Carnegie Mellon University

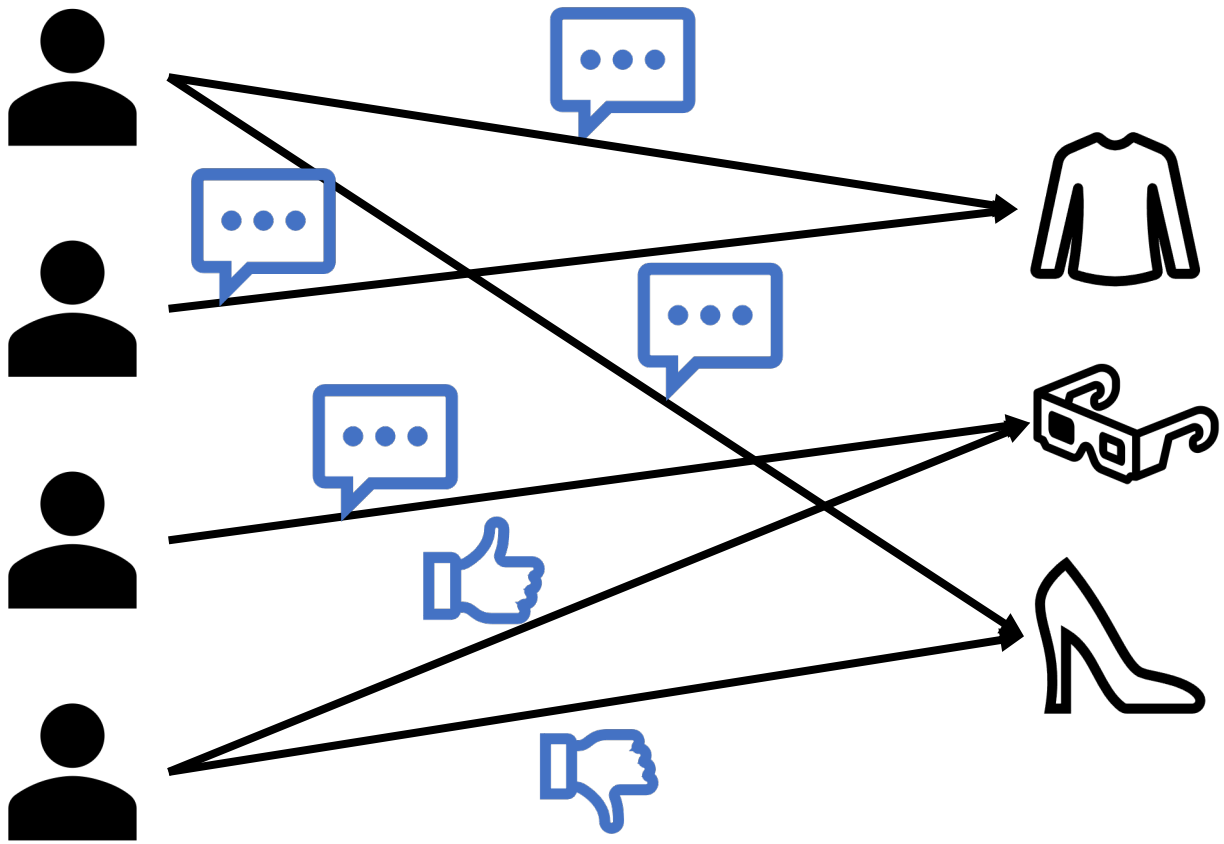
ECML PKDD 2021

Graph Fraud Detection based on Accessibility Score Distributions

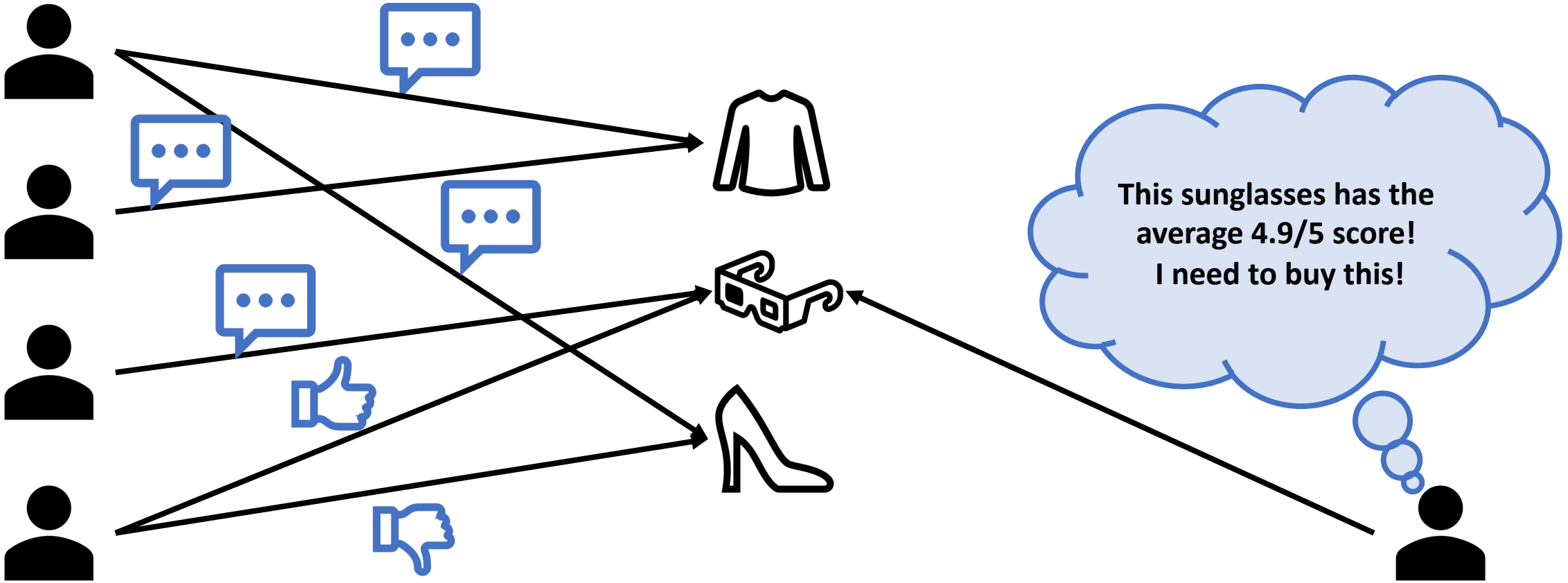
Minji Yoon

Carnegie Mellon University

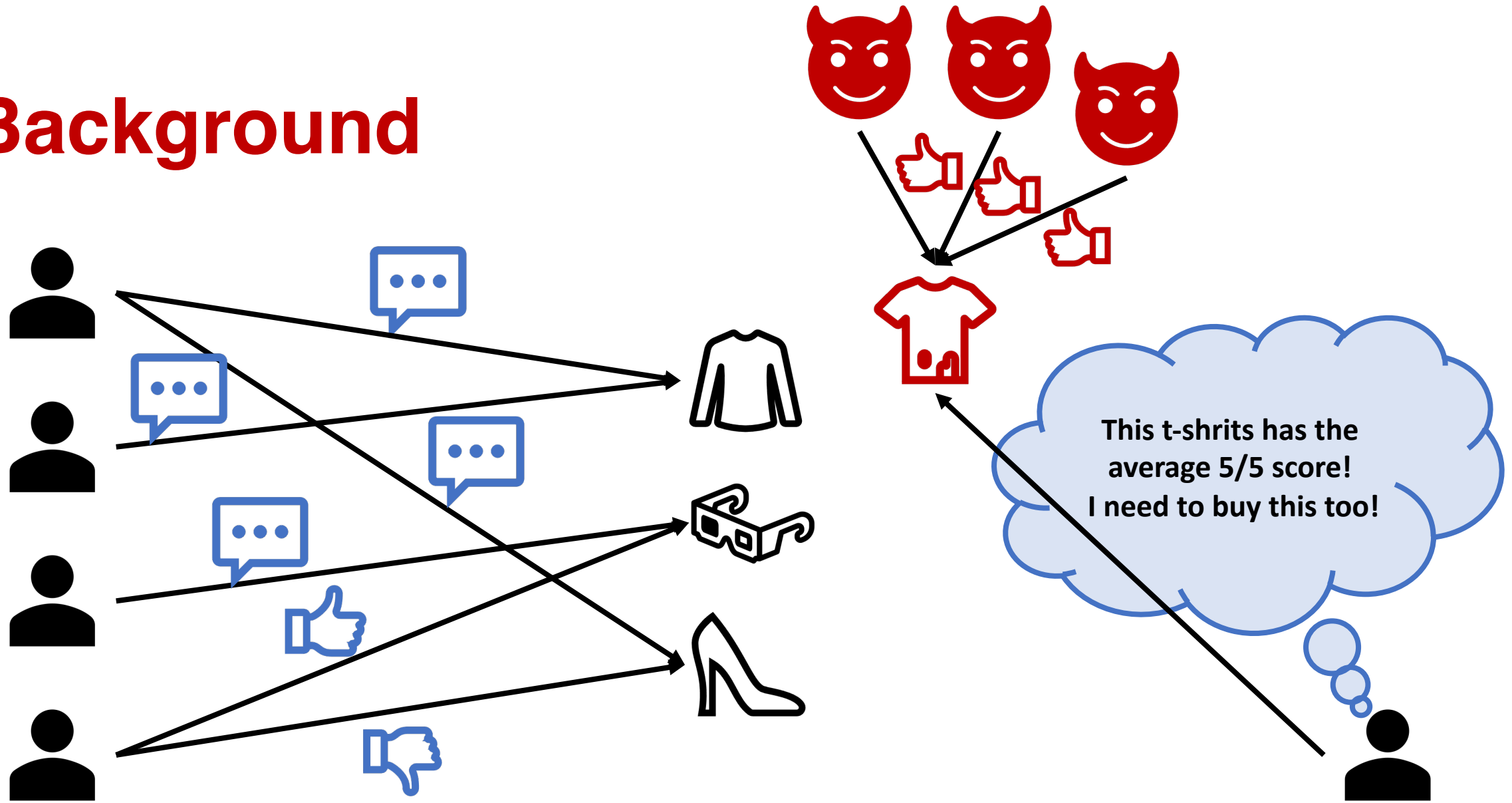
Background



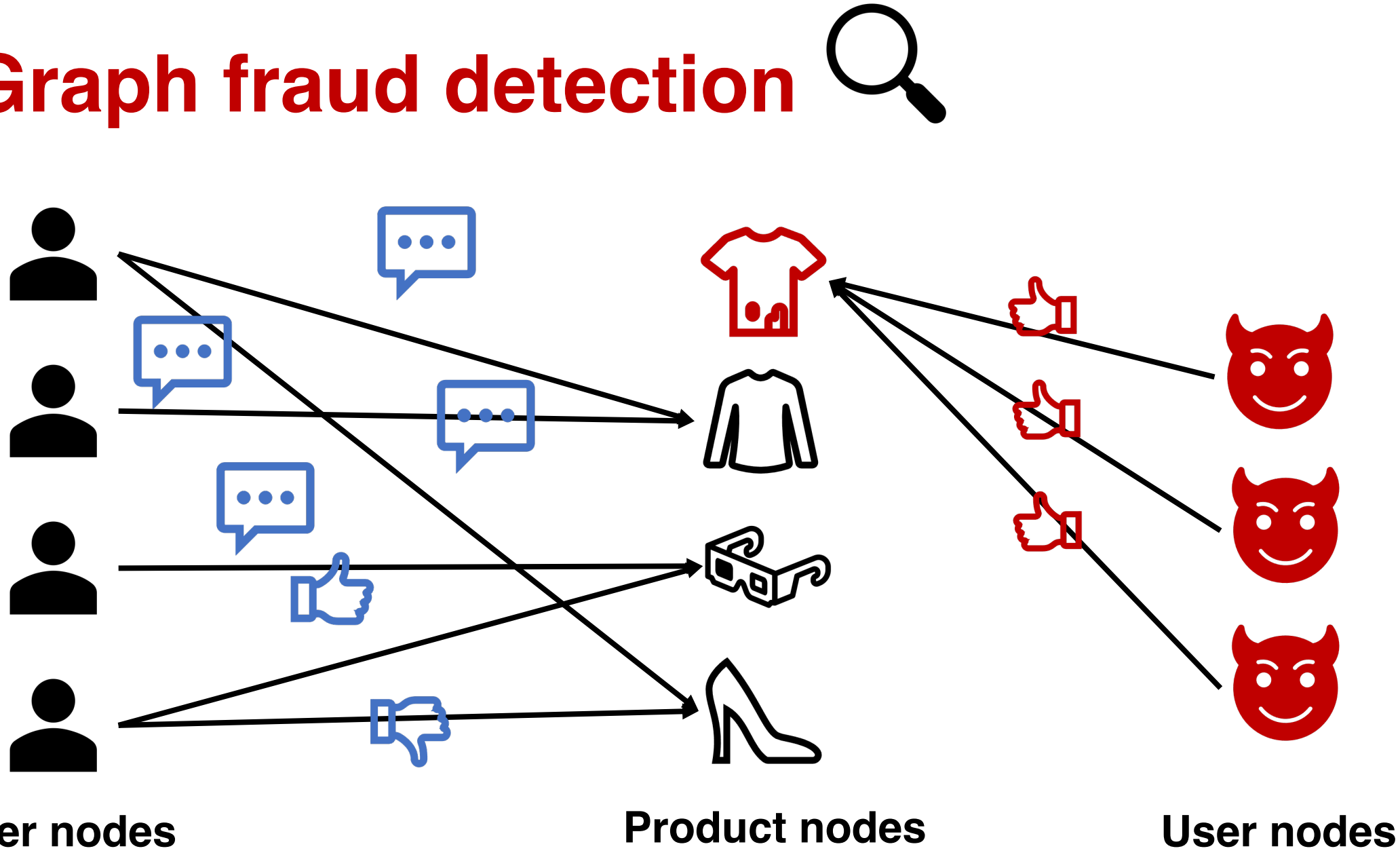
Background



Background

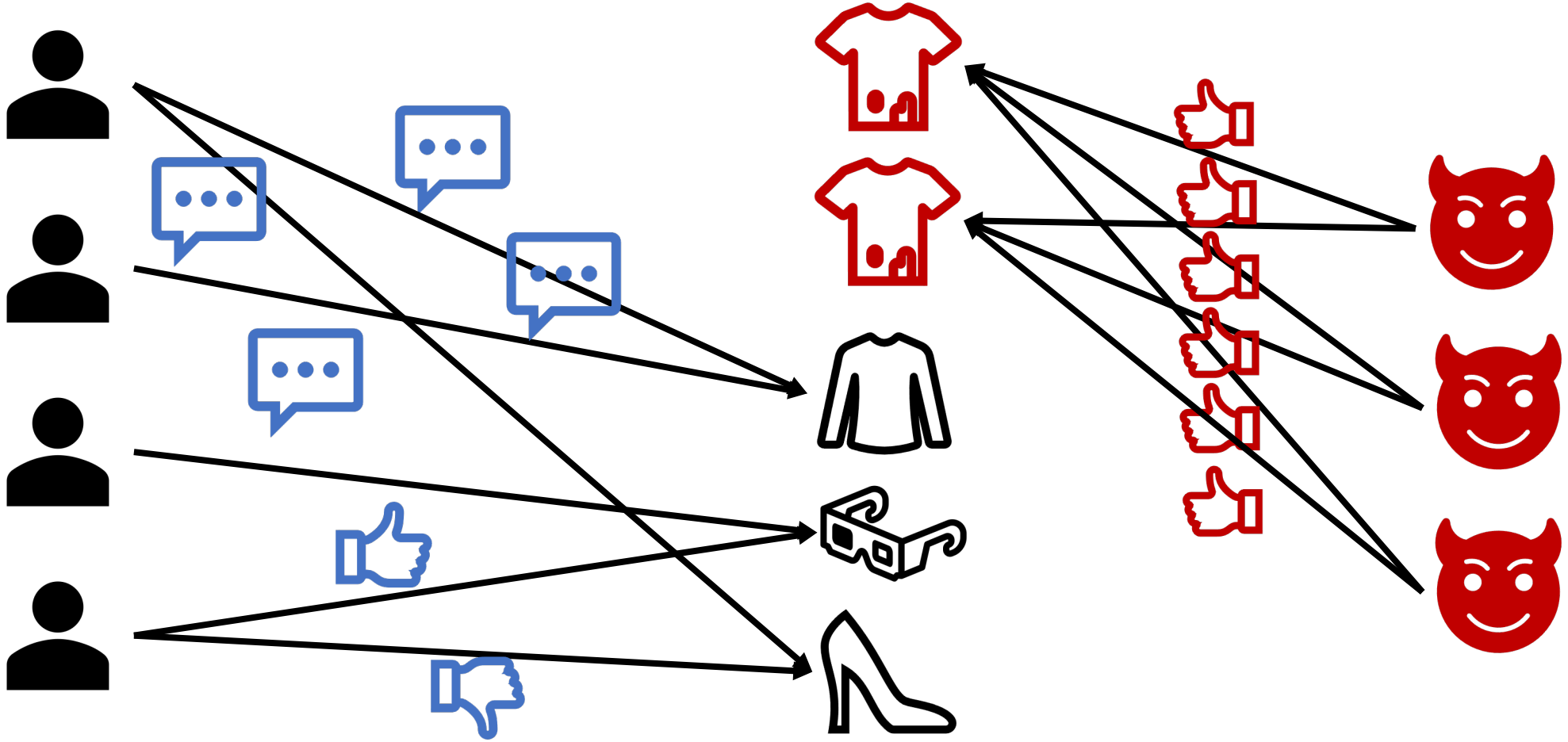


Graph fraud detection



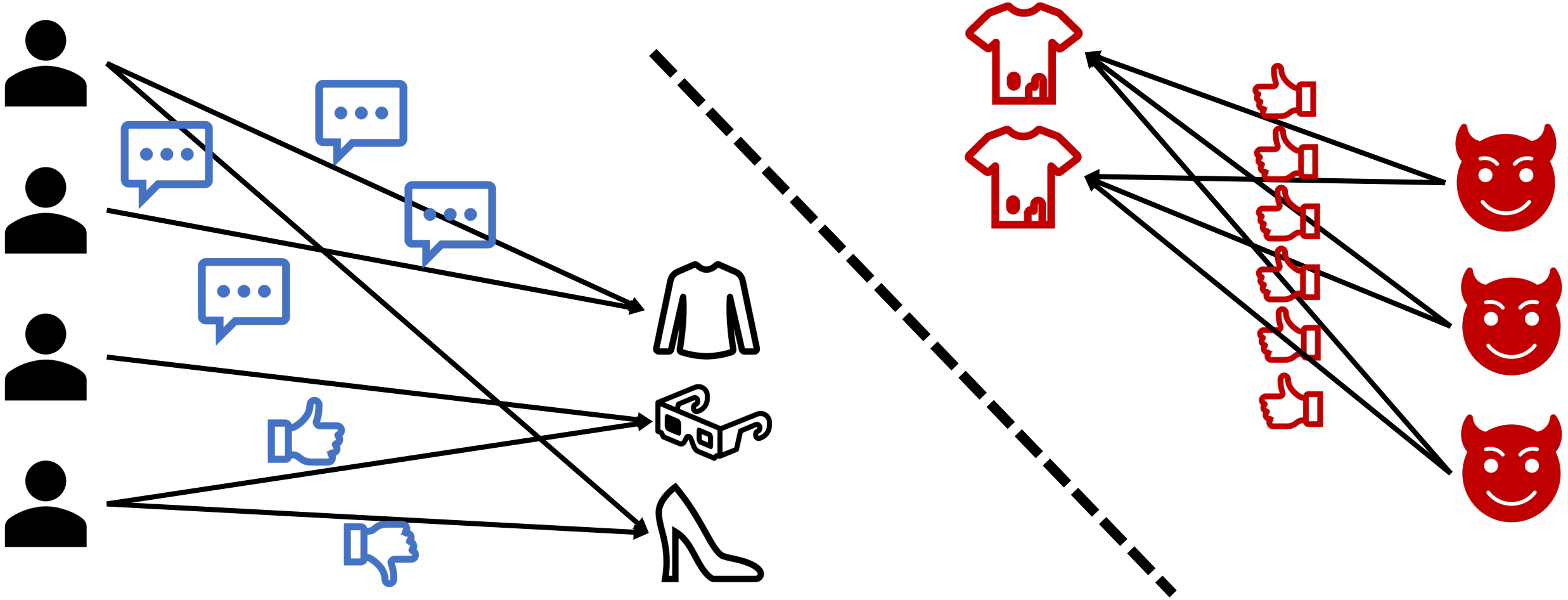
Previous graph fraud detection methods

- Type 1: detect dense interconnections among fraudsters



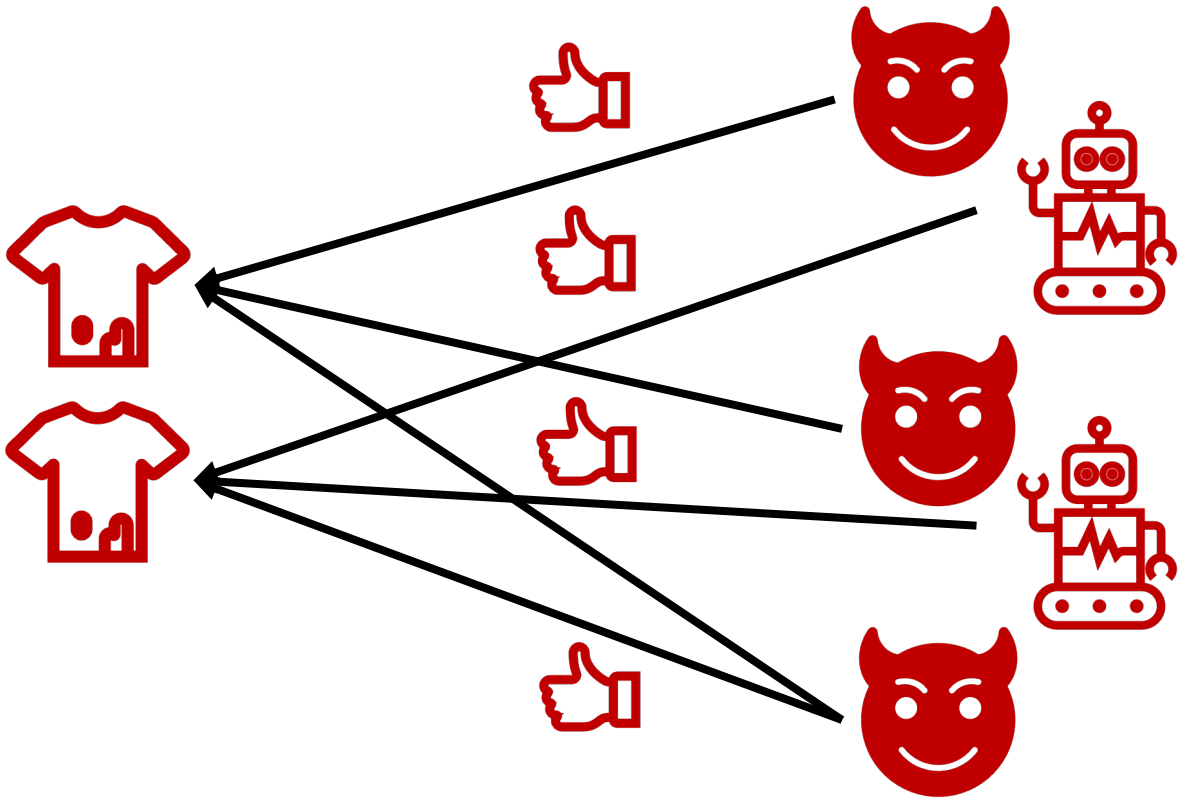
Previous graph fraud detection methods

- Type 2: detect isolated fraud communities



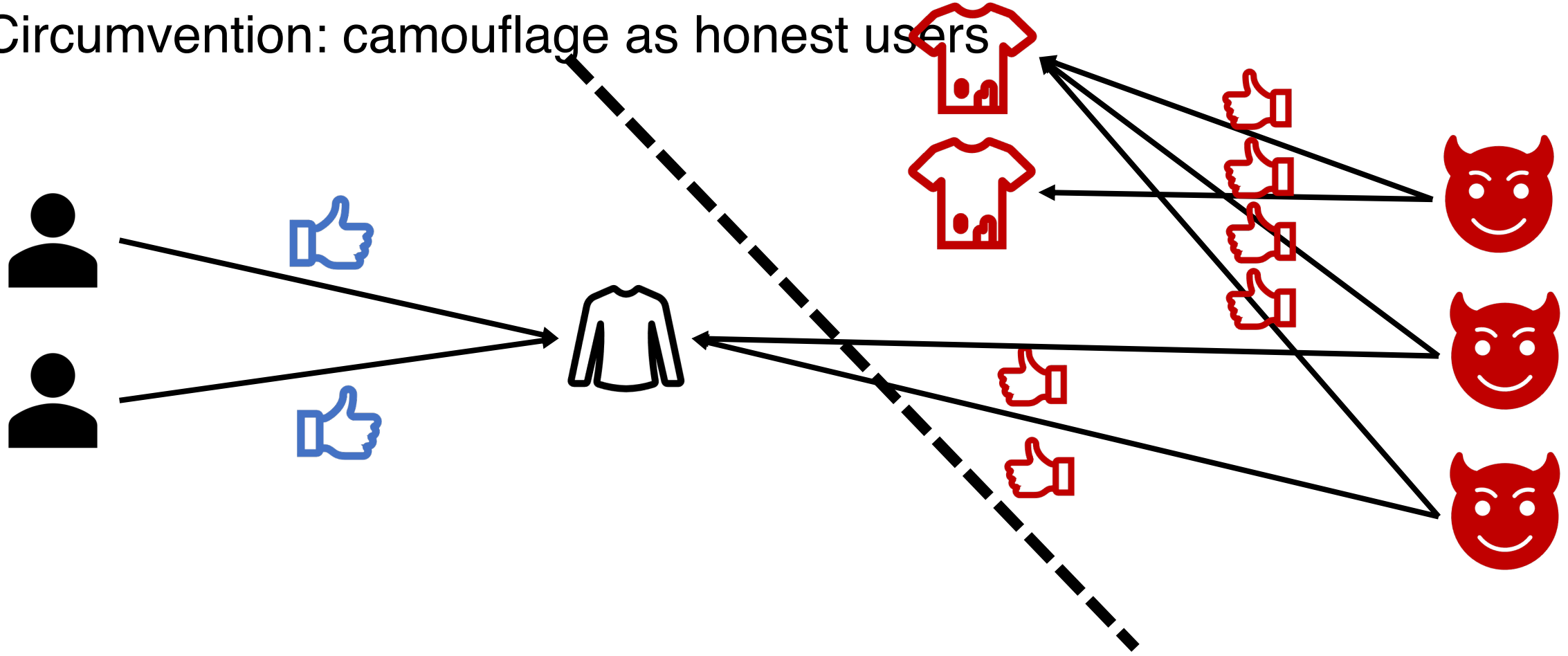
Fraudsters circumvent them easily

- Type 1: detect dense interconnections among fraudsters
- Circumvention: generate a number of bot accounts



Fraudsters circumvent them easily

- Type 2: detect isolated fraud communities
- Circumvention: camouflage as honest users

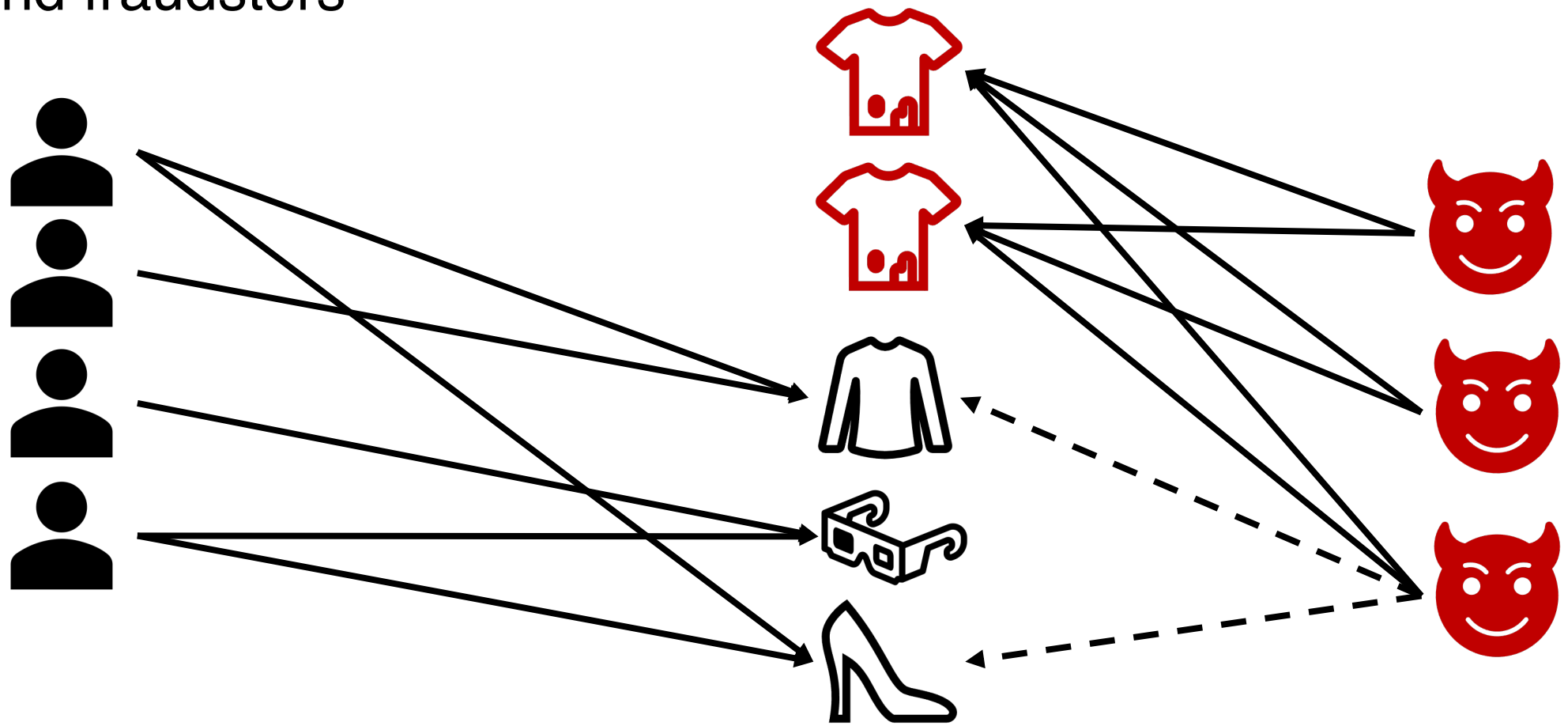


Our approach

- Characteristics that are *hard for frauds to manipulate*

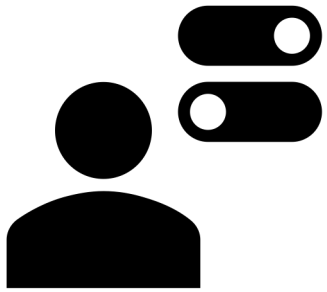
Our approach

- **Unidirectionality of communication** between honest users and fraudsters



Our approach

- **Unidirectionality of communication** between honest users and fraudsters
- This unidirectionality is generated by honest users
  - hard for fraudsters to manipulate like densities or connections

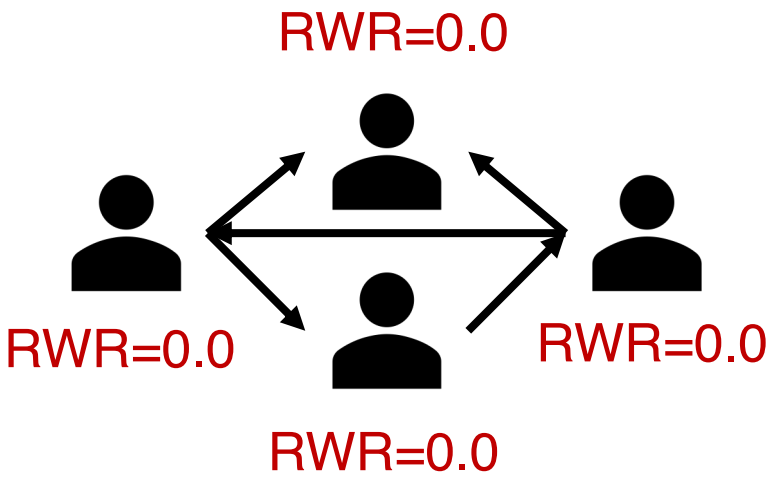


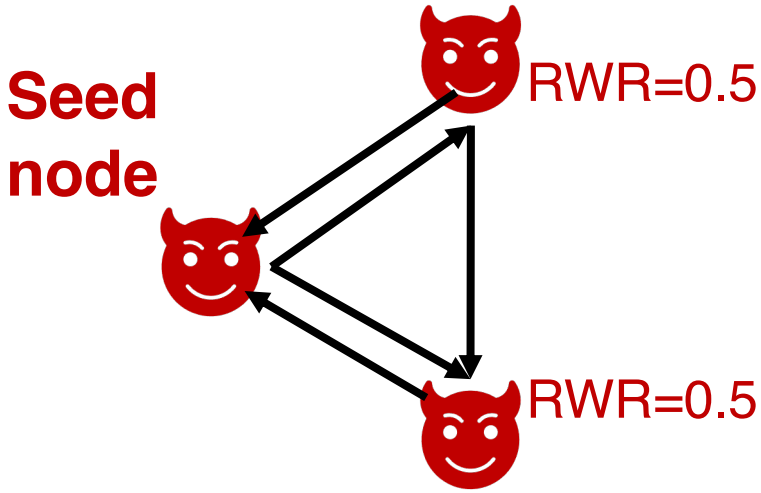
Our approach

1. Define an **accessibility score** to quantify the unidirectionality
2. Observation: unidirectionality makes fraudsters have **skewed accessibility score distributions**
3. Theoretical analysis
4. Novel graph fraud detection algorithm, **SkewA**

Accessibility score vector

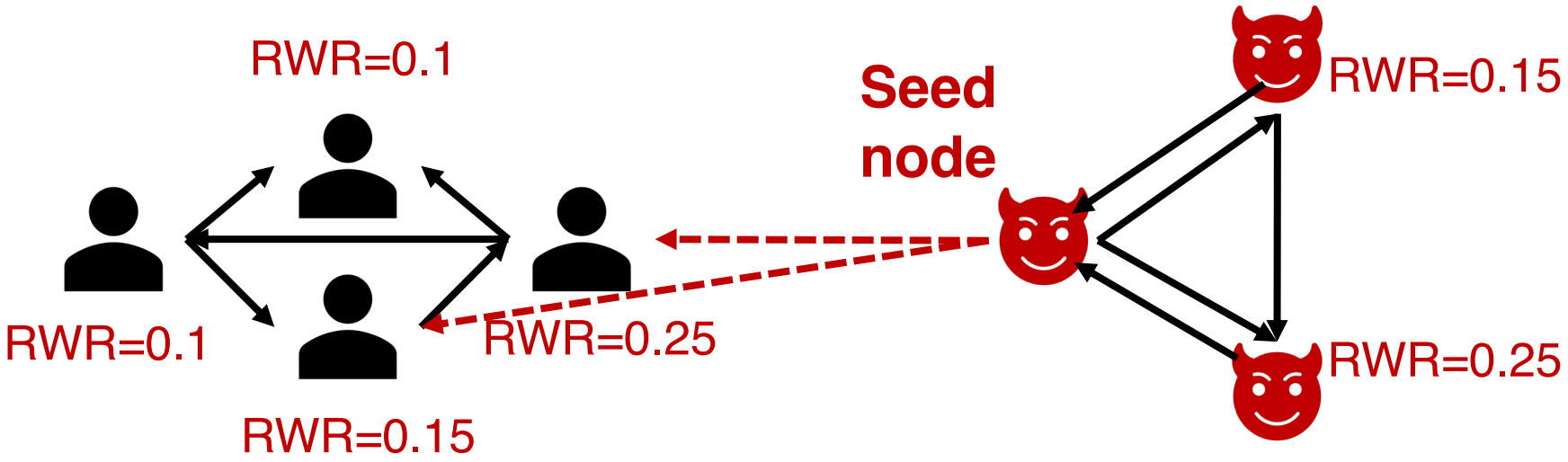
- RWR scores
  - How easily the seed node v_i could reach other nodes





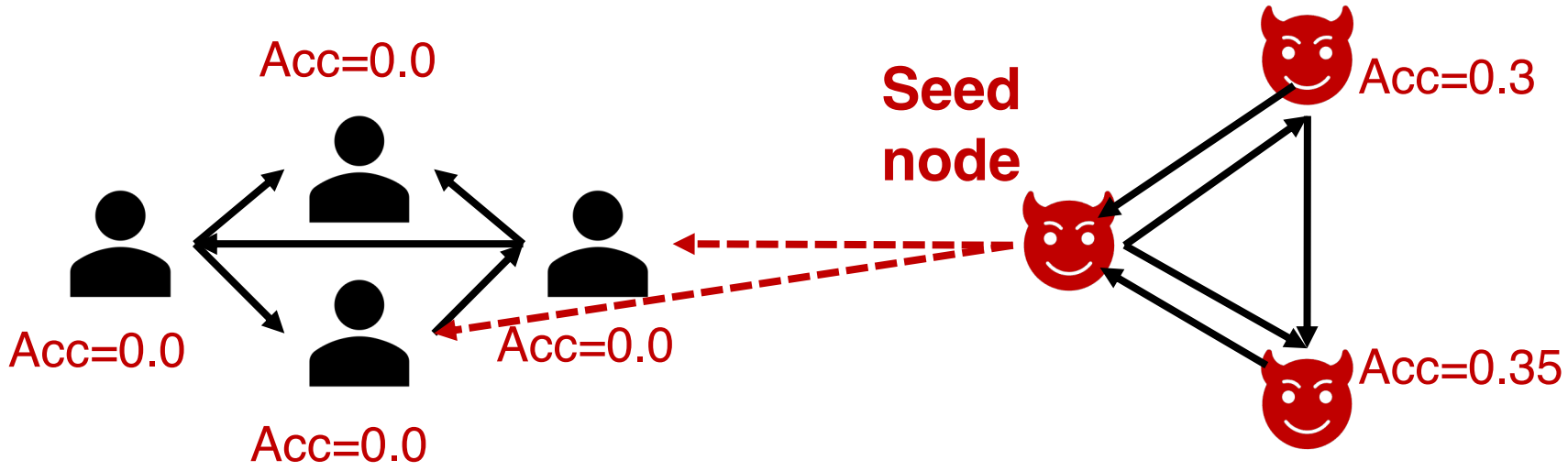
Accessibility score vector

- RWR scores
  - How easily the seed node v_i could reach other nodes
  - From the perspective of the seed node
  - Easily manipulated by the seed node by adding edges to target nodes to increase their RWR scores



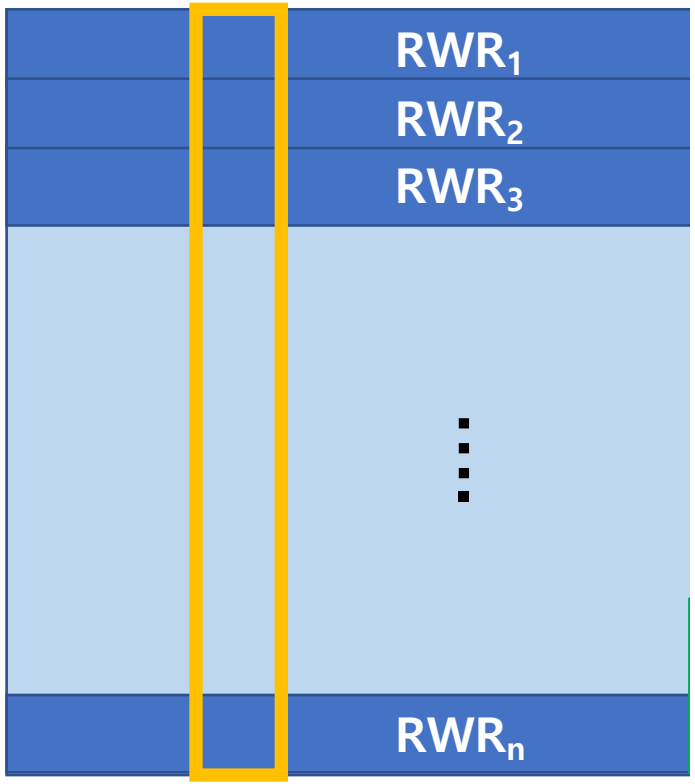
Accessibility score vector

- Accessibility scores
  - How easily other nodes could reach the seed node v_i
  - Estimated by target nodes and hard for the seed node to control.



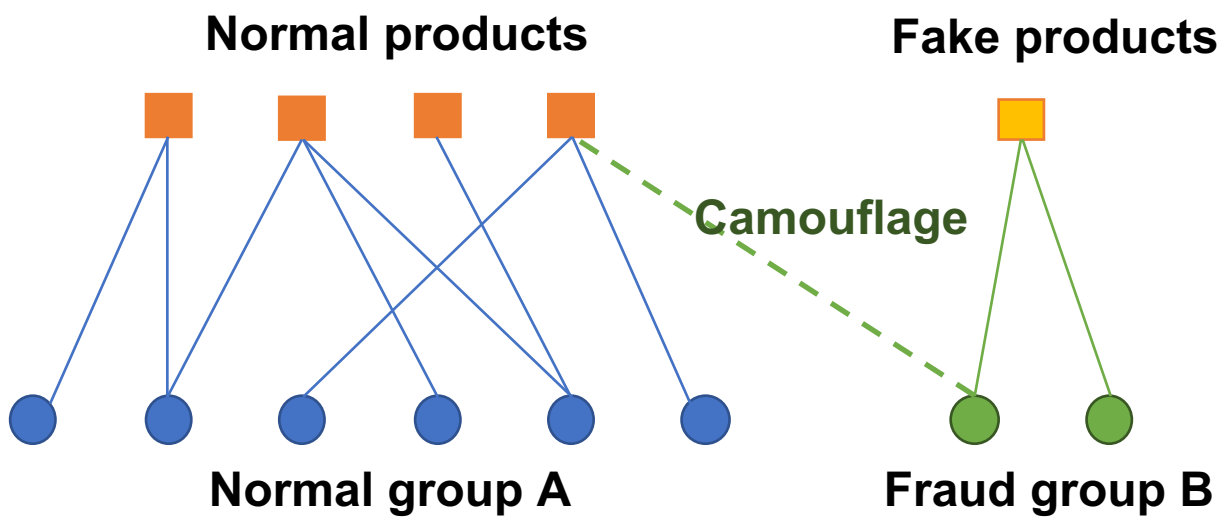
Accessibility score vector

- Accessibility score matrix is transpose to RWR score matrix



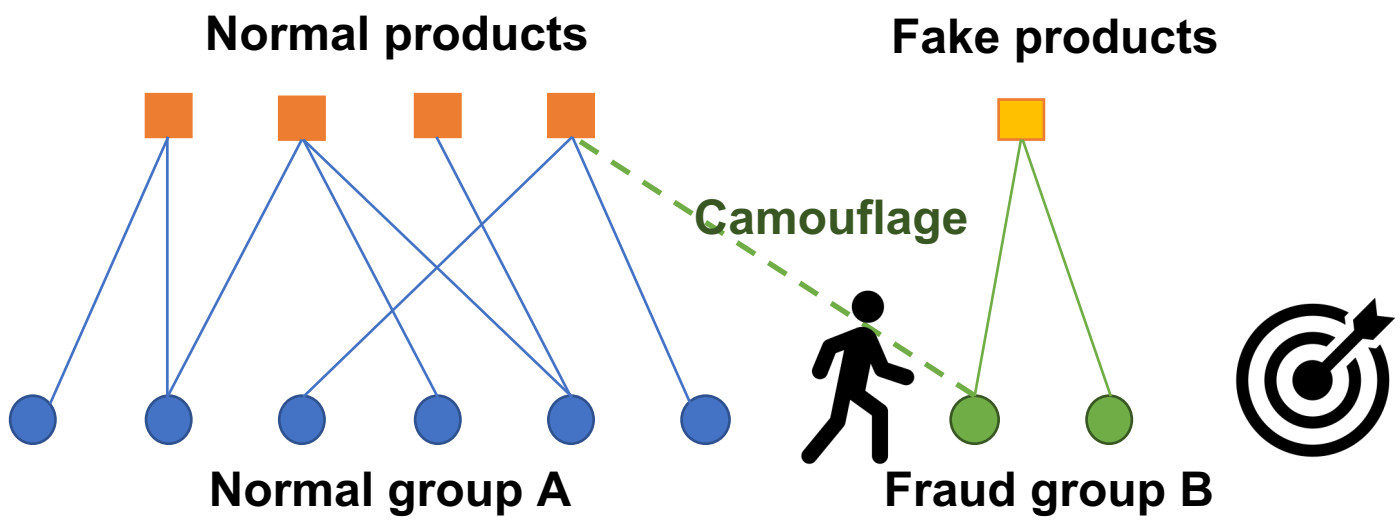
Skewed Accessibility Score Distributions

- **Fraudsters** have..
  - High accessibility scores from their fraudulent group
  - Low accessibility scores from the honest group



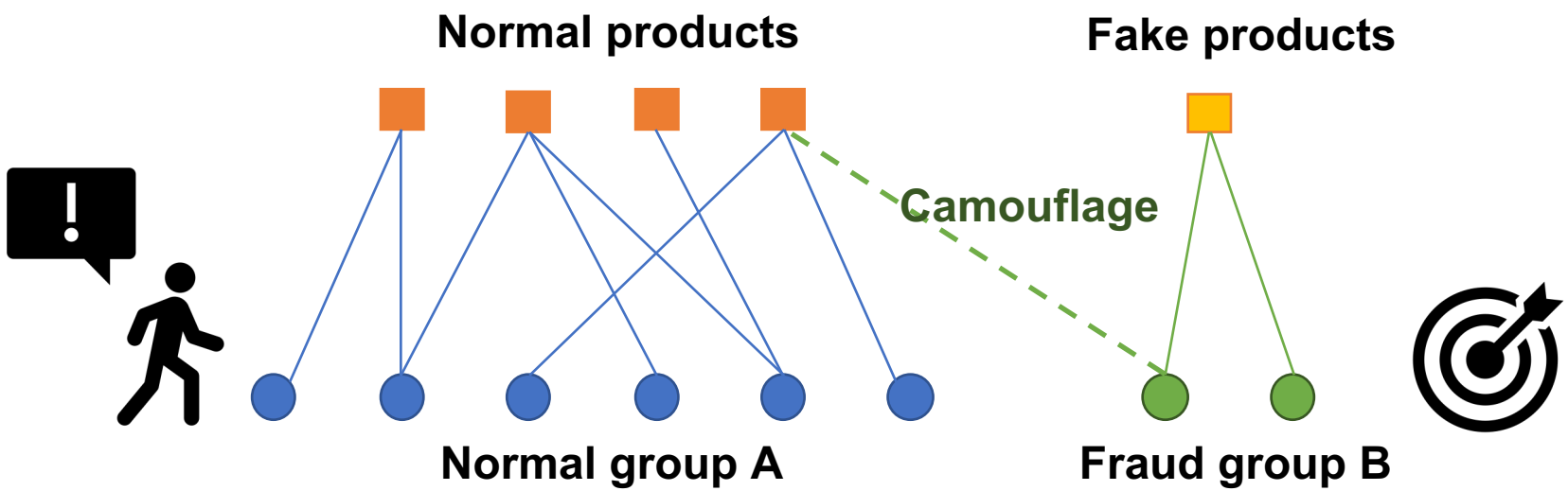
Skewed Accessibility Score Distributions

- **Fraudsters** have high accessibility scores from the fraud group



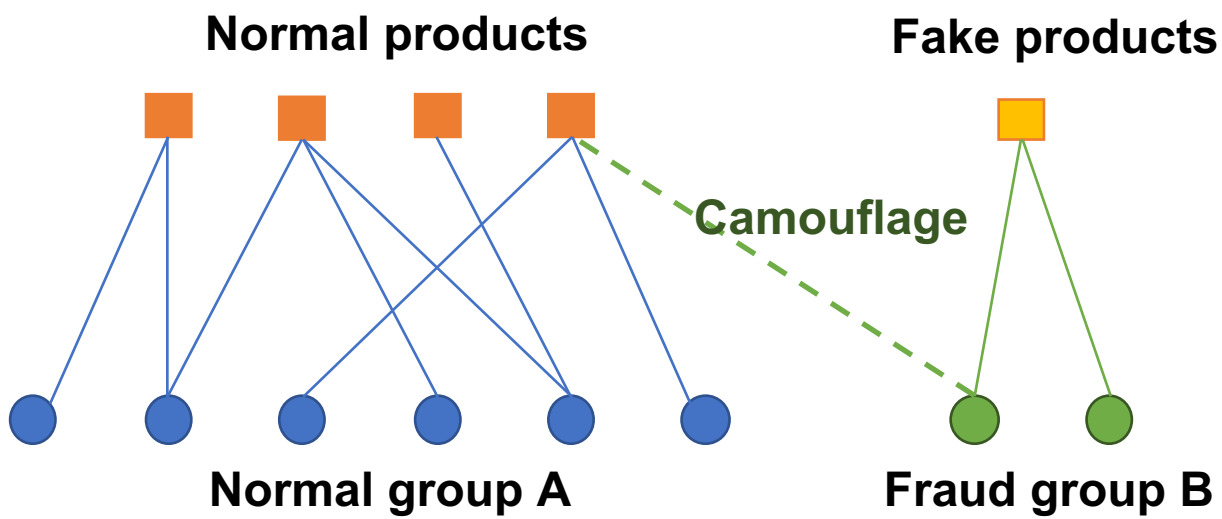
Skewed Accessibility Score Distributions

- **Fraudsters** have low accessibility scores from the honest group



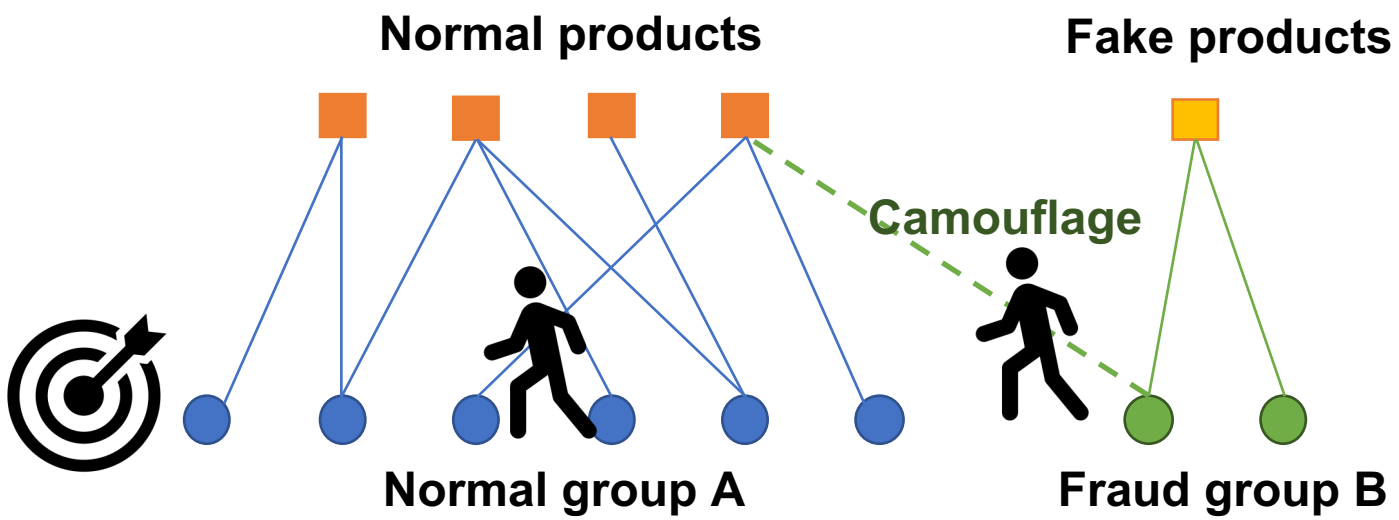
Skewed Accessibility Score Distributions

- **Honest users** have more even accessibility distributions



Skewed Accessibility Score Distributions

- **Honest users** have more even accessibility distributions

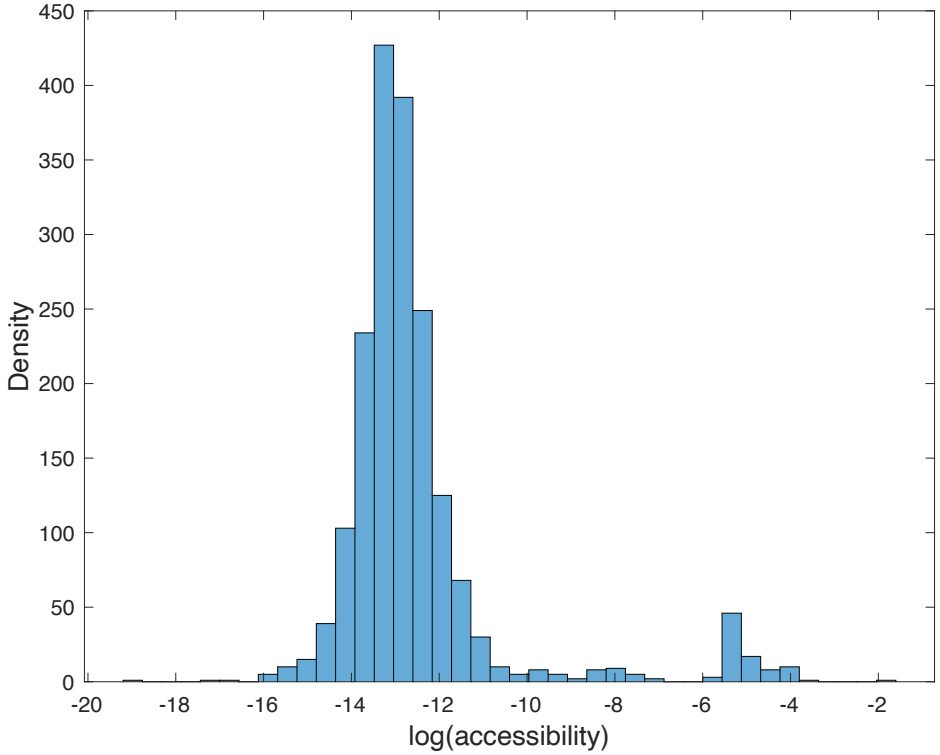


Skewed Accessibility Score Distributions

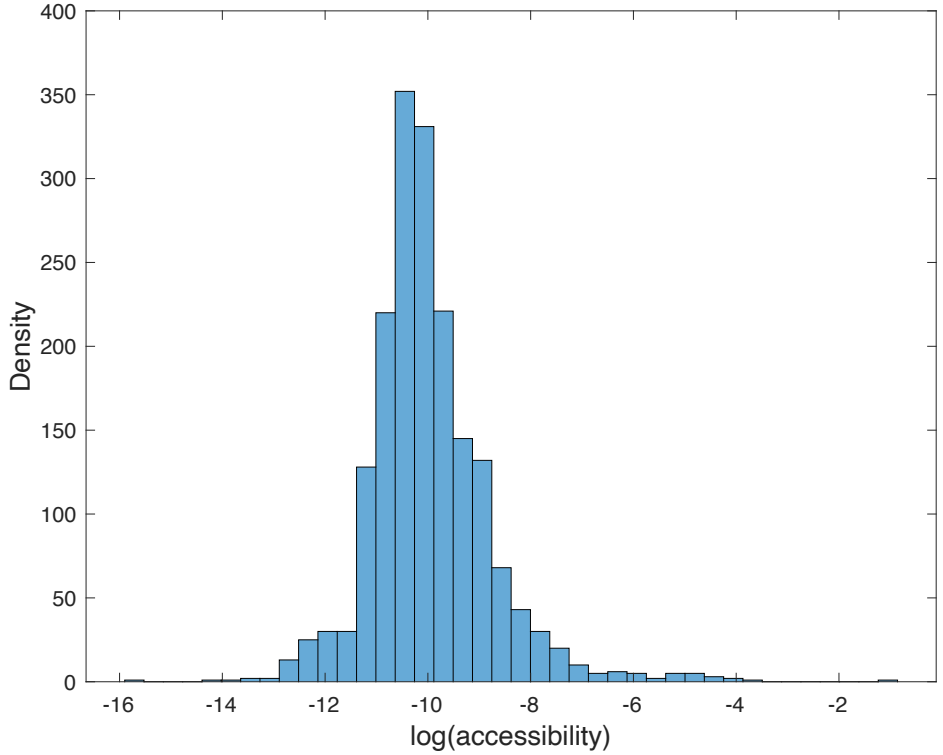
Theorem 2 (Ratio of Propagated Scores). *Given ratio of camouflage edges to honest edges ρ_a and ratio of camouflage edges to fake edges ρ_c, scores propagated into group S_1 and S_2 at the k-th iteration of RWR computation are:*

$$s_1(k) = (1-\rho_a)s_1(k-1) + \rho_a\rho_c s_1(k-1) + \rho_c s_2(k-1)$$
$$s_2(k) = \rho_a(1-\rho_c)s_1(k-1) + (1-\rho_c)s_2(k-1)$$

Skewed Accessibility Score Distributions



(a) Fraudster



(b) Honest user

Probability density function of accessibility scores

Proposed method: SkewA

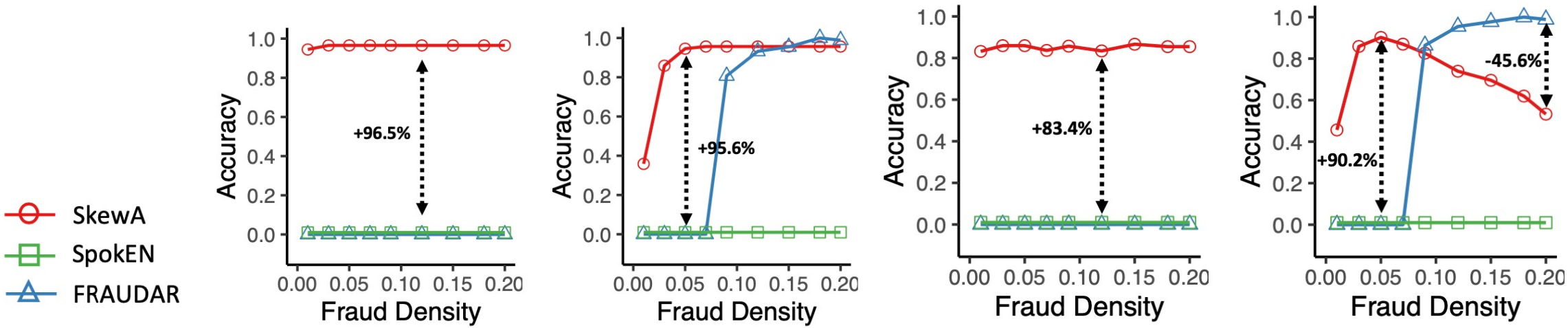
Algorithm 1: SkewA

Input: A bipartite graph G, Top k
Output: k fraudsters
Compute accessibility score matrix $\mathbf{A}_{acc}$;
Compute $\alpha = log(\frac{m}{n_1})$;
foreach *column vector* $\mathbf{a}$ *in* $\mathbf{A}_{acc}$ **do**
  ComputeHonesty($\mathbf{a}$, α)
return k *nodes with lowest honesty scores*

Algorithm 2: ComputeHonesty

Input: Accessibility score vector $\mathbf{a}$, parameter α
Output: Honesty score s_{honest}
Find local minimum in pdf;
Divide into S_1 and S_2 by the local minimum;
Compute sum and variance of S_1 and S_2;
$s_{\text{honest}} = (\text{var}_1 \text{var}_2)^{\frac{\alpha}{2}} (\text{sum}_2)^{-\frac{2}{\alpha}}$;
return s_{honest}

Experiment 1.
Robustness to sparse frauds



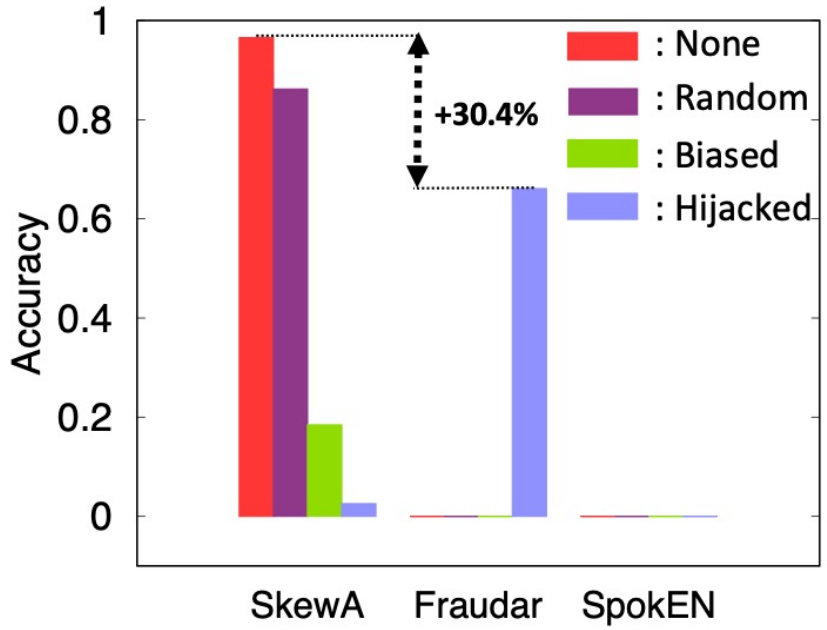
(a) Wiki-vote w/o camouflage

(b) TripAdvisor w/o camouflage

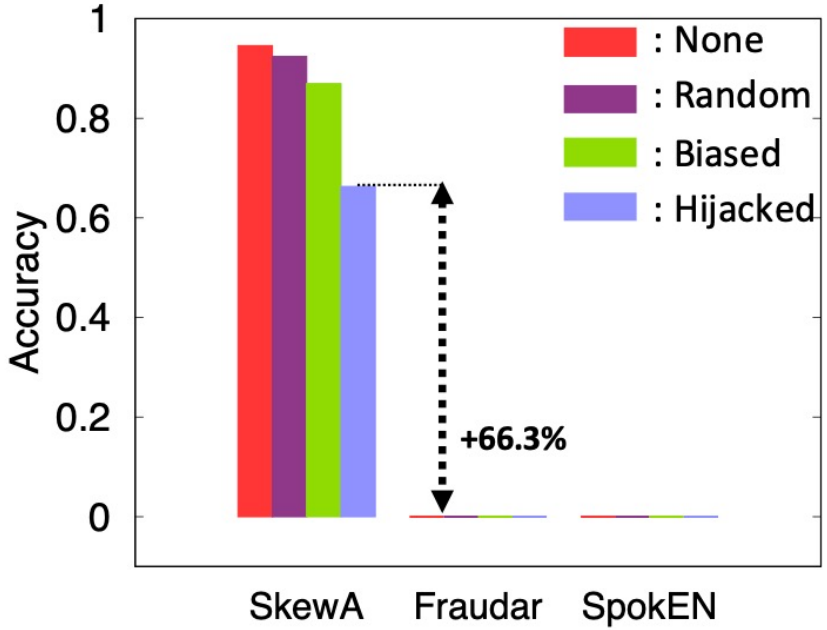
(c) Wiki-vote w/ camouflage

(d) TripAdvisor w/ camouflage

Experiment 2.
Camouflage-resistance



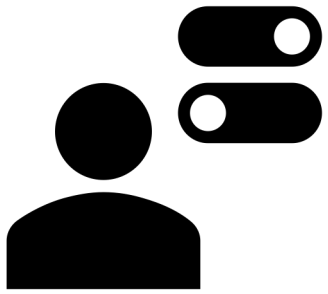
(a) Wiki-vote



(b) TripAdvisor

Conclusion

- Focus on **unidirectionality of communication**
  - Hard for fraudsters to manipulate
- Define **accessibility scores**
  - Measure the unidirectionality
- Analyze **skewed accessibility score distributions for fraudsters**



Conclusion

- Novel graph fraud detection algorithm: **SkewA**
  - Robust to sparse frauds
  - Robust to camouflaged frauds
  - Theoretical analysis
  - Presents up to 95.6% accuracy in public benchmarks

Carnegie Mellon University

ECML PKDD 2021

Thank you

Paper: https://minjiyoon.xyz/Paper/SkewA.pdf

Code: https://github.com/minjiyoon/PKDD21-SkewA