Sybilproof Reputation Mechanisms

Presented by Ohad Lutzky

A. Cheng[1] E. Friedman[2]

[1]Center for Applied Mathematics
Cornell University

[2]School of Operations Research and Industrial Engineering
Cornell University

Multi-Agent Systems - Seminars

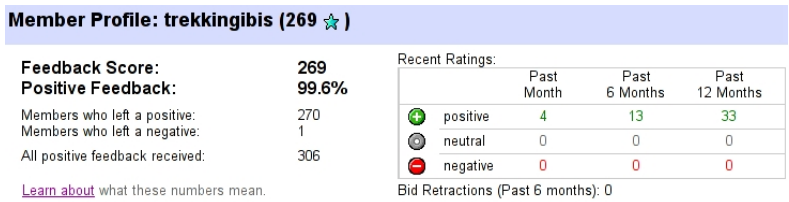
Outline

1. Problem Formulation
   - Motivation
   - Setting
   - Sybil strategies
   - Sybilproofness
2. Reputation Functions
   - Symmetric Reputations
   - Assymetric reputation functions

Outline

1. Problem Formulation
   - Motivation
   - Setting
   - Sybil strategies
   - Sybilproofness
2. Reputation Functions
   - Symmetric Reputations
   - Assymetric reputation functions

- In P2P networks (say Bittorrent, say Ebay), one interacts mostly with previously-unknown users, which makes trusting them difficult.
- A "Reputation", derived from other users' previous interactions, can help. (say PageRank, say "stars")



Member Profile: trekkingibis (269 ☆)

Feedback Score:	**269**
Positive Feedback:	**99.6%**
Members who left a positive:	270
Members who left a negative:	1
All positive feedback received:	306

Learn about what these numbers mean.

Recent Ratings:

	Past Month	Past 6 Months	Past 12 Months
positive	4	13	33
neutral	0	0	0
negative	0	0	0

Bid Retractions (Past 6 months): 0

Member Profile: trekkingibis (269 ★)

Feedback Score:	**269**
Positive Feedback:	**99.6%**
Members who left a positive:	270
Members who left a negative:	1
All positive feedback received:	306

Learn about what these numbers mean.

Recent Ratings:

	Past Month	Past 6 Months	Past 12 Months
positive	4	13	33
neutral	0	0	0
negative	0	0	0

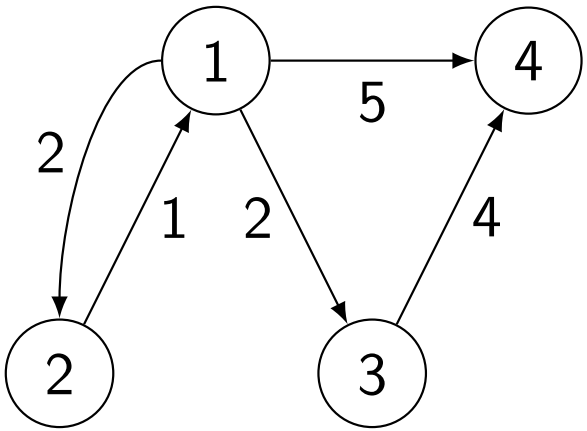
Bid Retractions (Past 6 months): 0

- In those same networks, it's often easy to create "dummy users", and falsify their interactions
- Such users can be used to artifically enhance one's "Reputation"
- How can we define "Reputation" so that this won't be possible?

Outline

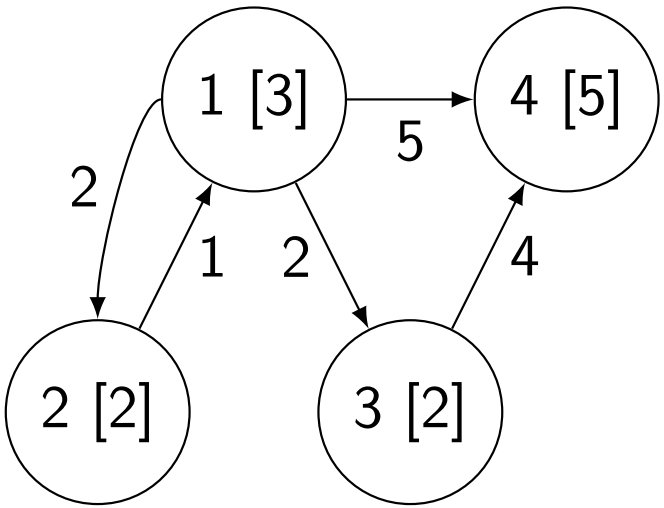
1. Problem Formulation
   - Motivation
   - Setting
   - Sybil strategies
   - Sybilproofness
2. Reputation Functions
   - Symmetric Reputations
   - Assymetric reputation functions

What-proof what-whats?



- Reputation will be computed based on peer interactions
- We will display those in a **finite** directed graph $G = (V, E)$
- V - each vertex is a user
- E - an interaction between users i, j is represented by an edge i, j, with outcome $c(i, j)$.
- The collection of all such graphs (outcomes included) will be labeled $\mathscr{G}$.

Reputation functions



Definition 1

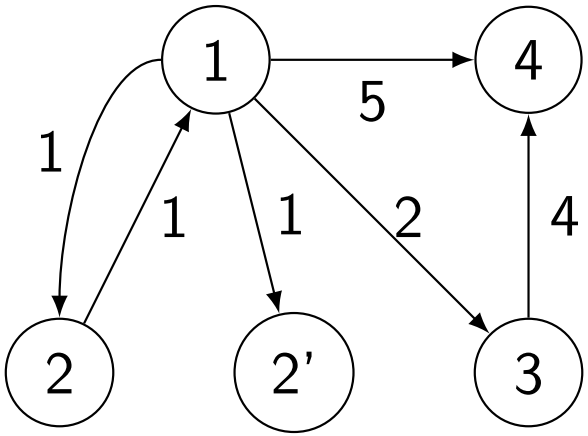
A "function" $f : \mathcal{G} \to \mathbb{R}^V$ is called a **reputation function**. We say that a node $i \in V$ in graph G has reputation $f(G)_i \in \mathbb{R}$.

That is, given a graph $G = (V, E) \in \mathcal{G}$, f assigns to each $v \in V$ its **reputation**, denoted $f(G)_i$. (Here in [square brackets])

Outline

1. Problem Formulation
   - Motivation
   - Setting
   - Sybil strategies
   - Sybilproofness
2. Reputation Functions
   - Symmetric Reputations
   - Assymetric reputation functions

Sybil strategies

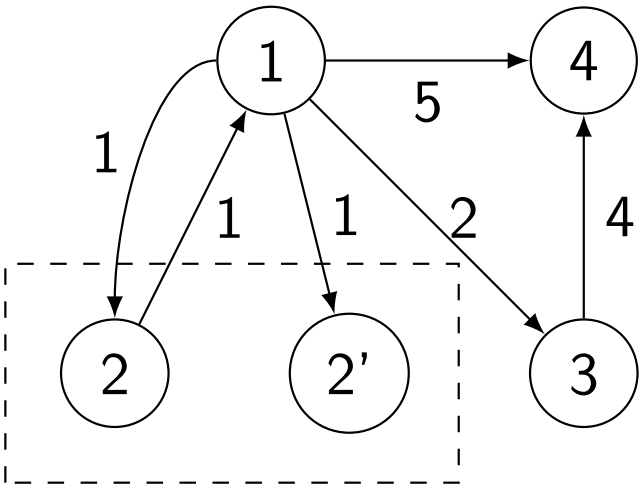


Definition 2

Given a graph $G = (V, E)$ and a user $i \in V$, we say that a graph $G' = (V', E')$ along with a subset $U' \subseteq V'$ is a **sybil strategy** for user i in the network $G = (V, E)$ if $i \in U'$ and collapsing U' into a single node with label i in G' yields G. We can refer to U as the **sybils** of i, and denote a sybil strategy by (G', U').

We assume the system is unable to tell the difference between a sybil and a user.

Sybil strategies



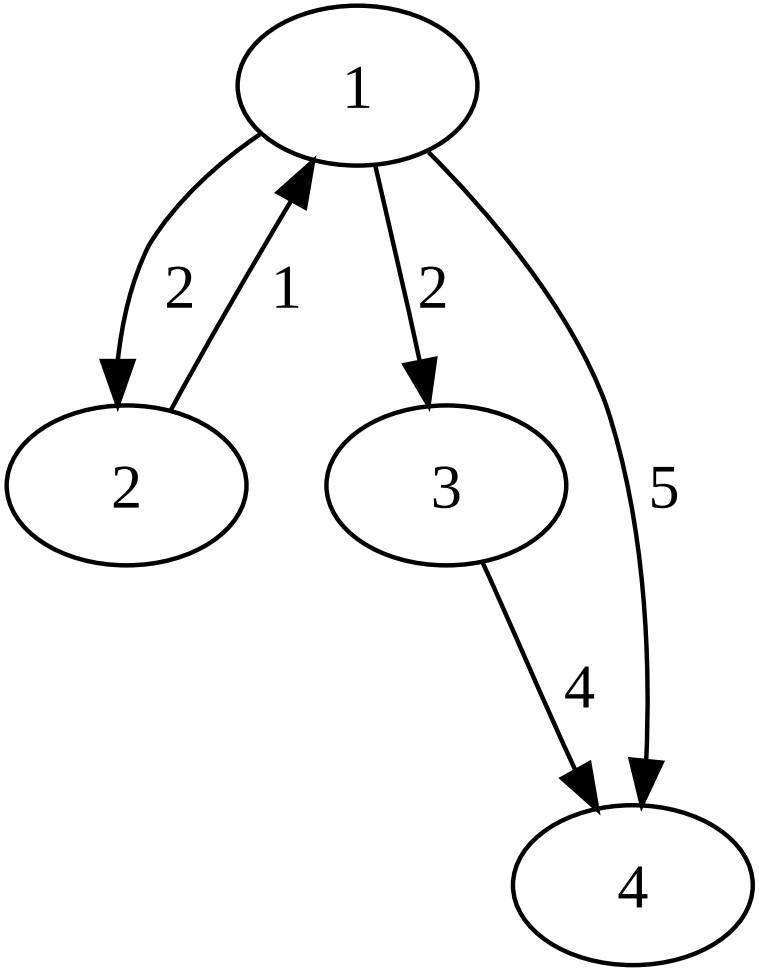
Sybil strategy of 2

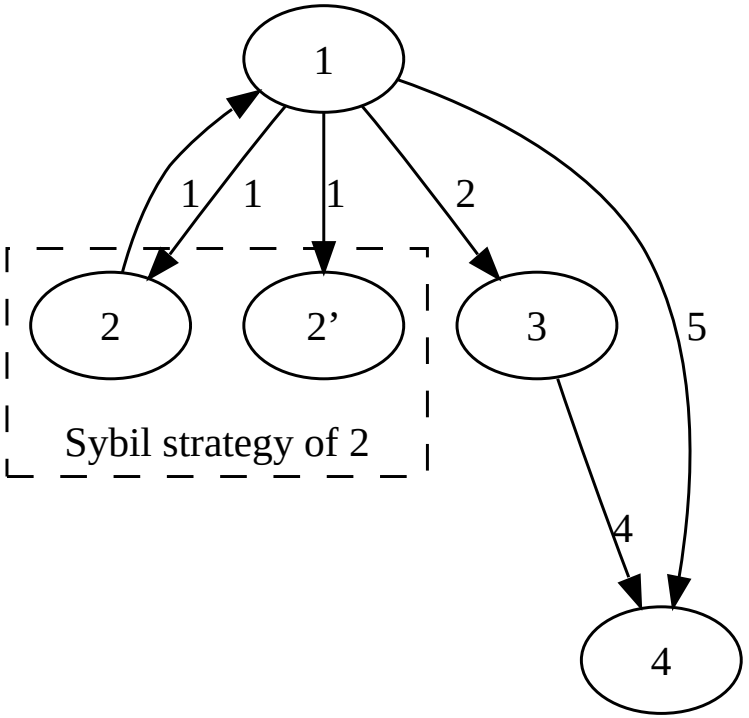
Definition 2

Given a graph $G = (V, E)$ and a user $i \in V$, we say that a graph $G' = (V', E')$ along with a subset $U' \subseteq V'$ is a **sybil strategy** for user i in the network $G = (V, E)$ if $i \in U'$ and collapsing U' into a single node with label i in G' yields G. We can refer to U as the **sybils** of i, and denote a sybil strategy by (G', U').

We assume the system is unable to tell the difference between a sybil and a user.

Example of a sybil strategy





Note that additive splitting is allowed.

Outline

1. Problem Formulation
   - Motivation
   - Setting
   - Sybil strategies
   - Sybilproofness
2. Reputation Functions
   - Symmetric Reputations
   - Assymetric reputation functions

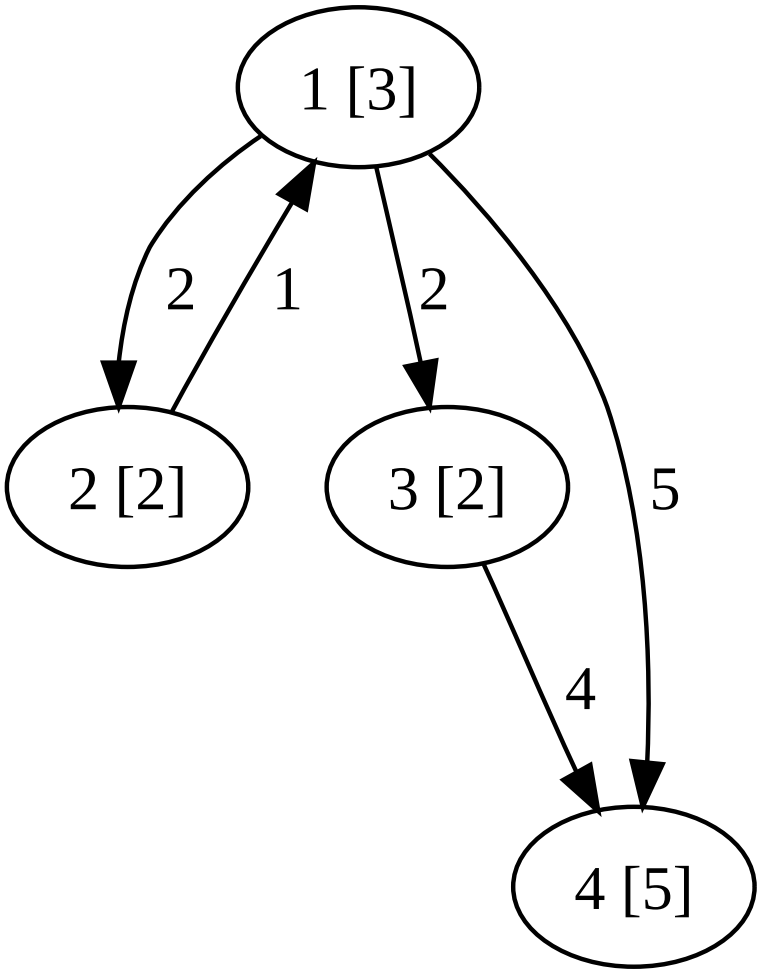
Rank sybilproofness

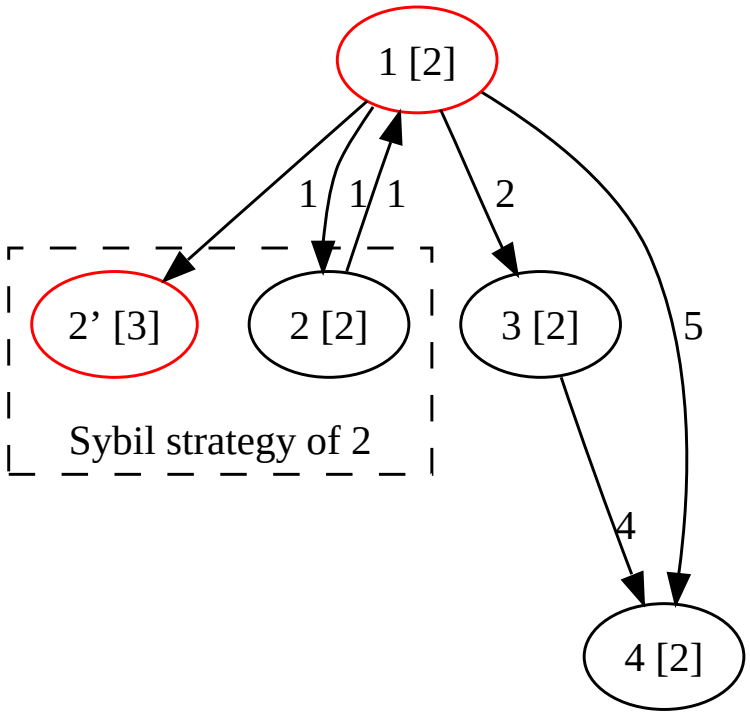
Definition 3

A reputation function f is **(rank) sybilproof** if for all graphs $G = (V, E)$, and all users $i \in V$, there is no sybil strategy for i, (G', U'), with $G' = (V', E')$ such that for some $u \in U', \exists j \in V$ such that $f(G)_j > f(G)_i$ and $f(G')_u \geq f(G')_j$

In other words, a reputation function is **not** sybilproof if one of the users can overtake another by using a sybil strategy.

Example of a non-sybilproof function





K-sybilproofness

Definition 4

We say that a reputation function is K-**sybilproof** if it is sybilproof over all possible sybil strategies (G', U') with $|U'| \leq K$.

Value sybilproofness

Sometimes the reputation value itself, and not the relative ranking of reputations, is of interest.

Definition 5

A reputation function f is **value sybilproof** if for all graphs $G = (V, E)$ and all users $i \in V$ there is no sybil strategy for i, (G', U') such that for some $u \in U'$, $f(G)_i < f(G')_u$.

In other words, no user can improve his reputation by using a sybil strategy.
Note that a function can be both value-sybilproof and sybilproof, just one, or neither.

Outline

1. Problem Formulation
   - Motivation
   - Setting
   - Sybil strategies
   - Sybilproofness
2. Reputation Functions
   - Symmetric Reputations
   - Assymetric reputation functions

Can we keep it fair?

- We might want our system to be fair, or anonymous
- We would ignore all "who is who" information
- The edge values would encapsulate all information

Definition 6

A reputation function f is **symmetric** if given a graph isomorphism[a] σ and a graph $G = (V, E)$, then for all $i \in V$, $f(G)_i = f(\sigma(G))_{\sigma(i)}$.

[a] Relabeling of vertices

That is - renaming the users would have no effect on a symmetric function.

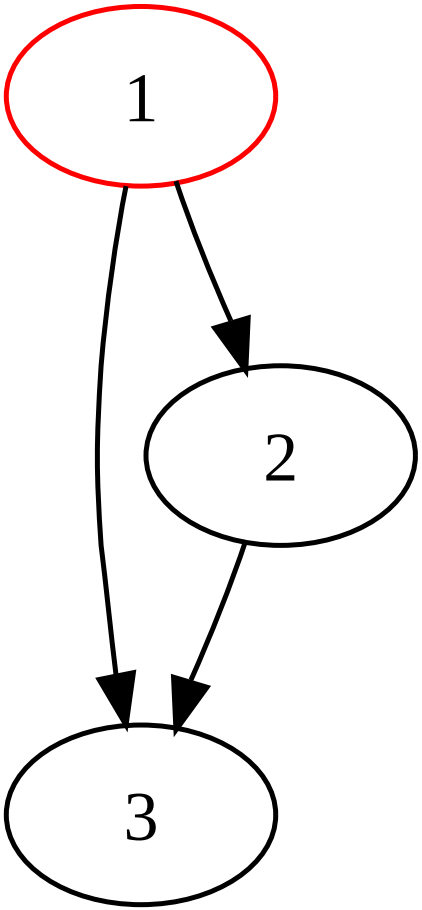
Bad news

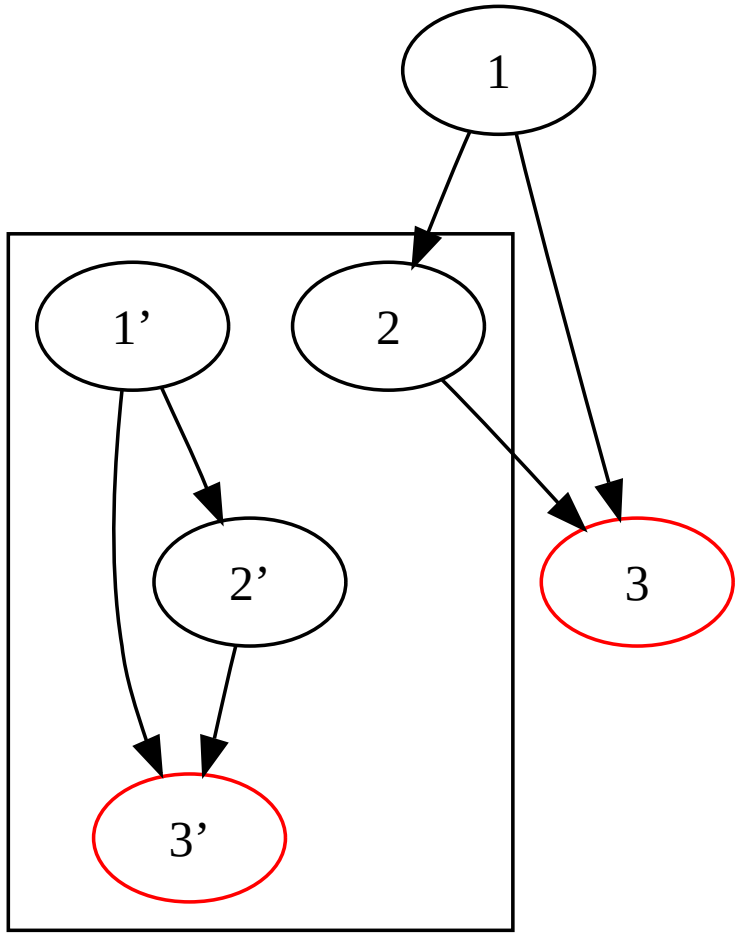
Theorem 7

There is no symmetric sybilproof nontrivial[a] reputation function.

[a] Not $f \equiv \text{const}$

Proof sketch for Theorem 7





For every node with a non-maximal rank, a successful sybil strategy would be duplicating the rest of the graph. Because of symmetry, the maximum of G' would be attained in both copies.

K-sybilproofness

- Are users really required (or able) to pull this sybil strategy off?
- They don't need to:

Theorem 8

There is no nontrivial symmetric k-sybilproof reputation function for any constant $k > 1$

Proof concept

It suffices to show for $k = 2$. Assume by contrast a sybilproof f exists, and create the graph-copy strategy step by step. The function is 2-sybilproof, so the strategy is unsuccessful at every step, but by the last step it must be - a contradiction.

K-sybilproofness

- Are users really required (or able) to pull this sybil strategy off?
- They don't need to:

Theorem 8

There is no nontrivial symmetric k-sybilproof reputation function for any constant $k > 1$

Proof concept

It suffices to show for $k = 2$. Assume by contrast a sybilproof f exists, and create the graph-copy strategy step by step. The function is 2-sybilproof, so the strategy is unsuccessful at every step, but by the last step it must be - a contradiction.

K-sybilproofness

- Are users really required (or able) to pull this sybil strategy off?
- They don't need to:

Theorem 8

There is no nontrivial symmetric k-sybilproof reputation function for any constant $k > 1$

Proof concept

It suffices to show for $k = 2$. Assume by contrast a sybilproof f exists, and create the graph-copy strategy step by step. The function is 2-sybilproof, so the strategy is unsuccessful at every step, but by the last step it must be - a contradiction.

K-sybilproofness

- Are users really required (or able) to pull this sybil strategy off?
- They don't need to:

Theorem 8

There is no nontrivial symmetric k-sybilproof reputation function for any constant $k > 1$

Proof concept

It suffices to show for $k = 2$. Assume by contrast a sybilproof f exists, and create the graph-copy strategy step by step. The function is 2-sybilproof, so the strategy is unsuccessful at every step, but by the last step it must be - a contradiction.

Value sybilproofness

For value sybilproofness, there are certain pathological functions we need to exclude:

Definition 9

Given a graph G, its **B-extension** with respect to i is the graph which is constructed by taking a copy of G and contracting the node $i \in V$ with its double in the copy of G.

Definition 10

A reputation function f is **B-Nontrivial** if there exists a graph $G = (V, E)$ and $i, j \in G$ such that $f(G)_j > f(G)_i$ and $\exists v \in V'$ such that $f(G')_v > f(G')_i$, where G' is the B-extension of G with respect to i.

Value sybilproofness contd.

Theorem 11

If a reputation function f is B-nontrivial[a], then it cannot be value sybilproof, or k-value sybilproof.

[a] And symmetric?

- Note that PageRank is B-nontrivial and symmetric, and thus is not sybilproof.
- A much better candidate would be Personalized Pagerank.

Value sybilproofness contd.

Theorem 11

If a reputation function f is B-nontrivial[a], then it cannot be value sybilproof, or k-value sybilproof.

[a] And symmetric?

- Note that PageRank is B-nontrivial and symmetric, and thus is not sybilproof.
- A much better candidate would be Personalized Pagerank.

Outline

1. Problem Formulation
   - Motivation
   - Setting
   - Sybil strategies
   - Sybilproofness
2. Reputation Functions
   - Symmetric Reputations
   - Assymetric reputation functions

We would like to define an assymetric reputation by computing reputation values with respect to some fixed node in the graph - a trusted user, or perhaps oneself.

- The root ("trusted") node will be labeled s.
- Let $\mathbb{P}_i$ be the **set** of all **collections** of edge-disjoint paths from s to i in G.
- We allow an edge of value $\alpha + \beta$ to split into two parallel edges with values α, β at will.
- Let g be a function from **paths** to $\mathbb{R}$.
- Let $\oplus$ be an "addition"-like operator on real numbers.

s-centric reputation

We will deal with reputation functions of this kind:

$$(f^s(G))_i := \max_{\mathscr{P}_{s,i} \in \mathbb{P}_i} \bigoplus_{P \in \mathscr{P}_{s,i}} g(P)$$

We set $f^s(G)_s = \infty$.

- With $\oplus = +$, $g(P) = \min\{c(e) | e \in P\}$, this is maximum flow.

Theorem 12

If f^s, as defined above, satisfies the following properties,

Diminishing returns *For all $s-i$ paths $'$, if an $s-j$ path P' is an extension of P, then $g(P') \leq g(P)$.*

Monotonicity *$\oplus$ is nondecreasing, and g is nondecreasing with repsect to the edge values.*

No splitting *Given a single $s-i$ path, if we split P into two $s-i$ paths P_1, P_2, then $g(P_1) \oplus g(P_2) \leq g(P)$.*

for all graphs $G = (V, E), s \in V$ and all $i \in V$, then f^s is value sybilproof.

Proof.

Let $G = (V, E)$ be a graph, let $s, i \in V$, $s \neq i$. Let (G', U') be a sybil strategy for i with respect to f^s, with collection of sybils U'. For $u \in U'$, there is some collection of disjoint $s - u$ paths $\mathscr{P}'$ in G' such that $f^s(G')_u = \bigoplus_{P' \in \mathscr{P}'} g(P')$. For each $P' \in \mathscr{P}'$, let P be the subpath starting from s and ending at the first node in U' appearing along the path. By D.R., $g(P) \geq g(P')$, and by definition of a sybil strategy, P' must correspond to some $s - i$ path **in** G. Let $\mathscr{P} = \{P | P' \in \mathscr{P}'\}$. $\mathscr{P}$ forms an edge disjoint collection of $s - i$ paths in G, so by definition of f^s,

$$f^s(G)_i \geq \bigoplus_{P \in \mathscr{P}} g(P) \geq \bigoplus_{P' \in \mathscr{P}'} g(P') = f^s(G')_u$$

- Under f^s satisfying these conditions, no node can increase their own reputation value.
- A node may still improve their rank by "ruining" another's reputation.
- This won't work if the only nodes who may be affected by i's sybil strategies have lower reputation than i.
- $\oplus = \max$ will give us this.

Theorem 13

If f^s satisfies the above properties and additionally, $\oplus = \max$, then f^s is sybilproof. Conversely, if g is such that for all paths P, there exists a strictly longer path P', $P \subsetneq P'$, such that $g(P) = g(P')$, then f^s being sybilproof implies that $\oplus = \max$.

Summary

- We have shown a possible framework for assessing a reputation mechanism's robustness to sybils
- We have shown that no nontrivial symmetric reputation function is sybilproof - for example, pagerank
- We have shown a class of reputation functions which are sybilproof.

- Outlook
  - Personalized pagerank **might** fall under this class. . .
  - Primary unsolved problem: Finding a more general set of sufficient conditions for sybilproofness.