Theoretical Computer Science 410 (2009) 4975–4988

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Navigable Small-World networks with few random bits ☆

Gennaro Cordasco *, Luisa Gargano

Dipartimento di Informatica ed Applicazioni "R.M. Capocelli", Università di Salerno, Via Ponte don Melillo, Fisciano 84084, Italy

ARTICLE INFO

Article history:
Received 13 February 2007
Received in revised form 14 July 2009
Accepted 17 July 2009
Communicated by D. Peleg

Keywords:
Small-World networks
Kleinberg's model
Greedy routing
Complex networks

ABSTRACT

We study Small-World graphs in the perspective of their use in the development of efficient as well as easy to implement network infrastructures. Our analysis starts from the Small-World model proposed by Kleinberg: a grid network augmented with directed long-range random links. The choices of the long-range links are independent from one node to another. In this setting greedy routing and some of its variants have been analyzed and shown to produce paths of polylogarithmic expected length. We start from asking whether all the randomness, used in Kleinberg's model for establishing the long-range contacts of the nodes, is indeed necessary to assure the existence of short paths. In order to deal with the above question, we impose (stringent) restrictions on the choice of long-range links and we show that such restrictions do not increase the average path length of greedy routing and its variations.

We are able to decrease the number of random bits, required to establish each node's long-range link, from $\Omega(\log n)$ to $O(\log\log n)$ on a network of size n. Diminishing the randomness in the choice of random links has several benefits; in particular, it implies an increase in the clustering of the graph, thus increasing the resilience of the network.

1. Introduction

In this paper we consider Small-World networks based on Kleinberg's model [14,15]. We investigate the possibility of diminishing the amount of randomness that nodes need in the choice of their long-range links, while keeping the short greedy routes of the original model.

1.1. Small-World (SW) networks

The study of many large–scale real-world networks shows that such networks exhibit a set of properties that cannot be totally captured by the traditional models: regular graphs, such as a Euclidean lattice, and random graphs (cf. Erdös and Rényi random graphs [7]). Indeed, many biological and social networks occupy a position which is intermediate between completely regular and random graphs. Such networks, commonly called *Small-World* networks, are characterized by the following main properties:

- they tend to be large, in the sense that they contain $n \gg 1$ nodes;
- they tend to be sparse, in the sense that each node is connected to an average of $\delta \ll n$ other nodes;

☆ An extended abstract of an early version of this work appeared in [Gennaro Cordasco, Luisa Gargano, "How Much Independent Should Individual Contacts be to Form a Small-World?", in: Proc. of 17th International Symposium on Algorithms and Computation, ISAAC'06, Dec. 2006, Kolkata, India]. [4].

* Corresponding author. Tel.: +39 089969307.
E-mail addresses: cordasco@dia.unisa.it (G. Cordasco), lg@dia.unisa.it (L. Gargano).

0304-3975/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2009.07.050

- they exhibit a small average shortest path length, typically logarithm of the number of nodes (as random graphs);
- they have a high clustering coefficient (unlike sparse random graphs).

We recall that the clustering coefficient C_v for a vertex v is the ratio of the number of links among v's neighbors over the number of links that could possibly exist among them. The graph clustering coefficient is the average clustering coefficient over all the nodes of the graph [27].

The study of SW graphs was pioneered by Milgram [18] in the 1960s. In his famous experiment, people were asked to send letters to unknown targets only through acquaintances. The result not only confirmed the belief that random pairs of individuals are connected by short chains of acquaintances but also that people were able to route letters in a few hops by forwarding them to one of their acquaintances.

The work by Milgram has been followed by several studies aimed to a better understanding and modeling of the SW phenomenon. In particular, Watts and Strogatz [27] noticed that, unlike random graphs, real data networks tend to be clustered. Following the notion that SW graphs interpolate between completely regular graphs and purely random graphs, Watts and Strogatz proposed a novel network model, which is based on randomly rewiring each edge, of a regular graph, according to a given probability p. For suitable values of p, this gives rise to networks having both the small average shortest path length and the high level of clustering that are observed in natural and artificial networks (e.g. the network of actors' cooperations, the neural wiring diagram of the worm *C. elegans*, food webs, power grids and so on).

1.2. Kleinberg's model

Recently, Kleinberg [14] reconsidered an important algorithmic aspect of Milgram's experiment: not only short length paths exist but individuals are also able to deliver messages to unknown targets using short routes. Kleinberg proposed a basic model that uses a two-dimensional grid as underlying interconnection, the grid is augmented with random links: each node has an undirected local link to each of its grid neighbors and one directed long-range random link (which reaches a node at lattice distance d with probability proportional to d^{-2}). Kleinberg proved that such graphs are indeed *navigable*, that is, a simple *greedy* algorithm finds routes between any source and target using only $O(\log^2 n)$ expected hops in a network of n nodes [14].

Notice that navigability is an interesting property for a graph. Navigable graphs, in fact, can be easily used in the development of efficient network infrastructures, such as for Peer-to-Peer systems, where neither flooding nor complex routing protocols are to be used. Indeed, augmenting an overlay network with random links is also at the base of some randomized Peer-to-Peer networks. Two examples are randomized Chord and Symphony [20].

Following [14], routing strategies that make use of an augmented topological awareness of the nodes have been investigated under the SW model. Papers [19] and [20] consider the improvements obtainable over greedy routing in the case in which the topological awareness of a node is augmented by the knowledge of the long-range contacts of all neighbors of the node (*Neighbor-of-Neighbor greedy routing*). Such routing reaches the optimal $O(\log n/\log\log n)$ expected number of hops in SW percolation networks and in the randomized version of the Chord network [26] (both having $O(\log n)$ degree). In [21] the authors consider the improvements obtainable if each node knows the long-range contacts of its closest $O(\log n)$ neighbors on the grid (*Indirect greedy routing*), obtaining $O(\log^{1+1/s} n)$ expected number of hops over an s-dimensional lattice. The $O(\log^{1+1/s} n)$ bound is also achieved in [10] by an oblivious greedy algorithm.

Recently, some work has been devoted to the study of networks obtained by adding long-range links to the nodes of a generic underlying graph [6,9,25]. In particular, papers [11] and [12] give an upper bound[1] of $\widetilde{O}(n^{1/3})$ and a lower bound of $\Omega\left(n^{1/\sqrt{\log n}}\right)$ on the expected number of hops for greedy routing in such a case.

1.3. Low randomness Small-World networks

In a Kleinberg's SW network, the additional long-range random links represent the chance of generating shortcuts, which plays a large role in creating short paths through the network as a whole. In the original Kleinberg's model, there is one long-range link per node without any restriction (that is, each node can have a long-range link to any other node.) Hence each node requires at least $\Omega(\log n)$ random bits to generate its long-range link. In this paper we consider the following question:

> *Do the long-range contacts really need to be completely random, or some "long-range clustering" could instead be envisaged in such a "navigable" network?*

In other words, we investigate the problem of whether all the independence assumed in Kleinberg's model among different long-range contacts is indeed necessary to assure the existence of short greedy paths. We show that, up to a certain extent, the answer to the above question is that the same (greedy) routing performances can be maintained if we limit the amount of randomness that nodes use in the choice of long-range contacts.

[1] Where the notation $\widetilde{O}$ ignores the polylogarithmic factors.

Why does one want to reduce randomization?

We first notice that, in the perspective of using SW graphs as a base for the development of network infrastructures, the use of randomization increases the difficulties in the implementation and testing of applications. As an example, while in a regular graph it can be assumed that the structure of the network is a global information, in a random graph the more randomness is injected, the less information about its structure is available.

Moreover, Watts and Strogatz showed that for a regular graph, where some randomization is introduced, it holds that the clustering coefficient is inversely proportional to the amount of injected randomization. Clustering is a very interesting concept that is found in many natural and artificial phenomena. Hence, it represents a fundamental feature that a network model, designed to describe complex networks, must posses.

As an example, in the perspective of the development of network infrastructures, we notice that besides diameter and degree, some very important properties are the resilience to simultaneous node failures and the tolerance to hot spot workload:

- The resilience of a network grows with the clustering coefficient of the underlying graph. Intuitively, a high clustering coefficient provides several alternative paths through which flow can pass, thus avoiding the failed component [17]. Also, in [2] and [23] clustering has been identified as a topological structure that facilitates resilience to cascading failure.
- Clustering is also useful in a flash crowd scenario (i.e., when a node catches the attention of a large number of other nodes, and gets an unexpected and overloading surge of traffic): high clustering implies an improved ability to handle heavy traffic workload, allowing fast self-organization and replication of popular data object in the network infrastructure [13]. Hence, high clustering coefficient implies that the underlying network can effectively provide object lookup even under heavy demands.

Therefore, it is worth investigating if, in analogy with real SW graphs [27], a SW interconnection can be obtained by using a limited amount of randomness. This would allow both SW requirements (small average path length and high clustering) to be obtained together with easy routing strategies.

Some work in this direction has been done in the context of Peer-to-Peer networks. Namely, paper [3] deals with the problem of speeding-up bootstrap in the randomized Chord ring; this is attained by limiting the amount of randomization needed for optimal routing in the network.

The extreme case of a deterministic augmentation strategy was considered in [8]. Such an augmentation provides a $O(\log n)$ oblivious diameter for paths, trees and s-dimensional grids, for a fixed value of s. The proposed approach builds a network where the added long-range links are exactly determined by nodes' positions in the original graph. This method deterministically designs a small-diameter network starting from the input graph and so the augmented network cannot be considered as a pure SW network. Moreover, it lacks of some interesting properties which hold for SW networks. Mainly, the proposed network is based on a hierarchical structure providing very unbalanced traffic; it is easily observed that nodes on the first levels are more congested than nodes on the latest levels. Hence, the network is not scalable in terms of congestion. Furthermore, the proposed construction provides an $O(\log n)$ oblivious diameter only for grids having a constant number of dimensions; namely, the diameter is $O(s^2 \log n)$ for an augmented s-dimensional grid.

1.4. Our results

We start from Kleinberg's model[2] and proceed in two steps. In the first step, the choice of long-range links of a node u is restricted to the nodes that differ from u in exactly one coordinate; namely, a node $u = \langle u_1, \ldots, u_s \rangle$ has its long-range contacts chosen from nodes $v = \langle v_1, \ldots, v_s \rangle$ for which there exists i $(1 \leq i \leq s)$ such that $v_i \neq u_i$ and $v_j = u_j$ for each $j \neq i$. We show that all the routing properties (for Greedy, Indirect and Neighbor-of-Neighbor routing strategies) immediately translate to this restricted model, sometimes with easier proofs. In the case of greedy routing, such a result was independently discovered in [12].

Our main result allows a stronger restriction on the long-range contacts of nodes to be imposed. We introduce the notion of *groups*: Keeping the restriction that long-range links can only connect two nodes differing in exactly one coordinate, the set of nodes is partitioned into (random) sets, called groups; all nodes belonging to the same group are subjected to the same additional restrictions in the choice of long-range contacts.

We show that a logarithmic (in the number of nodes) number of different groups is sufficient to ensure that the SW property holds for the resulting graph. Namely, we analyze the routing performances of Greedy, Indirect and Neighbor-of-Neighbor (NoN) routing strategies in dependence of the number of groups. In particular, when the number of groups is logarithmic in the size of the network, all the routing strategies attain the same performances as in Kleinberg's original model.

The proposed network model requires only $\log c + \log s = O(\log \log n)$ random bits to establish all the long-range contacts of each node – where n is the number of nodes in the network, s is network's dimension and c is the number of groups – instead of the $\Omega(\log n)$ random bits required in the original Kleinberg construction.

[2] The original work of Kleinberg considers an s-dimensional undirected grid augmented with long-range links. For the sake of simplicity throughout this paper we will focus on tori; however all results can be easily extend to grids.

Table 1
Performance of variants of greedy routing (see Section 2.1 for more details). Toroidal grids having n nodes, s dimensions and q long-range contacts are considered: $\mathcal{K}(n, s, q)$ (Kleinberg-Small-World network $\mathcal{K}(n, s, q, p)$ with probability $p(d)$ proportional to d^{-s}, cf. Definition 1), $\mathcal{R}(n, s, q)$ (Restricted-Small-World network, cf. Definition 2), $\mathcal{R}_c(n, s, q)$ (Small-World network with groups, cf. Definition 3).

Paper	Routing	Avg #steps	Network	Random bits
[14]	Greedy	$O((\log^2 n)/q)$	$\mathcal{K}(n, s, q)$	$\Omega(q \log n)$
[1,21]	Greedy	$\Omega((\log^2 n)/q)$	$\mathcal{K}(n, s, q)$	$\Omega(q \log n)$
[10,21]	IR	$O((\log^{1+1/s} n)/(q^{1/s}))$	$\mathcal{K}(n, s, q)$	$\Omega(q \log n)$
[10,21]	IR	$\Omega(\log^{1+1/s} n)$	$\mathcal{K}(n, s, q)$	$\Omega(q \log n)$
[19,20]	NoN	$O((\log^2 n)/(q \log q))$	$\mathcal{K}(n, s, q)$	$\Omega(q \log n)$
This paper	Greedy	$O\left(\frac{\log s}{s} \cdot \frac{\log^2 n}{q}\right)$	$\mathcal{R}(n, s, q)$	$O\left(q\left(\frac{\log n}{s} + \log\log n\right)\right)$
This paper	Greedy	$O\left(\frac{\log s}{s} \cdot \frac{\log^2 n}{q}\right)$	$\mathcal{R}_c(n, s, q)$	$\log c + \log s$
This paper	IR	$O((\log^{1+1/s} n)/(q^{1/s}))$	$\mathcal{R}(n, s, q)$	$O\left(q\left(\frac{\log n}{s} + \log\log n\right)\right)$
This paper	IR	$O((\log^{1+1/s} n)/(q^{1/s}))$	$\mathcal{R}_c(n, s, q)$	$\log c + \log s$
This paper	NoN	$O((\log^2 n)/(q \log q))$	$\mathcal{R}(n, s, q)$	$O\left(q\left(\frac{\log n}{s} + \log\log n\right)\right)$
This paper	NoN	$O((\log^2 n)/(q \log q))$	$\mathcal{R}_c(n, s, q)$	$\log c + \log s$

We also provide a lower bound on the performance of any network having degree $O(\log n)$ and c groups, showing that our construction is asymptotically optimal.

The main results presented in this paper, together with those known for the original Kleinberg model, are summarized in Table 1.

2. Preliminary notation and definitions

Consider an s-dimensional toroidal grid $\{0, \ldots, m-1\}^s$, having $n = m^s$ nodes $\langle u_1, \ldots, u_s \rangle$, with $u_i \in \{0, \ldots, m-1\}$ for $i = 1, \ldots, s$.

The undirected distance $d(u, v)$, between two points $u = \langle u_1, \ldots, u_s \rangle$ and $v = \langle v_1, \ldots, v_s \rangle$ on the toroidal grid, is defined as[3]

$$d(u, v) \stackrel{\text{def}}{=} \sum_{i=1}^{s} \min(v_i \ominus u_i, u_i \ominus v_i). \tag{1}$$

As said before, while the original Kleinberg's model was based on a bidirectional grid, we assume that the interconnection is formed by a bidirectional torus. In most cases, for the sake of simplicity, we will route using only one direction on each dimension; to this purpose we will make use of the directed distance metric

$$\vec{d}(u, v) \stackrel{\text{def}}{=} \sum_{i=1}^{s} v_i \ominus u_i.$$

All results can be easily re-obtained, by using for routing the appropriate undirected distance metric whenever $\vec{d}$ occurs, both on tori (using d as defined in (1)) and on grids (i.e. with no wrap-around).

Definition 1 (*Kleinberg-Small-World Network*). A Kleinberg-Small-World network $\mathcal{K}(n, s, q, p)$ is a bidirectional s-dimensional toroidal grid $\{0, \ldots, m-1\}^s$ where each node maintains two types of connections:

short-range contacts ($2s$ bidirectional connections): Each node has a connection to every other node at distance 1 on the grid;
long-range contacts (q unidirectional connections): Each node u establishes $q > 0$ pairwise independent directed links according to the probability distribution p (on the integers): each link has endpoint v with probability $p(d(u, v))$.

All reported results on Kleinberg's model assume $p(d)$ proportional to d^{-s} with normalization factor $\sum_v d(u, v)^{-s}$.

Observation 1. *Analyzing the probability distribution $p(d) = \left(d^s \sum_v d(u, v)^{-s}\right)^{-1}$, it is easy to observe that a Kleinberg-Small-World network requires $\Omega(\log n)$ random bits for establishing each long-range contact, that is, $\Omega(q \log n)$ random bits for each node.*

In the following we will denote by:

- $N(u)$ the neighborhood of u, that is, the set of the $2s$ neighbors of u on the s-dimensional toroidal grid;
- $N_r(u) = \{v \mid d(u, v) \le r\}$ the ball of radius r and center u on the s-dimensional toroidal grid;
- $L(u)$ the set of the q long-range contacts of node u;

3 Throughout this paper, $\oplus$ and $\ominus$ represent the math operations $+$ and $-$ modulo m.

- $L_r(u) = \bigcup_{v \in N_r(u)} L(v)$ the set of all long-range contacts of nodes in $N_r(u)$;
- $L^2(u) = \bigcup_{v \in L(u) \cup \{u\}} L(v)$ the set of long-range contacts of the long-range contacts of u.

It has been shown that $|N_r(u)|$ corresponds to the Delannoy number $d_{r,s}$ (which originally counts the number of paths from $(0, 0)$ to (r, s) on a two-dimensional lattice, in which only east, north, and northeast steps are allowed – i.e., $\rightarrow$, $\uparrow$, and $\nearrow$) [24]. Such values can be expressed as

$$|N_r(u)| = \frac{2^s}{s!} \cdot r^s + \nu(r), \tag{2}$$

where $\nu(r)$ is a positive polynomial of degree $s - 1$ [5,16].

2.1. Routing strategies

Consider a SW network $\mathcal{K}(n, s, q, p)$. We shortly review the routing strategies, adopted in the Small-World related literature, that will be subsequently used in this paper. We denote by u the node currently holding the message and by t the target node.

Greedy Routing uses only the local knowledge of u: a message is forwarded along the link (either short or long) that brings it to the node which is closest to the target t (with respect to some metric distance).

Indirect greedy routing [10,21] and *Neighbor-of-Neighbor greedy routing* [19,20] are obtained through an additional *topological awareness* given to the nodes: Each node u is in fact aware of the long-range contacts of some other nodes. Known results for these strategies are summarized in Fig. 1.

Indirect Greedy Routing (IR) assumes that node u is aware of the long-range contacts of each node at distance at most r from u on the grid, for some $r > 0$ (cf. Fig. 1). Formally, IR entails the following decision: Among the nodes in $L_r(u) \cup N(u)$ (e.g., among the nodes having distance at most r from u and their long-range contacts), assume that z is the closest to the target (with respect to the metric distance $\vec{d}$): If $z \in L(u) \cup N(u)$ then route the message from u **to** z directly; otherwise, let $z \in L(v)$ ($v \in N_r(u)$ for some $v \neq u$) and route the message from u **to** z **via** v.

Neighbor-of-Neighbor Greedy Routing (NoN) assumes that each node knows its long-range contacts and the long-range contacts of each of its neighbors. Here we restrict ourselves to consider only the long-range contacts of nodes in $L(u)$. Indeed, this will be sufficient to get the improvements over greedy routing assured by NoN. Formally, a NoN greedy step entails the following decision: Among the nodes in $L^2(u) \cup N(u)$, let z be the closest to the target (with respect to the metric distance $\vec{d}$): If $z \in L(u) \cup N(u)$ then route the message from u **to** z directly; otherwise, let $z \in L(v)$ ($v \in L(u)$ for some $v \neq u$) and route the message from u **to** z **via** v.

3. Restricted-Small-World networks

We allow each node to make long-range connections only with nodes that differ from it in exactly one coordinate. A connection between two nodes u and v is created with probability proportional to $(\vec{d}(u, v))^{-1}$. This probability is

$$p(\vec{d}(u, v)) = \frac{1}{\lambda \vec{d}(u, v)},$$

where λ is the normalized coefficient,

$$\lambda = s \sum_{j=1}^{m} \frac{1}{j} \approx \ln n. \tag{3}$$

Different connections are established by independent trials.

Definition 2 (*Restricted-Small-World Network*). A Restricted-Small-World network ($\mathcal{R}(n, s, q)$), is a network $\mathcal{K}(n, s, q, p)$ with probability distribution p such that, for any u and v, the probability of having a long-range link from u to v is

$$p(\vec{d}(u, v)) = \begin{cases} \frac{1}{\lambda \vec{d}(u,v)} & \text{if } u \text{ and } v \text{ differ in exactly one dimension} \\ 0 & \text{otherwise,} \end{cases}$$

where λ is defined in (3).

It is immediate to see that any outgoing link goes along dimension i with probability $1/s$, for $i = 1, \ldots, s$.

Analyzing the probability distribution defined above, it is easy to show that a Restricted-Small-World network requires at most $O\left(\frac{\log n}{s} + \log s\right)$ random bits for establishing each long-range contact.

We will show that all the results obtained on Kleinberg's SW networks [10,14,20,21] can be easily proved to remain valid in spite of the restrictions we impose on the long-range connections.

3.1. Greedy routing

Consider a generic node $u = \langle u_1, \dots, u_s \rangle \in \{0, \dots, m-1\}^s$ that holds a message addressed to the target node $t = \langle t_1, \dots, t_s \rangle$. For each $i = 1, \dots, s$, let d_i denote the directed distance between u and t on dimension i, that is, $d_i = (t_i \ominus u_i)$.

Part of the proofs of the following Lemma 1 and Theorem 1 follows the lines of the proof given in [19] in the case of greedy routing on a ring with long-range connections added according to Definition 1; we report the full proof for the sake of completeness and for later reference in this paper.

Lemma 1. *Consider a Restricted-Small-World network $\mathcal{R}(n, s, q)$. For any integer $k \geq 2$ such that $q \leq k \ln n$ and for a fixed value of i, $1 \leq i \leq s$, the probability that in one hop the node u, holding the message, is able to diminish the distance d_i on dimension i to at most d_i/k is lower bounded by $\frac{q}{2k \ln n}$.*

Proof. Fix a dimension i and an integer $k \geq 2$. Let ϕ denote the event that the current node is able to diminish the remaining distance, on dimension i, from d_i to at most d_i/k in one hop. We first evaluate the probability that event ϕ occurs. To this aim, we notice that the probability of diminishing the distance from d_i to at most d_i/k using *one* long-range contact is

$$\Psi = \sum_{j=d_i(1-1/k)}^{d_i} p(j) \geq \frac{d_i}{k} \cdot p(d_i) = \frac{d_i}{k} \cdot \frac{1}{d_i \ln n} = \frac{1}{k \ln n}. \tag{4}$$

Therefore, using $q \leq k \ln n$ long-range connections, the probability that event ϕ occurs is

$$\Phi \geq 1 - (1-\Psi)^q \geq \Psi \frac{q}{2} \geq \frac{q}{2k \ln n}. \quad \square$$

Theorem 1. *Consider a Restricted-Small-World network $\mathcal{R}(n, s, q)$. For any $q \leq 2 \ln n$, the average path length for greedy routing on $\mathcal{R}(n, s, q)$ using $\vec{d}$ as metric distance is $O\left(\frac{\log^2 n}{q}\right)$.*

Proof. Both grid neighbors and long-range contacts can diminish the distance only on one dimension and each greedy hop is independent of each other. Hence, we partition the whole sequence of hops into s subsequence so that the ith subsequence includes all the hops along dimension i, preserving their order in the original sequence. We will bound the number of hops by upper bounding the length of each subsequence, this can be done since each node chooses its long-range contacts according to the same probability distribution.

Consider a fixed dimension i and its corresponding sequence of hops. Let us first notice that if a suitable long-range is found, it is always preferred to u's local links that can only reduce the global distance by 1.
We say that a hop is successful on dimension i if the node currently holding the message is able to diminish in one hop the distance on dimension i from its current value d_i to at most d_i/k.

By Lemma 1, the expected number of hops, that are done on dimension i before a successful one occurs, is $O\left(\frac{k \ln n}{q}\right)$. Moreover, the number of needed successful hops on dimension i is at most $\log_k m$, since $d_i \leq m$ for each $i = 1, \dots, s$.

If follows that the number of hops that are done on each dimension is $O\left(\frac{k \ln n}{q} \cdot \log_k m\right)$. By summing up on all s dimensions, we have that the average total number of hops is

$$O\left(s \cdot \frac{k \ln n}{q} \cdot \frac{\log m}{\log k}\right) = O\left(\frac{k}{\log k} \cdot \frac{\log^2 n}{q}\right). \tag{5}$$

By choosing $k = 2$, the result follows. $\square$

3.1.1. Improving greedy routing on high-dimensional networks

In the following we give a different greedy strategy which allows us to obtain a significant improvement, with respect to Theorem 1, in the case of high-dimensional networks. The strategy is based on deciding each greedy hop according to the relative gains d_i/d'_i, where d'_i represents the remaining distance when one hop on dimension i is performed (rather than according to the *absolute* gains $d_i - d'_i$ considered in the standard greedy strategy).

Namely, at each step, the algorithm evaluates the gains $\rho_i = d_i/d'_i$, for $i = 1, \dots, s$, and decides to progress by making one hop on the dimension having the maximum gain, i.e., the dimension i^* such that

$$\rho_{i^*} = \max \rho_i,$$

where the maximum is taken on all $i = 1, 2, \dots, s$ such that $d_i > 0$.

Theorem 2. *Consider a Restricted-Small-World network $\mathcal{R}(n, s, q)$. For any q and s such that $qs \leq 4 \ln n$, the average path length for greedy routing with relative gains on $\mathcal{R}(n, s, q)$ is $O\left(\frac{\log s}{s} \cdot \frac{\log^2 n}{q}\right)$.*

Proof. For each $i = 1, \ldots, s$, we say that dimension i is *active* if $d_i > 2$. Since the number of active dimensions decreases (down to 0) during the routing toward the target t, we will perform the analysis in $\lfloor \log s \rfloor + 1$ stages, where stage ℓ lasts while the number of active dimensions belongs to the interval $\left[2^\ell, 2^{\ell+1}\right)$, for $\ell = \lfloor \log s \rfloor, \ldots, 0$.

Consider now any hop performed during stage ℓ, for some $0 \le \ell \le \lfloor \log s \rfloor$.
If we fix any active dimension i and apply Lemma 1 with $k = 2$, we get that the probability of having a gain $\rho_i \ge 2$, is lower bounded by $\frac{q}{4 \ln n}$.

It follows that the probability that there exists an *active* dimension j such that $\rho_j \ge 2$ is lower bounded by

$$1 - \left(1 - \frac{q}{4 \ln n}\right)^{2^\ell} \ge \frac{q 2^\ell}{8 \ln n},$$

where the inequality holds since $q2^\ell \le qs \le 4 \ln n$.

Therefore, the expected number of hops that are done before performing a successful hop, e.g. a hop giving a gain of 2 or more, is $O\left(\frac{\log n}{q 2^\ell}\right)$.

We notice now that during stage ℓ, the number of active dimensions can be at most $2^{\ell+1}$ and that at most $\log d_i \le \log m$ successful hops can be performed on each active dimension i. Therefore, we get that the number of successes, needed to complete stage ℓ, is at most $2^{\ell+1} \log m$; therefore, each stage will last for $O\left(\frac{\log m \log n}{q}\right)$ routing hops.

Summing up over all stages, we have that the total number of hops needed to route to the target t is $O\left(\frac{\log m \log n}{q} \log s\right)$ and the theorem holds. □

Remark 1. The above strategy based on relative gains does not provide any significant improvement for the Indirect greedy routing considered in the following. Indeed, the additional awareness, therein available to the nodes, provides an improvement which overcomes the improvement obtained above. Similar reasons seem to apply for the Neighbor-of-Neighbor greedy routing. Indeed, in Section 4.5 we will show that the Neighbor-of-Neighbor greedy routing approach (based on absolute gains) is already asymptotically optimal.

3.2. Indirect greedy routing

In order to simplify the analysis, we provide here a modified version of the Indirect greedy routing algorithm for $\mathcal{R}(n, s, q)$:

Let $u = \langle u_1, \ldots, u_s \rangle$ be the current node and $t = \langle t_1, \ldots, t_s \rangle$ the target node.

Among all the nodes $u + (z - v)$ such that $v \in N_r(u)$ and $z \in L(v)$, assume that w is the closest to the target t with respect to the metric distance $\vec{d}$.

The routing step (see Fig. 1) from u to w is as follows:

(1) If $w \in N(u) \cup L(u)$ then route the message from u **to** w directly,
(2) otherwise, let $w = u + (z - v)$ with $z \in L(v)$ for some $v = \langle u_1 \oplus r_1, u_2 \oplus r_2, \ldots, u_s \oplus r_s \rangle \in N_r(u)$,
  (2.a) first route the message from u **to** v (it takes $\sum_{i=1}^{s} |r_i| \le r$ hops),
  (2.b) then use v's long-range contact **to** z (1 hop),
  (2.c) finally go from z **to** the node $w = \langle z_1 \ominus r_1, z_2 \ominus r_2, \ldots, z_s \ominus r_s \rangle$ (at most r hops).

We remark that point (2.c) can move the message away from the target node; one could easily improve the algorithm by removing in (2.c) all the hops which go in the wrong direction. For the sake of simplicity we will analyze the algorithm as described above.

Theorem 3. *Consider a Restricted-Small-World network $\mathcal{R}(n, s, q)$. If $q \le e^s \ln n$ and each node knows the long-range contacts of its $\frac{e^s \ln n}{q}$ closest neighbors, then the average path length for Indirect greedy routing on $\mathcal{R}(n, s, q)$ is $O\left(\frac{\log^{1+1/s} n}{q^{1/s}}\right)$.*

Proof. Let us notice that each routing step diminishes the distance only on one dimension; indeed during one step either one makes one hop according to point (1) or the only change in the distance is due to the hop done at point (2b) using the long-range contact of node v (cf. Fig. 1). Therefore, as in Section 3.1, we can partition the whole sequence of steps (each having at most $2r + 1$ hops) into s subsequences so that the ith subsequence includes all the routing steps that move the message along dimension i (e.g. such that hop in (1) or (2.b) goes along dimension i), preserving their order in the original sequence.
We evaluate now the average path length.
Consider a fixed dimension i and its corresponding subsequence of steps. Let $d_i = t_i \ominus u_i$ be the directed distance on dimension i from the current node u to t. We first notice that in a routing step, made according to point (2) and leading from u to w (via v and z):

(a) We move to v (in at most r steps); at this point the target can be seen as the node $t_v = \langle t_1 \oplus r_1, t_2 \oplus r_2, \ldots, t_s \oplus r_s \rangle \in N_r(t)$. (cf. Fig. 1)

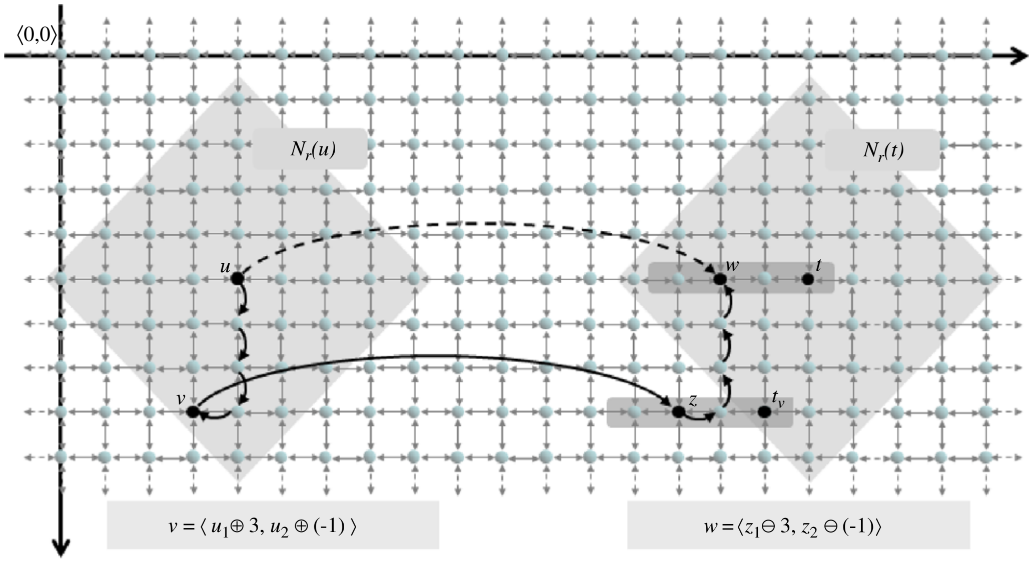

Fig. 1. An indirect greedy routing step on top of a two-dimensional Restricted-Small-World network.

(b) The distance of t_v from v on dimension i is exactly d_i. Hence, we can apply equation (4) between v and t_v. Thus (4) gives the probability that a link connects v to some node z at distance at most $d_i(1 - 1/k)$ from t_v (on dimension i).

(c) Finally, point (2c) performs a move from z to $w = \langle z_1 \ominus r_1, z_2 \ominus r_2, \ldots, z_s \ominus r_s \rangle$ (node w is s.t. $w_j = u_j$ for $j \neq i$ while on dimension i the gain on the distance from u to t is equal to that on the distance from v to t_v).

We can then repeat the proof of Theorem 1 by noticing that $q' = q|N_r(u)|$ long-range contacts are available at each routing step. Hence, as in equation (5) we get that, whenever $q' \leq k \ln n$, the average total number of indirect steps needed is $O\left(s \cdot \frac{k \ln n}{q'} \cdot \frac{\log m}{\log k}\right) = O\left(\frac{k \ln n}{q|N_r(u)|} \cdot \log_k n\right)$. By choosing the parameter k so that $q|N_r(u)| = k \ln n$, we have that the expected number of indirect steps to reach the destination is $O(\log_k n)$. Each step requires at most $2r + 1$ hops where, by (2), $r = O\left(s\left(\frac{k \ln n}{q}\right)^{1/s}\right)$.

Hence, the average path length is

$$O\left(r \log_k n\right) = O\left(s \cdot \left(\frac{k \ln n}{q}\right)^{1/s} \cdot \frac{\ln n}{\ln k}\right).$$

By choosing $k = e^s$, the result follows. □

Remark 2. Unlike in [10,21], where multiplicative factors in s are discarded in the asymptotic notation, our results are expressed also for a non-constant number s of dimensions. In particular the results in [10,21] are obtained using an awareness of $O\left(\frac{\log n}{q}\right)$ neighbors; this awareness allows us to obtain an average path length $O\left(s \cdot \frac{\log^{1+1/s} n}{q^{1/s}}\right)$.

3.3. NoN greedy routing

In the case of Neighbor-of-Neighbor greedy routing, we obtain the following result.

Theorem 4. *Consider a Restricted-Small-World network $\mathcal{R}(n, s, q)$. For any q such that $q \leq \log n$, the average path length for NoN routing on $\mathcal{R}(n, s, q)$ is $O\left(\frac{\log^2 n}{q \log q}\right)$.*

Proof. The proof of this theorem can be obtained following the lines of the more general (and complex) result of Theorem 7 in Section 4.4. □

4. Small-World networks with groups

In this section we impose stricter restrictions on the choice of long-range contacts held by the nodes in the network. Namely, we assume that nodes are partitioned into c categories, called *groups*. First of all, each node randomly chooses one of the groups to belong to; this choice is then used to establish the long-range contacts associated to that node.

In order to balance the number of long-range contacts per dimension, the number of long-range contacts, which will be drawn along each dimension, is determined as follows: each node picks at random $q \bmod s$ dimensions and assigns to each of them $\lceil q/s \rceil$ long-range contacts; on the remaining dimensions there will be exactly $\lfloor q/s \rfloor$ long-range contacts. In particular, when $q < s$ there is exactly 1 long-range contact on each of the q randomly chosen dimensions and no one on any of the remaining $s - q$ dimensions.

Once the number of long-range contacts per dimension has been established, the actual contacts are drawn; they are chosen according to (an approximation of) the 1-harmonic distribution, which depends on the group. Formally:

Definition 3 (*Small-World Network with Groups*). A Small-World network with groups ($\mathcal{R}_c(n, s, q)$), is a Kleinberg-Small-World network, where each node u computes its long-range contacts as follows:

(i) Node u chooses uniformly at random an integer σ in the set $[1, s]$. Then, for $i = 1, \ldots, s$, let

$$q_i = \begin{cases} \lceil q/s \rceil & \text{if } i = (\sigma + j) \bmod s, \text{ for } 1 \le j \le q \bmod s, \\ \lfloor q/s \rfloor & \text{otherwise.} \end{cases} \tag{6}$$

(ii) Node u chooses the group it belongs to by uniformly at random selecting an integer c_u in the interval $[0, c - 1]$.
(iii) For each dimension $i = 1, \ldots, s$, such that $q_i > 0$, and for each $\ell = 0, \ldots, q_i - 1$, the long-range contacts associated to the node u are the nodes $v_{i,\ell}$, such that, u and $v_{i,\ell}$ differ in dimension i only and $\vec{d}(u, v_{i,\ell}) = \left\lfloor \gamma_i^{\ell + \frac{c_u}{c}} \right\rfloor$, where γ_i denotes a real number satisfying $\ln \gamma_i = (\ln m)/q_i$.

Remark. A Small-World network with groups can be also seen as a network $\mathcal{K}(n, s, q, p)$ with probability distribution p such that for each node u and integer j, with $1 \le j \le q$, the jth long-range link from u has endpoint v at distance $\vec{d}(u, v)$ with probability

$$p(\vec{d}(u, v)) = \begin{cases} \frac{1}{q} & \text{if } u \text{ and } v \text{ differ in exactly dimension } i \text{ and } \vec{d}(u, v) = \left\lfloor \gamma_i^{\ell + \frac{c_u}{c}} \right\rfloor, \\ & \text{for some } 1 \le i \le s \text{ and } 0 \le \ell \le q_i - 1. \\ 0 & \text{otherwise.} \end{cases}$$

Observation 2. *By construction, each node of a Small-World network having c groups requires only* $\log c + \log s$ *random bits for establishing all of its q long-range contacts.*

4.1. Routing in Small-World networks with groups

In this section we show that the introduction of groups reduces the amount of randomness with no harm to the efficiency of the system.

The following result will be used in the analysis of the performances of the routing strategies described in Section 2.1.

Lemma 2. *Consider Small-World network with groups $\mathcal{R}_c(n, s, q)$. Denote by Φ the probability that a node is able to diminish in one hop the distance to the target, on dimension i, from d_i to at most d_i/k, for some fixed integer $k \le d_i/2$. If $q \le k \ln n$ and $c \ge \frac{4k \ln n}{q}$ then $\Phi \ge \frac{q}{16k \ln n}$.*

Proof. Suppose the message is at node u and the target is at directed distance $\vec{d}(u, t) = d$. Without loss of generality assume that $u = \langle u_1, \ldots, u_s \rangle = \langle 0, \ldots, 0 \rangle$ and $t = \langle d_1, \ldots, d_s \rangle$ (where $d = \sum_{i=1}^{s} d_i$).

For a fixed dimension i, we want to route in one hop the message to a node $w = \langle 0, \ldots, w_i, \ldots, 0 \rangle$ with $(d_i \ominus w_i) \le d_i/k$.

There are two cases to take into consideration, depending on the relative values of q and s.

Case 1: $q \ge s$.

According to (6), each dimension has at least one long-range contact.

Let $b = \left\lfloor \frac{c}{2k \ln \gamma_i} \right\rfloor$, that is,

$$2bk \ln \gamma_i \le c < 2(b + 1)k \ln \gamma_i. \tag{7}$$

Since $c \ge \frac{4k \ln n}{q} > 2k \ln \gamma_i$, we have $b \ge 1$. To see why the last inequality holds, we notice that by definition of γ_i (cf. Definition 3), it suffices to show that $\frac{4ks \ln m}{q} > \frac{2k \ln m}{q_i}$, or, that $q/(2s) < \lfloor q/s \rfloor$, and this always holds when $q \ge s$.

We denote by ℓ^* the smallest integer such that $\left\lfloor \gamma_i^{\ell^*+1} \right\rfloor > d_i$ and by a^* the smallest integer such that

$$\left\lfloor \gamma_i^{\ell^* + \frac{a^*}{c}} \right\rfloor > d_i. \tag{8}$$

Notice that $\ell^* < q_i$ (since $\gamma_i^{q_i} = m > d_i$) and $a^* \le c$.

Consider now the integers

$$\left\lfloor \gamma_i^{\ell^* + \frac{a^*-1}{c}} \right\rfloor, \left\lfloor \gamma_i^{\ell^* + \frac{a^*-2}{c}} \right\rfloor, \ldots, \left\lfloor \gamma_i^{\ell^* + \frac{a^*-b}{c}} \right\rfloor. \tag{9}$$

Each of these integers corresponds to the length of a long-range link of the current node u with probability $1/c$. Indeed, by (iii) of Definition 3 and the definitions of ℓ^* and a^*, we have that u has a long-range contact at distance $\left\lfloor \gamma_i^{\ell^*+\frac{a^*-\alpha}{c}} \right\rfloor$ if and only if its group c_u satisfies $c_u = (a^* - \alpha) \bmod c$.

We first show that

$$d_i\left(1 - \frac{1}{k}\right) \le \left\lfloor \gamma_i^{\ell^*+\frac{a^*-b}{c}} \right\rfloor \le \cdots \le \left\lfloor \gamma_i^{\ell^*+\frac{a^*-1}{c}} \right\rfloor \le d_i. \tag{10}$$

By definition of a^* (see (8)), we know that $\left\lfloor \gamma_i^{\ell^*+\frac{a^*-1}{c}} \right\rfloor \le d_i$. Hence, in order to prove (10), it suffices to show that the first inequality holds.

Assume by contradiction that

$$d_i\left(1 - \frac{1}{k}\right) > \left\lfloor \gamma_i^{\ell^*+\frac{a^*-b}{c}} \right\rfloor.$$

By using this and (8), we get,

$$1 - \frac{1}{k} = \frac{d_i(1 - 1/k)}{d_i} > \frac{\left\lfloor \gamma_i^{\ell^*+\frac{a^*-b}{c}} \right\rfloor}{d_i} \ge \frac{\gamma_i^{\ell^*+\frac{a^*-b}{c}} - 1}{d_i} > \frac{\gamma_i^{\ell^*+\frac{a^*-b}{c}}}{\gamma_i^{\ell^*+\frac{a^*}{c}}} - \frac{1}{d_i} = \gamma_i^{-\frac{b}{c}} - \frac{1}{d_i}.$$

Recalling that $k \le d_i/2$, we get

$$1 - \frac{1}{k} > \gamma_i^{-\frac{b}{c}} - \frac{1}{2k}. \tag{11}$$

From (11) it follows that $c < \frac{b \ln \gamma_i}{\ln\left(\frac{2k}{2k-1}\right)}$. Using the fact that $\ln\left(\frac{2k}{2k-1}\right) > \frac{1}{2k}$, we get $c < 2bk \ln \gamma_i$ which contradicts (7). Hence (10) holds.

We want now evaluate the probability Φ that a long-range contact of u diminishes the distance on a fixed dimension i, from d_i to at most d_i/k. Recalling that each integer in (9) corresponds to a long-range contact of u with probability $\frac{1}{c}$, we get

$$\Phi = 1 - \left(1 - \frac{1}{c}\right)^b \ge \frac{b}{2c} > \frac{b}{4(b+1)k \ln \gamma_i} \ge \frac{q}{16k \ln n}.$$

Case 2: $q < s$.

Consider a dimension i' with exactly one long-range contact.

Let $b' = \left\lfloor \frac{c}{2k \ln m} \right\rfloor$ (i.e. $2b'k \ln m \le c < 2(b'+1)k \ln m$). Since $c \ge \frac{4k \ln n}{q} > 2k \ln m$, we have that $b' \ge 1$.

Using the same arguments as in Case 1, one can show that the probability $\Psi_{i'}$ that the long-range contact of u diminishes the distance on dimension i', from $d_{i'}$ to at most $d_{i'}/k$ is

$$\Psi_{i'} = 1 - \left(1 - \frac{1}{c}\right)^{b'} \ge \frac{b'}{2c} > \frac{b'}{4(b'+1)k \ln m} \ge \frac{1}{8k \ln m}.$$

Overall, the probability Φ that the long-range contact diminishes the distance on fixed dimension i, from d_i to at most d_i/k is

$$\Phi = \frac{q}{s} \cdot \Psi_{i'} \ge \frac{q}{8k \ln n}. \quad \square$$

4.2. Greedy routing

Theorem 5. *Consider a Small-World network with groups $\mathcal{R}_c(n, s, q)$. If $q \le 2 \ln n$ and $c \ge \frac{8 \ln n}{q}$, then the average path length for greedy routing on $\mathcal{R}_c(n, s, q)$ is $O\left(\frac{\log^2 n}{q}\right)$.*

Proof. By following the same arguments as in the proof of Theorem 1, we only need to show that $\Phi \ge \frac{q}{\alpha \log n}$, for some constant α, where Φ denotes the probability that the current node is able to halve the distance on a fixed dimension i. Using Lemma 2 with $k = 2$ we obtain the desired value. $\square$

Using the same argument of the proof above one can easily adapt the proof of Theorem 2 in order to obtain the following result on greedy routing with relative gains.

Theorem 6. *Consider a Small-World network with groups $\mathcal{R}_c(n, s, q)$. For any q, s and c such that $qs \le 4 \ln n$ and $c \ge \frac{8 \ln n}{q}$, the average path length for greedy routing with relative gain on $\mathcal{R}_c(n, s, q)$ is $O\left(\frac{\log^2 n}{q} \cdot \frac{\log s}{s}\right)$.*

4.3. Indirect greedy routing

Consider an indirect greedy routing step as described in Section 3.2. Let u be the node currently holding the message. Consider an integer r such that $N_r(u)$ contains a set $N_c \subseteq N_r(u)$ of nodes which belong to pairwise different groups with $|N_c| = \Omega\left(\frac{k \ln n}{q}\right)$. Fix any dimension i for which the distance from u to the target on dimension i is $d_i > r$ (otherwise we route on this dimension using short-range links).

For each node in N_c, consider the event that one of its long-range contacts allows us to diminish in one indirect step the distance on dimension i from d_i to at most d_i/k. Then applying Lemma 2, with $c = \frac{4k \ln n}{q}$, the probability that a generic node in N_c is able to diminish, with one indirect step, the distance to the target on dimension i, from d_i to at most d_i/k is $\frac{q}{16k \ln n}$; since nodes in N_c belong to different groups, such events are independent. Therefore, the overall probability that at least one node in N_c is able to diminish, with one indirect step, the distance to the target on dimension i, from d_i to at most d_i/k is constant.

Moreover, if $c \geq \frac{4k \ln n}{q}$ and $|N_r(u)| \geq \frac{k \ln n}{q}$ (using Chernoff's bound) we have $|N_c| = \Omega\left(\frac{k \ln n}{q}\right)$ with probability larger than $1 - e^{-\frac{k \ln n}{\alpha q}}$, for some constant α. We can repeat the proof of Theorem 3 by noticing that $q|N_c|$ long-range contacts are available at each greedy step. Hence, we can reach the destination using $O(\log_k n)$ indirect greedy routing steps. Therefore, since $r = O\left(s\left(\frac{k \ln n}{q}\right)^{1/s}\right)$, choosing $k = e^s$, the average path length is the same as for Restricted-Small-World.

Corollary 1. *Consider a Small-World network with groups $\mathcal{R}_c(n, s, q)$. If $q \leq e^s \ln n$, $c \geq \frac{4e^s \ln n}{q}$ and each node knows the long-range contacts of its $\frac{e^s \ln n}{q}$ closest neighbors, then the average path length for Indirect greedy routing on $\mathcal{R}_c(n, s, q)$ is $O\left(\frac{\log^{1+1/s} n}{q^{1/s}}\right)$.*

Remark. We observe that in this case, in order to have only $O(\log \log n)$ random bits, the value of s should be assumed to be upper bounded by $\log \log n$. However, this is not a severe constraint since it makes no sense to consider the performance of a greedy routing approach when each node is aware of a big portion of the network (that is, more than polylogarithmic in the network size).

4.4. NoN greedy routing

In order to be able to analyze NoN greedy routing in SW networks with groups, we need the following auxiliary result.

Consider a node $u = \langle u_1, \ldots, u_s\rangle$ that holds a message addressed to the target node $t = \langle t_1, \ldots, t_s\rangle$ at directed distance $d = \sum_{i=1}^{s} d_i$, where d_i denotes the directed distance between u and t on dimension i, that is, $d_i = (t_i \ominus u_i)$, for each $i = 1, \ldots, s$.

Lemma 3. *Consider a Small-World network with groups $\mathcal{R}_c(n, s, q)$ with $1 < q \leq \log n$ and $c > \frac{8 \ln n}{\log q}$. Let $k = \left\lfloor \frac{q}{\log q} \right\rfloor$ and $R = \{j \mid 1 \leq j \leq s,\ d_j > 0\}$.*
Denote by ϕ the event that for a fixed value of $i \in R$, the node u holding the message is able to diminish the remaining distance d_i, on dimension i, to at most d_i/k in two hops.
If $\sum_{j \in R} \log d_j > \frac{\log n}{\log q}$, then the probability that the event ϕ occurs is $\Phi = \Omega\left(\frac{q}{\log n}\right)$.

Proof. Consider a fixed dimension $i \in R$.

Let Ψ denote the probability that the node u currently holding the message has a link to a node having distance at most d_i/k from t in dimension i. Recalling that $k = \lfloor q/\log q\rfloor$, we have $c > \frac{8 \ln n}{\log q} \geq \frac{8k \ln n}{q}$. By this and Lemma 2, we can easily derive that

$$\Psi \geq \frac{q}{16k \log n}. \tag{12}$$

Let $P = \{v \in L(u) \mid \vec{d}(v, t) \leq \vec{d}(u, t)\}$. By construction, each long-range contact $v \in P$ differs from u only on one dimension, hence following the long-range link from u to v one modifies the distance to the target only on one dimension. It follows that

$$P = \{v \in L(u) \mid \text{there exists } j \in R \text{ s.t. } \vec{d}(u, v) = v_j \ominus u_j \leq d_j\}.$$

First we will bound the size of P. There are two cases to consider depending on the relative values of q and s.
Case 1: $q \geq s$.

For each dimension $j \in R$ and integer $a \leq \lfloor c \log d_j/\log \gamma_j\rfloor$, let $v_{j,a}$ be the node differing from u in dimension j only such that $\vec{d}(u, v_{j,a}) = \left\lfloor \gamma_j^{a/c} \right\rfloor$. The following properties hold:

(a) If $a \leq c \log d_j/\log \gamma_j$ then $\vec{d}(u, v_{j,a}) = \left\lfloor \gamma_j^{a/c} \right\rfloor \leq \left(\gamma_j^{1/c}\right)^a \leq d_j$.
(b) $v_{j,a}$ is an u's long-range contact with probability $1/c$.

To see (b), we notice that by using (iii) of Definition 3, $v_{j,a}$ is a long-range contact of u iff $c_u = a \bmod c$; (b) follows since c_u is uniformly chosen among $0, \ldots, c-1$.

By (a) and (b), we have that the node $v_{j,a} \in P$ with probability $1/c$, for each $a = 0, \ldots, \lfloor c \log d_j / \log \gamma_j \rfloor$.
By summing up on each dimension $j \in R$, we have that the expected size of P satisfies

$$|P| \geq \sum_{j\in R} \frac{\lfloor c \log d_j/\log \gamma_j \rfloor + 1}{c} \geq \sum_{j\in R} \frac{c \log d_j/\log \gamma_j}{c} = \sum_{j\in R} \frac{\log d_j}{\log \gamma_j}$$
$$= \sum_{j\in R} \frac{q_j \log d_j}{\log m} \geq \sum_{j\in R} \frac{\lfloor q/s \rfloor \log d_j}{\log m} > \frac{q}{2\log n} \sum_{j\in R} \log d_j > \frac{q}{2\log q},$$

where the inequalities hold since we know that $\log \gamma_j = (\log m)/q_j$, $q_j \geq \lfloor q/s \rfloor$, and $q/(2s) < \lfloor q/s \rfloor$ when $q \geq s$.
Case 2: $q < s$.

For each dimension $j \in R$ holding one of u's long-range contacts, and for each $a \leq \lfloor c \log d_j / \log m \rfloor$ let $v_{j,a}$ be the node such that, u and $v_{j,a}$ differ in dimension j only and $\vec{d}(u, v_{j,a}) = \lfloor m^{a/c} \rfloor$.
We first notice that if $(m^{1/c})^a \leq d_j$ then $\vec{d}(u, v_{j,a}) = \lfloor m^{a/c} \rfloor \leq (m^{1/c})^a \leq d_j$.

Moreover, according to (6), node u has one of its long-range contacts along dimension j with probability q/s, for any $j = 1, \ldots, s$. Finally, by (iii) of Definition 3, we know that $v_{j,a}$ is an u's long-range contact iff $c_u = a$. Allover, we obtain that $v_{j,a} \in P$ with probability $\frac{q}{sc}$, for each $a = 0, \ldots, \lfloor c \log d_j / \log m \rfloor$.
Therefore, the expected number of u's long-range contacts on dimension j having distance at most d_j from t are at least $q(\lfloor c \log d_j / \log m \rfloor + 1)/sc$.

By summing up on each dimension $j \in R$ we have that the expected size of P satisfies

$$|P| \geq \sum_{j\in R} \frac{q}{sc} \left(\left\lfloor c \frac{\log d_j}{\log m} \right\rfloor + 1 \right) \geq \sum_{j\in R} \frac{q}{s} \frac{\log d_j}{\log m} \geq \frac{q}{\log n} \sum_{j\in R} \log d_j > \frac{q}{\log q}.$$

Consequently (using the Chernoff bound [22]) we have that $|P| = \Omega\left(\frac{q}{\log q}\right)$ with probability larger than $1 - e^{-\frac{q}{\alpha \log q}}$ for some constant α.

Let $P_c \subseteq P$ a maximal subset of nodes which belong to different groups, using $c > \frac{8 \ln n}{\log q}$ we have $|P_c| = \Omega\left(\frac{q}{\log q}\right)$.

We remark that when a long-range is drawn along dimension j the distance along any other dimension does not change. Therefore, if we consider a long-range contact in P_c both along distance i or any $j \neq i$ we have that the remaining distance along dimension i is at most d_i.

Since nodes in P_c belong to different groups, the events that one of them has a link at distance at most d_i/k from t in dimension i are independent from one node to another. Hence, we can apply equation (12) to have that the probability that there exist one node in P_c that has a link at distance at most d_i/k from t in dimension i is

$$\Phi \geq 1 - (1 - \Psi)^{|P_c|} \geq \Omega\left(\frac{q}{k \log n} \cdot \frac{q}{\log q}\right) = \Omega\left(\frac{q}{\log n}\right),$$

where the inequality holds since $\Psi \cdot |P_c| < 1$. □

Theorem 7. *Consider a Small-World network with groups $\mathcal{R}_c(n, s, q)$. If $1 < q \leq \log n$ and $c > \frac{8 \ln n}{\log q}$, then the average path length for Neighbor-of-Neighbor greedy routing on $\mathcal{R}_c(n, s, q)$ is $O\left(\frac{\log^2 n}{q \log q}\right)$.*

Proof. There are two cases to take into consideration depending on the distance $d = \sum_{j\in R} d_j$ from the node u currently holding the message to the target t.
Case 1: $\sum_{j\in R} \log d_j \leq \frac{\log n}{\log q}$.

The remaining distance can be covered using *greedy routing*. Since $c > \frac{8 \ln n}{\log q}$, using Lemma 2 with $k = 2$ we know that the probability that the current node is able to diminish the distance along the dimension j from d_j to $d_j/2$ in one hop is lower bounded by $\frac{q}{32 \log n}$.

Then the expected number of required hops is

$$\frac{32 \log n}{q} \sum_{j\in R} \log d_j = O\left(\frac{\log^2 n}{q \log q}\right).$$

Case 2: $\sum_{j\in R} \log d_j > \frac{\log n}{\log q}$. By Lemma 3, the expected number of nodes encountered before a successful event ϕ occurs is $O\left(\frac{\log n}{q}\right)$. Since on each dimension j the initial distance is d_j, the maximum number of times the remaining distance could possibly be diminished is $\log_k d_j$. Therefore since $d_j < m$, it follows that the average number of hops we need on each dimension is $O\left(\frac{\log m}{\log\left(\frac{q}{\log q}\right)} \frac{\log n}{q}\right)$ and consequently the average path length is $O\left(\frac{\log^2 n}{q \log q}\right)$. Thus, in $O\left(\frac{\log^2 n}{q \log q}\right)$ hops, the distance is decreased to d^*, where $\sum_{j\in R}\log d_i^* \le \frac{\log n}{\log q}$, and we have reduced case 2 into case 1. Hence, the average total number of hops needed is $O\left(\frac{\log^2 n}{q\log q}\right)$. $\square$

4.5. A lower bound

In this section we present a lower bound on the diameter and the average distance of any Small-World network with groups $\mathcal{R}_c(n, s, q)$ having c groups and $O(\log n)$ degree.

Definition 4 (*Uniform Grid Networks [3]*). Consider a bidirectional s-dimensional toroidal grid ($\{0, \ldots, m-1\}^s$) augmented with q long-range contacts for each node; call K the resulting graph. K is called *uniform* iff there exist q vectors (a.k.a. long-range contact distances) $x_i = \langle x_{i_1}, \ldots x_{i_s}\rangle$, such that, each node $v = \langle v_1, \ldots, v_s\rangle \in K$ has exactly q long-range contacts to nodes $w_i = \langle v_1 \oplus x_{i_1}, \ldots, v_s \oplus x_{i_s}\rangle$ for $i = 1, \ldots, q$.

The following Lemmas were proved in [3]

Lemma 4. *[3] Let c be any function such that $2 \le c \le \log n$. The diameter of any uniform network having degree $O(c\log n)$ is $\Omega(\log_c n)$.*

Lemma 5. *[3] The average path length of any uniform network having degree $\delta(n)$ and diameter $d(n) = O(\delta(n))$ is $\Omega(d(n))$.*

Theorem 8. *A Small-World network with groups ($\mathcal{R}_c(n, s, q)$), having $c \ge 2$ groups and degree $q + 2s = O(\log n)$ has both diameter and average (shortest) path length lower bounded by $\Omega(\log_c n)$.*

Proof. We consider a generic Small-World network having $c \ge 2$ groups, n nodes, and degree $\delta' = q + 2s$.

Let $\overline{K}$ be the network obtained by augmenting K in such a way that each node in $\overline{K}$ maintains $cq+2s$ connections. Namely, each node u has all the q connections corresponding to its membership to group ℓ, for each $\ell = 0, \ldots, c-1$.

We denote by δ and $d(n)$, respectively, the degree and the diameter of $\overline{K}$. Obviously $\delta = cq + 2s = O(c\log n)$.

It is easy to observe that: (a) $\overline{K}$ is a *uniform grid network* (all the nodes maintain δ symmetric connections); (b) $\overline{K}$ has diameter and average path length smaller than K ($\overline{K}$ is in fact obtained by adding some connection to K). Hence, in order to bound the diameter of K (with degree δ'), we can apply to $\overline{K}$ (with degree δ) the lower bound given in Lemma 4. Moreover, by Lemma 5, we get that the same bound holds on the average path length of K. $\square$

Corollary 2. *The optimal path length (that is $O(\log n/\log\log n)$) on a Small-World network with groups ($\mathcal{R}_c(n, s, q)$), having $q + 2s = O(\log n)$, can be obtained only using at least $\Omega(\log n/\log\log n)$ groups.*

5. Conclusions

Theorems 5 and 7 answer positively to our initial question: *Do the long-range contacts really need to be completely random, or some "long-range clustering" could instead be envisaged in such a "navigable" network?* In a sense, we show that it is not necessary to use a completely heterogeneous network in order to obtain a SW. Indeed, such result can be obtained using a limited amount of heterogeneity, namely only $O(\log n/\log q)$ groups. Theorem 8 shows that our construction is asymptotically optimal – when the degree is $O(\log n)$ – in the sense that in order to obtain the same performance as Kleinberg networks, any network requires at least $\Omega(\log n/\log\log n)$ groups.

Moreover, a Small-World network having c groups requires only $\log c + \log s$ random bits, for establishing all the q long-range contacts of each node – where n is the number of nodes in the network, s is network's dimension (notice that when s is close to $\log n$ it makes no sense to augment the network, since the diameter is already small) and c is the number of groups – instead of $\Omega(q\log n)$ required in the original Kleinberg construction.

In addition to their theoretical interest, such networks can be used toward the design of efficient, as well as easy to implement, network infrastructures based on the SW approach. Diminishing the amount of randomness used for random links increases the clustering of the network. Hence, one can get interconnected networks which, in addition to possessing convenient graph properties (such as low average path length and degree) and besides providing efficient and easy routing algorithms (as offered by Kleinberg's model) offer an increased resilience to simultaneous node failures and tolerance to hot spot workload (due to a higher clustering).

As a consequence of limiting the range in which a node can choose its long-range contacts, the analysis of some Small-World properties becomes easier. It also allows us to provide results without any hidden factor (depending on s) (cf. Remark 2).

Acknowledgments

The authors would like to thank the anonymous referees whose very helpful comments on a previous version of this paper allowed to significantly improve the presentation of their work and for suggesting the use of the strategy presented in Theorem 2.

References

[1] Lali Barrière, Pierre Fraigniaud, Evangelos Kranakis, Danny Krizanc, Efficient routing in networks with long range contacts, in: Proc. of 15th International Symposium on Distributed Computing, DISC'01, in: LNCS, vol. 2180, 2001, pp. 270–284. Oct..

[2] Markus Brede, John Finnigan, Growing networks with enhanced resilience to perturbation, in: ArXiv Condensed Matter e-prints, http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0405076, 2004.

[3] Giovanni Chiola, Gennaro Cordasco, Luisa Gargano, Mikael Hammar, Alberto Negro, Vittorio Scarano, Degree-optimal routing for P2P systems, Theory of Computing Systems (ToCS) (ISSN: 1432-4350) 45 (1) (2009) 43–63.

[4] Gennaro Cordasco, Luisa Gargano, How much independent should individual contacts be to form a small–world? in: Proc. of 17th International Symposium on Algorithms and Computation, ISAAC'06, Dec. 2006, Kolkata, India.

[5] John H. Conway, Neil J.A. Sloane, Low dimensional lattices VII: Coordination sequences, Proceedings of the Royal Society of London, Series A 453 (1997) 2369–2389.

[6] Philippe Duchon, Nicolas Hanusse, Emmanuelle Lebhar, Nicolas Schabanel, Could any graph be turned into a small-world? Theoretical Computer Science 355 (1) (2006) 96–103.

[7] Paul Erdös, Alfréd Rényi, On the evolution of random graphs, Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5 (1960) 17–61.

[8] Michele Flammini, Luca Moscardelli, Alfredo Navarra, Stéphane Pérennes, Asymptotically optimal solutions for small world graphs, Theory of Computing Systems (TOCS) 42 (4) (2008) 632–650.

[9] Pierre Fraigniaud, Greedy routing in tree-decomposed graphs, in: 13th Annual European Symposium on Algorithms (ESA), 2005, pp. 791–802.

[10] Pierre Fraigniaud, Cyril Gavoille, Christophe Paul, Eclecticism shrinks even small worlds, Distributed Computing 18 (4) (2006) 279–291.

[11] Pierre Fraigniaud, Cyril Gavoille, Adrian Kosowski, Emmanuelle Lebhar, Zvi Lotker, Universal augmentation schemes for network navigability: Overcoming the $\sqrt{n}$-Barrier, in: Proceedings of the Nineteenth Annual ACM Symposium on Parallel Algorithms and Architectures, SPAA'07, 2007, pp. 1–7.

[12] Pierre Fraigniaud, Emmanuelle Lebhar, Zvi Lotker, A doubling dimension threshold $\Theta(\log\log n)$ for augmented graph navigability, in: 14th Annual European Symposium on Algorithms (ESA), 2006, pp. 376–386.

[13] Ken Y.H. Hui, Jon C.S. Lui, David K.Y. Yau, Small world overlay P2P networks, in: Proc. of 12th Intl. Workshop on Quality of Service, IWQoS'04, 2004.

[14] Jon M. Kleinberg, The Small–World phenomenon: An algorithm perspective, in: Proc. of 32th ACM Symp. on Theory of Computing, STOC'00, May 2000, pp. 163–170.

[15] Jon M. Kleinberg, Complex networks and decentralized search algorithms, in: International Congress of Mathematicians (ICM), 2006.

[16] Emmanuelle Lebhar, Nicolas Schabanel, Close to optimal decentralized routing in long-range contact networks, Theoretical Computer Science 348 (2-3) (2005) 294–310.

[17] Dmitri Loguinov, Anuj Kumar, Vivek Rai, Sai Ganesh, Graph-theoretic analysis of structured peer-to-peer systems: Routing distances and fault resilience, in: Proc. of the ACM SIGCOMM '03 Conference, Aug. 2003, Karlsruhe, Germany, pp. 395–406.

[18] Stanley Milgram, The Small World Problem, Psychology Today (1967) 60–67. May.

[19] Gurmeet Singh Manku, Mayank Bawa, Prabhakar Raghavan, Symphony: Distributed hashing in a small world, in: Proc. of USENIX Symp. on Internet Technologies and Systems, USITS'03, Mar. 2003.

[20] Gurmeet Singh Manku, Moni Naor, Udi Wieder, Know thy neighbor's neighbor: The power of lookahead in randomized P2P networks, in: Proc. of 36th ACM Symp. on Theory of Computing, STOC'04, Jun. 2004, pp. 54–63.

[21] Chip Martel, Van Nguyen, Analyzing Kleinberg's (and other) Small–World models, in: Proc. of 23rd ACM Symposium on Principles of Distributed Computing, PODC'04, Jul. 2004, pp. 179–188.

[22] Rajeev Motwani, Prabhakar Raghavan, Randomized Algorithms, Cambridge University Press, 1995.

[23] David Newth, Jeff Ash, Evolving cascading failure resilience in complex networks, in: Proc. of 8th Asia Pacific Symp. on Intelligent and Evolutionary Systems, Dec. 2004.

[24] Robert A. Sulanke, Objects counted by the central delannoy numbers, Journal of Integer Sequences 6 (2003).

[25] Aleksandrs Slivkins, Distance estimation and object location via rings of neighbors, in: Proceedings of the 24th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing, 2005.

[26] Ion Stoica, Robert Morris, David Liben-Nowell, David R. Karger, M. Frans Kaashoek, Frank Dabek, Hari Balakrishnan, Chord: A scalable peer-to-peer lookup protocol for internet applications, IEEE/ACM Transactions on Networking 11 (1) (2003) 17–32.

[27] Duncan J. Watts, Steven H. Strogatz, Collective dynamics of 'smallworld' networks, in: Nature, vol. 393, Macmillan Publishers Ltd, 1998, pp. 440–442. Jun.