Supplementary Material for Z-GCNETs: Time Zigzags at Graph Convolutional Networks for Time Series Forecasting

Yuzhou Chen [1 2] **Ignacio Segovia-Dominguez** [3 4] **Yulia R. Gel** [3 2]

A. Additional Experimental Setting

On PeMSD4, we train our model using Adam optimizer with fixed learning rate $lr = 0.003$; on PeMSD8, we train our model using Adam optimizer with an initial learning rate $lr = 0.003$ and decay rate of $\rho = 0.3$; whilst we set learning rate lr of 0.001 in Bytom and Decentraland datasets. The length of Laplacianlink is set to 2 and 3 for transportation networks and token networks, respectively. Our Z-GCNETs is trained with batch sizes of 64 and 8 on PeMSD4 and PeMSD8, respectively. On Ethereum token networks, we set the batch size to 8. We run the experiments for 350 epochs and 100 epochs on transportation networks and Ethereum token networks, respectively. In all experiments, we set the grid size of ZPI to 100×100 and use CNN model to learn zigzag persistence representation. The CNN model consists of 2 CNN layers with number of filter set to 8, kernel size to 3, stride to 2, and the global max-pooling with the pool size of 5×5.

B. The Choice of Filtration

We now perform experiments on the impact of the filtration choice. Here, we consider the weighted-degree sublevel filtration and the weight rank clique filtration, with edge weights being induced either by transaction amounts or by transaction volume. Table 1 shows a subset of illustrative results for Ethereum token networks. For sparser graphs such as Bytom, all filtrations tend to yield similar results, and the weighted-degree sublevel filtration with transaction amounts as edge weight delivers even better performance than the power filtration which is reported in the main body of the paper. For more heterogeneous dynamic graphs with a richer topological structure, e.g., Decentraland, power filtration (see the main body of the paper) is the winner as it better captures evolution of the underlying graph organization. The next best result for Decentraland is delivered by the weight rank clique filtration with edge weight being induced by the transaction amounts.

The proposed methodology is compatible with any filtration.

Table 1. Z-GCNETs (MAPE) for different zigzag filtrations, i.e., weighted-degree sublevel set and weight rank clique filtrations with edge weights induced either by transaction amounts or by transaction volume.

Filtration	**Weighted-degree sublevel set**		**Weight rank clique**
Dataset\Edge Weight	Amount	Volume	Amount
Bytom	30.56	30.80	31.04
Decentraland	25.18	24.93	23.81

C. Ablation study on Ethereum token networks

To make sure that all the components of the Z-GCNETs perform well, we also conduct an ablation study on Ethereum token networks. Table 2 summarizes the results obtained on Bytom and Decentraland. The results demonstrate that our Z-GCNETs outperforms Z-GCNETs without zigzag persistence representation learning (zigzag learning), spatial graph convolution ($\text{GCN}_{\text{Spatial}}$), and temporal graph convolution ($\text{GCN}_{\text{Temporal}}$).

Table 2. Ablation study (MAPE) of Ethereum token networks.

Architecture	**Dataset**	
	Bytom	Decentraland
Z-GCNETs	**31.04%**	**23.81%**
W/o Zigzag learning	33.19%	24.24%
W/o $\text{GCN}_{\text{Spatial}}$	34.32%	25.22%
W/o $\text{GCN}_{\text{Temporal}}$	31.25%	24.62%

[1]Department of Statistical Science, Southern Methodist University, TX, USA [2]Energy Storage & Distributed Resources Division, Lawrence Berkeley National Laboratory, CA, USA [3]Department of Mathematical Sciences, University of Texas at Dallas, TX, USA [4]NASA Jet Propulsion Laboratory, CA, USA. Correspondence to: Yuzhou Chen <yuzhouc@smu.edu>, Ignacio Segovia Dominguez <ignacio.segoviadominguez@utdallas.edu>, Yulia R. Gel <ygl@utdallas.edu>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

D. Proof of Proposition 3.1

Let $\mathrm{DgmZZ}^1_{\nu_*}$ and $\mathrm{DgmZZ}^2_{\nu_*}$ be two zigzag persistence diagrams for some fixed scale parameter ν_*, with the corresponding zigzag persistence surfaces $\rho_{\mathrm{DgmZZ}^1_{\nu_*}}$ and $\rho_{\mathrm{DgmZZ}^2_{\nu_*}}$. Throughout the derivations, we denote the Gaussian kernel $\exp\left\{-||z-\mu||^2/2\vartheta^2\right\}$ as $\kappa_\mu(z)$. Hence, the zigzag persistence surface takes the form

$$\rho_{\mathrm{DgmZZ}_{\nu_*}} = \sum_{\mu\in\mathrm{DgmZZ}'_{\nu_*}} g\left(\mu\right)\kappa\left(\mu\right).$$

We also assume, without loss of generality, that the weighting g is piece-wise linear.

Let γ^* be a bijective map which delivers the infimum over all matchings γ between the zigzag persistence diagrams $\mathrm{DgmZZ}^1_{\nu_*}\cup\Delta$ and $\mathrm{DgmZZ}^2_{\nu_*}\cup\Delta$, where $\Delta=\{(t,t)|t\in\mathbb{R}\}$:

$$\begin{aligned} &d_{W_1}(\mathrm{DgmZZ}^1_{\nu_*},\mathrm{DgmZZ}^2_{\nu_*})\\ &\qquad= \inf_{\gamma}\Big(\sum_{x\in\mathrm{DgmZZ}^1_{\nu_*}\cup\Delta}||x-\gamma(x)||^1_\infty\Big).\end{aligned}$$

Here $||\cdot||_\infty$ is a norm in $\mathcal{L}^\infty$, i.e., $||z||_\infty=\max_i|z_i|$, and d_{W_1} is called a 1-Wasserstein metric.

Note that $\mathrm{DgmZZ}^1_{\nu_*}$ and $\mathrm{DgmZZ}^2_{\nu_*}$ may have different cardinalities. However, since we consider bijections among $\mathrm{DgmZZ}^1_{\nu_*}\cup\Delta$ and $\mathrm{DgmZZ}^2_{\nu_*}\cup\Delta$ (that is, the diagonal set Δ is included), the set of bijections is nonempty and γ^* exists.

Now, following Adams et al. (2017), we obtain

$$\begin{aligned} &||\rho_{\mathrm{DgmZZ}^1_{\nu_*}}-\rho_{\mathrm{DgmZZ}^2_{\nu_*}}||_\infty \qquad (1)\\ \leq& \sum_{\mu\in\mathrm{DgmZZ}^1_{\nu_*}}||g\left(\mu\right)\kappa\left(\mu\right)-g(\gamma^*\left(\mu\right))\kappa_{\gamma^*(\mu)}(z)||_\infty\\ \leq& C_1(||g||_\infty|\nabla\kappa|+||\kappa||_\infty|\nabla g|)\\ &\times\sum_{\mu\in\mathrm{DgmZZ}^1_{\nu_*}}||\mu-\gamma^*\left(\mu\right)||_\infty\\ =& C_1\big(||g||_\infty|\nabla\kappa|+||\kappa||_\infty|\nabla g|\big)\\ &\times d_{W_1}(\mathrm{DgmZZ}^1_{\nu_*},\mathrm{DgmZZ}^2_{\nu_*}),\end{aligned}$$

where $\mathrm{DgmZZ}'^{,1}_{\nu_*}(\mathrm{x},\mathrm{y})=(\mathrm{x},\mathrm{y}-\mathrm{x})$ and C_1 is a positive constant. Here the first inequality is due to the triangle inequality and the second inequality is due to the fundamental theorem of calculus (Krantz & Krantz, 1999) and norm equivalence in $\mathbb{R}^2$ (see also Lemma 1 and Theorem 1 of Adams et al. (2017)).

Again, using the argument of Adams et al. (2017) and definition of the zigzag persistence, for each pixel in the zigzag persistence image, we have

$$\begin{aligned} &|\mathrm{ZPI}^1_{\nu_*\,\mathrm{pixel}}-\mathrm{ZPI}^2_{\nu_*\,\mathrm{pixel}}|\\ &\qquad\leq S_{\mathrm{pixel}}||\rho_{\mathrm{DgmZZ}^1_{\nu_*\,\mathrm{pixel}}}-\rho_{\mathrm{DgmZZ}^2_{\nu_*\,\mathrm{pixel}}}||_\infty, \qquad (2)\end{aligned}$$

where S_{pixel} is the pixel area. Now, combining (1) and (2), we get

$$\begin{aligned} &||\mathrm{ZPI}^1_{\nu_*\,\mathrm{pixel}}-\mathrm{ZPI}^2_{\nu_*\,\mathrm{pixel}}||_\infty\\ &\qquad\leq C_1C_2\big(||g||_\infty|\nabla\kappa|+||\kappa||_\infty|\nabla g|\big)\\ &\qquad\qquad\times d_{W_1}(\mathrm{DgmZZ}^1_{\nu_*},\mathrm{DgmZZ}^2_{\nu_*}),\end{aligned}$$

where $C_2=\max_{\mathrm{pixel}\in\mathrm{ZPI}}\{S_{\mathrm{pixel}}\}$, which concludes the proof.

References

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., and Ziegelmeier, L. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18, 2017.

Krantz, S. G. and Krantz, S. G. *Handbook of complex variables*. Springer Science & Business Media, 1999.