Link prediction with hyperbolic geometry

Kitsak, Maksim; Voitalov, Ivan; Krioukov, Dmitri

DOI
10.1103/PhysRevResearch.2.043113

Publication date
2020

Document Version
Final published version

Published in
Physical Review Research

Link prediction with hyperbolic geometry

Maksim Kitsak,[1,2] Ivan Voitalov,[3,2] and Dmitri Krioukov[4,2]

[1]*Faculty of Electrical Engineering, Delft University of Technology, Mathematics and Computer Science, 2600 GA Delft, The Netherlands*
[2]*Network Science Institute, Northeastern University, 177 Huntington Avenue, Boston, Massachusetts 02115, USA*
[3]*Department of Physics, Northeastern University, 110 Forsyth Street, 111 Dana Research Center, Boston, Massachusetts 02115, USA*
[4]*Department of Physics, Department of Mathematics, Department of Electrical and Computer Engineering, Northeastern University, 110 Forsyth Street, 111 Dana Research Center, Boston, Massachusetts 02115, USA*

(Received 3 April 2019; revised 6 June 2020; accepted 18 August 2020; published 21 October 2020)

Link prediction is a paradigmatic problem in network science with a variety of applications. In latent space network models this problem boils down to ranking pairs of nodes in the order of increasing latent distances between them. The network model with hyperbolic latent spaces has a number of attractive properties suggesting it must be a powerful tool to predict links, but the past work in this direction reported mixed results. Here we perform a systematic investigation of the utility of latent hyperbolic geometry for link prediction in networks. We first show that some measures of link prediction accuracy are extremely sensitive with respect to inaccuracies in the inference of latent hyperbolic coordinates of nodes. This observation leads us to the development of a hyperbolic network embedding method, the HYPERLINK embedder, which we show maximizes the accuracy of such inference, compared to existing hyperbolic embedding methods. Applying this method to synthetic and real networks, we then find that when it comes to predicting obvious missing links hyperbolic link prediction—for short, HYPERLINK—is rarely the best but often competitive, compared to a multitude of other methods. However, HYPERLINK appears to be at its best, maximizing its competitive power, when the task is to predict less obvious missing links that are really hard to predict. These links include missing links in incomplete networks with large fractions of missing links, missing links between nodes that do not have any common neighbors, and missing links between dissimilar nodes at large latent distances. Overall these results suggest that the harder a specific link prediction task the more seriously one should consider using hyperbolic geometry.

DOI: 10.1103/PhysRevResearch.2.043113

I. INTRODUCTION

Link prediction is a paradigmatic example of forecasting network dynamics [1–4], with diverse applications including the reconstruction of networks based on partial data [5–7] and prediction of future social ties [1,8,9], protein interactions [10–12], and user ratings in recommender systems [13–16].

Latent space network models [17–21] offer an intuitive and simple approach to link prediction. In these models, network nodes are points in a latent space, while connections are established with probabilities that decrease with latent distances between nodes. Latent distances model similarity between nodes, and the main idea behind these models is to model homophily: more similar nodes are more likely to be linked. Link prediction then reduces to ranking unconnected node pairs in the order of increasing latent distances between them: the closer the two unlinked nodes in the latent space, the higher the probability of a missing link [4,22–24].

Among many latent space models considered in literature, only the one that assumes that the latent space is hyperbolic reproduces sparsity, self-similarity, scale-free degree distribution, strong clustering, the small-world property, and community structure [20,25–27]. All these properties are often observed in many real networks [28–30], and hyperbolic geometry captures them all. In addition, the hyperbolic network model is likely to be the simplest or parsimonious with respect to these properties, as in some of its limiting regimes it has been proven to be statistically unbiased, satisfying the maximum entropy principle [31,32].

Given the combination of these attractive properties, one could naturally expect that the hyperbolic latent space model must be a powerful tool in link prediction. Yet the previous studies on this subject reported mixed results [24,33–37].

Here we perform systematic investigation of the efficiency of link prediction using latent hyperbolic geometry. We organize the presentation of the results as follows.

In Sec. II we recall the definitions of the hyperbolic latent space network model, which for short we call random hyperbolic graphs (RHGs), and outline the basic idea behind link prediction based on this model. We also recall the definitions of the main measures of link prediction accuracy—AUC (area under receiver-operating characteristic), AUPR (area under precision-recall curve), and Precision—and discuss what these measures actually measure: while AUPR cares mostly about most obvious easy-to-predict missing links, AUC puts more weight on less obvious and harder-to-predict missing links between more dissimilar nodes, albeit with the cost of not caring that much about false positives.

Our main results are then given in Secs. III and IV. In Sec. III, we calculate analytically the AUC and AUPR on RHGs with *known* hyperbolic coordinates of all nodes. That is, the same coordinates are used both to generate RHGs and to predict missing links in them, an ideal situation yielding the upper bound for the link prediction accuracy using hyperbolic geometry. To understand the robustness of link prediction in the case where coordinates are inferred (Sec. IV), so that they are not equal exactly to the true coordinates, we add uniform noise to the true coordinates, and analyze the AUC, AUPR, and Precision as functions of the noise amplitude to find that (1) AUC is not that sensitive to noise, but (2) AUPR and Precision decrease quickly as noise grows. The latter result implies that the AUPR and Precision scores of link prediction using hyperbolic geometry in real networks can be high only if node coordinates are inferred with sufficiently high accuracy. This is because the most likely missing links candidates are those between similar nodes at small hyperbolic distances, which are most sensitive to coordinate inaccuracies.

To predict missing links in networks with unknown coordinates one first needs to infer these coordinates. Motivated by the results in Sec. III calling for high-accuracy coordinate inference, and given that no existing hyperbolic coordinate inference algorithm is sufficiently accurate, in Sec. IV we develop an alternative one, which we call the HYPERLINK embedder, the focus of which is on high precision in coordinate inference. We present its overview in Sec. IV, while all the details are delegated to Appendix F, where we also compare it to some existing inference algorithms to show that its accuracy is indeed higher. A software package implementing the HYPERLINK embedder is hosted by the Bitbucket repository [38].

We then apply the HYPERLINK embedder to a collection of RHGs with "forgotten" coordinates, and to real networks, calling the overall link prediction procedure the HYPERLINK method, and comparing it to a representative collection of other link prediction methods.

Section V contains both high-level (Tables I and II) and more detailed summaries of all the results. The results are definitely not that the HYPERLINK or any other method is a clear winner in all the considered scenarios according to all the considered link prediction accuracy measures. We discuss what methods are strong in what scenarios. The HYPERLINK appears to be the strongest in the most difficult link prediction tasks. That is, the more challenging a particular link prediction task/scenario, the better off is the HYPERLINK compared to other methods. We conclude the paper with an outline of open problems at the end of Sec. V.

These results emphasize that the HYPERLINK is definitely not the universally best link prediction method, which simply cannot exist as was recently shown in [24,39–41]. That is, there can exist no *one size fits all* solution for the link prediction problem. Different methods are good at predicting different types of links. Therefore, as far as a particular link prediction method is concerned, the best one can do is to document what particular link prediction scenarios the method is good at, that is, what types of links the method is good at predicting, which is exactly the subject of this paper.

II. METHODS

We begin the exposition by discussing the latent geometric link prediction framework and the null model that we utilize to predict missing links.

A. Link prediction with latent geometry

Link prediction with hyperbolic geometry is a two-step procedure. First, one needs to infer node coordinates in the hyperbolic space and calculate hyperbolic distances between node pairs. This coordinate inference procedure is often referred to as network mapping or embedding. The second step of the procedure is to identify most likely missing link candidates. This subsection focuses on the second step of this procedure. The technical details of the null geometric model and the network mapping algorithm constituting the first step are provided in Secs. II B and II C and Appendix F. We refer to the network mapping algorithm and the entire hyperbolic link prediction framework as the HYPERLINK embedder and the HYPERLINK, respectively.

The latent geometric link prediction framework is applicable to all latent geometric models, where connections are established independently with decreasing connection probability function $p(x)$. Intuitively, the smaller the latent distance between two nodes, the higher the probability of a link between them. Then, if two nodes located close to each other in the latent space are not connected, it is likely that there is a missing link between them.

Specifically, consider a latent geometric model where nodes are assigned positions $\{\mathbf{x}_i\}$ in a certain latent space $\mathcal{M}$, and every node pair $\{ij\}$ is connected with probability $p_{ij} = p(x_{ij})$, where $x_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$ is the latent distance between the nodes, and $p : \mathbb{R}_+ \to [0, 1]$ is the decreasing connection probability function specified by the model. After all connections are established, some links are removed with probabilities $1 - q_{ij}$. These pairs of nodes are referred to as missing links.

Any unconnected node pair $\{ij\}$ in the resulting network is either not connected in the network formation process or connected in the network formation and later removed with probability $1 - q_{ij}$. Therefore, the probability for an unconnected pair of nodes $\{ij\}$ separated by x_{ij} to be a missing link is

$$\tilde{p}(x_{ij}) = \frac{p(x_{ij})(1 - q_{ij})}{1 - p(x_{ij}) + p(x_{ij})(1 - q_{ij})}. \tag{1}$$

In the particular case of a decreasing connection probability function $p(x)$ and the random link removal process, $q_{ij} = q$,

$$\tilde{p}(x_{ij}) = \frac{(1 - q)p(x_{ij})}{1 - qp(x_{ij})} \tag{2}$$

is the decreasing function of x_{ij} for any $q > 0$. Thus, the most probable candidates for missing links are indeed unconnected node pairs located at small latent distances, as stated, and the latent geometric link prediction algorithm only needs to rank unconnected node pairs in the increasing order of latent distance between them.

It is important to note, however, that this approach is only guaranteed to work in the case the links are removed uniformly at random. In the general case, missing link probabilities in Eq. (1) depend both on latent distances $\{x_{ij}\}$ and missing link rates $\{1-q_{ij}\}$ and further information on the nature of $\{q_{ij}\}$ is needed to rank missing link candidates properly.

B. Random hyperbolic graphs

While the latent geometric framework described above is applicable to all latent space models, in our paper we use the RHG as a null model for link prediction.

RHGs have been extensively studied in the literature [20,34,35,42–46] and have been shown to reproduce common properties of many real networks including heterogeneous distributions of node degrees, strong clustering, as well as community structure [20,27,47].

The latent space of the RHG model is the two-dimensional hyperbolic disk of constant negative curvature $K=-1$ and radius R. The hyperbolic distance x between any two points in the hyperbolic disk is given by the hyperbolic law of cosines:

$$\cosh x = \cosh r \cosh r' - \sinh r \sinh r' \cos \Delta\theta, \tag{3}$$

where (r,θ) and (r',θ') are the hyperbolic coordinates of the two points within the disk and $\Delta\theta = \pi - |\pi - |\theta - \theta'||$ is the angle between them.

The RHG has three parameters—hyperbolic disk radius $R>0$, temperature $T\in[0,1)$, and node density parameter $\alpha > 1/2$—and is defined as follows.

(1) Draw node coordinates $\{r_i,\theta_i\}$, $i=1,2,\ldots,N$, from probability density functions:

$$\theta_i \leftarrow \rho(\theta) = 1/(2\pi),\ \theta_i \in [0,2\pi], \tag{4}$$

$$r_i \leftarrow \rho(r) = \frac{\sinh(\alpha r)}{\cosh(\alpha R)-1},\ r_i \in [0,R]. \tag{5}$$

(2) Compute distances $\{x_{ij}\}$ between all node pairs using Eq. (3).

(3) Connect node pairs with probability:

$$p(x_{ij}) = \frac{1}{1+e^{\frac{x_{ij}-R}{2T}}}. \tag{6}$$

We summarize basic RHG properties in Appendix C: parameter α controls the exponent $\gamma = 2\alpha+1$ of the power-law degree distribution, while clustering is a decreasing function of temperature T approaching zero in the $N\to\infty$ limit as $T\to 1$. In this limit, clustering is zero for any $T \geqslant 1$.

C. HYPERLINK embedder in a nutshell

To infer hyperbolic coordinates of nodes in a given network with random links removed, we aim to find the set of node coordinates $\{\mathbf{x}_i\} \equiv \{(r_i,\theta_i)\}$, $i=1,2,\ldots,N$, that maximize the posterior probability $\mathcal{L}(\{\mathbf{x}_i\}|a_{ij},\mathcal{P},q)$, which is the probability density function of coordinates $\{\mathbf{x}_i\}$ in an RHG with adjacency matrix a_{ij}, parameters $\mathcal{P}=\{\alpha,T,R\}$, and link removal probability $1-q$. By the Bayes rule this probability is

$$\mathcal{L}(\{\mathbf{x}_i\}|a_{ij},\mathcal{P},q) = \frac{\mathcal{L}(a_{ij}|\{\mathbf{x}_i\},\mathcal{P},q)\mathrm{Prob}(\mathbf{x}_i)}{\mathcal{L}(a_{ij}|\mathcal{P},q)}, \tag{7}$$

where $\mathcal{L}(a_{ij}|\{\mathbf{x}_i\},\mathcal{P},q)$ is the likelihood that network a_{ij} is generated as an RHG with subsequent random link removal with probability $1-q$, $\mathrm{Prob}(\mathbf{x}_i)$ is the prior probability of node coordinates generated by the RHG, and $\mathcal{L}(a_{ij}|\mathcal{P},q)$ is the probability that the network has been generated as the RHG with random link removal. Since node pairs are connected independently, this likelihood is

$$\mathcal{L}(a_{ij}|\{\mathbf{x}_i\},\mathcal{P},q) = \prod_{i<j}[\tilde{p}(x_{ij})]^{a_{ij}}[1-\tilde{p}(x_{ij})]^{1-a_{ij}}, \tag{8}$$

where $\tilde{p}(x_{ij})$ is the effective connection probability in the RHG generation process with subsequent random link removal:

$$\tilde{p}(x) \equiv q p(x), \tag{9}$$

where $p(x)$ is the RHG connection probability function in Eq. (6). Finally, in RHGs node coordinates $\{\mathbf{x}_i\}\equiv\{r_i,\theta_i\}$, and the prior probability is given by

$$\mathrm{Prob}(\mathbf{x}_i) = \frac{1}{(2\pi)^N}\prod_{i=1}^{N}\rho(r_i), \tag{10}$$

where $\rho(r_i)$ is as in Eq. (5).

The HYPERLINK embedder aims to find node coordinates $\hat{\mathbf{x}}_i$ that maximize the likelihood $\mathcal{L}(\{\mathbf{x}_i\}|a_{ij},\mathcal{P},q)$, or equivalently its logarithm:

$$\ln\mathcal{L}(\{\mathbf{x}_i\}|a_{ij},\mathcal{P},q) = K + \sum_{i=1}^{N}\ln\rho(r_i) + \sum_{i<j}\{a_{ij}\ln\tilde{p}(x_{ij}) + (1-a_{ij})\ln[1-\tilde{p}(x_{ij})]\}, \tag{11}$$

where constant K absorbs all terms independent of $\{\mathbf{x}_i\}$.

Similar to other maximum-likelihood estimation (MLE) based embedders [34,35,43,48], node coordinates $\hat{\mathbf{x}}_i$ are computed iteratively: starting with initial random coordinate configuration, the HYPERLINK embedder updates node coordinates at each iteration step to increase $\ln\mathcal{L}(\{\mathbf{x}_i\}|a_{ij})$ and stops when we arrive to a stable configuration. One feature of the HYPERLINK embedder which is different from other MLE-based embedders is that at each iteration step ℓ the embedder adds synthetic noise of variable magnitude $a(\ell)$ to angular node coordinates:

$$\hat{\theta}_i \leftarrow \hat{\theta}_i + a(\ell)X_i, \tag{12}$$

where X_i is a random number drawn from the uniform distribution on the circle $[0,2\pi]$. These coordinate perturbations allow the HYPERLINK embedder to avoid getting trapped for long time in local maxima of the log-likelihood function and to find (nearly) optimal solutions much faster, thus increasing the coordinate inference accuracy given the same amount of computational resources (see Appendix F for details).

D. Link prediction accuracy

We evaluate the accuracy of the HYPERLINK as well as other link prediction methods through random link removal experiments. To this end, we first remove existing links uniformly at random with probability $1 - q$ from the network of interest G. We refer to the remaining network as the *pruned* network and denote it by $\tilde{G}$. We refer to removed links as missing links and denote them by Ω_R. The set of remaining links in $\tilde{G}$ is referred to as Ω_E.

To test the link prediction method of interest we compute likelihood scores for all unconnected node pairs in $\tilde{G}$, $\overline{\Omega}_E$, which include both missing links Ω_R and true nonlinks Ω_N, so that $\overline{\Omega}_E = \Omega_R \cup \Omega_N$. We then rely on these scores to rank unconnected node pairs in the decreasing order of missing link likelihood and refer to them as missing link candidates. We denote the fraction $\lambda \in [0, 1]$ of most likely missing link candidates as set $\Omega_M(\lambda)$. In the case $\lambda = 0$, $\Omega_M(\lambda)$ is the empty set, while in the $\lambda = 1$ case $\Omega_M(\lambda) = \Omega_R \cup \Omega_N = \overline{\Omega}_E$.

In the case the exact number of missing links is known, the most direct way to assess link prediction accuracy is to consider the same number of the most likely missing link candidates and evaluate its intersection with the set of missing links. This metric is known as Precision and is formally defined as

$$\text{Precision} = \frac{|\Omega_R \cap \Omega_M(\lambda^*)|}{|\Omega_R|}, \tag{13}$$

where fraction $\lambda^* = 1 - q$ is chosen such that $|\Omega_M(\lambda^*)| = |\Omega_R|$. The Precision score is bounded by 0 and 1 with the upper bound corresponding to the ideal link predictor ranking all missing links in Ω_R higher than nonlinks in Ω_N.

In practical circumstances, however, the exact number of missing links is often unknown. Further, depending on the application, one might be interested to minimize the number of false positives in the prediction set, possibly by the expense of false negatives, or, vice versa, minimize the number of false negatives by the expense of false positives. One example of the former case where one is interested to minimize the number of false positives, i.e., good citizens misclassified as criminals, is the criminal justice system. This example is in contrast to cancer screening, where the number of false negatives, or not-identified cancer cases, should be minimized. In both cases one is interested to explore the performance of the link predictor for a range of $\Omega_M(\lambda)$ sizes.

A number of link prediction metrics have been developed to this end with the receiver operating characteristic (ROC) and the precision-recall (PR) being the most popular.

To formally introduce ROC and PR curves we first define the confusion matrix. The latter consists of four values—the numbers of *true positives* (TP), *false positives* (FP), *false negatives* (FN), and *true negatives* (TN), Fig. 1—and is extensively used in statistical classification problems. Link prediction is not a genuine classification problem since one is only interested to predict links and not their absence. Nonlink node pairs are predicted implicitly as unconnected node pairs that are not part of Ω_R.

In the context of link prediction, the number of true positives is the number of correctly identified missing links from $\Omega_M(\lambda)$, Eq. (14). The number of false negatives is the remaining number of missing links that are not part of the $\Omega_M(\lambda)$, Eq. (15). The number of false positives is the number of missing link candidates in $\Omega_M(\lambda)$ that are not correctly identified, Eq. (16). Finally, the number of true negatives is the number of unconnected node pairs that are neither true positives nor false positives nor false negatives [see Eq. (17) and Fig. 1]:

$$\text{TP}(\lambda) = |\Omega_R \cap \Omega_M(\lambda)|, \tag{14}$$

$$\text{FN}(\lambda) = |\Omega_R \setminus \Omega_M(\lambda)|, \tag{15}$$

$$\text{FP}(\lambda) = |\Omega_N \cap \Omega_M(\lambda)|, \tag{16}$$

$$\text{TN}(\lambda) = |\Omega_N \setminus \Omega_M(\lambda)|. \tag{17}$$

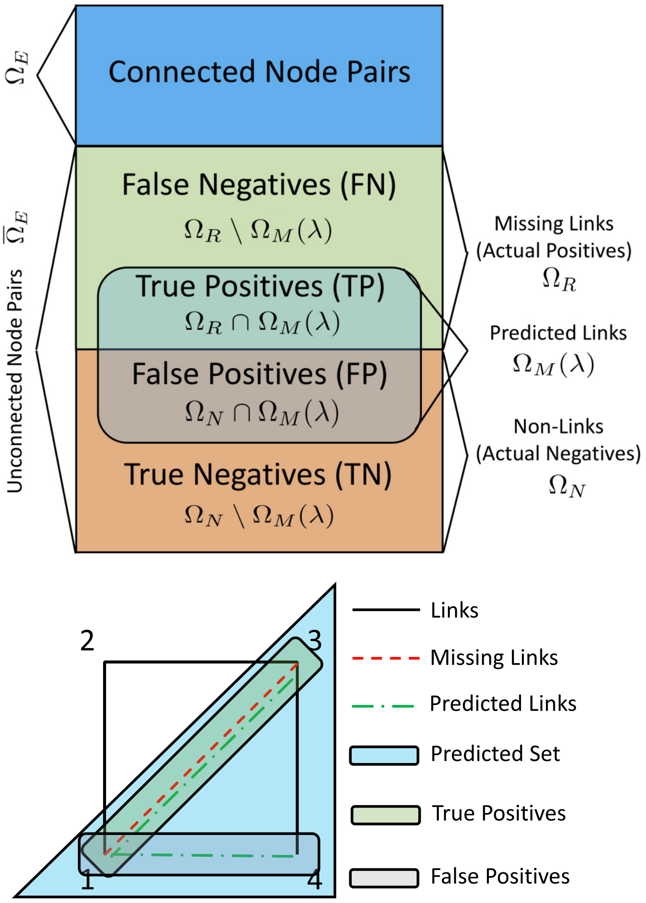

FIG. 1. Confusion matrix and a toy example of link prediction. Top: Confusion matrix for link prediction. Bottom: Toy link prediction example. Existing links are shown with solid black lines. Missing links, $\Omega_R = \{13\}$, are shown with red dotted lines, while predicted missing links, $\Omega(\lambda) = \{13, 14\}$, are shown with green dashed lines. In this example the sizes of the confusion matrix sets are $\text{TP} = 1$, $\text{FP} = 1$, $\text{FN} = 0$, and $\text{TN} = 1$.

Since network sizes vary, it is common to normalize confusion matrix elements, obtaining true positive, false positive, false negative, and true negative rates, formally defined as

$$\text{tpr}(\lambda) \equiv \frac{\text{TP}(\lambda)}{|\Omega_R|}, \tag{18}$$

$$\text{fnr}(\lambda) \equiv \frac{\text{FN}(\lambda)}{|\Omega_R|}, \tag{19}$$

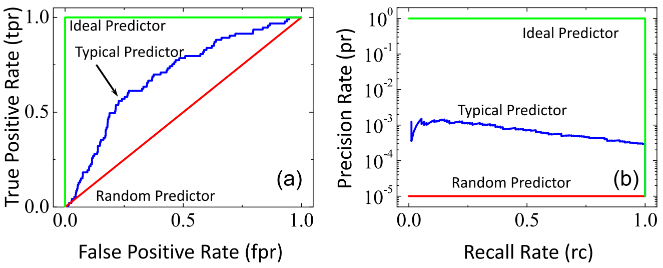

FIG. 2. Sketches of typical (a) ROC and (b) PR curves.

$$\mathrm{fpr}(\lambda) \equiv \frac{\mathrm{FP}(\lambda)}{|\Omega_N|}, \tag{20}$$

$$\mathrm{tnr}(\lambda) \equiv \frac{\mathrm{TN}(\lambda)}{|\Omega_N|}. \tag{21}$$

An ROC statistics or curve is defined as the parametric plot of the true positive rate $\mathrm{tpr}(\lambda)$ as a function of the false positive rate $\mathrm{fpr}(\lambda)$ obtained by varying the fraction of considered link candidates $\lambda \in [0, 1]$. The ideal predictor is expected to rank all node pairs corresponding to missing links, Ω_R, higher than nonlinks, Ω_N, resulting in unit true positive rate and zero false positive rate for $\lambda = 1 - q$, $\mathrm{tpr}(1 - q) = 1$, $\mathrm{fpr}(1 - q) = 0$. The corresponding ROC curve of the ideal predictor is thus a rectangle going through the upper left corner (0,1) of the ROC space. A fully random link predictor, on the other hand, will guess missing links at random from $\overline{\Omega}_E$ and is expected to yield equal true positive and false positive rates, $\mathrm{tpr}(\lambda) = \mathrm{fpr}(\lambda)$ for all λ values, resulting in the diagonal ROC curve, Fig. 2(a).

The standard way to quantify ROC-based prediction accuracy is through the AUC:

$$\mathrm{AUC} = \int_0^1 \mathrm{tpr}(\lambda)\mathrm{fpr}'(\lambda)d\lambda. \tag{22}$$

AUC values vary in between 0 and 1 with $\mathrm{AUC} = 0.5$ corresponding to a fully random predictor and $\mathrm{AUC} = 1.0$ corresponding to the perfect predictor.

The AUC score can be interpreted as the probability that a randomly chosen missing link is assigned a higher link prediction score than a randomly chosen unconnected node pair. ROC curves are easy to read and interpret, which is arguably the basic reason behind their popularity.

At the same time, there is a growing consensus that ROC curves and corresponding AUC scores are insensitive in class imbalance problems, where the size of the positives is disproportional to that of the negatives [49]. Link prediction in sparse networks is one example of class imbalance. Here the number of missing links is of the order of N and is significantly smaller than the number of nonlinks, which is of the order of N^2. Intuitively, in this situation the $\mathrm{tpr}(\lambda)$ rate grows much faster than the false positive rate since the latter is normalized by $|\Omega_N|$ and, as a result, most ROC curves tend to be substantially above the random baseline, yielding AUC scores close to 1.0, regardless of the link prediction method.

An alternative to the ROC curve is the PR characteristic, defined as the parametric plot of the precision rate $\mathrm{pr}(\lambda)$ as a function of the recall rate $\mathrm{rc}(\lambda)$ obtained by varying $\lambda \in [0, 1]$, where the two rates are defined by

$$\mathrm{pr}(\lambda) \equiv \frac{\mathrm{TP}(\lambda)}{|\Omega_M(\lambda)|}, \tag{23}$$

$$\mathrm{rc}(\lambda) \equiv \frac{\mathrm{TP}(\lambda)}{|\Omega_R|} = \mathrm{tpr}(\lambda). \tag{24}$$

That is, the recall rate is identical to the true positive rate, while the precision rate differs from the latter by a different normalization—to the number of predicted links versus the number of removed links.

In the case of an ideal predictor, the precision rate is maximized, $\mathrm{pr}(\lambda) = 1.0$ for $\lambda \leqslant 1 - q$, while the recall is growing from $\mathrm{rc}(0) = 0$ to $\mathrm{rc}(1 - q) = 1$, resulting in the rectangular PR curve going through the upper right corner (1,1) of the PR space. A fully random predictor, on the other hand, maintains constant precision rate equal to the ratio of the number of true missing links to the total number of unconnected node pairs, $\mathrm{pr}^{\mathrm{rand}}(\lambda) = \frac{|\Omega_R|}{|\Omega_R|+|\Omega_N|}$ for all λ values, Fig. 2(b). The standard metric quantifying PR-based prediction accuracy is the AUPR:

$$\mathrm{AUPR} = \int_0^1 \mathrm{pr}(\lambda)\mathrm{rc}'(\lambda)d\lambda. \tag{25}$$

AUPR values vary between $\frac{|\Omega_R|}{|\Omega_R|+|\Omega_N|}$ and 1 with the unit score corresponding to the ideal predictor. In the case of sparse networks $\Omega_R \ll \Omega_N$, leading to $\mathrm{AUPR} \ll 1$ in the case of a random predictor. Unlike ROC curves, PR characteristics do not directly depend on the number of true negatives and, as a result, do not suffer from the class imbalance problem in case of sparse networks.

E. AUC versus AUPR

While both AUC and AUPR quantify link prediction accuracy, they tend to weigh missing link candidates differently. AUPR scores tends to emphasize highly ranked missing links candidates, i.e., those corresponding to small λ values. AUC scores, on the other hand, put more weight on missing links candidates corresponding to larger λ values.

Indeed, AUPR averages precision rate $\mathrm{pr}(\lambda)$ over the recall rate $\mathrm{rc}(\lambda)$. Since the recall rate is given by $\mathrm{rc}(\lambda) = \frac{|\Omega_R \cap \Omega_M(\lambda)|}{|\Omega_R|}$, Eq. (24), good link predictors tend to reach $\mathrm{rc}(\lambda) = 1$ values when the size of missing link candidates set $\Omega_M(\lambda)$ becomes comparable to that of Ω_R: $|\Omega_M(\lambda)| \approx |\Omega_R| \ll |\Omega_N|$. The latter inequality holds in the case of sparse networks, where the number of links is much smaller than the number of nonlinks. Thus, $|\Omega_M(\lambda)| \ll |\Omega_N|$, which corresponds to $\lambda \ll 1$ values. Thus, AUPR link prediction scores are dominated by small λ fractions, i.e., by the most likely and, typically, most obvious missing link candidates in Ω_M.

AUC scores, on the other hand, average true positive rate $\mathrm{tpr}(\lambda)$ over false positive rate $\mathrm{fpr}(\lambda)$. The latter takes large values when $|\Omega_M(\lambda)|$ becomes comparable to $|\Omega_N|$, i.e., for λ values close to 1. AUC scores, thus, are emphasizing not only easy-to-predict links at small λ values but also harder to predict links in Ω_M at intermediate and large λ values.

In summary, AUC and AUPR scores complement each other by weighing missing link candidates in Ω_M differently. Thus, in our paper we compute both metrics to obtain a comprehensive view on the utility of hyperbolic geometry in link prediction. In addition to AUPR and AUC scores, we also compute Precision scores, which are the scores to use if the number of missing links is known exactly, although such knowledge is rarely the case in practice.

III. LINK PREDICTION WITH KNOWN COORDINATES

Before investigating link prediction accuracy in real networks, we conduct link prediction experiments with RHGs with known coordinates. In doing so we pursue several goals. The RHGs provide the upper bound for link prediction accuracy of the HYPERLINK if the same node coordinates are used both for the graph construction and for link prediction [24], so that we want to quantify this upper bound. Second, we want to measure link prediction accuracies of other methods, listed in Appendix B, and compare them to that of the HYPERLINK. Establishing these results provides a baseline for interpreting link prediction results on real networks. To achieve these goals, we first calculate analytically the AUC and AUPR in RHGs with known coordinates and with coordinates disturbed by noise of varying magnitude. The latter result allows us to quantify in a controlled environment the level of coordinate inaccuracy beyond which the HYPERLINK becomes essentially impuissant.

We start with the analysis of HYPERLINK accuracy in the case of randomly missing links in RHGs. After the generation of an RHG we visit each of its links and remove it with probability $1 - q$, arriving at a pruned network. We then rank missing link candidates using distances between all unconnected node pairs calculated with coordinates from which the network was originally generated.

As seen in Fig. 3, the predictive power of the HYPERLINK is maximized as $T \to 0$ and decreases as T increases. This result is expected. In the $T \to 0$ limit the RHG is deterministic since the connection probability in Eq. (6) becomes the Heaviside step function, $p(x) \to \Theta(R - x)$. As a result, all node pairs with $x < R$ are connected and other node pairs are not. Then, an unconnected pair of nodes at distance $x < R$ is guaranteed to be a true positive and all unconnected pairs at $x \geqslant R$ are true negatives. As T increases, connections are allowed at distances $x > R$ with increasing probability and, as a result, underlying geometry plays a smaller role in the formation of links, explaining the decreasing link prediction accuracy as a function of T, as quantified by all scores in Fig. 3.

Even though all scores, AUC, AUPR, and Precision, are decreasing functions of T, they behave differently. AUC scores remain constant in the $T \in (0, \frac{1}{2})$ interval and then exhibit a slow decay to AUC $= 0.95$ at $T = 0.9$. AUPR and Precision scores, on the other hand, decrease rapidly in the entire testing interval of $T \in [0, 0.9]$ from AUPR $= 1$ (Precision $= 1$) at $T = 0$ to AUPR $= 0.34$ (AUPR $= 0.29$) at $T = 0.9$.

We can predict these results analytically as we explain next.

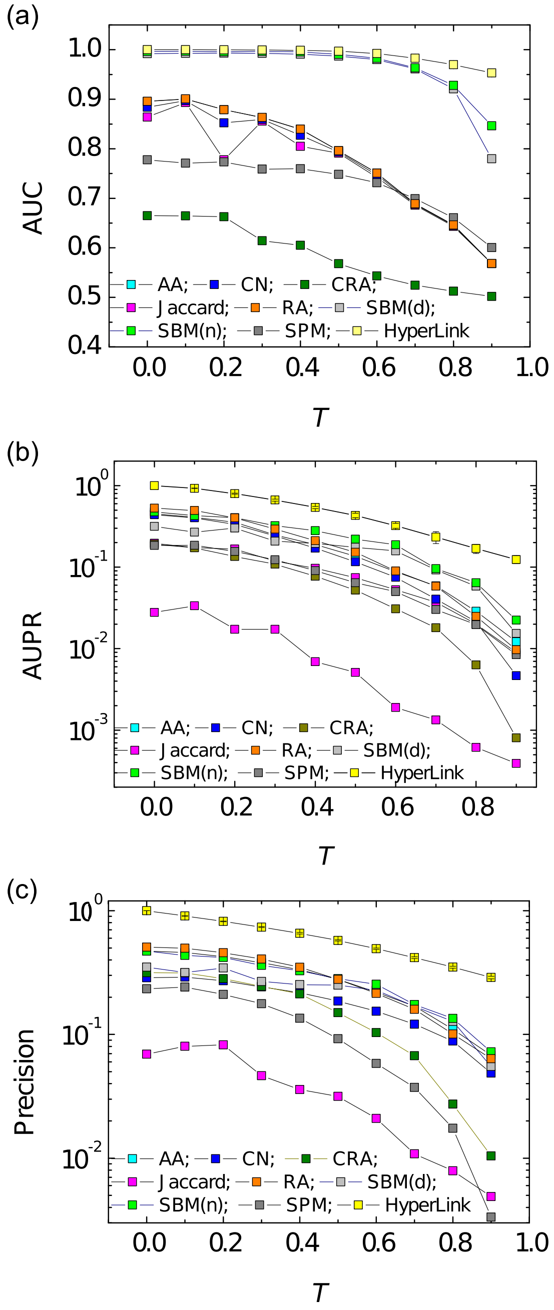

FIG. 3. Link prediction on RHGs with known coordinates. In all experiments we remove links uniformly at random with probability $1 - q = 0.5$. Then missing links are predicted using hyperbolic distances between unconnected node pairs. Link prediction accuracy is quantified using (a) AUC, (b) AUPR, and (c) Precision scores plotted as a function of RHG temperature T. All results correspond to RHGs with $N = 10^4$ nodes, $\gamma = 2.5$, and $\bar{k} = 10$. The HYPERLINK link prediction scores are compared to those of AA, CN, CRA, Jaccard, RA, SBM(d,n), and SPM methods (see Appendix B).

A. AUC

The AUC score in RHGs is

$$\mathrm{AUC} = \int_0^{2R} \mathrm{tpr}(x)\mathrm{fpr}'(x)dx, \tag{26}$$

where tpr(x) and fpr(x) are, respectively, distance-dependent true positive and false positive rates among node pairs separated by distances up to x. As seen from Fig. 4(a), the true positive rate grows exponentially for $x < R$ and saturates to tpr$(x) = 1$ as x approaches $2R$. This observation

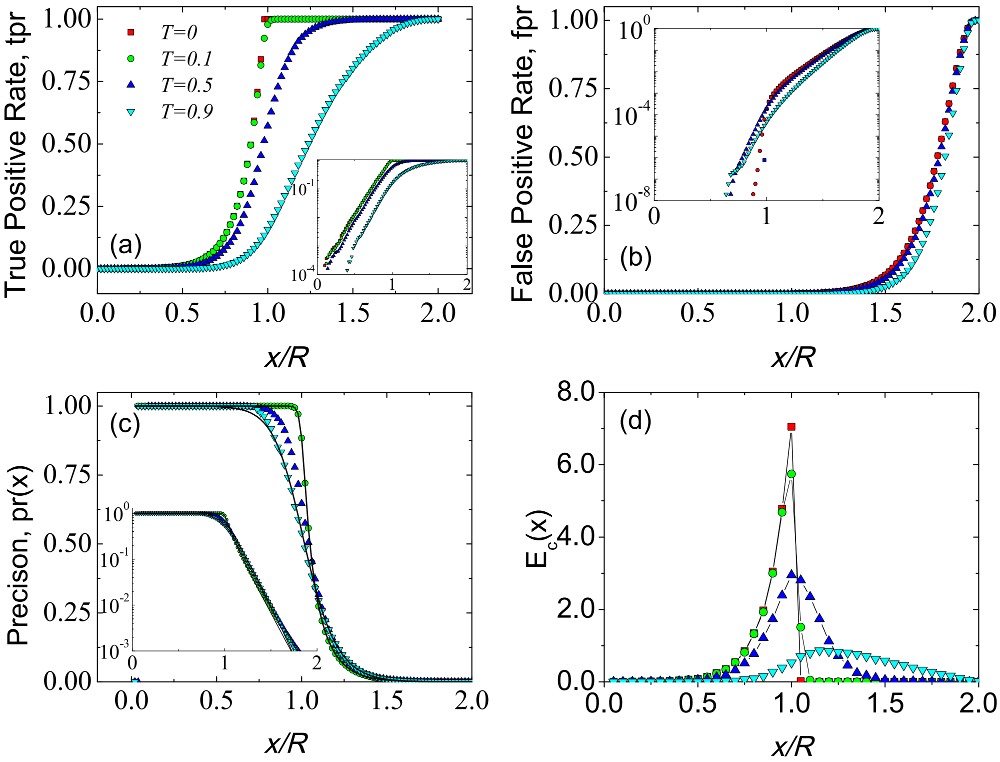

FIG. 4. Link prediction with known coordinates: (a) true positive rate tpr(x), (b) false positive rate fpr(x), (c) Precision pr(x), and (d) link density $n(x)p(x)$ as a function of hyperbolic distance x. In all experiments we remove links uniformly at random with probability $1 - q = 0.5$. Then missing links are predicted using hyperbolic distances between unconnected node pairs. All results correspond to RHGs with $N = 10^4$ nodes, $\gamma = 2.5$, and $\bar{k} = 10$. The insets display the same plots as the main panels but in log-linear format. Solid lines correspond to analytical estimates.

is easy to predict analytically. Let $n(x)$ be the distribution of hyperbolic distances x between node pairs in the RHG. It follows from the results in [50] that $n(x)$ can be approximated as

$$n(x) = \frac{4\alpha^2}{\pi(2\alpha - 1)^2} e^{x/2-R} \tag{27}$$

for $\alpha > \frac{1}{2}$. To be more specific, we note that R in the RHG is a function of network size N, given by Eq. (C3), and the approximation in Eq. (27) holds in the large N limit for any $x = cR$, where constant $c \in (0, 2)$, $\lim_{N\to\infty} \frac{n(x)}{n^{\text{true}}(x)} = 1$. Henceforth, we say $f(x) \approx g(x)$ if $\lim_{N\to\infty} \frac{f(x)}{g(x)} = 1$, and, more generally, $f(x) \sim g(x)$ if $\lim_{N\to\infty} \frac{f(x)}{g(x)} = K \neq 1$.

The connection probability $p(x)$ is close to unity for $x < R$, so that the number of true positives for $x < R$ grows proportional to the number of node pairs $N(x) \equiv \int_0^x n(y)dy$ in the hyperbolic disk, $\text{tpr}(x) \sim N(x) \sim e^{\frac{x}{2}}$ for $x < R$. In the $x > R$ regime connection probability $p(x)$ decays exponentially as $p(x) \sim e^{-\frac{x}{2T}}$ faster than the exponential growth of $n(x)$, leading to the saturation of $\text{tpr}(x) = 1$, Sec. III D.

The false positive rate remains negligible for $x < R$ and grows exponentially for $x \in (R, 2R)$, Fig. 4(b). We explain this observation using similar arguments. Since $p(x)$ is close to unity for $x < R$, and all unconnected node pairs with $x < R$ are almost guaranteed to be true positives, the false positive rate is negligible for $x < R$. In the $x > R$ regime, $p(x) \sim e^{-\frac{x}{2T}}$, and the number of unconnected node pairs is proportional to $N(x)$, resulting in $\text{fpr}(x) \sim e^{\frac{x}{2}}$ for $x > R$, Sec. III D.

Taken together, tpr(x) and fpr(x) rates provide a qualitative explanation for nearly perfect AUC scores observed in Fig. 3(a). The false positive rate fpr(x) takes large values only when x approaches $2R$. At the same time, as x approaches $2R$ the tpr(x) approaches 1.

Supporting this rough estimation, our more detailed analytical calculations in Sec. III D show that the AUC scores for RHGs with known coordinates converge to 1 in the large N limit as

$$1 - \text{AUC} \begin{cases} \sim N^{-1} & \text{if } T \in \left[0, \frac{1}{2}\right), \\ = \mathcal{O}\left(\frac{\ln N}{N}\right) & \text{if } T = \frac{1}{2}, \\ = \mathcal{O}\left(N^{1-\frac{1}{T}}\right) & \text{if } T \in \left(\frac{1}{2}, 1\right). \end{cases} \tag{28}$$

B. AUPR

To calculate the AUPR score we need to calculate the distance-dependent precision and recall rates pr(x) and rc(x) because

$$\text{AUPR} = \int_0^{2R} \text{pr}(x)\text{rc}'(x)dx. \tag{29}$$

Since $p(x)$ is close to 1 for $x < R$, all unconnected node pairs at $x < R$ are true positives, resulting in $\text{pr}(x) = 1$ [see Fig. 4(c) and Sec. III D]. The precision rate decays exponentially for $x > R$ since the true positive rate $\mathrm{tpr}(x)$ approaches 1 for $x > R$, while the number of unconnected node pairs $N_d(x)$ grows exponentially, $\binom{N}{2}\int_0^x n(y)[1-qp(y)]dy \sim e^{\frac{x}{2}}$ [see Fig. 4(c) and Sec. III D].

The dependence of AUPR on T arises from the recall function or its derivative, $rc'(x)$, quantifying the expected distance-dependent link density and, consequently, the density of missing links. As T increases, the missing links are more likely to be located at larger distances, Fig. 4(d), where precision $\mathrm{pr}(x)$ is smaller, resulting in lower AUPR scores, consistent with our observations in Fig. 3.

We also note that the AUPR score depends weakly on the node density parameter α and consequently on the degree distribution exponent $\gamma = 2\alpha + 1$. Indeed, the precision and recall rates depend on α only via the node pair distribution $n(x)$, Sec. III D, which depends on α only in subleading terms, as shown in [50].

C. Coordinate uncertainty and link prediction accuracy

While the HYPERLINK provides the upper bound for link prediction on RHGs, it is important to note that its accuracy is comparable to that of other link prediction methods, in particular, resource allocation (RA), adamic adar (AA), and stochastic block models SBM(d,n), Fig. 3. This observation motivates the question: *How accurately does one need to infer node coordinates to ensure the superior performance of the HYPERLINK?*

To answer this question we analyze the impact of node coordinate uncertainty on the HYPERLINK accuracy. To this end, we add synthetic noise to original angular node coordinates, while keeping radial node coordinates unchanged:

$$\hat{\theta}_i \leftarrow \theta_i + aX_i, \tag{30}$$

$$X_i \leftarrow U\left(-\tfrac{1}{2}, \tfrac{1}{2}\right), \tag{31}$$

where $a > 0$ is the noise amplitude. The effects of synthetic noise on the HYPERLINK accuracy are depicted in Fig. 5. Our results indicate that AUPR and Precision scores, Figs. 5(b) and 5(c), decrease rapidly as a function of noise amplitude, while AUC scores remain largely unchanged even at $a > 1$ rad values.

To better understand the effects of noise on link prediction accuracy we juxtapose HYPERLINK prediction results to those of the RA method, which is one of its leading competitors according to Fig. 3. We show RA accuracy with dashed lines of matching color in Fig. 5. Consistent with our earlier observations we find that HYPERLINK AUC scores are robust to noise, preserving its leading ranking among other link prediction methods, Fig. 5(a).

In contrast, as quantified by AUPR and Precision scores, the HYPERLINK is superior to the RA method only if coordinate uncertainty is sufficiently small. The maximum tolerable noise amplitude value a_c increases as T increases [see the inset of Figs. 5(b) and 5(c)]. While noise amplitude a does not exceed 10^{-2} rad in the case of $T = 0.1$, the noise tolerance in the case of $T = 0.9$ is significantly higher, $a_c \approx 0.5$ rad, suggesting, somewhat surprisingly, that the HYPERLINK is better off on networks characterized by larger T values or, equivalently, smaller clustering coefficient.

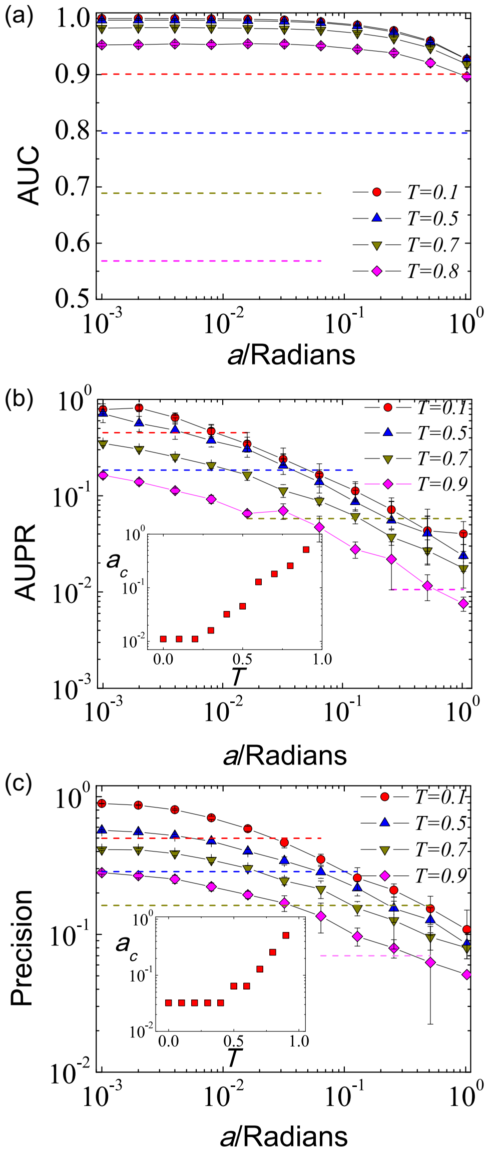

FIG. 5. Effects of synthetic noise on link prediction accuracy. HYPERLINK accuracy quantified using (a) AUC, (b) AUPR, and (c) Precision scores as a function of noise amplitude a for RHGs with different T values. All results correspond to RHGs with $N = 10^4$ nodes, $\gamma = 2.5$, and $\bar{k} = 10$. The HYPERLINK accuracy is compared to that of RA, i.e., its top competitor according to Fig. 3. Corresponding scores of the RA index are shown with dashed lines of matching color. The insets of panels (b) and (c) display the maximum tolerable coordinate noise amplitude as a function of T, i.e., the values of a corresponding to equal HYPERLINK and RA accuracy.

Qualitatively, the observed fast degradation of the AUPR and Precision scores is due to the sensitivity of the hyperbolic distance to the angular distance between the nodes $\Delta\theta$. It follows from Eq. (3) that even a small change in $\Delta\theta$ may significantly change the corresponding hyperbolic distance, adversely affecting the ranking of missing link candidates at small distances x, Appendix E. Since AUPR and Precision emphasize link prediction accuracy of most likely candidates, proper ranking of unconnected node pairs at small x values is crucial. AUC scores, on the other hand, place more emphasis on less obvious link candidates and are less affected by coordinate uncertainty. We find that the uniform synthetic noise adversely affects distance dependent true positive rate $\mathrm{TP}(x|a)$, which scales as

$$\mathrm{TP}(x|a) \sim \begin{cases} a^{1-2\gamma} & \text{if } x \leqslant R, \\ a^{1-2\gamma}\left(R + 2\ln\frac{a}{2}\right) & \text{if}x > R \end{cases} \tag{32}$$

(see Appendix E), leading to

$$\mathrm{AUPR}(a) \sim a^{2-4\gamma}\left(R + 2\ln\frac{a}{2}\right)^2. \tag{33}$$

The robustness of the AUC scores to synthetic noise in RHGs can be qualitatively explained by the fact that AUC scores emphasize the prediction of missing links at large x distances. Large hyperbolic distances are affected by synthetic noise to a lesser extent than small hyperbolic distances. This effect follows directly from Eq. (3) and can be observed in

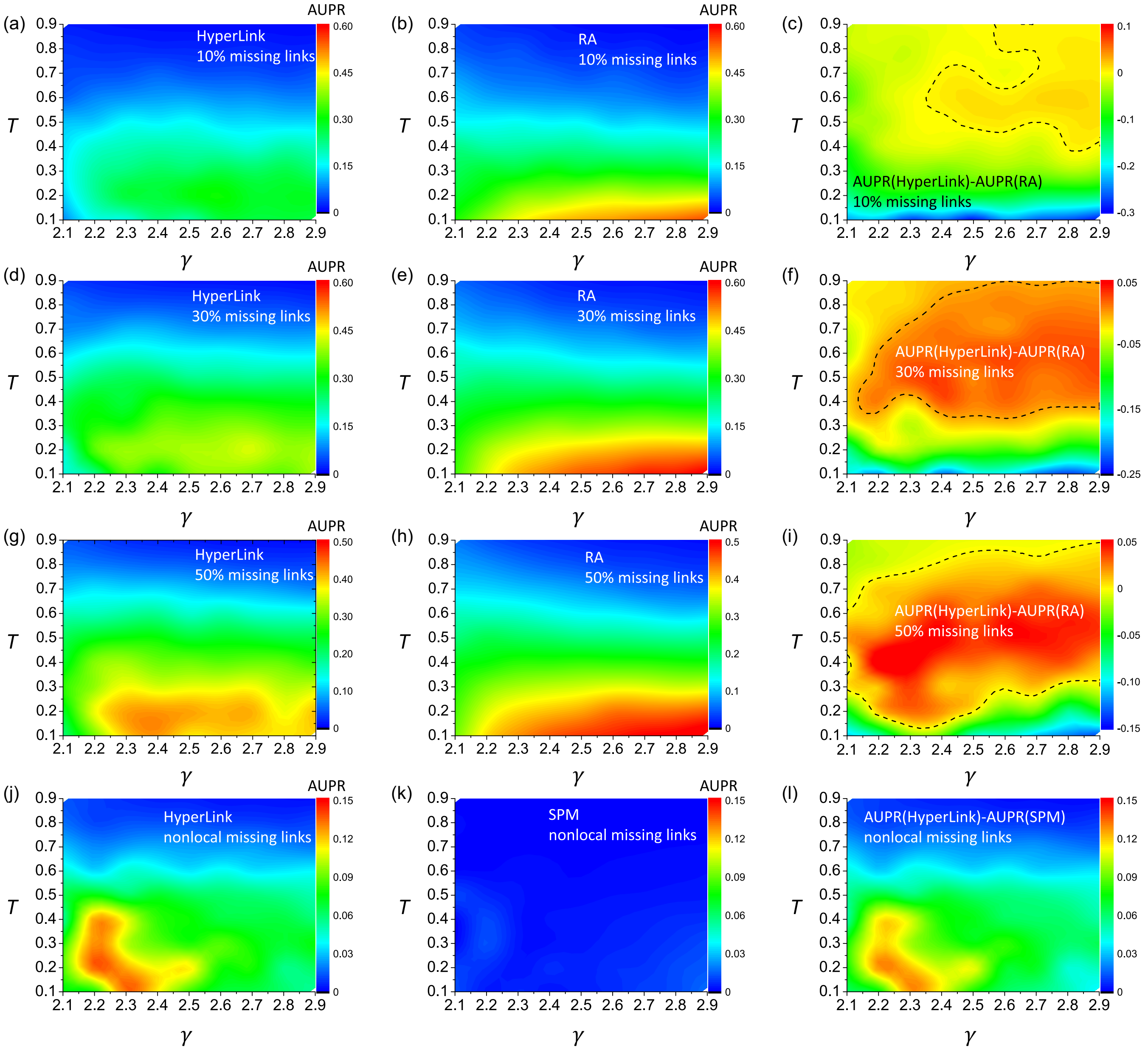

FIG. 6. Link prediction accuracy for RHGs with inferred coordinates: AUPR. (a)–(i) Random missing links. (j)–(l) Nonlocal missing links. Each panel is a heatmap displaying AUPR values as functions of T and $\gamma = 2\alpha + 1$ parameters of the RHG. We compare link prediction accuracy of the HYPERLINK to that of the RA and SPM methods, which are its leading competitors in cases of randomly missing links and nonlocal missing links, respectively. In each random missing link experiment links are removed uniformly at random with prescribed probabilities: (a)–(c) $1 - q = 0.1$, (d)–(f) $1 - q = 0.3$, and (g)–(i) $1 - q = 0.5$. (a), (d), (g) AUPR values for HYPERLINK. (b), (e), (h) AUPR values for RA. (j)–(l) AUPR values of HYPERLINK and SPM, as well as their difference, for nonlocal links, i.e., links connecting nodes with no common neighbors, which comprise a subset of randomly removed links with $1 - q = 0.5$. The dashed curves in panels (c), (f), (i), and (l) denote the regions in the γ-T parameter space where the HYPERLINK accuracy is higher than that of the competitive method.

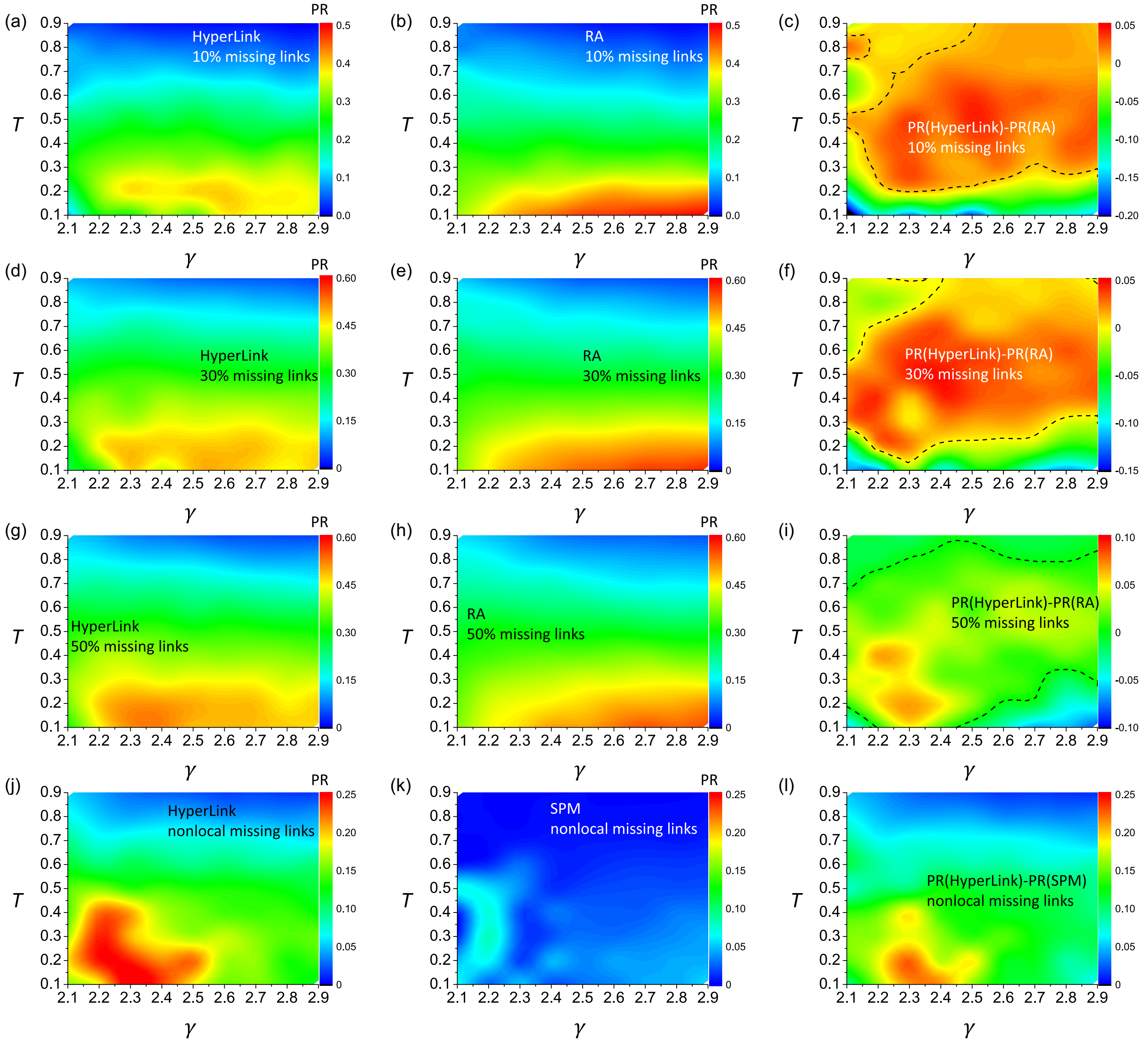

FIG. 7. Link prediction accuracy for RHGs with inferred coordinates: Precision. The legend is identical to that of Fig. 6.

Fig. 14(a), displaying the saturation of $\mathrm{TP}(x|a) \to 1$ as x approaches $2R$, regardless of noise amplitude a.

Our conclusions in this section are different for AUC and AUPR/Precision metrics.

The AUPR and Precision metrics emphasize prediction of the most likely missing link candidates and are highly sensitive to the accuracy of node coordinate inference. Synthetic noise added to original node coordinates smears hyperbolic distances among missing link candidates, adversely affecting the HYPERLINK accuracy. Our results suggest that one needs to maximize the accuracy of the network mapping in order to efficiently predict missing links. We also find that, as temperature T increases, the performance of other link prediction methods, as measured by AUPR and Precision, decreases faster than that of the HYPERLINK, suggesting that the latter has a competitive advantage on networks characterized by large T values.

AUC scores, on the other hand, emphasize less obvious link candidates that correspond to node pairs at larger hyperbolic distances. Since larger hyperbolic distances are affected by coordinate uncertainty to a lesser extent, the AUC scores of the HYPERLINK are robust to synthetic noise, suggesting that HYPERLINK is capable of predicting less obvious missing links even under less accurate mapping conditions.

IV. LINK PREDICTION WITH INFERRED COORDINATES

In this section we build upon our results obtained in the previous section to analyze the HYPERLINK accuracy on networks with unknown node coordinates. We first conduct systematic analysis of HYPERLINK accuracy on RHGs with unknown node coordinates and then apply HYPERLINK to several real networks. In both cases network coordinates are unknown

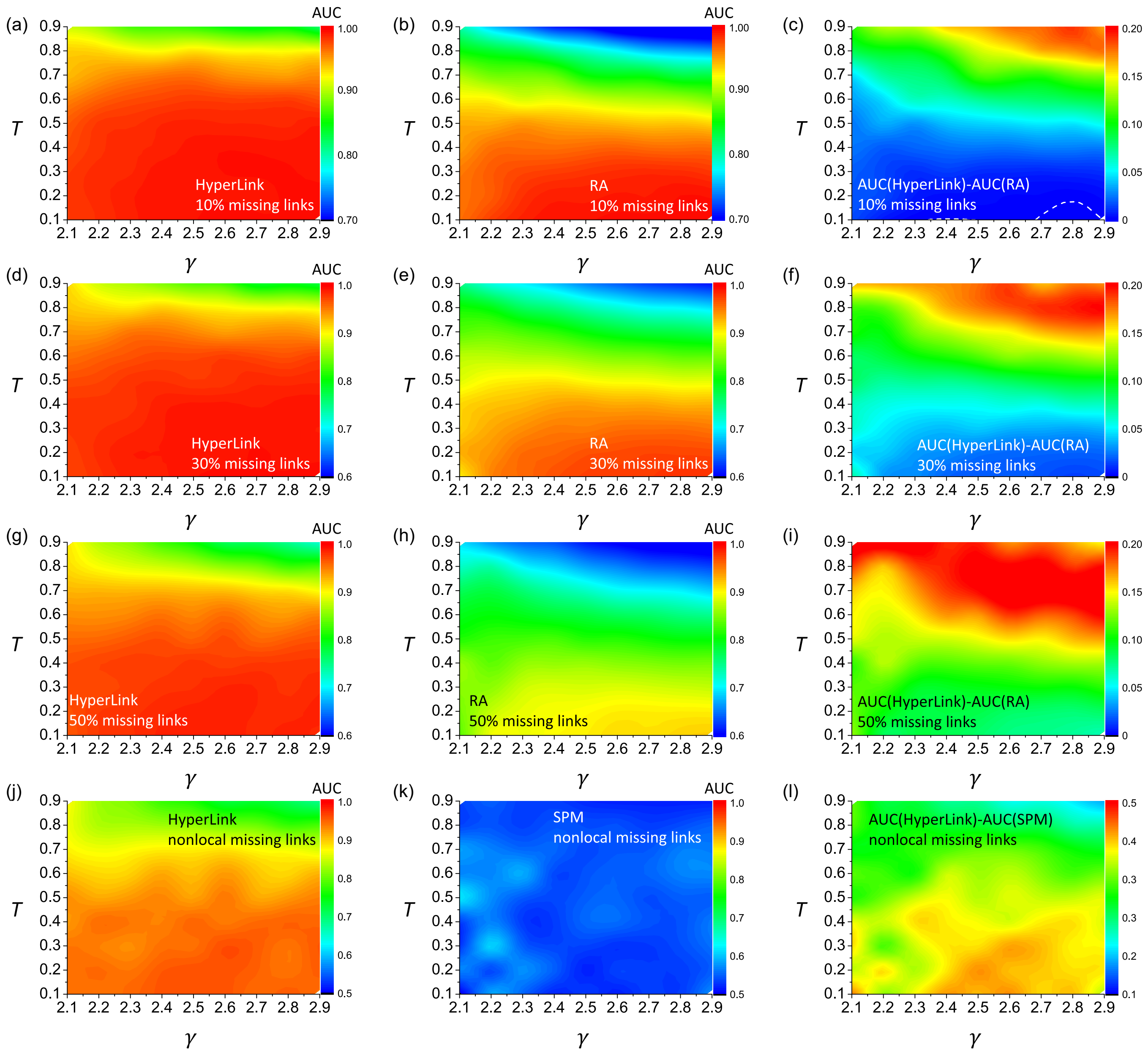

FIG. 8. Link prediction accuracy for RHGs with inferred coordinates: AUC. The legend is identical to that of Fig. 6.

and in order to predict missing links we first infer node coordinates by mapping networks of interest to the two-dimensional hyperbolic disk. To this end, we developed a mapping algorithm, which is tailored to the link prediction problem. This algorithm is referred to as the HYPERLINK embedder and is fully described in Appendix F.

A. Tests on RHGs with inferred coordinates

To evaluate the HYPERLINK accuracy on RHGs with unknown node coordinates we perform the following experiments. After generating an RHG we remove a fraction of existing missing links. As before, each existing link is removed with probability $1 - q$. Occasionally, after links are removed, the remaining network splits into several components. If this is the case, we limit our consideration to the largest connected component of the pruned network. We refer to the resulting connected component of the pruned network as the training network. To predict missing links we erase our knowledge of the true node coordinates and then infer node coordinates by mapping the training network to the hyperbolic disk using the HYPERLINK embedder (see Appendix F for details on the mapping procedure). After the mapping is complete, we use the inferred node coordinates to calculate distances between all unconnected node pairs in the training network and rank these pairs in the increasing order of distance.

Figures 6, 7, and 8 show the results for the AUPR, Precision, and AUC scores, respectively. Each panel in these

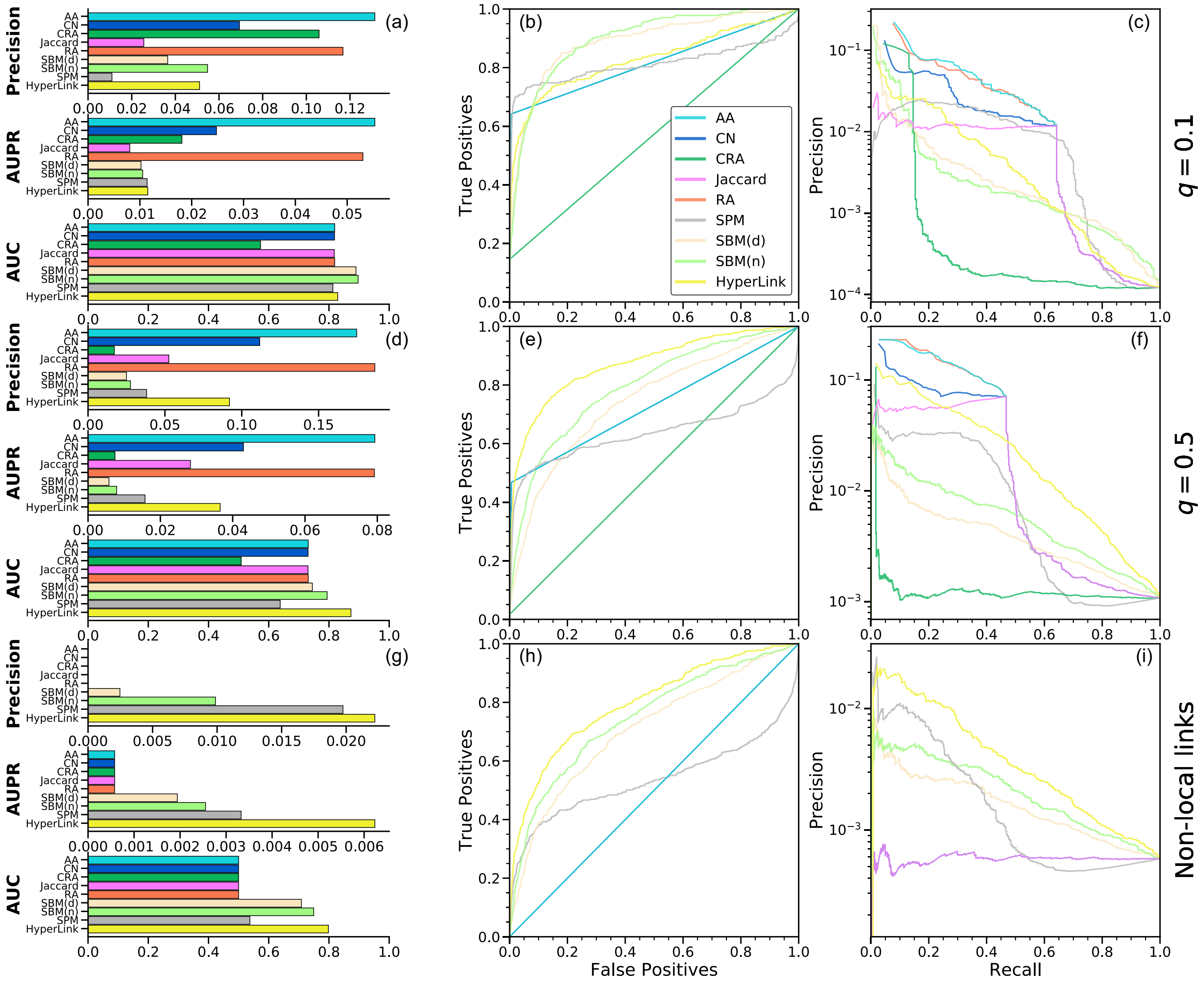

FIG. 9. Link prediction accuracy for the Metabolic network with (a)–(c) 10% ($q = 0.9$) randomly missing links, (d)–(f) 50% ($q = 0.5$) randomly missing links, and (g)–(i) nonlocal missing links, i.e., links connecting node pairs that have no common neighbors. Nonlocal links constitute 20% of the $q = 0.5$ missing links set. (a), (d), (g), (j) Precision, AUPR, and AUC link prediction scores. (b), (e), (h), (k) ROC curves. (c), (f), (i), (l) PR curves.

figures is a heatmap, aggregating the link prediction accuracy scores for RHGs with different $\gamma \in [2.1, 2.9]$ and $T \in [0.1, 0.9]$ values, which we change with an increment of 0.1 each. We compare the HYPERLINK to the RA method, which is its leading competitor in these experiments [cf. Figs. 3(a) and 3(b)].

The results for the AUPR and Precision scores are similar. Quantified by these scores, the HYPERLINK accuracy is nearly independent of the degree distribution exponent γ, and at the same time decreases rapidly as temperature T increases [see panels (a), (d), and (g) in Figs. 6 and 7]. This observation is consistent with our theoretical analysis in Sec. III, where we establish that AUPR scores decrease as T increases and do not strongly depend on γ.

Even though RA performs similar to HYPERLINK [panels (b), (e), and (h) in Figs. 6 and 7], we note that RA is more accurate at lower T values and less accurate than HYPERLINK for higher T values. To obtain the direct comparison of the two methods we plot the difference between their AUPR (Precision) scores in panels (c), (f), and (i) in Figs. 6 and 7. In agreement with our theoretical considerations in Fig. 5, we find that the HYPERLINK is superior to RA in the region of γ-T phase space corresponding to higher T values; these regions are denoted with dashed lines in panels (c), (f), and (i) in Figs. 6 and 7.

Compared to RA, the HYPERLINK yields better link prediction accuracy for larger fractions of missing links. In the case $1 - q = 0.1$, for instance, HYPERLINK is better than RA in a small upper right corner region of the γ-T phase space, Figs. 6(c) and 7(c). On the other hand, in the case 50% of links are missing, $1 - q = 0.5$, the HYPERLINK outperforms RA for the majority of γ-T values with the exception of smallest, $T = 0.1$, and largest, $T = 0.9$, temperature values, Figs. 6(i) and 7(i).

The better, compared to RA, performance of the HYPERLINK in Figs. 6(i) and 7(i) is the result of two effects. On

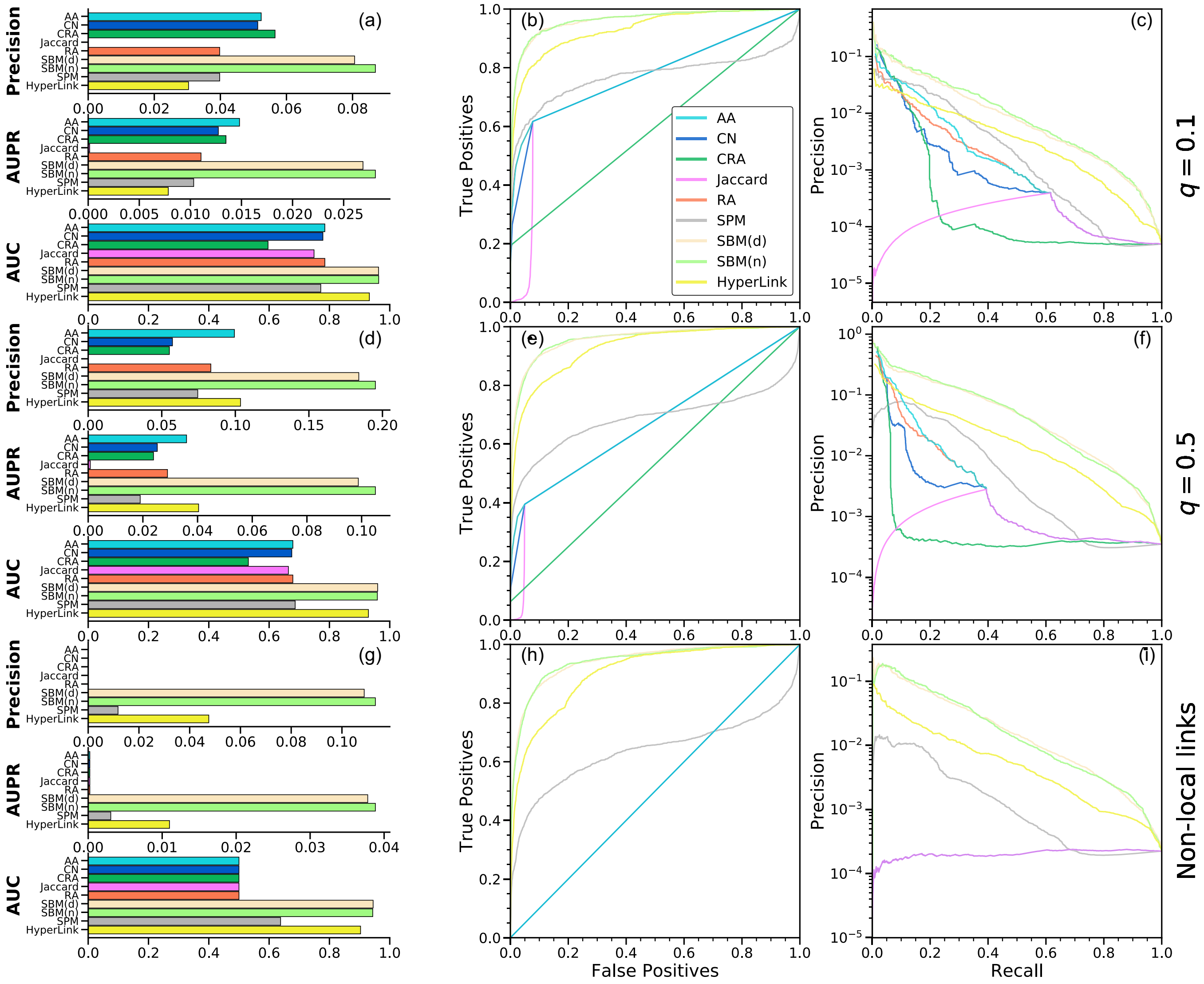

FIG. 10. Link prediction accuracy for the Internet. Nonlocal links constitute 32% of the $q = 0.5$ missing links set.

one hand, the HYPERLINK accuracy appears to increase as $1 - q$ increases. This effect is consistent with a recent observation in [24] that the upper bound of link predictability in edge-independent graphs increases with $1 - q$. On the other hand, as $1 - q$ increases, the accuracy of RA decreases. RA, as well as other similarity-based methods, e.g., RA, Cannistraci resource allocation (CRA), AA, common neighbors (CN), and Jaccard's index (JC), predict missing links based on the similarity of node neighborhoods, e.g, the number of common neighbors; the higher the similarity the higher the probability of a missing link, Appendix B. Neighborhood similarities are local measures, reflecting network structure in the network-based vicinity of the node pair of interest, and ignoring the structure of the remaining network. The larger the fraction of missing links, the smaller the fraction of links in the training network and, as a result, the poorer the link prediction results. While this is true for all link prediction methods, the similarity-based methods are the ones that suffer most. Since links are established independently in RHGs, and each link is removed with probability $p = 1 - q$, the number of common neighbors between any node pairs on average decreases proportionally to p^2. All extensive RHG properties, on the other hand, depend on p linearly. HYPERLINK as a global method uses the structure of the entire network to map it, so that it is less sensitive to network incompleteness.

An attractive feature of a global method is that it is capable of predicting *nonlocal missing links*, i.e., links between node pairs with no common neighbors. To quantify HYPERLINK accuracy for nonlocal links we consider the subset of nonlocal links within the set of links removed with probability $1 - q = 0.5$, Figs. 6(j) and 7(j), which comprise from 20% (for $\gamma = 2.1$, $T = 0.9$) to 86% (for $\gamma = 2.9$, $T = 0.1$) of all removed links.

Similarity-based methods, RA, AA, CN, and JC, cannot predict nonlocal missing links since corresponding node pairs have no common neighbors at all and, consequently, have zero similarity. Therefore, in nonlocal link prediction experiments we compare HYPERLINK to the structural perturbation method (SPM) index, which is a global method and the leading competitor to HYPERLINK for nonlocal links. As seen in panels (k)

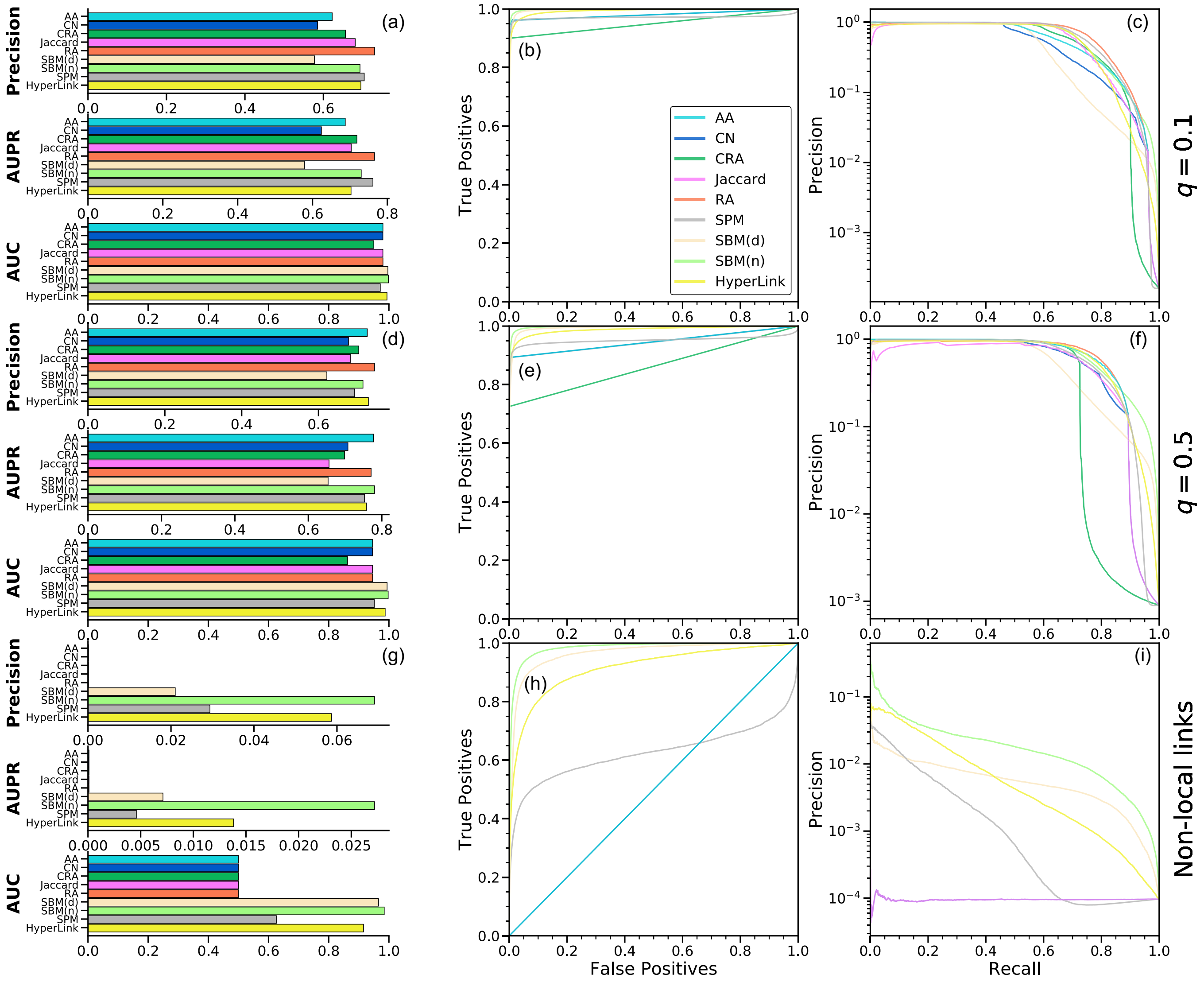

FIG. 11. Link prediction accuracy for the PGP network. Panels are identical to those of Fig. 9. Nonlocal links constitute 10% of the $q = 0.5$ missing links set.

and (l) in Figs. 6 and 7, the SPM index yields substantially lower link prediction accuracy than HYPERLINK for all the considered values of γ and T.

Overall, we observe that according to the AUPR and Precision scores HYPERLINK's competitive advantage is higher the more incomplete the network is, and the HYPERLINK is particularly strong in prediction of nonlocal links.

According to AUC scores, the HYPERLINK offers superior link prediction accuracy across the entire γ-T parameter space, surpassing its leading competitors—RA for all links, and SPM for nonlocal links, Fig. 8. This result is again consistent with our calculations in Sec. III showing that RHG-based AUC scores are robust with respect to coordinate uncertainty.

B. Tests on real networks

Finally, we apply the HYPERLINK to real networks: the network of human metabolism [51], the Internet at the autonomous system level [52], and the pretty-good-privacy (PGP) web of trust [53]. Basic properties of these networks as well as the data curation steps are documented in Appendix A.

Our link prediction experiments on real networks are performed identically to those on RHGs with inferred coordinates, and the results are shown in Figs. 9–11.

According to AUPR and Precision metrics, the HYPERLINK offers competitive performance in random link removal experiments, panels (a)–(f) in Figs. 9–11, but, at the same time, is not the most accurate. We do note that the relative performance of the HYPERLINK is better in cases of higher missing link rate, $1 - q = 0.5$, which is consistent with our results in Sec. III.

We also note that the HYPERLINK offers superior performance in prediction of nonlocal links where it is either the winner or runner-up, with the SBM methods being its leading competitors, panels (g)–(i) in Figs. 9–11. This observation comes in sharp contrast with nearly random performance of similarity based methods, RA, AA, CN, JC, and CRA, in nonlocal link prediction.

TABLE I. The summary of the results in Sec. III: HYPERLINK's measures of accuracy of link prediction in RHGs with known node coordinates as functions of the parameters in Sec. III.

Parameter	AUC	AUPR, Precision
Exponent $\gamma \in (2, 3)$	≈const	≈const
Temperature $T \in (0, 1)$	≈const	Decreasing
Fraction of missing links $1 - q$	Increasing	Increasing
Noise amplitude a	≈const	Decreasing

In contrast to AUPR-based rankings where the HYPERLINK is rarely the most accurate method, it is either the winner or runner-up in all the experiments according to the AUC metric, in agreement with all the AUC-related results above. In particular, it is the winner in predicting nonlocal links in the most challenging human metabolic network. This network is the most challenging because it is the sparsest and has the lowest clustering, Appendix A, thus providing the least amount of local information for link prediction.

V. SUMMARY, DISCUSSION, AND CONCLUSION

Tables I and II summarize the results in Secs. III and IV, respectively. We see that when it comes to predicting obvious missing links that are easy to predict employing hyperbolic geometry may be an overkill. In fact, one should consider using much simpler local methods instead of any global ones, according to the AUPR or Precision results presented here. This is because according to these results the local methods appear to be nearly as good as the global ones at predicting easy links. In particular, the HYPERLINK method cannot be the best at predicting the most obvious missing links because such links are the links between closest nodes in the latent hyperbolic space, and to rank them exactly at the top of the disconnected node pair list one has to infer the coordinates nearly exactly, Sec. III.

However, if the task is to identify missing links that are really hard to predict, then this is the situation where one should consider using global methods in general and hyperbolic geometry in particular. The most striking example is the prediction of missing links between the nodes that do not share any common neighbors. Here the HYPERLINK is either the winner or runner-up to the SBM methods, according to all the AUC, AUPR, and Precision measures, in all the considered real and synthetic networks. It is not surprising that local methods do a poor job in predicting such links—they are simply not designed to do so. In contrast, the HYPERLINK, SBM, and SPM are global methods that base their decisions on the global structure of the whole network, which helps enormously to predict nonlocal and other hard-to-predict links. The SBM and SPM methods were reported to outperform a vast collection of other methods [4,6,54]. Here we see that the HYPERLINK outperforms even these powerful methods in many cases. In particular, the HYPERLINK is the winner according to all the scores in the most challenging considered case, which is nonlocal links in the sparsest lowest-clustering network of human metabolic reactions.

TABLE II. The summary of the results in Sec. IV: HYPERLINK's measures of accuracy of link prediction in RHGs with inferred coordinates and in real networks, as well as those for nonlocal links, compared to other methods. The parameters are the same as in Table I.

Scenario	AUC	AUPR, Precision
RHGs with inferred coordinates	Winner	Winner if T, γ, or $1 - q$ is large
Real networks	Winner/runner-up	The more competitive, the larger the $1 - q$
Nonlocal links in RHGs and real networks	Winner/runner-up	Winner/runner-up

We also see that according to the AUC measure the HYPERLINK is either the winner or runner-up in all the considered situations. This is because the AUC does not care that much about false positives, and HYPERLINK achieves (nearly) the best balance between the true and false positive rates by finding missing links between highly dissimilar nodes located at large distances in the latent hyperbolic space.

We have also shown that the HYPERLINK is better off the weaker the clustering (the higher the T), and the larger the fraction of missing links $1 - q$ in RHGs with inferred coordinates. This does not mean that HYPERLINK's link prediction accuracy scores are getting better in these more difficult conditions; its scores do degrade. But the speeds of the degradation of these that the other methods experience are higher than HYPERLINK's.

Our results also resolve the controversy among earlier reports on link prediction using hyperbolic geometry [24,33–37]. These reports approached link prediction using different measures of link prediction accuracy. To reiterate, if applied to sparse networks, the AUPR emphasizes the prediction of a small fraction of the most likely missing links and, as a result, is extremely sensitive to inaccuracies in the node coordinate inference. On the other hand, the AUC is more robust to coordinate uncertainties as it emphasizes the prediction of less likely missing links between dissimilar nodes at large latent distances.

To maximize HYPERLINK's link prediction accuracy, we have developed a hyperbolic network mapping method, the HYPERLINK embedder, that maximizes the accuracy of coordinate inference. Its accuracy comes at the computational complexity cost of $O(n^2)$. While faster methods for hyperbolic mapping have been developed recently [55–59], an optimal balance between the accuracy and speed of hyperbolic mapping is still to be found. Ideally, it would be highly desirable to have a method that would be as accurate as at least the HYPERLINK embedder, and that would run in $O(n)$ time.

We emphasize that link prediction using latent hyperbolic geometry is expected to yield good results only if this geometry is there in a given network. That is, the network structure must be consistent with the existence of this geometry. It is well known that RHGs are characterized by sparsity, self-similarity, scale-free degree distributions, and strong clustering, meaning that these properties are necessary conditions

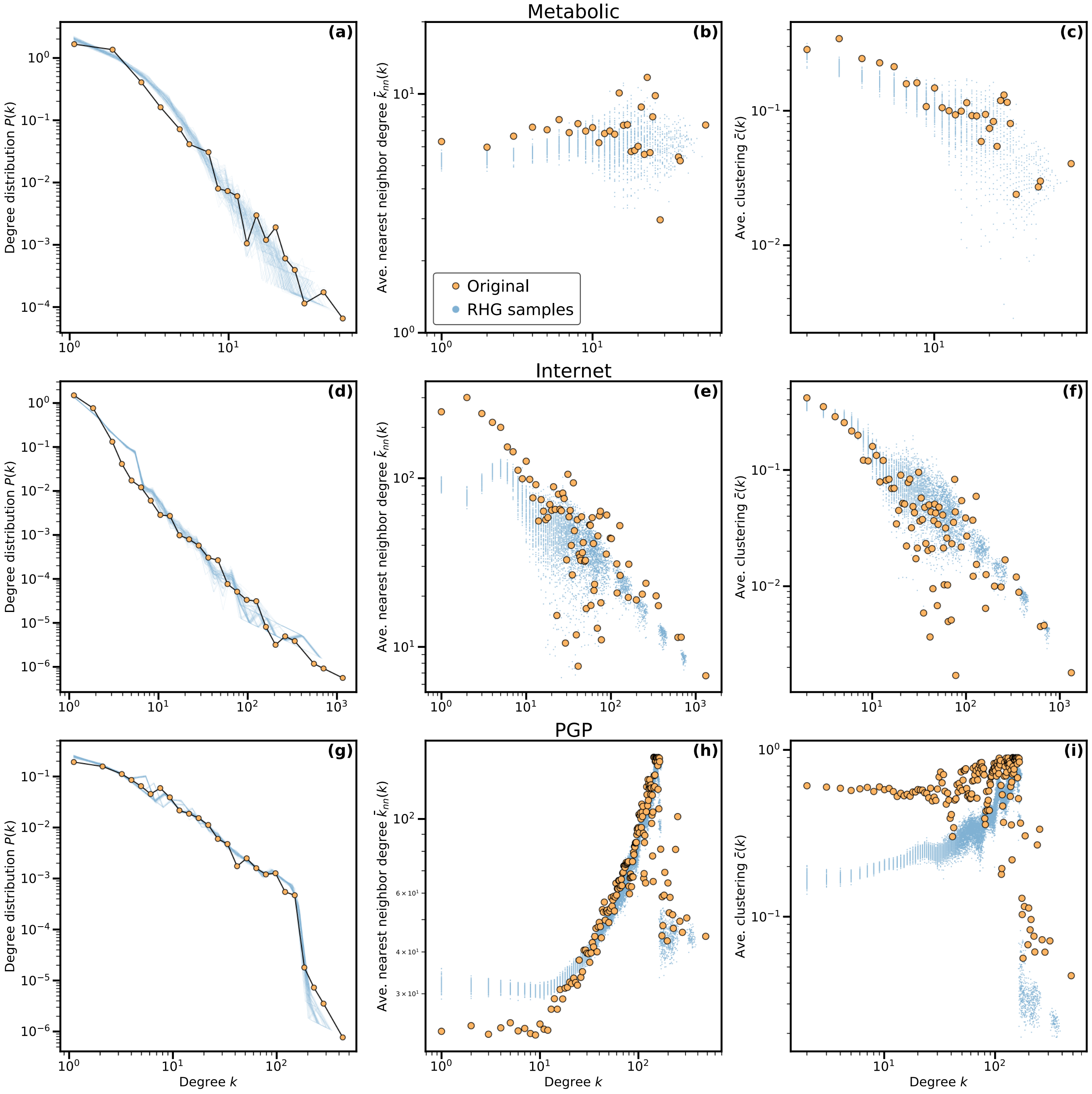

FIG. 12. The comparison of main structural properties of the three pruned real networks and corresponding 100 pruned RHGs generated using the hyperbolic coordinates learned by the HYPERLINK embedder. In both real and synthetic networks the pruning is the random link removal with rate $1-q=0.9$. To generate an RHG for a real network of interest we use its parameters R and T and node coordinates that are learned by the HL embedder. For each real network we generate 100 i.i.d. instances of RHGs by connecting node pairs with probabilities given by Eq. (F4), where $1=q=0.9$ and $p(x)$ is given by Eq. (6). (a), (d), (g) Degree distribution $P(k)$. (b), (e), (h) Degree-dependent average nearest-neighbor degree $\bar{k}_{\mathrm{NN}}(k)$. (c), (f), (i) Degree-dependent average local clustering coefficient $\bar{c}(k)$. Since RHG parameters R and T are inferred using the assumption of uniform angular coordinate distribution, $\rho(\theta)=1/2\pi$, which is not the case in real networks, we had to adjust the hyperbolic disk radius R in RHGs to match the average degrees in the pruned real and synthetic networks.

for hyperbolic geometry presence. It is also well known that many real networks do possess these properties as well. The results in [32] suggest that clustering is also a sufficient condition for network geometricity, but these results apply only to homogeneous large-world networks, and ignore coordinate entropy. That is, in theory, the detailed sufficient conditions for the presence of latent hyperbolic geometry are currently unknown, remaining a subject of ongoing research. Experimentally it is known, however, that random hyperbolic graphs are good descriptors of the structure of many real networks. In

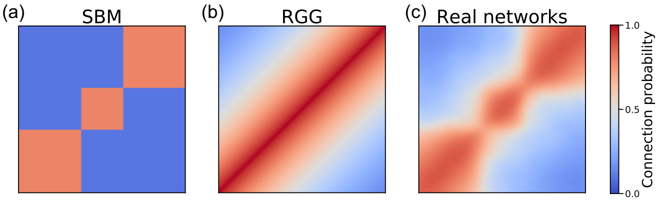

FIG. 13. Schematic illustration of the connection probabilities as functions of latent variables/coordinates of pairs of nodes in (a) the stochastic block model (SBM), (b) random geometric graphs (RGG), and (c) real networks.

particular, we are not aware of any other model capable of reproducing self-similarity of real networks, a highly nontrivial property [25]. As far as the more standard structural properties of real networks are concerned, the adequacy of hyperbolic geometry to model them has been documented many times, as early as in [26]. Here we report similar results in Fig. 12.

Overall, it appears that the harder a specific link prediction task the better the HYPERLINK is at this task. Yet the HYPERLINK is not always the winner even at such hard tasks. In particular, in application to real networks it is often a close runner-up to the stochastic block model methods. These results are consistent with the findings in [60], where the RHG and SBM were compared across a variety of properties. In the SBM the connection probability has a block structure, while in RHGs it is a function of the latent distance, Figs. 13(a) and 13(b). Clearly neither model can pretend to describe the connection probability in real networks exactly—at least because the RHGs have no communities, while the SBM has no clustering in the large-network limit. In view of the results in [60,61], the connection probability in real networks is likely to be some nontrivial mixture of the two pictures, Fig. 13(c), with geometry appearing as a mesoscopic structure gluing community blocks together. In short, the RHGs and SBM are complementary models capturing different aspects of the structure of real networks, and the link prediction accuracy of a model-based method depends on how prominent and prevalent the model's features are in a given real network.

ACKNOWLEDGMENTS

We thank R. Aldecoa, F. Papadopoulos, C. V. Cannistraci, M. Á. Serrano, T. Peixoto, and A. Clauset for useful discussions and suggestions. This work was supported by Army Research Office (ARO) Grant No. W911NF-17-1-0491 and National Science Foundation (NSF) Grant No. IIS-1741355. M. Kitsak was additionally supported by the NExTWORKx project.

APPENDIX A: REAL NETWORKS

1. Metabolic network

The metabolic network is based on the dataset of metabolic interactions of 107 organisms constructed by Ma and Zeng [51]. The original network is bipartite and consists of metabolites (top domain) connected to chemical reactions (bottom domain). We consider the unipartite projection of the network on the top domain. Basic properties of the metabolic network are summarized in Table III.

TABLE III. Basic properties of the considered real networks. N is the number of top nodes; E is the number of edges; $\bar{k}$ is the average degree; γ is the degree distribution exponent, which we estimated using methods from [64]; $\bar{c}$ is the average degree-dependent clustering coefficient; and T is the corresponding RHG temperature.

Network name	N	E	$\bar{k}$	γ	$\bar{c}$	T
Internet	6474	13 234	4.09	2.1	0.51	0.7
Metabolic network	2732	4040	2.96	2.9	0.29	0.6
PGP web of trust	14 138	160 080	22.65	2.1	0.66	0.8

2. Internet

The Internet network is a snapshot of the autonomous system level Internet taken from the University of Oregon Route Views Project [52]. The full dataset contains 733 daily instances which span an interval of 785 days from 8 November 1997 to 2 January 2000. Here we use a network instance as of 2 January 2000 [62].

3. PGP web of trust

PGP is a data encryption and decryption computer program that provides cryptographic privacy and authentication for data communication [53]. The data are collected and maintained by Cederlöf [63]. In the paper we use the PGP snapshot taken in April of 2003. The PGP web of trust is a directed network where nodes are certificates consisting of public PGP keys and owner information. A directed link in the web of trust pointing from certificate A to certificate B represents a digital signature by the owner of A endorsing the owner/public key association of B. We construct the undirected PGP graph by taking into account only bidirectional trust links between the certificates. Further, we only consider the giant connected component of the resulting undirected PGP web of trust network. Basic properties of the PGP network are summarized in Table III.

APPENDIX B: LINK PREDICTION—ALTERNATIVE METHODS AND SCORING TECHNIQUES

We compare the accuracy of the HYPERLINK link prediction method against the following set of link prediction methods: CN [65], AA [8], RA [66], CRA [67], JC [68], SPM [54], and SBM [4,6] methods.

All these methods, as well as the HYPERLINK, assign scores to (a subset of) all not directly connected pairs of nodes (nonlinks), and all such pairs are then ranked according to these scores from the most to least likely interaction prediction. To briefly describe these methods, it is thus sufficient to tell how these scores are calculated, for which we use the following notations: k_i is the degree of node i; $\Gamma(i)$ is the set of i's neighbors (directly connected nodes); $\gamma_{ij}(s)$ is the subset of all $\Gamma(s)$ that are neighbors of both i and j; e_i^j is i's *j-external degree*, the number of i's neighbors that are *not* j's neighbors; A is the network adjacency matrix.

1. Common neighbors

The score for a pair of nodes i and j is defined as the cardinality of the intersection of their sets of neighbors:

$$s_{ij}^{\mathrm{CN}} = |\Gamma(i) \cap \Gamma(j)|. \tag{B1}$$

2. Jaccard's index

The score is a normalized measure of the overlap of i's and j's sets of neighbors:

$$s_{ij}^{\mathrm{JC}} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|}. \tag{B2}$$

3. Adamic-Adar index

The score assigns more weight to the less-connected neighbors:

$$s_{ij}^{\mathrm{AA}} = \sum_{s\in\Gamma(i)\cap\Gamma(j)} \frac{1}{\log k_s}. \tag{B3}$$

4. Resource allocation index

The score is similar to the AA score, but punishes high-degree nodes more strongly:

$$s_{ij}^{\mathrm{RA}} = \sum_{s\in\Gamma(i)\cap\Gamma(j)} \frac{1}{k_s}. \tag{B4}$$

5. Cannistraci resource allocation index

The score is similar to the RA score, but takes into account the subset of nodes shared between nodes i, j and their common neighbors s:

$$s_{ij}^{\mathrm{RA}} = \sum_{s\in\Gamma(i)\cap\Gamma(j)} \frac{\gamma_{ij}(s)}{k_s}. \tag{B5}$$

6. Structural perturbation method

This method is based on repetitive perturbations of the adjacency matrix $\boldsymbol{A}$ by removals of small fractions of links that we denote by ΔE. The original adjacency matrix can then be written as $\boldsymbol{A} = \boldsymbol{A}' + \boldsymbol{\Delta A}$, where $\boldsymbol{A}'$ is the adjacency matrix of the network after removal of links ΔE, and $\boldsymbol{\Delta A}$ is the adjacency matrix constructed on the set of removed links ΔE. Denoting eigenvectors and eigenvalues of $\boldsymbol{A}'$ by x_k and λ_k, the perturbations of the original eigenvalues λ_k using the perturbation matrix $\boldsymbol{\Delta A}$ are

$$\Delta\lambda_k \approx \frac{x_k^T \boldsymbol{\Delta A} x_k}{x_k^T x_k}, \tag{B6}$$

so that the perturbed adjacency matrix is

$$\widetilde{\boldsymbol{A}} = \sum_{k=1}^{N} (\lambda_k + \Delta\lambda_k) x_k x_k^T. \tag{B7}$$

All nonlinks i, j are then ranked by $\widetilde{A}_{ij}$. In our experiments we repeat this perturbation procedure ten times, and then average perturbed matrices over these trials, thus obtaining an averaged perturbed matrix $\langle\widetilde{A}\rangle$, so that the SPM score is

$$s_{ij}^{\mathrm{SPM}} = \langle\widetilde{A}_{ij}\rangle. \tag{B8}$$

7. Stochastic block model

The stochastic block model is a generative network model designed to model community structure. Nodes are partitioned into groups (blocks) forming a node partition $\boldsymbol{b}$. The number of links between blocks is given by a matrix $\boldsymbol{e}$ the elements e_{rs} of which are the numbers of links between blocks r and s. If the observed degree sequence of a network $\boldsymbol{k}$ is used as an additional model parameter, the model is called *degree-corrected* SBM [69]. Moreover, if node blocks are themselves clustered into groups, and these groups are organized into higher-level groups, and so on recursively up to some nestedness level l, the model is called *nested* SBM. It can capture hierarchical and fine-grained structural properties of a given network [70]. In our experiments, we use both degree-corrected and nested SBMs denoted as *SBM(d)* and *SBM(n)* in the figures. We rely on the GRAPH-TOOL library [71] in the procedures below. Given the observed data (network adjacency matrix) $\boldsymbol{\mathcal{D}}$, and the prior probability density $P(\boldsymbol{A}, \boldsymbol{b})$ given by the network block structure $\boldsymbol{b}$ that produces a network with adjacency matrix $\boldsymbol{A}$ in the model, we reconstruct the full network using the posterior distribution:

$$P(\boldsymbol{A}, \boldsymbol{b}|\boldsymbol{\mathcal{D}}) = \frac{P(\boldsymbol{\mathcal{D}}|\boldsymbol{A})P(\boldsymbol{A}, \boldsymbol{b})}{P(\boldsymbol{\mathcal{D}})}, \tag{B9}$$

where $P(\boldsymbol{\mathcal{D}}|\boldsymbol{A})$ describes the measurement process of a network. We avoid computing the normalization factor $P(\boldsymbol{\mathcal{D}})$ by Markov Chain Monte Carlo (MCMC) sampling from the joint posterior distribution $P(\boldsymbol{A}, \boldsymbol{b}|\boldsymbol{\mathcal{D}})$ as described in [4]. In our experiments, we assume that each network link is observed and measured once. A possible link (i, j) then has marginal probability

$$\pi_{ij} = \sum_{\boldsymbol{A},\boldsymbol{b}} a_{ij} P(\boldsymbol{A}, \boldsymbol{b}|\boldsymbol{\mathcal{D}}). \tag{B10}$$

To sample over $(\boldsymbol{A}, \boldsymbol{b})$ configurations, the MCMC algorithm is initialized with the block structure obtained by the procedure from [72]. The MCMC is equilibrated using $10|E|$ equilibration steps, where E is the set of links in a given network. For the PGP network, due to its large size, we use only $2|E|$ MCMC equilibration steps. Then $T = 10$ epochs of MCMC iterations, 1000 swaps each, are performed to sample different block-network configurations. After each epoch, marginal link probabilities from Eq. (B10) are collected. These probabilities are then averaged over the epochs to obtain a single score used for link prediction:

$$s_{ij}^{\mathrm{SBM}} = \frac{1}{T} \sum_{t=1}^{T} \pi_{ij}^{(t)}. \tag{B11}$$

APPENDIX C: BASIC PROPERTIES OF THE RHG

The hyperbolic geometry inference algorithm relies on several properties of the RHG, which we review in this section.

1. Degree distribution

RHGs are characterized by scale-free degree distributions, $P(k) \sim k^{-\gamma}$, where $\gamma = 2\alpha + 1$. Indeed, the expected degree of a node located at (r, θ) is independent of its angular coordinate θ, $\bar{k}(r, \theta) = \bar{k}(r, 0) = \bar{k}(r)$, and is given by

$$\begin{aligned}\bar{k}(r) &= (N-1)\int dr'\rho(r')\int d\theta'\rho(\theta')p[x(r, 0, r', \theta')]\\ &\approx \frac{4N\alpha}{2\alpha-1}\frac{T}{\sin\pi T}e^{-r/2}\end{aligned} \tag{C1}$$

(see [20]). The average degree of the model is given by

$$\bar{k} = \int dr\rho(r)\bar{k}(r) = \frac{8N\alpha^2}{(2\alpha-1)^2}\frac{T}{\sin\pi T}e^{-R/2}. \tag{C2}$$

As seen from Eq. (C2), $\bar{k}$ in the most general case depends on the network size N.

To achieve sparse models with $\bar{k}$ independent of N one sets the radius of the hyperbolic disk to

$$R(N) = 2\ln(N/\nu), \tag{C3}$$

where $\nu > 0$ is the tuning parameter, directly related to $\bar{k}$. Indeed, with $R(N)$ given by (C3),

$$\bar{k} = \frac{8\nu\alpha^2}{(2\alpha-1)^2}\frac{T}{\sin\pi T}, \tag{C4}$$

prescribing the value of ν for the target values of $\bar{k}$, α, and T.

It has been shown in [73] that in the sparse limit the probability of a node located at (r, θ) to have k connections can be approximated with the Poisson distribution with the mean of $\bar{k}(r)$:

$$P(k|r) = e^{-\bar{k}(r)}\frac{[\bar{k}(r)]^k}{k!}. \tag{C5}$$

Then the degree distribution of the RHG is

$$P(k) = \int dr\rho(r)P(k|r) \sim k^{-\gamma}, \tag{C6}$$

$$\gamma = 2\alpha + 1. \tag{C7}$$

It follows from Eqs. (C1) and (C4) that model parameters α and R can be used to control degree distribution exponent γ and the average degree of the model, respectively.

a. Clustering coefficient

As seen from Eq. (6), connection probability $p(x)$ decreases exponentially for distances $x > R$ with the rate of $\frac{1}{2}T$. Thus, the temperature parameter T tunes the role of large distances in the formation of links: the higher the T the more likely are long-distance connections. As a result, T controls the clustering coefficient of the RHG. In the $T \to 0$ limit connections are only possible at hyperbolic distances $x < R$ and the clustering coefficient is maximized. Conversely, the clustering coefficient decreases as T increases and vanishes asymptotically in the $T \geqslant 1$ case [20].

APPENDIX D: HYPERLINK ACCURACY

In this section we calculate analytically the HYPERLINK accuracy, in terms of AUC and AUPR, on RHGs with known coordinates. Our results in this section are confirmed by the numerical experiments in Sec. III and build our intuition for Sec. IV, where we analyze HYPERLINK on RHGs and real networks with inferred coordinates.

1. AUC

To understand the behavior of AUC scores as a function of RHG parameters we define distance-dependent true positive $\mathrm{tpr}(x)$ and false positive $\mathrm{fpr}(x)$ rates as the fractions of true and false positives, respectively, contained among unconnected node pairs separated by distances up to x:

$$\mathrm{tpr}(x) = \frac{\mathrm{TP}(x)}{(1-q)E} = \frac{1}{E}\binom{N}{2}\int_0^x n(y)p(y)dy, \tag{D1}$$

$$\mathrm{fpr}(x) = \frac{\binom{N}{2}\int_0^x n(y)[1-p(y)]dy}{\binom{N}{2} - E}, \tag{D2}$$

where E is the true number of links in the network, $E = |\Omega_E \cup \Omega_R|$, $p(y)$ is the connection probability in the RHG given by Eq. (6), and $n(y)$ is the distance distribution for node pairs in the RHG, given by Eq. (27).

It is seen from Eqs. (D1) and (D2) that in the $T \to 0$ limit $p(y) = \Theta(R - y)$, resulting in $\mathrm{fpr}(x) = 0$ for $x \leqslant R$ and $\mathrm{tpr}(x) = 1$ for $x \geqslant R$, resulting in the ideal ROC curve, Fig. 2(a), and AUC = 1.

Using the expression for $n(y)$ from Eq. (27), we can evaluate true and false positive rates, up to the proportionality coefficient, as

$$\mathrm{tpr}(x) \approx \frac{4\alpha^2}{\pi(2\alpha-1)^2}\frac{N^2}{E}e^{-\frac{R}{2}}I(e^{\frac{x-R}{2}}; T), \tag{D3}$$

$$\mathrm{fpr}(x) \approx \frac{8\alpha^2}{\pi(2\alpha-1)^2}e^{-R}[e^{\frac{x}{2}} - e^{\frac{R}{2}}I(e^{\frac{x-R}{2}}; T)], \tag{D4}$$

where

$$I(z; T) \equiv \int_0^z \frac{dx}{1+x^{1/T}} = z\,{}_2F_1(1, T, 1+T, -z^{1/T}), \tag{D5}$$

and ${}_2F_1$ is the Gaussian hypergeometric function. In the $z \ll 1$ regime $I(z; T) \approx z$ and, thus, $\mathrm{tpr}(x) \sim e^{\frac{x-R}{2}}$ and $\mathrm{fpr}(x) \approx 0$ for $x < R$, Figs. 4(a) and 4(b).

In the $z \gg 1$ regime $I(z; T) \approx I(T)$, where $I(T) = \frac{\pi}{T\sin(\pi/T)}$, explaining the saturation of the true positive rate, $\mathrm{tpr}(x) \to 1$ as x approaches $2R$, and the exponential growth of the false positive rate, $\mathrm{fpr}(x) \sim e^{\frac{x}{2}}$ for $x > R$, Figs. 4(a) and 4(b).

To obtain the analytical estimate of the AUC as a function of RHG parameters we represent it as

$$\mathrm{AUC} = \int_0^{2R}\mathrm{tpr}(x)\mathrm{fpr}'(x)dx. \tag{D6}$$

By making use of Eqs. (D1) and (D2) we arrive at

$$\mathrm{AUC} = 1 - \Delta_1 - \Delta_2, \tag{D7}$$

$$\Delta_1 = \frac{E}{\binom{N}{2}}, \tag{D8}$$

$$\Delta_2 = -\frac{1}{E}\binom{N}{2}\int_0^{2R}[n^c(x)]^2 p'(x)dx, \tag{D9}$$

where $n^c(x) \equiv \int_0^x n(y)dy$.

In the case of sparse networks the first correction term $\Delta_1 \sim N^{-1}$ and can be ignored in the large N limit. The second correction term requires further analysis. It is straightforward to verify that in the $T \to 0$ limit $\Delta_2 \sim N^{-1}$ and can also be ignored. Indeed, in this case $p'(x) = -\delta(x-R)$, and

$$\Delta_2(T=0) = \frac{1}{E}\binom{N}{2}[n^c(R)]^2. \tag{D10}$$

Since $\binom{N}{2}[n^c(R)]$ equals the number of node pairs in the hyperbolic disk with distances up to R and all these node pairs are connected in the $T \to 0$ case, $\binom{N}{2}[n^c(R)] = E$, resulting in $\Delta_2(T=0) = \frac{E}{\binom{N}{2}} \sim N^{-1}$.

To estimate the behavior of Δ_2 in the case of $T > 0$ we need to understand the behavior of its integrand in Eq. (D9). Since $n^c(x) \sim e^{\frac{x}{2}}$ and $-p'(x) = \frac{1}{2T}\exp\left(\frac{x-R}{2T}\right)[p(x)]^2$, the integrand is sharply peaked at $x = R + 2T\ln\left(\frac{1+T}{1-T}\right)$ in the case of $T \in (0, \frac{1}{2})$, resulting in $\Delta_2 \sim N^{-1}$, similar to the $T \to 0$ case.

Conversely, the integrand in Eq. (D9) grows monotonously as a function of x in the case of $T \in (\frac{1}{2}, 1)$. The evaluation of Δ_2 in this regime is quite involved and is not informative. Instead, we elect to compute the upper bound for Δ_2, which also provides the lower bound for AUC scores. In doing so we note that the leading term behavior of $n(x)$ given by Eq. (27) is also its upper bound [50]. Then

$$\Delta_2 \leqslant \frac{2\alpha^2 e^{-R}\binom{N}{2}}{\pi T(2\alpha-1)^2 E}\int_0^{2R}\frac{e^{(x-R)\left(1+\frac{1}{2T}\right)}dx}{\left[1+e^{\frac{x-R}{2T}}\right]^2} \sim N^{1-\frac{1}{T}}, \tag{D11}$$

since $e^{\frac{R}{2}} \sim N$ in the case of sparse RHGs, Eq. (C3). In the case of $T = \frac{1}{2}$ Eq. (D11) simplifies to

$$\Delta_2 \leqslant \frac{4\alpha^2 e^{-R}\binom{N}{2}}{\pi(2\alpha-1)^2 E}\int_0^{2R}\frac{e^{2(x-R)}dx}{[1+e^{x-R}]^2} \sim \frac{\ln N}{N}. \tag{D12}$$

Taken together, the results above show that the AUC scores for RHGs with known coordinates converge to 1 in the large N limit as

$$1 - \text{AUC}\begin{cases} \sim N^{-1} & \text{if } T \in \left[0, \frac{1}{2}\right), \\ = \mathcal{O}\left(\frac{\ln N}{N}\right) & \text{if} T = \frac{1}{2}, \\ = \mathcal{O}\left(N^{1-\frac{1}{T}}\right) & \text{if} T \in \left(\frac{1}{2}, 1\right). \end{cases} \tag{D13}$$

2. AUPR

AUPR scores can be evaluated in a similar fashion:

$$\text{AUPR} = \int_0^{2R}\text{pr}(x)\text{rc}'(x)dx, \tag{D14}$$

where pr(x) and rc(x) are, respectively, distance-dependent precision and recall functions for hyperbolic distances up to x:

$$\text{pr}(x) \equiv \frac{\text{TP}(x)}{N_d(x)}, \tag{D15}$$

$$\text{rc}(x) \equiv \text{tpr}(x) = \frac{\text{TP}(x)}{(1-q)E}, \tag{D16}$$

where $N_d(x)$ is the number of disconnected node pairs with distances up to x:

$$N_d(x) = \binom{N}{2}\int_0^x n(y)[1-qp(y)]dy. \tag{D17}$$

Using Eqs. (D1) and (D17) we obtain

$$\text{pr}(x) = (1-q)\frac{\int_0^x n(y)p(y)dy}{\int_0^x n(y)[1-qp(y)]dy}, \tag{D18}$$

$$\text{rc}(x) = \frac{1}{E}\binom{N}{2}\int_0^x n(y)p(y)dy. \tag{D19}$$

In the $T \to 0$ limit pr$(x) = 1$ for all $x < R$, while rc$(x) = E(x)/E$, resulting, as expected, in AUPR $= 1$. Here $E(x)$ is the cumulative number of links between the node pairs with distances up to x.

In the $T > 0$ case we rely on Eqs. (D1), (27), and (D17) to obtain

$$\text{TP}(x) \approx \frac{4\alpha^2(1-q)}{\pi(2\alpha-1)^2}\binom{N}{2}e^{-\frac{R}{2}}I(e^{\frac{x-R}{2}}; T), \tag{D20}$$

$$\text{pr}(x) \approx \frac{(1-q)I(e^{\frac{x-R}{2}}; T)}{e^{\frac{x-R}{2}} - qI(e^{\frac{x-R}{2}}; T)}, \tag{D21}$$

where $I(z;T)$ is given by Eq. (D5).

In the $x \ll R$ regime $I(e^{\frac{x-R}{2}}; T) \sim e^{\frac{x-R}{2}}$ and pr$(x) \to 1$. In the $x \gg R$ case $I(e^{\frac{x-R}{2}}; T) \sim \frac{\pi}{T\sin(\pi/T)}$, and, as a result, precision decays exponentially, pr$(x) \sim e^{-x/2}$, independent of T, Fig. 4(c).

The dependence of AUPR on T arises from the recall function or its derivative, $rc'(x)$, quantifying the expected distance-dependent link density and, consequently, the density of missing links:

$$\text{rc}'(x) = \frac{1}{E}\binom{N}{2}n(x)p(x). \tag{D22}$$

$E_c(x)$ grows exponentially as $e^{x/2}$ for $x \ll R$ values and decays as $e^{x(1-\frac{1}{T})}$ for $x \gg R$, reaching the maximum at $x^* = R - 2T\ln\left(\frac{1}{T} - 1\right)$, Fig. 4(d). Thus, as T increases, the missing links are more likely to be located at larger distances where precision pr(x) is smaller, resulting in lower AUPR scores, consistent with the observations in Fig. 3.

APPENDIX E: EFFECTS OF COORDINATE UNCERTAINTY ON HYPERLINK ACCURACY

To understand the effects of coordinate uncertainties on HYPERLINK accuracy we model coordinate inference uncertainty as synthetic noise that we add to true angular coordinates of the RHG. In the following we first generate RHG as described in Sec. II B and then simulate uncertainties of angular coordinates by adding synthetic noise to original angular coordinates:

$$\hat{\theta}_i = \theta_i + aX_i, \tag{E1}$$

$$X_i \leftarrow U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \tag{E2}$$

where $a > 0$ is the noise amplitude. Further, we conduct link prediction experiments by calculating latent distances with uncertain coordinates:

$$\hat{x}_{ij} = x(r_i, \hat{\theta}_i, r_j, \hat{\theta}_j), \tag{E3}$$

where x is calculated according to the hyperbolic law of cosines, Eq. (3).

1. Link prediction with noise

In the case of synthetic noise, the AUPR scores are still given by Eq. (29) with effective precision and recall rates $\mathrm{pr}(y|a)$ and $\mathrm{rc}(y|a)$ evaluated in the presence of noise. To calculate these rates we start with the effective true positive rate $\mathrm{TP}(y|a)$.

To this end, we first define the subgraph G_y obtained from the RHG G by keeping only links between node pairs separated by distances at most y. Then, it is easy to realize that the true positive rate $\mathrm{TP}(y)$ is proportional to the expected degree $\bar{k}_y$ of the G_y:

$$\mathrm{TP}(y) = (1-q)\frac{N}{2}\bar{k}_y. \tag{E4}$$

$\bar{k}_y$ can be calculated using the hidden variable formalism:

$$\bar{k}_y = (N-1)\int\cdots\int_{x(r_1,\theta_1,r_2,\theta_2)\leqslant y} dr_1dr_2d\theta_1d\theta_2\,\rho(r_1)\rho(r_2) \times \rho(\theta_1)\rho(\theta_2)p[x(r_1,\theta_1,r_2,\theta_2)]. \tag{E5}$$

To account for noise we next define noisy subgraph $G_y(a)$ as follows. First, noise is added to node coordinates of the original RHG as prescribed by Eq. (30) and hyperbolic distances between nodes are recalculated using the updated coordinates. Second, $G_y(a)$ is formed from RHG by keeping connections at *recalculated* distances up to y. It is then easy to see that the thought true positive rate is given by

$$\mathrm{TP}(y|a) = (1-q)\frac{N}{2}\bar{k}_y(a), \tag{E6}$$

where $\bar{k}_y(a)$ is the average degree of noisy subgraph $G_y(a)$.

After a series of tedious calculations, which we detail in Appendix E 2, we obtain the leading-order behavior of $\bar{k}_y(a)$:

$$\bar{k}_y(a) \sim \begin{cases} Ng(y)a^{1-2\alpha} & \text{if} \frac{R}{2} \leqslant y \leqslant R, \\ Ng(y)a^{1-2\alpha}\left[R+2\ln\frac{a}{2}\right] & \text{if} y > R, \end{cases} \tag{E7}$$

where $\alpha \in (\frac{1}{2}, 1)$ is the radial node density parameter in Eq. (5) corresponding to degree distribution exponent $\gamma = 2\alpha + 1$. Similar to the noiseless case, $g(y)$ grows as $\exp(\frac{y}{2})$ for $y \leqslant R$ and saturates to a constant value, corresponding to $\bar{k}_y(a) = \bar{k}$ as $y \to 2R$, Fig. 14(a).

Using Eq. (D18) one can rewrite the distance-dependent precision function as

$$\mathrm{pr}(y|a) = \frac{\mathrm{TP}(y|a)}{\binom{N}{2}\int_0^y n(y'|a)dy' - \frac{q}{1-q}\mathrm{TP}(y|a)}, \tag{E8}$$

where $n(y|a)$ is the node pair distribution in the hyperbolic disk with coordinate noise.

Due to the uniform initial angular distribution $\rho(\theta)$, the node pair distribution is independent of noise, $n(y|a) = n(y)$, Fig. 14(b). Further, in the case of sufficiently large noise amplitude a, $\mathrm{TP}(y|a) \ll \binom{N}{2}\int_0^y n(y'|a)dy'$ and

$$\mathrm{pr}(y|a) \approx \frac{\mathrm{TP}(y|a)}{\binom{N}{2}\int_0^y n(y')dy'}. \tag{E9}$$

As a result, in the case $y \leqslant R$, $\mathrm{pr}(y|a) \sim a^{1-2\alpha}$, Fig. 14(c).

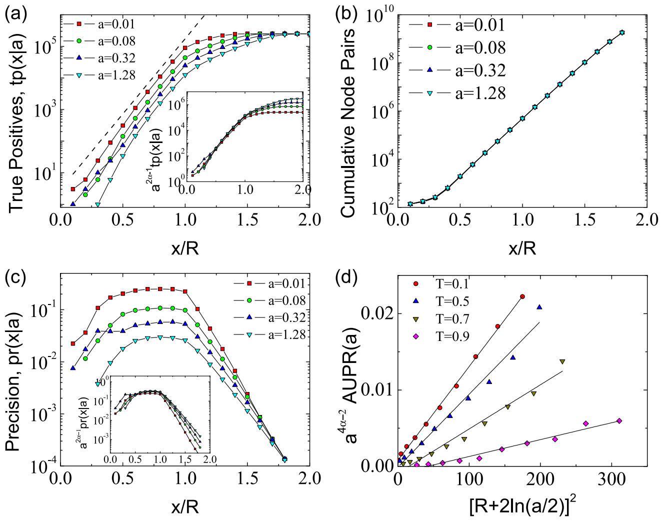

FIG. 14. HYPERLINK accuracy in case of coordinate uncertainty. All plots correspond to RHGs of $N = 10^5$ nodes, $\gamma = 2.5$ ($\alpha = 0.75$), $T = 0.1$, and $\bar{k} = 10$. (a) Distance-dependent true positive rate $\mathrm{TP}(x|a)$ evaluated for different noise amplitude values. For $x < R$, $\mathrm{TP}(x|a)$ grows as $e^{x/2}$ (see the dashed line for the reference). The inset tests the scaling of $\mathrm{TP}(x|a) \sim a^{1-2\alpha}$ for $x < R$. (b) The cumulative number of node pairs in the hyperbolic disk as a function of hyperbolic distance between the nodes. Note that the cumulative number of node pairs is independent of noise amplitude. (c) Distance-dependent precision rate $\mathrm{pr}(x|a)$ for different a values. $\mathrm{pr}(x|a)$ is nearly constant for $x < R$ since both $\mathrm{TP}(x|a)$ and $n(x|a)$ grow as $e^{x/2}$. $\mathrm{pr}(x|a)$ decays as $e^{-x/2}$ for $x > R$. The inset tests the scaling of $\mathrm{pr}(x|a) \sim a^{1-2\alpha}$ for $x < R$. (d) The scaling test for $\mathrm{AUPR}(a)$ of the RHG with $N = 5000$, $\gamma = 2.5$, and $\bar{k} = 10$. Note that $a^{4\alpha-2}\mathrm{AUPR}(a)$ grows linearly as a function of $(R + 2\ln\frac{a}{2})^2$, confirming Eq. (E11).

Since the distance-dependent recall function is proportional to the true positive rate,

$$\mathrm{rc}(y|a) = \frac{\mathrm{TP}(y|a)}{(1-q)E}. \tag{E10}$$

The resulting AUPR score scales as

$$\mathrm{AUPR}(a) \sim a^{2-4\alpha}\left[A + B\left(R + 2\ln\frac{a}{2}\right)^2\right], \tag{E11}$$

where

$$A = \frac{1-q}{E}\binom{N}{2}\int_{\frac{R}{2}}^{R}\frac{dyg(y)g'(y)}{n^c(y)}, \tag{E12}$$

$$B = \frac{1-q}{E}\binom{N}{2}\int_{R}^{2R}\frac{dyg(y)g'(y)}{n^c(y)} \tag{E13}$$

[see Fig. 14(d)].

This result suggests that the impact of coordinate uncertainty on link prediction is higher in RHG with larger $\gamma = 2\alpha + 1$ values. Intuitively, this is the case since networks with larger γ values have larger fractions of small degree nodes. Small degree nodes in the RHG are characterized by large radial coordinates, and the hyperbolic distance between the point with large radial coordinates is most affected by angular coordinate uncertainties.

2. The average degree of the noisy subgraph

Here we derive the leading term behavior of the average degree of the noisy subgraph $G_y(a)$ as a function of noise amplitude a.

As shown in the subsection above, the number of true positives $\mathrm{TP}(y|a)$ is related to the average degree of noisy subgraph $G_y(a)$. To define $G_y(a)$ we add uniform noise of amplitude a to original angular coordinates of the RHG and calculate noisy hyperbolic distances $\hat{x}_{ij}$ between all node pairs using noisy coordinates. $G_y(a)$ is the RHG subgraph formed by node pairs with noisy hyperbolic distances $\hat{x}_{ij} < y$. The average degree of $G_y(a)$ is given by

$$\bar{k}_y(a) = (N-1)\int\cdots\int_{x(r_1,\hat{\theta}_1,r_2,\hat{\theta}_2)\leqslant y} dr_1 dr_2 d\hat{\theta}_1 d\theta_1 d\hat{\theta}_2 d\theta_2 \times \rho(r_1)\rho(r_2)\rho(\hat{\theta}_1)\rho(\theta_1|\hat{\theta}_1)\rho(\hat{\theta}_2)\rho(\theta_1|\hat{\theta}_2) \times p[x(r_1,\theta_1,r_2,\theta_2)]. \tag{E14}$$

Here $\rho(r)$ is given by Eq. (5), and $\rho(\theta|\hat{\theta})$ is the conditional probability of the true angle θ, given inferred angle $\hat{\theta}$. In case of the uniform noise, $\rho(\theta|\hat{\theta})$ is also a uniform distribution centered at $\hat{\theta}$:

$$\rho(\theta|\hat{\theta}) = U(\hat{\theta}-a/2, \hat{\theta}+a/2), \tag{E15}$$

while

$$\rho(\hat{\theta}) = \rho(\theta) = \frac{1}{2\pi}. \tag{E16}$$

Throughout the calculation of $k_y(a)$ we will rely on the number of assumptions. We are primarily interested in RHGs with $2 < \gamma < 3$, which correspond to $\frac{1}{2} < \alpha < 1$. To identify leading terms we will also recall the scaling of R with the system size, $N \sim e^{\frac{R}{2}}$.

Since hyperbolic distance x in Eq. (3) depends on θ_1 and θ_2 only through their difference,

$$x(r_1,\theta_1,r_2,\theta_2) = x(r_1,r_2,\Delta\theta_{12}), \tag{E17}$$

$$\Delta\theta_{12} \equiv \pi - |\pi - |\theta_1-\theta_2||, \tag{E18}$$

and angles distributed uniformly on $[-\pi,\pi]$, $\rho(\hat{\theta}_{1,2}) = \frac{1}{2\pi}$, we can simplify Eq. (E14) as

$$\bar{k}_y(a) = \frac{N}{(2\pi)^2}\int\cdots\int_{x(r_1,r_2,\Delta\hat{\theta}_{12})\leqslant y} dr_1 dr_2 \rho(r_1)\rho(r_2) d\hat{\theta}_1 d\hat{\theta}_2 \times d\Delta\theta_{12}\tilde{\rho}(\Delta\theta_{12}|\Delta\hat{\theta}_{12}) p[x(r_1,r_2,\Delta\theta_{12})], \tag{E19}$$

where

$$\tilde{\rho}(\Delta\theta_{12}|\Delta\hat{\theta}_{12}) = \frac{1}{a^2}\Theta(a - |\Delta\theta_{12}-\Delta\hat{\theta}_{12}|), \tag{E20}$$

and $\Theta[x]$ is the Heaviside theta function. Similar to the calculation of $\bar{k}$ in the RHGs [20], we can rewrite Eq. (E19) as

$$\bar{k}_y(a) = \int_0^R dr_1 \rho(r_1)\bar{k}_y(r_1|a), \tag{E21}$$

where $\bar{k}_y(r|a)$ is the average degree of node with radial coordinate r in noisy subgraph $G_y(a)$:

$$\bar{k}_y(r_1|a) = \frac{N}{(2\pi)}\int\cdots\int_{x(r_1,r_2,\hat{\phi})\leqslant y} dr_2\rho(r_2) d\hat{\phi} d\phi \tilde{\rho}(\phi|\hat{\phi}) \times p[x(r_1,r_2,\phi)], \tag{E22}$$

and angles $\phi \equiv \Delta\theta_{12}$ and $\hat{\phi} \equiv \Delta\hat{\theta}_{12}$ are introduced to ease the notation.

To evaluate $\bar{k}_y(r_1|a)$ we note that the integration region in Eq. (E22) is given by intersection of two hyperbolic disks. The first one is of radius R and is centered at the coordinate system origin, (0,0). The second disk is of radius y and is centered at $(r_1, 0)$.

We perform the integration for the two regimes of $y \in [\frac{R}{2}, R]$ and $[R, 2R]$ separately. We do not perform the integration for the $y \in [0, \frac{R}{2}]$ regime since the number of true positives here is much smaller than that in the other two regimes. This is the case since $n(y)$ grows exponentially with y, $n(y) \sim e^{\frac{y}{2}}$. Consequently, the number of possible true positives in the $y \in [0, \frac{R}{2}]$ regime is much smaller than that in the $y \in [\frac{R}{2}, R]$ regime.

a. $y \in [\frac{R}{2}, R]$

To evaluate $\bar{k}_y(r_1|a)$ we perform the integration over r_1 and r_2 values over the domain shown in Fig. 15(d). Based on this domain, it is convenient to split the integration over r_1 into three regions, $0 \leqslant r_1 \leqslant R-y$, $R-y \leqslant r_1 \leqslant y$, and $y \leqslant r_1 \leqslant R$. However, due to specifics of the approximation techniques, it is more convenient to split the integration not into three but into five regions—(i) $0 \leqslant r_1 \leqslant \frac{R-y}{2} - \ln\frac{a}{2}$, (ii) $\frac{R-y}{2} - \ln\frac{a}{2} \leqslant r_1 \leqslant R-y$, (iii) $R-y \leqslant r_1 \leqslant y$, (iv) $y \leqslant r_1 \leqslant \frac{R+y}{2} - \ln\frac{a}{2}$, and (v) $\frac{R+y}{2} - \ln\frac{a}{2} \leqslant R$—which we depict for convenience in Fig. 15(d) with vertical dashed lines. We evaluate the contributions to $\bar{k}_y(r_1|a)$ from each of these five regions below.

Region I: $0 \leqslant r_1 \leqslant \frac{R-y}{2} - \ln\frac{a}{2}$. In this region the disk y is fully contained within the disk R. Further, since $y > R/2$, disk y is guaranteed to include the coordinate system origin for all $r_1 \in [0, R-y]$ values, Fig. 15(a). In this case the integral in $\bar{k}_y(r_1|a)$ can be evaluated as

$$\bar{k}_y(r_1|a) = \mathfrak{I}_1 + \mathfrak{I}_2, \tag{E23}$$

$$\mathfrak{I}_1 = \frac{N}{2\pi}\int_0^{y-r_1} dr_2\rho(r_2)\int_0^{2\pi} d\hat{\phi}\int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1,r_2,\phi)], \tag{E24}$$

$$\mathfrak{I}_2 = \frac{N}{\pi}\int_{y-r_1}^{y+r_1} dr_2\rho(r_2)\int_0^{\phi_y} d\hat{\phi}\int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1,r_2,\phi)], \tag{E25}$$

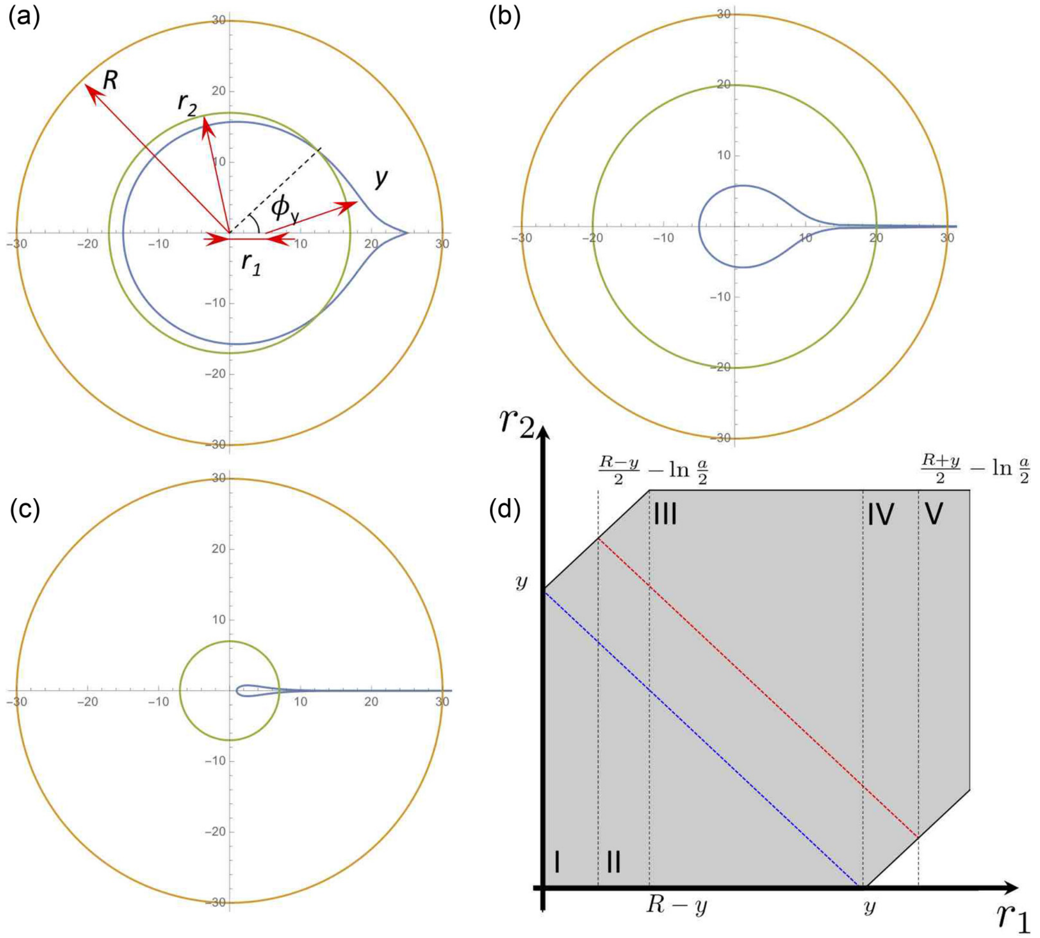

FIG. 15. Integration domain for $\bar{k}_y(a)$ in the case $y < R$. The integration is performed at the intersection of two hyperbolic disks. The first disk (yellow) corresponds to the latent space of the RHG, has radius R, and is centered at the origin. The second disk (blue) has radius y and is centered at $(r_1, 0)$. The third disk depicts the integration radius r_2 that sweeps the integration domain. Angle $\phi_y \approx 2e^{y-r_1-r_2}$ corresponds to the intersection of disks y and r_2. Based on R, y, and r_1 values we distinguish three configurations. (a) Disk y contains the origin and is fully contained within R, regions I and II. (b) Disk y contains the origin and is partially contained within R, region III. (c) Disk y does not contain the origin and is partially contained within R, regions IV and V. (d) The shaded region corresponds to the integration domain for $\bar{k}_y(a)$. Vertical dashed lines separate the five integration regions. Phase space below the blue dashed line corresponds to the case of the disk r_2 fully contained within the disk y. Phase space above the blue line corresponds to the case of disk r_2 intersecting disk y. The red dashed line is given by $r_2 + r_1 = R - 2\ln(\frac{a}{2})$ and corresponds to the loci of the integrand maxima in regions II, III, and IV.

where ϕ_y is the angle given by the intersection of the disk with radius r_2 centered at $r = 0$ and that of radius y, centered at $r = r_1$. To estimate ϕ_y we consider the triangle formed by the origin (0,0), disk y centered at $(r_1, 0)$, and the intersection of r_2 with y. The triangle has sides equal to r_1, r_2, and y with ϕ_y being the angle between r_1 and r_2. Thus, ϕ_y is given by the hyperbolic law of cosines:

$$\cosh y = \cosh r_1 \cosh r_2 - \sinh r_1 \sinh r_2 \cos\phi_y. \tag{E26}$$

In the case of sufficiently large r_1, r_2, and y values we can approximate $\cos\phi_y$ as

$$\cos\phi_y \approx 1 - 2e^{y-r_1-r_2}. \tag{E27}$$

Since $\hat{\phi}$ in the first integral sweeps the entire 2π angle, $\mathfrak{I}_1$ is given by

$$\mathfrak{I}_1 = \frac{N}{2\pi}\int_0^{y-r_1} dr_2 \rho(r_2) \int_0^{2\pi} d\hat{\phi}\, p[x(r_1, r_2, \hat{\phi})]. \tag{E28}$$

Then, since $x(r_1, r_2, \hat{\phi}) \leqslant r_1 + r_2 \leqslant R$, $p[x(r_1, r_2, \hat{\phi})] \approx 1$, leading to

$$\mathfrak{I}_1 = N e^{\alpha(y-r_1-R)}. \tag{E29}$$

The evaluation of $\mathfrak{I}_2$ is more involved and requires further approximations. We notice that $\phi_y \ll 1$ since $r_2 \in [y - r_1, y + r_1]$, which can be further approximated as

$$\phi_y \approx 2e^{\frac{y-r_1-r_2}{2}}. \tag{E30}$$

Then, for sufficiently large noise amplitudes $a \gg \phi_y$, we can approximate the integral $\int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi$ as $2\int_0^a d\phi$, resulting in

$$\mathfrak{I}_2 = \frac{2N}{\pi a^2}\int_{y-r_1}^{y+r_1} dr_2 \rho(r_2) \phi_y \int_0^a \frac{d\phi(a-\phi)}{1 + \exp\left(\frac{x(r_1, r_2, \phi) - R}{2T}\right)}. \tag{E31}$$

Since $r_1 < \frac{R-y}{2} - \ln\frac{a}{2}$, $r_2 < y + r_1$, and $y < R$, it follows that $x(r_1, r_2, \phi) < r_1 + r_2 + 2\ln\frac{a}{2} < R$, and, as a result, $\exp(\frac{x(r_1,r_2,\phi)-R}{2T}) \ll 1$, resulting in

$$\mathfrak{I}_2 = \frac{4\alpha N}{\pi(2\alpha-1)} e^{-\alpha R} e^{\alpha y} e^{(\alpha-1)r_1}. \tag{E32}$$

Since $\gamma > 2$ case ($\alpha > \frac{1}{2}$), $\mathfrak{I}_2 \gg \mathfrak{I}_1$, and

$$\bar{k}_y(r_1|a) \approx \mathfrak{I}_2 = \frac{4\alpha N}{\pi(2\alpha-1)} e^{-\alpha R} e^{\alpha y} e^{(\alpha-1)r_1}. \tag{E33}$$

Region II: $\frac{R-y}{2} - \ln\frac{a}{2} \leqslant r_1 \leqslant R - y$. Similar to region I, the hyperbolic disk y fully lies within disk R, Fig. 15(a). Thus, $\bar{k}_y(r_1|a)$ is given by the same expression:

$$\bar{k}_y(r_1|a) = \mathfrak{I}_3 + \mathfrak{I}_4, \tag{E34}$$

$$\mathfrak{I}_3 = \frac{N}{2\pi}\int_0^{y-r_1} dr_2\rho(r_2)\int_0^{2\pi} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1, r_2, \phi)], \tag{E35}$$

$$\mathfrak{I}_4 = \frac{N}{\pi}\int_{y-r_1}^{y+r_1} dr_2\rho(r_2)\int_0^{\phi_y} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1, r_2, \phi)]. \tag{E36}$$

The calculation of $\mathfrak{I}_3$ is identical to that of $\mathfrak{I}_1$, resulting in

$$\mathfrak{I}_3 = \mathfrak{I}_1 = Ne^{\alpha(y-r_1-R)}. \tag{E37}$$

Different from region I is the calculation of $\mathfrak{I}_4$. Indeed, in the case $r_1 \geqslant \frac{R-y}{2} + \ln\frac{a}{2}$, and $r_2 \in [y - r_1, y + r_1]$, hyperbolic distance $x(r_1, r_2, \phi)$ is no longer guaranteed to be smaller than R, and $p[x(r_1, r_2, \phi)]$ can no longer be approximated by unity. We first split $\mathfrak{I}_4$ into two parts and calculate them separately:

$$\mathfrak{I}_4 = \mathfrak{I}_{4,1} - \mathfrak{I}_{4,2}, \tag{E38}$$

where

$$\mathfrak{I}_{4,1} = \frac{N}{\pi a}\int_{y-r_1}^{y+r_1} dr_2\rho(r_2)\phi_y \times \int_0^a \frac{d\phi}{1+\exp(\frac{x(r_1,r_2,\phi)-R}{2T})}, \tag{E39}$$

$$\mathfrak{I}_{4,2} = \frac{N}{\pi a^2}\int_{y-r_1}^{y+r_1} dr_2\rho(r_2)\phi_y \times \int_0^a \frac{\phi d\phi}{1+\exp(\frac{x(r_1,r_2,\phi)-R}{2T})}. \tag{E40}$$

By approximating the hyperbolic law of cosines in Eq. (3) as $x(r_1, r_2, \phi) \approx r_1 + r_2 + 2\ln\frac{\phi}{2}$ and making use of Eq. (E30) we obtain for $\mathfrak{I}_{41}$

$$\mathfrak{I}_{4,1} = \frac{4\alpha N}{\pi a} e^{(\frac{1}{2}-\alpha)R} e^{\frac{y}{2}} e^{-r_1} \times \int_{y-r_1}^{y+r_1} dr_2 e^{(\alpha-1)r_2} I\left(\frac{a}{2}e^{\frac{r_1+r_2-R}{2}}; T\right), \tag{E41}$$

where $I(z;T) \equiv \int_0^z \frac{dx}{1+x^{\frac{1}{T}}}$ is the same function as in Eq. (D5).

Recall that for small $z \ll 1$ function $I(z;T) \approx z$, while, for $z \gg 1$, $I(z;T) \approx I(T) = \frac{\pi}{T\sin(\frac{\pi}{T})}$. With these approximations in mind we split the integration in $\mathfrak{I}_{41}$ into two subregions:

$$\int_{y-r_1}^{y+r_1} dr_2 = \int_{y-r_1}^{R-r_1-2\ln\frac{a}{2}} dr_2 + \int_{R-r_1-2\ln\frac{a}{2}}^{y+r_1} dr_2. \tag{E42}$$

In the first subregion, $r_2 \in [y - r_1, R - r_1 - 2\ln\frac{a}{2}]$, and $\frac{a}{2}e^{\frac{r_1+r_2-R}{2}} \leqslant 1$, which allows us to approximate $I(\frac{a}{2}e^{\frac{r_1+r_2-R}{2}}; T) \approx \frac{a}{2}e^{\frac{r_1+r_2-R}{2}}$. In the second subregion, $r_2 \in [R - r_1 - 2\ln\frac{a}{2}, y + r_1]$, $\frac{a}{2}e^{\frac{r_1+r_2-R}{2}} \geqslant 1$, and $I(\frac{a}{2}e^{\frac{r_1+r_2-R}{2}}; T) \approx I(T)$. Using these approximations we obtain, to the leading order,

$$\mathfrak{I}_{4,1} = \frac{2N\alpha}{\pi}\left[\frac{2}{2\alpha-1} + \frac{I(T)}{1-\alpha}\right] e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha}. \tag{E43}$$

Following the same approximation steps,

$$\mathfrak{I}_{4,2} = \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{2\tilde{I}(T)}{3-2\alpha}\right] e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha} \tag{E44}$$

where

$$\tilde{I}(T) \equiv \int_0^\infty \frac{xdx}{1+x^{\frac{1}{T}}} = \frac{\pi T}{\sin(2\pi T)} \tag{E45}$$

in the case $T < 1/2$.

Taken together, $\mathfrak{I}_{4,1}$ and $\mathfrak{I}_{4,2}$ result in

$$\mathfrak{I}_4 = \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{2I(T)}{1-\alpha} - \frac{8\tilde{I}(T)}{3-2\alpha}\right] \times e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha}. \tag{E46}$$

Finally, since $y < R$, we conclude that $\mathfrak{I}_3 \ll \mathfrak{I}_4$, resulting in

$$\bar{k}_y(r_1|a) \approx \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{I(T)}{1-\alpha} - \frac{2\tilde{I}(T)}{3-2\alpha}\right] \times e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha} \tag{E47}$$

for $\frac{R-y}{2} - \ln\frac{a}{2} \leqslant r_1 \leqslant R - y$.

Region III: $R - y \leqslant r_1 \leqslant y$. In this region disk y is partially contained within the disk R. Since $r_1 \leqslant y$, disk y still contains the coordinate system origin, Fig. 15(b). Similar to regions I and II, we split the calculation of $\bar{k}_y(r_1|a)$ into two parts:

$$\bar{k}_y(r_1|a) = \mathfrak{I}_5 + \mathfrak{I}_6, \tag{E48}$$

$$\mathfrak{I}_5 = \frac{N}{2\pi}\int_0^{y-r_1} dr_2\rho(r_2)\int_0^{2\pi} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1, r_2, \phi)], \tag{E49}$$

$$\mathfrak{I}_6 = \frac{N}{\pi}\int_{y-r_1}^{R} dr_2\rho(r_2)\int_0^{\phi_y} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1, r_2, \phi)], \tag{E50}$$

where $\phi_y \ll 1$ is the intersection angle of disk r_2 with that of y, Fig. 15(b), and is given by Eq. (E30).

We first note that the integral in $\mathfrak{J}_5$ is identical to those in $\mathfrak{J}_3$ and $\mathfrak{J}_1$:

$$\mathfrak{J}_5 = \mathfrak{J}_1 = Ne^{\alpha(y-r_1-R)}. \tag{E51}$$

The integration in $\mathfrak{J}_6$ is very similar to that in $\mathfrak{J}_4$ with the only difference in the upper integration bound of $r_2 \leqslant R$. The evaluation of $\mathfrak{J}_6$ is, therefore, straightforward and requires the same approximation steps as in $\mathfrak{J}_4$. A quicker estimate can be obtained by noting that the upper bound for r_2 in $\mathfrak{J}_4$ does not contribute to the leading term. The reason is that $\mathfrak{J}_{42}$ is dominated by r_2 in the vicinity of the $r_2 = R - r_1 - 2\ln\frac{a}{2}$ point.

Since $R > R - r_1 - 2\ln\frac{a}{2} > y - r_1$

$$\mathfrak{J}_6 = \int_{y-r_1}^{R-r_1-2\ln\frac{a}{2}} dr_2 + \int_{R-r_1-2\ln\frac{a}{2}}^{R} dr_2 \tag{E52}$$

with integrands identical to those of $\mathfrak{J}_{41}$ and $\mathfrak{J}_{42}$. Since the integrand in $\mathfrak{J}_{42}$ is dominated by smaller r_2 values we conclude that

$$\mathfrak{J}_6 = \mathfrak{J}_4. \tag{E53}$$

Finally, $\mathfrak{J}_6$ dominates $\mathfrak{J}_5$ for $\alpha > \frac{1}{2}$, resulting in

$$\overline{k}_y(r_1|a) \approx \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{I(T)}{1-\alpha} - \frac{2\tilde{I}(T)}{3-2\alpha}\right] \times e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha} \tag{E54}$$

for $R - y \leqslant r_1 \leqslant y$.

Region IV: $y \leqslant r_1 \leqslant \frac{R+y}{2} - \ln\frac{a}{2}$. In this region, hyperbolic disk y is partially contained within R and does not include the origin, Fig. 15(c). Therefore, in this region

$$\overline{k}_y(r_1|a) = \frac{N}{\pi}\int_{r_1-y}^{R} dr_2 \rho(r_2) \int_0^{\phi_y} d\hat{\phi} \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi \tilde{\rho}(\phi|\hat{\phi}) p[x(r_1, r_2, \phi)]. \tag{E55}$$

Using the arguments similar to that of region III, we obtain

$$\overline{k}_y(r_1|a) \approx \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{I(T)}{1-\alpha} - \frac{2\tilde{I}(T)}{3-2\alpha}\right] \times e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha} \tag{E56}$$

for $y \leqslant r_1 \leqslant \frac{R+y}{2} - \ln\frac{a}{2}$.

Region V: $\frac{R+y}{2} - \ln\frac{a}{2} \leqslant r_1 \leqslant R$. Similar to the situation in region IV, hyperbolic disk y intersects disk R and does not include the coordinate system origin. Different from region IV is the $r_2 = R - r_1 - 2\ln\frac{a}{2}$ point that lies outside the r_2 integration region and we can no longer relate $\overline{k}_y(r_1|a)$ to those in other regions.

To evaluate

$$\overline{k}_y(r_1|a) = \frac{N}{\pi}\int_{r_1-y}^{R} dr_2 \rho(r_2) \int_0^{\phi_y} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi \tilde{\rho}(\phi|\hat{\phi}) p[x(r_1, r_2, \phi)] \tag{E57}$$

we recall that $\phi_y \ll 1$, and for sufficiently large $a \gg \phi_y$ we obtain

$$\overline{k}_y(r_1|a) = \mathfrak{J}_7 - \mathfrak{J}_8, \tag{E58}$$

$$\mathfrak{J}_7 = \frac{N}{\pi a}\int_{r_1-y}^{R} dr_2 \rho(r_2)\phi_y \int_0^a \frac{d\phi}{1 + \exp\left(\frac{x(r_1,r_2,\phi)-R}{2T}\right)}, \tag{E59}$$

$$\mathfrak{J}_8 = \frac{N}{\pi a^2}\int_{r_1-y}^{R} dr_2 \rho(r_2)\phi_y \int_0^a \frac{\phi d\phi}{1 + \exp\left(\frac{x(r_1,r_2,\phi)-R}{2T}\right)}. \tag{E60}$$

After straightforward approximations we obtain

$$\overline{k}_y(r_1|a) = \frac{4\alpha N}{\pi a} e^{\left(\frac{1}{2}-\alpha\right)R} e^{\left(\frac{3}{2}-\alpha\right)y} e^{(\alpha-2)r_1}\left[\frac{I(T)}{1-\alpha} - \frac{4\tilde{I}(T)}{a(3-2\alpha)} e^{\frac{R+y}{2}-r_1}\right] \tag{E61}$$

for $\frac{R+y}{2} - \ln\frac{a}{2} \leqslant r_1 \leqslant R$

Merged together, Eqs. (E33), (E47), (E54), (E56), and (E61) provide the solution for $\overline{k}_y(r_1|a)$:

$$\overline{k}_y(r_1|a) \approx \begin{cases} \frac{4\alpha N}{\pi(2\alpha-1)} e^{-\alpha R} e^{\alpha y} e^{(\alpha-1)r_1} & \text{if } 0 \leqslant r_1 \leqslant \frac{R-y}{2} - \ln\frac{a}{2}, \\ \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{I(T)}{1-\alpha} - \frac{2I(\tilde{T})}{3-2\alpha}\right] e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha} & \text{if } \frac{R-y}{2} - \ln\frac{a}{2} \leqslant r_1 \leqslant \frac{R+y}{2} - \ln\frac{a}{2}, \\ \frac{4\alpha N}{\pi a} e^{\left(\frac{1}{2}-\alpha\right)R} e^{\left(\frac{3}{2}-\alpha\right)y} e^{(\alpha-2)r_1}\left[\frac{I(T)}{1-\alpha} - \frac{4\tilde{I}(T)}{a(3-2\alpha)} e^{\frac{R+y}{2}-r_1}\right] & \text{if} \frac{R+y}{2} - \ln\frac{a}{2} \leqslant r_1 \leqslant R. \end{cases} \tag{E62}$$

Using Eq. (E62) together with Eq. (E21) we finally obtain

$$k_y(a) \sim Ne^{-\left(\alpha+\frac{1}{2}\right)R} e^{\frac{y}{2}} a^{1-2\alpha}. \tag{E63}$$

b. $y \in [R, 2R]$

In the regime $y \geqslant R$ hyperbolic disk y always contains the origin, Fig. 16. To evaluate $\overline{k}_y(r_1|a)$ in this regime we need to distinguish two cases, (VI) $0 \leqslant r_1 \leqslant y - R$ and (VII) $y - R \leqslant r_1 \leqslant R$.

Region VI: $0 \leqslant r_1 \leqslant y - R$. In this regime hyperbolic disk R is fully contained within hyperbolic disk y, Fig. 16(a), and $\bar{k}_y(r_1|a) = \bar{k}(r_1)$, where $\bar{k}(r_1)$ is the average degree of a node at r_1 in the RHG. Indeed, radial coordinates of all points are within disk R, and all distances from point $(r_1, 0)$ to any point within disk R are guaranteed to be smaller than y, $x(r_1, 0, r_2, \theta) < y$ for any $\theta \in [0, 2\pi]$. Therefore in this regime

$$\bar{k}_y(r_1|a) = \frac{N}{2\pi}\int_0^R dr_2\rho(r_2)\int_0^{2\pi} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1, r_2, \phi)]. \quad \text{(E64)}$$

Since the integral over $\hat{\phi}$ sweeps the entire circle, $\hat{\theta} \in [0, 2\pi]$, synthetic noise does not affect the integration:

$$\bar{k}_y(r_1|a) = \frac{N}{2\pi}\int_0^R dr_2\rho(r_2)\int_0^{2\pi} d\phi p[x(r_1, r_2, \phi)] = \bar{k}(r_1), \quad \text{(E65)}$$

resulting in

$$\bar{k}_y(r_1|a) = \frac{4\alpha N I(T)}{(2\alpha - 1)\pi}e^{-\frac{r_1}{2}} \quad \text{(E66)}$$

in the case $0 \leqslant r_1 \leqslant y - R$.

Region VII: $R - y \leqslant r_1 \leqslant R$. In this regime hyperbolic disk R is partially contained within y and the calculation of $\bar{k}_y(r_1|a)$ splits into two integrals:

$$\bar{k}_y(r_1|a) = \Im_9 + \Im_{10}, \quad \text{(E67)}$$

$$\Im_9 = \frac{N}{2\pi}\int_0^{y-r_1} dr_2\rho(r_2)\int_0^{2\pi} d\phi p[x(r_1, r_2, \phi)] = \bar{k}(r_1), \quad \text{(E68)}$$

$$\Im_{10} = \frac{N}{\pi}\int_{y-r_1}^{R} dr_2\rho(r_2)\int_0^{\phi_y} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi\tilde{\rho}(\phi|\hat{\phi})p[x(r_1, r_2, \phi)], \quad \text{(E69)}$$

where ϕ_y is the angle of intersection of disks R and y, Fig. 16(b).

We note that the integration region for $\Im_9$ is identical to that of $\Im_1$. Different from the case of $\Im_1$ is the condition that $y > R$. In this case $x(r_1, r_2, \phi)$ is no longer guaranteed to be less than R, and $p[x(r_1, r_2, \phi)]$ cannot be approximated by 1. We start evaluating $\Im_9$ by performing the integration over ϕ, which leads to

$$\Im_9 = \frac{2\alpha N}{\pi}e^{-(\alpha-\frac{1}{2})R}e^{-\frac{r_1}{2}}\int_0^{y-r_1} dr_2 e^{(\alpha-\frac{1}{2})r_2} I\left(\frac{\pi}{2}e^{\frac{r_1+r_2-R}{2}}; T\right), \quad \text{(E70)}$$

where $I(z; T)$ is given by Eq. (D5). Recall that $I(z; T) \approx z$ if $z \ll 1$ and $I(z; T) \approx I(T)$ in case $x \gg 1$. Thus, to evaluate $\Im_9$ we split the integration over r_2 into two integrals, $\int_0^{y-r_1} = \int_0^{R-r_1-2\ln\frac{\pi}{2}} + \int_{R-r_1-2\ln\frac{\pi}{2}}^{y-r_1}$. In the first integral $\frac{\pi}{2}e^{\frac{r_1+r_2-R}{2}} < 1$ and we approximate $I(\frac{\pi}{2}e^{\frac{r_1+r_2-R}{2}}; T) \approx \frac{\pi}{2}e^{\frac{r_1+r_2-R}{2}}$, while in the second integral $\frac{\pi}{2}e^{\frac{r_1+r_2-R}{2}} > 1$ and $I(\frac{\pi}{2}e^{\frac{r_1+r_2-R}{2}}; T) \approx I(T)$. The

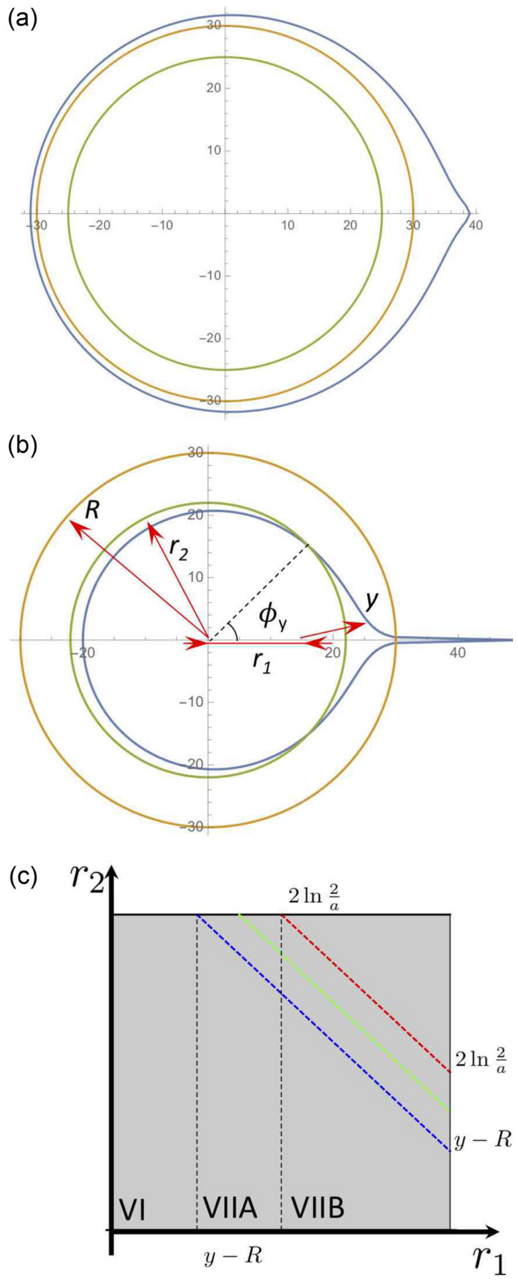

FIG. 16. Integration domain for $\bar{k}_y(a)$ at $y > R$. The integration is performed at the intersection of two hyperbolic disks. The first disk (yellow) corresponds to the latent space of the RHG, has radius R, and is centered at the origin. The second disk (blue) has radius y and is centered at $(r_1, 0)$. The third disk (green) depicts the integration radius r_2 that sweeps the integration domain. Angle ϕ_y corresponds to the intersection of disks y and r_2. Based on R, y, and r_1 values, we distinguish two configurations. (a) Disk y fully contains disk R, regions VI. (b) Disk y overlaps within R, region VII. (c) The integration domain $\bar{k}_y(a)$ is shown by the shaded region. Vertical dashed lines separate the domain into two integration regions, VI and VII. Region VII further splits into subregions VIIA and VIIB. Phase space below the blue dashed line corresponds to the case of disk r_2 fully contained within disks y and R. Phase space above the blue line corresponds to the case of disk r_2 intersecting disk y. The red dashed line is given by $r_2 + r_1 = R - 2\ln\frac{a}{2}$ and corresponds to the loci of the integrand maxima in region VII. The green dashed line corresponds to the $\tilde{R}(r_1)$ line. By construction, $\phi_y \ll 1$ for $r_2 \geqslant \tilde{R}(r_1)$.

remaining integration steps in $\mathfrak{I}_9$ are straightforward, resulting in

$$\mathfrak{I}_9 \approx \frac{2\alpha N}{\pi} e^{-(\alpha-\frac{1}{2})R} e^{-\alpha r_1} \left[\frac{1}{\alpha}\left(\frac{\pi}{2}\right)^{1-2\alpha} e^{R(\alpha-\frac{1}{2})} + \frac{2I(T)}{2\alpha-1} e^{y(\alpha-\frac{1}{2})} \right]. \tag{E71}$$

Finally, since $y > R$ and $\alpha > \frac{1}{2}$, we get

$$\mathfrak{I}_9 \approx \frac{4\alpha I(T)N}{(2\alpha-1)\pi} e^{(\alpha-\frac{1}{2})(y-R)} e^{-\alpha r_1}. \tag{E72}$$

In order to calculate $\mathfrak{I}_{10}$ we first need to estimate the cutoff angle ϕ_y, which is given by the intersection of disks R and y, and is given by Eq. (E27). ϕ_y takes values from $\phi_y \approx 2e^{\frac{y-2R}{2}}$ at $r_1 = r_2 = R$ to $\phi_y = \pi$ at $r_2 = y - r_1$. Thus, we can no longer use the $\phi_y \ll a$ approximation, as in $\mathfrak{I}_2$.

To proceed further we note that the integration domain in $\mathfrak{I}_{10}$ is given by the area above the $r_2 = y - r_1$ line, Fig. 16(c). We recall that the integration in the case $y < R$ is dominated by points in the vicinity of the $r_1 + r_2 = R - 2\ln\frac{a}{2}$ line [see red dashed line in Fig. 15(c)]. Let us assume that this is also the case in the $y \geqslant R$ regime [see red dashed line in Fig. 16(c)]. We next note that in the vicinity of the $r_1 + r_2 = R - 2\ln\frac{a}{2}$ line $\cos\phi_y \approx 1 - 2e^{y-R-2\ln\frac{a}{2}}$. For sufficiently small noise amplitude, such that $y < R - 2\ln\frac{a}{2}$, the cutoff angle $\phi_y \ll 1$ and can be approximated by Eq. (E30), and we can employ the same approximation techniques as in $\mathfrak{I}_2$.

Our strategy now is to split the integration domain of $\mathfrak{I}_{10}$ into two parts by the curve $r_2 = \tilde{R}(r_1)$ such that (i) this curve is below the $r_1 + r_2 = R - 2\ln\frac{a}{2}$ line and (ii) above this curve, $r_2 > \tilde{R}(r_1)$, the cutoff angle $\phi_y \ll 1$. One possibility for such a curve is the $\tilde{R}(r_1) = A - r_1$ line, where $A = \frac{y+R}{2} - \ln\frac{a}{2}$ [see green dashed curve in Fig. 16(c)].

Then region VII splits into two subregions, VIIA and VIIB, corresponding to $r_1 \in [y - R, 2\ln\frac{2}{a}]$ and $r_1 \in [2\ln\frac{2}{a}, R]$, respectively, Fig. 16(c). We expect the contribution to $k_y(a)$ from VIIA to be much smaller than that from VIIB since the latter contains the $r_1 + r_2 = R - 2\ln\frac{a}{2}$ line and the former does not. Therefore, we will estimate the upper bound for $k_y(r_1|a)$ in VIIA by replacing ϕ_y with π. In subregion VIIB we split the integration over r_2 into two intervals, $r_2 \in [0, \tilde{R}(r_1)]$ and $r_2 \in [\tilde{R}(r_1), R]$.

Subregion VIIA: $y - R \leqslant r_1 \leqslant 2\ln\frac{2}{a}$. Here the integral splits into

$$\bar{k}_y(r_1|a) = \mathfrak{I}_{11} + \mathfrak{I}_{12}, \tag{E73}$$

$$\mathfrak{I}_{11} = \frac{N}{2\pi}\int_0^{y-r_1} dr_2 \rho(r_2) \int_0^{2\pi} d\phi p[x(r_1, r_2, \phi)] = \bar{k}(r_1), \tag{E74}$$

$$\mathfrak{I}_{12} = \frac{N}{\pi}\int_{y-r_1}^{R} dr_2 \rho(r_2) \int_0^{\phi_y} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi \tilde{\rho}(\phi|\hat{\phi}) p[x(r_1, r_2, \phi)]. \tag{E75}$$

Following our strategy, we evaluate the upper bound for $\mathfrak{I}_{12}$ by replacing the integration limit of ϕ_y with π:

$$\mathfrak{I}_{12} \leqslant \frac{N}{\pi}\int_{y-r_1}^{R} dr_2 \rho(r_2) \int_0^{\pi} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi \tilde{\rho}(\phi|\hat{\phi}) p[x(r_1, r_2, \phi)]. \tag{E76}$$

Then,

$$\bar{k}_y(r_1|a) \leqslant \bar{k}(r_1) = \frac{4\alpha N I(T)}{(2\alpha-1)\pi} e^{-\frac{r_1}{2}} \tag{E77}$$

for $y - R \leqslant r_1 \leqslant 2\ln\frac{2}{a}$.

Subregion VIIB: $2\ln\frac{2}{a} \leqslant r_1 \leqslant R$. Here we distinguish three intervals:

$$\bar{k}_y(r_1|a) = \mathfrak{I}_{13} + \mathfrak{I}_{14} + \mathfrak{I}_{15}, \tag{E78}$$

$$\mathfrak{I}_{13} = \frac{N}{2\pi}\int_0^{y-r_1} dr_2 \rho(r_2) \int_0^{2\pi} d\phi p[x(r_1, r_2, \phi)], \tag{E79}$$

$$\mathfrak{I}_{14} = \frac{N}{\pi}\int_{y-r_1}^{\tilde{R}(r_1)} dr_2 \rho(r_2) \int_0^{\phi_y} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi \tilde{\rho}(\phi|\hat{\phi}) p[x(r_1, r_2, \phi)], \tag{E80}$$

$$\mathfrak{I}_{15} = \frac{N}{\pi}\int_{\tilde{R}(r_1)}^{R} dr_2 \rho(r_2) \int_0^{\phi_y} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi \tilde{\rho}(\phi|\hat{\phi}) p[x(r_1, r_2, \phi)], \tag{E81}$$

where $\tilde{R}(r_1) = \frac{y+R}{2} - \ln\frac{a}{2} - r_1$.

We evaluate the upper bound for $\mathfrak{I}_{14}$ by replacing the ϕ_y cutoff with π:

$$\mathfrak{I}_{14} \leqslant \frac{N}{2\pi}\int_{y-r_1}^{\tilde{R}(r_1)} dr_2 \rho(r_2) \int_0^{2\pi} d\hat{\phi} \times \int_{\hat{\phi}-a}^{\hat{\phi}+a} d\phi \tilde{\rho}(\phi|\hat{\phi}) p[x(r_1, r_2, \phi)], \tag{E82}$$

leading to

$$\mathfrak{I}_{13} + \mathfrak{I}_{14} \leqslant \frac{N}{2\pi}\int_0^{\tilde{R}(r_1)} dr_2 \rho(r_2) \int_0^{2\pi} d\phi p[x(r_1, r_2, \phi)]. \tag{E83}$$

After the same calculation steps as in $\mathfrak{I}_9$ we obtain

$$\mathfrak{I}_{13} + \mathfrak{I}_{14} \leqslant \frac{4\alpha I(T)N}{(2\alpha-1)\pi} e^{(\alpha-\frac{1}{2})(\tilde{R}(r_1)-R)} e^{-\alpha r_1}. \tag{E84}$$

To evaluate $\mathfrak{I}_{15}$ we use the $\phi_y \ll 1$ assumption, which enables us to use Eq. (E30). This approximation holds since $r_2 > \tilde{R}(r_1)$. Then, by following the same simplification steps as in $\mathfrak{I}_4$ we obtain

$$\mathfrak{I}_{15} = \mathfrak{I}_{151} - \mathfrak{I}_{152}, \tag{E85}$$

$$\mathfrak{I}_{151} = \frac{N}{\pi a}\int_{\tilde{R}(r_1)}^{R} dr_2 \rho(r_2) \phi_y \int_0^{a} \frac{d\phi}{1 + \exp\left(\frac{x(r_1, r_2, \phi) - R}{2T}\right)}, \tag{E86}$$

$$\mathfrak{I}_{152} = \frac{N}{\pi a^2}\int_{\tilde{R}(r_1)}^{R} dr_2\rho(r_2)\phi_y \int_0^a \frac{\phi d\phi}{1+\exp\left(\frac{x(r_1,r_2,\phi)-R}{2T}\right)}. \tag{E87}$$

Following the same evaluation steps as in $\mathfrak{I}_4$ we confirm that both $\mathfrak{I}_{151}$ and $\mathfrak{I}_{152}$ are dominated by points in the vicinity of $r_1 + r_2 = R - 2\ln\frac{a}{2}$, resulting in

$$\mathfrak{I}_{15} = \mathfrak{I}_4 = \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{I(T)}{1-\alpha} - \frac{2\tilde{I}(T)}{3-2\alpha}\right]$$

$$\times\, e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha}. \tag{E88}$$

By comparing Eqs. (E88) and (E84) we establish that $\mathfrak{I}_{15} \gg \mathfrak{I}_{13} + \mathfrak{I}_{14}$ since $R(\tilde{r_1}) < R$ and $y > R$, confirming our hypothesis and resulting in

$$\overline{k}_y(r_1|a) \approx \mathfrak{I}_4 = \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{2I(T)}{1-\alpha} - \frac{8\tilde{I}(T)}{3-2\alpha}\right]$$

$$\times\, e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha} \tag{E89}$$

in case $2\ln\frac{2}{a} \leqslant r_1 \leqslant R$.

Taken together, our results for regions VI and VII read

$$\overline{k}_y(r_1|a)\begin{cases} \approx \frac{4\alpha N I(T))}{\pi(2\alpha-1)} e^{-\frac{r_1}{2}} & \text{if } 0 \leqslant r_1 \leqslant y-R, \\ \leqslant \frac{4\alpha N I(T)}{(2\alpha-1)\pi} e^{-\frac{r_1}{2}} & \text{if } y-R \leqslant r_1 \leqslant 2\ln\frac{2}{a}, \\ \approx \frac{2N\alpha}{\pi}\left[\frac{1}{2\alpha-1} + \frac{I(T)}{1-\alpha} - \frac{2\tilde{I}(T)}{3-2\alpha}\right] e^{\frac{y-R}{2}} e^{-\alpha r_1}\left(\frac{a}{2}\right)^{1-2\alpha} & \text{if } 2\ln\frac{2}{a} \leqslant r_1 \leqslant R. \end{cases} \tag{E90}$$

Using Eq. (E90) together with Eq. (E21) we finally obtain

$$\overline{k}_y(a) = k_y^1(a) + k_y^2(a), \tag{E91}$$

$$k_y^1(a) \leqslant \frac{8\alpha^2 N I(T)}{\pi(2\alpha-1)^2} e^{-\alpha R}\left(\frac{a}{2}\right)^{1-2\alpha}, \tag{E92}$$

$$k_y^2(a) \approx \frac{2N\alpha^2}{\pi}\left[\frac{1}{2\alpha-1} + \frac{2I(T)}{1-\alpha} - \frac{8\tilde{I}(T)}{3-2\alpha}\right] e^{\frac{y-R}{2}} e^{-\alpha R}\left(\frac{a}{2}\right)^{1-2\alpha}\left[R + 2\ln\frac{a}{2}\right]. \tag{E93}$$

Finally, we conclude that $k_y^2(a) \gg k_y^1(a)$ since $y > R$, which allows us to establish

$$\overline{k}_y(a) \sim N e^{-(\alpha+\frac{1}{2})R} e^{\frac{y}{2}}\left(\frac{a}{2}\right)^{1-2\alpha}\left[R + 2\ln\frac{a}{2}\right] \tag{E94}$$

for $y > R$. Equation (E63) together with Eq. (E94) establish the baseline for calculation of AUPR(a) in Appendix E1.

APPENDIX F: HYPERLINK EMBEDDER

The original hyperbolic geometry inference algorithm was developed in [43] and is based on MLE. While the algorithm is rather slow with the overall computational complexity of $\mathcal{O}(N^3)$, it has been shown to accurately infer node coordinates in $\mathbb{H}^2$ leading to a number of promising applications ranging from interdomain Internet routing [43] to understanding the growth of large-scale networks [26].

In recent years hyperbolic geometry inference has become an active area of research and a collection of alternative inference methods has been developed by different research teams based on the MLE [55,56], Laplacian eigenmaps [57–59], and ISOMAP [59]. Even though most of these methods are characterized by relatively small computational complexity, $\mathcal{O}(N) - \mathcal{O}(N^2)$, their inference accuracy has not been well explored.

At the same time, our initial experiments indicate that even small node coordinate uncertainties drastically reduce link prediction accuracy (Fig. 5). Therefore, to optimize link prediction results one needs to maximize the accuracy of node coordinate inference. To this end, we developed an enhanced MLE-based geometry inference algorithm, which we outline below.

1. General MLE formulation of hyperbolic geometry inference

Given the real network of interest with randomly removed links, we aim to find the set of node coordinates $\{\mathbf{x}_i\} \equiv \{(r_i, \theta_i)\}$, $i = 1, 2, \ldots, N$, in the hyperbolic disk $\mathbb{H}^2$ maximizing the probability $\mathcal{L}(\{\mathbf{x}_i\}|a_{ij}, \mathcal{P}, q)$ that node coordinates take particular values in the case the network is generated as the RHG with a subsequent random link removal process. Here a_{ij} is the network's observed adjacency matrix, and $\mathcal{P}$ is the set of parameters of the RHG, $\mathcal{P} = \{\alpha, T, R\}$.

By the Bayes rule the thought probability is given by

$$\mathcal{L}(\{\mathbf{x}_i\}|a_{ij}, \mathcal{P}, q) = \frac{\mathcal{L}(a_{ij}|\{\mathbf{x}_i\}, \mathcal{P}, q)\mathrm{Prob}(\mathbf{x}_i)}{\mathcal{L}(a_{ij}|\mathcal{P}, q)}, \tag{F1}$$

where $\mathcal{L}(a_{ij}|\{\mathbf{x}_i\}, \mathcal{P}, q)$ is the likelihood that network a_{ij} is generated as RHG with subsequent random link removal, $\mathrm{Prob}(\mathbf{x}_i)$ is the prior probability of node coordinates generated by the RHG, and $\mathcal{L}(a_{ij}|\mathcal{P}, q)$ is the probability that the network has been generated as the RHG with random link removal.

In the following we assume the uniform prior probability

$$\mathrm{Prob}(\mathbf{x}_i) = \frac{1}{(2\pi)^N}\prod_{i=1}^{N}\rho(r_i), \tag{F2}$$

where $\rho(r_i)$ are given by Eq. (5). Since node pairs are connected independently, the likelihood is given by

$$\mathcal{L}(a_{ij}|\{\mathbf{x}_i\},\mathcal{P},q)=\prod_{i<j}[\tilde{p}(x_{ij})]^{a_{ij}}[1-\tilde{p}(x_{ij})]^{1-a_{ij}},\qquad \text{(F3)}$$

where $\tilde{p}(x_{ij})$ is the effective connection probability in the RHG generation process with subsequent random link removal:

$$\tilde{p}(x)\equiv qp(x),\qquad \text{(F4)}$$

where $p(x)$ is the RHG connection probability function prescribed by Eq. (6).

The MLE inference aims to find node coordinates $\hat{\mathbf{x}}_i$ maximizing the likelihood $\mathcal{L}(\{\mathbf{x}_i\}|a_{ij},\mathcal{P},q)$ or, equivalently, its logarithm:

$$\ln\mathcal{L}(\{\mathbf{x}_i\}|a_{ij},\mathcal{P},q)=K+\sum_{i=1}^{N}\ln\rho(r_i)+\sum_{i<j}\{a_{ij}\ln\tilde{p}(x_{ij})+(1-a_{ij})\ln[1-\tilde{p}(x_{ij})]\},\qquad \text{(F5)}$$

where constant K absorbs all terms independent of $\{\mathbf{x}_i\}$.

Our hyperbolic geometry inference procedure consists of three components: (1) finite-size effects and model parameter inference, (2) MLE-based inference of radial node coordinates, and (3) MLE-based inference of angular node coordinates.

2. Finite-size effects and model parameter inference

The RHG has four parameters: the number of nodes N, hyperbolic disk radius R, node density parameter α, and temperature T. To infer α we first estimate the degree distribution exponent γ through the inspection of the network degree distribution $P(k)$. Node density α is related to γ through Eq. (C7):

$$\alpha=\tfrac{1}{2}(\gamma-1).\qquad \text{(F6)}$$

The estimation of N and in R is less straightforward due to finite-size effects. First, in a real network one normally can only observe nodes with nonzero degrees. In contrast, the RHG may generate nodes of zero degree, which are accounted for in the calculation of the network's average degree, $\bar{k}$, Eq. (C2).

Second, due to finite-size effects, there is a cutoff value for the smallest node radius, R_0, affecting $\langle e^{-r/2}\rangle$ and, as a result, the observable $\bar{k}(r)$ and $\bar{k}$, Eqs. (C2) and (C1). Specifically, with the radius cutoff R_0

$$\langle e^{-r/2}\rangle(R_0)=\int_{R_0}^{R}e^{-r/2}\rho(r)dr=\langle e^{-r/2}\rangle\lambda(\alpha,R-R_0),\qquad \text{(F7)}$$

where $\lambda(\alpha,x)$ is the finite-size correction coefficient:

$$\lambda(\alpha,x)\equiv 1-e^{-(\alpha-1/2)x}.\qquad \text{(F8)}$$

In the thermodynamic limit $\lambda(\alpha,(R-R_0))\to 1$ as

$$1-\lambda(\alpha,(R-R_0))\sim N^{\frac{1-2\alpha}{2\alpha}}=N^{\frac{2-\gamma}{\gamma-1}}.\qquad \text{(F9)}$$

However, in networks with α close to 1/2 (γ close 2) the rate of λ convergence is slow and one needs to account for nonzero R_0.

Third, one needs to account for missing links that affect all observable properties of the RHG. In the particular case links are missing uniformly with probability $1-q$, the connection probability function $p(x)$ gets attenuated by the factor of q, Eq. (F4), affecting all observable network properties.

Taken together, zero degree nodes, minimum radius cutoff, and missing links affect observable network properties as follows:

$$\tilde{N}=N[1-P(0)],\qquad \text{(F10)}$$

$$\tilde{k}=\frac{q[\lambda(\alpha,R-R_0)]^2}{1-P(0)}\bar{k},\qquad \text{(F11)}$$

$$\tilde{k}_{\max}\approx q\lambda(\alpha,R-R_0)\bar{k}\frac{e^{-R_0/2}}{\langle e^{-r/2}\rangle},\qquad \text{(F12)}$$

where $\tilde{k}_{\max}$ is the maximum degree observed in the network and $P(0)$ is the fraction of zero degree nodes in the network. The latter can be estimated by averaging the conditional degree distribution $P(k=0|r)$ in Eq. (C5) over possible r values:

$$P(0)=2\alpha\tau^{2\alpha}\Gamma[-2\alpha,\tau],\qquad \text{(F13)}$$

$$\tau\equiv q[\lambda(\alpha,R-R_0)]\bar{k}\frac{e^{-R/2}}{\langle e^{-r/2}\rangle},\qquad \text{(F14)}$$

where $\Gamma[s,x]$ is the upper incomplete gamma function.

Equations (F10), (F11), (F12), (F13), and (F14) allow one to infer the RHG parameters R_0, R, N, as well as resulting $\bar{k}$, and $P(0)$ by measuring observables $\tilde{N}$, $\tilde{k}$, and $\tilde{k}_{\max}$. The caveat here is that parameter estimation presumes the knowledge of the missing link probability $1-q$. While this information is available in our synthetic experiments, it may not be available in real networks. In case the fraction of missing links is small, one can assume that $q=1$. The most general case of substantially incomplete networks where $q\ll 1$ is beyond the scope of this paper and will be studied elsewhere.

Finally, the temperature parameter T needs to be estimated numerically by finding the solution of

$$\bar{c}(T)=c_0,\qquad \text{(F15)}$$

where c_0 is the average clustering coefficient of the network of interest and $\bar{c}(T)$ is the average clustering coefficient of the RHG generated with temperature T. We utilize this approach to infer T of real networks in Sec. IV B, while in experiments with RHGs we use actual T values.

3. MLE-based inference of radial node coordinates

To infer radial node coordinates we extremize the logarithm of the likelihood function,

$$\frac{\partial}{\partial r_\ell}\ln\mathcal{L}(\{\mathbf{x}_i\}|a_{ij},\mathcal{P},q)=0,\qquad \text{(F16)}$$

obtaining

$$2\alpha T\coth(\alpha r_\ell)+\sum_j\left[\frac{1-p(x_{\ell j})}{1-qp(x_{\ell j})}[a_{\ell,j}-qp(x_{\ell,j})]\right]\frac{\partial x_{\ell j}}{\partial r_\ell}=0.\qquad \text{(F17)}$$

In the case of sufficiently large r values $\coth(\alpha r_\ell) \approx 1$. Further, one can approximate $x_{\ell j}$ as

$$x_{\ell j} = r_\ell + r_j + \ln \sin \theta_{\ell j}/2, \tag{F18}$$

resulting in $\frac{\partial x_{\ell j}}{\partial r_\ell} \approx 1$. Taken together, these approximations allow us to simplify Eq. (F17) as

$$2\alpha T + \sum_j a_{\ell,j} - q \sum_j p(x_{\ell,j}) = 0 \tag{F19}$$

for $1 - q \ll 1$. Note that the first summation in Eq. (F19) is the degree of node ℓ, $\sum_j a_{\ell j} = k_\ell$, while the second summation is the expected degree of the node with r_ℓ, $\tilde{k}(r_\ell) = q \sum_j p(x_{\ell,j})$. As a result, the value of $\hat{r}_\ell$ extremizing the likelihood is given by

$$\tilde{k}(\hat{r}_\ell) = k_\ell + 2\alpha T, \tag{F20}$$

where $\tilde{k}(r)$ is the observable expected degree of the node with radial coordinate r. Since the latter is given by

$$\tilde{k}(r) = q\lambda(\alpha, R - R_0)\bar{k}\frac{e^{-r/2}}{\langle e^{-r/2}\rangle}, \tag{F21}$$

one can estimate $\hat{r}_\ell$ as

$$\hat{r}_\ell = 2\ln\left[\frac{q\lambda(\alpha, R - R_0)\bar{k}}{(k_\ell + 2\alpha T)\langle e^{-r/2}\rangle}\right]. \tag{F22}$$

4. MLE inference of angular node coordinates

To infer angular node coordinates one needs to maximize the likelihood $\ln \mathcal{L}(\{\mathbf{x}_i\}|a_{ij}, \mathcal{P}, q)$ in Eq. (F5) with respect to angular coordinates $\{\theta_i\}$, given the MLE values for radial coordinates $\{\hat{r}_i\}$. Since the maximization of $\ln \mathcal{L}(\{\mathbf{x}_i\}|a_{ij}, \mathcal{P}, q)$ with respect to $\{\theta_i\}$ cannot be performed analytically, we have to rely on numerical approximations. To this end, we developed an MLE-based algorithm optimized for the linked prediction problem.

Conceptually, our algorithm is similar to the one developed in [43] but has several important differences. Following the exposition of [43], we make two observations based on the link independence in RHG. First, angular coordinates of any node subset $\mathbb{S}$ can be inferred independently (albeit, with lower accuracy) based only on the partial information contained in the graph $G_\mathbb{S}$ formed by these nodes. In other words, the inference of angular coordinates in $\mathbb{S}$ is possible by maximizing the $\mathbb{S}$-specific log likelihood:

$$\ln \mathcal{L}[G_\mathbb{S}] = \frac{1}{2}\sum_{\{i,j\}\in G_\mathbb{S}} \{a_{ij}\ln \tilde{p}(x_{ij}) + (1 - a_{ij})\ln[1 - \tilde{p}(x_{ij})]\}. \tag{F23}$$

Second, any log likelihood $\mathcal{L}[G_\mathbb{S}]$ can be represented as a sum of local contributions $\mathcal{L}[G_\mathbb{S}]_i$:

$$\ln \mathcal{L}[G_\mathbb{S}] = \frac{1}{2}\sum_i \ln \mathcal{L}[G_\mathbb{S}]_i, \tag{F24}$$

where

$$\ln \mathcal{L}[G_\mathbb{S}]_i = \sum_{j\neq i\in G_\mathbb{S}} \{a_{ij}\ln \tilde{p}(x_{ij}) + (1 - a_{ij})\ln[1 - \tilde{p}(x_{ij})]\}. \tag{F25}$$

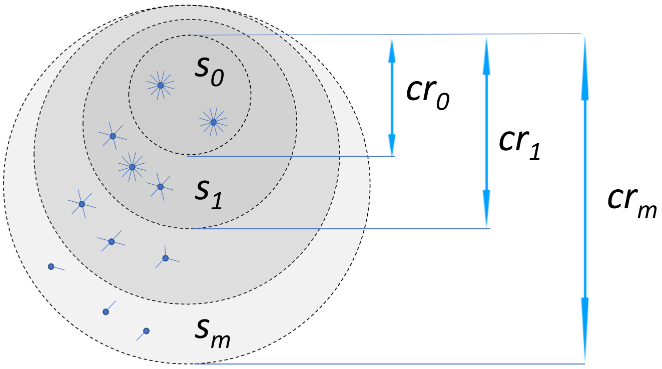

FIG. 17. Layered network structure for MLE inference. Nodes are sorted in the decreasing order of their degree and placed into logarithmically sized layers. The outer layer contains only $k = 1$ nodes.

Since the log-likelihood profile $\ln \mathcal{L}(\{\mathbf{x}_i\}|a_{ij}, \mathcal{P}, q)$ is nonconvex with abundant local maxima, we do not intend to find its global maximum by optimizing all angles at once. Instead, we proceed in a nested fashion by organizing network nodes into logarithmically sized layers with nodes of larger degree belonging to inner layers. To this end, we define the set $\mathbb{C}$ of all nodes with degrees $k > 1$. We then rank all nodes in $\mathbb{C}$ in the decreasing order of their degree value, and split the resulting node list into m layers with logarithmically growing sizes s_i, $i = 0, \dots, m - 1$:

$$s_{i+1} = \lfloor w \times s_i \rfloor, \tag{F26}$$

$$w = [N(k > 1)]^{1/m}, \tag{F27}$$

where $N(k > 1)$ is the number of nodes with degree $k > 1$, and $s_0 \ll N$. Unless otherwise noted, we set $s_0 = 20$. Finally, all $k = 1$ nodes are assigned to the outer layer s_m.

Complementary to layers $\{s_i\}$, we also define self-enclosed cores $\{\mathrm{cr}_i\}$, $i = 0, \dots, m$, such that core cr_i contains all layer with indices $j \leqslant i$, $\mathrm{cr}_i = \prod_{j=0}^{i} \bigcup s_j$, as well as the sequence of nested subgraphs $\{G_i\}$, $i = 0, \dots, m$, spanned by the nodes in corresponding cores, Fig. 17.

We start by inferring node angular coordinates $i \in \mathrm{cr}_0$ by maximizing G_0-specific likelihood $\ln\mathcal{L}[G_0]$. We then utilize the inferred angles $\{\theta_i\} \in \mathrm{cr}_0$ as initial approximation to maximize $\ln\mathcal{L}[G_1]$. We continue the angular coordinate inference procedure in the nested fashion to find angular values maximizing $\ln\mathcal{L}[G_m]$:

$$\ln\mathcal{L}[G_0] \to \ln\mathcal{L}[G_1] \to \dots \to \ln\mathcal{L}[G_m]. \tag{F28}$$

We maximize each log likelihood $\ln\mathcal{L}[G_\ell]$ iteratively by visiting G_ℓ nodes in rounds. At each round every node i in G_ℓ is visited once and placed at $\hat{\theta}_i$ maximizing its local log likelihood $\mathcal{L}[G_\mathbb{S}]_i$ with respect to the current angular values of other nodes in G_ℓ. The procedure is continued until we arrive at the stable angular configuration:

$$\max_{i\in G_\ell} \Delta\hat{\theta}_i < \epsilon, \tag{F29}$$

where $0 < \epsilon \ll 1$ is the precision parameter and $\Delta\hat{\theta}_i$ is the angular difference between angular positions of node i in two consecutive rounds. In our experiments we set $\epsilon = 10^{-4}$ rad.

Algorithm 1 Angular MLE Inference

```
organize network nodes into layers $\{s_i\}$ and cores $\{cr_i\}$, $i = 0, 1, ..., m$.
define the sequence of subgraphs $\{G_i\}$ spanned by nodes in $\{cr_i\}$.
for iter = 0 to max_iter do
    for $\ell = 0$ to $\lfloor m/2 \rfloor$ (first half) do
        assign random angle values, $\theta_i \leftarrow U[0, 2\pi]$, to nodes in $s_\ell$.
        Other nodes in $G_\ell$ retain their previous angular positions.
        $a(\ell) \leftarrow \frac{\pi}{4}(1 - \frac{\ell}{m}) + a_0$.
        for all nodes $i$ in $G_\ell$ do
            $X_i \leftarrow U(-\frac{\pi}{2}, \frac{\pi}{2})$.
            $\hat{\theta}_i \leftarrow \hat{\theta}_i + a(\ell)X_i$.
        end for
        repeat
            for all nodes $i$ in $G_\ell$ do
                $\hat{\theta}_i \leftarrow$ argmax $\ln \mathcal{L}[G_\ell]_i$, see Algorithm 2.
            end for
        until ($\max_{i \in G_\ell} \Delta\hat{\theta}_i < \epsilon$) or (# rounds > max_rounds)
    end for
    compute resulting log-likelihood $\ln \mathcal{L}[G_{\lfloor m/2 \rfloor}]$ value and save corresponding $\{\theta_i\}$ values.
end for
continue with $\{\theta_i\}$ values corresponding to the largest $\ln \mathcal{L}[G_{\lfloor m/2 \rfloor}]$.
for $\ell = \lfloor m/2 \rfloor + 1$ to $m$ (second half) do
    assign random $\{\theta_i\}$ values to nodes in $s_\ell$. Other nodes in $G_\ell$ retain their previous angular positions.
    repeat
        for all nodes $i$ in $G_\ell$ do
            $\hat{\theta}_i \leftarrow$ argmax $\ln \mathcal{L}[G_\ell]_i$, see Algorithm 2.
        end for
    until ($\max_{i \in G_\ell} \Delta\hat{\theta}_i < \epsilon$) or (# rounds > max_rounds)
    $a(\ell) \leftarrow \frac{\pi}{4}(1 - \frac{\ell}{m}) + a_0$.
    for all nodes $i$ in $G_\ell$ do
        $X_i \leftarrow U(-\frac{\pi}{2}, \frac{\pi}{2})$.
        $\hat{\theta}_i \leftarrow \hat{\theta}_i + a(\ell)X_i$.
    end for
end for
for 20 iterations do
    for all nodes $i$ in $G$
        $X_i \leftarrow U(-\frac{\pi}{2}, \frac{\pi}{2})$.
        $\hat{\theta}_i \leftarrow \hat{\theta}_i + a_0 X_i$.
    end for
    for all nodes $i$ in $G$ do
        $\hat{\theta}_i \leftarrow$ argmax $\ln \mathcal{L}[G]_i$, see Algorithm 2.
    end for
end for
```

The required total number of all-node visit rounds is typically small, of the order of the network average degree. In certain circumstances, e.g., in the case of the global $\ln \mathcal{L}[G_\ell]$ maximum close to the second largest maximum, the procedure may require a large number of rounds to converge. To avoid these scenarios we limit the maximum number of rounds to 10 per G_ℓ.

Our experiments indicate that the resulting HYPERLINK link prediction accuracy is highly sensitive to the correct placement of highest degree nodes. Thus, to further improve angular inference of the most connected nodes, we split the procedure into two parts, $\ell = 0, 1, \ldots, \lfloor m/2 \rfloor$ and $\lfloor m/2 \rfloor + 1, \ldots, m$, respectively. The first part is repeated independently for max_iter = 20 times, starting from different initial angle values. For each repetition the resulting $\ln \mathcal{L}[G_{\lfloor m/2 \rfloor}]$ value is computed. The second part is carried out only once using $\{\theta_i\}$ values corresponding to the iteration with largest $\ln \mathcal{L}[G_{\lfloor m/2 \rfloor}]$ value. Since $\ell = 0, 1, \ldots, \lfloor m/2 \rfloor$ cores are significantly smaller than $\ell = \lfloor m/2 \rfloor + 1, \ldots, m/2$ cores, the first part is carried out much faster than the second, despite the large number of repetitions.

After each round ℓ we perturb the angular coordinates $\hat{\theta}_i$, $i \in \mathrm{cr}_\ell$, by adding random noise:

$$\hat{\theta}_i \leftarrow \hat{\theta}_i + a(\ell)X_i, \tag{F30}$$

$$X_i \leftarrow U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \tag{F31}$$

with amplitude $a(\ell)$, which we decrease linearly as $a(\ell) = \frac{\pi}{4}(1 - \frac{\ell}{m}) + a_0$. These coordinate perturbations allow us to avoid getting trapped in local maxima of the log-likelihood function and to arrive to the optimal angles $\{\theta_i\}$ faster. We also stress the importance of the nonzero residual noise amplitude of a_0. In the final $\ell = m$ stage residual noise allows us to effectively "repel" $k = 1$ nodes connected to the same node. Without residual noise at the $\ell = m$ step, all $k = 1$ nodes connected to the same node are likely to be placed very close to each other and their common neighbor. As a result, pairs of these $k = 1$ nodes will be ranked as the most likely candidates for link prediction, and will adversely affect the HYPERLINK accuracy. Our experiments indicate that the HYPERLINK accuracy is not sensitive to specific a_0 values, as long as $a_0 \in [10^{-6}, 10^{-3}]$. In all our experiments we set $a_0 = 10^{-4}$ rad.

The final part of the embedder algorithm is the series of 20 coordinate perturbations, following local coordinate inferences in the entire network G. This last step often helps to further improve coordinate inference accuracy and, consequently, the accuracy of link prediction. The angular inference procedure is summarized in Algorithm 1.

Having sketched the angular inference procedure, we now focus on the individual node placement subroutine. We determine $\hat{\theta}_i$ for each node by maximizing the corresponding local log likelihood $\ln \mathcal{L}[G_\ell]_i$. To this end, we split the angular space $[-\pi, \pi]$ evenly into $\mathcal{O}(N_\ell)$ regions, where N_ℓ is the number of nodes in G_ℓ. By placing node i into each of these regions we then identify $\hat{\theta}_i$ maximizing its local likelihood. Since $\ln \mathcal{L}[G_\ell]_i$ calculation takes $\mathcal{O}(N_\ell)$ steps for each θ_i value, it takes $\mathcal{O}(N_\ell^3)$ steps to execute each round ℓ. As a result, the overall running-time complexity for m layers, $\mathcal{O}(mN^3)$, is prohibitive for large networks.

To reduce the running time complexity to $\mathcal{O}(m\langle k\rangle N^2)$, where $\langle k\rangle$ is the average degree of the entire network, we utilize the following approximation, first offered in [43]. If the number of nodes in G_ℓ is larger than or equal to 500, for each node we first obtain the rough estimate of $\hat{\theta}_i$ by taking into account only its neighboring nodes in G_ℓ. To this end we find the nearly optimal placement $\tilde{\theta}_i$ by maximizing

$$\ln \tilde{\mathcal{L}}[G_{\mathbb{S}}]_i = \sum_{j \neq i \in G_{\mathbb{S}}} a_{ij} \ln \tilde{p}(x_{ij}). \tag{F32}$$

Since the summation in Eq. (F32) goes only through node i neighbors, it now takes $\mathcal{O}(k_i N)$ steps to find $\tilde{\theta}_i$. Having obtained the initial approximation, we then look for the optimal angle $\hat{\theta}_i$ in the neighborhood of $\tilde{\theta}_i$ maximizing the full local likelihood $\ln \mathcal{L}[G_{\mathbb{S}}]_i$, which takes $\mathcal{O}(LN)$ steps, where L is the neighborhood centered at $\tilde{\theta}_i$. Specifically, we search for $\hat{\theta}_i$ within $L = 300\frac{N_\ell}{N}$ regions on both sides of $\tilde{\theta}_i$, which takes $\mathcal{O}(\frac{N_\ell^2}{N})$ steps, leading to the overall running time complexity of $\mathcal{O}(m\langle k\rangle N^2)$ steps. The individual node placement subroutine is summarized in Algorithm 2.

The outline of the HyperLink embedder above is its simplified description omitting a number of important details and presenting some of them slightly differently. The full detailed description of the algorithm exactly as used in this paper is included in its Bitbucket repository [38].

To validate the hyperbolic geometry inference algorithm we compare inferred coordinates in the RHG to its true coordinates. Parameters of the RHG are taken to be $N = 5000$, $\langle k\rangle = 10$, $T = 0.5$, and $\gamma = 2.5$. As seen from Figs. 18(a)–18(c), the accuracy of the angular coordinate inference does not decline significantly for small degree nodes. This is the case, mainly, due to the nested inference with inference cores cr_i covering all network nodes, in contrast to the original algorithm of [43], where cores only cover the most connected nodes.

As seen from Fig. 18(d), Eq. (F22) allows for accurate inference of small radial coordinates. At the same time, radial coordinates inference is less accurate for large radial

Algorithm 2 Individual node placement subroutine

```
if $N_\ell < 500$ then
    split the angular space $[-\pi, \pi]$ evenly into $\mathcal{O}(N_\ell)$ regions.
    for each region $r$ in $[-\pi, \pi]$ do
        assign $\theta_i(r)$ values to lower boundaries of each region $r$.
        compute $\ln \mathcal{L}[G_\ell]_i$ for $\theta_i(r)$, as defined in Eq. (F23).
    end for
    $\hat{\theta}_i \leftarrow \mathrm{argmax}_{r\in[-\pi,\pi]} \ln \mathcal{L}[G_\ell]_i$
else
    split the angular space $[-\pi, \pi]$ evenly into $\mathcal{O}(N_\ell)$ regions.
    for each region $r$ in $[-\pi, \pi]$ do
        sample $\theta_i(r)$ uniformly at random from region $r$.
        compute $\ln \tilde{\mathcal{L}}[G_\ell]_i$ for $\theta_i(r)$, as defined in Eq. (F32).
    end for
    $\tilde{\theta}_i \leftarrow \mathrm{argmax}_{r\in[-\pi,\pi]} \ln \tilde{\mathcal{L}}[G_\ell]_i$
    for each region $r$ in $[\|\tilde{\theta}_i - L\|, \|\tilde{\theta}_i + L\|]$ do
        assign $\theta_i(r)$ values to lower boundaries of each region $r$.
        compute $\ln \mathcal{L}[G_\ell]_i$ for $\theta_i(r)$.
    end for
    Identify $\hat{r}$ maximizing $\ln \mathcal{L}[G_\ell]_i$. $\hat{\theta}_i \leftarrow \theta_i(\hat{r})$
    $\hat{\theta}_i \leftarrow \mathrm{argmax}_{r\in[\|\tilde{\theta}_i - L\|, \|\tilde{\theta}_i + L\|]} \ln \mathcal{L}[G_\ell]_i$
end if
```

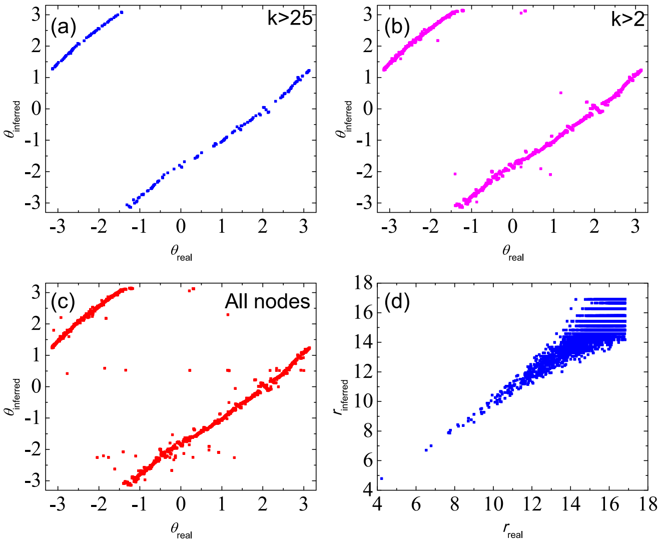

FIG. 18. Testing the hyperbolic geometry inference algorithm. Here we plot inferred vs original node coordinates for the RHG that we map to the hyperbolic space. All plots correspond to the same RHG of $N = 5000$, $\langle k\rangle = 10$, $T = 0.5$, and $\gamma = 2.5$. (a), (b) Angular coordinates for nodes with degrees $k > 25$ and 2, respectively. (c) Angular coordinates of all nodes. (d) Radial coordinates of all nodes in the graph.

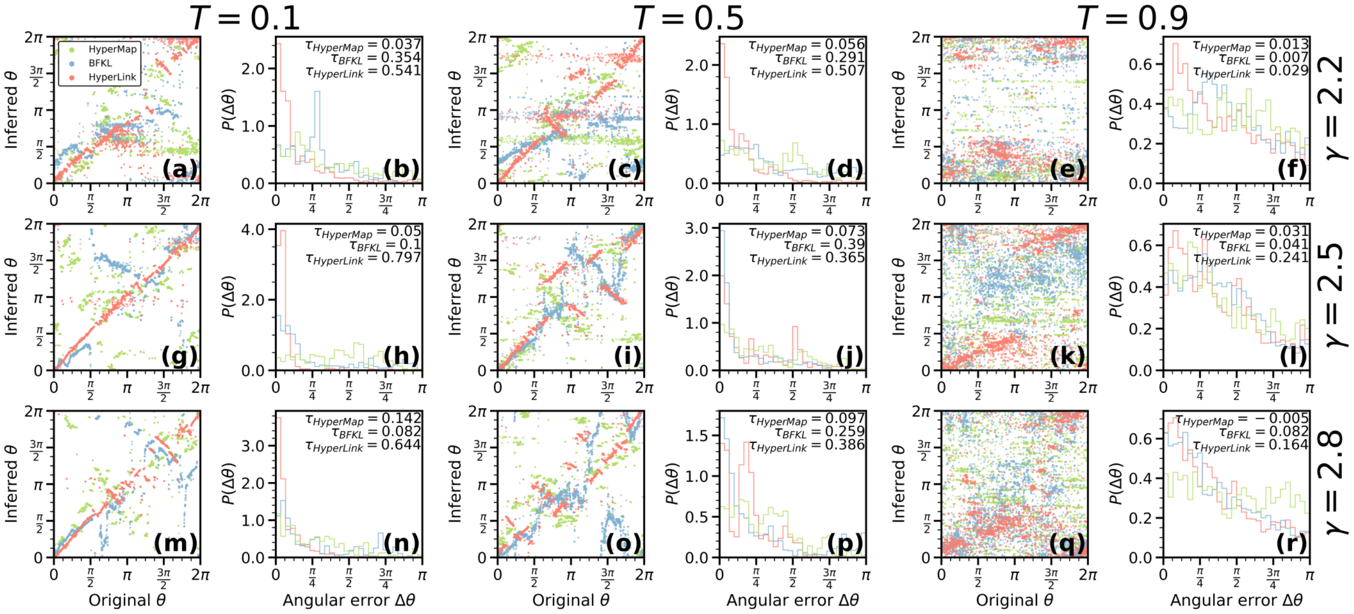

FIG. 19. HYPERLINK embedder accuracy compared to other embedding algorithms. RHGs are embedded to the hyperbolic disk by (red) HYPERLINK embedder, (blue) the algorithm by Bläsius *et al.* [55] (BFKL), and (green) the HYPERMAP [35] algorithm. All comparisons correspond to RHGs consisting of $N = 5000$ nodes, $\bar{k} = 10$, $1 - q = 0.5$ missing links, and various T and γ parameters. Panels are arranged according to T and γ parameters. (a), (c), (e), (g), (i), (k), (m), (o), (q) Scatter plots displaying inferred angular coordinates as a function of true angular coordinates. (b), (d), (f), (h), (j), (l), (n), (p), (r) To quantify the embedding accuracy, we plot the distributions of embedding errors, $P(\Delta\theta)$, where $\Delta\theta \equiv \pi - |\pi - |\theta_{\text{inferred}} - \theta_{\text{original}}||$. To quantify the association between the inferred and the original angular coordinates for each embedding we employ the U-statistic $\tau \in [-1, 1]$ [74]. The U-statistic τ quantifies the correlation between the ordering of the inferred and original and angular coordinates and ranges from $\tau = 1$, in the case the two orderings are the same, to $\tau = -1$ in the case the two orderings are inverted with respect to one another. The U-statistic τ is invariant under global shifts of the inferred coordinates. Our results indicate that the HYPERLINK accuracy is higher than that of the considered two algorithms in all cases, with the only exception of the $T = 0.5$, $\gamma = 2.5$ case, where BFKL is slightly better.

coordinates. To explain this observation we recall that the key assumption in Eq. (F22) is that the node degree in the RHG is fully determined by its radial coordinate. In other words, we assume that possible node degree values are narrowly distributed around its expected value, which is given by Eq. (F22). This is indeed the case since node degrees are distributed according to the Poisson distribution, Eq. (C5). The coefficient of variation of the Poisson distribution, however, is large for small mean values. This leads to significant variation in node degree values in the case of nodes with large radial coordinates, making Eq. (F22) inaccurate.

The HYPERLINK embedder allows for accurate node coordinate inference even in substantially incomplete networks in contrast to other mapping methods, e.g., HYPERMAP [34] and the algorithm by Bläsius *et al.* [55], which become less accurate in the case of large T values, Fig. 19.

As evidenced by Fig. 18 and, indirectly, by our link prediction results in Secs. III and IV, our hyperbolic inference algorithm is sufficiently accurate for the prediction of missing links on both synthetic and real networks. At the same time, the algorithm does have limitations. First, it is designed to map networks with links removed uniformly at random. The link presence rate q is the required parameter of the algorithm. In cases when the fraction of missing links is unknown, q needs to be estimated and this may lead to less accurate mapping. The second limitation is the algorithm's running time complexity of $\mathcal{O}(N^2)$ restricting its utility to networks of smaller size. Finally, the third limitation is the analytic estimation of radial coordinates, which is not accurate for small degree nodes. Addressing these limitations is the subject of future work that is expected to further improve the accuracy and the utility of link prediction with hyperbolic geometry.

[1] W. Peng, X. Baowen, W. Yurong, and Z. Xiaoyu, Link prediction in social networks: The state-of-the-art, Sci. China Inf. Sci. **58**, 011101 (2015).

[2] L. Lü and T. Zhou, Link prediction in complex networks: A survey, Phys. A Stat. Mech. Appl. **390**, 1150 (2011).

[3] A. K. Menon and C. Elkan, Link Prediction via Matrix Factorization, in *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2011*, Lecture Notes in Computer Science, Vol. 6912 (Springer, Berlin, Heidelberg, 2011), pp. 437–452.

[4] T. P. Peixoto, Reconstructing Networks with Unknown and Heterogeneous Errors, Phys. Rev. X **8**, 041011 (2018).

[5] D. J. Marchette and C. E. Priebe, Predicting unobserved links in incompletely observed networks, Comput. Stat. Data Anal. **52**, 1373 (2008).

[6] R. Guimerà and M. Sales-Pardo, Missing and spurious interactions and the reconstruction of complex networks, Proc. Natl. Acad. Sci. USA **106**, 22073 (2009).

[7] M. Kim and J. Leskovec, The Network completion problem: Inferring missing nodes and edges in networks, SIAM Int. Conf. Data Min., 47 (2011).

[8] L. A. Adamic and E. Adar, Friends and neighbors on the Web, Soc. Networks **25**, 211 (2003).

[9] M. E. J. Newman and A. Clauset, Structure and inference in annotated networks, Nat. Commun. **7**, 1 (2016).

[10] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork, Comparative assessment of large-scale data sets of protein-protein interactions, Nature (London) **217**, 399 (2002).

[11] H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A.-L. Barabási, J. Tavernier, D. E. Hill, M. Vidal, and M. Yldrm, High-quality binary protein interaction map of the yeast interactome network, Science **322**, 104 (2008).

[12] I. A. Kovács, K. Luck, K. Spirohn, Y. Wang, C. Pollis, S. Schlabach, W. Bian, D.-K. Kim, N. Kishore, T. Hao, M. A. Calderwood, M. Vidal, and A.-L. Barabási, Network-based prediction of protein interactions, Nat. Commun. **10**, 1240 (2019).

[13] T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E **76**, 046115 (2007).

[14] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, Recommender systems, Phys. Rep. **519**, 1 (2012).

[15] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, Recommender systems survey, Knowledge-Based Syst. **46**, 109 (2013).

[16] J. B. Schafer, J. Konstan, and J. Riedl, in *Proceedings of the First ACM Conference on Electronic Commerce*, 1999, pp. 158–166.

[17] E. N. Gilbert, Random plane networks, J. Soc. Ind. Appl. Math. **9**, 533 (1961).

[18] D. D. McFarland and D. J. Brown, Social distance as a metric: A systematic introduction to smallest space analysis, Bond. Plur. Form Subst. Urban Soc. Networks **6**, 213 (1973).

[19] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. **27**, 415 (2001).

[20] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, Phys. Rev. E **82**, 036106 (2010).

[21] M. E. J. Newman and T. P. Peixoto, Generalized Communities in Networks, Phys. Rev. Lett. **115**, 088701 (2015).

[22] A. Brew and M. Salter-Townshend, A Latent Space Mapping for Link Prediction, in *proceedings of Workshop on Networks Across Disciplines: Theory and Applications* (2010).

[23] L. Zhu, D. Guo, J. Yin, G. V. Steeg, and A. Galstyan, Scalable temporal latent space inference for link prediction in dynamic social networks, IEEE Trans. Knowl. Data Eng. **28**, 2765 (2016).

[24] G. García-Pérez, R. Aliakbarisani, A. Ghasemi, and M. Á. Serrano, Precision as a measure of predictability of missing links in real networks, Phys. Rev. E **101**, 052318 (2020).

[25] M. A. Serrano, D. Krioukov, and M. Boguñá, Self-Similarity of Complex Networks and Hidden Metric Spaces, Phys. Rev. Lett. **100**, 078701 (2008).

[26] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, Popularity versus similarity in growing networks, Nature (London) **489**, 537 (2012).

[27] K. Zuev, M. Boguñá, G. Bianconi, and D. Krioukov, Emergence of soft communities from geometric preferential attachment, Sci. Rep. **5**, 9421 (2015).

[28] E. Lazega, S. Wasserman, and K. Faust, Social Network Analysis: Methods and Applications, Rev. Française Sociol. **36**, 781 (2006).

[29] M. Newman, *Networks: An Introduction* (Oxford University, New York, 2010).

[30] A.-L. Barabási and M. Pósfai, *Network Science* (Cambridge University, Cambridge, England, 2016), p. 456.

[31] P. van der Hoorn, G. Lippner, and D. Krioukov, Sparse Maximum-Entropy Random Graphs with a Given Power-Law Degree Distribution, J. Stat. Phys. **173**, 806 (2018).

[32] D. Krioukov, Clustering Implies Geometry in Networks, Phys. Rev. Lett. **116**, 208302 (2016).

[33] M. Á. Serrano, M. Boguñá, and F. Sagués, Uncovering the hidden geometry behind metabolic networks, Mol. Biosyst. **8**, 843 (2012).

[34] F. Papadopoulos, C. Psomas, and D. Krioukov, Network mapping by replaying hyperbolic growth, IEEE/ACM Trans. Netw. **23**, 198 (2015).

[35] F. Papadopoulos, R. Aldecoa, and D. Krioukov, Network geometry inference using common neighbors, Phys. Rev. E **92**, 022807 (2015).

[36] A. Muscoloni and C. V. Cannistraci, Leveraging the nonuniform PSO network model as a benchmark for performance evaluation in community detection and link prediction, New J. Phys. **20**, 063022 (2018).

[37] A. Muscoloni and C. V. Cannistraci, Minimum curvilinear automata with similarity attachment for network embedding and link prediction in the hyperbolic space, arXiv:1802.01183.

[38] HYPERLINK embedder, https://bitbucket.org/dk-lab/2020_code_hyperlink/src/master/.

[39] T. Vallès-Català, T. P. Peixoto, M. Sales-Pardo, and R. Guimerà, Consistencies and inconsistencies between model selection and link prediction in networks, Phys. Rev. E **97**, 062316 (2018).

[40] A. Ghasemian, H. Hosseinmardi, and A. Clauset, Evaluating Overfit and Underfit in Models of Network Community Structure, IEEE Trans. Knowl. Data Eng. **32**, 1722 (2019).

[41] A. Ghasemian, H. Hosseinmardi, A. Galstyan, E. M. Airoldi, and A. Clauset, Stacking Models for Nearly Optimal Link Prediction in Complex Networks, PNAS **117**, 23393 (2020).

[42] D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguñá, Curvature and temperature of complex networks, Phys. Rev. E **80**, 035101(R) (2009).

[43] M. Boguñá, F. Papadopoulos, and D. Krioukov, Sustaining the Internet with hyperbolic mapping.Nat. Commun. **1**, 62 (2010).

[44] M. Kitsak, F. Papadopoulos, and D. Krioukov, Latent geometry of bipartite networks, Phys. Rev. E **95**, 032309 (2017).

[45] R. Aldecoa, C. Orsini, and D. Krioukov, Hyperbolic graph generator, Comput. Phys. Commun. **196**, 492 (2015).

[46] G. García-Pérez, M. Boguñá, and M. Á. Serrano, Multiscale unfolding of real networks by geometric renormalization, Nat. Phys. **14**, 583 (2018).

[47] A. Muscoloni and C. V. Cannistraci, A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities, New J. Phys. **20**, 052002 (2018).

[48] G. García-Pérez, A. Allard, M. Á. Serrano, and M. Boguñá, Mercator: Uncovering faithful hyperbolic embeddings of complex networks, New J. Phys. **21**, 123033 (2019).

[49] J. Davis and M. Goadrich, in *Proceedings of the 23rd International Conference on Machine Learning* (ACM, New York, 2006), pp. 233–240.

[50] G. Alanis-Lobato and M. A. Andrade-Navarro, Distance distribution between complex network nodes in hyperbolic space, Complex Syst. **25**, 223 (2016).

[51] H. Ma and A.-P. Zeng, Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms, Bioinformatics **19**, 270 (2003).

[52] University of Oregon Route Views Project, http://www.routeviews.org/routeviews/.

[53] The Open PGP Alliance, http://www.openpgp.org/.

[54] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, and H. E. Stanley, Toward link predictability of complex networks, Proc. Natl. Acad. Sci. USA **112**, 2325 (2015).

[55] T. Bläsius, T. Friedrich, A. Krohmer, and S. Laue, in *Proceedings of the 24th Annual European Symposium on Algorithms*, edited by P. Sankowski and C. Zaroliagis, Leibniz International Proceedings in Informatics (LIPIcs) (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2016), Vol. 57, p. 16.

[56] Z. Wang, Y. Wu, Q. Li, F. Jin, and W. Xiong, Link prediction based on hyperbolic mapping with community structure for complex networks, Phys. A Stat. Mech. Appl. **450**, 609 (2016).

[57] G. Alanis-Lobato, P. Mier, and M. A. Andrade-Navarro, Efficient embedding of complex networks to hyperbolic space via their Laplacian, Sci. Rep. **6**, 30108 (2016).

[58] G. Alanis-Lobato, P. Mier, and M. A. Andrade-Navarro, Manifold learning and maximum likelihood estimation for hyperbolic network embedding, Appl. Netw. Sci. **1**, 10 (2016).

[59] A. Muscoloni, J. M. Thomas, S. Ciucci, G. Bianconi, and C. V. Cannistraci, Machine learning meets complex networks via coalescent embedding in the hyperbolic space, Nat. Commun. **8**, 1615 (2017).

[60] A. Faqeeh, S. Osat, and F. Radicchi, Characterizing the Analogy Between Hyperbolic Embedding and Community Structure of Complex Networks, Phys. Rev. Lett. **121**, 098301 (2018).

[61] P. Colomer-de Simón, M. Á. Serrano, M. G. Beiró, J. I. Alvarez-Hamelin, and M. Boguñá, Deciphering the global organization of clustering in real complex networks, Sci. Rep. **3**, 2517 (2013).

[62] Further details on the dataset can be found at http://snap.stanford.edu/data/as.html.

[63] OpenPGP web of trust database, http://www.lysator.liu.se/~jc/wotsap/wots2/.

[64] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, Scale-free networks well done, Phys. Rev. Res. **1**, 033034 (2019).

[65] D. Liben-Nowell and J. Kleinberg, The Link Prediction Problem for Social Networks, in *Proceedings of the 12th Annual ACM International Conference on Informational Knowledge Management* (ACM, New York, 2003), pp. 556–559.

[66] T. Zhou, L. Lü, and Y. C. Zhang, Predicting missing links via local information, Eur. Phys. J. B **71**, 623 (2009).

[67] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks, Sci. Rep. **3**, 1613 (2013).

[68] P. Jaccard, Étude comparative de la distribution florale dans une portion des Alpes et des Jura, Bull. del la Société Vaudoise des Sci. Nat. **37**, 547 (1901).

[69] T. P. Peixoto, Nonparametric Bayesian inference of the microcanonical stochastic block model, Phys. Rev. E **95**, 012317 (2017).

[70] T. P. Peixoto, Hierarchical Block Structures and High-Resolution Model Selection in Large Networks, Phys. Rev. X **4**, 011047 (2014).

[71] T. P. Peixoto, The graph-tool PYTHON library, https://figshare.com (2014).

[72] T. P. Peixoto, Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models, Phys. Rev. E **89**, 012804 (2014).

[73] M. Boguñá and R. Pastor-Satorras, Class of correlated random networks with hidden variables, Phys. Rev. E **68**, 036112 (2003).

[74] N. I. Fisher and A. J. Lee, Nonparametric measures of angular-linear association, Biometrika **68**, 629 (1981).