Supplementary Information for Hidden geometric correlations in real multiplex networks

Kaj-Kolja Kleineberg,[1] Marián Boguñá,[1] M. Ángeles Serrano,[2,1] and Fragkiskos Papadopoulos[3]

[1] *Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*
[2] *Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, E-08010 Barcelona, Spain*
[3] *Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, 33 Saripolou Street, 3036 Limassol, Cyprus*

CONTENTS

I. REAL-WORLD MULTIPLEX NETWORK DATA

Here we provide details on the real-world multiplex network data we have considered. An overview of the data is given in Table I of the main text.

IPv4/IPv6 Internet. The IPv4 and IPv6 Autonomous Systems (AS) Internet topologies were extracted from the data collected by the Archipelago active measurement infrastructure (ARK) developed by CAIDA [1]. The connections in each topology are not physical but logical, representing AS relationships. An AS is a part of the Internet infrastructure administrated by a single company or organization. Pairs of ASs peer to exchange traffic. These peering relationships in the AS topology are represented as links between AS nodes. CAIDA's IPv4 and IPv6 [2] datasets provide regular snapshots of AS links derived from ongoing traceroute-based IP-level topology measurements. The IPv4 dataset consists of ASs that can route Internet packets with IPv4 destination addresses, while the IPv6 dataset consists of ASs that can route packets with IPv6 destination addresses. The considered IPv4 and IPv6 topologies were constructed by merging the AS link snapshots during the first 15 days of January 2015, which are provided at [3]. The IPv4 topology (Layer 1) consists of $N_1 = 37563$ nodes (ASs), and has a power law degree distribution with exponent $\gamma_1 = 2.1$, average node degree $\bar{k}_1 = 5.06$, and average clustering $\bar{c}_1 = 0.63$. The IPv6 topology (Layer 2) consists of $N_2 = 5163$ nodes, has a power law degree distribution with exponent $\gamma_2 = 2.1$, average node degree $\bar{k}_2 = 5.21$, and average clustering $\bar{c}_2 = 0.55$. There are 4819 common nodes in the two topologies, i.e., ASs that can route both IPv4 and IPv6 packets.

Air/Train. The Air/Train data is taken from [4]. The data contains the network of airports and the network of train stations in India, as well as the geographic distances between the airports and the train stations. For each airport, we aggregate all train stations that are within 50km from the airport into a supernode. Subsequently, we declare two supernodes connected if they have at least one train station in common, or if at least one train station from the one supernode is directly connected to a train station from the other supernode. If there are no train stations within 50km from an airport, we consider the nearest train station to the airport, which is considered a supernode on its own. Each supernode has the same id as its corresponding airport, i.e., it is considered to be the same node in the multiplex system. The idea behind this aggregation procedure is to relate train stations to the airports to which they are geographically close. The considered multiplex consists of the network of airports (Air) and the network of aggregated supern-

odes of train stations (Train). The two networks consist of $N_1 = N_2 = 69$ common nodes. The Air network (Layer 1) has average degree $\bar{k}_1 = 5.22$, maximum degree $k_1^{\max} = 42$, and average clustering $\bar{c}_1 = 0.79$. The Train network (Layer 2) has average degree $\bar{k}_2 = 9.33$, maximum degree $k_2^{\max} = 41$, and average clustering $\bar{c}_2 = 0.48$.

Drosophila Melanogaster. The Drosophila Melanogaster dataset is taken from [5, 6]. In this dataset, the networks represent protein–protein interactions and the layers correspond to interactions of different nature. Layer 1 in our multiplex corresponds to suppressive genetic interaction, while layer 2 corresponds to additive genetic interaction. More details on the data can be found in [5, 6]. Layer 1 has $N_1 = 838$ nodes, average degree $\bar{k}_1 = 4.43$, and average clustering $\bar{c}_1 = 0.28$. Its degree distribution can be approximated by a power law with exponent $\gamma_1 = 2.6$. Layer 2 has $N_2 = 755$ nodes, average degree $\bar{k}_2 = 3.77$, and average clustering $\bar{c}_2 = 0.29$. Its degree distribution can be approximated by a power law with exponent $\gamma_2 = 2.8$. There are 557 common nodes in the two layers.

C. Elegans Connectomme. The C. Elegans dataset is taken from [7, 8]. It corresponds to the neuronal network of the nematode Caenorhabditis Elegans. The nodes are neurons and each layer corresponds to a different type of synaptic connection: Electric (Layer 1), Chemical Monadic (Layer 2), and Chemical Polyadic (Layer 3). Layer 1 has $N_1 = 253$ nodes, average degree $\bar{k}_1 = 4.06$, and average clustering $\bar{c}_1 = 0.24$. Layer 2 has $N_2 = 260$ nodes, average degree $\bar{k}_2 = 6.83$, and average clustering $\bar{c}_2 = 0.21$. Layer 3 has $N_3 = 278$ nodes, average degree $\bar{k}_3 = 12.25$, and average clustering $\bar{c}_3 = 0.29$. The degree distribution in all three layers can be approximated by a power law with exponent $\gamma_1 = \gamma_2 = \gamma_3 = 2.9$.

Human Brain. The human brain data is taken from [9]. The data consists of a structural (anatomical) network, as well as a functional network obtained by an algebraic aggregation procedure. In both networks, nodes are brain regions—there are 90 different brain regions in the data. The structural network is obtained by Diffusion Magnetic Resonance Imaging (dMRI). For each pair of brain regions, the data gives the probability that these regions are connected. The connection probability is proportional to the density of the axonal fibers between the regions. In our multiplex, we declare two regions of the structural network connected if their connection probability is larger than a threshold $th_s = 0.92$. The functional network is obtained by BOLD fMRI resting state recordings for the same brain regions. The probability that two regions are connected here is proportional to a correlation coefficient between the fMRI time series of the region voxels [9]. In our multiplex, we declare two regions of the functional network connected if their correlation coefficient is larger than the threshold $th_f = 0.67$. The resulting structural network (Layer 1) consists of a giant connected component of 85 nodes, with average degree $\bar{k}_1 = 5.41$, maximum degree $k_1^{\max} = 12$, and average clustering $\bar{c}_1 = 0.49$. The resulting functional network (Layer 2) has a giant connected component of 78 nodes, average degree $\bar{k}_2 = 5.48$, maximum degree $\bar{k}_2^{\max} = 14$, and average clustering $\bar{c}_2 = 0.40$. The two layers have 77 nodes in common.

SacchPomb. The SacchPomb dataset is taken from [5, 6] and represents the multiplex genetic and protein interaction network of the Saccharomyces Pombe (fission yeast). The multiplex consists of 5 layers corresponding to 5 different types of interactions. Layer 1 corresponds to direct interaction, Layer 2 to colocalization, Layer 3 to physical association, Layer 4 to synthetic genetic interaction, and Layer 5 to association. More details on the data can be found in [5, 6]. The corresponding size N, average degree $\bar{k}$, average clustering $\bar{c}$, and power law exponent γ for each layer is shown in Table I.

Rattus. The Rattus dataset is taken from [5, 6] and represents the multiplex genetic and protein interaction network of the Rattus Norvegicus (common rat). Layer 1 corresponds to physical association and Layer 2 to direct interaction. The corresponding size N, average degree $\bar{k}$, average clustering $\bar{c}$, and power law exponent γ for each layer is shown in Table I.

arXiv. The arXiv data is taken from [10] and contains co-authorship networks from the free scientific repository arXiv. The nodes are authors that are connected if they have co-authored a paper. In arXiv, each paper is assigned to one or more relevant categories. The data considers only papers with the word "networks" in the title or abstract from different categories up to May 2014. We consider 8 different categories (layers). Layer 1 corresponds to the co-authorship network formed by the authors of papers in the "Biological Physics" (physics.bio-ph) category. Similarly, Layer 2 corresponds to "Disordered Systems and Neural Networks" (cond-mat.dis-nn), Layer 3 to "Physics and Society" (physics.soc-ph), Layer 4 to "Data Analysis, Statistics and Probability" (physics.data-an), Layer 5 to "Optimization and Control" (math.OC), Layer 6 to "Statistical Mechanics" (cond-mat.stat-mech), Layer 7 to "Molecular Networks" (q-bio.MN), and Layer 8 to "Social and Information Networks" (cs.SI). The corresponding size N, average degree $\bar{k}$, average clustering $\bar{c}$, and power law exponent γ for each layer is shown in Table I.

Physicians. The Physicians dataset is taken from [11] and contains 3 network layers corresponding to different types of relations among physicians in four US towns. Layers 1, 2, 3 correspond respectively to advice, discussion, and friendship relations among the physicians. The corresponding size N, average degree $\bar{k}$, average clustering $\bar{c}$, and power law exponent γ for each layer is shown in Table I.

II. HYPERBOLIC MAPPING

We map each layer of each real multiplex to its hyperbolic space using the *HyperMap* method [12, 13], whose implementation is available at [14]. On its in-

Layer	N	$\bar{k}$	$\bar{c}$	γ	m	T
Internet Layer 1	37563	5.06	0.63	2.1	1.0	0.5
Internet Layer 2	5163	5.21	0.55	2.1	1.0	0.5
Air/Train Layer 1	69	5.22	0.79	2.6	1.0	0.005
Air/Train Layer 2	69	9.33	0.48	2.9	1.0	0.4
Drosophila Layer 1	838	4.43	0.28	2.6	0.5	0.68
Drosophila Layer 2	755	3.77	0.29	2.8	0.5	0.65
C. Elegans Layer 1	253	4.06	0.24	2.9	2.0	0.65
C. Elegans Layer 2	260	6.83	0.21	2.9	3.4	0.7
C. Elegans Layer 3	278	12.25	0.29	2.9	6.1	0.65
Brain Layer 1	85	5.41	0.49	6.0	2.7	0.4
Brain Layer 2	78	5.48	0.40	6.0	1.0	0.5
arXiv Layer 1	2956	4.13	0.83	2.6	2.0	0.05
arXiv Layer 2	3506	4.19	0.81	2.6	2.0	0.05
arXiv Layer 3	1594	3.79	0.92	6.0	1.9	0.05
arXiv Layer 4	5465	5.30	0.83	2.5	2.0	0.05
arXiv Layer 5	1605	5.52	0.89	4.0	2.8	0.05
arXiv Layer 6	1451	3.56	0.88	4.0	1.8	0.05
arXiv Layer 7	1905	4.64	0.88	4.0	2.3	0.05
arXiv Layer 8	4946	4.69	0.85	2.5	2.0	0.05
SacchPomb Layer 1	936	2.85	0.18	2.9	1.0	0.75
SacchPomb Layer 2	346	2.14	0.28	3.3	1.0	0.57
SacchPomb Layer 3	2400	5.81	0.42	2.1	1.0	0.62
SacchPomb Layer 4	897	5.66	0.26	2.8	1.0	0.7
SacchPomb Layer 5	181	2.41	0.64	2.2	1.0	0.25
Rattus Layer 1	2032	2.73	0.26	2.05	1.0	0.68
Rattus Layer 2	1010	1.94	0.05	2.5	0.97	0.85
Physicians Layer 1	215	4.18	0.28	2.7	2.1	0.65
Physicians Layer 2	231	2.31	0.28	2.8	2.2	0.65
Physicians Layer 3	288	3.71	0.24	10	1.9	0.65

Table I. **Topological properties and HyperMap parameter values for the considered empirical multiplex networks.**

put the method takes the network adjacency matrix α_{ij} ($\alpha_{ij} = \alpha_{ji} = 1$ if there is a link between nodes i and j, and $\alpha_{ij} = \alpha_{ji} = 0$ otherwise), and the network parameters m, γ, T. It then computes radial and angular coordinates r_i, θ_i, for all nodes $i \leq N$ in the network. Parameter m is the expected minimum node degree, γ is the power law degree distribution exponent, and T is the temperature. The values of m, γ, T used to embed each layer are shown in Table I.

To estimate the values of m, γ, T for each layer, we use the Extended Popularity×Similarity Optimization (E-PSO) model described in [12]. The E-PSO model grows synthetic complex networks and it is equivalent to the hyperbolic $\mathbb{H}^2$ model [15]. It takes as input the final network size N, the average node degree $\bar{k}$, and the network parameters m, γ, T. We use the E-PSO model to construct synthetic networks with the same size N and average degree $\bar{k}$ as in each real layer, using different parameter values for m, γ, T. The estimated m, γ, T values for each layer are then the values that best match the degree distribution and average clustering between the layer and the corresponding synthetic network.

HyperMap is based on Maximum Likelihood Estimation. It finds the radial and angular coordinates r_i, θ_i for all nodes $i \leq N$, which maximize the likelihood

$$\mathcal{L} = \prod_{1 \leq j < i \leq N} p(x_{ij})^{\alpha_{ij}} \left[1 - p(x_{ij})\right]^{1-\alpha_{ij}}, \qquad (1)$$

where the product goes over all node pairs i, j in the network, x_{ij} is the hyperbolic distance between pair i, j,

$$\begin{aligned} x_{ij} =& \operatorname{arccosh}\left(\cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta\theta_{ij}\right) \\ \approx& r_i + r_j + 2 \ln \sin\left(\Delta\theta_{ij}/2\right), \\ \approx& r_i + r_j + 2 \ln\left(\Delta\theta_{ij}/2\right), \end{aligned} \qquad (2)$$

where $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$ and $p(x_{ij})$ is the Fermi-Dirac connection probability,

$$p(x_{ij}) = \frac{1}{1 + e^{\frac{1}{2T}(x_{ij} - R)}}, \quad (3)$$

where $R \sim \ln N$. To efficiently and accurately maximize the likelihood in Eq. (1) the method follows the techniques described in [12, 13]. Here, we have used the most recent version of the method described in [13].

The inferred radial coordinate of a node i depends on its observed degree in the network k_i via $r_i \sim \ln N - \ln k_i$ [12]. Therefore, if node degrees are correlated in different layers so will be the radial coordinates. Fig. S1 shows the conditional probability $P(r_2|r_1)$ for layer pairs of some of our real multiplexes, which is the probability that a node has radial coordinate r_2 in layer 2 given its radial coordinate r_1 in layer 1. We observe strong correlations among the radial coordinates of nodes in the different layers. Similar results hold for the rest of our multiplexes. In Section X, we quantify both the radial and angular correlations across the layers of all of our real multiplexes.

We note that a recent work also studying latent geometry inference problems in a multiplex setting is the one in [16]. However, that work addresses a very different problem (low distortion reconstruction of the latent space of individual categories/layers from an aggregated network), under certain assumptions (local category–disjointness, etc.).

III. DESTROYING TRANS-LAYER CORRELATIONS

To destroy the geometric trans-layer correlations in our real multiplexes, we randomly reshuffle the trans-layer node-to-node mappings. Specifically, for each real multiplex we select one of its layers and we interchange the id of each node of the layer with the id of a randomly selected node from the same layer. The idea behind this process is that if a node with id i is node n_1 in layer 1 and node n_2 in layer 2 with correlated coordinates (r_{n_1}, θ_{n_1}), (r_{n_2}, θ_{n_2}), then, after reshuffling layer 2, the node will become some other node n_2' in this layer, with coordinates $(r_{n_2'}, \theta_{n_2'})$ that will not be correlated with (r_{n_1}, θ_{n_1}). We note that this reshuffling process is just a random id interchange among the nodes of a layer and does not alter the layer's topology. We used this process to create the reshuffled counterparts of the real multiplexes in the main text and in the next section. The reshuffled counterparts serve as a null model for what one would expect if there were no geometric correlations among the layers.

IV. SIMILARITY COMMUNITIES AND TRANS-LAYER CONNECTION PROBABILITIES

Figs. S2–S4 show the distribution of nodes that have angular coordinates θ_i, θ_j in layers i, j of the rest of our real multiplexes as well as the corresponding distributions in their reshuffled counterparts. In cases where there are more than 3 layers in a multiplex, we present results for selected layer-pair combinations. Figs. S5–S7 show the empirical trans-layer connection probability, $P(i|j)$, that two nodes are connected in layer i of each multiplex, given their hyperbolic or angular distance in layer j. We can make similar observations as with the Internet, Drosophila, and arXiv multiplexes in the main text (cf. Figs. 2, 3). We also observe that trans-layer link prediction in some multiplexes is better compared to others. This is expected since the strength of geometric correlations is different in different multiplexes, see Table II in Section X where we quantify the radial and angular correlation strengths. For example, we can see from Table II that the Air/Train has the weakest angular correlations, which explains the relatively poorer trans-layer link prediction using only similarity distances (Fig. S5). We quantify the quality of trans-layer link prediction in all the consider real multiplexes in Sec. VI. Finally, we note that in all of our connection probability plots we use in the x-axis linear scale for hyperbolic distances and logarithmic scale for angular distances. This is because the hyperbolic distance is a logarithmic function of the angular distance ($x_{ij} \approx r_i + r_j + 2\ln(\Delta\theta_{ij}/2)$), making this the most natural representation option.

The two-dimensional similarity histograms (Figs. 2, 3 in the main text and Figs. S2–S4) provide a natural way to observe soft community structure simultaneously present in pairs of network layers, where by soft communities we mean groups of nodes located close to each other in the similarity space [17–20]. We note that we do not propose an automated community detection algorithm here—developing such an algorithm is beyond the scope of the present manuscript. However, our findings could serve as a starting point for the future development of such an algorithm, where one first embeds the n layers of a given multiplex, and then utilizes well-established algorithms from the area of spatial clustering (e.g., spectral clustering [21], DBSCAN [22], etc.) that would automatically detect multidimensional communities on a geometric basis (soft communities), and which are known to work well even in high dimensions.

As mentioned, the development of community detection techniques for multiplex systems is recently gaining attention [10, 23–27]. Different techniques introduce complementary notions of communities and complementary techniques for identifying them, often extending results from single-layer networks to multiplexes. For instance, Mucha et al. [24] derive a multilayer generalization of modularity that is based on a simultaneous random walk on all layers of a multiplex to detect communities. Stanley et al. [26] derive a strata multilayer

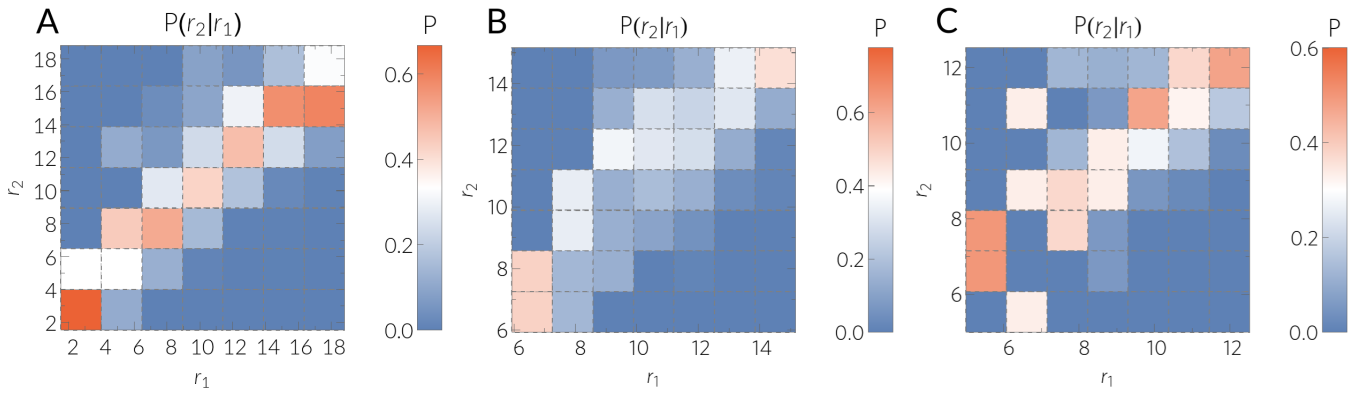

Figure S1. **Conditional probability $P(r_2|r_1)$ that a node has radial coordinate r_2 given its radial coordinate r_1. From left to right the plots correspond to the IPv4/IPv6 Internet (A), arXiv Layers 1, 2 (B), and Drosophila (C).**

stochastic block model approach that identifies and combines sets of layers with similarities in community structure (strata), and infers consensus node-to-community assignments. Furthermore, Taylor et al. [27] aggregate the layers of a multiplex using different operations and show using random matrix theory for stochastic block models, that by aggregating similar layers, it is possible to identify community structure that is undetectable in a single layer.

In contrast to previous approaches, where the boundaries of communities are determined by the network itself, soft communities are defined by the similarity coordinates of nodes, which manifest themselves indirectly via the nodes' connections and disconnections. Nodes closer in the similarity space form soft communities whose boundaries in our similarity histograms are determined by the bin size. The bin size can be seen as a community resolution parameter akin to [24]—the smaller the bin size the higher is the resolution and the smaller is the community size in the bin. However, we emphasize that our similarity histograms are just a natural and convenient way of observing soft community structure in layer pairs. They do not comprise an automated community detection method. We also note that soft communities have been shown to correspond well to actual and meaningful network communities, cf. [17–20] and Sec. V.

In this work, we independently embed the layers of a given multiplex and observe community structure in layer pairs. This approach allows one to observe communities of nodes that are simultaneously present in both layers, as well as communities of nodes that are present in one layer but not in the other (cf. Sec. V). In contrast, previous approaches, e.g., [26, 27], utilize different layers of a multiplex simultaneously in order to detect community structure across layers, that is, in order to infer consensus node-to-community assignments. In line with these studies, our embedding procedure can be extended in different directions. One direction would be to jointly embed the different layers of a given multiplex (or the layers belonging to the same stratum [26]), and infer single consensus radial and angular coordinates for each node. A second direction would be to aggregate the different layers using different operations as in [27] and then embed the aggregated network, again inferring single consensus coordinates for nodes. Finally, a third direction would be to infer the node coordinates in each layer as we do in this work, but by jointly embedding the layers (instead of independently). Whether these alternatives can lead to enhanced community detectability across layers as in [27], or improved performance in other applications such as greedy routing, are open research directions.

We note that the simultaneous embedding of layers that yields single consensus coordinates for each node will give a consensus of community membership for each node as in [26, 27]. This approach cannot identify how communities are spread between layers, e.g., it cannot identify if there are communities present in one layer but not in others. Therefore, we believe that all the embedding alternatives discussed above are complementary and depend on what aspect/details of community structure one is mostly interested in.

Finally, we note that the work in [26] groups layers of a multiplex together (in the same stratum) if their connection probability matrices are close to each other. A similar approach could be applied in our case using our connection probability matrices where connection probabilities are decreasing functions of hyperbolic distances among nodes. Also, a simpler approach might be to use the Normalized Mutual Information (NMI), see Sec. X, which quantifies the correlation strength among angular coordinates of layer pairs, and group layers together if their NMI is above a certain threshold.

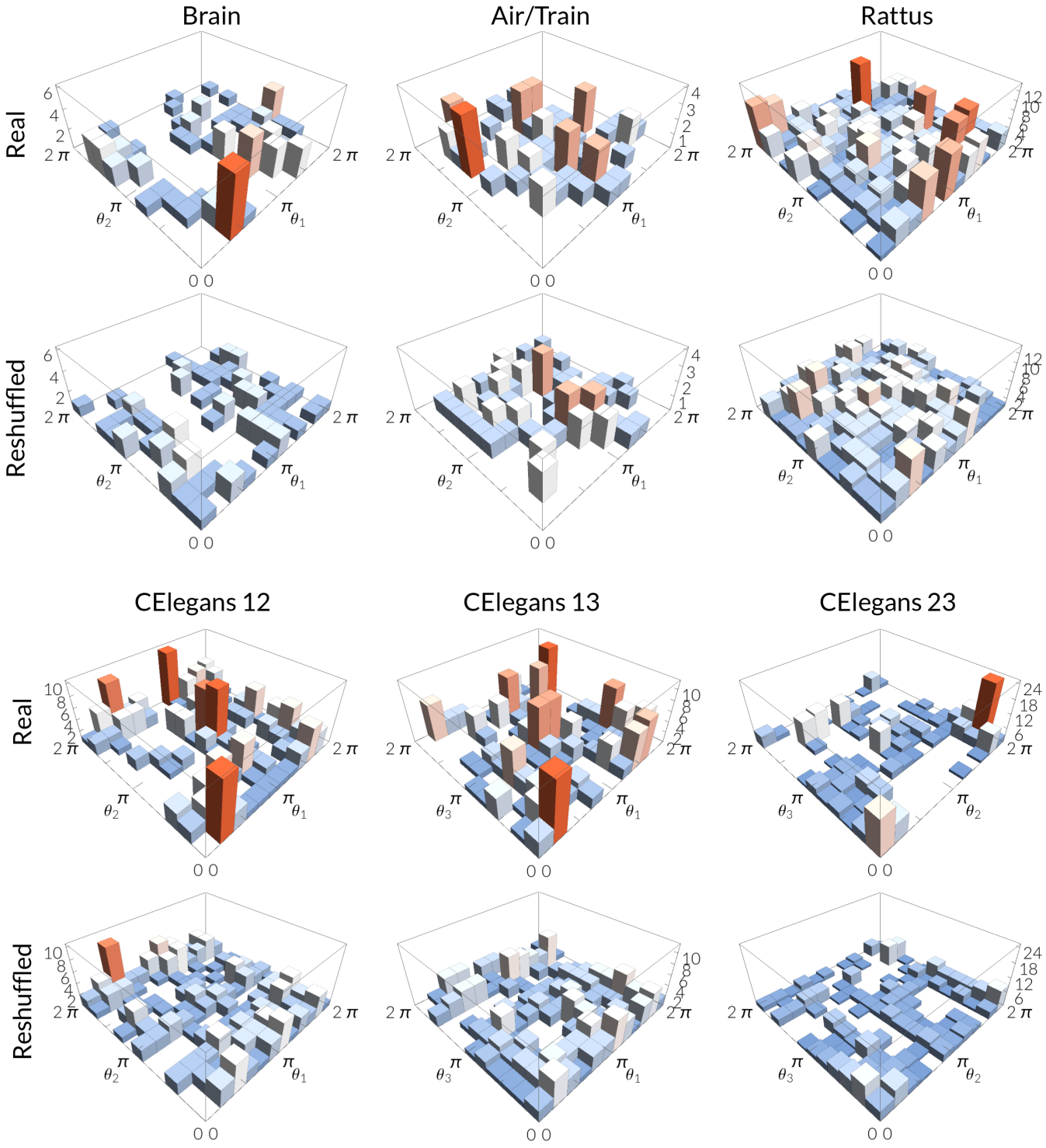

Figure S2. **Distribution of nodes in the two-dimensional similarity space of the human brain, Air/Train, and Rattus multiplexes (first row), and for all layer-pair combinations of the CElegans multiplex (third row).** In each case the plots correspond to nodes that exist in both layers. The angular similarity coordinate of a node in layers i, j is denoted by θ_i and θ_j, respectively. The histogram heights are equal to the number of nodes falling within each two-dimensional similarity bin, and the colors in each case denote the relative magnitude of the heights. **Second and fourth row:** The same distributions for the reshuffled counterparts of the real systems.

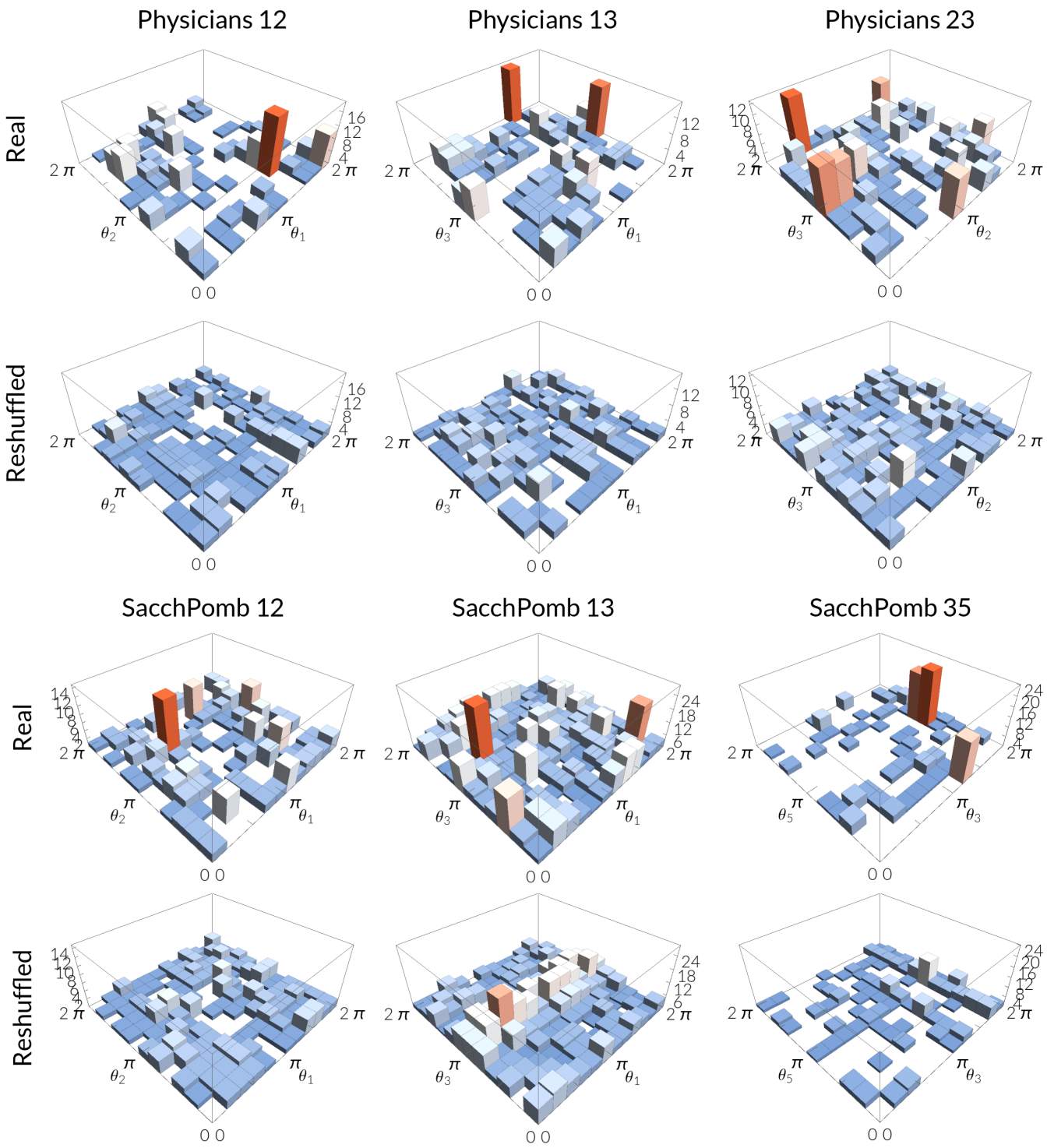

Figure S3. **Distribution of nodes in the two-dimensional similarity space of all layer-pair combinations of the Physicians multiplex (first row), and for different layer-pair combinations of the SacchPomb multiplex (third row).** In each case the plots correspond to nodes that exist in both layers. The angular similarity coordinate of a node in layers i, j is denoted by θ_i and θ_j, respectively. The histogram heights are equal to the number of nodes falling within each two-dimensional similarity bin, and the colors in each case denote the relative magnitude of the heights. **Second and fourth row:** The same distributions for the reshuffled counterparts of the real systems.

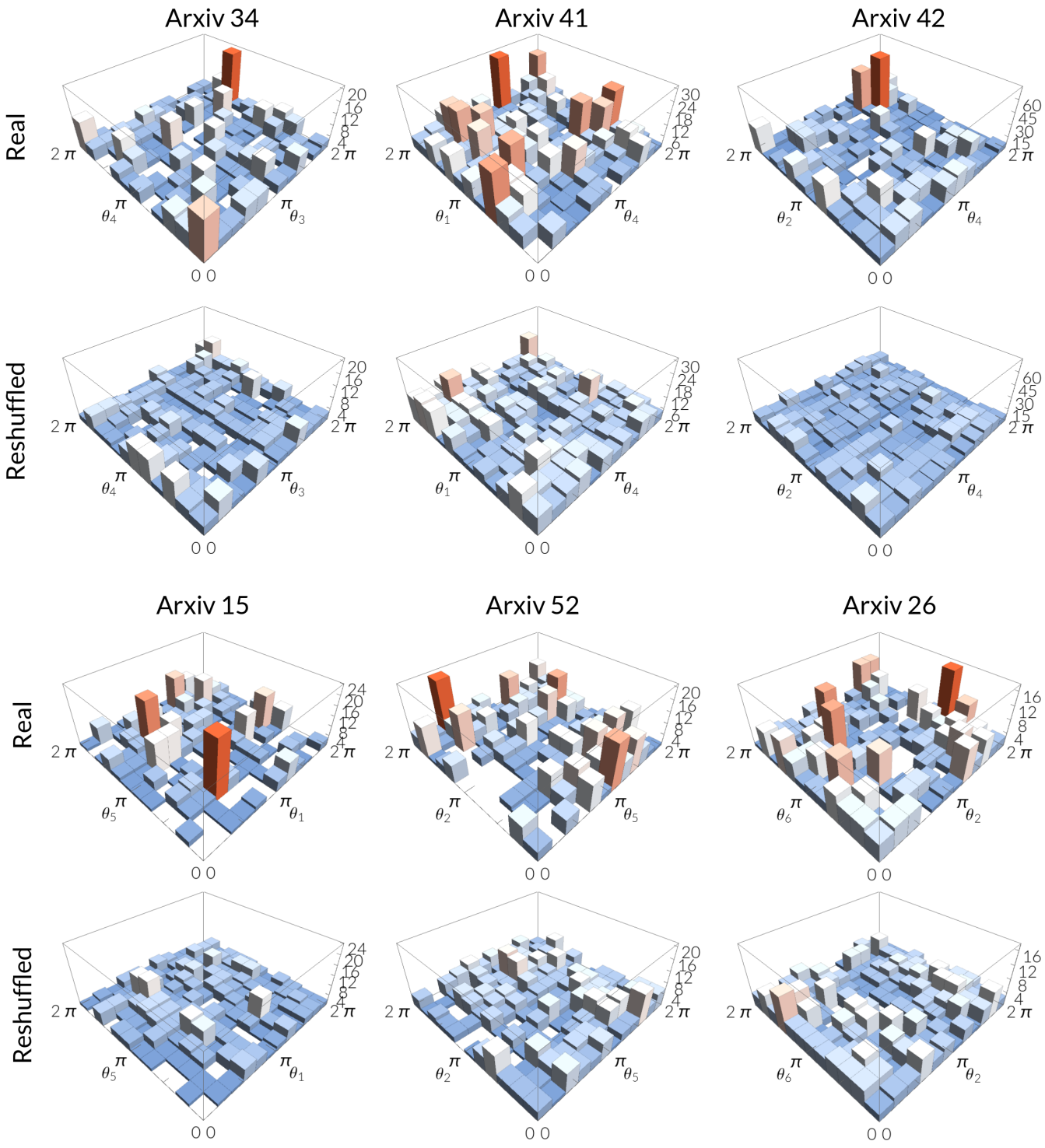

Figure S4. **Distribution of nodes in the two-dimensional similarity space of different layer-pair combinations of the arXiv multiplex (first and third row).** In each case the plots correspond to nodes that exist in both layers. The angular similarity coordinate of a node in layers i, j is denoted by θ_i and θ_j, respectively. The histogram heights are equal to the number of nodes falling within each two-dimensional similarity bin, and the colors in each case denote the relative magnitude of the heights. **Second and fourth row:** The same distributions for the reshuffled counterparts of the real systems.

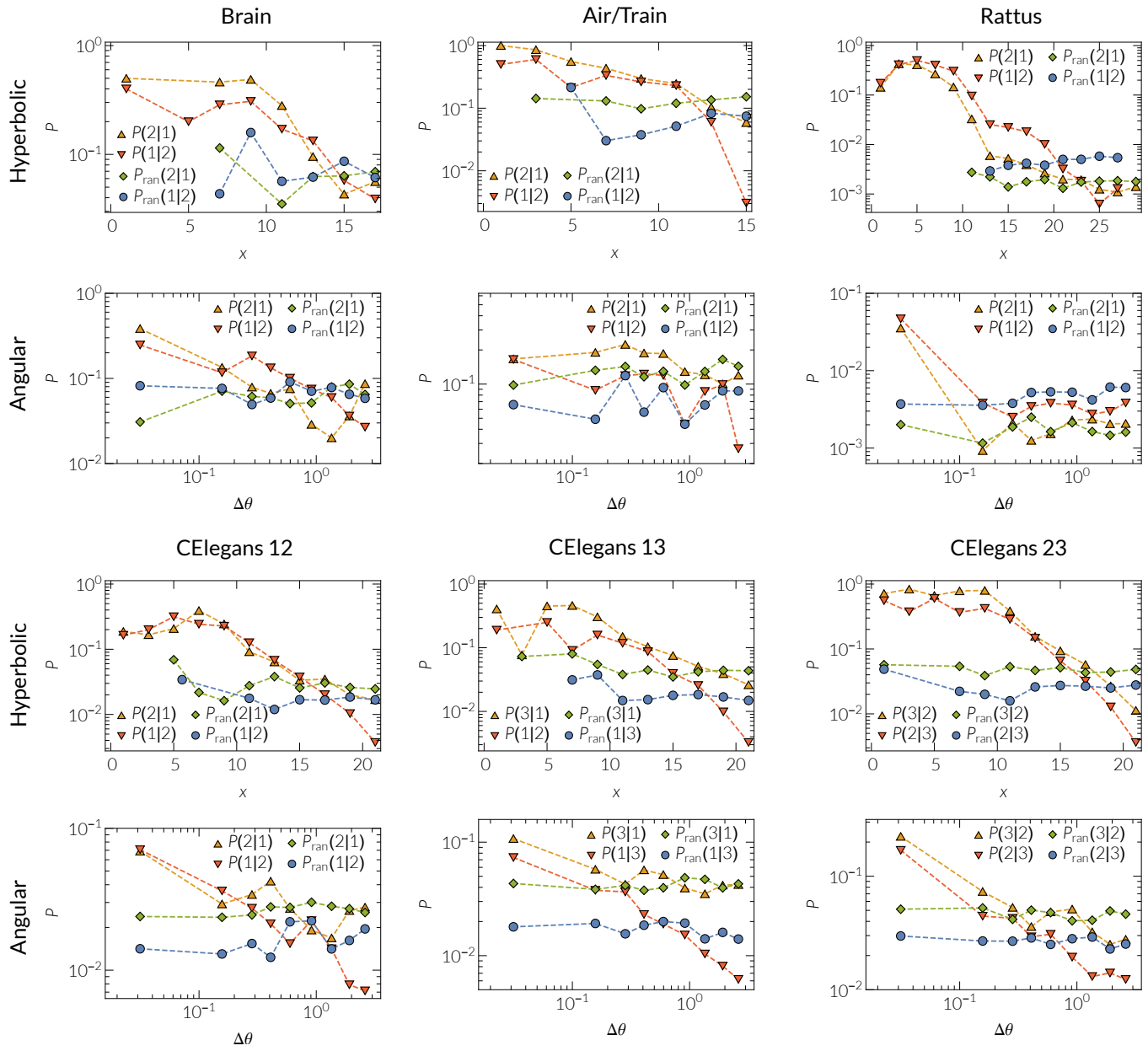

Figure S5. **Trans-layer connection probability in the human brain, Air/Train, and Rattus multiplexes and in different layer combinations of the CElegans. First and third row:** Trans-layer connection probability as a function of hyperbolic distance. $P(j|i)$ denotes the probability that a pair of nodes is connected in layer j given its hyperbolic distance x in layer i. $P_{\rm ran}(j|i)$ denotes the same probability for the reshuffled counterpart of each real system. **Second and fourth row:** Corresponding trans-layer connection probabilities when considering only the angular (similarity) distance between nodes, $\Delta\theta$.

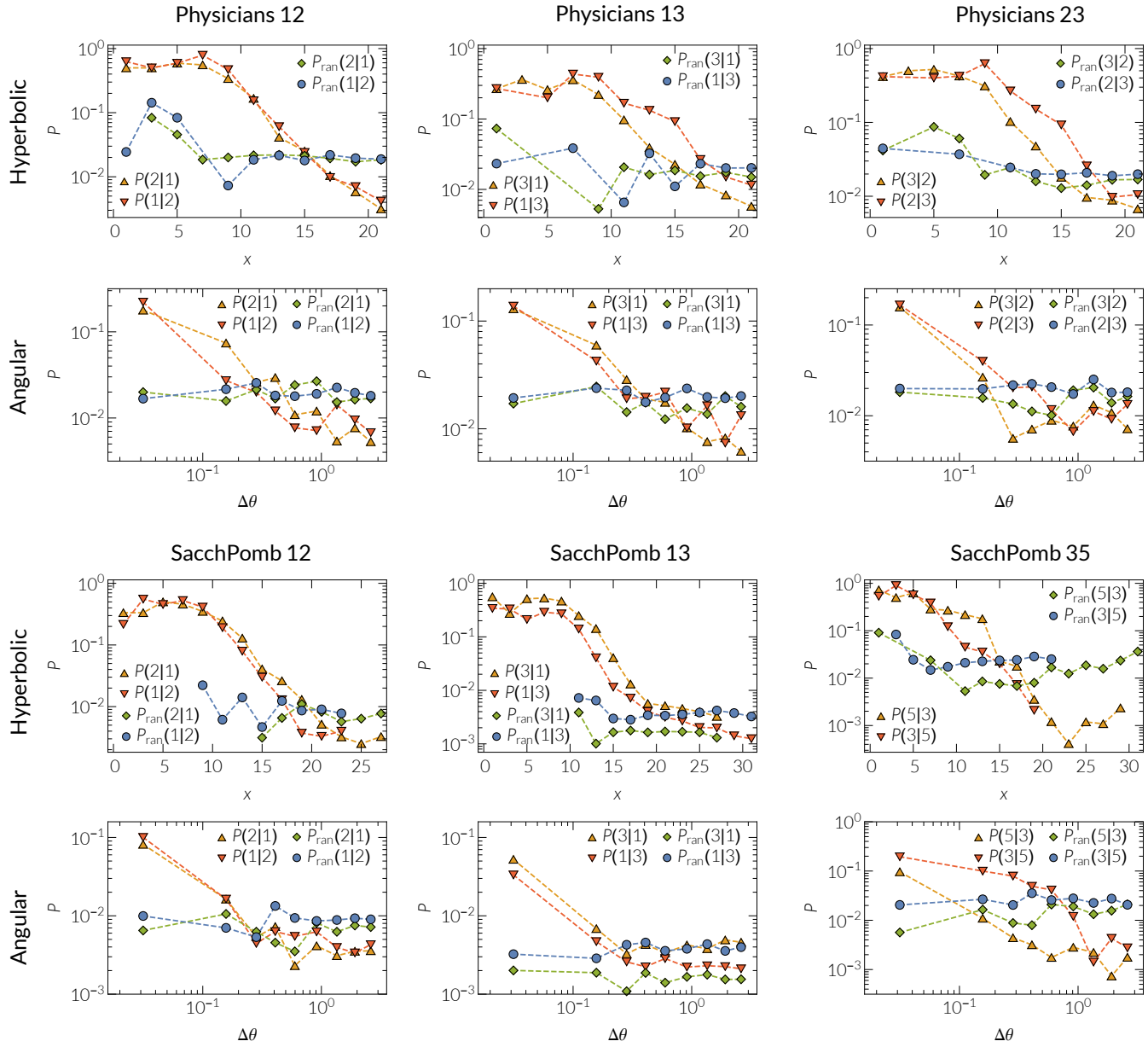

Figure S6. **Trans-layer connection probability in different layer combinations of the Physicians and SacchPomb multiplexes. First and third row:** Trans-layer connection probability as a function of hyperbolic distance. $P(j|i)$ denotes the probability that a pair of nodes is connected in layer j given its hyperbolic distance x in layer i. $P_{\text{ran}}(j|i)$ denotes the same probability for the reshuffled counterpart of each real system. **Second and fourth row:** Corresponding trans-layer connection probabilities when considering only the angular (similarity) distance between nodes, $\Delta\theta$.

V. SIMILARITY COMMUNITIES AND GEOGRAPHIC REGIONS IN THE IPV4/IPV6 INTERNET

In [12, 13, 17], we have considered the IPv4 Internet topology. We have shown that the mapping of the topology to its underlying hyperbolic space yields meaningful results, since ASs belonging to the same country are mapped close to each other. Specifically, for the majority of the countries, we have shown that their ASs are localized in narrow angular (similarity) regions. The reason for this effect is that ASs belonging to the same country are usually connected more densely to each other than to the rest of the world, and the mapping method (HyperMap) correctly places all such ASs in narrow regions close to each other. We note that other reasons besides geographic proximity may affect the connectivity between ASs, such as economical, political, and performance-related reasons. The mapping method does not favor any specific reason, but relies only on the connectivity

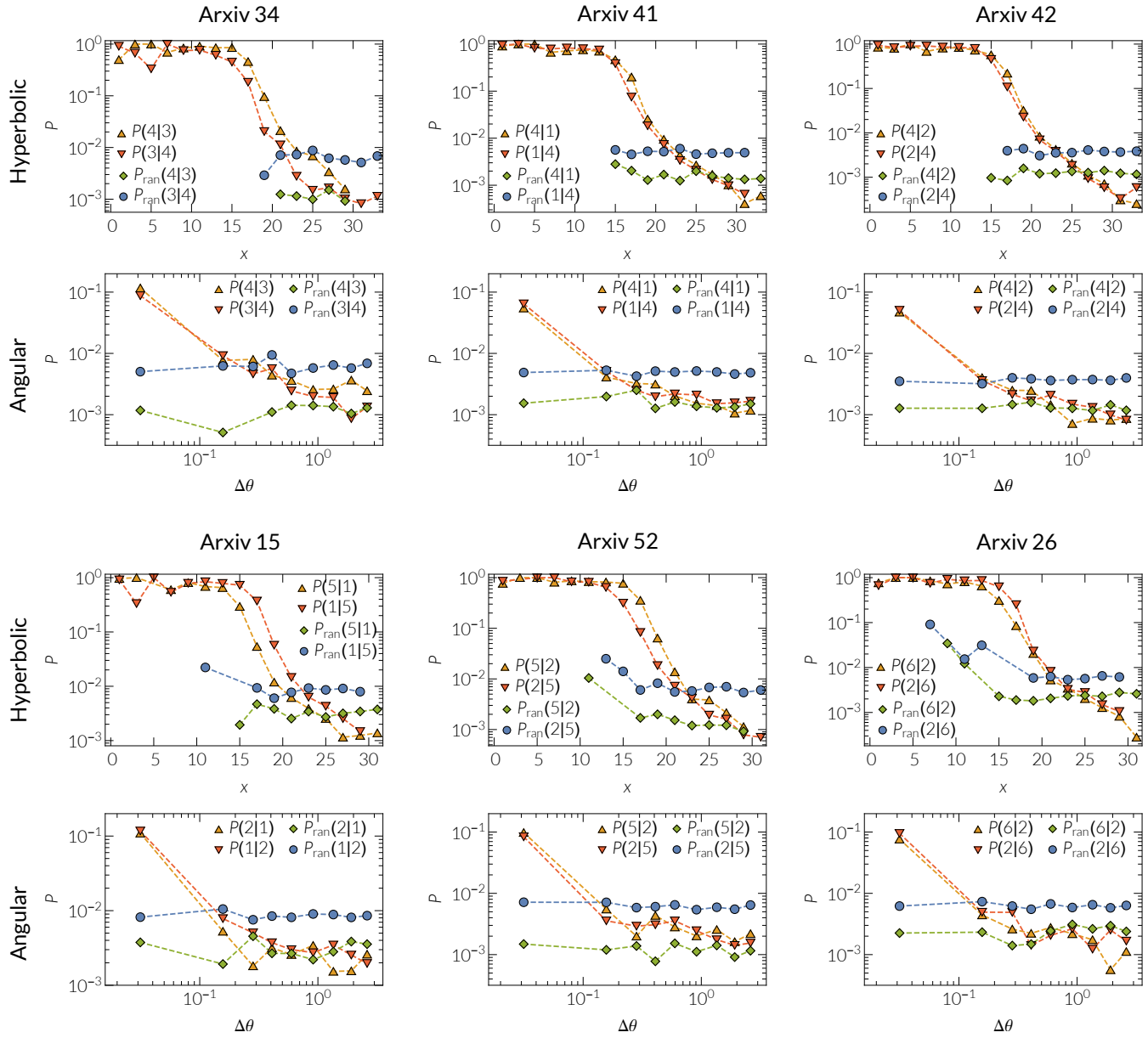

Figure S7. **Trans-layer connection probability in different layer combinations of the arXiv. First and third row:** Trans-layer connection probability as a function of hyperbolic distance. $P(j|i)$ denotes the probability that a pair of nodes is connected in layer j given its hyperbolic distance x in layer i. $P_{\rm ran}(j|i)$ denotes the same probability for the reshuffled counterpart of each real system. **Second and fourth row:** Corresponding trans-layer connection probabilities when considering only the angular (similarity) distance between nodes, $\Delta\theta$.

between nodes (ASs in this case) in order to place the nodes at the right angular (and hyperbolic) distances.

In Fig. S8, we observe a similar effect in the two-dimensional similarity space of the IPv4/IPv6 Internet. The figure shows the distribution of ASs belonging to different regions and countries. The AS-to-country mapping is taken from the CAIDA AS Organizations Dataset [28]. In Fig. S8, we can see ASs from regions/countries that are narrowly distributed in the two-dimensional similarity space, as well as ASs from regions/countries that are more widely spread. The former group of ASs are the ASs that form strong communities, i.e., that are densely connected to each other, in both the IPv4 and IPv6 topologies. In the figure, these are the ASs belonging to the Post-Soviet and South America regions (Fig. S8A), as well as the ASs belonging to some distinct counties such as Austria, Japan, China, Taiwan (Figs. S8B,C). By contrast, ASs belonging to the US and Europe are more widely spread in both the IPv4 and IPv6 similarity spaces (Figs. S8A,C). We note that Europe in Fig. S8A

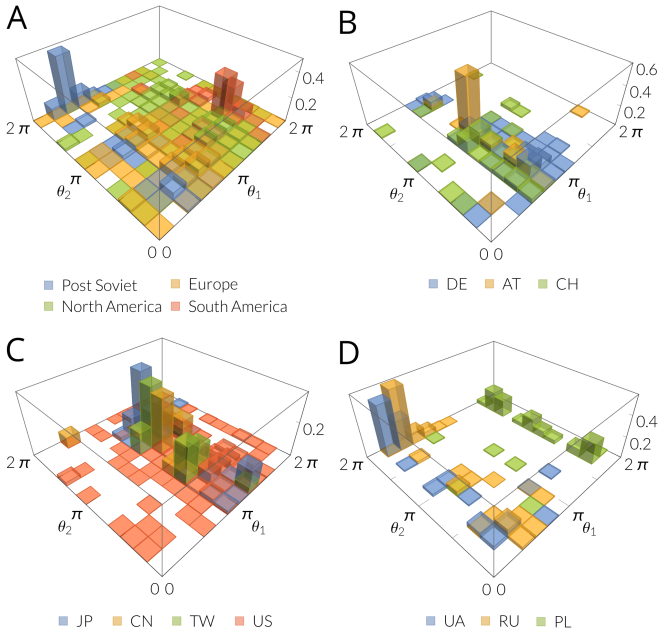

Figure S8. **Distribution of ASs of the same region/country in the two-dimensional similarity space of the IPv4/IPv6 Internet.** The plots correspond to ASs belonging to different regions/countries, which exist both in IPv4 and IPv6. The angular similarity coordinate of an AS in IPv4 is denoted by θ_1 and in IPv6 by θ_2. For each region/country, the histogram heights are normalized by the total number of ASs that belong to the region/country. In **A**, "Post Soviet" corresponds to the ASs belonging to Russia, Ukraine, Estonia, and Latvia; "Europe" corresponds to the ASs belonging to Germany, France, Spain, Finland, Austria, Netherlands, Sweden, Italy, and Greece; "North America" corresponds to the ASs belonging to the USA and Canada; and "South America" corresponds to the ASs belonging to Brazil, Uruguay, Argentina, and Colombia. The histograms in **B-D** correspond to the ASs belonging to 10 distinct countries. In **B**, the countries are Germany (DE), Austria (AT), and Switzerland (CH); in **C**, the countries are Japan (JP), China (CN), Taiwan (TW), and USA (US); and in **D**, the countries are Ukraine (UA), Russia (RU), and Poland (PL).

represents not one country but a collection of 9 different countries. Finally, we also observe that there can be ASs from countries that are narrowly distributed in the one similarity space, but not in the other. This is the case for example with the ASs belonging to Poland, which are narrowly distributed in the IPv4 space but not in the IPv6 (Fig. S8D). This suggests that these ASs do not form a strong community in IPv6, while they do in IPv4.

VI. QUALITY OF TRANS-LAYER LINK PREDICTION

Here, we quantify the quality of trans-layer link prediction across all layer-pairs of our real multiplexes. To this end, we use the *Area Under the Receiver Operating Characteristic Curve* (AUC) [29], which is a standard metric used to quantify the accuracy of a link prediction technique. Specifically, given two layers 1, 2, and the set of their common nodes, the AUC here is defined as the probability that a randomly selected link from the set of links among the common nodes in layer 2 is given a better score (i.e., a higher existence likelihood) than a randomly selected nonexistent link, where the "non-existent" links are the disconnected common node pairs in layer 2. The score between two common nodes i, j in layer 2 is the inferred hyperbolic distance x_{ij} between these nodes in layer 1. The smaller this score, i.e., the smaller the hyperbolic distance between two nodes in layer 1, the more likely it is that these nodes are connected in layer 2, since the connection probability (Eq. (3)) is a decreasing function of hyperbolic distance. The degree to which the AUC exceeds 0.5 indicates how much better the method performs than pure chance, while AUC = 1 is the best possible AUC.

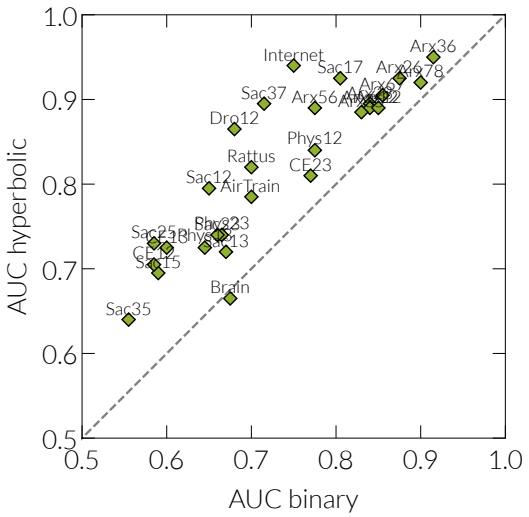

Figure S9. **Quality of trans-layer link prediction.** AUC hyperbolic vs. AUC binary.

We also compute the AUC for a simpler baseline binary link predictor, where if a link exists between two common nodes in layer 1, then we predict that this link will also exist in layer 2. Otherwise, if a link does not exist between two nodes in layer 1, then we predict that this pair of nodes will not be connected in layer 2 either.

The results are shown in Fig. S9, where the AUC value for each layer pair i, j is the average of the AUCs along the i to j and j to i prediction directions. The AUC hyperbolic is the AUC obtained when we use hyperbolic distances, while the AUC binary is the AUC obtained with the binary link predictor described above. From Fig. S9, we observe that in all cases apart from the brain (28 in total), trans-layer link prediction using hyperbolic distances performs better than the simple binary approach, and in several cases, e.g., Internet, Drosophila, impressively better.

VII. MODELING MULTIPLEX NETWORKS WITH GEOMETRIC CORRELATIONS—THE GEOMETRIC MULTIPLEX MODEL (GMM)

In this section we discuss the technical details of our framework, which constructs synthetic multiplex networks with geometric correlations. As mentioned, our framework constructs single-layer topologies using the $\mathbb{H}^2$ model [15], and allows for radial and angular coordinate correlations across the different layers. Instead of working directly with the $\mathbb{H}^2$ model, we make use of the $\mathbb{S}^1$ model [30] that is more convenient to work with, and which is isomorphic to the $\mathbb{H}^2$ model through a simple change of variables [15]. We first review the $\mathbb{S}^1$ model and its relation to the $\mathbb{H}^2$ model.

A. $\mathbb{S}^1/\mathbb{H}^2$ model of single-layer networks

Instead of radial and angular coordinates r_i, θ_i, each node i in the $\mathbb{S}^1$ model has hidden variables κ_i, θ_i. The hidden variable κ_i is the node's expected degree in the resulting network, while θ_i is the angular (similarity) coordinate of the node on a circle of radius $N/2\pi$, where N is the total number of nodes. To construct a network with the $\mathbb{S}^1$ model that has size N, average node degree $\bar{k}$, power law degree distribution with exponent $\gamma > 2$, and temperature $T \in [0, 1)$, we perform the following steps:

i. Sample the angular coordinates of nodes θ_i, $i = 1, 2, \ldots, N$, uniformly at random from $[0, 2\pi]$, and their hidden variables κ_i, $i = 1, 2, \ldots, N$, from the probability density function (PDF)

$$\rho(\kappa) = (\gamma - 1)\kappa^{\min\gamma-1}\kappa^{-\gamma}, \qquad (4)$$
$$\kappa^{\min} = \bar{k}\frac{\gamma-2}{\gamma-1},$$

where $\kappa^{\min}$ is the expected minimum node degree, which is a function of the average degree $\bar{k}$; [1]

ii. Connect every pair of nodes i, j with probability

$$r(\kappa_i, \theta_i; \kappa_j, \theta_j) = \frac{1}{1 + \left[\frac{d(\theta_i,\theta_j)}{\mu\kappa_i\kappa_j}\right]^{\frac{1}{T}}}, \qquad (5)$$
$$d(\theta_i, \theta_j) = \frac{N}{2\pi}\Delta\theta_{ij},\ \Delta\theta_{ij} = |\pi - |\pi - |\theta_i - \theta_j|||,$$
$$\mu = \frac{\sin T\pi}{2\bar{k}T\pi},$$

where $d(\theta_i, \theta_j)$ is the angular distance between nodes i, j on the circle.

[1] By sampling from a PDF $f(x)$ we mean that we first compute the CDF $F(x) = \int_{x_{\min}}^{x} \mathrm{d}x' \rho(x')$, where $x_{\min}$ is the minimum value of x, then generate a random number u_i uniformly at random from $[0, 1]$, and finally compute the value x_i such that $F(x_i) = u_i$. The value x_i is a sample from the PDF $\rho(x)$ (or the CDF $F(x)$).

The $\mathbb{S}^1$ model is equivalent to the $\mathbb{H}^2$ model after transforming the expected node degrees κ_i to radial coordinates r_i via

$$r_i = R - 2\ln\frac{\kappa_i}{\kappa^{\min}}, \qquad (6)$$

where R is the radius of the hyperbolic disc in the $\mathbb{H}^2$ model where all nodes reside,

$$R = 2\ln\frac{N}{c}, \qquad (7)$$
$$c = \bar{k}\frac{\sin T\pi}{2T}\left(\frac{\gamma-2}{\gamma-1}\right)^2.$$

It is easy to see that after the above change of variables the connection probability in Eq. (5) becomes the Fermi-Dirac connection probability in the $\mathbb{H}^2$ model,

$$p(x_{ij}) = \frac{1}{1 + e^{\frac{1}{2T}(x_{ij}-R)}}, \qquad (8)$$

where $x_{ij} \approx r_i + r_j + 2\ln\frac{\Delta\theta_{ij}}{2}$ is the hyperbolic distance between nodes i, j [15]. We note that without loss of generality, we use here a hyperbolic plane of curvature $K = -1$. See [15] for further details.

B. Two-layer multiplex model

We now describe our framework for constructing a two-layer multiplex system with geometric correlations. Each single-layer (layer 1, layer 2) is constructed according to the $\mathbb{S}^1$ model, and we account for correlations among the hidden variables of nodes in the two layers, whose strength can be tuned. The extension of the framework to more than two layers is straightforward and described in Section VII C. In a nutshell, our framework consists of the following steps:

i. Assignment of hidden variables $\kappa_{1,i}, \theta_{1,i}$ to each node i in layer 1 like in the $\mathbb{S}^1$ model (Eqs. (9), (11));

ii. Assignment of hidden variables $\kappa_{2,i}, \theta_{2,i}$ to each node i in layer 2, depending on the node's hidden variables in layer 1 (Eqs. (14), (24))—the assignment here is done such that the marginal (unconditional) distribution of $\kappa_{2,i}, \theta_{2,i}$ is still the one prescribed by the $\mathbb{S}^1$ model (Eqs. (12), (11));

iii. Creation of edges, by connecting node pairs in each layer with the corresponding $\mathbb{S}^1$ connection probability, which depends exclusively on the assigned hidden variables of nodes in each layer (Eqs. (26), (27));

iv. $\mathbb{S}^1$-to-$\mathbb{H}^2$ transformation, by mapping the hidden variables $\kappa_{1,i}, \kappa_{2,i}$ to radial coordinates $r_{1,i}, r_{2,i}$ (Eqs. (28), (29)).

Below, we describe these steps in detail. We assume that the two layers have the same number of nodes $N_1 = N_2 = N$. The extension of the framework to multiplexes with different layer sizes is given in Section VII D.

i. Assignment of hidden variables in layer 1. For each node $i = 1, 2, \ldots, N$ in layer 1 we sample its hidden variable $\kappa_{1,i}$ from the PDF

$$\rho_1(\kappa_1) = (\gamma_1 - 1)\kappa_1^{\min^{\gamma_1 - 1}} \kappa_1^{-\gamma_1}, \tag{9}$$

$$\kappa_1^{\min} = \bar{k}_1 \frac{\gamma_1 - 2}{\gamma_1 - 1}, \tag{10}$$

where $\bar{k}_1$ and $\gamma_1 > 2$ are respectively the target average degree and power law degree distribution exponent in layer 1. The angular coordinate $\theta_{1,i}$ of each node $i = 1, 2, \ldots, N$, is sampled from the uniform PDF

$$f(\theta) = \frac{1}{2\pi},\ \theta \in [0, 2\pi]. \tag{11}$$

ii. Assignment of hidden variables in layer 2. We now want to assign to each node $i = 1, 2, \ldots, N$ its hidden variable $\kappa_{2,i}$ in layer 2, conditioned on the value of its hidden variable $\kappa_{1,i}$ in layer 1. At the same time, we want the $\kappa_{2,i}$'s to satisfy the marginal (unconditional) PDF

$$\rho_2(\kappa_2) = (\gamma_2 - 1)\kappa_2^{\min^{\gamma_2 - 1}} \kappa_2^{-\gamma_2}, \tag{12}$$

$$\kappa_2^{\min} = \bar{k}_2 \frac{\gamma_2 - 2}{\gamma_2 - 1}, \tag{13}$$

where $\bar{k}_2$ and $\gamma_2 > 2$ are respectively the target average degree and power law degree distribution exponent in layer 2. Eq. (12) should be satisfied irrespectively of the correlation strength between the $\kappa_{2,i}$ and $\kappa_{1,i}$. To accomplish this, we sample the hidden variable $\kappa_{2,i}$ of each node $i = 1, 2, \ldots, N$, from the conditional cumulative distribution function (CDF)

$$F_\nu(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = e^{-(\varphi_1^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)})^{1-\nu}} \times \left[\varphi_1^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)}\right]^{-\nu} \frac{\varphi_1^{\nu/(1-\nu)} \kappa_1^{\min} \kappa_1^{\gamma_1}}{\kappa_1^{\min} \kappa_1^{\gamma_1} - \kappa_1^{\min^{\gamma_1}} \kappa_1},$$

$$\varphi_i = -\ln\left[1 - (\kappa_i^{\min}/\kappa_i)^{\gamma_i - 1}\right],\ \text{for } i = 1, 2, \tag{14}$$

where κ_1 is the value of the hidden variable of the node in layer 1, $\{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}$ are the network parameters defined earlier, and $\nu \in [0, 1]$ is the correlation strength parameter. The higher the value of ν the stronger is the correlation between $\kappa_{2,i}$ and $\kappa_{1,i}$. It is easy to see that when $\nu = 0$ (no correlation between $\kappa_{2,i}$ and $\kappa_{1,i}$), Eq. (14) becomes the marginal CDF of $\kappa_{2,i}$ given in Eq. (18) below. On the other hand, when $\nu \to 1$ (maximally correlated $\kappa_{2,i}$ and $\kappa_{1,i}$), Eq. (14) becomes

$$F_\nu(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \Theta\left[\kappa_2 - \kappa_2^{\min}\left(\frac{\kappa_1}{\kappa_1^{\min}}\right)^{(1-\gamma_1)/(1-\gamma_2)}\right], \tag{15}$$

where $\Theta[x]$ denotes the Heaviside step function. That is, when $\nu \to 1$

$$\kappa_{2,i} = \kappa_2^{\min}\left(\frac{\kappa_{1,i}}{\kappa_1^{\min}}\right)^{(1-\gamma_1)/(1-\gamma_2)}, \tag{16}$$

which yields $\kappa_{2,i} = \kappa_{1,i}$ if $\kappa_2^{\min} = \kappa_1^{\min}$ and $\gamma_1 = \gamma_2$.

To derive Eq. (14) we use copulas [31]. Copulas are multivariate probability distributions used to describe the dependence between random variables. In particular, any multivariate CDF $F(\kappa_1, ..., \kappa_n)$ of n random variables $\kappa_1, ..., \kappa_n$, can be written in the form $F(\kappa_1, ..., \kappa_n) = C(F_1(\kappa_1), ..., F_n(\kappa_n))$ where $F_1(\kappa_1), ..., F_n(\kappa_n)$ are the marginal CDFs of $F(\kappa_1, ..., \kappa_n)$, and C is called a copula. Each of the marginals of C is uniform in $[0, 1]$, and there are many parametric copula families available, which have parameters that control the strength of the dependence between the random variables [31].

In our case, the random variables are the node hidden variables κ_1, κ_2 in layers 1 and 2, whose marginal CDFs can be computed from Eqs. (9), (12),

$$F_1(\kappa_1) = 1 - \kappa_1^{(1-\gamma_1)} \kappa_1^{\min^{(\gamma_1 - 1)}}, \tag{17}$$

$$F_2(\kappa_2) = 1 - \kappa_2^{(1-\gamma_2)} \kappa_2^{\min^{(\gamma_2 - 1)}}. \tag{18}$$

For the copula function C, we use the bivariate Gumbel-Hougaard copula [31], defined as

$$C_\eta(u, v) = e^{-[(-\ln u)^\eta + (-\ln v)^\eta]^{1/\eta}}, \tag{19}$$

$$\eta \equiv \frac{1}{1 - \nu} \in [1, \infty).$$

Hence, our copula reads

$$C_\eta(F_1(\kappa_1), F_2(\kappa_2)) = e^{-[(-\ln F_1(\kappa_1))^\eta + (-\ln F_2(\kappa_2))^\eta]^{1/\eta}}. \tag{20}$$

The joint PDF of κ_1 and κ_2, $\rho_\eta(\kappa_1, \kappa_2, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$, can be obtained by differentiating the copula with respect to κ_1, κ_2,

$$\rho_\eta(\kappa_1, \kappa_2, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \frac{\partial^2 C_\eta(F_1(\kappa_1), F_2(\kappa_2))}{\partial \kappa_1 \partial \kappa_2}, \tag{21}$$

while the conditional PDF $\rho_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$ can be written as

$$\rho_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \rho_\eta(\kappa_1, \kappa_2, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) \frac{1}{\rho_1(\kappa_1)}. \tag{22}$$

The conditional CDF $F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$ can be therefore computed as

$$F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \int_{\kappa_2^{\min}}^{\kappa_2} \mathrm{d}\kappa'\, \rho_\eta(\kappa'|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \frac{\partial C_\eta(F_1(\kappa_1), F_2(\kappa_2))}{\partial \kappa_1} \frac{1}{\rho_1(\kappa_1)}, \tag{23}$$

which yields Eq. (14).

The angular coordinate $\theta_{2,i}$ of each node $i = 1, 2, \ldots, N$ in layer 2 is obtained by

$$\theta_{2,i} = \mod \left[\theta_{1,i} + \frac{2\pi l_i}{N}, 2\pi\right], \tag{24}$$

where $\theta_{1,i}$ is the angular coordinate of the node in layer 1, and l_i is a directed arc length on the $\mathbb{S}^1$ circle of radius $R = N/2\pi$, which is sampled from the zero-mean truncated Gaussian PDF

$$f_\sigma(l) = \frac{\frac{1}{\sigma}\phi\left(\frac{l}{\sigma}\right)}{\Phi\left(\frac{N}{2\sigma}\right) - \Phi\left(-\frac{N}{2\sigma}\right)}, \quad -\frac{N}{2} \leq l \leq \frac{N}{2}, \tag{25}$$
$$\sigma \equiv \sigma_0 \left(\frac{1}{g} - 1\right),$$

where $\sigma_0 = \min[100, N/(4\pi)]$ denotes the variance for $g = 0.5$ and $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $\Phi(x) = \int \mathrm{d}x\, \phi(x)$, $\sigma \in (0, \infty)$ is the variance of the PDF, and $g \in [0, 1]$ is the angular correlation strength parameter. The higher the value of g the stronger is the correlation between $\theta_{2,i}$ and $\theta_{1,i}$. When $g \to 0$, $\sigma \to \infty$, $f_\sigma(l)$ becomes the uniform PDF, and $\theta_{2,i}$, $\theta_{1,i}$ are not correlated. When $g = 1$, $\sigma = 0$, and $l_i = 0$, meaning that the angles of each node i are identical in the two layers, $\theta_{2,i} = \theta_{1,i}$.

iii. Creation of edges. Once all node hidden variables are assigned, we connect each node pair i, j in layers 1 and 2 with the corresponding $\mathbb{S}^1$ connection probabilities given in Eqs. (26), (27) below,

$$r_1(\kappa_{1,i}, \theta_{1,i}; \kappa_{1,j}, \theta_{1,j}) = \frac{1}{1 + \left[\frac{d_1(\theta_{1,i}, \theta_{1,j})}{\mu_1 \kappa_{1,i} \kappa_{1,j}}\right]^{\frac{1}{T_1}}}, \tag{26}$$
$$d_1(\theta_{1,i}, \theta_{1,j}) = \frac{N}{2\pi} \Delta\theta_{1,ij},$$
$$\Delta\theta_{1,ij} = |\pi - |\pi - |\theta_{1,i} - \theta_{1,j}|||,$$
$$\mu_1 = \frac{\sin T_1\pi}{2\bar{k}_1 T\pi},$$
$$r_2(\kappa_{2,i}, \theta_{2,i}; \kappa_{2,j}, \theta_{2,j}) = \frac{1}{1 + \left[\frac{d_2(\theta_{2,i}, \theta_{2,j})}{\mu_2 \kappa_{2,i} \kappa_{2,j}}\right]^{\frac{1}{T_2}}}, \tag{27}$$
$$d_2(\theta_{2,i}, \theta_{2,j}) = \frac{N}{2\pi} \Delta\theta_{2,ij},$$
$$\Delta\theta_{2,ij} = |\pi - |\pi - |\theta_{2,i} - \theta_{2,j}|||,$$
$$\mu_2 = \frac{\sin T_2\pi}{2\bar{k}_2 T_2\pi},$$

where $T_1 \in [0, 1), T_2 \in [0, 1)$ are the temperatures, which control clustering in each layer. We recall that the average node clustering is maximized at temperature $T = 0$, and nearly linearly decreases to zero with $T \in [0, 1)$.

iv. $\mathbb{S}^1$-to-$\mathbb{H}^2$ transformation. Finally, we map the node hidden variables $\kappa_{1,i}, \kappa_{2,i}$ in layers 1, 2, to radial coordinates $r_{1,i}, r_{2,i}$ using the relations below,

$$r_{1,i} = R_1 - 2\ln\frac{\kappa_{1,i}}{\kappa_1^{\min}},\ R_1 = 2\ln\frac{N}{c_1}, \tag{28}$$
$$c_1 = \bar{k}_1 \frac{\sin T_1\pi}{2T_1}\left(\frac{\gamma_1 - 2}{\gamma_1 - 1}\right)^2,$$
$$r_{2,i} = R_2 - 2\ln\frac{\kappa_{2,i}}{\kappa_2^{\min}},\ R_2 = 2\ln\frac{N}{c_2}, \tag{29}$$
$$c_2 = \bar{k}_2 \frac{\sin T_2\pi}{2T_2}\left(\frac{\gamma_2 - 2}{\gamma_2 - 1}\right)^2,$$

where $\kappa_1^{\min}, \kappa_2^{\min}$ are given in Eqs. (10), (13).

C. Modeling more than two layers

To construct a multilayer system consisting of n layers (layer 1, layer 2, ..., layer n), we work in the same way as with the two-layer system described above. Specifically, for each two consecutive layers $j-1, j$, for $2 \leq j \leq n$, we first fix their radial and angular correlation strength parameters $\nu_{j,j-1} \in [0, 1]$, $g_{j,j-1} \in [0, 1]$ to some desired values. Subsequently, we assign hidden variables $\kappa_{1,i}, \theta_{1,i}$ to nodes in layer 1 as described earlier (Eqs. (9), (11)), as well as hidden variables $\kappa_{2,i}, \theta_{2,i}$ to nodes in layer 2, conditioned on $\kappa_{1,i}, \theta_{1,i}$ (Eqs. (14), (24)). Then, we continue by assigning hidden variables $\kappa_{j,i}, \theta_{j,i}$ to nodes in layer $3 \leq j \leq n$, conditioned on the values of the hidden variables $\kappa_{j-1,i}, \theta_{j-1,i}$ of the nodes in layer $j-1$. This conditional assignment is done in exactly the same manner as the assignment of $\kappa_{2,i}, \theta_{2,i}$, which is conditioned on the values of $\kappa_{1,i}, \theta_{1,i}$. Once all node hidden variables in all layers are assigned, we create edges in each layer by connecting each node pair with the corresponding $\mathbb{S}^1$ connection probability (cf. Eqs. (26), (27)). Finally, for each layer $1 \leq j \leq n$, we map the node hidden variables $\kappa_{j,i}$ to radial coordinates $r_{j,i}$ as described earlier (cf. Eqs. (28), (29)). We have used this procedure to construct our three- and four-layer multiplexes in the main text and in Section XI, where the correlation strengths between subsequent layers are set to the same value, $\nu_{j,j-1} = \nu \in [0, 1]$, $g_{j,j-1} = g \in [0, 1]$, $\forall j \geq 2$.

D. Extension to multiplexes with different layer sizes

Here, we extend our framework to multiplexes with different layer sizes. Specifically, we consider a two-layer multiplex with layers 1, 2, which have number of nodes N_1, N_2. We assume that $N_1 > N_2$ and that there is a subset of N_{common} nodes in layer 1 that also exist in layer 2, $N_{\text{common}} \leq N_2$. To construct the two-layer multiplex we follow the steps below.

(i) Assignment of hidden variables in layer 1. For each node $i = 1, 2, \ldots, N_1$ in layer 1, we sample its hidden variable $\kappa_{1,i}$ as before, i.e., from the PDF $\rho_1(\kappa_1)$ in Eq. (9),

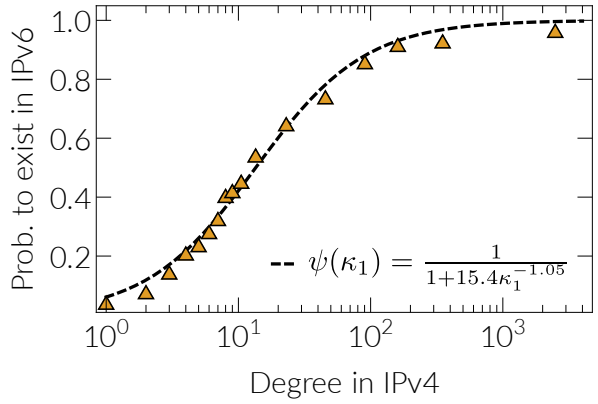

Figure S10. **Probability that a node (AS) exists in the IPv6 Internet given its degree in the IPv4 Internet.**

and its angular coordinate $\theta_{1,i}$ from the uniform PDF $f(\theta)$ in Eq. (11).

(ii) Determining the common nodes. We now need to decide the N_{common} nodes from layer 1 that will also be present in layer 2. The simplest approach is to randomly select (approximately) N_{common} nodes from layer 1, by sampling each node from layer 1 with the same probability ψ,

$$\psi = \frac{N_{\text{common}}}{N_1}, \tag{30}$$

and declaring each sampled node as a common node that will also exist in layer 2. However, this random sampling approach may not be realistic. Indeed, as we have mentioned in the main text, nodes with a higher degree in the IPv4 Internet (larger layer) have a higher probability to also exist in the IPv6 Internet (smaller layer). Fig. S10 shows the empirical probability for a node (AS) to exist in the IPv6 Internet given its degree in the IPv4 Internet. This probability can be approximated by

$$\psi(\kappa_1) = \frac{1}{1 + 15.4\kappa_1^{-1.05}}. \tag{31}$$

A dependence of the probability that a node exists in different layers on the degree of the node has also been observed in several other real multiplexes [32]. Therefore, a more realistic and general approach is to sample each node from layer 1 with a probability $\psi(\kappa_1)$ that is a function of its expected degree $\kappa_{1,i}$, such that

$$\int_{\kappa_1^{\min}}^{\infty} \mathrm{d}\kappa_1 \psi(\kappa_1)\rho_1(\kappa_1) = \frac{N_{\text{common}}}{N_1}, \tag{32}$$

where $\kappa_1^{\min}$ is given in Eq. (10). In the main text, we used the $\psi(\kappa_1)$ in Eq. (31) to sample nodes from layer 1 that also exist in layer 2 of the synthetic multiplex that best mimics the real IPv4/IPv6 Internet (Fig. 5 in the main text). The sampling yielded $N_{\text{common}} \approx 4800$, which is approximately equal to the number of common ASs (4819) in the real IPv4/IPv6 Internet.

iii. Assignment of hidden variables in layer 2. For the nodes i in layer 2 that do not exist in layer 1 (non-common nodes), we sample their $\kappa_{2,i}$'s from the unconditional PDF $\rho_2(\kappa_2)$ in Eq. (12), and their $\theta_{2,i}$'s from the uniform PDF $f(\theta)$ in Eq. (11). For the common nodes, we assign hidden variables $\kappa_{2,i}, \theta_{2,i}$, as described below.

We first compute the PDF $\tilde{\rho}_1(\kappa_1)$ of the hidden variables κ_1 of the common nodes,

$$\tilde{\rho}_1(\kappa_1) = \frac{\psi(\kappa_1)\rho_1(\kappa_1)}{\int_{\kappa_1^{\min}}^{\infty} \mathrm{d}\kappa_1 \psi(\kappa_1)\rho_1(\kappa_1)}, \tag{33}$$

and the CDF $\tilde{F}_1(\kappa_1)$,

$$\tilde{F}_1(\kappa_1) = \int_{\kappa_1^{\min}}^{\kappa_1} \mathrm{d}\kappa' \tilde{\rho}_1(\kappa'). \tag{34}$$

Then, we compute the conditional CDF $F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$ in exactly the same manner as in Section VII B, with the only difference that in place of $F_1(\kappa_1)$ in Eq. (17), we use the $\tilde{F}_1(\kappa_1)$ that we compute in Eq. (34), and instead of $\rho_1(\kappa_1)$ we use $\tilde{\rho}_1(\kappa_1)$. That is,

$$F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \frac{\partial C_\eta(\tilde{F}_1(\kappa_1), F_2(\kappa_2))}{\partial \kappa_1} \frac{1}{\tilde{\rho}_1(\kappa_1)}, \tag{35}$$

where $C_\eta(u, v)$ is the Gumbel-Hougaard copula (Eq. (19)) and $F_2(\kappa_2)$ is given in Eq. (18). The hidden variable $\kappa_{2,i}$ of each common node i is then sampled from the conditional CDF in Eq. (35). The angular coordinate of each common node $\theta_{2,i}$ is assigned using Eqs. (24), (25) with $N = N_2$.

iv. Creation of edges. The creation of edges in each layer is performed as before using the connection probabilities of the two layers in Eqs. (26), (27), with the only difference that now $d_1(\theta_{1,i}, \theta_{1,j}) = \frac{N_1}{2\pi}\Delta\theta_{ij}^1$ and $d_2(\theta_{2,i}, \theta_{2,j}) = \frac{N_2}{2\pi}\Delta\theta_{ij}^2$.

v. $\mathbb{S}^1$-to-$\mathbb{H}^2$ transformation. Finally, we again map the node hidden variables $\kappa_{1,i}, \kappa_{2,i}$ in layers 1, 2, to radial coordinates r_i^1, r_i^2 using Eqs. (28), (29), with the difference that now in these equations we have $R_1 = 2\ln\frac{N_1}{c_1}$ and $R_2 = 2\ln\frac{N_2}{c_2}$.

The above framework can be extended to more than two layers in the same manner as described in Section VII C. We note that when $\psi(\kappa_1)$ is the Internet's $\psi(\kappa_1)$ in Eq. (31), the conditional CDF in Eq. (35) can be approximated by

$$F_\nu(\kappa_2|\tilde{\kappa}, \{\gamma_2, \kappa_2^{\min}\}) = e^{-(\tilde{\varphi}^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)})^{1-\nu}} \times \left[\tilde{\varphi}^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)}\right]^{-\nu} \frac{\tilde{\varphi}^{\nu/(1-\nu)}\kappa_2^{\min}\tilde{\kappa}^{\gamma_2}}{\kappa_2^{\min}\tilde{\kappa}^{\gamma_2} - \kappa_2^{\min^{\gamma_2}}\tilde{\kappa}}, \tag{36}$$

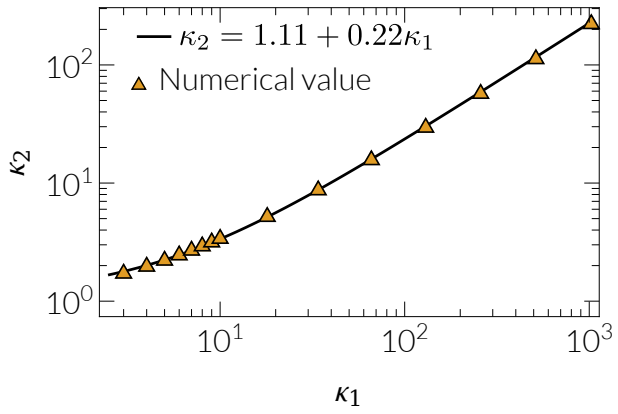

Figure S11. κ_2 **as a function of** κ_1 **at the limit** $\nu \to 1$**.** The triangles denote the values of κ_2 by numerically evaluating Eq. (35) using the $\psi(\kappa_1)$ in Eq. (31) and the IPv4/IPv6 Internet parameters $\gamma_1 = \gamma_2 = 2.1, \kappa_1^{\min} = 0.84, \kappa_2^{\min} = 1.30$. The best-fit line corresponds to $\kappa_2 = 1.11 + 0.22\kappa_1$.

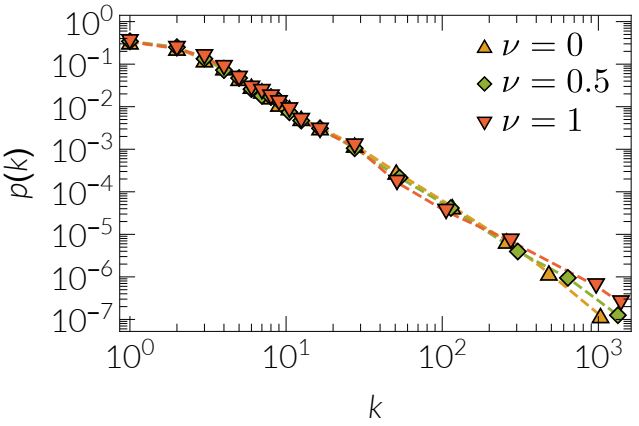

Figure S12. **Degree distribution (PDF) in layer 2 of our synthetic Internet-like multiplex (Fig. 5 in the main text) at correlation strengths** $\nu = 0, 0.5, 1$**.**

where

$$\tilde{\varphi} = -\ln\left[1 - \left(\kappa_2^{\min}/\tilde{\kappa}\right)^{\gamma_2 - 1}\right],$$
$$\varphi_2 = -\ln\left[1 - \left(\kappa_2^{\min}/\kappa_2\right)^{\gamma_2 - 1}\right],$$
$$\tilde{\kappa} = 1.11 + 0.22\kappa_1, \tag{37}$$

and $\kappa_2^{\min}$ is given in Eq. (13).

Eq. (37) is obtained by considering the maximally correlated case $\nu \to 1$, where Eq. (35) converges to a Heaviside step function $F_\nu(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \Theta[\kappa_2 - \tilde{\kappa}]$. That is, for $\nu \to 1$, $\kappa_2 \approx 1.11 + 0.22\kappa_1 \equiv \tilde{\kappa}$. This relation is the analogue of Eq. (16), and is obtained by numerically evaluating Eq. (35) at $\nu \to 1$, see Fig. S11. Once this result is known, then for $\nu \neq 1$, we can approximate Eq. (35) with Eq. (36), which results from Eq. (14) if in place of $\kappa_1, \gamma_1, \kappa_1^{\min}$, we use $\tilde{\kappa}, \gamma_2, \kappa_2^{\min}$. The idea behind this approximation is that instead of directly correlating the $\kappa_{2,i}$ with the $\kappa_{1,i}$ via Eq. (35), we correlate them via Eq. (36) with the corresponding values of $\kappa_{2,i}$ at the maximal correlations ($\nu \to 1$), which in our case are given by $\kappa_{2,i} = 1.11 + 0.22\kappa_{1,i}$.

In our synthetic Internet-like multiplex in the main text (Fig. 5), we sample the hidden variables $\kappa_{2,i}$ of the common nodes in layer 2 from the conditional CDF in Eq. (36). In Fig. S12, we show the marginal PDFs of the $\kappa_{2,i}$ of the common nodes at correlation strengths $\nu = 0$ (no correlations, where the $\kappa_{2,i}$ are sampled from their marginal CDF), $\nu = 0.5$ (partial correlations, where the $\kappa_{2,i}$ are sampled from the conditional CDF in Eq. (36)), and $\nu = 1$ (full correlations, where the $\kappa_{2,i}$ are directly obtained by Eq. (37)). In all cases, the marginal PDFs are nearly identical, validating the approximation described above.

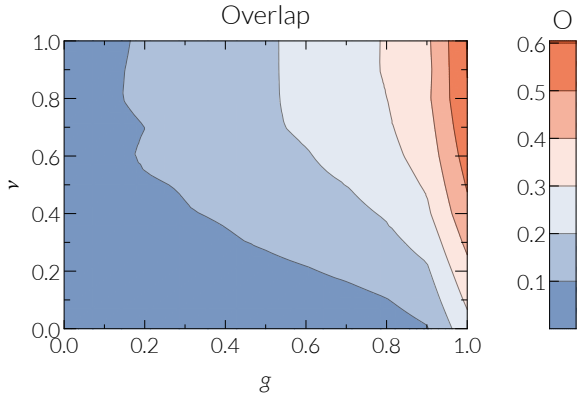

Figure S13. **Edge overlap** O **in a two-layer synthetic multiplex as a function of the radial (**ν**) and angular (**g**) correlation strengths.** Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}, \bar{k} = 10$, and temperature $T = 0.4$.

VIII. GEOMETRIC CORRELATIONS LEAD TO SIGNIFICANT EDGE OVERLAP

The radial and angular correlations across different layers naturally give rise to a significant amount of edge overlap between the layers, as observed in many real multiplexes [4, 33–36]. The edge overlap O between two layers (layer 1, layer 2) is formally defined as the ratio of the number of overlapping (i.e., common) edges between the layers, to the maximum possible number of common edges [37, 38],

$$O = \frac{\#(\text{overlapping edges})}{\min[\#(\text{edges in layer 1}), \#(\text{edges in layer 2})]}. \tag{38}$$

Fig. S13 shows the edge overlap in a synthetic two-layer multiplex as a function of the radial and angular correlation strength parameters ν, g. We observe that the

overlap increases as we increase the correlation strengths ν, g, and it is maximized at fully correlated coordinates, $\nu = 1, g = 1$. For uncorrelated coordinates, $\nu = 0, g = 0$, the overlap is minimized—it can be shown that when $\nu = 0, g = 0$, the overlap vanishes in the thermodynamic limit ($N \to \infty$), i.e., as the layer sizes increase.

The edge overlap also depends on the temperature of the layers. For fixed values of ν, g, a higher overlap is achieved when the temperature of the layers is lower. Specifically, if two layers have the same parameters $N, \gamma, \bar{k}, T$, and $\nu = g = 1$, i.e., the node coordinates in the two layers are identical, then at $T = 0$ the edge overlap is 100%, i.e., the topologies of the two layers are identical.

We note that although strong geometric correlations imply high edge overlap, the converse is not always true. That is, high edge overlap by itself, or in conjuction with degree (radial) correlations, does not necessarily imply that the hyperbolic geometries of the layers are strongly correlated. Only if two layers have high edge overlap, significant degree correlations *and* high clustering (i.e., strong metric structure [30]), then we indeed expect their hyperbolic geometries to be well correlated, although the exact strength of this correlation is still not obvious.

IX. ESTIMATION OF THE RADIAL AND ANGULAR CORRELATION STRENGTHS ν_E, g_E IN THE IPV4/IPV6 INTERNET

To estimate the empirical ν_E, we first compute the Pearson correlation coefficient [39] between the inferred radial coordinates of common nodes (ASs) in the IPv4/IPv6 Internet. We recall that the Pearson correlation is a measure of the linear correlation between two random variables X and Y, giving a value between $+1$ and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. Formally, the correlation coefficient $\rho_{X,Y}$ between X and Y is defined as

$$\rho_{X,Y} = \frac{E\left[(X - \mu_X)(Y - \mu_Y)\right]}{\sigma_X \sigma_Y}, \tag{39}$$

where μ_X, μ_Y are the expected values of X and Y and σ_X, σ_Y are their standard deviations. Then, we compute the same coefficient between the radial coordinates of common nodes in our synthetic Internet-like multiplex (Fig. 5B in the main text), at various radial correlation strengths $\nu \in [0, 1]$. The value of ν where the two coefficients are equal is the estimated value of ν_E. Fig. S14A shows the results, where we obtain $\nu_E \approx 0.4$.

To estimate g_E, we reconstruct IPv4-like and IPv6-like topologies as follows. We first consider all nodes in the real IPv4 topology with their inferred radial and angular coordinates, and connect each pair of nodes with the Fermi-Dirac connection probability $p(x_{ij})$ in Eq. (8), using the estimated temperature of the IPv4 topology, $T_1 = 0.5$, and such that the resulting network has the same average degree and power law degree distribution exponent as the real IPv4 topology, $\bar{k}_1 \approx 5$, $\gamma_1 = 2.1$. Subsequently, we consider all nodes in the real IPv6 topology. We assign to these nodes their inferred radial coordinates. For the nodes that also exist in IPv4, their angular coordinates are assigned using Eqs. (24), (25), and are correlated to their real angular coordinates in IPv4, using different correlation strengths $g \in [0, 1]$. The angular coordinates of the non-common nodes are set equal to their inferred angular coordinates. Then, we connect each pair of nodes in IPv6 with the Fermi-Dirac connection probability $p(x_{ij})$ in Eq. (8), using the estimated temperature of the IPv6 topology, $T_2 = 0.5$, and such that the resulting network has the same average degree and power law degree distribution exponent as the real IPv6 topology, $\bar{k}_2 \approx 5.2$, $\gamma_2 = 2.1$. For the different values of g, we evaluate the edge overlap O (Eq. (38)) between the reconstructed IPv4 and IPv6 topologies. The value of g that matches the edge overlap obtained when we reconstruct the IPv6 topology with all nodes having their inferred angular coordinates, is the estimated value of g_E. Fig. S14B shows the results, where we obtain $g_E \approx 0.4$.

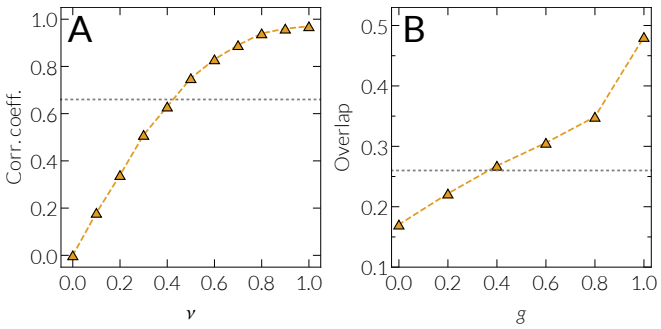

Figure S14. **Estimation of ν_E, g_E in the IPv4/IPv6 Internet. A.** Pearson correlation coefficient between the inferred radial coordinates of common ASs in the IPv4/IPv6 Internet (straight dashed line), and in our synthetic Internet-like multiplex at various radial correlation strengths ν. The two coefficients are equal at $\nu \approx 0.4 \equiv \nu_E$. **B.** Edge overlap O between reconstructed IPv4 and IPv6 topologies with inferred radial and angular coordinates (straight dashed line), and with synthetic angular coordinates for the common nodes in IPv6, at various correlation strengths g with their inferred IPv4 angles. The two overlaps are equal at $g \approx 0.4 \equiv g_E$.

X. QUANTIFYING GEOMETRIC CORRELATIONS

In this section we quantify the strength of geometric correlations in all the considered real multiplexes, using two different approaches. First, we consider all layer pairs and estimate the empirical radial and angular correlation strengths, ν_E, g_E, as we did in the previous section for the Internet. However, here we consider only the subgraphs from each layer consisting of the common nodes, as it

simplifies estimation. Second, we directly quantify the correlation strength among the inferred radial and angular coordinates of common nodes between layers. Specifically, we compute the Pearson correlation coefficient ρ between the common nodes' radial coordinates, as well as the mutual information I [40] between the common nodes' angular coordinates.

Mutual information is a more sophisticated measure than Pearson correlation, which can also measure nonlinear correlation relationships between variables. Formally, the mutual information between two random variables X and Y can be written as

$$I(X;Y) = \int_Y \int_X p(x,y) \ln \left(\frac{p(x,y)}{p(x)p(y)} \right) dxdy, \qquad (40)$$

where $p(x,y)$ is the joint probability density function of X and Y, and $p(x)$, $p(y)$ are the marginal probability density functions of X and Y. The higher the mutual information $I(X;Y)$ the stronger is the correlation between X and Y. To be able to compare angular correlations across different layer-pairs we compute the normalized mutual information NMI,

$$NMI = \frac{I(X;Y)}{\max\{I(X;X), I(Y;Y)\}}, \qquad (41)$$

which takes values in $[0,1]$. $NMI = 0$ ($I(X;Y) = 0$) means no correlation between X and Y, while $NMI = 1$ means perfect correlation. We note that $I(X;X), I(Y;Y)$ are the entropies of X and Y. To compute the mutual information between the angular coordinates we have used the k-nearest-neighbor mutual information estimator of Kraskov et al. [40], whose implementation is available at [41]. The estimator computes mutual information based on the distance between k nearest-neighbor data points, where k is an input parameter to the estimator. We have used $k = 3$, following [40] that proposes using $k =$ 2-4.

Table II shows the results for all layer pairs of all the considered real multiplexes. In all cases, we observe significant radial and angular correlations between the layers. We note that in the reshuffled counterparts of the real systems that do not exhibit geometric correlations, $\rho \approx 0, NMI \approx 0$. We observe that the agreement between ρ and ν_E is almost perfect (Fig. S15A), which is expected since ν_E is estimated using ρ (cf. Fig. S14A). We also observe a nice positive correlation between g_E and NMI (Fig. S15B), even though these measures are not directly dependent. Specifically, NMI is directly computed from the inferred angular coordinates, while the estimation of g_E is based on the edge overlap between the layers (cf. Fig. S14B), whose value depends not only on angular correlations but also on radial correlations and layer parameters. We note that for this reason, to reliably compare angular correlations across layer-pairs NMI should be used instead of g_E.

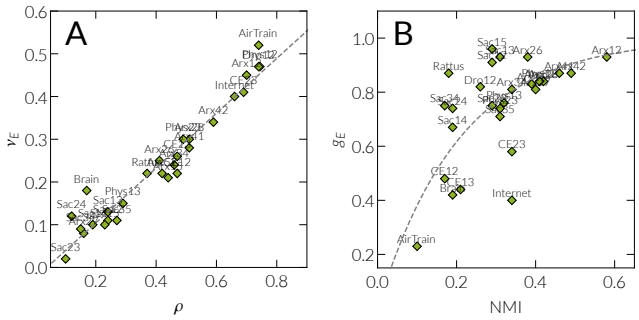

Figure S15. **Correlation measures for all layer pairs of the considered real multiplexes. A:** Pearson correlation ρ between radial coordinates vs. estimated radial correlation strength ν_E. **B:** Normalized mutual information NMI between angular coordinates vs. estimated angular correlation strength g_E. The dashed line in each plot corresponds to the best-fit line.

XI. PERFORMANCE OF MUTUAL NAVIGATION IN TWO-, THREE- AND FOUR-LAYER SYNTHETIC MULTIPLEXES.

Figs. S16...S18 show the success rate of angular MGR (Angular routing) and of MGR (Hyperbolic routing) in two-, three- and four-layer synthetic multiplexes as a function of the radial (ν) and angular (g) correlation strengths, and for different layer temperatures T.

In MGR with n layers (layer 1, ..., layer n), a node with a message first computes the distance between its neighbors and the destination of the message in layer 1, then it does the same for its neighbors and the destination in layer 2, and so on. The node then forwards the message to the neighbor that has the smallest distance to the destination across all computed distances. If a message is given back to a node it already visited, the delivery fails. The success rate is the percentage of messages that reach their destinations. In Figs. S16...S18 the success rate is evaluated across 10^5 randomly selected source-destination pairs. In MGR, hyperbolic distances (x_{ij}) between nodes are used, while in angular MGR only angular distances ($\Delta\theta_{ij}$) are used (Eq. (2)).

We note that in this work the reason we also study mutual navigation using only the node similarity coordinates is in order to investigate in more depth the extend to which similarity correlations across layers affect its performance, since it is the similarity of nodes that give to each layer an underlying geometry (that when combined with popularity becomes hyperbolic). Furthermore, there can be cases where nodes do not have access to the popularity of their neighbors and navigate the network using only the similarity between their neighbors and the target. This is, for instance, the case in Milgram's experiment [42]. Finally, similarity coordinates are also the ones responsible for the formation of community structure. Thus, given their importance, we study

Name	ν_E	g_E	ρ	NMI	I
arXiv Layers 1, 2	0.45	0.93	0.70	0.58	3.53
arXiv Layers 4, 2	0.34	0.87	0.59	0.49	2.93
arXiv Layers 4, 1	0.28	0.87	0.51	0.46	2.59
arXiv Layers 2, 8	0.30	0.84	0.51	0.42	2.33
Physicians Layers 1, 2	0.47	0.84	0.74	0.41	1.68
arXiv Layers 5, 2	0.22	0.84	0.42	0.41	2.06
arXiv Layers 2, 7	0.08	0.81	0.16	0.40	1.70
arXiv Layers 1, 5	0.21	0.83	0.44	0.39	1.88
arXiv Layers 2, 6	0.25	0.93	0.41	0.38	1.88
Internet Layer 1, 2	0.40	0.40	0.66	0.34	2.48
arXiv Layers 3, 4	0.24	0.81	0.46	0.34	1.62
C. Elegans Layers 2, 3	0.41	0.58	0.69	0.34	1.46
Physicians Layers 1, 3	0.15	0.76	0.29	0.32	1.27
Physicians Layers 2, 2	0.30	0.74	0.49	0.31	1.26
SacchPomb Layers 1, 3	0.13	0.93	0.24	0.31	1.64
SacchPomb Layers 3, 5	0.11	0.71	0.27	0.31	1.16
SacchPomb Layers 2, 3	0.02	0.75	0.10	0.29	1.30
SacchPomb Layers 1, 2	0.11	0.91	0.24	0.29	1.22
SacchPomb Layers 1, 5	0.10	0.96	0.23	0.29	1.97
Drosophila Layers 1, 2	0.47	0.82	0.74	0.26	1.31
C. Elegans Layers 1, 3	0.26	0.44	0.47	0.21	0.91
SacchPomb Layers 1, 4	0.10	0.67	0.19	0.19	0.94
SacchPomb Layers 2, 4	0.12	0.74	0.12	0.19	0.79
Brain Layers 1, 2	0.18	0.42	0.17	0.19	0.60
Rattus Layers 1, 2	0.22	0.87	0.37	0.18	0.90
C. Elegans Layers 1, 2	0.22	0.48	0.47	0.17	0.72
SacchPomb Layers 3, 4	0.09	0.75	0.15	0.17	0.72
Air/Train Layers 1, 2	0.52	0.23	0.74	0.10	0.30

Table II. **Estimated radial and angular correlation strengths ν_E, g_E, Pearson correlation ρ between radial coordinates, normalized mutual information NMI and mutual information I between angular coordinates, for all layer pairs of the considered real multiplexes.** The results are sorted in the decreasing order of NMI, i.e., from higher to lower angular (similarity) correlations.

them in isolation to provide both further intuitive insights and investigate their effect on applications. However, we note that once both the popularity and similarity coordinates of nodes are known, hyperbolic routing should be performed, which always performs better than angular routing.

XII. STRETCH AND POTENTIAL COSTS IN MUTUAL NAVIGATION

So far we have not considered potential costs that may exist by mutually navigating a multiplex system. Such costs can be of a very different nature in different real systems, e.g., ticket costs or traveling time by using the air network vs. using the train network, small processing times at Internet routers for switching from IPv4 to IPv6 and vice versa (see below), and so on. Also, there can be cases with no notion of cost at all, e.g., there is no obvious cost associated to friendship links in different online social networks through which information (e.g., news) can be mutually propagated. While MGR could be extended in order take into consideration some specific notions of cost depending on the system under study, e.g., layer switching cost when deciding to switch between layers, such extensions are beyond the scope of the present manuscript.

We note that in the Internet multiplex, the IPv4 Internet interoperates today with the IPv6 Internet via a process called *tunnelling*. This is a simple encapsulation/de-encapsulation procedure, described for instance in [43]. In this process, an IPv4 (IPv6) datagram that needs to enter an IPv6 (IPv4) network gets encapsulated as payload in an IPv6 (IPv4) datagram. When the IPv4

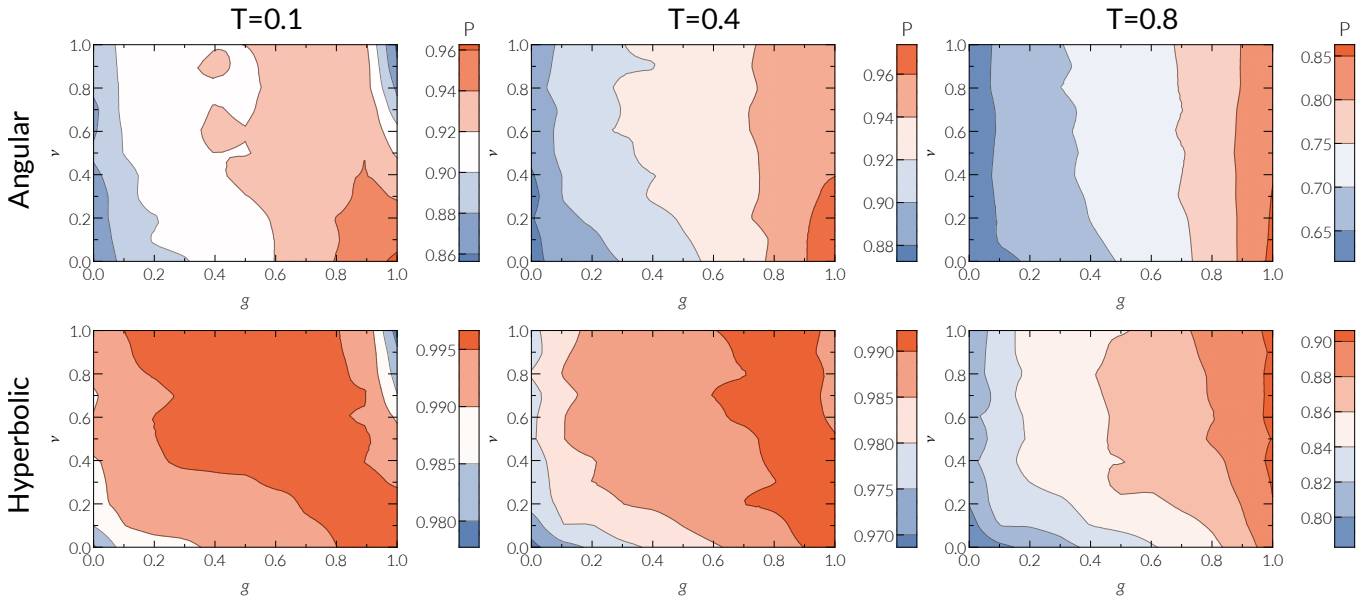

Figure S16. **Success rate of angular MGR (Angular routing, top row) and of MGR (Hyperbolic routing, bottom row) for a two-layer multiplex system as a function of the radial (ν) and angular (g) correlation strengths.** Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From left to right, $T = 0.1, 0.4, 0.8$.

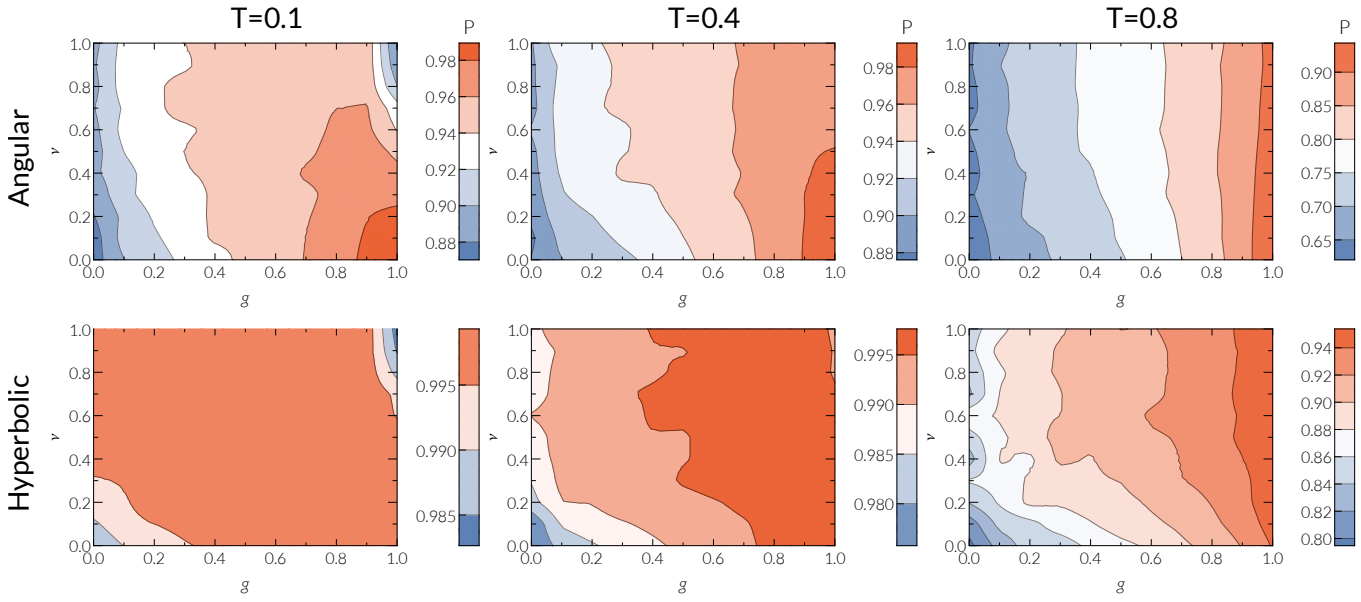

Figure S17. **Success rate of angular MGR (Angular routing, top row) and of MGR (Hyperbolic routing, bottom row) for a three-layer multiplex system as a function of the radial (ν) and angular (g) correlation strengths.** Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From left to right, $T = 0.1, 0.4, 0.8$.

(IPv6) datagram needs to exit the IPv6 (IPv4) network it gets de-encapsulated. The encapsulation/de-encapsulation overhead (cost) happens at border Internet Service Provider routers and is virtually insignificant (it is simply the addition of the size of an IPv4/IPv6 header of 20/40 bytes, cf. [44]).

Below, we perform a series of experiments that yield insights related to potential costs in mutual navigation. In general, the cost of MGR from a source to a destination is a combination of the number of layer switches and the number of hops that the process traverses in each layer, while the cost associated to each hop can in gen-

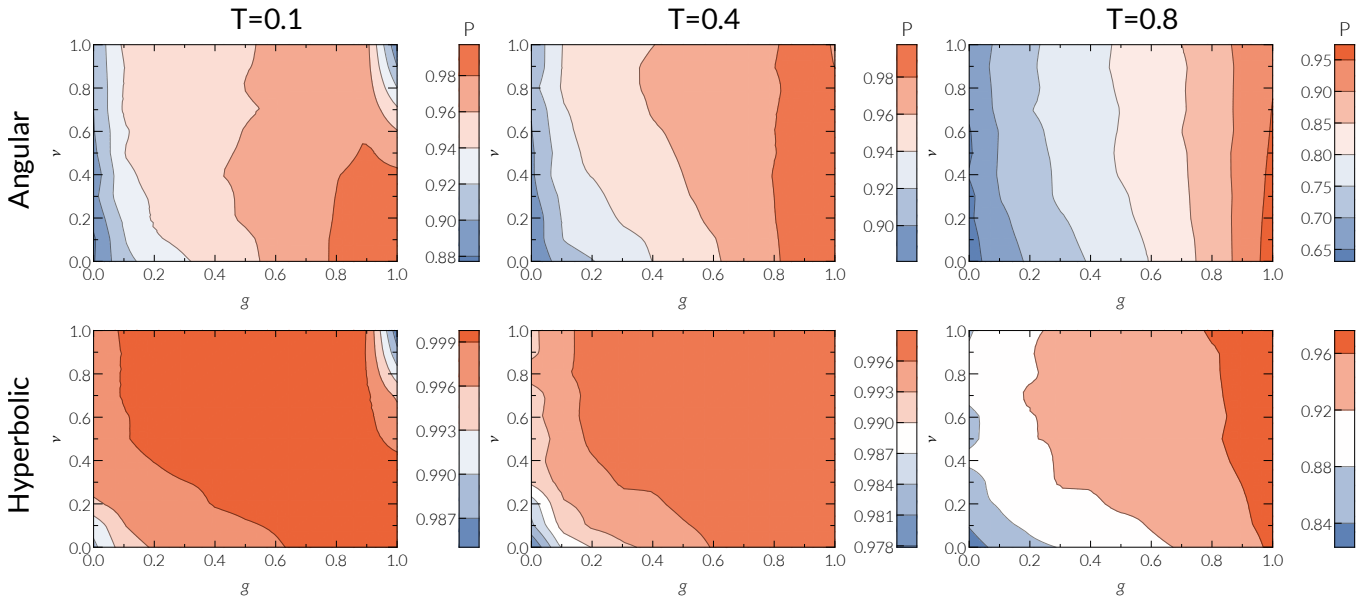

Figure S18. **Success rate of angular MGR (Angular routing, top row) and of MGR (Hyperbolic routing, bottom row) for a four-layer multiplex system as a function of the radial (ν) and angular (g) correlation strengths.** Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From left to right, $T = 0.1, 0.4, 0.8$.

eral be different in different layers. Therefore, here we consider two measures that contribute to the total cost of a delivery: (i) the number of layer switches along successful MGR paths; and (ii) MGR's stretch, which is the ratio between the hop-lengths of successful MGR paths and the corresponding topological shortest paths. The topological shortest paths are the shortest paths in the aggregated topology of the multiplex, obtained by aggregating the layers. A stretch equal to 1 means that MGR paths are shortest paths. We compute the mean, maximum, and standard deviation of these measures across 10^5 randomly selected source-destination pairs, and investigate how these measures behave as a function of the radial and angular correlation strengths ν and g.

Fig. S19 shows the results for our synthetic Internet-like multiplex (Fig. 5 in the main text), and Figs. S20–S22 show results for synthetic multiplexes with the same layer sizes and 2, 3, or 4 layers. From the figures, we observe the following. First, we observe that the mean number of layer switches is small in all cases, and in general it increases with increasing correlation strengths ν and g—the same behavior holds for the maximum and standard deviation of this measure. Second, we observe that the mean stretch is also small (close to 1) in all cases, and that it decreases with increasing correlation strengths ν and g—the same behavior holds for the maximum and standard deviation of this measure.

Our results suggest that while stronger geometric correlations tend to increase the number of layer switches, they also reduce stretch. While the former effect increases the cost of the MGR process, the latter decreases it. In other words, the cost of additional layer switching due to stronger geometric correlations is mitigated by reduced stretch, with the degree of mitigation depending on the relative weight/importance of each type of cost (layer switching vs. stretch). Our results also suggest that strong geometric correlations not only increase the success rate of MGR (as shown in the main text and in Sec. XI), but also make the successful MGR paths more efficient (shorter).

[1] Kimberly Claffy, Young Hyun, Ken Keys, Marina Fomenkov, and Dmitri Krioukov, "Internet Mapping: From Art to Science," *Proceedings of the 2009 Cybersecurity Applications & Technology Conference for Homeland Security*, CATCH '09, 205–211 (2009).

[2] "The IPv4 and IPv6 Topology Datasets," (2015), `http://www.caida.org/data/active/ipv4_routed_topology_aslinks_dataset.xml` and `https://www.caida.org/data/active/ipv6_allpref_topology_dataset.xml`.

[3] "IPv4 and IPv6 Topology Data," (January 2015), `http://data.caida.org/datasets/topology/ark/ipv6/as-links/2015/01/` and `http://data.caida.org/datasets/topology/ark/ipv4/as-links/`.

[4] Arda Halu, Satyam Mukherjee, and Ginestra Bianconi, "Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles," Phys. Rev. E **89**, 012806 (2014).

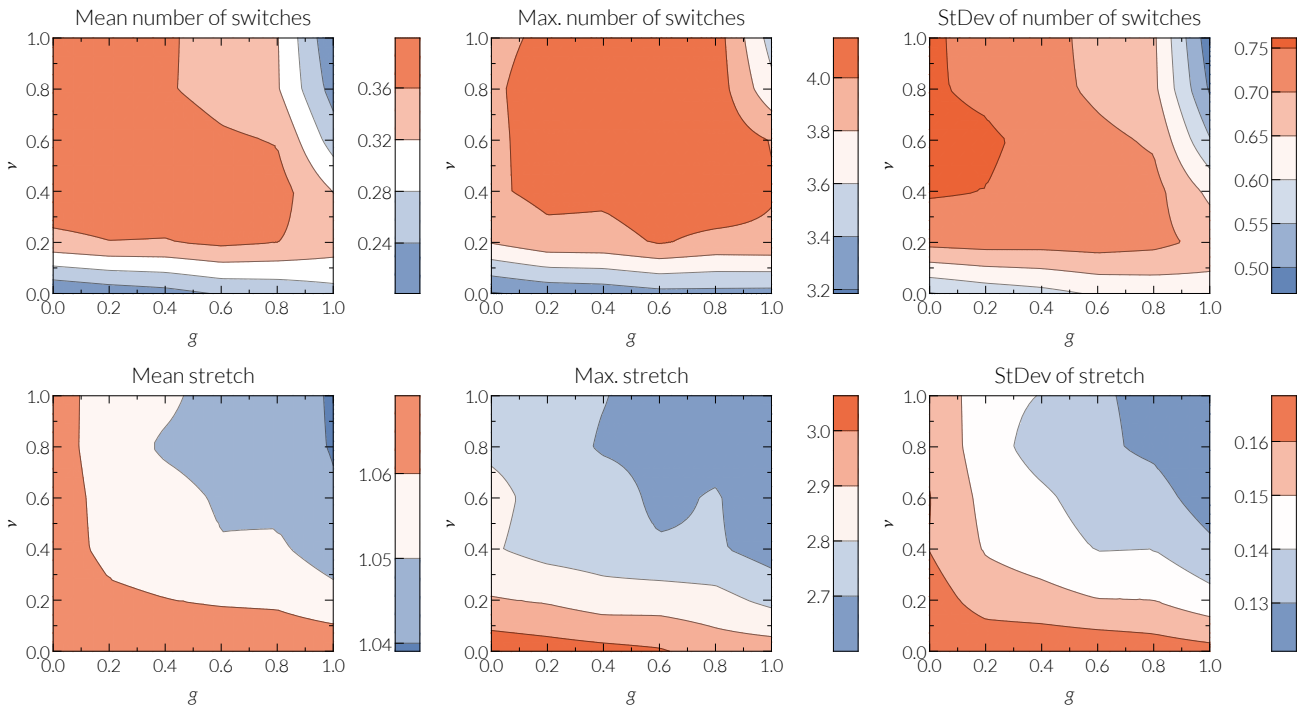

Figure S19. **Number of layer switches (first row) and stretch (second row) in MGR (Hyperbolic routing) in the synthetic Internet-like multiplex (Fig. 5 in the main text).** The first column shows mean values, the second column shows maximum values and the third column shows the standard deviation.

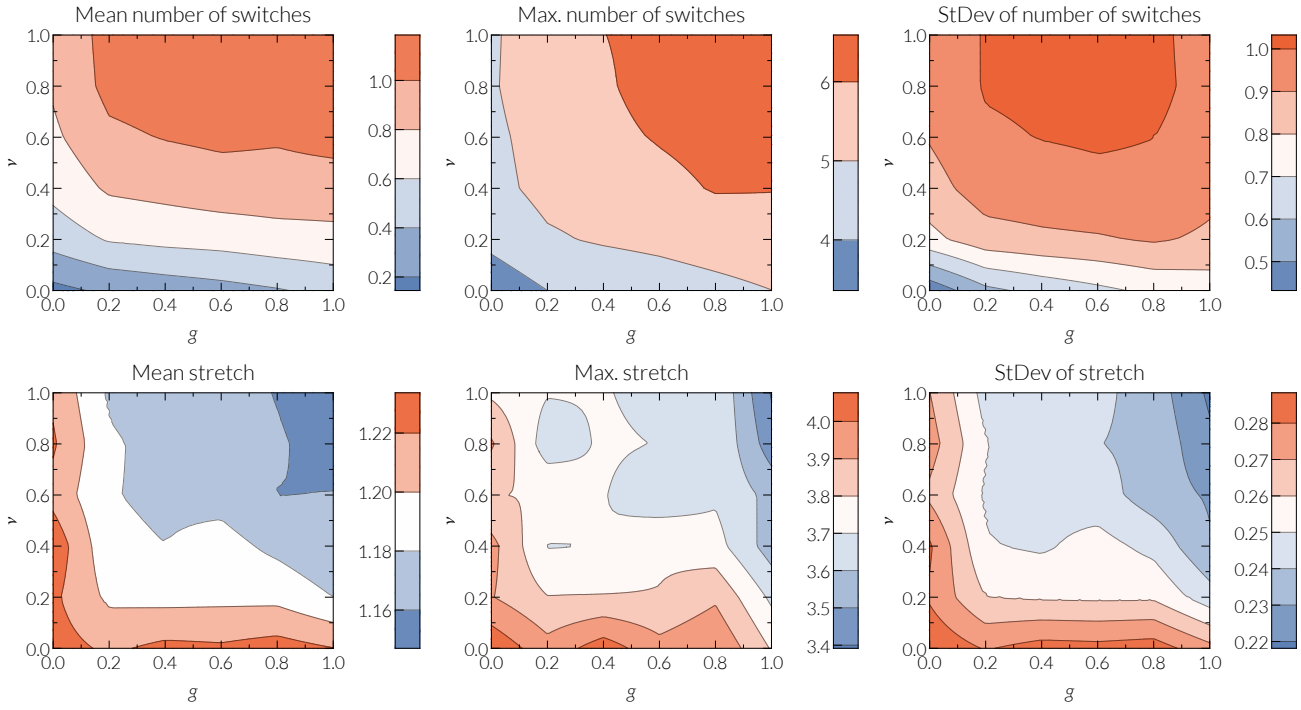

Figure S20. **Number of layer switches (first row) and stretch (second row) in MGR (Hyperbolic routing) in a two-layer synthetic multiplex.** Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter $T = 0.8$. The first column shows mean values, the second column shows maximum values and the third column shows the standard deviation.

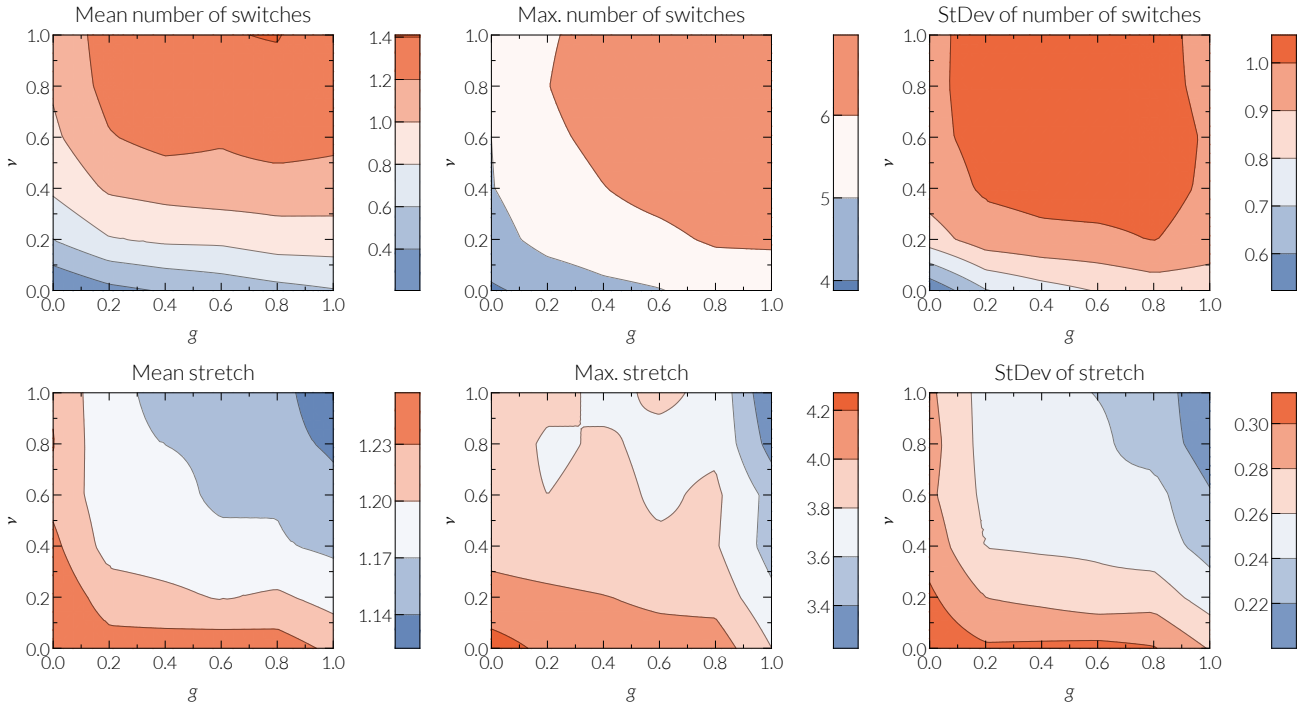

Figure S21. **Number of layer switches (first row) and stretch (second row) in MGR (Hyperbolic routing) in a three-layer synthetic multiplex.** Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter $T = 0.8$. The first column shows mean values, the second column shows maximum values and the third column shows the standard deviation.

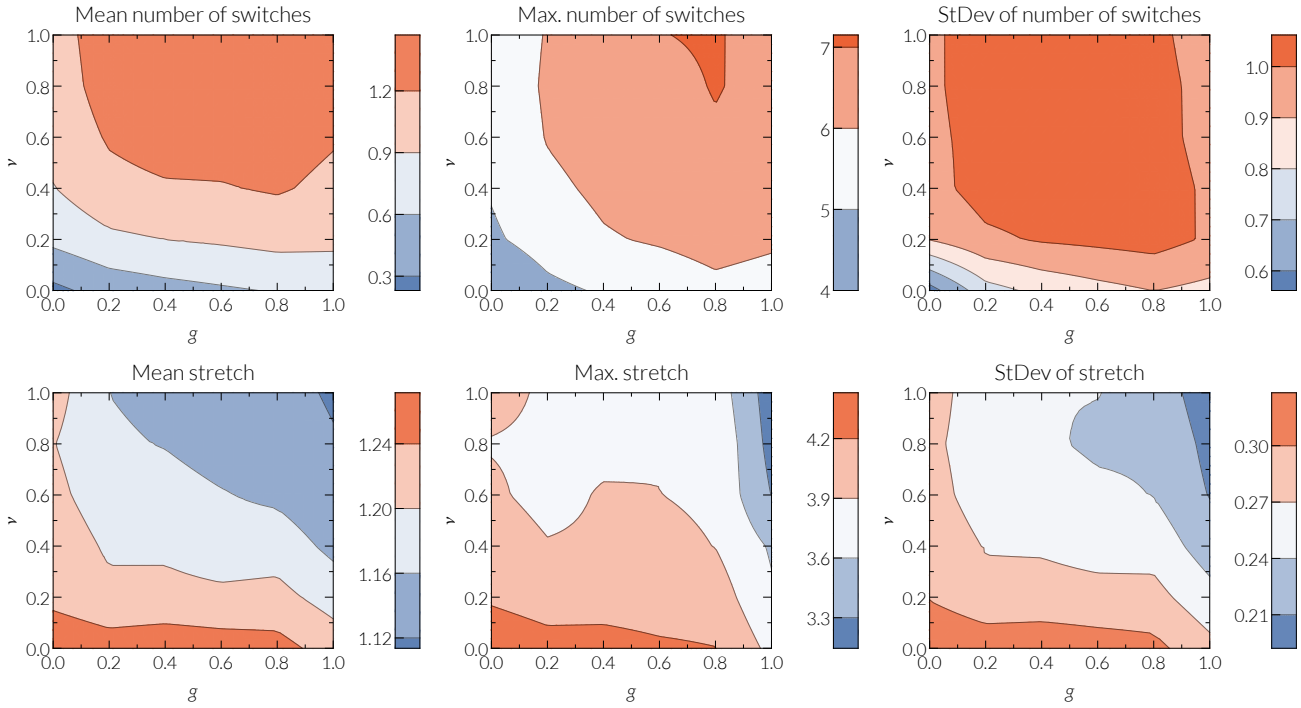

Figure S22. **Number of layer switches (first row) and stretch (second row) in MGR (Hyperbolic routing) in a four-layer synthetic multiplex.** Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter $T = 0.8$. The first column shows mean values, the second column shows maximum values and the third column shows the standard deviation.

[5] "Biogrid: a general repository for interaction datasets," Nucleic Acids Research **34**, D535–D539 (2006).

[6] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," Nature Communications **6**, 6864 (2015).

[7] Beth L. Chen, David H. Hall, and Dmitri B. Chklovskii, "Wiring optimization can relate neuronal structure and function," Proceedings of the National Academy of Sciences of the United States of America **103**, 4723–4728 (2006).

[8] Manlio De Domenico, Mason A. Porter, and Alex Arenas, "Muxviz: a tool for multilayer analysis and visualization of networks," Journal of Complex Networks **3**, 159–176 (2015).

[9] Tiago Simas, Mario Chavez, Pablo R. Rodriguez, and Albert Diaz-Guilera, "An algebraic topological method for multimodal brain networks comparison," Frontiers in Psychology **6**, 904 (2015).

[10] Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall, "Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems," Phys. Rev. X **5**, 011027 (2015).

[11] J. Coleman, E. Katz, and H. Menzel, "The diffusion of an Innovation among Physicians," Sociometry **20**, 253–270 (1957).

[12] Fragkiskos Papadopoulos, Constantinos Psomas, and Dmitri Krioukov, "Network mapping by replaying hyperbolic growth," IEEE/ACM Transactions on Networking **23**, 198–211 (2015).

[13] Fragkiskos Papadopoulos, Rodrigo Aldecoa, and Dmitri Krioukov, "Network geometry inference using common neighbors," Phys. Rev. E **92**, 022807 (2015).

[14] "HyperMap-CN Software Package," `https://bitbucket.org/dk-lab/2015_code_hypermap`.

[15] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá, "Hyperbolic geometry of complex networks," Phys. Rev. E **82**, 036106 (2010).

[16] Ittai Abraham, Shiri Chechik, David Kempe, and Aleksandrs Slivkins, "Low-distortion Inference of Latent Similarities from a Multiplex Social Network," SIAM Journal on Computing **44**, 617–668 (2015).

[17] Marián Boguñá, Fragkiskos Papadopoulos, and Dmitri Krioukov, "Sustaining the Internet with hyperbolic mapping." Nature communications **1**, 62 (2010).

[18] M. Ángeles Serrano, Marián Boguñá, and Francesc Sagués, "Uncovering the hidden geometry behind metabolic networks," Mol. BioSyst. **8**, 843–850 (2012).

[19] Fragkiskos Papadopoulos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá, and Dmitri Krioukov, "Popularity versus similarity in growing networks," Nature **489**, 537–540 (2012).

[20] Konstantin Zuev, Marián Boguñá, Ginestra Bianconi, and Dmitri Krioukov, "Emergence of soft communities from geometric preferential attachment," Scientific Reports **5**, 9421 (2015).

[21] Andrew Ng, Michael Jordan, and Yair Weiss, "On Spectral Clustering: Analysis and an algorithm," Advances in Neural Information Processing Systems 14 **2**, 849–856 (2002).

[22] Hans-Peter Kriegel, Peer Kröger, and Arthur Zimek, "Clustering High-dimensional Data: A Survey on Subspace Clustering, Pattern-based Clustering, and Correlation Clustering," ACM Trans. Knowl. Discov. Data **3**, 1:1–1:58 (2009).

[23] Guangyao Zhu and Kan Li, "A unified model for community detection of multiplex networks," *Web Information Systems Engineering – WISE 2014*, Lecture Notes in Computer Science, **8786**, 31–46 (2014).

[24] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela, "Community Structure in Time-Dependent, Multiscale, and Multiplex Networks," Science **328**, 876–878 (2010).

[25] Chuan Wen Loe and Henrik Jeldtoft Jensen, "Comparison of communities detection algorithms for multiplex," Physica A: Statistical Mechanics and its Applications **431**, 29–45 (2015).

[26] Natalie Stanley, Saray Shai, Dane Taylor, and Peter J. Mucha, "Clustering Network Layers With the Strata Multilayer Stochastic Block Model," Preprint at http://arxiv.org/abs/1507.01826 (2015).

[27] D. Taylor, S. Shai, N. Stanley, and P. J. Mucha, "Enhanced detectability of community structure in multilayer networks through layer aggregation," Preprint at https://arxiv.org/abs/1511.05271 (2015).

[28] "The CAIDA AS Organizations Dataset, 09-Jan-2015," (2015), `http://www.caida.org/data/as-organizations`.

[29] Linyuan Lü and Tao Zhou, "Link prediction in complex networks: A survey," Physica A: Statistical Mechanics and its Applications **390**, 1150–1170 (2011).

[30] M. Ángeles Serrano, Dmitri Krioukov, and Marián Boguñá, "Self-Similarity of Complex Networks and Hidden Metric Spaces," Phys. Rev. Lett. **100**, 078701 (2008).

[31] Roger B. Nelsen, *An Introduction to Copulas*, Springer Series in Statistics (Springer-Verlag New York, 2006).

[32] Vincenzo Nicosia and Vito Latora, "Measuring and modeling correlations in multiplex networks," Phys. Rev. E **92**, 032805 (2015).

[33] Michael Szell, Renaud Lambiotte, and Stefan Thurner, "Multirelational organization of large-scale social networks in an online world," Proceedings of the National Academy of Sciences **107**, 13636–13641 (2010).

[34] Ginestra Bianconi, "Statistical mechanics of multiplex networks: Entropy and overlap," Phys. Rev. E **87**, 062806 (2013).

[35] Davide Cellai, Eduardo López, Jie Zhou, James P. Gleeson, and Ginestra Bianconi, "Percolation in multiplex networks with overlap," Phys. Rev. E **88**, 052811 (2013).

[36] Giulia Menichetti, Daniel Remondini, and Ginestra Bianconi, "Correlations between weights and overlap in ensembles of weighted multiplex networks," Phys. Rev. E **90**, 062817 (2014).

[37] Federico Battiston, Vincenzo Nicosia, and Vito Latora, "Structural measures for multiplex networks," Phys. Rev. E **89**, 032804 (2014).

[38] Elisa Omodei, Manlio De Domenico, and Alex Arenas, "Characterizing interactions in online social networks during exceptional events," Frontiers in Physics **3** (2015).

[39] Theodore Wilbur Anderson, *An introduction to multivariate statistical analysis*, Wiley series in probability and mathematical statistics (J. Wiley, 1984).

[40] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger, "Estimating mutual information," Phys. Rev. E

69, 066138 (2004).

[41] "Mutual Information Least-dependent Component Analysis (MILCA)," https://www.ucl.ac.uk/ion/departments/sobell/Research/RLemon/MILCA/MILCA.

[42] Jeffrey Travers and Stanley Milgram, "An experimental study of the small world problem," Sociometry **32**, 425–443 (1969).

[43] James F. Kurose and Keith W. Ross, *Computer Networking: A Top-Down Approach*, 5th ed. (Addison-Wesley Publishing Company, 2009).

[44] "6in4," (2016), https://en.wikipedia.org/wiki/6in4.