EÖTVÖS LORÁND UNIVERSITY

FACULTY OF SCIENCE
STATISTICAL PHYSICS, BIOLOGICAL PHYSICS AND PHYSICS OF QUANTUM SYSTEMS
PROGRAMME AT THE DOCTORAL SCHOOL OF PHYSICS

Hyperbolic geometry of complex networks: models of network growth and embeddings of real networks

PHD THESIS BOOKLET

DOI: 10.15476/ELTE.2023.008

Author:
Bianka Kovács

Supervisor:
Gergely Palla, DSc

HEAD OF THE DOCTORAL PROGRAMME:
Gábor HORVÁTH, DSc

HEAD OF THE DOCTORAL SCHOOL:
Jenő GUBICZA, DSc

2023

Introduction

In the rapidly expanding research field of complex networks, more and more studies seek for hidden geometric structures of network topology. These rely on the assumption that based on a spatial arrangement of the network nodes in a given metric space, the patterns of node-node interactions can be expressed in terms of some geometric measures calculated for the nodes' position vectors. In a recent branch of the field of network geometry (established mainly by Ref. [1]), instead of the "flat" Euclidean space of zero curvature, hyperbolic spaces of negative curvature are associated with the geometric origin of real-like network structures. This idea originates from the observation that real networks are typically underlain by a tree-like structure abstracting some sort of hierarchical organization in the represented system [2], which tree can be interpreted as a discretized hyperbolic space [1].

From the presumption of an analogy between the hyperbolic geometry and the connection rules of real-world networks, two applications followed naturally: the generation of real-like artificial graphs using distance-dependent connection probabilities between nodes arranged in a hyperbolic space, and the hyperbolic embedding of real networks, where the aim is to find such a hyperbolic arrangement of the network nodes that well reflect the connection structure of the given network through the hyperbolic distances between the node position vectors. Simple hyperbolic network models – including the popularity-similarity optimization (PSO) model [3] of network growth on the hyperbolic plane – are well-known for being able to generate networks that serve as excellent synthetic benchmarks due to simultaneously possessing the small-world property, a scale-free degree distribution and a high average clustering coefficient too, which are all frequently mentioned as common features of many real networks. Hyperbolic embedding methods [4–6] create a low-dimensional vector representation of each network node, which can help in predicting links that are possibly missing or expected to emerge soon, can uncover the communities of the nodes, or can be used for navigating throughout the graph based on local neighborhood relations.

Objectives

My PhD dissertation deals with the fields of both hyperbolic network generation models and hyperbolic node embeddings. In order to further explore the achievable network features in hyperbolic models, besides broadening the investigation of the basic structural properties arising from the original PSO model [3] with the examination of the PSO networks' emergent community structure [T2], I also created more comprehensive versions of the model. On the one hand, I enabled in Ref. [T1] the simulation of the disappearance of links during the network growth in the so-called E-PSO variant [4]. On the other hand, I extended in Ref. [T3] the number of dimensions of the applied hyperbolic space beyond 2.

With the purpose of improving the performance of previous hyperbolic embedding methods, I studied in detail the common procedure of hyperbolic embeddings for creating radial node arrangements in accordance with the E-PSO model [4, 5, 7], finding a previously unexploited degree of freedom in the choice of the radial order of the network nodes, and generalizing the procedure to higher-dimensional cases using my d-dimensional PSO model from Ref. [T3]. Finally, assembling all my experiences regarding the behavior of networks generated in hyperbolic spaces and the well-known techniques of the hyperbolic embedding of undirected networks, I aimed in Ref. [T4] at the development of such hyperbolic embedding algorithms that do not assume any specific hyperbolic network model as the generator of the network to be embedded, allow the utilization of higher-dimensional hyperbolic spaces besides the hyperbolic plane and, most importantly, are capable of embedding not only undirected graphs but directed networks too, grasping the possible asymmetries of the connections between the network nodes.

Thesis points

1. While the popularity-similarity optimization (PSO) model [3] simulates network growth on the hyperbolic plane using solely external links (connecting the new node of the network to the previously appeared ones), the E-PSO model [4] also simulates the formation of so-called internal links (connecting previously appeared nodes to each other). In Ref. [T1], I expanded the E-PSO model with the possibility of the disappearance of previously appeared connections. I demonstrated by simulations that the average internal degree of the subgraphs spanning between nodes having a degree larger than a certain threshold can show strongly different tendencies as a function of the value of the applied smallest node degree, depending on whether the appearance or the disappearance of internal links dominates during the network growth.

2. In Ref. [T2], I studied how the network generation parameters affect the characteristics of the modular structures automatically emerging along the angular node coordinates in networks obtained from hyperbolic models despite the uniform angular distribution of the network nodes on the hyperbolic plane. My detailed numerical investigations revealed that a large variety of network generation parameter settings result in the formation of a strong and relevant community structure in hyperbolic networks.

3. I generalized the number of dimensions of the hyperbolic space in the PSO model above 2 in Ref. [T3].

    (a) I showed that the degree distribution of the networks produced by the new algorithm follows a power law of the theoretically expected exponent, and I also explored by simulations the impact of the different model parameters on the clustering and the community structure of the generated networks and gave intuitive explanations for the observed tendencies.

    (b) For transferring Euclidean embeddings of networks in the hyperbolic space, it is common practice [5, 7] to replace the

Euclidean radial coordinate of each network node with the value that is the most correspondent with the hyperbolic PSO model [4]. I extended this procedure to an arbitrary number of dimensions according to the newly introduced dPSO model [T3].

4. I pointed out [T1] for different embeddings [4, 5] of real networks on the hyperbolic plane that the likelihood maximizing procedure [4] based on the PSO model [3] does not yield a unique radial arrangement of the network nodes, and that the quality of the embeddings created with the obtained different radial arrangements shows a normal distribution both according to the logarithmic loss [4] and the greedy routing score [5] that measures the efficiency of the navigation along the embedded network relying on local information. I described in Ref. [T1] the function characterizing the relationship of these two measures with the number of embedding trials (i.e., the number of tested radial arrangements), the fitting of which to measured data enables the prediction of the expectable improvement of the embedding quality during the further trials.

5. I developed in Ref [T4] such node embedding algorithms, which operate in hyperbolic spaces of any number of dimensions, are capable of taking into account not only the weight but also the directedness of a network's links, and do not assume any specific network generating model behind the network to be embedded.

   (a) I introduced a mapping using which the position vectors of Euclidean embeddings that represent high connection probabilities with high inner products can be converted to position vectors in the hyperbolic space in such a way that in the obtained hyperbolic embedding the large connection probabilities become indicated by small hyperbolic distances. I presented the excellent applicability of the new conversion formula in the case of Euclidean node arrangements yielded by both the well-known HOPE (High-Order Proximity preserved Embedding) method [8] and my own Euclidean embedding algorithms: the HOPE-S and HOPE-R methods that eliminate the angular restriction of the HOPE algorithm, TREXPEN (TRansformation of EXponential shortest Path lengths to EuclideaN measures) that is similar to HOPE but starts from a new, exponential measure of the proximity along the graph, and also the methods named TREXPEN-S and TREXPEN-R.

(b) Starting from the hydra (hyperbolic distance recovery and approximation) [6] method that is designed to create node embeddings of undirected networks directly in a hyperbolic space, making the dimension reduction step suitable for handling directed graphs too and replacing the distance matrix applied by hydra with a new, exponential matrix that enables the embedding of even weakly connected directed networks, I constructed the TREXPIC (TRansformation of EXponential shortest Path lengths to hyperbolIC measures) algorithm, which arranges the nodes of undirected and directed networks in a hyperbolic space without preparing an intermediate Euclidean embedding.

(c) I examined the performance of the directed networks' embedding methods in three tasks: from the point of view of the relation between the shortest path lengths measured along the network and the different geometric quantities measured in the embeddings, the reconstructability of the links based on the different geometric measures, and the efficiency of the navigation along the embedded networks using local information (i.e., the so-called greedy routing [9–11]). I showed that the proposed hyperbolic embedding algorithms perform relatively well on real networks in all the tasks.

Publications during the doctoral training

Publications used in the dissertation

[T1] B. Kovács and G. Palla. Optimisation of the coalescent hyperbolic embedding of complex networks. *Scientific Reports* **11**, 8350, https://doi.org/10.1038/s41598-021-87333-5 (2021)

[T2] B. Kovács and G. Palla. The inherent community structure of hyperbolic networks. *Scientific Reports* **11**, 16050, https://doi.org/10.1038/s41598-021-93921-2 (2021)

[T3] B. Kovács, S. G. Balogh and G. Palla. Generalised popularity-similarity optimisation model for growing hyperbolic networks beyond two dimensions. *Scientific Reports* **12**, 968, https://doi.org/10.1038/s41598-021-04379-1 (2022)

[T4] B. Kovács and G. Palla. Model-independent embedding of directed networks into Euclidean and hyperbolic spaces. *Communications Physics* **6**, 28, https://doi.org/10.1038/s42005-023-01143-x (2023)

Other publications

[S1] S. G. Balogh, B. Kovács and G. Palla. Maximally modular structure of growing hyperbolic networks. Preprint at arXiv:2206.08773 [physics.soc-ph], accepted for publication in Communications Physics on March 16, 2023. https://doi.org/10.48550/arXiv.2206.08773

References

1. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. & Boguñá, M. Hyperbolic geometry of complex networks. *Phys. Rev. E* **82,** 036106. https://doi.org/10.1103/PhysRevE.82.036106 (2010).

2. Clauset, A., Moore, C. & Newman, M. E. J. Hierarchical structure and the prediction of missing links in networks. *Nature* **453,** 98–101. ISSN: 1476-4687. https://doi.org/10.1038/nature06830 (2008).

3. Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489,** 537–540. https://doi.org/10.1038/nature11459 (2012).

4. Papadopoulos, F., Psomas, C. & Krioukov, D. Network Mapping by Replaying Hyperbolic Growth. *IEEE/ACM Transactions on Networking* **23,** 198–211. ISSN: 1063-6692. https://doi.org/10.1109/TNET.2013.2294052 (2015).

5. Muscoloni, A., Thomas, J. M., Ciucci, S., Bianconi, G. & Cannistraci, C. V. Machine learning meets complex networks via coalescent embedding in the hyperbolic space. *Nature Communications* **8,** 1615. ISSN: 2041-1723. https://doi.org/10.1038/s41467-017-01825-5 (2017).

6. Keller-Ressel, M. & Nargang, S. Hydra: a method for strain-minimizing hyperbolic embedding of network- and distance-based data. *Journal of Complex Networks* **8.** cnaa002. ISSN: 2051-1329. https://doi.org/10.1093/comnet/cnaa002 (Feb. 2020).

7. Alanis-Lobato, G., Mier, P. & Andrade-Navarro, M. Efficient embedding of complex networks to hyperbolic space via their Laplacian. *Sci. Rep.* **6,** 301082. https://doi.org/10.1038/srep30108 (2016).

8. Ou, M., Cui, P., Pei, J., Zhang, Z. & Zhu, W. *Asymmetric Transitivity Preserving Graph Embedding* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, San Francisco, California, USA, 2016), 1105–1114. ISBN: 9781450342322. https://doi.org/10.1145/2939672.2939751.

9. Kleinberg, J. Navigation in a small world. *Nature* **406,** 845. https://doi.org/10.1038/35022643 (2000).
10. Boguñá, M., Krioukov, D. & Claffy, K. Navigability of complex networks. *Nature Phys.* **5,** 74–80. https://doi.org/10.1038/nphys1130 (2009).
11. Muscoloni, A. & Cannistraci, C. V. Navigability evaluation of complex networks by greedy routing efficiency. *Proceedings of the National Academy of Sciences* **116,** 1468–1469. ISSN: 0027-8424. eprint: https://www.pnas.org/content/116/5/1468.full.pdf. https://doi.org/10.1073/pnas.1817880116 (2019).