Supplemental Material for k-core structure of real multiplex networks

Saeed Osat,[1,*] Filippo Radicchi,[2] and Fragkiskos Papadopoulos[3]

[1] *Deep Quantum Labs, Skolkovo Institute of Science and Technology, Moscow 143026, Russia*
[2] *Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana 47408, USA*
[3] *Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, 33 Saripolou Street, 3036 Limassol, Cyprus*

CONTENTS

* saeedosat13@gmail.com

I. REAL-WORLD NETWORK DATA

Table I gives an overview of the analyzed single-layer networks, while Table II gives an overview of the considered multiplex systems that consist of network layers from Table I.

Network	N	$\bar{k}$	γ	Ref.
Airport	3397	11.32	1.88	[1–3]
Drosophila	1770	10.01	1.91	[3, 4]
Metabolic	1436	6.57	2.6	[3, 5]
Proteome	4100	6.52	2.25	[3, 6]
Enron	33696	10.73	2.66	[3, 7, 8]
Internet	23748	4.92	2.17	[3, 9]
Music	2476	16.66	2.27	[3, 10, 11]
Words	7377	11.99	2.25	[3, 12]
arXiv1 (physics.bio-ph)	2956	4.13	2.6	[13, 14]
arXiv2 (physics.data-an)	3506	4.19	2.6	
arXiv3 (physics.soc-ph)	1594	3.79	6.0	
arXiv4 (cond-mat.dis-nn)	5465	5.30	2.5	
arXiv5 (math.OC)	1605	5.52	4.0	
arXiv6 (cond-mat.stat-mech)	1451	3.56	4.0	
arXiv7 (q-bio.MN)	1905	4.64	4.0	
arXiv8 (cs.SI)	4946	4.69	2.5	
C.Elegans1 (Electric)	253	4.06	2.9	[14–16]
C.Elegans2 (Chemical Monadic)	260	6.83	2.9	
C.Elegans3 (Chemical Polyadic)	278	12.25	2.9	
Drosophila1 (Suppressive)	838	4.43	2.6	[14, 17, 18]
Drosophila2 (Additive)	755	3.77	2.8	
Internet1 (IPv4)	37563	5.06	2.1	[9, 14, 19]
Internet2 (IPv6)	5163	5.21	2.1	
Air	69	5.22	2.6	[14, 20]
Train	69	9.33	2.9	

TABLE I. **Single-layer networks.** The first column indicates the name of the network, while N, $\bar{k}$ and γ denote respectively the number of nodes in the network, the network's average degree, and the exponent γ of the power law that best approximates the network's degree distribution. More details on the data can be found in the references listed in the last column.

Multiplex network	N_{common}	$r_{k,k'}$	$\text{NMI}_{\theta,\theta'}$	Ref.
arXiv1 - arXiv2	911	0.82	0.46	[13, 14]
arXiv1 - arXiv4	1441	0.90	0.56	
arXiv1 - arXiv5	354	0.80	0.44	
arXiv2 - arXiv4	1323	0.86	0.48	
Internet1 - Internet2	4731	0.82	0.32	[9, 14]
C.Elegans2 - C.Elegans3	259	0.80	0.31	[14–16]
Drosophila1 - Drosophila2	500	0.83	0.26	[14, 17, 18]
Air - Train	69	0.80	0.08	[14, 20]

TABLE II. **Multiplex networks.** Each two-layer multiplex network is composed of layers from Table I. The first column indicates the two layers that constitute the multiplex. N_{common} is the number of common nodes between the two layers; $r_{k,k'}$ is the Pearson correlation coefficient among the degrees of the nodes in the two layers; and $\text{NMI}_{\theta,\theta'}$ is the normalized mutual information of the angular coordinates of the nodes in the two layers.

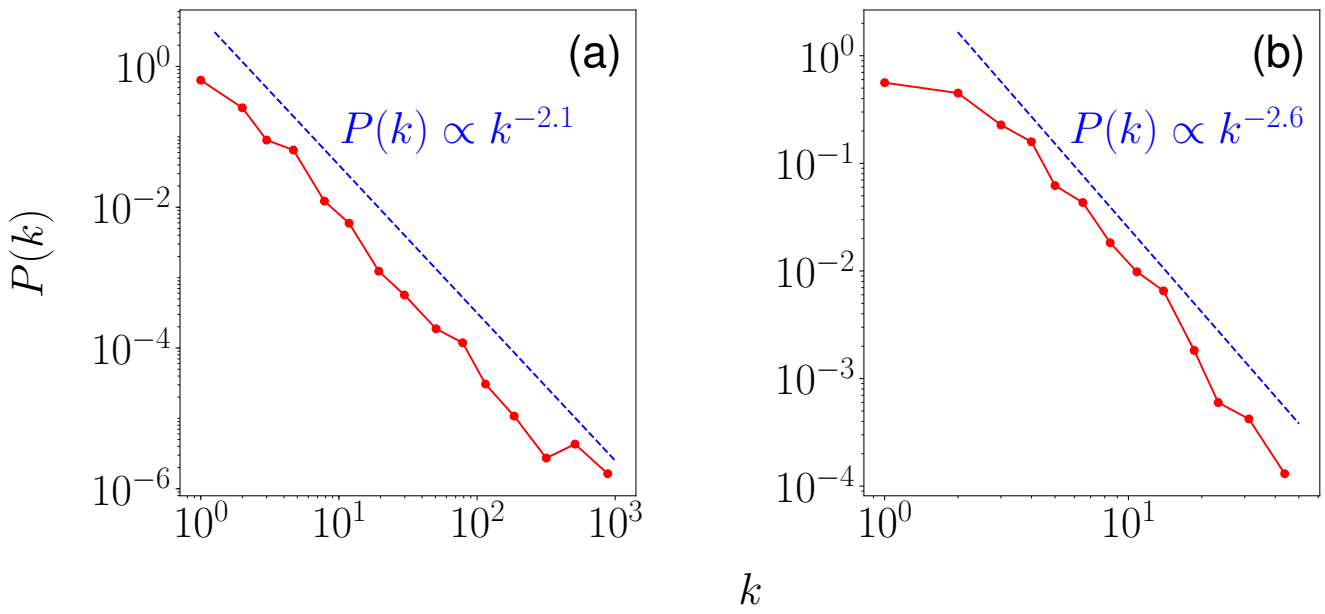

FIG. 1. **Degree Distribution.** Degree distribution of (**a**) the IPv6 Internet and (**b**) the arXiv co-authorship network (physics.bio-ph) considered in the main text. The red lines are the empirical distributions, while the dotted blue lines are power laws with exponents $\gamma = 2.1, 2.6$.

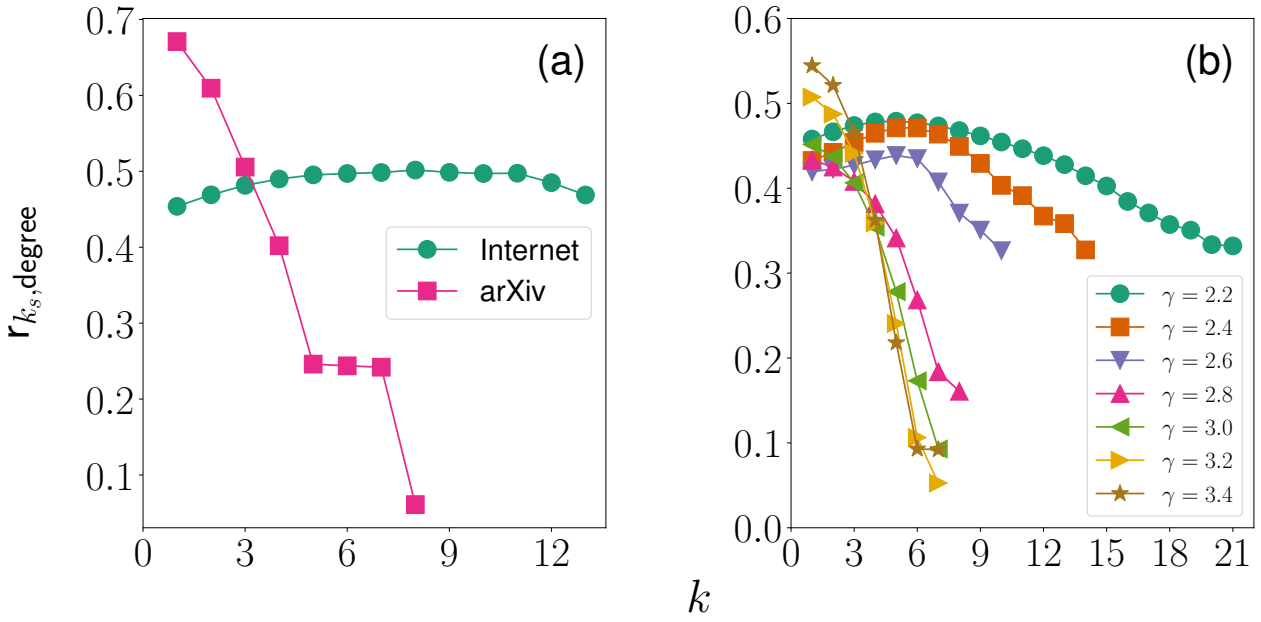

FIG. 2. **Correlation between node degree and k-shell index.** Pearson correlation coefficient between node degree and k-shell index, $r_{k_s,\text{degree}}$, for nodes with k-shell index $k_s \geq k$ where k as indicated in the x-axis. Panel (**a**) shows the results for the IPv6 Internet and the arXiv co-authorship network considered in the main text. Panel (**b**) shows the results for synthetic networks constructed with the $\mathbb{S}^1$-model (Methods section C in the main text) for various values of the degree exponent γ. The synthetic networks have $N = 10,000$ nodes, average degree $\bar{k} = 6$, and temperature parameter $T = 0.5$, while the results are averages over 100 network realizations. Results for a value of k are shown only if there were samples for that value in at least 20% of the realizations.

II. k-CORE STRUCTURE OF REAL-WORLD NETWORKS

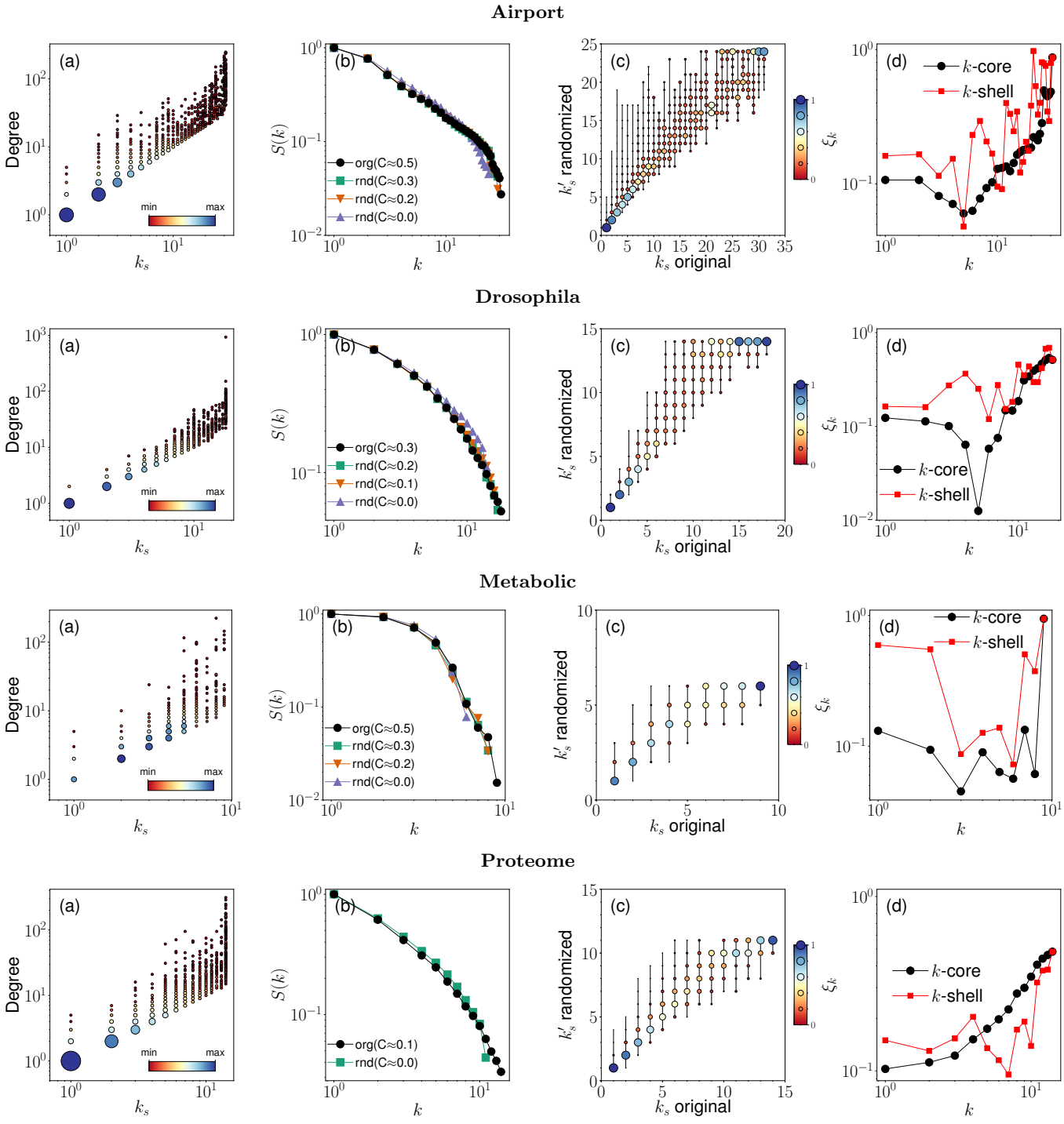

FIG. 3. **k-core structure of real-world networks.** Each row corresponds to a single-layer network from Table I. (**a**) Scatter plot of node degrees vs. coreness. The size of the symbols is proportional to the number of nodes having each specific degree and k-shell index values. (**b**) Relative size $S(k)$ of the k-core in the real networks (labeled as "org") and their randomized counterparts (labeled as "rnd"). Randomized networks are obtained by shuffling random pairs of edges while controlling for the average value of the clustering coefficient C. (**c**) k-shell index of nodes before and after network randomization (obtained for $C \approx 0$). The size of the symbols is proportional to the percentage of nodes whose coreness changed from k_s in the original network to k'_s in the reshuffled network. (**d**) Angular coherence ξ_k of the nodes belonging to each k-core and k-shell.

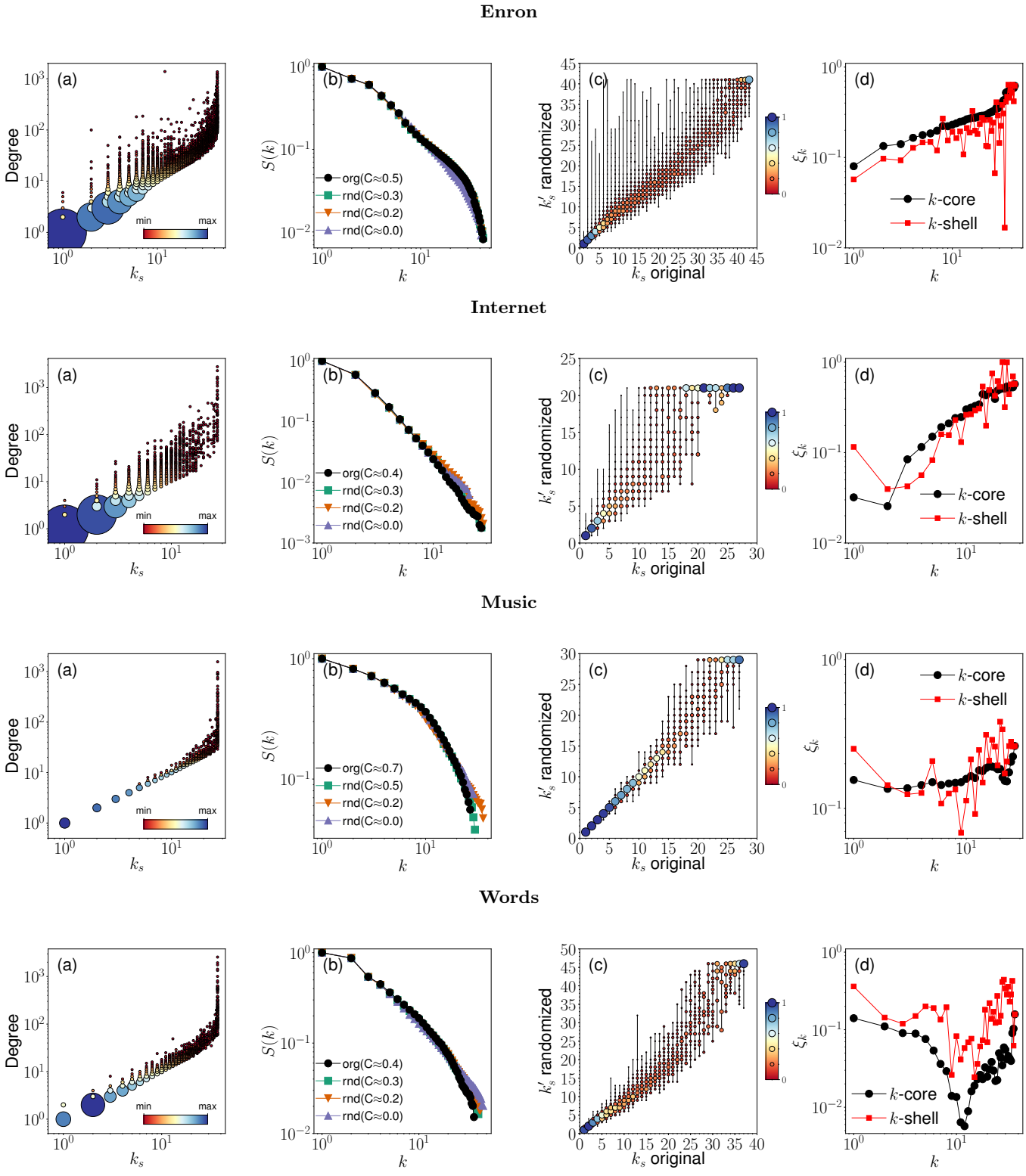

FIG. 4. ***k*-core structure of real-world networks.** Same as in Figure 3, but for a different set of networks from Table I.

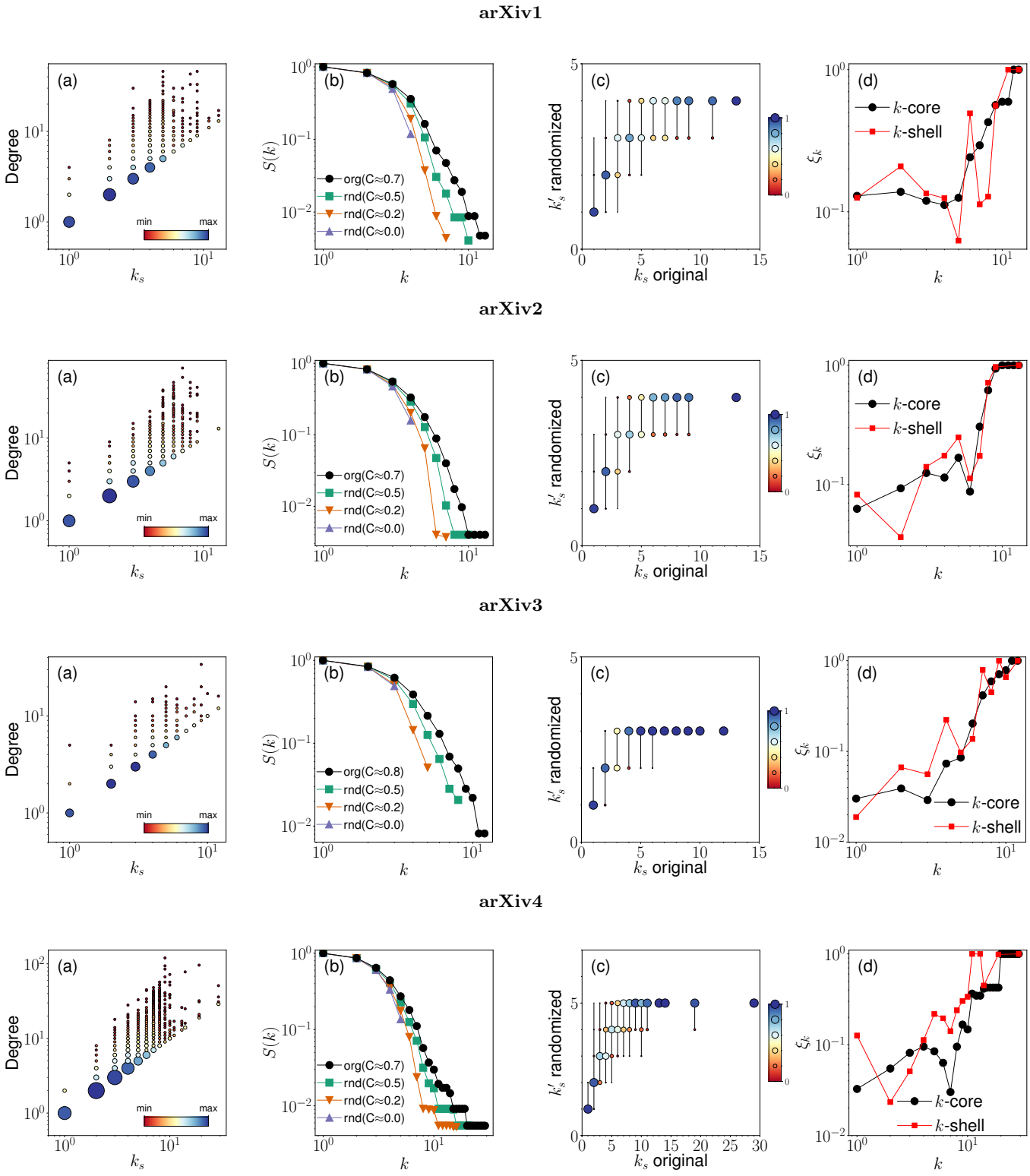

FIG. 5. **k-core structure of real-world networks.** Same as in Figure 3, but for a different set of networks from Table I.

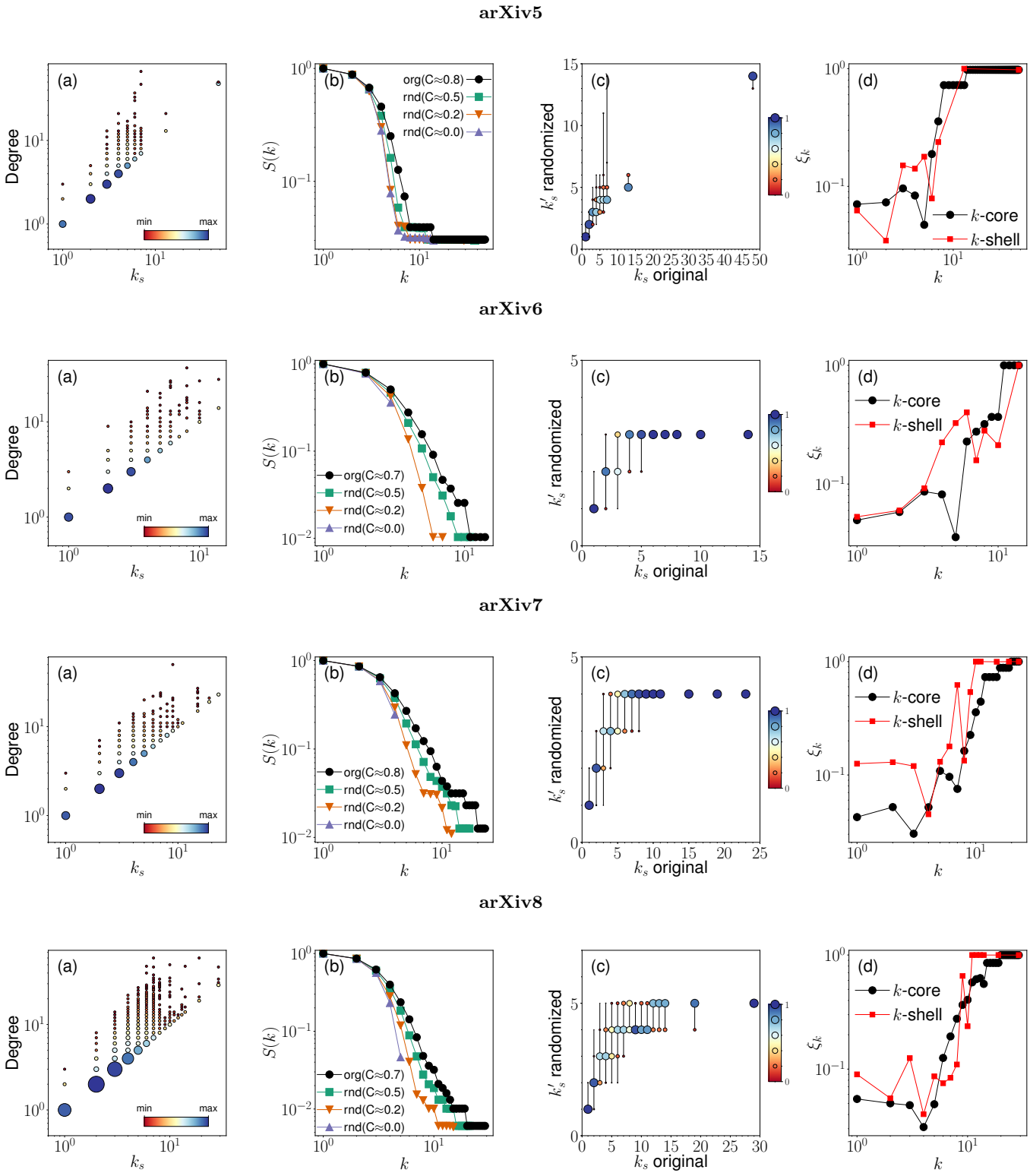

FIG. 6. ***k*-core structure of real-world networks.** Same as in Figure 3, but for a different set of networks from Table I.

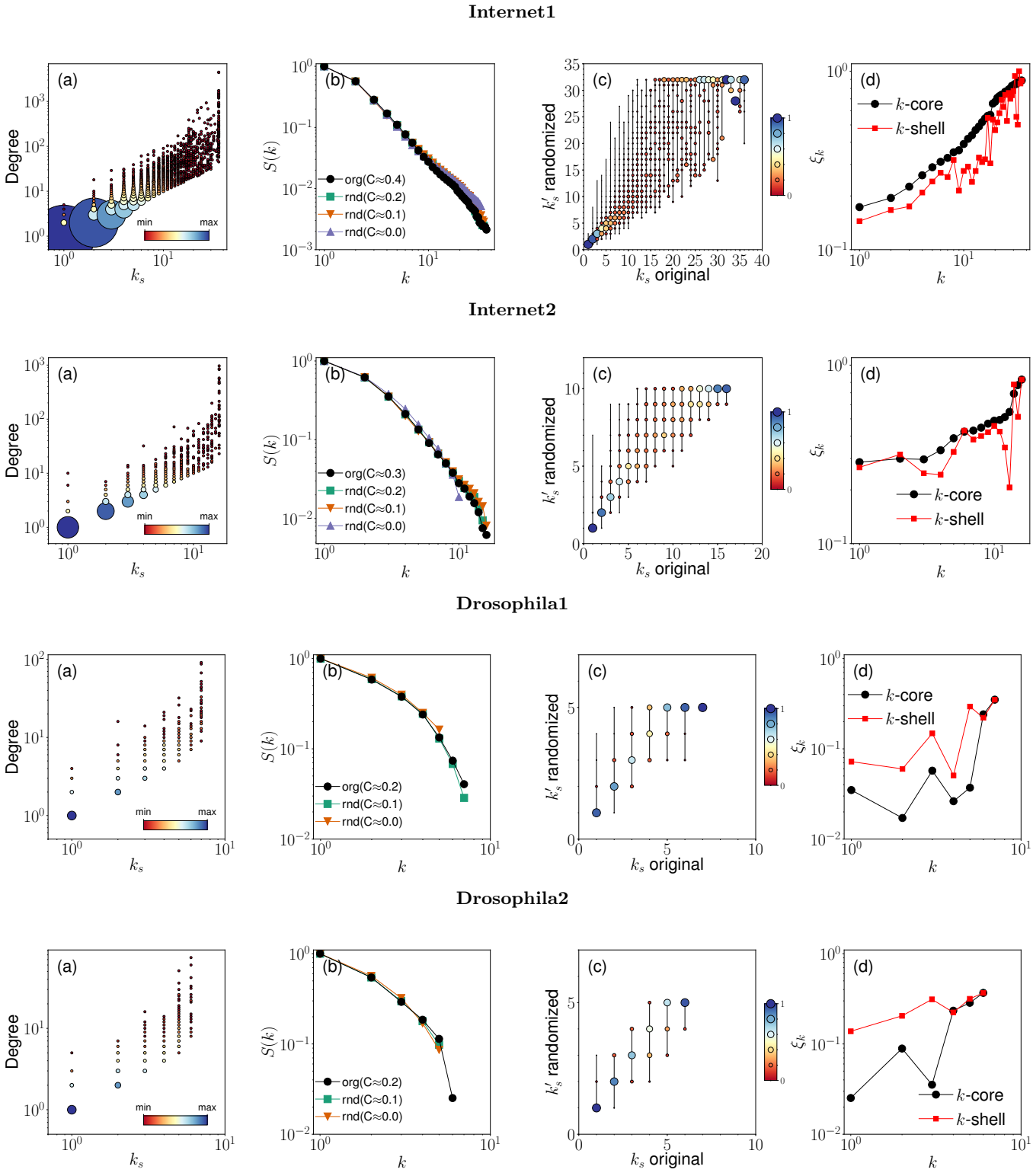

FIG. 7. ***k*-core structure of real-world networks.** Same as in Figure 3, but for a different set of networks from Table I.

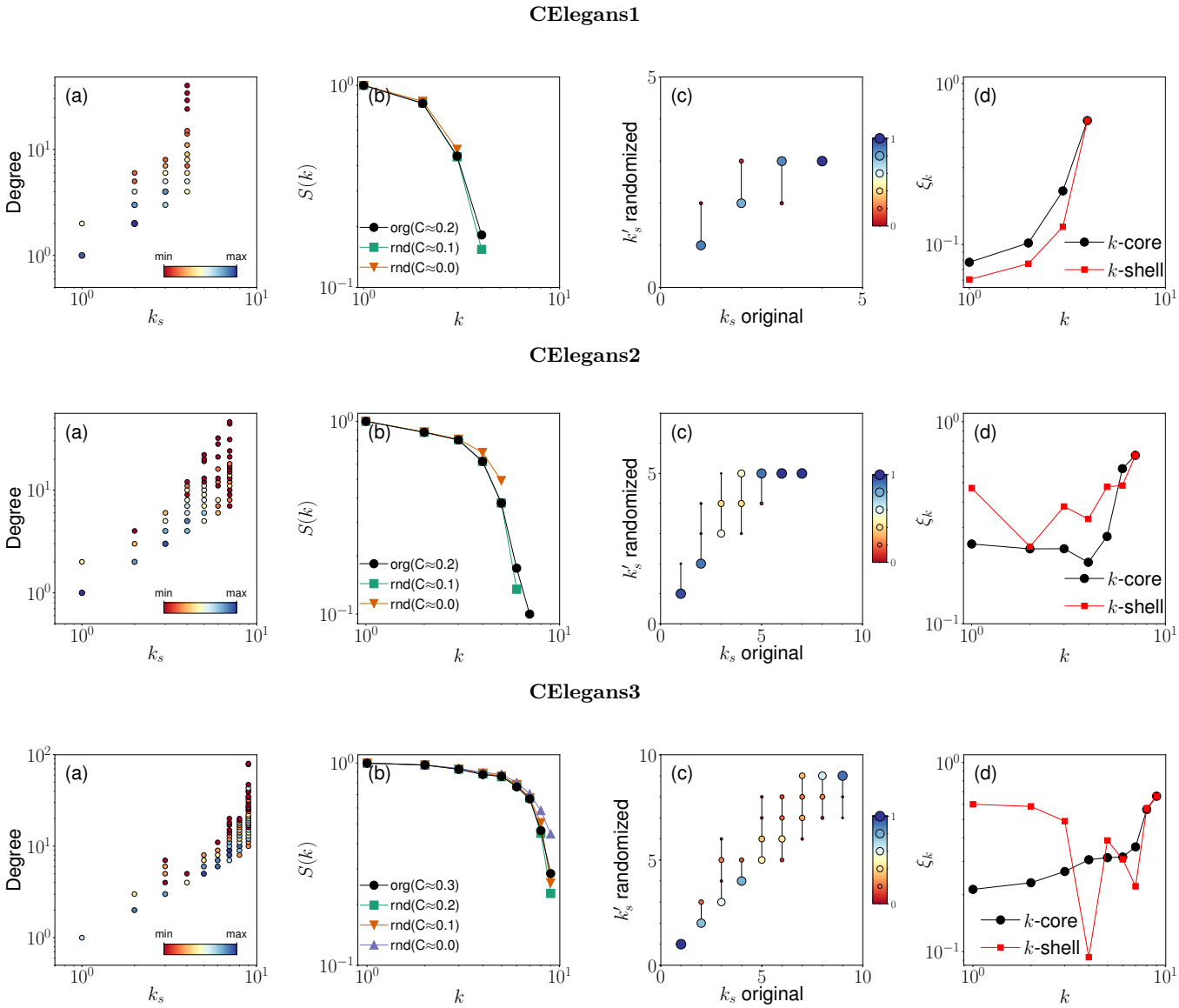

FIG. 8. **k-core structure of real-world networks.** Same as in Figure 3, but for a different set of networks from Table I.

III. k-CORE STRUCTURE OF REAL-WORLD MULTIPLEX NETWORKS

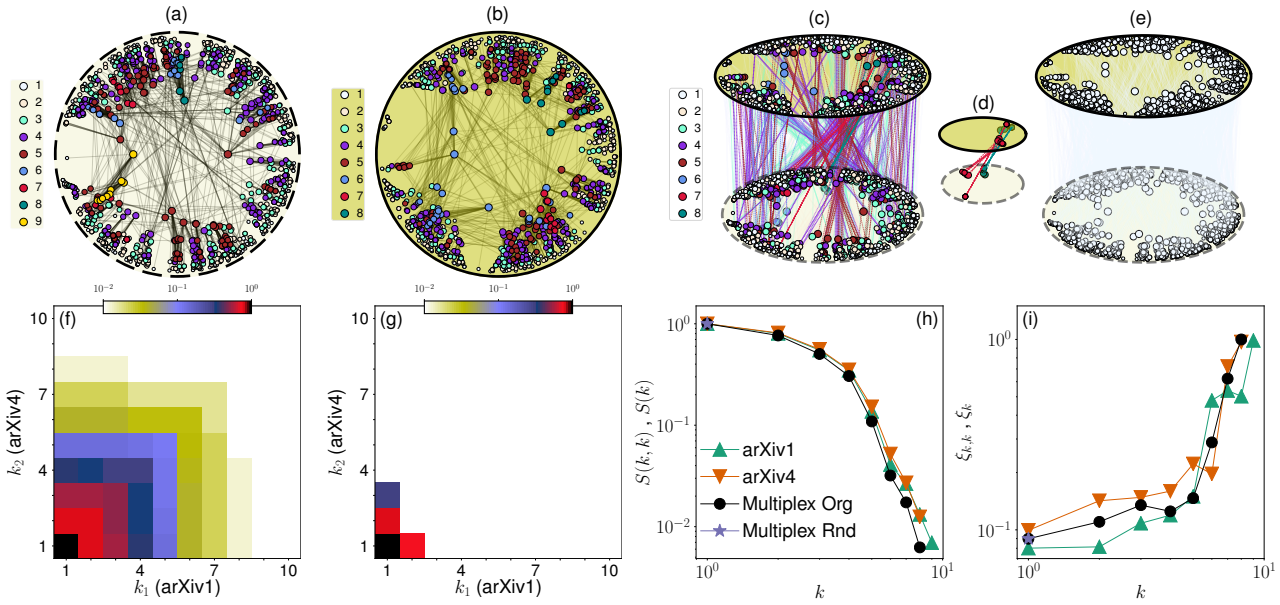

FIG. 9. **k-core of real-world multiplex networks.** (**a** and **b**) Hyperbolic embedding of the arXiv1-arXiv4 multiplex network. Panel a refers to the layer arXiv1, while panel b to the layer arXiv4. The position of the nodes in the disk is determined by their hyperbolic coordinates, and only nodes that exist in both layers are shown; different colors serve to differentiate nodes depending on their k-shell index value. (**c**) Correspondence among nodes belonging to the (k,k)-shells of the arXiv1-arXiv4 multiplex network. (**d**) Same as in panel c, but for $k \geq 7$. (**e**) Same as in panel c, but for the randomized version of the multiplex where the node labels of one of the two layers are randomly reshuffled. (**f**) Relative size $S(k_1,k_2)$ of the (k_1,k_2)-core for the arXiv1-arXiv4 multiplex network. (**g**) Same as in panel f, but for the randomized version of the multiplex network. (**h**) Relative size $S(k,k)$ of the (k,k)-core for the arXiv1-arXiv4 multiplex network, and its randomized version. These curves are compared with those of the relative size $S(k)$ of the k-core of the individual layers. (**i**) Same as in panel h, but for the metrics of angular coherence $\xi_{k,k}$ and ξ_k.

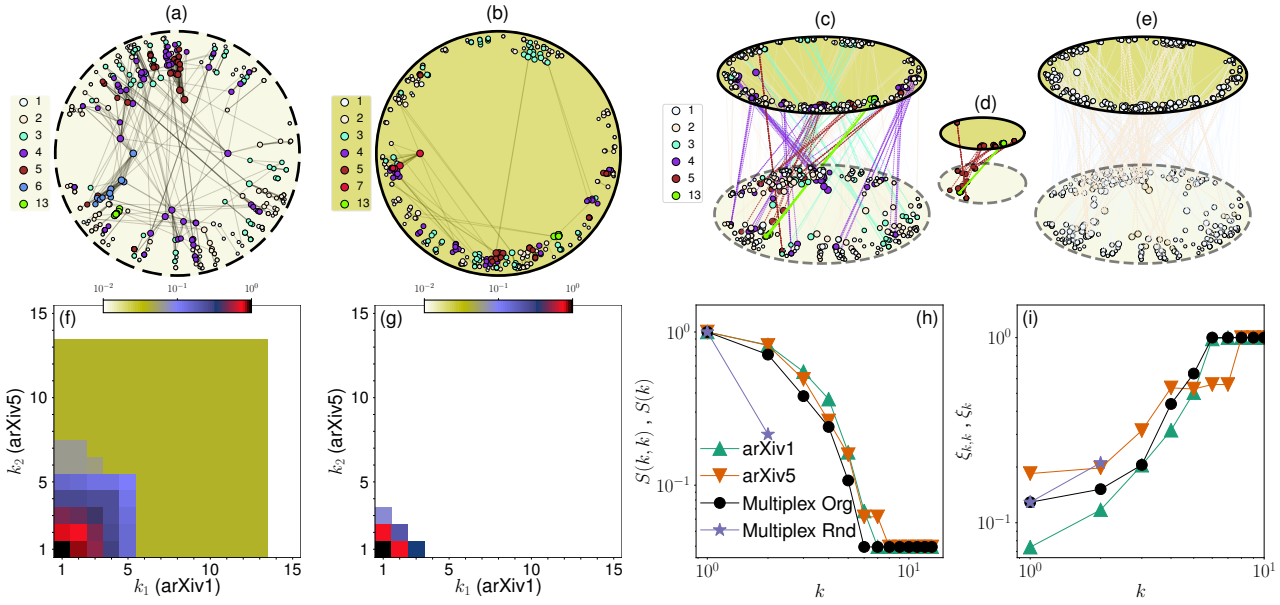

FIG. 10. **k-core of real-world multiplex networks.** Same as in Figure 9, but for the multiplex consisting of the layers arXiv1 (a) and arXiv5 (b). Panel (**d**) shows the correspondence among nodes belonging to the (k,k)-shells with $k \geq 5$.

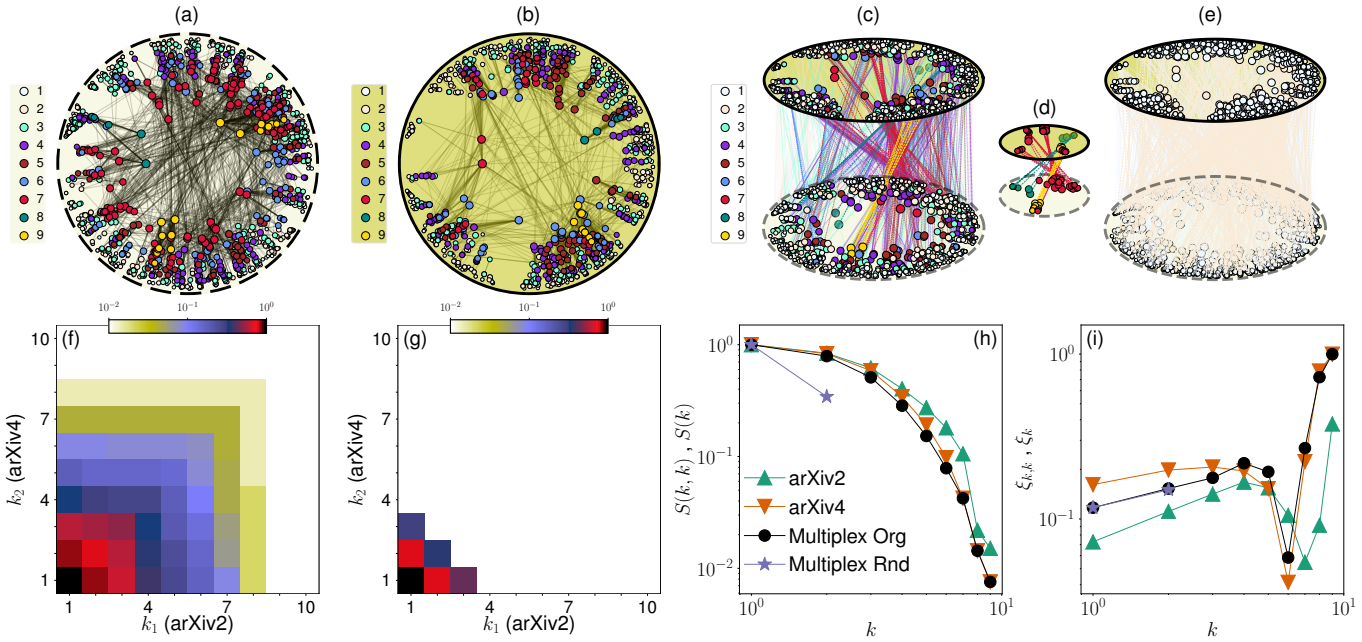

FIG. 11. **k-core of real-world multiplex networks.** Same as in Figure 9, but for the multiplex consisting of the layers arXiv2 (a) and arXiv4 (b). Panel (**d**) shows the correspondence among nodes belonging to the (k,k)-shells with $k \geq 7$.

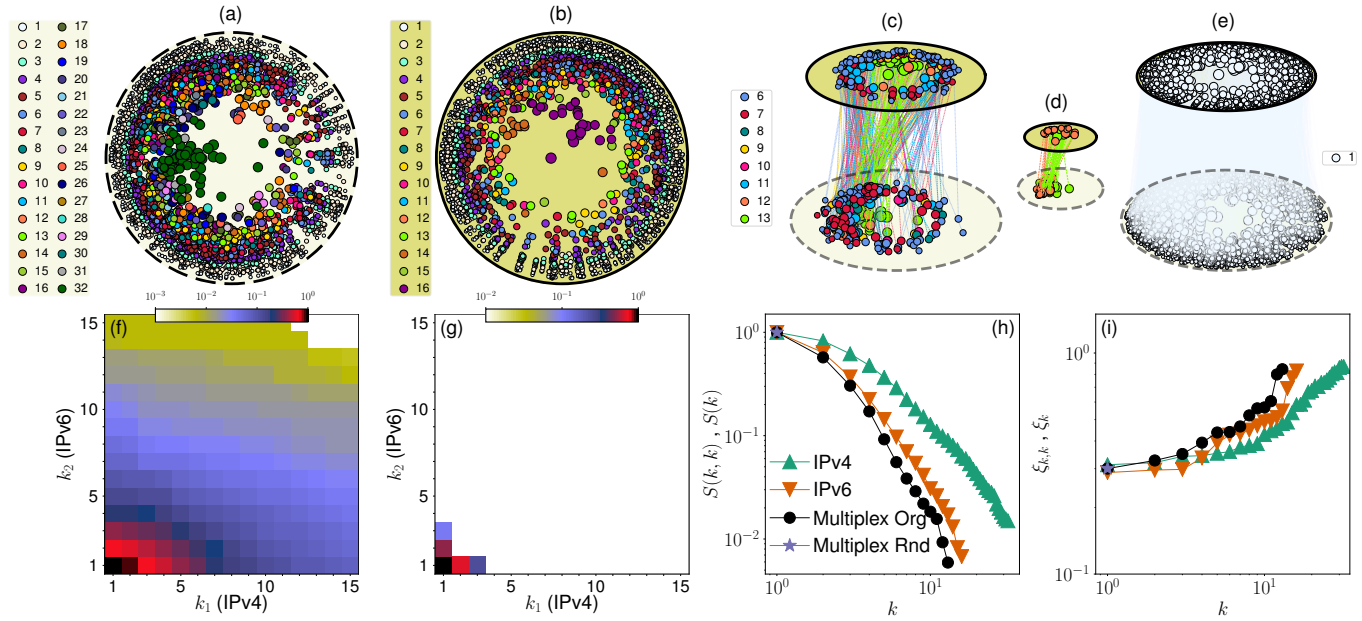

FIG. 12. **k-core of real-world multiplex networks.** Same as in Figure 9, but for the multiplex consisting of the layers IPv4 Internet (a) and the IPv6 Internet (b). Panel (**c**) shows the correspondence among nodes belonging to the (k,k)-shells with $k \geq 6$. (d) Same as in panel c, but for $k \geq 12$.

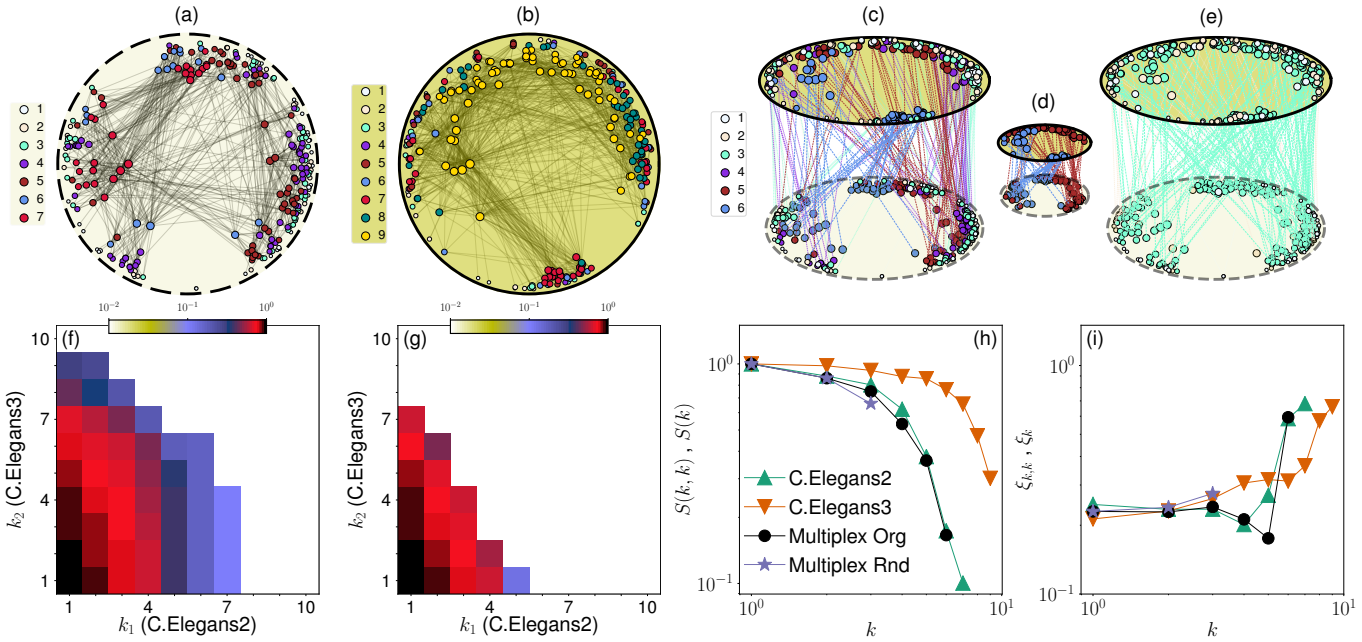

FIG. 13. **k-core of real-world multiplex networks.** Same as in Figure 9, but for the multiplex consisting of the layers C.Elegans2 (a) and C.Elegans3 (b). Panel (**d**) shows the correspondence among nodes belonging to the (k,k)-shells with $k \geq 5$.

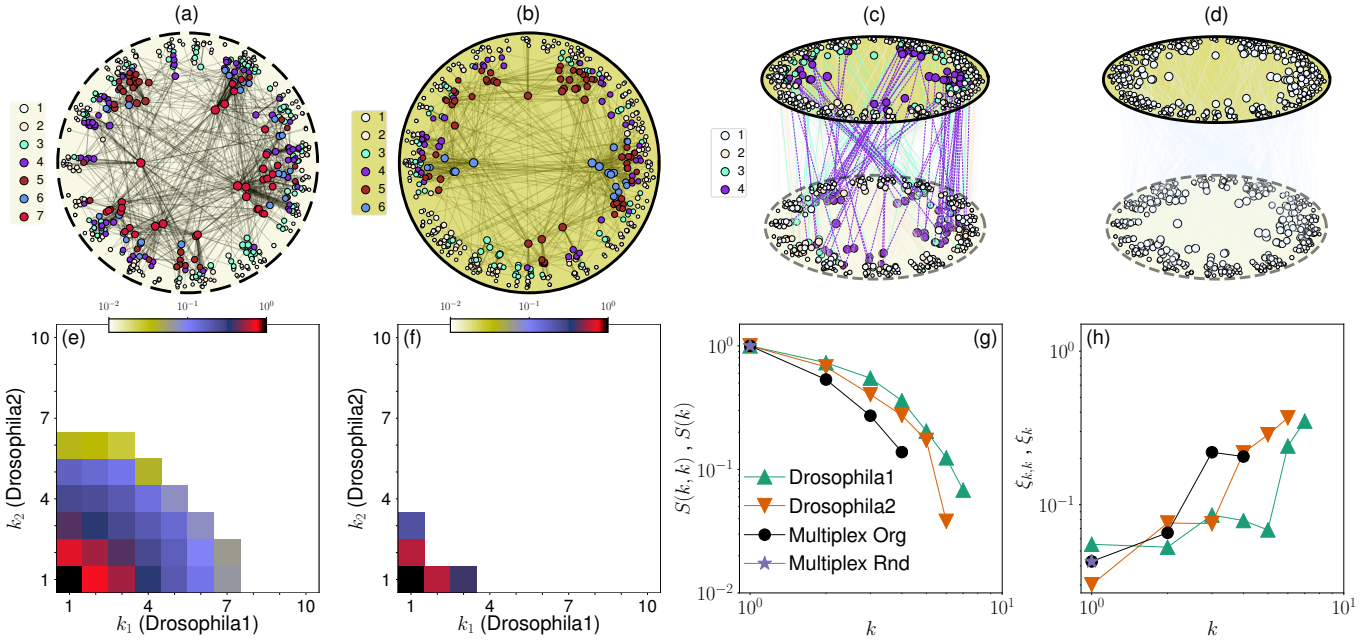

FIG. 14. **k-core of real-world multiplex networks.** Same as in Figure 9, but for the multiplex consisting of the layers Drosophila1 (a) and Drosophila2 (b). This figure does not contain a zoom-in of inner shells.

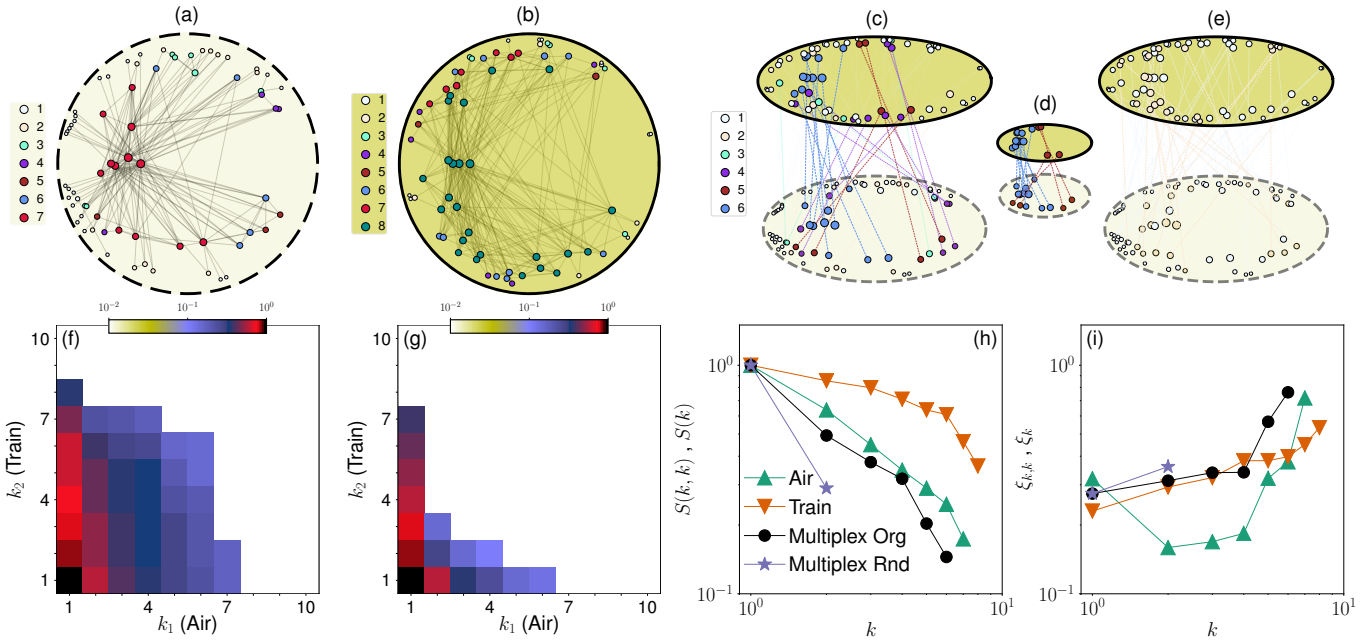

FIG. 15. **k-core of real-world multiplex networks.** Same as in Figure 9, but for the multiplex consisting of the layers Air (a) and Train (b). Panel (**d**) shows the correspondence among nodes belonging to the (k, k)-shells with $k \geq 5$.

IV. DESTROYING INTER-LAYER DEGREE AND SIMILARITY CORRELATIONS

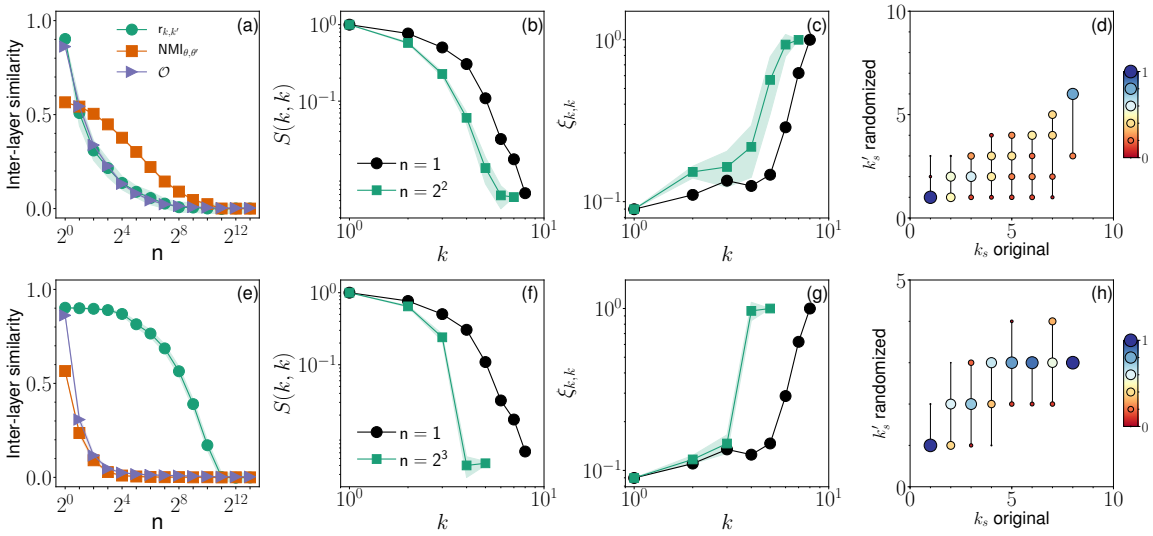

FIG. 16. **Inter-layer correlations and the k-core structure of the arXiv1-arXiv4 multiplex network.** We analyze the multiplex network consisting of arXiv1 and arXiv4. (**a**) Different metrics of inter-layer similarity as a function of the group size n used to randomize node labels, thus breaking inter-layer degree correlations. For $n = 1$, node labels of the network are not randomized; full shuffle of node labels is obtained for large n values. We focus here on the case where degree inter-layer correlation is broken, but we preserve inter-layer correlation among nodes' angular coordinates. Metrics of similarities considered here are the Pearson correlation coefficient $r_{k,k'}$ among the degrees of nodes in the two layers; normalized mutual information $\mathrm{NMI}_{\theta,\theta'}$ of the angular coordinates of the nodes in the two layers; and edge overlap $\mathcal{O}$ among the two layers. (**b**) Relative size $S(k,k)$ of the (k,k)-core. The results of the original multiplex network ($n = 1$) are compared with those valid for $n = 4$. The results for $n = 4$ are average values obtained on 100 independent randomizations. Shaded areas identify the region corresponding to one standard deviation away from the average. (**c**) Same as in panel b, but for the angular coherence $\xi_{k,k}$. (**d**) Scatter plot of the (k,k)-shell index of nodes in the original vs. the randomized multiplex network. The size of the symbols is proportional to the percentage of points in the scatter plot. (**e**, **f**, **g** and **h**) Same as in panel a, b, c and d, respectively. We consider here the case where inter-layer correlation among nodes' angular coordinates is destroyed, but inter-layer correlation among nodes' degrees is preserved. The results of the original network are compared with those obtained for $n = 8$.

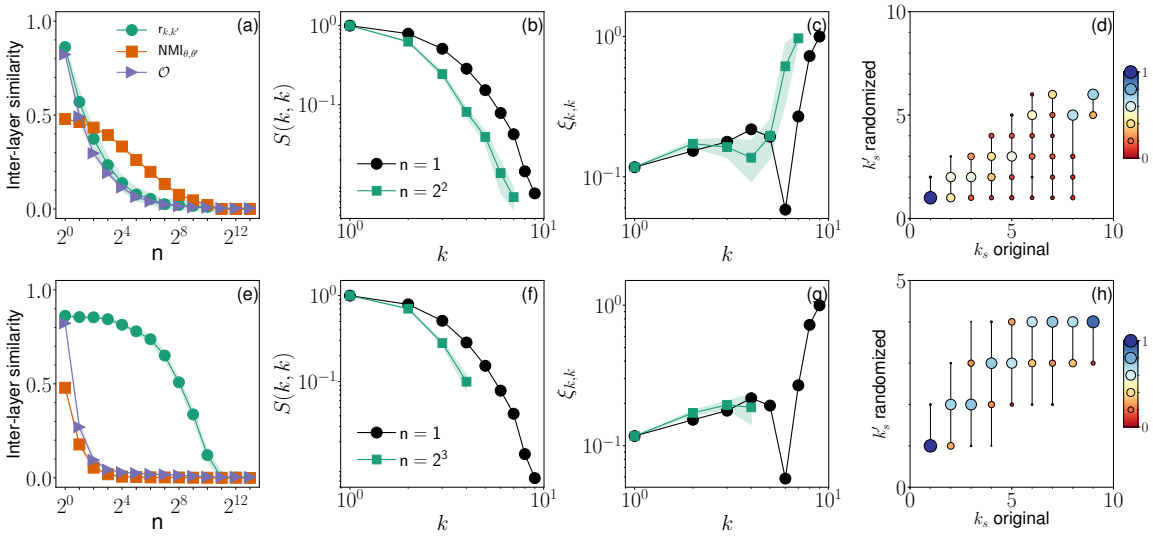

FIG. 17. **Inter-layer correlations and the k-core structure of the arXiv2-arXiv4 multiplex network.** Same as in Figure 16, but for the multiplex consisting of arXiv2 and arXiv4.

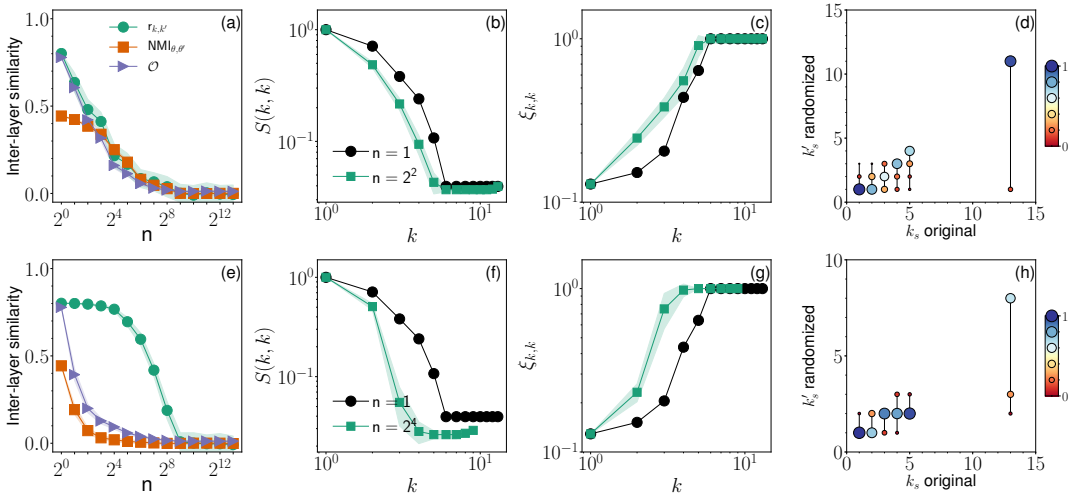

FIG. 18. **Inter-layer correlations and the k-core structure of the arXiv1-arXiv5 multiplex network.** Same as in Figure 16, but for the multiplex consisting of arXiv1 and arXiv5.

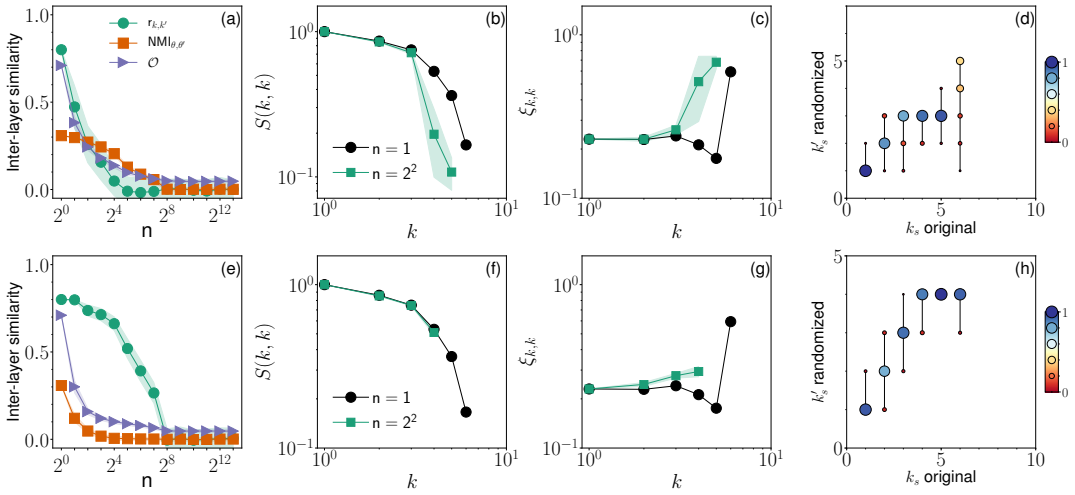

FIG. 19. **Inter-layer correlations and the k-core structure of the C.Elegans2-C.Elegans3 multiplex network.** Same as in Figure 16, but for the multiplex consisting of C.Elegans2 and C.Elegans3.

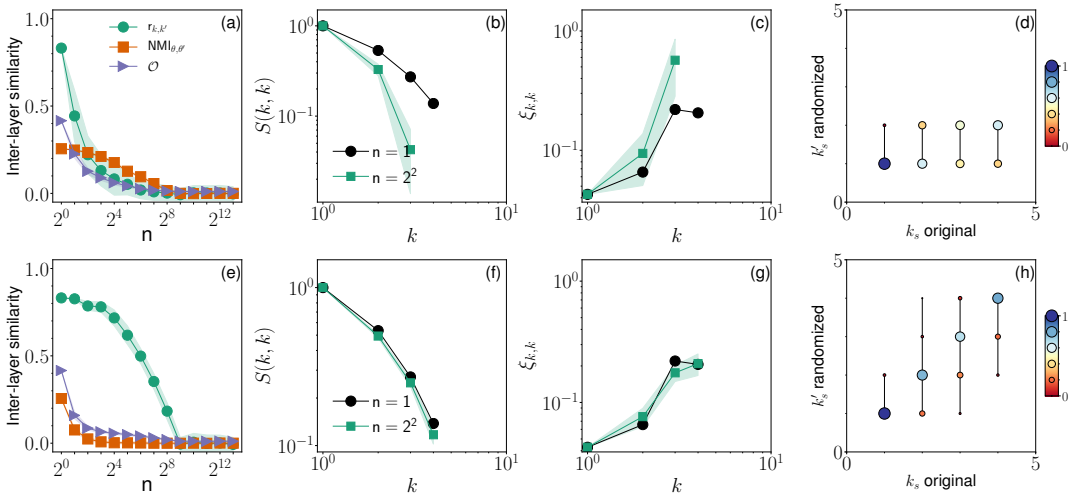

FIG. 20. **Inter-layer correlations and the k-core structure of the Drosophila1-Drosophila2 multiplex network.** Same as in Figure 16, but for the multiplex consisting of Drosophila1 and Drosophila2.

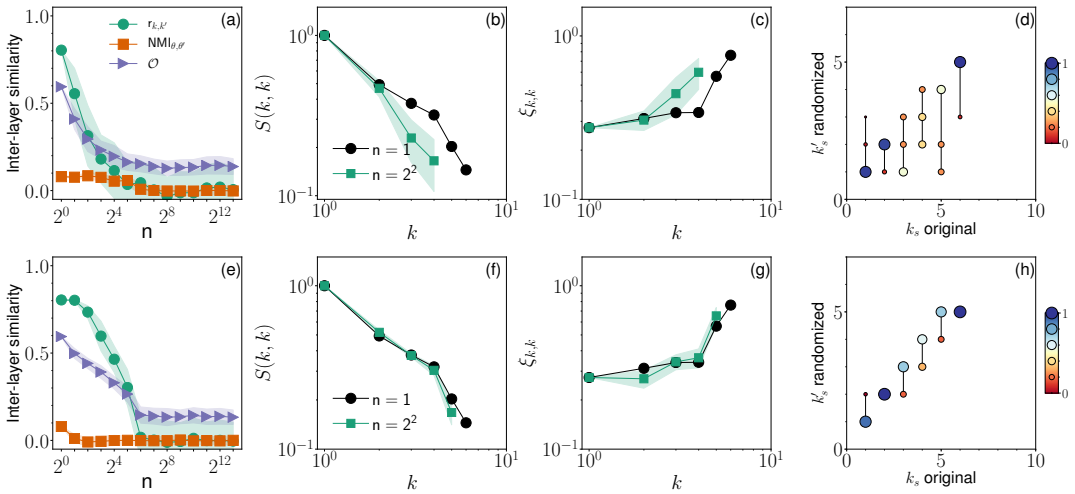

FIG. 21. **Inter-layer correlations and the k-core structure of the Air-Train multiplex network.** Same as in Figure 16, but for the multiplex consisting of Air and Train.

V. k-CORE STRUCTURE OF SYNTHETIC MULTIPLEX NETWORKS

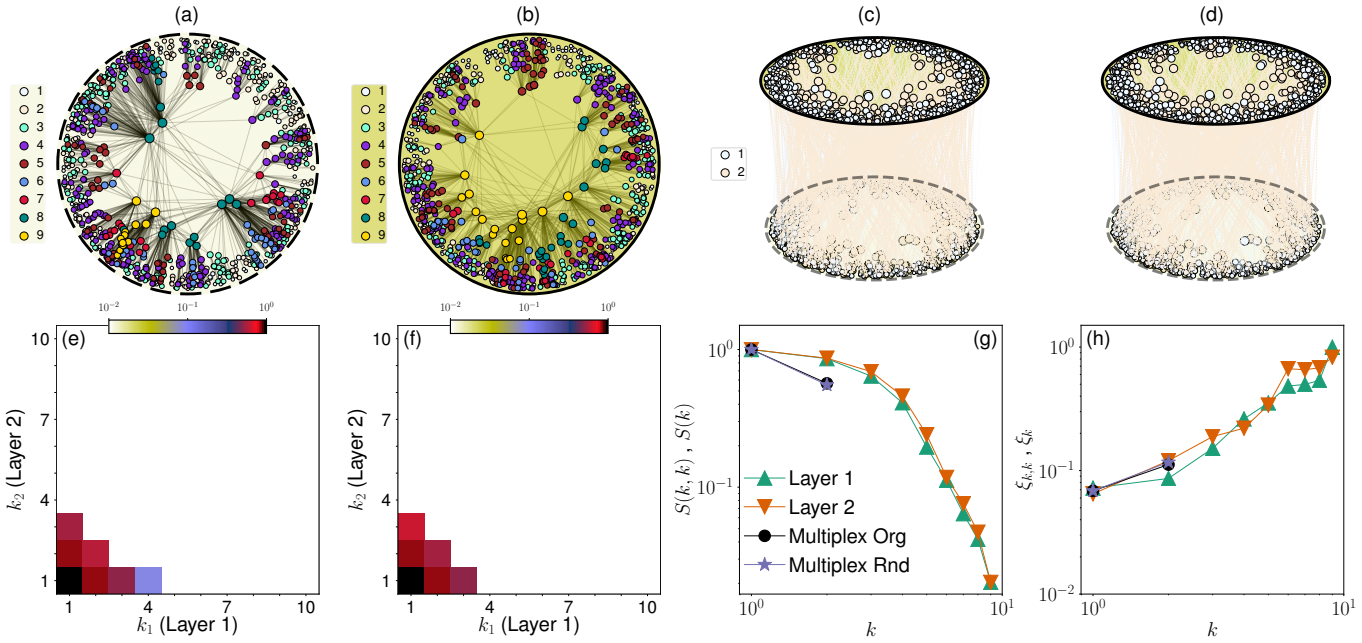

FIG. 22. **k-core structure of synthetic multiplex networks.** Same as in Figure 9, but for a two-layer synthetic multiplex network constructed according to the Geometric Multiplex Model (GMM) with no inter-layer degree and angular correlations ($\nu = g = 0$). Each layer of the multiplex has $N = 1000$ nodes, power-law degree distribution with exponent $\gamma = 2.6$, average degree $\bar{k} \approx 6$, and temperature $T = 0.1$.

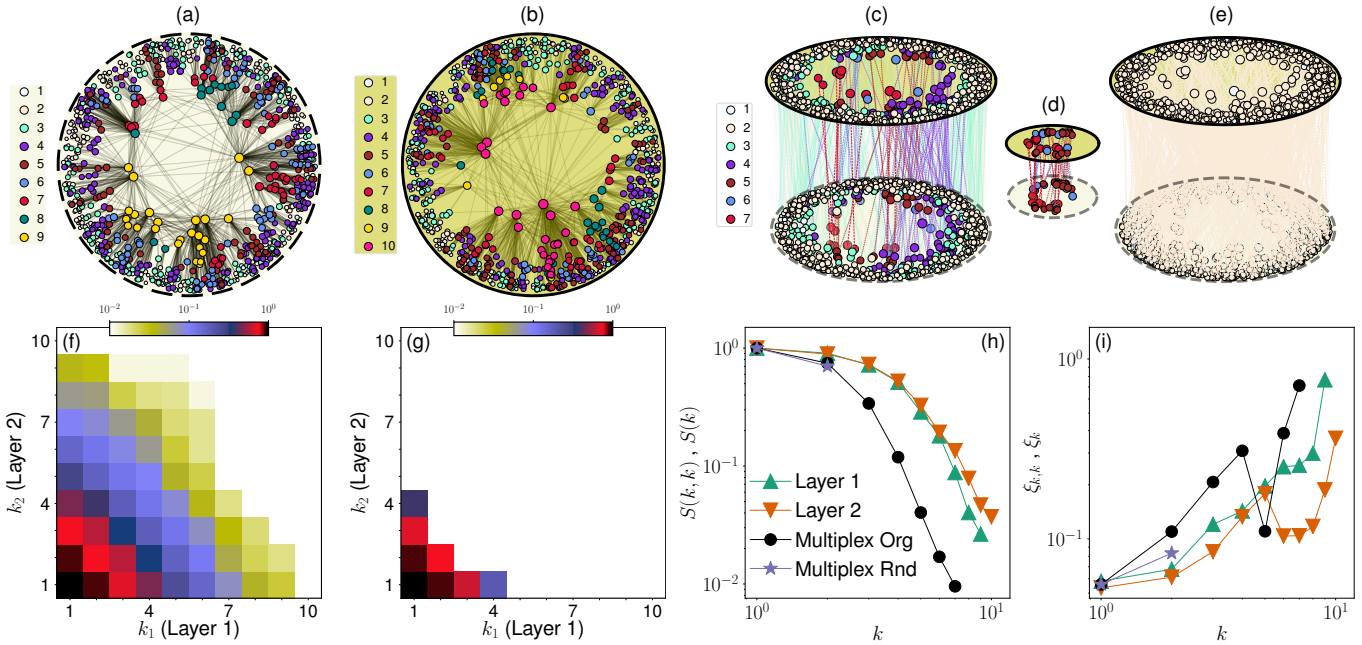

FIG. 23. **k-core structure of synthetic multiplex networks.** Same as in Figure 9, but for a two-layer synthetic multiplex network constructed according to the GMM with intermediate inter-layer degree and angular correlations ($\nu = g = 0.5$). Each layer of the multiplex has the same parameters as in Figure 22. Panel (**d**) shows the correspondence among nodes belonging to the (k,k)-shells with $k \geq 5$.

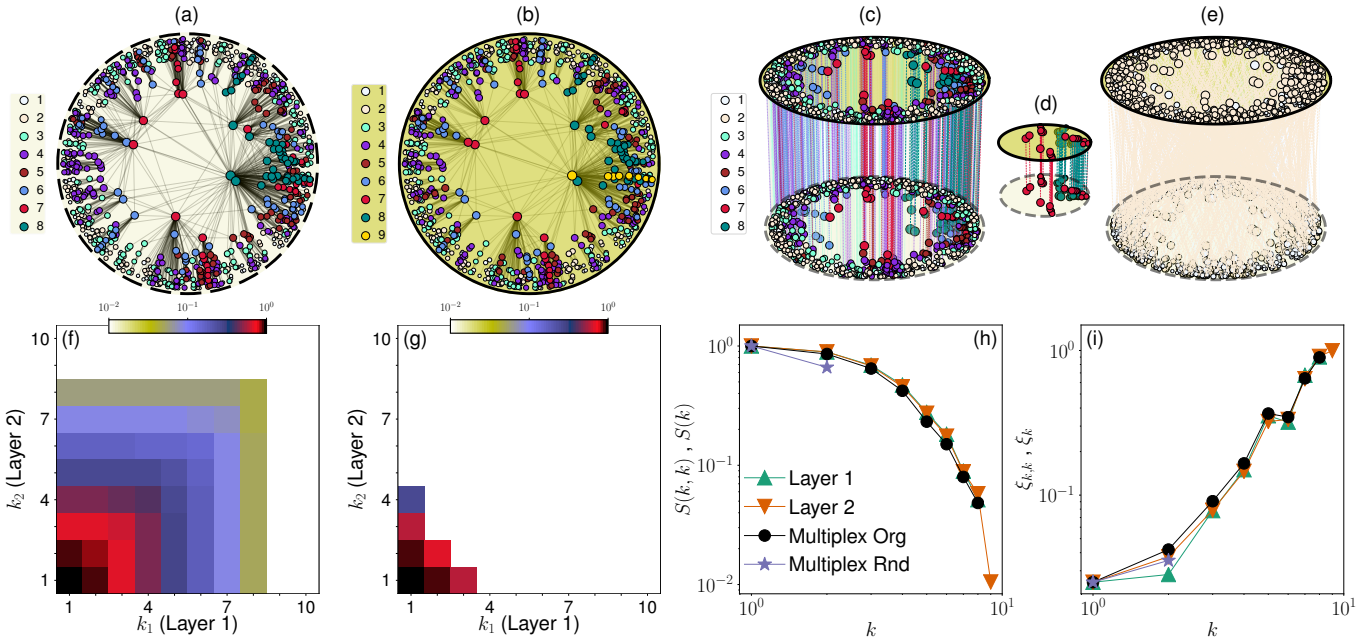

FIG. 24. **k-core structure of synthetic multiplex networks.** Same as in Figure 9, but for a two-layer synthetic multiplex network constructed according to the GMM with maximal inter-layer degree and angular correlations ($\nu = g = 1$). Each layer of the multiplex has the same parameters as in Figure 22. Panel (**d**) shows the correspondence among nodes belonging to the (k, k)-shells with $k \geq 7$.

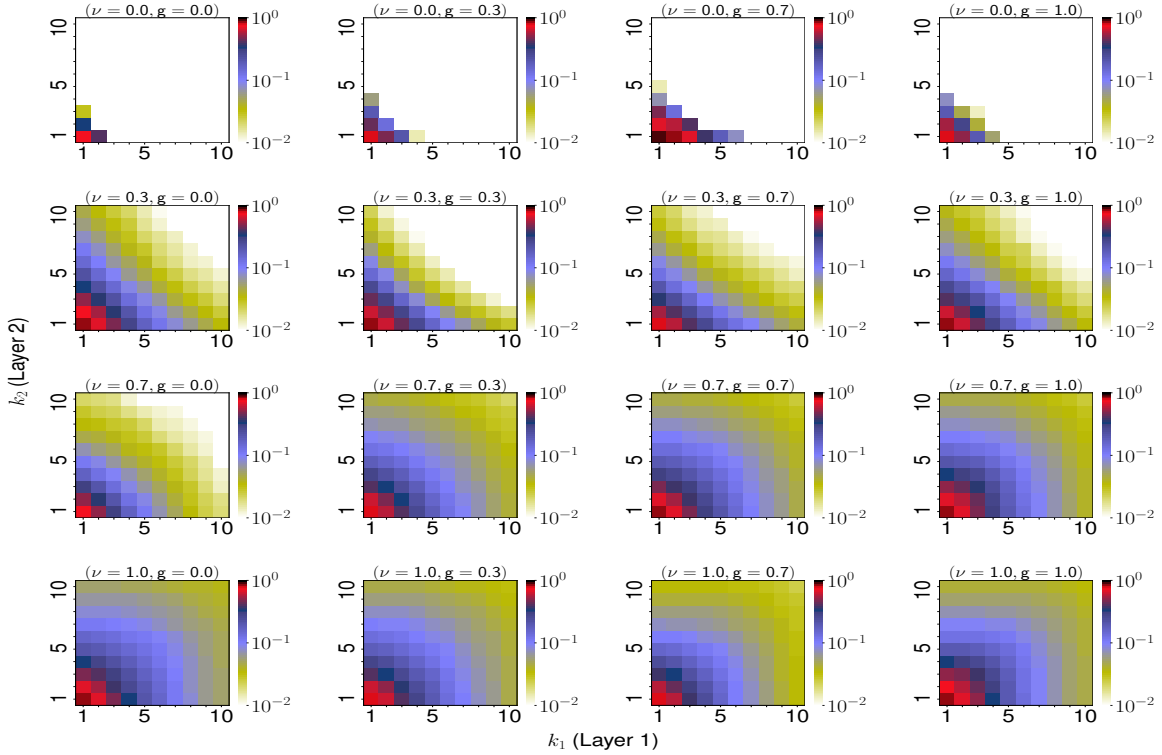

FIG. 25. **k-core of synthetic multiplex networks.** We study here the effect of degree and angular correlations on the relative size $S(k_1, k_2)$ of the (k_1, k_2)-core, in two-layer synthetic multiplexes constructed according to the GMM. Results are shown for different combinations of the inter-layer degree and angular correlation strength parameters $\nu \in [0, 1]$ and $g \in [0, 1]$. Each layer has $N = 1000$ nodes, power-law degree distribution with exponent $\gamma = 2.2$, average degree $\bar{k} \approx 6$, and temperature $T = 0.5$.

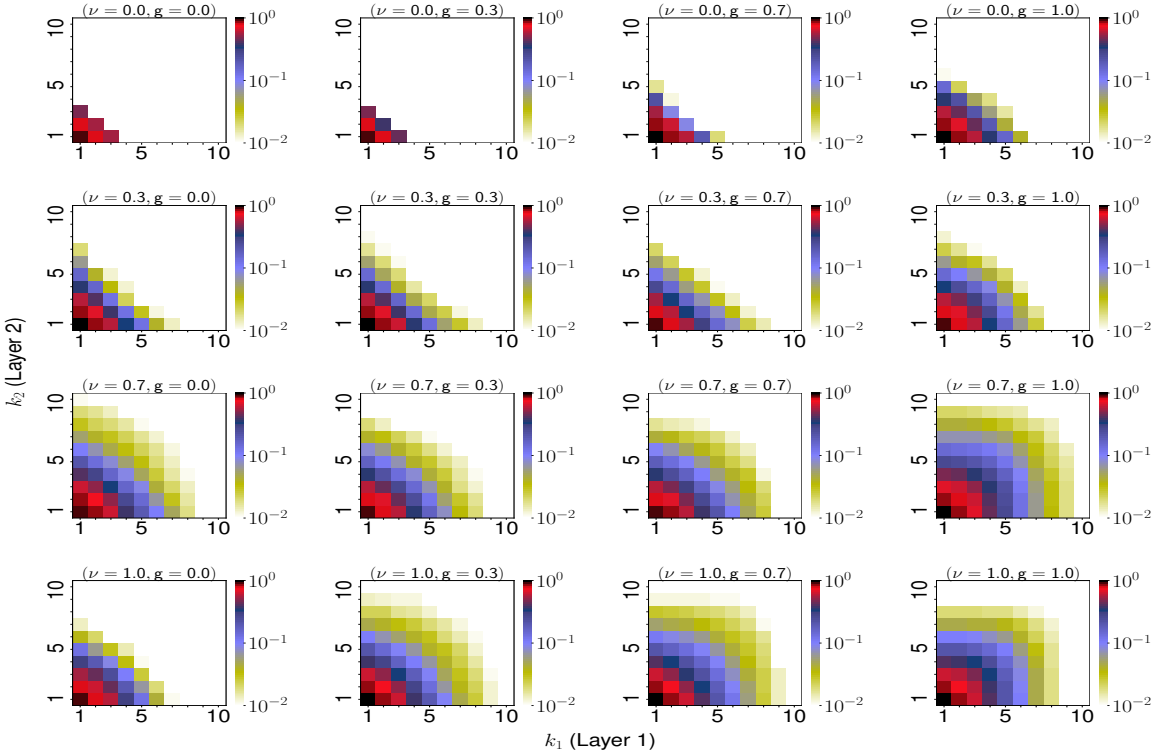

FIG. 26. **k-core of synthetic multiplex networks.** Same as in Figure 25, but for a different value of the degree exponent $\gamma = 2.6$. All other model parameters are identical to those used in Figure 25.

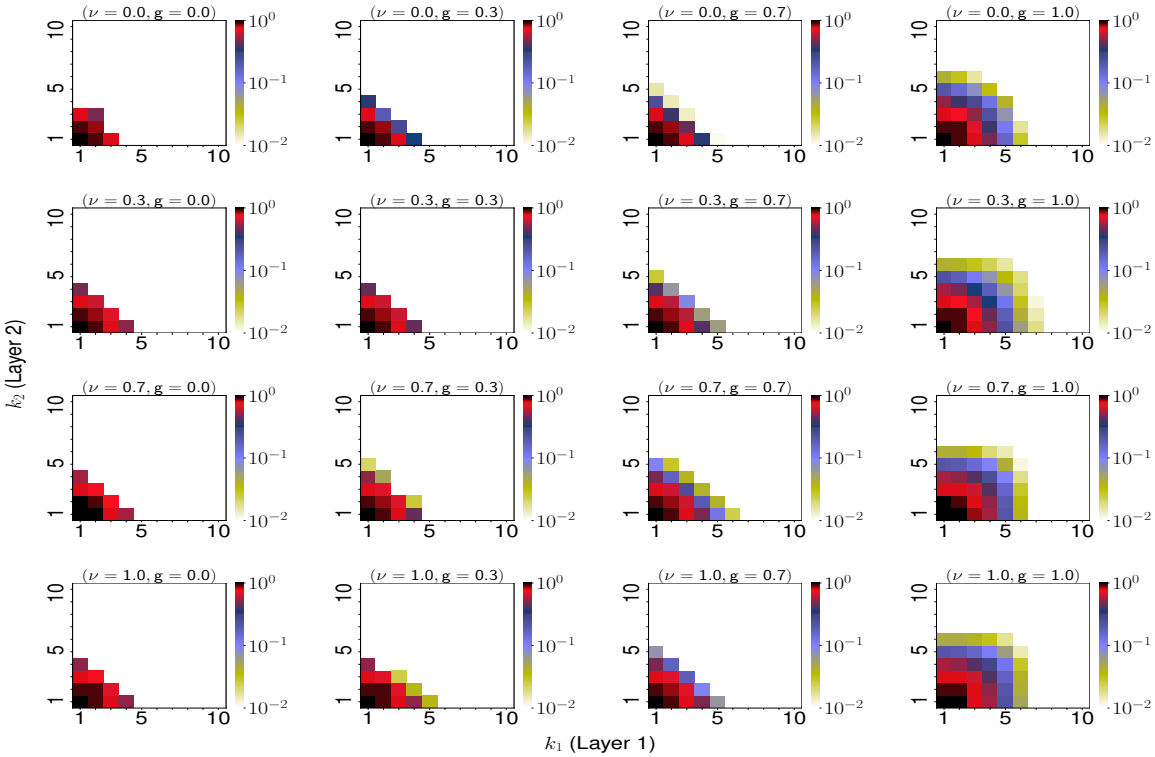

FIG. 27. **k-core of synthetic multiplex networks.** Same as in Figure 25, but for a different value of the degree exponent $\gamma = 3.5$. All other model parameters are identical to those used in Figure 25.

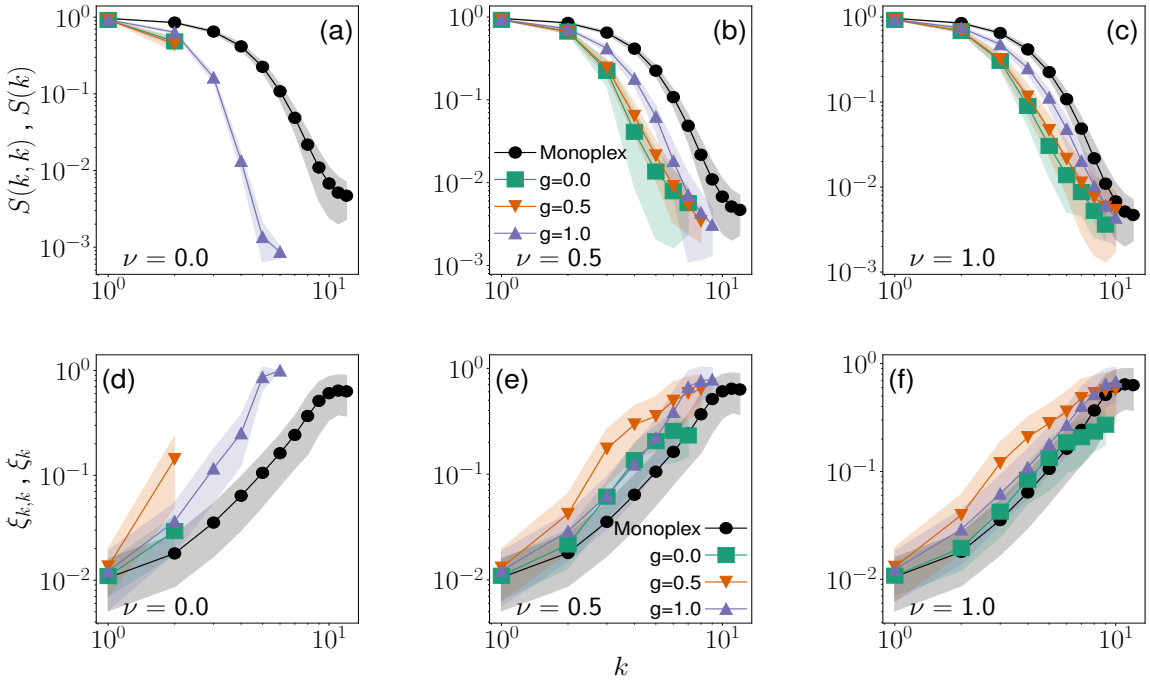

FIG. 28. **k-core structure of synthetic multiplex networks.** We study here the effect of degree and angular correlations on the size of the (k,k)-core $S(k,k)$ and its coherence $\xi_{k,k}$, in two-layer synthetic multiplex networks constructed according to the GMM. Each layer of the multiplex has $N = 10000$ nodes, power-law degree distribution with exponent $\gamma = 2.6$, average degree $\bar{k} \approx 6$, and temperature $T = 0.5$ (i.e., average clustering coefficient $\bar{c} = 0.4$). We consider various combinations of the degree and angular correlation strength parameters ν and g. Results in each case are obtained by taking the average value over 100 realizations. Shaded areas denote regions corresponding to one standard deviation away from the average. (**a**) Relative size $S(k,k)$ of the (k,k)-core as a function of the threshold k. The curve corresponding to the monoplex is obtained by measuring $S(k)$ for the k-core of the individual layers, and then taking the average value. (**b** and **c**) Same as in panel a, but for different choices of the model parameters. (**d**, **e** and **f**) We consider the same data as in panels a, b, and c, respectively, but we monitor the metrics of angular coherence $\xi_{k,k}$ and ξ_k as functions of the threshold value k.

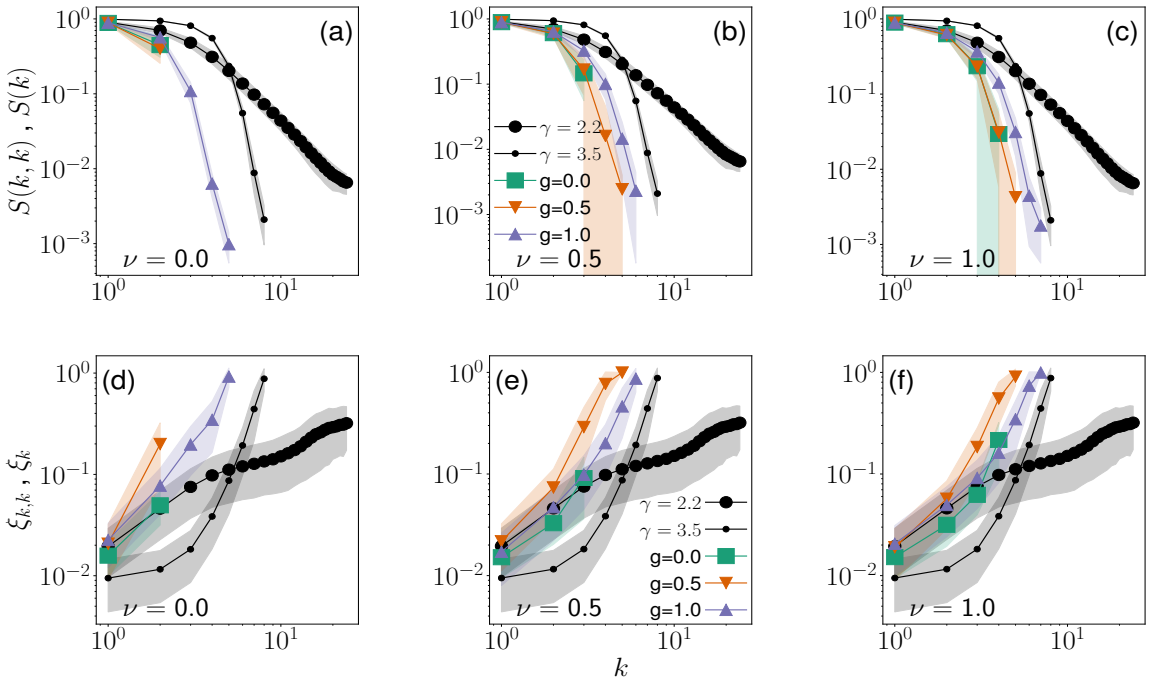

FIG. 29. **k-core structure of synthetic multiplex networks.** Same as in Figure 28, but for power-law degree distribution with exponent $\gamma = 2.2$ in one layer and $\gamma = 3.5$ in the other layer.

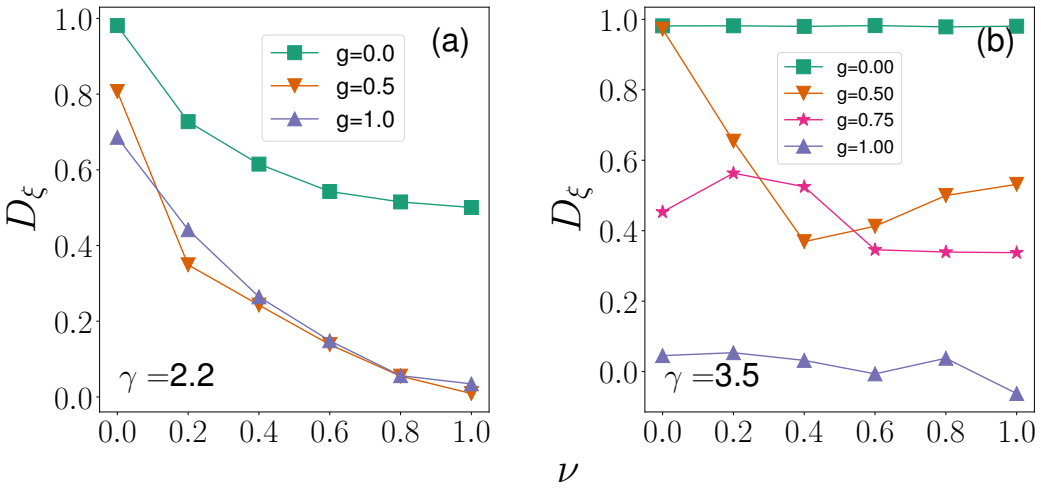

FIG. 30. **Quantifying the effect of inter-layer degree and similarity correlations in the k-core structure of synthetic multiplex networks.** (**a** and **b**) Relative difference $D_\xi = \left[\sum_k \xi_k - \sum_k \xi_{k,k}\right] / \sum_k \xi_k$ between the angular coherences ξ_k and $\xi_{k,k}$ for the networks constructed in Figures 9a and 9b in the main text. Error bars are not shown for clarity.

[1] "KONECT – openflights network dataset," `http://konect.uni-koblenz.de/networks/openflights` (2016).

[2] J. Kunegis, in Proc. Int. Conf. on World Wide Web Companion, pp. 1343–135 (2013).

[3] Guillermo García-Pérez, Marián Boguñá, and M. Ángeles Serrano, "Multiscale unfolding of real networks by geometric renormalization," Nature Physics **14**, 583–589 (2018).

[4] Shin-ya Takemura, Arjun Bharioke, Zhiyuan Lu, Aljoscha Nern, Shiv Vitaladevuni, Patricia K. Rivlin, William T. Katz, Donald J. Olbris, Stephen M. Plaza, Philip Winston, Ting Zhao, Jane Anne Horne, Richard D. Fetter, Satoko Takemura, Katerina Blazek, Lei-Ann Chang, Omotara Ogundeyi, Mathew A. Saunders, Victor Shapiro, Christopher Sigmund, Gerald M. Rubin, Louis K. Scheffer, Ian A. Meinertzhagen, and Dmitri B. Chklovskii, "A visual motion detection circuit suggested by drosophila connectomics," Nature **500**, 175 EP – (2013).

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

[6] Thomas Rolland, Murat Taşan, Benoit Charloteaux, Samuel J. Pevzner, Quan Zhong, Nidhi Sahni, Song Yi, Irma Lemmens, Celia Fontanillo, Roberto Mosca, Atanas Kamburov, Susan D. Ghiassian, Xinping Yang, Lila Ghamsari, Dawit Balcha, Bridget E. Begg, Pascal Braun, Marc Brehme, Martin P. Broly, Anne-Ruxandra Carvunis, Dan Convery-Zupan, Roser Corominas, Jasmin Coulombe-Huntington, Elizabeth Dann, Matija Dreze, Amélie Dricot, Changyu Fan, Eric Franzosa, Fana Gebreab, Bryan J. Gutierrez, Madeleine F. Hardy, Mike Jin, Shuli Kang, Ruth Kiros, Guan Ning Lin, Katja Luck, Andrew MacWilliams, Jörg Menche, Ryan R. Murray, Alexandre Palagi, Matthew M. Poulin, Xavier Rambout, John Rasla, Patrick Reichert, Viviana Romero, Elien Ruyssinck, Julie M. Sahalie, Annemarie Scholz, Akash A. Shah, Amitabh Sharma, Yun Shen, Kerstin Spirohn, Stanley Tam, Alexander O. Tejeda, Shelly A. Trigg, Jean-Claude Twizere, Kerwin Vega, Jennifer Walsh, Michael E. Cusick, Yu Xia, Albert-László Barabási, Lilia M. Iakoucheva, Patrick Aloy, Javier De Las Rivas, Jan Tavernier, Michael A. Calderwood, David E. Hill, Tong Hao, Frederick P. Roth, and Marc Vidal, "A proteome-scale map of the human interactome network," Cell **159**, 1212 – 1226 (2014).

[7] Bryan Klimt and Yiming Yang, "Introducing the enron corpus," in *CEAS 2004 - First Conference on Email and Anti-Spam, July 30-31, 2004, Mountain View, California, USA* (2004).

[8] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," Internet Mathematics **6**, 29–123 (2009), https://doi.org/10.1080/15427951.2009.10129177.

[9] k. claffy, Y. Hyun, K. Keys, M. Fomenkov, and D. Krioukov, *IEEE DHS Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH)* (2009) pp. 205–211.

[10] Joan Serrà, Álvaro Corral, Marián Boguñá, Martín Haro, and Josep Ll Arcos, "Measuring the evolution of contemporary western popular music," Scientific Reports **2**, 521 EP – (2012), article.

[11] M. Ángeles Serrano, Marián Boguñá, and Alessandro Vespignani, "Extracting the multiscale backbone of complex weighted networks," Proceedings of the National Academy of Sciences **106**, 6483–6488 (2009), https://www.pnas.org/content/106/16/6483.full.pdf.

[12] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon, "Superfamilies of evolved and designed networks," Science **303**, 1538–1542 (2004), https://science.sciencemag.org/content/303/5663/1538.full.pdf.

[13] 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).

[14] Kaj-Kolja Kleineberg, Marián Boguñá, M. Ángeles Serrano, and Fragkiskos Papadopoulos, "Hidden geometric correlations in real multiplex?networks," Nature Physics **12**, 1076 EP – (2016).

[15] 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 **103**, 4723–4728 (2006), https://www.pnas.org/content/103/12/4723.full.pdf.

[16] 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 (2014), http://oup.prod.sis.lan/comnet/article-pdf/3/2/159/1070864/cnu038.pdf.

[17] Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers, "BioGRID: a general repository for interaction datasets," Nucleic Acids Research **34**, D535–D539 (2006), http://oup.prod.sis.lan/nar/article-pdf/34/suppl_1/D535/3925435/gkj109.pdf.

[18] Manlio De Domenico, Vincenzo Nicosia, Alexandre Arenas, and Vito Latora, "Structural reducibility of multilayer networks," Nature Communications **6**, 6864 (2015).

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

[20] 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).