Detecting Web Spam through Backward Propagation of Distrust

CS315-Web Search and Mining

Anti-Spam Lessons from Society

What would you do if you realize that you should not trust a member of your trust network?

Anti-Propagandistic Lessons for Web

- How do you deal with propaganda in real life?
- Backwards propagation of distrust
- The recommender of an untrustworthy message becomes untrustworthy
- Can you transfer this technique to the web?

An Anti-Propagandistic Algorithm

- Start from untrustworthy site $s$
  - $S = \{s\}$
- Using BFS for depth $D$ do:
  - Find the set $U$ of sites linking to sites in $S$
    - (using the Google API for up to 8 b-links/site)
    - Ignore blogs, directories, edu's
    - $S = S + U$
  - Find the bi-connected component BCC of $S$ that includes $s$
- BCC shows multiple paths to boost the reputation of $s$
Backwards Propagation of Distrust

- Start from untrustworthy site $s$
- $S = \{s\}$
- Using BFS for depth $D$ do:
  - Find the set $U$ of sites linking to sites in $S$ (using the Google API for up to $B$ b-links/site).
  - Ignore blogs, directories, edu's.
  - $S = S + U$
- Find the bi-connected component $BCC$ of $S$ that includes $s$

$BCC$ shows multiple paths to boost the reputation of $s$

---

BCC vs Periphery

- Since the $BCC$ reveals multiple paths to boost the reputation of $s$, we expect it to contain a higher percentage of untrustworthy sites.
- The Periphery of the $BCC$ on the other hand, should have significantly lower percentage of untrustworthy sites.

---

Explored neighborhoods

- The trustworthiness of starting site is a very good predictor for the trustworthiness of $BCC$ sites.
- The $BCC$ is significantly more predictive of untrustworthiness than the Periphery.

---

Evaluated Experimental Results