Assessing the Navigational Effects of Click Biases and Link Insertion on the Web

Florian Geigl
KTI, Graz University of
Technology
florian.geigl@tugraz.at

Kristina Lerman
ISI, University of Southern
California
lerman@isi.edu

Simon Walk
IICM, Graz University of
Technology
simon.walk@tugraz.at

Markus Strohmaier
University of Koblenz-Landau
and GESIS
markus.strohmaier@gesis.org

Denis Helic
KTI, Graz University of
Technology
dhelic@tugraz.at

ABSTRACT

Websites have an inherent interest in steering user navigation in order to, for example, increase sales of specific products or categories, or to guide users towards specific information. In general, website administrators can use the following two strategies to influence their visitors' navigation behavior. First, they can introduce *click biases* to reinforce specific links on their website by changing their visual appearance, for example, by locating them on the top of the page. Second, they can utilize *link insertion* to generate new paths for users to navigate over. In this paper, we present a novel approach for measuring the potential effects of these two strategies on user navigation. Our results suggest that, depending on the pages for which we want to increase user visits, optimal link modification strategies vary. Moreover, simple topological measures can be used as proxies for assessing the impact of the intended changes on the navigation of users, even before these changes are implemented.

Keywords

Click Biases, Link Insertion, Random Surfer, Stationary Distribution

1. INTRODUCTION

Millions of people use the Web on a daily basis to buy products in online shops, perform financial transactions via online banking, or simply browse information systems, media libraries or online encyclopedias, such as IMDb, Netflix or Wikipedia. To find and access relevant information on the Web, people either search, navigate, or combine these two activities. A recent study [10] found that 35% of all visits to a website can be attributed to teleports, which are the direct result of clicks on search-engine results, navigation through manually typed URLs, or clicks on browser book-

marks. The remaining 65% of the clicks can be attributed to the task of navigating a webpage. In this paper, we direct our attention towards these 65% of actions and tackle the question what potential effects we can expect if we influence the link selection process of website visitors by simple link modifications. In particular, we are interested in the effects of different link modification strategies on (stochastic) models of Web navigation.

Problem. By inserting new links between webpages of a website, we alter the link structure. This has the potential to change user browsing behavior, since new links create new paths for users to explore the website. Alternatively, without changing the link structure of the website, we might be able to influence the link selection process of visitors. Studies have shown that the decisions of users for where to navigate next can be influenced by the layout and the position of the links on a webpage. In particular, due to position bias [20] users are more likely to select links higher up on webpages [5, 6, 24]. As a result, inducing click biases, such as repositioning links on a webpage, highlighting the links, or even making them visually more appealing, can affect the users' decision of where to click next on a website, similar to the way that adding new links affects browsing.

In this paper we are particularly interested in investigating and comparing the potential consequences of inserting new links and modifying already existing links on the navigational behavior of users. These newly obtained insights are of a significant practical relevance for website owners, as they can be used, for example, by owners of media libraries to increase visits of specific media files in order to reduce the number of different files that need to be cached on fast storage devices. Another example includes online encyclopedias, where operators may want to guide users towards articles of a specific category over some period of time (e.g., the birthday of an inventor). In some of these cases, link insertion might be more time-consuming than simply changing the layout of the website to increase visibility of specific links and vice versa. Theoretically, we would like to analyze and compare the effects of such link modification endeavors. Practically, new tools are needed to assist website operators in deciding which of the two strategies they should deploy to achieve the desired effects.

Methods. In this paper we study the impact of link modifications on the random surfer, which we apply as a proxy

for real user behavior. In the past, a user's decision to click on a link on a webpage was successfully modeled using the random surfer [4, 15, 30]. In this model, a user selects one of the links on a webpage uniformly at random and navigates to the page to which the link points. Apart from the huge success of the Google search engine, whose ranking algorithm is based on the random surfer model, empirical studies have shown that this model provides a very precise approximation of real browsing behavior in many situations and for a variety of applications [4, 9]. An important property of a random surfer is its *stationary distribution*, which is the probability distribution of finding a random surfer at a specific webpage in the limit of large number of steps.

In particular, we investigate how the random surfer's stationary distribution of a subset of pages (i.e., *target pages*) of a given website changes as a consequence of (i) modifying already existing links towards them, (ii) introducing new links towards them, or by (iii) combining these two approaches. To that end, we introduce a *click bias*, and a *link insertion* strategy. We model the effects of click biases on the intrinsic attractiveness of a link to the user by increasing the weight of that link. In practice, we may introduce such click biases, for example, by locating the corresponding link on the top of a page. With link insertion, we simply introduce new links between webpages of a website, for example, by linking towards a given target page from the starting page.

We introduce quantitative measures that allow us to address the following research questions:
Navigational Boost. How stable is the stationary distribution with respect to the proposed modification strategies, and what are the limits of stationary distributions that can be achieved for a given set of webpages? Is it (theoretically) possible to achieve a given stationary probability distribution for an arbitrary subset of webpages of a website? What is the connection between simple topological measures of the website network and stationary probability?
Influence Potential. What is the relative gain of the stationary probabilities compared to their unmodified counterparts. This provides us with an answer to the "guidance" potential of a set of webpages, defining to what extent it is possible to increase the relative stationary probabilities as compared to the initial unmodified values.
Combinations. Finally, we are interested how combinations of the two proposed link modification strategies perform in terms of increased stationary probabilities of selected subpages. In particular, we investigate the performance of certain combinations across several different networks and/or selected subpages.
Contributions & Findings. We find that intuitions about how either modification strategy affects navigation are not always correct. Further, our experiments show that the size of a set of targeted subpages is not always a good predictor for the observed effects. Rather, other topological features often better reflect the consequences of a modification. Practically, we provide an open source framework[1] for website administrators to estimate the effects of link modifications on their website.

2. RELATED WORK

The random surfer model has received much attention from the research community [22, 31]. While the model is

---

[1] https://github.com/floriangeigl/RandomSurfers

very simple, it became well-established over the last years. It was applied to a variety of problems from graph generators over graph analysis to modeling user navigation. Furthermore, the model has been applied to calculate structural node properties in large networks. HITS [18] and PageRank [4, 26] rank network nodes according to their values in the stationary distribution of the random surfer model. Especially for the later there exists a detailed analysis ranging from the efficiency of its calculation towards its robustness [3, 19]. Bianchini et al. [3] provided an in-depth analysis of how to tweak the cumulative PageRank of a community of websites. They found that splitting up the content of pages onto more highly interlink pages increases the community's cumulative PageRank—since the community is larger it consists of more pages which are able to trap the random surfer for a longer period of time. Moreover, they suggest to avoid dangling webpages (i.e., pages without links to other pages). In this paper we are also interested in the sum of the random surfers visit probabilities in a community, however we do not use (i) teleportation as in the PageRank model, and (ii) do not modify the network in its size (i.e., number of pages). On the contrary we modify the transition probabilities of certain links and insert new links into the network. Moreover, since all our datasets are strongly connected, we do not face the problem of unwanted high visit probabilities of usually unimportant pages (i.e., dangling nodes) [3].

A random surfer can be steered towards specific nodes in the network by increasing the probability of traversing links towards those nodes. This can be accomplished by biasing random surfer's link selection strategy so that it is not uniformly random anymore, but biased towards specific nodes. For instance, in the field of information retrieval Richardson et al. [28] successfully applied biased random surfers to increase the quality of search results compared to those achieved using a simple PageRank. At the same time Haveliwala [13, 14] biased PageRank towards topics retrieved from a search query to rank the query results. Utilizing this technique the results where more accurate than those produced using a single, generic PageRank. Moreover, Gyongyi et al. [12] successfully used trust as bias to detect and filter out spam pages of search results. However, Al-Saffar and Heileman [1] showed that biased PageRank algorithms generate a considerable overlap in top results with a simple PageRank. Concerning this problem their main suggestion was to use external biases which do not rely onto the underlying link structure of the network. In our paper we randomly decide towards which nodes we bias the random surfer. This allows us to explore the borders of changes in stationary distributions caused by a bias.

In 2013, Helic et al. [15] compared click trails characteristics of stochastically biased random surfers with those of humans. Their conclusion was, that biased random surfers can serve as valid models of human navigation. Further, Geigl et al. [9] validated this by showing that the result vector of PageRank and clickdata biased PageRank have a strong correlation in an online encyclopedia. This is especially interesting, since it creates the connection of our simulation to real human navigation on the web. Additionally, Lerman and Hogg [20] already showed that it is possible to bias the link selection of users. In particular, they came to the conclusion that users are subject to a *position bias*, making the selection of links higher up on webpages up to a factor of 3.5 more likely [5, 20, 24]. Hence, it is of practical relevance

to investigate also the effects of *biases* in the link selection process onto the stationary distribution.

Concerning link insertion there already exists work in literature which makes use of statistical methods to suggest new links in network structures to, for instance, increase the performance of chip architectures [25]. In particular, the authors use a standard mesh and insert long-range links, converting the network into a small-world network. This reduced packet latency results in a major improvement in throughput. Another field of research where link insertion is of interest are recommender systems for social friendship networks [2, 21, 23, 29]. For example, Xie et al. [32] characterized interests of users in two dimensions (i.e., context and content) and exploited this information to efficiently recommend potential new friends in an online social network. In this paper we focus on the effects of inserted links onto the typical whereabouts of the random surfer.

3. METHODOLOGY

We base our methodology on the calculations of the stationary distribution of a random surfer on the original and manipulated networks. The networks consist of nodes, which represent webpages and directed links between nodes, which represent hyperlinks between webpages. We first calculate the transition matrix and the stationary distribution for the original network as a baseline for comparing the effects of link modifications. Second, we increase the statistical weight of a random surfer visiting a set of predefined nodes (i.e., *target pages* or *target nodes*). We do that either by increasing the link weights towards selected nodes (click bias) or by adding new links pointing towards those nodes (link insertion). Third, we compute the corresponding transition matrix for the modified network. Fourth, we calculate the stationary distribution of the new transition matrices. Finally, we compare the modified stationary distribution with the original stationary distribution to gain insights into the effects of the different link modifications. Figure 1 illustrates these steps on a toy example.

3.1 Preliminaries

In what follows we formalize our approach algebraically. We represent a website as a directed network with a weighted adjacency matrix $\boldsymbol{W} \in \mathbb{R}^{n \times n}$, where n is the number of webpages in the website under investigation. We define the element W_{ij} of the weighted adjacency matrix $\boldsymbol{W}$ as the sum of edge weights of all links pointing from node j to node i. For example, $W_{ij} = 1$ if there is a single link from page j to page i with weight 1, and $W_{ij} = 3$ if there are three links pointing from page j to page i each with weight 1.

For our analysis we introduce *target nodes* as the nodes whose stationary probability we want to increase. We use vector $\boldsymbol{t} \in \mathbb{R}^n$ to specify them:

$$t_i = \begin{cases} 1 & \text{if } i \text{ is a } target\ node \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

We further define ϕ as a fraction of target nodes with respect to the total number of nodes n:

$$\phi = \frac{\sum_i t_i}{n} \quad (2)$$

Hence, $\phi = 0.1$ means that 10% of nodes from the network are target nodes.

3.2 Stationary Distribution

The stationary distribution is a probability distribution over nodes that assigns a probability of finding the random surfer on a given node in the limit of large number of steps. To compute the stationary distribution we first need to construct a diagonal out-degree matrix $\boldsymbol{D}$, with the weighted node out-degrees on its diagonal. Using $\text{diag}(\boldsymbol{v})$ to denote diagonal matrices with elements of a vector $\boldsymbol{v}$ on their diagonal we define $\boldsymbol{D}$ as:

$$\boldsymbol{D} = \text{diag}\left(\sum_{i=1}^n W_{ij}\right). \quad (3)$$

Using $\boldsymbol{D}$ matrix we can calculate the transition matrix $\boldsymbol{P}$, which is a left stochastic matrix of $\boldsymbol{W}$ as $\boldsymbol{P} = \boldsymbol{W}\boldsymbol{D}^{-1}$ (in fact this is PageRank matrix without teleportation). The stationary distribution $\boldsymbol{\pi}$ now satisfies the (right) eigenvalue equation for the matrix $\boldsymbol{P}$: $\boldsymbol{\pi} = \boldsymbol{P}\boldsymbol{\pi}$.

3.3 Click Bias

To introduce click biases that influence the link selection strategy of the random surfer, we reweigh the links pointing towards target nodes by multiplying their weight by a constant scalar b, which we call bias strength. For example, a bias strength of $b = 2$ doubles the weight of all links towards target nodes. The final probability of the random surfer to traverse a link is then directly proportional to its weight.

Algebraically, we induce biases with a diagonal bias matrix $\boldsymbol{B}$ which we define as $\boldsymbol{B} = \boldsymbol{I} + (b - 1) \cdot \text{diag}(\boldsymbol{t})$. The adjacency matrix of a biased network is $\boldsymbol{W}' = \boldsymbol{B}\boldsymbol{W}$. To compute the stationary distribution of the biased network, we first calculate the new transition matrix $\boldsymbol{P}' = \boldsymbol{W}'\boldsymbol{D}'^{-1}$ and then its stationary distribution $\boldsymbol{\pi}'$.

Please note that from the technical perspective, inducing a bias is the same as inserting parallel links towards target nodes—it increases the value of specific elements (i.e., those representing links towards target nodes) in the adjacency matrix. The total weight of newly added parallel links $l(b)$ due to an induced bias b is given by:

$$l(b) = \underbrace{\sum_{ij} W'_{ij}}_{\#\ \text{links in } \boldsymbol{W'}} - \underbrace{\sum_{ij} W_{ij}}_{\#\ \text{links in } \boldsymbol{W}} \quad (4)$$

To allow for a fair comparison between the click bias and the link insertion strategy we insert exactly $l(b)$ new links with weight 1 in the latter case.

3.4 Link Insertion

The second link modification strategy consists of inserting new links towards the target nodes from a given set of source nodes. This strategy represents the case where a website administrator inserts links towards target nodes from important subpages of their website. We define the importance of a webpage as its stationary probability in the original network.

To insert a given number $l(b)$ of new links we proceed as follows. We start by sorting nodes by their stationary probability in a descending order. In the next step we insert new links from the top $l(b)/(n \cdot \phi)$ nodes to all target nodes. Here $n \cdot \phi$ is the number of target nodes and we always *ceil* the calculated number of source nodes to ensure that there are enough pairs of nodes. If one of the target nodes
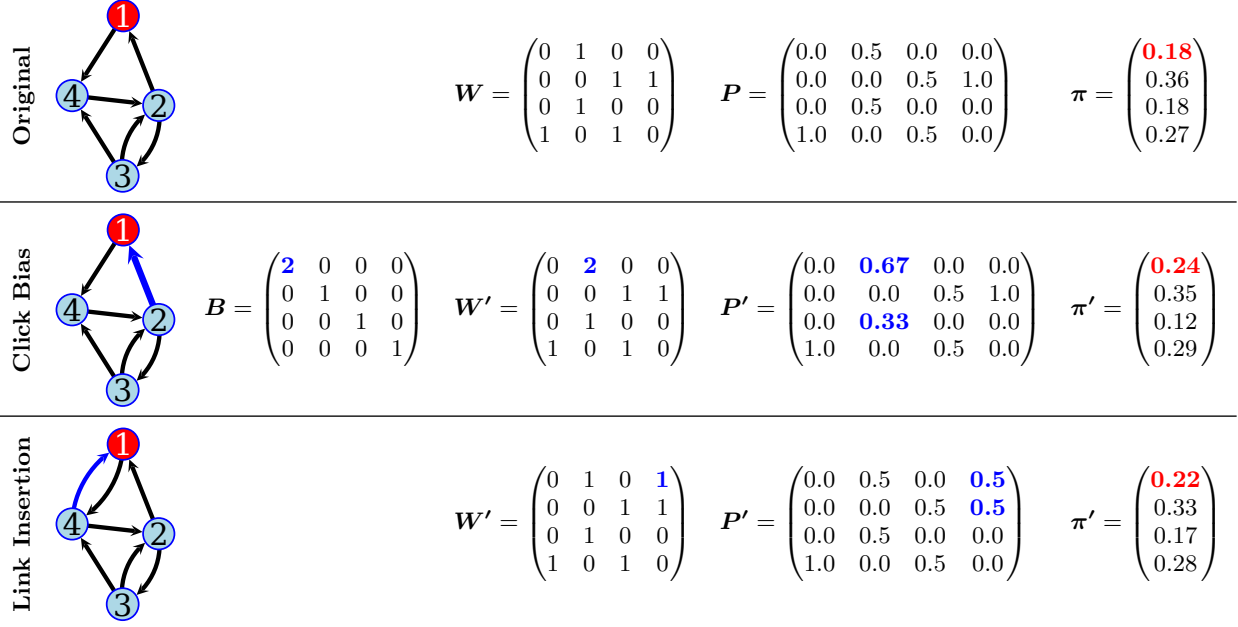
Figure 1: **Modeling Click Bias and Link Insertion - Illustrative Example.** We intend to use different link modification strategies to steer the random surfer towards the red colored node 1 more often. Hence, our target nodes vector consists of only one node: node 1 ($\boldsymbol{t} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}}$). In each row we visualize the corresponding network of the website, where nodes represent webpages and links represent hyperlinks between the webpages. Further, for each of them we show how we calculate its stationary distribution. This involves (from left to right) the weighted adjacency matrix $\boldsymbol{W}$ (unmodified network) or $\boldsymbol{W}'$ (modified networks), the corresponding transition matrix $\boldsymbol{P}$ (unmodified network) or $\boldsymbol{P}'$ (modified networks) and finally the corresponding stationary distribution $\boldsymbol{\pi}$ (unmodified network) or $\boldsymbol{\pi}'$ (modified networks). The blue links in the graphs and the blue matrix elements in bold show the link modifications and their effects on the adjacency and the transition matrix. The red vector elements show the effects of the modifications on the stationary probability (energy) of node 1. **Top row.** Here we depict the original and unmodified network. **Middle row.** We modify the network with a *click bias*. We double the statistical weights of links towards target nodes (bias strength $b = 2$). To calculate the modified adjacency matrix we first construct the diagonal bias matrix $\boldsymbol{B}$ and then compute $\boldsymbol{W}' = \boldsymbol{BW}$. We see an increase in energy of node 1 from 0.18 in the unmodified network to 0.24. **Bottom row.** We insert a new link from node 4 to 1 (i.e., blue link in graph and blue element in $\boldsymbol{W}'$) into the original network. Due to the link insertion the energy of node 1 increases from 0.18 in the unmodified network to 0.22 in the modified network. Thus, in this toy example the effects of the click bias are stronger than those of link insertion. Additionally, we see that also elements in the out-component of node 1 (i.e., node 4) profit of an increased energy of node 1 since a significant amount of 1's increased energy flows into node 4.

is itself designated as a source node we do not insert self-loops—from the practical point of view, it does not make sense to link a webpage to itself. In the rare case where we have connected all possible combinations of source and target nodes but did not reach the required number of links, we simply reiterate the list of the source nodes resulting in parallel links between nodes. Please note that we insert parallel links if a link between a source and a target node has already existed in the original network. However, this happens extremely rarely because all of our networks are sparse. In fact, in all our experiments the fraction of inserted parallel links was on average less than 1%.

3.5 Combinations

Finally, we can combine the two link modification strategies and study the effects of such combinations on the stationary distribution and investigate if an optimal combination of strategies exists, which outperforms the individual approaches. From the practical point of view this means that for optimally steering website users, we combine both, the click bias and link insertion mechanisms.

To create a combined link modification method we first introduce $\alpha \in [0,1]$, which we call the mixing factor. The mixing factor determines how many of the $l(b)$ links are inserted by the click bias. Then, $1-\alpha$ defines how many links are inserted by the link insertion strategy:

$$l(b) = \underbrace{\alpha \cdot l(b)}_{\text{\# biased links}} + \underbrace{(1-\alpha) \cdot l(b)}_{\text{\# inserted links}} \qquad (5)$$

With a combined strategy we cannot bias all links towards target nodes—again, we need to select a subset of links towards target nodes. In analogy to the link insertion method we again preferably select links between nodes having higher stationary probability in the unmodified network. Thus, we first compute the probability distribution over the eligible links in the form of matrix $\boldsymbol{L}$, where $\sum_{ij} L_{ij} = 1$. We define matrix $\boldsymbol{L}$ as:

$$\boldsymbol{L} = \text{diag}(\boldsymbol{\pi}) \cdot \text{diag}(\boldsymbol{t}) \cdot \boldsymbol{W} \cdot \text{diag}(\boldsymbol{\pi}). \qquad (6)$$

The probability of selecting a link is directly proportional to the product of the unmodified stationary probability of its source and target node. Note that due to the multiplicative factor $\text{diag}(\boldsymbol{t}) \cdot \boldsymbol{W}$ only links towards target nodes have a

non-zero probability. With L in place we sample $\alpha \cdot l(b)$ links without replacement and multiply their value in W' by b to induce the click bias. To insert the remaining $(1 - \alpha) \cdot l(b)$ links we adopt the link insertion strategy on the matrix W' as described previously.

3.6 Measuring the Effects

To measure the effects of link modification strategies we quantify how the stationary probabilities of given target nodes change as a function of the modification. In the remainder of this paper we will refer to a node's stationary probability using the, in the literature established, term **energy** [3]. To that end, we calculate the *energy of target nodes* (π'_t), which is the sum of the modified stationary probabilities of target nodes, as following:

$$\pi'_t = \sum_i \pi'_i \cdot t_i \,, \qquad (7)$$

where π' is the stationary distribution of the modified adjacency matrix.

We further measure the *influence potential*, which is the relative increase in the energy of target nodes due to the modification, as a factor τ:

$$\tau = \frac{\pi'_t}{\pi_t}, \qquad (8)$$

where π_t is the energy of target nodes of the unmodified network (i.e., $\pi_t = \sum_i \pi_i \cdot t_i$).

4. DATASETS

For our experiments we use three datasets: an online encyclopedia Wikipedia for Schools[2] (*W4S*) and two online media libraries ORF TVthek[3] (*ORF*) and Das Erste Mediathek[4] (*DEM*).

We collected the data by crawling the corresponding websites. Starting from the main page of a website we recursively crawled all subpages by following all outgoing links from a given webpage. Note that we did not follow external links, meaning that we skipped links to pages not belonging to a given website. Further, we did not follow links generated via Flash, AJAX or any other client-rendered content.

After collecting the data, we removed self-loops, which are links from a webpage to itself, and special links such as "log-in", "write a review", and all other links that require a session-id. In the next step, we represented each dataset as a directed network—webpages are represented as nodes connected by directed links. For calculating the stationary distribution, we extracted the largest strongly connected component (SCC) of each network, so that in the final network it is possible to navigate from any given node to any other node in the network. These final networks have $4,051$ nodes and $111,795$ links (*W4S*), $9,799$ nodes and $301,844$ links (*ORF*), and $70,063$ nodes and $3,448,513$ links (*DEM*).

5. EXPERIMENTAL SETUP

To investigate the effects of manipulating links we first generate sets of target nodes. For this purpose we draw the desired number of nodes uniformly at random from the network without replacement, creating a synthetic set of nodes

of a specified size. Note that those sets can consist of unconnected webpages. We conduct all of our experiments with the same initially generated target nodes to reduce the influence of the random node selection process. For making the number of webpages selected as target nodes comparable between datasets we refer to the size of target nodes as ϕ, which is the fraction of target nodes. To generate target nodes we use several values for ϕ which range from 0.01 to 0.2. For each dataset and each ϕ we generate 100 different synthetic sets of nodes (i.e., target nodes).

Limiting (High) Bias Behavior. In our first experiment we are interested in analyzing the impact of an increasing bias strength on the energy of target nodes using either a click bias on already existing links or inserting new links in an informed way. We use bias strengths reaching from $b = 2$ to $b = 200$ to investigate their effects. Note that for the link insertion strategy the number of inserted links is defined by the bias strength b using Equation 4. This ensures a fair comparison between the two methods.

Realistic (Lower) Bias Strengths. In this experiment we investigate practically relevant [17,20] values for the bias strength b. In particular, we iterate over the range 2 to 15 as bias strengths. With this experiment we gain insights into the effects of the proposed modifications, which can be implemented in websites. After the modification of the adjacency matrix we measure the energy of target nodes π'_t. This allows us to investigate the efficiency of both methods for a given bias strength.

Relative Increase in Stationary Probability. With the previous experiments we analyze changes in the energy of target nodes in absolute terms. For instance, we may learn that for a given set of target nodes we may achieve an energy of $\pi'_t = 0.5$. However, we do not know what the relative increase in their energy is. For example, the set of target nodes may have had $\pi_t = 0.49$ in the unmodified network rendering our efforts futile in *relative* terms. Thus, in this experiment we use τ to measure the influence potential. A higher value for τ means a larger *relative* increase in the energy of target nodes. Again, we compare the results for a given bias strength between our two methods.

Combination of Strategies. Finally, we are interested in investigating if and to what extent the energy of target nodes changes if we combine click biases and link insertion. We vary the mixture factor α from 0 to 1 in steps of 0.1 and measure the energy of target nodes π'_t of the modified networks.

6. RESULTS & DISCUSSION

6.1 Saturation

Figure 2 depicts the effects of link modifications in our datasets with increasing values of bias strength b and varying fractions of target nodes ϕ (0.01, 0.1 and 0.2).

In the case of click bias we observe the following situation. For small values of b the energy of target nodes π'_t increases very quickly (navigational boost phase, which we analyze in more detail in Section 6.2)—this energy saturates for larger values of b (i.e., $b > 35$). This holds for larger ϕ values (0.1 and 0.2), whereas for a smaller ϕ, for example $\phi = 0.01$, the initial growth as well as the saturation are significantly slower and lower respectively. Further, for higher ϕ (0.1 and 0.2) π'_t saturates at an almost identical and very high level
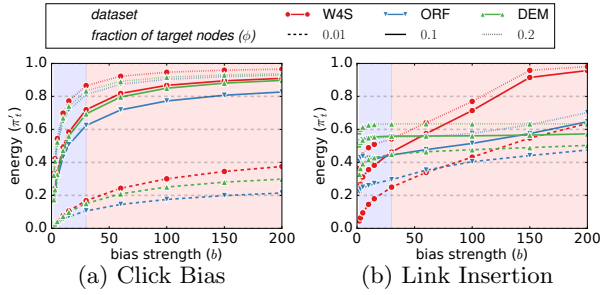
Figure 2: **Saturation.** The plots depict the connection between bias strength (x-axis) and the increased energy of target nodes due to an induced click bias (left) or link insertion (right). Each marker type and color refers to one dataset. Dashed, solid and dotted line styles refer to fraction of target nodes ϕ 0.01, 0.1 and 0.2 respectively. We can observe that both link modification strategies reach a certain level of saturation—meaning that further increases in bias strength do not result in an increase in energy of target nodes. Therefore, for both strategies we identify two phases: a (i) *navigational boost* phase in which we observe a rapid increase of the stationary probability (blueish region with small values of the bias strength), and a (ii) *saturation* phase (reddish region with larger values of the bias strength).

(>0.8)—if the click bias is strong enough we can increase the energy of any fraction of target nodes larger than $\phi > 0.1$.

An interesting question in this respect is the height of the energy saturation level. Theoretically, this level is close to 1.0 but as Figure 2 shows, in empirical networks this level can not be fully reached. Essentially, due to the directed nature of the network, the target nodes out-component (i.e., the nodes with incoming links from target nodes) will always act as a drain that will take some energy from the target nodes. That amount depends on the size of the out-component as well as its connectivity with other parts of the network—in particular the existence of back-links towards target nodes. This situation is depicted in our toy example Figure 1 in the middle row. Node 4, which has an incoming link from node 1, profits from an induced click bias towards node 1 (cf. original $\pi_4 = 0.27$ and modified $\pi_4' = 0.29$). Thus, although π_1' increases with increasing bias strength, node 1 would never reach energy values close to 1.0 because node 4 attracts a certain amount energy to itself.

In the case of the link insertion strategy the results are more diverse (cf. Figure 2b). For DEM dataset we observe a quick saturation for all values of ϕ. Differently from the click bias the saturation level is significantly lower for this dataset (i.e., 0.6). For the ORF dataset we do not observe saturation but a monotonous increase in the energy of target nodes for increasing values of ϕ. Finally, for the W4S dataset and larger ϕ (0.1 and 0.2) we can observe saturation at levels higher than 0.9.

As previously, the size of the out-component of the target nodes, combined with the size of their in-component (i.e., the source nodes which point towards target nodes), as well as the ratio of these two quantities provide a possible explanation for this behavior. Basically, we can calculate the average number of newly inserted links as $l(b) = \overline{d} \cdot n \cdot \phi \cdot b$, where $\overline{d}$ is the average degree (i.e., in a directed network average degree $\overline{d}$ corresponds to both the average in-degree as well as average out-degree) and n, ϕ, b are as before. Thus, in the networks with a higher average degree we in-

sert more new links. For smaller values of bias strength (blueish region in Figure 2b) these new links lead to a navigational boost, resulting in a quick increase in the energy π_t' of target nodes. The navigational boost is higher in networks with a higher average degree—we observe the highest increase in π_t' in DEM with $\overline{d} = 49.22$, the second highest in ORF with $\overline{d} = 30.8$, and the lowest in W4S with $\overline{d} = 27.6$. As mentioned before, in Section 6.2 we analyze this navigational boost in more detail. However, for larger values of bias strength (reddish region in Figure 2b) the effects of the drain due to the larger size of the out-component become visible—the networks with a higher increase for smaller bias strengths lose their energy now more quickly. Thus, the ordering of the saturation levels for higher bias strengths is reversed to the navigational boost in energy for lower bias strengths, resulting in W4S to now have the highest saturation level, followed by ORF and then by DEM.

To confirm our intuition about the saturation for the link insertion strategy we performed the following analysis. First, we calculated some structural properties for the target nodes. In particular, based on the insights of Ding et al. [7,8], we define the *in-degree* of target nodes as the sum of the weights of links pointing towards target nodes $d_t^- = \sum_{ij} (\text{diag}(\boldsymbol{t}) \cdot \boldsymbol{W})_{ij}$. The *out-degree* of target nodes is the sum of the weights of outgoing links of target nodes $d_t^+ = \sum_{ij} (\boldsymbol{W} \cdot \text{diag}(\boldsymbol{t}))_{ij}$. Finally, the *degree ratio* of target nodes is a ratio between the previous two measurements (i.e., $d_t^r = d_t^+ / d_t^-$). Although, it has been shown that properties, such as the simple count of in-links of a node, are bad approximations for PageRank on a large scale [27], they proved to be a good indicator for the random surfer behavior on our datasets.

In our experiments, DEM has on average by one order of magnitude higher both target node in-degree and out-degree than the other two datasets. This explains a quick increase of π_t' for smaller bias strengths. However, degree ratio is typically larger in DEM target nodes than in ORF or W4S target nodes and this explains a higher drain of energy and a lower saturation level in the DEM dataset (cf. Figure 3).

Finding. For larger fractions ϕ of target nodes their energy π_t' achieved through a click bias quickly saturates across all datasets at very high levels (> 0.8). Boost and saturation of the energy is significantly slower for smaller fractions ϕ. The saturation level is determined by the out-degree of the target nodes and reciprocity of outgoing links from the target nodes. For link insertion saturation existence, speed, and levels vary between datasets and ϕ values. The average degree of the original networks as well as the ratio between out-degree and in-degree of target nodes significantly influences those effects.

Implications. In case of medium ($\phi = 0.1$) and large ($\phi = 0.2$) fractions of target nodes we reach high saturation levels with both link modification methods even with small bias strengths. For example, if we would like to increase visibility of a large category in, for example Wikipedia, we can achieve this by either slightly increasing the font size of the links towards the articles of that category or by simple creating some new links towards those articles. Click bias reaches very high visibility levels consistently across several different datasets, whereas link insertion is dependent on the network structure—in datasets with a smaller average number of links we can achieve larger changes. This follows our intuition—in a network with a smaller number of links, each new link affects the network more significantly. How-
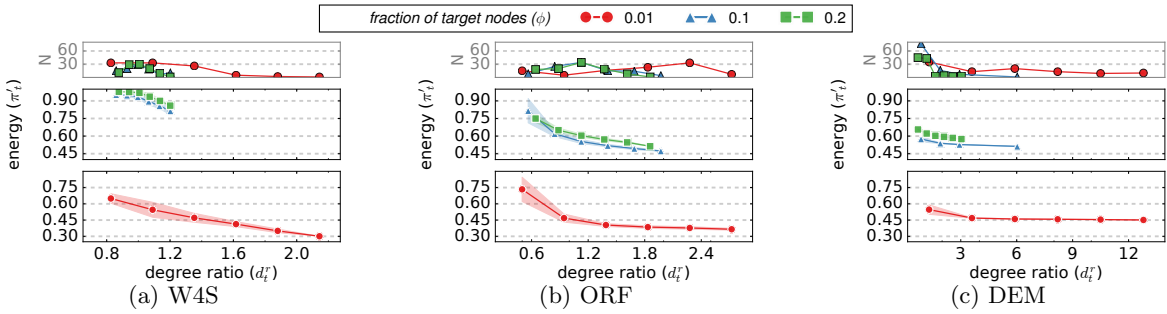
Figure 3: **Influence of Target Nodes Degree Ratio onto the Saturation of their Energy.** The plots depict the due to link insertion achieved energy of target nodes pi'_t as the function of their degree ratio. Each line depicts the results for a given ϕ. For increased readability we group data points into six equally sized bins according to their degree ratio (x-axes). Values on the y-axes represent the averages of the data points falling into the corresponding bin. **Top row.** In the top row we show the distribution of the target nodes degree ratios (in bins; x-axes)—the y-axes denote the number of data points (N) falling into each bin. **Middle row.** Here, we depict the results for medium and large fractions of target nodes. **Bottom row.** For readability we depict small fractions of target nodes separately. Over all datasets and all ϕ we consistently observe a negative correlation between degree ratio and energy π'_t. This means that with an increasing ratio—an increasing out-degree and a decreasing in-degree—the drain of energy increases and this leads to the saturation of the energy of target nodes.

ever, to match the effects of the click bias we need to insert a very large amount of new links. On the other hand, in case of small ($\phi = 0.01$) fractions of target nodes, we can achieve larger changes by using link insertion—we are able to reach higher saturation levels more consistently and more quickly regardless of the dataset. Again, we can explain this intuitively—small fractions of target nodes have, on average, only few links pointing towards them. Hence, inserting a new link from a top webpage achieves larger changes than highlighting an existing (and probably negligible) link.

6.2 Navigational Boost

The blueish region from Figure 2b corresponds to smaller and more realistic bias strengths. In practice, increasing the visibility of a link (e.g., by repositioning or highlighting) by more than a factor of 15, meaning that it would receive 15 times more clicks than before, seems quite unrealistic. In particular, users position bias is estimated to be lower than 3.5 [17,20]. Hence, we focus on bias strengths ranging from 2 to 15 (the blueish region in Figure 2b) where we can observe a phase of quick increase in the energy of target nodes. We call this phase *navigational boost* phase. The results for all bias strengths from 2 to 15 are quite similar and therefore we report only the results for bias strength $b = 5$.

For click bias we observe a robust performance across datasets, see Figure 4. The energy of target nodes π'_t increases almost linearly with the fraction ϕ of target nodes. However, at higher ϕ (i.e., $0.15 \le \phi \le 0.2$) the linear trend tends to flatten. This is due to a transition to the stationary phase (cf. Section 6.1). Further, we observe a rather high variance of π'_t over ϕ and different sets of target nodes. For example, we measure the following average standard deviations over ϕ: W4S = 0.023, ORF = 0.103 and DEM = 0.068. This high variance can be attributed to situations in which smaller fractions of target nodes are often able to outperform larger ones. We depict one such extreme situation of two outlier samples marked as A and B in Figure 4. Target nodes depicted with A with $\phi = 0.1$ reach an energy that is almost twice as high as those of the target nodes depicted with B with $\phi = 0.2$.

One potential explanation for these observations is that if the energy of target nodes of the unmodified network is already quite high, that is, the target nodes include one or more nodes with a substantial energy, then the click bias acts as an *amplifier* further magnifying the energy of target nodes. On the other hand, target nodes with a small unmodified energy receive indeed the amplifying effect but are never able to reach the same (high) levels of the modified energy. Therefore, it is possible for smaller fractions of target nodes with one or more nodes with high starting energy to outperform larger fractions of target without such nodes. This can be further attributed to the target nodes structural properties, such as out-degree, in-degree and degree ratio, which we introduced in the previous sections. Basically, starting energy positively correlates with in-degree of target nodes, and therefore we can expect that the click bias is able to amplify target nodes with a higher in-degree more than the target nodes with a lower in-degree. In particular, to confirm this finding we conducted a similar correlation experiment as depicted in Figure 3, but used a combination of the target nodes in-degree and energy achieved due to a click bias. However, due to limitations in space, we do not report the experimental details here.

Finding. The fraction ϕ of target nodes does not have a decisive effect on navigational boost. Often, smaller ϕ exhibit larger effect sizes. Click bias acts as an *amplifier* that only magnifies what is already present in the target nodes.

In the case of link insertion, navigational boost appears to be highly dataset dependent (see Figure 4b). However, the variance of each dataset individually is very low with average standard deviations of 0.017 for W4S, 0.029 for ORF and 0.034 for DEM. Across all datasets we can observe a quick increase in the energy of target nodes with an increasing fraction of target nodes, which then experience a transition towards a stable saturation phase.

To explain the difference in performance between different datasets we have plotted the Lorenz curves of the stationary distributions of our datasets (see Figure 4c). We see that for W4S, a very small fraction of top nodes (0.01) only possesses 0.4 of energy. Diversely, for ORF and DEM the same fraction of top nodes already possesses energy higher than 0.85. As the out-component of a specific set of nodes acts as a drain for the energy of source nodes, connecting source
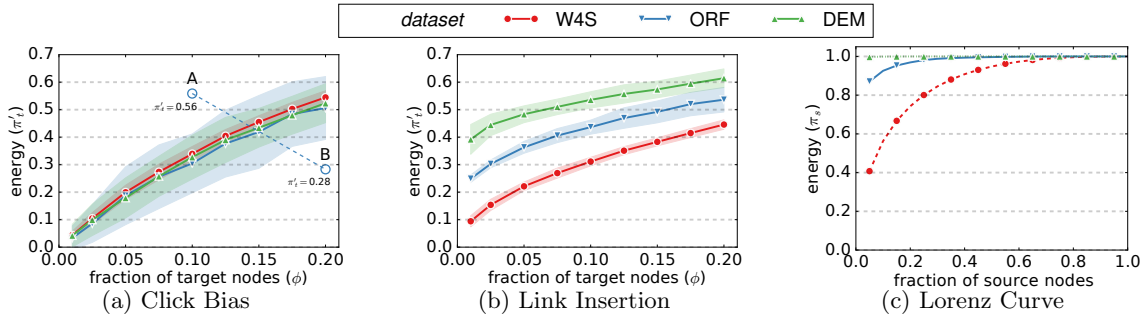
Figure 4: **Navigational Boost.** Left center figures depict the energy of target nodes after modifying the network through either inducing a click bias or link insertion respectively. The x-axes correspond to the fraction ϕ of target nodes, whereas on the y-axes we denote the energy of target nodes π_t'. Each line represents the average of 100 samples for each ϕ of one dataset. The areas filled with the same color denote the corresponding standard deviations. **Left.** Inducing a click bias is robust across datasets. However, the variability within values of ϕ is rather high. The high variance is caused by the presence or absence of one or more nodes with high original energy in the target nodes. Thus, in cases where such nodes are present in target nodes (depicted as point A in the plot) even smaller fractions of target nodes are able to outperform larger fractions of target nodes without such a top node (depicted as point B in the plot). **Center.** On the contrary, the performance of link insertion varies over datasets but is stable across various ϕ values, which is signified by the low standard deviation suggests over ϕ. **Right.** We plot the Lorenz curve of the datasets' original stationary distributions. We can observe that for different datasets these distributions are differently skewed. In particular, in DEM the energy of just a few nodes is close to 1, whereas in ORF and W4S we need far more nodes to reach the same level (i.e., 0.4 and 0.7 respectively). This explains why we can achieve the highest effect with link insertion in DEM, followed by ORF and W4S. Thus, the performance of link insertion depends on the initial stationary distribution of the network, whereas click biases are robust across datasets. Moreover, for smaller fractions of target nodes link insertion constantly outperforms click biases.

nodes with high energy to target nodes leads to a flow of energy from those source nodes towards target nodes. Thus, the initial energy of source nodes plays a crucial role in this process. Through link insertion from top source nodes towards target nodes we attach the target nodes as drains to such top nodes. Consequently, target nodes receive a huge amount of energy and experience a large navigational boost (i.e., ORF and DEM). In other words, we can say that link insertion induces *diffusion* of the energy of top nodes towards target nodes. Given the average degree of the network and the fraction of source nodes (which increases with the fraction ϕ of target nodes), we can use the Lorenz curves to approximately predict the point where the performance across datasets becomes similar. For example, the Lorenz curves of DEM and ORF meet around a fraction of 0.4 of source nodes and we can expect that the performance of those two datasets will become similar for all fractions of source nodes larger than 0.4. In the case of W4S, we need a larger fraction of source nodes (0.7) to reach a similar behavior (cf. Figure 4c).

Comparing link insertion with click bias we find that the former outperforms the latter for smaller fractions ϕ of target nodes. For example, in the DEM dataset, link insertion reaches four times higher energy values for the target nodes with $\phi = 0.01$. However, for higher values of ϕ the click bias exhibits a similar performance as link insertion. Further, in the case of the W4S dataset, the click bias even outperforms link insertion (see W4S in Figure 4a and Figure 4b) at $\phi = 0.2$.

Finding. The performance of link insertion varies across the datasets and depends on the skewness of the initial stationary distribution in a dataset. Inserting links from other important webpages towards a given set of webpages results in a higher navigational boost than with the click bias. This is due to the induced *diffusion* of the energy from top nodes towards target nodes.

Implications. If it is possible to insert new links on a website (especially if the fraction of target nodes is small) we should prefer the link insertion over the click bias. However, creation and insertion of such links may be problematic in practice. For example, on Wikipedia it may be difficult and semantically unjustified to insert new links to completely unrelated articles since this may have opposite and contrasting effects on the navigational behavior of users, such as confusion and dissatisfaction. In those cases we may rather choose to increase the transition probability of an already existing link by, for example, highlighting that link (i.e., using CSS[5]) or repositioning it to the webpage's top area. In some other scenarios (i.e., birthdays of famous inventors) implementing a banner which contains links towards a given set of webpages may be an easy way to insert thousands of new links instantly. In those cases, such user interface modifications may prove to have higher lasting effects on the stationary probability than, for example, highlighting links.

6.3 Influence Potential

Figure 5 depicts the effects of link modifications strategies on the relative increase of the energy of target nodes (i.e., influence potential). Again, we concentrated in this experiment on realistic settings for the bias strength from the interval $[2, 15]$. Since we got comparable results over that complete interval we present only the results for bias strength $b = 5$.

The performance of the click bias is robust across datasets and different ϕ with a low variance in both dimensions (cf. Figure 5a). We observe a negative correlation between influence potential and fraction ϕ of target nodes, meaning the smaller fractions of target nodes profit more from an induced click bias than larger fractions. Our calculations of the influence potential confirm once more the results from the previous section, in which smaller fractions with top en-
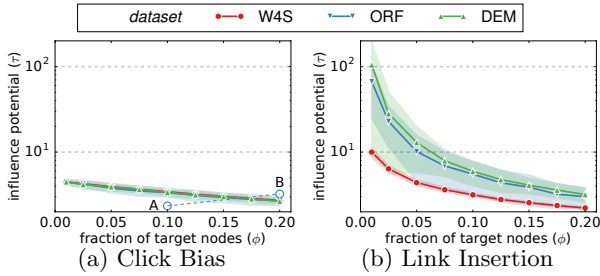
---

[5]cascading stylesheets

Figure 5: **Influence Potential.** The figure depicts the relative increase in the energy of target nodes τ with a fixed bias strength of $b = 5$ over different ϕ and datasets. **Left.** Inducing a *click bias* performs robustly and similarly over all datasets (curves for different datasets overlap each other in the plot). Influence potential correlates negatively with the fraction of target nodes, that is, the relative increase in energy is higher for small fractions of target nodes than for large fractions. **Right.** With *link insertion*, we find a significant variance in performance across our datasets. This confirms our findings from the previous section—the skewness of the original stationary distribution determines the effectiveness of the link insertion strategy in a dataset. Similarly to the click bias, the influence potential decays with an increasing fraction of target nodes.

ergy nodes are able to outperform larger fractions of target nodes without top nodes. We once more depict two such examples from Figure 4a. Target nodes depicted by A with $\phi = 0.1$ reach an energy that is almost twice as high as those depicted by B with $\phi = 0.2$. However, nodes A start with a larger initial energy and nodes B with a smaller one. Therefore, in relative terms nodes B have a higher influence potential than nodes A (cf. Figure 5a).

Performance of link insertion is again strongly dependent of the dataset. However, similarly to the click bias we observe over all datasets that smaller fractions of target nodes profit significantly more from the link insertion than the larger ones. For example, in DEM dataset for $\phi = 0.01$ we measure an average influence potential of more than 100, whereas for $\phi = 0.2$ influence potential is less than 4 (cf. Figure 5b). A similar decay, although not as pronounced as in DEM can be seen in the other two datasets. Similarly to the navigational boost this high influence potential of smaller fractions of target nodes in the case of link insertion can be explained through the skewness of the initial stationary distributions (cf. Figure 4c).

As previously, we investigated more closely the relation between influence potential of small fractions of target nodes and their structural properties such as in-degree, out-degree and degree ratio. Target nodes with a high degree ratio (i.e., a small in-degree, a large out-degree or both) have the largest influence potential. Intuitively, such target nodes start with a very small initial energy and therefore can achieve a significant relative increase. On contrary, in absolute terms such target nodes keep a rather small energy even after the modification, whereas target nodes with a large initial energy (a low degree ratio) are experiencing a significant navigational boost in absolute terms but possess relatively low influence potential.

Finding. The influence potential of small fractions of target nodes is very high regardless of the link modification strategy. For click bias the influence potential is limited by the bias strength, whereas for link insertion we do not observe

such a limit and influence potential can become as high as 100. With increasing fraction of target nodes the influence potential decays drastically.

Implications. As previously, if possible we should prefer link insertion over click bias in cases where we are interested in utilizing the influence potential of the target nodes. Our findings suggest that in practice there is a trade-off that we need to make between optimizing for influence potential and for navigational boost. For the former, we need to aim at target nodes with a high degree ratio and for the latter at target nodes with a low degree ratio.

6.4 Combinations

In the previous experiments we found that in some situations link insertion should be preferred over click bias (e.g., small fraction ϕ of target nodes), whereas sometimes the opposite represents an optimal approach (e.g., large ϕ). For that reason we want now to shed more light onto combinations of both strategies, that is, we are interested in the navigational effects of simultaneously applying click bias and link insertion to varying extent. Figure 6 depicts the results of this experiment. We find consistent best performing mixtures over all datasets. In particular, we observe that for small fractions ϕ of target nodes, exclusive link insertion outperforms any other combination (see Figure 6a). For medium sized target nodes (i.e., $\phi = 0.1$) we observe a shift of best performing combinations towards $\alpha = 0.9$ for higher bias strengths (i.e., $b = 5$ and $b = 15$). This combination consist of 90% click bias and 10% link insertion. For combinations of large fractions of target nodes (i.e., $\phi = 0.2$) and small bias strengths ($b = 2$) the best performing combination is around $\alpha = 0.5$ (50% click bias and 50% link insertion) and further shifts towards $\alpha = 0.9$ (90% click bias and 10% link insertion) with an increased bias strength.

These results confirm our insights from the previous experiments. Thus, click biases act as an amplifier and only work well if target nodes initially possess valuable incoming links. This is highly likely for larger and medium sized fractions of target nodes, and very unlikely for the case of smaller fractions of target nodes. On the other hand, link insertion diffuses a large portion of the energy of top nodes towards target nodes. Hence, it works especially well for combinations of small fractions of target nodes and datasets with a highly skewed stationary distribution.

Finding. For small fractions of target nodes with initially low energy, pure link insertion should be preferred over any other combination. However, with increasing bias strength and larger fraction of target nodes, combinations consisting of 90% click bias and 10% link insertion performs best.

Implications. Smaller sets of webpages (i.e., small ϕ) should focus on introducing new links to achieve the highest browsing guidance. The bigger the set of webpages and the used bias strength becomes, the more this preference shifts towards a combination of 0.9, meaning that 90% of the modifications should be invested in increasing the transition probability of already existing links towards target nodes (e.g., highlighting in the user interface). The remaining 10% should be used to insert new links towards target nodes.

6.5 Stationary vs. Transient User Behavior

The random surfer which navigates forever (stationary behavior) may look like a rather unrealistic behavior of users. More realistically, a single user visits a website clicks a cou-
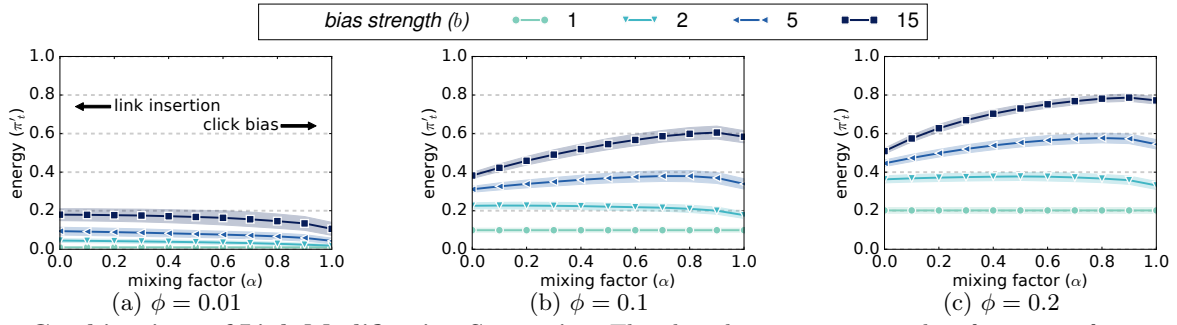
Figure 6: **Combinations of Link Modification Strategies.** The plots depict average results of 100 sets of target nodes of W4S for three ϕ: 0.01 small on the left, 0.1 medium in the middle, and 0.2 large on the right. On the x-axes we denote α, which defines the combination of the two link modification strategies, whereas on the y-axes we denote the energy of target nodes. We see that for smaller values of ϕ $\alpha = 0$ (100% link insertion) outperforms all the others over all used bias strengths. However, for medium and large ϕ values with higher bias strengths this sweet spot shifts towards higher combinations ($\alpha = 0.7$). In the other two datasets we can observe similar results.

ple of times on various links and leaves the website again (transient behavior). However, our calculations of the stationary distribution show that, at least on the networks that we have investigated in this paper these two behaviors are quite similar to each other.

The stationary distribution is calculated with the power-iteration method [11]. Thus, we initialize a probability vector representing an initial probability to find a random surfer on each particular node in the network. We initialize this vector (i) with a uniform distribution and (ii) by setting the visit probability of the home page to 1. The former initialization accounts for the assumption that initially each page is equally likely to be visited by users, whereas the latter models users entering the website over the home page. Afterwards, we iterate by recalculating the probabilities for the next click of the random surfer. Thus, one iteration step of the power-iteration method can be interpreted as a step or a click performed by the random surfer moving from the current node to one of its neighbors. Hence, the number of iteration steps that are needed until there are no significant changes in the node probabilities, that is, the convergence rate of the power-iteration method, can be interpreted as the number of clicks needed to model the stationary user behavior. In other words the random surfer does not need to navigate forever—it only needs to navigate through the network until the point where the next click does not change the observed stationary distribution.

In all our datasets, all networks that we generated and modified for these datasets, all combinations of fractions of target nodes ϕ and the bias strength b our calculations converge within 8 iterations regardless of the initialization. Thus, the stationary user behavior is in fact a behavior of users who navigate 8 pages in a website at most. We believe that these 8 clicks are within realistic boundaries for user behavior in the cases in which users decide to explore and browse a website. However, since many users leave a website immediately upon arrival or within only a single or a small number of clicks this still represents a limitation in our work. This limitation can be easily remedied by introducing a small teleportation probability of jumping to an arbitrary page without following the underlying network structure (i.e., calculating PageRank vector instead of the stationary distribution). We have already experimented with the calculations of PageRank and our first results are quite similar to results that we have presented in this paper.

However, we plan to address this question in more details in our future work.

7. CONCLUSIONS

In this paper we have analyzed the effects of two link modification strategies used to influence the typical whereabouts of the random surfer. We investigated how an induced click bias towards a set of webpages changes the stationary distribution (i.e., energy) of those pages. Additionally, we compared those effects with the consequences of altering the network structure by inserting new links. We find that both strategies have a high potential to modify the stationary distribution and that for certain situations there exist constantly high performing link modification strategy. In particular, click biases work well on sets of webpages containing already highly visible webpages, whereas link insertion should be preferred for sets of webpages consisting of pages with low visibility. Further, we showed that a simple structural property of target nodes, namely degree ratio, provides a valuable basis for the estimation of the effects of both link modification strategies.

Assuming that the random surfer is a realistic model of user behavior on the Web—which previous studies seem to confirm [9,15]—website operators can use our approach and open source framework to determine the best strategy for their settings without having to implement and test all the different strategies. Such strategies include but are not limited to altering link positions [20] (bias) or creating new links using a recommender systems [16] (link insertion).

An important practical issue that we have not addressed in this paper is usability. Usability considerations limit the number of new links that we can insert or how we can reposition links. In future, we plan to account for usability by extending our model and investigating limitations induced by various usability restrictions. Also, including the existing user link selection bias derived from user clickstreams into the model would further improve the practical relevance of our method.

Acknowledgments

8. REFERENCES

[1] S. Al-Saffar and G. Heileman. Experimental bounds on the usefulness of personalized and topic-sensitive pagerank. In *Web Intelligence, ACM International Conference on*, pages 671–675. IEEE, 2007.

[2] L. Bian and H. Holtzman. Online friend recommendation through personality matching and collaborative filtering. *Proc. of UBICOMM*, pages 230–235, 2011.

[3] M. Bianchini, M. Gori, and F. Scarselli. Inside pagerank. *ACM Trans. Internet Technol.*, 5(1):92–128, Feb. 2005.

[4] S. Brin and L. Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833, 2012.

[5] G. Buscher, E. Cutrell, and M. R. Morris. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 21–30. ACM, 2009.

[6] D. Dimitrov, P. Singer, F. Lemmerich, and M. Strohmaier. Visual Positions of Links and Clicks on Wikipedia. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16 Companion, New York, NY, USA, 2016. ACM.

[7] C. Ding, X. He, P. Husbands, H. Zha, and H. D. Simon. Pagerank, hits and a unified framework for link analysis. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 353–354. ACM, 2002.

[8] C. H. Ding, H. Zha, X. He, P. Husbands, and H. D. Simon. Link analysis: hubs and authorities on the world wide web. *SIAM review*, 46(2):256–268, 2004.

[9] F. Geigl, D. Lamprecht, R. Hofmann-Wellenhof, S. Walk, M. Strohmaier, and D. Helic. Random surfers on a web encyclopedia. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, i-KNOW '15, pages 5:1–5:8, New York, NY, USA, 2015. ACM.

[10] D. F. Gleich, P. G. Constantine, A. D. Flaxman, and A. Gunawardana. Tracking the random surfer: empirically measured teleportation parameters in pagerank. In *Proceedings of the 19th international conference on World wide web*, pages 381–390. ACM, 2010.

[11] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

[12] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with trustrank. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 576–587. VLDB Endowment, 2004.

[13] T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526. ACM, 2002.

[14] T. H. Haveliwala. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *Knowledge and Data Engineering, IEEE Transactions on*, 15(4):784–796, 2003.

[15] D. Helic, M. Strohmaier, M. Granitzer, and R. Scherer. Models of human navigation in information networks based on decentralized search. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 89–98. ACM, 2013.

[16] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5–53, 2004.

[17] T. Hogg and K. Lerman. Disentagling the effects of social signals. *Human Computation Journal*, 2(2):189–208, 2015.

[18] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[19] A. N. Langville and C. D. Meyer. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2004.

[20] K. Lerman and T. Hogg. Leveraging position bias to improve peer recommendation. *PLoS ONE*, 9(6):e98914, 06 2014.

[21] N. Li and G. Chen. Multi-layered friendship modeling for location-based mobile social networks. In *MobiQuitous, 2009. 6th Annual International*, pages 1–10, 2009.

[22] L. Lovász. Random walks on graphs: A survey. *Combinatorics, Paul erdos is eighty*, 2(1):1–46, 1993.

[23] M. Moricz, Y. Dosbayev, and M. Berlyant. Pymk: Friend recommendation at myspace. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, SIGMOD '10, pages 999–1002, New York, NY, USA, 2010. ACM.

[24] J. Murphy, C. Hofacker, and R. Mizerski. Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2):522–535, 2006.

[25] U. Y. Ogras and R. Marculescu. " it's a small world after all": Noc performance optimization via long-range link insertion. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 14(7):693–706, 2006.

[26] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[27] G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. In *Computing and Combinatorics*, pages 330–339. Springer, 2002.

[28] M. Richardson and P. Domingos. The intelligent surfer: Probabilistic combination of link and content information in pagerank. In *NIPS*, pages 1441–1448, 2001.

[29] N. Silva, I.-R. Tsang, G. Cavalcanti, and I.-J. Tsang. A graph-based friend recommendation system using genetic algorithm. In *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pages 1–7, 2010.

[30] R. West and J. Leskovec. Automatic versus human navigation in information networks. In *ICWSM*, 2012.

[31] W. Woess. Random walks on infinite graphs and groups-a survey on selected topics. *Bulletin of the London Mathematical Society*, 26(1):1–60, 1994.

[32] X. Xie. Potential friend recommendation in online social network. In *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on Int'l Conference on Cyber, Physical and Social Computing (CPSCom)*, pages 831–835, 2010.