Maximizing PageRank with new Backlinks

Martin Olsen

Center for Innovation and Business Development
Institute of Business and Technology*
Aarhus University
Birk Centerpark 15, DK-7400 Herning, Denmark
martino@hih.au.dk

Abstract. For a given node t in a directed graph $G(V_G, E_G)$ and a positive integer k we study the problem of computing a set of k new links pointing to t – so called backlinks to t – producing the maximum increase in the PageRank value of t. This problem is known as *Link Building* in the www context. We present a theorem describing how the topology of the graph comes in to play when evaluating potential new backlinks. Based on the theorem we show that no FPTAS exists for Link Building under the assumption NP≠P and we also show that Link Building is W[1]-hard.

1 Introduction

The founders of Google introduced the PageRank algorithm [1] that computes an estimate of the popularity of each web page based solely on the link structure of the web graph – these estimates are the so called *PageRank values*. A page will achieve one of the top spots of search results if it has a high PageRank value and matches the search criteria for the actual Google search. For a company it is extremely important that its web page does well in the ranking and *Search Engine Optimization* (SEO) is a billion dollar industry [2] helping companies to achieve this goal. The problem of obtaining optimal new backlinks in order to achieve good search engine rankings is known as *Link Building* (a backlink to a page t is a link pointing to t from another page). This paper addresses the problem of *identifying* new backlinks to t maximizing the PageRank value for t.

You might not find it realistic to obtain specific backlinks but Link Building attracts much attention in the SEO literature [3–6] and leading experts from the SEO industry consider Link Building to be an important aspect of SEO [7]. It should also be noted that there is a market for commercial link analysis software such as *LinkScape* offering users the ability to "*judge the quality of potential links*" to their sites according to a quote from `www.seomoz.org/linkscape`. There is even professional on line services for buying/selling links with `www.textlinkbrokers.com` as an example.

* Some of the research was carried out visiting NICTA, Sydney, as a PhD student at MADALGO, Department of Computer Science, Aarhus University.

Avrachenkov and Litvak [8] study the effect on PageRank if a given page establishes one or more links *to* other pages and show that an optimal linking strategy for a page is to establish links only to pages in the *community* of the page. When Avrachenkov and Litvak speak about a web community they mean "*... a set of Web pages that a surfer can reach from one to another in a relatively small number of steps*". It should be stressed that Avrachenkov and Litvak look for optimal links in $\{t\} \times V_G$ for a given page t where V_G is the nodes in the directed graph under consideration and that they conclude that t "*... cannot significantly manipulate its PageRank value by changing its outgoing links*". Kerchove et al. [9] study the more general problem of maximizing *the sum* of PageRank values for a set of pages T by adding links from $T \times V_G$. In [10] the author of this paper shows that computing k new links from $V_G \times V_G$ maximizing *the minimum* PageRank value for a given set of nodes T is NP-hard.

1.1 Contribution and Outline of the Paper

In this paper we obtain stronger intractability results compared to [10] for computing k new links in $V_G \times \{t\}$ maximizing the PageRank value of t which is a natural way to formulate the problem of *Link Building* from the perspective of a web master – the max-min–formulation in [10] is more artificial.

In Sect. 2 we briefly introduce the mathematical background and notation for the paper. In Sect. 3 we develop a theorem expressing among other things how the topology of the graph determines the PageRank potential for a set of backlinks to t. Based on the theorem we show in Sect. 4 that NP$\neq$P implies that no FPTAS exists for Link Building and we also show that Link Building is W[1]-hard. In the rest of this section we briefly explain the acronyms PTAS and FPTAS and the complexity classes FPT and W[1]:

PTAS and FPTAS: Consider a maximization problem "$\arg\max_x f(x)$" with solution x^*. A FPTAS (Fully Polynomial Time Approximation Scheme) can compute an x such that $f(x) \geq (1-\varepsilon)f(x^*)$ in time polynomial in $\frac{1}{\varepsilon}$ and the size of the instance. Some NP-hard problems allow a FPTAS (for example the Knapsack problem) and some do not. If there is no FPTAS for a problem there is still a chance for a PTAS (Polynomial Time Approximation Scheme) where we can obtain x in polynomial time for any *fixed* ε. As an example an algorithm with running time $n^{\frac{1}{\varepsilon}}$ counts as a PTAS but not as an FPTAS for a problem with instance size n.

FPT and W[1]: We will say that a problem with instance size n involving a parameter k is *fixed parameter tractable* if it can be solved in time $f(k)n^c$ where f is some function and c is independent of k. The class FPT contains the *decision* problems with this property. We will write A $\leq$ B if the problem A can be reduced to the problem B preserving fixed parameter tractability in the sense that B $\in$ FPT $\Rightarrow$ A $\in$ FPT. Consider the problems VERTEX COVER and INDEPEN-

DENT SET where we have to decide whether a graph contains a vertex cover[1] of size k or an independent set[2] of size k respectively. FPT is contained in the complexity class W[1] = {P : P $\leq$ INDEPENDENT SET}. Even though VERTEX COVER is NP-complete it has been solved for large n and $k = 400$ [11]. The reason is that VERTEX COVER $\in$ FPT with moderate f and c. A corresponding breakthrough is believed to be impossible for INDEPENDENT SET since there is strong evidence in the literature that FPT$\neq$W[1] so hardness for W[1] is accepted as evidence that a problem is fixed parameter *intractable*. According to a recent paper [12] then the currently "best algorithm" for INDEPENDENT SET runs in time $O(n^{0.792k})$ where the exponent of n increases dramatically with k. For more information on FPT and W[1] we refer to [13].

2 Mathematical Background

This section gives the mathematical background for the PageRank algorithm. We refer to [14] for more details on the PageRank algorithm. All vectors throughout this paper are column vectors. Let $G(V_G, E_G)$ denote a directed graph and let $|V_G| = n$. The nodes V_G and links E_G could as an example represent the pages and links in the web graph respectively – multiple links from u to v are treated as *one* link so G is an unweighted graph. A *random surfer* visits the nodes in V_G according to the following rules: When visiting u the surfer picks a link $(u, v) \in E_G$ uniformly at random and visits v. If u is a sink – a node not linking to any node – the next node to visit is chosen uniformly at random from V_G. The sequence of visited pages is a Finite Markov Chain with state space V_G and transition probability matrix $P = \{p_{uv}\}$ given by $p_{uv} = \frac{1}{\text{outdeg}(u)}$ if $(u, v) \in E_G$ and 0 otherwise where outdeg(u) is the out degree of u. If outdeg$(u) = 0$ then $p_{uv} = \frac{1}{n}$.

Now we modify the behavior of the random surfer so that he behaves as described above with probability $\alpha < 1$ when visiting u but *zaps* with probability $1 - \alpha$ to a node v chosen uniformly at random from V_G. In this paper we will assume that α is a *fixed constant*. If E is the matrix with each entry equal to 1 then the transition probability matrix Q for the modified Markov Chain is given by $Q = \frac{1-\alpha}{n}E + \alpha P$. The powers $w^T Q^i$ converge to the same probability distribution π^T for any initial probability distribution w on V_G as i tends to infinity – implying $\pi^T Q = \pi^T$. The vector $\pi = \{\pi_v\}_{v \in V_G}$ is known as the *PageRank vector* and the PageRank value for a node v is the probability that a random surfer visits v after i steps for large i.

The matrix $I - \alpha P$ is invertible and entry z_{uv} in $Z = (I - \alpha P)^{-1}$ is the *expected* number of visits – preceding the first zapping event – to page v for a random surfer starting at page u [8]. If $u = v$ then the initial visit is also included in the count. The entries in Z induce a sort of distance measure on the nodes in

[1] A vertex cover is a subset of the nodes satisfying that every edge has at least one endpoint in the set.

[2] A set of nodes in a graph is independent if no edge connects two of the nodes.

V_G: Two nodes u and v that are "close" to each other will have relatively large entries z_{uv} and z_{vu}. The following identity expresses the connection between π and Z [8] where e is the vector with all entries equal to 1:

$$\pi^T = \frac{1-\alpha}{n} e^T Z \ . \tag{1}$$

In this paper we look at the PageRank vector for the graph we obtain if we add a set of links E' to $G(V_G, E_G)$. We will let $\tilde{\pi}_v(E')$ denote the PageRank value of v in $G(V_G, E_G \cup E')$. The argument E' may be omitted if E' is clear from the context.

3 The Effect of Receiving Links

Avrachenkov and Litvak [8] study the effect on PageRank of adding new links *with the same origin* to the web graph. Avrachenkov and Litvak establish a Theorem that expresses the new PageRank vector $\tilde{\pi}$ by means of the "old" PageRank vector π and the "old" version of Z. We present a theorem showing the effect of adding new links *pointing to the same page*. Without loss of generality we will assume that each of the pages 2 to $k+1$ establishes a link *to* page 1. The techniques used in the proof are similar to the techniques used in [8].

Theorem 1. *Let each of the pages 2 to $k+1$ create a link to page 1. If $\tilde{\pi}_p$ denotes the updated PageRank value for page p for $p \in \{1, \dots, n\}$ then we have:*

$$\tilde{\pi}_p = \pi_p + \left[\pi_2 \ \pi_3 \ \dots \ \pi_{k+1} \right] M^{-1} q \ .$$

where $M = \{m_{ij}\}$ is a $k \times k$ matrix and q is a k-dimensional column vector given by

$$m_{ij} = \delta_{ij} k_{i+1} + z_{i+1j+1} - \alpha z_{1j+1} \ .$$

$$q_i = \alpha z_{1p} - z_{i+1p} + \delta_{i+1p} \ .$$

Here $k_i = outdeg(i)$ prior to the update and $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Theorem 1 shows how to express the *increase* (or *decrease*) in the PageRank value for the page p as a product of two factors: Roughly the first factor concerns the PageRank values of the nodes involved and the second factor $M^{-1}q$ concerns the "distances" between the nodes involved in the update.

Proof. Let e_i denote the n-dimensional column vector with a 1 at coordinate i and 0'es elsewhere and e denote the n-dimensional column vector with all 1'es. Let b_i denote the k-dimensional column vector with all 0'es except a 1 at coordinate i. Let $\tilde{P}$ denote the updated version of the matrix P. Then we have $\tilde{P} = P + \Delta_P$ where

$$\Delta_P = \sum_{i=2}^{k+1} e_i \frac{1}{k_i + 1} (e_1^T - e_i^T P) \ .$$

The corresponding change of $I - \alpha P$ is

$$(I - \alpha\tilde{P}) - (I - \alpha P) = -\alpha\Delta_P \ .$$

We will use the Woodbury formula [15] to compute $\tilde{Z} = (I - \alpha\tilde{P})^{-1}$ – the updated version of Z. In order to do this we find matrices S, T and U with dimensions $n \times k$, $k \times k$ and $k \times n$ respectively such that

$$-\alpha\Delta_P = STU \ .$$

We will use

$$S = -\sum_{i=2}^{k+1} e_i b_{i-1}^T \ .$$

$$T = \sum_{i=2}^{k+1} b_{i-1} b_{i-1}^T \frac{1}{k_i + 1} \ .$$

$$U = \sum_{i=2}^{k+1} \alpha b_{i-1}(e_1^T - e_i^T P) \ .$$

According to the Woodbury formula we have the following

$$\tilde{Z} = Z - ZS(T^{-1} + UZS)^{-1}UZ \ . \tag{2}$$

Since $(I - \alpha P)Z = I$ we have that $\alpha PZ = Z - I$ and consequently

$$UZ = \sum_{i=2}^{k+1} b_{i-1}(\alpha e_1^T Z - e_i^T(Z - I)) \ .$$

Now we can calculate UZS:

$$\begin{aligned} UZS &= \sum_{i=2}^{k+1}\sum_{j=2}^{k+1} b_{i-1}(e_i^T(Z - I) - \alpha e_1^T Z)e_j b_{j-1}^T \\ &= \sum_{i=2}^{k+1}\sum_{j=2}^{k+1} b_{i-1}(z_{ij} - \delta_{ij} - \alpha z_{1j})b_{j-1}^T \ . \end{aligned}$$

The entry in row i and column j in the $k \times k$ matrix $M = T^{-1} + UZS$ is

$$\begin{aligned} m_{ij} &= \delta_{ij}(k_{i+1} + 1) + z_{i+1j+1} - \delta_{ij} - \alpha z_{1j+1} \\ &= \delta_{ij} k_{i+1} + z_{i+1j+1} - \alpha z_{1j+1} \ . \end{aligned}$$

Now we multiply (2) with $\frac{1-\alpha}{n} e^T$ from the left and e_p from the right. By using (1) we get

$$\tilde{\pi}_p = \pi_p - \pi^T S M^{-1} U Z e_p \ .$$

The i'th entry in the k-dimensional column vector $q = UZe_p$ is

$$q_i = \alpha z_{1p} - z_{i+1p} + \delta_{i+1p} \ .$$

The i'th entry in the k-dimensional row vector $-\pi^T S$ is π_{i+1}. □

In the next section we will use Theorem 1 to prove results on the computational complexity of Link Building.

4 Link Building is W[1]-hard and Allows no FPTAS

We now formally define the problem of LINK BUILDING:

Definition 1. *The* LINK BUILDING *problem:*

- *Instance: A triple* (G, t, k) *where* $G(V_G, E_G)$ *is a directed graph,* $t \in V_G$ *and* $k \in \mathbb{Z}^+$.
- *Solution: A set* $S \subseteq V_G \setminus \{t\}$ *with* $|S| = k$ *maximizing* $\tilde{\pi}_t(S \times \{t\})$.

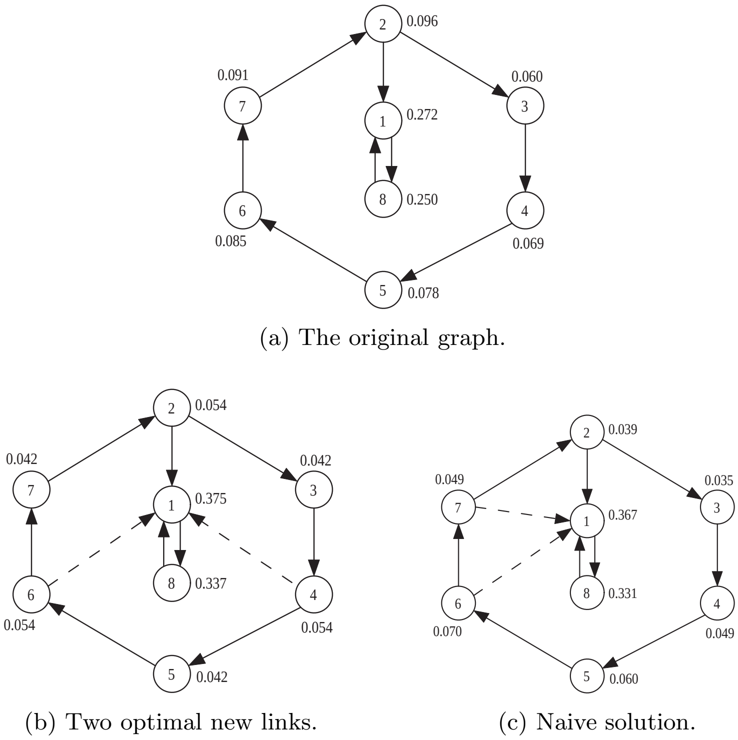

Fig. 1: Link Building examples. The PageRank values for the *modified* graphs are shown besides the nodes.

Example: Now consider an instance of the LINK BUILDING problem given by the graph in Fig. 1a together with $t = 1$ and $k = 2$. In other words we are looking for two optimal sources for links to node 1. The solution $S = \{4, 6\}$ is shown in Fig. 1b. You might think that the solution would be $\{6, 7\}$ as shown in Fig. 1c because nodes 6 and 7 are the two most popular nodes in $\{3, 4, 5, 6, 7\}$ prior to the modification. The nodes $\{2, 4, 6\}$ is an independent set with relatively low

z-values between them making the set a good choice of sources for links to node 1 according to Theorem 1.

We will show how to embed *any* regular[3] graph in a graph containing a node t such that obtaining links from nodes in an independent set of the regular graph will be optimal for t. All nodes in a regular graph have the same PageRank value and out degree so the z-values become crucial. We just have to make sure that z_{uv} is *very* big if u and v are neighbors compared to the case where they are not neighbors – this is the intuition for the proof below. To put it more formally we show that no FPTAS for LINK BUILDING exists under the assumption NP≠P by reduction from the independent set problem restricted to undirected regular graphs. This problem is known to be NP-complete even for 3-regular graphs [16, 17]. We need a couple of definitions to clarify matters:

Definition 2. *The* REGULAR INDEPENDENT SET *problem:*

- *Instance: An undirected regular graph $H(V_H, E_H)$ and an integer $k \geq 2$.*
- *Question: Does H contain an* independent set *of size k?*

Definition 3. *Let S^* be a solution to the LINK BUILDING problem. A* FPTAS *for the LINK BUILDING problem is an algorithm that given input (G, t, k, ε) computes a feasible solution S to the LINK BUILDING problem satisfying*

$$\tilde{\pi}_t(S \times \{t\}) > (1 - \varepsilon)\tilde{\pi}_t(S^* \times \{t\})$$

in time polynomial in $\frac{1}{\varepsilon}$ and the size of (G, t, k).

We now state our first theorem on the computational complexity of LINK BUILDING:

Theorem 2. *If NP≠P then there is no FPTAS for LINK BUILDING.*

Proof. We show how to solve an instance of the REGULAR INDEPENDENT SET problem in polynomial time if we have a FPTAS to the LINK BUILDING problem at our disposal.

Now let the regular graph $H(V_H, E_H)$ and the number $k \in \mathbb{Z}^+$ represent an instance of the REGULAR INDEPENDENT SET problem and let d denote the degree of all nodes in H. From $H(V_H, E_H)$ we now construct the graph $G(V_G, E_G)$ shown in Fig. 2 in polynomial time in the following way:

1. The nodes in G are all the nodes in H together with two new nodes named 1 and r and a clique[4] consisting of y nodes with $y > \frac{4kx}{1-\alpha}$ where $x = \left\lceil \frac{2d^2k^3}{1-\alpha^2} \right\rceil$.
2. For every node in V_H we add links to $x - d$ nodes in the clique. We add the links such that no node in the clique has more than one incoming link from V_H (we choose $y > (x - d)|V_H|$).
3. For every edge $\{u, v\} \in E_H$ we add two links (u, v) and (v, u) to E_G.
4. Finally, we add links $(1, r)$ and $(r, 1)$.

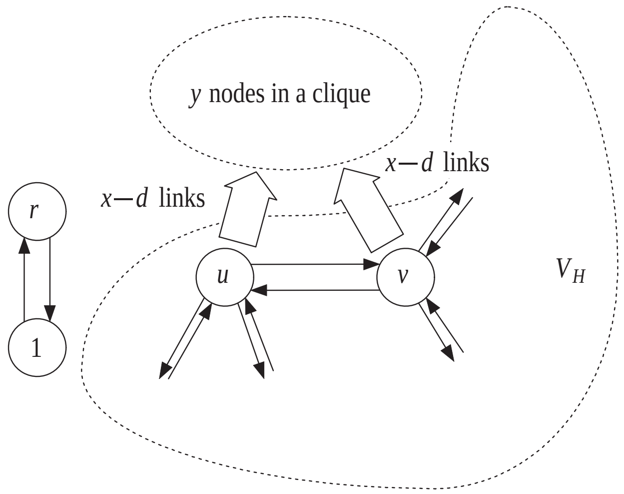

Fig. 2: The graph $G(V_G, E_G)$.

The graph G is constructed such that one step transitions between the nodes in V_H are much more likely to happen compared to transitions of two steps or more. This will make independent sets in H superior sets of sources for links to node 1 since all nodes in V_H have the same out degree and PageRank value. The following lemma quantifies the advantage for node 1 for obtaining backlinks from independent sets in H.

Lemma 1. *Let $S_1 \subseteq V_H$ be an independent set in H and let $S_2 \subseteq V_G \setminus \{1\}$ be a set with $|S_1| = |S_2| = k$. If S_2 is* not *an independent set in H then the following holds:*

$$\tilde{\pi}_1(S_1 \times \{1\}) - \tilde{\pi}_1(S_2 \times \{1\}) > \rho n^{-1} d^{-6} k^{-9} \ , \tag{3}$$

where $n = |V_G|$ and ρ is a positive constant only dependent on α.

The proof of Lemma 1 is presented after the proof of Theorem 2 in an attempt to make the structure of the proof of Theorem 2 clear to the reader.

Now assume that H contains an independent set of size k. From Lemma 1 we conclude that any solution S^* to the LINK BUILDING problem $(G, t = 1, k)$ must be independent and that a constant ρ exists such that the following holds for any feasible solution S which is *not* an independent set in H:

$$\tilde{\pi}_1(S \times \{1\}) < (1 - \rho n^{-1} d^{-6} k^{-9}) \tilde{\pi}_1(S^* \times \{1\}) \ . \tag{4}$$

This shows that we can decide whether an independent set exists by activating our LINK BUILDING FPTAS with input $(G, t = 1, k, \varepsilon = \rho n^{-1} d^{-6} k^{-9})$ and

[3] A regular graph is a graph where all nodes have the same degree.

[4] A clique is a set of nodes where all nodes link to all other nodes

check whether the solution from the FPTAS is independent. Thus we can solve the REGULAR INDEPENDENT SET problem in polynomial time using the LINK BUILDING FPTAS implying NP=P. □

Before presenting the second theorem on the computational complexity of LINK BUILDING we need to prove Lemma 1 used in the proof of Theorem 2:

Proof (Lemma 1). Now we will apply Theorem 1 and examine what happens if we add links to node 1 from an independent set S_1 in H and another set S_2 both consisting of k nodes. First, we will consider the case where $S_2 \subseteq V_H$ is a non-independent set in H. In order to align the proof with Theorem 1 we will refer to the nodes in S_1 and S_2 as nodes 2, 3, ..., $k+1$.

According to Theorem 1 we have the following:

$$\tilde{\pi}_1 - \pi_1 = \left[\, \pi_2 \; \pi_3 \; \ldots \; \pi_{k+1} \,\right] M^{-1} q \; . \tag{5}$$

Let $B = \{b_{ij}\}$ be the $k \times k$ matrix defined by the following identities:

$$b_{ij} = \frac{z_{i+1j+1}}{x+1} \text{ if } i \neq j \text{ and } b_{ii} = \frac{z_{i+1i+1} - 1}{x+1} \quad . \tag{6}$$

Now we have

$$M = (x+1)(I+B) \; .$$

If $\bar{b}$ is an upper bound on the entries in B then it is not hard to show that $k^{s-1}\bar{b}^s$ is an upper bound on the entries in B^s:

$$0 \leq B^s \leq k^{s-1}\bar{b}^s E = k^{-1}(k\bar{b})^s E \; . \tag{7}$$

For S_1 we can use the following upper bound:

$$\bar{b}_1 = \frac{1}{x+1}\left(\frac{d\alpha}{x}\right)^2 \frac{1}{1-\alpha^2} \leq \left(\frac{d^2}{x^3}\right)\left(\frac{\alpha^2}{1-\alpha^2}\right) \quad . \tag{8}$$

Here we use that $\left(\frac{d\alpha}{x}\right)$ is the probability of following a link and staying in V_H for a random surfer starting in V_H. Because S_1 is independent then $\left(\frac{d\alpha}{x}\right)^2$ is an upper bound on the probability of reaching j from node i without zapping. We also use that $1 + \alpha^2 + \alpha^4 + \alpha^6 + \cdots = \frac{1}{1-\alpha^2}$ is an upper bound on z_{jj}.

We also get an upper bound for S_2:

$$\bar{b}_2 = \frac{1}{x+1}\left(\frac{d\alpha}{x}\right) \frac{1}{1-\alpha^2} \leq \left(\frac{d}{x^2}\right)\left(\frac{\alpha}{1-\alpha^2}\right) \quad . \tag{9}$$

For $x = \left\lceil \frac{2d^2k^3}{1-\alpha^2} \right\rceil$ we have $k\bar{b} < 1$ and hence we have the following:

$$M^{-1} = \frac{1}{x+1}(I - B + B^2 - B^3 + B^4 - \cdots) = \frac{1}{x+1}\sum_{s=0}^{\infty}(-1)^s B^s \quad . \tag{10}$$

Now consider a probability distribution w on V_G with the same probability mass for each entry corresponding to a node in V_H. All entries in $w^T Q^i$ corresponding to nodes in V_H will have the same probability mass for any i because H is regular. The limiting distribution π^T will also have this property. This means that a number β exists such that:

$$\left[\pi_2\ \pi_3\ \dots\ \pi_{k+1}\right] = \beta e^T \ . \tag{11}$$

We now use Theorem 1 for $p = 1$ and get the following identity for q where we use that $z_{11} = \frac{1}{1-\alpha^2}$ and $z_{i1} = 0$ for $i \in \{2, 3, \dots, k+1\}$:

$$q = \frac{\alpha}{1-\alpha^2} e \ . \tag{12}$$

We now insert the results from (10), (11) and (12) in (5):

$$\tilde{\pi}_1 - \pi_1 = \frac{\alpha\beta}{(x+1)(1-\alpha^2)} \sum_{s=0}^{\infty} (-1)^s e^T B^s e \ . \tag{13}$$

We now use (7) to establish a lower bound of the factor $\sum_{s=0}^{\infty}(-1)^s e^T B^s e$ for S_1:

$$\sum_{s=0}^{\infty}(-1)^s e^T B^s e \geq k(1 - \bar{b}_1 k - (\bar{b}_1 k)^3 - (\bar{b}_1 k)^5 - \cdots) = k\left(1 - \frac{\bar{b}_1 k}{1-(\bar{b}_1 k)^2}\right) \ . \tag{14}$$

We will now develop an upper bound for $\sum_{s=0}^{\infty}(-1)^s e^T B^s e$ for S_2. There are two nodes u and v in S_2 such that $(u, v), (v, u) \in E_G$. The probability of reaching v for a random surfer starting at u – preceding the first zapping event – is greater than $\frac{\alpha}{x}$:

$$b_{vu}, b_{uv} \geq \frac{1}{x+1}\frac{\alpha}{x} \geq \frac{1}{x^2}\frac{\alpha}{2} \ . \tag{15}$$

Now we can construct the desired upper bound:

$$\sum_{s=0}^{\infty}(-1)^s e^T B^s e \leq k(1 - \frac{1}{k}(b_{uv} + b_{vu}) + (\bar{b}_2 k)^2 + (\bar{b}_2 k)^4 + (\bar{b}_2 k)^6 + \cdots) \tag{16}$$

$$= k\left(1 - \frac{1}{k}(b_{uv} + b_{vu}) + \frac{(\bar{b}_2 k)^2}{1-(\bar{b}_2 k)^2}\right) \ .$$

By inserting the lower bound from (14) and the upper bound from (16) in (13) we now conclude that

$$\tilde{\pi}_1(S_1 \times \{1\}) - \tilde{\pi}_1(S_2 \times \{1\}) \geq$$
$$\frac{\alpha\beta}{(x+1)(1-\alpha^2)} k\left(\frac{1}{k}(b_{uv} + b_{vu}) - \frac{(\bar{b}_2 k)^2}{1-(\bar{b}_2 k)^2} - \frac{\bar{b}_1 k}{1-(\bar{b}_1 k)^2}\right) \ . \tag{17}$$

For $x = \left\lceil \frac{2d^2k^3}{1-\alpha^2} \right\rceil$ we have that $(\bar{b}_1 k)^2$ and $(\bar{b}_2 k)^2$ are both less than $\frac{1}{2}$ which implies the following where we also use (8), (9) and (15):

$$\frac{1}{k}(b_{uv} + b_{vu}) - \frac{(\bar{b}_2 k)^2}{1 - (\bar{b}_2 k)^2} - \frac{\bar{b}_1 k}{1 - (\bar{b}_1 k)^2} \geq$$

$$\frac{1}{k}(b_{uv} + b_{vu}) - 2(\bar{b}_2 k)^2 - 2\bar{b}_1 k \geq$$

$$\frac{1}{k}\frac{\alpha}{x^2} - 2\left(\frac{d}{x^2}\right)^2 \left(\frac{\alpha}{1-\alpha^2}\right)^2 k^2 - 2\left(\frac{d^2}{x^3}\right)\left(\frac{\alpha^2}{1-\alpha^2}\right) k \geq$$

$$k^{-1}x^{-2}\left(\alpha - \frac{1}{2}\alpha^2 d^{-2} k^{-3} - \alpha^2 k^{-1}\right) \geq$$

$$k^{-1}x^{-2}\left(\alpha - \frac{1}{16}\alpha^2 - \frac{1}{2}\alpha^2\right) .$$

We now use this inequality together with $\beta > \frac{1-\alpha}{n}$ and $2x > x+1$ to replace the lower bound in (17):

$$\tilde{\pi}_t(S_1 \times \{t\}) - \tilde{\pi}_t(S_2 \times \{t\}) \geq$$

$$\frac{\alpha(1-\alpha)}{2(1-\alpha^2)} x^{-1} n^{-1} k \cdot k^{-1} x^{-2}\left(\alpha - \frac{1}{16}\alpha^2 - \frac{1}{2}\alpha^2\right) =$$

$$\frac{\alpha(1-\alpha)}{2(1-\alpha^2)} n^{-1} x^{-3}\left(\alpha - \frac{1}{16}\alpha^2 - \frac{1}{2}\alpha^2\right) .$$

which shows that (3) holds.

Up till now we have shown that (3) holds if $S_1 \subseteq V_H$ is an independent set from H and $S_2 \subseteq V_H$ is *not* an independent set from H. It is a bit technical but not hard to show that $\tilde{\pi}_1(S_1 \times \{1\}) > \tilde{\pi}_1(S_2 \times \{1\})$ holds if S_1 is *any* subset of V_H and $S_2 \subseteq V_G \setminus \{1\}$ is a subset of V_G such that $|S_1| = |S_2| = k$ and $|S_2 \cap V_H| < k$. Informally, we show that node 1 should prefer links from V_H compared to links from the clique – partly due to the relatively high out degree of the nodes in the clique. This part of the proof is omitted for the sake of brevity. □

The REGULAR INDEPENDENT SET problem is W[1]-complete [18] so we immediately get the following theorem because k is preserved in the reduction in the proof of Theorem 2 and because the construction of G and the check of independence runs in polynomial time:

Theorem 3. *If W[1]≠FPT then LINK BUILDING is not fixed parameter tractable.*

Acknowledgments The author of this paper thanks Joachim Gudmundsson, NICTA, and Gerth Brodal, Aarhus University, for constructive criticism.

References

1. Brin, S., Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems 30, 107–117 (1998)
2. Olsen, P.R.: A Future in Directing Online Traffic. The New York Times, http://www.nytimes.com/2009/01/11/jobs/11starts.html (2009)
3. Blankson, S.: SEO. How to Optimize your Web Site for Internet Search Engines. Blankson Enterprises Ltd, London (2008)
4. Kent, P.: Search Engine Optimization For Dummies. Wiley, Indianapolis (2006)
5. Ledford, J.L.: Search Engine Optimization Bible. Wiley, Indianapolis (2008)
6. Smith, J.: Get into Bed with Google. The Infinite Ideas Company Limited, Oxford (2008)
7. SEOmoz: Search engine ranking factors 2009. http://www.seomoz.org/article/search-ranking-factors (2009)
8. Avrachenkov, K., Litvak, N.: The Effect of new Links on Google Pagerank. Stochastic Models 22, 319–331 (2006)
9. de Kerchove, C., Ninove, L., van Dooren, P.: Maximizing PageRank via outlinks. Linear Algebra and its Applications 429, 1254–1276 (2008)
10. Olsen, M.: The Computational Complexity of Link Building. In: Computing and Combinatorics, 14th International Conference, COCOON 2008. LNCS, vol. 5092, pp. 119–129. Springer, Heidelberg (2008)
11. Cheetham, J., Dehne, F., Rau-chaplin, A., Stege, U., Taillon, P.J.: Solving Large FPT Problems on Coarse Grained Parallel Machines. Journal of Computer and System Sciences 67, 691–706 (2002)
12. Chen, J., Meng, J.: On Parameterized Intractability: Hardness and Completeness. Comput. J. 51, 39–59 (2008)
13. Downey, R., Fellows, M.: Parameterized Complexity. Springer, New York (1999)
14. Langville, A.N., Meyer, C.D.: Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton (2006)
15. Harville, D.A.: Matrix Algebra From a Statistician's Perspective. Springer (1997)
16. Elias, I.: Settling the Intractability of Multiple Alignment. Technical Report TRITA-NA-0316, Nada, KTH (2003)
17. Garey, M., Johnson, D.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman, New York (1979)
18. Cai, L.: Parameterized Complexity of Cardinality Constrained Optimization Problems. Comput. J. 51, 102–121 (2008)