A Ablation Studies

We conduct a series of ablation studies to evaluate the impact of different design choices and components on localization performance.

A.1 Node Features, Graph Construction, and Augmentations

Table 6 evaluates the impact of various node features in static and dynamic settings using the GCN model on a validation set. Incorporating Cartesian coordinates and weighted centroid distances improves RMSE, while azimuth-based features do not provide benefits. Local noise statistics, particularly mean noise deviation, significantly enhance accuracy. Furthermore, we incorporate unique IDs [19], with which we observe a degradation in performance, possibly due to a loss in the generalization ability of the model [32]. In dynamic scenarios, direction vectors and temporal noise variations further reduce RMSE, highlighting the importance of motion-aware attributes.

Table 6: Node feature ablation study.

(a) Static.

Node Feature	**RMSE (m)**
Baseline	$67.4_{\pm 0.1}$
+ $\mathbf{x}_i^{\text{cart}}$	$66.0_{\pm 0.1}$ ↓ 2.1%
+ d_i^{wcent}	$66.0_{\pm 0.1}$ - 0.0 %
+ $\mathbf{x}_i^{\text{wcent}}$	$63.6_{\pm 0.2}$ ↓ 3.6%
+ Azimuth to centroid	$66.1_{\pm 0.2}$ ↑ 0.2%
+ d_i^{wcent}	$63.3_{\pm 0.2}$ ↓ 0.5%
+ Azimuth to WC	$65.5_{\pm 0.5}$ ↑ 3.5%
+ median($\eta_{\mathcal{N}_{k(i)}}$)	$61.3_{\pm 0.2}$ ↓ 3.2%
+ max($\eta_{\mathcal{N}_{k(i)}}$)	$59.4_{\pm 0.2}$ ↓ 3.2%
+ $\Delta\eta_i$	$57.3_{\pm 0.0}$ ↓ 3.5%
+ Random feature	$57.5_{\pm 0.1}$ ↑ 0.3%

(b) Dynamic.

Node Feature	**RMSE (m)**
Baseline	$27.4_{\pm 0.3}$
+ Moving avg. signal	$27.5_{\pm 0.3}$ ↑ 0.4%
+ Path distance	$32.7_{\pm 0.4}$ ↑ 19.3%
+ $\mathbf{d}_i$	$26.2_{\pm 0.2}$ ↓ 4.4%
+ $\Delta\eta_i^{\text{temp}}$	$25.6_{\pm 0.3}$ ↓ 6.6%

A.2 Data Augmentations

In this section, we present additional experiments related to data augmentations, including feature corruption, rotation augmentation, DropNode [5], and random cropping. Feature corruption [4] applies Gaussian noise to a randomly selected subset of node features. Up to half of the total features in each vector are chosen, and a mask determines which node values receive noise with probability p. Rotation augmentation rotates the graph by a random angle between 0 and 360 degrees to encourage orientation invariance. DropNode removes nodes with a drop rate of p. The scaling factor is deliberately excluded to maintain the integrity of physical relationships. Random cropping selects a portion of the graph by identifying the three nodes with the highest noise levels and extending to a randomly chosen endpoint. Although we recognize the preexisting

GraphCrop augmentation[27], this method is not relevant for our application which is characterized by uniform node degree and localized areas of importance in the graph. Table 4 presents the results of these experiments, showing that the best performance is achieved with DropNode at $p = 0.2$. Consequently, we adopt this setting exclusively during training, as discussed in Section 5.

Table 7: Performance comparison of different augmentation techniques.

Augmentation	p	RMSE
Feature Noise	1.0	$55.0_{\pm 0.1}$
Feature Noise	0.5	$54.6_{\pm 0.1}$
Feature Noise	0.1	$54.7_{\pm 0.1}$
DropNode	0.5	$54.7_{\pm 0.2}$
DropNode	0.2	$\mathbf{52.9_{\pm 0.1}}$
DropNode	0.1	$53.3_{\pm 0.1}$

We also evaluate combined augmentation strategies where Crop+DN refers to cropping followed by DropNode, DN+Feat Noise represents DropNode followed by feature corruption, and Crop+Feat Noise denotes cropping followed by feature corruption. Table 4d illustrates that combining these augmentations does not significantly improve performance. The best result is achieved with DN + Feat Noise, reinforcing our decision to use DropNode with $p = 0.2$ exclusively.

Table 8: MAE in jammer localization for static scenarios, averaged over three trials with different seeds. Results are split by sampling geometry.

	Method	**Jammer within ($\mathbf{x}_j \in \mathcal{R}$)**					**Jammer outside ($\mathbf{x}_j \in \mathcal{A} \setminus \mathcal{R}$)**					**Mean**
		C	**T**	**R**	**RD**	**Mean**	**C**	**T**	**R**	**RD**	**Mean**	
MAE	WCL	37.7	40.2	27.2	35.7	35.2	138.6	169.0	170.1	164.2	160.5	97.9
	PL	122.2	83.3	90.6	82.0	94.5	289.4	247.3	290.6	267.4	273.7	184.1
	MLE	70.5	65.5	63.4	67.4	66.7	208.4	197.9	203.5	230.3	210.0	138.4
	MLAT	121.6	89.6	75.3	64.8	87.8	282.2	259.9	275.2	235.8	263.3	175.6
	LSQ	237.9	138.5	92.0	94.1	140.6	410.5	329.7	351.1	310.9	350.6	245.6
	MLP	37.1	31.3	24.7	29.5	30.7	60.1	78.2	78.5	84.1	75.2	53.0
	GCN	35.5	30.7	25.9	35.3	31.9	57.5	75.5	77.4	85.4	74.0	53.0
	PNA	34.3	27.7	21.9	26.7	27.7	56.5	72.7	74.1	79.4	70.7	49.2
	GAT	33.3	27.4	21.5	27.2	27.4	55.8	72.0	73.5	77.9	69.8	48.6
	CAGE	**28.4**	**23.6**	**19.8**	**25.0**	**24.2**	**46.5**	**61.6**	**65.1**	**68.8**	**60.5**	**42.5**

B Further Results and Analysis

Figure 5 illustrates the impact of varying parameters, such as jammer power, network density, and shadowing effects, on localization performance for CAGE, GAT, and WCL on the static experiment data. Across all conditions, CAGE

Table 9: MAE in jammer localization along dynamic trajectories, reported across distance intervals to the jammer and averaged over three trials. The final column reports the mean MAE across the full trajectory.

	Method	Distance to the Jammer					Mean
		$d > 500$	$d \in [500, 200]$	$d \in [200, 100]$	$d \in [100, 50]$	$d \in [50, 0]$	
MAE	WCL	316.8	220.6	95.2	34.6	7.7	28.2
	PL	323.8	275.3	153.5	74.0	28.9	56.3
	MLE	245.1	130.0	54.2	36.5	1381.8	1006.7
	MLAT	237.8	192.5	122.2	85.8	57.7	73.1
	MLP	140.5±9.7	83.7±4.9	35.5±1.5	21.8±0.9	14.8±0.8	20.7±0.4
	GCN	123.0±3.8	65.4±1.0	25.2±0.1	14.3±0.2	8.1±0.2	13.1±0.1
	PNA	162.1±5.2	95.1±2.6	40.9±2.2	26.3±1.1	19.5±1.0	25.7±1.0
	GAT	88.7±10.6	49.3±1.4	23.1±0.7	13.3±0.4	7.3±0.4	11.4±0.5
	CAGE	**73.1**±0.7	**36.7**±0.9	**17.1**±0.6	**9.9**±0.1	**4.2**±0.1	**7.6**±0.1

consistently achieves the lowest RMSE, highlighting its robustness. Stronger shadowing (σ) degrades performance across all methods due to increased signal uncertainty [21]. An inverse relationship between jammer power and localization accuracy emerges, likely because lower power forces nodes closer to the jammer, improving localization (see Tables 8 and 1). As shown in Appendix D, Table 11, the 2D graphs predominantly consists of instances with a small number of nodes. This imbalance limits the model's exposure to such graphs, leading to degraded performance on larger graphs. The lack of sufficient representation may contribute to feature explosions, a phenomenon consistent with prior findings [3].

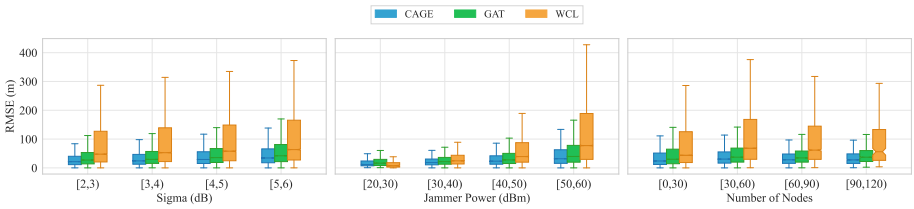

Fig. 5: RMSE analysis of the CAGE, GAT and WCL models across all topologies under the LDPL environment with varying shadowing effects, jammer transmit powers, and number of nodes per graph.

C Training Details and Hyperparameter Tuning

Each model is trained for 300 epochs using the AdamW optimizer, with weights initialized via Xavier initialization [26]. The learning rate is linearly warmed up over the first 20% of the total epochs and then follows a cosine annealing schedule. Dropout is set to 0.5 [20] and is applied to graph representations before the final linear layers for all models, except for CAGE, where it is set to 0, as this configuration was found to yield the best results. Weight decay is fixed at 10^{-5} for all models.

We perform hyperparameter tuning using the Tree-structured Parzen Estimator (TPE) sampler and Hyperband pruner for MLP, GCN, and GAT. PNA hyperparameters follow [2], utilizing mean, maximum, and standard deviation as node feature aggregators. The minimum aggregator was excluded as it degraded performance. The hyperparameters for each model are summarized in Table 10, where "LR" denotes learning rate, "WD" refers to weight decay, "BS" represents batch size, "DO" indicates dropout rate, "NH" corresponds to the number of attention heads (for applicable models), "NL" is the number of layers, and "HC/OC" represents the number of hidden and output channels, respectively.

Table 10: Model hyperparameter tuning results, detailing the optimized learning rates (LR), weight decay (WD), batch sizes (BS), dropout rates (DO), number of attention heads (NH), number of layers (NL), and hidden/output channels (HC/OC) for each model.

Model	LR	WD	BS	DO	NH	NL	HC/OC
MLP	0.0004	0.00001	16	0.5	–	8	128 / 64
GCN	0.0005	0.00001	32	0.5	–	2	512 / 256
PNA	0.001	0.00001	128	0.5	–	6	64 / 64
GAT	0.0007	0.00001	8	0.5	4	8	128 / 128

D Signal Propagation and Interference Modeling

This section complements the problem definition by describing how the RSSI, noise floor, and jammer interference are modeled to generate instances for the jammer localization problem. The environment modeling incorporates signal propagation models and the effects of the jammer on the network.

RSSI Modeling The RSSI ($P_{\text{rx},i}$) at a receiving device i from a transmitting device is calculated as:

$$P_{\text{rx},i} = P_t + G_t + G_r - \text{PL}(d_i),$$

where P_t is the transmit power of the transmitting device (in dBm), G_t and G_r are the antenna gains of the transmitting and receiving devices (in dBi), and $\text{PL}(d_i)$ is the path loss (in dB) as defined in Equation (16), determined by the distance d_i between the transmitting and receiving devices and the adopted propagation model.

The path loss $\text{PL}(d_i)$ is modeled for complex NLOS propagation environments using the LDPL model [18]:

$$\text{PL}_{\text{LDPL}}(d) = PL_0 + 10\gamma \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma, \tag{16}$$

where PL_0 is the reference path loss at a reference distance $d_0 = 1$ m, γ is the path loss exponent, and X_σ is the shadowing effect modeled as a Gaussian random variable.

Noise Floor Modeling The noise floor (η_i^t) at device i is the combination of baseline environmental noise and jammer interference, determined using the relevant path loss model. It is expressed in dBm as:

$$\eta_i^t = 10 \cdot \log_{10}(N_{\text{ambient}} + P_{\text{jam},i}),$$

where N_{ambient} is the ambient noise level, and $P_{\text{jam},i}$ is the jammer RSSI at device i, both represented in mW. The ambient noise level is assumed to be -100 dBm. The jammer's RSSI is computed as:

$$P_{\text{jam},i} = P_t^{\text{jam}} + G_t^{\text{jam}} + G_r - \text{PL}(d_{\text{jam},i}),$$

where P_t^{jam} is the jammer's transmit power, G_t^{jam} is its transmitting antenna gain, and $d_{\text{jam},i}$ is the distance from the jammer at location $\mathbf{x}_j$ to device i.

Data Generation We generate 2D (static) and 3D (dynamic) datasets to evaluate localization algorithms. For the static, we create 25,000 samples per topology per jammer placement strategy and 1000 instances for the dynamic dataset, both utilizing LDPL to model signal propagation. Urban and shadowed urban areas

are modeled, with path loss exponents ranging from 2.7 to 3.5 and 3.0 to 5.0 [1], respectively, and moderate to strong shadowing effects between 2 to 6 dB [21], representing physical phenomena which impact the accuracy of signal-based localization techniques [1]. In the static dataset, nodes are placed within the geographic area $\mathcal{A} = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x, y \leq 1500\}$, with a communication range of 200m. The nodes are arranged in circular, triangular, rectangular, and uniformly random topologies, where the circular and triangular topologies feature a perimeter-bounded node distribution, while a surface-covering node distribution characterizes the rectangular and random topologies. The number of nodes is determined using a beta-distributed scaling factor within a calculated minimum and maximum range, allowing for variability in network densities to simulate different levels of congestion or sparsity. The dataset captures scenarios of both fully jammed or partially jammed networks, ensuring at least one node records a noise floor above -80 dBm, and at least three nodes are jammed to distinguish jamming effects from natural signal fading. For the dynamic dataset, the geographical area of the trajectory is $\mathcal{A} = \{(x, y, z) \in \mathbb{R}^3 \mid 0 \leq x, y \leq 500,\ 0 \leq z \leq 50\}$, though device motion may extend beyond these bounds during simulation. Further information regarding dataset statistics can be found in 11. For reproducibility, we share the data at *https://www.kaggle.com/datasets/daniaherzalla/network-jamming-simulation-dataset*.

Table 11: Datasets statistics reported for number of nodes per graph (Nodes), noise floor (η), jammer transmit power (P_t^{jam}), shadowing (X_σ), path loss exponent (γ), and distance of graph centroid to jammer (D_{jam}).

Data	Feature	Mean	Min/Max
Static	Nodes	$33.6_{\pm 19.1}$	3.0/122.0
	η	$-63.9_{\pm 13.3}$	-100.0/72.4
	P_t^{jam}	$51.1_{\pm 6.3}$	20.0/59.0
	X_σ	$4.2_{\pm 1.2}$	2.0/6.0
	γ	$3.1_{\pm 0.3}$	2.7/5.0
	D_{jam}	$674.6_{\pm 392.6}$	0.1/2117.4
Dynamic	Nodes	$5809.0_{\pm 5527.7}$	1112.0/37144.0
	η	$-56.6_{\pm 18.6}$	-80.0/19.9
	P_t^{jam}	$39.3_{\pm 11.9}$	20.0/60.0
	X_σ	$4.0_{\pm 1.1}$	2.0/6.0
	γ	$3.1_{\pm 0.2}$	2.7/3.5
	D_{jam}	$295.2_{\pm 349.5}$	3.0/2184.7