# Dealing with Dead Ends: Efficient Routing in Darknets

STEFANIE ROOS and THORSTEN STRUFE, TU Dresden

Darknets, membership-concealing peer-to-peer networks, suffer from high message delivery delays due to insufficient routing strategies. They form topologies restricted to a subgraph of the social network of their users by limiting connections to peers with a mutual trust relationship in real life. Whereas centralized, highly successful social networking services entail a privacy loss of their users, Darknets at higher performance represent an optimal private and censorship-resistant communication substrate for social applications.

Decentralized routing so far has been analyzed under the assumption that the network resembles a perfect lattice structure. Freenet, currently the only widely used Darknet, attempts to approximate this structure by embedding the social graph into a metric space. Considering the resulting distortion, the common greedy routing algorithm is adapted to account for local optima. Yet the impact of the adaptation has not been adequately analyzed.

We thus suggest a model integrating inaccuracies in the embedding. In the context of this model, we show that the Freenet routing algorithm cannot achieve polylog performance. Consequently, we design *NextBestOnce*, a provable poylog algorithm based only on information about neighbors. Furthermore, we show that the routing length of *NextBestOnce* is further decreased by more than a constant factor if neighbor-of-neighbor information is included in the decision process.

Categories and Subject Descriptors: C.2.2 [**Computer-Communication Networks**]: Network Protocols—*Routing Protocols*; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Routing and Layout*; C.4 [**Performance of Systems**]: *Modeling Techniques*; G.3 [**Probability and Statistics**]: *Stochastic Processes*

General Terms: Algorithms, Performance, Security, Theory

Additional Key Words and Phrases: Darknets, censorship resilience, routing algorithm

**ACM Reference Format:**
Stefanie Roos and Thorsten Strufe. 2016. Dealing with dead ends - Efficient routing in darknets. ACM Trans. Model. Perform. Eval. Comput. Syst. 1, 1, Article 4 (February 2016), 30 pages.
DOI: http://dx.doi.org/10.1145/2809779

## 1. INTRODUCTION

Centralized communication platforms, such as online social networking (OSN) services, are partial to giving away data due to economical or political pressure. Encryption—if permitted by the service provider—does not provide protection against tracking and tracing of online habits. Distributed services such as Diaspora[1] or peer-to-peer OSNs [Buchegger et al. 2009; Cutillo et al. 2009] avoid data collection at one central point, but arbitrary participants can track others. Darknets by design protect users from being tracked by a foreign party, be it a governmental or commercial institution or a

[1] http://www.joindiaspora.com.

This work was partly supported by the Deutsche Forschungsgemeinschaft (DFG), grant STR 1131/2-1.

Authors' addresses: S. Roos and T. Strufe, TU Dresden, Fakultät Informatik, Nöthnitzerstr. 46, 01187 Dresden, Germany; email: {firstname.lastname}@tu-dresden.de.

DOI: http://dx.doi.org/10.1145/2809779

curious individual. Devices of users, henceforth called *nodes*, only establish connections if their owners share a mutual trust relationship in the real world. In this manner, Darknets provide anonymity; that is, a node should not be linkable to its actions such as requesting certain content, and membership concealment against untrusted parties, that is, the presence of a node in the network, should remain unknown to anyone but its trusted contacts. Whereas currently deployed Darknets such as Freenet [Clarke et al. 2010] and GnuNet [Evans and Grothoff 2011] focus on content storage and retrieval, Darknets present a suitable communication substrate for all social applications such as chat or email.

More precisely, Darknets are commonly characterized by three main criteria: (1) *restricted topology with only local knowledge*: the overlay topology corresponds to the real-world trust graph and each node is only aware of its immediate neighbors, (2) *hop-by-hop anonymization*: queries are forwarded by trusted links only and each contacted node resets the source tag to itself to obfuscate the original source, and (3) *use of end-to-end cryptography, steganography, and anonymization*. Anonymity is hence achieved by the rewriting of the source tag combined with suitable anonymization techniques such as mixing [Chaum 1981]. Membership concealment is achieved by restricting a node's direct communication of the network to its trusted contacts and applying steganographic techniques such as Moghaddam et al. [2012], which obfuscates Tor traffic within Skype by adjusting the pattern of the traffic, to hide the typical traffic patterns of the Darknet from a passive global adversary.

Locating objects, be it data items or individual users, in a Darknet is extremely challenging. Conventional approaches such as DHTs can only be established in a Darknet by adding an additional layer. The goal is to nevertheless achieve acceptable latency while at the same time providing membership concealment and anonymity. Asymptotically, the latency is directly proportional to the routing length, that is, the number of edges on the route between the source and destination. Hence, for maintaining acceptable latencies in a large-scale Darknet, the expected routing length is required to increase only slightly with the network. In this article, we thus aim to provide a routing algorithm for which the expected routing length scales at most polylog with the network size.

Freenet, the most widely known Darknet, assigns node identifiers in order to approximate a ring topology with additional links to allow for polylog routing [Clarke et al. 2010]. However, polylog routing has only been proven under the assumption of a greedy embedding; that is, each nonterminal node on the routing path has a neighbor closer to the destination. Freenet cannot achieve a greedy embedding, so it remains unclear if the routing is indeed efficient.

In this article, we suggest a novel small-world model that allows analyzing routing in nongreedy embeddings. We model the accuracy of the embedding, that is, its closeness to a connected lattice, by a parameter $C$. Each node is required to have at least one neighbor $C$ in each direction. The standard model by Kleinberg [2000] then corresponds to the case $C = 1$. The model accounts for the ability of Darknet embeddings to reflect topological closeness, albeit in a nongreedy fashion. We then show that the Freenet algorithm *$D^2$-DFS* cannot provide polylog routing length. We suggest *NextBestOnce* to overcome the discovered weaknesses. For a scale-free degree distribution with exponent $\alpha$, *NextBestOnce* has an expected routing length of $\mathcal{O}(\log^{\alpha-1} n \log\log n + C^3 \log n)$. Furthermore, we compare *NextBestOnce* to *NextBestOnce-NoN*, which considers neighbor-of-neighbor (*NoN*) information for the routing rather than only information about the direct neighbors. We find that the asymptotic upper bound on *NextBestOnce-NoN* is strictly lower than the lower bound on *NextBestOnce*. So, the expected routing length is improved by considering neighbors of neighbors, though the average degree is constant.

We start by introducing related work on analyzing and designing routing algorithms in Section 2. In Section 3, we present our model, algorithms, and notation. In Sections 4, 5, and 6, we give performance bounds for *$D^2$-DFS*, *NextBestOnce*, and *NextBestOnce-NoN*, respectively, before concluding in Section 7.

## 2. RELATED WORK

In this section, we describe the state of the art with regard to the theoretical analysis of routing algorithms and provide an overview of existing approaches for Darknets. Due to our focus on routing in connectivity-restricted networks rather than on anonymization, we do not consider P2P-based anonymization services requiring connections to arbitrary peers such as I2P[2] or Torsk [McLachlan et al. 2009] in detail.

### 2.1. Routing Analysis

In this section, we describe the state of the art with regard to the theoretical analysis of routing algorithms. When analyzing decentralized routing algorithms, the most intensively studied property is the expected routing length. Let $V$ be the set of nodes and $R^A(s,t)$ denote the number of steps needed to route from node $s$ to node $t$ using algorithm $A$. The maximal expected routing length is then given by $\max_{s,t\in V} \mathbb{E}(R^A(s,t))$. The expected routing length is similarly defined as $\frac{1}{|V|(|V|-1)} \sum_{s\neq t\in V} \mathbb{E}(R^A(s,t))$. The analyzed models vary with regard to the locally available information as well as the degree distribution of the nodes. In general, nodes are aware of all identifiers in their $k$-neighborhood for a small $k$, most commonly $k = 1$. The degree distribution, that is, the probability distribution of a node to have a certain number of neighbors, is often assumed to be constant (e.g., in Kleinberg [2000] and Martel and Nguyen [2003]), but also more general degree distributions frequently observed in complex networks have been considered (e.g., in Fraigniaud and Giakkoupis [2009]).

In Kleinberg's famous model for routing small-world networks, nodes are placed on an $m$-dimensional lattice. Each node $v$ then is connected to all nodes within distance $p \geq 1$ and additionally has $q \geq 1$ long-range contacts. A long-range contact $u$ is chosen with probability antiproportional to $d^r$ for some $r > 0$, where $d$ is the distance of $v$ to $u$. The routing length of the standard algorithm with respect to the described topology model is polylog if and only if $r = m$ [Kleinberg 2000]. The result for the case $r = m$ has been extended in various ways: it has been shown that the standard routing algorithm has expected routing length $\Theta(\log^2 n)$ steps. Since the diameter is logarithmic, this is not asymptotically optimal. Consequently, extensions of the routing algorithm using the information of $\lceil \log n \rceil$ nodes in each step have been proposed, which reduce the expected routing length to $\Theta(\log^{1+1/m} n)$ [Martel and Nguyen 2003]. Similar alternative routing algorithms, considering a larger neighborhood before choosing the next hop, have been discussed in Lebhar and Schabanel [2004] and Giakkoupis and Schabanel [2011]. Though achieving close-to-optimal or optimal performance, these algorithms are designed considering a constant degree distribution. Furthermore, they are based on additional knowledge about the network size, which is not supposed to be known in a privacy-preserving embedding. More closely related to the topic of Darknets, Fraigniaud and Giakkoupis analyzed greedy routing for a scale-free distribution with exponent $\alpha$. The expected routing length for directed scale-free graphs is asymptotically the same as in the original model, but in case of undirected links, it is reduced to $\mathcal{O}(\log^{\alpha-1} n \log\log n)$ [Fraigniaud and Giakkoupis 2009]. In their generative model for the undirected graphs, long-range links are first created as directed edges and then the reverse edges are added. However, the resulting graphs are not truly undirected and

[2] https://geti2p.net/en/.

hence fail to provide some essential properties. In particular, the number of neighbors of neighbors given the degree of the node is distributed differently than in undirected graphs, and thus the model is not suitable for analyzing neighbor-of-neighbor routing in undirected graphs. Our model provides undirected graphs by design. The case of using NoN information for routing has been treated in Manku et al. [2004], who found that with $\Theta(\log n)$ neighbors per node, the expected routing length is asymptotically equal to the diameter of $\mathcal{O}(\frac{\log n}{\log \log n})$. However, general degree distributions are not considered. All these results assume a greedy embedding on a lattice. Thus, they are only of limited applicability to Darknet topologies, which in general do not offer the required lattice structure. In the next section, we show how the assumptions on the underlying structure can be loosened to better model restricted topologies.

### 2.2. Darknet Routing

Early routing approaches for Darknets, for example, Turtle [Popescu et al. 2006], use flooding, and hence are aimed at rather small network sizes. Probabilistic search has been implemented in OneSwarm [Isdal et al. 2010], a Darknet protocol for BitTorrent. Both approaches can lead to large overhead, low success rates, and long routes in case of rare files and sparse topologies. GNUnet, an anonymous publication system with a Darknet mode, uses recursive Kademlia for routing, restricting the neighbors to trusted contacts [Evans and Grothoff 2011]. It requires a high replication rate to still locate content. All of these approaches have mainly been proposed for anonymous file sharing with a high replication rate for popular files. They are not designed to provide social networking or real-time communication services.

Second-level virtual overlays have been proposed to decrease latency and overhead: MCON [Vasserman et al. 2009] hence implements structured peer-to-peer systems by connecting the closest neighbors in the namespace through tunnels of trusted nodes. However, their original design, restricting a node's knowledge to its direct neighbors, exhibited a low resilience to failures and churn. Aiming to improve the resilience, the authors hence proposed a *robust* routing algorithm including NoN. Their NoN algorithm differs from ours in that messages are actually sent to all neighbors of the next hop, whereas we only include information about the neighbor's IDs in a namespace. As a consequence, their robust scheme reveals the actual identity of the NoNs by establishing connections between them, whereas we only reveal their pseudonymous IDs in the namespace, which prevents direct identification. Furthermore, sending a message to each neighbor of the next hop for all hops drastically increases the number of messages required for routing in MCON, whereas our scheme reduces the message overhead at the price of slightly increased computation costs at each hop. An essential drawback of MCON is the high cost of maintaining the tunnels produced by the use of flooding for overlay neighbor discovery. In contrast, X-Vine [Mittal et al. 2012] establishes tunnels by leveraging the overlay routing of the social contacts of a newly joined or otherwise disconnected node. Though X-Vine is primarily designed as a Sybil defense for distributed hash tables (DHTs), its design is based on a social graph. The performance of X-Vine over a longer period of joins and departures has not been considered in Mittal et al. [2012], so it remains unclear if sufficiently short tunnels can be maintained over time. Indeed, recent work shows that tunnels of a polylogarithmic length can only be maintained at a high cost [Roos and Strufe 2015], indicating that the tunnel length in X-Vine is bound to increase beyond a suitable length eventually. Thus, virtual overlays such as MCON and X-Vine are by design costly with regard to either routing or maintenance overhead when considering their long-term behavior.

The Darknet modus of Freenet, the only widely used P2P-based system for censorship-resilient and anonymous communication, makes uses of an *embedding*.

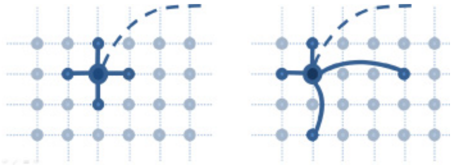

Fig. 1. Kleinberg model (left) with full neighborhood connectivity and one long-range link (dashed line), opposed to our topology model with connectivity within $C$-neighborhood and a long-range link (right, further long-range links omitted).

Nodes are assigned identifiers (*IDs*) in a metric space (the *ID space*). The distances in the ID space are supposed to mirror the structure of the social graph. In general, *embeddings* are designed to be *greedy*: a stateless *greedy routing*, so that each node on the path contacts the neighbor closest to the destination, is supposed to always find its destination. However, existing greedy embeddings disclose topology information and have been shown to suffer from a highly unbalanced load distribution [Höfer et al. 2013]. Thus, the embedding deployed in Freenet is not greedy. Nodes are usually not adjacent to the closest nodes in the ID space. The standard routing algorithm hence fails and has to be adapted to deal with local optima during the routing process. Freenet suggests a *distance-directed depth-first search* $D^2$*-DFS* to mitigate inaccuracies in the embedding but fails to provide a proof that the modified algorithm achieves the desired polylog routing steps [Clarke et al. 2000].

## 3. MODEL AND ALGORITHM DESIGN

In this section, we define our model of a Darknet topology. Afterward, the considered routing algorithms are introduced. We conclude this section by giving an informal overview of the proof ideas for the later sections and discussing the impact of changing the routing algorithm on the anonymity. The model has been used as a basis for the results in Roos and Strufe [2012] and Roos and Strufe [2013].

### 3.1. Model

We use a model for restricted topologies with nongreedy embeddings, extending Kleinberg's small-world model [Kleinberg 2000]. Though Kleinberg's model offers an explanation of how short paths are found in small-world networks, it is only of restricted use with respect to Darknets due to the assumed greedy embedding. Our Darknet model introduces an additional parameter $C$ to characterize the maximal distance to the closest neighbor in nongreedy embeddings. Furthermore, Kleinberg's model is extended to allow for arbitrary degree distributions and undirected graphs.

Throughout the article, $P$ denotes a probability measure, and whp is used to denote with high probability. Note that we generally drop the brackets indicating sets for improved readability.

A graph of the class $\mathcal{D}(n, m, C, L)$ consists of $n^m$ nodes, arranged in an $m$-dimensional hypercube of side length $n$, so nodes are given unique identifiers (IDs) in $\mathbb{Z}_n^m$. In the following, we use the name $v$ of a node synonymously with its identifier $id(v) = (v_1, \dots, v_m)$. The distance between two nodes $u$ and $v$ is given by the Manhattan distance with wraparound, that is, $dist(v, u) = \sum_{i=1}^{m} \min\{|v_i - u_i|, n - |v_i - u_i|\}$.

The parameter $C$ is a measure for the accuracy of the embedding and gives the maximal distance to the closest neighbor in each principal direction. Figure 1 illustrates the difference in the choice of short-range links between Kleinberg's and our model for the case $d = 2$ and $C = 3$: the node in the center indeed has short-range links to two of the four nodes closest to it, but the links to the other two are replaced by longer links in the same principal direction. Formally, each node $v = (v_1, \dots, v_m)$ is given short-range links to neighbors $a_1^v, \dots, a_m^v, b_1^v, \dots, b_m^v$. Here, $a_j^v$ is chosen from the set

$$A_j^v = \{u = (u_1, \dots, u_m) \in V : u_i = v_i \text{ for } i \neq j, 1 \leq \min\{u_j - v_j, n + u_j - v_j\} \leq C\}.$$

Analogously, $b_j^v$ is chosen from

$$B_j^v = \{u = (u_1, \dots, u_m) \in V : u_i = v_i \text{ for } i \neq j, 1 \leq \min\{v_j - u_j, n + v_j - u_j\} \leq C\}.$$

The random variable $L$ governs the degree distribution, an inherent property of the trust graph. In addition to the short-range links, long-range links are chosen in a two-step process:

(1) Choose a label $l_v \in \mathbb{N}$, distributed according to $L$, for each node $v \in V$.
(2) Connect nodes $u$, $v$ with probability

$$P(l(u, v)|l_u = l_1, l_v = l_2, dist(u, v) = d) = 1 - e^{-\frac{l_1 l_2}{d^m \gamma}}, \tag{1}$$

where $\gamma$ is a normalization constant chosen such that

$$2 \sum_{d=1}^{n/2} \sum_{l_1=1}^{\infty} \left(1 - e^{-\frac{l_1}{d^m \gamma}}\right) P(L = l_1) = 1; \tag{2}$$

that is, the expected number of long-range links of a node $v$ with label $l_v = 1$ is 1.

We denote all long-range neighbors of a node $u$ by $LN(u)$, and short-range neighbors by $SN(u)$.

The following lemma provides a general result about the model needed for the rest of the article.

Lemma 3.1. *Let the average degree $\mathbb{E}(L)$ be bound by a constant. The probability that a long-range link is at least of length $2\sqrt{n}$ is constant, that is,*

$$P(dist(u, v) \geq 2\sqrt{n}|l(u, v)) = \Omega(1).$$

Proof. The proof is a direct application of Bayes' Theorem combined with a summation over all possible distances, that is,

$$\begin{aligned} P(dist(u, v) \geq 2\sqrt{n}|l(u, v)) &= \sum_{d=2\sqrt{n}}^{n} P(dist(u, v) = d|l(u, v)) \\ &= \sum_{d=2\sqrt{n}}^{n} \frac{P(l(u, v)|dist(u, v) = d)P(dist(u, v) = d)}{P(l(u, v), dist(u, v) = d)}. \end{aligned}$$

We start by deriving $P(l(u, v)|dist(u, v) = d)$. Recall that $\gamma = \Theta(\log n)$ is the normalization constant in Equation (2). First, consider that the probability that $u$ and $v$ are neighbors given their distance is

$$P\left(l(u, v)|dist(u, v) = d\right) = \sum_{l_1=1}^{\infty} \sum_{l_2=1}^{\infty} \left(1 - e^{-\frac{l_1 l_2}{d\gamma}}\right) P\left(L = l_1\right) P\left(L = l_2\right)$$

$$= \sum_{l_1=1}^{\infty} \sum_{l_2=1}^{\infty} \Theta\left(\frac{l_1 l_2}{d\gamma} P\left(L = l_1\right) P\left(L = l_2\right)\right) = \Theta\left(\frac{1}{\gamma d} \mathbb{E}(L)^2\right) = \Theta\left(\frac{1}{\gamma d}\right).$$

The last step holds because $E(L)$ is bound by a constant. Note that the probability that two randomly selected nodes on a ring of length $n$ have at least distance $\sqrt{n}$ converges to 1. The claim now follows from

$$\begin{aligned} P(dist(u,v) \geq 2\sqrt{n} | l(u,v)) &= \frac{P(l(u,v)|dist(u,v) \geq 2\sqrt{n})}{P(l(u,v))} P(dist(u,v) \geq 2\sqrt{n}) \\ &= \Omega\left(\frac{P(l(u,v)|dist(u,v) \geq 2\sqrt{n})}{P(l(u,v))}\right) \\ &= \Omega\left(\frac{\sum_{d=2\sqrt{n}}^{n/2} \frac{1}{\gamma d} \cdot \frac{2}{n}}{\sum_{d=1}^{n/2} \frac{1}{\gamma d} \cdot \frac{2}{n}}\right) = \Omega\left(\frac{\sum_{d=2\sqrt{n}}^{n/2} \frac{1}{d}}{\sum_{d=1}^{n/2} \frac{1}{d}}\right) \\ &= \Omega\left(\frac{\log(n/2) - \log(2\sqrt{n})}{\log(n/2)}\right) = \Omega\left(\frac{1/2 \log n - 4}{\log n}\right) \\ &= \Omega(1). \end{aligned}$$

The second last step holds because $\sum_{i=1}^{n} \frac{1}{i} = \Theta(\log n)$. □

In the following, we assume that labels and hence the node degrees are chosen according to a scale-free distribution $S_\alpha$ with exponent $2 < \alpha < 3$ and a maximum $\mu$, that is,

$$P(S_\alpha = k) \propto \frac{1}{k^\alpha}, \quad k = 1 \ldots \mu. \tag{3}$$

Scale-free degree distribution is common in various complex networks, especially social networks. Theorem 4.1 in Section 4 does not assume a scale-free degree distribution, so the inefficiency of the Freenet routing algorithm holds regardless of the degree distribution. For the remaining results, the actual bounds depend on the degree distribution, but the fact that the algorithms allow polylog performance holds for arbitrary degree distributions as explained in greater detail in the respective sections. Furthermore, the set $B_d(v) = \{u : dist(v, u) < d\}$ contains all nodes at distance less than $d$ of $v$. For an event $A$, we denote its complement by $A^\perp$.

For reasons of presentation, results are given for $m = 1$ dimensions but can analogously be derived for multidimensional identifier spaces.

### 3.2. Algorithms

For deterministic routing based on a nongreedy embedding, state information is needed to avoid loops and dead ends. In this section, we introduce several algorithms that can deal with local optima with regard to the distance to the target.

These algorithms are based on two principles: (1) backtracking is used in case a node has no suitable neighbor to forward the message to, and (2) nodes are *marked* when they should only be contacted for backtracking in the future. The order by which nodes are *marked* is crucial for the routing performance. The straightforward approach, currently implemented in Freenet, is a distance-directed depth-first search $D^2$-DFS. Here, a node $u$ is *marked* the first time it receives a message. $u$ selects the neighbor $v$ that is closest to the target as the next hop and has not been contacted before. If $v$ is already *marked*, it returns a backtrack message, and $u$ either contacts the next closest node or sends a backtracking message to its predecessor if all neighbors have been contacted. The routing is considered failed if the source node cannot forward the message to any neighbor. We show in Section 4 that $D^2$-DFS does not achieve a polylog routing length. Thus, we present *NextBestOnce**, a generic algorithm that can be applied to multiple scenarios. These scenarios differ by the amount of information each node maintains. *NextBestOnce**'s performance is based on the fact that it always chooses the neighbor that offers the least fallback regardless of whether this neighbor has seen the message

**ALGORITHM 1:** NextBestOnce*(Node v, p, ID t, Set $B$, Boolean b)

```
 1: # input: v: message holder, p: predecessor, t: target
 2: #        $B$: *marked* nodes, b: backtracking flag
 3: # $N_v$: neighbors of v
 4: # $IDS(u)$: set of identifiers associated with $u$
 5: **if** id(v) == t **then**
 6:    routing successful; terminate
 7: **end if**
 8: **if** !backtrack **then**
 9:    v.predecessor.add(p);
10: **end if**
11: $S = \{u \in N_v : !B.contains(u)\}$
12: **if** $S$ NOT EMPTY **then**
13:    nextNode = $argmin_{u \in S} dist(IDS(u)), t)$
14:    b = false;
15:    **if** $dist(nextNode, t) \geq dist(v, t)$ **then**
16:       B.add(v)
17:    **end if**
18: **else**
19:    B.add(v)
20:    nextNode = v.predecessor.pop();
21:    b = true;
22: **end if**
23: **if** nextNode != null **then**
24:    NextBestOnce(v, t, nextNode, B, b)
25: **else**
26:    routing failed; terminate
27: **end if**
```

before or not. A node is only *marked* after all its neighbors closer to the target have been *marked*. In the basic Darknet scenario, the only available information is the ID of all neighbors. However, when the average degree in a graph is low, this has been shown to entail long routing paths and frequent backtracking [Schiller et al. 2011; Roos and Strufe 2012, 2013]. Hence, *NextBestOnce** allows the possibility to include additional topology information. Each neighbor $u$ is mapped to a set *IDS*($u$) rather than only one $ID$. In this article, we explicitly consider the case that *IDS*($u$) consists of the ID of u and the IDs of its neighbors. However, *NextBestOnce** is designed to work in more general situations.

*NextBestOnce** is detailed in Algorithm 1. The input of *NextBestOnce** consists of the current message holder $v$, the predecessor $p$ of $v$, the target ID $t$, the set $B$ of *marked* nodes, and a flag $b$ indicating if the routing is in the backtracking phase. Note that $B$ can be realized in a privacy-preserving manner, for example, by relying on a Bloom filter, and is not decisive for the asymptotic routing length. Each node keeps a stack of predecessors for backtracking, which are contacted if $v$ has only *marked* neighbors (ll. 19–20).

If at least one neighbor is not *marked*, $v$ selects the not *marked* neighbor $u$ so that the distance to one of the identifiers in $IDS(u)$ is minimal (l. 12).

After determining the next node $u$ on the path, $v$ is *marked* if $u$ is at a larger distance to $t$ (l. 15) or backtracking starts (l.18). To guarantee termination, only one representative ID of $u$ is considered for the decision of *marking* $v$.

In the remainder of the article, we first show that the distance-directed depth-first search $D^2$-*DFS* as deployed in Freenet cannot achieve a polylog hop count. Afterward, we analyze the performance of two routing algorithms based on *NextBestOnce**. The

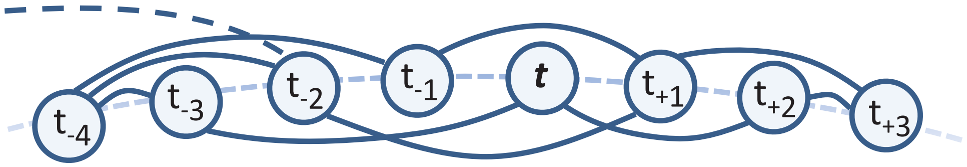

Fig. 2. Exemplary adverse connectivity for $D^2$-*DFS*.

first one, *NextBestOnce,* has been proposed in Roos and Strufe [2012] and only uses the identifiers of the direct neighbors, that is, $IDs(u) = id(u)$ in Algorithm 1. The second algorithm, *NextBestOnce-NoN*, uses information about neighbors of neighbors, that is, $IDs(u) = \{id(u)\} \cup \bigcup_{v \in N(u)}\{id(v)\}\}$, where $N(u)$ is the set of neighbors of node $u$. The number of hops required by $D^2$-*DFS*, *NextBestOnce*, and *NextBestOnce-NoN* to find a path from source $s$ to destination $t$ are denoted by $R^{DFS}$, $R^{NBO}(s,t)$, and $R^{NoN}(s,t)$, respectively.

In the following sections, we first show that *NextBestOnce* achieves a polylog routing length, while the Freenet algorithm $D^2$-*DFS* does not scale polylog with the network size. The principal reason for $D^2$-*DFS*'s inadequate performance lies in the fact that nodes are not considered twice before the backtracking phase. However, due to the inaccuracies of the embedding, a node $u$'s closest neighbor $v$ to the target $t$ in terms of distance *dist* might not be as close to $t$ in terms of the hops as a neighbor $w$ at a slightly higher distance. For $D^2$-*DFS*, $u$ forwards the message to $v$ rather than $w$. $v$ in turn chooses the closest of its not-yet-contacted neighbors to $t$ despite them potentially being at a larger distance to $t$ than $v$. Now, $w$ cannot be contacted by $u$ until the backtracking phase. For the routing to terminate in a timely fashion, we thus have to discover either a path to $w$ not containing $u$ or a path to $t$ not containing $w$. In the proof, we show that both possibilities are sufficiently unlikely for the expected routing length not to be polylog. Figure 2 shows an exemplary scenario for which the routing does not converge within an acceptable number of steps with high probability. *NextBestOnce* removes this principal weakness of $D^2$-*DFS* by allowing $u$ to be contacted several times before the backtracking phase as long as $u$ has not contacted all neighbors offering a potential improvement. In this scenario, *NextBestOnce* allows $v$ to return the message to $u$ when not in possession of any other suitable neighbor rather than contacting a node at a large distance to $t$. Then, $u$ can contact $w$, which has indeed a short path to $t$. So, a timely termination of the routing can be guaranteed.

In Section 6, we prove that *NextBestOnce-NoN* offers more than a constant improvement over *NextBestOnce* by taking the IDs of neighbors of neighbors into consideration. The proof is based on two observations: high-degree neighbors are more likely to have a neighbor close to $t$ because they have more neighbors to choose from. By allowing a node to consider the coordinates of the neighbor's neighbor, a node is more likely to choose the high-degree neighbor and achieve a large improvement in two hops. In contrast, when only considering direct neighbors, the next hop is chosen independently of its degree and thus might have little potential to improve the routing further.

### 3.3. Impacts on Anonymity

One of the key concerns of routing in Darknets is to maintain the anonymity and membership concealment. Recent work describes how the Freenet routing algorithm $D^2$-*DFS* can undermine sender anonymity due to its loop detection. Furthermore, the authors show that algorithms like *NextBestOnce* that do not reroute when encountering a loop enhance the anonymity [Tian et al. 2014]. Both *NextBestOnce* and $D^2$-*DFS* reveal information about the receiver of the route by including a target ID in the routing.

The identifier clearly allows a node to estimate the probability that a neighbor is the receiver, that is, the closest node to the target identifier. However, the problem is widely known for anonymous routing in structured P2P systems and is out of the scope of the current work. Possible solutions are described in, for example, Panchenko et al. [2009] and Mittal and Borisov [2009].

So, while *NextBestOnce* actually increases the privacy of the routing in contrast to $D^2$*-DFS*, *NextBestOnce-NoN* is clearly more privacy invasive. By being aware of the IDs of neighbors of neighbors, a node is given additional information about the social graph, namely, which of its trusted contacts also share a trust relationship. Furthermore, the probability for nodes in the 2-hop neighborhood to be the receiver can be estimated. The question of whether such information should be available for an increased performance is highly scenario dependent. In the current version of Freenet, neighbor-of-neighbor information is hence only revealed if nodes allow it. In this article, we only show that such information indeed increases the performance considerably, concluding that it should be revealed when considered acceptable.

## 4. ANALYSIS OF $D^2$-DFS

In this section, we analyze the performance of $D^2$*-DFS* in the context of $\mathcal{D}(n, 1, C, L)$. The only requirement with regard to the degree distribution is that the degree of a node is bound by a constant $T$ with probability $r$, meaning that the degree of a certain percentage of nodes does not increase with the network size. A variant of this proof is presented in Roos and Strufe [2013].

THEOREM 4.1. *Let $L$ be such that the degree $D_u$ of node $u$ is bound by a constant $T \in \mathbb{N}$ with constant probability $r \in \mathbb{R}_+$, that is,*

$$P(D_u \leq T) \geq r > 0, \tag{4}$$

*and $C > 2$. Then $D^2$-DFS does not have polylog expected routing length; that is, for any $\rho > 0$,*

$$\frac{1}{n(n-1)} \sum_{s \neq t \in V} \mathbb{E}(R^{DFS}(s,t)) = \Omega(\log^{\rho} n). \tag{5}$$

We identify the main drawback of $D^2$-DFS: if all short-range neighbors of a node have received the message, the message is forwarded via a long-range link. Hence, whp a node at a high distance to the target is contacted, so the progress up to this point is nullified. We show that in this manner, polylog routing complexity is not possible. The idea of the proof is based on two steps: we first show that with constant probability, the target can only be reached via a long-range link to a node in its vicinity or after the backtracking phase has started (Lemma 4.3). Second, we show that with high probability, both conditions are not fulfilled within a polylog number of steps (Lemma 4.4). The result then follows from the fact that the routing length is not polylog with constant probability and hence the average routing length is not polylog either.

We start by proving that nodes closer to the target are more likely to be contacted. An intuitive explanation for this fact is that the improvement toward the target decreases with the distance, so that the density of contacted nodes increases as routing nears the termination.

LEMMA 4.2. *Let $P_t \subset V$ be the nodes contained on the routing path of $R^{DFS}$ with target $t$. The probability that a node $v$ is contained in $P_t$ is bound from above by*

$$P(v \in P_t | dist(v,t) \geq d) = \mathcal{O}\left(\frac{|P_t|}{d}\right). \tag{6}$$

*In particular, for two nodes $u$, $v$ with $dist(u, t) < dist(v, t)$, the probability that $u$ is contained in the path is of the same order as the probability that $v$ is part of the path, that is,*

$$P(u \in P_t | u \in P_t \cup v \in P_t) = \Omega(1). \tag{7}$$

Proof. We first show Equation (7). Let $u$ and $v$ be nodes such that $dist(u, t) < dist(v, t)$. Let $w$ be node a on the path. If $w$ has edges to both $u$ and $v$, it is more likely to forward the message to $u$. It remains to show that the probability to have an edge to $u$ is at least as high as the probability to have an edge to $v$ for any distance $dist(w, t)$. Recall that $P(l(u, v)|dist(u, v) = d) = \theta(\frac{1}{\gamma d})$ for the normalization constant $\gamma$. Note that the distance between $w$ and a node $z$ is either $|dist(w, t) - dist(z, t)|$ (if $z$ and $w$ are on the same side of $t$) or $dist(w, t) + dist(z, t)$ (if $z$ and $w$ are on opposite sides of $t$). Hence,

$$\begin{aligned} &P(l(w, z)|dist(z, t) = d_z \cup dist(w, t) = d_w) \\ &\quad = \theta\left(\frac{1}{\gamma}\left(\frac{1}{|d_z - d_w|} + \frac{1}{d_z + d_w}\right)\right) = \theta\left(\frac{1}{\gamma(d_z + d_w)}\right), \end{aligned}$$

and as a result

$$P(l(w, u)|dist(u, t) = d_u \cup dist(w, t) = d_w) = \Omega(P(l(w, v)|dist(v, t) = d_v \cup dist(w, t) = d_w))$$

because $1/(dist(u, t) + dist(w, t)) \geq 1/(dist(v, t) + dist(w, t))$. Equation (7) follows. In order to derive Equation (6), consider that each node on the path is more likely to be one of $2(d - 1)$ nodes that are closer to $t$ than $v$. So, the probability of any node on the path to be $v$ is at most $1/(2d - 1)$. Equation (6) follows by a union bound. □

We now describe a neighbor selection in the vicinity of $t$ that entails that $D^2$-DFS chooses a path, such that $t$ can only be reached by a long-range link or backtracking. Furthermore, we show that the probability of this event is constant in $n$. In general, we denote the vicinity of $t$ by $S_t = \{t_{-m_1}, \ldots, t, \ldots, t_{m_2}\}$, containing $m_1$ nodes with lower and $m_2$ nodes with higher IDs. Our proof considers the case of $m_1 = 4$ and $m_2 = 3$. During the proof, we will use the fact that all nodes in $S_t$ but $t_{-m_1}$ and $t_{m_2}$ can have short-range links to nodes within $S_t$ only. For brevity, we denote the set of these nodes by $S_t^I = S_t \setminus \{t_{-m_1}, t_{-m_2}\}$. Furthermore, let $\Lambda_t$ be the set of nodes that are reached by long-range links and $B_{S_t}$ the set of nodes that are first contacted after at least one node in $S_t$ has been backtracked to. We abbreviate the event $E = \cup_{v \in S_t^I} v \in \Lambda_t \cup t \in B_{S_t}$ that any node in $S_t^I$ is reached by long-range link or $t$ is reached by backtracking.

Lemma 4.3. *For a source $s$ and a destination $t$ with local neighborhood $S_t$ and $s \notin S_t$, it holds that $P(E) = \Omega(1)$.*

Proof. We consider the event $A$ of link selection presented in Figure 2. We first show that the probability of $A$ is constant. Then we demonstrate that if $S_t$ is not first reached by a long-range link, $D^2$-DFS forwards the message from a node in $S_t^I$ to a node not in $S_t$ with high probability, so that afterward nodes in $S_t$ can only be contacted during backtracking or via long-range links. The nodes have the depicted short-range link selection, and in addition, $t_{-2}$ is required to have one long-range link. In this lemma, we do not make any assumptions about the number of long-range links of other nodes. We condition

$$P(E) \geq P(E|A)P(A).$$

First, note that the probability for $A$ is only dependent on $C$ and $r$, and hence, $P(A) = \Omega(1)$[3] Furthermore,

$$P\left(\cup_{v\in S_t^I} v \in \Lambda_t \cup t \in B_{S_t} \middle| A\right)$$
$$= P\left(\cup_{v\in S_t^I} v \in \Lambda_t \middle| A\right) + P\left(t \in B_{S_t} \middle| A \cap \left(\cup_{v\in S_t^I} v \in \Lambda_t\right)^{\perp}\right) \cdot P\left(\left(\cup_{v\in S_t^I} v \in \Lambda_t\right)^{\perp} \middle| A\right).$$

Because for all events $Y$, $P(Y) + P(Y^{\perp}) = 1$ and hence $P(Y) + \Omega(1)P(Y^{\perp}) = \Omega(1)$, it remains to show

$$P\left(t \in B_{S_t} \middle| A \cap \left(\cup_{v\in S_t^I} v \in \Lambda_t\right)^{\perp}\right) = \Omega(1).$$

If the first contacted node $X$ in $S_t^I$ is not reached by a long-range link, $X$ is either $t_{-4}$ or $t_{+3}$. If $t_{+3}$ is contacted, $t_{+1}, t_{-1}$, $t_{-4}$, and $t_{-2}$ are contacted. Afterward, the message is forwarded to a long-range contact of $t_{-2}$, and nodes in $S_t$ can only be reached by long-range links or backtracking. So a lower bound on the probability to backtrack given that $S_t$ is not reached by a long-range link is

$$P\left(t \in B_{S_t} \middle| \left(\cup_{v\in S_t^I} v \in \Lambda_t\right)^{\perp} \cap A\right) \geq P(X = t_{+3}) = \Omega(1),$$

where the last step follows from Lemma 4.2. This completes the proof. □

The second step in the proof gives a bound on the number of hops to reach a node in $S_t$ via a long-range link or during backtracking.

LEMMA 4.4. *The probability that the routing takes at least $M \log^{\rho} n$ hops is bound from below by*

$$P(R^{DFS}(t, s) \geq M \log^{\rho} n \,|E) = \Omega(1),$$

*given the event E that nodes in $S_t^I$ are reached via long-range links or t is reached during backtracking.*

PROOF. We first show that the routing length exceeds $M \log^{\rho} n$ whp if $S_t$ is reached by a long-range link. Then we show that the same holds for backtracking. The claim follows. For both cases, we condition on the number of long-range links and their lengths, showing that whp all long-range contacts of nodes in $S_t$ are at a high distance to $t$. Thus, it is very unlikely that they are found during the routing because only few nodes at a high distance are contacted. For backtracking to start, all nodes reachable from the node $v \notin S_t$ contacted by $u \in S_t$ need to have received the message. Again, this is unlikely if $v$ is far from $t$. We now formalize the previous proof idea. Recall that $P(D_u \leq T) \geq r$ for some constant $r > 0$ and $p_l = P(dist(u, v) \geq 2\sqrt{n}|l(u, v))$. We first determine the probability of the event

$$A = \cap_{v\in S_t}(|LN(v)| \leq T \cap \ \cap_{u\in LN(v)} dist(u, v) \geq 2\sqrt{n})$$

that all long-range links of nodes in $S_t$ have length of at least $2\sqrt{n}$ and there are at most $|S_t|T$ long-range links into $S_t$. By Equation (4), Lemma 3.1, and a union bound, we have $P(A) \geq (p_l^T r)^{|S_t|} = \Omega(1)$ because all involved quantities are bound by a constant. So it remains to show that

$$P(R^{DFS}(t, s) \geq M \log^{\rho} n \,|E \cap A) = \Omega(1). \tag{8}$$

[3] The exact probability is computed in Roos and Strufe [2013].

By Lemma 4.2, we can bound the probability that a long-range link into $S_t$ has been encountered in $M \log^{\rho} n$ steps

$$\begin{aligned} P\big(R^{DFS}(t,s) \geq M \log^{\rho} n \,\big|\cup_{v \in S_t^I} v \in \Lambda_t \cap A\big) &= \Omega\left(\left(1 - \frac{M \log^{\rho} n}{\sqrt{n}}\right)^{|S_t|T}\right) \\ &= \Omega\left(1 - \frac{|S_t| T M \log^{\rho} n}{\sqrt{n}}\right) \\ &= \Omega(1). \end{aligned} \tag{9}$$

Backtracking is only applied if after following a long-range link from $S_t$ to a node $v$, all nodes reachable from $v$ have been contacted. In particular, these nodes include at least $M \log^{\rho} n$ nodes that are on the path following consecutive short-range links starting at $v$ and at distance $d > dist(v,t)$ to $t$. If $A$ holds, then $dist(v,t) \geq 1/2 dist(v, S_t) \geq \sqrt{n}$ also holds. We hence get by Lemma 4.2 with $|P_t| = M \log^{r} n$ the lower bound

$$\begin{aligned} &P(R^{DFS}(t,s) \geq M \log^{\rho} n \,|t \in B_{S_t} \cap A) \\ &= P(R^{DFS}(t,s) \geq M \log^{\rho} n \,|t \in B_{S_t} \cap dist(v,t) \geq \sqrt{n}) \\ &= \Omega\left(\left(1 - \frac{M \log^{\rho} n}{\sqrt{n}}\right)^{M \log^{\rho} n}\right) = \Omega(1). \end{aligned} \tag{10}$$

Equation (8) is a direct consequence of Equations (9) and (10). □

The proof of Theorem 4.1 follows directly.

PROOF. For the expected routing length not to be polylog, it suffices to show that it is not polylog conditioned on the event that $S_t$ is either reached by a long-range link or by backtracking. Formally, for $s, t \in V$ with $s \notin S_t$, we have

$$\begin{aligned} &\mathbb{E}(R^{DFS}(s,t)) \geq M \log^{\rho} n \; P(R^{DFS}(s,t) \geq M \log^{\rho} n) \\ &\geq M \log^{\rho} n \; P(E) \, P(R^{DFS}(t,s) \geq M \log^{\rho} n \,|E) \\ &= \Omega(\log^{\rho} n). \end{aligned}$$

The last step holds by Lemmas 4.3 and 4.4. Because $P(s \notin S_t) = \Omega(1)$, the claim

$$\frac{1}{n(n-1)} \sum_{s \neq t \in V} \mathbb{E}(R^{DFS}(s,t)) = \Omega(\log^{\rho} n)$$

is proven. □

We have shown that the routing algorithm in Freenet cannot achieve polylog routing length if $C > 2$. In the following, we prove that *NextBestOnce*, in contrast, can achieve polylog routing length.

## 5. ANALYSIS OF NEXTBESTONCE

We present upper and lower bounds for the performance of *NextBestOnce*. The routing length increases at least linearly with $C$, the maximal distance to a local neighbor. If $C$ is constant, the bounds are the same for a lattice [Fraigniaud and Giakkoupis 2009].

### 5.1. Upper Bound

A simplified version of Theorem 5.1 for constant $C$ has been shown in Roos and Strufe [2012].

THEOREM 5.1. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$ and two nodes $s, t \in V$ at distance $d = dist(s, t)$, an upper bound on the expected routing length of* NextBestOnce *is given by*

$$\mathbb{E}(R^{NBO}(s, t)) = \mathcal{O}(\log^{\alpha-1} d \log\log d + C^3 \log n). \tag{11}$$

*The maximal expected routing length is consequently*

$$\max_{s,t\in V} \mathbb{E}(R^{NBO}(s, t)) = \mathcal{O}(\log^{\alpha-1} n \log\log n + C^3 \log n).$$

The upper bound on *NextBestOnce*'s routing length is derived by dividing the routing into two phases: the number of steps $R_1^{NBO}(s, t)$ needed to reach a node $v$ within distance $C$ of $t$ and the number of steps $R_2^{NBO}(s, t)$ to reach $t$ from $v$. The result for the first phase follows directly from the corresponding result for the lattice (Lemma 5.2). The idea of the proof for the second phase is to show that (1) whp the message is not forwarded to a node that is not in $t$'s proximity, and (2) the number of hops needed to find $t$ by only contacting nodes in its proximity is $\mathcal{O}(C^3 \log n)$. We first show that there exist polylog paths between two nodes within distance $C^2 \log n$ (Lemma 5.3). Afterward, we prove that these paths are discovered by *NextBestOnce* in the claimed number of hops (Lemma 5.4 and Lemma 5.5).

LEMMA 5.2. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$ and two nodes $s, t \in V$ with $d = dist(s, t)$, the expected routing length of* NextBestOnce *during the first phase is $\mathbb{E}(R_1^{NBO}(s, t)) = \mathcal{O}(\log^{\alpha-1} d \log\log d)$.*

Since the distance to $t$ decreases by at least 1 in each step during the first phase, the previous lemma is essentially treated in Fraigniaud and Giakkoupis [2009], Theorem 2.4, which gives the upper bound of standard greedy routing on a lattice with a scale-free degree distribution.

Lemma 5.3, showing the existence of paths between nodes at distance at least $C^2 \log n$, requires the concept of a *greedy* path, a path with nodes at a monotone decreasing distance to some node $v$. Formally, a greedy path is a sequence of neighboring nodes $u_0, u_1, \ldots, u_{l+1}$, so that $dist(u_i, v) < dist(u_{i-1}, v)$ for $i = 0, \ldots, l$. Let $g(v, w)$ indicate if $v$ and $w$ are connected by a greedy path.

LEMMA 5.3. *For two nodes $w, v \in V$ with $dist(w, v) \geq C^2 \log n$, the probability that $w, v$ are connected by a greedy path is*

$$P(g(w, v)) = \Omega\left(1 - \frac{1}{n}\right).$$

PROOF. The idea of the proof is to show that each node has probability at least $1/C$ to be on a greedy path. The probability that two greedy paths intersect on a segment of length $C^2 \log n$ is thus derived as the probability that one of those $C^2 \log n$ nodes is on both paths. Recall from Section 3 that each node $u$ has two short-range neighbors $a_1^u$ and $b_1^u$ chosen independently of each other. They are both within distance $C$ of $u$, but in opposite directions. For all pairs $(v, w) \in V \times V$, there is a path of short-range links of length at most $C$ originating at $v$ leading to a node within distance $C$ of $w$ and vice versa. A greedy path between $v$ and $w$ exists if those two paths intersect (see Figure 3). Denote by $g^a(u_0, u_{l+1})$ the event that $u_0$ and $u_{l+1}$ are connected by a greedy path and $a_1^{u_i} = u_{i+1}$ for $i = 0, \ldots, l$. $g^b(u_0, u_{l+1})$ is defined analogously. Without loss of generality, $w$ is "above" $v$ in the ID space, that is, $dist(w, v) = w - v \mod n$. Now, we show that $P(g^a(v, u)) \geq \frac{1}{C}$ for all $u \in U := \{u \in V : dist(v, w) = dist(v, u) + dist(u, w)\}$ by induction. If $dist(v, u) \leq C$, then $P(g^a(v, u)) \geq P(a_1^v = u) = \frac{1}{C}$

because all nodes have one short-range link to a node within distance $C$. Otherwise, there exists a node $z$ such that $g^a(v,z)$ holds and $dist(v,u) - C \leq dist(v,z) < dist(v,u)$. It follows that $P(g^a(v,u)) \geq P(a_1^z = u) = \frac{1}{C}$. Similarly, $P(g^b(w,u)) \geq \frac{1}{C}$ holds for all $u \in U$. Because $a_1^u$ and $b_1^u$ are chosen independently, the probability that the two paths intersect is at least the probability of one node being on greedy paths from both $w$ and $v$, that is,

$$P(g(v,w)) \geq P(\cup_{u \in U}(g^a(v,u) \cap g^b(w,u))) \geq 1 - \left(1 - \frac{1}{C^2}\right)^{C^2 \log n} \geq 1 - e^{-\frac{C^2 logn}{C^2}} = 1 - \frac{1}{n}.$$

The third inequality follows from $1 - x \leq e^{-x}$ for $x \in [0, 1]$. □

As a second step, a worst-case bound on the routing length of *NextBestOnce* is needed.

LEMMA 5.4. *Let $G = (V, E)$ be an undirected graph that is embedded in $\mathbb{Z}_{|V|}$, so that all $v \in V$ are connected to nodes within distance $C$ in each direction. The expected routing length of* NextBestOnce *on $G$ is bound by*

$$\max_{s,t \in V} \mathbb{E}(R^{NBO}(s,t)) = \mathcal{O}(C|V|).$$

PROOF. The algorithm definitely terminates after every node has been *marked*. We show that for every circle, on average at least every $C + 1$-th node is *marked*. So in expectation, all nodes are marked after $\mathcal{O}(C|V|)$ hops. First, note that the maximal increase per hop is $C$: each node $u$ has a short-range link to a node $v$, so that the $dist(v,t) \leq dist(u,t) + C$. $v$ is not yet *marked*, because a node is only *marked* after all neighbors closer to the destination, including the current message holder $u$, have been *marked*. Therefore, *NextBestOnce* can always choose a successor within distance $C$. The maximal path length without producing a circle is $|V|$. Assume the algorithm produces a circle of length $l$. Then at least $l/(C+1)$ nodes on the circle are *marked*. To see this, recall that an increase of the distance to $t$ implies that a node is declared *marked*. In case of a circle, the sum of the distance changes per hop equals zero, so the distance is increased in at least $\frac{1}{C+1} = \frac{minDecrease}{maxIncrease+minDecrease}$ of all hops of the circle. So, after a maximum number $|V|$ of hops without circles, at least every $C + 1$-th node is marked on average, so that

$$\max_{s,t \in V} \mathbb{E}(R^{NBO}(s,t)) \leq |V| + (C+1)|V| = \mathcal{O}(C|V|). \quad \square$$

It follows that the maximal number of steps is linear in the network size if the maximal increase to the destination is restricted by a parameter $C$ independent of $n$. For arbitrary graphs, the algorithm terminates after $\mathcal{O}(n^2)$ steps by Lemma 5.4. The last two lemmata enable us to bound the complexity of *NextBestOnce* during the second phase.

LEMMA 5.5. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$ and two nodes $s, t \in V$, the expected routing length of* NextBestOnce *in the second phase is bound by*

$$\mathbb{E}(R_2^{NBO}(s,t)) = \mathcal{O}(C^3 \log n).$$

PROOF. The idea of the proof is to determine a lower bound for the probability that $C$ consecutive nodes at distance $C^2 \log n$ to $C^2 \log n + C - 1$ all have greedy paths to $t$. Then, the maximum increase of *NextBestOnce* being $C$, nodes at distance $C^2 \log n + C$ or higher are not contacted anymore. The claim follows from Lemma 5.4.

Let $A$ denote the event that nodes at a distance exceeding $C^2 \log n + C$ to the set $\{v, t\}$, with $v$ being the first node contacted within distance $C$ of $t$, are not contacted after $v$

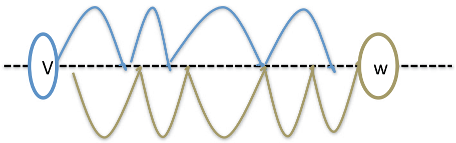

Fig. 3. Path of edges of length of at most $C$ originating from $v$ ($w$, respectively). A greedy path between the $v$ and $w$ exists, because the two paths intersect.

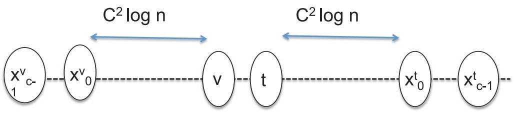

Fig. 4. Nodes $v$ and $t$ and the corresponding set $X$: No node at distance exceeding $C^2 \log n + C$ is contacted with high probability.

has been reached:

$$\begin{aligned}&\mathbb{E}(R_2^{NBO}(s,t))\\ &\quad = P(A)\cdot\mathbb{E}(R_2^{NBO}(s,t)|A) + (1-P(A))\cdot\mathbb{E}(R_2^{NBO}(s,t)|A^{\perp})\\ &\quad = P(A)\cdot\mathcal{O}(C^3\log n) + (1-P(A))\cdot\mathcal{O}(Cn).\end{aligned}$$

The last step follows from applying Lemma 5.4 to the subgraph of size $C^2 \log n$ as well as to the whole graph $G$. It remains to determine $P(A)$. If $v = t$, the claim holds. Otherwise, let $x_i^v$ for $i = 0, \ldots, C-1$ be the node such that $dist(x_i^v, t) = C^2 \log n + i$ and $dist(x_i^v, t) > dist(x_i^v, v)$. Analogously, $x_i^t$ denotes the node such that $dist(x_i^v, t) = C^2 \log n + i$ and $dist(x_i^t, v) > dist(x_i^v, t)$. The set $X = \{x_i^u : u \in \{v,t\}, 0 \le i < C\}$ consists of two sets of consecutive nodes at distance $C^2 \log n$ to $C^2 \log n + C - 1$ from the set $\{v, t\}$ (see Figure 4). If a node at higher distance than $C^2 \log n$ is reached after $v$, at least one node in $X$ needs to be on the path as well, because the maximal regression per hop is bound by $C$. Recall that *NextBestOnce marks* a node $u$ if all its neighbors closer to $t$ have been *marked*. It follows recursively that if a successor at a higher distance than the current node $u$ is chosen, all nodes reachable from $u$ by paths along which the distance to $t$ decreases monotonously have been *marked*. Consequently, a node $u$ with $dist(u,t) \ge C^2 \log n + C$ can only be on the path if all nodes $x \in X$ do *not* have a greedy path to $t$.

Lemma 5.3 is applied to bound $P(A)$ by the probability that all $2C$ nodes in $X$ have a greedy path to $t$, that is,

$$P(A) \ge P(\cap_{x\in X} g(t,x)) = \Omega\left(\left(1-\frac{1}{n}\right)^{2C}\right) = \Omega\left(1-\frac{2C}{n}\right).$$

The last step holds since $(1-x)^k \ge 1 - kx$ for $0 < x < 1$ and $k > 1$. Finally, we get

$$\mathbb{E}(R_2^{NBO}(s,t)) = \Omega\left(1-\frac{2C}{n}\right)\cdot\mathcal{O}(C^3\log n) + \mathcal{O}\left(\frac{2C}{n}\right)\cdot\mathcal{O}(Cn) = \mathcal{O}(C^3\log n). \quad \square$$

Theorem 5.1 is a direct consequence.

Proof. For a source-destination pair $(s,t)$ at distance $d = dist(s,t)$, the expected routing length of *NextBestOnce* is bound by

$$\begin{aligned}\mathbb{E}(R^{NBO}(s,t)) &= \mathbb{E}\big(R_1^{NBO}(s,t)\big) + \mathbb{E}\big(R_2^{NBO}(s,t)\big)\\ &= \mathcal{O}(\log^{\alpha-1} d \log\log d + C^3 \log n)\end{aligned}$$

by Lemmas 5.2 and 5.5. The distance between two nodes is at most $n/2$, so

$$\max_{s,t\in V}\mathbb{E}(R^{NBO}(s,t)) = \mathcal{O}(\log^{\alpha-1} n \log\log n + C^3\log n),$$

as claimed. □

### 5.2. Lower Bound

THEOREM 5.6. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$ with $C < \frac{1}{4}n^{1/4}$ and two nodes $s, t \in V$, a lower bound on the expected routing length of* NextBestOnce *is given by*

$$\frac{1}{n(n-1)} \sum_{s \neq t \in V} \mathbb{E}(R^{NBO}(s,t)) = \Omega(\log^{\alpha-1} n + C). \tag{12}$$

As for the upper bound, the proof is done by dividing the routing process into two phases. Let $R_1^{NBO}(s,t)$ be the number of nodes contacted to reach a node within distance $C$ and $R_2^{NBO}(s,t)$ be the number of steps needed to get from this node to $t$. The result for the first phase follows from the respective result for the lattice. The proof idea for the second phase is very similar to that of Theorem 4.1, but technically more demanding. We consider the case that the target node $t$ and its short-range neighbors only have long-range contacts at high distance. Thus, each of them is whp contacted by a short-range link. We show that it takes at least $C$ hops to find the correct link.

LEMMA 5.7. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$ and two nodes $s, t \in V$, the expected routing length for the first phase is*

$$\mathbb{E}\big(R_1^{NBO}(s,t)\big) = \Omega(\log^{\alpha-1} n).$$

The proof of Lemma 5.7 is very similar to the one presented in Fraigniaud and Giakkoupis [2009] for Theorem 2.4, giving the upper bound of standard greedy routing on a lattice with a scale-free degree distribution.

In order to analyze the second phase, a preparatory result about the number of neighbors at a given distance is needed. By this, we can bound the probability for a node $v$ to have a link to $t$ or its short-range neighbors.

LEMMA 5.8. *The expected number of nodes $Q$ in $V \setminus B_{\sqrt{n}}(t)$ that have a neighbor in $B_d(t)$ for any $d < \sqrt{n}$ is*

$$\mathbb{E}(Q) = \Omega\,(d).$$

PROOF. The claim follows from the fact that $P(l(u,v)|dist(u,v) = d) = \Theta(\frac{1}{d \log n})$ for any pair of nodes $(u, v)$, so that

$$\begin{aligned}\mathbb{E}(Q) &= \sum_{d_1=\sqrt{n}}^{n/2} \sum_{d_2=0}^{d} \Theta\left(\frac{1}{(d_1-d_2)\log n}\right) + \sum_{d_1=\sqrt{n}}^{n/2} \sum_{d_2=1}^{d} \Theta\left(\frac{1}{(d_1+d_2)\log n}\right)\\ &= \Omega\left(\sum_{d_1=\sqrt{n}}^{n/2} \sum_{d_2=1}^{d} \frac{2}{2d_1 \log n}\right) = \Omega\left(\sum_{d_1=\sqrt{n}}^{n/2} \frac{2d}{2d_1 \log n}\right) = \Omega\,(d)\,.\end{aligned}$$

The last step uses $\sum_{d_1=\sqrt{n}}^{n/2} \frac{1}{d_1 \log n} = \Omega(1)$ as shown in the proof of Lemma 3.1. □

We can now derive a lower bound on $R_2^{NBO}$.

LEMMA 5.9. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$ with $C < \frac{1}{4}n^{1/4}$ and two nodes $s, t \in V$, the expected routing length for the second phase is $\mathbb{E}(R_2^{NBO}(s,t)) = \Omega(C)$.*

PROOF. We first show that whp $t$ has few short-range neighbors, and all its long-range contacts are at high distance. Based on this event, the probability that routing takes at least $C$ more steps after reaching a node within distance $C$ of $t$ is constant.

Formally, we prove the claim that $P(R_2^{NBO}(s,t) \geq C/4|A) = \Omega(1)$ for a suitable event $A$ with $P(A) = \Omega(1)$. It follows directly that $\mathbb{E}(R_2^{NBO}(s,t))$ grows at least linearly with $C$. Recall that $t$ has at least two short-range neighbors $a_1^t$ and $b_1^t$ within distance $C$ of $t$. The set of short-range neighbors of a node $u$ is denoted by $SN(u)$, whereas $LN(u)$ is the set of long-range neighbors. Furthermore, we abbreviate $U = LN(t) \cup LN(a_1^t) \cup LN(b_1^t)$. The event $A = A_1 \cap A_2 \cap A_3 \cap A_4$ is the intersection of the following events:

—$A_1 = \{v \notin \{t, a_1^t, b_1^t\}\}$: the first contacted node within distance $C$ of $t$ is not $t$, $a_1^t$, or $b_1^t$.
—$A_2 = \{SN(t) = 2\}$: $a_1^t$ and $b_1^t$ are $t$'s only short-range neighbors.
—$A_3 = \{|LN(t)| \leq 1\} \cap \{|LN(a_1^t)| \leq 1\} \cap \{|LN(b_1^t)| \leq 1\}$: $t$ and its two short-range neighbors have maximally one long-range neighbor.
—$A_4 = \cup_{u \in U}\{dist(u,t) \geq \sqrt{n}\}$: $t$ and its short-range neighbors have only long-range neighbors at distance at least $\sqrt{n}$ to $t$.

Before showing that $P(A) = \Omega(1)$, note that indeed $P(R_2^{NBO}(s,t) \geq C/4|A) = \Omega(1)$. *NextBestOnce* increases the distance to $t$ by at most $C$ in each step; hence, by conditioning on $A_3$ (and recalling that $C \cdot C/4 < \sqrt{n}$), $t$, $a_1^t$, and $b_1^t$ cannot be contacted by a long-range neighbor in fewer than $C/4$ steps. Therefore, $t$ can only be found in fewer than $C/4$ steps if a node on the path contacts either $a_1^t$ or $b_1^t$ via a short-range link (by event $A_2$ and $A_3$). The probability that a node $u \in \{a_1^t, b_1^t\}$ is a short-range neighbor of a node $w$ on the routing path $X$ is

$$\begin{aligned} &P(u \in SN(w)|w \in X) \\ &\quad = P(u \in \{a_1^w, b_1^w\} \cup w \in \{a_1^u, b_1^u\}|w \in X) \\ &\quad \leq 2P(u \in \{a_1^w, b_1^w\}|dist(u,w) \leq C) = \frac{2}{C}. \end{aligned} \tag{13}$$

The second-to-last step holds because the probability that two nodes are short-range neighbors is maximal when their distance is at most $C$. Applying a union bound, the probability that none of the first $C/4$ nodes on the path after reaching $v$ has an edge to either $a_1^t$ or $b_1^t$ is bound by

$$P\big(R_2^{NBO}(s,t) \geq C/4|A\big) = \Omega\left(\left(1 - \frac{4}{C}\right)^{C/4}\right) = \Omega(1). \tag{14}$$

The last step holds, because $(1 - 1/x)^x$ converges to $1/e$ for $x \to \infty$.

It remains to show that $P(A) = \Omega(1)$. Using independence of edge selection, we can rewrite:

$$\begin{aligned} P(A) &= P(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= P(A_1|A_2 \cap A_3 \cap A_4)P(A_2)P(A_3)P(A_4|A_3). \end{aligned} \tag{15}$$

We now derive that each factor in Equation (15) has constant probability. The individual steps are a bit lengthy but only rely on basic properties about the model. Note that the event $A$ only considers the link selection and the probability is hence independent of the routing algorithm. For determining $P(A_1|A_2 \cap A_3 \cap A_4)$, we distinguish between short-range and long-range neighbors of nodes in $B_C(t)$. We define $W_L^d = \{w \in V \setminus B_d(t) : l(w, B_C(t))\}$, the set of all nodes with long-range links to a node at distance at least $d$ that have links into $B_C(t)$. Similarly, let $W_S = \{w \in V \setminus B_C(t) : SN(w) \cap B_C(t) \neq \emptyset\}$ be the set of nodes with short-range links into $B_C(t)$. Denote the predecessor of $v$ on the routing path by $w$. Note that when conditioning on $A_3$, $W_L^C = W_L^{\sqrt{n}}$. We consider the complement of $A_1$ to derive the desired bound

$$\begin{aligned}
&P\big(A_1^{\perp}|A_2 \cap A_3 \cap A_4\big)\\
&\quad\leq P\big(A_1^{\perp} \cap w \in W_L^C|A_2 \cap A_3 \cap A_4\big)\\
&\quad+ P\big(A_1^{\perp} \cap w \in W_S|A_2 \cap A_3 \cap A_4\big)\\
&\quad\leq P\big(A_1^{\perp} \cap w \in W_L^{\sqrt{n}}|A_2 \cap A_3 \cap A_4\big)\\
&\quad+ P\big(A_1^{\perp} \cap w \in W_S|A_2 \cap A_3 \cap A_4\big)\\
&\quad= P\big(A_1^{\perp}|A_2 \cap A_3 \cap A_4 \cap w \in W_L^{\sqrt{n}}\big)\\
&\quad\cdot P\big(w \in W_L^{\sqrt{n}}|A_2 \cap A_3 \cap A_4\big)\\
&\quad+ P\big(A_1^{\perp}|A_2 \cap A_3 \cap A_4 \cap w \in W_S\big)\\
&\quad\cdot P\big(w \in W_S|A_2 \cap A_3 \cap A_4\big)\\
&\quad\leq P\big(A_1^{\perp}|A_2 \cap A_3 \cap A_4 \cap w \in W_L^{\sqrt{n}}\big)\\
&\quad+ P\big(A_1^{\perp}|A_2 \cap A_3 \cap A_4 \cap w \in W_S\big)\\
&\quad= \mathcal{O}\left(\frac{3}{C}\right) + \mathcal{O}\left(\frac{2}{C}\right).
\end{aligned} \tag{16}$$

The first summand in Equation (16) is derived by first dropping condition $A_2$ because the short-range links of $t$ do not influence $A_1$ given $w \in W_L^{\sqrt{n}}$. By Lemma 5.8, there are $\Omega(C)$ nodes in $V \setminus B_{\sqrt{n}(t)}$ with edges into $B_C(t)$. Conditioning on $A_3$ and $A_4$, at most three of these long-range links are incidents to $t$, $a_1^t$, and $b_1^t$. The second summand $\frac{2}{C}$ in Equation (16) is derived as in Equation (13). Note that $A_3$ and $A_4$ do not influence the event, given that $w \in W_S$. Consequently,

$$P(A_1|A_2 \cap A_3 \cap A_4) = 1 - \mathcal{O}\left(\frac{3}{C} + \frac{2}{C}\right) = \Omega\,(1)\,.$$

$P(A_2)$ corresponds to the probability that none of the $2(C-1)$ potential short-range neighbors but $a_1^t$, $b_1^t$ have chosen $t$ as a neighbor. So

$$P(A_2) = \left(1 - \frac{1}{C}\right)^{2(C-1)} = \Omega(1).$$

Similarly to Equation (14), the last bound follows from $(1-1/x)^{2x} \to e^{-2}$. Long-range edges are selected independently, and hence,

$$P(A_3) = P(|LN(t)| \leq 1)^3 = \Omega\,(1)\,.$$

The last step follows by conditioning on $S_\alpha = 1$ and Equation (2).

For calculating the last factor $P(A_4|A_3)$, denote the long-range neighbor of $t$, $a_1^t$, $b_1^t$ by $t_l$, $a_l$, and $b_l$, respectively. It follows from $dist(a, a_l) > 2\sqrt{n}$ that $dist(t, a_l) > \sqrt{n}$ because $C < \sqrt{n}$, and hence,

$$P(A_4|A_3) = (P(dist(t_l, t) > 2\sqrt{n}|l(t_l, t)))^3.$$

Now, $P(A_4|A_3) = \Omega\,(1)$ is a direct consequence from Lemma 3.1. The previous results confirm that indeed,

$$P(A) = P(A_1|A_2 \cap A_3 \cap A_4)P(A_2)P(A_3)P(A_4|A_3) = \Omega(1).$$

Thus, we have shown that

$$P\big(R_2^{NBO}(s,t) \geq C/4\big) \geq P\big(R_2^{NBO}(s,t) \geq C/4|A\big)P(A) = \Omega(1).$$

Consequently, the expected routing length grows at least linearly in $C$ as claimed, that is, $\mathbb{E}(R_2^{NBO}(s,t)) = \Omega(C)$. □

Theorem 5.6 follows from Lemmas 5.7 and 5.9 because

$$\mathbb{E}(R^{NBO}(s,t)) = \mathbb{E}(R_1(s,t)) + \mathbb{E}(R_2(s,t)) = \Omega(\log^{\alpha-1} n + C).$$

We have provided upper and lower bounds on the expected routing length of *NextBestOnce*. Theorems 5.1 and 5.6 assume a scale-free degree distribution. Because some Darknets such as MCON artificially restrict the degree to reduce the dependency on central nodes [Vasserman et al. 2009], this assumption might not be valid in practice. If the maximal degree is indeed bound by a constant $K$, the first phase of the routing is bound by the results for Kleinberg's original model, which provide an upper bound of $\mathcal{O}(\log^2 n)$ [Kleinberg 2000] and $\Omega(\log^2 n)$ [Martel and Nguyen 2004]. Note that the lower bound is stated for a degree of 1 but the proof can be trivially extended to any constant bound $K$. Our bound for the second phase of the routing is independent of the degree distribution, such that for a constant maximal degree $K$, the routing length of *NextBestOnce* is asymptotically bound by $\mathcal{O}(\log^2 n + C^3 \log n)$ and $\Omega(\log^2 n + C)$.

In the next section, we show that including information about neighbors of neighbors increases the performance by more than a constant factor. More precisely, we show that the upper bound on *NextBestOnce-NoN* is strictly better than the lower bound on *NextBestOnce* if $C$ is small.

## 6. ANALYSIS OF NEXTBESTONCE-NON

In this section, we provide an upper bound on the expected routing length of *NextBestOnce-NoN* as a function of the scale-free degree distribution's exponent $\alpha$.

THEOREM 6.1. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$, an upper bound on the maximal expected routing length of* NextBestOnce-NoN *is given by*

$$\max_{s,t\in V} \mathbb{E}(R^{NoN}(s,t)) = \mathcal{O}\big(\log^{\delta(\alpha)(\alpha-1)} n \log\log n + C^3 \log n\big) \tag{17}$$

$$\text{for } \delta(\alpha) = 1 - \frac{(\alpha-2)(3-\alpha)}{\alpha}.$$

As for the earlier proofs, the routing is divided into two phases $R_1^{NoN}(s,t)$ and $R_2^{NoN}(s,t)$. For the first phase, we determine a lower bound on the probability to halve the distance of the currently contacted node's closest neighbor to $t$ in the next two steps. The distance of the closest neighbor is used because it is decreasing during the routing in the first phase in contrast to the distance of the current message holder. This monotonicity then allows us to apply well-known results about decreasing integer-valued random processes. The probability of halving the distance is obtained by (1) considering the probability of contacting a high-degree neighbor and (2) the probability that such a neighbor has a neighbor at half its distance to the target. The bound for the second phase can easily be derived from the bound on *NextBestOnce* in Theorem 5.1. In the end, the transition point between the two phases is determined as the result of an extremal value problem. The expected routing length is then the minimal value of said extremal value problem.

Formally, we first fix $1/2 \leq r \leq 1$ and $0 \leq k \leq \alpha - 2$. $R_1^{NoN}(s,t)$ gives the number of steps needed to get within distance $e^{\log^r n}$ of $t$. $R_2^{NoN}(s,t)$ is the number of steps to cover the remaining distance. For the proof, we assume that the maximum value of $S_\alpha$

is $\mu = \Theta(\log n)$. Restricting the degree is obviously a relaxation, which avoids further case distinctions.

We now derive the expected length of the second phase.

LEMMA 6.2. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$; two nodes $s, t \in V$; and $1/2 \leq r \leq 1$, the expected routing length of* NextBestOnce-NoN *after reaching a node within distance $e^{\log^r n}$ of $t$ is*

$$\mathbb{E}\big(R_2^{NoN}(s,t)\big) = \mathcal{O}\big(\log^{r(\alpha-1)} n \log\log n + C^3 \log n\big).$$

PROOF. *NextBestOnce-NoN* is in expectation at least as fast as *NextBestOnce* using the same procedure, only with additional information. Let $u$ be the first node on the routing path with $dist(t,u) \leq e^{\log^r n}$. By Theorem 5.1, the expected routing length to get from $u$ to $t$ is

$$\begin{aligned}\mathbb{E}\big(R_2^{NoN}(s,t)\big) &= \mathcal{O}\big(\mathbb{E}\big(R^{NBO}(u,t)\big)\big)\\ &= \mathcal{O}(\log^{\alpha-1} e^{\log^r n} \log\log e^{\log^r n} + C^3 \log n)\\ &= \mathcal{O}\big(\log^{r(\alpha-1)} n \log\log n + C^3 \log n\big).\end{aligned}$$

This proves the claim. □

We now bound the number of hops needed during the first phase. The proof is divided into three parts: We first determine the probability that nodes of a certain degree are neighbors, because contacting nodes of a sufficiently high degree is essential. Afterward, we use that result to determine the probability that the distance is halved by considering the following scenario: the current node contacts a neighbor of degree at least $\log^k n$, which has a neighbor of degree at least $\log n$. This node then has a neighbor within half the distance of the first neighbor to $t$. In the last step, we compute the expected routing length for the first phase based on the aforementioned bound. In the following, assume $C < e^{log^{1/2} n}$. Otherwise, $\mathbb{E}(R^{NoN}(s,t)) = \mathcal{O}(C^3 \log n)$ holds by Theorem 5.1.

LEMMA 6.3. *Consider a node $u$ with $d = dist(u,t) > e^{\log^{1/2} n}$ and a set $W \subset V$, so that $dist(r,t) > d$ for all $w \in W$. Denote by $V_{d'}^a = \{v \in V : v \in B_{d'}(t), l_v \geq a\}$ the set of all nodes within distance $d' \leq d$ of the destination and label at least $a$. Furthermore, assume $\frac{|W|\mu}{(d-d')\log n} < 1/2$, $\mu$ being the maximum degree. The probability that $u$ is adjacent to a node in $V_{d'}^a$, conditioned on $l_u$ and the absence of edges between $W$ and $B_{d'}(t)$, is bound by*

$$P\left(l(u, V_{d'}^a) | l_u = l \cap l(W, B_{d'}(t))^{\perp}\right) = \Omega\left(\frac{l}{\log n}[\log(d+d'-1) - \log(d-d'+1)]a^{2-\alpha}\right).$$

PROOF. We first show that it suffices to derive the expected number of neighbors of $u$ in $V_{d'}^a$. Then, we determine said expected value as a sum of the probabilities of all nodes in $V_{d'}^a$ to be adjacent to $u$. Each summand can then be derived based on elementary properties of the model.

We start by showing that the result follows from determining the expected number $Q$ of nodes $v$ in $V_{d'}^a$ that are adjacent to $u$. For each $v \in B_{d'}(t)$, the random variable $Q_v$ is 1 if $v \in V_{d'}^a$ and adjacent to $u$. Otherwise, $Q_v$ is 0. We need to derive a lower bound on $P(Q=1)$ for the sum $Q = \sum_{v \in B_{d'}(t)} Q_v$. It holds that $P(Q=1) \geq 1 - e^{-E(Q)}$ because

$$P(Q=0) = \prod_{v \in B_{d'}(t)} P(Q_i = 0) = \prod_{v \in B_{d'}(t)} (1 - E(Q_i)) \leq \prod_{v \in B_{d'}(t)} e^{-E(Q_i)} = e^{-E(Q)}.$$

The inequality in the second-to-last step follows from $1 - x \leq e^{-x}$ for $x \in [0, 1]$. For small $E(Q)$, we have $1 - e^{-E(Q)} = \Theta(E(Q))$. Hence, a lower bound on $P(Q = 1)$ can be obtained from deriving $\mathbb{E}(Q)$.

In the following, $E(Q)$ is computed by summarizing over all nodes in $B_{d'}(t)$. Note that a node within distance $d' - 1$ to $t$ has a distance between $d - d' + 1$ and $d + d' - 1$ to $u$. Denote the event $\{l_u = l \cap l(W, B_{d'}(t))^{\perp} \cap v \in B_{d'}(t) \cap dist(v, u) = i\}$ by $E_i$, so that

$$\begin{aligned} E(Q) &= \sum_{i=d-d'+1}^{d+d'-1} P(l(u, v) \cap v \in V_{d'}^{a} | E_i) \\ &= \sum_{i=d-d'+1}^{d+d'-1} \sum_{j=a}^{\infty} P(l(u, v) \cap l_v = j | E_i) \\ &= \sum_{i=d-d'+1}^{d+d'-1} \sum_{j=a}^{\infty} P(l_v = j | E_i) P(l(u, v) | E_i \cap l_v = j). \end{aligned} \tag{18}$$

We abbreviate $t_{i,j}^{1} = P(l_v = j | E_i)$ and $t_{i,j}^{2} = P(l(u, v) | E_i \cap l_v = j)$. We now show $t_{i,j}^{1} = \Omega(P(l_v = j)) = \Omega(j^{-\alpha})$. By substituting $E_i$ and using $P(A_1 | A_2) = P(A_2 | A_1) \frac{P(A_1)}{A_2}$ and $P(A_1 \cap A_2) = P(A_1 | A_2) P(A_2)$ for all events $A_1, A_2$ with nonzero probability, we get

$$\begin{aligned} t_{i,j}^{1} &= P(l_v = j | l(W, v)^{\perp} \cap v \in B_{d'}(t) \cap dist(v, u) = i) \\ &= \frac{P(l(W, v)^{\perp} | l_v = j \cap v \in B_{d'}(t) \cap dist(v, u) = i)}{P(l(W, v)^{\perp} | v \in B_{d'}(t) \cap dist(v, u) = i)} P\left(l_v = j\right). \end{aligned}$$

It remains to show that the numerator in the last line can be bound by $\Omega(1)$ and hence, since the denominator is at most 1, we get

$$\frac{P(l(W, v)^{\perp} | l_v = j \cap v \in B_{d'}(t) \cap dist(v, u) = i)}{P(l(W, v)^{\perp} | v \in B_{d'}(t) \cap dist(v, u) = i)} = \Omega(1),$$

and due to the scale-free degree distribution,

$$t_{i,j}^{1} = \Omega(j^{-\alpha}).$$

Recall from Equation (1) in Section 3.1 that two nodes $v$, $w$ are adjacent with probability

$$\begin{aligned} &P(l(v, w) | dist(v, w) = i \cap l_v = l_1 \cap l_w = l_2) \\ &\quad = 1 - e^{-\frac{l_1 l_2}{i\gamma}} = \Theta\left(\frac{l_1 l_2}{i\gamma}\right) = \Theta\left(\frac{l_1 l_2}{i \log n}\right). \end{aligned} \tag{19}$$

The last steps holds due to Equation (2). The probability that a node $v \in B_{d'}$ is adjacent to a node in $W$ is obtained by a union bound. We abbreviate $F_i = \{dist(t, W) > d \cap v \in B_{d'}(t) \cap dist(v, u) = i\}$ to obtain

$$\begin{aligned} &P(l(v, W) | l_v = j \cap F_i) \\ &\quad \leq \sum_{w \in W} P(l(v, w) | l_v = j \cap F_i) \\ &\quad \leq \sum_{w \in W} P(l(v, w) | dist(v, w) > d - d' \cap l_v = j \cap dist(v, u) = i) \\ &\quad = \Theta\left(\frac{j|W|}{(d - d') \log n}\right). \end{aligned}$$

The last step holds by Equation (19). By assumption, $\frac{|W|\mu}{(d-d')\log n} < 1/2$, and hence indeed,

$$P(l(W,v)^{\perp}|l_v = j \cap v \in B_{d'}(t) \cap dist(v,u) = i) = \Omega(1).$$

This completes the derivation of $t^1_{i,j} = \Omega(j^{-\alpha})$.

Since edges are chosen independently, the event $l(W, B_{d'}(t))^{\perp}$ does not influence $t^2_{i,j}$. So

$$t^2_{i,j} = P(l(u,v)|l_u = l \cap l_v = j \cap dist(u,v) = i) = \Theta\left(\frac{l \cdot j}{i\gamma}\right)$$

is a consequence of Equation (19). Substituting $t^1_{i,j}$ and $t^2_{i,j}$ in Equation (18), we obtain the desired result:

$$\begin{aligned} E(Q) &= \sum_{i=d-d'+1}^{d+d'-1} \sum_{j=a}^{\infty} t^1_{i,j} \cdot t^2_{i,j} \\ &= \sum_{i=d-d'+1}^{d+d'-1} \sum_{j=a}^{\infty} \Omega\left(\frac{l}{i \log n} j^{1-\alpha}\right) \\ &= \Omega\left(\frac{l}{\log n}[\log(d+d'-1) - \log(d-d'+1)]a^{2-\alpha}\right). \end{aligned}$$

We have now given a lower bound on the expected number of neighbors. Hence, the probability to have at least one such neighbor is indeed as claimed. □

In the following, we model the routing process as an integer-valued random process $X_1, X_2, \ldots$, such that $X_i$ gives the distance of the closest neighbor of the $i$th node on the path to $t$. The distance of the closest neighbor to $t$ decreases in each step until a node within distance $C$ is reached. Let $Z_i$ denote the set of all nodes on the path before the $i$th node and their neighbors. All events need to be conditioned on the fact that all nodes within distance $d = X_i$ are not adjacent to a node in $Z_i$, that is, the event $l(B_d, Z_i)^{\perp}$. The next result is the main part of the proof, because it enables us to derive the expected routing length from a common result about integer-valued decreasing processes.

LEMMA 6.4. *Let $X_i$ be the distance of the closest neighbor of the ith node on the routing path, and let*

$$1/2 \le r \le 1, \quad 0 \le k \le \alpha - 2, \quad |Z_i| < 1/2\sqrt{d}\log n.$$

*The probability that the distance to t is halved in the next two steps is*

$$P\left(X_{i+2} \le \frac{d}{2}|Z_i \cap X_i = d\right) = \Omega\left(\frac{\log d \cdot \log^{r+k(3-\alpha)} n}{\log^{\alpha} n}\right).$$

PROOF. Let $u$ be the $i$th node on the path. We first condition that the distance between $u$ and the neighbor closest to $t$ is not too small (in order to apply Lemma 6.3 with a sufficiently high $d - d'$) or too large (so that the decrease in distance is not so unlikely). We then derive the result by distinguishing two cases: $l_u < \log^k n$ and $l_u \ge \log^k n$. For the first case, we design four events that lead to halving the distance and derive their probability. For the second case, we design two events and point out that their probability is at least as high as for the first four events.

We start by conditioning on the distance of $u$ to $t$. The distance $\Delta$ of $u$ is not captured by the random process $X_i$, which only provides the distance of a neighbor of $u$. In the following, we condition on the event $G = \{d + \sqrt{d} \leq \Delta \leq 2d\}$, which we show to be of constant probability and which allows us to determine the probability to halve the distance. The first inequality in the definition of $G$ is necessary to apply Lemma 6.3 with $\frac{|Z_i|}{(d+\sqrt{d}-d)\log n} < 1/2$. The bound $\Delta \leq 2d$ ensures that $dist(u, t)$ needs to be at most quartered to have $X_{i+2} \leq d/2$. For a lower bound on the event $A$ of halving the distance, $P(A) \geq P(A|G)P(G)$ can be applied. If $P(G) = \Omega(1)$, $P(A) = \Omega(P(A|G))$ holds. It remains to show $P(G) = \Omega(1)$. The lower bound $\Delta \geq d + \sqrt{d}$ holds with probability $\Theta(1)$ by Lemma 3.1. The upper bound $\Delta \leq 2d$ holds with probability $\Omega(1)$ as well, as can be seen from the proof of Theorem 2.4 in Fraigniaud and Giakkoupis [2009]: the probability that an arbitrary node has a neighbor at half its distance to the destination is shown to be $\mathcal{O}(\frac{1}{\log^{\epsilon} n})$ for some $\epsilon > 0$. Thus, the probability of not having such a neighbor is $\Omega(1 - \frac{1}{\log^{\epsilon} n}) = \Omega(1)$, because $\frac{1}{\log^{\epsilon} n} < 1/2$ for $n$ big enough. So indeed, $P(G) = \Omega(1)$.

Assume $l_u < \log^k n$. The following events result in $X_{i+2} \leq d/2$:

—A neighbor $v \in B_\Delta(t)$ of $u$ has label $l_v \geq \log^k n$ .
—$v$ has a neighbor $w \in B_\Delta(t)$ with label $l_w \geq \log n$.
—$w$ has a link into $B_{d/2}$.
—$v$ is the node $u$ chooses as the next node on the routing path; denote this event by $\{Z = v\}$.

All events are conditioned on $F = l(B_d, Z_i)^\perp) \cap l_u \leq \log^k n \cap G$. Formally, the probability is determined by

$$\begin{aligned}
&P\big(l\big(u, V_\Delta^{\log^k n}\big) \cap l\big(v, V_\Delta^{\log n}\big) \cap l(w, B_{d/2}) \cap Z = v|F\big)\\
&\quad = P\big(l\big(u, V_\Delta^{\log^k n}\big)|F\big) \cdot P\big(l\big(v, V_\Delta^{\log n}\big)|l\big(u, V_\Delta^{\log^k n}\big) \cap F\big)\\
&\quad\quad \cdot P\big(l(w, B_{d/2})|l\big(v, V_\Delta^{\log n}\big) \cap l\big(u, V_\Delta^{\log^k n}\big) \cap F\big)\\
&\quad\quad \cdot P\big(Z = v|l(w, B_{d/2}) \cap l\big(v, V_\Delta^{\log n}\big) \cap l\big(u, V_\Delta^{\log^k n}\big) \cap F\big)\\
&\quad := q_1 q_2 q_3 q_4. \qquad (20)
\end{aligned}$$

We now subsequently derive $q_1$, $q_2$, $q_3$, and $q_4$. $q_1$, the probability that $u$ has a link to a node of degree at least $v$ within distance $\Delta$ of $t$, can be derived using Lemma 6.3 with $d = d' = \Delta$. Note that the probability of having a link is minimal for a node $u$ with $l_u = 1$, so that

$$\begin{aligned}
q_1 &= P\big(l\big(u, V_\Delta^{\log^k n}\big)|l(B_d, Z_i)^\perp \cap l_u \leq \log^k n \cap G\big)\\
&\geq P\big(l\big(u, V_\Delta^{\log^k n}\big)|l(B_d, Z_i)^\perp \cap l_u = 1 \cap G\big)\\
&= \Omega\left(\frac{1}{\log n}[\log(2\Delta - 1) - 0] \log^{k(2-\alpha)} n\right)\\
&= \Omega\left(\frac{\log d}{\log n} \log^{k(2-\alpha)} n\right).
\end{aligned}$$

The last step uses $d \leq \Delta \leq 2d$. Since links are selected independently, the events $l(u, V_\Delta^{\log^k n})$ and $l(v, V_\Delta^{\log n})$ are independent. Furthermore, because labels are selected independently, $l_u \leq \log^k n$ does not influence $q_2$ or $q_3$. Hence, $q_2$, the probability that $v$ has a neighbor of degree at least $\log n$ within distance $\Delta$ to $t$, is derived similarly to $q_1$

$$q_2 = \Omega\left(\frac{\log d}{\log n} \log^k n \log^{(2-\alpha)} n\right).$$

To determine $q_3$, Lemma 6.3 can be applied because $B_{d/2} = V^1_{d/2}$, Furthermore, the function $\log(x + d/2) - \log(x - d/2)$, being a monotonously decreasing function for $x > d/2$, assumes its minimum in the interval $[d, 2d]$ at $\Delta = 2d$, so that

$$\begin{aligned} q_3 &= P(l(w, B_{d/2})|l(B_d, Z_i)^{\perp} \cap l_w \geq \log n \cap G) \\ &\geq P(l(w, B_{d/2})|l(B_d, Z_i)^{\perp} \cap l_w = \log n \cap G) \\ &= \Omega\left(\frac{\log n}{\log n}[\log(\Delta + d/2 - 1) - \log(\Delta - d/2 + 1)]\right) \\ &= \Omega(\log(2d + d/2 - 1) - \log(2d - d/2 + 1)) = \Omega\,(1)\,. \end{aligned}$$

If u has $\mathcal{O}(\log^k n)$ neighbors at distance at least d from t, the probability that $u$ has a link to a certain node is asymptotically at most as high as the probability that one of $\log^k n$ arbitrary nodes have such a neighbor. Hence, the probability for the event $L_1$ that $v$ has a neighbor closest to $t$ and the event $L_2$ that any of the remaining neighbors has the closest neighbor to $t$ are of the same order, that is, $q_4 = \Omega\,(1)$.

Combining the results for the individual terms, we get a bound for halving the distance in case of $l_u \leq \log^k n$:

$$\begin{aligned} &P(X_{i+2} \leq \frac{d}{2}|Z_i \cap X_i = d) \\ &= \Omega\,(q_1 q_2 q_3 q_4) \\ &= \Omega\left(\frac{\log d \log^{k(3-\alpha)} n}{\log^{\alpha} n} \log d\right) \\ &= \Omega\left(\frac{\log d \log^{k(3-\alpha)} n}{\log^{\alpha} n} \log^r n\right). \end{aligned}$$

The last step uses that $\log d > \log e^{\log^r n} = \log^r n$. If $l_u \geq \log^k n$, we only need to consider the event that (1) $u$ has a neighbor $w$ within distance $\Delta$ of $t$ with degree at least $\log n$, and (2) $w$ has a neighbor in $B_{d/2}(t)$. This corresponds to the second and third events for $l_u \leq \log^k n$. So, they are already bound by $q_2$ and $q_3$, respectively, and the bound for $l_u \leq \log^k n$ holds for $l_u \leq \log^k n$ as well. So, we can halve the distance in one step with probability $\Omega(\frac{\log d}{\log n} \log^k n \log^{(2-\alpha)} n)$. As a consequence, we indeed obtain the claimed result:

$$P\left(X_{i+2} \leq \frac{d}{2}|Z_i \cap X_i = d\right) = \Omega\left(\frac{\log d \log^{k(3-\alpha)} n}{\log^{\alpha} n} \log^r n\right). \quad \square$$

Now, the expected length of the first phase can be bound.

LEMMA 6.5. *For a graph $G = (V, E) \in \mathcal{D}(n, 1, C, S_\alpha)$ with $C < e^{\log^{1/2} n}$, and two nodes $s, t \in V$, the expected routing length of* NextBestOnce-NoN *to reach a node within distance $e^{\log^r n}$ of $t$ is*

$$\mathbb{E}\big(R_1^{NoN}(s, t)\big) = \mathcal{O}\big(\log^{\alpha - r - k(3-\alpha)} n \log\log n\big) \tag{21}$$

*for all $1/2 \leq r \leq 1$ and $0 \leq k \leq \alpha - 2$.*

PROOF. We show that with high probability, $|Z_i|$ is small enough to apply Lemma 6.4 and then obtain the desired bound on the first phase. By Lemma 6.4, the probability to halve the distance during the next two steps is given by

$$P\left(X_{i+2} \leq \frac{d}{2}|X_1, X_2, \ldots, X_i = d\right) = \Omega\left(\frac{\log d \cdot \log^{r+k(3-\alpha)} n}{\log^{\alpha} n}\right) = \Omega\left(\frac{\log d}{\log^{\alpha-r-k(3-\alpha)} n}\right) \quad (22)$$

as long as $d > e^{\log^r n}$ and $|Z_i| < \sqrt{d}/2$. The latter holds with probability at least $1 - \frac{1}{n}$, as can be seen from the proof for the upper bound of the standard greedy algorithm on the lattice [Fraigniaud and Giakkoupis 2009], which is also applicable for the first routing phase of *NextBestOnce-NoN*. It is shown that routing needs at most $\mathcal{O}(\log^3 n)$ steps with probability $\Omega(1 - 1/n)$. Since we assume the maximal degree $\mu = \theta(\log n)$ to be bound logarithmically, $|Z_i| \leq K \log^4 n$ for some constant $K$ follows. Lemma 5.2 from Fraigniaud and Giakkoupis [2009] gives the expected number of steps necessary for an integer-valued decreasing random process defined by Equation (22) to reach a value $\lambda = e^{\log^r n}$, resulting in

$$\begin{aligned}
&\mathbb{E}\big(R_1^{NoN}(s,t)\big)\\
&= P(|Z_i| \leq K \log^4 n)\mathbb{E}\big(R_1^{NoN}(s,t) \; ||Z_i| \leq K \log^4 n\,\big)\\
&\quad + (1 - P(|Z_i| \leq K \log^4 n))\mathbb{E}\big(R_1^{NoN}(s,t) \; ||Z_i| > K \log^4 n\,\big)\\
&= \mathcal{O}\big(\log^{\alpha-r-k(3-\alpha)} n \log\log n\big) + \mathcal{O}\left(\frac{1}{n}\right)\mathcal{O}(n)\\
&= \mathcal{O}\big(\log^{\alpha-r-k(3-\alpha)} n \log\log n\big).
\end{aligned}$$

The second-to-last step holds since $e^{\log^r n} > C$, so that the distance is guaranteed to decrease in each step. As a consequence, at most $n/2 - e^{\log^r n} = \mathcal{O}(n)$ hops are needed to complete the first phase. □

Theorem 6.1 can now be shown solving a two-dimensional extremal value problem.

Proof. It follows from Lemmas 6.5 and 6.2 that for all $(k, r) \in [0, \alpha - 2] \times [1/2, 1]$,

$$\mathbb{E}(R^{NoN}(s,t)) = \mathcal{O}\big(\log^{\alpha-r-k(3-\alpha)} n \log\log n + \log^{r(\alpha-1)} n \log\log n + C^2 \log^{\epsilon} n\big)$$

Since the last summand does not depend on $r$ or $k$, our minimal bound can be found as the minimum of the function

$$f(k,r) = \log^{\alpha-r-k(3-\alpha)} n + \log^{r(\alpha-1)} n. \quad (23)$$

Computing the gradient of f gives

$$Df = \begin{pmatrix} -(3-\alpha)(\log n)^{\alpha-r_{min}-k_{min}(3-\alpha)} \\ -(\log n)^{\alpha-r_{min}-k_{min}(3-\alpha)} + (\alpha-1)(\log n)^{r_{min}(\alpha-1)} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

So the function $f$ takes its minimum on the border of the $[0, \alpha - 2]x[1/2, 1]$, for example, if either $k_{min} = 0$, $k_{min} = \alpha - 2$, $r_{min} = 0.5$, or $r_{min} = 1$.

When $k_{min} = 0$, a node of degree at least 1 needs to be contacted first. This leads essentially to the same scenario used to obtain the bound for *NextBestOnce*, and cannot have an improved complexity. The same goes for the case $r_{min} = 1$, because $e^{\log^1 n} = n$, so only the second phase, for which the complexity is bounded by that of *NextBestOnce,* is considered. As for $r_{min} = 0.5$, observe the exponent of the first summand of f:

$$\alpha - 0.5 - k \cdot (3 - \alpha) \geq \alpha - 0.5 - (\alpha - 2) \cdot (3 - \alpha) > 1.$$

The last step uses $2 < \alpha < 3$, and thus $\alpha - 2 < 1$, $3 - \alpha < 1$, and at least one of the two factors is maximally 0.5. So, an improved bound with regard to *NextBestOnce* can only be obtained for $k_{min} = \alpha - 2$.

We determine $r_{min}$ by minimizing:

$$g(r) = (\log n)^{\alpha-(\alpha-2)(3-\alpha)-r} + (\log n)^{r(\alpha-1)}.$$

The first derivative of g is

$$g'(r) = -\log\log n(\log n)^{\alpha-(\alpha-2)(3-\alpha)-r} + (\alpha-1)\log\log n(\log n)^{r(\alpha-1)}.$$

Setting $g'(r_{min}) = 0$, we get that

$$\begin{aligned}(\alpha-1)\log\log n(\log n)^{r_{min}(\alpha-1)} &= \log\log n(\log n)^{\alpha-(\alpha-2)(3-\alpha)-r_{min}}\\ (\log n)^{r_{min}(\alpha-1)+r_{min}} &= \frac{1}{\alpha-1}(\log n)^{\alpha-(\alpha-2)(3-\alpha)}\\ (\log n)^{r_{min}^{\alpha}} &= \frac{1}{(\alpha-1)}(\log n)^{\alpha-(\alpha-2)(3-\alpha)}\\ (\log n)^{r_{min}} &= \frac{1}{(\alpha-1)^{1/\alpha}}(\log n)^{1-\frac{(\alpha-2)(3-\alpha)}{\alpha}}.\end{aligned}$$

Finally, we get

$$\begin{aligned}r_{min} &= \frac{\log\frac{1}{(\alpha-1)^{1/(\alpha)}} + (1-\frac{(\alpha-2)(3-\alpha)}{\alpha})\log\log n}{\log\log n}\\ &= \frac{\log\frac{1}{\alpha-1)^{1/(\alpha)}}}{\log\log n} + \left(1-\frac{(\alpha-2)(3-\alpha)}{\alpha}\right).\end{aligned} \tag{24}$$

This is indeed a minimum since

$$g''(r_{min}) = (\log n)^{\alpha-(\alpha-2)(3-\alpha)-r_{min}} + (\alpha-1)^2(\log n)^{r_{min}(\alpha-1)} > 0.$$

Consider that for $a = \log\frac{1}{(\alpha-1)^{1/(\alpha)}}$,

$$(\log n)^{\frac{a}{\log n}} = 2^{\frac{a\log n}{\log n}} = 2^a.$$

By this, the first summand in Equation (24) does not contribute to the asymptotic complexity. By the second summand, the asymptotic value of $r^*_{min}$ of $r_{min}$ is

$$r^*_{min} = 1 - \frac{(\alpha-2)(3-\alpha)}{\alpha}$$

for the routing bound in Theorem 6.1.

The upper bound on *NextBestOnce-NoN* is then obtained as

$$\begin{aligned}&\mathcal{O}(f(k_{min}, r_{min})\log\log n + C^3\log n)\\ &\quad= \mathcal{O}(\log^{\delta(\alpha)(\alpha-1)} n\log\log n + C^3\log n)\end{aligned}$$

for $\delta(\alpha) = 1 - \frac{(\alpha-2)(3-\alpha)}{\alpha}$. This completes the remaining steps in the proof of Theorem 6.1. □

We have now proven that using neighbor-of-neighbor information decreases the expected routing length to roughly the $\delta(\alpha)$-th ($<1$) power. So, including neighbor-of-neighbor information can significantly decrease the routing length in undirected scale-free graphs despite a constant average degree. As for the bounds in Section 5, our proof relies on the existence of a scale-free degree distribution. If this assumption is not met, the first phase of the routing can be bound by $\mathcal{O}(\log^2 n)$ by Kleinberg

[2000], and the total expected routing length is bound by $\mathcal{O}(\log^2 n + C^3)$. In contrast to a scale-free degree distribution, a bound degree distribution should not result in a lower asymptotic bound for *NextBestOnce-NoN* than for *NextBestOnce*. Because the expected number of neighbors of neighbors is bound by a constant $K^2$, the expected routing length for *NextBestOnce-NoN* cannot be less than the expected routing length for *NextBestOnce* with a constant bound of $K^2$ on the degree. So, the asymptotic advantage of *NextBestOnce-NoN* can only be shown for a nonconstant degree distribution, though we can expect an improvement by a factor for a constant degree as well. For this reason, we here focus on scale-free degree distributions, which are exhibited in social graphs and apply to Darknets without degree restrictions.

## 7. CONCLUSION

Darknets provide privacy by design by (1) limiting overlay connections to trusted contacts, (2) obfuscating communicating parties, and (3) encrypting all communication. In this manner, Darknets offer protection against governmental and industrial parties collecting personal data. However, deployed Darknets fail to provide an adequate performance. Due to their restricted connectivity and inherent topology obfuscation, Darknets cannot be structured using common techniques such as DHTs. In addition, the conventional methods for analyzing routing performance in distributed systems cannot be applied to provide performance bounds. The consequential lack of a suitable model for Darknet topologies has severely impeded the assessment and improvement of Darknet routing techniques.

We thus presented three major contributions to the design and analysis of Darknets. First, we extended Kleinberg's model to account for the inherent inaccuracy of Darknet embeddings. Second, we analyzed the Freenet routing algorithm and found that it does not achieve polylog routing length. The result fortifies the experimental validation in previous work that Freenet routing is frequently slow or even unsuccessful [Vasserman et al. 2009]. Hence, our third contribution is the design and analysis of *NextBestOnce**, a generic algorithm with polylog routing length under weak assumptions on the accuracy of the underlying embedding. We gave concrete bounds on the efficiency of two of its variants in terms of the embedding accuracy.

A direct consequence of the model is the need of a polylog embedding accuracy $C$ for polylog routing. In the future, we thus aim to focus on the embedding accuracy. For this purpose, we identified two important lines of work. First, we plan to derive upper and lower bounds on the potential embedding accuracy of social graphs. These bounds inherently depend on the primal properties of the embedding algorithm, such as dimensionality, embedding costs, and knowledge available to the algorithm, and should thus be expressed in terms of these properties. Second, we aim to develop new routing and embedding algorithms. Here, we expect the insights provided by our model to guide our design process and entail improved results for both synthetic and real-world datasets. Our newly developed Darknet routing algorithms are then to be integrated in a real-world network, so that users benefit from the highly privacy-preserving nature of Darknets without suffering from their current lack of efficiency.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments.

## REFERENCES

Sonja Buchegger, Doris Schiöberg, Le Hung Vu, and Anwitaman Datta. 2009. PeerSoN: P2P social networking. In *Social Network Systems*.

David L. Chaum. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM* 24, 2 (1981), 84–90.

Ian Clarke, Oskar Sandberg, Matthew Toseland, and Vilhelm Verendel. 2010. Private Communication Through a Network of Trusted Connections: The Dark Freenet. Retrieved from http://freenetproject.org/papers.html.

Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W. Hong. 2000. Freenet: A distributed anonymous information storage and retrieval system. In *Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability*.

Leucio-Antonio Cutillo, Refik Molva, and Thorsten Strufe. 2009. Privacy preserving social networking through decentralization. In *Proceedings of the 6th International Conference on Wireless On-demand Network Systems and Services (WONS'09)*. 145–152.

Nathan S. Evans and Christian Grothoff. 2011. R5N: Randomized recursive routing for restricted-route networks. In *Proceedings of the 5th International Conference on Network and System Security (NSS'11)*.

Pierre Fraigniaud and George Giakkoupis. 2009. The effect of power-laws on the navigability of small worlds. In *Proceedings of the 23rd Annual ACM Symposium on Principles of Distributed Computing, (PODC'09)*.

George Giakkoupis and Nicolas Schabanel. 2011. Optimal path search in small worlds: Dimension matters. In *Proceedings of the 43rd Symposium on Theory of Computing (STOC'11)*.

Andreas Höfer, Stefanie Roos, and Thorsten Strufe. 2013. Greedy embedding, routing and content addressing for darknets. In *Proceedings of the Conference on Networked Systems (NetSys'13*).

Tomas Isdal, Michael Piatek, Arvind Krishnamurthy, and Thomas E. Anderson. 2010. Privacy-preserving P2P data sharing with OneSwarm. In *Proceedings of the ACM SIGCOMM 2010 Conference (SIGCOMM'10)*.

Jon Kleinberg. 2000. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the 32nd Symposium on Theory of Computing (STOC'00)*.

Emmanuelle Lebhar and Nicolas Schabanel. 2004. Almost optimal decentralized routing in long-range contact networks. In *Proceedings of the 30th International Colloquium on Automata, Languages and Programming (ICALP'04)*.

Gurmeet Singh Manku, Moni Naor, and Udi Wieder. 2004. Know thy neighbor's neighbor: The power of lookahead in randomized P2P networks. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC'04)*.

Chip Martel and Van Nguyen. 2003. *The Complexity of Message Delivery in Kleinberg's Small-World Model*. Technical Report. UC Davis Department of Computer Science.

Chip Martel and Van Nguyen. 2004. Analyzing Kleinberg's (and other) small-world models. In *Proceedings of the 33rd Annual ACM Symposium on Principles of Distributed Computing (PODC'04*).

Jon McLachlan, Andrew Tran, Nicholas Hopper, and Yongdae Kim. 2009. Scalable onion routing with Torsk. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*. ACM, 590–599.

Prateek Mittal and Nikita Borisov. 2009. ShadowWalker: Peer-to-peer anonymous communication using redundant structured topologies. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*. ACM.

Prateek Mittal, Matthew Caesar, and Nikita Borisov. 2012. X-Vine: Secure and pseudonymous routing using social networks. In *Proceedings of the 19th Annual Network & Distributed System Security Symposium (NDSS'12)*.

Hooman Mohajeri Moghaddam, Baiyu Li, Mohammad Derakhshani, and Ian Goldberg. 2012. Skypemorph: Protocol obfuscation for tor bridges. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*. ACM, 97–108.

Andriy Panchenko, Stefan Richter, and Arne Rache. 2009. NISAN: Network information service for anonymization networks. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*. ACM, 141–150.

Bogdan C. Popescu, Bruno Crispo, and Andrew S. Tanenbaum. 2006. Safe and private data sharing with turtle: Friends team-up and beat the system. In *Proceedings of the 12th International Workshop Security Protocols*. Springer.

Stefanie Roos and Thorsten Strufe. 2012. Provable polylog routing for darknets. In *Proceedings of the 4th Workshop on Hot Topics in Peer-to-Peer Computing and Online Social Networking (HotPOST'12)*.

Stefanie Roos and Thorsten Strufe. 2013. A contribution to darknet routing. In *Proceedings of the 32nd IEEE International Conference on Computer Communications (INFOCOM'13)*.

Stefanie Roos and Thorsten Strufe. 2015. On the impossibility of efficient self-stabilization in virtual overlays with churn. In *INFOCOM*. IEEE.

Benjamin Schiller, Stefanie Roos, Andreas Höfer, and Thorsten Strufe. 2011. Attack resistant network embeddings for darknets. In *Proceedings of the 30th Symposium on Reliable Distributed Systems Workshops (SRDSW'11)*.

Guanyu Tian, Zhenhai Duan, Todd Baumeister, and Yingfei Dong. 2014. Reroute on loop in anonymous peer-to-peer content sharing networks. In *Proceedings of the 2014 IEEE Conference on Communications and Network Security (CNS'14)*. IEEE, 409–417.

Eugene Vasserman, Rob Jansen, James Tyra, Nicholas Hopper, and Yongdae Kim. 2009. Membership-concealing overlay networks. In *Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS'09)*.

Received October 2014; revised May 2015; accepted July 2015