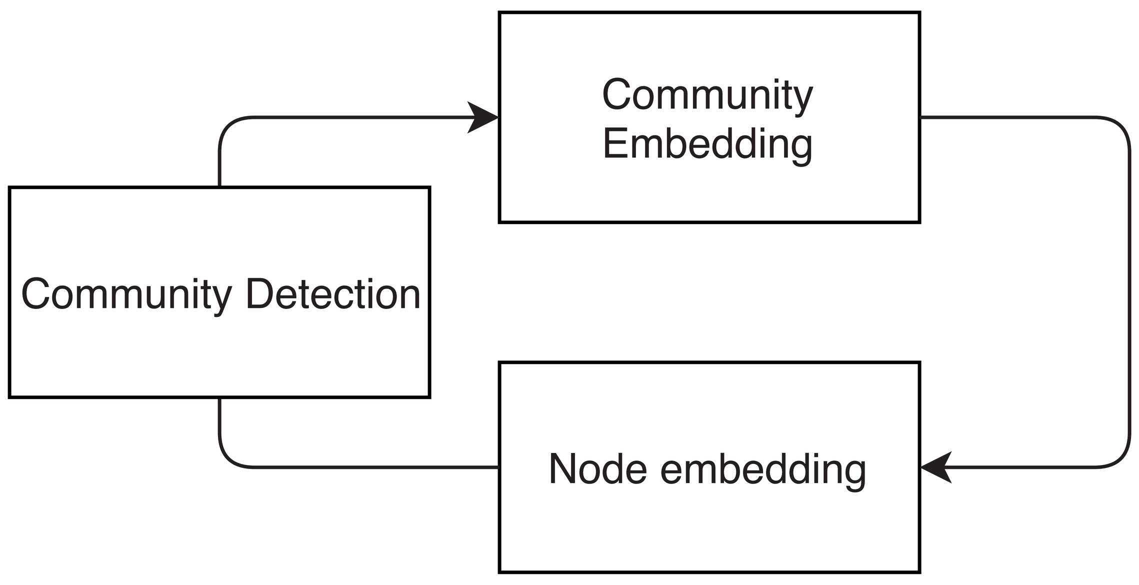
SystemX Institute of Research and Technology
Boosting Digital Transformation

A Hyperbolic approach for learning communities on graphs

Hatem Hajri, Thomas Gerald and Hadi Zaatiti

Tools

Community Learning Approach [1]



Poincaré Embedding [2]

The Poincaré embedding of a binary graph $\mathcal{G} = \{(U, V)\}$ maximises

$$\mathcal{L}(\Theta) = \sum_{(U,V)} \log \left(\frac{e^{-d(u,v)}}{\sum_{v' \in \mathcal{N}(u)} e^{-d(u,v')}} \right)$$

Riemannian Gaussian Distribution [3]

- The density of the Gaussian distribution on $\mathbb{D}$:

$$p(z|\bar{z}, \sigma) = \frac{1}{\zeta(\sigma)} \exp \left[-\frac{d^2(z, \bar{z})}{2\sigma^2} \right]$$

Riemannian EM clustering [3]

Algorithm 1

Initialise **Randomly**
for iter **do**
  $\hat{\varpi}_\mu \leftarrow N_\mu(\hat{\vartheta}) \big/ N$
  $\hat{z}_\mu \leftarrow \text{argmin}_z \sum_{n=1}^{N} \omega_\mu(z_n, \hat{\vartheta})\, d^2(z, z_n)$ ▷ computed using Riemannian gradient descent
  $\hat{\sigma}_\mu \leftarrow \Phi \left(N_\mu^{-1}(\hat{\vartheta}) \times \sum_{n=1}^{N} \omega_\mu(z_n, \hat{\vartheta})\, d^2(\hat{z}_\mu, z_n) \right)$
end for
return $\{(\hat{\varpi}_\mu, \hat{Y}_\mu, \hat{\sigma}_\mu)\}$

Euclidean and Hyperbolic approaches for learning communities on Large Graphs

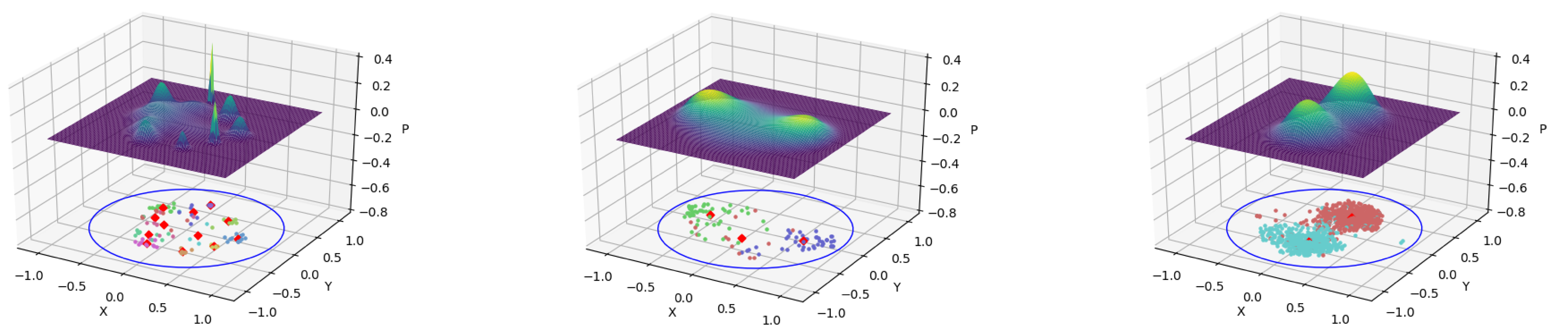
Loss function sum of first, second-order proximities and community losses. :

$$L = \alpha.O_1 + \beta.O_2 + \gamma.O_3$$

		Precision			Conductance			NMI		
Dataset	m	H-KM	H-EM	ComE	H-KM	H-EM	ComE	H-KM	H-EM	ComE
DBLP	2	78.5±1.8	**78.6**±4.8	60.4±5.5	6.8±4.2	**6.7**±4.4	14.1±15.8	66.1±3.4	**66.2**±5.8	50.5±2.1
	5	79.6±2.4	**81.2**±2.1	78.0±4.7	**4.6**±3.4	4.8±3.8	7.4±4.3	**71.3**±2.4	69.7±2.1	63.6±4.4
	10	71.4±13.1	**81.5**±0.1	78.2±1.2	6.0±4.6	**5.2**±4.0	6.5±4.2	65.4±8.9	**69.3**±0.6	62.5±0.8
Wikipedia	2	8.6±0.8	**16.1**±4.0	8.8±0.2	96.6±3.0	96.6±5.1	**94.9**±3.2	**6.9**±0.8	5.5±2.1	6.5±0.
	5	9.7±0.4	**10.1**±0.7	**10.1**±0.2	93.7±3.4	93.8±4.4	**92.9**±3.7	8.8±0.3	8.6±0.3	**10.1**±0.5
	10	9.3±0.3	**11.9**±1.1	**12.1**±0.7	91±4.1	**90.5**±4.7	91.4±4.3	8.6±0.0	8.6±0.1	9.3±0.2
BlogCatalog	2	8.1±0.2	**9.8**±0.3	7.0±0.1	**92.5**±6.0	93.1±8.1	93.4±4.3	**4.4**±0.0	4.1±0.0	3.3±0.
	5	**13.4**±0.3	12.6±0.7	12.3±0.3	88.4±6.6	**87.8**±7.5	87.9±8.8	**10.4**±0.5	10.1±0.5	**10.5**±0.3
	10	**18.9**±0.6	16.5±0.8	15.9±0.5	85.9±7.5	**84.7**±7.8	87.4±8.6	**14.6**±0.2	14.0±0.3	13.7±0.1
Flickr	2	8.0±0.2	**12.9**±0.8	6.2±0.	**93.5**±10.0	94.4±13.0	96.4±2.9	**24.8**±0.6	**24.7**±0.5	21.7±0.
	5	13.0±0.1	**13.4**±0.1	10.2±0.2	**89.7**±12.8	**89.7**±13.9	91.2±10.6	**31.8**±0.1	**31.8**±0.1	29.6±0.2
	10	13.8±0.1	14.0±0.2	13.2±0.	89.5±12.1	**88.2**±14.5	**88.8**±11.8	32.7±0.1	32.7±0.1	**33.0**±0.

Table 1: Performances for learning communities with Hyperbolic K-means (H-KM) and EM (H-EM) in comparison with *ComE* in terms of Precision and NMI: measure the precisions of predictions (the higher the better), Conductance: measures the number of shared graph edges between clusters (the lower the better).

Visualisations of the Approach on Small Graph Datasets



(a) Football. (b) Polbooks. (c) Polblogs.

Figure 1: Visualisations of the Hyperbolic approach to learn communities of the Football, Polbooks and Polblogs datasets.

SPIGL'20

References

[1] Sandro Cavallari et al. Learning community embedding with community detection and node embedding on graphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017*.

[2] Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems 30*, pages 6338–6347. Curran Associates, Inc., 2017.

[3] Salem Said, Hatem Hajri, Lionel Bombrun, and Baba C. Vemuri. Gaussian distributions on Riemannian symmetric spaces: Statistical learning with structured covariance matrices. *IEEE Trans. Information Theory*, 64(2), 2018.