# Bending the Future: Autoregressive Modeling of Temporal Knowledge Graphs in Curvature-Variable Hyperbolic Spaces

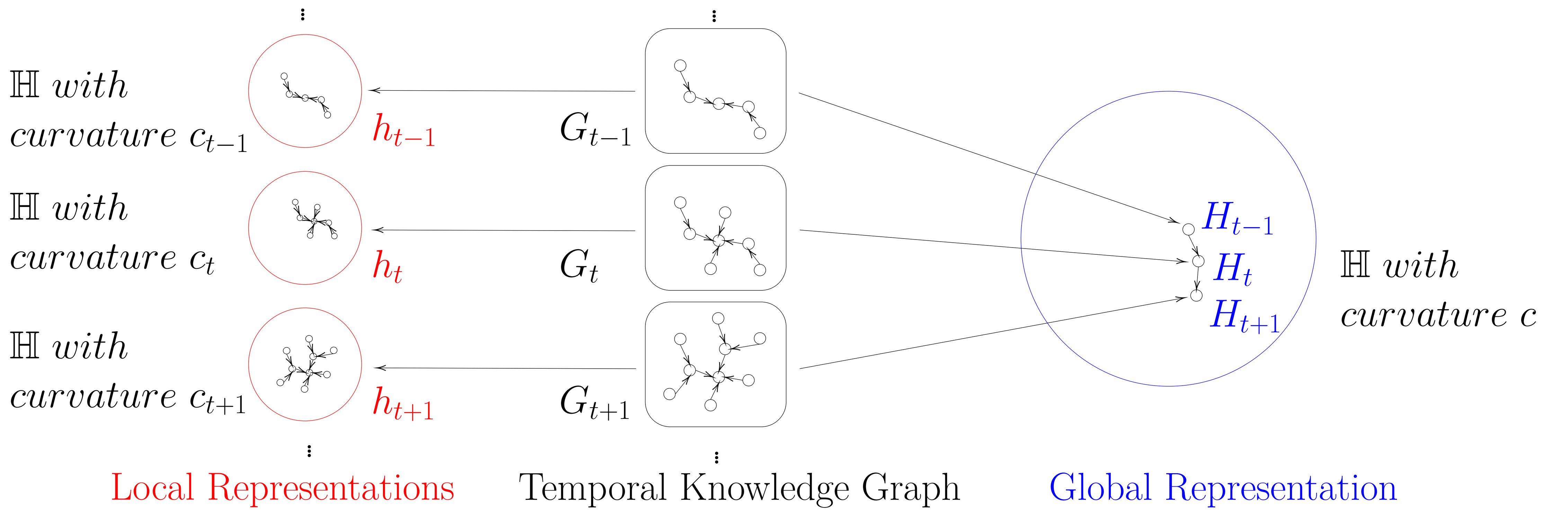
Jihoon Sohn[1], Mingyu Derek Ma[2], Muhao Chen[1]

University of Southern California[1] and University of California, Los Angeles[2]

## Highlights of our method (HyperVC)

1. `HyperVC`: a novel hyperbolic reasoning method for event forecasting in Temporal Knowledge Graphs (TKGs).
2. Modeling the changing hierarchy of TKGs with a learnable function to control the varied curvatures.
3. A hyperbolic RRN performs autoregressive event prediction.
4. Showing consistent improvement on benchmarks, particularly when there are more hierarchical relations.

## Motivation




- Hierarchies can be derived chronologically from TKGs (global representation).
- Each KG at one timestamp has a different level of hierarhcies (local representation).

## Model Components

For a TKG $\{G_t\}_{t\in\mathcal{T}}$, `HyperVC` implements two representations:

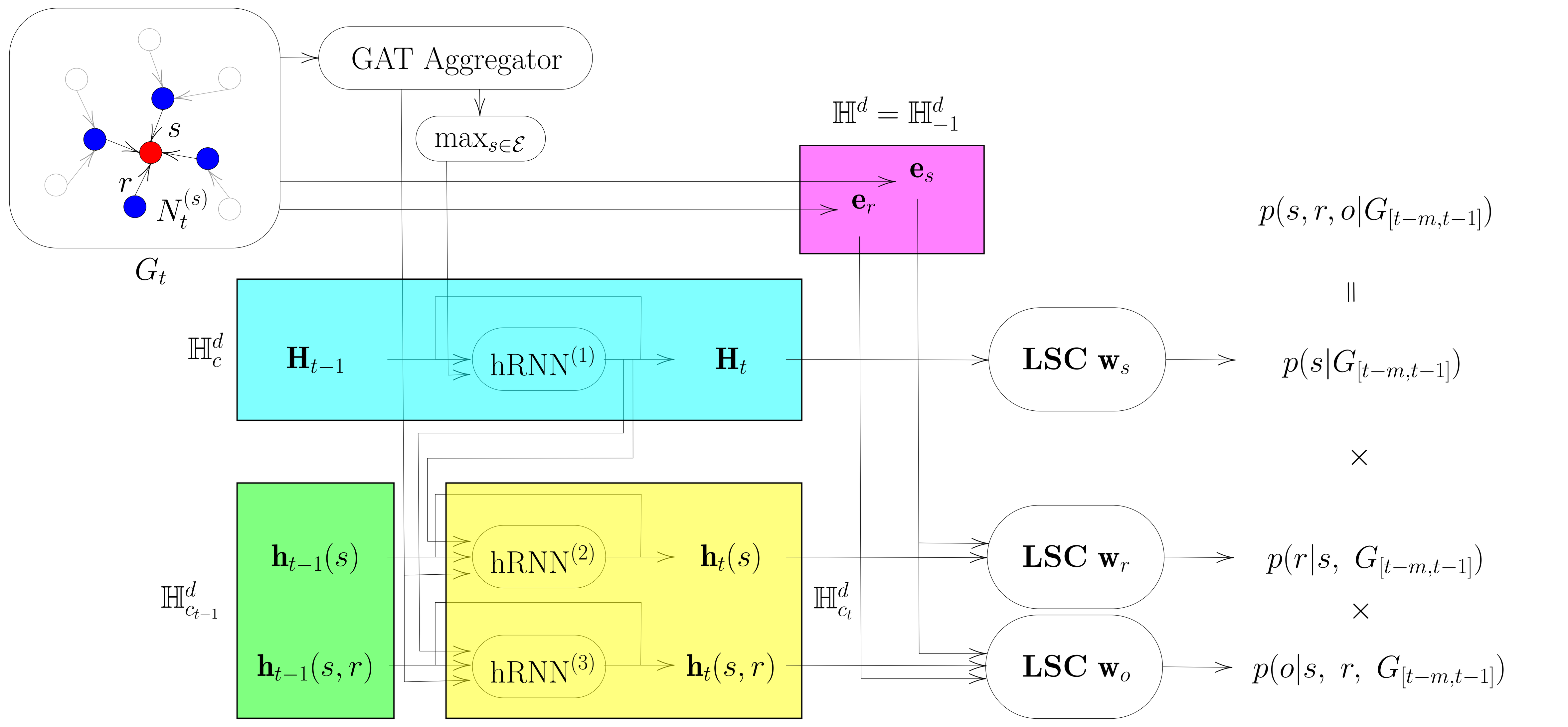
- the global representation $\mathbf{H}_t \in \mathbb{H}^d_c$ with common learnable curvature $c$.
- the local representations $\mathbf{h}_t \in \mathbb{H}^d_{c_t}$, curvatures $c_t$ are optimized for each $G_t$ (different hierarchical levels).

$$\mathbf{H}_t = \text{hRNN}^{(1)}\left(\mathcal{T}^c\left(g'(G_t)\right), \mathbf{H}_{t-1}\right),$$
$$\mathbf{h}_t(s) = \text{hRNN}^{(2)}\left(\mathcal{T}^{c_t}\left(g(N_t^{(s)})\right), \mathcal{T}^{c_t}_c(\mathbf{H}_t), \mathcal{T}^{c_t}_{c_{t-1}}(\mathbf{h}_{t-1}(s))\right),$$
$$\mathbf{h}_t(s,r) = \text{hRNN}^{(3)}\left(\mathcal{T}^{c_t}\left(g(N_t^{(s)})\right), \mathcal{T}^{c_t}_c(\mathbf{H}_t), \mathcal{T}^{c_t}_{c_{t-1}}(\mathbf{h}_{t-1}(s,r))\right),$$

where

- hRNN are the hyperbolic Recurrent Neural Networks (HRNNs),
- $N_t^{(s)}$ is the neighborhood of $s$ in $G_t$,
- $g$, $g'$ are the neighborhood aggregator with graph attention layer, and
- $\mathcal{T}^{c_2}_{c_1}$ is the transformation of curvatures between hyperbolic spaces.

## Architecture of HyperVC (Code)




Through the cyan box of the **global** representations, the green and yellow boxes of the **local** representations, and the pink box of time-consistent hyperbolic embeddings of entities and relations, we calculate probability of triplet $(s, r, o)$ at timestamp $t$. **LSC** refers to the linear softmax classifier.

## Results

| Model | YAGO MRR | YAGO H@1 | YAGO H@3 | YAGO H@10 | WIKI MRR | WIKI H@1 | WIKI H@3 | WIKI H@10 | ICEWS18 MRR | ICEWS18 H@1 | ICEWS18 H@3 | ICEWS18 H@10 | GDELT MRR | GDELT H@1 | GDELT H@3 | GDELT H@10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RE-Net | 65.16 | 63.29 | 65.63 | 68.08 | 51.97 | 48.01 | 52.07 | 53.91 | 42.93 | 36.19 | 45.47 | 55.80 | 40.42 | 32.43 | 43.30 | 53.70 |
| CyGNet | 63.47 | 64.26 | 65.71 | 68.95 | 45.50 | 50.48 | 50.79 | 52.80 | <u>46.69</u> | <u>40.58</u> | <u>49.82</u> | <u>57.14</u> | 50.92 | <u>44.53</u> | <u>54.69</u> | <u>60.99</u> |
| SeDyT-CONV | 66.88 | −− | 67.05 | 68.73 | 52.90 | −− | 52.96 | 54.00 | 45.91 | −− | 45.86 | 49.54 | **54.86** | −− | 54.68 | 58.14 |
| HIP Network | **67.55** | <u>66.32</u> | **68.49** | **70.37** | **54.71** | **53.82** | **54.73** | **56.46** | **48.37** | **43.51** | **51.32** | **58.49** | <u>52.76</u> | **46.35** | **55.31** | **61.87** |
| `HyperVC` | <u>67.52</u> | **66.46** | <u>67.52</u> | <u>69.28</u> | <u>53.02</u> | <u>51.98</u> | <u>53.36</u> | <u>54.55</u> | 41.38 | 34.21 | 44.25 | 55.17 | 40.08 | 32.98 | 42.84 | 53.26 |

Our approach strengthened the performance on the more hierarchical data (i.e., WIKI, YAGO) whereas, with less hierarchical data (i.e., GDELT and ICEWS18), ours did not outperform the earlier models.

## Learnable curvatures

The learnable local curvature $c_t$ is trained as a function of two variables: times and Krackhardt hierarchical scores.

- $c_t^1 = c$ (learnable constant)
- $c_t^2 = -\sigma\left(\alpha\sin(\omega t) + (\beta t + \gamma)\right)$ (time only, additive time series)
- $c_t^3 = -\sigma(f(Khs_{G_t}))$ (hierarchical score only)
- $c_t^4 = -\sigma\left(\alpha\sin(\omega t) + (\beta t + \gamma) + f(Khs_{G_t})\right)$ (combination of two variables)

| Curv. | YAGO MRR | YAGO H@1 | YAGO H@3 | YAGO H@10 |
|---|---|---|---|---|
| $c_t^1$ | 67.49 | **66.46** | 67.49 | 69.12 |
| $c_t^2$ | **67.52** | **66.46** | **67.52** | **69.28** |
| $c_t^3$ | 67.49 | 65.89 | 67.11 | 68.61 |
| $c_t^4$ | 66.79 | 65.57 | 67.10 | 68.54 |

LaTeX TikZposter