Predicting Popularity of Twitter Accounts through the Discovery of Link-Propagating Early Adopters

Daichi Imamori

Kyoto University

(CyberZ, inc.)

Keishi Tajima

Kyoto University

BACKGROUND

In social media, such as 

- New useful users frequently appears.
- We want to detect such new useful users.
- Popularity-based methods, e.g. , HITS and PageRank, do not work well for new users that have not established their reputation yet.

We propose a method of estimating prospective popularity of new users.

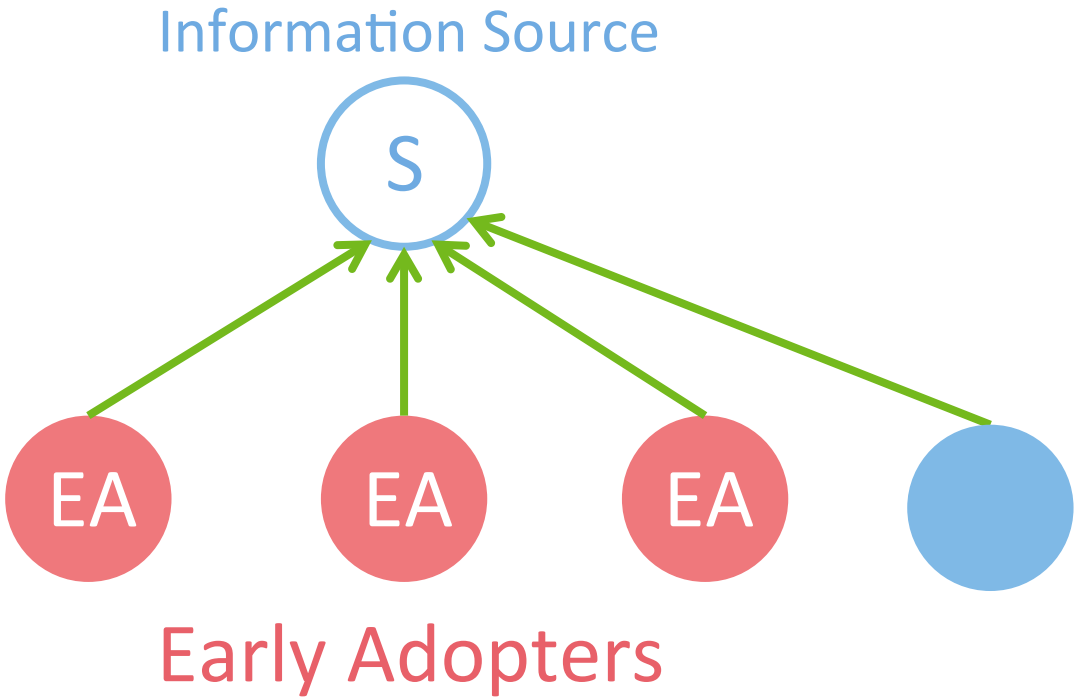
OUR APPROACH

We first detect early adopters

※ Early adopters = The users who are good at finding new good information sources earlier than others.

- The new users followed by early adopters are probably good information sources even if they have few followers at this point.
- We can find good information sources by detecting early adopters.

OUR APPROACH



DETECTION OF EARLY ADOPTERS

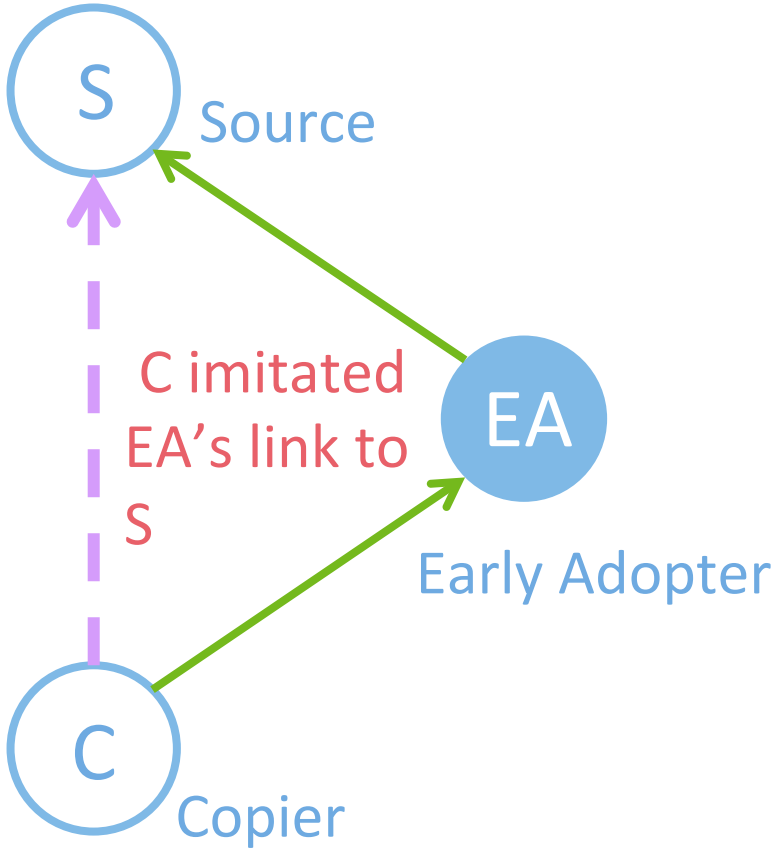
How do we detect early adopters?

Assumption

Early adopters

↕

The users whose follow links are imitated by many followers.



DETECTION OF EARLY ADOPTERS

How do we detect early adopters?

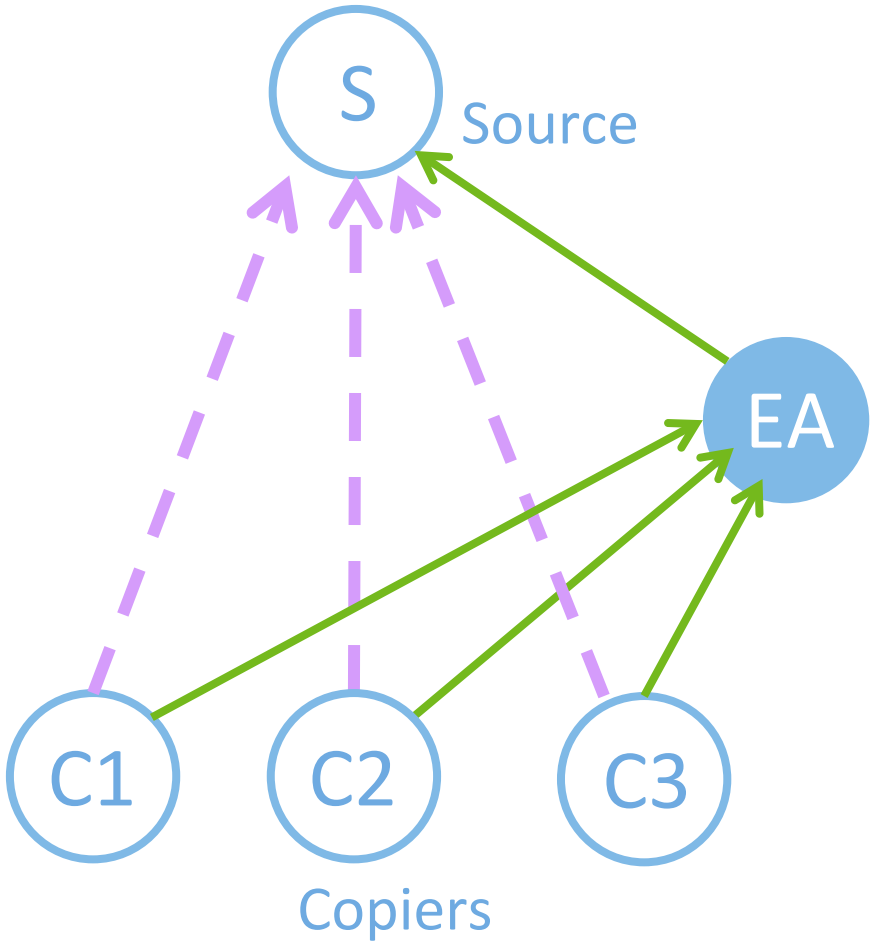
Assumption

Early adopters

↕

The users whose follow links are imitated by many followers.

We can detect early adopters based on the frequency of link imitation.

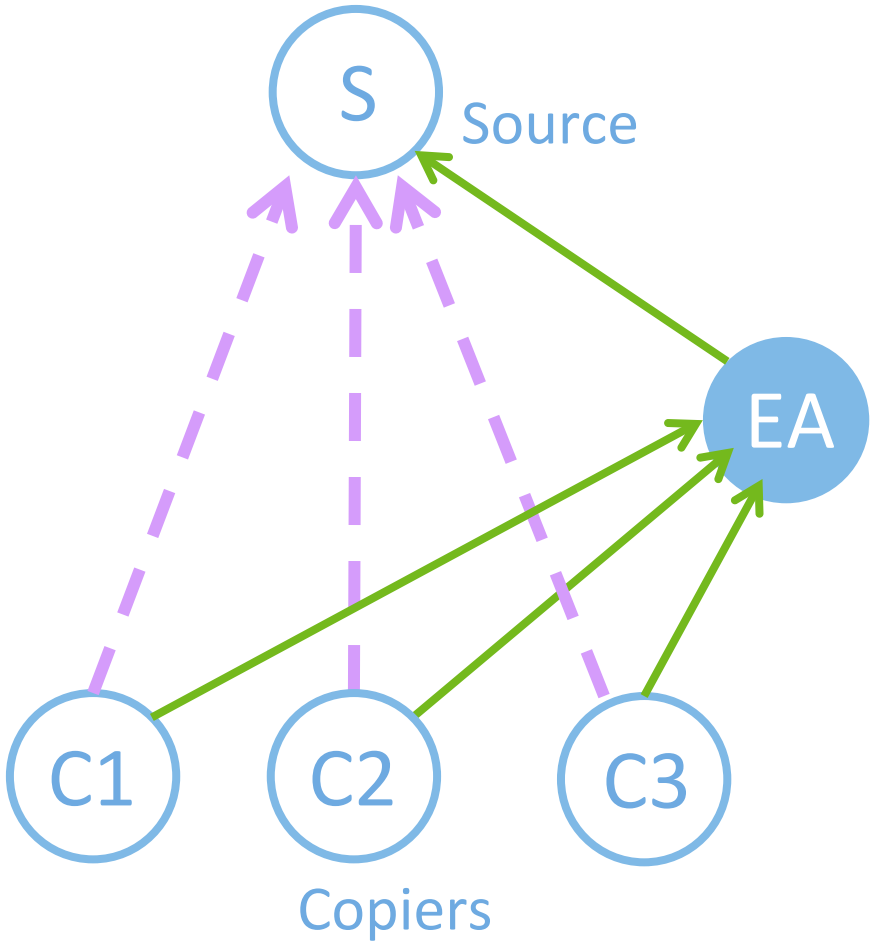


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

3. Calculate ***early adopter score*** from the number of link imitation.

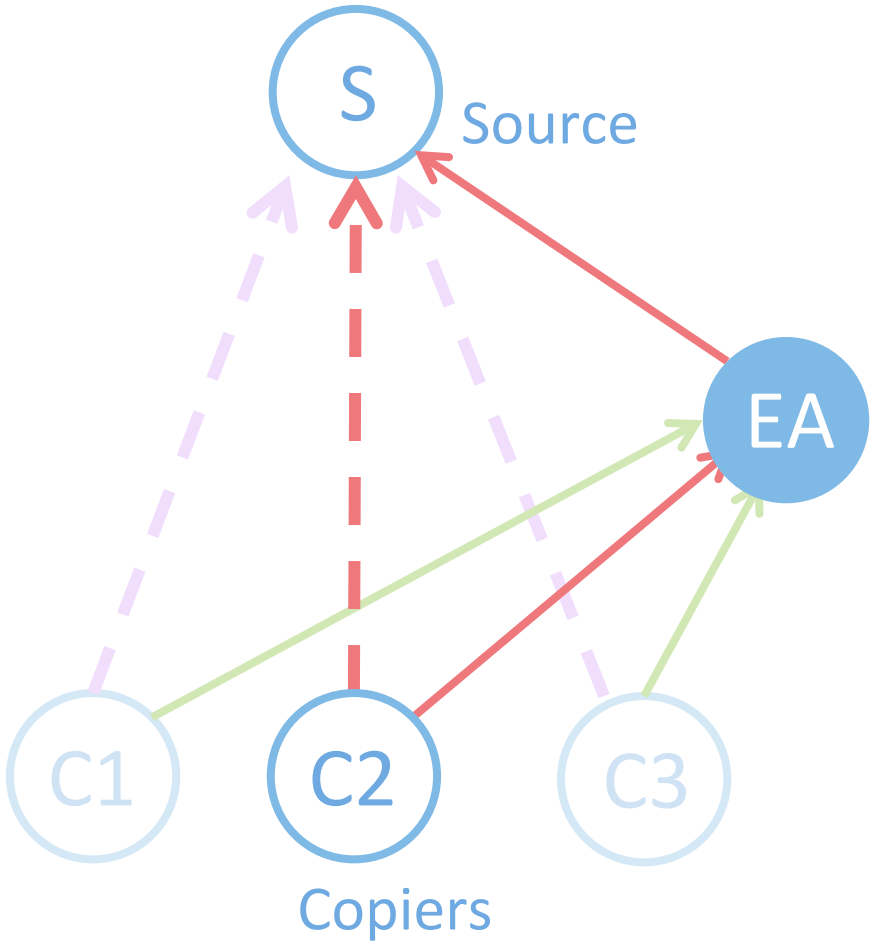


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

3. Calculate ***early adopter score*** from the number of link imitation.

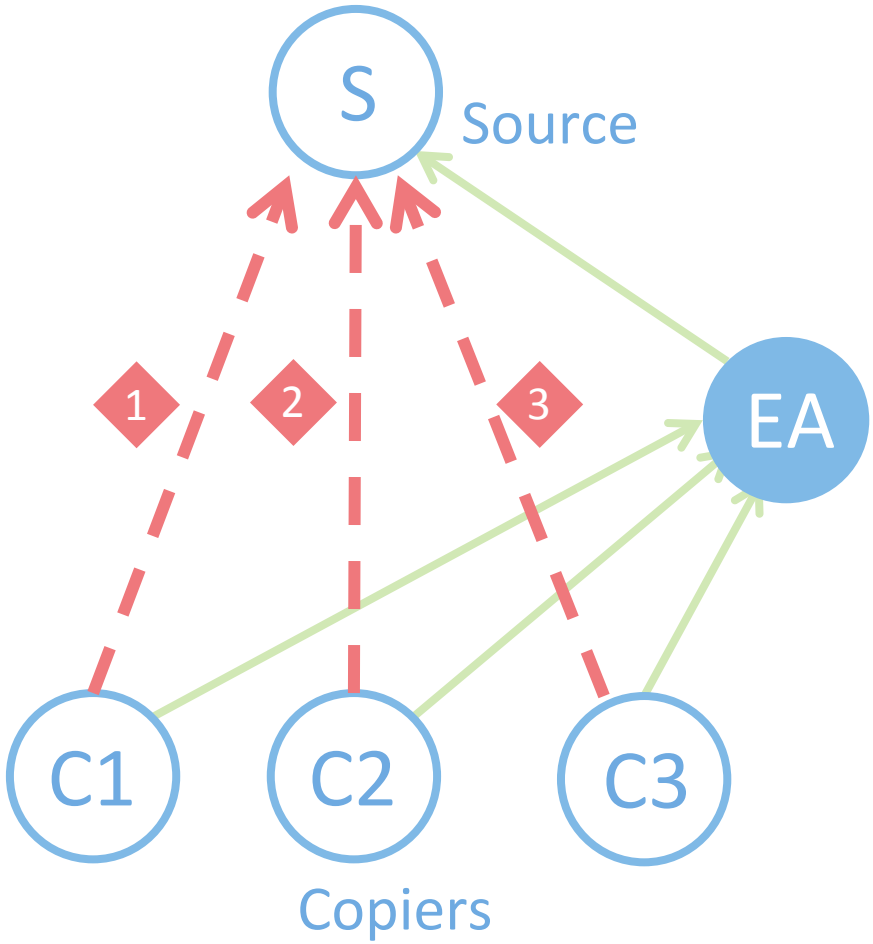


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

3. Calculate ***early adopter score*** from the number of link imitation.

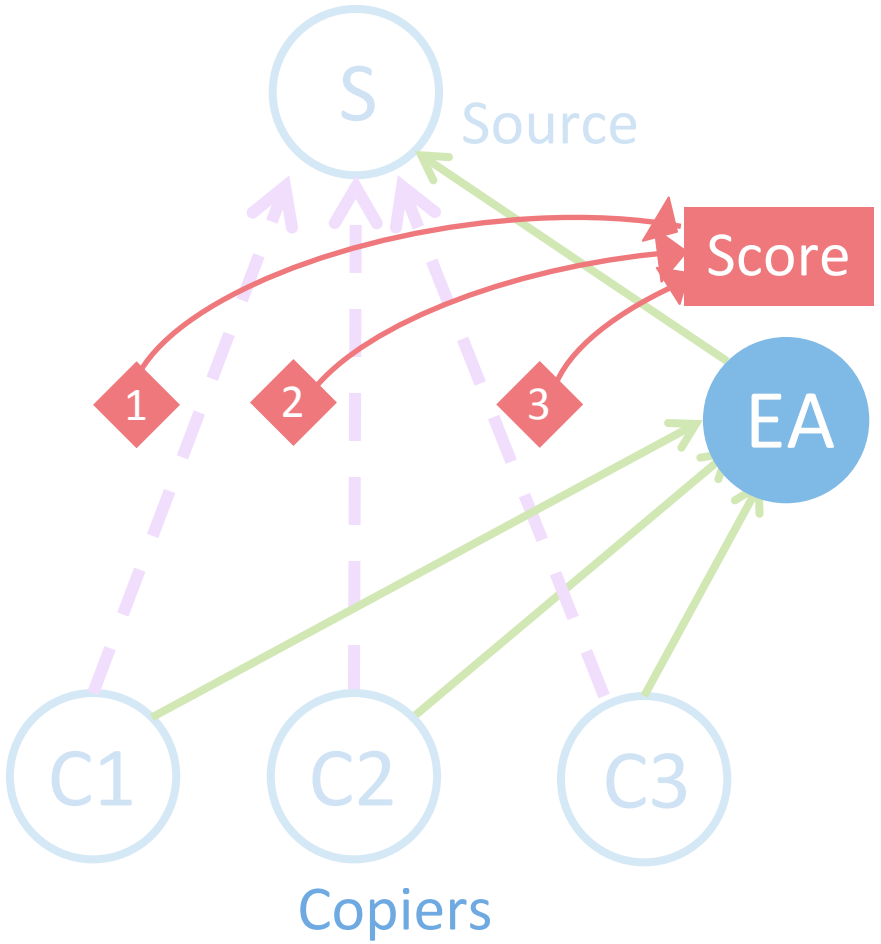


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

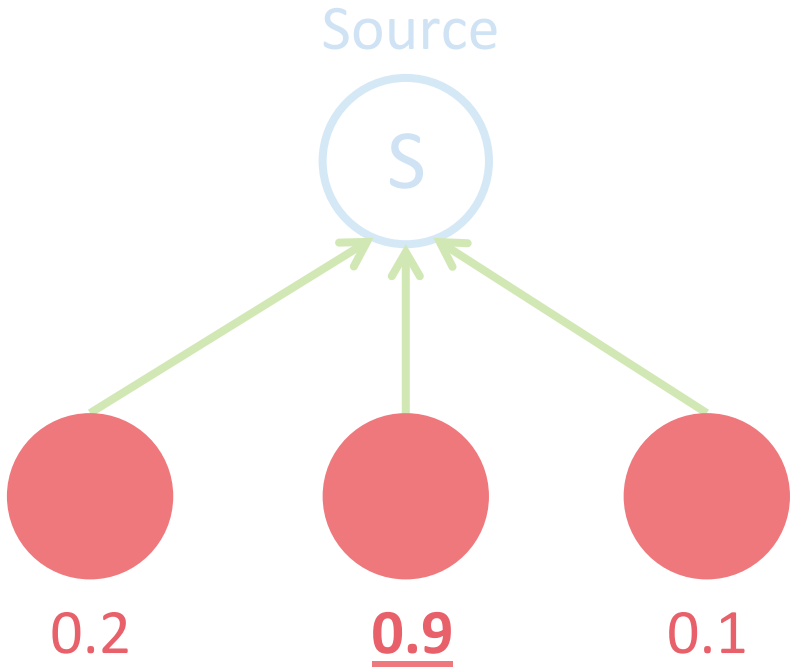
3. Calculate ***early adopter score*** from the number of link imitation.



ROADMAP

4. Calculate ***all users***' early adopter score.

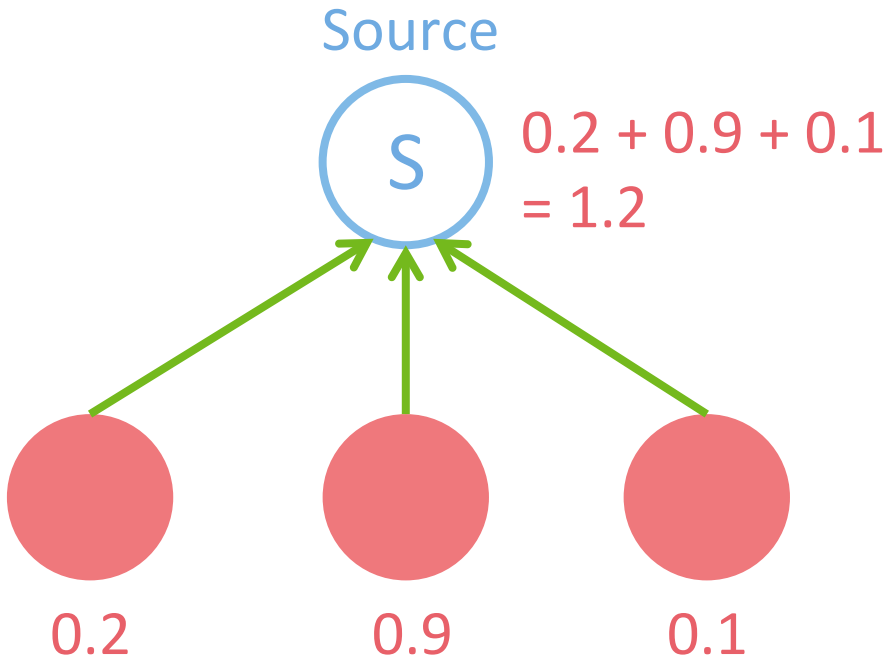
5. Calculate ***future popularity score*** from followers' early adopter score.



ROADMAP

4. Calculate ***all users***' early adopter score.

5. Calculate ***future popularity score*** from followers' early adopter score.

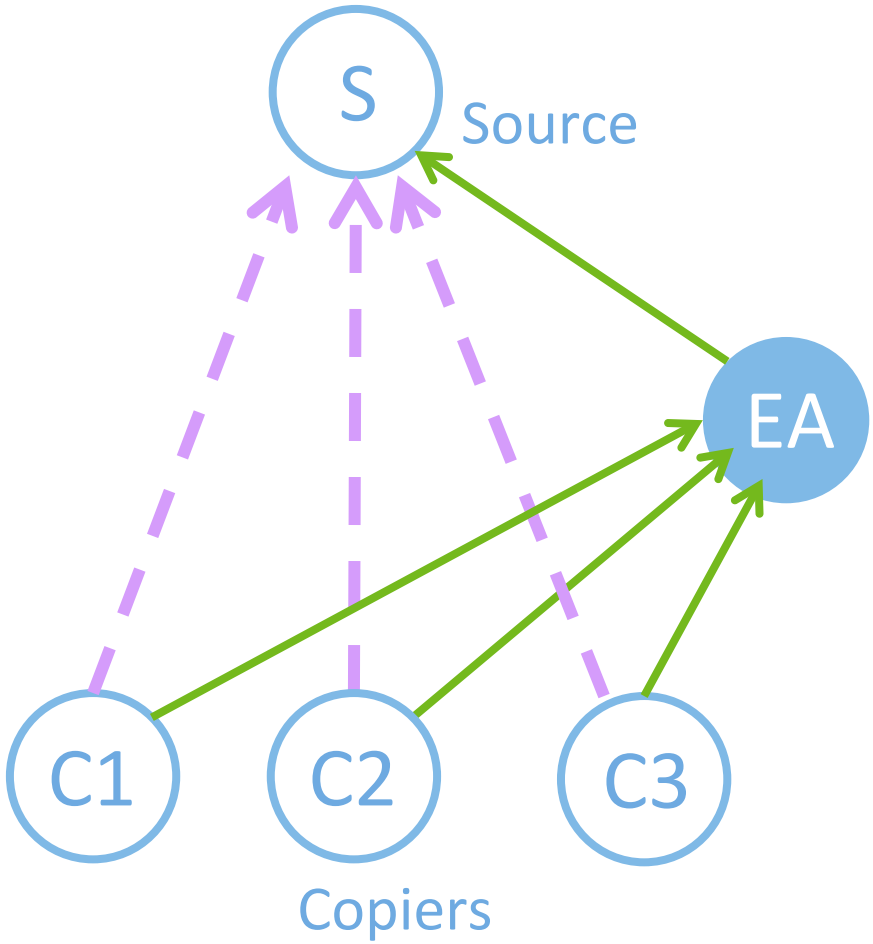


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

3. Calculate ***early adopter score*** from the number of link imitation.

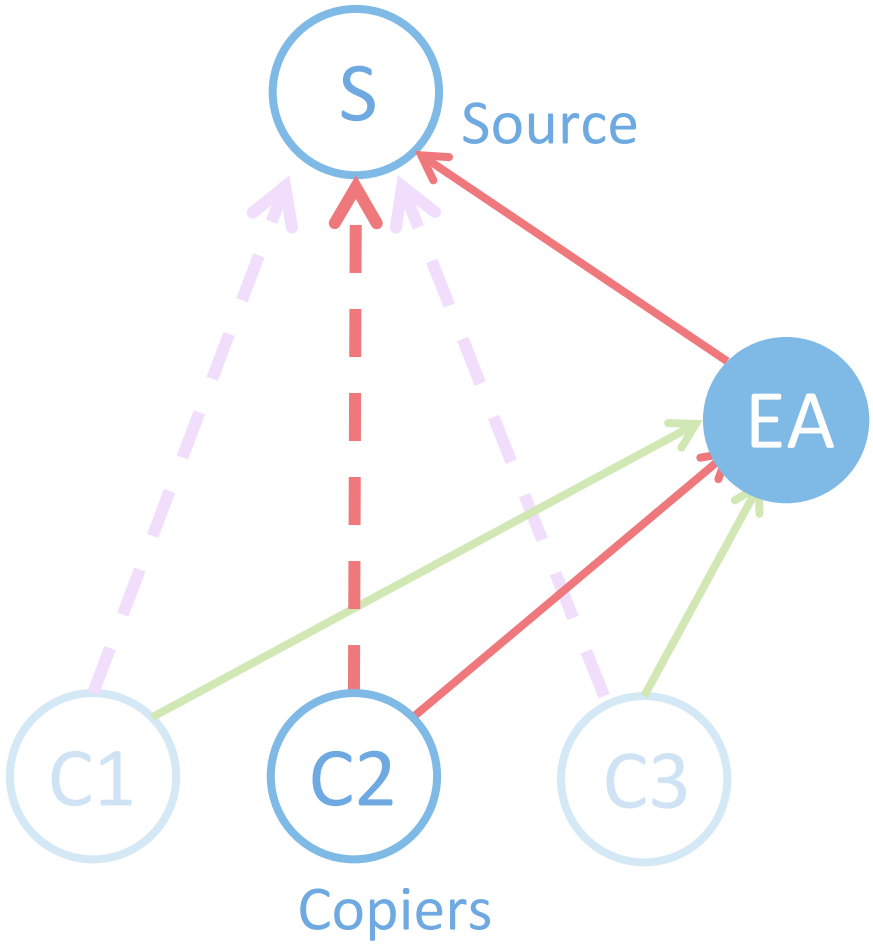


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

3. Calculate ***early adopter score*** from the number of link imitation.



DETECTION OF LINK IMITATION

We cannot immediately know
imitation of follow links.

DETECTION OF LINK IMITATION

We cannot immediately know imitation of follow links.

Structures

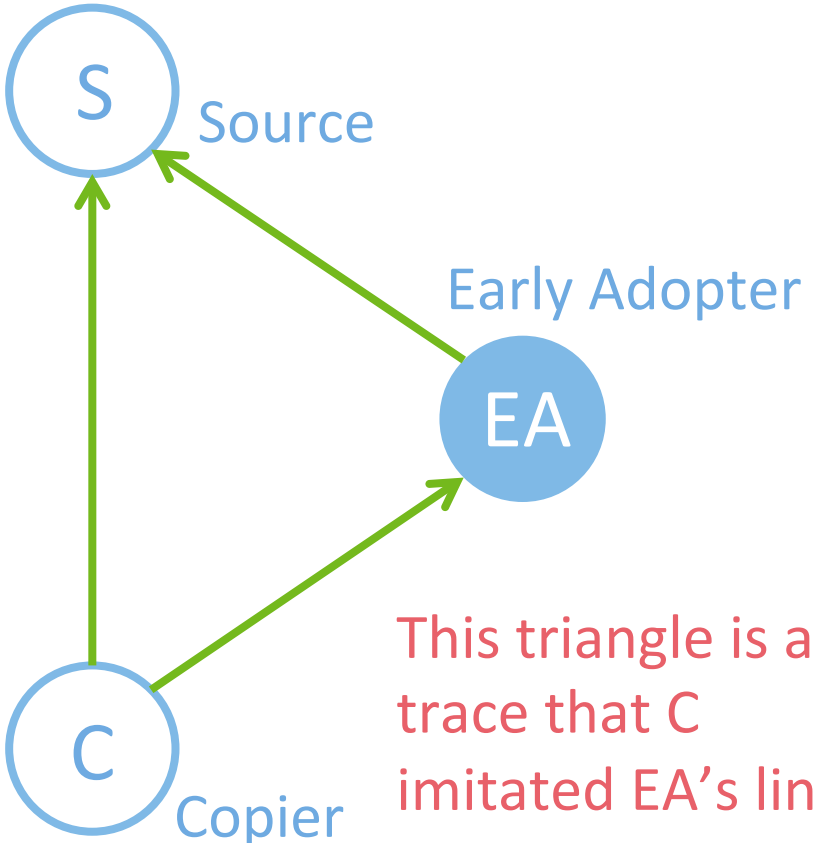
- triangle
- Non-reciprocity

DETECTION OF LINK IMITATION

We cannot immediately know imitation of follow links.

Structures

- triangle
- Non-reciprocity

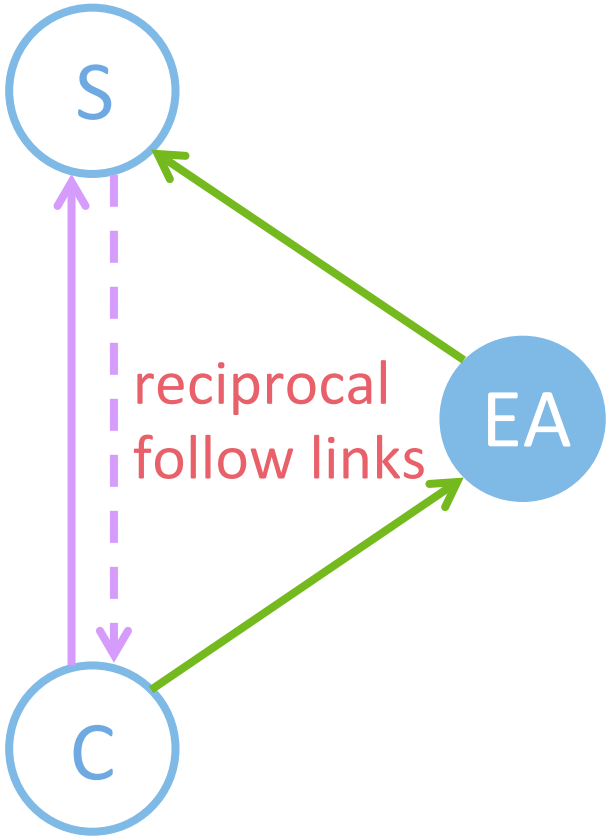


DETECTION OF LINK IMITATION

We cannot immediately know imitation of follow links.

Structures

- triangle
- Non-reciprocity



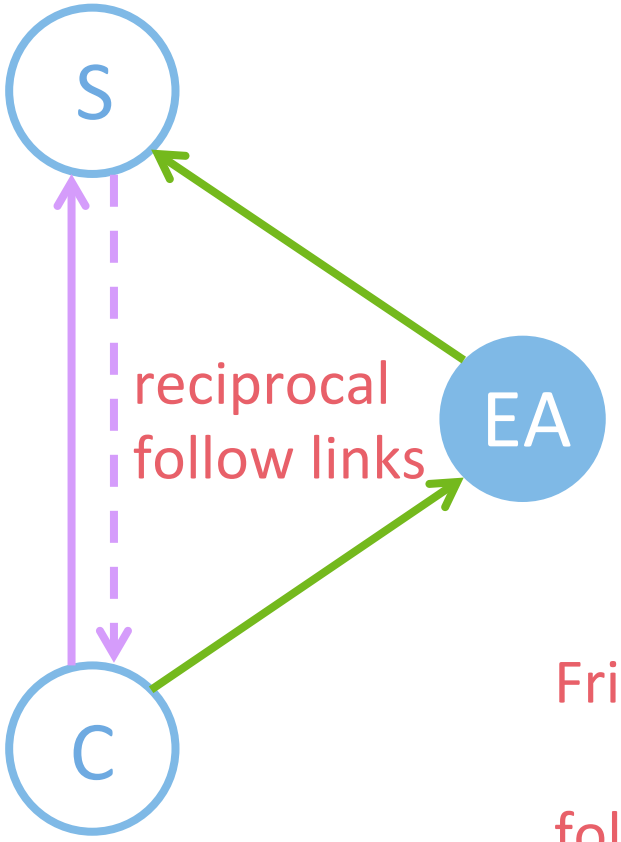
DETECTION OF LINK IMITATION

We cannot immediately know imitation of follow links.

Structures

- triangle
- Non-reciprocity

It is important whether this link of triangle is reciprocal or non-reciprocal.



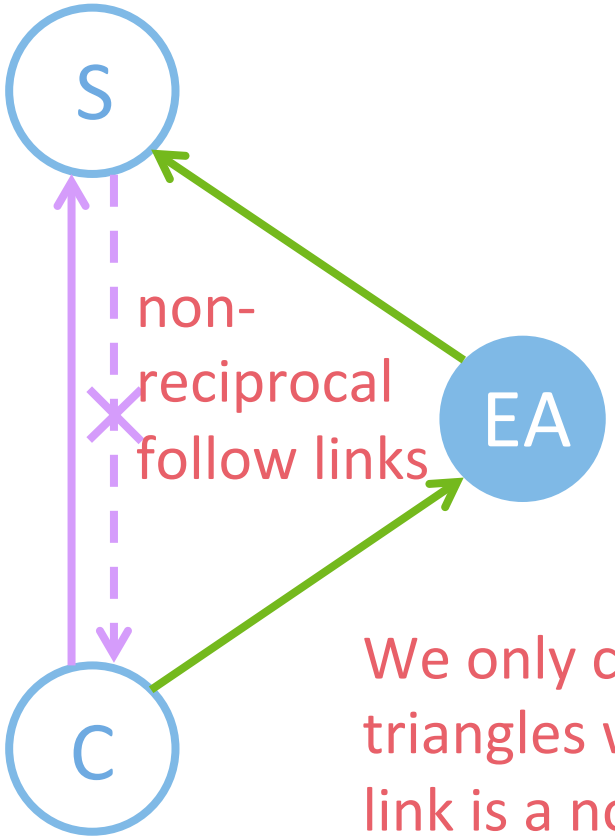
Friendship
or
followback

DETECTION OF LINK IMITATION

We cannot immediately know imitation of follow links.

Structures

- triangle
- Non-reciprocity



We only count triangles where the link is a non-reciprocal.

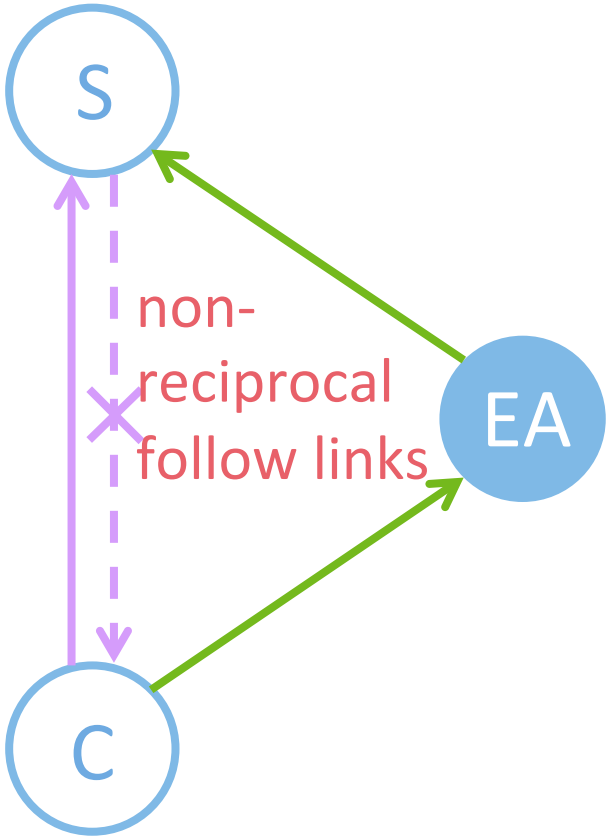
DETECTION OF LINK IMITATION

We cannot immediately know imitation of follow links.

Structures

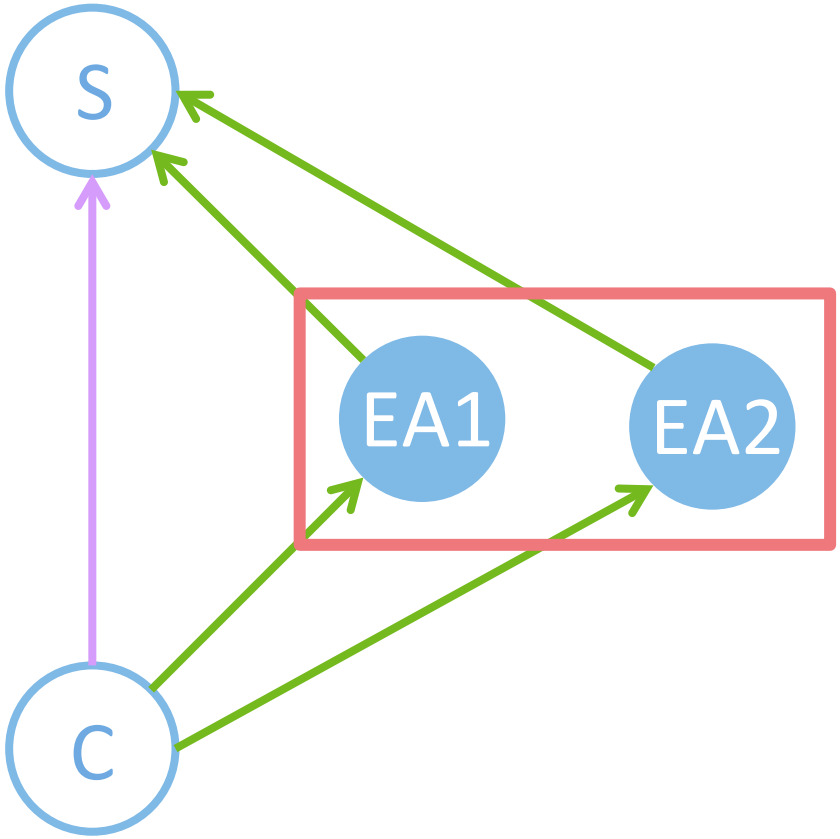
- triangle
- Non-reciprocity

We can detect links created through imitation



WHEN THERE ARE MULTIPLE CANDIDATES

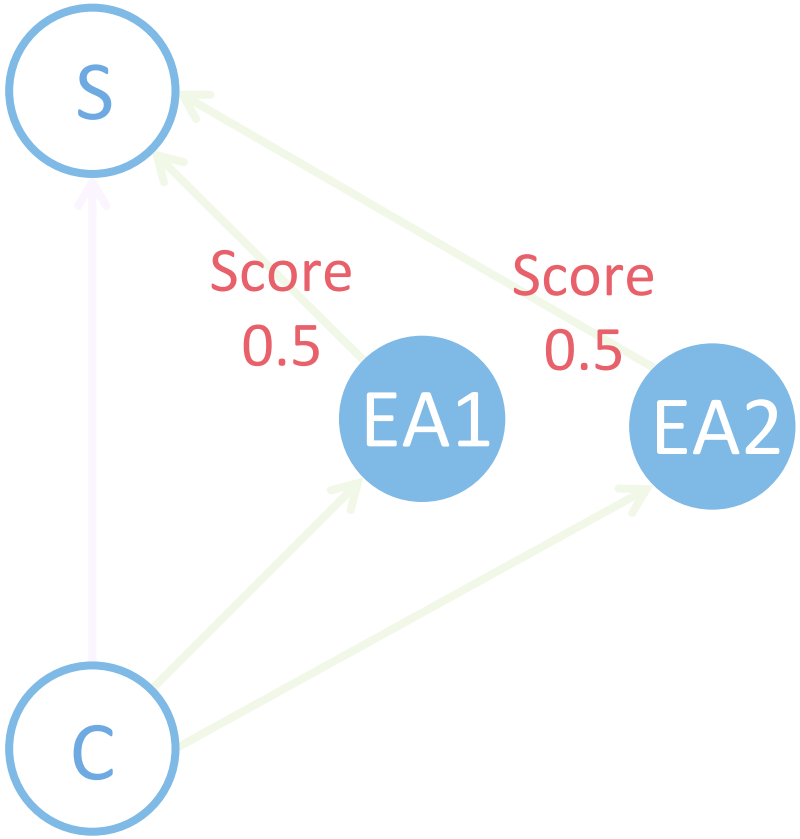
However, in the right figure, it is difficult to determine which was imitated by C, EA1 or EA2



WHEN THERE ARE MULTIPLE CANDIDATES

However, in the right figure, it is difficult to determine which was imitated by C, EA1 or EA2

Each candidates are given a score equally.

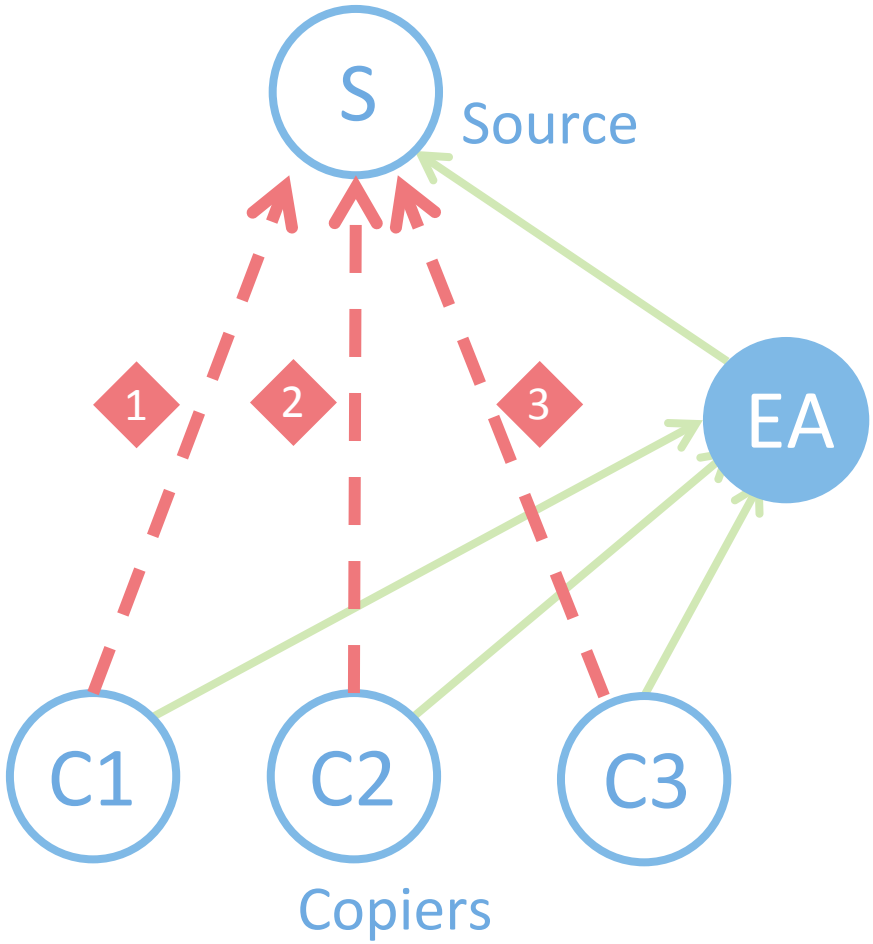


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

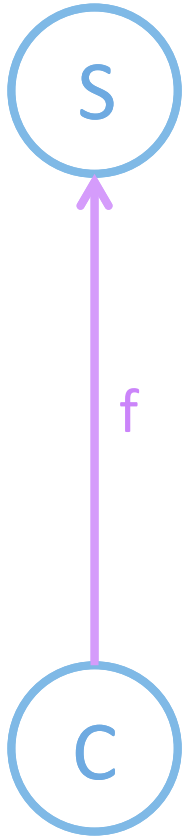
3. Calculate ***early adopter score*** from the number of link imitation.



HOW TO COUNT NUMBER OF IMITATION

Our method process all edges in the graph one by one.

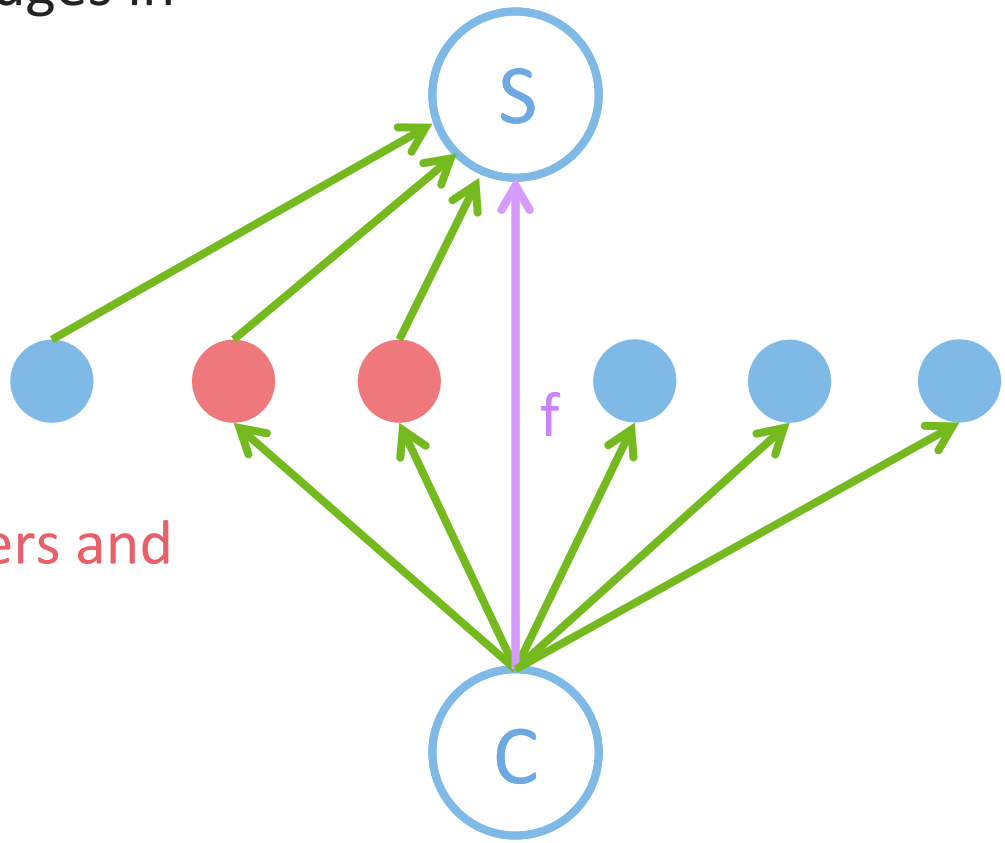
Each follow link f in a graph



HOW TO COUNT NUMBER OF IMITATION

Our method process all edges in the graph one by one.

Each follow link f in a graph

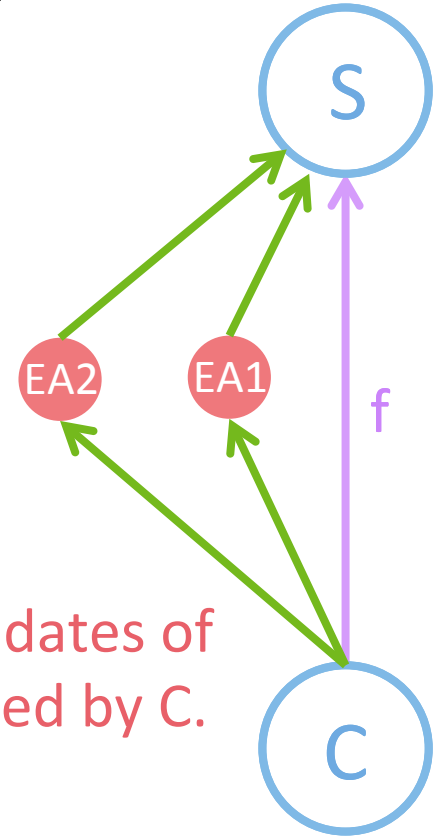


Intersection of S's followers and C's followees

HOW TO COUNT NUMBER OF IMITATION

Our method process all edges in the graph one by one.

Each follow link f in a graph



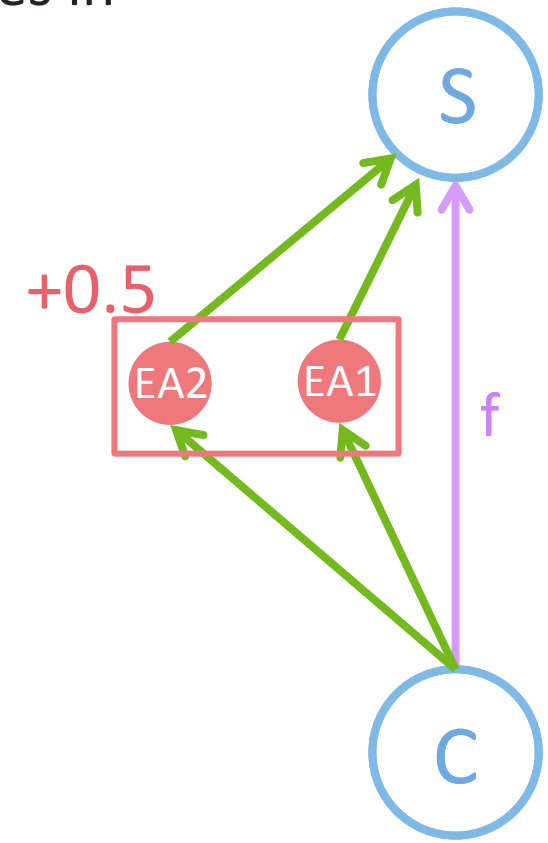
These intersection users are candidates of users whose links to S were imitated by C.

HOW TO COUNT NUMBER OF IMITATION

Our method process all edges in the graph one by one.

Accumulating the score for each candidate

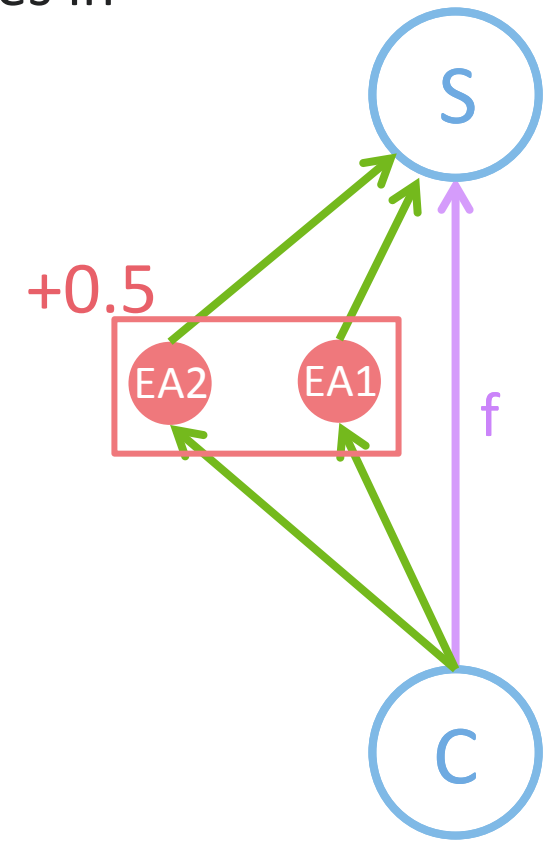
Each follow link f in a graph



HOW TO COUNT NUMBER OF IMITATION

Our method process all edges in the graph one by one.

Each follow link f in a graph



Accumulating the score for each candidate

EA1's accumulated score = 0.5

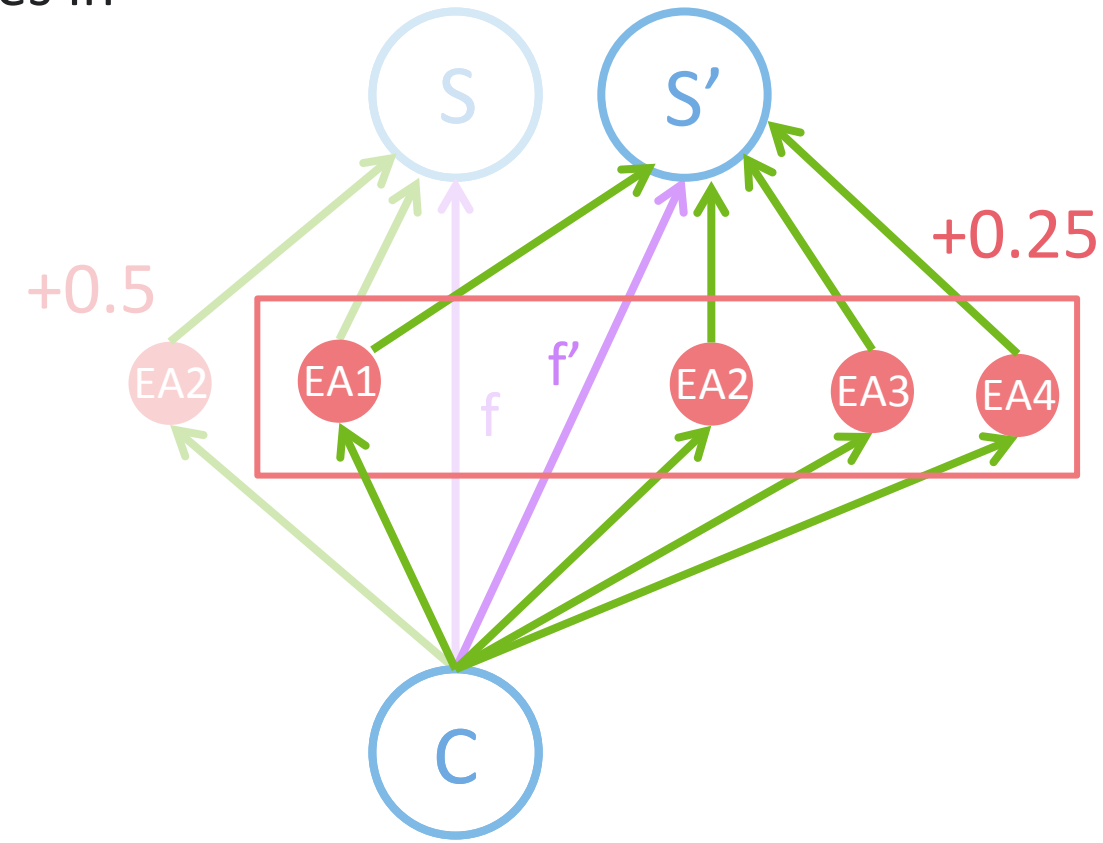
HOW TO COUNT NUMBER OF IMITATION

Our method process all edges in the graph one by one.

Accumulating the score for each candidate

EA1's accumulated score
= 0.5 + 0.25
= 0.75

Each follow link f in a graph

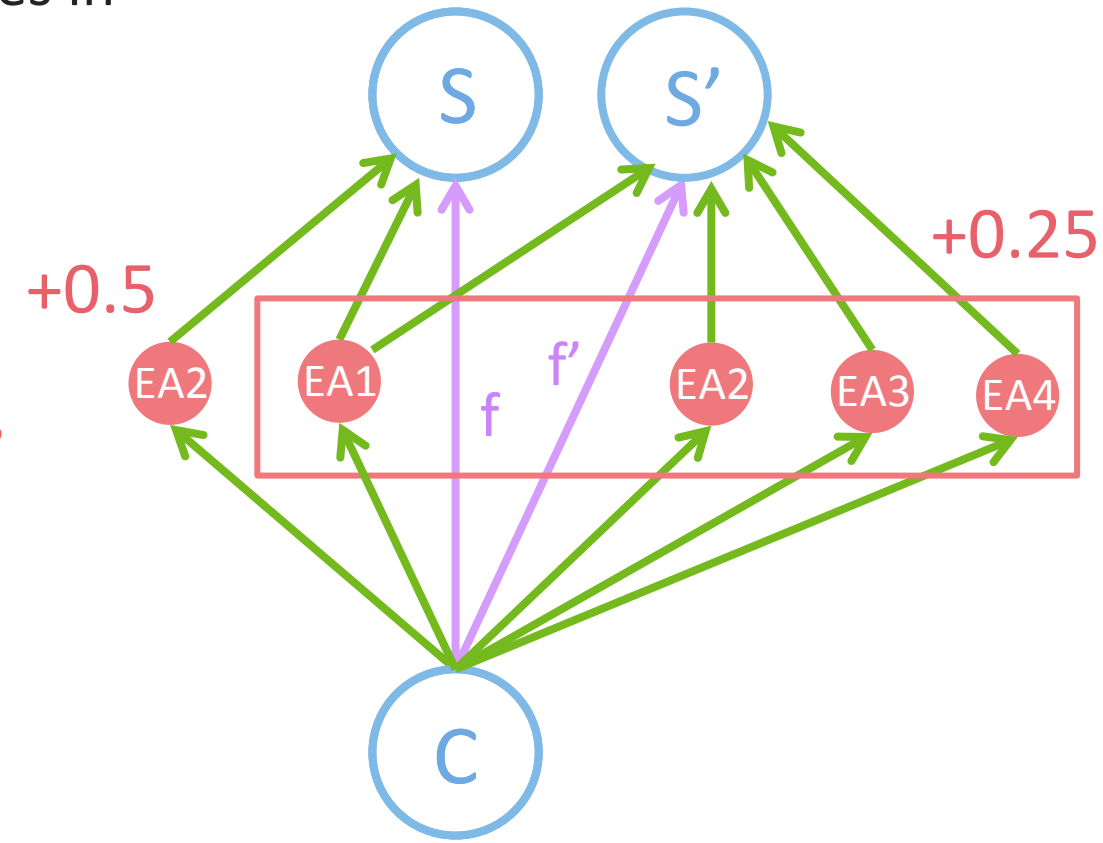


HOW TO COUNT NUMBER OF IMITATION

Our method process all edges in the graph one by one.

After processing all links in the graph, the scores accumulated to each user is the expected number of times that user has been imitated.

Each follow link f in a graph

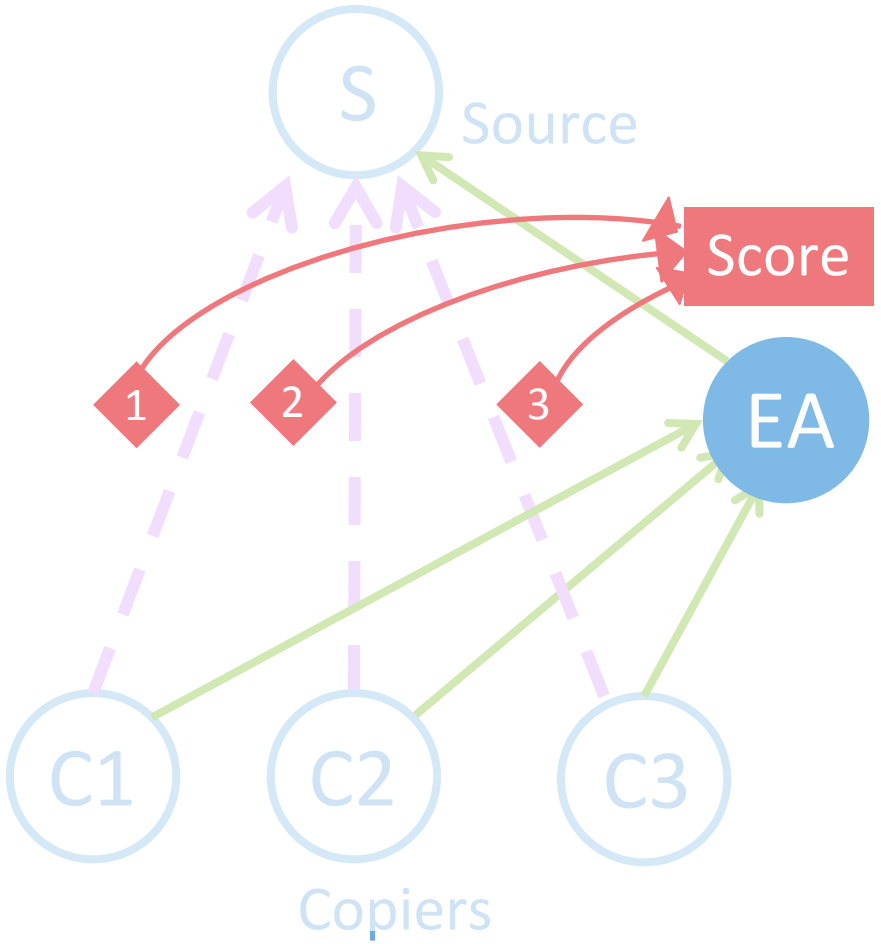


ROADMAP

1. Detect links created through imitation.

2. Count number of link imitation.

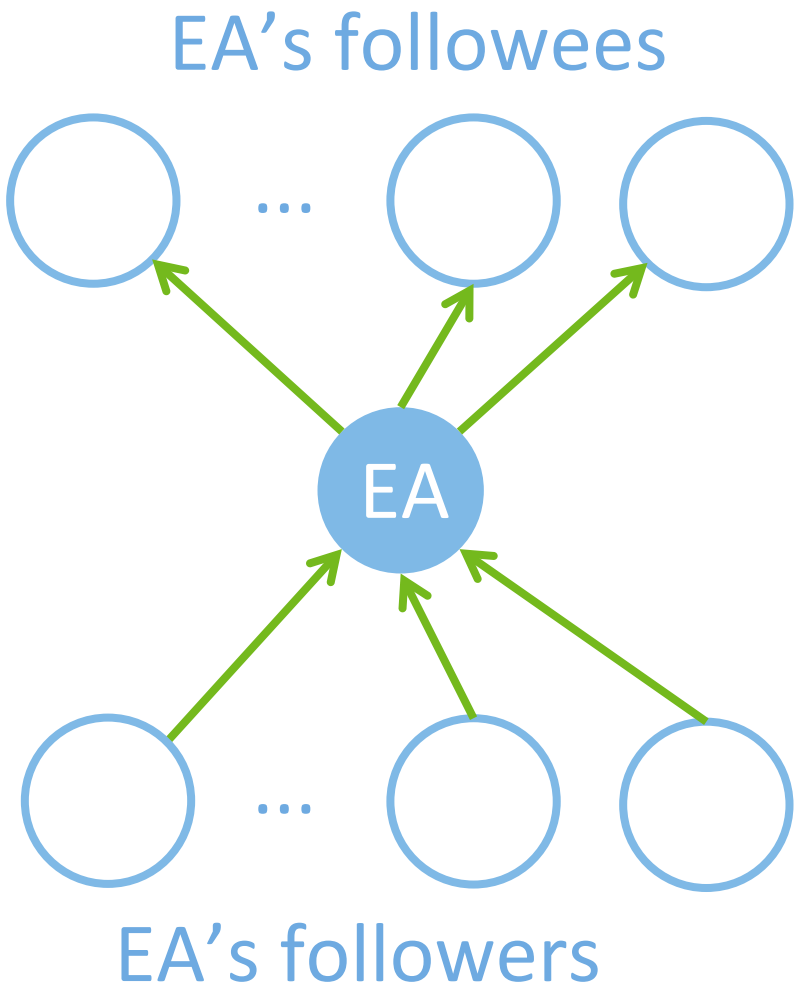
3. Calculate ***early adopter score*** from the number of link imitation.



EARLY ADOPTER SCORE

Early adopter score

$$= \frac{\text{Number of link imitation}}{|\text{followees}| \times |\text{followers}|}$$

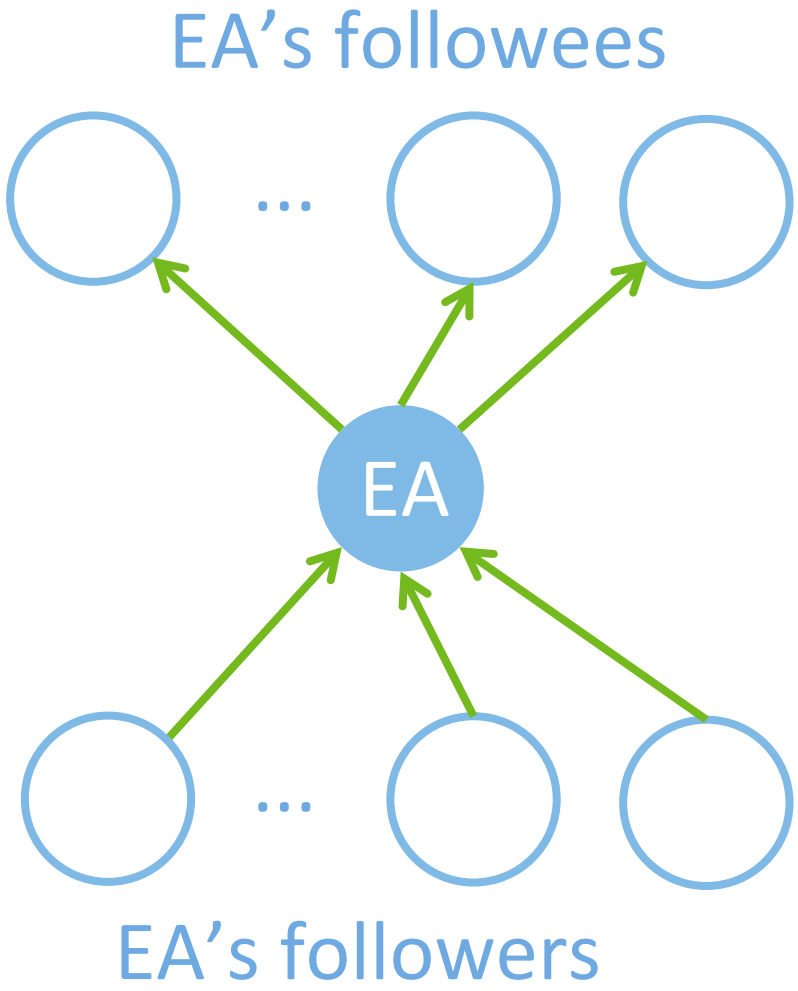


EARLY ADOPTER SCORE

Early adopter score

$$= \frac{\text{Number of link imitation}}{|\text{followees}| \times |\text{followers}|}$$

This denominator corresponds to the maximum number of times this user could be imitated.

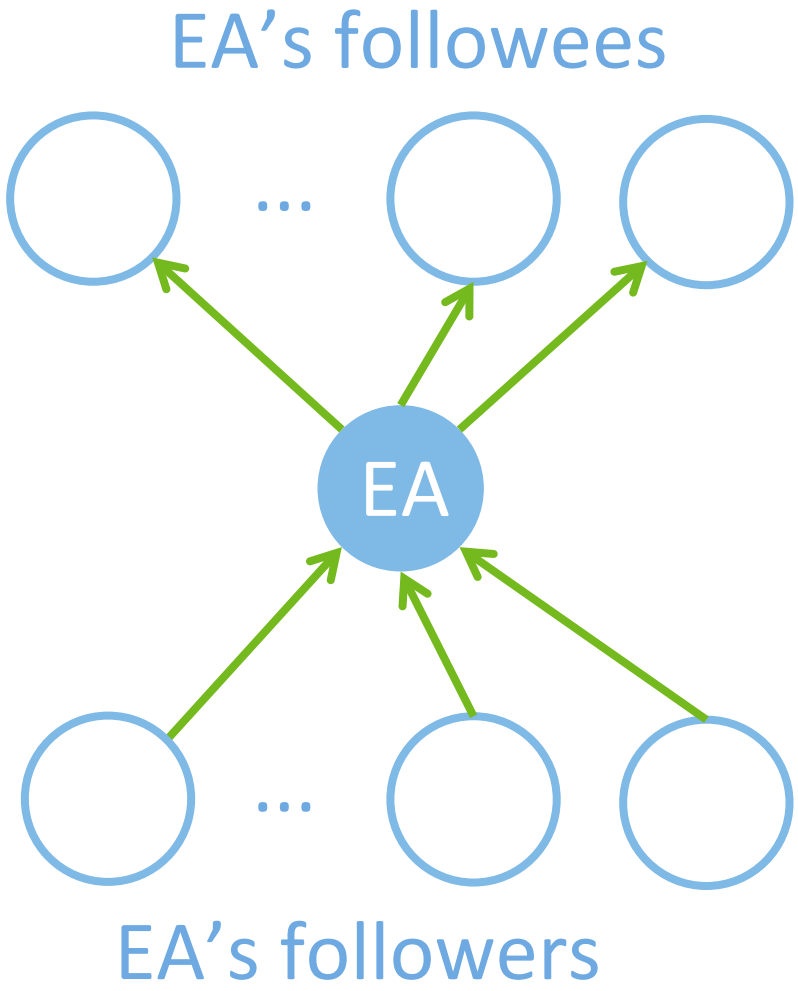


EARLY ADOPTER SCORE

Early adopter score

$$= \frac{\text{Number of link imitation}}{|\text{followees}| \times |\text{followers}|}$$

This fraction corresponds to the imitation ratio of this user.

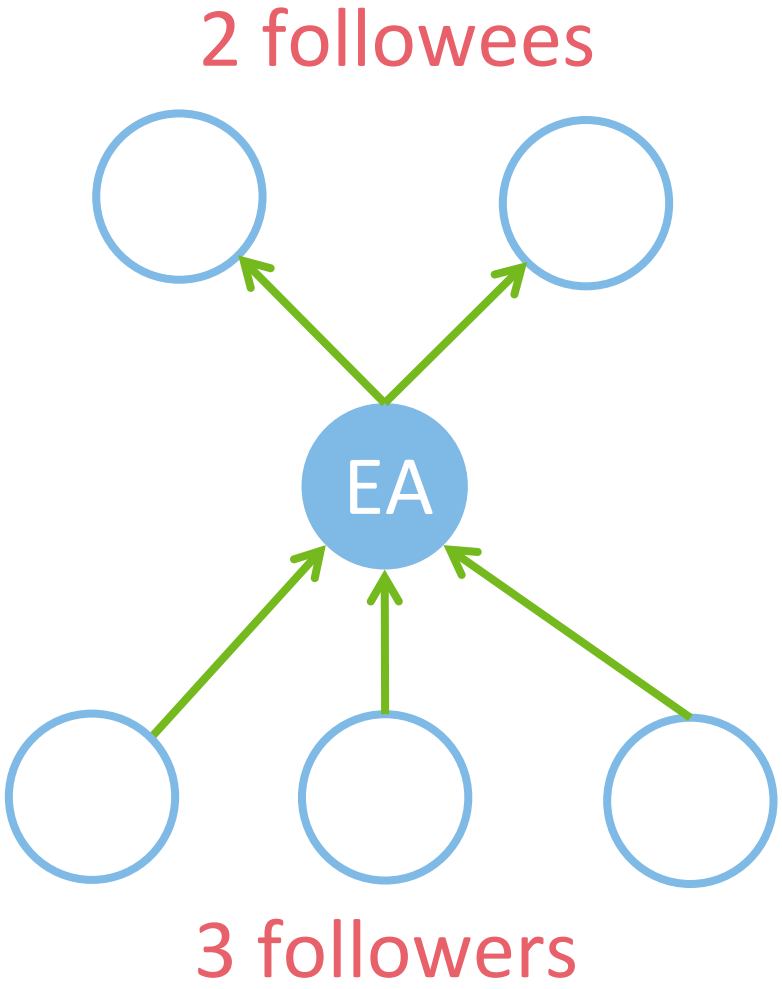


EARLY ADOPTER SCORE

Early adopter score

$$= \frac{\text{Number of link imitation}}{|\text{followees}| \times |\text{followers}|}$$

$$= \frac{\quad}{2 \times 3} = \frac{\quad}{6}$$

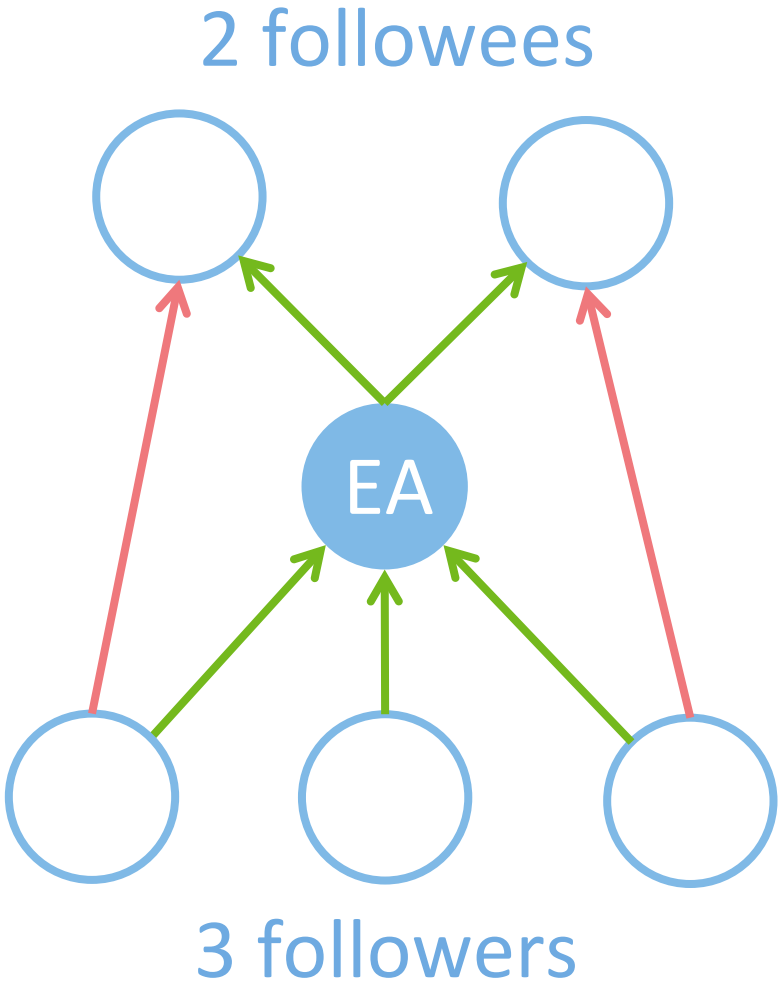


EARLY ADOPTER SCORE

Early adopter score

$$= \frac{\text{Number of link imitation}}{|\text{followees}| \times |\text{followers}|}$$

$$= \frac{2}{2 \times 3} = \frac{2}{6}$$

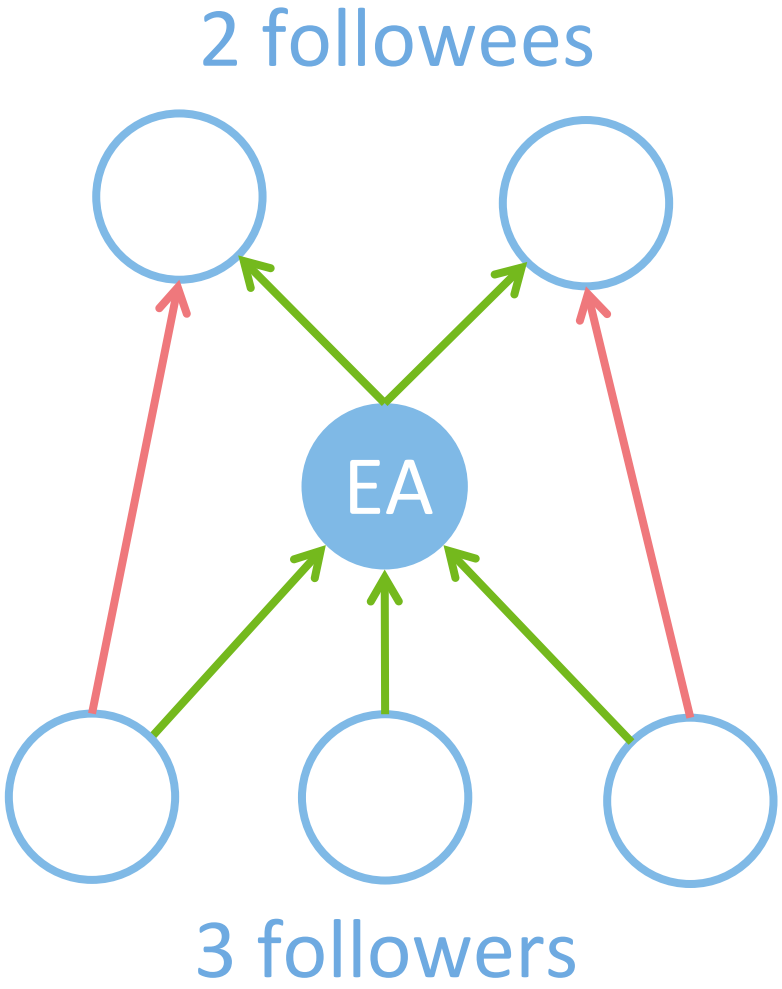


EARLY ADOPTER SCORE

Early adopter score

$$= \frac{\text{Number of link imitation}}{|\text{followees}| \times |\text{followers}|}$$

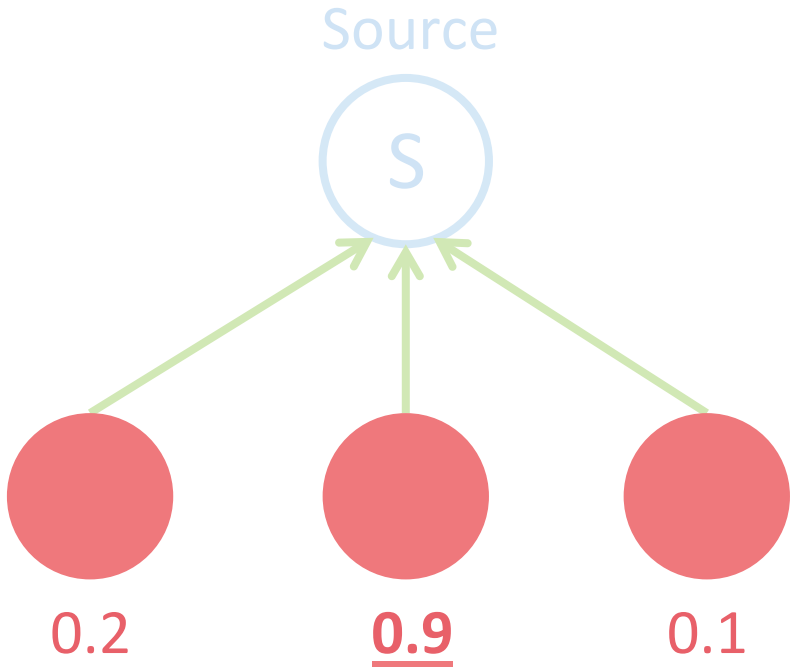
$$= \frac{2}{2 \times 3} = \frac{2}{6} = \frac{1}{3}$$



ROADMAP

4. Calculate ***all users***' early adopter score.

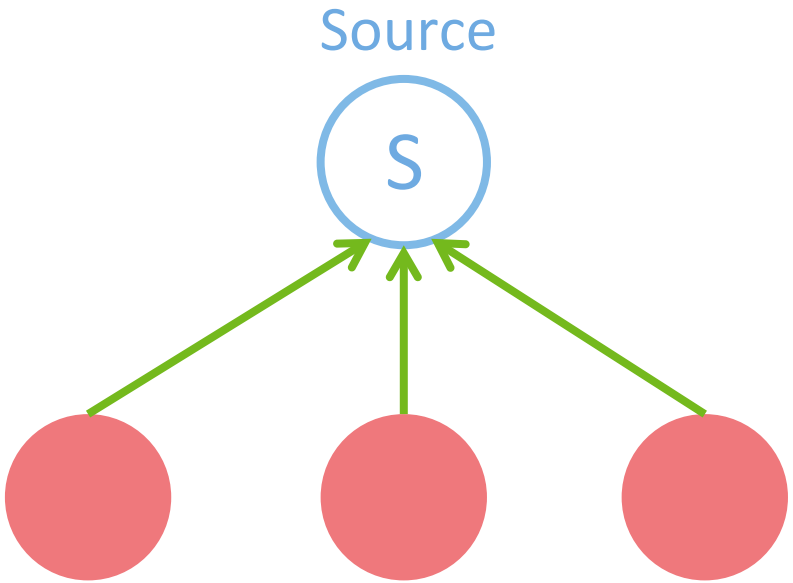
5. Calculate ***future popularity score*** from followers' early adopter score.



ROADMAP

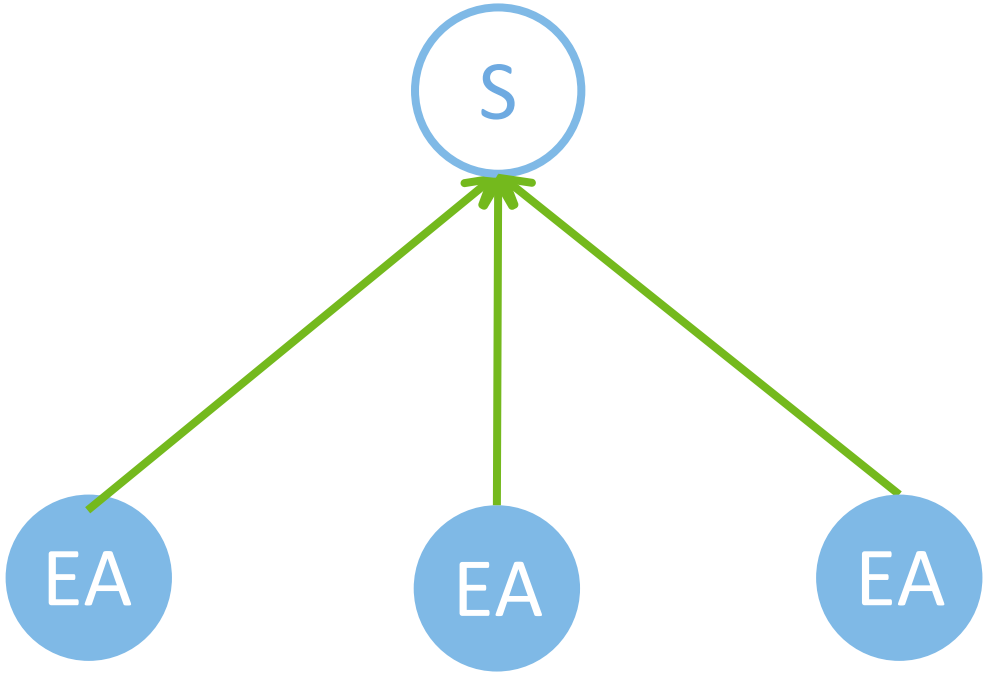
4. Calculate ***all users***' early adopter score.

5. Calculate ***future popularity score*** from followers' early adopter score.



FUTURE POPULARITY SCORE

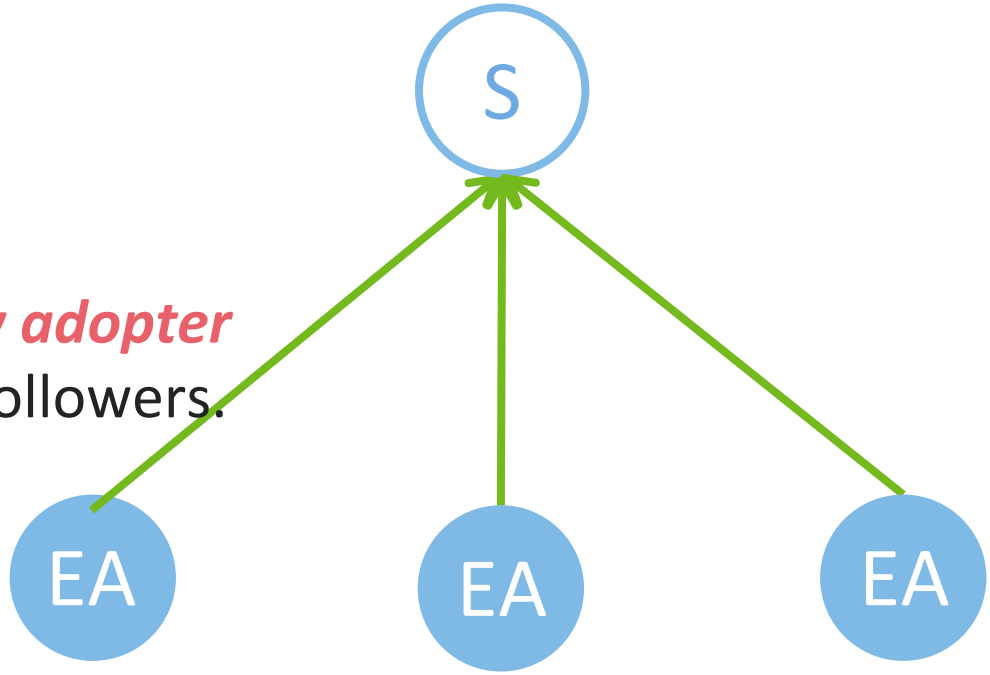
We call the estimated new user's prospective popularity ***future popularity score***.



FUTURE POPULARITY SCORE

We call the estimated new user's prospective popularity ***future popularity score***

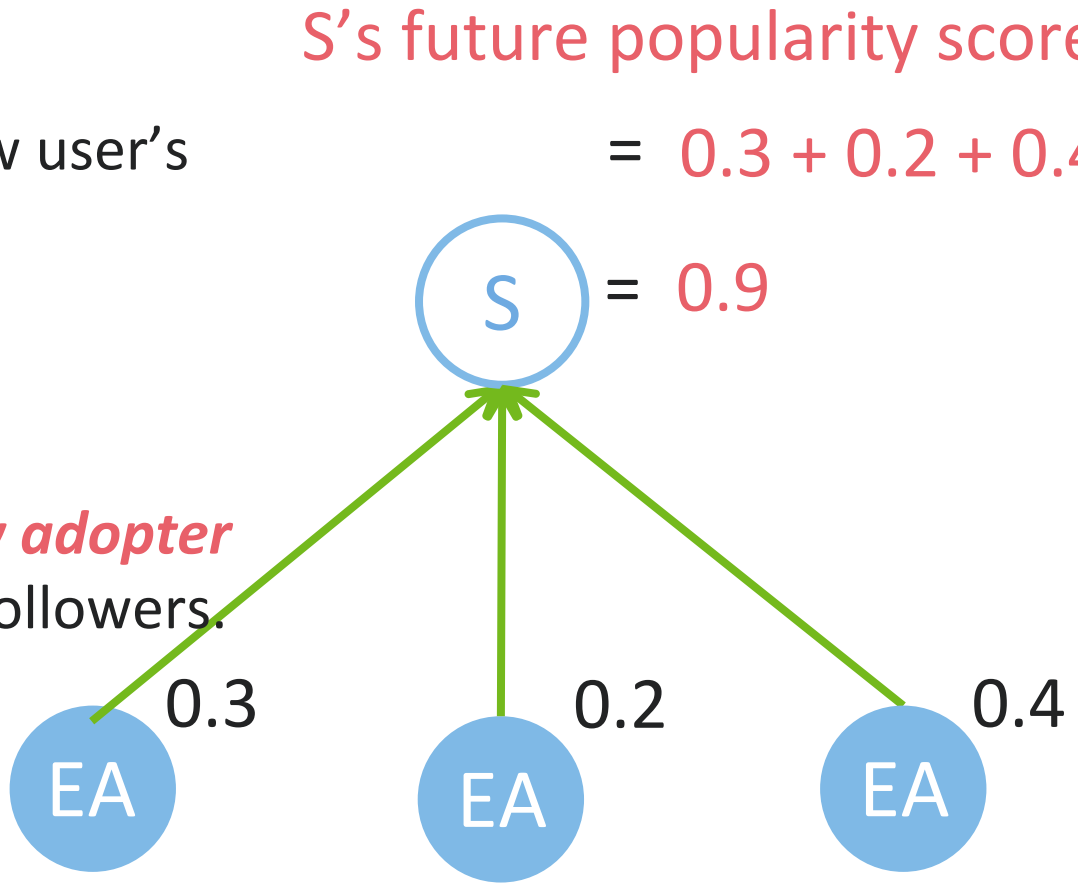
Future popularity score is computed as ***sum of early adopter scores*** of the new user's followers.



FUTURE POPULARITY SCORE

We call the estimated new user's prospective popularity ***future popularity score***

Future popularity score is computed as ***sum of early adopter scores*** of the new user's followers.



OUR EXPERIMENT

- Dataset [Li et al., KDD 2012]
  - A sub-graph of Twitter crawled in 2011
  - About 20,000,000 users
  - About 300,000,000 follow links
- Target users
  - T_n^w: we select then-new users that are
    - within *w* weeks after the creation, and
    - have more than *n* followers

OUR EXPERIMENT

- Evaluation
  - We rank users in T_n^w by our methods and baselines.
  - Ground truth: we rank users by their number of non-reciprocal followers as of 2015.
  - Compute Spearman's ρ

OUR EXPERIMENT

- Baseline methods
  - FW: Number of followers in May 2011
  - PR_{nr}: PageRank scores on the graph consisting only of non-reciprocal links
  - $HITS_{nr}$: HITS scores on the graph consisting only of non-reciprocal links
  - AD: Adamic-Adar index
- Our methods
  - FPS: Feature popularity score
  - LR: The linear regression of FPS and some baselines

OUR EXPERIMENT

Method	T^4_{10}	T^2_{10}	T^4_{20}	T^2_{20}	T^4_{30}	T^2_{30}
data size	6921	1515	2259	431	979	165
FW	0.18	0.23	0.15	0.07	0.19	0.00
$HITS_{nr}$	0.26	0.31	0.30	0.35	0.38	0.46
PR_{nr}	0.16	0.09	0.21	0.20	0.30	0.32
AD	-0.21	-0.13	-0.30	-0.46	-0.27	-0.50
FPS	0.39	0.41	0.39	0.45	0.40	0.47
LR	0.43	0.46	0.43	0.50	0.45	0.58

OUR EXPERIMENT

Method	T^4_{10}	T^2_{10}	T^4_{20}	T^2_{20}	T^4_{30}	T^2_{30}
data size	6921	1515	2259	431	979	165
FW	0.18	0.23	0.15	0.07	0.19	0.00
$HITS_{nr}$	0.26	0.31	0.30	0.35	0.38	0.46
PR_{nr}	0.16	0.09	0.21	0.20	0.30	0.32
AD	-0.21	-0.13	-0.30	-0.46	-0.27	-0.50
FPS	0.39	0.41	0.39	0.45	0.40	0.47
LR	0.43	0.46	0.43	0.50	0.45	0.58

green: best within baselines

- HITS works best in most cases.
- AD is the best in some cases.

OUR EXPERIMENT

Method	T^4_{10}	T^2_{10}	T^4_{20}	T^2_{20}	T^4_{30}	T^2_{30}
data size	6921	1515	2259	431	979	165
FW	0.18	0.23	0.15	0.07	0.19	0.00
$HITS_{nr}$	0.26	0.31	0.30	0.35	0.38	0.46
PR_{nr}	0.16	0.09	0.21	0.20	0.30	0.32
AD	-0.21	-0.13	-0.30	-0.46	-0.27	-0.50
FPS	0.39	0.41	0.39	0.45	0.40	0.47
LR	0.43	0.46	0.43	0.50	0.45	0.58

- FPS is the best in most cases among all the methods excluding LR
- LR is the best for all cases. It means that FPS captures some aspects that are not captured by other methods.

red: best
blue: best excluding LR

CONCLUSION

- We proposed a method of estimating prospective popularity of new users.
- Our method estimate it through the discovery of early adopters.
- Experiment by using sub-graph of Twitter.
- Our method outperforms baselines.