Optimal Link Bombs are Uncoordinated

Sibel Adali

Tina Liu

Malik Magdon-Ismail

Rensselaer Polytechnic Institute

Pagerank

- Pagerank algorithm models the behavior of a random surfer when at a specific web page v will either
  - Jump to a random page with probability (1-α), or
  - Choose a link from page v uniformly and follow this link with probability α.
- The pagerank p_i of a page v_i then models the probability of being at that page. It satisfies the following equation:

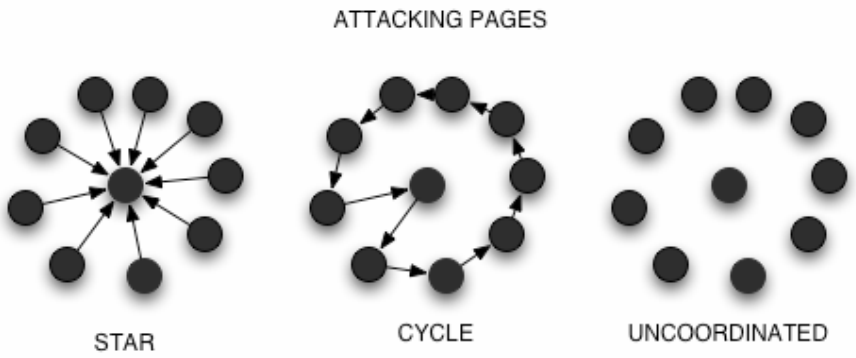
$$p_i = \alpha \sum_{(v_j, v_i) \in E} \frac{p_j}{outdeg(v_j)} + \frac{1-\alpha}{N}$$

Link bombing

- A set of pages $A = \{v_1,\ldots,v_k\}$ would like to boost the prominence of a page $v_0 \notin A$.
  - The score of page v_0 with respect to a keyword query Q is computed by a combination of
    - the number of times keywords in Q appear in page v_0,
    - the number of times keywords in Q appear in links pointing to v_0, and
    - the pagerank of v_0.
  - Pages are then sorted with respect to their scores and ranks are computed.
  - The only thing that the attacking pages can control is their own content and links.
- Problem: what is the best way to boost the rank of v_0 ?

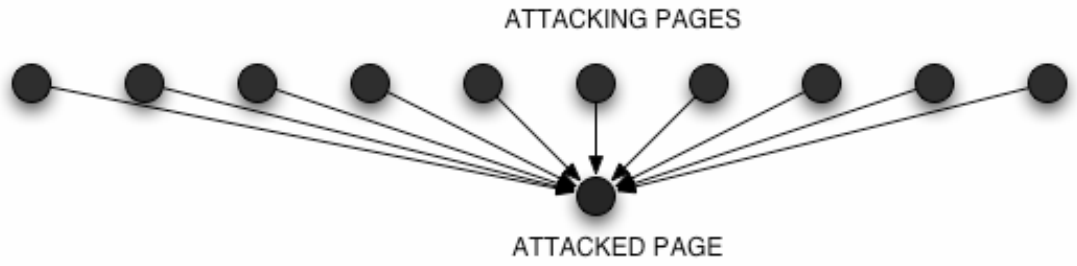
Coordination of the attack

- To improve the rank of page v_0 for query Q, add links with keyword Q to page v_0.
- What is the best link structure for attack?
  - Is there a benefit to adding additional links among attackers to improve their pagerank?
  - How many links should be added to the attacked page?



Optimal attack

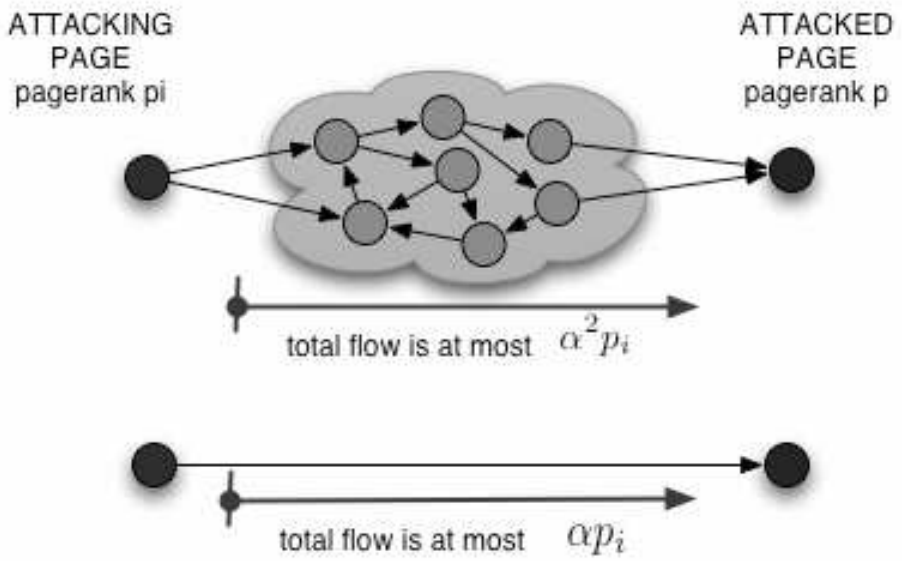
- The attack that optimally boosts the rank of a page with respect to pagerank is uncoordinated!
  - Attackers do not improve the effectiveness of their attack by adding links among themselves.
  - Attack improves as more attacking pages link to the attacked page.
  - The best attack by any page is to remove all outgoing links and only point to the attacked page. The number of links per page is not important in this case.
  - If there are other outgoing links, then as more links are added to the attacked page, the effectiveness of the attack will improve.



Why?

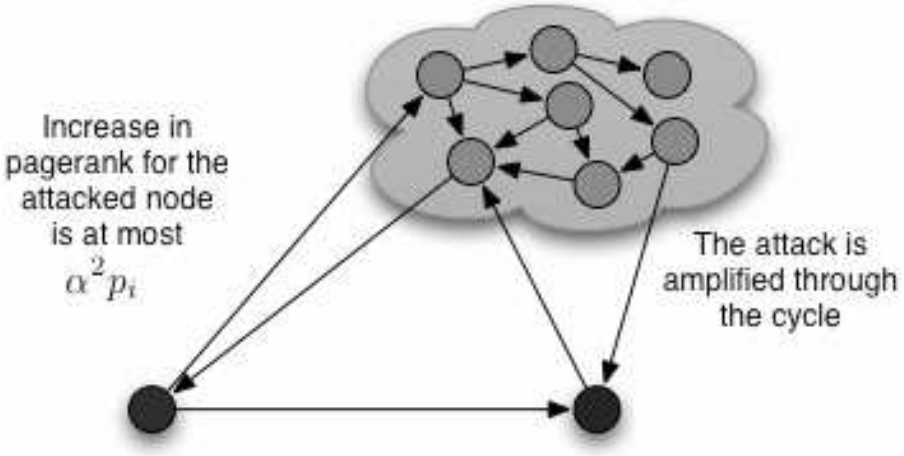
$$p_i = \alpha \sum_{(v_j, v_i) \in E} \frac{p_j}{outdeg(v_j)} + \frac{1-\alpha}{N}$$

- Each new link (v_j , v_i) introduces a new flow
  - Directs the pagerank of v_j to v_i
  - Any other outgoing link from v_j diverts a portion of the flow away from v_i
  - The highest flow is achieved with shortest path from the attacking page to the attacked page



Cycles

- Cycles improve the pagerank of a page due to the iterative nature of the algorithm
  - It is possible to visit the same page multiple times through cycles
  - The amplification of the pagerank at the attacked page through cycles is a monotonically increasing function of the increase in flow at the attacked page
  - Optimize the total flow to the attacked page with shortest route and least number of outgoing links

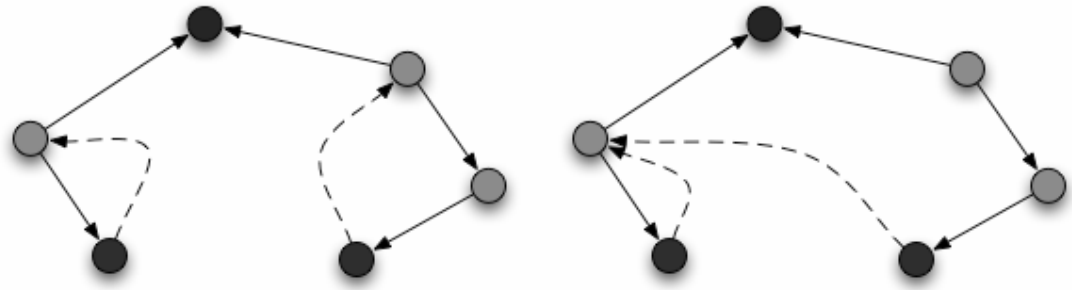


Rank

- The uncoordinated attack is also optimal in improving the rank of an attacked page with respect to its pagerank
  - The direct attack is best for page v_i independent of what other attackers do
  - Suppose by contradiction the rank of attacked page v_0 is less than some other page u in another attack type A, then it is for the uncoordinated attack.
    - It must be that pagerank of u is higher than pagerank of v_0 for attack A, but it is less in the uncoordinated attack.
    - But this is not possible since the uncoordinated attack maximizes the pagerank increase of v_0.

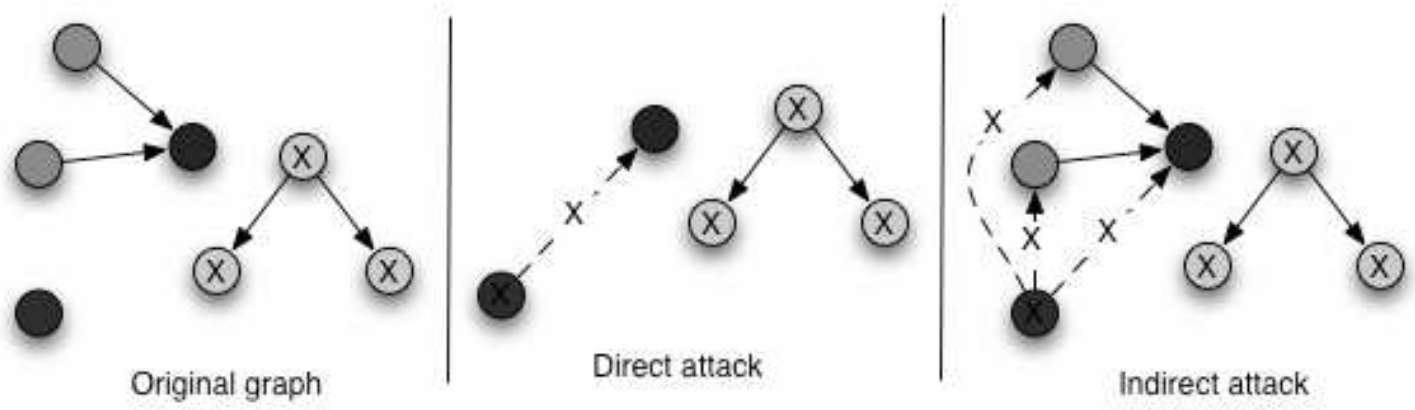
Optimal disguised attack

- Suppose attackers want to hide by not pointing directly to the attacked page.
  - Choose among the pages with required anchor text to point to.
  - If the objective is to be a distance of L hops away from the attacked page, choose a single page to point to L-1 hops away that maximized the flow of pagerank from the attacked page to the victim.
  - The individual optimal attack is not necessarily the same as optimal joint attack.



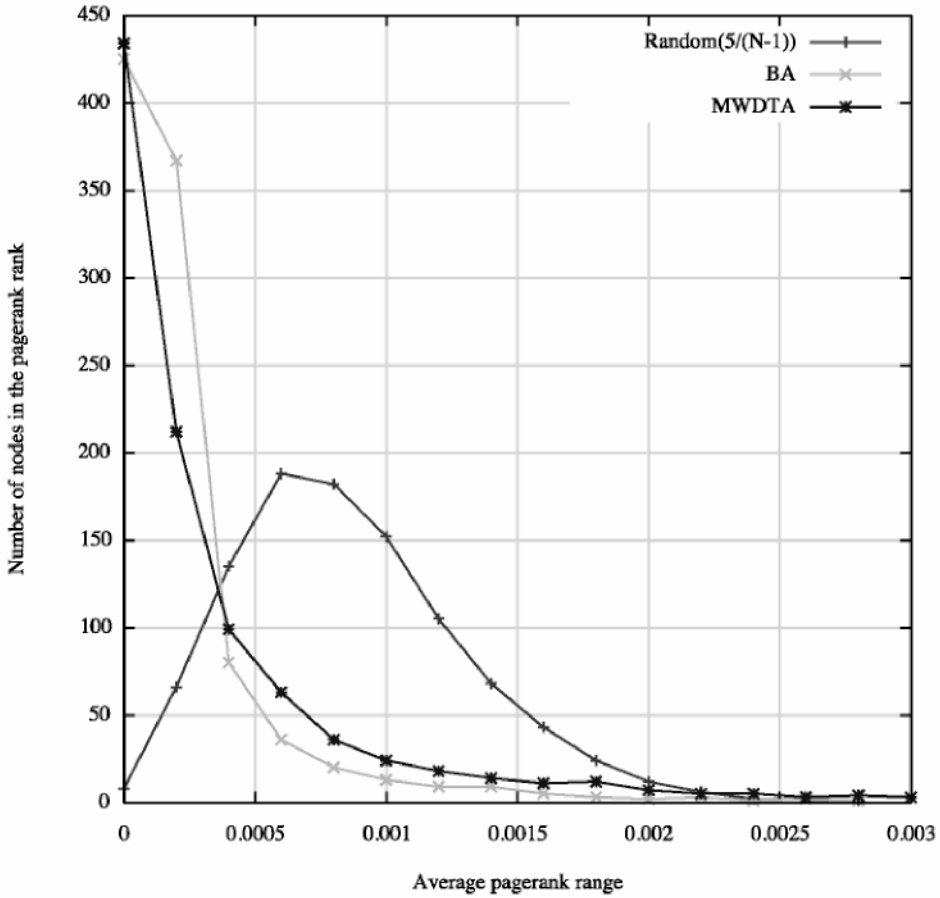
Effect of keywords

- Bombing improves pagerank for all keyword queries.
  - When keywords in links are considered in the ranking, then the link bombing is particularly effective.
  - In general, the probability of the same text appearing in links pointing to the same page may be low, but much higher for attacks.
- What if pagerank was computed only for the graph induced by the given query?
  - The optimal attack is no longer the direct individual attack.



Experimental results

- Different graph types

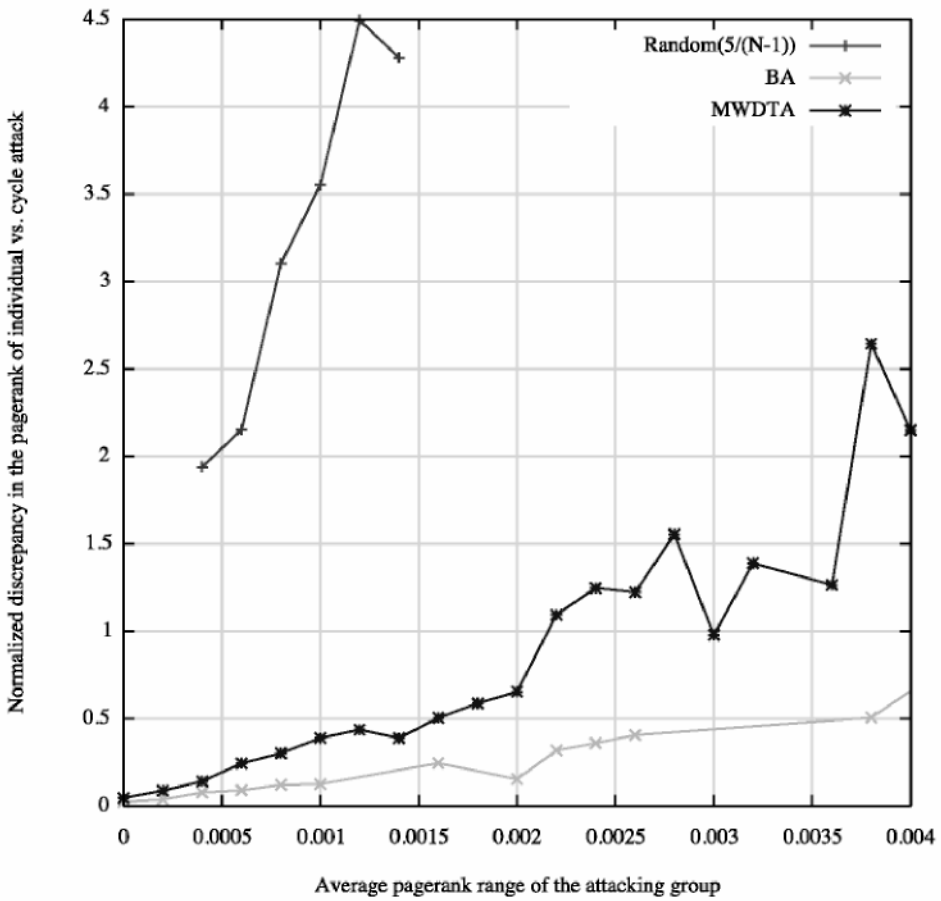


Random: Erdös-Reyni type random graph with edge probability 5(N-1)

BA: Barabási-Albert, preferential attachment

MWDTA: “Winners don’t take all”, BA higher probability of nodes with significant indegree and one outgoing link per node

Pagerank effectiveness of the uncoordinated attack



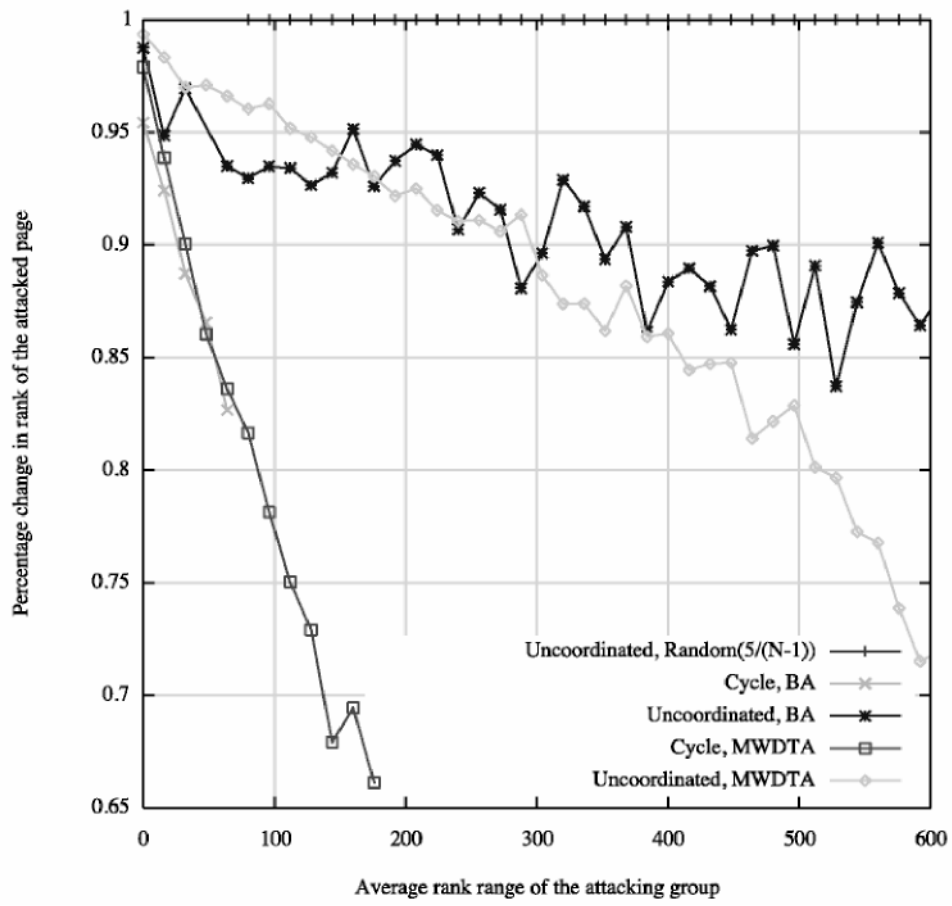
Normalized discrepancy=

$(\Delta p^{Uncoordinated} / \sigma_p) - (\Delta p^{Cycle} / \sigma_p)$

$\Delta p^{AttackType}$: the pagerank change of the attacked page

σ_p: standard deviation of the pagerank distribution

Rank effectiveness of the uncoordinated attack



Conclusions

- Uncoordinated attack is best for pagerank
  - Any additional coordination reduces the impact of the attack
  - Participants in an attack may have no relationship with each other, making it harder to detect and prove
  - A ranking algorithm that favors hierarchical attacks would mean small groups should participate in a group structure for an effective attack
- Conditions resistant to attack
  - Dense, power-law graphs, victims with high rank, attackers with low rank

Conclusions

- Assumptions made by pagerank revisited
  - Random jump to any page while user does not know about them, pages with no outgoing links accumulate pagerank [Eiron, McCurley, Tomlin]
  - The probability to navigate from a page may be proportional to the page's pagerank
  - The probability to use a link may be proportional to the pagerank of the destination page or to the page text properties [Chakrabarti et. al.]
- How is an attack different than the popular opinion of the web citizens?
  - Size of the attacking group and their overall influence
  - How likely is it that a small number of unrelated pages use the same text in their link to the same page?
  - The analysis of group structure in attacks provide a new way of discussing the resistance of an algorithm to attacks.