

Mercator: uncovering faithful hyperbolic embeddings of complex networks

— Supplementary Material —

Guillermo García-Pérez,^{1,2,*} Antoine Allard,^{3,4,*} M. Ángeles Serrano,^{5,6,7} and Marián Boguñá^{5,6,†}

¹*QTF Centre of Excellence, Turku Centre for Quantum Physics,
Department of Physics and Astronomy, University of Turku, FI-20014 Turun Yliopisto, Finland*

²*Complex Systems Research Group, Department of Mathematics and Statistics,
University of Turku, FI-20014 Turun Yliopisto, Finland*

³*Département de physique, de génie physique et d'optique,
Université Laval, Québec (Québec), Canada G1V 0A6*

⁴*Centre de modélisation mathématique, Université Laval, Québec (Québec), Canada G1V 0A6*

⁵*Departament de Física de la Matèria Condensada,
Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain*

⁶*Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Barcelona, Spain*

⁷*Institució Catalana de Recerca i Estudis Avançats (ICREA),
Passeig Lluís Companys 23, E-08010 Barcelona, Spain*

This document contains supplemental results to the results presented in the main text.

CONTENTS

I. Community detection	S1
A. Inhomogeneous $\mathbb{S}^1$ model	S1
B. Comparison with the nPSO model	S3
II. Embedding real networks	S9
References	S19

I. COMMUNITY DETECTION

A. Inhomogeneous $\mathbb{S}^1$ model

We tested the performance of Mercator to detect community structure with 4 synthetic networks generated with the $\mathbb{S}^1$ model. A community structure was induced in each of these networks by a nonuniform distribution of the angular positions (θ), see Fig. S1, and the hidden variables κ were drawn from the probability density function

$$\rho(\kappa) \propto \kappa^{-2.5} , \quad (\text{S1})$$

with the proportionality constant chosen such that $\langle k \rangle \simeq \langle \kappa \rangle \simeq 10$. We used $\beta = 3.2$ and fixed $\mu = \frac{\beta}{2\pi\langle k \rangle} \sin \frac{\pi}{\beta}$. We then isolated the largest connected component which yielded the 4 networks of approximately $N = 1000$ nodes used in this section.

Once we obtained the inferred positions from Mercator, we applied two techniques to identify angular communities. The first one is the geometric critical gap method developed in [5] and later in [6] in which two nodes, i and j , belong to the same community if their angular separation, $\Delta\theta_{ij}$, is less than

$$\Delta\theta_c = \frac{2\pi[\ln(N) + \gamma]}{N} , \quad (\text{S2})$$

where N is the number of nodes in the network and $\gamma \simeq 0.57721$ is the Euler-Mascheroni constant. The critical gap in Eq. (S2) is the expected maximal angular separation between two consecutive nodes found in a uniform distribution of N points in a circle, which is taken as a null model without communities. Thus, a group of nodes separated

* Both authors contributed equally to this work

† Corresponding author: marian.boguena@ub.edu

by two gaps larger than $\Delta\theta_c$ correspond to a deviation from this null model, thereby defining a community. Notice, however, that this definition is based only on geometry alone, without any direct information coming from the network structure, which is anyway included in the embedding itself.

The second method, the topological critical gap method [7], is similar to the previous one but uses modularity to determine what is the optimal critical gap to be used to partition the set of nodes. Our algorithm goes as follows. We set $\Delta\theta_c$ at a value larger than the largest angular gap between two angularly consecutive nodes; there is therefore only one single community. We then iteratively decrease $\Delta\theta_c$ and split any community into two smaller ones at the point where two consecutive nodes are separated by an angle larger than the new value of $\Delta\theta_c$. Note that we ignore any split that would create a community containing 5 nodes or less. The number of communities thus increases as the value of $\Delta\theta_c$ decreases until it becomes smaller than the smallest angular gap between consecutive nodes. This last situation corresponds to the case where every node belongs to its own community. At the end of this process, we preserve the community structure corresponding to the value of $\Delta\theta_c$ for which the modularity is maximal.

It is important to mention that these two methods will not recover the same partition used to generate the 4 original networks. Indeed, the topology of the networks is subject to fluctuations due to the distribution of the hidden variables κ and θ as well as to the network generation process itself (recall that the S^1 model only prescribes the probability for the links to exist). Besides, some of the blocks generated are small and/or very close to nearby blocks. Consequently, we do not expect these two methods to detect the exact same number of communities as the number of angular blocks used to generate the networks. The aim of this analysis is rather to assess whether or not Mercator is able to infer an embedding whose organization is coherent with the original angular structure.

We measured this coherence by computing the normalized mutual information between the angular partition used to generate the networks (each block on Fig. S1 corresponds to a community) and the communities identified by the geometrical and topological critical gap methods. Table S1 shows the results. As expected, neither of the two methods detect the exact same original angular groups, however we see that Mercator is able to infer angular positions that are extremely congruent with the original angular positions. Figures S2–S5 (right column) further illustrate the community structures detected with both methods and compare them with the original angular structure.

The left column of Figs. S2–S5 show the embedding obtained with Mercator when $g_i = g_i^{\text{LE}}$ at Eq. (A15) in the main text, i.e., when the initial positions are only given by the LE method without enforcing a minimal angular distance between nodes, which is set in Mercator as the expected angular distance two consecutive nodes in the S^1 model. Our results show that although the angular positions provided by the LE method lead to embeddings where the original angular blocks can be identified to some extent, the minimal angular distance coupled with the likelihood maximization step allow to infer an embedding for which distances between nodes can be translated into topological information in the form of a probability of connection.

Network	NMI (fast mode)	NMI (refined mode)
Network 1 (Geom. CGM)	0.88	0.88
Network 1 (Topol. CGM)	0.89	0.80
Network 2 (Geom. CGM)	0.57	0.80
Network 2 (Topol. CGM)	0.97	0.96
Network 3 (Geom. CGM)	0.80	0.80
Network 3 (Topol. CGM)	0.85	0.78
Network 4 (Geom. CGM)	1.00	0.97
Network 4 (Topol. CGM)	0.87	0.87

TABLE S1. Communities detected using the topological and geometrical critical gap methods on the embeddings obtained with the fast and refined modes of Mercator for the 4 synthetic networks described in the text. The overlap between the original angular blocks (see Fig. S1) and the detected communities is measured using the normalized mutual information (NMI).

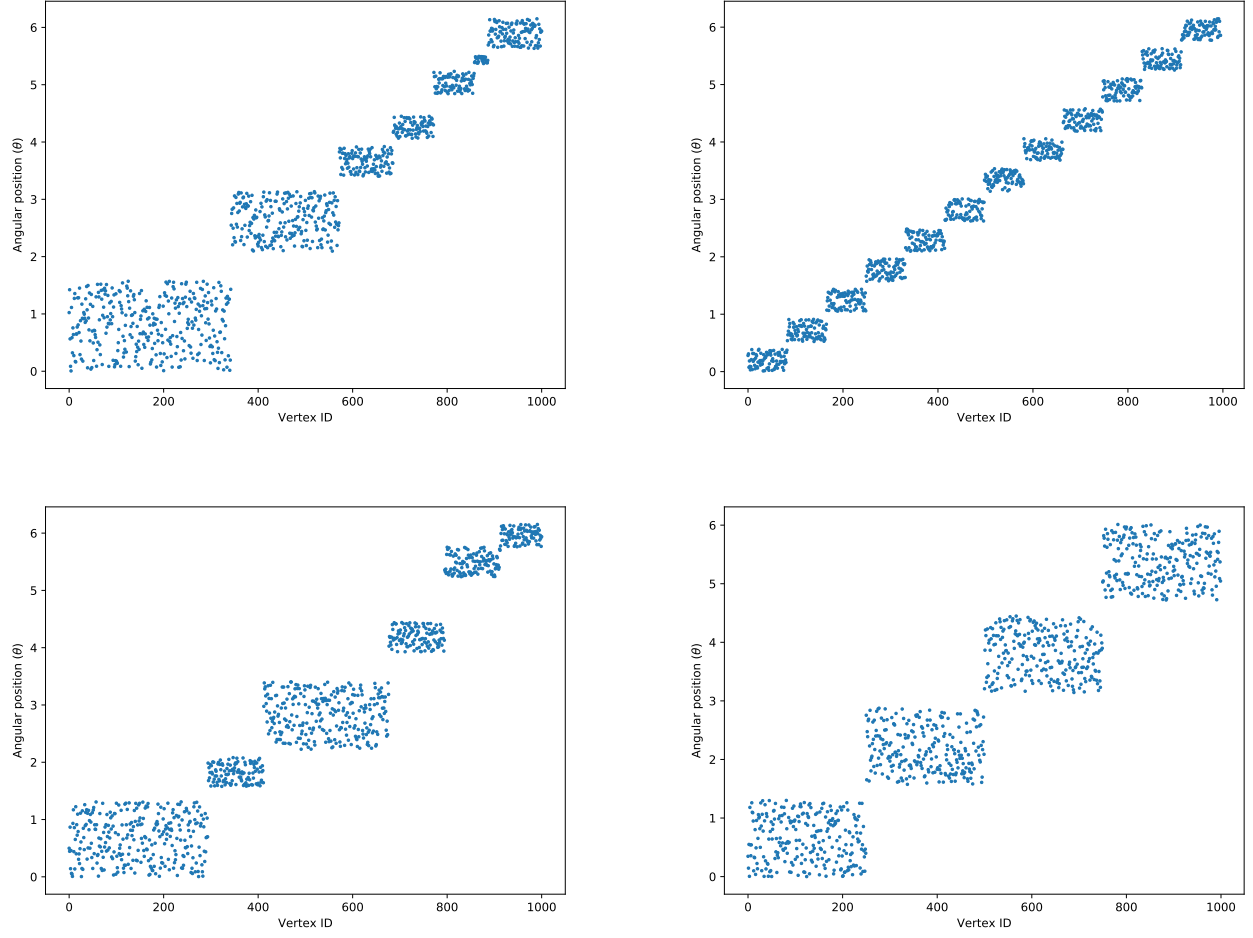


FIG. S1. Angular blocks used to generate the 4 synthetic networks with community structure. These blocks are used as the original communities in the calculation of NMI on Table S1.

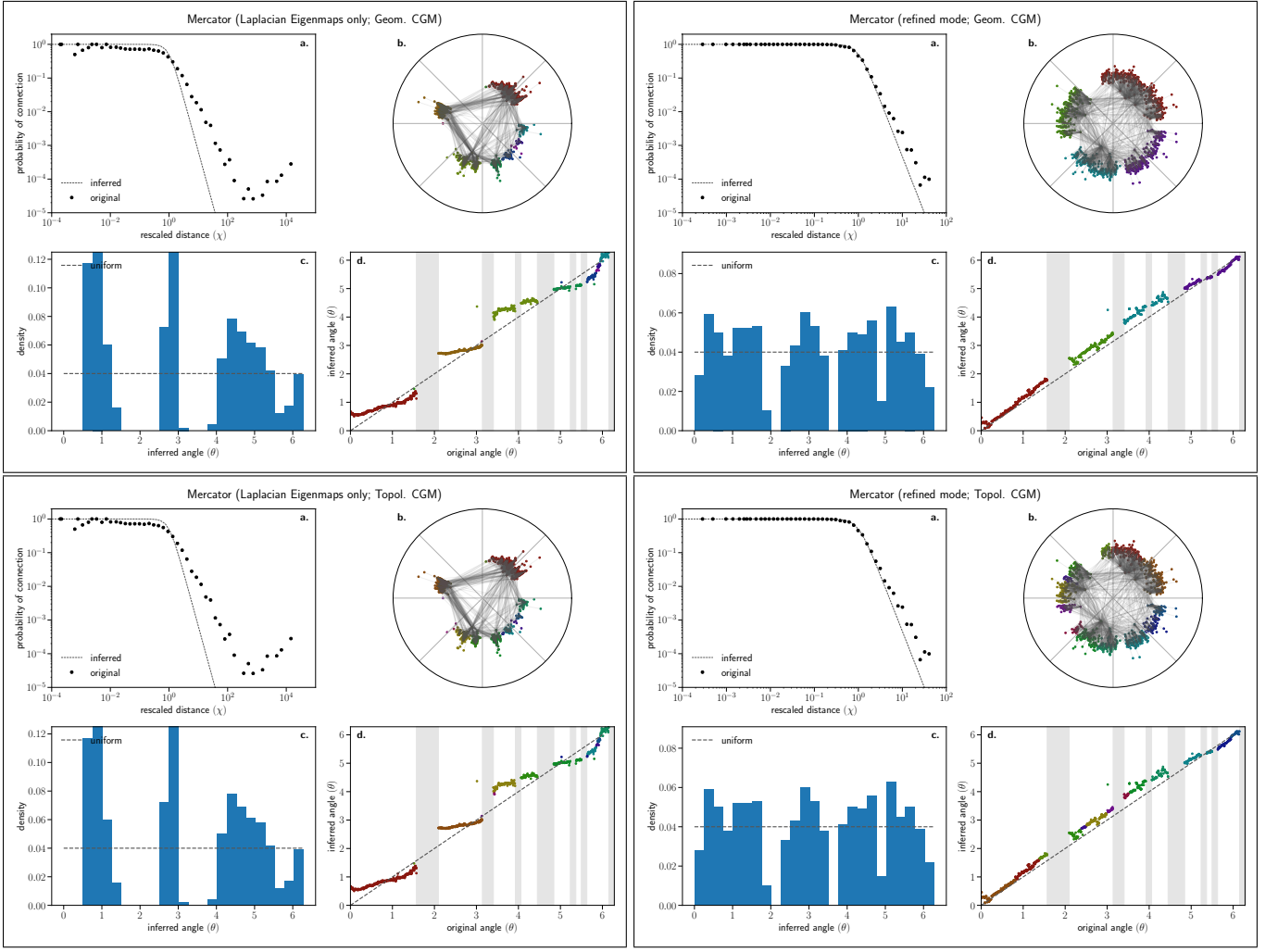


FIG. S2. Network 1. Left column shows the embedding obtained with the fast mode of Mercator when the angular positions are obtained directly from the LE method, i.e. $g_i = g_i^{\text{LE}}$ on Eq. (A15) in the main text. Right column shows the outcome of the refined mode of Mercator, i.e. using Eq. (A15) in its actual form. First row shows communities detected with the geometrical critical gap method and the second row with topological critical gap method. Nodes are colored according to the communities detected by the corresponding method. The grey areas indicate the position of the gaps separating the original angular blocks.

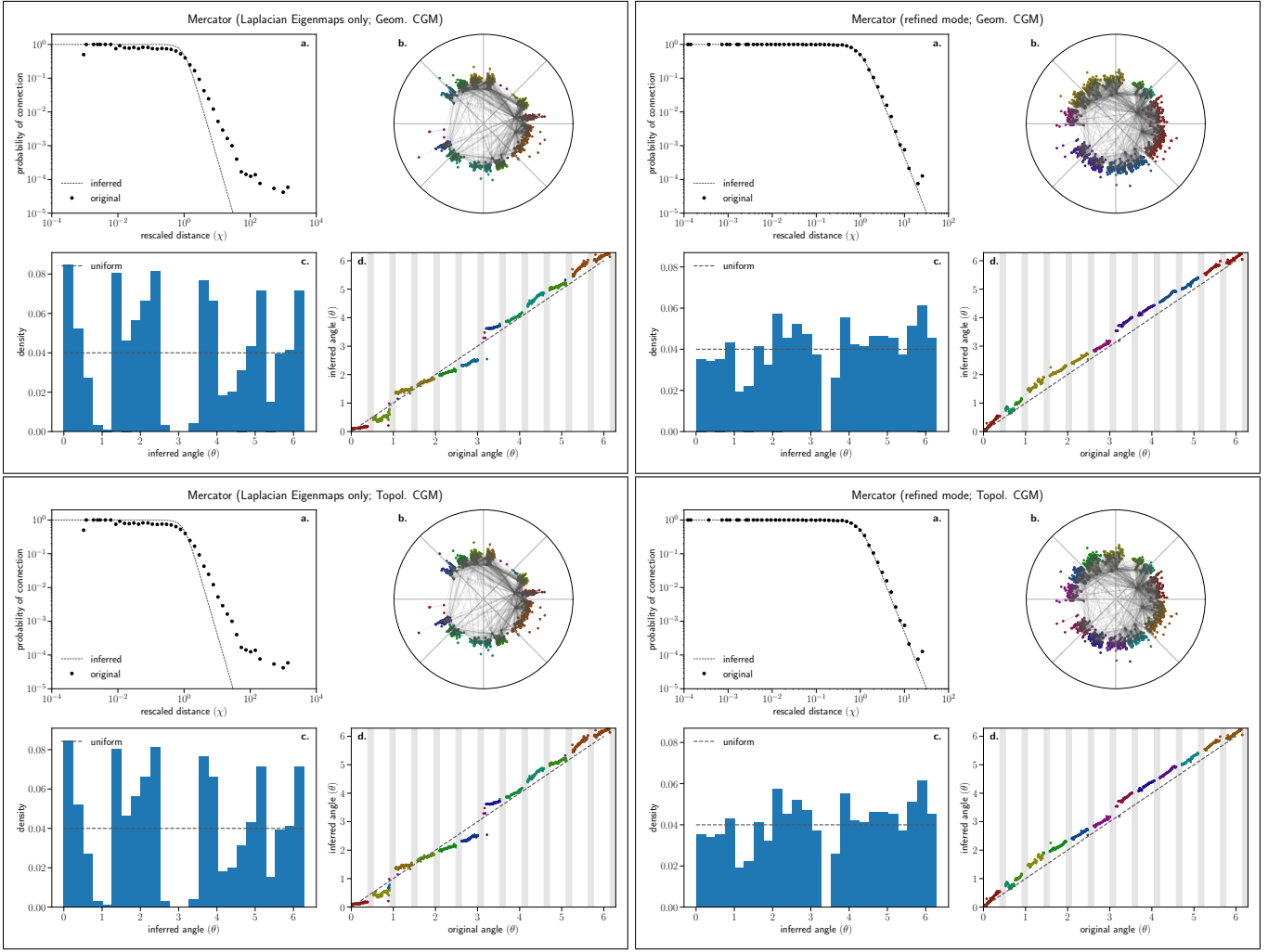


FIG. S3. Network 2. Left column shows the embedding obtained with the fast mode of Mercator when the angular positions are obtained directly from the LE method, i.e. $g_i = g_i^{\text{LE}}$ on Eq. (A15) in the main text. Right column shows the outcome of the refined mode of Mercator, i.e. using Eq. (A15) in its actual form. First row shows communities detected with the geometrical critical gap method and the second row with topological critical gap method. Nodes are colored according to the communities detected by the corresponding method. The grey areas indicate the position of the gaps separating the original angular blocks.

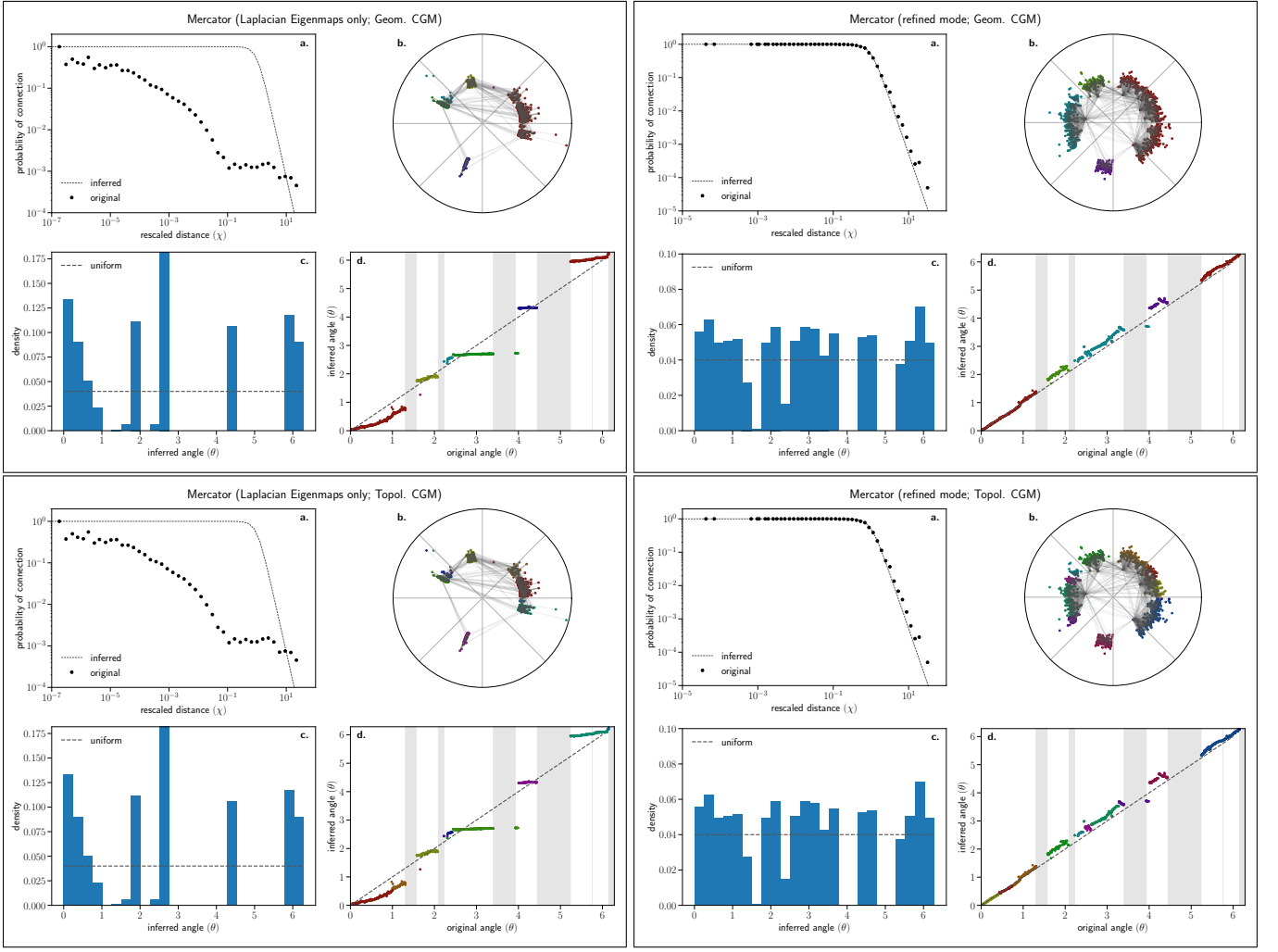


FIG. S4. Network 3. Left column shows the embedding obtained with the fast mode of Mercator when the angular positions are obtained directly from the LE method, i.e. $g_i = g_i^{\text{LE}}$ on Eq. (A15) in the main text. Right column shows the outcome of the refined mode of Mercator, i.e. using Eq. (A15) in its actual form. First row shows communities detected with the geometrical critical gap method and the second row with topological critical gap method. Nodes are colored according to the communities detected by the corresponding method. The grey areas indicate the position of the gaps separating the original angular blocks.

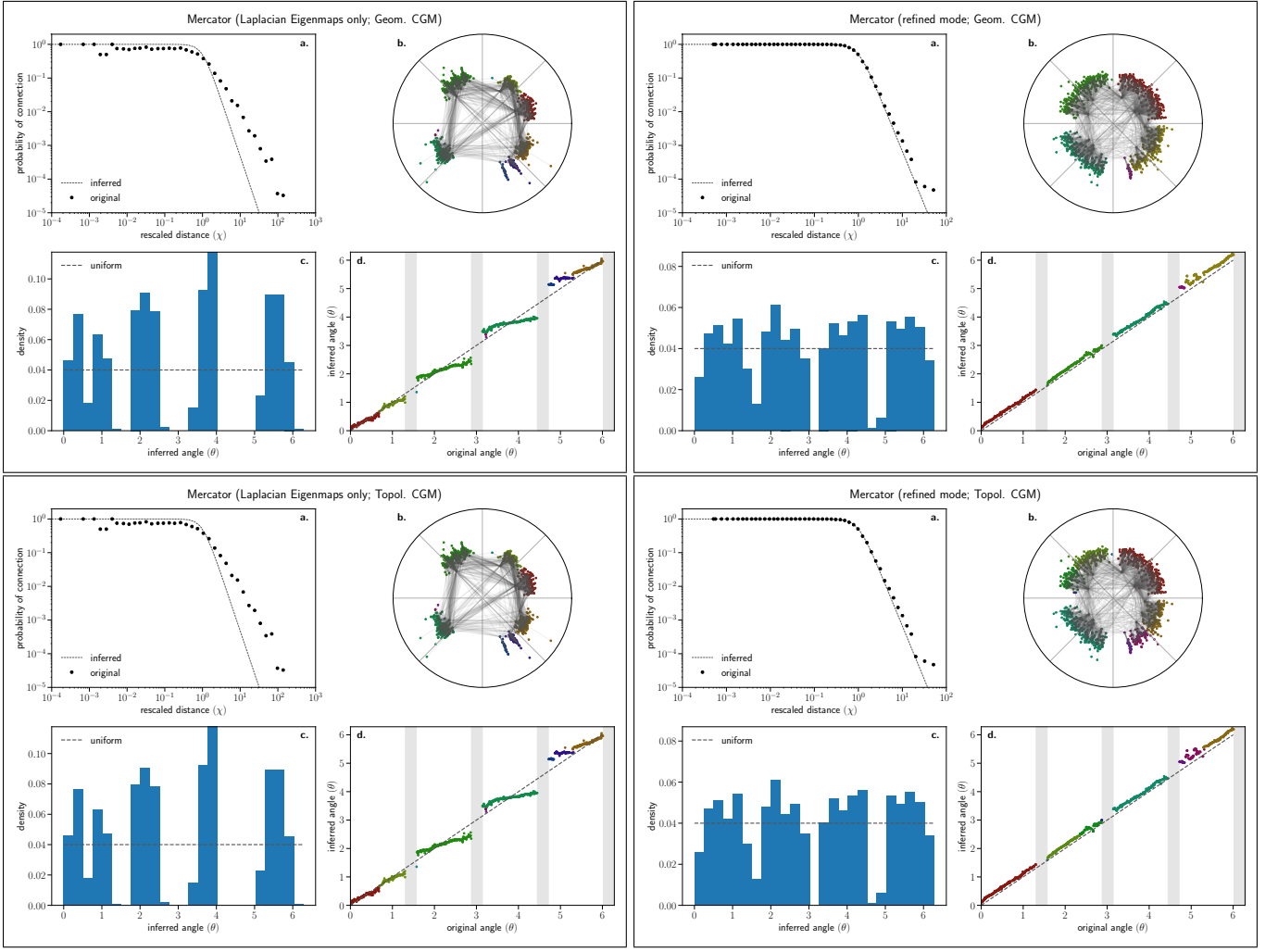


FIG. S5. Network 4. Left column shows the embedding obtained with the fast mode of Mercator when the angular positions are obtained directly from the LE method, i.e. $g_i = g_i^{\text{LE}}$ on Eq. (A15) in the main text. Right column shows the outcome of the refined mode of Mercator, i.e. using Eq. (A15) in its actual form. First row shows communities detected with the geometrical critical gap method and the second row with topological critical gap method. Nodes are colored according to the communities detected by the corresponding method. The grey areas indicate the position of the gaps separating the original angular blocks.

B. Comparison with the nPSO model

We compare Mercator with the Coalescent embedding algorithm proposed in Ref. [1] for synthetic networks generated with the non-uniform Popularity-Similarity-Optimization (PSO) model [3] —a growing geometric model generating complex networks in hyperbolic space with inhomogeneous angular distributions. Figure S6 shows results for networks generated with $\gamma = 2.5$, $N = 1000$, and $C = 9$ different geometric communities.

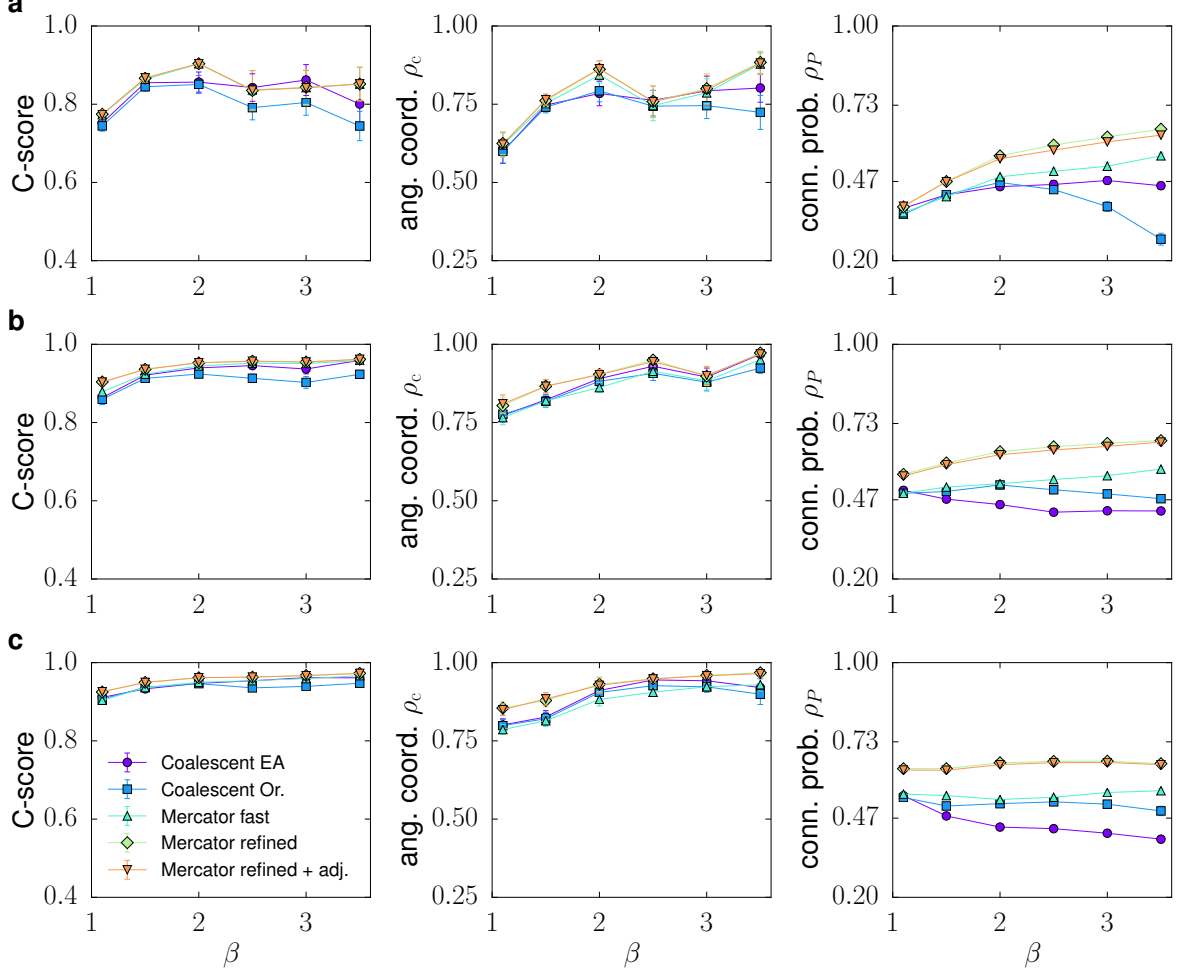


FIG. S6. Comparison between the angular coordinates inferred by different versions of Mercator and the Coalescent embedding method (with and without equidistant adjustment—EA and Or., respectively) in terms of the C-score (left) and the Pearson correlation coefficient between original and inferred angular coordinates (middle), as a function of the clustering parameter β , for networks generated with the nPSO model with $C = 9$ geometric communities. The right column shows the Pearson correlation coefficient between the original and the inferred connection probabilities of all pairs of nodes. Every row corresponds to a different value of $\langle k \rangle$: **a.** 4, **b.** 8, and **c.** 12. For every value of the parameters, we generated and embedded 10 synthetic networks of size $N = 1000$ and with power-law degree distribution exponent $\gamma = 2.5$. The plots show the resulting averages and standard deviations.

II. EMBEDDING REAL NETWORKS

Figures S7–S16 show the quality of the embeddings obtained with Mercator in the refined mode for the 9 other real complex networks mentioned in the main text.

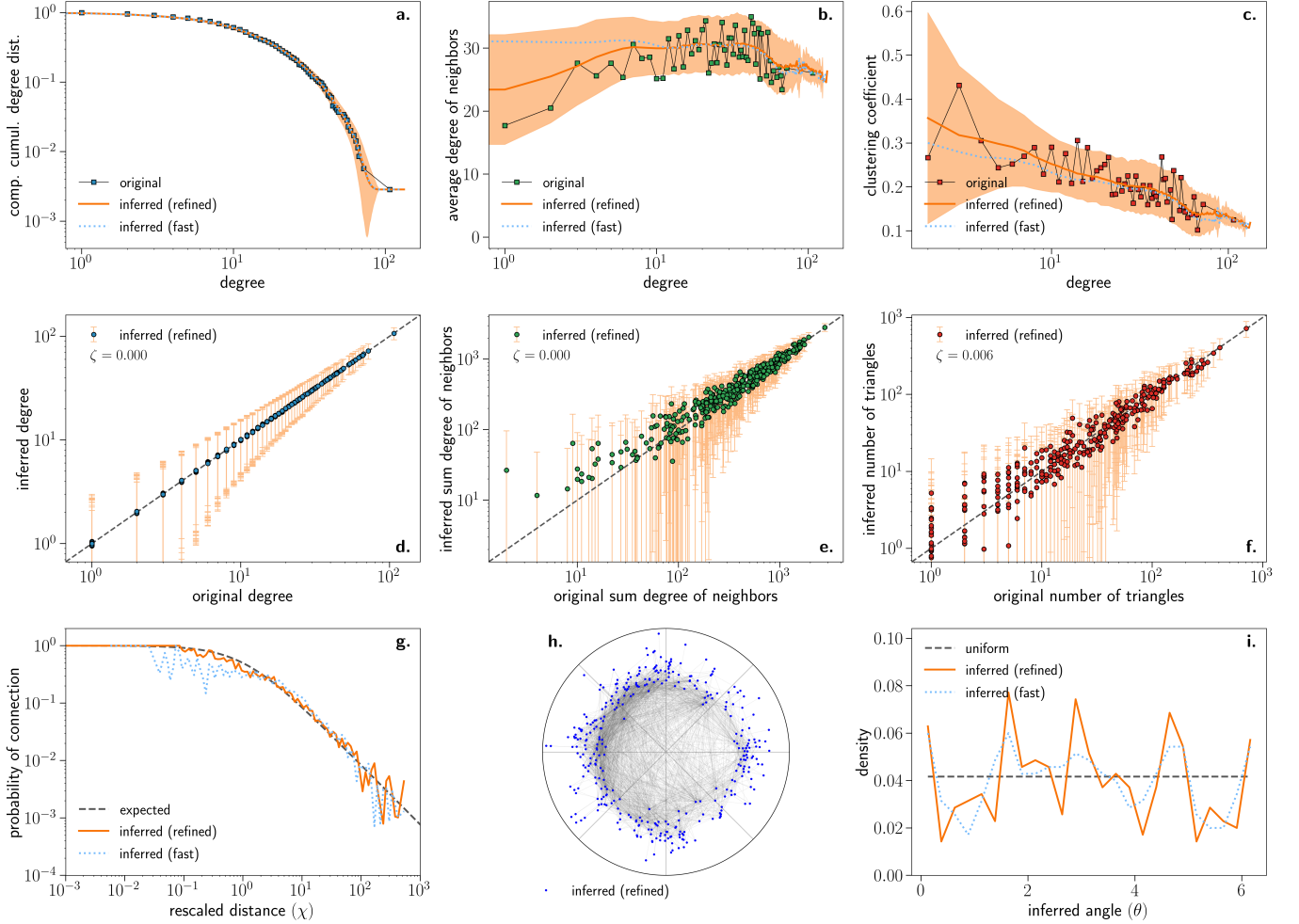


FIG. S7. Topological validation of the embedding of the neural network of the visual cortex of the *Drosophila Melanogaster* at the neuron level [8]. The first row shows **a.** the complementary cumulative degree distribution, **b.** the average nearest neighbors degree, $\bar{k}_{nn}(k)$, and **c.** the clustering spectrum, $\bar{c}(k)$. Symbols correspond to the value of these quantities in the original network, whereas the red lines indicate an estimate of their expected values in the ensemble of random networks inferred by Mercator. This ensemble was sampled by generating 100 synthetic networks with the $\mathbb{S}^1$ model and the inferred parameters and positions by Mercator. The orange regions correspond to an estimate of the 2σ confidence interval around the expected values. The second row shows scattered plots of **d.** the degree of every nodes, **e.** the sum of the degrees of their neighbors, and **f.** the number of triangles to which they participate. The plots show the estimated values of these three measures in the same ensemble of random networks considered above versus the corresponding values in the original network. The error bars show the 2σ confidence interval around the expected values. The quantity ζ corresponds to the fraction of nodes for which the value measured on the original network lies outside the 2σ confidence interval. **g.** Comparison of the expected connection probability based on the inferred value of β (expected) and the actual connection probability computed with the inferred hidden variables $\{k_i, \theta_i\}$ (inferred). **h.** Visualization of the embedding in hyperbolic space using the $\mathbb{H}^2$ model. Only links whose probability of existence in the $\mathbb{S}^1$ model is above 0.5 are shown to avoid cluttering the plot. **i.** Density of the inferred angular coordinates $\{\theta_i\}$. A horizontal dashed line is shown to indicate the uniform density for comparison.

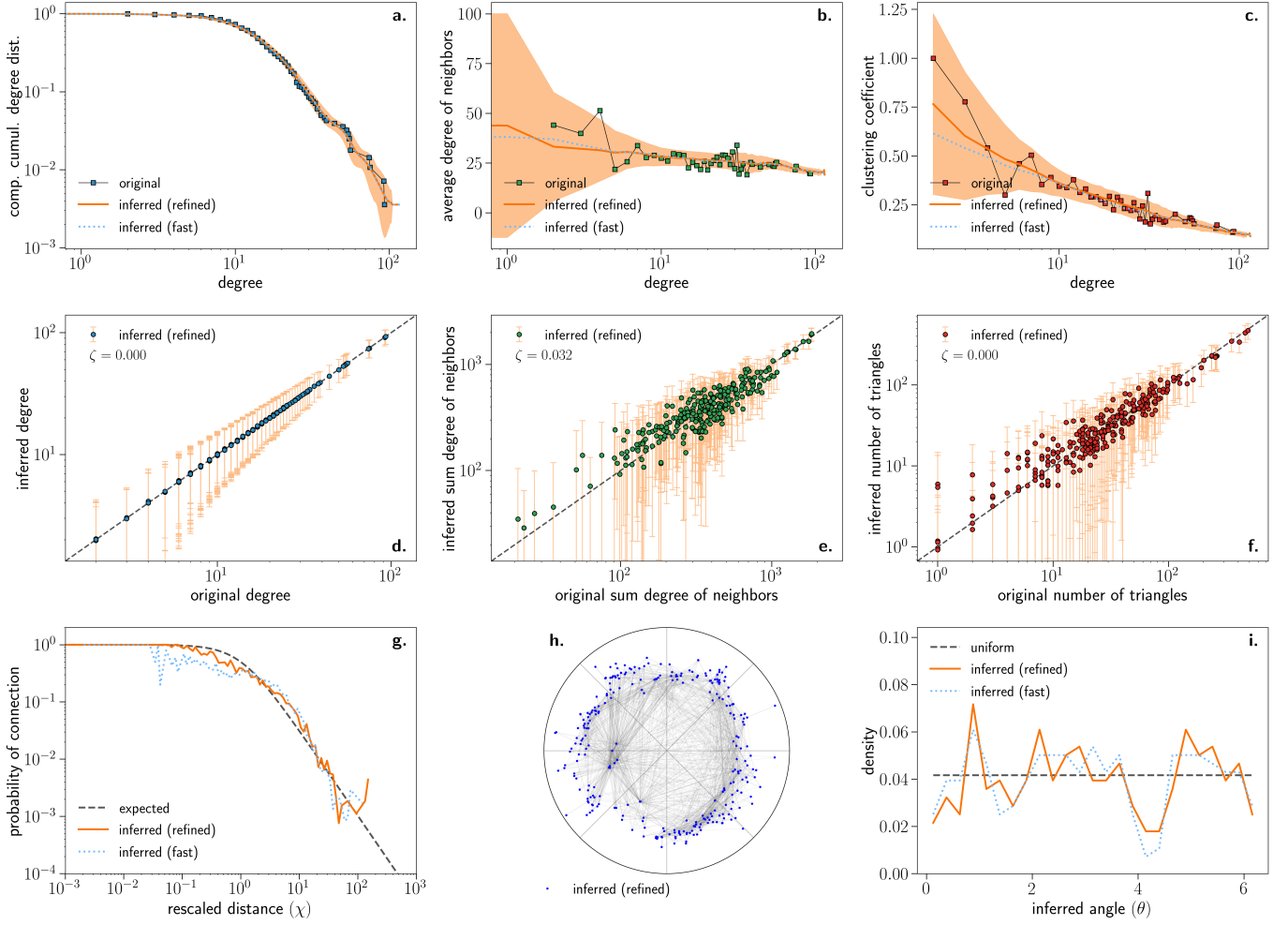


FIG. S8. Topological validation of the embedding of the neural network of the C. Elegans worm [9]. See the caption of Fig. S7 for details.

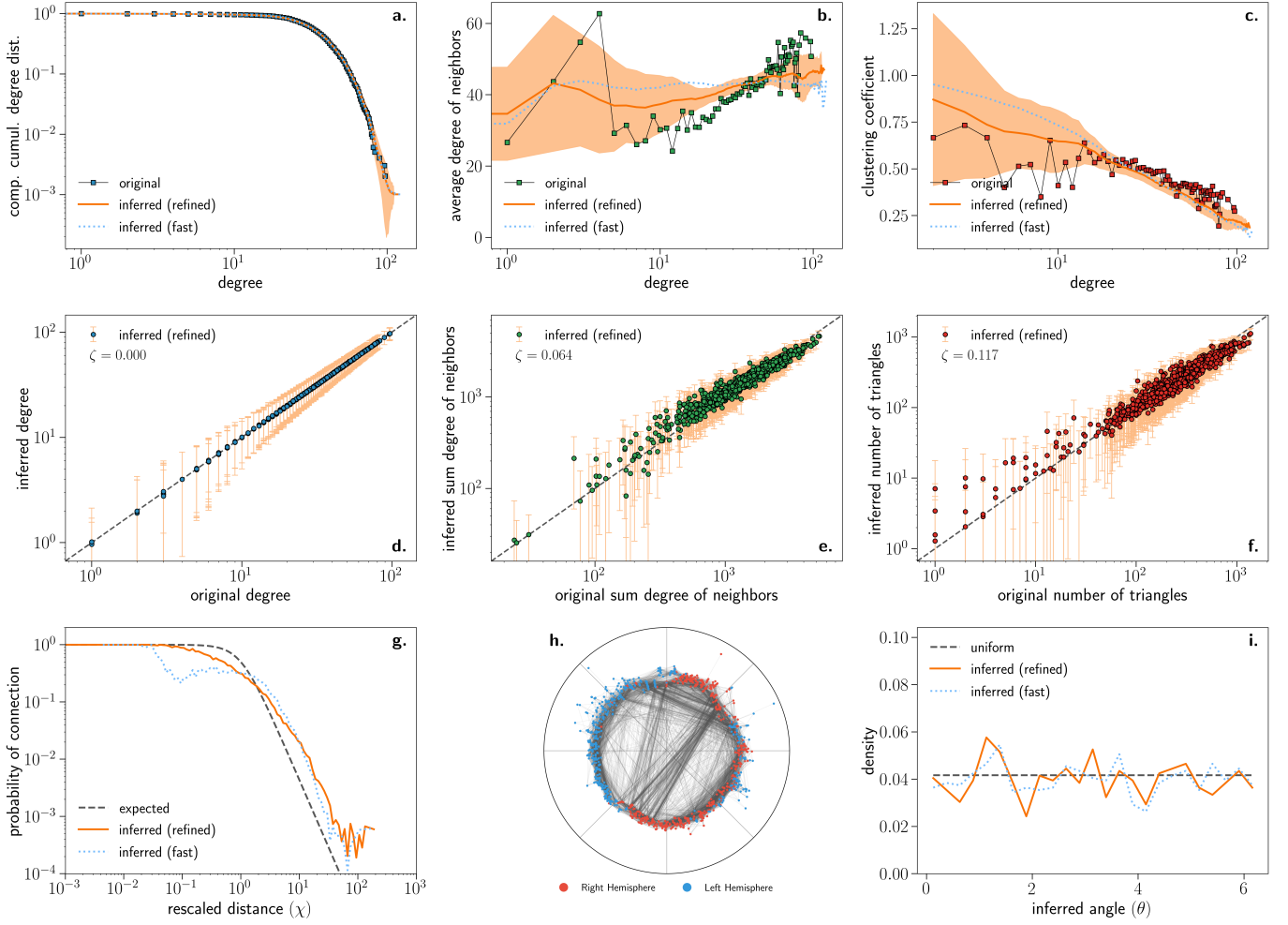


FIG. S9. Topological validation of the embedding of a human connectome [10, 11]. Nodes are colored according to the hemisphere in which they are located, information that is not used to obtain the embedding. See the caption of Fig. S7 for details.

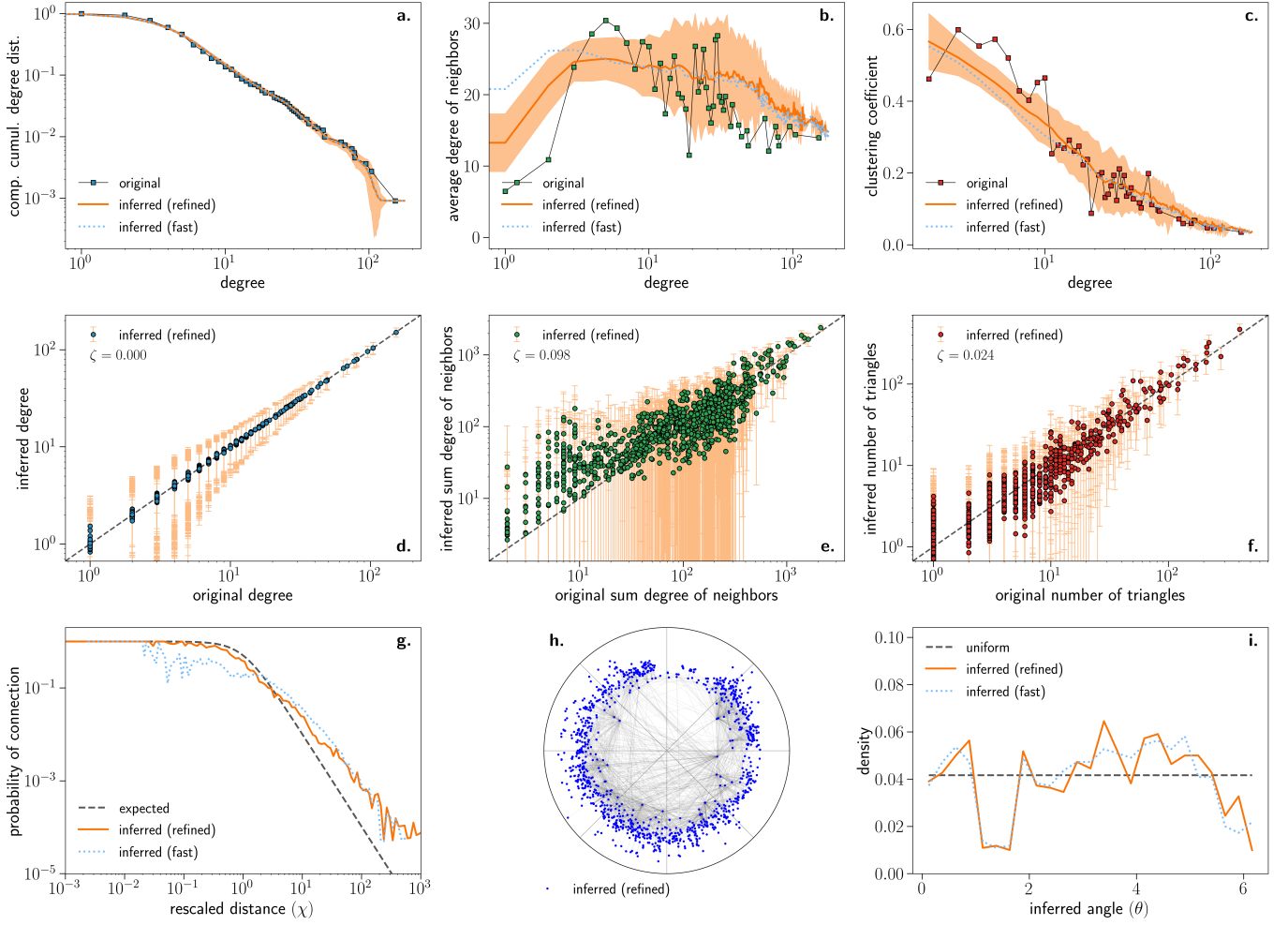


FIG. S10. Topological validation of the embedding of the metabolic network of the bacterium *E. Coli* [5, 12]. See the caption of Fig. S7 for details.

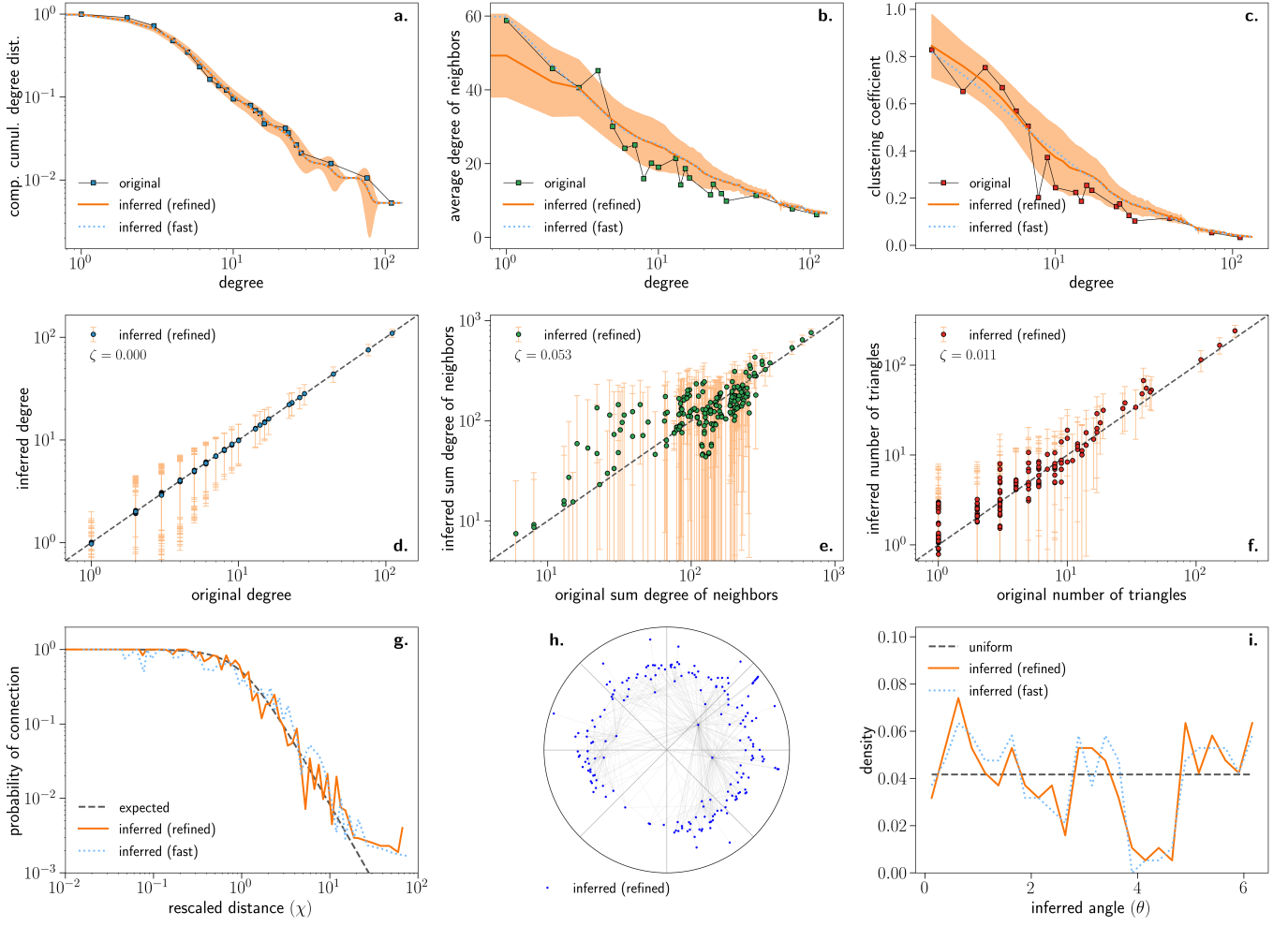


FIG. S11. Topological validation of the embedding of the world trade web [7]. See the caption of Fig. S7 for details.

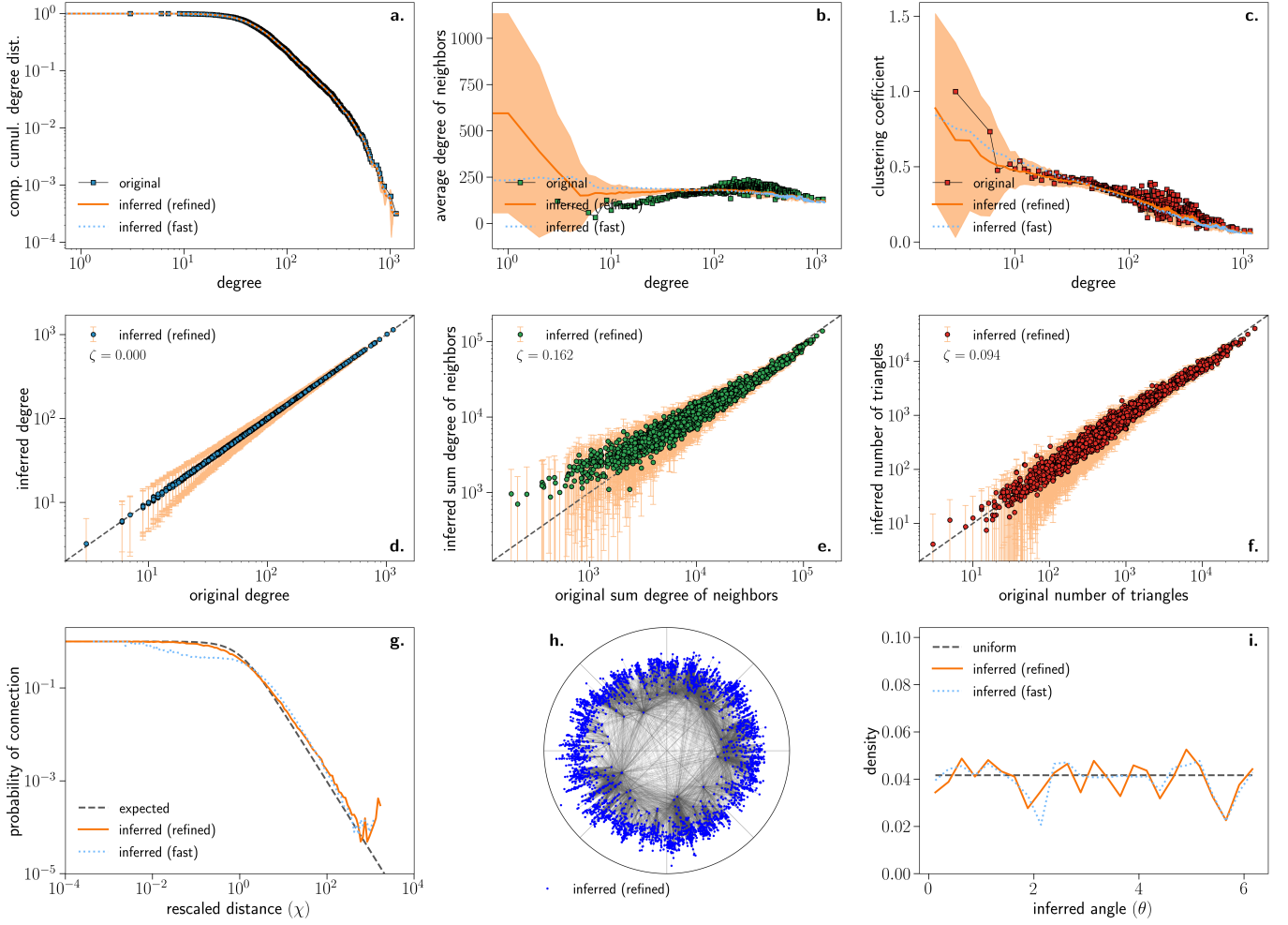


FIG. S12. Topological validation of the embedding of a US commute network [13]. See the caption of Fig. S7 for details.

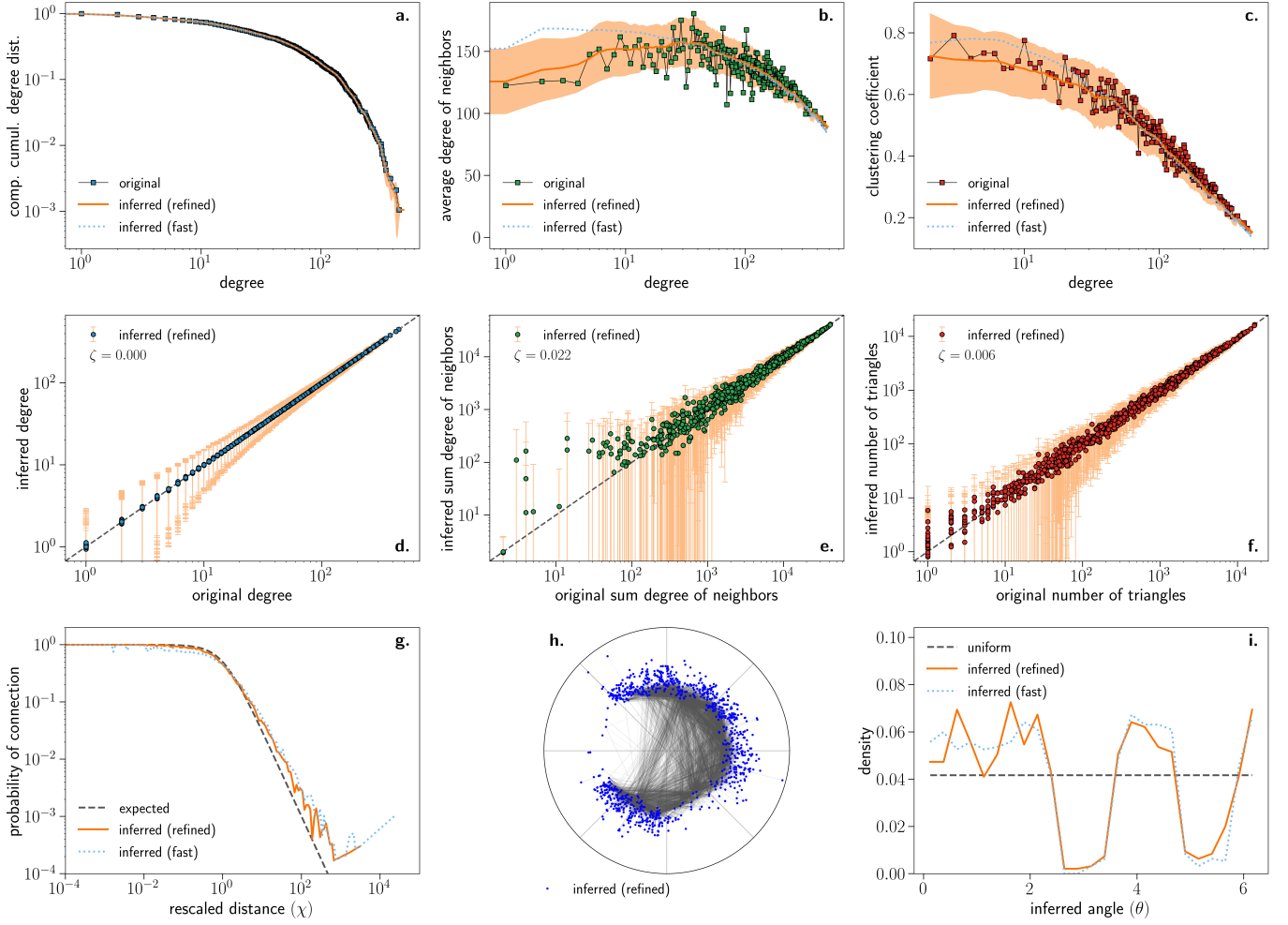


FIG. S13. Topological validation of the embedding of a cargo ships network [14]. See the caption of Fig. S7 for details.

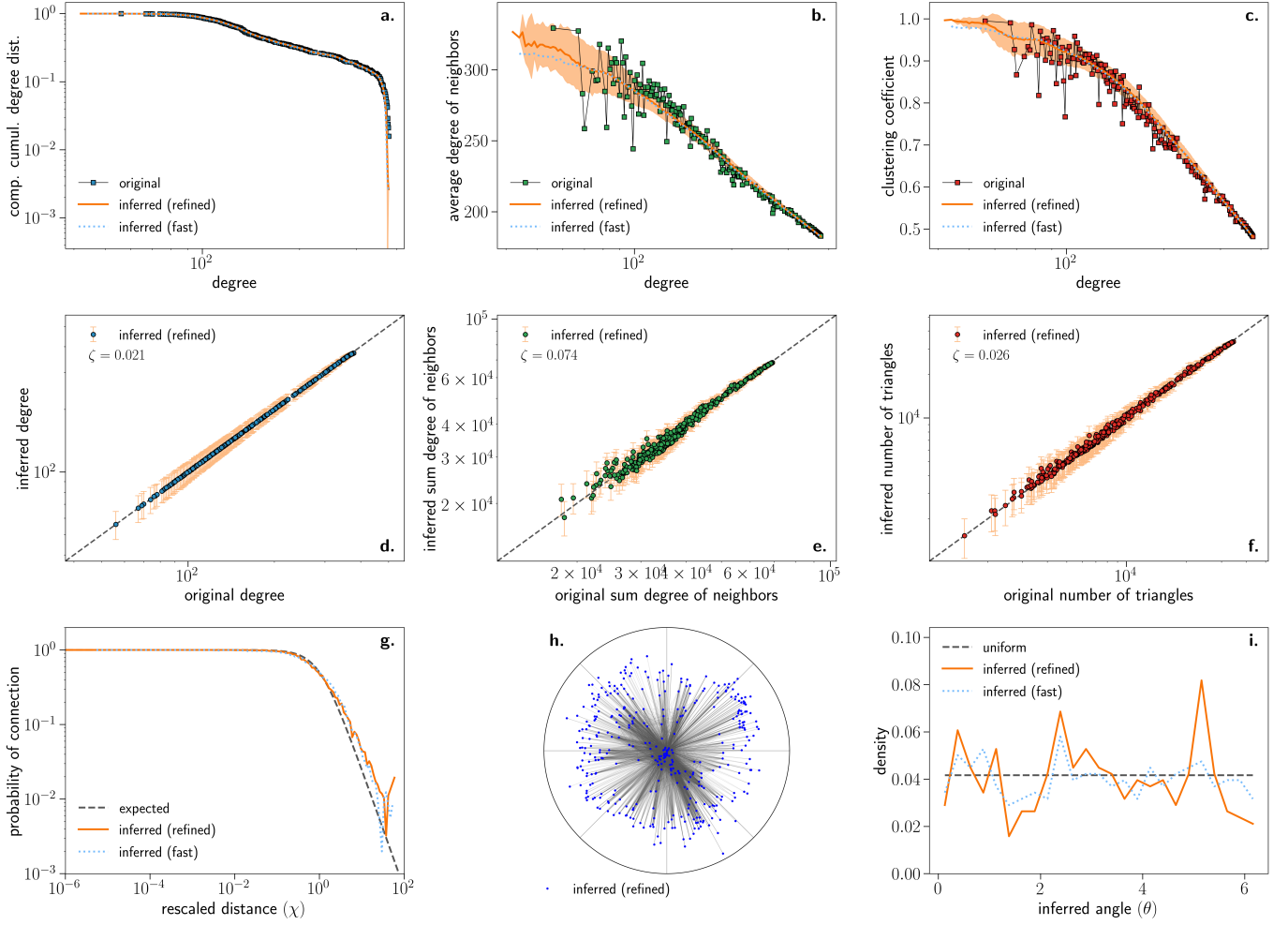
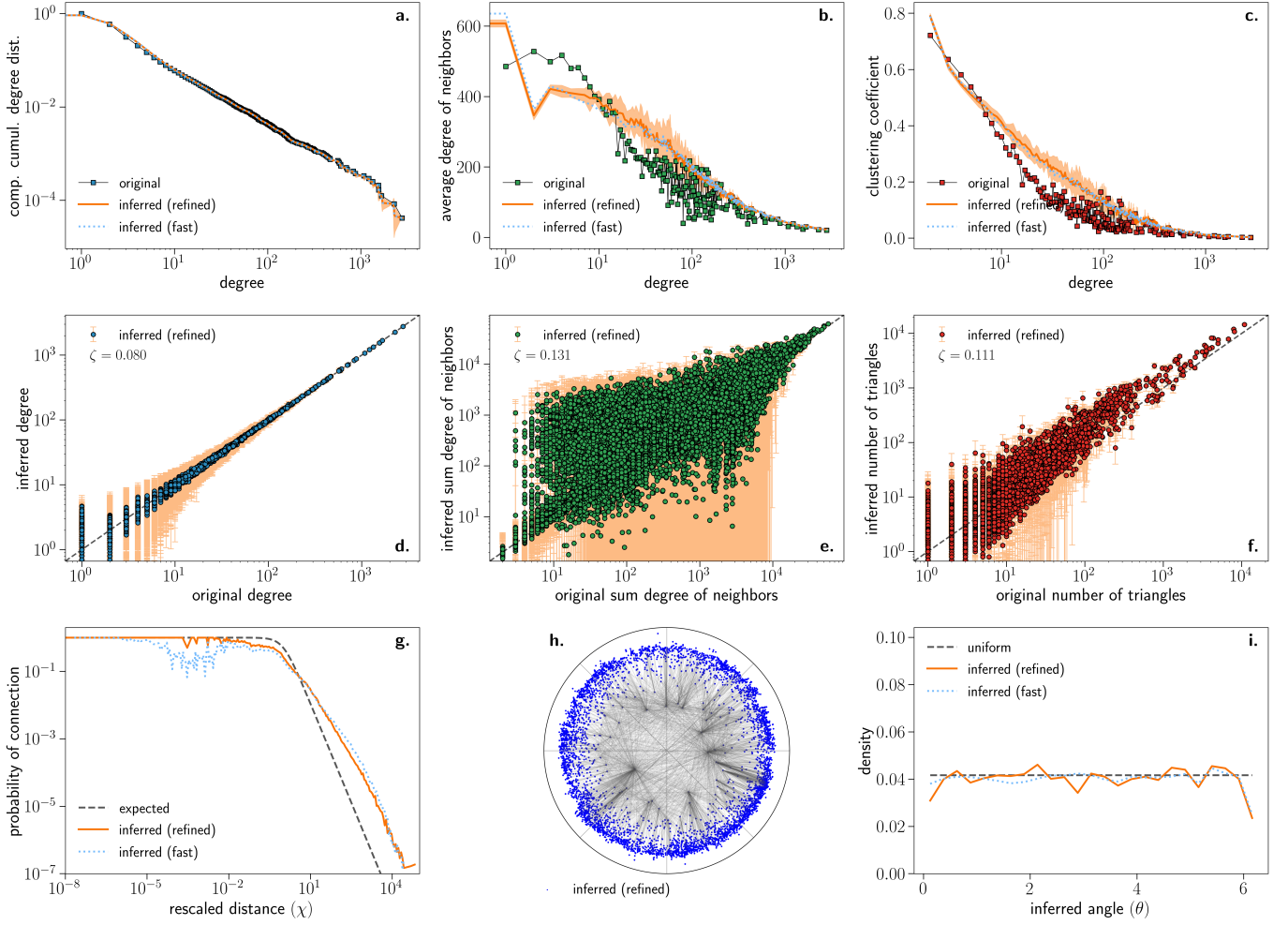


FIG. S14. Topological validation of the embedding of a US commodities network [13]. Only the 30% most likely links according to the $\mathbb{S}^1$ model are shown on **h** to avoid clustering the plot. See the caption of Fig. S7 for details.



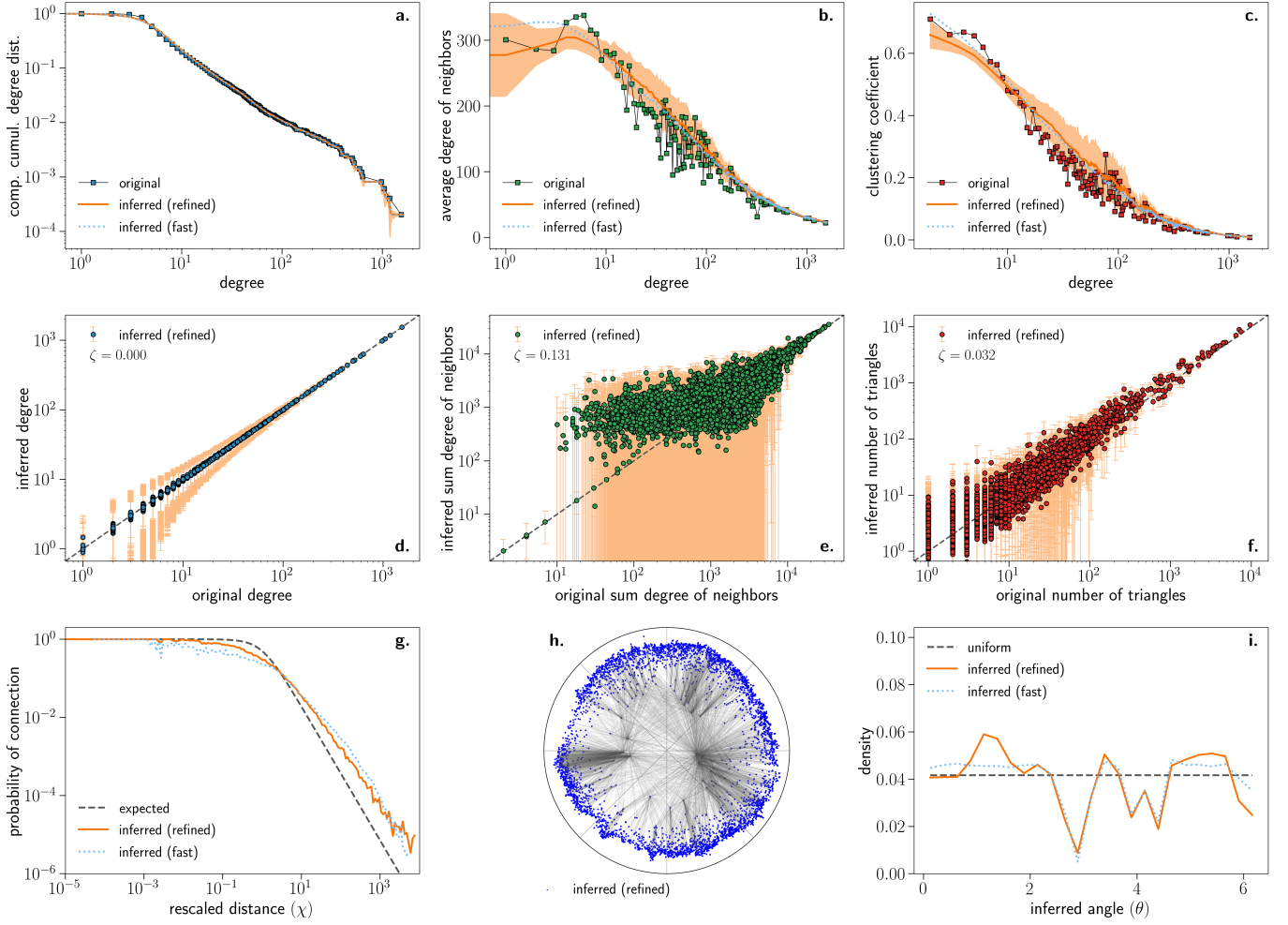


FIG. S16. Topological validation of the embedding of the Internet at the Autonomous Systems level [15] using nodes with an original degree larger than 3 (before removing nodes). Only nodes with degree greater than 3 are shown to avoid cluttering the plot. See the caption of Fig. S7 for details.

-
- [1] 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).
 - [2] Fragkiskos Papadopoulos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá, and Dmitri Krioukov, “Popularity versus similarity in growing networks,” *Nature* **489**, 537–540 (2012).
 - [3] Alessandro Muscoloni and Carlo Vittorio Cannistraci, “A nonuniform popularity-similarity optimization (npso) model to efficiently generate realistic complex networks with communities,” *New Journal of Physics* **20**, 052002 (2018).
 - [4] M. Ángeles Serrano, Dmitri Krioukov, and Marián Boguñá, “Self-similarity of complex networks and hidden metric spaces,” *Physical Review Letters* **100**, 078701 (2008).
 - [5] M. Á. Serrano, M. Boguñá, and F. Sagués, “Uncovering the hidden geometry behind metabolic networks,” *Molecular BioSystems* **8**, 843–850 (2012).
 - [6] K. Zuev, M. Boguñá, G. Bianconi, and D. Krioukov, “Emergence of soft communities from geometric preferential attachment,” *Scientific reports* **5**, 9421 (2015).
 - [7] G. García-Pérez, M. Boguñá, A. Allard, and M. Á. Serrano, “The hidden hyperbolic geometry of international trade: World Trade Atlas 1870-2013,” *Scientific Reports* **6**, 33441 (2016).
 - [8] S.-Y. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, T. Zhao, J. A. Horne, R. D. Fetter, S.-Y. Takemura, K. Blazek, L.-A. Chang, O. Ogundeyi, M. A. Saunders, V. Shapiro, C. Sigmund, G. M. Rubin, L. K. Scheffer, I. A. Meinertzhagen, and D. B. Chklovskii, “A visual motion detection circuit suggested by *Drosophila* connectomics,” *Nature* **500**, 175–181 (2013).
 - [9] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii, “Structural Properties of the *Caenorhabditis elegans* Neuronal Network,” *PLoS Comput. Biol.* **7**, e1001066 (2011).
 - [10] A. Avena-Koenigsberger, B. Misic, and O. Sporns, “Communication dynamics in complex brain networks,” *Nature Reviews Neuroscience* **19**, 17 (2018).
 - [11] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, “Mapping the Structural Core of Human Cerebral Cortex,” *PLoS Biol.* **6**, e159 (2008).
 - [12] J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist, and B. O. Palsson, “A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011,” *Mol. Syst. Biol.* **7**, 535 (2011).
 - [13] D. Grady, C. Thiemann, and D. Brockmann, “Robust classification of salient links in complex networks,” *Nat. Commun.* **3**, 864 (2012).
 - [14] P. Kaluza, A. Kolzsch, M. T. Gastner, and B. Blasius, “The complex network of global cargo ship movements,” *J. R. Soc. Interface* **7**, 1093–1103 (2010).
 - [15] M. Boguñá, F. Papadopoulos, and D. Krioukov, “Sustaining the internet with hyperbolic mapping,” *Nature Communications* **1**, 62 (2010).