Mean Eccentricities of de Bruijn Networks

Jean-Claude Bermond
I3S, CNRS-Université de Nice-Sophia Antipolis
Bât 4, 250 avenue A. Einstein, Sophia Antipolis
06560 Valbonne, France

Zhen Liu
INRIA, Centre de Sophia Antipolis
2004 route des Lucioles
06560 Valbonne, France

Michel Syska*
I3S, CNRS-Université de Nice-Sophia Antipolis
Bât 4, 250 avenue A. Einstein, Sophia Antipolis
06560 Valbonne, France

August 6, 1993

Abstract: Given a graph $G = (V, E)$ we define $\bar{e}(X)$, the mean eccentricity of a vertex X, as the average distance from X to all the other vertices of the graph. The computation of this parameter appears to be nontrivial in the case of the de Bruijn networks. In this paper we consider upper and lower bounds for $\bar{e}(X)$. For the directed de Bruijn network, we provide tight bounds as well as the extremal vertices which reach these bounds. These bounds are expressed as the diameter minus some constants. In the case of undirected networks, the computation turns out to be more difficult. We provide lower and upper bounds which differ from the diameter by some small constants. We conjecture that the vertices of the form $a \cdots a$ have the largest mean eccentricity. Numerical computations indicate that the conjecture holds for binary de Bruijn networks with diameters up to 18. We prove that the asymptotic difference, when the diameter goes to infinity, between the mean eccentricities of an arbitrary vertex and that of $a \cdots a$ is smaller than a small constant tending to zero with the degree. We also provide a simple recursive scheme for the computation of the asymptotic mean eccentricity of the vertices $a \cdots a$. A by-product of our analysis is that in both directed and undirected de Bruijn networks, most of the vertices are at distance near from the diameter and that all of the mean eccentricities tend to the diameter when the degree goes to infinity.

Keywords: Interconnection network, de Bruijn network, distances, mean eccentricities, bounds.

*This author is currently supported by INRIA while visiting Simon Fraser University, B.C. Canada

1 Introduction and Notation

Graphs are widely used in the design and analysis of computer networks. A vertex in the graph denotes a node or processor in the corresponding network, and an edge denotes a communication link between two nodes. If a network is unidirectional, i.e., the communication links in the network are unidirectional, a digraph (i.e., directed graph) is used. Whereas for a bidirectional network, an undirected graph (or simply, graph) is used.

Let $G = (V, E)$ be a (strongly) connected graph (or digraph), where V denotes the set of vertices, and E the set of edges (or arcs for digraphs). We will denote by $N = |V|$ the number of vertices in G.

The vertex X is a *neighbor* of the vertex Y if $(X, Y) \in E$ or $(Y, X) \in E$. The *degree* of a vertex is the number of its neighbors. The *degree* of a graph is the maximum degree of the vertices. In case of a digraph, we can distinguish the predecessors and successors of vertex X, which correspond to the neighbors Y which satisfy $(Y, X) \in E$ and $(X, Y) \in E$, respectively. The number of arcs entering (resp. going out from) a vertex X is called the in-degree (resp. out-degree) of X. The in-degree (resp. out-degree) of a graph is the maximum in-degree (resp. out-degree) of the vertices.

A *path* (resp. a *dipath*) between two vertices X and Y (resp. from X to Y) in a graph (resp. a digraph) G is a sequence of vertices $\{X = X_1, X_2, \cdots, X_k = Y\}$ such that two consecutive vertices in the sequence are joined by an edge (resp. an arc). The *length* of a path is the number of edges on this path. The length of a shortest path (resp. dipath) between X and Y (from X to Y) is called the *distance* and is denoted by $d(X, Y)$. Note that in the case of digraphs it is not a classical distance as $d(X, Y)$ might be different from $d(Y, X)$. The *diameter* of a graph is the maximum distance in the graph.

The *de Bruijn* digraph (resp. graph), denoted by $B(d, D)$ (resp. $UB(d, D)$), has $N = d^D$ vertices with diameter D and in-degree or out-degree d (resp. degree $2d$). The vertices correspond to the words of length D over an alphabet of d symbols. The arcs (or edges) correspond to the shift operations: Given a word $X = x_1 \cdots x_D$ on an alphabet $\mathcal{A}$ of d letters, where $x_i \in \mathcal{A}$, $i = 1, 2, \cdots, D$, and given $\lambda \in \mathcal{A}$, the operation:

- $x_1 \cdots x_D \longrightarrow x_2 \cdots x_D \lambda$ is called a left shift;
- $x_1 \cdots x_D \longrightarrow \lambda x_1 \cdots x_{D-1}$ is called a right shift.

In the de Bruijn digraph $B(d, D)$, the successors are obtained by left-shift operations, whereas in the de Bruijn graph $UB(d, D)$, the neighbors are obtained by either left or right shift operations. An example of a de Bruijn digraph is given in Figure 1. The corresponding undirected de Bruijn graph is obtained by transforming arcs to edges (i.e., removing the directions of the arcs) and

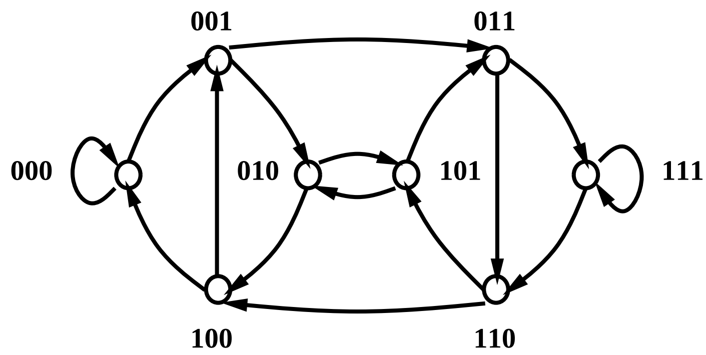

Figure 1: Example of a de Bruijn digraph $B(2,3)$.

removing the redundant edges (i.e., those with multiple occurrences in the graph, or those linking the same vertices). The multiprocessor system which can be modeled by a de Bruijn graph is called de Bruijn network.

The reader can see that in $B(d,D)$, each vertex has in-degree and out-degree d, and there are Nd arcs. Whereas in $UB(d,D)$, there exist $N-d^2$ vertices of degree $2d$, d^2-d vertices of degree $2d-1$ and d vertices of degree $2d-2$.

These networks have been discovered by many authors and are named after de Bruijn [7]. They are sometimes also called Good graphs [9].

They present many attractive features. In particular they are among the best known networks for a given degree and diameter (see the survey [2] for more details on this problem known as the (d,D) or (Δ,D) graph problem). They have a good vulnerability, being able to tolerate up to $d-1$ faults in the directed case and $2d-2$ in the undirected case, while the diameter can still be left small (see [3]). They are adequate for various applications as one can embed in them linear arrays, rings, and complete binary trees. They can also emulate without loss of time shuffle-exchange or hypercubes for the class of ascend-descend algorithms.

They have also many others interesting properties like easy greedy routing. This greedy algorithm is a simple D-step routing which consists in uni-directional shifts or "bit-erosion" of the destination address. Namely, in order to go from $X = x_1 \cdots x_D$ to $Y = y_1 \cdots y_D$, we apply D left shifts (and also right shifts in the undirected case) introducing successively the letters $y_1, y_2, \cdots, y_D$, corresponding to the dipath

$$x_1 \cdots x_D \rightarrow x_2 \cdots x_D y_1 \rightarrow x_3 \cdots x_D y_1 y_2 \rightarrow \cdots \rightarrow y_1 \cdots y_D.$$

We refer the reader to one of the two recent surveys concerning de Bruijn networks written by Bermond and Peyrat [4] and Samatham and Pradhan [13] or to the recent book of Leighton [10].

In this paper, we analyze the mean eccentricity of these graphs. The *eccentricity* of a vertex X is defined [6] as the distance to the farthest node from this vertex: $e(X) = \max\{d(X,Y) : Y \in V\}$. We define the *mean eccentricity* of a vertex X, denoted $\bar{e}(X)$, as the average distance from X to all the others vertices:

$$\bar{e}(X) = \frac{1}{N-1} \sum_{Y \in V - \{X\}} d(X,Y). \tag{1.1}$$

The computation of this parameter appears to be nontrivial in the case of the de Bruijn networks. In this paper we consider the upper and lower bounds of $\bar{e}(X)$.

In Section 2, we analyze the directed de Bruijn networks. We provide the tight bounds as well as the extremal vertices which reach these bounds. These bounds are expressed as the diameter minus some constants.

In Section 3, we analyze the undirected networks. The computation turns out to be more difficult. We provide lower and upper bounds which differ from the diameter by some small constants. We conjecture that the vertices of the form $a \cdots a$ have the largest mean eccentricity. Numerical computations indicate that the conjecture holds for binary de Bruijn networks with diameters up to 18. We will show that the asymptotic difference, when the diameter goes to infinity, between the mean eccentricities of an arbitrary vertex and that of $a \cdots a$ is smaller than 0.22. We also provide a simple recursive scheme for the computation of the asymptotic mean eccentricity of the vertices $a \cdots a$.

A by-product of our analysis is that in both directed and undirected de Bruijn networks, most of the vertices are at distance near from the diameter and that all of the mean eccentricities tend to the diameter when the degree goes to infinity.

Our results also imply that optimal routing algorithms are in most of the cases not advantageous due to their overheads. Instead, one can use the simple D-step routing algorithm as described above.

The following formulae will be extensively used in this paper:

$$\sum_{k=0}^{p} d^k = \frac{d^{p+1}-1}{d-1}. \tag{1.2}$$

$$\sum_{k=1}^{p} k d^{k-1} = \frac{(p+1)d^p}{d-1} - \frac{d^{p+1}-1}{(d-1)^2} = \frac{pd^{p+1} - (p+1)d^p + 1}{(d-1)^2}. \tag{1.3}$$

2 Directed case

In the case of a digraph $B(d, D)$, it is well known (see e.g. Fiol et al. [8]) that there is a unique shortest dipath from a given vertex $X = x_1 \cdots x_D$ to a vertex $Y = y_1 \cdots y_D$. To find the distance $d(X,Y)$ and the shortest dipath one has to find the smallest i such that $x_{i+1} \cdots x_D = y_1 \cdots y_{D-i}$.

The distance is then i, and the shortest path is obtained by doing the left shifts introducing successively $y_{D+1-i}, \cdots, y_D$. This fact allows one to compute easily $d(X,Y)$ and so $\bar{e}(X)$ for any X. But unfortunately that does not give a closed formula.

In this section, we provide upper and lower bounds of the mean eccentricities of $B(d,D)$, and show that these bounds are reached.

Our analysis will need some notions on trees (or more precisely, the outtrees). The *level* of a vertex in a tree is the distance from the root to the vertex, where by convention, the root is at level 0. The *weight* of a tree is the sum of the levels of all vertices. We call a *shortest path tree* of a (di)graph G rooted at vertex X a spanning tree of G with root X such that the (di)path in the tree from X to any vertex is a shortest (di)path in G. It can be obtained, e.g., using the "breadth first" search algorithm. Observe that for any given G and X, this tree is not unique. However, the vertices at level l of a shortest path tree of G rooted at vertex X are exactly all the vertices that are at distance l from X in G. Therefore, the weights of the shortest path trees rooted at vertex X are identical. The mean eccentricity of a vertex is in fact the weight of a shortest path tree rooted at this vertex divided by $N-1$.

Let $l_k(X)$ be the *width*, i.e., the number of vertices, at level k in the shortest path trees rooted at vertex X. Then,

$$\bar{e}(X) = \frac{1}{N-1}\sum_{k=1}^{D} k \cdot l_k(X). \tag{2.1}$$

We first look at the minimal mean eccentricity. We will need the following on the comparison result.

Lemma 2.1 *Let n be a positive integer, and $m_1, \cdots, m_n$ positive real numbers such that*

$$1 \le n \le D-1, \qquad \sum_{k=1}^{n} m_k \le N-1.$$

If for all $1 \le k \le n$, $l_k(X) \le m_k$, then

$$\bar{e}(X) \ge n+1-\frac{1}{N-1}\left(\sum_{k=1}^{n}(n+1-k)m_k\right). \tag{2.2}$$

Proof: Let $m_{n+1} = N-1-\sum_{k=1}^{n} m_k$, and $m_i = 0$ for all $n+1 < i \le D$. It then follows that

$$N-1 = \sum_{k=1}^{D} l_k(X) = \sum_{k=1}^{D} m_k,$$

and that for all $2 \le i \le D$,

$$\sum_{k=i}^{D} l_k(X) = \sum_{k=1}^{D} l_k(X) - \sum_{k=1}^{i-1} l_k(X) \ge \sum_{k=1}^{D} m_k - \sum_{k=1}^{i-1} m_k = \sum_{k=i}^{D} m_k.$$

Thus, according to (2.1),

$$\begin{aligned}
\bar{e}(X) &= \frac{1}{N-1}\sum_{k=1}^{D} k \cdot l_k(X) \\
&= \frac{1}{N-1}\sum_{i=1}^{D}\sum_{k=i}^{D} l_k(X) \\
&\ge \frac{1}{N-1}\sum_{i=1}^{D}\sum_{k=i}^{D} m_k \\
&= \frac{1}{N-1}\sum_{k=1}^{D} k m_k \\
&= \frac{1}{N-1}\left(\sum_{k=1}^{n} k m_k + (n+1)(N-1-\sum_{k=1}^{n} m_k)\right) \\
&= n+1-\frac{1}{N-1}\left(\sum_{k=1}^{n}(n+1-k)m_k\right).
\end{aligned}$$

∎

Proposition 2.1 *For any $d \ge 2$ and $D \ge 2$, and for all $X = x_1 \cdots x_D \in B(d,D)$,*

$$\bar{e}(X) \ge D - \frac{d}{(d-1)^2} + \frac{d}{d-1} \cdot \frac{D}{N-1}. \tag{2.3}$$

Moreover, the equality holds if and only if $x_i \ne x_D$ for all $i = 1, \cdots, D-1$.

Proof: The minimal mean eccentricity corresponds to the shortest path tree with the minimal weight, which, owing to Lemma 2.1, requires that as many as possible vertices should be put in the smallest levels. Since $B(d,D)$ has out-degree d, a vertex has at most d vertices at distance 1, and so d^2 vertices at distance 2, $\cdots$, and d^k vertices at distance k. Therefore, in the shortest path trees, there are at most d^k vertices at level k, $k = 1, 2, \cdots, D-1$. As $1 + d + d^2 + \cdots + d^{D-1} = \frac{d^D - 1}{d-1}$ we have still at least $x = d^D - \frac{d^D-1}{d-1}$ vertices which are at level D. Figure 2 illustrates such a tree for the case $d = 2$.

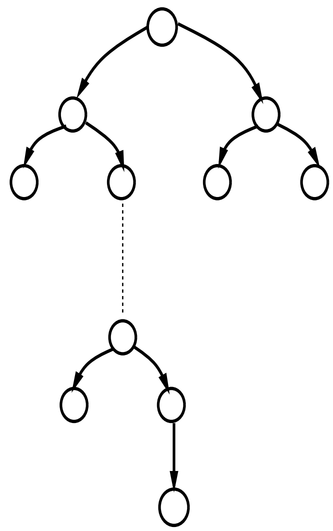

Figure 2: Shortest path tree with the minimal weight

Therefore, we can apply Lemma 2.1 with $n = D-1$ and $m_k = d^k$ for $1 \leq k \leq D-1$.

$$\begin{aligned}
\bar{e}(X) &\geq D - \frac{1}{N-1}\left(\sum_{k=1}^{D-1}(D-k)d^k\right) \\
&= D - \frac{1}{N-1}\left(D\frac{d^D - d}{d-1} - \frac{Dd^D}{d-1} + \frac{d^{D+1}-d}{(d-1)^2}\right) \\
&= D - \frac{d}{(d-1)^2} + \frac{d}{d-1}\cdot\frac{D}{N-1}.
\end{aligned}$$

Assume now the vertex $X = x_1 \cdots x_D$ is such that $x_i \neq x_D$ for all $i = 1, \cdots, D-1$. We want to show the equality in (2.3). Indeed, for all $1 \leq k \leq D-1$, the vertices at distance k from $x_1 \cdots x_D$ are of the form: $x_{k+1} \cdots x_D y_1 \cdots y_k$, where $y_1 \cdots y_k$ are arbitrary letters of the alphabet $\mathcal{A}$. For all $1 \leq j < k$, the vertices at distance j from $x_1 \cdots x_D$ are of the form: $x_{j+1} \cdots x_{D-k+j} \cdots x_D z_1 \cdots z_j$, where $z_1 \cdots z_j$ are arbitrary letters of the alphabet $\mathcal{A}$. Since $x_{D-k+j} \neq x_D$, none of the vertices $x_{k+1} \cdots x_D y_1 \cdots y_k$ are at distance $j < k$. Thus, the width of the shortest path tree of X at level k, $1 \leq k \leq D-1$, is d^k, so that the equality in (2.3) holds.

Assume now that for some vertex $X = x_1 \cdots x_D$, the equality in (2.3) holds. Then, necessarily, the width of its shortest path tree at level $D-1$ is d^{D-1}, so that all the vertices at level $D-1$ which have the form $x_D y_1 \cdots y_{D-1}$, should not already be at levels $k = 0, \cdots, D-2$, where $y_1 \cdots y_{D-1}$ are arbitrary letters of the alphabet $\mathcal{A}$. Thus, for all $k = 0, \cdots, D-2$, we should have $x_{k+1} \neq x_D$, which completes the proof. ∎

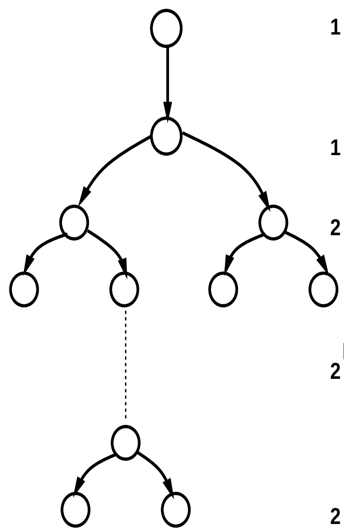

Figure 3: Shortest path tree with the maximal weight

We now look at the upper bound of the mean eccentricities. We will need the following property of the shortest path tree with the maximal weight. An example is given for the case $d = 2$.

Lemma 2.2 *An outtree of out-degree d with d^D vertices and D levels is such that:*

$$\sum_{k=i}^{D} l_k \leq \sum_{k=i}^{D} d^{k-1}(d-1) \quad \forall i,\ 1 \leq i \leq D$$

where l_k is the number of vertices at distance k from the root.

Proof: Suppose the lemma is not true and let i_0 be the largest $i \geq 1$ such that $\sum_{k=i}^{D} l_k > \sum_{k=i}^{D} d^{k-1}(d-1)$. Then, $l_{i_0} > d^{i_0-1}(d-1)$, which in turn, as the out-degree is d, implies $l_{i_0-1} > d^{i_0-2}(d-1)$, which further implies that $l_{i_0-2} > d^{i_0-3}(d-1)$, and so on. Therefore,

$$l_i \geq d^{i-1}(d-1)+1, \qquad 1 \leq i \leq i_0 - 1,$$

which, together with the fact that $\sum_{k=i_0}^{D} l_k \geq 1 + \sum_{k=i_0}^{D} d^{k-1}(d-1)$, imply that

$$\sum_{k=1}^{D} l_k \geq i_0 + \sum_{k=1}^{D} d^{k-1}(d-1) = i_0 + d^D - 1 \geq d^D.$$

This contradicts the fact that the tree has d^D vertices. (Note that the root which is at level 0 is not included in the above summation.) ∎

Proposition 2.2 *For any $d \geq 2$ and $D \geq 2$, and for all $X \in B(d, D)$,*

$$\bar{e}(X) \leq D - \frac{1}{d-1} + \frac{D}{N-1}. \tag{2.4}$$

Moreover, the inequality becomes equality if and only if $X = a \cdots a$, where $a \in \mathcal{A}$.

Proof: Note first that

$$\bar{e}(X) = \frac{1}{N-1} \sum_{k=1}^{D} k \cdot l_k(X) = \frac{1}{N-1} \sum_{i=1}^{D} \sum_{k=i}^{D} l_k(X).$$

It follows from Lemma 2.2 that

$$\begin{aligned}
\sum_{i=1}^{D} \sum_{k=i}^{D} l_k(X) &\leq \sum_{i=1}^{D} \sum_{k=i}^{D} d^{k-1}(d-1) \\
&= (d-1) \sum_{k=1}^{D} k d^{k-1} \\
&= (D+1) d^D - \frac{d^{D+1} - 1}{d-1} \\
&= D d^D - \frac{d^D - 1}{d-1} \\
&= D(N-1) - \frac{N-1}{d-1} + D.
\end{aligned}$$

Therefore,

$$\bar{e}(X) \leq D - \frac{1}{d-1} + \frac{D}{N-1}.$$

We now show that

$$\bar{e}(X) = D - \frac{1}{d-1} + \frac{D}{N-1}$$

if and only if $X = a \cdots a$. Firstly, the vertices at distance k from $a \cdots a$ are of the form: $a_1 \cdots a_D$, with $a_1 = \cdots = a_{D-k} = a$, $a_{D-k+1} = \bar{a}$ and $a_{D-k+1}, \cdots, a_D$ can be any letter of the alphabet $\mathcal{A}$, where $\bar{a} \neq a$. None of these vertices are at distance $j < k$. Their number is $d^{k-1}(d-1)$, so that all the inequalities in the above derivation of the proof become equalities. Secondly, by lemma 2.2, a necessary condition for $\bar{e}(X)$ to reach the bound is that $l_1 = d - 1$. Only the the vertices $a \cdots a$ have this property. ∎

The following properties are immediate consequences:

Corollary 2.1 *For any $d \geq 2$ and $D \geq 2$, and for all $X \in B(d, D)$,*

$$D - \frac{d}{(d-1)^2} + \frac{d}{d-1} \cdot \frac{D}{N-1} \leq \bar{e}(X) \leq D - \frac{1}{d-1} + \frac{D}{N-1}. \tag{2.5}$$

In particular, when $d = 2$,

$$D - 2 + \frac{2D}{N-1} \leq \bar{e}(X) \leq D - 1 + \frac{D}{N-1}. \tag{2.6}$$

Corollary 2.2

$$\lim_{d \to \infty} \bar{e}(X) = D, \qquad D \geq 2.$$

$$\frac{1}{d-1} \leq \liminf_{D \to \infty} (D - \bar{e}(X)) \leq \limsup_{D \to \infty} (D - \bar{e}(X)) \leq \frac{d}{(d-1)^2}, \qquad d \geq 2.$$

Our results indicate that the upper and the lower bounds of the mean eccentricities are both close to the diameter. This is especially true when the degree of the de Bruijn network is large. In [11], a linear (in time and in space with respect to the diameter) algorithm was proposed for the optimal routing of directed de Bruijn networks. However, even if we use D shifts to route instead of the optimal routing, the global communication delay of the network should not be affected too much. The routing scheme in D steps corresponds to a simple "bit-erosion" of the destination address, a very simple routing function that could be implemented with only a very few VLSI components on chip.

3 Undirected case

We now consider the mean eccentricities of $UB(d, D)$. It turns out that the computation of bounds is much more difficult. This is partially due to the facts that the sortest paths are in general not unique and that the determination of the distances between the vertices is more complicate.

In what follows, we will first study some properties of the shortest paths in undirected de Bruijn networks. We then derive the lower and upper bounds respectively.

3.1 Characterization of the Shortest Paths

Since the neighborhood in $UB(d, D)$ can be defined by the left and right shift operations, a path can be described by a sequence of corresponding shifts. The length of a path connecting X to Y is equal to the cardinality of this sequence of shifts. Formally, let L be the operation consisting of a sequence of l left shifts, introducing a suffix $B = b_1, \cdots, b_l$ with $|L| = l$, and let R be the operation consisting in a sequence of r right shifts, introducing a prefix $A = a_r \cdots a_1$ with $|R| = r$. Then, applying L to a vertex $X = x_1 \cdots x_D$ yields vertex $x_{l+1} \cdots x_D b_1 \cdots b_l$, which corresponds to the path

$$x_1 \cdots x_D \rightarrow x_2 \cdots x_D b_1 \rightarrow x_3 \cdots x_D b_1 b_2 \rightarrow \cdots \rightarrow x_{l+1} \cdots x_D b_1 \cdots b_l.$$

Applying R to the vertex $X = x_1 \cdots x_D$ yields vertex $a_r \cdots a_1 x_1 \cdots x_{D-r}$, which corresponds to the path

$$x_1 \cdots x_D \rightarrow a_1 x_1 \cdots x_{D-1} \rightarrow a_2 a_1 x_1 \cdots x_{D-2} \rightarrow \cdots \rightarrow a_r \cdots a_1 x_1 \cdots x_{D-r}.$$

In general, any path of the de Bruijn graph can be denoted by $L_1 R_1 L_2 R_2 \cdots L_n R_n$ with $|L_i| = l_i$ and $|R_i| = r_i$, meaning that we successively apply L_1, then R_1, then L_2 and so on. The length of the path is equal to $l_1 + r_1 + l_2 + r_2 + \cdots l_n + r_n$. Note that one can have $l_1 = 0$ or $r_n = 0$.

We will use the following results established by Bond [5, p92] and Peyrat [12, annexe 3.2]. For sake of completeness, we give a slightly simpler proof. We will also give a more precise property pertaining to the shortest paths from vertices $a \cdots a$ (cf. Proposition 3.2 below).

Lemma 3.1 *If a path of the form LRL' is a shortest path, where $l > 0$, $r > 0$, $l' > 0$, then $l + l' \leq r$. Similarly, if a path of the form RLR' is a shortest path, where $r > 0$, $l > 0$, $r' > 0$, then $r + r' \leq l$.*

Proof: We consider the shortest paths of the form LRL'. The case of RLR' is analogous. We first prove that $l < r$ and $l' < r$.

Given a path of the form LRL' from $X = x_1 \cdots x_D$ to Y, suppose that $l \geq r$, then after the application of L to X we reach vertex $x_{l+1} \cdots x_D b_1 \cdots b_l$. The application of R leads to the vertex $a_r \cdots a_1 x_{l+1} \cdots x_D b_1 \cdots b_{l-r}$. The first shift of L' leads to the vertex $a_{r-1} \cdots a_1 x_{l+1} \cdots x_D b_1 \cdots b_{l-r} \beta$.

But the vertex $a_{r-1} \cdots a_1 x_{l+1} \cdots x_D b_1 \cdots b_{l-r} \beta$ could have been reached by a shorter path $L_1 R_1$ with $l_1 = l$ and $r_2 = r - 1$ (L_1 introduces the word $b_1 \cdots b_{l-r} \beta \cdots$ and R_1 the word $a_{r-1} \cdots a_1$). Therefore, LRL' is not a shortest path.

Thus, we have necessarily $l < r$, and, by symmetry (in considering the reverse path from Y to X, which is also a shortest path), $l' < r$.

Now, the vertices obtained by LRL' are of the form

$$a_{r-l'} \cdots a_1 x_{l+1} \cdots x_{D-(r-l)} b_1 \cdots b_{l'},$$

where $a_1 \cdots a_{r-l'}$ and $b_1 \cdots b_{l'}$ are arbitrary letters. These vertices can also be obtained via the sequence of shifts $R_1 L_1 R_2$ with

$$r_1 = r - l, \quad l_1 = r, \quad r_2 = r - l'.$$

As LRL' is a shortest path, we

$$r_1 + l_1 + r_2 = 3r - l - l' \geq l + r + l',$$

which implies that $l + l' \leq r$. ∎

Proposition 3.1 *A shortest path of $UB(d, D)$ is made of the concatenation of at most 3 directed paths, that is the shortest path is of the form LRL' or RLR', where the cardinalities of these shift sequences can be zero.*

Proof: Suppose that a shortest path from X to Y is made of more than 3 directed paths, then a part of this shortest path is either $L_1R_1L_2R_2$ or $R_1L_1R_2L_2$, with $l_1 > 0$, $r_1 > 0$, $l_2 > 0$ and $r_2 > 0$. Let us assume that it is the first case. It then follows that $L_1R_1L_2R_2$ is a shortest path, so are $L_1R_1L_2$ and $R_1L_2R_2$. By Lemma 3.1, we get $l_2 < r_1$ and $r_1 < l_2$, which is a contradiction. ∎

Proposition 3.2 *A shortest path between a vertex X and the vertex $a \cdots a$ is of the form LR (with $0 \leq r < l$) or RL (with $0 \leq l < r$).*

Proof: Without loss of generality, suppose that the path from $a \cdots a$ to X is of the form LRL'. Suppose $l \leq r$, which is in particular the case if $l' > 0$ by Lemma 3.1. The vertex reached after the operation LR is $a_r \cdots a_1 a \cdots a$, which could have been reached directly by the path R which is shorter. ∎

3.2 Lower Bound

We now consider the lower bound of the mean eccentricity for general $UB(d, D)$.

Proposition 3.3 *For any $D \geq 2$, and any $X \in UB(d, D)$,*

$$\bar{e}(X) \geq \begin{cases} D - 3 - \frac{9}{8} & > \quad D - 4.2, \qquad d = 2; \\ D - 1 - \frac{8}{9} & > \quad D - 1.9, \qquad d = 3; \\ D - 1 - \frac{25}{72} & > \quad D - 1.4, \qquad d = 4; \\ D - \frac{2(d+1)^2}{d(d-1)^2} & \geq \quad D - 0.9, \qquad d \geq 5. \end{cases} \tag{3.1}$$

Proof: Consider the shortest path tree rooted at $X = x_1 \cdots x_D$. As in the case of directed de Bruijn networks, the more vertices are at the first levels of the shortest path tree, the smaller is the mean eccentricity (cf. Lemma 2.1). We will use the fact that the shortest paths are made of the concatenation of at most 3 directed paths to obtain upper bounds of the widths of the first levels in the shortest path tree.

At level k, $1 \le k \le D-1$, we have 3 kinds of vertices:

(a) those obtained with a directed path containing only left (resp. right) shifts. Their number is at most $2A_k$, with

$$A_k = d^k, \qquad 1 \le k \le D-1. \tag{3.2}$$

(b) those obtained with the concatenation of 2 directed paths corresponding to a sequence of LR (resp. RL) shifts with $l > 0$, $r > 0$ and $l + r = k$, where $l = |L|$ and $r = |R|$.

Consider the vertices obtained with a path LR. They are of the form

$$\begin{cases} a_r \cdots a_1 x_{l+1} \cdots x_D b_1 \cdots b_{l-r}, & \text{if } l \ge r; \\ a_r \cdots a_1 x_{l+1} \cdots x_{D-(r-l)} & \text{if } l < r, \end{cases}$$

where $a_r \cdots a_2$ and $b_1 \cdots b_{l-r}$ are arbitrary letters, and $a_1 \ne x_l$. Indeed, if $a_1 = x_l$, these vertices would have been reached with a path L_1R_1 with $l_1 = l-1$ and $r_1 = r-1$, and so are already at level $k-2$. Thus, there are at most $(d-1)d^{m-1}$ vertices reached by such paths, where $m = \max(l, r)$.

Hence, for fixed k, $2 \le k \le D-1$, the number of vertices reached by the concatenation of 2 directed paths is at most $2B_k$, with

$$\begin{aligned} B_k &= \sum_{l=1}^{k-1} (d-1) d^{\max(l,k-l)-1} \\ &= 2(d-1)\left(\sum_{l=\lceil \frac{k}{2} \rceil}^{k-1} d^{l-1}\right) - (d-1)d^{\lceil \frac{k}{2} \rceil - 1}\mathbf{1}_{\{k \bmod 2 = 0\}} \\ &= 2d^{k-1} - 2d^{\lceil \frac{k}{2} \rceil - 1} - (d-1)d^{\lceil \frac{k}{2} \rceil - 1}\mathbf{1}_{\{k \bmod 2 = 0\}}, \end{aligned} \tag{3.3}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, and $\mathbf{1}_{\{\bullet\}}$ is the indicator function. Therefore,

$$B_k \le 2d^{k-1}, \qquad 2 \le k \le D-1. \tag{3.4}$$

(c) those obtained with the concatenation of 3 directed paths corresponding to a sequence of LRL' (resp. RLR') shifts. Consider the path LRL'. We have $l > 0$, $r > 0$, $l' > 0$ and

$l + r + l' = k$, where $l = |L|$, $r = |R|$ and $l' = |L'|$. According to Lemma 3.1, $l + l' \leq r$ so that $k \geq 4$.

Consider the vertices obtained with a path LRL'. They are of the form

$$a_{r-l'} \cdots a_2 a_1 x_{l+1} \cdots x_{D-(r-l)} b_1 b_2 \cdots b_{l'},$$

where $a_{r-l'} \cdots a_2$ and $b_2 \cdots b_{l'}$ are arbitrary letters, and $a_1 \neq x_l$, $b_1 \neq x_{D-(r-l)+1}$ (otherwise, these vertices are already at a preceding level). Thus, there are at most $(d-1)^2 d^{r-2}$ vertices reached by such paths.

Hence, for fixed k, $4 \leq k \leq D-1$, the number of vertices reached by the concatenation of 3 directed paths is at most $2C_k$, with

$$C_k = \sum_{r=\lceil \frac{k}{2} \rceil}^{k-2} \sum_{l=1}^{k-r-1} (d-1)^2 d^{r-2}.$$

A simple calculation yields

$$\begin{aligned}
C_k &= \sum_{r=\lceil \frac{k}{2} \rceil}^{k-2} (k-r-1)(d-1)^2 d^{r-2} \\
&= (k-1)(d-1)^2 \sum_{r=\lceil \frac{k}{2} \rceil}^{k-2} d^{r-2} - \frac{(d-1)^2}{d} \sum_{r=\lceil \frac{k}{2} \rceil}^{k-2} r d^{r-1} \\
&= (k-1)(d-1)(d^{k-3} - d^{\lceil \frac{k}{2} \rceil - 2}) \\
&\quad - \frac{1}{d}\left[(k-2)d^{k-1} - (k-1)d^{k-2} - (\lceil \frac{k}{2} \rceil - 1)d^{\lceil \frac{k}{2} \rceil} + \lceil \frac{k}{2} \rceil d^{\lceil \frac{k}{2} \rceil - 1}\right] \\
&= d^{k-2} - \lfloor \frac{k}{2} \rfloor d^{\lceil \frac{k}{2} \rceil - 1} + (\lfloor \frac{k}{2} \rfloor - 1) d^{\lceil \frac{k}{2} \rceil - 2},
\end{aligned} \tag{3.5}$$

where $\lfloor x \rfloor$ denotes the integer part of x. Hence,

$$C_k \leq d^{k-2}. \tag{3.6}$$

For $1 \leq k \leq D-1$, let

$$m_k = 2(d^k + 2d^{k-1} + d^{k-2}) = 2(d+1)^2 d^{k-2} \tag{3.7}$$

It follows from relations (3.2,3.4,3.6) that

$$l_k(X) \leq 2A_k + 2B_k \mathbf{1}_{\{k \geq 2\}} + 2C_k \mathbf{1}_{\{k \geq 4\}} < 2(d^k + 2d^{k-1} + d^{k-2}) = m_k, \qquad 1 \leq k \leq D-1, \tag{3.8}$$

where $l_k(X)$ is the width of level k in the shortest path trees rooted at vertex X.

Let n_d be the greatest integer n such that for all $D \geq 2$,

$$\sum_{k=1}^{n} m_k \leq N - 1. \tag{3.9}$$

Note that $N = d^D$ and

$$\sum_{k=1}^{n} m_k = \frac{2(d+1)^2}{d(d-1)}(d^n - 1).$$

Hence, it is readily checked that

$$n_d \geq \begin{cases} D-4, & d = 2; \\ D-2, & d = 3, 4; \\ D-1, & d \geq 5. \end{cases} \tag{3.10}$$

In fact, we have equality in (3.10) as soon as $D \geq 7$ for $d = 2$, $D \geq 4$ for $d = 4$, and any $D \geq 2$ for other values of d.

Applying now Lemma 2.1 implies that

$$\begin{aligned} \bar{e}(X) &\geq n_d + 1 - \frac{1}{N-1}\left(\sum_{k=1}^{n_d}(n_d + 1 - k)m_k\right) \\ &= n_d + 1 - \frac{2(d+1)^2}{d^D - 1}\left((n_d+1)\frac{d^{n_d} - 1}{d(d-1)} - \frac{(n_d+1)d^{n_d}}{d(d-1)} + \frac{d^{n_d+1} - 1}{d(d-1)^2}\right) \\ &= n_d + 1 - \frac{2(d+1)^2}{d^D - 1} \cdot \frac{d^{n_d+1} - (n_d+1)d + n_d}{d(d-1)^2} \end{aligned}$$

Therefore

$$\bar{e}(X) \geq n_d + 1 - \frac{2(d+1)^2}{d(d-1)^2} \cdot \frac{d^{n_d+1}}{d^D}. \tag{3.11}$$

It is now a simple calculation from (3.10) and (3.11) to obtain relation (3.1). ∎

Observe that for small values of d and D, closer lower bounds can be obtained using the exact values of A_k, B_k and C_k (see (3.2), (3.3), (3.5)) in

$$m_k = 2A_k + 2B_k\mathbf{1}_{\{k\geq 2\}} + 2C_k\mathbf{1}_{\{k\geq 4\}}.$$

Note also that for $X = a\cdots a$, the lower bounds of Proposition 3.3 can be improved. Indeed, in that case, it follows from Proposition 3.2 that there exist only two kinds of vertices: type (a) with $A_k = d^k$ and type (b) with $B_k \leq d^{k-1}$ (as $l > r$ in the paths LR). Hence, we can do the computation with $m_k = 2(d+1)d^{k-1}$. This gives

$$\bar{e}(a \cdots a) \geq n'_d + 1 - \frac{2(d+1)}{(d-1)^2} \cdot \frac{d^{n'_d+1}}{d^D}. \tag{3.12}$$

where

$$n'_d \geq \begin{cases} D-3, & d = 2; \\ D-2, & d = 3; \\ D-1, & d \geq 4. \end{cases} \tag{3.13}$$

As a simple corollary of Proposition 3.3, we obtain the asymptotic mean eccentricity when d goes to infinity.

Corollary 3.1 *For any $D \geq 2$, and any $X \in UB(d, D)$,*

$$\lim_{d \to \infty} \bar{e}(X) = D.$$

The above property shows that all the mean eccentricities are close to the diameter when the degree of the de Bruijn network is large. Therefore, in this case, we can simply use D shifts to route instead of the sophisticated optimal routing of the shortest path [11] . The global communication delay would even be decreased due to the simplicity of the routing algorithm.

3.3 Upper Bound

It is clear that the mean eccentricity of any vertex X in the undirected de Bruijn network $UB(d, D)$ is smaller than the mean eccentricity of the same vertex in the directed de Bruijn network $B(d, D)$. Therefore, Proposition 2.2 still holds for undirected de Bruijn network $UB(d, D)$, i.e., for any $d \geq 2$ and $D \geq 2$, and for all $X \in UB(d, D)$,

$$\bar{e}(X) \leq D - \frac{1}{d-1} + \frac{D}{N-1}. \tag{3.14}$$

This bound is not tight. In the remainder of the paper, we will study upper bounds in more detail.

In order to get some intuition, we start with numerical computations for the case of small diameters. In Table 1, we provide the numerical results of the mean eccentricities in the binary de Bruijn network $UB(2, D)$. In particular, we present, for diameter $D = 2, 3, \cdots, 18$, the average

D	average $\bar{e}(X)$	maximum $\bar{e}(X)$ vertex $a \cdots a$	minimum $\bar{e}(X)$	vertices
2	1.1667	1.3333	1.0000	01 10
3	1.6429	2.0000	1.4286	001, 011 110, 100
4	2.1417	2.6667	1.8667	0011 1100
5	2.7540	3.4516	2.5484	00011, 00111 11100, 11000
6	3.4534	4.2698	3.2063	001011 110100
7	4.2148	5.1654	3.9685	0001011, 0010111 1110100, 1101000
8	5.0280	6.0706	4.8078	00010111 11101000
9	5.8844	7.0098	5.6888	001111010, 010111100 110000101, 101000011
10	6.7737	7.9589	6.5689	0010111100, 0011110100 1101000011, 1100001011
11	7.6886	8.9253	7.4934	00101111100, 00111110100 11010000011, 11000001011
12	8.6232	9.8960	8.4308	001101011100, 001110101100 110010100011, 110001010011
13	9.5733	10.8764	9.3770	0011010111100, 0011110101100 1100101000011, 1100001010011
14	10.5351	11.8594	10.3377	00101111110100 11010000001011
15	11.5063	12.8473	11.3107	001011111101100, 001101111110100 110100000010011, 110010000001011
16	12.4844	13.8372	12.2936	0010111111101100, 0011011111110100 1101000000010011, 1100100000001011
17	13.4678	14.8297	13.2790	00101101111110100, 00101111110110100 11010010000001011, 11010000001001011
18	14.4554	15.8233	14.2669	001011111010100011, 001110101000001011 110100000101011100, 110001010111110100

Table 1: Mean Eccentricities of $UB(2, D)$

(taken over all vertices) of the mean eccentricities, the maximal and the minimal weights of the shortest path trees as well as the vertices which exhibit these weights.

While the minimal weights are reached by different vertices, the maximal weights are obtained by the vertices $0\cdots 0$ and $1\cdots 1$ in all our experimentation. This leads us to the following conjecture:

Conjecture: *For any $d \geq 2$ and $D \geq 2$, the vertices $a\cdots a$, $a \in \mathcal{A}$, have the maximal mean eccentricity in $UB(d,D)$.*

This conjecture is numerically verified for the binary de Bruijn network $UB(2,D)$ with diameters up to 18. However, we were unable to prove it for the general case. Unlike the case of de Bruijn digraphs where we got closed formulae for the mean eccentricities of some extremal vertices, in the undirected de Bruijn graphs, we were not even able to obtain a closed formula for the mean eccentricity of the probably simplest vertices $a\cdots a$.

Numerical computation (see Table 2) indicates that the widths in the shortest path trees of $UB(2,D)$ rooted at $a\cdots a$ are quite regular. It is possible to prove that for any $1 \leq k \leq D/2$, the width at level k of the shortest path trees of $UB(2,D)$, denoted $l_k(2,D)$, can be expressed as

$$l_k(2,D) = \begin{cases} 2^k + 2^{k-2} + 2^{k-3} + \cdots + 2^{2p} + 2^{2p-2} & = & 2^k + 2^{k-1} - 2^{2p} + 2^{2p-2}, & k = 3p; \\ 2^k + 2^{k-2} + 2^{k-3} + \cdots + 2^{2p} & = & 2^k + 2^{k-1} - 2^{2p}, & k = 3p+1; \\ 2^k + 2^{k-2} + 2^{k-3} + \cdots + 2^{2p+1} & = & 2^k + 2^{k-1} - 2^{2p+1}, & k = 3p+2. \end{cases}$$

Unfortunately, it is nontrivial to characterize $l_k(2,D)$ for $k > D/2$. These numbers are not so regular as to yield simple recursive equations. It seems that for any fixed diameter, the widths are unimodal with the maximum at $D-1$. It also seems that $l_{D+1-k}(2,D+1) \leq 2l_{D-k}(2,D)$ for $k = 0,1,2$, and $l_{D+1-k}(2,D+1) \geq 2l_{D-k}(2,D)$ for $k \geq 3$.

In what follows, we will first consider the asymptotic value of $\bar{e}(a\cdots a)$ when D tends to infinity. More precisely, we will prove that the limit of

$$\Delta(X) \stackrel{\text{def}}{=} D - \bar{e}(X)$$

exists when $X = a\cdots a$, and we will provide a numerical recursive scheme for its computation. We will also establish an upper bound (better than (3.14)) for arbitrary X. As a consequence of these two results, we show that for any fixed degree $(2d)$, the asymptotic difference between $\bar{e}(X)$ and $\bar{e}(a\cdots a)$ when the diameter goes to infinity is smaller than 0.22.

Let $X \in UB(d,D)$ and $0 \leq h \leq D$. Define $E^h(X)$ to be the set of vertices that are at distance no greater than $D-h$ from X:

$$E_h(X) = \{Y |\ d(X,Y) \leq D-h\}.$$

k	$l_k(2, D)$				
	$UB(2, 18)$	$UB(2, 19)$	$UB(2, 20)$	$UB(2, 21)$	$UB(2, 22)$
1	2	2	2	2	2
2	4	4	4	4	4
3	9	9	9	9	9
4	20	20	20	20	20
5	40	40	40	40	40
6	84	84	84	84	84
7	176	176	176	176	176
8	352	352	352	352	352
9	720	720	720	720	720
10	1469	1471	1472	1472	1472
11	2926	2936	2941	2943	2944
12	5865	5911	5934	5944	5949
13	11648	11846	11945	11991	12014
14	22444	23268	23680	23878	23977
15	41559	45081	46764	47588	48000
16	68474	83382	90549	93976	95660
17	79558	136638	166578	180959	187888
18	26793	158954	273450	333665	362679
19		53393	317504	547235	668397
20			106351	633746	1093347
21				212347	1266714
22					423855

Table 2: Widths in the shortest path trees of $UB(2, D)$ rooted at $a \cdots a$

By definition, $E_0(X)$ contains all the vertices of $UB(d, D)$ and $E_D(X)$ is a singleton. Let $e_h(X) = |E_h(X)|/N$ be the proportion of vertices that are at distance no greater than $D - h$ from X. Then, $e_0(X) = 1$ and $e_D(X) = 1/N$.

Lemma 3.2 *For any $d \geq 2$, $D \geq 2$, and any $X \in UB(d, D)$,*

$$\bar{e}(X) = D + \frac{D-1}{N-1} - \frac{N}{N-1} \sum_{h=1}^{D-1} e_h(X). \tag{3.15}$$

Proof: For any $0 \leq h \leq D - 1$, the set of vertices that are exactly at distance $D - h$ from X is $E_h(X) - E_{h+1}(X)$. Since $E_{h+1}(X) \subseteq E_h(X)$, we obtain that the number of vertices that are at distance $D - h$ from X is $Ne_h(X) - Ne_{h+1}(X)$. Therefore,

$$\begin{aligned}
\bar{e}(X) &= \frac{1}{N-1} \sum_{h=0}^{D-1} (D-h) N (e_h(X) - e_{h+1}(X)) \\
&= \frac{N}{N-1} (De_0(X) - e_1(X) - e_2(X) - \cdots - e_D(X)) \\
&= \frac{N}{N-1} \left(D - \frac{1}{N} - \sum_{h=1}^{D-1} e_h(X) \right) \\
&= D + \frac{D-1}{N-1} - \frac{N}{N-1} \sum_{h=1}^{D-1} e_h(X).
\end{aligned}$$

■

We will first consider the vertices $a \cdots a$. For simplicity of notation, we arbitrarily fix a letter $a \in \mathcal{A}$, and denote $E_h \equiv E_h(a \cdots a)$, and $e_h \equiv e_h(a \cdots a)$.

Let

$$\begin{aligned}
P_{h,k} &= \{X \mid X = UAV,\ A = a \cdots a,\ |A| = k + h,\ |U| = k\}, \\
S_{h,k} &= \{X \mid X = UAV,\ A = a \cdots a,\ |A| = k + h,\ |V| = k\}.
\end{aligned}$$

Lemma 3.3 *For any $0 \leq h \leq D$,*

$$E_h = \bigcup_{0 \leq k \leq \lfloor \frac{D-h}{2} \rfloor} E_{h,k},$$

where $E_{h,k} = P_{h,k} \bigcup S_{h,k}$.

Proof: Let X be a vertex in $P_{h,k}$, where $0 \le k \le (D-h)/2$. Then $X = UAV$, with $A = a \cdots a$, $|A| = k+h$ and $|U| = k$. Thus, X can be reached from $a \cdots a$ by first $D-(k+h)$ left shifts applied to $a \cdots a$ introducing the suffix VV', where V' is an arbitrary word with length k, and then k right shifts introducing prefix U. The length of the corresponding path is $D-(k+h)+k = D-h$. Hence $X \in E_h$. Similarly, we can show that if $X \in R_{h,k}$, then $X \in E_h$. Therefore,

$$E_h \supseteq \bigcup_{0 \le k \le \lfloor \frac{D-h}{2} \rfloor} E_{h,k}. \tag{3.16}$$

Suppose now that X is a vertex in E_h. Then, by Proposition 3.2, there exists a path (not necessarily a shortest one) between $a \cdots a$ and X of the type LR or RL with $r+l \le D-h$, where $r = |R|$, $l = |L|$. Without loss of generality, suppose that this path is of the form LR. Thus X can be written as $X = UAV$ with $U = a_r \cdots a_1$, $A = a \cdots a$, $V = b_1 \cdots b_{l-r}$, where $l \ge r$ (cf. Proposition 3.2, where the equality holds when $l = r = 0$). It then follows that

$$|A| = D - r - (l-r) \ge D - (D-h) + r = r + h.$$

Let A' be the prefix of A with size $r+h$. Then X can be written as $X = UA'V'$ so that $X \in P_{h,r}$. Since $|V'| = D - r - (r+h) \ge |V| \ge 0$, we get that $r \le (D-h)/2$. Therefore,

$$E_h \subseteq \bigcup_{0 \le k \le \lfloor \frac{D-h}{2} \rfloor} E_{h,k}. \tag{3.17}$$

Combining relations (3.16) and (3.17) readily implies the assertion of the lemma. ∎

For $h \ge 1$ and $k \ge 0$, let

$$e_{h,k} = \frac{|E_{h,k} - \bigcup_{j=0}^{k-1} E_{h,j}|}{N}. \tag{3.18}$$

It then follows that

$$e_h = \sum_{k=0}^{\lfloor \frac{D-h}{2} \rfloor} e_{h,k}. \tag{3.19}$$

Lemma 3.4 *For any fixed $d \ge 2$ and any $p, q \le \lfloor D/3 \rfloor$,*

$$\sum_{h=1}^{p} \sum_{k=0}^{q} e_{h,k} \le \sum_{h=1}^{D-1} e_h \le \left(\sum_{h=1}^{p} \sum_{k=0}^{q} e_{h,k} \right) + 2 \left(\frac{d^{-q} + d^{1-p}}{(d-1)^2} \right) \tag{3.20}$$

Proof: Note first that

$$e_{h,k} \leq 2\frac{|P_{h,k}|}{N} = 2d^{-h-k}, \qquad 0 \leq k, h \leq D, \quad 2k + h \leq D, \tag{3.21}$$

which implies that

$$e_h = \sum_{k=0}^{\lfloor\frac{D-h}{2}\rfloor} e_{h,k} \leq \sum_{k=0}^{\lfloor\frac{D-h}{2}\rfloor} 2d^{-h-k} \leq \sum_{k=0}^{\infty} 2d^{-h-k} = \frac{2d^{1-h}}{d-1}. \tag{3.22}$$

Thus, for any $p, q \leq \lfloor D/3 \rfloor$,

$$\begin{aligned}
\sum_{h=1}^{D-1} e_h &= \sum_{h=1}^{p}\sum_{k=0}^{q} e_{h,k} + \sum_{h=1}^{p}\sum_{k=q+1}^{\lfloor\frac{D-h}{2}\rfloor} e_{h,k} + \sum_{h=p+1}^{D-1} e_h \\
&\leq \sum_{h=1}^{p}\sum_{k=0}^{q} e_{h,k} + \sum_{h=1}^{p}\sum_{k=q+1}^{\lfloor\frac{D-h}{2}\rfloor} 2d^{-h-k} + \sum_{h=p+1}^{D-1} \frac{2d^{1-h}}{d-1} \\
&\leq \sum_{h=1}^{p}\sum_{k=0}^{q} e_{h,k} + \sum_{h=1}^{\infty}\sum_{k=q+1}^{\infty} 2d^{-h-k} + \sum_{h=p+1}^{\infty} \frac{2d^{1-h}}{d-1} \\
&= \left(\sum_{h=1}^{p}\sum_{k=0}^{q} e_{h,k}\right) + 2\left[\frac{d^{-q} + d^{1-p}}{(d-1)^2}\right].
\end{aligned}$$

On the other hand, it is easy to see that

$$\sum_{h=1}^{D-1} e_h \geq \sum_{h=1}^{n}\sum_{k=0}^{\lfloor\frac{D-h}{2}\rfloor} e_{h,k} \geq \sum_{h=1}^{p}\sum_{k=0}^{q} e_{h,k},$$

hence the result. ∎

Proposition 3.4 *For any fixed $d \geq 2$,*

$$\lim_{D\to\infty} \Delta(a\cdots a) = \sum_{h=1}^{\infty}\sum_{k=0}^{\infty} e_{h,k} \leq \frac{2d}{(d-1)^2}. \tag{3.23}$$

Proof: Let $p = q = \lfloor D/3 \rfloor$ in (3.20), and let D go to ∞ readily implies that the limit of $\sum_{h=1}^{D-1} e_h$ exists when $D \to \infty$ and that

$$\lim_{D\to\infty} \sum_{h=1}^{D-1} e_h = \sum_{h=1}^{\infty}\sum_{k=0}^{\infty} e_{h,k}.$$

Applying Lemma 3.2 entails that

$$\lim_{D\to\infty} \Delta(a\cdots a) = \sum_{h=1}^{\infty}\sum_{k=0}^{\infty} e_{h,k}.$$

Using again inequality (3.21) yields

$$\sum_{h=1}^{\infty}\sum_{k=0}^{\infty} e_{h,k} \le 2\sum_{h=1}^{\infty}\sum_{k=0}^{\infty} d^{-h-k} = \frac{2d}{(d-1)^2}.$$

The proof is thus completed. ∎

Observe that the relations (3.26), (3.27) and (3.31) provide a recursive scheme for the computation of $\lim_{D\to\infty}\Delta(a\cdots a)$. The convergence of the series $\sum_{h=1}^{\infty}\sum_{k=0}^{\infty} e_{h,k}$ is very fast. Good precisions are obtained with computations stopped at quite small indices h, k. Indeed, Lemma 3.4 and Proposition 3.4 imply that for any $n \ge 1$,

$$\sum_{h=1}^{n}\sum_{k=0}^{n} e_{h,k} \le \lim_{D\to\infty}\Delta(a\cdots a) < \left(\sum_{h=1}^{n}\sum_{k=0}^{n} e_{h,k}\right) + \frac{2(d+1)}{(d-1)^2}d^{-n}. \tag{3.24}$$

In view of the above results, we see that the computation of limit of $\Delta(a\cdots a)$ can be carried out on the values of $e_{h,k}$ for small h and k. We derive below a recursive computational scheme for $e_{h,k}$. This will be based on counting the proportion of words (of length D over the alphabet $\mathcal{A}$) fulfilling some conditions. This approach and some of the ideas can be found in [1].

Let C be a condition on the words of length D over $\mathcal{A}$. Let $\overline{C}$ be the opposite condition of C. Two conditions C_1 and C_2 are said to be *independent* if they concern disjoint subgroups of letters of the word. Let $p(C)$ denote the proportion of words satisfying a condition C.

We recall the following *basic computation rules* which will be used in the remainder of the paper.

- *The proportion of words satisfying $\overline{C}$ is $1 - p(C)$, i.e., $p(\overline{C}) = 1 - p(C)$.*
- *If C_1 and C_2 are independent, then the proportion of words satisfying both C_1 and C_2 is $p(C_1)p(C_2)$, i.e., $p(C_1, C_2) = p(C_1)p(C_2)$.*

For $h \ge 1$ and $k \ge 0$, let

$$\eta_{h,k} \quad = \quad \frac{|P_{h,k+1} - P_{h,k}|}{N} \equiv \frac{|S_{h,k+1} - S_{h,k}|}{N},$$

$$\gamma_{h,k} = \frac{\left|\bigcup_{j=0}^{k} P_{h,j}\right|}{N} \equiv \frac{\left|\bigcup_{j=0}^{k} S_{h,j}\right|}{N},$$

$$\delta_{h,k} = \frac{\left|P_{h,k+1} - \bigcup_{j=0}^{k} P_{h,j}\right|}{N} \equiv \frac{\left|S_{h,k+1} - \bigcup_{j=0}^{k} S_{h,j}\right|}{N}.$$

For $k = -1$, $\eta_{h,-1} = \delta_{h,-1} = \frac{|P_{h,0}|}{N} = d^{-h}$. For $k \leq -1$, $\gamma_{h,k} = 0$. It is also clear that $\gamma_{h,0} = d^{-h}$.

Lemma 3.5 *For all $h \geq 1$ and $k \geq 0$ such that $2k + 2 + h \leq D$,*

$$\eta_{h,k} = (d-1)d^{-(k+h+2)}. \tag{3.25}$$

Proof: The words in $P_{h,k+1}$ are of the form $x_1 \cdots x_{k+1} A_{k+1+h} V$, where $x_1, \cdots, x_{k+1}$ are arbitrary letters, A_i is a word of length i with letter a, V is an arbitrary word. The words in $P_{h,k}$ are of the form $x_1 \cdots x_k A_{k+h} V'$. Therefore, the words in $P_{h,k+1} - P_{h,k}$ are of the form $x_1 \cdots x_k \bar{a} A_{k+1+h} V$, where $\bar{a}$ is an arbitrary letter in $\mathcal{A} - \{a\}$. Thus, $\eta_{h,k} = d^{-(k+h+1)}(d-1)/d$. ■

Lemma 3.6 *For all $h \geq 1$ and $k \geq 0$ such that $2k + 2 + h \leq D$,*

$$\gamma_{h,k+1} = \gamma_{h,k} + \delta_{h,k}. \tag{3.26}$$

Proof: The assertion comes from the fact that

$$\left|\bigcup_{j=0}^{k+1} P_{h,j}\right| = \left|\bigcup_{j=0}^{k} P_{h,j}\right| + \left|P_{h,k+1} - \bigcup_{j=0}^{k} P_{h,j}\right|.$$

■

Lemma 3.7 *For all $h \geq 1$ and $k \geq -1$ such that $2k + 2 + h \leq D$,*

$$\delta_{h,k} = \left(1 - \gamma_{h,\lfloor\frac{k-h}{2}\rfloor}\right) \eta_{h,k}, \qquad k \geq -1; \tag{3.27}$$

$$= \left(1 - \gamma_{h,\lfloor\frac{k-h}{2}\rfloor}\right) (d-1)d^{-(k+h+2)}, \qquad k \geq 0. \tag{3.28}$$

Proof: As seen in the proof of Lemma 3.5, the words in $P_{h,k+1} - P_{h,k}$ are of the form $x_1 \cdots x_k \bar{a} A_{k+1+h} V$, where $\bar{a}$ is an arbitrary letter in $\mathcal{A} - \{a\}$. It then follows that these words will not appear in $P_{h,j}$ for all j such that $j < k$ and $2j + h > k$. Therefore,

$$P_{h,k+1} - P_{h,k} = P_{h,k+1} - \bigcup_{j=\lfloor\frac{k-h}{2}\rfloor+1}^{k} P_{h,j}, \tag{3.29}$$

which, in turn, implies that

$$P_{h,k+1} - \bigcup_{j=0}^{k} P_{h,j} = P_{h,k+1} - P_{h,k} - \bigcup_{j=0}^{\lfloor \frac{k-h}{2} \rfloor} P_{h,j}. \tag{3.30}$$

It is easily verified that the conditions for words to be in $P_{h,k+1} - P_{h,k}$ and in $\bigcup_{j=0}^{\lfloor \frac{k-h}{2} \rfloor} P_{h,j}$ are independent (the former is on $x_{k+1} \cdots x_{2k+2+h}$ of word $X = x_1 \cdots x_D$, whereas the latter is on $x_1 \cdots x_k$). Therefore, relation (3.30) implies relation (3.27) according to the basic computation rule. Relation (3.28) follows from those of (3.25) and (3.27). ■

Proposition 3.5 *For all $h \geq 1$ and $k \geq 0$ such that $2k + h \leq D/2$,*

$$e_{h,k} = (2 - \gamma_{h,k-1} - \gamma_{h,k})\delta_{h,k-1}, \qquad k \geq 0; \tag{3.31}$$

$$e_{h,k} = (2 - \gamma_{h,k-1} - \gamma_{h,k}) \left(1 - \gamma_{h,\lfloor \frac{k-h-1}{2} \rfloor}\right) (d-1) d^{-(k+h+1)}, \qquad k \geq 1. \tag{3.32}$$

Proof:

$$\begin{aligned} e_{h,k} &= \frac{\left|P_{h,k} \bigcup S_{h,k} - \bigcup_{j=0}^{k-1} E_{h,j}\right|}{N} \\ &= \frac{\left|P_{h,k} - \bigcup_{j=0}^{k-1} E_{h,j}\right|}{N} + \frac{\left|S_{h,k} - P_{h,k} - \bigcup_{j=0}^{k-1} E_{h,j}\right|}{N} \\ &= \frac{\left|P_{h,k} - \bigcup_{j=0}^{k-1} P_{h,j} - \bigcup_{j=0}^{k-1} S_{h,j}\right|}{N} + \frac{\left|S_{h,k} - \bigcup_{j=0}^{k-1} S_{h,j} - \bigcup_{j=0}^{k} P_{h,j}\right|}{N} \end{aligned}$$

Under the assumption that $2k + h \leq D/2$, the conditions for words to be in $P_{h,k} - \bigcup_{j=0}^{k-1} P_{h,j}$ and in $\bigcup_{j=0}^{k-1} S_{h,j}$ are independent (the former is on the left-half part of the word, whereas the latter is on the right-half part of the word). Therefore,

$$\frac{\left|P_{h,k} - \bigcup_{j=0}^{k-1} P_{h,j} - \bigcup_{j=0}^{k-1} S_{h,j}\right|}{N} = \delta_{h,k-1}(1 - \gamma_{h,k-1}).$$

Similarly,

$$\frac{\left|S_{h,k} - \bigcup_{j=0}^{k-1} S_{h,j} - \bigcup_{j=0}^{k} P_{h,j}\right|}{N} = \delta_{h,k-1}(1 - \gamma_{h,k}).$$

Hence, relation (3.31) holds. Relation (3.32) follows from (3.27) and (3.31). ■

$e_{h,k}$	k=0	k=1	k=2	k=3	k=4	k=5	$\sum_{k=1}^{5} e_{h,k}$
h=1	0.75000	0.10938	0.02246	0.01050	0.00381	0.00188	0.89802
h=2	0.43750	0.08984	0.04110	0.01524	0.00752	0.00342	0.59552
h=3	0.23438	0.05371	0.02612	0.01288	0.00560	0.00279	0.33547
h=4	0.12109	0.029053	0.01434	0.00713	0.00355	0.00166	0.17683
h=5	0.06152	0.01508	0.00750	0.00373	0.00186	0.00093	0.09062
h=6	0.03101	0.00768	0.00383	0.00191	0.00095	0.00048	0.04585
$\sum_{h=1}^{6} e_{h,k}$	1.63550	0.30473	0.11624	0.05140	0.02330	0.01116	2.14232
$\hat{e_k}$	1.66667	0.31250	0.11914	0.05235	0.02388	0.01146	2.18600

Table 3: $\Delta(a\cdots a)$ in $UB(2,D)$

Lemmas 3.5, 3.6, 3.7 and proposition 3.5 enable us to compute recursively $e_{h,k}$. In Table 3, we indicate the values of $e_{h,k}$ in the case $d=2$ for $1 \leq h \leq 6$ and $0 \leq k \leq 5$. This gives already a lower bound of 2.14 for $\Delta(a\cdots a)$. In fact we have computed all the $e_{h,k}$ for $d=2$, $h \leq 30$ and $k \leq 85$. Results for e_h are given in Table 4. The values are given with 10 digits but we have computed them with infinity precision. Recall that having the values of e_h we obtain the number of vertices at distance exactly $D-h$. For instance we have $1-e_1 \simeq 10.03\%$ of vertices at distance D from $(a\cdots a)$ in the case $d=2$. This table shows that most of the vertices are at distance $D-h$ with h small.

Finally, let us note that one can get closed formulae for $\hat{e}_k$ where :

$$\hat{e}_k = \sum_{h=1}^{\infty} e_{h,k}. \tag{3.33}$$

We give below this formulae for $0 \leq k \leq 6$.

$$\begin{aligned}
\hat{e}_0 &= \frac{2d+1}{d^2-1} \\
\hat{e}_1 &= \frac{2d^2-2d+1}{d^4} \\
\hat{e}_2 &= \frac{2d^6-4d^5+4d^4+d^3-3d^2+1}{d^9} \\
\hat{e}_3 &= d^{-16}(2d^{12}-4d^{11}+4d^{10}-2d^9+d^8+2d^7-3d^6+2d^5-d^4-2d^3+d^2+2d-1) \\
\hat{e}_4 &= d^{-22}(2d^{17}-4d^{16}+2d^{15}+4d^{14}-4d^{13}+d^{12}-2d^{11}-d^{10}+7d^9-6d^7-d^6+5d^5 \\
&\quad -4d^4+2d^3-d^2+2d-1)
\end{aligned}$$

h	e_h	$\sum_{i=1}^{h} e_i$
1	0.8997111342	0.899711134
2	0.5988445366	1.498555671
3	0.3381424180	1.836698089
4	0.1784738644	2.015171953
5	0.0915215694	2.106693523
6	0.0463216922	2.153015215
7	0.0232996659	2.176314881
8	0.0116843542	2.187999235
9	0.0058507839	2.193850019
10	0.0029275407	2.196777560
11	0.0014643071	2.198241867
12	0.0007322877	2.198974155
13	0.0003661774	2.199340332
$\vdots$		$\vdots$
30	0.0000000027	2.199706529
$\vdots$		$\vdots$
60	0.260×10^{-17}	2.199706532

Table 4: $\Delta(a \cdots a)$ in $UB(2, D)$

$$\begin{aligned}\hat{e}_5 &= d^{-28}(2d^{22} - 4d^{21} + 2d^{20} + 4d^{19} - 6d^{18} + 4d^{17} + d^{16} - 6d^{15} + d^{14} + 8d^{13} - 2d^{12} \\ &\quad -10d^{11} + 11d^{10} - 2d^9 - 3d^8 - d^6 + 2d^5 - d^2 + 2d - 1) \\ \hat{e}_6 &= d^{-34}(-1 + 2d + 8d^{10} - d^2 + 2d^{24} + 3d^{20} - 2d^4 + 6d^5 - 3d^6 + 5d^{18} + d^8 - 2d^7 - 3d^{12} \\ &\quad -2d^9 - 2d^{16} - 8d^{11} - 2d^{15} - 2d^{14} + 8d^{13} - 4d^{19} + 2d^{17} - 4d^{21} + 2d^{25} - 4d^{26} + 2d^{27})\end{aligned}$$

In Table 3, we also provided exact values of $\hat{e}_k$ for $d = 2$. Again, we note that using only $\sum_{k=0}^{6} \hat{e}_k$ gives a very good estimation of $\Delta(a \cdots a)$.

Indeed it follows from Lemma 3.4 and Proposition 3.4 that for any $q \geq 0$,

$$\sum_{k=0}^{q} \hat{e}_k \leq \lim_{D \to \infty} \Delta(a \cdots a) < \left(\sum_{k=0}^{q} \hat{e}_k \right) + \frac{2d^{-q}}{(d-1)^2}. \tag{3.34}$$

The following proposition gives values of $\lim_{D \to \infty} \Delta(a \cdots a)$ for $d \leq 7$. There again we give only truncated values with 5 decimals. The values given in parenthesis are approximation using $\sum_{k=0}^{6} \hat{e}_k$.

Proposition 3.6 *In the de Bruijn network $UB(d, D)$,*

$$\lim_{D \to \infty} \Delta(a \cdots a) = \begin{cases} 2.19970 & d = 2 \quad (2.19151) \\ 1.09605 & d = 3 \quad (1.09556) \\ 0.72359 & d = 4 \quad (0.72353) \\ 0.53752 & d = 5 \quad (0.53751) \\ 0.42652 & d = 6 \quad (0.42652) \\ 0.35304 & d = 7 \quad (0.35303) \end{cases}$$

We now turn back to upper bounds for arbitrary vertex X. Let us consider an arbitrarily fixed vertex $X = x_1 \cdots x_D$. Let

$$\begin{aligned} P_{h,k}(X) &= \{Z \mid Z = UX_{h,k}V,\ X_{h,k} = x_{D-h-k+1}x_{D-h-k+2} \cdots x_D,\ |U| = k\}, \\ S_{h,k}(X) &= \{Z \mid Z = UX^s_{h,k}V,\ X^s_{h,k} = x_1 x_2 \cdots x_{h+k},\ |V| = k\}, \\ E_{h,k}(X) &= P_{h,k}(X) \bigcup S_{h,k}(X), \end{aligned}$$

Lemma 3.8 *For any $0 \leq h \leq D$,*

$$E_h(X) \supseteq \bigcup_{0 \leq k \leq \lfloor \frac{D-h}{2} \rfloor} E_{h,k}(X).$$

Proof: The proof is analogous to the first part of the proof of Lemma 3.3, and is thus omitted. ■

Proposition 3.7 *For any fixed $d \geq 2$, and $D \geq 2$,*

$$\sum_{h=1}^{D-1} e_h(X) > \theta(d) - 22d^{-\frac{D}{4}}, \tag{3.35}$$

where

$$\theta(d) = \frac{2d+1}{d^2-1} + \frac{2d-1}{d^3-(d-1)^2} - \frac{1}{d^3+1} = \frac{2d^6-d^5+2d^4-d^3+d^2+d-1}{d^7-2d^6+3d^5-2d^4-d^3+3d^2-3d+1}. \tag{3.36}$$

Proof: For $1 \leq h \leq D/2$ and $0 \leq k \leq (D-2h)/4$, let

$$\begin{aligned} P'_{h,k}(X) &= \{Z = a_1 \cdots a_k x_{D-h-k+1} x_{D-h-k+2} \cdots x_D V b'_{h+2k-2} \cdots b'_k b_{k-1} \cdots b_1 \mid \\ &\quad \forall 1 \leq i \leq k: \; a_i \neq x_{D-h-i+2}, \;\; \forall 1 \leq j \leq k-1: \; b_j \neq x_{h+j-1}, \;\; b'_{h+2k-2} \cdots b'_k \neq x_1 \cdots x_{h+k-1}\} \\ S'_{h,k}(X) &= \{Z = a_1 \cdots a_k a'_{k+1} \cdots a'_{h+2k} V x_1 x_2 \cdots x_{h+k} b_k b_{k-1} \cdots b_1 \mid \\ &\quad \forall 1 \leq i \leq k: \; a_i \neq x_{D-h-i+2}, \; b_i \neq x_{h+i-1}, \;\; a'_{k+1} \cdots a'_{h+2k} \neq x_{D-h-k+1} \cdots x_D\} \end{aligned}$$

It is clear that $P'_{h,k}(X) \subseteq P_{h,k}(X)$, $S'_{h,k}(X) \subseteq S_{h,k}(X)$. Therefore, it follows from Lemma 3.8 that

$$E_h(X) \supseteq \bigcup_{0 \leq k \leq \lfloor \frac{D-h}{2} \rfloor} E_{h,k}(X) \supseteq \bigcup_{0 \leq k \leq \lfloor \frac{D-2h}{4} \rfloor} \left(P'_{h,k}(X) \cup S'_{h,k}(X) \right).$$

Let

$$e'_h(X) = \frac{\left| \bigcup_{0 \leq k \leq \lfloor \frac{D-2h}{4} \rfloor} \left(P'_{h,k}(X) \cup S'_{h,k}(X) \right) \right|}{N}.$$

It then follows that

$$\sum_{h=1}^{D-1} e_h(X) \geq \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} e'_h(X). \tag{3.37}$$

For any fixed $1 \leq h \leq D/2$, the sets $P'_{h,k}(X)$ and $S'_{h,k}(X)$, $0 \leq k \leq (D-2h)/4$, are disjoint:

$$P'_{h,k}(X) \bigcap P'_{h,j}(X) = \emptyset, \qquad j < k, \tag{3.38}$$

$$S'_{h,k}(X) \bigcap S'_{h,j}(X) = \emptyset, \qquad j < k, \tag{3.39}$$

$$P'_{h,k}(X) \bigcap S'_{h,j}(X) = \emptyset, \qquad j \leq k. \tag{3.40}$$

$$S'_{h,k}(X) \bigcap P'_{h,j}(X) = \emptyset, \qquad j \leq k. \tag{3.41}$$

Indeed, (3.38) comes from the constraint on $P'_{h,k}(X)$ that for $a_{j+1} \neq x_{D-h-j+1}$, so that for all $Z \in P'_{h,k}(X)$, $Z \notin P'_{h,j}(X)$. Relation (3.39) comes from the constraint on $S'_{h,k}(X)$ that for $b_{j+1} \neq x_{h+j}$, so that for all $Z \in S'_{h,k}(X)$, $Z \notin S'_{h,j}(X)$. Last, concerning relation (3.40), if $j < k-1$, then for all $Z \in P'_{h,k}(X)$, $b_{j+1} \neq x_{h+j}$, so that $Z \notin P'_{h,j}(X)$. If $j = k-1$, then for all $Z \in P'_{h,k}(X)$, $b'_{h+2k-2} \cdots b'_k \neq z_1 \cdots z_{h+k-1}$, so that $Z \notin P'_{h,k-1}(X)$. Finally, if $j = k$, then for all $Z \in S'_{h,k}(X)$, $a'_{k+1} \cdots a'_{h+2k} \neq z_{D-h-k+1} \cdots z_D$, so that $Z \notin P'_{h,k}(X)$. Relation (3.41) can be proved exactly in the same manner.

Let $\alpha_{h,k} = |P'_{h,k}(X)|/N$, $\beta_{h,k} = |S'_{h,k}(X)|/N$. It is now clear that

$$e'_h(X) = \sum_{k=0}^{\lfloor \frac{D-2h}{4} \rfloor} (\alpha_{h,k} + \beta_{h,k}) \tag{3.42}$$

It is also simple to see from the definitions of $P'_{h,k}(X)$ and $S'_{h,k}(X)$ that

$$\begin{aligned}
\alpha_{h,0} &= d^{-h} && (3.43)\\
\alpha_{h,k} &= \left(\frac{d-1}{d}\right)^k d^{-(h+k)} \left(\frac{d-1}{d}\right)^{k-1} \left(1 - d^{-(h+k-1)}\right) \\
&= d^{-(h+3k-1)}(d-1)^{2k-1}\left(1 - d^{-(h+k-1)}\right), && (3.44)\\
\beta_{h,0} &= d^{-h}\left(1 - d^{-h}\right) && (3.45)\\
\beta_{h,k} &= \left(\frac{d-1}{d}\right)^k \left(1 - d^{-(h+k)}\right) d^{-(h+k)} \left(\frac{d-1}{d}\right)^k \\
&= d^{-(h+3k)}(d-1)^{2k}\left(1 - d^{-(h+k)}\right). && (3.46)
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} e'_h(X) &= \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=0}^{\lfloor \frac{D-2h}{4} \rfloor} (\alpha_{h,k} + \beta_{h,k}) \\
&= \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \left\{ d^{-h} + d^{-h}\left(1 - d^{-h}\right) \right\} \\
&\quad + \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=1}^{\lfloor \frac{D-2h}{4} \rfloor} \left\{ d^{-(h+3k-1)}(d-1)^{2k-1}\left(1 - d^{-(h+k-1)}\right) + d^{-(h+3k)}(d-1)^{2k}\left(1 - d^{-(h+k)}\right) \right\} \\
&= \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \left\{ 2d^{-h} - d^{-2h} \right\}
\end{aligned}$$

$$
\begin{aligned}
& + \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=1}^{\lfloor \frac{D-2h}{4} \rfloor} d^{-(h+3k)}(d-1)^{2k} \left(\frac{d}{d-1} + 1 \right) \\
& - \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=1}^{\lfloor \frac{D-2h}{4} \rfloor} d^{-(2h+4k)}(d-1)^{2k} \left(\frac{d^2}{d-1} + 1 \right) \\
\stackrel{\text{def}}{=} \; & F_1 + F_2 + F_3
\end{aligned}
\tag{3.47}
$$

where

$$
\begin{aligned}
F_1 & \stackrel{\text{def}}{=} \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \left\{ 2d^{-h} - d^{-2h} \right\} \\
F_2 & \stackrel{\text{def}}{=} \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=1}^{\lfloor \frac{D-2h}{4} \rfloor} d^{-(h+3k)}(d-1)^{2k} \left(\frac{d}{d-1} + 1 \right) \\
F_3 & \stackrel{\text{def}}{=} - \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=1}^{\lfloor \frac{D-2h}{4} \rfloor} d^{-(2h+4k)}(d-1)^{2k} \left(\frac{d^2}{d-1} + 1 \right)
\end{aligned}
$$

Some simple algebra yields

$$
\begin{aligned}
F_1 & = \frac{2 - 2d^{-\lfloor \frac{D}{2} \rfloor}}{d-1} - \frac{1 - d^{-2\lfloor \frac{D}{2} \rfloor}}{d^2-1} \\
& = \frac{2}{d-1} - \frac{1}{d^2-1} - \frac{2d^{-\lfloor \frac{D}{2} \rfloor}}{d-1} + \frac{d^{-2\lfloor \frac{D}{2} \rfloor}}{d^2-1}
\end{aligned}
$$

Hence,

$$
F_1 > \frac{2d+1}{d^2-1} - \frac{2d^{-\lfloor \frac{D}{2} \rfloor}}{d-1} \tag{3.48}
$$

We can also obtain a lower bound for F_2:

$$
\begin{aligned}
F_2 & = \frac{2d-1}{d-1} \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=1}^{\lfloor \frac{D-2h}{4} \rfloor} d^{-h} \left(\frac{(d-1)^2}{d^3} \right)^k \\
& = \frac{2d-1}{d-1} \cdot \frac{(d-1)^2}{d^3-(d-1)^2} \cdot \left(\sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} d^{-h} - \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} d^{-h} \left(\frac{(d-1)^2}{d^3} \right)^{\lfloor \frac{D-2h}{4} \rfloor} \right) \\
& > \frac{(2d-1)(d-1)}{d^3-(d-1)^2} \cdot \left(\sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} d^{-h} - \sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} d^{-h} \left(\frac{(d-1)^2}{d^3} \right)^{\frac{D-2h}{4} - 1} \right)
\end{aligned}
$$

$$= \frac{(2d-1)(d-1)}{d^3-(d-1)^2} \cdot \left(\frac{1-d^{-\lfloor \frac{D}{2} \rfloor}}{d-1} - \left(\frac{(d-1)^2}{d^3} \right)^{\frac{D}{4}-1} \cdot \frac{1-\left(\frac{\sqrt{d}}{d-1}\right)^{\lfloor \frac{D}{2} \rfloor}}{\frac{d-1}{\sqrt{d}}-1} \right)$$

$$= \frac{2d-1}{d^3-(d-1)^2} - \frac{(2d-1)d^{-\lfloor \frac{D}{2} \rfloor}}{d^3-(d-1)^2} - \frac{(2d-1)(d-1)}{d^3-(d-1)^2} \cdot \left(\frac{(d-1)^2}{d^3} \right)^{\frac{D}{4}-1} \cdot \frac{1}{\frac{d-1}{\sqrt{d}}-1}$$

$$+ \frac{(2d-1)(d-1)}{d^3-(d-1)^2} \cdot \left(\frac{(d-1)^2}{d^3} \right)^{\frac{D}{4}-1} \cdot \frac{\left(\frac{\sqrt{d}}{d-1}\right)^{\lfloor \frac{D}{2} \rfloor}}{\frac{d-1}{\sqrt{d}}-1}$$

It is easy to see that if $d = 2$, then,

$$\begin{aligned} F_2 &> \frac{2d-1}{d^3-(d-1)^2} - \frac{(2d-1)d^{-\lfloor \frac{D}{2} \rfloor}}{d^3-(d-1)^2} - \frac{(2d-1)(d-1)}{d^3-(d-1)^2} \cdot \left(\frac{(d-1)^2}{d^3} \right)^{\frac{D}{4}-1} \cdot \frac{\left(\frac{\sqrt{d}}{d-1}\right)^{\lfloor \frac{D}{2} \rfloor}}{1-\frac{d-1}{\sqrt{d}}} \\ &= \frac{2d-1}{d^3-(d-1)^2} - \frac{3}{7} d^{-\lfloor \frac{D}{2} \rfloor} - \frac{3}{7} d^{-\frac{3D}{4}+3+\lfloor \frac{D}{2} \rfloor/2} \cdot \frac{\sqrt{2}}{\sqrt{2}-1} \\ &> \frac{2d-1}{d^3-(d-1)^2} - 12 d^{-\frac{D}{4}}. \end{aligned}$$

If, however, $d \geq 3$,

$$\begin{aligned} F_2 &> \frac{2d-1}{d^3-(d-1)^2} - \frac{(2d-1)d^{-\lfloor \frac{D}{2} \rfloor}}{d^3-(d-1)^2} - \frac{(2d-1)(d-1)}{d^3-(d-1)^2} \cdot \left(\frac{(d-1)^2}{d^3} \right)^{\frac{D}{4}-1} \cdot \frac{1}{\frac{d-1}{\sqrt{d}}-1} \\ &= \frac{2d-1}{d^3-(d-1)^2} - \frac{(2d-1)d^{-\lfloor \frac{D}{2} \rfloor}}{d^3-(d-1)^2} - \frac{(2d-1)(d-1)}{d^3-(d-1)^2} \cdot \frac{d^3}{(d-1)^2} \cdot \frac{1}{\frac{d-1}{\sqrt{d}}-1} \cdot d^{-\frac{D}{4}} \\ &> \frac{2d-1}{d^3-(d-1)^2} - \frac{2d}{d^3-d^2} d^{-\lfloor \frac{D}{2} \rfloor} - \frac{d^3}{d^3-(d-1)^2} \cdot \frac{2d-1}{d-1} \cdot \frac{1}{\frac{d-1}{\sqrt{d}}-1} \cdot d^{-\frac{D}{4}} \\ &> \frac{2d-1}{d^3-(d-1)^2} - \frac{1}{3} d^{-\lfloor \frac{D}{2} \rfloor} - \frac{27}{23} \cdot \frac{3}{2} \cdot \frac{1}{\frac{2}{\sqrt{3}}-1} \cdot d^{-\frac{D}{4}} \\ &> \frac{2d-1}{d^3-(d-1)^2} - 20 d^{-\frac{D}{4}} \end{aligned}$$

Therefore, for any $d \geq 2$,

$$F_2 > \frac{2d-1}{d^3-(d-1)^2} - 20 d^{-\frac{D}{4}} \tag{3.49}$$

In a similar way, we obtain a lower bound for F_3:

$$F_3 = -\sum_{h=1}^{\lfloor \frac{D}{2} \rfloor} \sum_{k=1}^{\lfloor \frac{D-2h}{4} \rfloor} d^{-(2h+4k)} (d-1)^{2k} \left(\frac{d^2}{d-1} + 1 \right)$$

$$
\begin{aligned}
&> \quad -\frac{d^2+d-1}{d-1}\sum_{h=1}^{\infty}\sum_{k=1}^{\infty} d^{-(2h+4k)}(d-1)^{2k} \\
&= \quad -\frac{d^2+d-1}{d-1}\cdot\frac{(d-1)^2}{d^4-(d-1)^2}\sum_{h=1}^{\infty} d^{-2h} \\
&= \quad -\frac{d^2+d-1}{d-1}\cdot\frac{(d-1)^2}{d^4-(d-1)^2}\cdot\frac{1}{d^2-1} \\
&= \quad -\frac{1}{d^3+1}
\end{aligned}
$$

Hence,

$$
F_3 > -\frac{1}{d^3+1} \tag{3.50}
$$

Combining relations (3.37), (3.47), (3.48), (3.49) and (3.50) readily entails (3.35). ∎

As a corollary of Lemma 3.2 and Proposition 3.7, we obtain

Corollary 3.2 *For any $d \geq 2$, $D \geq 2$, and any $X \in UB(d, D)$,*

$$
\bar{e}(X) < D - \theta(d) + 22d^{-\frac{D}{4}} + \frac{1}{N-1}\cdot\left(D - 1 - \theta(d) + 22d^{-\frac{D}{4}}\right), \tag{3.51}
$$

where $\theta(d)$ is defined in (3.36).

Proof: It follows from (3.15) and (3.35) that

$$
\begin{aligned}
\bar{e}(X) &= D - \sum_{h=1}^{D-1} e_h(X) + \frac{1}{N-1}\cdot\left(D - 1 - \sum_{h=1}^{D-1} e_h(X)\right) \\
&< D - \theta(d) + 22d^{-\frac{D}{4}} + \frac{1}{N-1}\cdot\left(D - 1 - \theta(d) + 22d^{-\frac{D}{4}}\right)
\end{aligned}
$$

∎

As a consequence of Proposition 3.4 and the above corollary, we obtain

Proposition 3.8 *For any fixed $d \geq 2$,*

$$
\limsup_{D\to\infty} \left(\bar{e}(X) - \bar{e}(a\cdots a)\right) < \left(\sum_{h=1}^{\infty}\sum_{k=0}^{\infty} e_{h,k}\right) - \theta(d), \tag{3.52}
$$

where $\theta(d)$ is defined in (3.36).

Proof: It follows from Proposition 3.4 that

$$\begin{aligned}
&\limsup_{D\to\infty}(\bar{e}(X) - \bar{e}(a\cdots a)) \\
&= \lim_{D\to\infty} \Delta(a\cdots a) - \liminf_{D\to\infty} \Delta(X) \\
&\geq \left(\sum_{h=1}^{\infty}\sum_{k=0}^{\infty} e_{h,k}\right) - \theta(d),
\end{aligned}$$

where the inequality comes from (3.51). ∎

Numerical results for small d are given below:

Proposition 3.9

$$\limsup_{D\to\infty} (\bar{e}(X) - \bar{e}(a\cdots a)) < \begin{cases} 0.2155 & d=2 \\ 0.0393 & d=3 \\ 0.0117 & d=4 \\ 0.0045 & d=5 \\ 0.0021 & d=6 \\ 0.0011 & d=7 \end{cases}$$

4 Conclusions

In this paper, we have provided upper and lower bounds of the mean eccentricities of the de Bruijn networks. For the directed de Bruijn network, we have presented tight bounds as well as the extremal vertices which reach these bounds. For the undirected de Bruijn network, we have provided lower and upper bounds which differ from the diameter by some small constants. We conjecture that the vertices of the form $a\cdots a$ have the largest mean eccentricity. This conjecture has been verified by numerical computations for binary de Bruijn networks with diameters up to 18. We have shown that the asymptotic difference, when the diameter goes to infinity, between the mean eccentricities of an arbitrary vertex and that of $a\cdots a$ is a small constant tending to zero with the degree. We have also provided a simple recursive scheme for the computation of the asymptotic mean eccentricity of the vertices $a\cdots a$.

We have proved that in both directed and undirected de Bruijn networks, all the mean eccentricities tend to the diameter when the degree goes to infinity. This implies that when the degree is large, the simple routing algorithm which consists of doing left shifts all the time or right shifts all the time is more efficient than optimal routing algorithms. In general, the gain of applying optimal routing algorithms is quite limited, even without considering their overheads.

References

[1] J-C. Bermond, J. Bond, W. Fernandez de la Vega, S. Rudich and M. Santha, "The radius of graphs on alphabets", preprint *Graph and Combinatorics*, 1988.

[2] J-C. Bermond, C. Delorme, and J-J. Quisquater, "Strategies for interconnection networks: Some methods from graph theory", *JPDC*, **3**, pp. 433–449, 1986.

[3] J-C. Bermond, N. Homobono, and C. Peyrat, "Large fault-tolerant interconnection networks", *Graph and Combinatorics*, **5**, pp. 107-123, 1989.

[4] J-C. Bermond and C. Peyrat, "De Bruijn and Kautz networks: a competitor for the hypercube?", In *Hypercube and Distributed Computers*, J. P. Verjus and F. André, (Eds.), pp. 279–294, North-Holland, 1989.

[5] I. Bond, *Constructions de grands réseaux d'interconnexion.* PhD thesis, Université de Paris-Sud, Centre d'Orsay, 1984.

[6] F. Buckley and F. Harary, *Distance in Graphs.* Addison Wesley, 1990.

[7] N. G. de Bruijn, "A combinatorial problem", In *Koninklijke Nederlands Akademie van Wetenschappen Proceedings*, **49-2**, pp. 758–764, 1946.

[8] M. A. Fiol, J. L. A. Yebra, I. Alegre de Miquel, "Line Digraph Iterations and the (d,k) Digraph Problem," *IEEE Trans. on Computers*, **C-33**, No. 5, pp. 400-403, 1984.

[9] S. W. Golomb, *Shift Register Sequences.* Holden-Day Inc, 1967.

[10] F. T. Leighton, *Introduction to Parallel Algorithms and Architectures : Arrays, Trees, Hypercubes.* Morgan Kaufmann, 1991.

[11] Z. Liu, "Optimal Routing in the De Bruijn Networks." *Proc. of the Tenth International Conference on Distributed Computing Systems*, May 1990, Paris, France, pp. 537-544.

[12] C. Peyrat, *Vulnérabilité dans les grand réseaux d'interconnexion.* PhD thesis, Université de Paris-Sud, Centre d'Orsay, 1984.

[13] M.R. Samatham and D.K. Pradhan, "The de Bruijn multiprocessor network: a versatile parallel processing and sorting network for VLSI", *IEEE Trans. on Comp.*, **38**, No. 4, pp. 567–581, 1989.