On the α-Sensitivity of Nash Equilibria in PageRank-Based Network Reputation Games

Wei Chen[1], Shang-Hua Teng[2⋆], Yajun Wang[3], and Yuan Zhou[4]

[1] Microsoft Research, weic@microsoft.com
[2] Microsoft Research, t-shaten@microsoft.com
[3] Microsoft Research, yajunw@microsoft.com
[4] Tsinghua University, timzhouyuan@gmail.com

Abstract. Web search engines use link-based reputation systems (e.g. PageRank) to measure the importance of web pages, giving rise to the strategic manipulations of hyperlinks by spammers and others to boost their web pages' reputation scores. Hopcroft and Sheldon [10] study this phenomenon by proposing a network formation game in which nodes strategically select their outgoing links in order to maximize their PageRank scores. They pose an open question in [10] asking whether all Nash equilibria in the PageRank game are insensitive to the restart probability α of the PageRank algorithm. They show that a positive answer to the question would imply that all Nash equilibria in the PageRank game must satisfy some strong algebraic symmetry, a property rarely satisfied by real web graphs. In this paper, we give a negative answer to this open question. We present a family of graphs that are Nash equilibria in the PageRank game only for certain choices of α.

1 Introduction

Many web search engines use link-based algorithms to analyze the global link structure and determine the reputation scores of web pages. The popular PageRank [3] reputation system is one of the good examples. It scores pages according to their stationary probability in a random walk on the graph that periodically jumps to a random page. Similar reputation systems are also used in peer-to-peer networks [11] and social networks [8].

A common problem in reputation systems is manipulation: strategic users arrange links attempting to boost their own reputation scores. On the web, this phenomenon is called link spam, and usually targets at PageRank. Users can manage to obtain in-links to boost their own PageRank [6], and can also achieve this goal by carefully placing out-links [2, 4, 7]. Thus PageRank promotes certain link placement strategies that undermine its original premise, that links are placed organically and reflect human judgments on the importance and relevance of web pages.

⋆ Affiliation starting from the Fall of 2009: Department of Computer Science, University of Southern California, Los Angeles, CA, U.S.A.

There are a number of works focusing on understanding, detecting [6, 2, 4, 7] and preventing link manipulations [5, 9, 12]. Other works investigate the consequences of manipulation. In [10], Hopcroft and Sheldon introduce a network formation game called the *network reputation game*, where players are nodes attempting to maximize their reputation scores by strategically placing out-links. They mainly focus on the game with the PageRank reputation system, which we refer to directly as PageRank game, and study the Nash equilibria of the game. Each Nash equilibrium is a directed graph, where no player can further improve his own PageRank by choosing different out-links. They present several properties that all Nash equilibria hold, and then provide a full characterization of α-insensitive Nash equilibria, those graphs that are Nash equilibria for all restart probability α, a parameter in the PageRank algorithm. They show that all α-insensitive Nash equilibria must satisfy some strong algebraic symmetry property, which is unlikely seen in real web graphs.

However, the work of [10] leaves an important question unanswered, which is whether all Nash equilibria are α-insensitive. In this paper, we give a negative answer to this open question. In particular, we construct a family of graphs and prove that for α close to 0 and 1 they are not Nash equilibria. At the same time, by applying the mean value theorem, we argue that for every graph in the family, there must exist an α for which the graph is a Nash equilibrium.

In Section 2, we give the definition of PageRank functions and some known properties of α-random walk and best response in PageRank games. In Section 3, we present the construction of a family of graphs and prove that they are α-sensitive Nash equilibria.

2 Preliminaries

2.1 PageRank

Let $G = (V, E)$ be a simple directed graph, i.e., without multiple edges and self-loops. Furthermore, we assume each node has at least one out-link. Let $V = [n] = \{1, 2, \ldots, n\}$. We denote (i, j) as the directed edge from node i to node j.

The PageRank of the nodes in G can be represented by the stationary distribution of a random walk as follows. Let $\alpha \in (0, 1)$ be the restart probability. The initial position of the random walk is uniformly distributed over all nodes in the graph. In each step, with probability α, the random walk restarts, i.e., jumps to a node uniformly at random. Otherwise, it walks to the neighbors of the current node, each with the same probability. The PageRank of a node u is the probability mass of this random walk in its stationary distribution. We refer to this random walk as the PageRank random walk. Note that this random walk is ergodic, which implies the existence of the unique stationary distribution.

Mathematically, let A be the adjacency matrix of G, such that the entry $A_{ij} = 1$ if and only if there exists an edge from node i to j in G. Let d_i be the out degree of node i in G. Define matrix P as $P = AD^{-1}$, where $D =$

$diag\{d_1, d_2, \ldots, d_n\}$. Thus P is the transition matrix for the random walk on G without the random restart. By defining $M_\alpha = I - (1-\alpha)P$, it is well known (e.g. see [10]) that the PageRank (row) vector π_α is:

$$\begin{aligned}\pi_\alpha &= \frac{\alpha}{n}[1, 1, \cdots, 1](I - (1-\alpha)P)^{-1}\\ &= \frac{\alpha}{n}[1, 1, \cdots, 1]M_\alpha^{-1}.\end{aligned}$$

The i-th entry of π_α is the PageRank value of node i in G. In particular, the PageRank of node i is:

$$\pi_\alpha[i] = \frac{\alpha}{n}\sum_{j=1}^{n} {M_\alpha^{-1}}_{ji}. \tag{1}$$

Returning Probability

It is easy to see that $M_\alpha^{-1} = I + \sum_{i=1}^{\infty}[(1-\alpha)P]^i$. Therefore the entry ${M_\alpha^{-1}}_{ij}$ is the expected number of times to visit j starting from i before the next restart, in the PageRank random walk.

Suppose that the random walk currently stops at i. We define ϕ_{ij} to be the probability of reaching j in the random walk before the next restart. We set $\phi_{ii} = 1$. Let ϕ_{ii}^{+} be the probability of returning to i itself from i before the next restart. The following lemma holds.

Lemma 1 (Hopcroft and Sheldon [10]). *Let* $i, j \in [n]$. *For any simple directed graph* G *and restart probability* $0 < \alpha < 1$, *we have*

$$\begin{aligned}&\forall\, i \neq j,\, {M_\alpha^{-1}}_{ij} = \phi_{ij}{M_\alpha^{-1}}_{jj}\\ &\forall\, j, \qquad {M_\alpha^{-1}}_{jj} = \frac{1}{1-\phi_{jj}^{+}}.\end{aligned}$$

2.2 PageRank Game and Best Responses

In the PageRank game, each node is building out-links. Let $E_i = \{(i,j) \,|\, j \in V \setminus \{i\}\,\}$ be the set of possible out-links from node i. The strategy space of node i is all subsets of E_i. The payoff function for each node is its PageRank value in the graph.

A best response strategy for node i in the PageRank game is a strategy that maximizes its PageRank in the graph, given the strategies of all other nodes. By Lemma 1 and Equation (1), the best response for i is maximizing the following statement in the new graph:

$$\sum_{j=1}^{n} \frac{\phi_{ji}}{1-\phi_{ii}^{+}}.$$

Note that ϕ_{ji} is the probability of reaching node i from j before the next restart in the random walk, which is independent of the out-links from i. ($\phi_{ii} = 1$ by definition.) Based on this observation, following lemma from [10] holds.

Lemma 2 (Lemma 1 of Hopcroft and Sheldon [10]). *In the PageRank game, a best response strategy for node i is a strategy that maximizes ϕ^+_{ii} in the new graph.*

By definition, ϕ^+_{ii} is the probability of returning to i from i before the next restart in the random walk. Therefore,

$$\phi^+_{ii} = \frac{\alpha}{|N(i)|} \sum_{j \in N(i)} \phi_{ji}.$$

Corollary 1. *In the PageRank game, a best response strategy for node i is a nonempty set $S \subseteq E_i$ that maximizes*

$$\frac{1}{|S|} \sum_{j \in S} \phi_{ji}.$$

Note that the out-links of node i does not affect ϕ_{ji} for $j \neq i$. Since Corollary 1 means that the best response of node i is to maximize the average ϕ_{ji} of its outgoing neighbors, it is clear that i should only select those nodes j with the maximum ϕ_{ji} as its outgoing neighbors. This also implies that in the best response S of i, we have for all $j, k \in S$, $\phi_{ji} = \phi_{ki}$.

2.3 Algebraic Characterization of a Nash Equilibrium

A graph G is a Nash equilibrium if for any node i, its out-links in G is a best response. The following lemma is a directly consequence of Corollary 1.

Lemma 3. *If a strongly-connected graph $G = (V, E)$ is a Nash equilibrium, the following conditions hold:*

(1) for all different $i, j, k \in V$, ${M^{-1}_\alpha}_{ji} < {M^{-1}_\alpha}_{ki} \implies (i, j) \notin E$; and

(2) $\forall (i, j), (i, k) \in E, {M^{-1}_\alpha}_{ji} = {M^{-1}_\alpha}_{ki}$.

It has been shown that any strongly connected Nash equilibrium should be bidirectional [10].

Lemma 4 (Theorem 1 of Hopcroft and Sheldon [10]). *If a strongly-connected graph $G = (V, E)$ is a Nash equilibrium, then for any edge $(i, j) \in E$, we have $(j, i) \in E$.*

The proof of Lemma 4 uses the fact that only the nodes having out-links to vertex i can achieve the maximum value of ${M^{-1}_\alpha}_{\cdot,i}$. The following lemma states that the necessary conditions given in Lemma 3 (2) and Lemma 4 are actually sufficient to characterize a Nash equilibrium.

Lemma 5 (Equivalent condition for a Nash equilibrium). *A strongly-connected graph $G = (V, E)$ is a Nash equilibrium if and only if G is bidirectional and $\forall (i, j), (i, k) \in E, {M^{-1}_\alpha}_{ji} = {M^{-1}_\alpha}_{ki}$.*

Proof. The necessary conditions are given by Lemma 3 (2) and Lemma 4. We focus on the sufficient condition. Since ${M_\alpha^{-1}}_{ji} = {M_\alpha^{-1}}_{ki}$ is true for all nodes $j, k \in V \setminus \{i\}$ that have out-links to i by our condition, we have $\phi_{ji} = \phi_{ki}$ from Lemma 1. By Lemma 2 of Hopcroft and Sheldon [10], only the incoming neighbors j of i can achieve the maximum ϕ_{ji}. Therefore, all incoming neighbors of i actually achieve this maximum. By Corollary 1, node i selecting all incoming neighbors as outgoing neighbors is certainly a best response, which implies that G is a Nash equilibrium. □

3 Existence of α-sensitive Nash Equilibria

A Nash equilibrium is *α-insensitive* if it is a Nash equilibrium for all possible parameter $0 < \alpha < 1$. Otherwise, we say that it is *α-sensitive*. Hopcroft and Sheldon asked the following question about the α-insensitivity property for Nash equilibria.

Question 1 (Hopcroft and Sheldon's question on α-insensitivity [10]). Are all the Nash equilibria α-insensitive or there exist other α-sensitive Nash equilibria in the PageRank game?

We are interested in this question because if all the Nash equilibria are α-insensitive, the set of Nash equilibria can be characterized by the set of Nash equilibria at some specific α. In particular, we can characterize the set of Nash equilibria when α is arbitrary close to 1 (e.g. $1-1/n^3$). However, if the answer to this question is positive, all Nash equilibria must satisfy some strong algebraic symmetry as illustrated in [10], making them less likely the right choice for modeling web graphs. We show below that there exist α-sensitive Nash equilibria, which implies that the set of equilibrium graphs in the PageRank game is much richer.

Theorem 1 (Answer to Question 1). *There exist an infinite number of α-sensitive Nash equilibria in the PageRank game.*

The rest part of this section is devoted to the construction of such α-sensitive Nash equilibria.

3.1 Construction of $G_{n,m}$

For all $n, m \in \mathbb{Z}^+$, we construct an *undirected* graph $G_{n,m}$ as follows.

Definition 1 ($G_{n,m}$). *As shown in Figure 1, the vertex set of graph $G_{n,m}$ consists of U and V. $U = \{u_1, u_2, \cdots, u_n\}$ and $V = \{v_{i,j} \mid i \in [m], j \in [n]\}$. Each $V_i = \{v_{ij} \mid j \in [n]\}$ is fully connected. For each node $u_i \in U$, there are m edges $\{(u_i, v_{j,i})\}$ adjacent to u_i for $j \in [m]$.*

Note that we define $G_{n,m}$ as an undirected graph for the simplicity of presentation. This is because each Nash equilibrium is bidirectional. Nevertheless, we can change $G_{n,m}$ to a directed graph, by replacing each edge with two direct edges. All the following treatment assumes $G_{n,m}$ is *directed*.

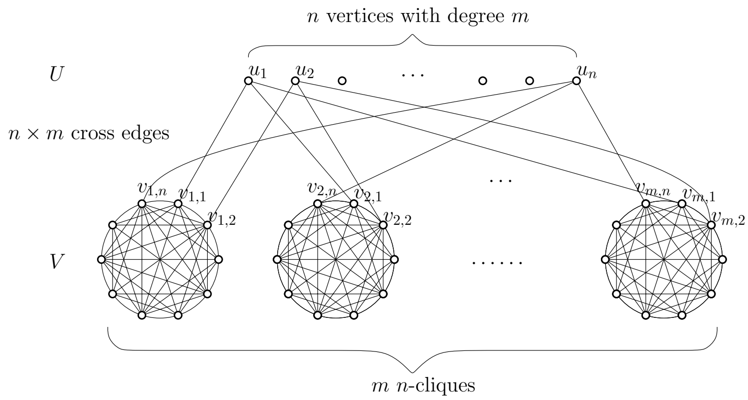

Fig. 1. The graph $G_{n,m}$.

3.2 Equivalent condition of $G_{n,m}$ being Nash equilibria

By Lemma 5, given α, n and m, $G_{n,m}$ is a Nash equilibrium if and only if the following statements hold

$$\forall \text{ different } i, i' \in [m],\, j \in [n], \quad {M_\alpha^{-1}}_{v_{i,j},u_j} = {M_\alpha^{-1}}_{v_{i',j},u_j} \tag{2}$$

$$\forall\, i \in [m], \text{different } k, k', j \in [n],\ {M_\alpha^{-1}}_{v_{i,k},v_{i,j}} = {M_\alpha^{-1}}_{v_{i,k'},v_{i,j}} \tag{3}$$

$$\forall\, i \in [m], \text{different } k, j \in [n], \quad {M_\alpha^{-1}}_{v_{i,k},v_{i,j}} = {M_\alpha^{-1}}_{u_j,v_{i,j}} \tag{4}$$

It is easy to see that equations in (2) and (3) hold for any α, n, m by symmetry. Moreover, for all $i \in [m]$ and different $k, j \in [n]$, ${M_\alpha^{-1}}_{v_{i,k},v_{i,j}}$ has the same value, and for all $i \in [m]$ and $j \in [n]$, ${M_\alpha^{-1}}_{u_j,v_{i,j}}$ has the same value in $G_{n,m}$. We define two functions based on this observation:

$$f_{n,m}(\alpha) = {M_\alpha^{-1}}_{v_{i,k},v_{i,j}}, \text{ for } i \in [m] \text{ and } j,k \in [m],\ j \neq k,$$
$$g_{n,m}(\alpha) = {M_\alpha^{-1}}_{u_j,v_{i,j}}, \quad \text{for } i \in [m] \text{ and } j \in [n].$$

The above argument together with Lemma 5 implies the following lemma.

Lemma 6. *Given α, n and m, $G_{n,m}$ is a Nash equilibrium for the PageRank game with parameter $\alpha \in (0,1)$ if and only if*

$$f_{n,m}(\alpha) - g_{n,m}(\alpha) = 0 \tag{5}$$

3.3 α-sensitivity of $G_{n,m}$: Proof outline

Given n, m, by Lemma 6, to prove $G_{n,m}$ is α-sensitive, we only need to show that there is some α satisfying Equation (5), while there is also some other α that does not. Note that $f_{n,m}(\alpha)$ (resp. $g_{n,m}(\alpha)$) is the expected number of times the random walk starting from $v_{i,k}$ (resp. u_j) visiting $v_{i,j}$ before the next restart. We use this interpretation to both explain the intuitive idea of the proof and carry out the detailed proof.

Intuitively, when α is very close to 1 (i.e. the random walk restarts with a very high probability), we only need to compute the first few steps of the random walk to give a good estimate of ${M_\alpha^{-1}}_{i,j}$, since the random walk is very likely to restart after a few steps and the contribution of the longer steps of the walk is negligible. Thus, in this case, by estimating the first step, for all $i \in [m]$, different $k, j \in [n]$,

$$f_{n,m}(\alpha) = {M_\alpha^{-1}}_{v_{i,k},v_{i,j}} \approx \frac{1-\alpha}{n}$$
$$g_{n,m}(\alpha) = {M_\alpha^{-1}}_{u_j,v_{i,j}} \approx \frac{1-\alpha}{m}.$$

When $m < n$, there exists α such that $f_{n,m}(\alpha) - g_{n,m}(\alpha)$ is indeed negative. The following lemma formalizes the intuition. We defer its proof to the next section.

Lemma 7. *Given* $1 \leq m \leq n-2$, *if* $\alpha \geq \frac{n-3}{n-2}$, $f_{n,m}(\alpha) - g_{n,m}(\alpha) < 0$ *and* $G_{n,m}$ *is not a Nash equilibrium for the PageRank game with parameter* α.

On the other hand, when α is very close to 0, random walks starting from both u_j and $v_{i,k}$ tend to be very long before the next restart. In this case, the random walk starting at $v_{i,k}$ has an advantage to be in the same clique as the target $v_{i,j}$, such that the expected number of time it hits $v_{i,j}$ before the walk jumps out of the clique is close to 1. The random walk starting at u_j, however, has $(m-1)/m$ chance to walk into a clique different from the i-th clique where the target $v_{i,j}$ is in, in which case it cannot hit $v_{i,j}$ until it jumps out of its current clique. After both walks jump out of their current clique for the first time, their remaining stochastic behaviors are close to each other and thus contributing to $f_{n,m}(\alpha)$ and $g_{n,m}(\alpha)$ approximately the same amount. As a result, $g_{n,m}(\alpha)$ gains about $(\frac{1}{m} - \frac{1}{n})$ over $f_{n,m}(\alpha)$ in the case where the random walk from u_j hit the target $v_{i,j}$ in the first step, but $f_{n,m}(\alpha)$ gains about 1 over $g_{n,m}(\alpha)$ in the rest cases. Therefore, $f_{n,m}(\alpha)$ is approximately $1 + \frac{1}{n} - \frac{1}{m}$ larger than $g_{n,m}(\alpha)$ when m is large. The following lemma gives the accurate statement of this intuition while the next section provides the detailed and rigorous analysis.

Lemma 8. *Assume* $4 \leq m < n$. *If* $\alpha \in (0, \frac{1}{n(n-1)})$, $f_{n,m}(\alpha) - g_{n,m}(\alpha) > 0$.

Lemma 9. $f_{n,m}(\alpha)$ *and* $g_{n,m}(\alpha)$ *are continuous with the respect of* $\alpha \in (0,1)$.

Proof. For any graph, the corresponding M_α is invertible for $\alpha \in (0,1)$. Note that each entry of M_α is a continous function with respect to $\alpha \in (0,1)$. By Cramer's rule,

$$M_\alpha^{-1} = \frac{M_\alpha^*}{|M_\alpha|},$$

where $|M_\alpha|$ is the determinant and M^*_α is the co-factor matrix of M_α. Therefore, ${M_\alpha^{-1}}_{i,j}$ is continuous with respect to $\alpha \in (0,1)$, since it is a fraction of two finite-degree polynomials and $|M_\alpha|$ is non-zero. □

With the continuity of $f_{n,m}(\alpha)$ and $g_{n,m}(\alpha)$ and the results of Lemma 7 and 8, we can apply the mean value theorem on $f_{n,m}(\alpha) - g_{n,m}(\alpha)$ and know that it must have a zero point. With Lemma 6, we thus have

Corollary 2. *Given $4 \le m \le n-2$, there exists $\alpha \in (\frac{1}{n(n-1)}, \frac{n-3}{n-2})$ such that $G_{n,m}$ is a Nash equilibrium for the PageRank game with parameter α.*

Lemma 7 and Corollary 2 immediately imply Theorem 1.

3.4 Proof details

In this section, we provide the complete proofs to the results presented in last section. To make our intuitive idea accurate, the actual analysis is sufficiently more complicated.

Proof. [of Lemma 7] By Lemma 1, $f_{n,m}(\alpha) - g_{n,m}(a) < 0$ if and only if $\phi_{v_{i,j'},v_{i,j}} < \phi_{u_j,v_{i,j}}$.

By symmetry of the graph, we have $\phi_{v_{i,j'},v_{i,j}} = \frac{1-\alpha}{n} + \frac{(n-2)(1-\alpha)}{n}\phi_{v_{i,j'},v_{i,j}} + \frac{1-\alpha}{n}\phi_{u_{j'},v_{i,j}}$ for $j' \ne j$. Since $u_{j'}$ is at 2 hops away from $v_{i,j}$, $\phi_{u_{j'},v_{i,j}} \le (1-\alpha)^2$. From $\alpha \ge \frac{n-3}{n-2}$, we have $n - (n-2)(1-\alpha) \ge n-1$. Therefore, we have

$$\phi_{v_{i,j'},v_{i,j}} \le \frac{(1-\alpha) + (1-\alpha)^3}{n-1}.$$

Finally, since $\alpha \ge \frac{n-3}{n-2} = 1 - \frac{1}{n-2} \ge 1 - 1/\sqrt{n-2}$, we have $(1-\alpha)^2 \le 1/(n-2)$, and thus

$$\phi_{v_{i,j'},v_{i,j}} \le \frac{(1-\alpha) + (1-\alpha)^3}{n-1} \le \frac{1-\alpha}{n-2}.$$

On the other hand, $\phi_{u_j,v_{i,j}} > \frac{1-\alpha}{m}$. Therefore, the lemma holds when $m \le n-2$. □

Definition 2. *To simplify the notation, by symmetry of the graph, we define*

1. $r = {M_\alpha^{-1}}_{v_{i,j},v_{i,j}}$,
2. $x = {M_\alpha^{-1}}_{v_{i',j},v_{i,j}}$, $i' \ne i$,
3. $y = {M_\alpha^{-1}}_{v_{i,j'},v_{i,j}}$, $j' \ne j$,
4. $z = {M_\alpha^{-1}}_{v_{i',j'},v_{i,j}}$, $i' \ne i$ *and* $j' \ne j$,
5. *and* $\mu = {M_\alpha^{-1}}_{u_j,v_{i,j}}$.

By definition, $r \ge \max\{x, y, z, u\}$. The proof of Lemma 8 requires the following result.

Lemma 10. *Assume* $1 < m \leq n$. *If* $\alpha \in (0, \frac{1}{n(n-1)})$, *we have*

1. $y - z \in (\frac{n}{n+1}, \frac{n}{n-1})$,
2. $r - y < \frac{n}{n-1}$,
3. $x - z < \frac{2}{n-1}$.

Proof. For any path $L = (i_1, i_2, \cdots, i_k)$, let $Pr_\alpha(L)$ be the probability of completing the path L exactly before next restart when starting from i_1. Therefore, $Pr_\alpha(L) = \alpha(1-\alpha)^{k-1}\prod_{j=1}^{k-1} P_{i_j,i_{j+1}}$, where P is the transition matrix of the graph. Let $V(L)$ be the set of nodes of L.

Let $Count_L(i)$ be the number of occurrence of node i in L. Denote by $\mathcal{L}(i)$ the set of paths of finite length starting from vertex i. Therefore,

$$M_\alpha^{-1}{}_{ij} = \sum_{L \in \mathcal{L}(i)} Pr_\alpha(L) Count_L(j)$$

For the first claim, we define function $F_{i\to i'}^{(j)} : \mathcal{L}(v_{i,j}) \to \mathcal{L}(v_{i',j})$ to map any path L starting from $v_{i,j}$ to a path L' starting from $v_{i',j}$, by replacing the longest prefix of L in V_i with a corresponding path in the $V_{i'}$. It is easy to see that $F_{i\to i'}^{(j)}$ always outputs valid paths, and is a bijection. By symmetry, we have $Pr_\alpha(L) = Pr_\alpha(F_{i\to i'}^{(j)}(L))$ for all $L \in \mathcal{L}(v_{i,j})$. For a path L such that $V(L) \subseteq V_i$, we define $Pr_\alpha^{(i)}(L)$ be the probability that the random walk exactly finishes L before leaving V_i or restarting.

Then,

$$\begin{aligned}
y - z &= M_\alpha^{-1}{}_{v_{i,j'},v_{i,j}} - M_\alpha^{-1}{}_{v_{i',j'},v_{i,j}} \\
&= \sum_{L \in \mathcal{L}(v_{i,j'})} Pr_\alpha(L) Count_L(v_{i,j}) - \sum_{L \in \mathcal{L}(v_{i',j'})} Pr_\alpha(L) Count_L(v_{i,j}) \\
&= \sum_{L \in \mathcal{L}(v_{i,j'})} Pr_\alpha(L)(Count_L(v_{i,j}) - Count_{F_{i\to i'}^{(j')}(L)}(v_{i,j})) \\
&= \sum_{L \in \mathcal{L}(v_{i,j'}) \wedge V(L) \subseteq V_i} Pr_\alpha^{(i)}(L) Count_L(v_{i,j})
\end{aligned}$$

The last equality is because (a) $Count_{F_{i\to i'}^{(j')}(L)}(v_{i,j})$ is zero before the walk on L leaves clique $V_{i'}$ for the first time, and after it leaves $V_{i'}$, it cancels with $Count_L(v_{i,j})$; and (b) $Pr_\alpha^{(i)}(L)$ with $V(L) \subseteq V_i$ is the aggregate probability of all paths with L as the prefix.

Therefore, $y - z = \sum_{L \in \mathcal{L}(v_{i,j'}) \wedge V(L) \subseteq V_i} Pr_\alpha^{(i)}(L) Count_L(v_{i,j})$ is the expected number of times for a random walk visiting $v_{i,j}$ from $v_{i,j'}$ before leaving V_i or restarting. Since the probability for a random walk leaving V_i or restarting is $\alpha + (1-\alpha)/n$, the expected length of a such path is $\frac{1}{\alpha + (1-\alpha)/n}$. Let the expected number of times $v_{i,j}$ appears in such paths is t_j for $j \neq j'$. Let t be the expected number of times $v_{i,j'}$ appears in such paths except the starting node. Note that $t + \sum_{j \neq j'} t_j = \frac{1}{\alpha + (1-\alpha)/n}$ and $\forall\, j, k \neq j'$, $t_j = t_k$. Furthermore, $t = \frac{(1-\alpha)(n-1)}{n} t_j < t_j$. We have

$$\frac{n}{n+1} < \frac{1}{(\alpha + (1-\alpha)/n)(n)} < t_j = y - z < \frac{1}{(\alpha + (1-\alpha)/n)(n-1)} < \frac{n}{n-1}.$$

Consider the second claim. We have

$$r = 1 + \frac{1-\alpha}{n}\mu + \frac{(n-1)(1-\alpha)}{n} y \leq 1 + \frac{1-\alpha}{n} r + \frac{(n-1)(1-\alpha)}{n} y.$$

Therefore, $r \leq \frac{1}{1-(1-\alpha)/n} + (n-1)(1-\alpha)y/(n-(1-\alpha))$ and $r - y < \frac{n}{n-(1-\alpha)} \leq \frac{n}{n-1}$.

For the third claim,

$$\begin{aligned} x - z &\leq \left(\frac{(1-\alpha)(n-1)}{n} z + \frac{1-\alpha}{n}\mu \right) - (1-\alpha) z \\ &= \frac{1-\alpha}{n}(\mu - z) \leq \frac{1-\alpha}{n}(r - z) \\ &< \frac{1-\alpha}{n}\left(\frac{n}{n-1} + (y - z)\right) < \frac{2}{n-1}. \end{aligned}$$

□

We are ready to present the proof of Lemma 8.

Proof. [of Lemma 8] By definition, $f_{n,m}(\alpha) = \mu$ and $g_{n,m}(\alpha) = y$. We want to show that $\mu < y$. By Lemma 10,

$$y - x = (y - z) + (z - x) > \frac{n}{n+1} - \frac{2}{n-1}.$$

We have

$$\begin{aligned} \mu &= \frac{1-\alpha}{m} r + \frac{(m-1)(1-\alpha)}{m} x < \frac{1}{m} r + \frac{m-1}{m} x \\ &< \frac{1}{m}\left(y + \frac{n}{n-1}\right) + \frac{m-1}{m}\left(y - \frac{n}{n+1} + \frac{2}{n-1}\right) \\ &= y + \frac{n}{m(n-1)} - \frac{(m-1)n}{m(n+1)} + \frac{2(m-1)}{m(n-1)} \\ &\leq y + \frac{2}{m} - \frac{(m-1)n}{m(n+1)} + \frac{2}{n-1} < y \end{aligned}$$

The last inequality holds when $4 \leq m < n$. □

References

1. M. S. Bartlett, An Inverse Matrix Adjustment Arising in Discriminant Analysis, *Annals of Mathematical Statistics*, Vol. 22, No. 1, pp. 107-111 (1951) MR40068.
2. Monica Bianchini, Marco Gori, and Franco Scarselli. Inside PageRank. *ACM Trans. Inter. Tech.*, 5(1):92C128, 2005.
3. Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1C7):107C117, 1998.
4. Alice Cheng and Eric Friedman. Manipulability of PageRank under sybil strategies. In *Proceedings of the First Workshop of Networked Systems (NetEcon06)*, 2006.
5. Eric Friedman, Paul Resnick, and Rahul Sami. Manipulation-resistant reputation systems. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory.* Cambridge University Press, 2007.
6. Zoltán Gyöngyi and Hector Garcia-Molina. Web spam taxonomy. In *First International Workshop on Adversarial Information Retrieval on the Web*, 2005.
7. Zoltán Gyöngyi and Hector Garcia-Molina. Link spam alliances. In *Proceedings of the 31st International Conference on Very Large Databases*, pages 517C528. ACM, 2005.
8. T. Hogg and L. Adamic. Enhancing reputation mechanisms via online social networks. *Proceedings of the 5th ACM Conference on Electronic Commerce*, pages 236C237, 2004.
9. J. Hopcroft and D. Sheldon. Manipulation-Resistant Reputations Using Hitting Time. *Lecture Notes in Computer Science*, 4863:68, 2007.
10. John Hopcroft, Daniel Sheldon, Network Reputation Games, *Manuscript*, eCommons@Cornell, http://ecommons.library.cornell.edu/handle/1813/11579, Oct. 2008.
11. Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in P2P networks. In *WWW'03: Proceedings of the 12th international World Wide Web Conference*, pages 640C651, New York, NY, USA, 2003. ACM Press.
12. Éva Tardos and TomWexler. Network formation games and the potential function method. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory.* Cambridge University Press, 2007.