



Combating Web Spam with TrustRank

Zoltán Gyöngyi
Hector Garcia-Molina
Jan Pedersen

"Kaiser pharmacy"

[kaiser pharmacy online](#)

kaiser pharmacy online. Stop surfing! Only here you can get info about **kaiser pharmacy** online. Click here now to read info about ...

[careers.dulcineasystems.net/ kaiser-pharmacy-online/kaiser-pharmacy-online.html](#) - 2k - [Cached](#) - [Similar pages](#)

[pretty lady at kaiser pharmacy - m4w](#)

... pretty lady at **kaiser pharmacy** - m4w. Reply to: anon-37941915@craigslist.org Date: 2004-07-30, 2:03PM PDT Seeing you really brightened up my day. ...

[www.craigslist.org/pen/mis/37941915.html](#) - [Similar pages](#)

[The Beautiful Pharmacy Girls at Kaiser](#)

... So to all of the women that work in the **Kaiser Pleasanton pharmacy**, I'd just like to let you know that the guy with the itchy nuts thinks all of you are ...

[www.craigslist.org/about/best/sfo/2705982.html](#) - [Similar pages](#)

[[More results from www.craigs](#)

[Kaiser Permanente](#)

... In addition to outpatient, inpatient, and ambulatory care, **Kaiser Permanente** maintains specialized **pharmacy** services, automated refill **pharmacies**, regional ...

[www.kaiserpermanentejobs.org/ workinghere/specialty-pharmacy.asp](#) - 25k -

[Cached](#) - [Similar pages](#)

[Kaiser pharmacy](#)

KAISER PHARMACY. Education. help with debt. purchase phentermine. Consulting. Venetian Casino. extended warranty. creatine. Free Poker Games. Phone System. ...

[www.virtuell-bilden.de/pharmacies--kaiser-pharmacy/](#) - 30k - [Cached](#) - [Similar pages](#)

[Getting Started at Kaiser Permanente](#)

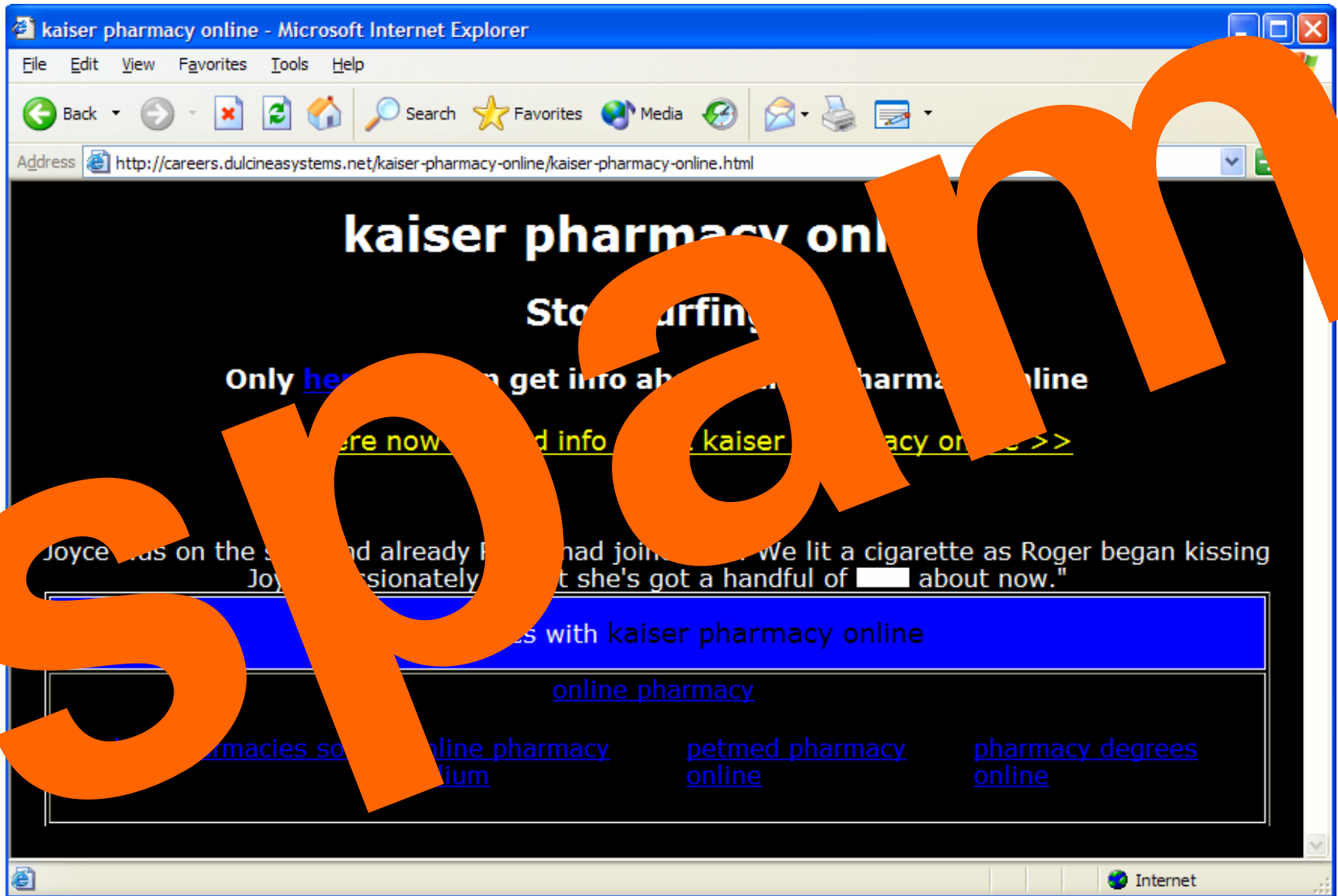
... As a member, you may transfer a prescription from a non-**Kaiser Permanente pharmacy** to any of our **pharmacies**. Be ready to tell the ...

[www.kaisersantaclara.org/choose/newmember.html](#) - 13k - [Cached](#) - [Similar pages](#)

[kaiser pharmacy online](#)

kaiser pharmacy online.
Stop surfing! Only here
you can get info about
kaiser pharmacy online.
Click here to read info
about...

"Kaiser pharmacy" (CONT'D)



Roadmap

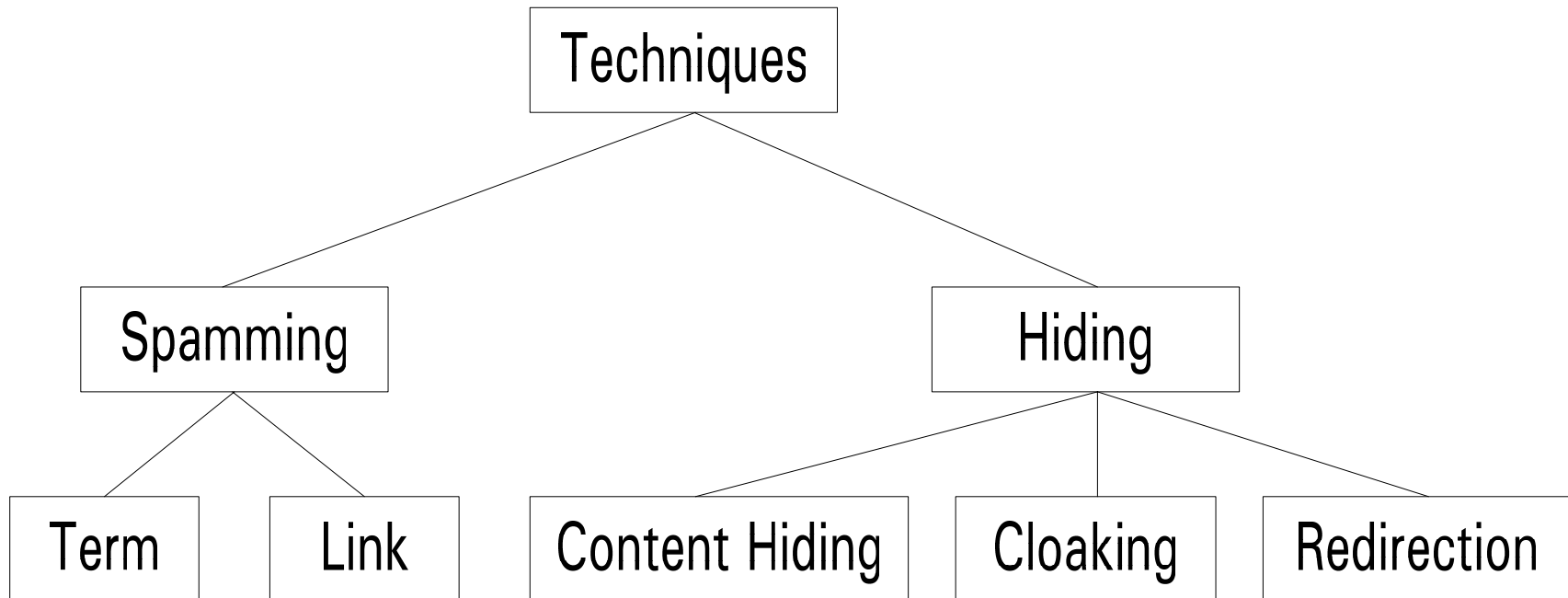
- Web spam
 - Definition, techniques
 - Observation
 - Objective
- TrustRank
 - Idea, related work
 - Score propagation
 - Algorithm
- Experiments

Roadmap

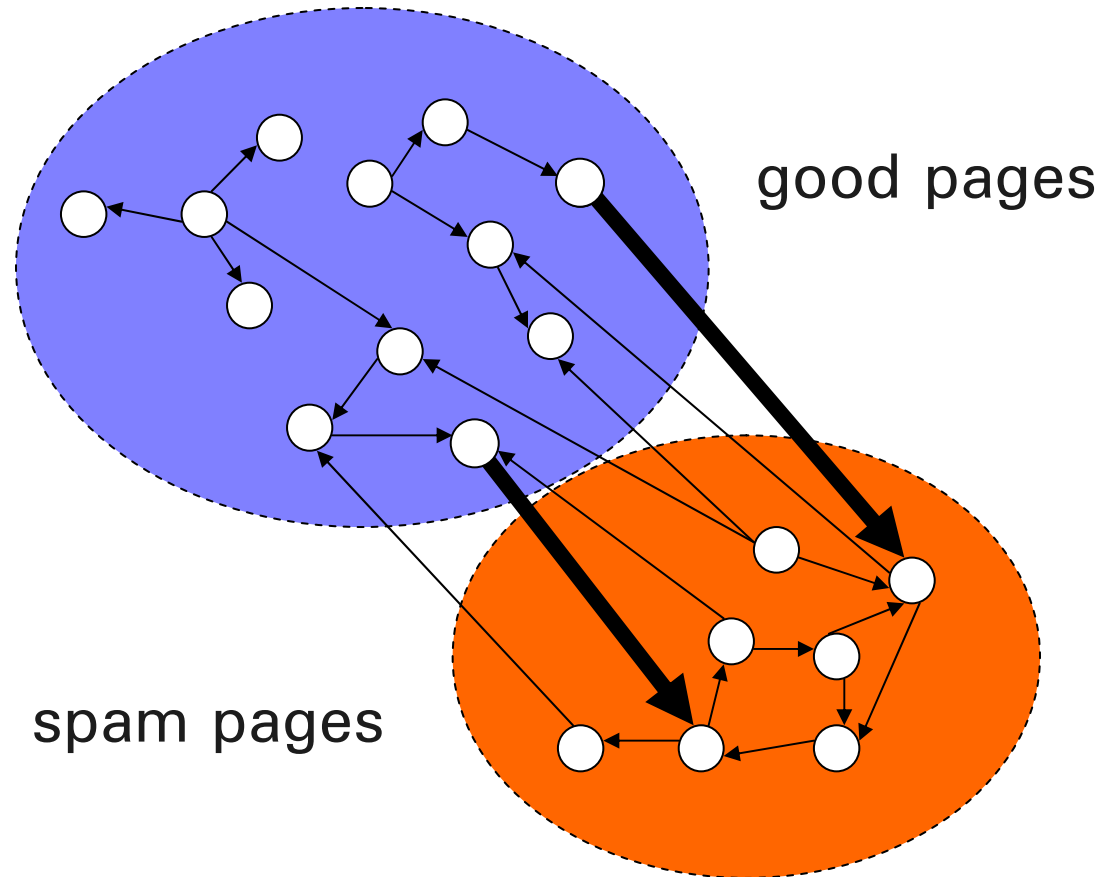
- Web spam
 - Definition, techniques
 - Observation
 - Objective
- TrustRank
 - Idea, related work
 - Score propagation
 - Algorithm
- Experiments

Web Spam / Definition

Spamming = misleading search engines to obtain higher-than-deserved ranking



Web Spam / Observation



***Approximate isolation of good pages:
good pages seldom point to spam***

Web Spam / Objective

*Separate good pages from
spam pages*

What?

Assign high scores to very good pages

How?

Propagate scores from known good pages

When?

Use results in ranking

Roadmap

- Web spam
 - Definition, techniques
 - Observation
 - Objective
- TrustRank
 - Idea, related work
 - Score propagation
 - Algorithm
- Experiments

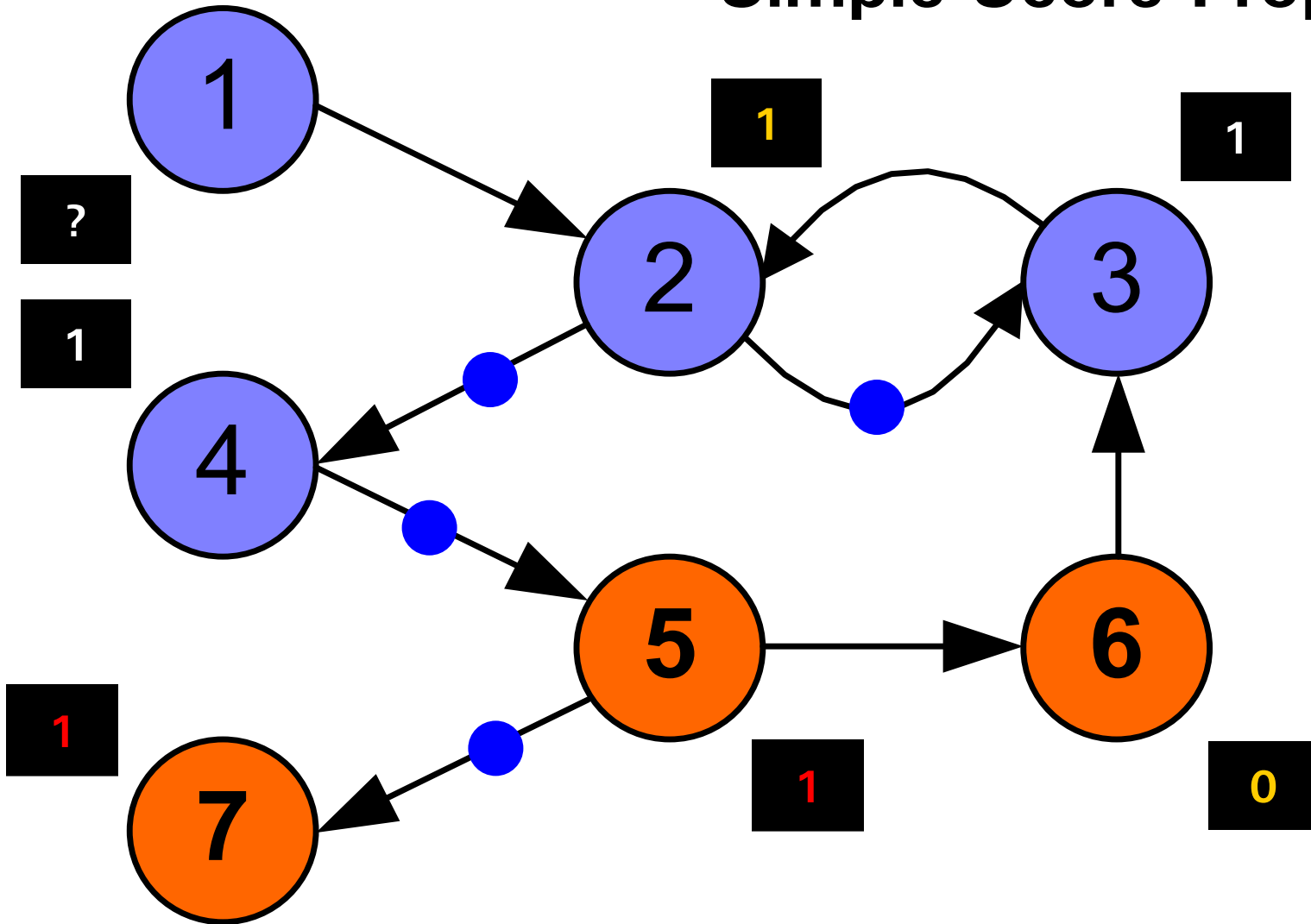
TrustRank / Idea

***TrustRank =
Selection + Propagation***

- Our approach: spam demotion
- Related work: spam detection
 - Text → machine learning
 - Link → graph clustering

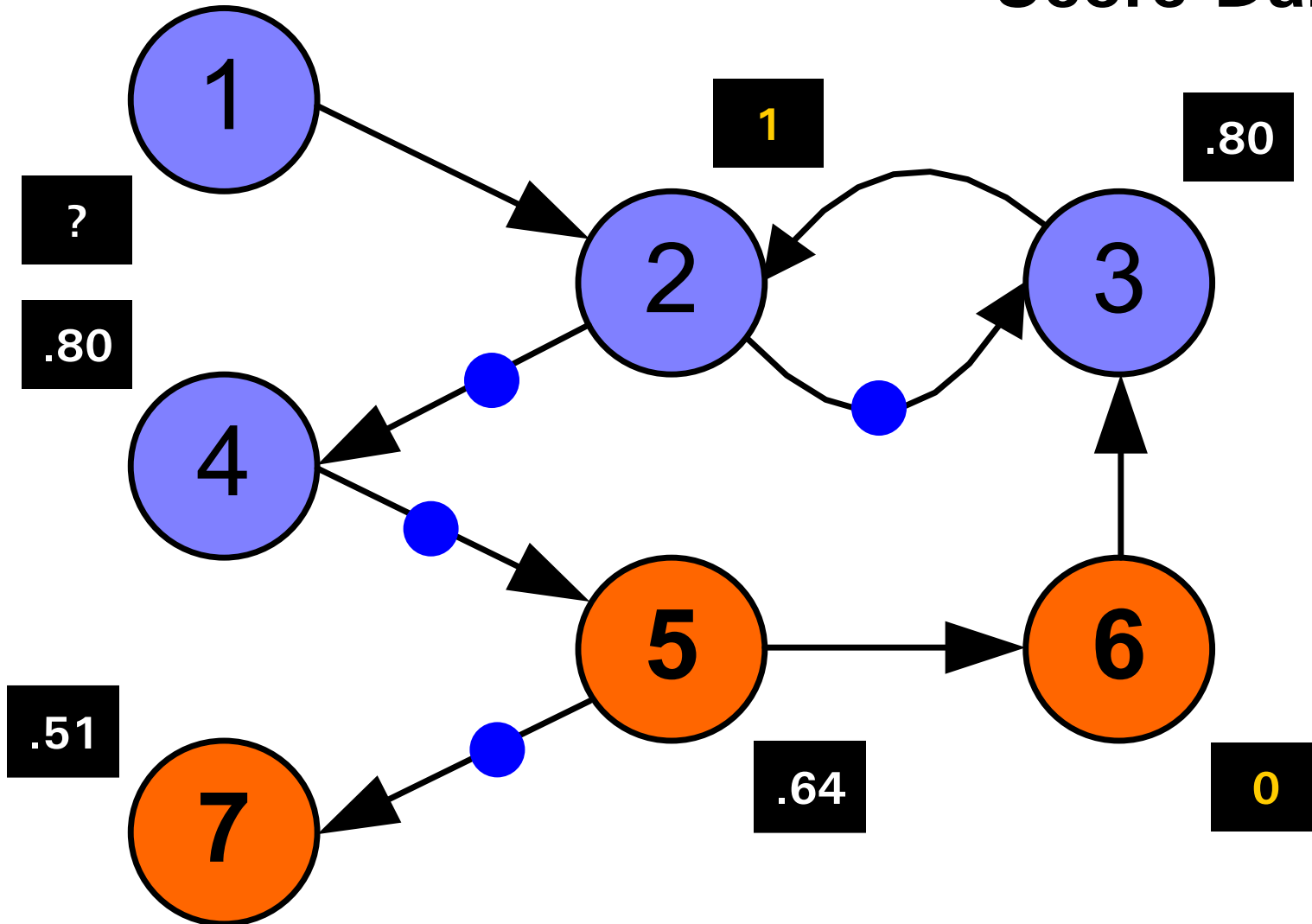
TrustRank / Propagation

Simple Score Propagation



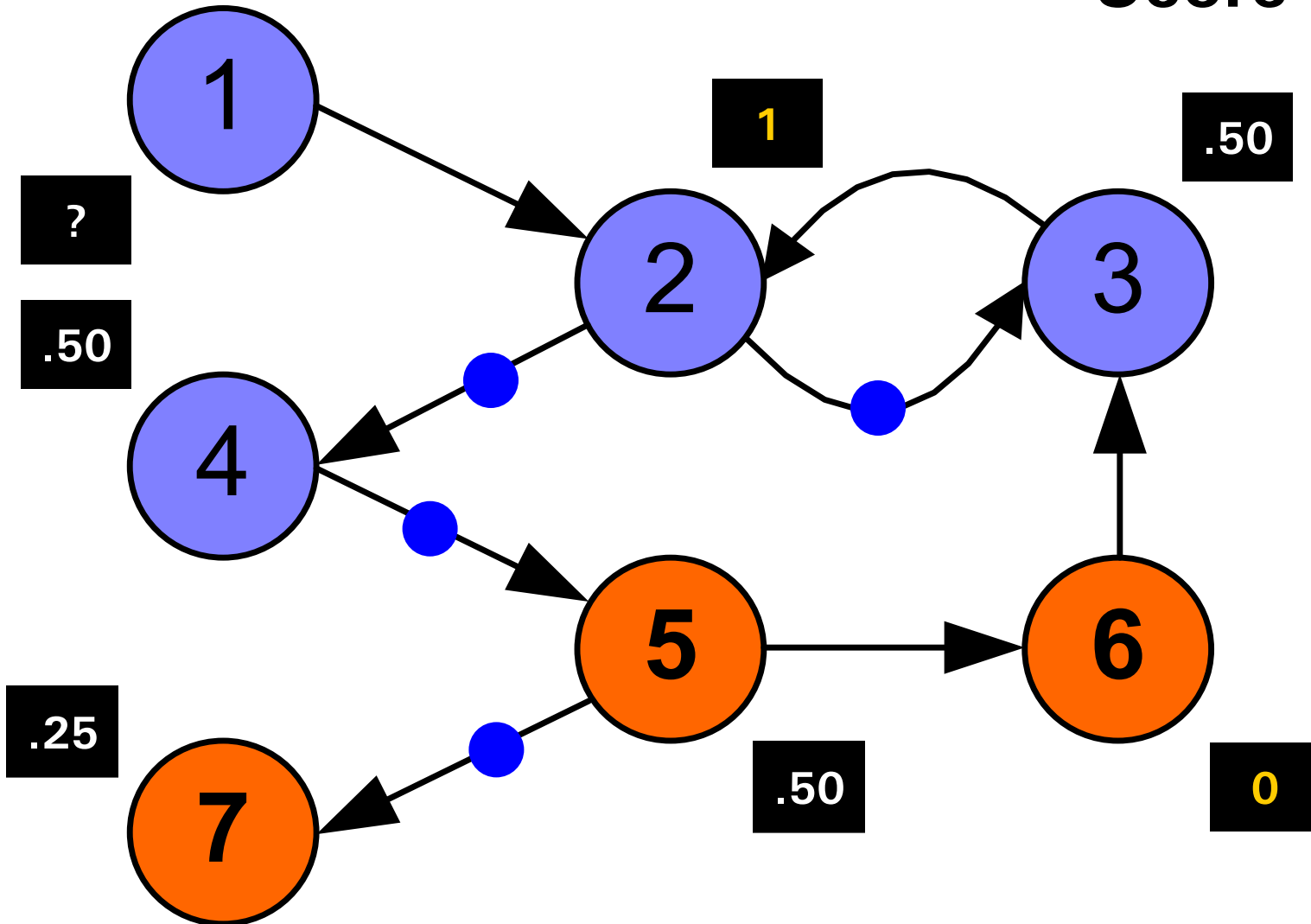
TrustRank / Propagation (CONT'D)

Score Dampening

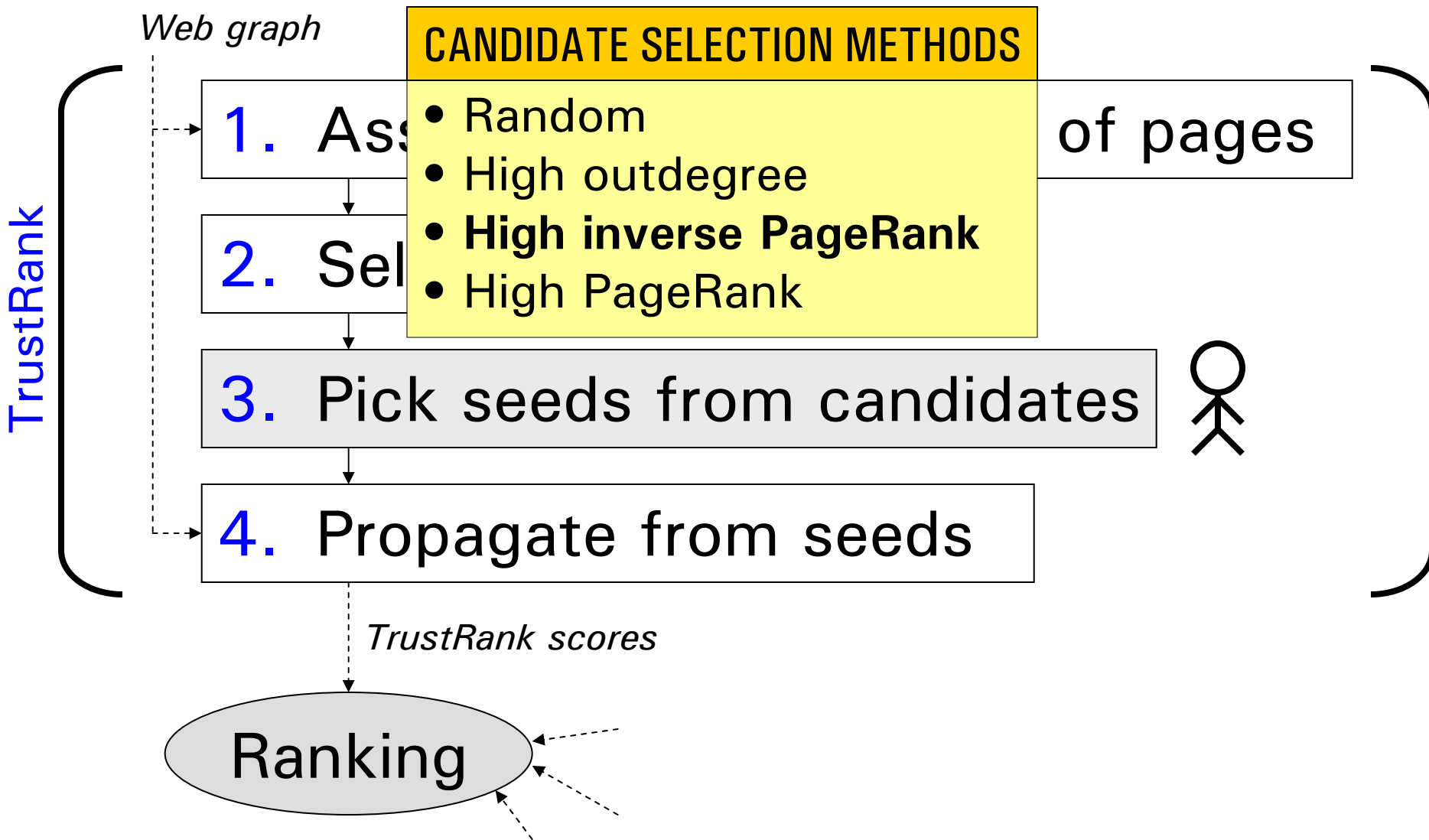


TrustRank / Propagation (CONT'D)

Score Splitting



TrustRank / Algorithm



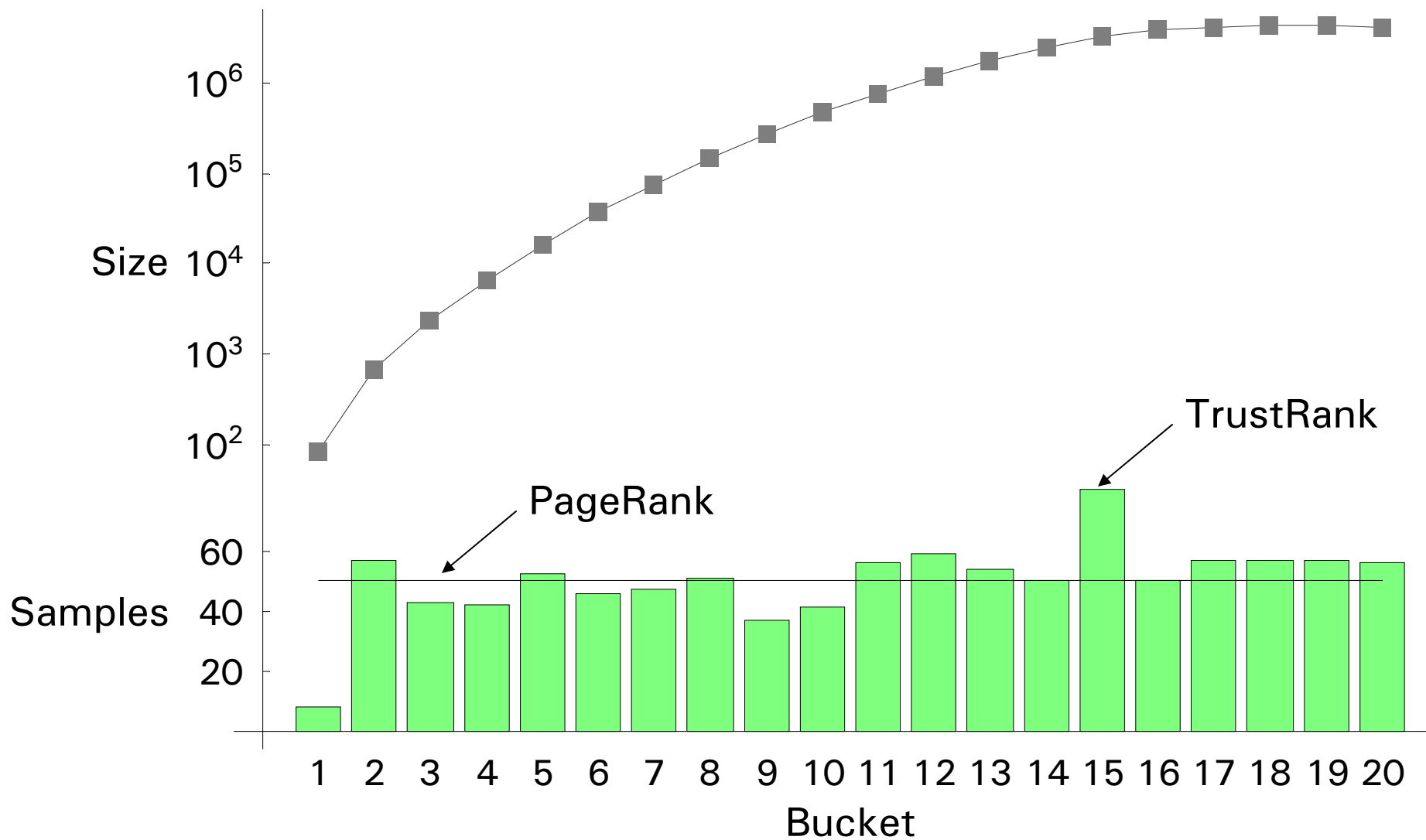
Roadmap

- Web spam
 - Definition, techniques
 - Observation
 - Objective
- TrustRank
 - Idea, related work
 - Score propagation
 - Algorithm
- Experiments

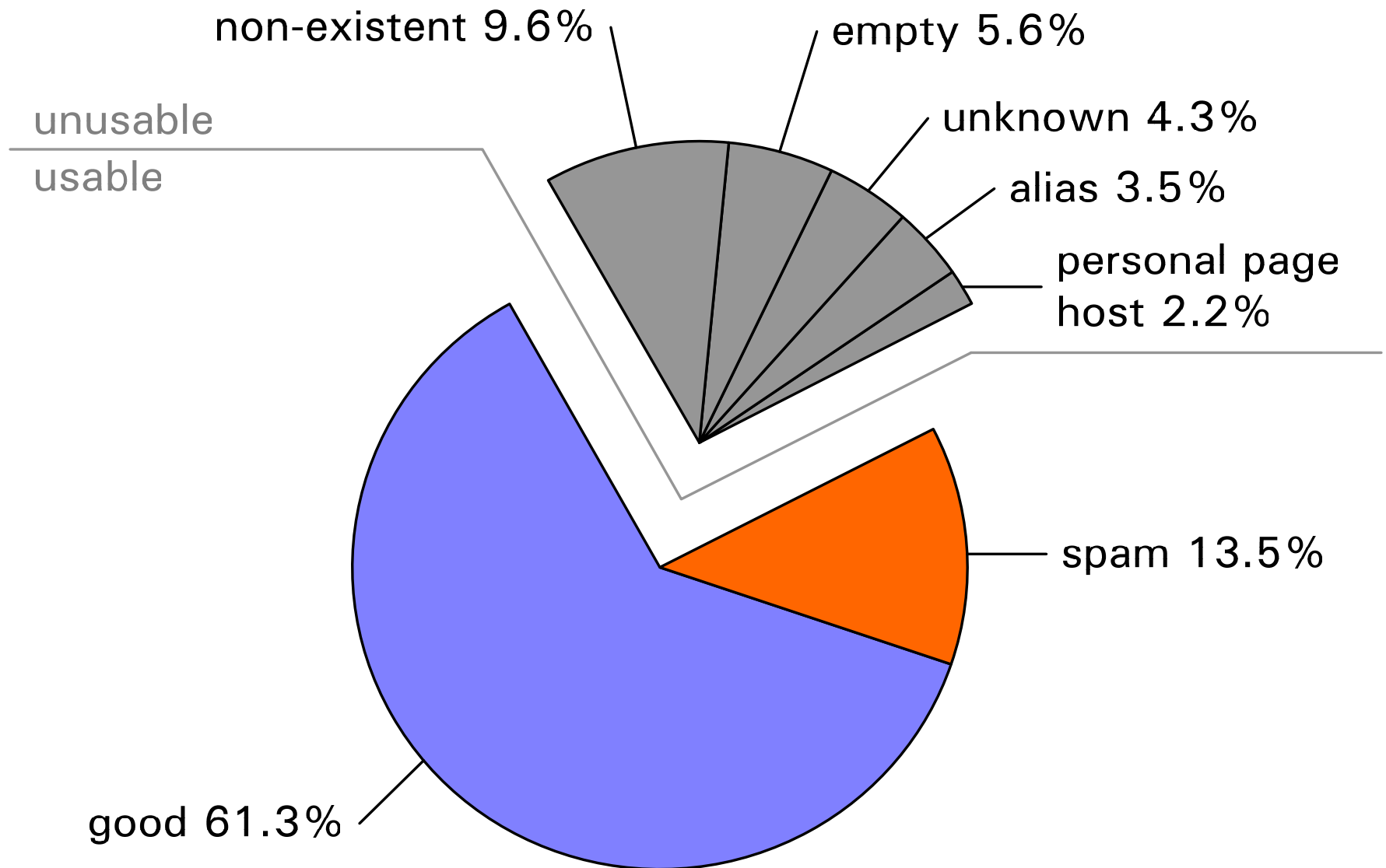
Experiments / Data

- Web data
 - Entire AltaVista index (June 2003)
 - Site-level web graph
 - 31M nodes
 - 13M without inlinks
- Seed set
 - 2500 candidates
 - 178 selected high-quality sites
- Evaluation sample
 - 1000 manually tagged sites

Experiments / Sampling

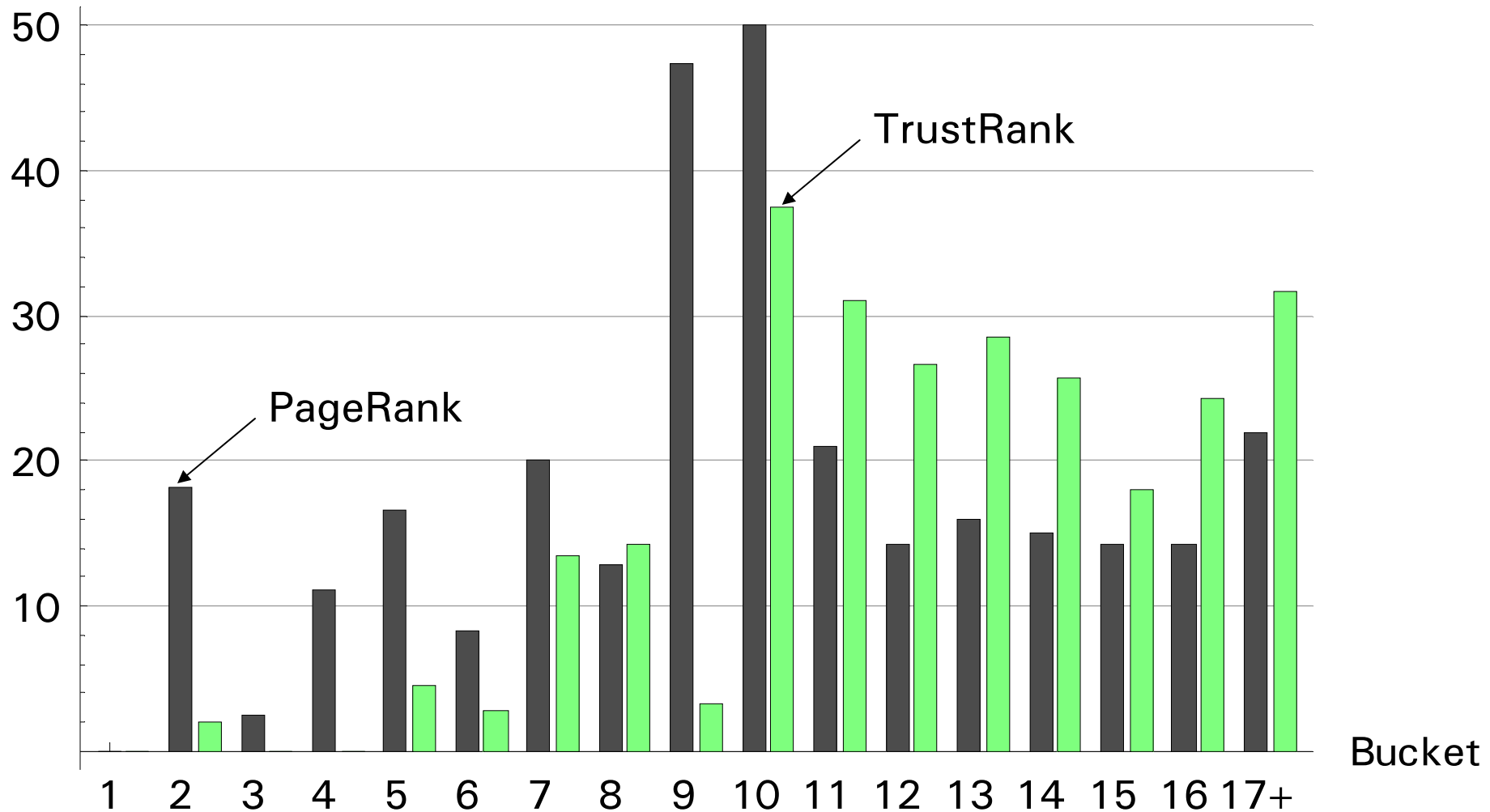


Experiments / Sample Composition

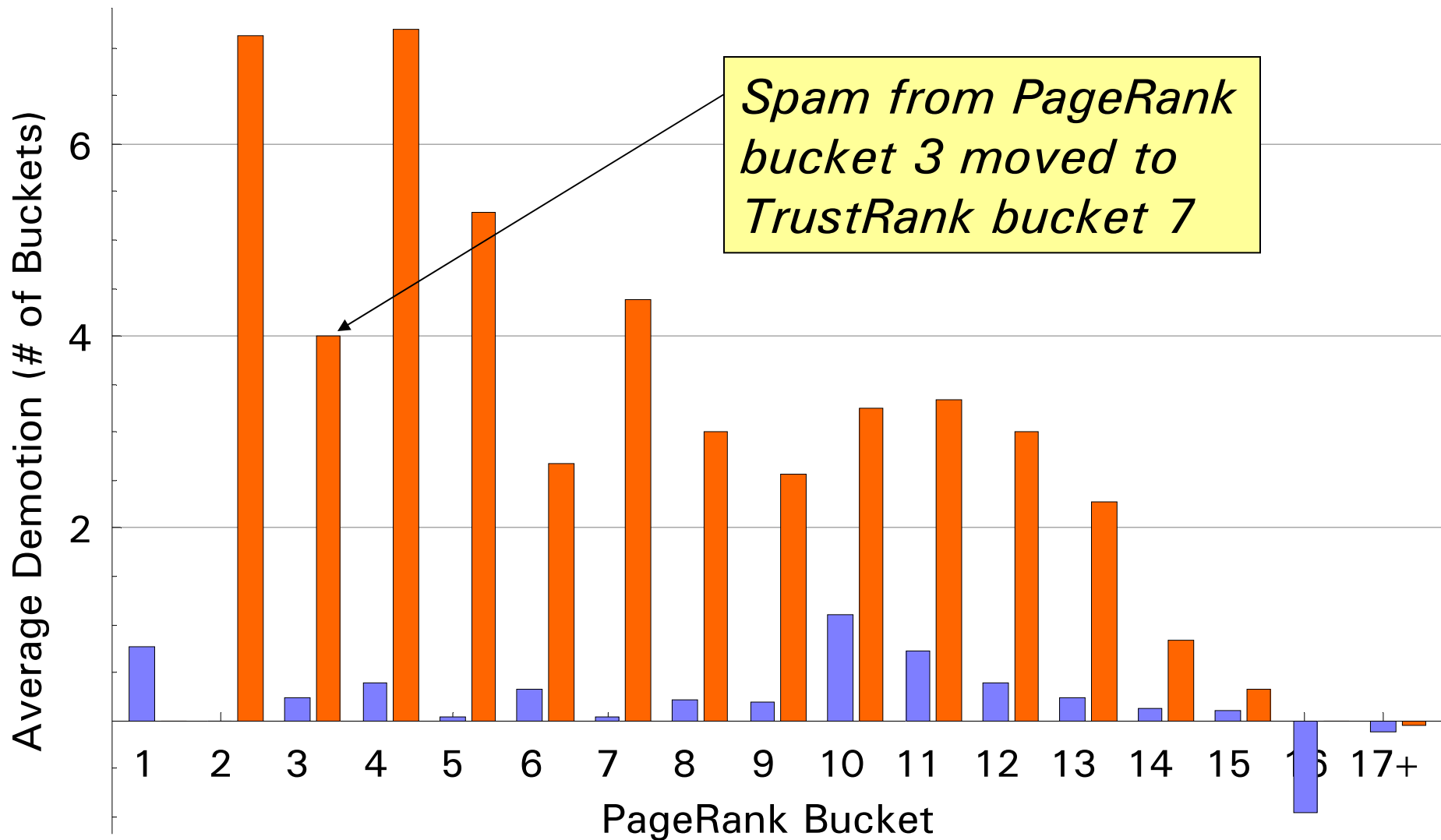


Experiments / Spam per Bucket

% Spam



Experiments / Spam Demotion

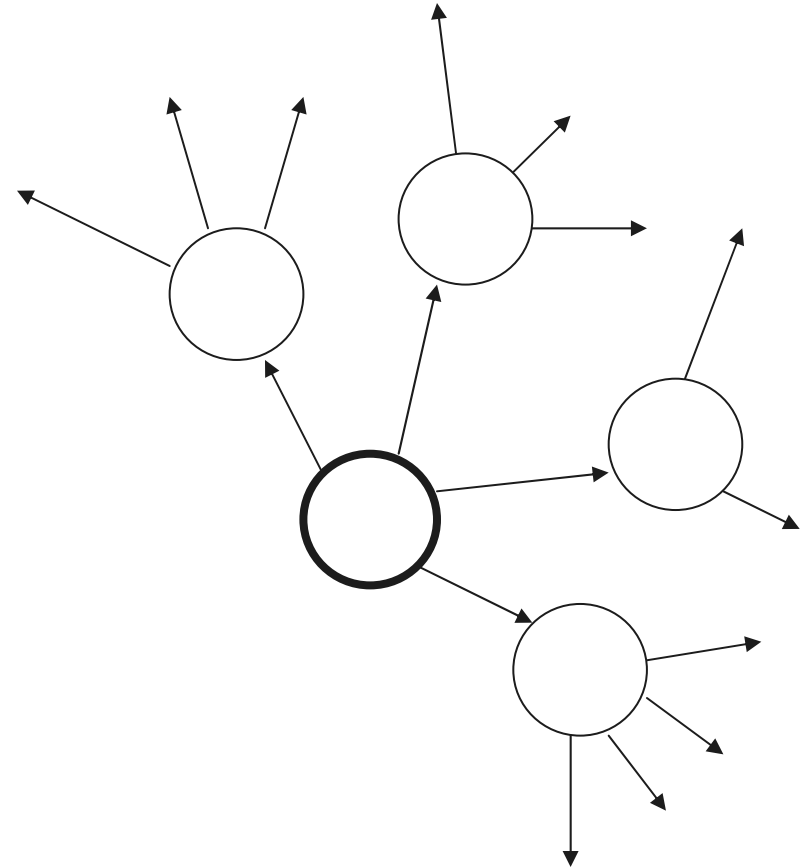
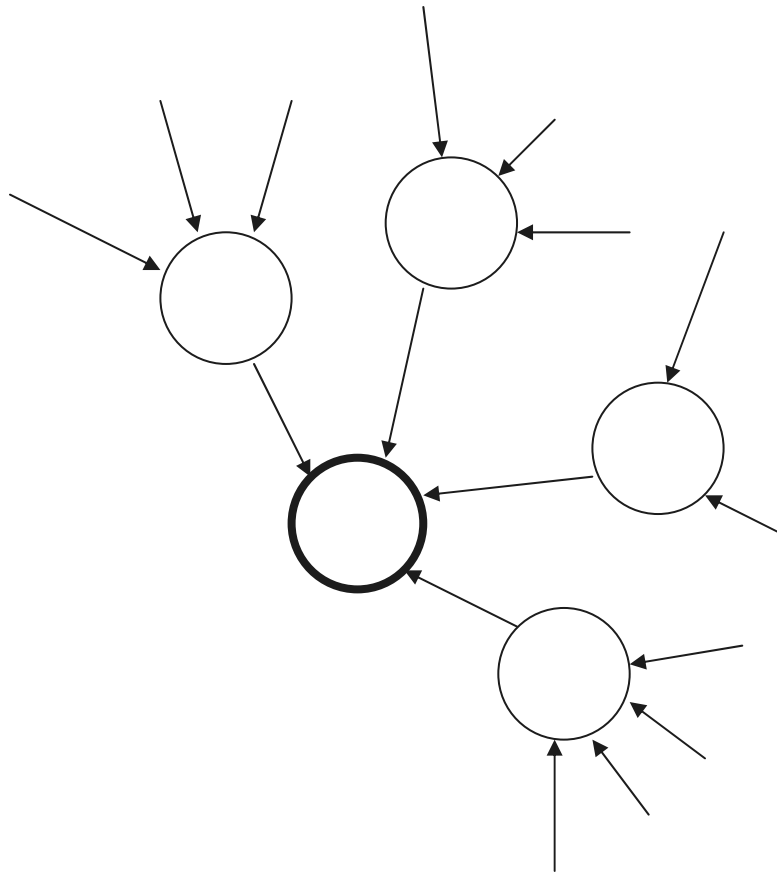


Conclusions

- TrustRank
 - **Selects** good seed pages
 - **Propagates** scores to other good pages
 - **Separates** good from bad
- Experiments
 - Spam demoted by TrustRank
- Future work
 - Different seed selection and score propagation methods
- Contact: zoltan@cs.stanford.edu

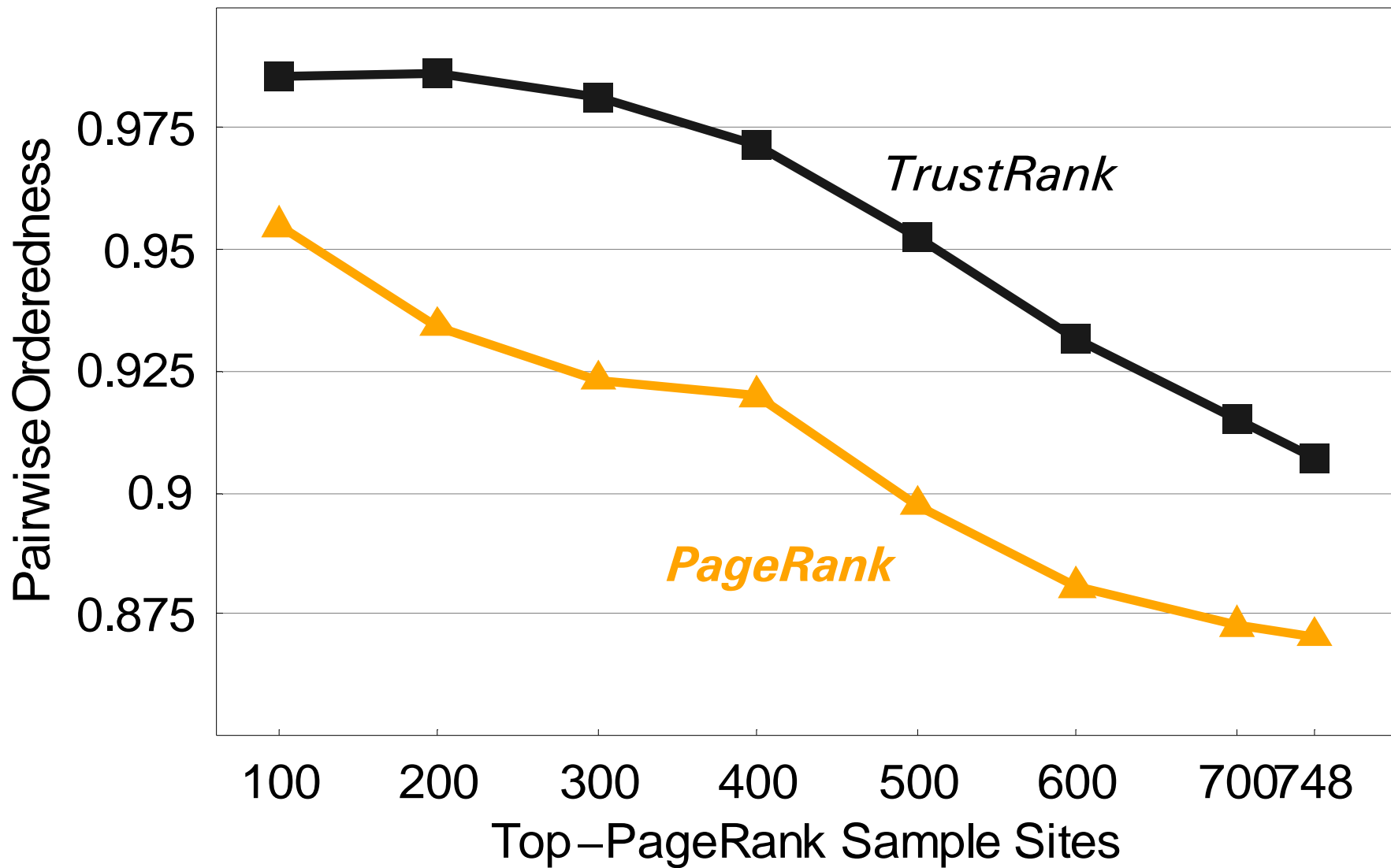
Inverse PageRank

PageRank ~ inlinks



**inverse PageRank
~ outlinks**

Experiments / Pairwise Orderedness



Experiments / Precision and Recall

