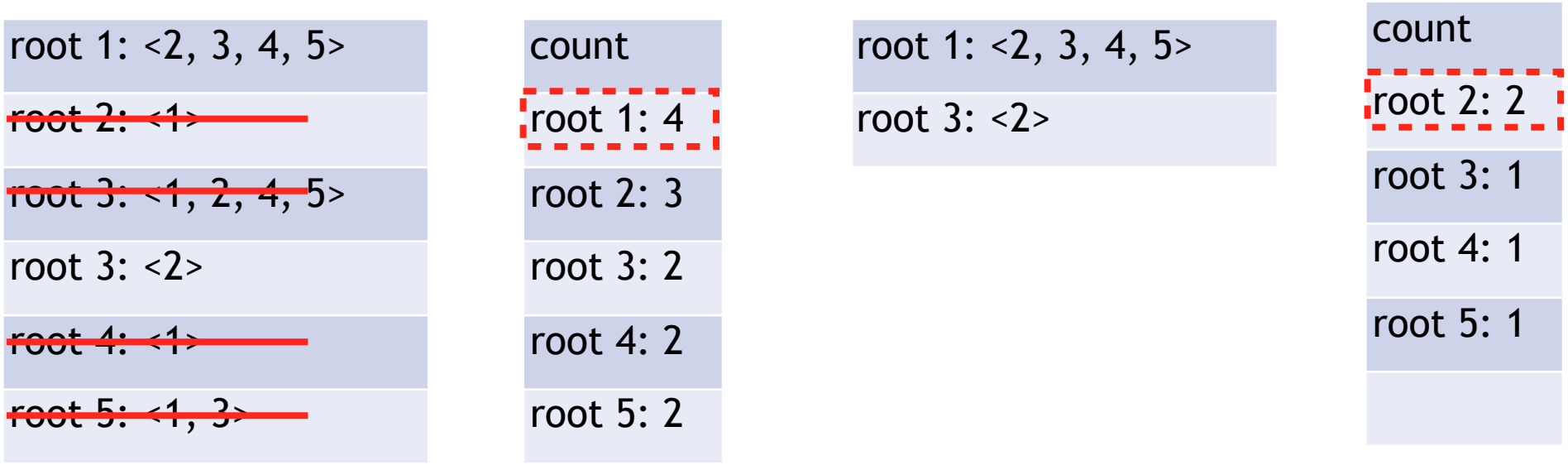
Optimization: Reverse Reachable Set



Monto-Carlo

Reverse-Reachable

Optimization: Reverse Reachable Set



We implement all multi-round model with reverse reachable set algorithm.

Speed up 1000 times!!!


Concept
Creativity
Development
Mobile Apps Design
Social network
Browser Respective Design
SEO

Experiment

Dataset	Description	node	edge
Flixster	movie social	29357	212614
NetHEPT	author/co-author	15233	62774

Table 1: The performance of influence spread on NetHEPT.

Method/Simulations	Round				
	1	2	3	4	5
SG (R = 10000)	290.1 [288.8, 291.4]	505.7 [504.0, 507.3]	688.6 [686.6, 690.4]	868.2 [866.2, 870.2]	1027.3 [1025.2, 1029.4]
SG-R (R = 10000)	289.5 [288.2, 290.8]	516.3 [514.6, 518.0]	714.0 [712.0, 716.0]	884.9 [882.7, 887.1]	1042.0 [1039.7, 1044.2]
E-WR-Greedy (R = 10000)	290.7 [289.4, 292.0]	528.9 [527.2, 530.6]	738.8 [736.9, 740.8]	930.2 [928.0, 932.3]	1097.6.9 [1095.3, 1099.8]
WR-IMM (R = 10000)	290.9 [289.7, 292.3]	532.8 [531.1, 534.5]	745.3 [743.2, 747.3]	930.1 [928.0, 932.2]	1093.1 [1090.8, 1095.3]
CR-Greedy (R = 10000)	267.8 [266.5, 269.1]	528.7 [527.2, 530.4]	730.4 [728.5, 732.4]	938.5 [933.7, 937.8]	1121.3 [1119.0, 1123.5]
CR-IMM (R = 10000)	283.0 [281.7, 284.2]	517.4 [515.7, 519.2]	721.9 [720.0, 723.9]	931.6 [929.4, 933.7]	1129.7 [1127.7, 1131.9]
AdaGreedy (R = 150)	288.3 [276.7, 299.7]	533.4 [519.4, 547.3]	758.1 [743.6, 772.7]	960.1 [943.9, 976.3]	1141.5 [1123.7, 1160.0]
AdaIMM (R = 150)	**291.8** [281.3, 302.4]	**544.4** [531.6, 557.2]	**761.8** [746.6, 776.9]	**965.8** [949.7, 982.0]	**1146.3** [1129.1, 1163.5]

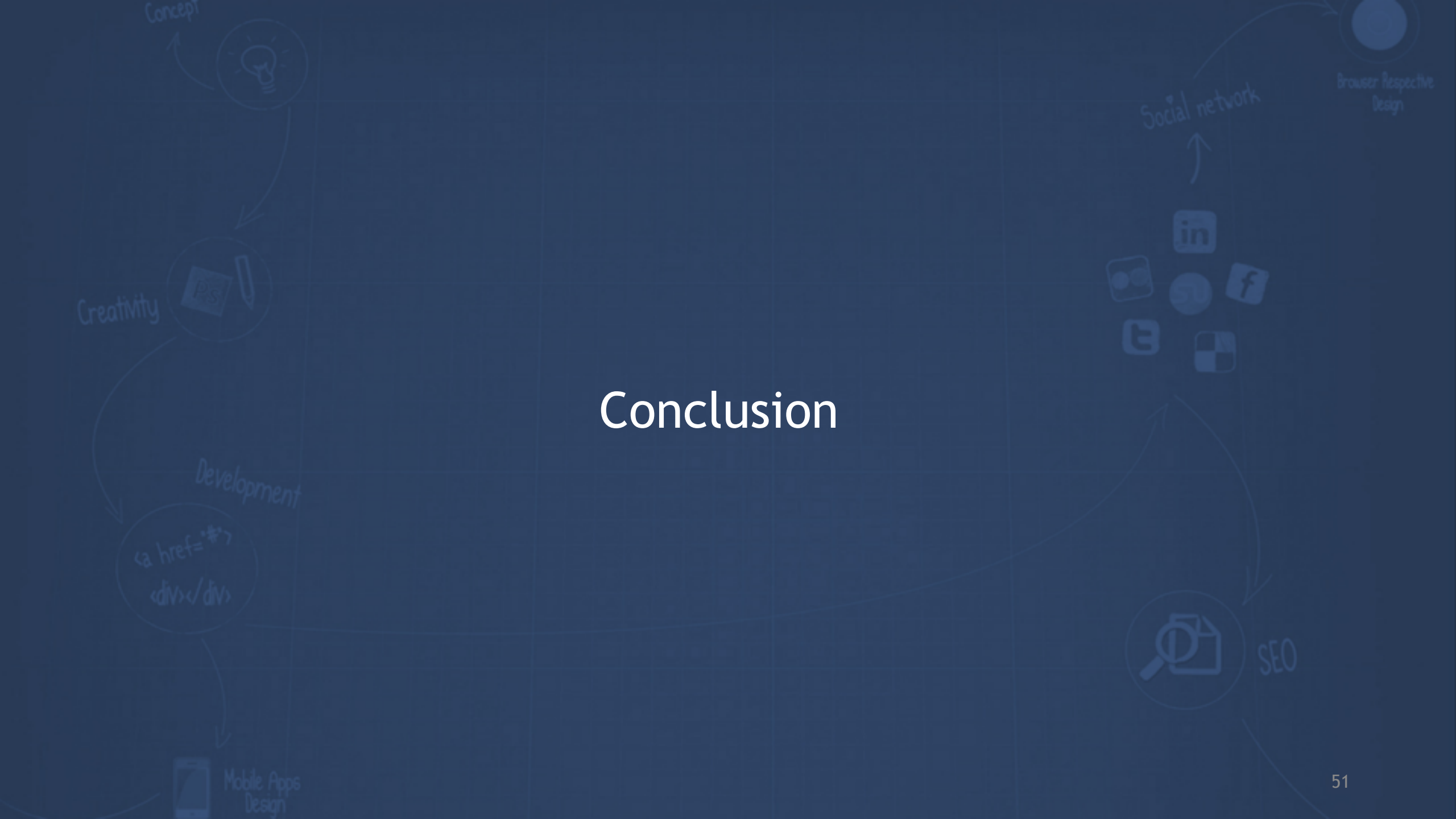
Table 3: Influence spread with different adaptive degree.

Num. of Rounds	1	2	5	10
Num. of Seeds	50	25	10	5
AdaIMM (R = 100)	883.0 [856.0, 910.1]	1040.3 [1022.6, 1058.1]	1141.0 [1119.3, 1162.6]	1204.7 [1178.2, 1231.3]

Table 4: Running time of the algorithms, in seconds.

	SG	SG-R	E-WR-Greedy	WR-IMM
NetHEPT (R = 5)	439.2 [407, 470.94]	87.8 [81.5, 94.2]	551.2 [527.9, 574.4]	1.97 [1.91, 2.03]
Flixster (R = 5)	4862.3 [4773.3, 4951.3]	972.5 [990.3,954.7]	2478.9 [2422.4, 2535.5]	3.16 [3.14, 3.18]

	CR-Greedy	CR-IMM	AdaGreedy	AdaIMM
NetHEPT (R = 5)	2105.6 [2036.2, 2175.0]	2.13 [2.05, 2.21]	465.4 [473.8, 457.0]	2.01 [1.93, 2.09]
Flixster (R = 5)	9587.6 [9145.3.10029.9]	3.61 [3.59, 3.63]	2305.5 [2161.0, 2450.0]	3.23 [3.16, 3.30]


Concept
Creativity
Development
Mobile Apps Design
Social network
Browser Respective Design
SEO

Conclusion

Summary and Future Plan

lsun29@uic.edu

- Non-adaptive model
  - cross-round seed selection (good performance)
  - within-round seed selection (fast)
- Adaptive model
  - within-round seed selection (fast, good performance, feedbacks)

- Multi-round is not only a new model, but an idea brings more research opportunities.


Thank
you