On the Expandability and Fidelity of Distributed Line Graphs

Yiming Zhang[1], Tian Tian[1], Dongsheng Li[1], Jiaxin Li[1], Huiba Li[1], Ping Zhong[2], Ling Liu[3]
[1]NUDT, [2]Central South University, [3]Georgia Institute of Technology
sdiris@gmail.com

***Abstract*—The design of maintenance mechanisms of distributed hash tables (DHTs) is usually specific to their initial graphs, and thus it is complicated and error-prone. Zhang and Liu propose in [4] the "distributed line graphs" (DLG) mechanism, a universal technique for designing DHTs based on arbitrary regular graphs while preserving the main features of the initial graphs. However, two important properties of DLG, the expandability and fidelity, have not been studied with detailed explanations or analysis. In this paper, we study the above properties of DLG transformations, and prove that (i) the DLG transformations are incrementally expandable, and (ii) DLG transformations from G_i to G_{i+1} keep fidelity.**

Keywords—DLG Transformations; Expandability; Fidelity; Kautz Graphs; Line Graphs

I. INTRODUCTION

Distributed hash tables (DHTs) [1,2] have been widely used in providing cloud services. The maintenance mechanisms of DHTs are usually designed based on a specific type of *regular graphs*, where all nodes have the same number of edges. These mechanisms usually require complicated (and error-prone) design [1~3], tightly coupled with the static graphs on which they are based.

Zhang and Liu propose the "*distributed line graphs*" (DLG) [4], a universal technique for designing DHTs based on arbitrary regular graphs. The main features of the initial graphs are preserved. They prove that in a DLG-enabled, *N*-node DHT, the out-degree is *d*, the in-degree is between 1 and 2*d*, and the diameter is less than $2(\log_d N - \log_d N_0 + D_0 + 1)$, where d, D_0 and N_0 represent the degree, diameter and number of nodes of the initial graph, respectively. This diameter reaches the lower bound $\Omega(\log_d N)$ [2] of constant-degree DHTs.

DLG is inspired by the *line graph* (LG) iteration [5], which has been proposed as a universal technique for designing multiprocessor networks. These techniques need global knowledge of the network topology and centralized control, which are not practical in large-scale distributed networks. Although DLG is a novel technique for designing DHTs on different kinds of initial graphs, two important properties of DLG, the *expandability* and *fidelity*, have not yet been well studied.

Expandability is equal to the number of nodes that will be added to the network after a transformation. When the *expandability* is 1 we say that the transformation is *incrementally expandable*. Clearly DLG must be incrementally expandable to be used in DHTs since any node may join/leave at any time.

Fidelity implies that the degree of similarity between the two graphs before and after a DLG transformation. If the properties (like node degree, diameter, routing, etc.) are preserved after any number of DLG transformations, we say that the transformations keep *fidelity*; otherwise we say that they do not keep fidelity.

II. PRELIMINARLIES

A. Concepts

The network *topology* [4,5] is modeled by a *graph* $G = (V,E)$ whose *vertices* $V = V(G)$ and *edges* $E = E(G)$ represent, respectively, the processing elements (nodes in the network) and the links between them. In this paper only unidirectional links and directed graphs [5] are considered. For simplicity, in the following we use the same term "node" to refer both to a vertex in a graph and to a node in a network [6]. If $a = [x, y]$ is an edge from node *x* to node *y*, we say that *x* (or *a*) is *adjacent to y* and *y* is an *out-neighbor* of *x*, and that *y* (or *a*) is *adjacent from x* and *x* is an *in-neighbor* of *y* [5]. We also say that *a* is an *out-edge* (resp. *in-edge*) of *x* (resp. *y*). The number of nodes in graph *G* is called the *order* of *G*. Let $\Gamma_G^-(x)$ and $\Gamma_G^+(x)$ denote the nodes adjacent to and from *x*. Their cardinalities are the *in-degree*, $\delta_G^-(x) = \left|\Gamma_G^-(x)\right|$, and the *out-degree*, $\delta_G^+(x) = \left|\Gamma_G^+(x)\right|$, of *x* [4]. The degree of *x* is the sum of its out-degree and in-degree. The graph *G* is *d-out-regular* (resp. *d-in-regular*) if $\delta_G^+(x) = d$ (resp. $\delta_G^-(x) = d$) for all $x \in G$. *G* is *d-regular* [4] if it is *d-out-regular* and *d-in-regular*. The *distance* from *x* to *y*, $d_G(x, y)$, is the length of a shortest path from *x* to *y*. The *diameter* of *G*, denoted as $D(G)$, is the largest distance over all the pairs of nodes [5].

B. Line Graphs

Let the initial graph G_0 (i.e., the network topology at the beginning) be a *d*-regular graph. In the line graph [5] of G_i, denoted as $G_{i+1} = L(G_i)$, each node represents an edge of G_i, i.e., $V(G_{i+1})=\{uv\,|\,[u,v]\in E(G_i)\}$; and a node uv is adjacent to a node wz iff $v=w$, i.e., $[uv,wz]\in E(G_{i+1})$ when $[u,v]$ is adjacent to $[w,z]$ in G_i.

Line graphs can be defined iteratively as $G_i = L(G_{i-1}) = \ldots = L^i(G_0)$. The line graph (shortly LG) iterations simply refer to the process of generating graphs by iteratively applying the L operator, where G_i is said to be *derived from* G_0. Clearly graphs G_i are d-regular. LG iterations preserve the main feature of the initial graph, such as degree, diameter and routing algorithm. Many graphs can be defined iteratively by LG iterations. E.g., Kautz graph $K(d,D) = L(K(d, D\text{-}1)) = \ldots = L^{D\text{-}1}(K(d,1))$, and de Bruijn graph $B(d,D) = L(B(d, D\text{-}1)) = \ldots = L^{D\text{-}1}(B(d,1))$.

However, the series of graphs generated by LG iterations are **not** *incrementally expandable* [5], i.e., they cannot accommodate arbitrary number of nodes. Many variations were proposed to address this problem, e.g., PLG (partial line graphs) [6], necklaces [7], and factorization [8]. For example, to support arbitrary number of nodes, Fiol and Llado [6] proposed the PLG iteration. Let $G=(V,E)$ be a graph. Let $E'\subseteq E$ be a subset of edges which are adjacent to all nodes of G, that is, $\{v\,|\,[u,v]\in E'\}=V$. A graph $G'=PL(G,E')$ is said to be a *partial line* (*PL*) *graph* of G if its nodes represent the edges of E', that is,

$$V(G')=\{uv\,|\,[u,v]\in E'\}, \tag{1}$$

and a node uv is adjacent to the nodes $v'w$, for each $w\in\Gamma_G^+(v)$, where

$$v'=v, \text{if } [v,w]\in E', \text{ or} \tag{2α}$$

$$v'=\text{any other node of } \Gamma_G^-(w), \text{otherwise}. \tag{2β}$$

Clearly PLG is *incrementally expandable*. However, as discussed in Section III.A, PLG needs global topology information, i.e., all the nodes and edges of the old graph to generate a new graph, and thus PLG is not suitable for P2P scenarios.

C. Distributed Line Graphs

Let $u = u_1u_2\ldots u_m$, $v = v_1v_2\ldots v_n$, $m\geq n$. The conjunction operator [4] $\circ$ is defined as Formula (3):

$$u\circ v=u_{m-n+1}v_1v_2\ldots v_n=u_{m-n+1}v\,. \tag{3}$$

The *distributed line graphs* are defined as follows.

Definition 1 [4]. Let the initial graph $G_0=(V,E)$ be a d-regular graph. A series of graphs $G_{i+1}=DL(G_i,v^{(i)})$ where node $v^{(i)}\in V(G_i)$ satisfies

$$\forall u\in\Gamma_{G_i}^-(v^{(i)})\cup\Gamma_{G_i}^+(v^{(i)}),\ \left|v^{(i)}\right|\leq|u|, \tag{4α}$$

is called a family of distributed line (DL) graphs derived from the initial graph G_0, if the following conditions hold.

$$V(G_{i+1})=V(G_i)-\{v^{(i)}\}+\{u\circ v^{(i)}\,\big|\,u\in\Gamma_{G_i}^-(v^{(i)})\} \tag{4b}$$

$$\begin{aligned}E(G_{i+1})=E(G_i)&-\{[x,v^{(i)}]\,\big|\,x\in\Gamma_{G_i}^-(v^{(i)})\}-\{[v^{(i)},y]\,\big|\,y\in\Gamma_{G_i}^+(v^{(i)})\}+\\ &\{[u,u\circ v^{(i)}]\,\big|\,u\in\Gamma_{G_i}^-(v^{(i)})\}+\{[u\circ v^{(i)},w]\,\big|\,u\in\Gamma_{G_i}^-(v^{(i)}),w\in\Gamma_{G_i}^+(v^{(i)})\}\end{aligned} \tag{4χ}$$

The above transformation is called distributed line (DL) iteration, and node $v^{(i)}$ is called responsible node. As discussed in [4], (4a) restricts that the identifier length of $v^{(i)}$ is no greater than any of its direct neighbors; (4b) gives the identifiers of the new nodes generated by old edges; and (4c) shows how to generate new edges. Fig. 1(a) ~ (c) show a DL iteration $G_1=DL(G_0,v^{(0)})$ with $v^{(0)}=1$. Fig. 1(d) shows another example with $v^{(1)}=4$.

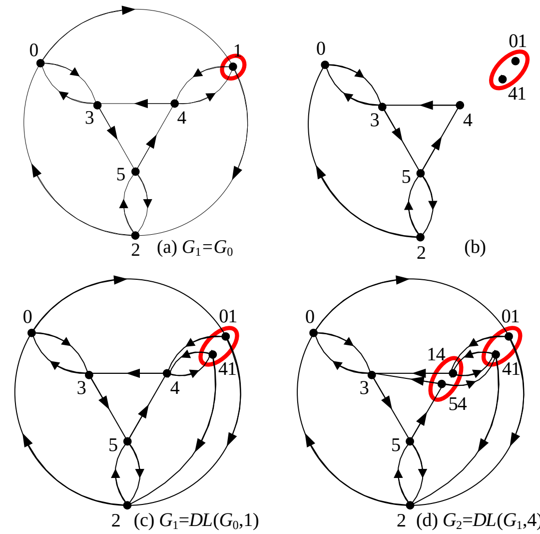

Fig. 1. Examples of DL iterations.

After each DL iteration there would be $d-1$ more nodes in the new graph. Clearly when d = 1 the DL iterations are incrementally expandable. But a series of DL graphs with base $d>2$ are not incrementally expandable and thus cannot be directly applied to building DHTs. E.g., suppose that G_0 is a 3-regular graph $G_0=K(3,2)$. After the iteration $G_1=DL(G_0,1)$, there will be two more nodes in G_1 than in G_0. A simple approach is proposed to address this problem: the newly generated nodes are divided into *two* groups together with their edges, each group merging to one node and resulting in a DL+ graph [4] with one more node than the old graph. The DL iteration and the grouping operation are called together “DLG transformations” [4].

III. Expandability and Fidelity of DLG

A. DLG Vs. LG/PLG

Although DLG is inspired by the LG/PLG (partial line graph) iterations, they have little in common, and DLG is much more than a distributed version of PLG. Consequently, they have totally different routing algorithms, maintenance mechanisms, and properties.

A sub-branch of the DLG technique, called optimal DL iterations, could be viewed as a distributed version of PLG. It can be proved (in Theorem 2 in the next Section) that the *optimal* DL iterations are equivalent to PLG iterations. That is, if both started with a common regular graph G_0, any graph obtained by a series of *optimal* DL iterations has its isomorphic counterpart obtained by a series of PLG iterations, and vice versa. The core difference between the *optimal* DL iterations and the normal DLG is that, in *optimal* DL iterations the candidate responsible nodes are required to have globally shortest identifiers, while in normal DL iterations they are only required to have locally shortest identifiers (see (4a)).

Let us compare the result graphs of DLG and PLG. Although they have the same results at the first few iterations, they will have different results as the network evolves, since DLG does not try to strictly emulate interconnection networks. For example, assume G_5 is shown in Fig. 2(a). Now, node 01 satisfies the requirements for being a responsible node. Let $G_6 = DL(G_5, 01)$, as shown in the above Fig. 2(b), then in G_6 there would be two nodes with length 3 (nodes 201 and 301), and a node with length 1 (node 5). Clearly G_6 can never be achieved by any series of PLG iterations from G_0, since every PLG iteration always changes the node with the *globally* shortest identifier.

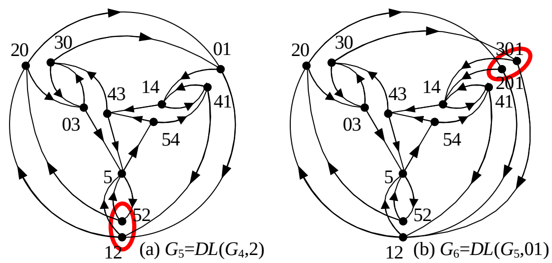

Fig. 2. DL iteration that never happens in PLG iterations.

More important, the inside processing of DLG is totally different from that of PLG. Consider the DL iteration $G_1 = DL(G_0, 1)$ shown in Fig. 1(a), (b) and (c). This iteration turns node 1's two in-edges [0,1] and [4,1] into two new nodes 01 and 41, respectively, and assigns node 1's in-neighbors to the two new nodes: in-neighbor 0 is assigned to 01 and 4 to 41. The out-neighbors of the new nodes are the same as those of node 1: nodes 2 and 4. Here, the number of involved edges is O(1).

B. DLG Expendability

This subsection presents the following Theorems to show DLG is incrementally expandable.

Theorem 1. Distributed line graph transformations are incrementally expendable.

Proof. For clarity, we first prove that in a distributed line graph, the label of a node must not be a suffix of the label of another. Obviously it holds initially for G_0. Suppose it holds for G_i with $i \ge 0$.

Let $v = v_1 v_2 ... v_m$. By the definition of DL graphs, the new nodes in $G_{i+1} = DL(G_i, v)$ are in the form of $\alpha v_1 v_2 ... v_m$ with length $m+1$, and there is no node v in G_{i+1}. Let y be a new node in G_{i+1}. Let u be an old node other than v both in G_i and G_{i+1}. It is easy to see that in G_{i+1}: node y is not a suffix of other new nodes; node u is not a suffix of other old nodes; and neither node y nor u is a suffix of the other. Therefore, it holds for G_{i+1} and thus it holds for all DL transformations.

Define $|x|$ as the identifier length of node x. Suppose that $x = x_1 ... x_m$. If there is some node $y \in \Gamma_G^-(x)$ satisfying $|y| < |x|$, by Theorem 1 in [4], we have $y = x_1 ... x_{m-1}$. If $\delta_G^-(x) > 1$, node x must have another in-neighbor of the form $y' = \alpha_1 ... \alpha_r x_1 ... x_{m-1}$ with $1 \le r \le 2$. Then y will be a suffix of y', which is contradictory to the above conclusion. So if there is some $y \in \Gamma_G^-(x)$ satisfying $|y| \ge |x|$, then for any $w \in \Gamma_G^-(x)$, we have $|w| \ge |x|$. Therefore, a node has exactly 1 in-neighbor if and only if its label length is 1-letter longer than that of the in-neighbor, and thus it cannot be the responsible node in a DL iteration. Then according to (4b), after each DL iteration there would be d new nodes in the new graph, and according to Section III.A the d new nodes are divided into 2 groups merging into 2 nodes, one for the deleted node and the other for the newly added node. Therefore, the DLG transformations are incrementally expandable.

C. DLG Fidelity

Next we use the definitions [4] of *optimal DL iteration* and ϕ *mapping* to bridge the DL graphs with traditional line graphs, in order to prove the following Theorem 2 for the property of DLG fidelity. Let the initial graph $G_0 = (V, E)$ be a d-regular graph. For a series of DL iterations $G_{i+1} = DL(G_i, v^{(i)})$ is said to be *optimal*, if the constraint for $v^{(i)}$ in Definition 1 is strengthened to

$$\forall u \in V(G_i),\ \left|v^{(i)}\right| \le |u| . \tag{5}$$

Clearly by (5) the candidate responsible nodes are required to have the globally shortest identifiers, while by (4a) they are only required to have the locally shortest identifiers compared with their neighbors.

Let $u = u_1 u_2 ... u_m$. Define

$$\phi(u,i) = \begin{cases} u_{m-i+1} ... u_m, & \text{if } i \le m \\ u, & \text{otherwise.} \end{cases} \tag{6α}$$

Define ϕ mapping of graph G as $G' = \phi(G, m)$, where

$$V(G') = \{\phi(v,m) \mid v \in V(G)\} , \tag{6β}$$

$$E(G') = \{[\phi(u,m), \phi(v,m)] \mid u \in V(G), v \in V(G), [u,v] \in E(G)\} . \tag{6χ}$$

Theorem 2. DLG transformations keep fidelity.

Proof. Let graph G be a DL+ graph with base d. From [4], for all $x \in V(G)$, the out-degree is d, the in-degree is between 1 and $2d$, and the average degree is $2d$. The diameter satisfies $D(G) < 2(\log_d N - \log_d N_0 + D_0 + 1)$. So the properties of degree, diameter and routing are kept after DLG transformations. Next we will show a series of DLG graphs are *consanguineous*. Let the order of G_0 be N_0. By the definition of *optimal DL iterations*, for each $v^{(i)} \in G_i$ we have

$$v^{(i)} \in G_0 \text{ where } i < N_0 . \tag{7}$$

Suppose that $E^{(0)} \subset E(G_0)$ satisfies

$$\{v \mid [u,v] \in E^{(0)}\} = V(G_0) \text{, and} \tag{8α}$$

$$|E^{(0)}| = N_0 . \tag{8β}$$

Then the PL graph $G_0^* = PL(G_0, E^{(0)})$ satisfies

$$\phi(G_0^*, 1) = G_0 . \tag{9}$$

Suppose that $E^{(i+1)} \subseteq E(G_0)$ with $0 \le i < N_0$ satisfies

$$E^{(i+1)} = E^{(i)} + \{[u, v^{(i)}] \mid u \in \Gamma_{G_0}^-(v^{(i)})\} . \tag{10}$$

Then, PL graphs $G_i^* = PL(G_0, E^{(i)})$ are isomorphic with DL graphs $G_i = DL(G_{i-1}, v^{(i-1)})$ generated by *optimal* DL iterations, and they satisfy

$$\phi(G_i^*,1) = \phi(G_i,1) . \tag{11}$$

with $0 < i \le N_0$. On the other hand, clearly we have

$$\phi(G_i^*,1) = G_0 . \tag{12}$$

with $0 < i \le N_0$. Then we have

$$\phi(G_{i+1},1) = G_0 \text{ with } 0 \le i < N_0 . \tag{13}$$

Similarly, for each $x \ge 2$ and integer *i* satisfying

$$N_0 \times \sum_{j=0}^{x-2} d^j \le i < N_0 \times \sum_{j=0}^{x-1} d^j , \tag{14}$$

and we have

$$\phi(G_{i+1}, x) = L^{x-1}(G_0) . \tag{15}$$

So all nodes in $G_i^* = PL(G_0, E^{(i)})$ have the same identifier length $x+1$, and the shortest node identifier in G_{i+1} be of length *x*. Next, let G_i be a series of DL graph derived from the initial graph G_0 by DL iterations. Let the shortest node identifier in G_i be of length m_i. Let

$$G_i^* = \phi(G_i, m_i + 1) . \tag{16}$$

Clearly we have

$$G_0^* = G_0 . \tag{17}$$

For each $G_{i+1} = DL(G_i, v^{(i)})$, there are two cases:

(i) If $|v^{(i)}| = m_i$, that is, $v^{(i)}$ is of the shortest identifier in G_i, then we have

$$G_{i+1}^* = DL(G_i^*, v^{(i)}) . \tag{18}$$

Note that $v^{(i)}$ also has the shortest identifier in G_i^*.

(ii) Otherwise we have

$$G_{i+1}^* = G_i^* . \tag{19}$$

Therefore, the graphs G_i^* $(i = 0,1,...)$ are a series of DL graphs derived from the initial graph G_0 by optimal DL iterations. Then by (15) we have

$$\phi(G_i^*, m_i) = L^{m_i - 1}(G_0^*) . \tag{20}$$

On the other hand, since $G_i^* = \phi(G_i, m_i + 1)$, we have

$$\phi(G_i, m_i) = \phi(G_i^*, m_i) . \tag{21}$$

Then, by (21), (20) and (17), a series of DLG graphs are *consanguineous*. Since line graphs keep fidelity [5], DLG transformations also keep fidelity.

IV. Evaluation

In this section we demonstrate the incremental expendability of DLG by evaluating the diameter (maximum routing path length) of a DLG-enabled, Kautz graph-based DHT called DLG-Kautz (DK) [4].

We evaluate the diameter of DK, as a function of the number of nodes. The base of DK is fixed to *d* = 2. In each experiment we choose an initial node uniformly at random and the initial node sends a message which is then routed following the bit-correct routing algorithm. The initial graph is *K*(2,2) as depicted in Fig. 1(a). The result is shown in Fig. 3, where the number of nodes varies from 6 to 1500. Clearly the expandability is 1 and DLG is incrementally expandable.

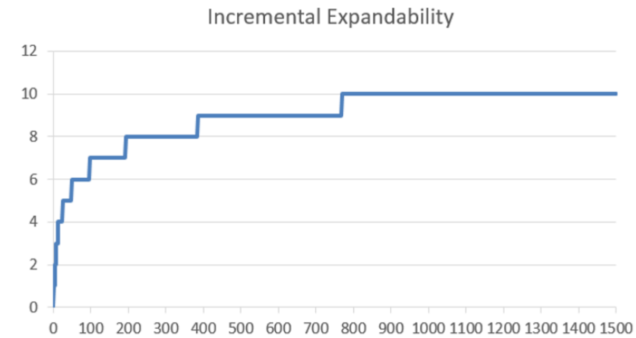

Fig. 3. Incrementally expandable evolution of topology.

V. Conclusion

This paper focuses on the two properties of expandability and fidelity in the DLG transformations [4], and proves that (i) DLG transformations are incrementally expandable; and (ii) DLG transformations keep fidelity.

Acknowledgment

This work was supported by the National Basic Research Program of China under Grant No.2014CB340303, the National Natural Science Foundation of China (NSFC) under Grant No. 61379055, 61379053 and 61222205. Ping Zhong is the corresponding author.

References

[1] I. Stoica, R. Morris, D. L. Nowell, et al, "Chord: A Scala-ble Peer-to-Peer Lookup Protocol for Internet Applica-tions," IEEE/ACM Trans. Networking, 11(1) 2003.

[2] 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-SAC, vol. 22, no. 1, pp. 151-163, Jan. 2004.

[3] D. Li, X. Lu, and J. Wu, "FissionE: A Scalable Constant Degree and Low Congestion DHT Scheme Based on Kautz Graphs," Proc. IEEE INFOCOM '05.

[4] Y. Zhang, L. Liu. "Distributed Line Graphs: A Universal Technique for Designing DHTs Based on Arbitrary Regular Graphs," IEEE TKDE, 24(9), September 2012.

[5] M. A. Fiol, J. L. A. Yebra, and I. Alegre, "Line digraph iterations and the (d, k) digraph problem," IEEE Trans. Computers, vol. 33, no. 5, pp. 400–403, May 1984.

[6] M. A. Fiol and A. S. Llado, "The Partial Line Digraph Technique in the Design of Large Interconnection Networks," IEEE Trans. Computers, vol. 41, no. 7. 1992.

[7] P.Tvrdik, "Kautz Necklaces," Technical Report 94-08, LIP ENSL, 69364 Lyon, France, Mar. 1994.

[8] P.Tvrdik, "Factoring and Scaling Kautz Digraphs," Technical Report 94-15, LIP ENSL, 69364. 1994.