SUPPLEMENTARY APPENDIX FOR OPEN WORLD LEARNING GRAPH CONVOLUTION FOR LATENCY ESTIMATION IN ROUTING NETWORKS

A. Tuning Latency Estimation to Node Level Task

In this proof, we will address how we turn the latency estimation task per trajectory (path level task) into per hop's occupancy estimation (node level task). We summarise the additional notations we used beyond the paper content in Table I:

Given any OD pair (src, dst) formulated path $p_{i,j} \in P_i$, with its latency measurement $q_{i,j} = Q(p_{i,j})$, $p_{i,j}$ consisting communication link representation $v_k \in V_i^{\mathcal{L}}$ for each hop, in line graph $G_i^{\mathcal{L}}$. With its respective representation in G_i being $e_k \in E_i$ (i.e: $\mathcal{L}^{-1}(v_k) = e_k$). We will take attributes from $Y_i(e_k)$ and transfer into corresponding $Y_i(v_k)$, according line graph transformation.

- Per communication link queue size $q(Y_i(v_k))$, denote the upper limit packet storage in v_k before being sent to its next hop.
- Per communication link joint packet size distribution $ps(Y_i(v_k))$.
- Per communication link joint packet arrival rate[1] as $\lambda(Y_i(v_k))$.
- Per communication link capacity $c(Y_i(v_k))$.
- Per communication link summed traffic $t^s(i, v_k)$

Also, denote the current occupancy in v_k to be $O(i, v_k) \in Q_i$

If there exist scenarios either *(i)* Arrival packets with their total quantity larger than $q(Y_i(v_k))$ in v_k. *(ii)* Introducing the time-integral definition of effective packet size :

$$ps(Y_i(v_k))^{rms} = \int_t^{t+\Delta} ps(Y_i(v_k))(t) * \lambda(Y_i(v_k))(t)dt \tag{1}$$

at any moment t with short period Δ. If

$$ps(Y_i(v_k))(t) > q(Y_i(v_k)) * ps(Y_i(v_k))^{rms} * (1 - O(i, v_k)) \tag{2}$$

In either (i, ii), we consider packet drop happening. In case of packet drop, re-transmission can happen that introduces redundant hops in OD pair's trajectories. This only takes up to 1% in our training set. Thus, we start from assuming there is no packet drop, to formulate the later proof.

Based on the above assumption, latency $q_{i,j}$, one can decompose the target latency into sum of augmented variables describing per-hop latency of node $v_k \in V_i^{\mathcal{L}}$, as q_{i,v_k}

$$q_{i,j} = \sum_{v_k; v_k \in p_i} q_{i,v_k} \tag{3}$$

To compute q_{i,v_k}:

$$q_{i,v_k} = \frac{q(Y_i(v_k)) * ps(Y_i(v_k))^{rms} * O(i, v_k)}{c(Y_i(v_k))} \tag{4}$$

Since all variable in Eq. 4 is known except $O(i, v_k)$. We modify the problem formulation into estimating occupancy per-link, which is a node attribute in $G_i^{\mathcal{L}}$.

B. From Service Utility Factor to Occupancy

This section discusses the queuing theory explanation of feature formulation in Sec. **??**, where we denote per node feature to be summed traffic divided by its capacity. This also provides a theoretically proof why such feature formulation helps to DGCN to learn effectively.

Consider a multi-OD pair case, where we have a set of OD pairs P'_i, with source set **src**, all shared two neighboring hops in the $G_i^{\mathcal{L}}$, former as v_k, latter as v_{k+1}.

One can consider the arrival rate $\lambda(Y_i(v_k))$ have the following inequality. At any time t:

$$\lambda(Y_i(v_k))(t) \geq \lambda(Y_i(v_{k+1}))(t) \tag{5}$$

The inequality is sharp i.f.f the packet drop happens in v_k. Otherwise, we assume:

$$\lambda(Y_i(v_k))(t) = \lambda(Y_i(v_{k+1}))(t) = \ldots = \sum_{\{j; src \in \mathbf{src}, p_{i,j} \in P'_i\}} \lambda_j(Y_i(v_{src}))(t) \tag{6}$$

[1]In particular, the rate of injection of new packets into the network, which is called the arrival rate. This is usually model as Poisson distribution

where $\lambda_j(Y_i(v_{src}))(t)$ denote the arrival rate of OD pair $p_{i,j} \in P_i$ measured at the source node. Based on 6, we can first turn Eq. 1 to be:

$$ps(Y_i(v_k))^{rms} = \int_t^{t+\Delta} \sum_{\{j; src \in \textbf{src}, p_{i,j} \in P_i'\}} \lambda_j(Y_i(v_{src}))(t) * ps_j(Y_i(v_{src}))(t)dt \tag{7}$$

Note that $ps_j(Y_i(v_{src}))(t)$ is the packet size of OD pair $p_{i,j} \in P_i$ measured at the source node. This is different with $ps(Y_i(v_{src}))(t)$ as the joint distribution of all packet from OD pair x, $p_{i,x} \in P_i$.

$$ps(Y_i(v_k))^{rms} = \int_t^{t+\Delta} \sum_{\{j; p_{i,j} \in P_i\}} T_i dt = t^s(i, v_k)\Delta \tag{8}$$

Second, we introduce service utility factor ρ. its precise definition is:

$$\rho(Y_i(v_k)) = \frac{\int_t^{t+\Delta} ps(Y_i(v_k))(t) * \lambda(Y_i(v_k))(t)dt}{c(Y_i(v_k))} \tag{9}$$

Due to $\lambda(Y_i(v_k))(t)$ follows a Poisson distribution, the probability of at any given time, there are n packet in the queue of v_k is:

$$Prob(Y_i(v_k), n) = \frac{1}{\rho(Y_i(v_k)) + \frac{\rho(Y_i(v_k))}{1-\rho(Y_i(v_k))}} * \rho(Y_i(v_k))^n \tag{10}$$

One can use $Prob(Y_i(v_k), n)$ to inference $O(i, v_k)$:

$$O(i, v_k) = \frac{Prob(Y_i(v_k), n)}{q(Y_i(v_k))} \tag{11}$$

Thus, we combine Eq. 11, and Eq. 8, into Eq. 4, with $\Delta = 1$:

$$q_{i,v_k} = (q(Y_i(v_k)) * O(i, v_k)) * \frac{ps(Y_i(v_k))^{rms}}{c(Y_i(v_k))} \tag{12}$$

Suggest a function $\mathcal{F}$:

$$q_{i,v_k} = Prob(Y_i(v_k), n) * \frac{t^s(i, v_k)}{c(Y_i(v_k))} = \mathcal{F}(\rho(Y_i(e_k))) * \frac{t^s(i, v_k)}{c(Y_i(v_k))} \tag{13}$$

While for its next hop v_{k+1}, with its own capacity $c(Y_i(v_{k+1}))$:

$$\rho(Y_i(e_k))\frac{c(Y_i(v_k))}{c(Y_i(v_{k+1}))} = \rho(Y_i(v_{k+1})) \tag{14}$$

This reflects our edge weight formulation in $G_i^{\mathcal{L}}$ between 2 links v_k, v_{k+1}, referencing Eq. **??**.

In Eq. 13, $\frac{t^s(i,v_k)}{c(Y_i(v_k))}$ is per node attributes, while F is model we proposed. Yet one can not precisely compute $\rho(Y_i(v_k))$, as either $\lambda_j(Y_i(v_k))(t)$ or $ps_j(Y_i(v_k))(t)$ for any specific $p_{i,j}$ is not computable (except at its source), nor the joint distribution $\lambda(Y_i(v_k))(t), ps(Y_i(v_k))(t)$ of all $p_{i,j}$ passing through v_k is given. Thus, we come up with 3 estimation of $\rho(Y_i(v_k))$: $\rho_T(Y_i(v_k))$, $\rho_{in}(Y_i(v_k))$ and $\rho_{out}(Y_i(v_k))$.

C. Using Formulated Feature to Estimate Service Utility Factor

From now we will gradually remove the 'no packet drop' assumption. in a step-by-step manner. A first step to estimation of service utility factor ρ for node v_k holds under the condition of Eq. 6. but this assumes there is no packet drop for all $p_{i,j}$ passing through v_k, which is too strong. The estimation only precise if $v_k = v_{src}$

$$\rho_T(Y_i(v_k)) = \frac{t^s(i, v_k)}{c(Y_i(v_k))} \tag{15}$$

Second, we expand to more than one optional 'next hop' scenario. If we expect more than one possible 'next hop' for v_k other than v_{k+1}, we can suggest a vector denotes all possible optional next-hop nodes $V_{i[O]}^{\mathcal{L}}$, then we can modify Eq. 14 into:

$$\rho_{out}(Y_i(v_k))A_{out(k)} = ||_{v_o \in V_{i[O]}^{\mathcal{L}}} \rho(Y_i(v_o)) \tag{16}$$

Where $A_{out(k)}$ denote weighted out-adjacency row of node v_k in A_{out}, $||$ denotes concatenation in column axis. The condition to ensure Eq. 16 holds is there in no packet drop in v_k.

Third, if we expand to multiple optional 'previous hop' sets $V_{i[I]}^{\mathcal{L}}$ for v_k:

$$||_{v_i \in V_{i[I]}^{\mathcal{L}}}(\rho(Y_i(v_i))A_{in(k)}^T = \rho_{in}(Y_i(v_k)) \tag{17}$$

Table I: Additional Notation Table for Appendix

Given $\mathcal{L}^{-1}(v_k) = e_k, v_k \in V_i^{\mathcal{L}}, e_k \in E_i$**:**	
$q(Y_i(v_k))$	Queue size for $v_k \in V_i^{\mathcal{L}}$
$ps(Y_i(v_k))$	Joint packet size distribution for $v_k \in V_i^{\mathcal{L}}$
$\lambda(Y_i(v_k))$	Joint packet arrival rate for $v_k \in V_i^{\mathcal{L}}$
$c(Y_i(v_k))$	Link capacity for $v_k \in E_i$
$t^s(i, v_k)$	Summed traffic for all $p_{i,j}$ passing through node $v_k \in V_i^{\mathcal{L}}$
$O(i, v_k)$	v_k's queue occupancy state (ground truth), for $v_k \in V_i^{\mathcal{L}}$
$ps(Y_i(v_k))^{rms}$	Effective packet size arrived at v_k. Introduced in equ.1
$ps(Y_i(v_k))(t)$	Packet size arrived at time t. Introduced in equ.2
$ps_j(Y_i(v_k))(t)$	Packet size arrived at v_k, by OD pair $p_{i,j}$, at time t
$q_{i,j}$	Latency of $p_{i,j}$
q_{i,v_k}	Contribution of latency of $p_{i,j}$ on v_k, $v_k \in p_{i,j}$. Introduced in equ. 3
$\rho(Y_i(v_k))$	Service utility factor on v_k
$Prob(Y_i(v_k), n)$	Probability of at any given time, there are n packet in the queue of v_k. Introduced in equ.11
$\rho_T(Y_i(v_k))$	Estimation of $\rho(Y_i(v_k))$, from global traffic. Introduced in equ.15
$\rho_{in}(Y_i(v_k))$	Estimation of $\rho(Y_i(v_k))$, from prior hops' traffic. Introduced in equ.17
$\rho_{out}(Y_i(v_k))$	Estimation of $\rho(Y_i(v_k))$, from subsequent hops' traffic. Introduced in equ.16
$V_{i[I]}^{\mathcal{L}}$	Set of nodes denoting all prior hops before v_k for all $p_{i,j}$, $p_{i,j} \in P_i$ and $v_k \in p_{i,j}$.
$V_{i[O]}^{\mathcal{L}}$	Set of nodes denoting all subsequent hops after v_k for all $p_{i,j}$, $p_{i,j} \in P_i$ and $v_k \in p_{i,j}$.
$A_{in(k)}$	Adjacency formed by $V_{i[I]}^{\mathcal{L}}$
$A_{out(k)}$	Adjacency formed by $V_{i[O]}^{\mathcal{L}}$

Where $A_{in(k)}^T$ denote weighted in-adjacency column of v_k. The condition to ensure Eq. 17 holds is there in no packet drop in all $v_i \in V_{i[I]}^{\mathcal{L}}$.

Through Eq. 15, 16 and 17, we have come up with three estimation of the service utility, under different constrains. Denoting node attribute same as Eq. **??**:

$$X_i^{\mathcal{L}}(v_k) = \frac{t^s(i, v_k)}{c(Y_i(v_k))} \tag{18}$$

We can turn Eq. 13 into:

$$q_{i,v_k} = F(\rho_T(Y_i(v_k)), ||_{v_i \in V_{i[I]}^{\mathcal{L}}} \rho_{in}(Y_i(v_i)), ||_{v_0 \in V_{i[O]}^{\mathcal{L}}} \rho_{out}(Y_i(v_o))) * X_i^{\mathcal{L}}(v_k) \tag{19}$$

One could see this formulation can be easily approximated with DGCN (refer Eq. **??**) message passing aggregation:

$$h_{v_k}^{N+1} = \sigma(\mathbf{W}(\alpha m_{\mathbf{N}_{in}(v_k)}, \beta m_{\mathbf{N}_{out}(v_k)}, m_{\mathbf{N}_{all}(v_k)}) h_{v_k}^N) \tag{20}$$

where $m_{\mathbf{N}_{in}(v_k)}$ denotes message from v_k's in-degree neighbors $\mathbf{N}_{in}(v_k)$, $m_{\mathbf{N}_{out}(v_k)}$ denotes message from v_k's out-degree neighbors $\mathbf{N}_{out}(v_k)$. Similar formulation applies on $m_{\mathbf{N}_{all}(v_k)}$ as well. $h_{v_k}^N$ denotes targeting node attribute embedding at layer N. α, β and $\mathbf{W}$ are the trainable parameters.

D. Dataset Analysis

In this section, we give a detailed illustration of its attributes distribution in Fig. 1.Note we have been using Rolx[2] to extract role automatically from $G_i^{\mathcal{L}}$. As we mentioned in Fig. **??** previously, one can observe extrapolated feature distribution pdf on test data. Besides larger graph size, *(i)* Many network features and the ground truth (OD-pair latency) is considered imbalanced for test set comparing to training set we are trained with; *(ii)* There exists unseen values in the test set features (Link capacity, expected traffic summed per link $t^s(i, v)$ and number of OD-pair sharing the same link) that is considered to be out-of-distribution; *(iii)* Relational dependencies (i.e.: OD-pair Trajectory length) expand longer in test data.
Comparing with expansion of other traffic-related network[3, 4], we consider SDN posed unique extrapolation pattern towards its features: *(i)* Unlike human mobility traffic network, little social impact and land-of-use pattern is considered in routing network expansion. *(ii)* According to Hasslinger and Schnitter [1], having more users in the network could pose a global impact on the network status, regardless of the location of introducing the users. *(iii)* Expanding per-link capacity is a verified practice in network expansion, while similar approach (expanding lane number) can be a source of congestion in other mobility network.

Reference for Appendix

[1] C Hasslinger and S Schnitter. Ip network expansion for growing traffic demand with shortest path routing compared to traffic engineering. In *11th International Telecommunications Network Strategy and Planning Symposium. NETWORKS 2004,*, pages 81–86. IEEE, 2004.

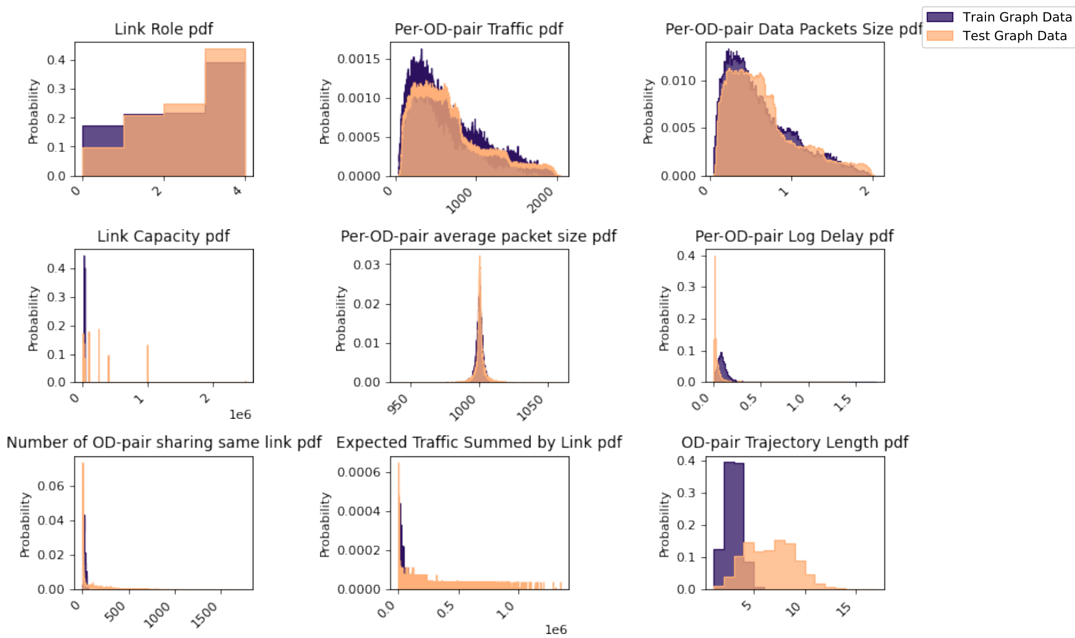

Figure 1: Features and Ground Truth Probability Density Function(pdf)

[2] Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Hanghang Tong, Sugato Basu, Leman Akoglu, Danai Koutra, Christos Faloutsos, and Lei Li. Rolx: structural role extraction & mining in large graphs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1231–1239, 2012.

[3] Yikang Rui and Yifang Ban. *Urban Growth Modeling with Road Network Expansion and Land Use Development*, volume 2, pages 399–412. 01 2011. ISBN 978-3-642-19213-5. doi: 10.1007/978-3-642-19214-2_27.

[4] Yena Song, Hyun Kim, Keumsook Lee, and Kwangwon Ahn. Subway network expansion and transit equity: A case study of gwangju metropolitan area, south korea. *Transport Policy*, 72:148–158, 2018.