# Mitigating multiple identity attacks on content rating systems

Arash Molavi Kakhki[†$] Aniko Hannak[†$] Alan Mislove[†] Ravi Sundaram[†]

[†]*Northeastern University* [$]*Student*

## MOTIVATION

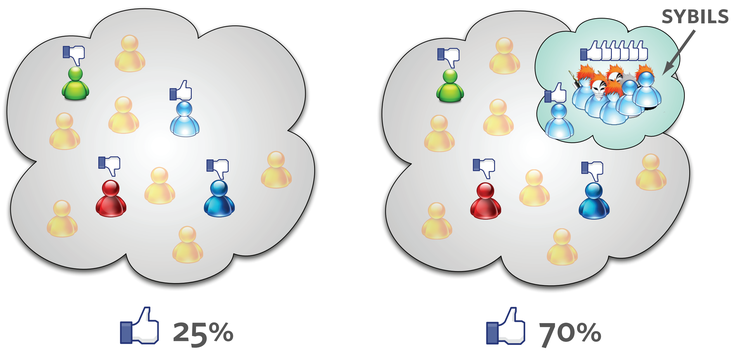
Content sharing sites allow users to find and share content
- Examples: news articles (Digg), videos (YouTube), URLs

All have basic mechanisms:
- Creating accounts
- Declaring friendships
- Uploading and rating content (voting)
- Locating content via aggregated votes

But, accounts are often not verified and free to create
- Usually only require email address + CAPTCHA
- Multiple identities referred to as Sybils [IPTPS'02]

Sybils can be used to manipulate content rating
- Vote multiple times with multiple accounts

Can make fraudulent content appear highly rated
Or, can make legitimate content appear poorly rated

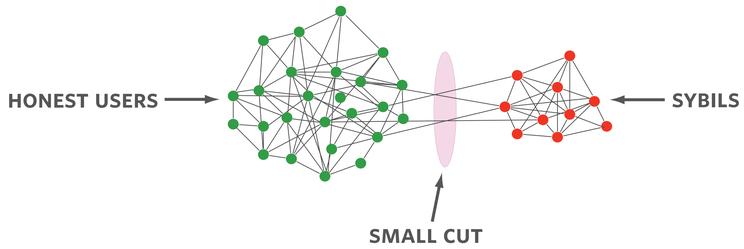



Sybil attacks observed on real-world sites
- Digg [NSDI'09]
- TripAdvisor [NYT, 08/20/2011]
- Labor markets [USENIX SEC.'11]
- Automated software
  `http://www.tubeautomator.com/`


## GOAL AND ASSUMPTIONS

Goal: Create system where users gain no additional influence by creating Sybils
- Key idea: Levarage underlying social network
- Social network often already exists on these sites

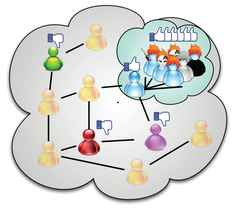
Assumption: Social links to honest users take effort to form and maintain
- Malicious user cannot obtain arbitrary links to honest users
- Introduces topological feature in social network




## DESIGN

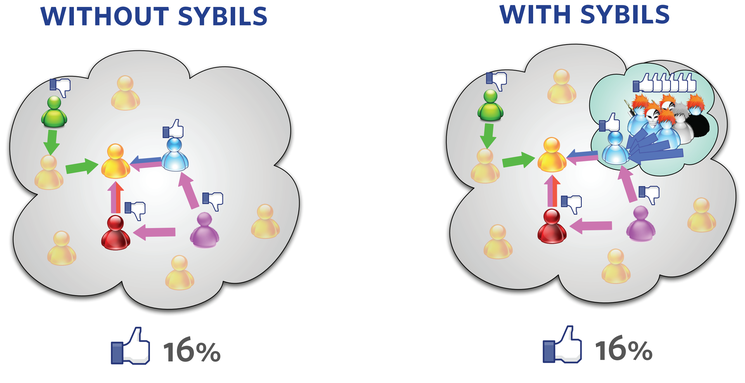
Our approach is to assign weights to votes; not all votes are counted equally

Goal: Assign weights so that user's aggregate weight does not depend on number of identities they possess
- Naturally mitigates the effect of Sybils
- Challenge is choosing weight assignment algorithm



We use flow over the social network to assign weights
- Every voter is a source; each link has unit capacity
- User asking for rating is the sink (vote collector)

Model problem as multi-commodity max flow problem
- Users "compete" to push flow to vote collector; determines vote weight
- User influence only dependent on number of real links; Sybils don't help

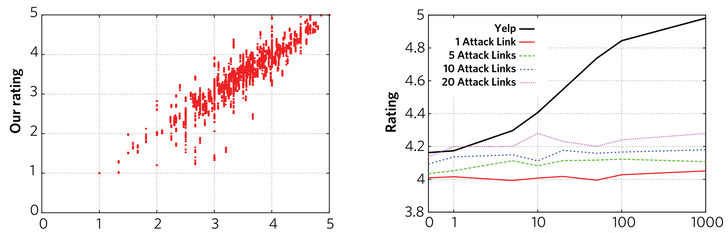



## PRELIMINARY RESULTS

Evaluate on Yelp data (65K users, 6.9K business, 152K reviews)
- Social network: Link between users with 3 common ratings

Figure a: How do our ratings compare?
- Our rating vs. Yelp's; existing ratings are largely similar

Figure b: Do we prevent Sybil voting attacks?
- Simulate Sybils by connecting Sybil network with attack links
- Sybils all rate business 5 stars.
- Result: Rating is constant, regardless of number of identities




## RELATED WORK

DSybil [OAKLAND'09] finds trusted guides (users who have a similar voting history)

Assumes all users provide enough feedback to find guides
- Many users don't vote/feedback in practice

We only require a subset of users to vote
- Others can just declare friends (feedback not required)

SumUp [NSDI'09] uses social network; inspired our design
- Defines a trusted "envelope", where all votes are counted
- Nodes outside must compete.

Vulnerable to Sybils in envelope
- Can introduce more identities into envelope

Nodes incentivized to "split" identity
- Receive more chances to find path to the vote collector
- Result: Sybils outside envelope can get more votes