Manipulationsresistente Reputationssysteme

eine spieltheoretische Einführung

Stephan Beyer

Hauptseminar AFS
Institut für Theoretische Informatik
Fakultät für Informatik und Automatisierung
Technische Universität Ilmenau

26. Juni 2008

Worum geht es?

Reputationssysteme

„Definition“

Ein **Reputationssystem** sammelt, verwaltet und verbreitet Reputationen eines jeden Teilnehmers (*Subjekt*/*Knoten*).

Beispiele:

- Meinungen über (Ver-)Käufer bei Online-Auktionen
- PageRank von Webseiten
- Beurteilungen über Verfügbarkeit von Peers in P2P-Systemen
- Zitationen in wissenschaftlichen Publikationen

Verarbeitung und Darstellung

. . . der Reputation

„Definition“

Die **Reputation** eines Knotens stellt die in irgendeiner Form zusammengefasste Art und Weise der *Geschichte* (bisherigen Mitwirkung) des Subjekts im System dar.

Auflistung der Geschichte (Taten, Bewertungen, . . .) des Subjekts

- erlaubt komplexe Beurteilung

Zusammenfassung zu numerischen Werten

- erlaubt direkten Vergleich
- ermöglicht Rankings
- Aggregationsprozedur notwendig, oft nicht trivial

Wichtigkeit Reputationssysteme

...im Internet

Problem:

- Anonymität
- „Anarchie“

sichtbare Geschichte des Subjekts eröffnet Anderen

- Informationen über „Befähigung“ des Subjekts zur Qualität bei der Mitwirkung,
- damit Urteile,
- bessere Entscheidungen bei der Auswahl des Gegenübers.

Subjekt wird von Betrug und geringem Bemühen abgeschreckt und bekommt Anreize sich „anzustrengen“.

Wichtigkeit manipulationsresistenter Reputationssysteme

...im Internet

Manipulationen:

- Neuanfangen im System (engl. *Whitewashing*)
  - Anlegen einer neuen Identität
  - Neuanfang mit „reiner Weste“
  - führt jedes Reputationssystem *ad absurdum*
- Phantomfeedback (engl. *Sybil Attacks*)
  - Anlegen weiterer Identitäten um anderen eigenen Identitäten gutes Feedback zu geben
- Unehrlichkeit oder Zurückhalten von Feedback

Teil I

Der Effekt von Reputation

Gefangenendilemma

das Spiel, einfach[1]

- Spieler 1 und 2 ziehen verdeckt
- Züge:
  - C = mit anderem kooperieren
  - D = den anderen verraten
- Punkteverteilung:
  - $\pi_1(C, C) = \pi_2(C, C) = 1$
  - $\pi_1(D, D) = \pi_2(D, D) = 0$
  - $\pi_1(D, C) = \pi_2(C, D) = 2$
  - $\pi_1(C, D) = \pi_2(D, C) = -1$

π_i	Gegner wählt C	Gegner wählt D
Ich wähle C	1	-1
Ich wähle D	2	0

[1] siehe auch Einführungsvortrag

Strategien

Einzelspiel-Überlegung:

- C resultiert entweder in 1 oder -1
- D resultiert entweder in 2 oder 0
- trivial: wähle D

Wiederhole Spiel unendlich oft:

- wird immer (D, D) gespielt, so bleibt jeder Spieler bei 0 Punkten
- wird immer (C, C) gespielt, so gewinnt jeder Spieler in jeder Runde einen Punkt
- der, der dann zu erst D spielt, erringt zwar Vorsprung, aber zugleich *schlechten Ruf*
- also: Geschichte des Handelns der einzelnen Spieler wichtig

Beispielstrategien

zufällig

-1	(C, D)	2
0	(C, C)	3
2	(D, C)	2
3	(C, C)	3
4	(C, C)	4
6	(D, C)	3
7	(C, C)	4
7	(D, D)	4
9	(D, C)	3
8	(C, D)	5
8	(D, D)	5
8	(D, D)	5
9	(C, C)	6

Grim

1	(C, C)	1
2	(C, C)	2
3	(C, C)	3
4	(C, C)	4
5	(C, C)	5
4	(C, D)	7
4	(D, D)	7
4	(D, D)	7
4	(D, D)	7
4	(D, D)	7
4	(D, D)	7
4	(D, D)	7
4	(D, D)	7

Tit-for-Tat

1	(C, C)	1
2	(C, C)	2
1	(C, D)	4
1	(D, D)	4
1	(D, D)	4
3	(D, C)	3
4	(C, C)	4
3	(C, D)	6
5	(D, C)	5
6	(C, C)	6
5	(C, D)	8
5	(D, D)	8
7	(D, C)	7

Strategie: Grim

Definition

Ein Spieler spielt die **Grim**-Strategie (verbittert, engl. auch *spite*), gdw. er solange C wählt, bis ein Spieler D gewählt hat.

- spielen beide Grim, so bedeutet das ewiges (C, C)
- kein Spieler will dann von der Grim-Strategie abweichen

$\Rightarrow$ NASH-Equilibrium

Bewertungsmodell

Definiere

π_i^t: Gewinn für Spieler i in Runde t allein

δ: Diskontfaktor, $0 < \delta < 1$
Je kleiner, desto früher sollte man hohen Gewinn erzielen.

$\pi_i^t \, \delta^t$: diskontierter Gewinn

$\overline{\pi_i}$: diskontierter Gewinn pro Runde im Durchschnitt,
nach unendlich vielen Runden.
Es gilt

$$\overline{\pi_i} = (1 - \delta) \sum_{t \geq 0} \pi_i^t \, \delta^t$$

Bemerkung: $\delta = 0 \quad \Rightarrow \quad \overline{\pi_i} = \pi_i^0$,
d. h. $\overline{\pi_i}$ entspricht Gewinn eines einmaligen Spiels; uninteressant.

Analyse Grim-Strategie

Beide Spieler spielen Grim, aber Spieler i weicht in Runde T ab:

- $T = 0$: $\overline{\pi_i} = (1-\delta)\,(2 + 0\,\delta + 0\,\delta^2 + 0\,\delta^3 + 0\,\delta^4 + \ldots)$
- $T = 1$: $\overline{\pi_i} = (1-\delta)\,(1 + 2\,\delta + 0\,\delta^2 + 0\,\delta^3 + 0\,\delta^4 + \ldots)$
- $T = 2$: $\overline{\pi_i} = (1-\delta)\,(1 + 1\,\delta + 2\,\delta^2 + 0\,\delta^3 + 0\,\delta^4 + \ldots)$
- $T = 3$: $\overline{\pi_i} = (1-\delta)\,(1 + 1\,\delta + 1\,\delta^2 + 2\,\delta^3 + 0\,\delta^4 + \ldots)$

Allgemein

$$\begin{aligned}
\overline{\pi_i} &= (1-\delta)\left(\delta^T + \sum_{0 \le t \le T} 1\,\delta^t\right) \\
&= (1-\delta)\left(\delta^T + \frac{1-\delta^{T+1}}{1-\delta}\right) \\
&= 1 + \delta^T - 2\,\delta^{T+1}
\end{aligned}$$

Analyse Grim-Strategie (2)

Spieler i weicht in Runde T ab: $\overline{\pi_i} = 1 + \delta^T - 2\,\delta^{T+1}$

Spieler i weicht nie ab:

$$\begin{aligned}
\overline{\pi_i} &= (1-\delta) \sum_{t \geq 0} 1\,\delta^t \\
&= (1-\delta)\,(1-\delta)^{-1} \\
&= 1
\end{aligned}$$

Für welche δ lohnt es sich, nicht abzuweichen?

$$\begin{aligned}
1 &\geq 1 + \delta^T - 2\,\delta^{T+1} \\
0 &\geq \delta^T\,(1 - 2\,\delta) \qquad \delta \neq 0 \\
0 &\geq 1 - 2\,\delta \\
\delta &\geq \frac{1}{2}
\end{aligned}$$

Gefangenendilemma mit N Spielern

Grim-Strategien

N Spieler, N gerade und für jede Runde zufällige Auswahl zweier Spieler

Personalized Grim (persönlich verbittert):

- Spieler i wählt gegen Spieler j solange C, bis j einmal gegenüber i D gewählt hat
- NASH-Equilibrium für $\delta \geq 1 - \frac{1}{2(N-1)}$

Reputational Grim (verbittert durch Reputation):

- jeder Spieler startet mit guter Reputation
- ein Spieler erhält eine schlechte Reputation, sobald er D gegenüber einem Spieler mit guter Reputation spielt
- Spieler i wählt C gegen Spieler mit guter Reputation und D gegen Spieler mit schlechter Reputation
- NASH-Equilibrium für $\delta \geq \frac{1}{2}$

Whitewashing

Allgemein:

- Teilnehmer mit schlechter Reputation registrieren sich neu im System (andere Identität)
- gerade im Internet leicht

Im Gefangenenmodell:

- D wählen und dann neue Identität
- klappt bei allen betrachteten Strategien

$\Rightarrow$ Reputationssystem nutzlos

Eintrittsgebühr

Eintrittsgebühr f verhindert Whitewashing

- wenn hinreichend hoch – *wie hoch?*
- *Reputational Grim*:
  $\overline{\pi} = (1-\delta)\,(-f + 1 + \delta + \delta^2 + \delta^3 + \ldots)$
- Abweichung ist wegen δ am wertvollsten in der ersten Runde
- *einmalige Abweichung in Runde* 1 *mit Whitewashing*:

$$\begin{aligned}\overline{\pi}' &= (1-\delta)\,(-f + 2 + (-f+1)\,\delta + \delta^2 + \delta^3 + \ldots) \\ &= (1-\delta)\,(-f - f\,\delta + 1 + 1 + \delta + \delta^2 + \delta^3 + \ldots)\end{aligned}$$

- Für welche f ist $\overline{\pi} \geq \overline{\pi}'$?

$$\begin{aligned}-f &\geq -f - f\,\delta + 1 \\ f &\geq \frac{1}{\delta}\end{aligned}$$

- Bspw. $\delta = \frac{1}{2} \;\Rightarrow\; f = 2$

Strategie PYD

Pay Your Dues!

Unterscheidung der Teilnehmer in

- *Neulinge* – spielen zum ersten Mal
  - spielen C
- *Veteranen* – haben schon einmal gespielt
  - spielen D gegen einen Neuling
  - spielen C gegen einen Veteranen, gdw. er dieser Strategie folgt.

Neulinge zahlen Eintrittsgebühr, wenn sie gegen Veteranen spielen.

- Mißtrauen gegenüber Neulingen
- Strategie ist Nash-Equilibrium
- aber: Summe der Gewinne aller Spieler kleiner als bei Reputational Grim ohne Whitewashing

Teil II

Ehrlichkeit und Objektivität

Unterversorgung und Unehrlichkeit

weitere Probleme

Typische Herausforderungen bei Systemen, die auf Meinungen basieren:

- Unterversorgung
  - Formulieren und Berichterstatten ist Aufwand
  - benötigt Zeit
  - „bevor schlechte Bewertung, lieber keine“
- Unehrlichkeit
  - Nettigkeitsbedürfnis, Bösartigkeit
  - Furcht vor Vergeltung
  - Interessenskonflikte

$\Rightarrow$ verzerrtes Bild

Lösungsansätze

baldige Verfügbarkeit objektiver Informationen:

- Belohnungssystem für Ehrlichkeit
- Beispiel: Wettervorhersagen, Sportwetten

Aber objektive Informationen meist nicht vorhanden!
Beispiel: Produktbewertung
Oder nicht öffentlich:
Beispiel: reale Ausfallhäufigkeit eines Produkts

- Vergleich zwischen Bewertern
- Übereinstimmung belohnen
- Problem: wahre Ausnahmen
- Beispiel: Verkäufer hat sehr gute Bewertungen bei Online-Auktion
  ein Käufer macht schlechte Erfahrung
  wird dies nicht zugeben

Simultaneous Reporting Game (SRG)

Spiel des gleichzeitigen Bewertens

- Spieler $i \in \{1, \ldots, N\}$ (auch $N = \infty$) beurteilt ein Produkt
- Produkt hat Qualitätsstufe (**Typ**) $t \in \{1, \ldots, T\}$, $T < \infty$
- Spieler i nimmt Qualität (**Signal**) $S^i \in S = \{s_1, \ldots, s_M\}$ wahr
- Spieler i sendet Bewertung $w^i \in S$ an Zentrum Z
  - $w^i = S^i \;\Rightarrow\;$ Spieler ist ehrlich
  - $w^i \neq S^i \;\Rightarrow\;$ Spieler ist unehrlich
  - $w^i_m :=$ Bewertung von Spieler i, wenn $S^i = s_m$
- sind alle N Bewertungen $w = \left(w^1, \ldots, w^N\right)$ bei Z, werden diese offengelegt und Z verteilt Punkte
  - $\pi_i(w) =$ Punktzahl an Spieler i
  - $\pi(w) = (\pi_1(w), \ldots, \pi_N(w))$ ist Vektor der Punkte an alle Spieler

die Peer-Prediction-Methode (PPM)

Einleitung

- realisiert Vergleichen der Peers
- Vergleich von Wahrscheinlichkeiten der möglichen Bewertungen und der tatsächlichen Bewertung eines anderen Bewerters (**Referenzbewerter**)
- Punkte sollen motivieren ehrlich zu bewerten

Annahmen und Definitionen

- $\mathbf{Pr}_0(t)$ = Ausgangswahrscheinlichkeit für Typ t;
  $\mathbf{Pr}_0(t) > 0, \quad \sum_{t=1}^{T} \mathbf{Pr}_0(t) = 1$
- Signal von Spieler i ist Zufallsvariable S^i
- $f(s_m, t) := \mathbf{Pr}(S^i = s_m \mid t)$ unabhängig und gleichverteilt;
  $f(s_m, t) > 0, \quad \sum_{m=1}^{M} f(s_m, t) = 1$ für festes t
- $\mathbf{Pr}_0(t)$ und $f(s_m, t)$ darf allen Spielern bekannt sein

Optimale Strategie

Spieler i spielt die *optimale Bewertungsstrategie*, wenn für jedes $m = 1, \ldots, M$ und alle Bewertungen $\hat{w}^i \in S$ gilt:

$$\sum_{j \leq 0, j \neq i} \pi_i \left(w_m^1, w_m^2, \ldots, w_m^i, \ldots, w_m^N \right) \mathbf{Pr}(P = \pi_i (\ldots) \mid S^i = s_m) \geq$$

$$\sum_{j \leq 0, j \neq i} \pi_i \left(w_m^1, w_m^2, \ldots, \hat{w}^i, \ldots, w_m^N \right) \mathbf{Pr}(P = \pi_i (\ldots) \mid S^i = s_m)$$

d. h.

- unter der Bedingung $S^i = s_m$
- wird die (bedingte) erwartete Punktzahl maximiert

NASH-Equilibrium, wenn Ungleichung für $i = 1, \ldots, N$ erfüllt

Punkteregel

- Punkteregel $\mathcal{T}(s, w^i) \in \mathbb{R}$ gibt an, wieviele Punkte für Bewertung s gegeben werden, wenn Bewertung w^i eines anderen Spielers vorliegt.
- **streng geeignet**, wenn Spieler seine erwartete Punktzahl maximiert durch *wahrheitsgemäße* Bewertung
- Beispiel: logarithmische Regel
  - $\mathcal{T}(s, w^i) = \ln \mathbf{Pr}(S^j = s \mid S^i = w^i)$
  - bestraft Spieler durch den Logarithmus (≤ 0) der (vom Spieler gegebenen) W'keit des Ereignisses, das eingetreten ist.
  - Varianten: Multiplikation/Addition mit Konstante
- Beispiel: quadratische Regel
  - $\mathcal{T}(s, w^i) = 2\,\mathbf{Pr}(S^j = s \mid S^i = w^i) - \sum_{x \in S} \mathbf{Pr}(S^j = x \mid S^i = w^i)^2$
- Beispiel: sphärische Regel
  - $\mathcal{T}(s, w^i) = \frac{\mathbf{Pr}(S^j = s \mid S^i = w^i)}{\sqrt{\sum_{x \in S} \mathbf{Pr}(S^j = x \mid S^i = w^i)^2}}$

Punktevergabe

Satz

Für $\pi_i^*(w^i, w^{r(i)}) = \mathcal{T}(w^{r(i)}, w^i)$, streng geeignetem $\mathcal{T}$ und beliebiger Zuordnung r mit $r(i) \neq i$ ist wahres Bewerten ein Nash-Equilibrium.

Beweisskizze.

- $r(i)$ bewertet wahrheitsgemäß
- Spieler i will $\sum_{s' \in S} \mathcal{T}(s', w^i)\, \mathbf{Pr}(S^{r(i)} = s' \mid S^i = s^*)$ maximieren
- $\mathcal{T}$ ist streng geeignet

$\Rightarrow$ Summe wird maximiert durch $w^i = s^*$ (Wahrheit) $\square$

Bemerkung: $r(i)$ ist **Referenzspieler** von i

Beispiel: Wahrscheinlichkeiten

- Typen A, B mit $\mathbf{Pr}_0(A) = \mathbf{Pr}_0(B) = 0.5$
- Signale g (gut) und s (schlecht) mit
  $f(g, A) = 0.85, f(s, A) = 0.15, f(g, B) = 0.45, f(s, B) = 0.55$
- $\mathbf{Pr}(S^i = g) = \mathbf{Pr}_0(A) \cdot f(g, A) + \mathbf{Pr}_0(B) \cdot f(g, B) = 0.65$ und
  $\mathbf{Pr}(S^i = s) = \mathbf{Pr}_0(A) \cdot f(s, A) + \mathbf{Pr}_0(B) \cdot f(s, B) = 0.35$
- W'keit, dass β Signal von Sp. j ist, wenn α das Signal von Sp. i ist:

$$\mathbf{Pr}(S^j = \beta \mid S^i = \alpha) = \sum_{t \in \{A,B\}} f(\beta, t) \frac{f(\alpha, t)\, \mathbf{Pr}_0(t)}{\mathbf{Pr}(S^i = \alpha)}$$

$$\begin{aligned}
\mathbf{Pr}(S^j = g \mid S^i = g) &= 0.7115 \\
\mathbf{Pr}(S^j = s \mid S^i = g) &= 0.2885 \\
\mathbf{Pr}(S^j = g \mid S^i = s) &= 0.5357 \\
\mathbf{Pr}(S^j = s \mid S^i = s) &= 0.4643
\end{aligned}$$

Beispiel: logarithmische Punktevergabe

logarithmisch: erwartete Punktzahl $p(w^i, s^i)$

$$
\begin{aligned}
p(s,s) &= \ln \mathbf{Pr}(S^j = g \mid S^i = s)\,\mathbf{Pr}(S^j = g \mid S^i = s) + \\
&\quad \ln \mathbf{Pr}(S^j = s \mid S^i = s)\,\mathbf{Pr}(S^j = s \mid S^i = s) \\
&= -0.624 \cdot 0.5357 - 0.767 \cdot 0.4643 \\
&= -0.69 \\
p(g,s) &= \ln \mathbf{Pr}(S^j = g \mid S^i = g)\,\mathbf{Pr}(S^j = g \mid S^i = s) + \\
&\quad \ln \mathbf{Pr}(S^j = s \mid S^i = g)\,\mathbf{Pr}(S^j = s \mid S^i = s) \\
&= -0.34 \cdot 0.5357 - 1.243 \cdot 0.4643 \\
&= -0.76 \\
p(g,g) &= \ln \mathbf{Pr}(S^j = g \mid S^i = g)\,\mathbf{Pr}(S^j = g \mid S^i = g) + \ldots \\
&= -0.60 \\
p(s,g) &= \ln \mathbf{Pr}(S^j = g \mid S^i = s)\,\mathbf{Pr}(S^j = g \mid S^i = g) + \ldots \\
&= -0.67
\end{aligned}
$$

Teil III

Reputation aus transitivem Vertrauen

Vertrauenstransitivität

Grundlagen

Grundidee:

- Bewertungen von Spieler j sind wertvoll.
- Bekommt Spieler i positives Feedback von j, dann sind auch Bewertungen von i wertvoll.
- Vereinfachung: zeitliche Ordnung wird ignoriert

Definitionen:

- $t(i,j)$ ist Vertrauen, dass i zu j hat bzw. angibt
- $G = (V, E, t)$ ist (gerichteter) **Vertrauensgraph**, $t : E \to \mathbb{R}_{>0}$
- $F : \mathcal{G} \to \mathbb{R}^{|V|}$ ist **Aggregationsfunktion**
  - $F_v(G) \in \mathbb{R}$ ist **Reputationswert** oder **Rang** von $v \in V$
  - **monoton**, gdw. neue eingehende Kante (u, v) verringert $F_v(G)$ nicht
  - **symmetrisch**, gdw. alle Knoten v sind gleichwertig bezgl. F

Ehrlichkeit

F ist **strategiesicher**, gdw. $\forall G = (V, E, t)\ \forall v \in V$:
v kann ihren Rang durch strategische Bewertungen nicht erhöhen
(auch nicht relativ zu anderen)

$\Rightarrow$ kein Grund falsch zu bewerten

- nicht trivial zu erreichen in symmetrischen Systemen
- nicht möglich in monotonen, symmetrischen Systemen
  (Beweis: entferne ausgehende Kante zu höher-bewertetem Spieler)

Phantomfeedback

Problem: viele Schwindel-Identitäten (Phantome, *Sybils*) erzeugen gute Bewertungen über Spieler v.

- (G', U') ist **Phantomstrategie** für Spieler $v \in V$ mit $G = (V, E, t)$, gdw. $G' = (V', E', t')$, $v \in U' \subseteq V'$ und Kontraktion von U' zu v erzeugt G
- $U' - \{v\}$ sind **Phantome** von v
- F ist **wert-phantomsicher**, gdw. $\forall G = (V, E, t)\ \forall v \in V$: $\nexists$Phantomstrategie (G', U') für v, sodass $\exists u \in U' : F_u(G') > F_v(G)$
  d. h. Spieler v kann in G' stärker sein als in G
- F ist **rang-phantomsicher**, gdw. $\forall G = (V, E, t)\ \forall v \in V$: $\nexists$Phantomstrategie (G', U') für v, sodass $\exists u \in U', w \in V - \{v\} : F_u(G') \geq F_w(G')$, wenn $F_v(G) < F_w(G)$,
  d. h. Spieler v kann in G' stärker werden als Spieler w, obwohl es in G umgekehrt ist.

Beispiel: PageRank

PageRank (Page/Brin) ist ursprünglicher Rankingmechanismus von Google:

- $v \in V$ ist Webseite, $(v, w) \in E \;\Leftrightarrow\; v$ hat Hyperlink zu w
- $t(v, w) = \frac{F_{v'}(G)}{out(v)}$
- PageRank grob: $F_v(G) = d \sum_{v' | (v', v) \in E} t(v', v) + \frac{1-d}{|V|}$
- d Dämpfungsfaktor, $0 < d < 1$
- symmetrisch, monoton $\Rightarrow$ nicht strategiesicher!
- nicht rang-phantomsicher, nicht wert-phantomsicher

Beispiel: OpenPGP – Web of Trust

dezentrales System zur Sicherstellung, dass öffentliche Schlüssel ihren angeblichen Eigentümern gehören

- $v \in V$ sind Schlüssel
- $(v, w) \in E \;\Leftrightarrow\; v$ hat w signiert
- $t(v, w) =$ Distanz von v zu w
- *Strong set* $= S \subseteq V$:
  $\forall s_1, s_2 \in S : \exists$Pfad von s_1 zu $s_2 \wedge \exists$Pfad von s_2 zu s_1
- *mean shortest distance*, $\mathrm{MSD}_G(v) = \frac{\sum_{s \in S} t(s,v)}{|S|}$ minimieren
- $F_v(G) = -\mathrm{MSD}_G(v)$
- monoton, symmetrisch, nicht strategiesicher
- wert-phantomsicher, rang-phantomsicher

Beispiel: Advogato

Online-Community für Entwickler freier Software

- Mitglieder ($v \in V$) zertifizieren sich gegenseitig:
  Level $l \in \{$Apprentice, Journeyer, Master$\}$
- drei Graphen je Level:
  - G_M enthält nur Master,
  - G_J enthält Journeyer und Master,
  - G_A enthält alle Mitglieder
- Hilfsgraph $H = (\{v_0, v_1^-, \ldots, v_N^-, v_1^+, \ldots, v_N^+\}, E_H, t)$
  - *Seed*-Knoten v_0, hat feste Kanten zu vertrauenswürdige Mitglieder
  - v_i^-, v_i^+ entstehen aus originalen v_i
  - $t(v_i^-, v_i^+) =$ Kapazitäten
  - Kapazitäten $cap(v)$ basieren auf Distanz (Breitensuche!)
  - Distanz klein, Kapazität groß
  - $(v_i, v_j) \in E \;\Rightarrow\; t(v_i^+, v_j^-) = \infty$

- $F_v(H) =$ maximaler Fluß von Startknoten v_0 zu v.
- monoton, nicht symmetrisch, strategiesicher
- wert-phantomsicher, nicht rang-phantomsicher

Beispiel: P2P-Systeme

Allgemein:

- $v \in V$ ist Peer,
- $(v, w) \in E \;\Leftrightarrow\; v$ hat mit w interagiert,
- $t(v, w)$ ist Vertrauensgrad

PathRank:

- $F_v(G) =$ kürzester Weg (bezgl. t^{-1}) von $v_0 \in V$ zu v.
- monoton, nicht symmetrisch, strategiesicher
- wert-phantomsicher, rang-phantomsicher

Rang-Phantomsicherheit

Satz

Es gibt kein (nicht-konstantes) symmetrisches, rang-phantomsicheres F.

Folgt direkt aus stärkerer Aussage:

Satz

Es gibt kein (nicht-konstantes) symmetrisches, rang-phantomsicheres F, wenn $U' = \{v, u\}$, d. h. wenn nur ein Phantom existiert.

- d. h. PageRank ist nicht rang-phantomsicher und Spam-Sites brauchen nur eine weitere Seite um ihren Rang zu erhöhen
- einziger Ausweg: Symmetrie brechen durch
  - gesichtete Knoten mit a-priori-Vertrauen
  - Community-Analysen
  - ...

1-Rang-Phantomsicherheit (1)

Beweis, konstruiert Gegenbeispiel.

Sei $G_1 = (V, E_1, t_1)$ beliebig, F symmetrisch, nicht konstant:

- Sei $V = \{v_1, v_2, \ldots, v_r\}, v = v_1, w = v_r$
- Sei $F_w(G) > F_v(G)$,
- Sei G' Kopie von G mit Knoten $U = \{u_1, \ldots, u_r\}, v' = u_1, w' = u_r$
- Sei $G^k = (V \cup \{u_1, u_2, \ldots, u_k\}, E'_k, t'_k)$ und $G' := G^r$

$\Rightarrow$ Dann $F_w(G^0) > F_v(G^0)$ und $F_{w'}(G^r) = F_w(G^r)$ (Symmetrie)

$\Rightarrow$ $\exists k_0, 0 \leq k_0 < r$:

$$\begin{aligned} F_w(G^{k_0}) &> \max_{i \in \{v, u_1, \ldots, u_{k_0}\}} (F_i(G^{k_0})) \quad \wedge \\ F_w(G^{k_0+1}) &\leq \max_{i \in \{v, u_1, \ldots, u_{k_0+1}\}} (F_i(G^{k_0+1})) \end{aligned}$$

1-Rang-Phantomsicherheit (2)

Beweis (Fortsetzung).

- Sei $m = \underset{i \in \{v, u_1, \ldots, u_{k_0}\}}{\operatorname{argmax}} (F_i(G^{k_0+1}))$ Knoten mit größtem Ruf in G^{k_0+1}

$\Rightarrow$ $F_m(G^{k_0+1}) \geq F_w(G^{k_0+1})$ oder $F_{k_0+1}(G^{k_0+1}) \geq F_w(G^{k_0+1})$

$\Rightarrow$ Phantom u_{k_0+1} erhöht (in beiden Fällen) $F_v(G)$

$\Rightarrow$ Hinzufügen von lediglich u_{k_0+1} ist erfolgreiche Phantomstrategie für m in G^{k_0}

$\Rightarrow$ Es gibt Graphen, bei denen ein Phantom von v ausreicht, um $F_v(G)$ zu erhöhen $\square$

Wert-Phantomsicherheit flußbasierter F

Satz

Ist $F_v(G)$ der maximale Fluß in $G = (V, E, t)$ von einem v_0 zu v mit Kapazitäten $t(e) \forall e \in E$, so ist das auf F basierende Reputationssystem wert-phantomsicher.

Beweis.

Folgt aus *Max-Flow-Min-Cut*-Theorem, denn

- Seien C_1, C_2 die Knotenmengen des minimalen Schnitts
- $v_0 \in C_1$, $v \in C_2$
- es muss: $U' \subseteq C_2$
- $\forall u \in U' : F_u(G) \leq \sum_{v_1 \in C_1, v_2 \in C_2} t(v_1, v_2) = F_v(G)$

$\square$

Rang-Phantomsicherheit flußbasierter F

- auf maximalem Fluß basierendes F ist *nicht rang-phantomsicher*
- v kann $F_w(G)$ verringern für w, die auf Weg des maximalen Flußes liegen, indem es diesen Weg unterbricht (Kante wegnehmen)

F basierend auf kürzesten Wegen

Satz

Ist $F_v(G)$ der kürzeste Weg in $G = (V, E, t)$ von einem v_0 zu v mit Kantenlängen $\frac{1}{t(e)} \forall e \in E$, so ist das auf F basierende Reputationssystem wert- und rang-phantomsicher.

Beweis.

Wert-phantomsicher:

- $u \in U'$ kann kürzesten Weg nicht verlängern, sondern nur verkürzen

Rang-phantomsicher:

- v kann F_w nur genau dann beeinflußen, wenn v auf kürzestem Weg von v_0 zu w liegt
- dann ist der Weg von v_0 zu v aber kürzer als zu w, d. h. $F_v(G) > F_w(G)$ □

Schlusswort

„Ist der Ruf erst ruiniert,
lebt es sich ganz ungeniert.“

„Ein guter Ruf, der fünfzig Jahre währt,
wird oft durch eine schlechte Tat entehrt.“