Manipulating PageRank via Spam Farms: Structural Strategies and Theoretical Insights

Chenkai Wang
Email: wangck2022@mail.sustech.edu.cn

Abstract

This essay examines the vulnerability of the PageRank algorithm to manipulation through spam farms. A synthetic directed network is generated using the "networkx"[1], and various spam farm structures are introduced to evaluate their effects on the PageRank of a target node. Both empirical findings and theoretical analysis illustrate how different configurations influence the ranking dynamics. Notably, a saturation effect and even a reversal of ranking gains are observed under certain structural conditions. These results offer a detailed understanding of how link-based ranking algorithms respond to structural changes.

Keywords: Page Rank Algorithm, Spam Farm, Ranking Dynamics

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Contents

1 Description of Homework

1.1 Homework 1: Design Spam Farm

- Generate an undirected graph with 100 nodes, ensuring a minimum degree of 2 and a maximum degree of 20. The degree distribution of the nodes follows $P(k) \sim k^{-2.5}$.
- Randomly assign directions to the edges of the undirected graph to convert it into a directed graph.
- Implement the PageRank algorithm with $\beta = 0.85$ on the directed graph.
- Design a Spam Farm to improve the PageRank of the target webpage.

1.2 Homework 2: Investigate the Spam Farm

Under the configuration of Homework 1:

- Attempt to maximize the PageRank of the target webpage. What key observations or conclusions can you draw from your results?
- Provide a theoretical analysis to support your findings (optional).

2 Introduction

Link-based ranking algorithms such as PageRank have become central to information retrieval systems by quantifying the relative importance of webpages based on their hyperlink structure. In PageRank, the value of a node depends not only on the number of incoming links, but also on the quality of the linking sources. However, this dependence renders the algorithm susceptible to adversarial manipulation through artificially constructed link schemes—commonly known as spam farms.

A spam farm typically comprises a set of low-quality pages that direct outbound links to a single target page. These structures are designed to concentrate ranking value and thereby inflate the target's PageRank score. Such manipulations distort the intended semantics of link-based ranking and compromise the reliability of search engine results. A deeper understanding of the structural and dynamical properties of spam farms is therefore essential for designing more resilient ranking mechanisms.

To investigate this phenomenon, a synthetic directed network is constructed with a power-law degree distribution $p(k) \sim k^{-2.5}$ and converted using a degree-aware edge orientation method. The standard PageRank algorithm is implemented on this network, followed by the introduction of various spam farm topologies—including complete versus incomplete and unidirectional versus bidirectional configurations. The resulting effects on the PageRank of a designated target node are measured and analyzed.

Both empirical observations and theoretical derivations are provided to characterize how spam farms influence ranking dynamics. Notably, in the case of large complete farms with bidirectional links, an initially increasing PageRank can exhibit a decline as the spam farm grows—a reversal effect that highlights the nontrivial interaction between link structure and rank propagation. Additionally, convergence rates under different spam farm configurations are examined, showing consistent geometric convergence across scenarios.

The remainder of this report is organized as follows: Section 3 introduces the network generation procedure; Section 4 outlines the PageRank algorithm; Section 5 details the construction of spam farms; and Sections 6 - 7 present the analytical results and comparative findings.

3 Network Generation and Conversion

3.1 Undirected Network Construction

Several approaches exist for generating networks with prescribed degree distributions.

The **Configuration Model** samples a degree sequence from a power-law distribution $P(k) \propto k^{-\gamma}$ and pairs stubs randomly to form edges. While it precisely controls node degrees, it often lacks higher-order structure such as clustering or modularity.

The **Static Model** (Goh, Kahng, and Kim) assigns each node a weight $w_i \propto i^{-\alpha}$, with edge probabilities defined by $p_{ij} \propto w_i w_j$. This yields networks with power-law degree distributions but no explicit community structure.

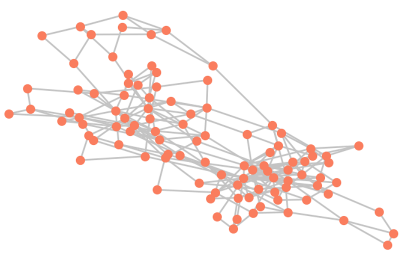
The **Barabási-Albert (BA) Model** uses preferential attachment, where a new node connects to existing nodes with probability proportional to their degree. This produces a scale-free network with exponent $\gamma \approx 3$, though further modification is required for greater flexibility.

In contrast, the **Lancichinetti-Fortunato-Radicchi (LFR) Model** offers greater realism by generating networks that exhibit both heavy-tailed degree distributions and explicit community structures. This makes it particularly suitable for simulating web-like environments where modular organization and local clustering are prevalent.

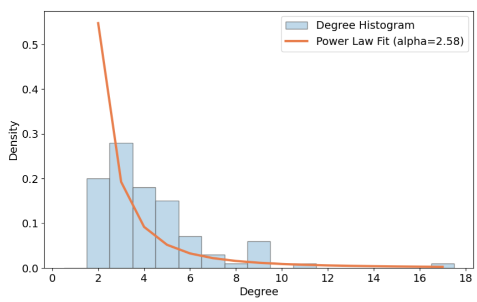
In this study, an undirected network with $N = 100$ nodes is generated using the LFR model, targeting a power-law degree distribution $p(k) \sim k^{-2.5}$, with degree bounds $2 \leq k \leq 20$. The model parameters are summarized below:

- $\tau_1 = 2.5$ (degree exponent), $\tau_2 = 1.5$ (community size exponent).
- Minimum and maximum degree: 2 and 20.
- Mixing parameter $\mu = 0.1$, promoting strong intra-community density.
- Minimum community size: 4; maximum set to $N/2$.
- Random seed fixed at 42 for reproducibility.

Self-loops are removed post-generation. Figure 1 displays the resulting topology and confirms that the degree sequence aligns with the expected power-law behavior.



(a) LFR Network Topology



(b) Degree Distribution Fit

Figure 1: Undirected network and power-law verification

3.2 Conversion to Directed Graph

To simulate hyperlink directions, the undirected graph is converted into a directed one via a degree-aware orientation algorithm. The goal is to preserve high-connectivity among influential nodes while allowing directionality in less critical regions.

The conversion algorithm proceeds as follows:

- For each undirected edge (u, v), if either u or v has degree exceeding a threshold D, assign bidirectional links.
- Otherwise, assign bidirectional edges with probability P; if not chosen, assign a single directed edge uniformly at random.

The pseudocode is provided in Appendix 8. The parameters are fixed at $D = 5$, $P = 0.3$ without further tuning.

Figure 2: Directed graph after orientation

The final directed graph contains 100 nodes and 212 edges, with degree bounds consistent with the original undirected network.

4 PageRank Algorithm

The PageRank algorithm [2] assigns a numerical relevance score to each node in a directed network, based on the link structure of the graph. It models the behavior of a random web surfer who either follows outbound links or jumps uniformly to a random node with a fixed probability.

Formally, the PageRank score $PR(v)$ of a node v is defined as

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in \mathcal{B}(v)} \frac{PR(u)}{L_u}, \tag{1}$$

where $d \in (0,1)$ is the damping factor, N is the total number of nodes, $\mathcal{B}(v)$ denotes the set of nodes linking to v, and L_u is the out-degree of node u. The damping factor d typically takes the value 0.85.

To ensure well-defined behavior in the presence of *dangling nodes*—nodes with zero out-degree—their PageRank mass is redistributed uniformly across all nodes at each iteration. This maintains the stochasticity of the underlying transition matrix.

Under mild connectivity assumptions and for any $0 < d < 1$, the iteration converges to a unique stationary distribution regardless of initialization.

Iterative Algorithm

The standard iterative procedure for computing PageRank is outlined below:

Algorithm 1 PageRank Algorithm

Input: Directed graph $G = (V, E)$, damping factor $d \in (0,1)$, number of iterations T

Output: PageRank vector PR

1: Initialize $PR(v) \leftarrow \frac{1}{|V|}$ for all $v \in V$
2: **for** $t = 1$ to T **do**
3: $\quad PR_{\text{new}}(v) \leftarrow \frac{1-d}{|V|}$ for all $v \in V$
4: $\quad$ **for** each edge $(u \to v) \in E$ **do**
5: $\quad\quad PR_{\text{new}}(v) \mathrel{+}= d \cdot \frac{PR(u)}{L_u}$
6: $\quad$ Redistribute dangling mass uniformly across all v
7: $\quad$ Update $PR(v) \leftarrow PR_{\text{new}}(v)$ for all v
8: **return** PR

Matrix Representation

Let $\mathbf{p} \in \mathbb{R}^N$ be the vector of PageRank scores. Let $A \in \mathbb{R}^{N \times N}$ be the column-stochastic adjacency matrix defined by

$$A_{ij} = \begin{cases} \frac{1}{L_j}, & \text{if } j \to i, \\ 0, & \text{otherwise.} \end{cases}$$

To accommodate dangling nodes, any column corresponding to $L_j = 0$ is replaced with a uniform vector $1/N$.

The PageRank equation then becomes:

$$\mathbf{p} = \frac{1-d}{N} \cdot \mathbf{1} + d \cdot A^\top \mathbf{p}, \tag{2}$$

which can be solved iteratively via:

$$\mathbf{p}^{(k+1)} = \frac{1-d}{N} \cdot \mathbf{1} + d \cdot A^\top \mathbf{p}^{(k)}, \tag{3}$$

until convergence, where $\mathbf{1}$ is the all-ones vector. The iteration converges at a geometric rate determined by d.

5 Directed Network with Spam Farm

To evaluate the impact of adversarial link injection on PageRank, we construct a directed network based on the converted graph from Section 3.2. A target node, denoted by p_{50}, is randomly selected from the original network. Around this node, a set of K auxiliary nodes is appended to form a *spam farm*.

The spam farm is characterized by two key structural dimensions: (1) the directionality of connections between the spam nodes and the target node, and (2) the internal connectivity among the spam nodes themselves.

- **Unidirectional case:** Each spam node has a directed edge pointing to the target node t, with no return edges from t.
- **Bidirectional case:** Each spam node connects to t via bidirectional edges, creating mutual links between the farm and the target.

The internal connectivity of the spam farm is further classified as follows:

- **Complete farm:** Every spam node is connected to every other spam node, forming a fully connected subgraph.
- **Incomplete farm:** No links exist among the spam nodes; they only connect externally to the target node.

Figure 3 illustrates a bidirectional network configuration. The core topology remains unchanged, while K external nodes are injected to artificially inflate the PageRank of the designated target. This setup mimics common adversarial strategies on the web, where low-quality pages are fabricated to boost the ranking of a specific page.

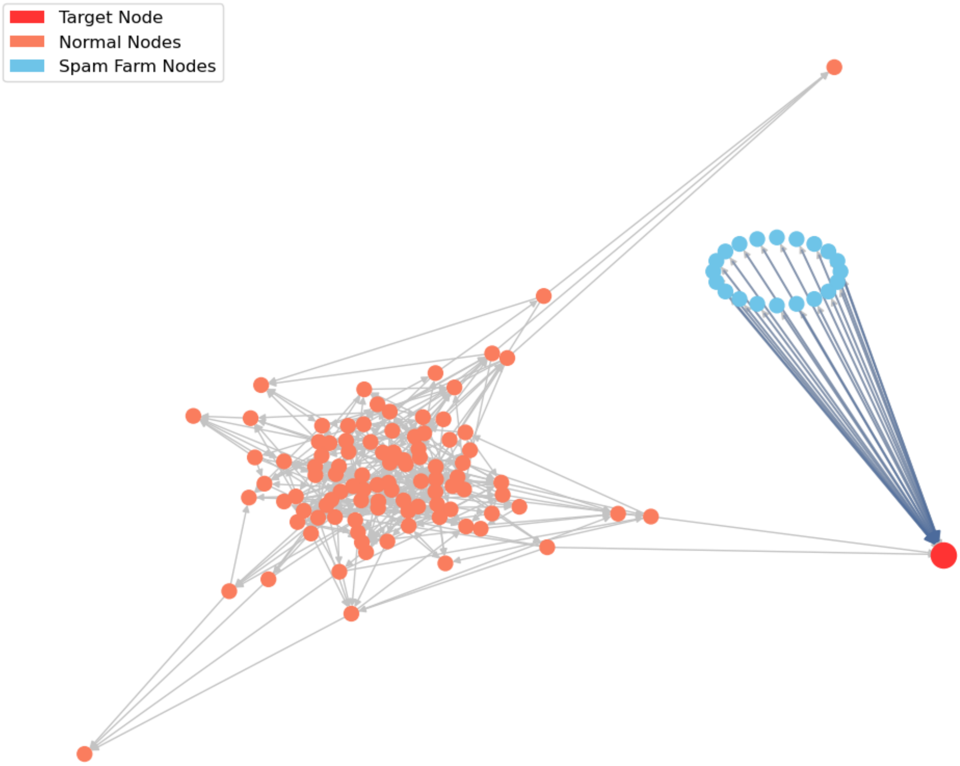
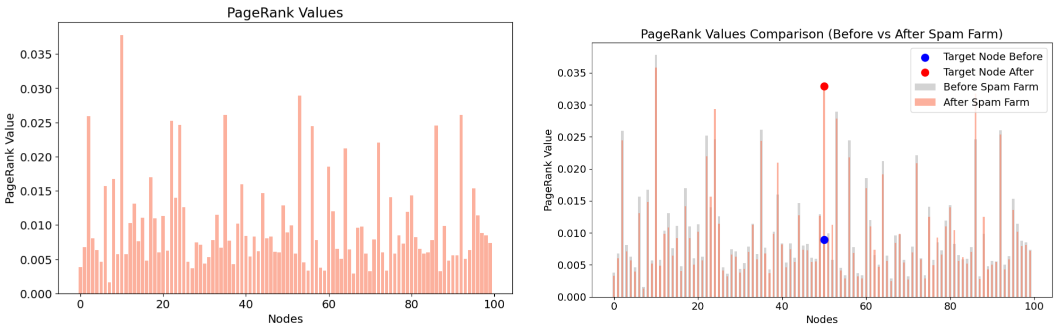


Figure 3: Illustration of the original network with an appended incomplete, bidirectional spam farm

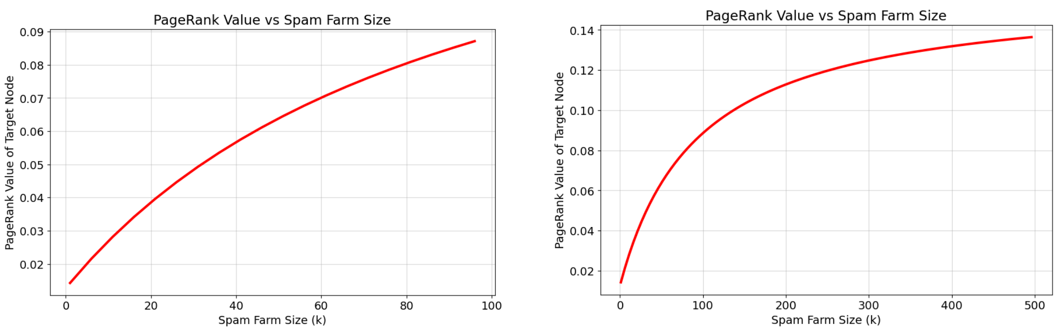
In this section, we adopt the *incomplete* and *unidirectional* setting for experiments: each spam node links to the target node with no return and no intra-farm edges are present. To quantify the impact of this structure, PageRank values are computed before and after spam farm injection. In the first experiment, the spam farm size is fixed at $K = 20$. As shown in Figure 4, the PageRank of the target node p_{50} increases significantly, while the scores of most other nodes remain largely unchanged. All values are globally recomputed following the structural modification, and results are reported from a single network realization.



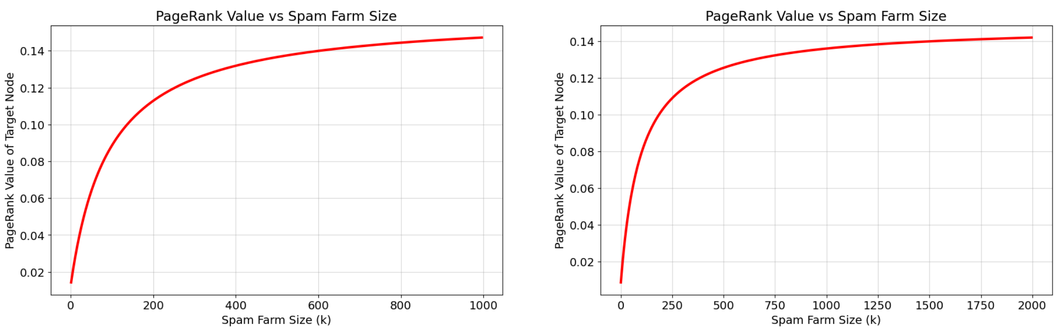
(a) Without spam farm (b) With spam farm

Figure 4: Amplification of the target node's PageRank after spam farm injection

To examine how the scale of the spam farm influences the manipulation effect, we vary K across $100, 500, 1000, 2000$ and plot the resulting PageRank of the target node. Figure 5 shows a monotonic increase in the target node's PageRank with K, although the marginal gains diminish as K grows. This saturation effect reflects the damping behavior of the PageRank algorithm, which imposes a natural limit on the influence a node can acquire solely through inbound links.



(a) Target PageRank ($K = 100$) (b) Target PageRank ($K = 500$)



(c) Target PageRank ($K = 1000$) (d) Target PageRank ($K = 2000$)

Figure 5: Evolution of the target node's PageRank as spam farm size K increases

These results confirm that incomplete spam farms can substantially amplify the importance of a target node. However, their effectiveness saturates with size, highlighting the intrinsic regularization imposed by the PageRank formulation.

6 Dynamics of Spam Farm Impact

To investigate how the PageRank of a designated node responds to adversarial manipulation, we study the dynamics of PageRank variation as the spam farm size K increases.

Consider a directed graph G of size N, augmented by a spam farm consisting of K auxiliary nodes targeting a specific node t. The node t has out-degree $L_t \geq 0$, and the PageRank is governed by the damping factor $\beta \in (0,1)$. The PageRank score p_u of each node u satisfies:

$$p_u = \frac{1-\beta}{N+K} + \beta \sum_{v \in \mathcal{B}(u)} \frac{p_v}{L_v}, \tag{4}$$

where $\mathcal{B}(u)$ is the set of nodes linking to u, and L_v denotes the out-degree of node v. We adopt a row-stochastic normalization for the transition matrix throughout.

We now present a collection of theorems that characterize how different structural designs of spam farms—unidirectional vs. bidirectional, complete vs. incomplete—affect the PageRank of the target node t and the convergence of the iterative process.

6.1 Dynamics in a Incomplete Spam Farm

Theorem (PageRank with Arbitrary Out-Degree and Possibly Incomplete Spam Farm)**.** *Let G be a directed graph of size N with an additional spam farm of size K targeting a node t. The node t has arbitrary out-degree $L_t \geq 0$. Then:*

a) ***Unidirectional case.*** *If the K spam farm nodes only point to t (and t does not link back to them), then*

$$p_t^{unidirectional} = \frac{1-\beta}{N+K} + \beta \cdot \frac{K}{\max(L_t, 1)}. \tag{5}$$

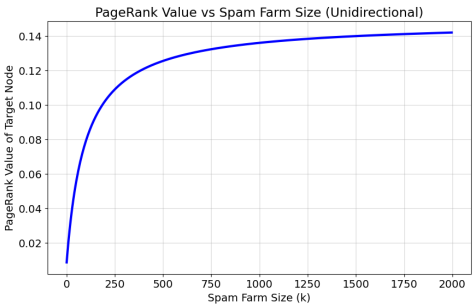
b) ***Bidirectional case.*** *If the spam farm nodes link bidirectionally with t, let $S = \sum_{i=1}^{K} p_i$ be the total PageRank of those K nodes. Then*

$$p_t^{bidirectional} = \frac{1-\beta}{N+K} + \beta \cdot \frac{S}{\max(L_t, 1)}. \tag{6}$$

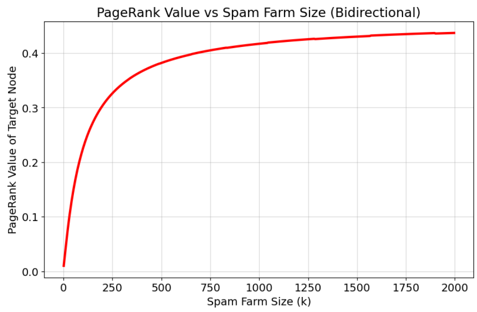
Moreover, if each farm node gains additional rank from t in the bidirectional scenario, we often have

$$p_t^{bidirectional} > p_t^{unidirectional}. \tag{7}$$

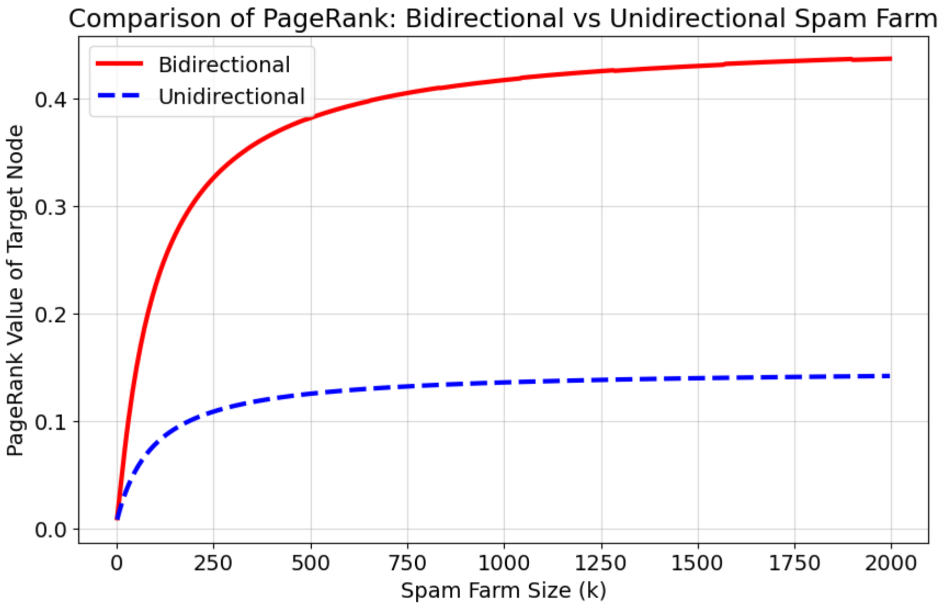
Remark 1. *The inequality $p_t^{bidirectional} > p_t^{unidirectional}$ typically relies on either small or moderate K (and possibly low L_t). In certain large-scale or highly interconnected farms, other factors (like t's outflow) can invert this conclusion; see Section 7.*



(a) Unidirectional



(b) Bidirectional



(c) Comparison

Figure 6: Evolution of the target node's PageRank under unidirectional and bidirectional incomplete spam farms

6.2 Dynamics in a Complete Spam Farm

Theorem (PageRank Dynamics in a Complete Spam Farm)**.** *Let G be a directed graph of size N with a* complete *spam farm of size K targeting a specific node t. Then the PageRank p_t exhibits:*

- *an initial increase as K grows from small values,*
- *a subsequent decrease as $K \to \infty$.*

Proof. We may write

$$p_t = \frac{1-\beta}{N+K} + \beta \sum_{i=1}^{K} \frac{p_i}{L_t}, \tag{8}$$

where L_t is the out-degree of t, and p_i is the PageRank of the i-th spam node. In a complete spam farm, each spam node i has out-degree $K-1$, distributing its PageRank across the other $K-1$ spam nodes.

Initial Increase: When K is small, the internal dissipation is negligible, so each spam node has roughly $p_i \approx \frac{1}{N+K}$. Hence, the contribution to t increases with K, causing p_t to rise initially.

Subsequent Decrease: For large K, each spam node has out-degree $K-1$, so a significant portion of its PageRank is recirculated among the K spam nodes. This internal recycling dilutes each p_i. The final contribution to t becomes on the order of $\frac{K}{K(K-1)} = \frac{1}{K-1}$, which decreases as $K \to \infty$. Therefore p_t eventually falls for large K. □

Theorem (Upper and Lower Bounds for PageRank in a Complete Spam Farm)**.** *Let G be of size N, and let there be a complete spam farm of size K pointing to a node t. Then p_t satisfies:*

$$p_t^{lower} = \frac{1-\beta}{N+K}, \quad p_t^{upper} = \frac{1-\beta}{N+K} + \beta \cdot \frac{K}{N+K}. \tag{9}$$

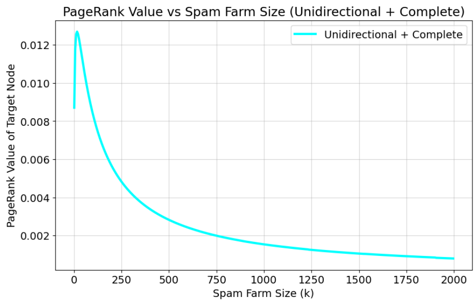
Proof. **Lower Bound:** As $K \to \infty$, the internal links among the spam farm cause heavy dissipation of PageRank within the farm, leaving a negligible net flow to t. Thus p_t approaches the baseline

$$\frac{1-\beta}{N+K}. \tag{10}$$

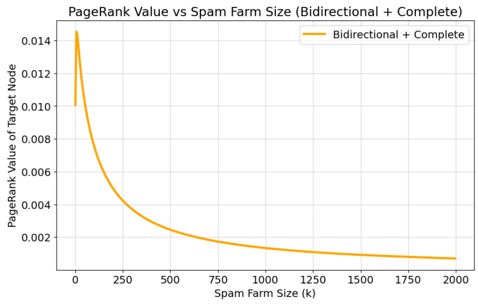
Upper Bound: If there were no internal dissipation (i.e., each spam node gave all of its PageRank to t), then the maximum extra contribution to t would be $\beta \cdot \frac{K}{N+K}$. Hence

$$p_t^{\text{upper}} = \frac{1-\beta}{N+K} + \beta \cdot \frac{K}{N+K}. \tag{11}$$

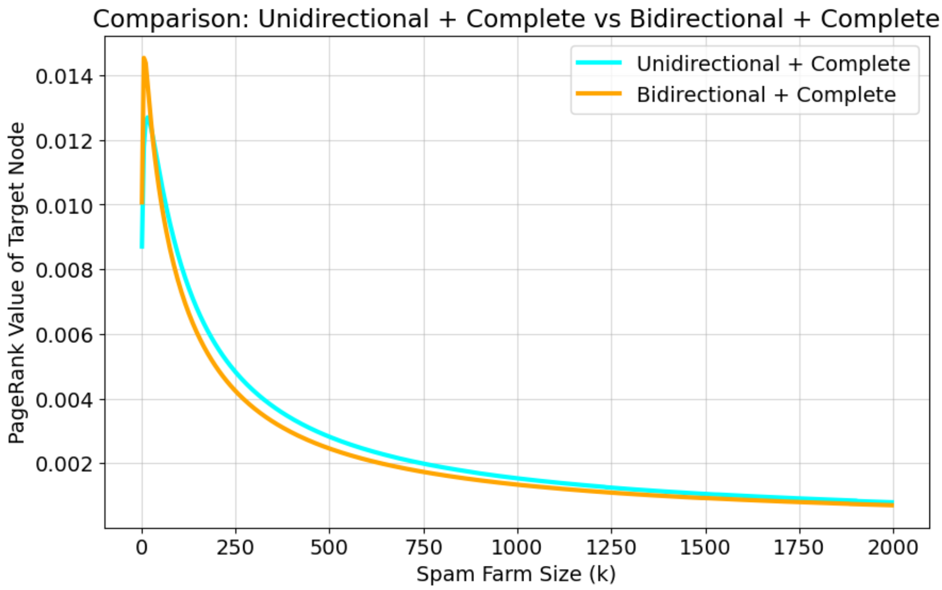
□



(a) Unidirectional



(b) Bidirectional



(c) Comparison

Figure 7: Evolution of the target node's PageRank under unidirectional and bidirectional incomplete spam farms

6.3 Comparison: Complete vs. Incomplete Spam Farms

Theorem (Complete vs. Incomplete Spam Farm). *Let G be a directed graph of size N with a spam farm of size K targeting a node t.*

- *In a **complete farm**, each of the K spam nodes links to every other spam node, so each node distributes its PageRank among $K-1$ neighbors.*
- *In an **incomplete farm** (no internal links, only links to t), every spam node funnels all its PageRank directly to t.*

As $K \to \infty$, the ratio *of t's total increment from a complete farm versus that from an incomplete farm tends to*

$$\frac{\Delta p_t^{complete}}{\Delta p_t^{incomplete}} \sim \frac{1}{K}, \quad \textit{thus an incomplete farm is more effective for large } K. \tag{12}$$

Remark 2. *Notably, in an **incomplete farm**, each spam node has out-degree 1, pointing only to t. Even if K becomes large, no internal cycles form, so there is* no reversal *of t's benefit from the farm. By contrast, a **complete farm** can exhibit a "rise and fall" in p_t with K.*

6.4 Convergence Speed Analysis

Theorem (Convergence Rates in Four Scenarios)**.** *Consider the standard PageRank iteration:*

$$\mathbf{p}^{(k+1)} = \beta M\,\mathbf{p}^{(k)} + \frac{1-\beta}{N+K}\,\mathbf{1}, \tag{13}$$

where M is the row-stochastic transition matrix (of dimension $(N+K)\times(N+K)$), $\mathbf{1}$ is the all-ones vector, and $\beta \in (0,1)$. Let $\mathbf{p}^*$ be the limit. The spectral radius of βM is $\beta < 1$, ensuring geometric convergence in all the following scenarios:

*a) **Unidirectional Farm**: The farm nodes only point to t.*

*b) **Bidirectional Farm**: The farm nodes and t link mutually.*

*c) **Complete Farm**: Every spam node is fully interconnected (many cycles).*

*d) **Incomplete Farm**: No internal links among spam nodes.*

In each case, the iteration converges at a geometric rate $\mathcal{O}(\beta^k)$.

Remark 3. *All the theorems show that while the target node t may gain or lose PageRank in different farm structures, the global iteration speed remains geometric as long as $\beta < 1$.*

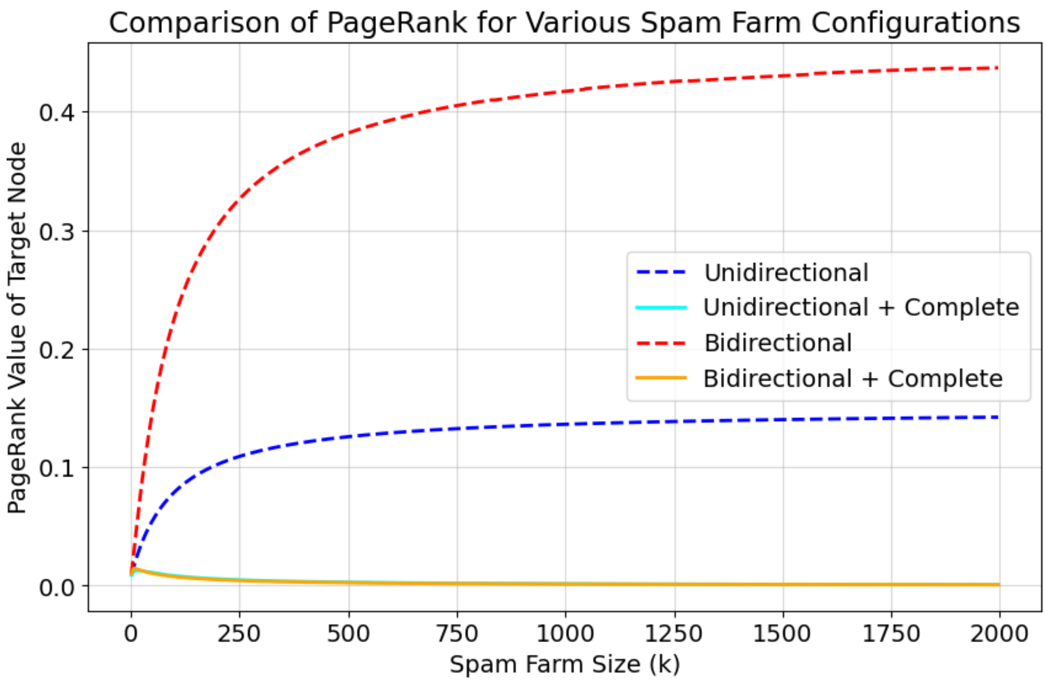


Figure 8: Visualization of different spam farm structures for PageRank convergence speed comparison

7 Reversal of Bidirectional vs Unidirectional in Large Complete Farms

We now state a result indicating that, for a *complete* and *large* spam farm with t having nonzero out-degree, the initially beneficial bidirectional arrangement can *reverse* to become worse than the unidirectional one.

Proposition 1 (Reversal of Bidirectional vs. Unidirectional in Large Complete Farms)**.** *Let G be a directed graph of size N, and let there be a* complete *spam farm of size K targeting a node t. Suppose the target node t has out-degree $L_t > 0$, and let $\beta \in (0, 1)$ be the damping factor in the standard PageRank model. We compare two scenarios:*

- ***Unidirectional + Complete****: Each spam node points to t, and spam nodes are fully interlinked among themselves. Meanwhile, t does* not *link back to any spam node, so t's total out-degree is exactly L_t.*
- ***Bidirectional + Complete****: Each spam node points to t* and *is pointed to by t. The spam nodes are again fully interconnected, but now t gains K extra out-links (one to each spam node), so t's out-degree is $L_t + K$.*

Denote by $p_t^{(U)}(K)$ the steady-state PageRank of t under Unidirectional + Complete, and by $p_t^{(B)}(K)$ the steady-state under Bidirectional + Complete. ***Then there exists a threshold K_0 (depending on L_t, β, and other parameters) such that for all $K > K_0$,***

$$p_t^{(B)}(K) \;<\; p_t^{(U)}(K).$$

In other words, once K becomes large enough, the bidirectional *scenario yields strictly lower PageRank for t than the* unidirectional *one, exhibiting the "reversal" phenomenon.*

Proof. We outline the argument in five steps: constructing the linear systems, applying symmetry among spam nodes, performing a partial analytical (asymptotic) expansion for large K, concluding with a sign-change/continuity argument, and addressing consistency with smaller K.

1. Construction of the Two Linear Systems. We use the usual PageRank model with damping factor β. Let

- $\mathbf{p}^{(U)}(K)$ be the stationary vector for Unidirectional + Complete,
- $\mathbf{p}^{(B)}(K)$ for Bidirectional + Complete.

They each satisfy

$$\mathbf{p} \;=\; \frac{1-\beta}{N+K}\,\mathbf{1} \;+\; \beta\,M\,\mathbf{p}, \tag{14}$$

where M is the row-stochastic transition matrix of size $(N+K) \times (N+K)$.

- In **Unidirectional + Complete**, each spam node s_i (for $1 \leq i \leq K$) has out-degree K (that is, one link to t plus links to the other $K-1$ spam nodes), but t has no links back to them.
- In **Bidirectional + Complete**, each spam node still has out-degree K, but now t has out-degree $L_t + K$.

Hence, we can write down the entries of $M^{(U)}(K)$ and $M^{(B)}(K)$ explicitly. The only difference lies in how t and the farm nodes are interconnected.

2. Symmetry: All Spam Nodes Have Equal Rank. Since the spam farm is *complete*, it is natural to assume (and can be formally justified by a standard symmetry argument) that in the steady state each spam node s_i has the same PageRank value; denote that value by

$$p_s^{(U)} \quad \text{(unidirectional case)}, \qquad p_s^{(B)} \quad \text{(bidirectional case)},$$

respectively. Likewise, let

$$p_t^{(U)}(K), \quad p_t^{(B)}(K)$$

be the respective PageRank of the target node t in each scenario.

3. Partial Analytical Expansions for Large K. We now compare how p_t depends on K in each scenario when K is large. Below we illustrate the core ideas:

- $p_t = \frac{1-\beta}{N+K} + \beta \times$ (inflow from spam & others)
- $p_s = \frac{1-\beta}{N+K} + \beta \times$ (inflow from t & other farm nodes)

Unidirectional + Complete:

- t has out-degree L_t. It does *not* link to the farm, so all farm $\to t$ flows are essentially one-way.
- Each spam node s points to t and the other $K-1$ spam nodes, so out-degree is K. In the steady state, we may approximate

$$p_s^{(U)}(K) \sim \frac{1}{K}, \quad p_t^{(U)}(K) \sim \frac{\beta\, p_s^{(U)}}{L_t} \sim \frac{\beta}{L_t\, K}.$$

We also add the baseline random-jump term $\frac{1-\beta}{N+K} \approx \frac{1}{K}$. Altogether, $p_t^{(U)}(K)$ remains on the order of $\frac{1}{K}$.

Bidirectional + Complete:

- t has out-degree $L_t + K$. It points back to the spam nodes, so some of t's PageRank flows into the farm.
- Each spam node still has out-degree K. Now it receives an additional fraction $\frac{p_t^{(B)}}{L_t+K}$ from t. Meanwhile, these farm nodes redistribute heavily among themselves.
- A more explicit system is

$$\begin{cases} p_t^{(B)} = \dfrac{1-\beta}{N+K} + \beta\, \dfrac{K \cdot \frac{p_s^{(B)}}{K}}{L_t + K}, \\ p_s^{(B)} = \dfrac{1-\beta}{N+K} + \beta \left(\dfrac{(K-1)\, p_s^{(B)}}{K} + \dfrac{p_t^{(B)}}{L_t + K} \right). \end{cases}$$

Solving (or approximating) for large K shows that $p_t^{(B)}(K)$ is also of order $\frac{1}{K}$, but with a smaller prefactor if $L_t > 0$. A key reason is that t distributes part of its rank to the K spam nodes, which get "stuck" in internal recycling.

Hence, in the limit $K \to \infty$, we typically find

$$p_t^{(B)}(K) \; < \; p_t^{(U)}(K),$$

since the "net outflow" from t in the bidirectional scenario is not fully compensated by the "return flow" from the large complete farm.

4. Sign-Change Argument & Threshold Existence. Define the difference

$$D(K) \; = \; p_t^{(B)}(K) \; - \; p_t^{(U)}(K).$$

Empirically (and by small-K expansions), one finds $D(K)$ is often > 0 for small or moderate K; hence early on, the mutual linking can benefit t. However, from the large-K asymptotics above (and numerical checks), we see that $D(K)$ eventually becomes negative for sufficiently big K. By a continuity or sign-change argument in K (treating K as an integer but conceptually as a parameter over which we can track $D(K)$), there must exist at least one crossing K_0 so that:

$$D(K_0) = 0, \quad D(K) > 0 \text{ for } K < K_0, \quad D(K) < 0 \text{ for } K > K_0.$$

Equivalently,

$$p_t^{(B)}(K) > p_t^{(U)}(K) \quad \text{for } K < K_0, \qquad p_t^{(B)}(K) < p_t^{(U)}(K) \quad \text{for } K > K_0.$$

Thus beyond the threshold K_0, the bidirectional case indeed yields strictly lower PageRank for t than the unidirectional case.

5.Consistency with Incomplete Farms. In an **incomplete farm** (no internal links), each spam node has out-degree 1 pointing solely to t. Even if K is large, t does *not* link back to the farm; thus no high-degree cycles occur and no "net outflow" from t re-enters a huge feedback loop. Therefore, *incomplete farms do not exhibit* the reversal phenomenon. By contrast, the large **complete** farm's intense internal cycles can dilute the return flow to t so much that $p_t^{(B)}(K)$ becomes smaller than $p_t^{(U)}(K)$. This completes the proof. □

8 Conclusion

This essay investigates the impact of adversarial link structures—commonly referred to as spam farms—on the PageRank algorithm. By constructing synthetic directed networks with controlled spam farm configurations, we demonstrate how the target node's PageRank is affected under various settings. Notably, incomplete farms consistently amplify PageRank more effectively than complete ones, and bidirectional links can introduce a rise-and-fall dynamic depending on the farm size. Among all configurations, incomplete farms with bidirectional links yield the strongest PageRank amplification. Theoretical results are provided to explain these patterns, along with convergence rate analysis. These findings offer preliminary insights into how different structural configurations of spam farms may influence link-based ranking behavior, as observed in PageRank.

Appendix

Algorithm for Converting Undirected Graph to Directed Graph

Algorithm 2 Convert an Undirected Graph to a Directed Graph Based on Node Degree

Input: Undirected graph $G = (V, E)$, degree threshold D, probability of bidirectional edges P

Output: Directed graph $G_d = (V, E_d)$

Initialize $G_d \leftarrow (V, \emptyset)$ $\triangleright$ Empty directed graph with nodes from G
for each edge $(u, v) \in E$ **do**
    Compute degree of u, $d_u \leftarrow \text{degree}(u)$
    Compute degree of v, $d_v \leftarrow \text{degree}(v)$
    if $d_u > D$ or $d_v > D$ **then**
        Add bidirectional edges $(u \to v)$ and $(v \to u)$ to E_d
    else
        Generate a random number $r \in [0, 1]$
        if $r < P$ **then**
            Add bidirectional edges $(u \to v)$ and $(v \to u)$ to E_d
        else
            Generate another random number $r' \in [0, 1]$
            if $r' < 0.5$ **then**
                Add a unidirectional edge $(u \to v)$ to E_d
            else
                Add a unidirectional edge $(v \to u)$ to E_d
return G_d