Graph Theoretical Analysis on Distributed Line Graphs for Peer-to-Peer Networks

Akiko Manada, *Member, IEEE*, and Hiroyoshi Morita, *Senior Member, IEEE*

Abstract—Distributed line graphs were introduced by Zhang and Liu as an overlay for Peer-to-Peer networks. Distributed line graphs have some useful properties as a network topology, such as out-regular and small diameter, where the former implies that each user possesses a constant size of routing table and the latter means that a reasonably small number of hops is necessary to reach a target user. In this paper, we newly introduce distributed line sharp graphs, as an extension of distributed line graphs, so as to further expand the possibility of finding better network topologies and theoretically analyze their properties. The analysis on the distributed line sharp graphs is based on the analysis on distributed line graphs presented by Zhang and Liu. In particular, we further examine the diameter and the in-degree of a vertex, and improve the previous results together with detailed proofs. We also show that the diameter of a certain distributed line sharp graph is equal to the diameter of its corresponding distributed line graph.

Index Terms—Distributed line graph, distributed line sharp graph, diameter, in/out degree, vertex merging

1 Introduction

The study of Peer-to-Peer (P2P) networks [2] has been gaining particular interest these days due to the broadbandization of internet and the improvement of computers. Distributed Hash Tables (DHTs) are structured overlay for P2P networks. Typical examples for DHTs are CAN [3], Chord [4], and Kademlia [5]. Indeed, those DHTs are mainly used in file sharing networks.

For many DHTs, when there are N users in a network, either 1) the routing table size (which corresponds to the out-degree) is $\mathcal{O}(\log N)$ and the diameter (the worst number of hops required to reach a target user) is $\mathcal{O}(\log N)$ (e.g., Chord, Kademlia, Pastry [6], Tapestry [7]), or 2) the routing table size is $\mathcal{O}(d)$ and the diameter is $\mathcal{O}(N^{1/d})$ (e.g., CAN). It is obvious that there exists a fundamental tradeoff between the routing table size and the diameter [8]. Indeed, it has been shown that $\Omega(\log N/\log\log N)$ and $\Omega(\log N)$ are the lower bounds when routing table sizes are no more than $\log N$ and d, respectively [9]. Thus, typical DHTs are close to optimal (in the sense of satisfying the lower bounds above) but not exactly optimal, and therefore, DHTs satisfying the optimality would be desired.

One of the main purposes of this paper is to propose Distributed Line Sharp (DL$^{\#}$) graphs as an overlay of DHT-based P2P networks. The class of DL$^{\#}$ graphs contains the class of Distributed Line (DL) graphs and that of DL plus (DL^{+}) graphs introduced by Zhang and Liu [10], [11]. DL graphs and DL^{+} graphs are generated from regular graphs, including De Bruijn graphs and Kautz graphs [12], [13], with an aim to locally handle user joins and user leaves. More precisely, DL^{+} graphs are considered to be an extension of DL graphs, based on the vertex merging in graph theory, so that both user joins and user leaves in a proper manner. DL^{+} graphs also possess additional conditions for the vertices to maintain the balance of networks. DL$^{\#}$ graphs, on the other hand, relax the conditions in order to consider more general networks, including unbalanced networks. This relaxation not only expands the possibility of finding better network topologies, but also covers more practical cases. In addition, it can also reduce the complexity of the user joins and user leaves.

For DL graphs and DL^{+} graphs graphs, it is observed in [11] that the out-degree is constant and the diameter is $\mathcal{O}(\log N)$, such as Viceroy [14] and Koorde [15] (both of which are based on De Bruijn graphs). Furthermore, routing process is simple and straightforward, and a path from the source to the target can be obtained from their IDs. It implies that a new path can be easily found even though some edges are missing by accident (i.e., edge-fault tolerance). Indeed, DL graphs and DL^{+} graphs are considered to be a generalization of SKY [16], a DHT based on Kautz graphs which has remarkable effectiveness compared with other Kautz-based DHT algorithms such as FissionE [17] and Moore [18]. DL$^{\#}$ graphs also possess such desired properties. Therefore, DL$^{\#}$ graphs can be a suitable candidate for a DHT-based P2P network topology.

It is important to analyze properties of network topologies based on DL$^{\#}$ graphs. Indeed, theoretical results on DL$^{\#}$ graphs guarantee the performance of DL$^{\#}$-based P2P networks. Since DL$^{\#}$ graphs are based on DL graphs, the study on DL graphs is the core of the study of DL$^{\#}$ graphs. However, many results are provided only with the outline

• *The authors are with the Graduate School of Informatics and Engineering, University of Electro-Communications, Tokyo 182-8585, Japan. E-mail: {amanada, morita}@uec.ac.jp.*

Manuscript received 24 May 2017; revised 16 Oct. 2017; accepted 5 Jan. 2018. Date of publication 11 Jan. 2018; date of current version 11 Sept. 2019. (Corresponding author: Akiko Manada.) Recommended for acceptance by J. Tang. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TNSE.2018.2792483

of proofs. The results are based on an intuitive assumption that DL graphs are suffix-free (so any ID is not a suffix of any other IDs), which gives us the central concept of routing in DL graphs. We therefore first present a mathematically sound proof of suffix-free (Proposition 2). We then show important properties of DL graphs such as the in-degree of a vertex in a DL graph (Theorem 4) which is more accurate than the result in [11]. We also focus on deriving the lower bound of the diameter (Theorem 8), and the explicit diameter of a DL graph generated from a complete graph (Collorary 12).

Another main purpose of this paper is to show that the diameter of a DL$^{\#}$ graph is equal to the diameter of its corresponding DL graph when an initial graph is a complete graph (Theorem 13). The result implies that we can directly apply existing theoretical results on the diameters of DL graphs to DL$^{\#}$ graphs. Thus, combining this result with results on DL graphs can strongly support effectiveness of DL$^{\#}$ graphs for P2P network topologies.

The rest of the paper is organized as follows. In Section 2, we go over fundamental background on graph theory and languages which are the core of the analysis. In Section 3, we introduce the precise definition of DL graphs and DHTs, together with examples. In Section 4, we show useful properties of DL graphs. Based upon the properties, we first discuss in Section 5 the in-degrees of vertices in DL graphs and present some simulation results. We next present in Section 6 a main theorem regarding the diameter of a DL graph. We then move our focus to DL$^{\#}$ graphs in Section 7, and consider the diameters of DL$^{\#}$ graphs. Section 8 discusses simulation results on DL$^{\#}$ graphs to support efficiency of our contributions. In particular, we focus on the diameters, the average path length, and the betweenness centrality. We terminate this paper with conclusion and future works in Section 9.

2 Preliminaries

2.1 Graph Theory

We begin with fundamental background on graph theory, based on [19], [20]. Let $G = (V, E)$ be a directed graph with vertex set V and edge set $E \subset V \times V$. Throughout this paper, we assume that G is *simple*; that is, there are no multiple edges or self-loops.

For a directed graph G, an edge $e = (\mathbf{u}, \mathbf{v})$ in G is called an *incoming edge* of $\mathbf{v}$, and an *outgoing edge* of $\mathbf{u}$. For a vertex $\mathbf{v}$ in G, an *in-neighbour* (*resp. out-neighbour*) of $\mathbf{v}$ is a vertex $\mathbf{u}$ such that $(\mathbf{u}, \mathbf{v}) \in E$ (*resp.* $(\mathbf{v}, \mathbf{u}) \in E$), and the sets of in-neighbours and out-neighbours are denoted by $N_G^{IN}(\mathbf{v})$ and $N_G^{OUT}(\mathbf{v})$, respectively. We further define the neighbour set $N_G(\mathbf{v})$ of $\mathbf{v}$ to be $N_G(\mathbf{v}) := N_G^{IN}(\mathbf{v}) \cup N_G^{OUT}(\mathbf{v})$.

The *in-degree* $\delta_{out}(\mathbf{v})$ and the *out-degree* $\delta_{out}(\mathbf{v})$) of $\mathbf{v}$ satisfy $\delta_{in}(\mathbf{v}) = card(N_G^{IN}(\mathbf{v}))$ and $\delta_{out}(\mathbf{v}) = card(N_G^{OUT}(\mathbf{v}))$ when G is simple. For $d \in \mathbb{N}$, we call a graph G *d-in-regular* (*resp. d-out-regular*) if $\delta_{in}(\mathbf{v}) = d$ (*resp.* $\delta_{out}(\mathbf{v}) = d$) for each vertex $\mathbf{v}$ in G. In particular, G is called *d-regular* if G is d-in-regular and d-out-regular. A typical example of a d-regular graph would be the *complete graph* $K_{d+1} = (V, E)$, consisting of $d + 1$ vertices, such that for any distinct vertices $\mathbf{u}, \mathbf{v}$ in K_{d+1}, $(\mathbf{u}, \mathbf{v}) \in E$; that is, each vertex in K_{d+1} is adjacent to any other vertices in K_{d+1}.

A (directed and non-empty) *path* π from $\mathbf{u}$ to $\mathbf{v}$ is a sequence of vertices $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_\ell$ (for some $\ell \geq 1$) such that $\mathbf{x}_0 = \mathbf{u}$, $\mathbf{x}_\ell = \mathbf{v}$ and $(\mathbf{x}_i, \mathbf{x}_{i+1}) \in E$ for each $0 \leq i \leq \ell - 1$. In particular, a path is called a *cycle* if $\mathbf{x}_0 = \mathbf{x}_\ell$. To clarify the transitions in π, we represent the path π as $\pi : \mathbf{u} = \mathbf{x}_0 \to \mathbf{x}_1 \to \cdots \to \mathbf{x}_\ell = \mathbf{v}$. The *length* of π is the number of transitions in π (that is, ℓ) and it is denoted by $l(\pi)$. A directed graph G is called *irreducible* (or *strongly connected*) if for any pair of vertices $\mathbf{u}$ and $\mathbf{v}$ in G, there exists a path from $\mathbf{u}$ to $\mathbf{v}$. We always assume that a directed graph G is irreducible if not stated.

2.2 Languages

We next go over the fundamental background on languages, based on [21], [22]. Let Σ be an *alphabet*, a finite set of symbols. A *word* $\boldsymbol{w} = w_1 w_2 \ldots w_n$ (for some $n \in \mathbb{N}$) is a finite-length sequence over Σ, and n is the *length* of $\boldsymbol{w}$, denoted by $|\boldsymbol{w}|$. The *empty word* ϵ is the word such that $|\epsilon| = 0$. We set $\boldsymbol{w} = w_1 w_2 \ldots w_n = \epsilon$ when $n = 0$.

For a word $\boldsymbol{w} = w_1 w_2 \ldots w_n$, a *subword* $\boldsymbol{x} = x_1 x_2 \ldots x_m$ is a word with $m \leq n$ such that $\boldsymbol{x} = w_i w_{i+1} \ldots w_{i+m-1}$ for some $1 \leq i \leq n - m + 1$. In particular, $\boldsymbol{x}$ is called a *prefix* of $\boldsymbol{w}$ if $i = 1$, and a *suffix* of $\boldsymbol{w}$ if $i = n - m + 1$. We use the notations $\boldsymbol{x} \prec_p \boldsymbol{w}$ and $\boldsymbol{x} \prec_s \boldsymbol{w}$ to denote that $\boldsymbol{x}$ is a prefix of $\boldsymbol{w}$ and a suffix of $\boldsymbol{w}$, respectively.

Given words $\boldsymbol{w}$ and $\boldsymbol{w}'$, $\boldsymbol{w}\boldsymbol{w}'$ is the word generated by concatenating $\boldsymbol{w}$ followed by $\boldsymbol{w}'$. We define by convention that $\boldsymbol{w}\epsilon = \epsilon\boldsymbol{w} = \boldsymbol{w}$ for any word $\boldsymbol{w}$. In addition, define $\boldsymbol{w} \diamond \boldsymbol{w}'$ to be the longest word z such that $z \prec_s \boldsymbol{w}$ and $z \prec_p \boldsymbol{w}'$.

Example 1. Let $\boldsymbol{w} = 2101$ and $\boldsymbol{w}' = 1011$. The prefixes of $\boldsymbol{w}'$ are $\epsilon, 1, 10, 101, 1011$ and the suffices of $\boldsymbol{w}$ are $\epsilon, 1, 01, 101, 2101$. Therefore, the longest word z such that $z \prec_p \boldsymbol{w}$ and $z \prec_s \boldsymbol{w}'$ is 101, and thus, $\boldsymbol{w} \diamond \boldsymbol{w}' = 101$.

3 Distributed Line Transform and Distributed Line Graphs

In this section, we go over the definition of a distributed line transform (DLT) and distributed line graphs (DL graphs) obtained under DLTs. Throughout this paper, a word over Σ is assigned to each vertex as its ID, and we identify a vertex with its ID. Thus, for vertex $\mathbf{v}$ with ID $\boldsymbol{v} = v_1 v_2 \ldots v_n$, we consider $\mathbf{v} = \boldsymbol{v} = v_1 v_2 \ldots v_n$.

Suppose that a d-regular graph $G_0 = (V_0, E_0)$ is given as an initial graph, and assign distinct symbols to the vertices. Define an alphabet Σ to be the set of symbols assigned to the vertices of G_0; that is, $\Sigma = V_0$. Without loss of generality, set $\Sigma := \{0, 1, \ldots, q-1\}$ when there are q vertices in G_0.

We first introduce the definition of DLT. Zhang and Liu [11] use the *conjunction operation* to define DLT, but we modify the definition to clarify its properties.

Definition 1 (modification of [11, Definition 1]). *Suppose that a d-regular graph $G_0 = (V_0, E_0)$ is given as an initial graph. Let $G = (V, E)$ be a directed graph, and $\mathbf{r} = r_1 r_2 \ldots r_\ell$ be a vertex in G (a word over Σ) such that $|\mathbf{r}| \leq |\mathbf{x}|$ for any $\mathbf{x} \in N_G(\mathbf{r})$. A DLT is the procedure to generate a new graph G' from G as follows.*

Step 1: Delete $\mathbf{r}$ and edges attached to $\mathbf{r}$ from G.

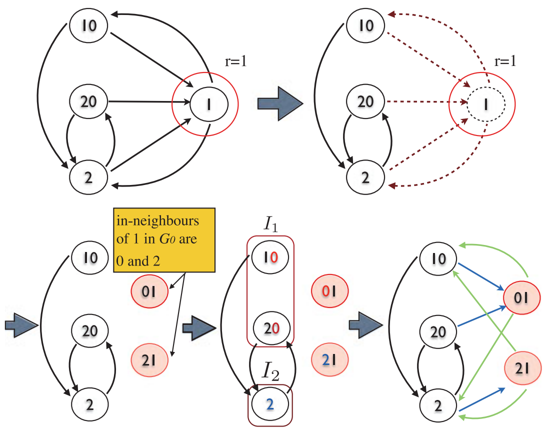

Fig. 1. An example of DLT w.r.t. $\mathbf{r} = 1$.

Step 2: Add new vertices $a_1\mathbf{r}, a_2\mathbf{r}, \ldots, a_d\mathbf{r}$ *to* G, *where* $a_i \in N^{IN}_{G_0}(r_1)$ *for each* $1 \le i \le d$.

Step 3: Partition $N^{IN}_G(\mathbf{r})$ *into* d *disjoint sets* $I_1, I_2, \ldots, I_d$ *so that*

$$\mathbf{u} = u_1 u_2 \ldots u_{|\mathbf{u}|} \in I_i \Longleftrightarrow u_{|\mathbf{u}|-|\mathbf{r}|+1} = a_i.^1$$

Step 4: Assign edges from vertices in I_i *to* $a_i\mathbf{r}$ *for each* $1 \le i \le d$.

Step 5: Assign edges from $a_i\mathbf{r}$ *to vertices in* $N^{OUT}_G(\mathbf{r})$ *for each* $1 \le i \le d$.

We denote $G' = DL(G, \mathbf{r})$ *to emphasize that* G' *is generated from* G *by applying the DLT with respect to (w.r.t.)* $\mathbf{r}$. *The vertex* $\mathbf{r}$ *is called the responsible vertex of the DLT.*

Definition 2 (modification of [11, Definition 1]). *Let* G_0 *be an initial* d*-regular graph. A (base-*d*) DL graph* G *is a graph generated from* G_0 *by applying DLT finite times. We denote by* $G_i = (V_i, E_i)$ *a DL graph generated from* G_0 *by applying DLT* i *times. When a DL graph* $\tilde{G}$ *can be generated from* G *by a proper iteration of DLTs, then we say that* $\tilde{G}$ *is obtainable from* G.

Example 2. Let an initial graph G_0 be the complete graph K_3; that is, there are three vertices labelled 0, 1, 2 in K_3 and they are adjacent each other.

Let G be the leftmost DL graph in Fig. 1 generated from G_0. Suppose that we generate new DL graph G' from G applying the DLT w.r.t. $\mathbf{r} = 1$. First delete vertex 1 and edges attached to 1 from G *(Step 1)*. Since $N^{IN}_{G_0}(r_1) = N^{IN}_{G_0}(1) = \{0, 2\}$, add $a_1\mathbf{r} = 01$ and $a_2\mathbf{r} = 21$ to G *(Step 2)*. For vertex 2 in $N^{IN}_G(1) = \{2, 10, 20\}$, focus the first symbol of 2 (which is 2) since $|2| - |\mathbf{r}| + 1 = 1$. Similarly, focus the second symbols of 10 and 20 in $N^{IN}_G(1)$ (which are both 0) since $|10| - |\mathbf{r}| + 1 = |20| - |\mathbf{r}| + 1 = 2$. Then partition $N^{IN}_G(1)$ into $I_1 = \{10, 20\}$ and $I_2 = \{2\}$ *(Step 3)* based on these symbols. Finally, for each $i \in \{1, 2\}$, assign edges from vertices in I_i to $a_i\mathbf{r}$ *(Step 4)* and edges from $a_i\mathbf{r}$ to vertices in $N^{OUT}_G(1)$ *(Step 5)*. The resulting graph G' is the rightmost DL graph in Fig. 1.

Example 3. Fig. 2 represents an iteration of DLTs, where the initial graph G_0 is the complete graph K_4; so $d = 3$ and $\Sigma = \{0, 1, 2, 3\}$. The responsible vertex of the first DLT is 0 and that of the second DLT is 1.

1. You will see that we can always find such a_i from Lemma 6.

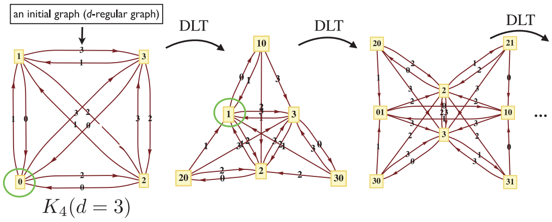

Fig. 2. An iteration of DLTs.

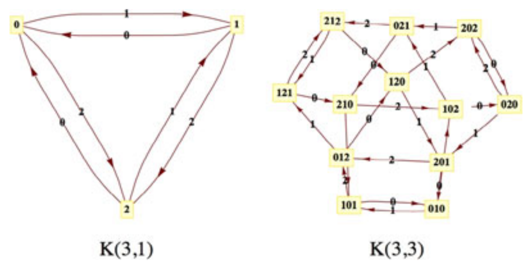

Fig. 3. Kautz graphs; The left is $K(1, 3)$ and the right is $K(3, 3)$.

If $G_0 = K_q$ for some $q \in \mathbb{N}$ and if we always pick a vertex with the shortest length as the responsible vertex of each DLT, then we obtain following graphs (Fig. 3) called Kautz graphs $K(\ell, q)$, where the vertex set of $K(\ell, q)$ is the set of all words $\boldsymbol{v} = v_1 v_2 \ldots v_\ell$ over $\Sigma = \{0, 1, \ldots, q-1\}$ with constant length ℓ such that any adjacent symbols v_i and v_{i+1} (for $1 \le i \le \ell - 1$) are distinct.

Remark 1. Let G_0 be an initial d-regular graph. Then for each base-d DL graph G_i generated from G_0, we have the following.

1) The out-degree $\delta^{OUT}_{G_i}(\mathbf{v}) = d$ for each $\mathbf{v} \in V_i$. That is, the out-degree is always constant even though the number of vertices in a graph gets larger.
2) Let $i \ge 1$ and suppose $G_i = DL(G_{i-1}, \mathbf{r})$. Then

$$V_i = (V_{i-1} \setminus \{\mathbf{r}\}) \cup \{a\mathbf{r} \mid a \in N^{IN}_{G_0}(r_1)\},$$

where r_1 is the first symbol of $\mathbf{r}$.
3) If G_0 is irreducible, then G_i is irreducible.

4 Useful Properties of DL Graphs

In this section, we introduce some useful properties on DL graphs which will be used later in proving main results. These propositions focus on suffixes of vertices in DL graphs.

Proposition 1. *Let* G_0 *be an initial graph and consider two DL graphs* $G = (V, E)$ *and* $\tilde{G} = (\tilde{V}, \tilde{E})$ *generated from* G_0. *If* $\tilde{G}$ *is obtainable from* G, *then for any vertex* $\tilde{\mathbf{w}} \in \tilde{V}$, *there exists a vertex* $\mathbf{w} \in V$ *such that* $\mathbf{w} \prec_s \tilde{\mathbf{w}}$.

Proof. It is enough to show that the statement holds when $\tilde{G} = DL(G, \mathbf{r})$ for some responsible vertex $\mathbf{r} \in V$.

Recall that $\tilde{V} = (V \setminus \{\mathbf{r}\}) \cup \{a\mathbf{r} \mid a \in N^{IN}_{G_0}(r_1)\}$. Hence, for a vertex $\tilde{\mathbf{w}} \in \tilde{V}$, if $\tilde{\mathbf{w}} \in (V \setminus \{\mathbf{r}\})$, then set $\mathbf{w} = \tilde{\mathbf{w}}$. If $\tilde{\mathbf{w}} \in \{a\mathbf{r} \mid a \in N^{IN}_{G_0}(r_1)\}$, then set $\mathbf{w} = \mathbf{r}$. Then $\mathbf{w}$ is in V and $\mathbf{w}$ is a suffix of $\tilde{\mathbf{w}}$. $\square$

The following proposition states that a DL graph is a *suffix-free*, which is a core of the analysis of DL graphs.

Proposition 2. *Let G be a DL graph generated from an initial graph G_0. Then G is suffix-free; that is, for each vertex $\mathbf{x}$ in G, there is no vertex $\mathbf{y}$ in G such that $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{y} \prec_s \mathbf{x}$.*

Proof. Let $G = G_i$ and we prove it by mathematical induction on i. The statement is true when $i = 0$ (for the initial graph G_0). So assume that the statement is true when $i = k$ and consider the case when $i = k + 1$.

Let $G_{k+1} = DL(G_k, \mathbf{r})$. Pick distinct vertices $\mathbf{x}, \mathbf{y} \in V_{k+1}$, arbitrarily. Then (without loss of generality) there are three cases for the choice of $\mathbf{x}$ and $\mathbf{y}$; (1) $\mathbf{x}, \mathbf{y} \in V_{k+1} \cap V_k$, (2) $\mathbf{x} \in V_{k+1} \cap V_k$ and $\mathbf{y} \in V_{k+1} \setminus V_k$, and (3) $\mathbf{x}, \mathbf{y} \in V_{k+1} \setminus V_k$.

For (1), clearly $\mathbf{x} \not\prec_s \mathbf{y}$ and $\mathbf{y} \not\prec_s \mathbf{x}$ since G_k is suffix-free by assumption. For (2), observe that $\mathbf{y} = a\mathbf{r}$ for some $a \in N_{G_0}^{IN}(r_1)$. If $\mathbf{y} \prec_s \mathbf{x}$, then we have $\mathbf{r} \prec_s \mathbf{x}$. Conversely, if $\mathbf{x} \prec_s \mathbf{y}$, then $|\mathbf{x}| < |\mathbf{y}|$ since $\mathbf{x} \neq \mathbf{y}$. As $\mathbf{y} = a\mathbf{r}$, we have $\mathbf{x} \prec_s \mathbf{r}$. In any case, we have a contradiction since $\mathbf{x}, \mathbf{r} \in V_k$. For (3), since $\mathbf{x} = b\mathbf{r}$ and $\mathbf{y} = a\mathbf{r}$ for some distinct symbols $a, b \in N_{G_0}^{IN}(r_1)$, it is obvious that $\mathbf{x} \not\prec_s \mathbf{y}$ and $\mathbf{y} \not\prec_s \mathbf{x}$ hold. Hence, G_{k+1} is also suffix-free as required. □

5 In-Degrees of Vertices in DL Graphs

We have already seen that when a d-regular graph is given as an initial graph G_0, each DL graph G obtainable from G_0 is d-out-regular (whatever the number of vertices in G), which is equivalent to say that the size of a routing table each user possesses is always constant. On the other hand, G is not always d-in-regular; that is, the in-degree of a vertex can vary. In this section, we precisely determine the value of the in-degree for a vertex in a DL graph.

We first remark that Zhang and Liu introduced the following theorem regarding the in-degree of a vertex in a DL graph.

Theorem 3 (Zhang and Liu, [11]). *Let G be a base-d DL graph. Then for each vertex $\mathbf{v}$ in G, its in-degree $\delta_G^{IN}(\mathbf{v})$ satisfies*

$$1 \leq \delta_G^{IN}(\mathbf{v}) \leq d^2.$$

However, only an outline of the proof is given, so more detailed explanations will be necessary for full understanding. Therefore, we provide the following theorem which is more explicit than the previous result, together with its complete proof.

Theorem 4. *Let G be a base-d DL graph generated from an initial graph G_0. Then for each vertex $\mathbf{v}$ in G, its in-degree $\delta_G^{IN}(\mathbf{v})$ is given by*

$$\delta_G^{IN}(\mathbf{v}) = 1 + (d-1)t,$$

for some integer $0 \leq t \leq d+1$.

The theorem does not go counter to Theorem 3, but it shows that the in-degree takes the value at intervals of $d-1$. To prove the theorem, we need the following lemmas. These results are also described in [11] only with outline of proofs, so here, we introduce them together with complete proofs.

Lemma 5. *Let G be a DL graph generated from an initial graph G_0. For a vertex $\mathbf{v}$ and its in-neighbour $\mathbf{u} \in N_G^{IN}(\mathbf{v})$, we have*

$$-1 \leq |\mathbf{v}| - |\mathbf{u}| \leq 1.$$

That is, $|\mathbf{u}|$ is $|\mathbf{v}| - 1$, $|\mathbf{v}|$, or $|\mathbf{v}| + 1$.

Proof. Let $G = G_i$, and prove it by mathematical induction on i. The statement holds when $i = 0$ since each vertex in G_0 has length 1. Suppose the statement is true when $i = k$ and consider $G_{k+1} = DL(G_k, \mathbf{r})$.

Since $\mathbf{r}$ is the responsible vertex, we only need to focus on the relationship between new vertices $a\mathbf{r}$ for some $a \in N_{G_0}^{IN}(r_1)$ and $\mathbf{x} \in N_{G_k}(\mathbf{r})$. Recall that $|\mathbf{r}| \leq |\mathbf{x}|$ holds for each $\mathbf{x} \in N_{G_k}(\mathbf{r})$ from the definition of DLT, and therefore, $|\mathbf{r}| = |\mathbf{x}| - 1$ or $|\mathbf{r}| = |\mathbf{x}|$ from the assumption on G_k. Hence, for an out-neighbour $\mathbf{y}$ of $\mathbf{r}$ in G_k (so $\mathbf{r}$ is an in-neighbour of $\mathbf{y}$ in G_k), $\mathbf{y}$ has in-neighbours $a\mathbf{r}$ of length $|\mathbf{y}|$ or $|\mathbf{y}| + 1$. Similarly, for a vertex $a\mathbf{r}$ in G_{k+1}, its in-neighbour has length $|\mathbf{r}| + 1 = |a\mathbf{r}|$ or $|\mathbf{r}| = |a\mathbf{r}| - 1$. Hence, we can show that the statement is true for G_{k+1} as well. □

Lemma 6. *Let G be a DL graph generated from an initial graph G_0. For a vertex $\mathbf{v} = v_1 v_2 \ldots v_n$ and its in-neighbour $\mathbf{u} = u_1 u_2 \ldots u_m \in N_G^{IN}(\mathbf{v})$, we have*

$$\mathbf{u} = \begin{cases} v_1 v_2 \ldots v_{n-1} & (m = n-1) \\ a v_1 v_2 \ldots v_{n-1} \text{ for some } a \in N_{G_0}^{IN}(v_1) & (m = n) \\ a' a v_1 v_2 \ldots v_{n-1} \text{ for some } a \in N_{G_0}^{IN}(v_1) & \\ \text{and } a' \in N_{G_0}^{IN}(a) & (m = n+1) \end{cases}$$

Proof. It is clear that there are only three choices for $m (= |\mathbf{u}|)$ from Lemma 5. Let $G = G_i$, and prove it by mathematical induction on i. The statement holds when $i = 0$ (for the initial graph G_0). Suppose the statement is true when $i = k$ and consider $G_{k+1} = DL(G_k, \mathbf{r})$.

Since $\mathbf{r} = r_1 r_2 \ldots r_{|\mathbf{r}|}$ is the responsible vertex, we only need to focus on the relationship between new vertices $\hat{a}\mathbf{r}$ for some $\hat{a} \in N_{G_0}^{IN}(r_1)$ and $\mathbf{x} \in N_{G_k}(\mathbf{r})$. Recall that $|\mathbf{x}| = |\mathbf{r}| + 1$ or $|\mathbf{r}| = |\mathbf{x}|$ for each $\mathbf{x} \in N_{G_k}(\mathbf{r}) = N_{G_k}^{IN}(\mathbf{r}) \cup N_{G_k}^{OUT}(\mathbf{r})$.

If $\mathbf{x} \in N_{G_k}^{IN}(\mathbf{r})$, then $\mathbf{x} = a' a r_1 r_2 \ldots r_{|\mathbf{r}|-1}$ or $\mathbf{x} = a r_1 r_2 \ldots r_{|\mathbf{r}|-1}$ by the assumption on G_k. For $\hat{a}\mathbf{r}$, recall from the definition of DLT that $\mathbf{x} \in N_{G_{k+1}}^{IN}(\hat{a}\mathbf{r})$ if and only if $a = \hat{a}$, and hence, we have that each in-neighbour of $a\mathbf{r}$ in G_{k+1} satisfies the form in the statement. Similarly, if $\mathbf{x} = x_1 x_2 \ldots x_{|\mathbf{x}|} \in N_{G_k}^{OUT}(\mathbf{r})$ (so $\mathbf{r}$ is an in-neighbour of $\mathbf{x}$ in G_k), then we have $\mathbf{r} = b x_1 x_2 \ldots x_{|\mathbf{x}|-1}$ for some $b \in N_{G_0}^{IN}(x_1)$ or $\mathbf{r} = x_1 x_2 \ldots x_{|\mathbf{x}|-1}$, and hence, each vertex $\hat{a}\mathbf{r}$ satisfies the form in the statement as an in-neighbour of $\mathbf{x}$ in G_{k+1}. □

We are now in a position of proving Theorem 4.

Proof of Theorem 4. From Lemmas 5 and 6, recall that for a vertex $\mathbf{v} = v_1 v_2 \ldots v_n$ in G and its in-neighbour $\mathbf{u} = u_1 u_2 \ldots u_m \in N_G^{IN}(\mathbf{v})$, $|m - n| \leq 1$ and

$$\mathbf{u} = \begin{cases} v_1 v_2 \ldots v_{n-1} & (m = n-1) \\ a v_1 v_2 \ldots v_{n-1} \text{ for some } a \in N_{G_0}^{IN}(v_1) & (m = n) \\ a' a v_1 v_2 \ldots v_{n-1} \text{ for some } a \in N_{G_0}^{IN}(v_1) & \\ \text{and } a' \in N_{G_0}^{IN}(a) & (m = n+1) \end{cases}$$

First observe that if there exists an in-neighbour $\mathbf{u}$ of length $n-1$, then $\mathbf{u}$ must be a unique in-neighbour of $\mathbf{v}$, since otherwise, in-neighbours of length n and $n+1$ have $\mathbf{u}$ as a suffix which contradicts suffix-free property.

Now suppose that each in-neighbour has length n or $n+1$. Consider a symbol $a \in N_{G_0}^{IN}(v_1)$. Then from

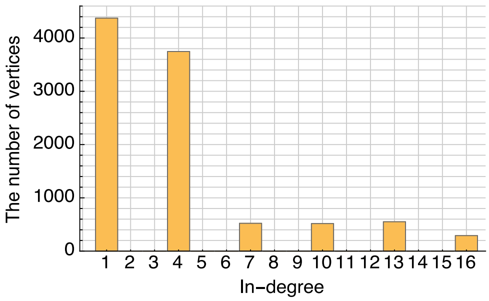

Fig. 4. The distribution of in-degrees.

suffix-free property, if a is the first symbol of an in-neighbour of length n, then it cannot be the second symbol of any in-neighbour of length $n+1$, and vice versa.

Let t' be the number of a's used as the second symbol of an in-neighbour of length $n+1$. Then there are $d-t'$ a's used as the first symbol of an in-neighbour of length n. Furthermore, since there are d choices of $a' \in N_{G_0}^{IN}(a)$ for each a used as the second symbol, we have

$$\begin{aligned} \delta_G^{IN}(\mathbf{v}) &= (d-t') + dt' \\ &= 1 + (d-1)(t'+1), \end{aligned} \tag{1}$$

where t' ranges from 0 to d. Substituting $t'+1$ in (1) with t, and noting that $\delta_G^{IN}(\mathbf{v}) = 1$ can be given when $t=0$, we obtain the theorem. □

Example 4. Fig. 4 represents the distributions of in-degrees for a DL graph consisting of 10,000 vertices,[2] where an initial graph G_0 is the complete graph K_5 (4-regular, thus $d=4$). It can be confirmed that in-degrees take the value only at $1+3t(=1+(d-1)t)$ for some t, which fits the statement of Theorem 4.

6 The Diameter of a DL Graph

For a directed graph $G=(V,E)$, define $d_G(\mathbf{u},\mathbf{v})$ to be the length of a shortest path from vertex $\mathbf{u}$ to vertex $\mathbf{v}$. The *diameter* of a graph G, denoted by $D(G)$, is given by

$$D(G) = \max_{\mathbf{u},\mathbf{v}\in V} d_G(\mathbf{u},\mathbf{v}).$$

Observe that $D(G) < \infty$ when G is irreducible, and $D(G) = \infty$ otherwise.

The diameter is an important value since it gives us how closely vertices are connected each other in a network. Indeed, it is preferable to have a graph with small diameter as a topology of a network, and there have been many existing studies regarding diameters and their comparisons (see, for example, [23]). Many typical networks have diameters $\mathcal{O}(\log N)$ with $\mathcal{O}(\log N)$ neighbours. On the other hand, DL graphs have diameters $\mathcal{O}(\log N)$ with a constant number of neighbours. In this section, we further analyze the diameters of DL graphs based on the routing scheme for DL graphs, and present an explicit diameter when an initial graph is a complete graph.

6.1 Routing Scheme for DL Graphs

Let G be a DL graph generated from an initial graph G_0. From Lemma 6, we have that for a vertex $\mathbf{x} = x_1x_2\ldots x_n$, an out-neighbour $\mathbf{y} = y_1y_2\ldots y_m$ of $\mathbf{x}$ is $\mathbf{y} = x_3\ldots x_nc$, $\mathbf{y} = x_2x_3\ldots x_nc$ or $\mathbf{y} = x_1x_2\ldots x_nc$ for some $c \in N_{G_0}^{OUT}(x_n)$. That is, a transition along an outgoing edge from the present vertex reaches a vertex with only one new symbol attached to the end. Using the property, the routing scheme (i.e., how to get to the target from the source) for G is described as follows. The scheme is intuitive and elementary, but it is supported by the suffix-free property in Proposition 2.

Routing scheme for DL graphs. Let G be a DL graph, and suppose that we want to find a path from the source $\mathbf{s} = s_1s_2\ldots s_m$ to the target $\mathbf{t} = t_1t_2\ldots t_n$ in G.

Step 1: Compute $z = \mathbf{s} \diamond \mathbf{t}$; the longest word z such that $z \prec_s \mathbf{s}$ and $z \prec_p \mathbf{t}$.

Step A: If $z \neq \epsilon$ and $|z| = k \geq 1$, then find a path $\pi : \mathbf{s} = \mathbf{p}_0 \to \mathbf{p}_1 \to \cdots \to \mathbf{p}_{n-k}$ such that the jth transition in π (the outgoing edge from $\mathbf{p}_{j-1}$ to $\mathbf{p}_j$) reaches a vertex with t_{j+k} at the end; that is, vertex $\mathbf{p}_j$ ends with t_{j+k}. Such a path π is uniquely determined, and the terminal vertex $\mathbf{p}_{n-k}$ is $\mathbf{t}$.

Step B: If $z = \epsilon$, then it is necessary to first reach a vertex $\mathbf{w}$ ending with t_1 from $\mathbf{s}$. To do so, find a path $\hat{\rho} : s_m = q_0 \to q_1 \to \cdots \to q_h = t_1$ in G_0. Then, based on $\hat{\rho}$, we can find a path ρ of length $l(\hat{\rho})$ in G from $\mathbf{s}$ to $\mathbf{w}$. After that, follow Step A to find a path π from $\mathbf{w}$ to $\mathbf{t}$.

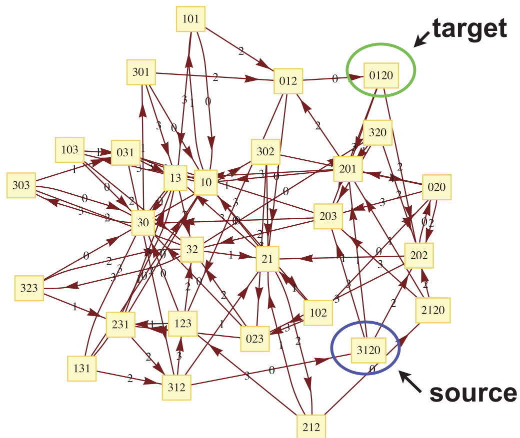

Fig. 5. An example of routing from $\mathbf{s} = 3120$ to $\mathbf{t} = 0120$.

Example 5. Given a DL graph in Fig. 5 with the initial graph K_4, and suppose that we want to find a path from the source $\mathbf{s} = 3120$ to the target $\mathbf{t} = 0120$.

First observe that $z = \mathbf{s} \diamond \mathbf{t} = 0 \neq \epsilon$, and $|z| = 1$. Then by following the routing scheme, we can find a path π such that

$$\pi : \mathbf{s} = 3120 \to 201 \to 012 \to 0120 = \mathbf{t},$$

where the jth transition of π $(1 \leq j \leq 3)$ reaches a vertex ending with t_{j+1}. This is a shortest path from $\mathbf{s}$ to $\mathbf{t}$ and hence, $d_G(\mathbf{s},\mathbf{t}) = 3$.

From the routing scheme, we can obtain the following lemma.

Lemma 7. *Let G_0 be a d-regular graph, and G be a base-d DL graph generated from G_0. Then the length of a shortest path from the source $\mathbf{s} = s_1s_2\ldots s_m$ to the target $\mathbf{t} = t_1t_2\ldots t_n$ is given by*

2. There are, to be exact, 10,001 vertices in the graph.

$$d_G(\mathbf{s},\mathbf{t}) = \begin{cases} |\mathbf{t}| - |\mathbf{s} \diamond \mathbf{t}| & (\mathbf{s} \diamond \mathbf{t} \neq \epsilon) \\ d_{G_0}(s_m, t_1) + |\mathbf{t}| - 1 & (\text{otherwise}) \end{cases}$$

Proof. When $\mathbf{s} \diamond \mathbf{t} \neq \epsilon$, it is obvious from the routing scheme that $d_G(\mathbf{s},\mathbf{t}) \leq |\mathbf{t}| - |\mathbf{s} \diamond \mathbf{t}|$ holds. Furthermore, recall that a transition along an outgoing edge from the present vertex reaches a vertex with only one new symbol attached to the end. Hence, for each path π from $\mathbf{s}$ to $\mathbf{t}$, we have $l(\pi) \geq |\mathbf{t}| - |\mathbf{s} \diamond \mathbf{t}|$. Therefore, $d_G(\mathbf{s},\mathbf{t}) = |\mathbf{t}| - |\mathbf{s} \diamond \mathbf{t}|$ as required.

When $\mathbf{s} \diamond \mathbf{t} = \epsilon$, consider paths from $\mathbf{s}$ to vertices $\mathbf{w}$ ending with t_1. Let η be a path of the shortest length amongst those paths. Then clearly $l(\eta) = d_{G_0}(s_m, t_1)$. Furthermore, for the vertex $\mathbf{w}'$ reachable from $\mathbf{s}$ by following η, we have $|\mathbf{w}' \diamond \mathbf{t}| = 1$ since otherwise, contradicts the assumption that $\mathbf{s} \diamond \mathbf{t} = \epsilon$ or the assumption on η. Therefore, $d_G(\mathbf{s},\mathbf{t}) = d_{G_0}(s_m, t_1) + d_G(\mathbf{w}', \mathbf{t}) = d_{G_0}(s_m, t_1) + |\mathbf{t}| - 1$, which completes the proof. □

6.2 The Diameter of a DL Graph

Zhang and Liu state in [11] (and it is also clear from Lemma 7) that the diameter $D(G)$ of a DL graph G generated from an initial graph G_0 satisfies

$$D(G) \leq D(G_0) + \ell^G_{\max} - 1, \tag{2}$$

where $\ell^G_{\max}$ is the length of a longest vertex in G. On the other hand, lower bounds on the diameter have not been well discussed. The main result of this section is the following which denotes an explicit lower bound of a DL graph, using the lengths of IDs in the graph.

Theorem 8. *Set $d \geq 2$, and consider a DL graph G generated from an initial d-regular graph G_0. Then the diameter $D(G)$ of G satisfies*

$$\ell^G_{\max} \leq D(G),$$

where ℓ^G_{max} is the length of a longest ID (vertex) in G.

We first show some lemmas and a corollary which will be used in the proof of Lemma 8.

Lemma 9. *Set $d \geq 2$ and consider a d-out-regular graph G. Then for each vertex $\mathbf{v}$ in G, there exists a cycle γ not containing $\mathbf{v}$.*

Proof. First remark that for a directed graph such that $\delta_{out}(\mathbf{v}) \geq 1$ for each vertex $\mathbf{v}$, a cycle does exist. This property can be easily verified using Pigeonhole principle.

Now fix a vertex $\mathbf{v}$ in G, and consider a subgraph G' of G generated by deleting $\mathbf{v}$ and its attached edges from G. Observe that (since G is simple) each vertex $\mathbf{v}'$ in G' has out-degree $\delta_{out}(\mathbf{v}') \geq d - 1 = 1$. Hence, from the remark above, G' has a cycle γ. Clearly, this γ does not contain vertex $\mathbf{v}$, and γ is a cycle in G. Therefore, we can conclude that for each vertex $\mathbf{v}$ in G, there exists a cycle γ not containing $\mathbf{v}$. □

Lemma 10. *Let $G = (V, E)$ be a DL graph generated from an initial graph $G_0 = (V_0, E_0)$, such that each vertex has the same length ℓ. Then for a word $w = w_1 w_2 \dots w_\ell \in \Sigma^\ell$, vertex $\mathbf{w} = w$ is in V if and only if there exists a path $\pi : w_1 \to w_2 \to \cdots \to w_\ell$ in G_0.*

Proof. We first show the existence of vertex implies the existence of path. Suppose that there exists a vertex $\mathbf{w} = w_1 w_2 \dots w_\ell \in V$. Then, from the definition of DLT, $w_i \in N^{IN}_{G_0}(w_{i+1})$ for each i. It automatically implies the existence of path π as desired.

We next show the existence of path implies the existence of vertex. Suppose that there exists a path $\pi : w_1 \to w_2 \to \cdots \to w_\ell$ in G_0. Now consider a chain of graphs $\mathcal{C} : G_0 \rightharpoonup G_1 \rightharpoonup \cdots \rightharpoonup G_k = G$ such that $G_i = DL(G_{i-1}, \mathbf{r}_i)$. To obtain graph G from G_0, there must exist a unique integer i_1, $1 \leq i_1 \leq k$, such that $G_{i_1} = DL(G_{i_1 - 1}, w_\ell)$. Observe that the word $w_{\ell-1} w_\ell$ of length 2 is a vertex in G_{i_1} as $w_{\ell-1} \in N^{IN}_{G_0}(w_\ell)$. Similar argument yields that there exists a unique sequence of integers $i_1, i_2, \dots, i_{\ell-1}$ with $1 \leq i_1 < i_2 < \cdots < i_{\ell-1} \leq k$ such that $G_{i_j} = DL(G_{i_j - 1}, w_{\ell-j} w_{\ell-j+1} \dots w_\ell)$ for $1 \leq j \leq \ell - 1$. Hence, $G_{i_{\ell-1}}$ has $\mathbf{w}$ as a vertex, and therefore (as we do not apply DLT w.r.t. $\mathbf{w}$ afterwords to obtain G from $G_{i_{\ell-1}}$), G has $\mathbf{w}$ as a vertex as well. □

Corollary 11. *Let $G = (V, E)$ and $G' = (V', E')$ be DL graphs generated from an initial graph G_0 such that any vertex in V and that in V' has the same length ℓ for some $\ell \in \mathbb{N}$; that is, $|\mathbf{v}| = |\mathbf{v}'| = \ell$ for each vertex $\mathbf{v} \in V$ and $\mathbf{v}' \in V'$. Then G and G' are the same.*

Proof. From Lemma 10, we have $V = V'$. Furthermore, from the allocations of edges between vertices, we also have $E = E'$. □

Corollary 11 implies that for each $\ell \in \mathbb{N}$, a graph H_ℓ with the constant vertex length ℓ uniquely exists. We name the graph H_ℓ the *ℓ-proper graph* generated from G_0. Observe that $H_\ell = K(\ell, q)$ if and only if the initial graph G_0 is the complete graph K_q.

We are now in a position of proving Theorem 8.

Proof of Theorem 8. Let $\ell^G_{\max} = \ell$. To show the statement, it is enough to prove the existence of vertices $\mathbf{x}$ and $\mathbf{y}$ in G such that

1) $|\mathbf{y}| = \ell^G_{\max} = \ell$; and
2) $\mathbf{x} \diamond \mathbf{y} = \epsilon$.

Indeed, if there exist such vertices, then we have $d_G(\mathbf{x}, \mathbf{y}) \geq \ell$ by Lemma 7, which automatically implies $D(G) \geq \ell$.

Now consider the ℓ-proper graph graph H_ℓ. First observe that H_ℓ is obtainable from G by choosing a vertex with the shortest length as the responsible vertex of each DLT.

Now take a vertex $\mathbf{y} = y_1 y_2 \dots y_\ell$ in G such that $|\mathbf{y}| = \ell$. Clearly, $\mathbf{y}$ is a vertex in H_ℓ. For this $\mathbf{y}$, from Lemmas 9 and 10, we can find a vertex $\mathbf{w}$ in H_ℓ such that $\mathbf{w}$ does not contain y_1 at any position, by considering a cycle γ in G_0 not containing y_1. Also, since H_ℓ is obtainable from G, from Proposition 1, we can find a vertex $\mathbf{x}$ in G such that $\mathbf{x} \prec_s \mathbf{w}$. For this $\mathbf{x}$, we have $\mathbf{x} \diamond \mathbf{y} = \epsilon$, since otherwise, contradicts the fact that $\mathbf{w}$ does not contain y_1 at any position. Therefore, we can show the existence of such $\mathbf{x}$ and $\mathbf{y}$ as desired. □

Theorem 8 and Inequality (2) imply the following corollary which explicitly computes the diameter of a DL graph when an initial graph G_0 is the complete graph K_d.

Corollary 12. *Let an initial graph G_0 be the complete graph K_{d+1} with $d \geq 2$. Then for a DL graph G generated from G_0, we have*

$$D(G) = \ell_{\max}^G.$$

Proof. Theorem 8 and Inequality (2) imply

$$\ell_{\max}^G \leq D(G) \leq D(G_0) + \ell_{\max}^G - 1.$$

Furthermore, since $D(G_0) = 1$ as G_0 is the complete graph, we have $D(G) = \ell_{\max}^G$. □

Suppose a complete graph is given as an initial graph. The corollary also implies that if a vertex $\mathbf{v}$ with the shortest length is chosen as the responsible vertex at each DLT, then for each $i \geq 0$, the resulting graph G_i^* after applying DLT i times in that manner has the smallest diameter amongst all DL graphs G_i's which can be generated from G_0 by applying DLT i times.

7 DL$^\#$ Graphs and Their Diameters

As we have seen so far, DL graphs have desirable properties and can be suitable candidates for network topologies. However, as we can see from 2) in Remark 1, each DLT generates new d vertices, even though only fewer vertices are necessary for new users joining a network. Also, it is necessary to design graphs so that not only user join but user leave should be properly handled. In this section, we introduce *DL$^\#$ graphs* and present a result on the diameter of a DL$^\#$ graph.

7.1 Vertex Merging and DL$^\#$ Graphs

In graph theory, vertex merging is, roughly speaking, a process to identify two or more vertices with one vertex. More precisely, vertex merging is e defined as follows.

Definition 3. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be k vertices in a graph G. Vertex merging of these k vertices is a procedure to identify them with one vertex $\mathbf{v}$. More precisely,*

$$N_{\widehat{G}}^{OUT}(\mathbf{v}) = \bigcup\nolimits_{i=1}^{k} N_G^{OUT}(\mathbf{v}_i); \textit{ and}$$
$$N_{\widehat{G}}^{IN}(\mathbf{v}) = \bigcup\nolimits_{i=1}^{k} N_G^{IN}(\mathbf{v}_i),$$

where $\widehat{G}$ is the resulting graph after vertex merging. We call vertex $\mathbf{v}$ a physical vertex, *and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$* logical vertices *of $\mathbf{v}$. For a physical vertex $\mathbf{v}$, the number of its logical vertices is denoted by $||\mathbf{v}||$. One of the IDs of logical vertices is randomly chosen and assigned as the ID of $\mathbf{v}$, but we assume that $\mathbf{v}$ possesses all of its logical vertices' IDs.*

It is possible to apply vertex merging to any vertices, but throughout this paper, we apply vertex merging only for the vertices generated at the same DLT, which we call *sibling vertices*. More precisely, if a d-regular graph is given as an initial graph, the d sibling vertices have the form $a_1\mathbf{r}, a_2\mathbf{r}, \ldots, a_d\mathbf{r}$, where $\mathbf{r}$ is the responsible vertex of some DLT. Based on DLTs and vertex merging for sibling vertices, DL$^\#$ graphs are defined as follows.

Definition 4. *Let G_0 be an initial d-regular graph. A (base-d) DL$^\#$ graph $G^\#$ is a graph generated from G_0 by an iteration of DLTs followed by a series of vertex merging for sibling vertices w.r.t. vertices $\mathbf{v}$ satisfying*

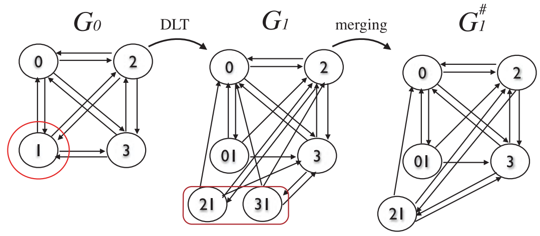

Fig. 6. An example of generating a DL$^\#$ graph.

$$|\mathbf{v}| \geq |\mathbf{u}| \text{ for any neighbour } \mathbf{u} \text{ of } \mathbf{v}. \tag{3}$$

It is clear from the definition that the set of DL graphs is (properly) included in the class of DL$^\#$ graphs. In other words, DL graphs are special DL$^\#$ graphs such that $||\mathbf{v}|| = 1$ for each vertex $\mathbf{v}$. We also note that the vertex splitting (the reverse process of the vertex merging) can be considered in order to represent user joins. However, graphs after the vertex splitting also belong to the class of DL$^\#$ graphs, so we only focus on the vertex merging hereafter.

In DL$^\#$ graphs, user joins are represented by DLTs or vertex splitting, and user leaves are represented by vertex merging. It has to be mentioned that Zhang and Liu [11] introduce DL$^+$ graphs not only to handle user joins and user leaves, but also to maintain the network balance. More precisely, conditions on the number of merged logical vertices and the choice of sibling vertices to be merged must be considered. DL$^\#$ graphs, on the other hand, are obtained by relaxing the conditions of the vertex merging as (3) above, with an aim to cover more general and practical cases.

Example 6. Let an initial graph G_0 be the complete graph K_4. Fig. 6 is an example of generating a DL$^\#$ graph.

Suppose that DLT is applied w.r.t. $\mathbf{r} = 1$ in G_0. Then we obtain DL graph G_1 with vertices 0,2,3,01,21,31. Since 21 and 31 are siblings, we apply vertex merging to 21 and 31, and the resulting graph is a DL$^\#$ graph $G_1^\#$ with (physical) vertices 0,2,3,01,21, where $||0|| = ||2|| = ||3|| = ||01|| = 1$ and $||21|| = 2$.

Restricting vertex merging only for sibling vertices gives us important and interesting remarks, which will be key points on the analysis of DL$^\#$ graphs.

Remark 2. Let $G^\#$ be a base-d DL$^\#$ graph. Then we have the followings.

1) *Invariance on out-neighbour sets.*

   Suppose that vertex merging is applied to k sibling vertices $a_1\mathbf{r}, a_2\mathbf{r}, \ldots, a_k\mathbf{r}$ in $G^\#$, and new vertex $b\mathbf{r}$ is generated in the resulting DL$^\#$ graph $H^\#$. Observe that

   $$N_{H^\#}^{OUT}(b\mathbf{r}) = \bigcup\nolimits_{i=1}^{k} N_{G^\#}^{OUT}(a_i\mathbf{r}) = N_{G^\#}^{OUT}(a_1\mathbf{r})$$

   since sibling vertices have the same out-neighbour set.

2) *The maximum number of logical vertices.*

   We can always assume that for any vertex $\mathbf{v}$ in $G^\# ||\mathbf{v}|| < d$. Indeed, if there exists a vertex $\mathbf{v}$ such that $||\mathbf{v}|| = d$, then we can substitute $\mathbf{v}$ with its longest proper suffix $\mathbf{v}'$ (i.e., $\mathbf{v}' \prec_s \mathbf{v}$ and $|\mathbf{v}'| = |\mathbf{v}| - 1$) and for $\mathbf{v}'$, we have $||\mathbf{v}'|| = 1$.

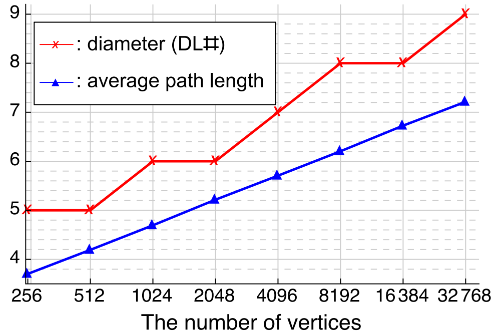

Fig. 7. The diameters and the average path lengths for DL$^{\#}$ graphs.

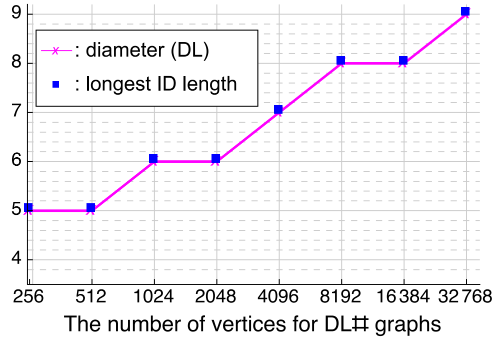

Fig. 8. The diameters for corresponding DL graphs of Fig. 7.

3) *The existence of sibling vertices.*
   There can exist a vertex with no sibling vertices. Indeed, such a vertex exists if and only if other sibling vertices are used as responsible vertices of DLTs. It implies that vertices $\mathbf{u}$ satisfying $|\mathbf{u}| = \ell^{G^{\#}}_{\max}$ has a sibling vertex.

7.2 The Diameter of a DL$^{\#}$ Graph

Recall that the diameter is an important value to measure the effectiveness of network topologies. In this section, we observe the diameter of a DL$^{\#}$ graph $G^{\#}$ using its *corresponding DL graph* G, which is obtained from $G^{\#}$ by replacing all physical vertices with their logical vertices. The corresponding DL graph is uniquely determined by 2) of Remark 2.

From the definitions on DL graphs and DL$^{\#}$ graphs, it is straightforward to obtain the followings regarding relationships between a DL$^{\#}$ graph and its corresponding DL graph.

Remark 3. Let $G^{\#}$ be a DL$^{\#}$ graph and G be its corresponding DL graph. Then

1) $\ell^{G}_{\max} = \ell^{G^{\#}}_{\max}$; and
2) The routing scheme of $G^{\#}$ is based on the routing scheme of G.

The main theorem in this section is the following. It implies that we only have to focus on the diameters of DL graphs when an initial graph is a complete graph.

Theorem 13. *Let an initial graph G_0 be the complete graph K_{d+1} with $d \geq 2$, and $G = (V, E)$ be a DL graph obtained from G_0. Then for any DL$^{\#}$ graph $G^{\#} = (V_{\#}, E_{\#})$ whose corresponding DL graph is G, we have*

$$D(G^{\#}) = D(G).$$

Proof. Throughout this proof, define a map $f : V \longrightarrow V_{\#}$ so that for vertex $\mathbf{v} \in V$, $f(\mathbf{v})$ is the physical vertex in $G^{\#}$ having $\mathbf{v}$ as its logical vertex. The map f is surjective; so $V_{\#} = f(V) = \{f(\mathbf{v}) \ : \ \mathbf{v} \in V\}$.

The statement is trivial when $G^{\#} = G_0$ since the corresponding DL graph G of G_0 is G_0 itself. So we hereafter assume that $G^{\#} \neq G_0$. Observe that this assumption gives us $\ell^{G^{\#}}_{\max} \geq 2$.

We first show that $D(G^{\#}) \leq D(G)$. For arbitrary vertices $\mathbf{s}, \mathbf{t} \in V$, let $\rho : \mathbf{s} \to \mathbf{m}_1 \to \cdots \to \mathbf{t}$ be any path from $\mathbf{s}$ to $\mathbf{t}$ in G. Then path $f(\rho) : f(\mathbf{s}) \to f(\mathbf{m}_1) \to \cdots \to f(\mathbf{t})$, which is obtained by replacing all vertices $\mathbf{m}$ in ρ with $f(\mathbf{m})$, is a path from $f(\mathbf{s})$ to $f(\mathbf{t})$ in $G^{\#}$. Assuming ρ to be a shortest path from $\mathbf{s}$ to $\mathbf{t}$, we have $d_{G^{\#}}(f(\mathbf{s}), f(\mathbf{t})) \leq d_G(\mathbf{s}, \mathbf{t})$. Therefore, together with the fact that f is surjective, we can conclude that $D(G^{\#}) \leq D(G)$.

We next show that $D(G) \leq D(G^{\#})$. Take an arbitrary vertex $\mathbf{v}^{\#} \in V_{\#}$ such that $|\mathbf{v}^{\#}| = \ell^{G^{\#}}_{\max}$. From 3) in Remark 2, there exists a (physical) sibling vertex $\mathbf{u}^{\#}$ of $\mathbf{v}^{\#}$ in $G^{\#}$.

Now let $u^{\#}_1$ and $u^{\#}_2$ be the first and the second symbols of $\mathbf{u}^{\#}$, respectively. Since the initial graph G_0 is complete graph K_{d+1}, there exists a cycle $\gamma : u^{\#}_1 \to u^{\#}_2 \to u^{\#}_1$ in G_0. Following Lemma 10 and the similar argument in the proof of Theorem 8, we can find a vertex $\mathbf{w}$ in V which is a suffix of $(u^{\#}_1 u^{\#}_2)^k$ (the word generated by concatenating $u^{\#}_1 u^{\#}_2$ k times) for some positive integer k. As $u^{\#}_1 \neq u^{\#}_2$ and $u^{\#}_2$ is the second symbol of $\mathbf{v}^{\#}$ as well, for any logical vertex $\mathbf{x}$ of $\mathbf{v}^{\#}$, $\mathbf{w} \diamond \mathbf{x} = \epsilon$ holds. Observing that

$$d_{G^{\#}}(f(\mathbf{w}), \mathbf{v}^{\#}) = \min_{\mathbf{x}} d_G(\mathbf{w}, \mathbf{x})$$

we have

$$\begin{aligned} D(G^{\#}) &\geq d_{G^{\#}}(f(\mathbf{w}), \mathbf{v}^{\#}) = \min_{\mathbf{x}} d_G(\mathbf{w}, \mathbf{x}) && (4) \\ &= |\mathbf{x}| \\ &= D(G), && (5) \end{aligned}$$

where the equality at (3) is from Lemma 7, and the equality at (4) comes from the fact that $|\mathbf{x}| = |\mathbf{v}^{\#}| = \ell^{G}_{\max}$ and Corollary 12. □

Note that the first half of the proof of Theorem 13 shows that $D(G^{\#}) \leq D(G)$ holds, whatever initial graphs are given. We believe that further analysis provides the concrete value of $D(G^{\#})$ for general case.

8 Simulations

In this section, we will show some simulation results to support our theoretical contributions. We focus on

1) the diameters and the average path lengths; and
2) the betweenness centrality of a vertex.

The simulations are conducted using Gephi, an open source software for analyzing networks [24]. Throughout the simulations, the initial graph is set to be the complete graph K_5.

8.1 The Diameters and the Average Path Lengths

Recall that the diameter of a graph is the length of the shortest path between distinct vertices. Similarly, the average

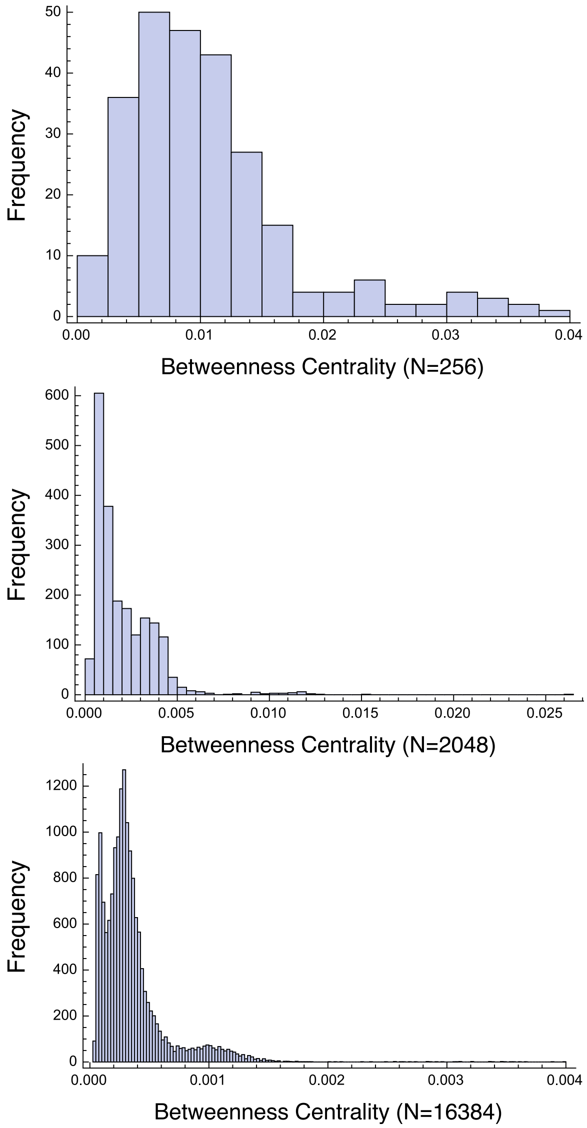

Fig. 9. The betweenness centralities of DL$^{\#}$ graphs.

path length is the average length of the shortest path between two distinct vertices.

Fig. 7 shows the diameters and the average path lengths for DL# graphs with various number of vertices. From this figure, we can observe that the diameters and the average path lengths are reasonably small. Indeed, by comparing Fig. 8a in [11], we can confirm that they are quite similar to the diameters and the average path lengths of DL^{+} graphs (about 3 percent better for the average path lengths). Thus, the relaxation of the conditions on vertices does not at least deteriorate the performance of path lengths. In addition, we can say that they are much smaller than CAN, Koore and FissionE; that is, other DHT-based networks.

Fig. 8 shows the diameters of corresponding DL graphs for DL$^{\#}$ graphs in Fig. 7, together with the longest ID lengths amongst vertices. Observe that Figs. 7 and 8 support Corollary 12 and Theorem 13. Indeed, we can confirm that the diameter of a DL$^{\#}$ graph is equal to the diameter of its corresponding DL graph. In addition, the diameter is equal to the length of the longest ID amongst all vertices.

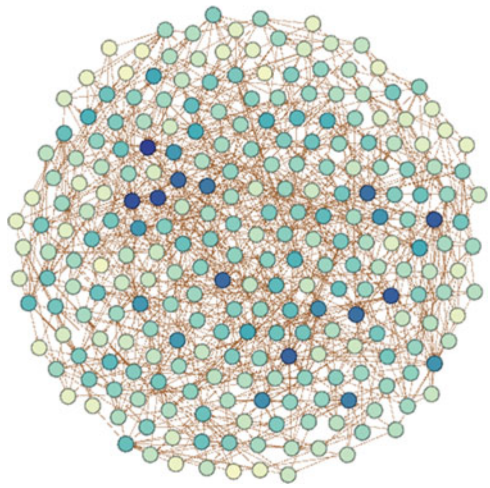

Fig. 10. DL$^{\#}$ graph of 256 vertices generated from 100 DLTs followed by vertex merging representing 49 user leaves.

TABLE 1
Average, Variance and Maximum Value for Each Graph

N	Average	Variance	Maximum value
256	1.0644×10^{-2}	5.0970×10^{-5}	0.039694
2,048	2.0607×10^{-3}	3.0833×10^{-6}	0.026475
16,384	3.4957×10^{-4}	8.3469×10^{-8}	0.003999

8.2 The Betweenness Centrality

Given a graph $G = (V, E)$, the *betweenness centrality* of vertex $\mathbf{v}$ implies the ratio of $\mathbf{v}$ being used within a shortest path between some vertices. More precisely, the betweenness centrality $c(\mathbf{v})$ for vertex $\mathbf{v}$ is defined to be

$$c(\mathbf{v}) = \sum_{\mathbf{s} \neq \mathbf{v} \neq \mathbf{t}} \frac{\sigma(\mathbf{s}, \mathbf{t} | \mathbf{v})}{\sigma(\mathbf{s}, \mathbf{t})},$$

where $\sigma(\mathbf{s}, \mathbf{t})$ is the number of shortest paths from $\mathbf{s}$ to $\mathbf{t}$ and $\sigma(\mathbf{s}, \mathbf{t}|\mathbf{v})$ is the number of those paths that pass through $\mathbf{v}$.

The betweenness centrality of a vertex gives us the importance of the vertex in the graph. Also, the distribution of the betweenness centralities is used to determine the uniformity of a network. See, for example, [25], [26] for further details.

Fig. 9 shows the histograms of normalized betweenness centralities of graphs with 256, 2048 and 16384 vertices, respectively, where the centralities for a graph of N vertices are normalized by multiplying $\frac{1}{(N-1)(N-2)}$. Fig. 10 depicts the DL$^{\#}$ graph of 256 vertices for which the vertices of higher betweenness centralities are painted in darker color. The average, the variance and the maximum value for each graph are presented in Table 1. From these results, we can observe that the the average and the variance of the betweenness centralities are quite small at each graph, especially for bigger graphs. It indicates that the importance of each vertex is well balanced. It also implies that shortest paths in the network can be maintained even though some vertices fail, which is a desired property as a network topology.

9 Conclusion

In this paper, we proposed DL$^{\#}$ graphs, an extension of DL graphs, as a candidate of an overlay for P2P networks. We then presented theoretical results of DL graphs and DL$^{\#}$ graphs, especially diameters and in-degrees of those graphs. More precisely, we showed the values of in-degrees for DL graph which are more specific than existing result, together with some simulation results. We also presented a tight lower bound of the diameter of DL graphs, which gives us the exact diameter when an initial graph is a complete graph. We further proved that the diameter of a certain DL$^{\#}$ graph can be obtained from its corresponding DL graph. To support theoretical results and efficiency of DL$^{\#}$ graphs, we also presented simulation results on the diameter, the average of path length, and the betweenness centrality.

As a future work, we aim to find some properties for a DL$^{\#}$ graph to have a short average path length, together with an algorithm to obtain such a DL$^{\#}$ graph. We also analyze the complexities on user joins and user leaves for DL$^{\#}$ -based networks .

Acknowledgments

The authors would like to thank Gergely Huszák and Shuhei Kaneko for their great help in simulations on this research. The authors also would like to thank anonymous reviewers for their valuable comments to revise the paper. This paper was presented in part at International Symposium on Information Theory and Its Applications (ISITA) 2014, Melbourne, Australia, Oct. 26-29, 2014 [1].

References

[1] A. Manada and H. Morita, "On some properties of distributed line graphs," in *Proc. Int. Symp. Inf. Theory Appl.*, Oct. 2014, pp. 373–377.
[2] R. Steinmetz and K. Wehrle, "Peer-to-Peer systems and applications," *Lecture Notes in Computer Science*. Berlin, Germany: Springer, 2005, vol. 3485.
[3] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in *Proc. ACM SIGCOM Conf.*, 2001, pp. 161–172.
[4] I. Stoica, et al., "Chord: A scalable peer-to-peer lookup protocol for internet applications," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 17–32, Feb. 2003.
[5] P. Maymounkov and D. Mazières, "Kademlia: A peer-to-peer information system based on the XOR metric," *First Int. Workshop Peer-to-Peer Syst.*, pp. 53–65, 2002.
[6] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems," in *Proc. FIP/ACM Int. Conf. Distrib. Syst. Platforms Open Distrib. Process.*, 2001, pp. 329–350.
[7] B. Zhao, L. Huang, J. Stribling, S. Rhea, A. D. Joseph, and J. Kubiatowicz, "Tapestry: A resilient global-scale overlay for service deployment," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 1, pp. 41–53, Jan. 2004.
[8] S. Ratnasamy, I. Stoica, and S. Shenker, "Routing algorithms for DHTs: Some open questions," *First Int. Workshop Peer-to-Peer Syst.*, pp. 45–52, 2002.
[9] J. Xu, A. Kumar, and X. Yu, "On the fundamental tradeoffs between routing table size and network diameter in peer-to-peer networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 1, pp. 151–163, Jan. 2004.
[10] Y. Zhang, L. Liu, D. Li, and X. Lu, "Distributed line graphs: A universal framework for building DHTs based on arbitrary constant-degree graphs," in *Proc. 28th Int. Conf. Distrib. Comput. Syst.*, Jun. 2008, pp. 152–159.
[11] Y. Zhang and L. Liu, "Distributed line graphs: A universal technique for designing DHTs based on arbitrary regular graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 9, pp. 1556–1569, Sep. 2012.
[12] M. A. Fiol, J. L. A. Yebra, and I. A. De Miquel, "Line digraph iterations and the (d, k) digraph problem," *IEEE Trans. Comput.*, vol. C-33, no. 5, pp. 400–403, May 1984.
[13] M. A. Fiol and A. S. Llado, "The partial line digraph technique in the design of large interconnection networks," *IEEE Trans. Comput.*, vol. 41, no. 7, pp. 848–857, Jul. 1992.
[14] D. Malkhi, M. Naor, and D. Ratajczak, "Viceroy: A scalable and dynamic emulation of the butterfly," in *Proc. 21st Annu. Symp. Principles Distrib. Comput.*, 2002, pp. 183–192.
[15] M. Kaashoek and D. Karger, "Koorde: A simple degree-optimal distributed hash table," *Lecture Notes Comput. Sci.*, vol. 2735, pp. 98–107, 2003.
[16] Y. Zhang, X. Lu, and D. Li, "SKY: Efficient peer-to-peer networks based on distributed Kautz graphs," *Sci. China Ser. F: Info. Sci.*, vol. 52, no. 4, pp. 588–601, 2009.
[17] D. Li and X. Lu, "A novel constant degree and constant congestion DHT scheme for peer-to-peer networks," *Sci. China Ser. F: Info. Sci.*, vol. 48, no. 4, pp. 421–436, 2005.
[18] D. Guo, J. Wu, H. Chen, and X. Luo, "Moore: An extendable peer-to-peer network based on incomplete Kautz digraph with constant degree," in *Proc. 26th Int. Conf. Comp. Commun.*, May 2007, pp. 821–829.
[19] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms and Applications*. 2nd ed. Berlin, Germany: Springer, 2009.
[20] D. B. West, *Introduction to Graph Theory (second edition)*. Upper Saddle River, NJ, USA: Prentice Hall, 2000.
[21] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge University Press, 1995.
[22] M. Lothaire, *Combinatorics on Words*. Cambridge, U.K.: Cambridge University Press, 1997.
[23] X. Shen, H. Yu, J. Buford, and M. Akon, *Handbook of Peer-to-Peer Networking*. Berlin, Germany: Springer, 2009.
[24] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Proc. 3rd Int. AAAI Conf. Weblogs Soc. Media*, 2009, pp. 361–362.
[25] C. Jiang, Y. Chen, Y. Ren, and K. J. R. Liu, "Maximizing network capacity with optimal source selection: A network science perspective," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 938–942, Jul. 2015.
[26] N. Kourtellis, G. D. F. Morales, and F. Bonchi, "Scalable online betweenness centrality in evolving graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 9, pp. 2494–2506, Sep. 2015.

Akiko Manada received the MSci degree from Tsuda College, Tokyo, Japan, in 2004, and the PhD degree in mathematics from Queen's University, Kingston, Canada, in 2009. She then worked in Claude Shannon Institute, University College Dublin, Ireland, as a postdoctoral fellow from 2009 through 2011. Since February 2012, she has been an assistant professor with the University of Electro-Communications, Tokyo, Japan. Her research interests include discrete mathematics (especially in graph theory) and its applications towards coding theory. She is a member of the IEEE.

Hiroyoshi Morita received the BE, ME, and DE degrees from Osaka University, Osaka, Japan, in 1978, 1980 and 1983, respectively. In 1983, he joined Toyohashi University of Technology, Aichi, Japan as a research associate in the School of Production System Engineering. In 1990, he joined the University of Electro-Communications (UEC), Tokyo, Japan, first an assistant professor in the Department of Computer Science and Information Mathematics, where from 1992, he was an associate professor. He was then with the Graduate School of Information Systems, University of Electro-Communications since 1995, where from 2005, he is a professor. Since 2016, he has been with Graduate School of Informatics and Engineering, University of Electro-Communications. He was a visiting fellow in the Institute of Experimental Mathematics, University of Essen, Essen, Germany during 1993–1994. His research interests include the combinatorial theory, information theory, and coding theory, with applications to the digital communication systems. He is a senior member of the IEEE.